Title
Structural and functional characterization of contact-dependent growth inhibition systems

Author
Morse, Robert Paul

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Structural and functional characterization of contact-dependent growth inhibition systems

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Robert Paul Morse

Dissertation Committee:
Associate Professor Celia Goulding, Chair
Professor Thomas Poulos
Professor Sheryl Tsai

2014

DEDICATION

This thesis is dedicated to

my wife, Stephanie,

for your constant love and support

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Most importantly, I would like to thank Dr. Celia Goulding for giving me the opportunity to perform my dissertation research in her lab, and for being an excellent mentor. In Dr. Goulding's lab, I received exceptional scientific training, and grew as a scientist and as an individual. Dr. Goulding was supportive of my scientific interests, and put me in a position to succeed. I would also like to thank my great lab mates in the Goulding lab. In particular, Dr. Nicholas Chim, who is a patient teacher, amazing scientific resource, and friend, and Angelina Iniguez, who is always willing to help with even the smallest problems. Additionally, I am grateful for the hard work of my undergraduates Elias Gerrick and Danna Yang. Their dedication and work ethic will lead them far in their careers.

I extend my gratitude to my dissertation and advancement committees for their input on my research.

I would also like the thank our collaborators at UC Santa Barbara, Christopher Hayes and David Low, for the opportunity to hone my X-ray crystallography skills on a fascinating biological system.

I also extend thanks to the groups at UC Irvine who have provided equipment, reagents, and advice: the Poulos lab, the Tsai lab, the White lab, and the Nowick lab. Special gratitude goes to Dr. Mandy Zheng of the Nowick lab for the synthesis of a molecule used in this dissertation.

Thank you to Elsevier for their permission to reprint copyrighted material within chapter 4 of this dissertation.

Finally, thank you to my family and friends. My parents instilled a scientific curiosity within me, which has helped bring me to this pivotal moment. They have been supportive of all my endeavors, and I couldn't be more thankful. Lastly, thank you to my wonderful wife, Stephanie. You challenge me and make me a better person. I am so incredibly grateful and lucky to have you in my life, and I can't wait for what our future holds.

CURRICULUM VITAE

ROBERT PAUL MORSE
rmorse@uci.edu

Education

University of California, Irvine (2009-2014)
Doctor of Philosophy in Biological Sciences
Thesis Title: Structural and functional characterization of contact-dependent growth inhibition systems
Advisor: Celia W. Goulding, Ph.D

University of California, Santa Cruz (2005-2009)
B.S., Biochemistry and Molecular Biology, *cum laude*, honors

Research Experience

Graduate Student Research (Advisor: Dr. Celia Goulding)

- Investigation of bacterial Contact-Dependent growth Inhibition (CDI)
  - Structurally characterized novel bacterial toxin/immunity protein complexes by X-ray crystallography
  - Characterized the binding of various toxin/immunity protein complexes
  - Elucidated enzymatic functions of various CDI toxins
- Characterization of MhuD, a Mycobacterium tuberculosis heme-degrading protein
  - Determined the crystal structure of MhuD in its active conformation
  - Investigated heme-binding and degradation by mutational analysis

Undergraduate Student Research (Advisor: Dr. William Saxton)

- Used Drosophila as a model to study neurodegenerative disease
  - Genetically manipulated Drosophila and employed confocal microscopy to investigate defects in motor neurons
- Title of undergraduate honors thesis: Impairment of Kinesin-1 in the motor neurons of *Drosophila* causes the symptoms of a Hereditary Spastic Paraplegia

Awards

William F. Holcomb Scholarship for excellence in biomedical research (UC Irvine School of Biological Sciences) – 2013

Robert Warner Award for outstanding achievement in nucleic acid biochemistry (UC Irvine School of Biological Sciences) – 2012

American Crystallographic Association Travel Grant – 2012

Dean's and Chancellor's awards for exceptional achievement in an undergraduate research project (UC Santa Cruz) - 2012

Publications

Beck, C., **Morse, R. P**., Cunningham, D.A., Iniguez, A, Low, D.A. Goulding, C. W., Hayes, C. S. (2014) CdiA from Enterobacter cloacae delivers a toxic ribosomal RNase into target bacteria, Structure, 22, 707-718. * joint primary authorship

Morse, R. P., Nikolakakis, K. C., Willett, J. L., Gerrick, E., Low, D. A., Hayes, C. S., and Goulding, C. W. (2012) Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems, Proc Natl Acad Sci U S A 109, 21480-21485. * joint primary authorship

Tullius, M. V., Harmston, C. A., Owens, C. P., Chim, N., **Morse, R. P**., McMath, L. M., Iniguez, A., Kimmey, J. M., Sawaya, M. R., Whitelegge, J. P., Horwitz, M. A., and Goulding, C. W. (2011) Discovery and characterization of a unique mycobacterial heme acquisition system, Proc Natl Acad Sci U S A 108, 5051-5056.

Chim, N., Habel, J. E., Johnston, J. M., Krieger, I., Miallau, L., Sankaranarayanan, R., **Morse, R**. P., Bruning, J., Swanson, S., Kim, H., Kim, C. Y., Li, H., Bulloch, E. M., Payne, R. J., Manos-Turvey, A., Hung, L. W., Baker, E. N., Lott, J. S., James, M. N., Terwilliger, T. C., Eisenberg, D. S., Sacchettini, J. C., and Goulding, C. W. (2011) The TB Structural Genomics Consortium: a decade of progress, Tuberculosis (Edinb) 91, 155-172.

Graves, A. B., **Morse, R. P**., Iniguez, A., Goulding, C. W., and Liptak, M. D. The Mycobacterium tuberculosis MhuD Active Site Stabilizes a Ruffled Heme with a Complex Electronic Structure. 2014, *in review*

Presentations

Structural and functional insights into bacterial Contact-dependent Growth Inhibition Coast. West Coast Protein Crystallography Workshop, Monterey, CA, 2013 (oral presentation).

Structural and Functional Analysis of Two Contact-dependent Growth Inhibiting (CDI) complexes. American Crystallographic Association Annual Meeting, Boston, MA, 2012 (poster presentation).

Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems. UC Irvine Pharmaceutical Sciences Retreat, Laguna Beach, CA, 2012 (poster presentation).

Impairment of Kinesin-1 in the motor neurons of Drosophila causes the symptoms of Hereditary Spastic Paraplegia. 12th Annual Undergraduate Research Poster Symposium, Santa Cruz, CA, 2009 (poster presentation).

Teaching Experience

Molecular Biology (Springs Quarters 2011-2013)

- Held discussion sections, graded exams

Experimental Microbiology Lab (Spring Quarters 2011-2013)

- Led weekly laboratory sections, graded lab reports

Undergraduate Lab Mentoring

Trained and supervised two undergraduate students

ABSTRACT OF THE DISSERTATION

Structural and functional characterization of contact-dependent growth inhibition systems

By

Robert P. Morse

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2014

Professor Celia W. Goulding, Chair

Bacteria have developed complex mechanisms to thrive in their environments. Contact-dependent growth inhibition (CDI), a novel inter-bacterial competition mechanism was recently discovered in gram-negative bacteria, and is used to block growth of neighboring cells. The CDI growth inhibitory activity is contained in the C-terminal domain of CdiA effector proteins (CdiA-CT, toxin), and is delivered into target bacterial cells. All CDI^+ bacteria also produce small CdiI immunity proteins that protect them from autoinhibition. CdiA-CT and CdiI from different species exhibit high sequence variability, which predicts diverse structures and/or interaction interfaces. Here we describe the initial structural characterization of the CDI system, providing functional insights into distinct CDI mechanisms.

We have solved X-ray crystal structures of CDI toxin/immunity complexes from *Escherichia coli* (EC869), *Burkholderia pseudomallei* (1026b), and *Enterobacter cloacae* (ECL), representing three distinct families of CDI proteins. The interfaces between the toxin/immunity pairs are distinct, providing important insights into CdiA-CT/CdiI binding specificity. The EC869 CdiA-CT/CdiI binding interface is especially intriguing and consists of a unique β-augmentation interaction, by which $CdiA\text{-}CT^{EC869}$ inserts a short β-hairpin structural element into a cavity on

the surface of CdiI^{EC869}. This interface is conserved among EC869 family CdiA-CT/CdiI complexes, but heterologous toxin and immunity proteins within this family do not interact with high affinity. Additionally, the EC869, 1026b, and ECL toxin structures helped elucidate and characterize three distinct CdiA-CT growth-inhibiting activities. EC869 and 1026b CdiA-CTs are structurally similar and contain DNase and tRNase activities, respectively. In contrast, ECL CdiA-CT shares no structural homology with EC869 or 1026b CdiA-CTs, but is structurally and functionally similar to colicin E3, a toxin with rRNase activity from a distinct bacterial system. We also report the crystal structure of the CdiA-CT toxin from uropathogenic *E. coli* strain 536 in complex with its permissive factor, CysK, representing a unique step in CDI. The complex formed mimics the interaction found in the cysteine synthase complex. Finally, we solved the structure of a domain from the large, non-toxin region of *E. coli* CdiA, providing a starting point towards future research into the CdiA-CT cleavage and translocation mechanisms.

CHAPTER 1

An overview of inter-bacterial communication strategies

Since the discovery of quorum sensing in the 1960s and 1970s, it has become clear that bacteria, like humans and other organisms, utilize complex strategies to communicate with one another and thrive in their environments (1, 2). Additionally, some forms of bacterial communication can be lethal, such as the use of secreted toxic protein molecules, called bacteriocins, for the purpose of killing other bacteria and gaining a competitive advantage (3). Understanding the different interaction and communication mechanisms utilized by bacteria, be they cooperative or competitive, is important as bacterial survival has many implications in human health. Recent research has revealed the importance of the human microbiota in a number of important medical problems including obesity, allergic disease, neuropsychiatric disease, and colonization and infection of multidrug-resistant organisms (4-8). While much progress has been made towards understanding how bacteria communicate and interact via secreted molecules, new contact-dependent communication mechanisms have been recently discovered (9, 10). This chapter offers a brief review of bacterial communication via secreted molecules, and a thorough introduction to a newly characterized contact-dependent bacterial communication system that has implications regarding bacterial survival, and thus could provide an avenue for new antimicrobial therapies.

Secreted signaling molecules

Bacteria often use secreted molecules to interact with each other, facilitating both cooperative and competitive advantages. One such strategy, quorum sensing, is a well-studied

cell-to-cell communication mechanism used by bacteria for the purpose of monitoring population. This “cooperative” mechanism involves secretion of either acyl homoserine lactones or peptides (by gram-negative and gram-positive bacteria, respectively), which are recognized by specific membrane receptors. In some cases, the secreted peptides are imported into neighboring cells and bind specific intracellular receptor proteins. Quorum sensing allows groups of bacteria to coordinate and alter gene expression, thus uniformly changing their behavior in various ways such as initiation of biofilm formation, expression of virulence genes, and fluorescence (1, 12, 13). Importantly, since quorum sensing plays a role in virulence of important human pathogens such as *Staphylococcus aureus, Pseudomonas aeruginosa*, and *Escherichia coli,* inhibition of various bacterial quorum sensing systems is being investigated as an avenue for developing new antimicrobial therapies (14).

In an effort to compete for space and nutrients, some types of secreted molecules or peptides cause death to neighboring bacterial cells. The first such family of secreted toxic proteins to be discovered was the colicin family, named for *E. coli*. These toxic, secreted proteins have since been identified in a variety of bacterial species and been given the general name of “bacteriocins” (3). Colicins are often genetically linked to lysis proteins that facilitate the toxin release from the cell, however this process kills the toxin-releasing cell. Over 20 different colicins have been identified and characterized, and are categorized into two groups, depending on which system (Tol vs Ton) they use to translocate into a target cell (15). Imported colicins efficiently inhibit the growth of nearby bacterial cells by targeting important cellular components such as DNA, RNA, and the cell envelope (3). Cells that produce colicins often also produce associated “immunity” proteins, capable of preventing the toxic activity of the specific “cognate” colicin produced by the cell.

Contact-dependent growth inhibition (CDI)

A new method of inter-bacterial competition, contact-dependent growth inhibition (CDI), was recently discovered in *E. coli* strain EC93, and has since been identified in a wide range of gram-negative bacteria including α-proteobacteria, β-proteobacteria, and γ-proteobacteria (16-18). *E. coli* EC93 was the predominant strain isolated from rat feces, suggesting the presence of survival mechanisms to outcompete other *E. coli* strains. Indeed, the isolated *E. coli* EC93 showed the ability to outcompete a variety of *E. coli* laboratory strains in a mechanism mediated by cell-to-cell contact, and not secreted soluble proteins, such as colicins. Importantly, CDI systems are prevalent in pathogenic bacteria including *Escherichia, Burkholderia, Moraxella, Serratia, Pectobacterium, Photorhabdus, Bartonella, Pseudomonas, Dickeya, Neisseria, Acinetobacter,* and *Yersinia* species. Despite the presence of CDI in pathogenic bacteria, CDI systems are not classical virulence factors, as they do not target eukaryotic hosts. CDI may play an interesting role in host colonization as *Dickeya dadantii* cells expressing CDI have been shown to outcompete CDI^- cells on a chicory plant host (17).

The EC93 survival ability was determined to be mediated by three genes, *cdiB, cdiA*, and *cdiI*, which together make up the *cdiBAI* locus (18) (Figure 1.1). In some species such as *Burkholderia*, the genetic organization is different (*cdiAIB)* (19). Although widely distributed among gram-negative bacteria, not all strains of a given bacterial species have *cdiBAI* homologs. In contrast some bacterial isolates encode multiple CDI modules. Together, CdiB and CdiA are predicted to be members of the type Vb, or two-partner secretion (TPS) protein family, in which CdiB is an outer membrane β-barrel protein that facilitates export of the large CdiA exoprotein (3,000-6,000 amino acids) to the cell surface. Type Vb secretion is distinct from type Va because the membrane “translocator” protein (CdiB) and the “passenger” protein (CdiA) are translated as two separate polypeptides (20, 21). The growth inhibiting activity is localized to the carboxy terminal domain of CdiA proteins (CdiA-CT), which are imported into

neighboring cells upon cell-to-cell contact (Figure 1.1). The third gene in the *cdiBAI* locus encodes for a small immunity protein, CdiI, capable of inactivating the cognate CdiA-derived toxin (17).

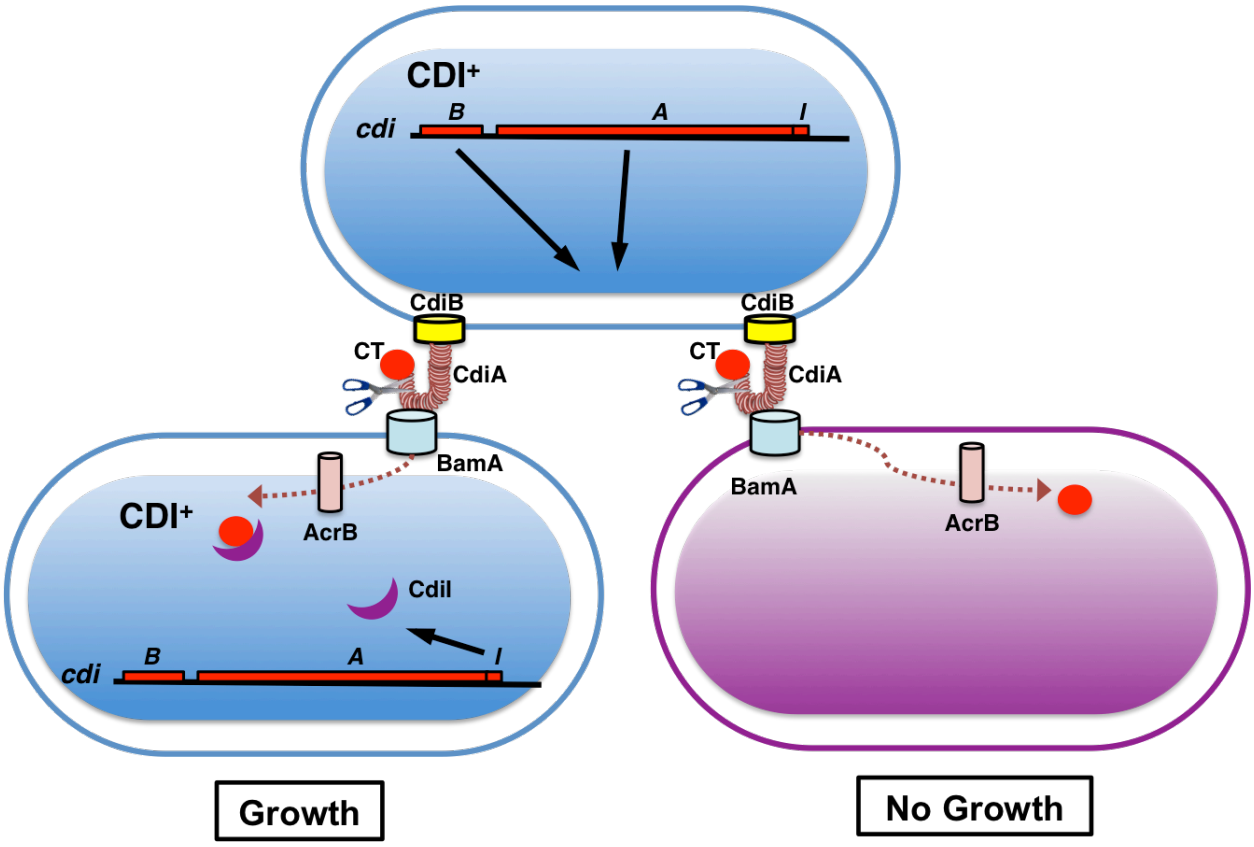

Figure 1.1. The CDI pathway. CDI$^+$ cells containing the *cdiBAI* gene cluster express CdiB and CdiA at the cell surface. The extended β-helical CdiA structure is depicted as a pink coil. Contact between CdiA and the BamA receptor on the surface of target cells results in delivery of the CdiA-CT toxin into the target cell. The mechanisms of toxin translocation are not understood, but BamA (22) and inner membrane components such as AcrB are hypothesized to mediate transport in at least some CDI systems. Cells carrying the identical CDI system (depicted as blue cells) are protected from growth inhibition by the CdiI immunity protein, which specifically binds and inactivates the CdiA-CT toxin. Non-immune cells are inhibited by the CdiA-CT toxin (depicted as purple cells). Figure is reproduced from Morse *et al.*, 2012 (23).

CdiA effector proteins

CdiA proteins contain a TPS domain in the N-terminal portion of the protein, which is essential for export via the TPS system, and a hemagglutinin-repeat region, which has been seen in other proteins utilizing TPS (18, 24). The hemagglutinin region is predicted to form an extended β-helical structure, projecting several hundred angstroms from the cell surface, allowing CdiA proteins to interact with target cells (25, 26) (Figure 1.1). Sequence alignments of CdiA proteins across bacterial species show a high level of sequence identity, suggesting that the overall CdiA structure is conserved within the CDI system.

However, CdiA structural conservation does not include the 250-300 residues encompassing the CdiA-CT domains, which are polymorphic, containing extremely low sequence homology between CDI expressing bacteria (17). Interestingly, the growth inhibitory activity of CDI resides within the CdiA-CT (17). The CdiA-CT toxins are demarcated from the large "non-toxin" region of CdiA proteins in most bacteria by a conserved VENN peptide, and a (Q/E)LYN motif in *Burkholderia* species (17) (Figure 1.2). Over 60 protein sequence families CdiA-CT toxins have been identified, suggesting that bacteria utilizing CDI have different mechanisms for toxicity. Indeed, the CdiA-CT from *D. dadantii* (CdiA-CT^{dd3937}) possesses robust DNAse activity (17), the CdiA-CT from uropathogenic *E. coli* strain 536 (CdiA-CT^{UPEC536}) is capable of cleaving transfer-RNA (tRNA) (11), and the CdiA-CT from *E. coli* EC93 (CdiA-CT^{EC93}) dissipates the proton motive force (27). Because many CdiA-CTs contain nuclease activity, these toxins must be delivered across the target cell envelope to confer toxicity. This was confirmed with recent research showing that the entire CdiA^{UPEC536} protein is transferred to the target cell surface, while CdiA-CT^{UPEC536} is cleaved and transferred into the target cell (28). Additionally, a chimeric CdiA effector protein, in which the CdiA-CT of CdiA^{EC93} was replaced with CdiA-CT^{dd3937}, also effectively delivered CdiA-CT^{dd3937} into the target cell cytosol. These results suggest that the "non-toxin" region of CdiA proteins may facilitate cell-to-cell contact and

that CdiA-CT translocation is independent of which CdiA-CT family toxin "caps" the CdiA protein.

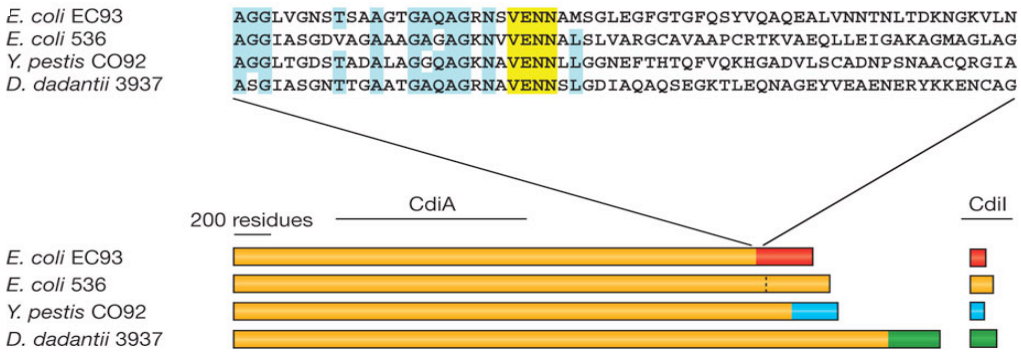

Figure 1.2. The modular structure of CdiA proteins. The protein sequences of CdiA proteins diverge following the conserved VENN motif (highlighted yellow), which demarcates the beginning of the CdiA-CT toxin domain. Figure is reproduced from Aoki *et al*., 2010 (17).

CdiA-CT entry into target cells

The mechanisms for CdiA target cell recognition and CdiA-CT translocation into target cells are not well understood, however genetic approaches have helped identify two membrane proteins that may facilitate cell-to-cell contact and toxin translocation in the EC93 CDI system. Selecting for *E. coli* mutants resistant to the EC93 CDI system resulted in the identification of two genes, *bamA* (formerly known as *yaeT*) and *acrB* (22) (Figure 1.1). BamA is a highly conserved, essential outer membrane β-barrel protein that together with four other proteins (BamB-E) forms the Bam complex, which is required for β-barrel membrane protein assembly (29, 30). Mutational analysis of BamB-E revealed that only BamA is required for CDI (22). Antibodies directed against BamA surface epitopes prevented both binding of $CdiA^{EC93}$ to target cells, and CDI mediated growth inhibition, showing that BamA is the surface receptor for the *E. coli* EC93 CDI system (22).

Recent work has shown that the EC93 CDI system cannot inhibit the growth of non *E. coli* enterobacterial species unless these other bacteria are supplemented with *E. coli bamA* (31). However, CdiA^{EC93} can utilize these non *E. coli* BamA proteins as receptors if two of their extracellular loops (loops 6 and 7) are engineered to contain the *E. coli* BamA protein sequence (31). Interestingly, while the BamA protein sequence is well conserved across bacterial species, these predicted extracellular loops show substantial sequence divergence. These results suggest that CdiA-BamA interactions may provide a means for CDI$^+$ bacteria to distinguish closely related bacteria from other bacterial species. Additionally, BamA does not appear to be a ubiquitous membrane receptor for all CdiA proteins, as CdiA from UPEC536 *E. coli* recognizes a different outer membrane protein (unpublished data from Hayes lab, UCSB).

The other protein identified in the aforementioned genetic screen was the inner-membrane multidrug transport protein, AcrB (22). Since AcrB is localized to the bacterial inner membrane, it does not appear to be a CdiA receptor. Unlike BamA, AcrB is required for growth inhibition when CdiA-CT^{EC93} is directly expressed within *E. coli* cells, suggesting that AcrB may contribute to both the toxin activity and/or translocation into the target cell. Furthermore, *acrB* mutants are not resistant to CDI mediated by non-EC93 CdiA-CTs, suggesting that AcrB may be specifically involved in EC93 mediated growth inhibition. This could indicate that various CdiA-CTs could hijack distinct proteins within the target cell to achieve their toxic activity. Interestingly, both BamA and AcrB are part of large, multi-component machines. While BamA is part of the Bam complex, AcrB interacts with two other proteins, AcrA and TolC, to form an efflux pump (32). Both *acrA* and *tolC* mutants showed no resistance to CDI, indicating that only AcrB is required for the EC93 CDI system. Together, both BamA and AcrB function within CDI appears to be independent to their previously defined roles in outer-membrane protein assembly and multidrug transport, respectively.

CdiI immunity proteins

To prevent growth inhibition of neighboring "self" cells, CDI^+ cells produce small CdiI immunity proteins (encoded by the third gene of the *cdiBAI* locus) to neutralize the imported CdiA-CT toxin (17) (Figure 1.1). CdiI immunity proteins only confer immunity to cognate CdiA-CT toxins, and not toxins from other bacterial species (Figure 1.3). Cognate CdiA-CT/CdiI pairs form stable protein complexes, while no stable interactions have been observed between heterologous protein pairs. However, there is some evidence of weak interactions between heterologous toxins and immunities from the same family (33). Interestingly, similar to CdiA-CT toxins, CdiI proteins are also highly polymorphic. The high sequence variability between CdiA-CT and CdiI proteins across bacterial species likely drives the specificity between cognate CdiA-CT and CdiI proteins. In general, CdiI sequence diversification is thought to be faster than that of CdiA-CTs, and is likely to be the driving force for continued specificity between cognate CdiA-CT/CdiI pairs (16). CdiA-CT evolution is slowed by the necessity for these proteins to retain their toxic activities.

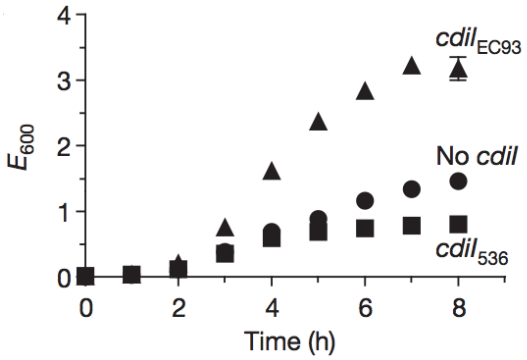

Figure 1.3. Cognate CdiI can prevent target cells from CDI. Growth of *E. coli* cells expressing $CdiA\text{-}CT^{EC93}$ was monitored (measured as attenuance at 600 nm (E_{600})). Co-expression of $CdiI^{EC93}$, but not heterologous $CdiI^{UPEC536}$, protected cells from growth inhibition. Figure is reproduced from Aoki *et al*., 2010 (17).

Unique case of the UPEC536 CDI system

The diversity of CdiA-CT and CdiI proteins is especially intriguing in the unique toxicity mechanism displayed in the UPEC536 CDI system (Figure 1.4). Recently, it has been shown that a target-cell "permissive" factor is required for the tRNase activity of CdiA-CT^{UPEC536} *in vitro* and *in vivo* (11) (Figure 1.5). The CdiA-CT^{UPEC536} toxin is only capable of cleaving the anticodon loop of tRNA^{Arg} if the toxin is bound to the biosynthetic enzyme CysK (*O*-acetylserine sulfhydrylase A). CysK, which is one of two enzymes (the other being CysM) that catalyze the final step of the synthesis of cysteine from serine, is required in the cytosol of the target cell for the UPEC536 CDI system to achieve toxicity. CdiA-CT^{UPEC536} and CysK form a stable complex mediated by the C-terminal GYGI peptide of CdiA-CT^{UPEC536}. The interaction appears to be remarkably similar to that observed in the cysteine synthase complex (CSC), formed between CysK and the C-terminal GDGI peptide of CysE (serine *O*-acetyltransferase) (34-41). It is currently unclear how the CdiA-CT^{UPEC536}/CysK interaction activates the toxin. This interaction presents an added layer of complexity to the CDI system, as they appear to have evolved interesting mechanisms to promote toxic activity. Presently CysK is the only CDI permissive factor that has been identified, but the diversity of CdiA-CT toxins, as well as the identification of over 60 CdiA-CT families, presents the likely possibility that more will be identified in the future.

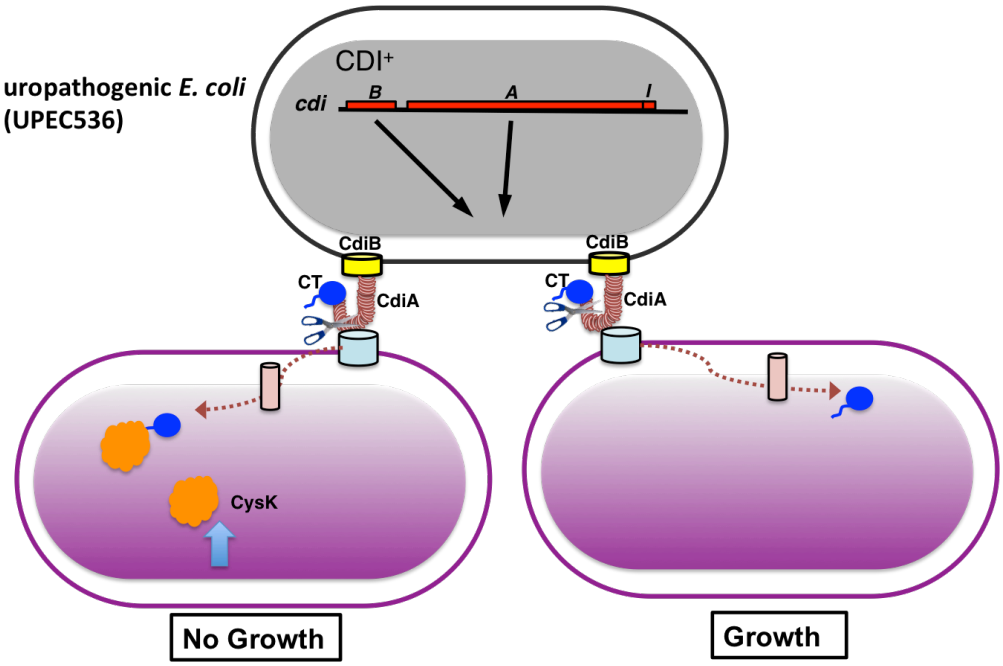

Figure 1.4. The UPEC536 CDI pathway. Translocation of the CdiA-CT^{UPEC536} toxin (blue) into target cells is not sufficient to inhibit cell growth (bottom right). CdiA-CT^{UPEC536} requires binding to the protein CysK in the target cell cytosol to activate its tRNase activity (bottom left). Cells not producing the CysK protein are immune to CdiA-CT^{UPEC536} (11).

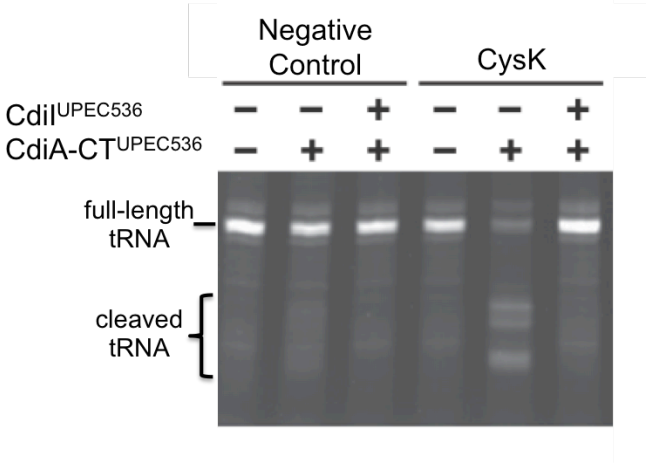

Figure 1.5. CysK is necessary for CdiA-CT^{UPEC536} tRNase activity. Purified tRNA was incubated with CdiA-CT^{UPEC536} in combination with CdiI^{UPEC536} and CysK, where indicated, and analyzed by gel electrophoresis and ethidium bromide staining. Negative control contains no added CysK. Figure is reproduced from Diner *et al.*, 2012 (11).

CDI is distinct from other toxic bacterial systems

Traditional toxin-antitoxin (TA) systems are well characterized and are found throughout eubacteria and archaea (42-44). These TA systems are one mechanism used by bacteria to cope with changes in their physical and chemical environment, and often act as stress response regulators. TA modules have other defining characteristics including gene organization, co-transcription of TA genes, and stability; the toxin protein is stable while the antitoxin is labile and often proteolytically susceptible. For example, in the RelBE system from *E. coli*, the RelB labile antitoxin is degraded in response to stress, allowing the RelE toxin to disrupt cell growth by suppressing protein synthesis (45). Additionally, some TA systems utilize functional RNA, and not proteins, as antitoxins (46). In contrast to antitoxins in TA systems, thus far CDI immunity proteins appear to be very stable. In CDI, toxins are not released from their immunity protein in response to stress. Moreover, TA toxins are not secreted, while CDI toxins transfer between bacteria by a novel contact-dependent mechanism. These major differences indicate that TA systems and CDI function by unique mechanisms and serve very different functions within bacteria.

The CDI system is also distinct from, but shares some similarities with the well-studied colicin family of bacteriocins from *E. coli*. Both CDI and colicins are bacterial mechanisms for toxin delivery in which all necessary components can be found in a single genetic locus. Colicins are smaller than CdiA proteins (30-70 kDa vs. 250-600 kDa), but contain a similar modular structure. Colicins have three domains, with each domain carrying out a specific role in the pathway. The central domain binds to specific receptors on the surface of the target cell, the N-terminal domain mediates transport across the target cell membrane, and similar to CDI, the C-terminal domain contains the toxin. Many different colicin C-terminal toxins have been characterized, with many of the activities similar to CdiA-CT toxins, including DNase, tRNase, and cell depolarization (3). Finally, like CDI, each colicin has an associated immunity protein

that can bind to and neutralize its activity.

Despite the similarities, there are key and notable differences between the CDI and colicin systems. Most importantly is the mechanism for protein secretion and toxin delivery. In CDI, CdiA proteins are secreted via the TPS, and toxin delivery requires cell-to-cell contact. Colicins are not secreted via protein machinery and instead require cell lysis, and therefore, cell death for secretion. Colicins are expressed and synthesized in response to stress, and expression can be stimulated by treating cells with DNA damaging agents (47). In contrast, the EC93 CDI system is expressed under normal growth conditions, which is possible because the "inhibiting-cell" is not lysed by invoking the system. Finally, colicins are known to utilize either the Ton or Tol systems for import into target cells (15), while the import of EC93 CdiA-CTs does not depend on these systems (22). Taken together, CDI is distinct from the colicin system, containing varying mechanisms for toxin export and delivery, as well as varying roles in bacterial competition and survival.

Scope of Thesis

The diversity of CDI presents a unique opportunity to investigate how bacterial species have evolved similar competition mechanisms, by structural and functional analysis. This thesis will describe the initial structural characterization of proteins in the CDI system by X-ray crystallography, as well as a functional characterization of CDI proteins. Specifically, our structures elucidate mechanisms for how various CdiA-CT toxins inhibit cell growth, how CdiI proteins neutralize toxicity, and why CdiA-CT/CdiI interactions are highly specific for cognate pairs. This research provides novel and important insights into the newly discovered CDI toxin-immunity pathway, which has implications regarding bacterial survival.

References

1. Nealson KH, Platt T, & Hastings JW (1970) Cellular control of the synthesis and activity of the bacterial luminescent system. *Journal of bacteriology* 104(1):313-322.
2. Tomasz A (1965) Control of the competent state in Pneumococcus by a hormone-like cell product: an example for a new type of regulatory mechanism in bacteria. *Nature* 208(5006):155-159.
3. Cascales E*, et al.* (2007) Colicin biology. *Microbiology and molecular biology reviews : MMBR* 71(1):158-229.
4. Khanna S & Tosh PK (2014) A clinician's primer on the role of the microbiome in human health and disease. *Mayo Clinic proceedings* 89(1):107-114.
5. Hornig M (2013) The role of microbes and autoimmunity in the pathogenesis of neuropsychiatric illness. *Current opinion in rheumatology* 25(4):488-795.
6. Tosh PK & McDonald LC (2012) Infection control in the multidrug-resistant era: tending the human microbiome. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 54(5):707-713.
7. Russell SL & Finlay BB (2012) The impact of gut microbes in allergic diseases. *Current opinion in gastroenterology* 28(6):563-569.
8. Ley RE (2010) Obesity and the human microbiome. *Current opinion in gastroenterology* 26(1):5-11.
9. Konovalova A & Sogaard-Andersen L (2011) Close encounters: contact-dependent interactions in bacteria. *Molecular microbiology* 81(2):297-301.
10. Hayes CS, Aoki SK, & Low DA (2010) Bacterial contact-dependent delivery systems. *Annual review of genetics* 44:71-90.
11. Diner EJ, Beck CM, Webb JS, Low DA, & Hayes CS (2012) Identification of a target cell permissive factor required for contact-dependent growth inhibition (CDI). *Genes Dev* 26(5):515-525.
12. Ng WL & Bassler BL (2009) Bacterial quorum-sensing network architectures. *Annual review of genetics* 43:197-222.
13. Waters CM & Bassler BL (2005) Quorum sensing: cell-to-cell communication in bacteria. *Annual review of cell and developmental biology* 21:319-346.
14. Antunes LC, Ferreira RB, Buckner MM, & Finlay BB (2010) Quorum sensing in bacterial virulence. *Microbiology* 156(Pt 8):2271-2282.
15. Kleanthous C (2010) Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nature reviews. Microbiology* 8(12):843-848.
16. Ruhe ZC, Low DA, & Hayes CS (2013) Bacterial contact-dependent growth inhibition. *Trends in microbiology* 21(5):230-237.
17. Aoki SK*, et al.* (2010) A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* 468(7322):439-442.
18. Aoki SK*, et al.* (2005) Contact-dependent inhibition of growth in Escherichia coli. *Science* 309(5738):1245-1248.
19. Nikolakakis K*, et al.* (2012) The toxin/immunity network of Burkholderia pseudomallei contact-dependent growth inhibition (CDI) systems. *Molecular microbiology* 84(3):516-529.
20. Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, & Ala'Aldeen D (2004) Type V protein secretion pathway: the autotransporter story. *Microbiology and molecular biology reviews : MMBR* 68(4):692-744.

21. Jacob-Dubuisson F, Locht C, & Antoine R (2001) Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Molecular microbiology* 40(2):306-313.
22. Aoki SK*, et al.* (2008) Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Molecular microbiology* 70(2):323-340.
23. Morse RP*, et al.* (2012) Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems. *Proceedings of the National Academy of Sciences of the United States of America* 109(52):21480-21485.
24. Locht C, Bertin P, Menozzi FD, & Renauld G (1993) The filamentous haemagglutinin, a multifaceted adhesion produced by virulent Bordetella spp. *Molecular microbiology* 9(4):653-660.
25. Clantin B*, et al.* (2004) The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway. *Proceedings of the National Academy of Sciences of the United States of America* 101(16):6194-6199.
26. Makhov AM*, et al.* (1994) Filamentous hemagglutinin of Bordetella pertussis. A bacterial adhesin formed as a 50-nm monomeric rigid rod based on a 19-residue repeat motif rich in beta strands and turns. *Journal of molecular biology* 241(1):110-124.
27. Aoki SK, Webb JS, Braaten BA, & Low DA (2009) Contact-dependent growth inhibition causes reversible metabolic downregulation in Escherichia coli. *Journal of bacteriology* 191(6):1777-1786.
28. Webb JS*, et al.* (2013) Delivery of CdiA nuclease toxins into target cells during contact-dependent growth inhibition. *PLoS One* 8(2):e57609.
29. Wu T*, et al.* (2005) Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. *Cell* 121(2):235-245.
30. Voulhoux R, Bos MP, Geurtsen J, Mols M, & Tommassen J (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science* 299(5604):262-265.
31. Ruhe ZC, Wallace AB, Low DA, & Hayes CS (2013) Receptor polymorphism restricts contact-dependent growth inhibition to members of the same species. *mBio* 4(4).
32. Nikaido H & Zgurskaya HI (2001) AcrAB and related multidrug efflux pumps of Escherichia coli. *Journal of molecular microbiology and biotechnology* 3(2):215-218.
33. Poole SJ*, et al.* (2011) Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. *PLoS genetics* 7(8):e1002217.
34. Kumaran S, Yi H, Krishnan HB, & Jez JM (2009) Assembly of the cysteine synthase complex and the regulatory role of protein-protein interactions. *The Journal of biological chemistry* 284(15):10268-10275.
35. Kumaran S & Jez JM (2007) Thermodynamics of the interaction between O-acetylserine sulfhydrylase and the C-terminus of serine acetyltransferase. *Biochemistry* 46(18):5586-5594.
36. Zhao C*, et al.* (2006) On the interaction site of serine acetyltransferase in the cysteine synthase complex from Escherichia coli. *Biochemical and biophysical research communications* 341(4):911-916.
37. Mino K*, et al.* (2000) Effects of bienzyme complex formation of cysteine synthetase from escherichia coli on some properties and kinetics. *Bioscience, biotechnology, and biochemistry* 64(8):1628-1640.
38. Mino K*, et al.* (2000) Characteristics of serine acetyltransferase from Escherichia coli deleting different lengths of amino acid residues from the C-terminus. *Bioscience, biotechnology, and biochemistry* 64(9):1874-1880.

39. Zhu X, Yamaguchi T, & Masada M (1998) Complexes of serine acetyltransferase and isozymes of cysteine synthase in spinach leaves. *Bioscience, biotechnology, and biochemistry* 62(5):947-952.
40. Droux M, Ruffet ML, Douce R, & Job D (1998) Interactions between serine acetyltransferase and O-acetylserine (thiol) lyase in higher plants--structural and kinetic properties of the free and bound enzymes. *European journal of biochemistry / FEBS* 255(1):235-245.
41. Kredich NM, Becker MA, & Tomkins GM (1969) Purification and characterization of cysteine synthetase, a bifunctional protein complex, from Salmonella typhimurium. *The Journal of biological chemistry* 244(9):2428-2439.
42. Hayes F & Van Melderen L (2011) Toxins-antitoxins: diversity, evolution and function. *Crit Rev Biochem Mol Biol* 46(5):386-408.
43. Fozo EM*, et al.* (2010) Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. *Nucleic acids research* 38(11):3743-3759.
44. Pandey DP & Gerdes K (2005) Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic acids research* 33(3):966-976.
45. Christensen SK, Mikkelsen M, Pedersen K, & Gerdes K (2001) RelE, a global inhibitor of translation, is activated during nutritional stress. *Proceedings of the National Academy of Sciences of the United States of America* 98(25):14328-14333.
46. Gerdes K & Wagner EG (2007) RNA antitoxins. *Current opinion in microbiology* 10(2):117-124.
47. Herschman HR & Helinski DR (1967) Comparative study of the events associated with colicin induction. *Journal of bacteriology* 94(3):691-699.

CHAPTER 2

Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems

Summary

Contact-dependent growth inhibition (CDI) systems encode polymorphic toxin/immunity proteins that mediate competition between neighboring bacterial cells. We present the first crystal structures of CDI toxin/immunity complexes from *Escherichia coli* EC869 and *Burkholderia pseudomallei* 1026b. Despite sharing little sequence identity, the toxin domains are structurally similar and have homology to endonucleases. The EC869 toxin is a Zn^{2+}-dependent DNase capable of completely degrading the genomes of target cells, whereas the Bp1026b toxin cleaves the aminoacyl acceptor stems of tRNA molecules rather than having DNase activity. Each immunity protein binds and inactivates its cognate toxin in a unique manner. The EC869 toxin/immunity complex is stabilized through an unusual β- augmentation interaction. In contrast, the Bp1026b immunity protein exploits shape and charge complementarity to occlude the toxin active site. These structures represent the initial glimpse into the CDI toxin/immunity network, illustrating how sequence-diverse toxins adopt convergent folds yet retain distinct binding interactions with cognate immunity proteins. Moreover, we present the first visual demonstration of CDI toxin delivery into a target cell.

Introduction

Bacteria employ a variety of strategies to compete and communicate with one another in the environment. Contact-dependent growth inhibition (CDI) is a mechanism that allows some Gram- negative bacteria to block the growth of neighboring cells (1, 2). CDI is mediated by the CdiB/CdiA family of two-partner secretion proteins. CdiB is a predicted outer membrane β-barrel protein required for secretion of the CdiA effector protein (2). CdiA exoproteins are very large (250 – 650 kDa) and comprised of an N-terminal transport domain followed by a variable number of hemagglutinin repeats (1). The hemagglutinin-repeat region is predicted to form an extended β-helical filament capable of projecting several hundred Å from the inhibitor cell surface (3). The current model of CDI postulates that CdiA binds to receptors on the surface of susceptible bacteria (4), initiating the proposed delivery of a CdiA-derived toxin into the target cell. The CDI toxin activity is contained within the C-terminal 250 – 300 residues of CdiA proteins – a region collectively termed "CdiA-CT" (1). CdiA-CT sequences are highly variable between CDI systems, but these toxin regions are typically demarcated by a conserved peptide motif: (Q/E)LYN in *Burkholderia* species (5) and VENN in most other bacteria (1). There are more than 60 CdiA-CT families based on sequence homology, suggesting that CDI^+ bacteria deploy a wide variety of toxins. CdiA-CTs can dissipate the proton motive force (6), degrade DNA (1), and cleave tRNA molecules (5, 7), with each activity sufficient to inhibit cell growth. CDI is only active against bacteria, and therefore CDI^+ cells must produce a CdiI immunity protein to protect themselves from autoinhibition. CdiI proteins are also highly variable and bind their cognate CdiA-CTs to block toxin activity. Because CdiA-CT/CdiI binding interactions are highly specific, immunity proteins provide no protection from the toxins deployed by other CDI systems (1, 5, 8). Thus, intercellular competition is thought to drive the diversification of CDI toxin/immunity pairs. Here, we describe the crystal structures of two different CdiA-CT/CdiI complexes, which provide the first insights into CDI diversity and mechanisms of toxicity and

immunity.

Materials and Methods

Purification of CdiA-CT/CdiI complexes

E. coli O157:H7 strain 869 (EC869) CdiA-CT/CdiI and *B. pseudomallei* 1026b (Bp1026b) CdiA- CT/CdiI (containing residues 123-294 of full length CdiA-CT) protein complexes were overexpressed in a pET21d plasmid containing the appropriate *cdiA-CT/cdiI* genes using *E. coli* BL21(DE3) cells (Novagen). Cells were grown aerobically at 37 °C in LB medium containing 100 μg/mL ampicillin. CdiA- CT/CdiI protein complex expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactosidase at an OD600 ~0.8 and grown for a further 4 hr before harvesting. Cells were collected by centrifugation at 5,100 rpm for 25 min and then washed with resuspension buffer [20 mM Tris-HCl (pH 7.9), 500 mM NaCl]. Cells were broken by sonication on ice in resuspension buffer containing 10 mg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride. Unbroken cells and debris were removed by centrifugation at 13,000 rpm for 30 min followed by filtration through 0.22 μm filter. Clarified lysates were loaded onto a Ni^{2+}-charged HiTrap column (5 mL, GE Healthcare) and washed with resuspension buffer supplemented with 10 mM imidazole. CdiA-CT/CdiI complexes were eluted with a linear gradient of imidazole (10 – 500 mM) in resuspension buffer. Fractions were collected, combined and concentrated to a volume of ~ 500 μl using a 10 kDa centrifugal concentrator (Centricon, Millipore). Complexes were further purified by gel filtration on a Superdex 200 column (GE Healthcare) equilibrated with 20 mM Tris-HCl (pH 7.4), 150 mM NaCl using an AKTA FPLC. The selenomethionine-derivatized EC869 (Se-EC869) CdiA-CT/CdiA complex was grown in M9 minimal medium supplemented with amino acids supplements (leucine, isoleucine and valine at 50 mg/L; phenylalanine, lysine and threonine at 100 mg/L; and selenomethionine at 75 mg/L) as described (10). The Se-EC869 CdiA-CT/CdiI was purified as described above for the native

protein complex. Se-EC869 CdiA-CT/CdiI was concentrated to 8 mg/mL in 20 mM Tris-HCl (pH 7.4), 150 mM NaCl for crystallization trials.

Purification of isolated CdiA-CT and CdiI-His_6 proteins

The individual His_6-tagged CdiI immunity proteins were overexpressed from pET21S-derived plasmids and purified by Ni^{2+}-affinity chromatography as described above. The purified CdiI-His_6 proteins were dialyzed against 20 mM Tris-HCl (pH 7.4), 150mM NaCl and concentrations determined by absorbance at 280 nm. The CdiA-CT proteins were isolated from their His_6-tagged cognate immunity proteins by Ni^{2+}-affinity chromatography under denaturing conditions. CdiA-CT/CdiI-His_6 complexes were denatured in 6 M urea, 150 mM NaCl, 20 mM Tris-HCl (pH 7.4) or 6 M guanidine-HCl (pH 7.9), then subjected to Ni^{2+}-affinity chromatography in the same buffers. Denatured CdiA-CTs were collected in the flow-through fractions, and dialyzed first against refolding buffer [1.3 M urea, 0.1 M Tris-HCl (pH 7.8), 0.1 M glycine] overnight at 4 °C, followed by dialysis into 20 mM Tris-HCl (pH 7.4), 150 mM NaCl.

Crystallization, data collection, structure determination and refinement

SeMet-derivatized CdiA-CT_{o11}^{EC869}/$CdiI_{o11}^{EC869}$ crystals were grown over two months at room temperature by hanging drop-vapor diffusion with a reservoir containing 0.1 M sodium acetate (pH 5.5), 0.2 M NaCl, 18% PEG-6000 and an 8 mg/mL protein solution containing 10 mM YCl_3. The SeMet- derivatized complex crystallized in space group $C222_1$ with unit cell dimensions of 81.7 Å × 103.6 Å × 125.173 Å and one complex per asymmetric unit. Crystals were mounted and collected under cryoconditions with the addition of 40% glycerol as cryoprotectant to the reservoir condition. A Se-single anomalous dispersion (SAD) dataset was collected at 70K at the Se absorption edge (0.979 Å) on beamline 8.2.1 at ALS. Data reduction was carried out with the HKL2000 suite (9), resulting in a 100% complete dataset up to 2.35 Å and significant anomalous differences up to 2.85 Å resolution. Eight Se sites were located

using CRANK (10) from CCP4i. An initial model was obtained with a custom pipeline that includes AFRO/CRUNCH2 for initial phasing (11), BP3 for substructure refinement (10), SOLOMON for density modification (12), BUCCANEER for model building (13) and PHENIX using Autobuild (14). The final model was built through iterative manual building in Coot (15) and refined through phenix.refine (16). The final model includes residues Val85 – Lys297 of CdiA-CT_{o11}^{EC869} and residues Ala2 – Arg164 of $CdiI_{o11}^{EC869}$ with a final R_{work}/R_{free} (%) 18.0/22.9 and 97.1% of residues in the favorable allowed regions. Residue Lys192 of CdiA-CT_{o11}^{EC869} (chain A), and $CdiI_{o11}^{EC869}$ residues Lys5, Glu78, Lys79, Glu82, Glu93, and Arg136 (chain B) were modeled as alanines due to lack of observable side chain density. Data collection and refinement statistics are presented in Table 2.1.

The Bp1026b CdiA-CT/CdiA complex with truncated CdiA-CT containing residues 123-294 of full length CdiA-CT was purified as above and concentrated to 12.5 mg/mL in 20 mM Tris-HCl (pH 7.4), 150 mM NaCl for crystallization trials. Diffraction quality crystals were as above with crystallization conditions consisting of 0.49 M sodium phosphate monobasic, 0.96 M potassium phosphate dibasic, and the protein solution containing 10 mM triethylene glycol and chymotrypsin (1:1000 protease:protein ratio), were grown over the period of two weeks. Crystals were swiped in reservoir condition buffer containing 20% 2-methyl-2,4-pentanediol and 0.5 M NaBr for 30 s and then flash frozen. The Bp1026b complex crystallized in space group F222 with four complexes per asymmetric unit cell of 151.9 Å × 173.6 Å × 174.8 Å. The bromide-soaked crystal diffracted to 2.65 Å with data collected at 0.919 Å for Br-SAD phasing. The dataset was indexed, integrated and scaled in HKL2000 (9). Experimental phasing and initial model building was performed using AutoSol and Autobuild in PHENIX (17), in which 20 Br– ions were located per asymmetric unit. The final model was built through iterative manual building in Coot (15) and refined with phenix.refine (16). The final model includes residues Gly163 – Pro294 of CdiA-$CT_{II}^{Bp1026b}$ and Ala2 – Arg101 of $CdiI_{II}^{Bp1026b}$ with R_{work}/R_{free} (%)

20.4/24.5 and 97.2% of the residues in the favorable allowed region. The side chains of CdiA-$CT_{II}^{Bp1026b}$ residues Lys168, Leu208, Lys279, Lys293, Pro294 (chains A, C, E G), and Glu206 (chains E, G) were modeled as alanine residues for lack of electron density. Similarly, the side chains of $CdiI_{II}^{Bp1026b}$ residues Glu32, Asn87 (chains B, D, F and H), Lys34 and Asp84 (chains B and D) were also modeled as alanines. CdiA-$CT_{II}^{Bp1026b}$ residues Gly123 – Lys155 were removed by chymotrypsin digestion during crystallization, and no electron density was observed for residues Thr156 – Thr162. Data collection and refinement statistics are presented in Table 2.1.

Differential scanning fluorimetry

The thermal stability of CdiA-CT/CdiI complexes and individual proteins was assessed using an Mx3005P QPCR machine (Agilent Technologies). Each 50 µL sample contained CdiA-CT, CdiI-His_6 or CdiA-CT/CdiI-His_6 complex at 5 µM in 100 mM potassium phosphate (pH 7.4) supplemented with 40 µM SYPRO orange dye. Fluorescence (Ex/Em wavelengths: 492/610 nm) was recorded from 25 – 95 °C with a temperature gradient of 1 °C/min (18). All samples were tested in duplicate and each experiment was repeated independently at least three times. The data were fitted using a nonlinear regression function (GraphPad Prism) and T_m values were extracted by the inflection point of the fluorescence curves.

Determination of CdiA-CT and CdiI binding affinities

Binding affinities of Bp1026b and EC869 complexes were determined by bio-layer interferometry (BLI) (19) using a BLitz™ (ForteBio Inc.), operating at 25 °C. All experiments were performed in 20 mM Tris pH 7.4, 150 mM NaCl. CdiI-His_6 immunity proteins were immobilized on Ni-NTA biosensors and exposed to different concentrations of cognate CdiA-CT toxins (0.4-2.5 µM). A reference was subtracted from all binding curves prior to curve fitting. Curve fitting and data processing was performed using BLitz Pro™ software (ForteBio Inc.).

Table 2.1. X-ray diffraction data and atomic refinement for Bp1026b and EC869 CdiA-CT/CdiI complexes

	Br-CdiA-CT/CdiI$_{II}^{Bp1026b}$	**Se-CdiA-CT/CdiI$_{o11}^{EC869}$**
Space group	F222	$C222_1$
Unit cell dimensions, Å	151.9 × 173.6 × 174.8	81.7 × 103.6 × 125.173
pH of crystallization condition	7.2	5.5
Protein concentration, mg/mL	12.5	8
Dataset		
Wavelength, Å	0.919	0.979
Resolution range	50–2.65	50–2.35
Unique reflections (total)	34,071 (2,642,611)	22,419 (787,310)
Completeness, %*	99.47 (96)	99.72 (100)
Redundancy*	29.6 (30.1)	14.3 (14.6)
Rmerge*,†	12.5 (39.1)	17.6 (48.2)
I/σ*	32.8 (12.2)	9.0 (10.3)
FOM	0.286	0.312
No. of Br/Se sites	20 Br	8 Se
Other metals		1 Zn, 3 Y
NCS copies	4	1
Model refinement		
Resolution range, Å	43.4–2.65	39.9–2.35
No. of reflections (working/free)	33,975/1,720	22,333/1,145
No. of protein atoms	6,942	2,950
No. of water molecules	33	55
Missing residues	123–162 (CdiA)	6–84 (CdiA)
	1, 102, 103 (CdiI)	1, 165–169 (CdiI)
Rwork/Rfree, %‡	20.4/24.5	18.0/22.9
Rms deviations		
Bond lengths, Å	0.01	0.007
Bond angles, °	1.2	1
Ramachandran plot (%)		
Most favorable region	97.2	97.1
Additional allowed region	2.8	2.9
Disallowed region	0	0
PDB ID code	4G6V	4G6U

*Statistics for the highest resolution shell are given in (brackets)

†$R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma I$

‡$R_{work}=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$ R_{free} was computed identically except where all reflections belong to a test set of 5% randomly selected data.

Additional Materials and Methods (Experiments performed by the Hayes lab, UCSB)

CdiA-CT/CdiI expression constructs and toxin/immunity protein complex purification have been described previously (5, 8). Site-directed mutagenesis and construction of the chimeric EC93-EC869 CDI system are outlined in the SI methods of Morse *et al* (20). Nuclease activity assays were performed essentially as described (5) with modifications outlined in SI Methods of Morse *et al* (20). Growth competitions were carried as described previously (1) except CDI^+ inhibitor and target cells were mixed at a 1:1 ratio and incubated at 37 °C with shaking for 6 h. Cells from the CDI competition experiments were visualized by fluorescence microscopy as described in SI Methods of Morse *et al* (20).

Results

CdiA-CT/CdiI crystallization and structure determination

To explore the structural diversity of CDI toxin/immunity proteins, we focused on CdiA-CT/CdiI pairs from *Burkholderia pseudomallei* 1026b (Bp1026b) and *Escherichia coli* O157:H7 strain EC869 (EC869). The CdiA-$CT_{II}^{Bp1026b}$/$CdiI_{II}^{Bp1026b}$ proteins are derived from the CDI locus on chromosome II of Bp1026b (5), and the CdiA-CT_{o11}^{EC869}/$CdiI_{o11}^{EC869}$ complex is encoded by the 11th "orphan" (o11) module of *E. coli* EC869. Orphan *cdiA-CT/cdiI* modules are toxin/immunity gene pairs that have been displaced from full-length *cdiA* genes (8). Tandem arrays of these modules are often associated with CDI systems and are thought to represent reservoirs of toxin/immunity diversity. We co-expressed each CdiA-CT together with a His_6-tagged version of its immunity protein, and the resulting CdiA-CT/CdiI- His_6 complexes were purified to near homogeneity (Figure 2.1). The CdiA-CT_{o11}^{EC869}/$CdiI_{o11}^{EC869}$ complex was stable, however the N-terminus of the CdiA-$CT_{II}^{Bp1026b}$ showed significant degradation after purification, suggesting that this region is sensitive to proteolysis. Therefore, we generated a truncated version of CdiA-$CT_{II}^{Bp1026b}$ beginning at residue Gly123 (numbered from Glu1 of the ELYN

motif), which still binds to the $CdiI_{II}^{Bp1026b}$ immunity protein and retains full toxin activity (5).

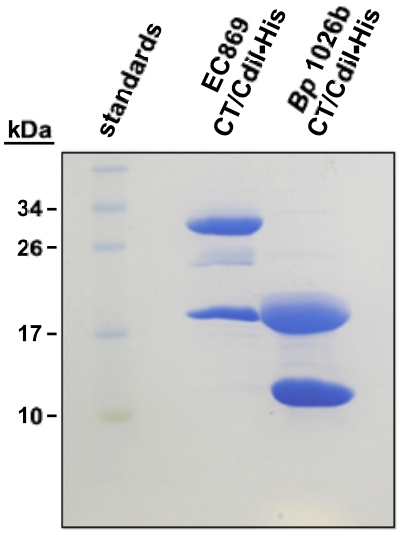

Figure 2.1. CdiA-CT/CdiI complex purification. CdiA-CT/CdiI-His6 complexes were purified by Ni^{2+}-affinity and size-exclusion chromatography and then analyzed by SDS/PAGE and Coomassie-blue staining. The gel migration positions of molecular mass standards are indicated.

The $CdiA\text{-}CT_{o11}^{EC869}/CdiI_{o11}^{EC869}$ crystal structure was solved to 2.35 Å resolution by Se-SAD phasing. The crystal space group was $C222_1$ with one complex per asymmetric unit. The structural model contains $CdiA\text{-}CT_{o11}^{EC869}$ residues Val85 – Lys297 (numbered from Val1 of the VENN motif) and Ala2 – Arg164 of $CdiI_{o11}^{EC869}$. In addition, 55 water molecules, three Y^{3+} ions and one Zn^{2+} ion were included in the final model resulting in an R_{work}/R_{free} of 18.0/22.9 (Table 2.1). The Bp1026b toxin/immunity complex contains no internal methionine residues for Se-Met incorporation, so crystals were soaked with bromide and the structure was solved to 2.65 Å resolution by Br-SAD phasing. The $CdiA\text{-}CT_{II}^{Bp1026b}/CdiI_{II}^{Bp1026b}$ complex crystallized in space group F222 with four complexes per asymmetric unit. The structural model contains $CdiA\text{-}CT_{II}^{Bp1026b}$ residues Gly163 – Pro294 and residues Ala2 – Arg101 of $CdiI_{II}^{Bp1026b}$. In addition, 33 water molecules were included in the final model to yield an R_{work}/R_{free} of 20.4/24.5 (Table 2.1).

Structure of the CdiA-CT$_{o11}^{EC869}$/CdiI$_{o11}^{EC869}$ complex

The CdiA-CT$_{o11}^{EC869}$ is comprised of two domains. Residues Val85 – Arg149 form an N-terminal four α-helical bundle (α1*-α4*); and residues Thr153 – Lys297 form a C-terminal ellipsoidal α/β domain containing one 310-helix, four α-helices (α1-α4) and seven β-strands (Figures 2.2a & 2.3a). The central mixed β-sheet (β2, β3, β6, β7, β1) of the C-terminal domain forms a half β-barrel-like structure. Two helices (α3, α4) are located on the outside of this half-barrel, and the C-terminal end of α1 interacts with its central core. A β-hairpin (β4, β5) protrudes from the C-terminal domain near β2 and the extended loop region (L1). The CdiI $_{o11}^{EC869}$ immunity protein consists of five α-helices and eight β- strands that form two β-sheets (Figures 2.2a & 2.3b). The central six-stranded antiparallel β-sheet (β3a′&b′, β2′, β1′, β4′, β5′, β8′) is decorated with four α-helices (α1′, α2′, α3′, α4′) inserted between strands β3′ and β4′. A fifth C-terminal helix (α5′) runs parallel to the central β-sheet, and a short two- stranded β-sheet (β6′, β7′) connects β5′ and β8′ of the central β-sheet.

The CdiA-CT$_{o11}^{EC869}$/CdiI$_{o11}^{EC869}$ binding interaction is mediated by β-augmentation, in which the toxin donates its β-hairpin (β4, β5) to the immunity protein to produce a six-stranded antiparallel β-sheet. The augmented sheet consists of CdiI$_{o11}^{EC869}$ β6′ and β7′, followed by the β4-β5 hairpin from CdiA-CT$_{o11}^{EC869}$, and completed by CdiI$_{o11}^{EC869}$ β3a′ and β2′ (Figure 2.2b). The interface is stabilized by ion-pairs between the toxin β−hairpin and the immunity central β-sheet and α2′ (Figure 2.3c & Table 2.2). Additionally, there are contributions by the toxin L1 loop region interacting with α2′ of CdiI$_{o11}^{EC869}$ facilitated by ion-pairs and hydrophobic contacts (Figures 2.2c & 2.3c). The toxin/immunity interface buries 1996 Å^2 of the surface area, approximately 12% and 10% of the solvent accessible surface area of CdiA-CT$_{o11}^{EC869}$ and CdiI$_{o11}^{EC869}$, respectively. The EC869 toxin/immunity proteins have high affinity for one another (K_d = 17.8 ± 7 nM), and the complex has greater thermal stability (T_m 65.1 ± 0.9°C) than isolated CdiA-CT$_{o11}^{EC869}$ (T_m 53.8 ± 1.4°C) and CdiI$_{o11}^{EC869}$ (T_m 50.1 ± 0.9°C) (Figure 2.3d).

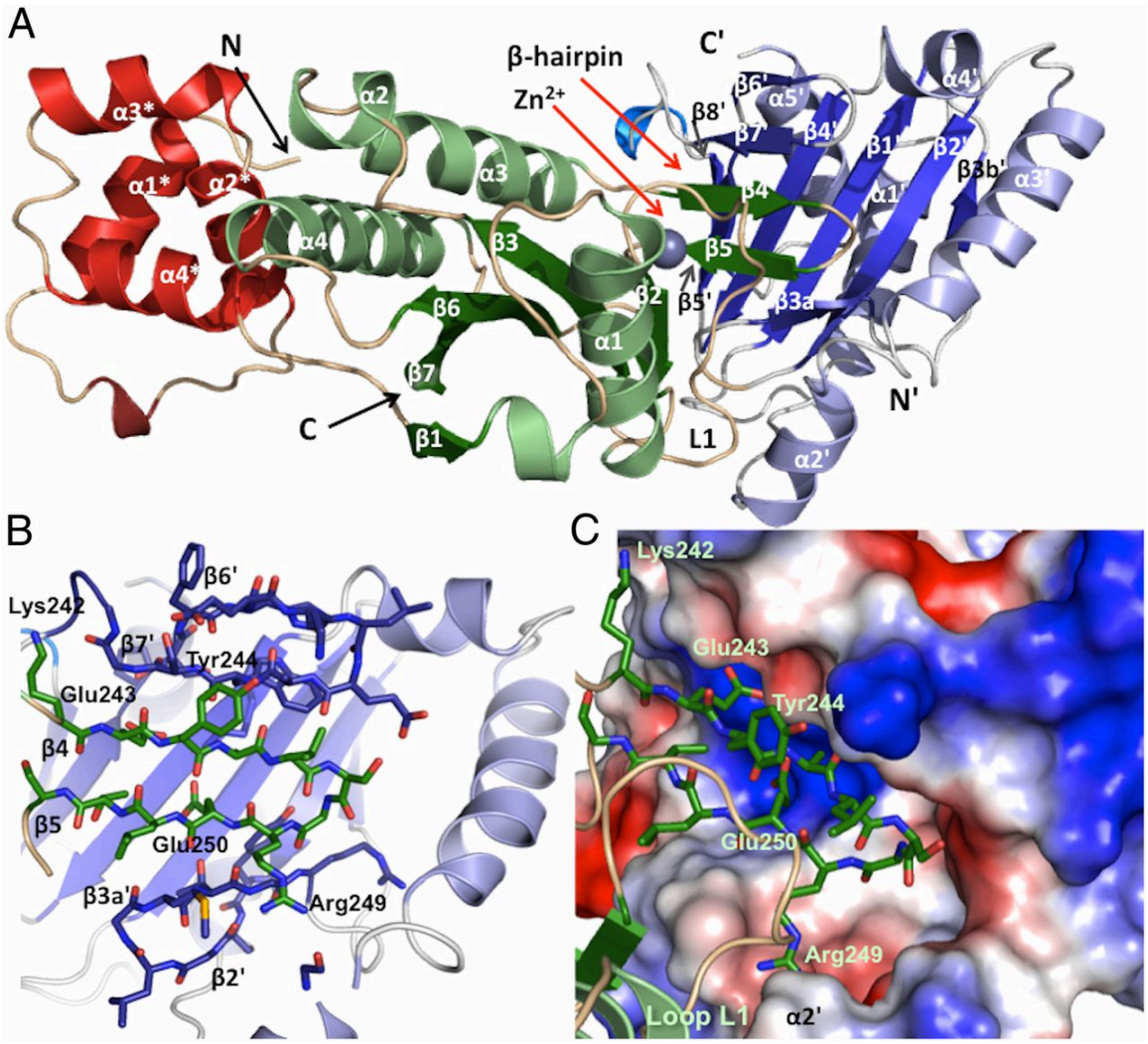

Figure 2.2. Structure of the EC869 CdiA-CT_{o11}/$CdiI_{o11}$ complex. (A) Ribbon representation of the CdiA- CT_{o11}^{EC869}/$CdiI_{o11}^{EC869}$complex. CdiA-CTo11EC869 contains two domains, an N-terminal α-helical bundle (red) and C-terminal α/β nuclease domain (green). The four helices marked with asterisks (*) form the N-terminal helical bundle of CdiA-CT_{o11}^{EC869}. The $CdiI_{o11}^{EC869}$ immunity protein (blue) is comprised of a single α/β domain. The secondary structure elements of each protein are identified and their N- and C- termini indicated. All immunity protein elements are denoted with a prime symbol (′) to differentiate them from the toxin secondary structure elements. The active site Zn^{2+} ion is depicted as a purple sphere. (B) The CdiA-CT_{o11}^{EC869} and $CdiI_{o11}^{EC869}$ proteins interact through β-augmentation. The β4-β5 hairpin of CdiA-CT_{o11}^{EC869} (carbon atoms, green) inserts into the $CdiI_{o11}^{EC869}$ immunity protein (carbon atoms, blue) to form a six-stranded antiparallel β-sheet. β-hairpin residues and ion-pairs are represented as sticks (where the oxygen, nitrogen and sulfur atoms are colored red, royal blue and yellow, respectively). (C) CdiA-CT_{o11}^{EC869} β-hairpin (green sticks) along with the extended loop region L1 fits snugly into the molecular surface representation of $CdiI_{o11}^{EC869}$. White surfaces represent hydrophobic regions, and the red and blue surfaces indicate negative and positive electrostatic potential, respectively.

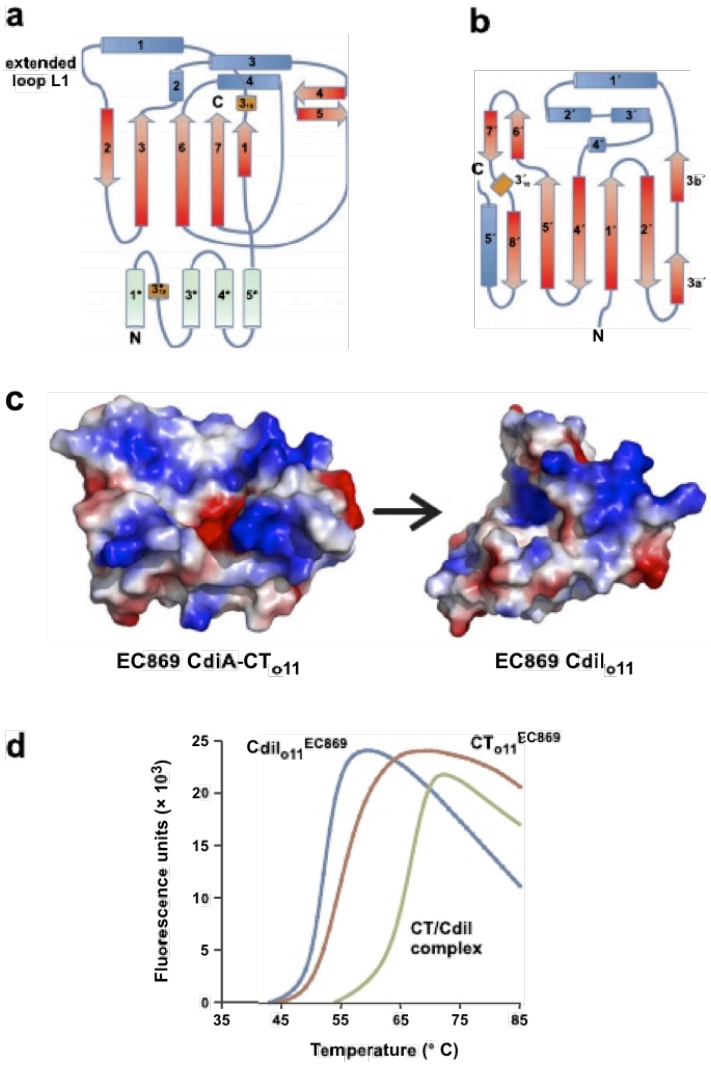

Figure 2.3. Topology and stability of the CdiA-CT$_{o11}^{EC869}$/CdiI$_{o11}^{EC869}$ complex. (A) CdiA-CT$_{o11}^{EC869}$ topology. Secondary structure elements are numbered from N to C terminus. Blue and light green cylinders represent α-helixes, red arrows represent β-strands, and 310 helices are depicted as short orange cylinders. The five helices marked with asterisks (*) form the N-terminal helical bundle of CdiA-CT$_{o11}^{EC869}$. (B) CdiI$_{o11}^{EC869}$ topology. Secondary structure elements are numbered and color-coded as in A. All immunity elements are denoted with a prime symbol (′) to differentiate them from the toxin secondary structure elements. (C) Surface electrostatic representation of CdiA-CT$_{o11}^{EC869}$ and CdiI$_{o11}^{EC869}$. Positive and negative electric isopotential of +70 kT/e and −70 kT/e are shown in blue and red, respectively. The arrow indicates how the two proteins fit together to form a heterodimeric complex. (D) Thermostability of the CdiA-CT$_{o11}^{EC869}$/CdiI$_{o11}^{EC869}$ complex. Differential scanning fluorimetry curves are presented for isolated CdiI$_{o11}^{EC869}$ (blue) and CdiA-CT$_{o11}^{EC869}$ (orange) proteins, as well as the CdiA-CT$_{o11}^{EC869}$/CdiI$_{o11}^{EC869}$ complex (green). All experiments were performed at least three times and representative curves are presented.

Table 2.2. Hydrogen bonds and ion pairs between CdiA-CT/CdiI toxin/immunity proteins

Species	CdiA-CT toxin	CdiI immunity	Distance (Å)
EC869o11			
	Tyr244 OH	Lys128 NZ	3.99
	Glu250 OE1	Arg122 NH2	3.05
	Glu250 OE2	Asn12 ND2	3.38
	Glu243 OE1	Arg122 NE	3.25
	Glu243 OE2	Arg122 NE	3.85
	Glu245 OE1	Lys109 NZ	2.98
	Lys242 NZ	Ile137 O	2.76
	Lys242 NZ	Gly134 O	3.33
	Arg249 NH2	Phe75 O	2.92
	Ser247 OG	Glu130 OE1	2.77
Bp1026b			
	Arg233 NH2	Glu49 OE2	2.8
	Arg233 NE	Glu49 OE2	3.46
	Arg233 NE	Glu49 OE1	3.51
	Ser244 OG	Arg41 NH1	3.81
	Ser245 OG	Arg41 NE	3.71
	Asn239 ND2	Glu53 OE1	3.0
	Lys242 NZ	Asp4 OD1	2.94
	Lys242 NZ	Asp96 OD1	3.07
	Asn217 ND2	Asn67 OD1	3.92
	Asn217 ND2	Asn67 O	2.79
	Asp214 OD1	Lys69 NZ	2.82
	Asn204 OD1	Lys69 NZ	3.03
	Glu187 OE1	Lys69 NZ	3.36
	Pro210 O	Asn67 ND2	2.87
	Lys242 O	Ala2 N	3.11
	Ala228 N	Glu98 OE1	2.76
	Ala228 N	Glu98 OE2	3.08
	Thr238 OG	Asn97 ND2	2.83
	Asp177 OD2	Arg101 NH2	2.57
	Asp177 OD2	Arg101 NH1	3.36
	Asp177 OD1	Arg101 NH2	3.26
	Thr180 OG1	Arg101 NH2	3.75
	Thr180 OG1	Gln95 NE2	3.41

Structure of the CdiA-CT$_{II}^{Bp1026b}$/CdiI$_{II}^{Bp1026b}$ complex

The CdiA-CT$_{II}^{Bp1026b}$ toxin consists of a seven-stranded mixed β-sheet and three α-helices. Like the C-terminal domain of the EC869 toxin, the central β-sheet of CdiA-CT$_{II}^{Bp1026b}$ forms a half β-barrel-like structure with the C-terminal end of a long α-helix (α1) running through its central cavity. The remaining α-helices (α2, α3) decorate the outside of the half-barrel (Figures 2.4a & 2.5a). The CdiI$_{II}^{Bp1026b}$ immunity protein has a simple topology with a five-stranded antiparallel β-sheet decorated with three α-helices (Figures 2.4a & 2.5b). The CdiA-CT$_{II}^{Bp1026b}$/CdiI$_{II}^{Bp1026b}$ complex interface is dominated by electrostatic interactions via residue

side-chains (Figures 2.4b & 2.5c). These interactions are mediated by residues located within the N-terminal half of the long α-helix α1, α2 and extended L1 loop of the toxin domain that interact with residues located at one end of the β-sheet and α2′ of the immunity protein. The interaction network is extensive, with at least 20 ion-pairs and direct hydrogen bonds between the toxin and immunity proteins (Table 2.2). In addition, a network of water-mediated hydrogen bonds also contributes to the CdiA-$CT_{II}^{Bp1026b}$/$CdiI_{II}^{Bp1026b}$ interface. The CdiA-$CT_{II}^{Bp1026b}$/$CdiI_{II}^{Bp1026b}$ interface buries 2044 Å2, which corresponds to 17% and 22% of CdiA-$CT_{II}^{Bp1026b}$ and $CdiI_{II}^{Bp1026b}$ total surface area (respectively). The Bp1026b complex has a dissociation constant of 21.1 ± 9 nM and a melting temperature (T_m) of 70.4 ± 0.7 °C (Figure 2.5d). The T_m of CdiA-$CT_{II}^{Bp1026b}$ is 52.3 ± 0.7 °C and $CdiI_{II}^{Bp1026b}$ is 60.9 ±1.2 °C (Figure 2.5d), again demonstrating that the complex is more stable than the isolated toxin and immunity proteins.

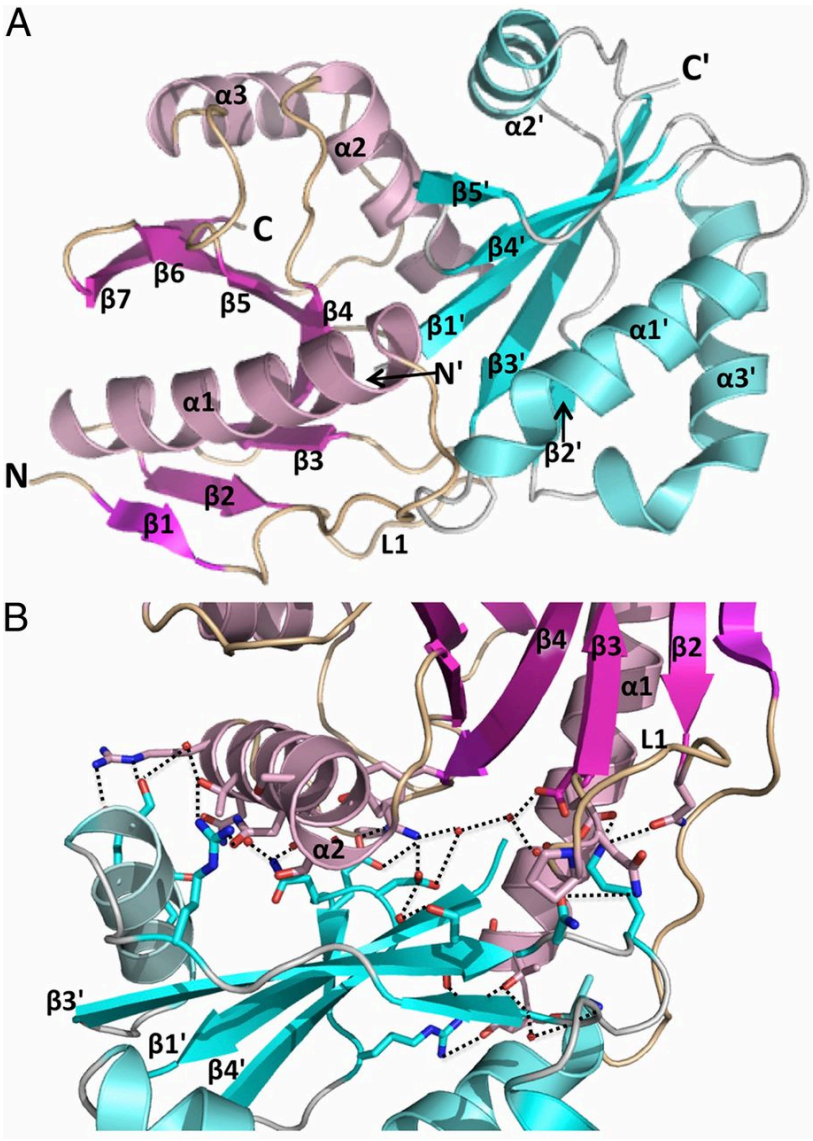

Figure 2.4. Structure of the Bp1026b CdiA-CT$_{II}$/CdiI$_{II}$ complex. (A) The CdiA-CT$_{II}^{Bp1026b}$ toxin (pink) and CdiI$_{II}^{Bp1026b}$ immunity protein (cyan) are depicted in ribbon representation with secondary structure elements indicated. All immunity protein elements are denoted with a prime symbol (′) to differentiate them from the toxin secondary structure elements. (B) The interface between CdiA-CT$_{II}^{Bp1026b}$ (pink) and CdiI$_{II}^{Bp1026b}$ (cyan) is formed by an extensive network of ion-pairs and hydrogen-bonds. Within the network, interacting residue side-chains are represented as sticks (oxygen and nitrogen atoms colored red and blue, respectively), water molecules as red spheres, and interacting bonds as black dotted lines.

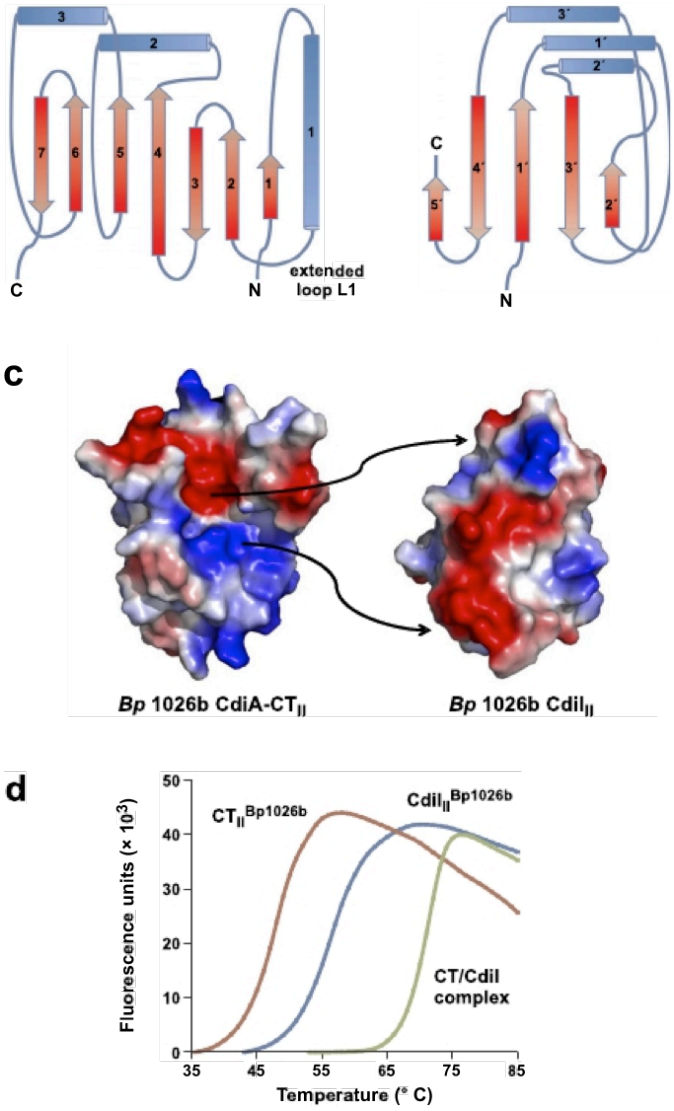

Figure 2.5. Topology and stability of the CdiA-CT$_{II}^{Bp1026b}$/CdiI$_{II}^{Bp1026b}$ complex. (A) CdiA-CT$_{II}^{Bp1026b}$ topology. Secondary structure elements are numbered from N to C terminus. Blue cylinders represent α-helixes and red arrows represent β-strands. (B) CdiI$_{II}^{Bp1026b}$ topology. Secondary structure elements are numbered and color-coded as in A. All immunity elements are denoted with a prime symbol (′) to differentiate them from the toxin secondary structure elements. (C) Surface electrostatic representation of CdiA-CT$_{II}^{Bp1026b}$ and CdiI$_{II}^{Bp1026b}$. Positive and negative electric isopotential of +70 kT/e and −70 kT/e are shown in blue and red, respectively. The arrows indicate the electrostatic complementarity of the two proteins as they come together to form a complex. (D) Thermostabilty of the CdiA-CT$_{II}^{Bp1026b}$/CdiI$_{II}^{Bp1026b}$ complex. Differential scanning fluorimetry curves are presented for isolated CdiI$_{II}^{Bp1026b}$ (blue) and CdiA-CT$_{II}^{Bp1026b}$ (orange) proteins, as well as the CdiA CT$_{II}^{Bp1026b}$/CdiI$_{II}^{Bp1026b}$ complex (green). All experiments were performed at least three times and representative curves are presented.

Comparison of CdiA-CT/CdiI complex structures

Although the two toxin domains share only ~15% sequence identity and have distinct topologies, their three-dimensional structures superimpose with an rmsd of 3.9 Å and Z-score of 5.8 (21) (Figures 2.6a & 2.7a). Notably, CdiA-$CT_{II}^{Bp1026b}$ lacks the β-hairpin element found in the CdiA-CT_{o11}^{EC869} toxin. Both CdiA-CT C-terminal domains are structurally similar (21) to type IIS restriction endonucleases (22), suggesting that the toxins have metal-dependent DNase activity. Furthermore, metal K- edge absorption analysis revealed that native CdiA-CT_{o11}^{EC869}/$CdiI_{o11}^{EC869}$ crystals have significant zinc content. Based on structural homology (21) to the BspD6I endonuclease, CdiA-CT_{o11}^{EC869} residues Glu177, Asp198, Ser209 and Lys211 are predicted to form the nuclease active site (Table 2.3). Additionally, extra electron density within the active site vicinity was modeled as a Zn^{2+} ion, which is coordinated by Glu177, Asp198 and three water molecules in a ββα–metal motif (β2, β3, α1) (Figure 2.6b) (23). Similarly, CdiA-$CT_{II}^{Bp1026b}$ residues Glu187, Asp214, Asp223 and Lys242 are predicted to form an active site and coordinate a catalytic metal ion within a ββα–metal motif (β3, β4, α1) (Figure 2.6a & Table 2.3). However, there is no density attributable to an active-site cation in the CdiA-$CT_{II}^{Bp1026b}$/$CdiI_{II}^{Bp1026b}$ complex, presumably because direct hydrogen-bonds between the immunity protein and active site residues preclude metal binding. These predictions are supported by our previous work showing that CdiA-$CT_{II}^{Bp1026b}$ is a Mg^{2+}-dependent tRNase and that its nuclease activity is ablated by the Asp214Ala mutation (5).

Despite their common α/β fold, the $CdiI_{o11}^{EC869}$ and $CdiI_{II}^{Bp1026b}$ immunity proteins share little sequence identity (~12%) or structural homology (Z-score of 0.2) (Figure 2.7b). Moreover, each CdiI protein binds its cognate toxin at a completely different location (Figure 2.6c), consistent with the specificity of CDI immunity. The $CdiI_{II}^{Bp1026b}$ protein binds directly over the central core of CdiA-$CT_{II}^{Bp1026b}$ to produce a 'closed clam' structure. This structure provides a mechanism for immunity because $CdiI_{II}^{Bp1026b}$ occludes the predicted nuclease active site

(Figure 2.6c), and presumably prevents the toxin from binding substrate. In contrast, the $CdiI_{o11}^{EC869}$ immunity protein binds to the C-terminal side of the toxin domain in a 'lock-and-key' type manner producing an elongated complex that buries little of the toxin's central core (Figure 2.6c). Because the CdiA-CT_{o11}^{EC869} active site is solvent exposed in the complex, it is not immediately clear how $CdiI_{o11}^{EC869}$ neutralizes the toxin. It is possible that $CdiI_{o11}^{EC869}$ prevents nucleic acid binding, or alternatively, the conformation of the toxin could be altered upon binding the immunity protein.

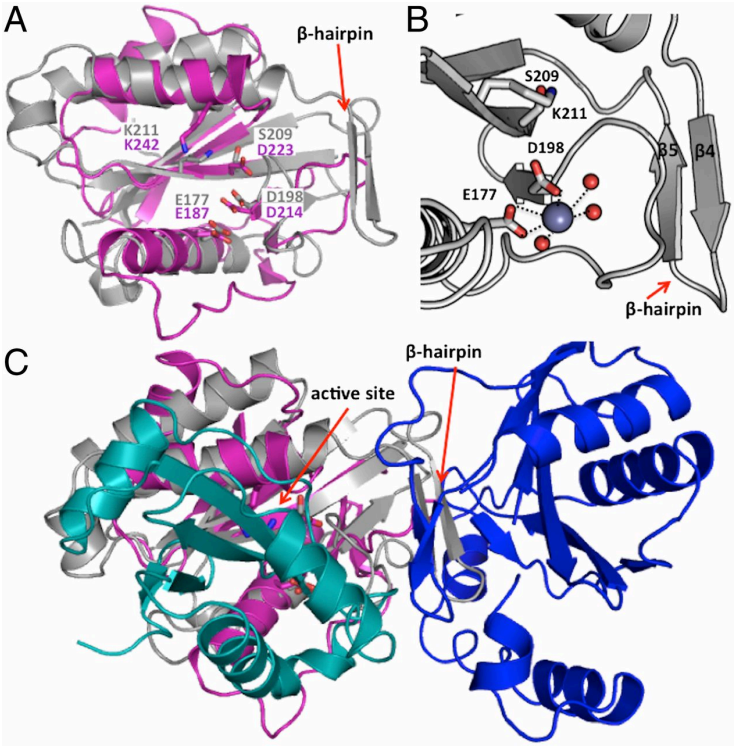

Figure 2.6. Structural superimposition of EC869 and Bp1026b CdiA-CT/CdiI protein complexes. (A) Predicted active site residues of the EC869 and Bp1026b toxin domains. The two toxin domains are superimposed and active site residues are rendered as stick representations. EC869 and Bp1026b carbon atoms are colored grey and pink, respectively; and oxygen and nitrogen atoms colored red and blue, respectively. (B) Coordination of Zn^{2+} within the CdiA-CT_{o11}^{EC869} active site. The Zn^{2+} ion is depicted as a purple sphere, ordered waters as smaller red spheres, and interacting bonds with Zn^{2+} are depicted as black dotted lines. (C) Superimposition of the EC869 and Bp1026b CdiA-CT/CdiI protein complexes. Ribbon representations of CdiA-$CT_{II}^{Bp1026b}$ and $CdiI_{II}^{Bp1026b}$ are colored pink and cyan (respectively), and the C-terminal domain of CdiA-CT_{o11}^{EC869} and $CdiI_{o11}^{EC869}$ are colored grey and blue (respectively). The C-terminal toxin domains superimpose upon one another, whereas the immunity proteins do not. The N-terminal α-helical domain of CdiA-CT_{o11}^{EC869} has been omitted for clarity.

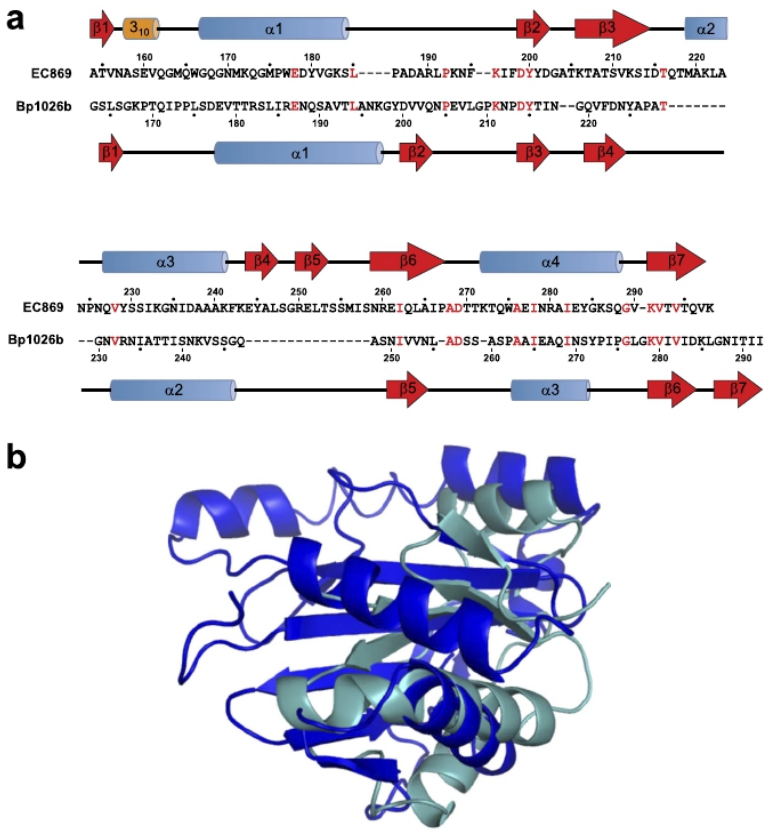

Figure 2.7. Structure-based alignment of CdiA-CT toxins and CdiI immunity proteins. (A) The sequences for the CdiA-CT_{o11}^{EC869} and CdiA-$CT_{II}^{Bp1026b}$ toxin domains were aligned based on structural homology. The corresponding secondary structure elements are shown above and below the sequences, with α-helixes depicted as blue cylinders and β-strands as orange arrows. Amino acid residues are given in one-letter code and red residues are identical between the two toxins. (B) The $CdiI_{II}^{Bp1026b}$ (blue) and $CdiI_{o11}^{EC869}$ (cyan) immunity proteins are depicted as ribbons and superimposed upon one another.

Table 2.3. Predicted active site residues in CdiA-CT_{o11}^{EC869} and CdiA-$CT_{II}^{Bp1026b}$ based on structural homology to type IIS restriction endonucleases

Protein	**Active site residues**			
CdiA-CT_{o11}^{EC869}	Glu177	Asp198	Ser209	Lys211
CdiA-$CT_{II}^{Bp1026b}$	Glu187	Asp214	Asp223	Lys242
R.BspD6I2*	Glu22	Asp60	Glu73	His93
N.BspD6I†	Glu418	Asp456	Glu496	His489
FdxN element excision factor‡	Glu9	Glu58	Lys60	Gln78

*PDB ID code 2PI4. †PDB ID code 2EWK. ‡PDB ID code 2OKF.

CdiA-CT toxins have distinct nuclease activities (Experiments performed by Hayes lab, UCSB)

To test whether CdiA-CT_{o11}^{EC869} is a DNase, we isolated the domain from its immunity protein and assayed the purified toxin for nuclease activity in vitro. CdiA-CT_{o11}^{EC869} converted supercoiled plasmid DNA into an open-circular form in the presence of Mg^{2+} (Figure 2.8a), consistent with "nickase" activity in which only one strand of double-stranded DNA is cleaved (22). Because Zn^{2+} is coordinated in the predicted CdiA-CT_{o11}^{EC869} active site, we also tested nuclease activity with this cation. Remarkably, Zn^{2+} greatly stimulated DNase activity, leading to the complete degradation of both supercoiled and linear plasmid substrates (Figure 2.8a). CdiA-CT_{o11}^{EC869} activity was completely blocked by $CdiI_{o11}^{EC869}$ in reactions supplemented with Mg^{2+}, but nickase activity was still apparent in the presence of Zn^{2+} (Figure 2.8a). In contrast, non-cognate $CdiI_{II}^{Bp1026b}$ immunity protein had no effect on DNase activity (Figure 2.8a). We also mutated two of the predicted active site residues in CdiA-CT_{o11}^{EC869} to test their role in catalysis. Both Glu177Ala and Asp198Ala mutants of CdiA-CT_{o11}^{EC869} co-purified with His_6-tagged $CdiI_{o11}^{EC869}$, indicating that the toxin variants retain their native fold, but neither mutant exhibited DNase activity in vitro (Figure 2.8b). We previously reported that CdiA-$CT_{II}^{Bp1026b}$ is a tRNase (5), but its structural resemblance to CdiA- CT_{o11}^{EC869} suggests that it may also possess Zn^{2+}-dependent DNase activity. However, purified CdiA- $CT_{II}^{Bp1026b}$ showed no nuclease activity on plasmid DNA in the presence of either Zn^{2+} or Mg^{2+}, but readily cleaved tRNA under the same conditions (Figure 2.8c). Similarly, the CdiA-CT_{o11}^{EC869} toxin is specific for DNA, with no nuclease activity detected on tRNA substrates. We next sought to determine where the CdiA-$CT_{II}^{Bp1026b}$ toxin cleaves its tRNA target. Digested tRNA appears to be nearly the same size as full-length tRNA (Figure 2.8c), suggesting the toxin cleaves near either the 5′- or 3′-ends of the molecules. S1 nuclease protection analysis revealed that CdiA-$CT_{II}^{Bp1026b}$ cleaves *E. coli* $tRNA_2^{Arg}$ after residues A70 and U71 in the aminoacyl acceptor stem. These sites suggest that

the toxin binds double-stranded RNA, but cuts only one strand of the duplex to inactivate tRNA. Together, these observations demonstrate that each CdiA-CT has a distinct nuclease activity and metal requirement despite sharing a common fold.

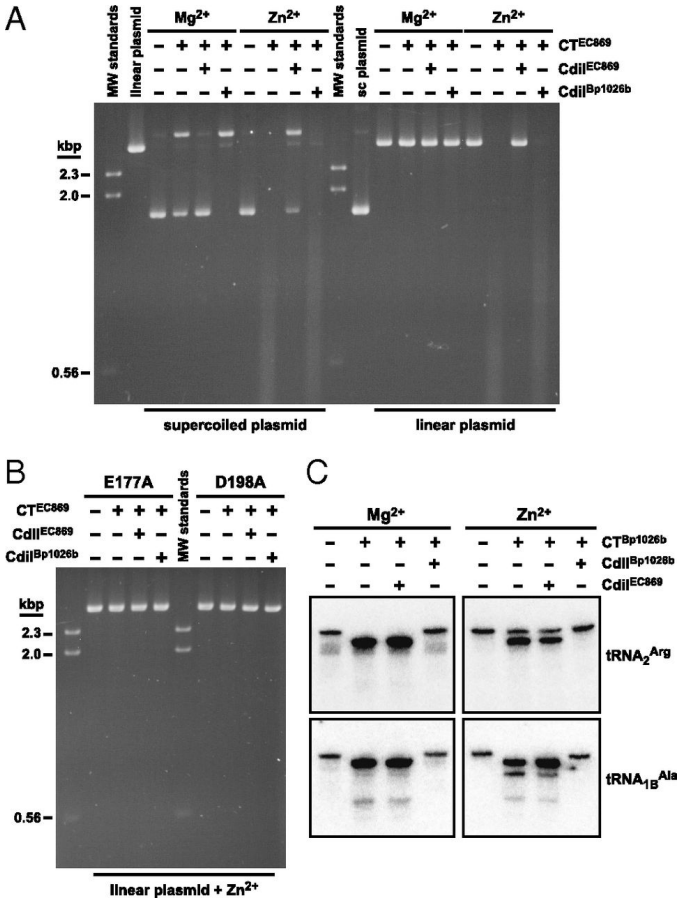

Figure 2.8. CdiA-CT toxins have distinct nuclease activities. (A) DNase activity of the CdiA-CT_{o11}^{EC869} toxin on supercoiled and linear plasmid substrates. Plasmid DNA was incubated with purified CdiA- CT_{o11}^{EC869} in the presence of either Mg^{2+} or Zn^{2+} and reactions analyzed by agarose gel electrophoresis and ethidium bromide staining. Reactions also included either purified $CdiI_{o11}^{EC869}$ or $CdiI_{II}^{Bp1026b}$ immunity proteins where indicated. Untreated supercoiled linear plasmid substrates were included as controls for the migration of undigested DNA. The migration positions of linear molecular weight (MW) DNA standards are indicated in kilo base-pairs (kbp). (B) Mutation of predicted active site residues ablates CdiA-CT_{o11}^{EC869} DNase activity. Linear plasmid DNA was incubated with purified CdiA- CT_{o11}^{EC869} containing the Glu177Ala (E177A) and Asp198Ala (D198A) mutations in buffer supplemented with Zn^{2+}. Reactions also contained $CdiI_{o11}^{EC869}$ or $CdiI_{II}^{Bp1026b}$ immunity proteins where indicated. (C) The CdiA-$CT_{II}^{Bp1026b}$ toxin has tRNase activity. Purified *E. coli* tRNA was treated with CdiA-$CT_{II}^{Bp1026b}$ toxin in reactions supplemented with Mg^{2+} or Zn^{2+}. Reactions contained $CdiI_{o11}^{EC869}$ or$CdiI_{II}^{Bp1026b}$ immunity proteins where indicated, and were run on denaturing polyacrylamide gels and analyzed by Northern blot hybridization using radiolabeled probes to $tRNA_2^{Arg}$ and $tRNA_{1B}^{Ala}$.

The CdiA-CT$_{o11}^{EC869}$ toxin degrades target cell DNA during CDI (Experiments performed by Hayes lab, UCSB)

We next asked whether the CdiA-CT$_{o11}^{EC869}$ toxin also displays DNase activity when expressed inside cells. We also examined a truncated CdiA-CT$_{o11}^{EC869}$ protein (residues Ala142 – Lys297) in these experiments to determine whether the N-terminal α-helical bundle (Figure 2.2a) is required for DNase activity. This toxin-encoding sequence could not be cloned in the absence of the cognate immunity gene; therefore, we used controlled proteolysis of CdiI$_{o11}^{EC869}$ to activate the CdiA-CT$_{o11}^{EC869}$ toxin inside *E. coli* cells (8, 24). Briefly, the C-terminus of CdiI$_{o11}^{EC869}$ was tagged with the ssrA(DAS) peptide, which targets the immunity protein for degradation by the ClpXP protease, thereby liberating the CdiA-CT to exert its toxic activity. Visualization of DAPI-stained cells showed that chromosomal DNA was lost after three hours of toxin activation (see supplemental of Morse *et al.* (20)). This avid DNase activity is consistent with the Zn^{2+}-dependent activity observed in vitro and strongly suggests that Zn^{2+} is the relevant cation for in vivo activity. In contrast, CdiA-CT$_{o11}^{EC869}$ carrying the Asp198Ala active site mutation had no effect on cellular DAPI staining (see supplemental of Morse *et al.* (20)). These results demonstrate that the C-terminal α/β domain of CdiA-CT$_{o11}^{EC869}$ is sufficient for DNase activity. Presumably, the α-helical bundle domain and the remainder of the N- terminal region perform another function during CDI.

Finally, we tested whether CdiA-CT$_{o11}^{EC869}$ DNase activity is responsible for growth inhibition during cell-mediated CDI. We generated a cosmid-borne chimeric CDI system, in which the *cdiA-CT*$_{o11}^{EC869}$/*cdiI*$_{o11}^{EC869}$ coding sequences are fused to the *E. coli* EC93 *cdiA* gene at the VENN-encoding sequence. The resulting EC93-EC869$_{o11}$ chimeric system was introduced into GFP-labeled *E. coli* to produce green fluorescent inhibitor cells. The inhibitor cells were then co-cultured with DsRed-labeled target cells, allowing the two cell populations to be distinguished by fluorescence microscopy. Upon initial mixing, the green inhibitor and red

target cells both have the same nucleoid morphology as assessed by DAPI staining (Figure 2.9). However, target cells underwent substantial changes in morphology and lost DAPI staining after six hr of co-culture with inhibitor cells (Figure 2.9a). These changes were paralleled by a dramatic loss of target cell viability during co-culture (Figure 2.9f). Target cells expressing the $CdiI_{o11}^{EC869}$ immunity protein retained genomic DNA during co-culture with inhibitor cells and suffered no loss of viability (Figure 2.9c), but the non-cognate $CdiI_{II}^{Bp1026b}$ immunity protein provided no protection (Figure 2.9d). Moreover, introduction of the CdiA-CT_{o11}^{EC869} Asp198Ala active site mutation into EC93-$EC869_{o11}$ chimera system resulted in a loss of growth inhibition and DNase activity (Figure 2.9e). Together, these results indicate that DNase activity is responsible for growth inhibition and that the CdiA-CT_{o11}^{EC869} toxin domain is translocated into the target cell cytoplasm during CDI.

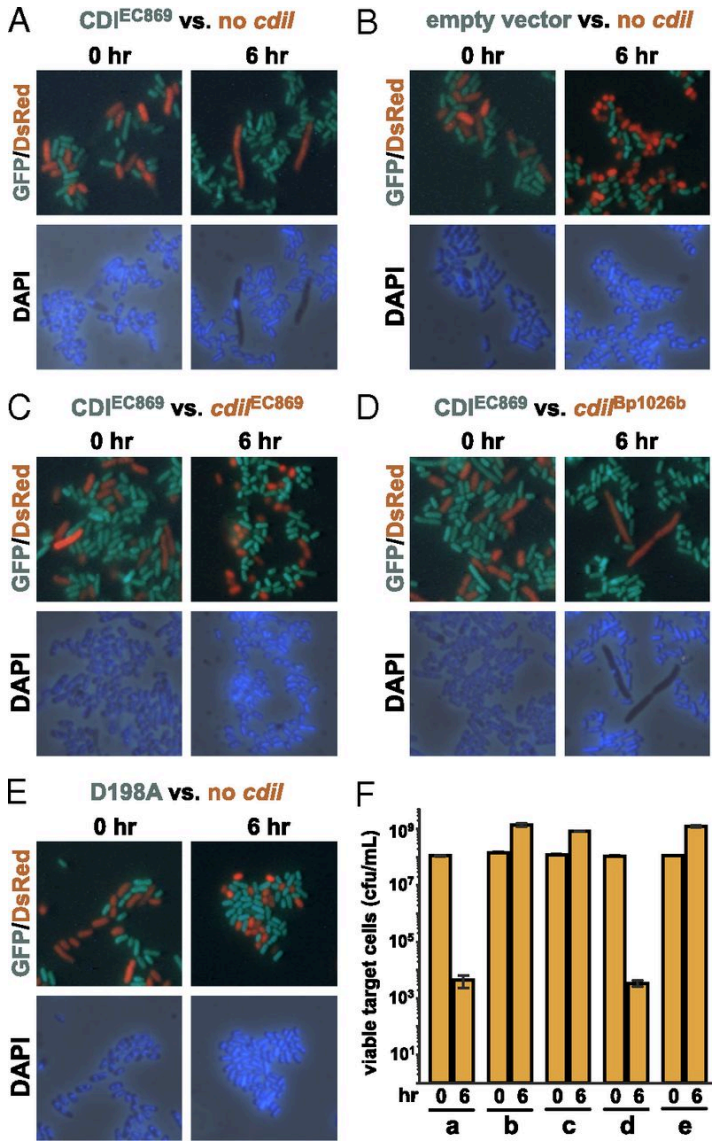

Figure 2.9. The CdiA-CT_{o11}^{EC869} toxin degrades DNA during contact-dependent growth inhibition (CDI). GFP-labeled *E. coli* inhibitor cells (green) were mixed with DsRed-labeled target cells (red) and grown in shaking broth cultures. Co-cultures were sampled at 0 hr and 6 hr, and stained with DAPI to visualize cellular DNA by fluorescence microscopy. (A) EC93-$EC869_{o11}$ inhibitor cells versus targets that lack an immunity gene. (B) Mock inhibitor cells (carrying an empty vector cosmid) versus targets that lack an immunity gene. (C) EC93-$EC869_{o11}$ inhibitors versus target cells that carry the cognate $cdiI_{o11}^{EC869}$ gene. (D) EC93-$EC869_{o11}$ inhibitors versus target cells that carry the non-cognate $cdiI_{II}^{Bp1026b}$ immunity gene. (E) EC93-$EC869_{o11}$ inhibitors carrying the Asp198Ala (D198A) missense mutation versus target cells that lack an immunity gene. (F) Quantification of viable target cells during CDI. The number of viable target cells at 0 and 6 hr were determined as colony forming units (cfu) per mL. The data from competitions corresponding to panels (A) through (E) are indicated. Values are the mean ± SEM for three independent experiments.

Discussion

The CdiA-CT/CdiI structures presented here provide the first glimpse into the CDI toxin/immunity protein network. These complexes bear no resemblance to the well-studied toxin/antitoxin 'addiction module' proteins (25) and are only distantly related to the colicin toxin/immunity family (26). Though structurally distinct, the CDI complexes share some general features with colicin nuclease domains and their immunity proteins. The extensive shape and charge complementarity of the Bp1026b complex interface is reminiscent of the interactions between colicins E5 and D and their cognate immunity proteins (27, 28). Moreover, $CdiI_{o11}^{EC869}$ appears to inactivate its toxin in a manner analogous to a number of colicin systems (e.g. E3, E7/E8/E9), in which the immunity protein binds an 'exosite' adjacent to the toxin active site (29-31). However, the elegant β-augmentation interaction between CdiA-CT_{o11}^{EC869} and $CdiI_{o11}^{EC869}$ has not been reported for any other toxin/immunity complex. Homotypic β-augmentation has been observed in viral capsid assembly (32) and appears to be the underlying mechanism of β-sheet expansion in amyloid diseases (33). Additionally, some signal transduction pathways exploit β-augmentation to mediate heterodimeric interactions. For example, the PDZ domain of neuronal nitric oxide synthase (nNOS) extends a β-hairpin that docks into the peptide-binding groove of syntrophin to produce an expanded β-sheet (34). Though the nNOS-syntrophin interface resembles the EC869 complex, we note that β-augmentation interactions during signal transduction are dynamic and transient. In contrast, the CdiA-CT_{o11}^{EC869}/$CdiI_{o11}^{EC869}$ complex appears to be the first example of a stable heterodimeric interface mediated by β-augmentation.

Comparative sequence analysis suggests that many CdiA-CTs are composites built from two variable domains (5, 7). The structures also indicate that each CdiA-CT is comprised of at least two domains, with the C-terminal nuclease domain forming a stable complex with its cognate immunity protein. The C-terminal domains are also sufficient to inhibit growth when expressed in *E. coli* cells (5), suggesting that they constitute the functional CDI toxin. In

contrast, the CdiA-CT N-terminal regions are not fully resolved in the structures and their functional significance remains unclear. The N-terminal regions are exceptionally labile to proteolysis, suggesting these domains are flexible and perhaps partially disordered. Intrinsically flexible domains are critical for colicin toxin import (35, 36), so perhaps the N-terminal region mediates CdiA-CT transport across the target cell envelope. The N-terminal α-helical bundle of CdiA-CT_{o11}^{EC869} has weak structural homology to diverse membrane-associated proteins, consistent with the translocation hypothesis, but the function of these domains in CDI remains to be determined. Our results together with previous predictions (37) also suggest that many other CdiA-CT toxin domains may have similar structures despite sharing very little sequence identity. However, we note that some CDI toxin family members must possess other folds because the *E. coli* EC93 toxin form pores in target cell membranes (6), and CdiA-CTs from *B. pseudomallei* K96243 and *Erwinia chrysanthemi* EC16 share significant sequence identity with colicins E5 and E3, respectively (5, 6). We are actively solving the structures of several other CdiA-CT/CdiI complexes to determine the spectrum of CDI toxin structural diversity.

Acknowledgements

We thank the Advanced Light Source (ALS) at Berkeley National Laboratories and the Stanford Synchrotron Radiation Lightsource (SSRL) for their invaluable help in data collection. This work was supported by grants AI099687 (C.W.G & C.S.H), GM078634 (C.S.H.) and GM102318 (C.W.G., C.S.H & D.A.L,) from the National Institutes of Health, and UCI CORCL funds (C.W.G.). J.L.E.W. is supported by an NSF Graduate Research fellowship (DGE-1144085). Finally, we thank Angelina Iniguez for technical assistance, Stephanie Aoki for providing cosmid pDAL930, Bruce Braaten for strain DL4259, and Drs. Tom Poulos and Nicholas Chim for critical reading of the manuscript.

References

1. Aoki SK*, et al.* (2010) A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* 468(7322):439-442.
2. Aoki SK*, et al.* (2005) Contact-dependent inhibition of growth in Escherichia coli. *Science* 309(5738):1245-1248.
3. Kajava AV*, et al.* (2001) Beta-helix model for the filamentous haemagglutinin adhesin of Bordetella pertussis and related bacterial secretory proteins. *Molecular microbiology* 42(2):279-292.
4. Aoki SK*, et al.* (2008) Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Molecular microbiology* 70(2):323-340.
5. Nikolakakis K*, et al.* (2012) The toxin/immunity network of Burkholderia pseudomallei contact-dependent growth inhibition (CDI) systems. *Molecular microbiology* 84(3):516-529.
6. Aoki SK, Webb JS, Braaten BA, & Low DA (2009) Contact-dependent growth inhibition causes reversible metabolic downregulation in Escherichia coli. *Journal of bacteriology* 191(6):1777-1786.
7. Diner EJ, Beck CM, Webb JS, Low DA, & Hayes CS (2012) Identification of a target cell permissive factor required for contact-dependent growth inhibition (CDI). *Genes Dev* 26(5):515-525.
8. Poole SJ*, et al.* (2011) Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. *PLoS genetics* 7(8):e1002217.
9. Otwinowski Z & Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* 276:307-326.
10. Ness SR, de Graaff RA, Abrahams JP, & Pannu NS (2004) CRANK: new methods for automated macromolecular crystal structure solution. *Structure* 12(10):1753-1761.
11. de Graaff RA, Hilge M, van der Plas JL, & Abrahams JP (2001) Matrix methods for solving protein substructures of chlorine and sulfur from anomalous data. *Acta crystallographica. Section D, Biological crystallography* 57(Pt 12):1857-1862.
12. Abrahams JP & Leslie AG (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta crystallographica. Section D, Biological crystallography* 52(Pt 1):30-42.
13. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta crystallographica. Section D, Biological crystallography* 62(Pt 9):1002-1011.
14. Terwilliger TC*, et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta crystallographica. Section D, Biological crystallography* 64(Pt 1):61-69.
15. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 4):486-501.
16. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2):213-221.
17. Terwilliger TC*, et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta crystallographica. Section D, Biological crystallography* 65(Pt 6):582-601.

18. Niesen FH, Berglund H, & Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2(9):2212-2221.
19. Concepcion J*, et al.* (2009) Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Combinatorial chemistry & high throughput screening* 12(8):791-800.
20. Morse RP*, et al.* (2012) Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems. *Proceedings of the National Academy of Sciences of the United States of America* 109(52):21480-21485.
21. Holm L, Kaariainen S, Rosenstrom P, & Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* 24(23):2780-2781.
22. Kachalova GS*, et al.* (2008) Structural analysis of the heterodimeric type IIS restriction endonuclease R.BspD6I acting as a complex between a monomeric site-specific nickase and a catalytic subunit. *Journal of molecular biology* 384(2):489-502.
23. Sokolowska M, Czapinska H, & Bochtler M (2009) Crystal structure of the beta beta alpha-Me type II restriction endonuclease Hpy99I with target DNA. *Nucleic acids research* 37(11):3799-3810.
24. McGinness KE, Baker TA, & Sauer RT (2006) Engineering controllable protein degradation. *Mol Cell* 22(5):701-707.
25. Yamaguchi Y, Park JH, & Inouye M (2011) Toxin-antitoxin systems in bacteria and archaea. *Annual review of genetics* 45:61-79.
26. Cascales E*, et al.* (2007) Colicin biology. *Microbiology and molecular biology reviews : MMBR* 71(1):158-229.
27. Luna-Chavez C, Lin YL, & Huang RH (2006) Molecular basis of inhibition of the ribonuclease activity in colicin E5 by its cognate immunity protein. *Journal of molecular biology* 358(2):571-579.
28. Graille M, Mora L, Buckingham RH, van Tilbeurgh H, & de Zamaroczy M (2004) Structural inhibition of the colicin D tRNase by the tRNA-mimicking immunity protein. *The EMBO journal* 23(7):1474-1482.
29. Mate MJ & Kleanthous C (2004) Structure-based analysis of the metal-dependent mechanism of H-N-H endonucleases. *The Journal of biological chemistry* 279(33):34763-34769.
30. Ko TP, Liao CC, Ku WY, Chak KF, & Yuan HS (1999) The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure* 7(1):91-102.
31. Kleanthous C*, et al.* (1999) Structural and mechanistic basis of immunity toward endonuclease colicins. *Nat Struct Biol* 6(3):243-252.
32. Gronenborn AM (2009) Protein acrobatics in pairs--dimerization via domain swapping. *Current opinion in structural biology* 19(1):39-49.
33. Yamasaki M, Li W, Johnson DJ, & Huntington JA (2008) Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* 455(7217):1255-1258.
34. Hillier BJ, Christopherson KS, Prehoda KE, Bredt DS, & Lim WA (1999) Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. *Science* 284(5415):812-815.
35. Loftus SR*, et al.* (2006) Competitive recruitment of the periplasmic translocation portal TolB by a natively disordered domain of colicin E9. *Proceedings of the National Academy of Sciences of the United States of America* 103(33):12353-12358.
36. Housden NG, Loftus SR, Moore GR, James R, & Kleanthous C (2005) Cell entry mechanism of enzymatic bacterial colicins: porin recruitment and the thermodynamics of receptor binding. *Proceedings of the National Academy of Sciences of the United States of America* 102(39):13849-13854.

37. Zhang D, Iyer LM, & Aravind L (2011) A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic acids research* 39(11):4532-4552.

CHAPTER 3

Insights into the β-augmentation interface of *E. coli* EC869 family toxin and immunity proteins

Summary

Contact-dependent frown inhibition (CDI) is a phenomenon used by gram-negative bacteria to inhibit the growth of neighboring cells. The C-terminal toxin domains of cell surface CdiA proteins (CdiA-CT) are cleaved upon contact with a target bacterium and enter the target's cytosol, disrupting growth, unless neutralized by a cognate CdiI immunity protein. CdiA-CT from *Escherichia coli* EC869 is a robust Zn^{2+}-dependent DNase, unless inhibited by binding to $CdiI^{EC869}$. The EC869 CdiA-CT/CdiI complex is formed by the interaction of the β-hairpin of $CdiA\text{-}CT^{EC869}$ with $CdiI^{EC869}$ to form an anti-parallel β-sheet with strands denoted by both toxin and immunity proteins. Here we show that formation of this β-augmentation interface is a requirement for all members of the EC869-family of toxin/immunity complexes, and that the specific amino acid sequence of the toxin β-hairpin contributes to cognate immunity specificity. Structure determination of an additional EC869-family CdiA-CT/CdiI complex member demonstrates conservation of the β-augmentation interface. While the overall protein folds of the toxin and immunity proteins are conserved, heterologous EC869 family CdiA-CT and CdiI members cannot form stable complexes. Finally, we solved the structure of $CdiI^{EC869}$ in complex with a macrocyclic peptide that mimics the β-hairpin of the toxin, illustrating that these types of molecules could potentially be used to prevent CdiA-CT/CdiI complex formation.

Introduction

Bacteria have evolved complex strategies to compete and communicate in their environments. Contact-dependent growth inhibition (CDI), an inter-bacterial competition mechanism, was recently discovered in gram-negative bacteria, and is used to inhibit the growth of neighboring cells (1, 2). CDI requires cell-to-cell contact, and is mediated by the CdiB/CdiA two-partner secretion system. In the first step of CDI, the outer-membrane protein CdiB facilitates transport of the large exoprotein CdiA (250-600 kDa) to the cell surface. CdiA proteins displayed on the cell surface are thought to extend out and interact with specific receptors on the target cell surface, which initiates delivery of a CdiA-derived toxin into the target cell (3-5). The CDI toxin is located at the C-terminus of CdiA (CdiA-CT), and contains 250-300 residues. To protect against auto-inhibition, CDI^+ cells produce small CdiI immunity proteins that block CdiA-CT activity. Over 60 families of CdiA-CT/CdiI complexes have been identified (2), suggesting a wide variety of CDI toxins. Indeed, different CdiA-CTs have been shown to inhibit the growth of target cells by employing distinct toxin activities, such as degrading DNA (2, 6), cleaving tRNA (7, 8) or rRNA (9) molecules, and dissipating the proton motive force (10). CdiA-CT/CdiI interactions are highly specific between their cognate pairs, thus CdiI proteins offer no protection against CdiA-CT toxins from other bacterial species/strains (7, 11). Despite this, there is some evidence of weak interactions between heterologous toxins and immunities from the same CDI family (11).

In a previous study, high-resolution X-ray crystal structures were determined of CdiA-CT/CdiI complexes from *Escherichia coli* EC869 (EC869) and *Burkholderia pseudomallei* 1026b (Bp1026b) (6). The Bp1026b complex was derived from the CDI locus on chromosome II of Bp1026b, while the EC869 complex is encoded by the 11th “orphan” module of *E. coli* EC869. Remarkably, despite little sequence identity (~15%) between the two CdiA-CT toxin domains, their tertiary structures superimpose with an rmsd of 3.9 Å and are structurally similar to type IIS

restriction endonucleases. The interfaces between the toxin and immunity proteins of EC869 and Bp1026b complex structures are distinct. Notably, within the EC869 CdiA-CT/CdiI complex, the toxin has two additional β-strands (β4 and β5) that form a protruding β-hairpin that inserts into $CdiI^{EC869}$, producing a six-stranded antiparallel β-sheet containing β-strands from both toxin and immunity proteins. The Bp1026b toxin structure contains no such β-hairpin. This β-augmentation interaction between $CdiA\text{-}CT^{EC869}$ and $CdiI^{EC869}$ appears to be unique to all previously solved toxin/antitoxin and colicin toxin/immunity complex structures (12, 13). A primary sequence search identifies a family of "EC869-like" CdiA-CT/CdiI pairs, suggesting that this unique interaction occurs across this protein family. Surprisingly, the most variable region of the "EC869-like" CdiA-CT toxins lays within the β-hairpin sequences (β4 & β5), suggesting that the dominant diversification within this family is located within the toxin/immunity interface, and the CdiA-CT β-hairpin region may be important for CdiA-CT/CdiI binding specificity (14) (Figure 3.1).

Through structural homology to type IIS restriction endonucleases, we identified the $CdiA\text{-}CT^{EC869}$ active site residues consisting of Glu177, Asp198, Ser209, and Lys211 (6). These residues are highly conserved among EC869 family CdiA-CT toxins (Figure 3.1). The EC869 β-augmentation interaction results in $CdiI^{EC869}$ binding a $CdiA\text{-}CT^{EC869}$ "exosite," leaving the predicted toxin active site solvent exposed. In the EC869 complex structure, active site residues (Glu177 and Asp198) and three water molecules coordinate a Zn^{2+} ion, suggesting that the $CdiA\text{-}CT^{EC869}$ toxin could be a Zn^{2+}-dependent endonuclease. Remarkably, $CdiA\text{-}CT^{EC869}$ contains robust *in vitro* DNase activity in the presence of Zn^{2+}, completely degrading plasmid DNA substrates, while the toxin only contains "nickase" activity in the presence of Mg^{2+}. Mutating either Glu177 or Asp198 to alanine abolished the *in vitro* DNase activity of $CdiA\text{-}CT^{EC869}$, confirming the importance of these residues in activity.

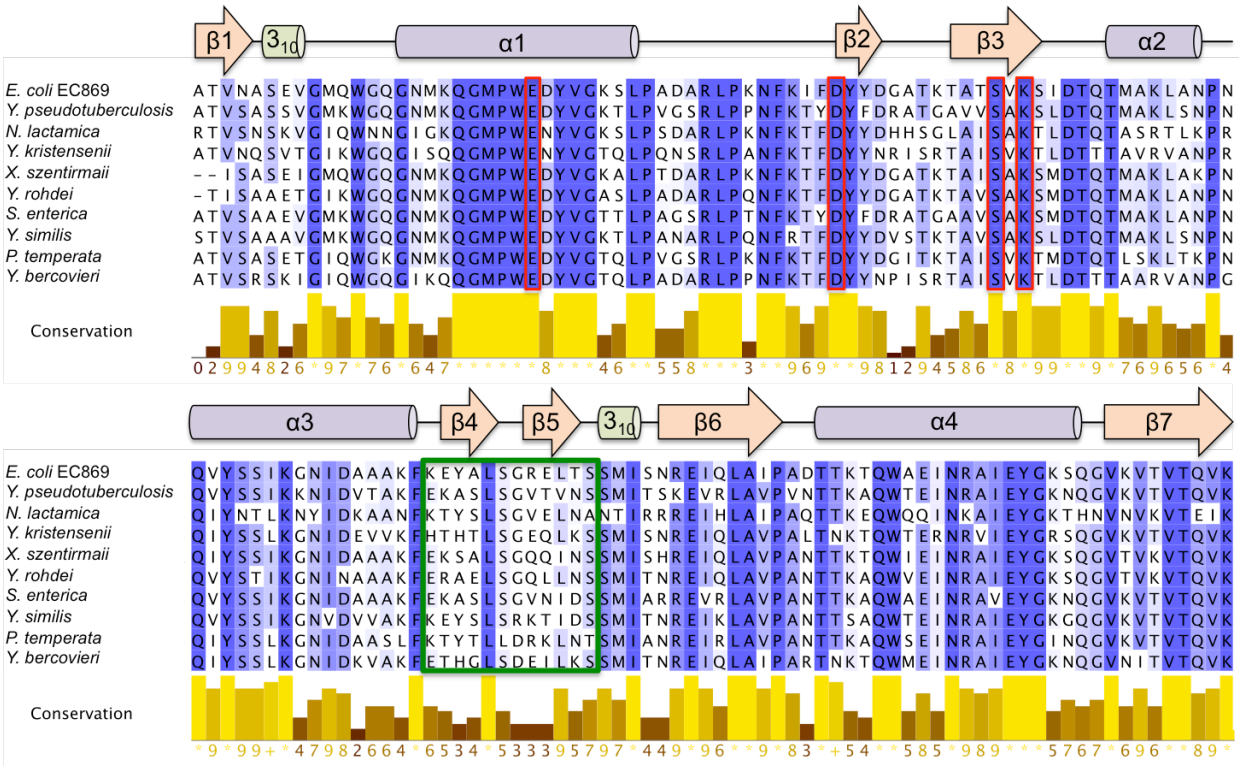

Figure 3.1. Protein sequence alignment of the CdiA-CT^{EC869} catalytic domain and homologs. The alignment was prepared using Jalview, with progressively darker shades of purple indicating greater residue conservation. Predicted toxin active site residues are outlined in red boxes, and the β-hairpin that mediates interactions with the CdiI^{EC869} immunity protein is outlined in a green box. The secondary structure elements shown are from the CdiA-CT^{EC869}/CdiI^{EC869} complex structure (PDB ID 4G6U) .

To further our understanding of the "EC869–like" family of toxin and immunity pairs, we have undertaken a comprehensive structural and biochemical characterization of protein complexes from this family. First, we establish that the CdiA-CT β-hairpin is required form a stable CdiA-CT^{EC869}/CdiI^{EC869} complex. Second, we determined the crystal structures of immunity protein from *Yersinia kristensenii* (YK), showing the availability of the CdiA-CT β-hairpin binding pocket in the absence of toxin, and the CdiA-CT/CdiI complex from *Yersinia pseudotuberculosis* (YP413) demonstrating the conservation of the β-augmentation interaction. High structural homology between the EC869 and YP413 toxin/immunity complexes suggests structural conservation across the "EC869-like" protein family. Third, we showed that despite structural homology, homologs of EC869 immunity protein cannot protect against EC869 toxin

activity *in vivo* and that binding affinities between heterologous "EC869–like" family CdiA-CT/CdiI pairs are drastically reduced compared to the cognate pair. Interestingly, swapping the CdiA-CT^{EC869} β-hairpin for that of heterologous EC869 family CdiA-CTs, also has a significant effect on CdiI binding. Finally, we solved the crystal structure of CdiI^{EC869} in complex with a macrocyclic peptide that mimics the CdiA-CT^{EC869} β-hairpin, where this may be the starting point for inhibitor design to interrupt CdiA-CT/CdiI complex formation.

Materials and Methods

Plasmid Constructs

All expression constructs with the exception of plasmid whereby the β –hairpin was deleted from toxin of the EC869 complex (EC869 CdiA-CT$^{\Delta\beta4\beta5}$/CdiI^{EC869}) were designed and constructed in the laboratory of Christopher Hayes (UC Santa Barbara). Construction of the EC869 CdiA-CT$^{\Delta\beta4\beta5}$/CdiI^{EC869} plasmid was carried out by manipulation of the pET21d CdiA-CT^{EC869}/CdiI^{EC869} expression plasmid (6). Primers were designed to amplify coding regions flanking the toxin β-hairpin (Table 3.1), whereby the first flanking region had restriction sites *Nco*I and *BamH*I engineered at the 5' and 3'ends, respectively, and the second flanking region had restriction sites *BamH*I and *Xho*I sites at the 5' and 3' ends, respectively. Following PCR the products were cut with *BamH*I and ligated together using T4 DNA ligase to form a combined insert. The combined insert was then amplified using PCR and inserted into a pET21d expression vector previously digested with *Nco*I and *Xho*I. The correct sequence was verified by DNA sequencing (Laguna).

Table 3.1. Primers for deleting the CdiA-CT^{EC869} β-hairpin

Primer	Sequence
β-deletion For1	GCCCAATGGGCACAAACCAGTCTCTGACCTTCGAT
β-deletion Rev1	GCGGATCCGCTTTTAAACTTAGCCGCAGCATCGATG
β-deletion For2	GCGGATCCGGCACTTCATCAATGATCTCTAACAGGG
β-deletion Rev2	GCCTCGAGACTAGTACCTTTGCAGCGACTCAAG

Protein expression and purification for crystallography

Y. pseudotuberculosis (YP413) CdiA-CT/CdiI complex, *Y. kristensenii* (YK) and *E. coli* (EC869) CdiI, were overexpressed in pET21d plasmids containing the appropriate *cdiA-CT/cdiI* genes using *E. coli* BL21 (DE3) Gold cells (Strategene). Cells were grown aerobically at 37 °C in LB medium containing 50 μg/mL ampicillin. YP413 CdiA-CT/CdiI expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactosidase at an OD600 ≈ 0.8 and grown for a further 3-4 h before harvesting. Cells were collected by centrifugation at 5,500 × g for 25 min and then washed with resuspension buffer (20 mM sodium phosphate, pH 7.0, 150 mM NaCl). Cells were disrupted by sonication on ice in resuspension buffer containing 10 mg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride. Cell debris was removed by centrifugation at 18,000 × g for 30 min followed by filtration through a 1.0-μm filter. Clarified lysates were either loaded onto a Ni^{2+}-charged HiTrap column (5 mL; GE Healthcare) or nickel-nitrilotriacetic acid (Ni^{2+}-NTA) agarose resin (MCLAB) and washed with resuspension buffer supplemented with 10 mM imidazole. Proteins were were eluted with a linear gradient of imidazole (10-500 mM) in resuspension buffer. Fractions were collected, combined, and concentrated to a volume of ~500 μL using a 10-kDa centrifugal concentrator (Centricon; Millipore). Proteins were further purified by gel filtration on a Superdex 200 column, for the YP413 complex, or Superdex 75 column, for the individual immunity proteins (GE Healthcare), equilibrated with 20 mM sodium phosphate, pH 7.0 and 150 mM NaCl using an AKTA FPLC. Purification of CdiI^{YK} and CdiI^{EC869}

followed the same protocol, except all buffers contained 20 mM Tris, pH 7.4, instead of sodium phosphate. YP413 CdiA-CT/CdiI, CdiI^{YP}, and CdiI^{EC869} were concentrated to 10, 12.5, and 7.5 mg/mL, respectively, for crystallization trials.

General Structure Determination Pipeline

All protein crystals were grown by hanging drop vapor diffusion, with drops containing a 1:1 ratio of protein:reservoir. Crystals were mounted and collected under cryoconditions with the addition of 40% (vol/vol) glycerol as cryoprotectant to the reservoir condition. Datasets were collected at 70K at a wavelength of 1.0 Å and images were indexed, integrated, and reduced using either iMOSFLM (CdiA-CT^{YP413}/CdiI^{YP413}) (15) or the HKL2000 suite (CdiI^{YK} and CdiI$^{CT\text{-}MAC}$) (16). The initial phases were determined by molecular replacement by autoMR in PHENIX utilizing the CdiA-CT^{EC869}/CdiI^{EC869} complex structure (CdiA-CT^{YP413}/CdiI^{YP413}) or CdiI^{EC869} alone (CdiI^{YK} and CdiI$^{CT\text{-}MAC}$) as a search model (PDB ID 4G6U). Initial model building was performed by Autobuild in PHENIX. The final models were built through iterative manual building in Coot and refined with phenix.refine. Data collection and refinement statistics are presented in Table 3.2. All molecular graphics were prepared with PyMOL (17).

Crystallization and Refinement

YP413 CdiA-CT/CdiI complex crystals were grown over 1 week with a reservoir containing 50 mM HEPES, pH 7.0, 20% (wt/vol) PEG 3350, and 1% (wt/vol) tryptone and a 10 mg/mL protein solution. The protein complex crystallized in space group C2 with unit cell dimensions 65.51 Å × 65.51 Å × 71.49 Å and one complex per asymmetric unit. The final model contains CdiA-CT^{YP413} and CdiI^{YP413} residues 174-297 and 3-176, respectively, and 148 water molecules resulting in an R_{work}/R_{free} (%) of 20.5/25.6, with 95.25% and 4.75% of residues in the favorable allowed, and allowed regions, respectively. CdiA-CT^{YP413} residues Lys182, Lys220, Lys240, and Lys297, were modeled as alanines due to lack of observable side-chain density.

Similarly, CdiI^{YP413} residues Asp3, Lys108, Lys118, Lys148, and Lys176 were modeled as alanines.

CdiI^{YK} crystals were grown over 1 week with a reservoir containing 0.2 M ammonium fluoride and 20% (wt/vol) PEG 3350, and a 12.5 mg/mL protein solution. The crystal space group was $P3_1$ with unit cell dimensions 54.448 Å × 54.448 Å × 54.472 Å and one molecule per asymmetric unit. The final model contains CdiI^{YK} residues 1-165, and 130 water molecules, resulting in an R_{work}/R_{free} (%) of 18.1/22.1, with 98.77% and 1.23% of residues in the favorable allowed, and allowed regions, respectively. Residues Lys4, Glu67, Lys96, Lys126, and Lys136 were modeled as alanines due to lack of observable side-chain density. Additionally, a mutation in our expression plasmid resulted in an A84T mutation in our purified protein.

A macrocyclic β-peptide mimic was designed and synthesized in the laboratory of James Nowick (UC Irvine). To obtain co-crystals (CdiI$^{CT\text{-}MAC}$), 200 µl of 7.5 mg/mL CdiI^{EC869} was added to 2 mg of lyophilized peptide, creating a mixture with a ratio of ≈ 10:1 peptide:protein. CdiI$^{CT\text{-}MAC}$ crystals were grown over the period of 2 days in 0.2 M sodium acetate, pH 5.6, 0.1 M bis-tris propane, pH 6.9, and 20% (wt/vol) PEG 3350, using the protein/peptide mixture described. Initial crystals were of poor quality, resulting in highly mosaic diffraction data. Microseeding successfully improved crystal quality (18). Briefly, crystals were harvested into 80 µL of crystallization solution and a seed stock was generated using a seed bead (Hampton). Following extensive optimization, suitable diffraction quality crystals were generated using a hanging drop containing 1 µL of seed stock, and 1 µL of the protein/peptide mixture following a 3-fold dilution. CdiI$^{CT\text{-}MAC}$ crystallized in space group $P2_1$, with unit cell dimensions 34.776 Å × 128.166 Å × 44.953 Å, and contained two CdiI$^{CT\text{-}MAC}$ complexes per asymmetric unit. The final CdiI$^{CT\text{-}MAC}$ model contains CdiI^{EC869} residues 1-168, two macrocyclic peptides, and 132 waters, resulting in an R_{work}/R_{free} (%) of 18.4/23.1, with 97.13% and 2.87% of residues in the favorable allowed, and allowed regions, respectively. CdiI^{EC869} residues Lys5, Gln43, Glu78, Lys85,

Glu93 (chain D only), Asp117, and Glu139 were modeled as alanines due to lack of observable side-chain density.

Table 3.2. X-ray diffraction data and atomic refinement

	CdiA-CT^{YP413}/CdiI^{YP413}	**CdiI^{YK}**	**CdiI$^{CT\text{-}MAC}$**
Space Group	C2	$P3_1$	$P2_1$
Unit cell dimensions (Å)	65.51 × 65.51 × 71.49	54.448 × 54.448 × 54.472	34.776 × 128.166 × 44.953
pH of crystallization condition	7.0	7.4	6.0
Protein concentration (mg/mL)	10	12.5	7.5
Data set			
Wavelength (Å)	1.0	1.0	1.0
Resolution range	46.49-2.09	50.0-1.8	50.0-2.0
Unique reflections (total)	18152	16618	23561
Completeness (%)*	99.4 (98.3)	99.7 (100)	96.5 (92.1)
Redundancy *	3.5 (3.5)	5.4 (5.5)	3.1 (3.1)
R_{merge}*,†	0.059 (0.460)	0.161 (0.607)	0.076 (0.485)
I/σ*	10.0 (1.6)	9.48 (3.1)	11.1 (3.1)
NCS copies	1	1	2
Other ions	2 Cl^-	-	2 Cl^-
Model refinement			
Resolution range (Å)	46.49-2.09	23.58-1.80	34.81-2.00
No. of reflections (working/free)	18149/1851	16580 (1673)	23526 (2004)
No. of protein atoms	2308	1283	2596
No. of water molecules	148	130	132
No. of CT-MAC atoms	-	-	222
Missing residues	1-173, 298 (CdiA) 1-2, (CdiI)	-	1, 169
R_{work}/R_{free}‡, %	20.5/25.6	18.1/22.1	18.4/23.1
R.m.s deviations			
Bond lengths (Å)	0.003	0.007	0.009
Bond angles (degrees)	0.694	0.998	1.222
Ramachandran Plot			
Most favorable region (%)	95.25	98.77	97.13
Additional allowed region (%)	4.75	1.23	2.87
Disallowed region	0	0	0
PDB ID Code			

*Statistics for the highest resolution shell are given in (brackets)

†$R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma I$

‡$R_{work}=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$ R_{free} was computed identically except where all reflections belong to a test set of 10% randomly selected data.

Protein Purification of Isolated CdiA-CT and CdiI-His_6 proteins

The individual His_6-tagged CdiI proteins were overexpressed in pET21d and purified as described above for $CdiI^{EC869}$. CdiA-CT proteins were isolated from co-expressed His_6-tagged CdiI proteins by two methods, depending on whether the two proteins co-eluted following Ni^{2+}-affinity chromatography. CdiA-CT/CdiI-His_6 complexes that co-eluted were denatured overnight in 6 M urea and then subjected to Ni^{2+}-affinity chromatography in buffers containing 6 M urea. Denatured CdiA-CTs collected in the flow through were refolded by dialysis (6). However, in cases when co-expression of a CdiA-CT protein with a CdiI-His_6 protein did not result in complex formation, CdiA-CTs could be found in the flow-through and wash (see figures 3.2 and 3.4). The fractions containing the CdiA-CT were combined and dialyzed into low salt buffer (20 mM Tris, pH 8, 10 mM NaCl), and then loaded onto an anion-exchange column (HiTrap Q HP, 5 mL) and eluted with a salt gradient, yielding 95% pure CdiA-CT protein. This was then passed over an S75 gel filtration column [equilibrated with 20 mM Tris (pH 7.4) and 150 mM NaCl] to complete the purification.

Circular Dichroism

The secondary structure of purified CdiA-CTs was analyzed by CD spectroscopy on a Jasco J-720 spectropolarimeter using 0.1-cm pathlength cells and 0.1-mg/mL protein samples. The purified proteins were buffer exchanged by a desalting column (GE Healthcare) into 20 mM Tris, pH 7.4, prior to the measurements. The bandwidth was set to 2 nm; the response time was equal to 4 s; scanning speed was 20 nm/min, and data pitch was 0.5 nm, and a total of 3 consecutive scans were accumulated for analysis.

CdiA-CT/CdiI binding affinities

The toxin and immunity protein binding affinities were determined by biolayer interferometry (BLitz; ForteBio Inc.), as previously described (6). All reactions were performed at 25°C in 20 mM Tris, pH 7.4, 150 mM NaCl. Briefly, CdiI-His_6 immunity proteins were immobilized on Ni^{2+}-NTA biosensors, and exposed to various concentrations of cognate or heterologous CdiA-CT toxins. CdiA-CT concentrations used are proportional to the complex binding affinities and ranged from 0.5-300 μM, depending on the particular affinity being measured. A reference was subtracted from all binding curves before curve fitting. Curve fitting and data processing were performed using BLitz Pro software (ForteBio Inc.).

Results

EC869 CdiA-CT β-hairpin is required for interaction with CdiI

During CdiA-CT^{EC869}/$CdiI^{EC869}$ complex formation, the CdiA-CT^{EC869} β-hairpin (residues 242-252) inserts into the $CdiI^{EC869}$ binding pocket, forming a six-stranded antiparallel β-sheet that spans both proteins (Figure 3.2a). This β-augmentation results in an extensive network of 18 hydrogen-bond (H-bond) and ion pair interactions between the eleven β-hairpin toxin residues and immunity protein, which includes both backbone and side-chain interactions. Additionally, there are a number of interactions between the CdiA-CT^{EC869} loop regions and CdiI. The large network of interactions facilitated via the toxin β-hairpin implies that it is required for complex formation, and thus we sort to test this hypothesis. First, a mutant was constructed where the encoded toxin β-hairpin was replaced by a four-residue (SGSG) loop within CdiA-CT^{EC869}, which we will refer to as CdiA-$CT^{\Delta\beta4\beta5}$. Then, $CdiI^{EC869}$-His_6 and CdiA-$CT^{\Delta\beta4\beta5}$ lacking a tag were co-expressed and purified by Ni^{2+}-affinity chromatography under non-denaturing conditions, and compared to the purification profile of $CdiI^{EC869}$-His_6 and wildtype CdiA-CT^{EC869}. The CdiA-$CT^{\Delta\beta4\beta5}$ was observed in the flow-through and did not co-purify with $CdiI^{EC869}$, while $CdiI^{EC869}$-

His_6 and wildtype CdiA-CT^{EC869} co-purify, suggesting that indeed the toxin β-hairpin is required for high-affinity complex formation (Figure 3.2b). After further purifying CdiA-$CT^{\Delta\beta4\beta5}$ by anion-exchange chromatography, we determined the binding affinity of CdiA-$CT^{\Delta\beta4\beta5}$ to CdiI-His_6 utilizing biolayer interferometry (BLitz; ForteBio Inc); however no detectable binding was observed between CdiA-$CT^{\Delta\beta4\beta5}$ and CdiI-His_6. Circular dichroism (CD) spectroscopy was used to ensure that CdiA-$CT^{\Delta\beta4\beta5}$ was folded. The resulting spectrum of CdiA-$CT^{\Delta\beta4\beta5}$ superimposes with that of wild-type CdiA-CT^{EC869}, suggesting that the mutant was folded with similar secondary structure elements as the wild-type toxin (Figure 3.2c). Taken together, these results suggest that the CdiA-CT^{EC869} β-hairpin is essential for complex formation with $CdiI^{EC869}$.

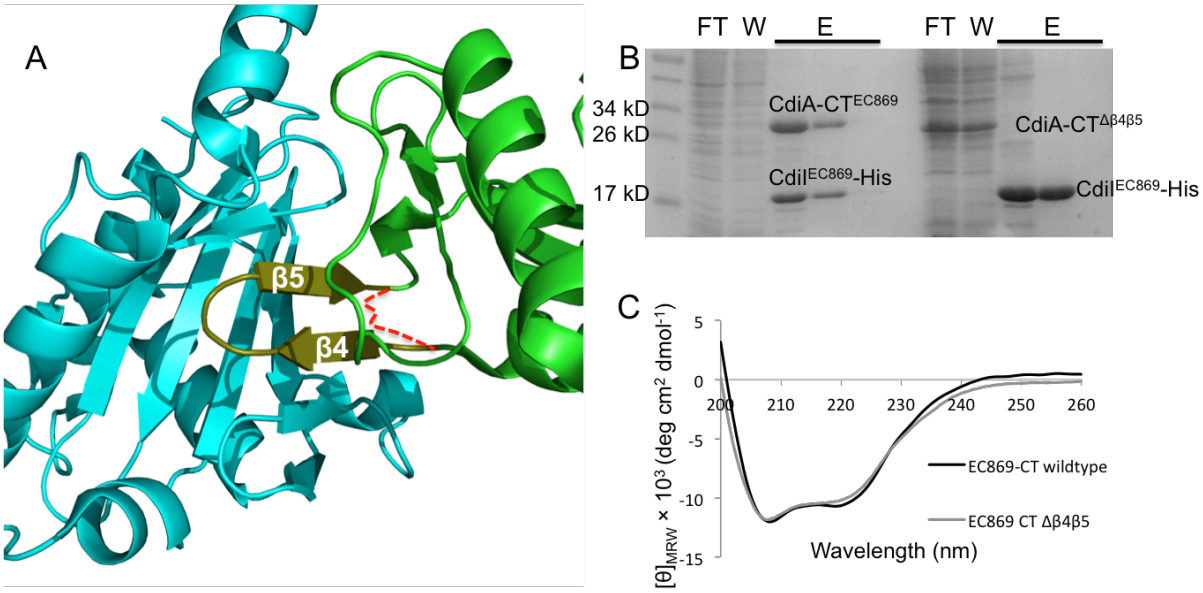

Figure 3.2. Deletion of the CdiA-CT^{EC869} β-hairpin disrupts complex formation. (A) Ribbon representation of the EC869 complex interface, with CdiA-CT^{EC869}, $CdiI^{EC869}$ and the CdiA-CT^{EC869}-β-hairpin colored in green, cyan, and olive, respectively. A representation of the short loop created in CdiA-$CT^{\Delta\beta4\beta5}$ is shown with a dashed red line. (B) $CdiI^{EC869}$-His_6 was co-expressed with wildtype CdiA-CT^{EC869} or CdiA-$CT^{\Delta\beta4\beta5}$, followed by purification via Ni^{2+}-affinity chromatography and then analysis by SDS/PAGE gel. Relevant molecular weight standards are labeled. FT, W, and E indicate column flow-through, wash, and elution, respectively. (C) CD spectrum of purified CdiA-$CT^{\Delta\beta4\beta5}$ is similar to that of wildtype toxin.

Structure of the CdiA-CT^{YP413}/CdiI^{YP413} complex reveals conservation of the β-augmentation complex interface

To investigate structural variance of the toxin/immunity complex within the "EC869-like" family, we solved the structure of the CdiA-CT/CdiI complex from *Y. pseudotuberculosis* (YP413) to 2.1 Å resolution by MR, using the CdiA-CT^{EC869}/CdiI^{EC869} complex structure as a search model (PDB ID 4G6U) (6). CdiA-CT ^{YP413}/CdiI^{YP413} crystallized in space group C2 with one complex per asymmetric unit. The final model contains CdiA-CT^{YP413} residues 174-297 (numbered from Val1 of the VENN motif), and CdiI^{YP413} residues 3-176 (Figure 3.3a). Similar to our previous CdiA-CT/CdiI structures (6, 9), the structural model only contains the C-terminal catalytic domain of CdiA-CT^{YP413}, as the CdiA-CT N-terminal regions are thought to be unstructured and thus susceptible to degradation. The final model CdiA-CT ^{YP413}/CdiI^{YP413} included 148 water molecules resulting in an R_{work}/R_{free} (%) of 20.5/25.6 (Table 3.2).

On a primary sequence level, CdiA-CT^{YP413} and CdiI^{YP413} are 70.4 and 49.1% identical to their EC869 counterparts, respectively, and not surprisingly the toxins and immunities share significant structural homology (Figures 3.3b, c). CdiA-CT^{YP413} and CdiI^{YP413} superimpose over CdiA-CT^{EC869} and CdiI^{EC869}, respectively, with rmsds of 0.84 Å (over 101 of 123 α-carbons) and 1.01 Å (over 133 of 173 α-carbons). CdiA-CT^{YP413}/CdiI^{YP413} complex formation is also mediated by β-augmentation, suggesting conservation of this interface across the "EC869-like" toxin/immunity family. The YP413 β-augmentation mediates a network of only 12 H-bond and ion pair interactions, less than the 18 interactions observed within the EC869 complex. Furthermore, within the YP413 β-hairpin only two residue side-chains, Glu242 and Lys243, interact with CdiI^{YP413}, while for the EC869 β-hairpin six residue side-chains interact with the immunity protein. In the YP413 structure, one additional salt bridge is formed just outside of the β-hairpin region between CdiA-CT^{YP413} Lys195 and CdiI^{YP413} Glu137. Of note, CdiI^{YP413} has a 10-residue insertion that results in an extended L1 loop (Figure 3.3b). Further, the L1 loop is stabilized by a salt bridge between CdiI^{YP413} Arg69 and CdiA-CT^{YP413} Asp201. Lastly, CdiA-

CT^{YP413} shares the same network of predicted active site residues as CdiA-CT^{EC869} (Figure 3.3c). Within the EC869 toxin active site, an ordered Zn^{2+} ion was observed, however no metal ion is observed in the structure of the YP413 toxin active site. This high structural homology suggests that CdiA-CT^{YP413} likely functions as a DNase as seen for CdiA-CT^{EC869}, but the catalytic metal of CdiA-CT^{YP413} remains unclear from the structure alone.

Despite the differing networks of interactions observed in the CdiA-CT^{YP413}/$CdiI^{YP413}$ complex compared to CdiA-CT^{EC869}/$CdiI^{EC869}$, the binding affinities are similar. To determine the binding affinity of the toxin and immunity protein, the complex dissociation constant (K_d) was determined by biolayer interferometry, where the K_d for the YP413 complex (K_d 16.3 ± 1 nM) is similar to that of the EC869 complex (K_d, 17.8 ± 7 nM). Of note, the CdiA-CT^{EC869} toxin has a weak binding affinity for $CdiI^{YP413}$, with a K_d of 13 ± 2 μM. A similar binding affinity is also determined between CdiA-CT^{YP413} and $CdiI^{EC869}$, with a K_d of 13 ± 6 μM. Regardless of these significantly reduced micromolar binding affinities, these weak interactions show that toxin and immunity interactions are not reserved only to cognate protein pairs.

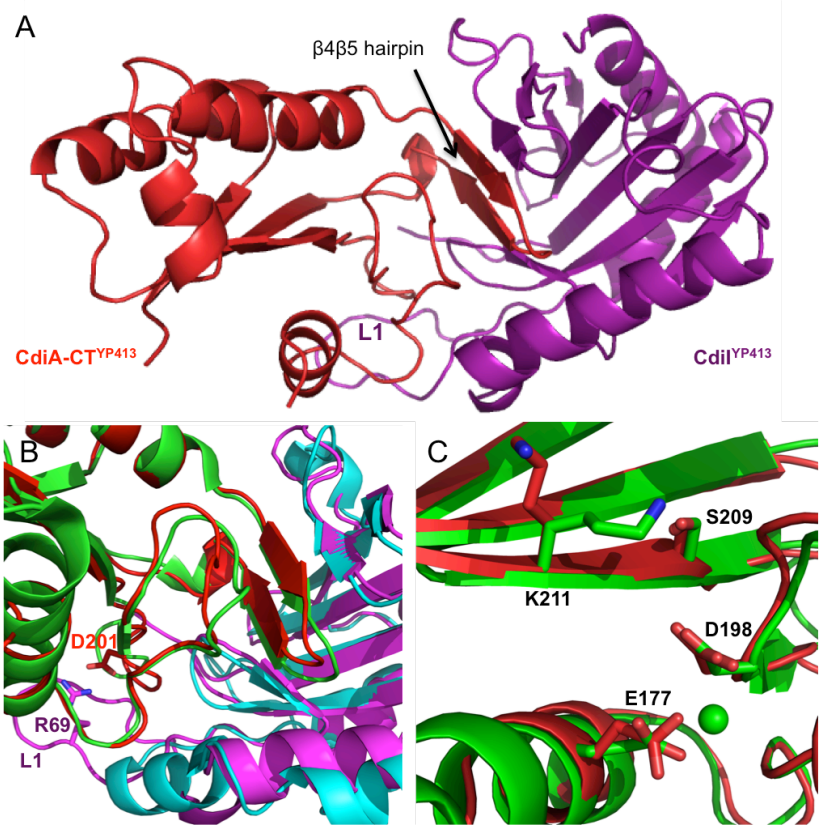

Figure 3.3. The structure of the CdiA-CT^{YP413}/CdiI^{YP413} complex. (A) Ribbon representation of the CdiA-CT^{YP413}/CdiI^{YP413} complex with CdiA-CT^{YP413} and CdiI^{YP413} colored in red and purple, respectively. Location of the CdiA-CT^{YP413} β4β5 hairpin is indicated. (B) Structural superimposition of the β4β5-hairpin binding region of the YP413 and EC869 CdiA-CT/CdiI complexes. CdiA-CT^{EC869} and CdiI^{EC869} are colored green and cyan, respectively. YP413 residues that form a salt bridge via loop L1 are depicted as sticks. (C) Predicted active site residues of CdiA-CT^{EC869} (green) and CdiA-CT^{YP413} (red). Residue labels correspond to both toxins. Zn^{2+}, which is part of the CdiA-CT^{EC869} structure, is shown in green. Variable loop L1 of CdiI^{YP413} is labeled in panels A and B.

The β-hairpin of the EC869 family toxins contributes to immunity binding specificity

We have established that the toxin β-hairpin is important for CdiA-CT^{EC869}/CdiI^{EC869} complex formation, and the β-augmentation interface appears to a conserved across this family of CDI complexes. We next tested the effect of specific β–hairpin sequences upon toxin and immunity complex formation. To carry out this study, we constructed CdiA-CT^{EC869}/CdiI^{EC869}-His_6 expression plasmids, such that the EC869 toxin β-hairpin sequence was interchanged with those from *N. lactamica* (CdiA-CT$^{EC869\text{-}NL}$) and *Y. kristensenii* (CdiA-CT$^{EC869\text{-}YK}$). Of note,

chimeric CdiA-CTs also contain the D198A mutation, which has previously been shown to inactivate the toxin (6). These two toxin β-hairpin regions show different degrees of similarity to that of EC869 (Figure 3.4a). In the structure of the EC869 complex, H-bond or ion pair interactions are observed via the β-hairpin residue side-chains of K242, E243, Y244, S247, R249, and E250. Within the NL β-hairpin compared to that of EC869, there are only two substitutions within these residues, E243T and R249V, resulting in loss of two charged side-chains, while there are four substitutions within the YK β-hairpin, K242H, Y244H, R249E, and E250Q. Thus, we propose that CdiA-CT$^{EC869\text{-}YK}$ will have more diverse consequences upon complex formation with CdiI^{EC869} compared to CdiA-CT$^{EC869\text{-}NL}$.

To investigate interactions of the chimeric toxin with immunity, co-expression and purification studies followed by determination of dissociation constants of purified proteins were carried out. CdiI^{EC869}-His$_6$ was co-expressed with either CdiA-CT$^{EC869\text{-}NL}$ or CdiA-CT$^{EC869\text{-}YK}$, followed by purification utilizing Ni^{2+}-affinity chromatography. The toxin with the less divergent β-hairpin sequence from *N. lactamica*, CdiA-CT$^{EC869\text{-}NL}$, co-purified with CdiI^{EC869}-His$_6$, while the toxin with the more diverse β-hairpin sequence from *Y. kristensenii,* CdiA-CT$^{EC869\text{-}YK}$, was observed in the flowthrough and CdiI^{EC869}-His$_6$ eluted alone (Figure 3.4b). These results suggest that CdiA-CT$^{EC869\text{-}NL}$ binds with relatively high affinity to CdiI^{EC869}, while CdiA-CT$^{EC869\text{-}YK}$ does not bind to CdiI^{EC869} with sufficient affinity to produce a stable complex for purification.

To quantify binding affinities of CdiI^{EC869} to EC869 chimeric toxins, the proteins were individually purified and bio-layer interferometry was utilized to determine dissociation constants. CdiA-CT$^{EC869\text{-}NL}$ was purified by denaturing the CdiA-CT$^{EC869\text{-}NL}$/CdiI^{EC869} complex in 6 M Urea, followed by separation of the two proteins by Ni^{2+}-affinity chromatography. The flowthrough containing CdiA-CT$^{EC869\text{-}NL}$ and elutant containing CdiI^{EC869}-His$_6$ were refolded separately by dialysis into native buffer. CdiA$^{EC869\text{-}YK}$ and CdiI^{EC869} were purified under native conditions by Ni^{2+}-affinity chromatography. CdiA$^{EC869\text{-}YK}$ was observed in the flowthrough while

pure CdiI^{EC869} was observed in the elutant. CdiA$^{EC869\text{-}YK}$ was further purified to homogeneity by anion exchange chromatography. We have previously demonstrated that refolding of CdiA-CT toxins results in correctly folded proteins as they can form tight complexes with purified cognate CdiI proteins (6). To ensure that the chimeric toxins were correctly folded CD spectroscopy was utilized. The CD spectra of the chimeric toxins are similar to that of CdiA-CT^{EC869}, suggesting that they were properly folded (Figure 3.4c). The dissociation constants for CdiI^{EC869} to CdiA-CT$^{EC869\text{-}NL}$ and CdiA$^{EC869\text{-}YK}$ were 175 nM ± 100 and 45.6 μM ± 36, respectively (Table 3.3). Taken together, our data strongly suggests that specific toxin β-hairpin sequences effect stable complex formation in the "EC869-like" family of toxin/immunity complexes.

Finally, we tested whether the chimeric toxins, CdiA-CT$^{EC869\text{-}NL}$ and CdiA-CT$^{EC869\text{-}YK}$, would promote stable complex formation to their cognate immunity proteins. To investigate this, we purified CdiI from *N. lactamica* (CdiI^{NL}) and *Y. kristensenii* (CdiI^{YK}), and assessed binding by bio-layer interferometry (Table 3.3). CdiI^{NL} and CdiI^{YK} share varying levels of sequence identity to CdiI^{EC869} (50.6 and 68.5%, respectively). Despite this homology, none of the CdiI^{EC869} residues that contain side chains that interact with CdiA-CT^{EC869} are completely conserved among the immunity proteins (Figure 3.5). Two of the eight H-bond or ion-pair interacting CdiI^{EC869} residues, Lys128 and Glu130, are conserved in two of the three homologs investigated in this study. The backbone of CdiI^{EC869} makes an additional nine H-bonds with CdiA-CT^{EC869}, and the heterologous immunity proteins could retain these H-bonds. Interestingly, CdiA-CT$^{EC869\text{-}NL}$ and CdiA-CT$^{EC869\text{-}YK}$ showed the ability to interact with both CdiI^{NL} and CdiI^{YK}, however with varying levels of affinity. The dissociation constants are very similar to what was observed for CdiI^{EC869} to CdiA-CT$^{EC869\text{-}NL}$ and CdiA-CT$^{EC869\text{-}YK}$. This suggests that the CdiA-CT β-hairpin sequences alone are not fully responsible for the tight interactions observed in cognate complexes. Finally, while no observed binding was as tight as the CdiA-CT^{EC869}/CdiI^{EC869} cognate pair, we did observe interactions between CdiA-CT^{EC869} and both heterologous

immunity proteins tested. This supports the hypothesis that the β-hairpin binding interface allows for weak interactions to occur between heterologous toxin/immunity pairs, however the overall sequence diversity of the proteins contributes strongly for high binding affinities between cognate pairs.

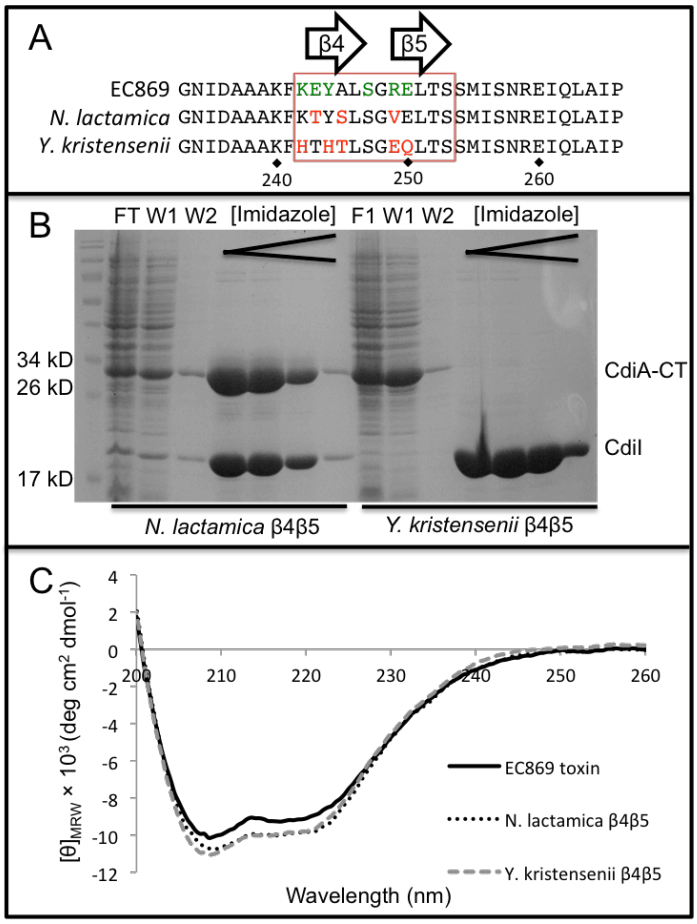

Figure 3.4. CdiA-CT β-hairpin sequence contributes to complex binding affinity. (A) Protein sequence alignment of the β-hairpin region (red box) of CdiA-CT from EC869, *N. lactamica* and *Y. kristensenii* . Residues colored green indicate CdiA-CT^{EC869} residues that interact with CdiI^{EC869}. Residues colored red indicate a difference in *N. lactamica* and *Y. kristensenii* compared to CdiA-CT^{EC869}. (B) CdiI^{EC869}-His$_6$ was co-expressed with CdiA-CT^{EC869} containing the β-hairpins from either *N. lactamica* or *Y. kristensenii.* Then purified by Ni^{2+} affinity chromatography and analyzed by SDS/PAGE. The molecular weight standards are indicated in kD. FT and W1 and W2 indicate column flow-through, first and second wash, respectively, followed by elution with imidazole gradient (C) CD spectra of purified CdiA-CT$^{EC869\text{-}NL}$ and CdiA-CT$^{EC869\text{-}YK}$ show similar secondary structure elements as compared to wildtype toxin.

Table 3.3. Dissociation constants of CdiA-CT and CdiI interactions

	EC869 toxin with *N. lactamica* β-hairpin	EC869 toxin with *Y. kristensenii* β-hairpin	EC869 toxin
Y. kristensenii CdiI	5.8 μM ± 3	23.5 μM ± 14	4.96 μM ± 0.4
N. lactamica CdiI	171 nM ± 85	3.2 μM ± 2	82.5 μM ± 18
EC869 CdiI	175 nM ± 100	45.6 μM ± 36	17.8 nM ± 7

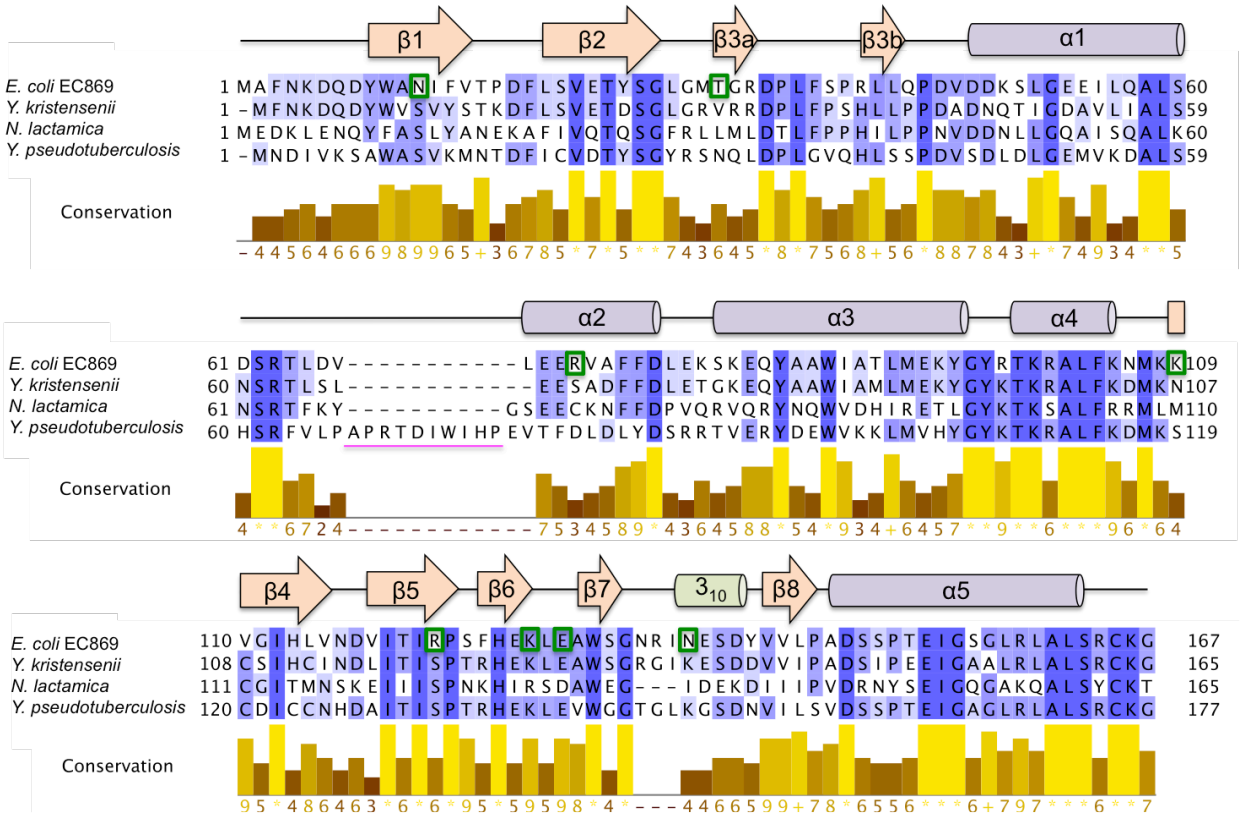

Figure 3.5. Alignment of CdiI^{EC869} immunity homologs. The CdiI^{EC869} sequence is aligned with related immunity proteins from the indicated bacteria. The alignment was prepared using Jalview, with progressively darker shades of purple indicating greater residue conservation. Secodary structure elements shown correspond to CdiI^{EC869}. The CdiI^{EC869} residues that contain sidechains that form H-bond or ion-pair interactions with CdiA-CT^{EC869} are marked with green boxes. The location of CdiI^{YP413} loop L1 is indicated with a magenta bar.

Only cognate CdiI^{EC869} can protect cells from chimerically delivered diA-CT^{EC869} *in vivo* (Experiments carried out by the Hayes lab, UCSB)

Our collaborators, Chris Hayes' laboratory at UCSB, tested whether EC869 family immunity proteins can confer immunity against CdiA-CT^{EC869} that is delivered to target cells during competition. Using a chimeric CDI system in which the *cdiA-CT*^{EC869}/*cdiI*^{EC869} coding sequences are fused to the *E. coli* EC93 *cdiA* gene at the VENN-encoding sequence (6), they delivered CdiA-CT^{EC869} to target cells expressing EC869 family CdiI proteins. None of the EC869 family CdiI proteins were able to protect against the delivered CdiA-CT^{EC869} as these cells showed significant growth inhibition. Only cognate CdiI^{EC869} conferred immunity. These results strongly suggest that the weak interactions we observed in our affinity studies between CdiA-CT^{EC869} and heterologous CdiI proteins are not strong enough to neutralize the toxin *in vivo*.

CdiA-CT^{EC869} family toxins exhibit varying metal specificities (Experiments carried out by the Hayes lab, UCSB)

Our previous structure of CdiA-CT^{EC869}/CdiI^{EC869} revealed a Zn^{2+} ion in the CdiA-CT^{EC869} active site, and we determined the activity of that toxin to be a Zn^{2+}-dependent DNase capable of completely degrading the genomes of target cells. In contrast, CdiA-CT^{EC869} only has DNA nicking activity in the presence of Mg^{2+} *in vitro*. While we predict similar activities due to the structural homology across members of the EC869 family toxins, the metal dependence of each member may be distinct. To test this (our collaborators Chris Hayes' laboratory at UCSB carried out these experiments), CdiA-CT^{YP413} and CdiA-CT^{NL} were purified, and their ability to degrade plasmid DNA *in vitro* was tested in the presence of various metals. These two toxins share the exact same network of predicted active site residues as CdiA-CT^{EC869}. Surprisingly, CdiA-CT^{YP413} showed robust DNase activity, completely degrading plasmid DNA in the presence of Mg^{2+}, Mn^{2+}, and Co^{2+}; however, CdiA-CT^{YP413} was inactive in the presence of Zn^{2+}. In contrast,

CdiA-CT^{NL} showed activity in the presence of all four metals. The reasons behind the loose restraints on metal dependence for CdiA-CT^{YP413} and CdiA-CT^{NL} compared to CdiA-CT^{EC869} is unclear. The ability to utilize multiple catalytic metals could be beneficial for the toxins to cope with the varying cellular conditions of the target cells.

The structure of CdiI^{YK} is not perturbed in the absence of toxin

To gain further insight into EC869 family CdiA-CT/CdiI complex formation, the structure of the CdiI^{YK} immunity protein was solved in the absence of toxin. The structure of *Y. kristensenii* CdiI alone was solved to 2.1 Å resolution by MR, using the CdiI^{EC869} structure as a search model. CdiI^{YK} crystallized in space group $P3_1$ with one molecule per asymmetric unit. The final structural model contains CdiI^{YK} residues 1-165, and 130 water molecules, resulting in an R_{work}/R_{free} of 18.1/22.1 (Table 3.2).

CdiI^{YK} shares significant structural homology with CdiI^{EC869} and CdiI^{YP413}, superimposing with rmsds of 0.626 and 0.984 Å (over all CdiI ^{YK} α-carbons), respectively (Figure 3.5). This is not surprising given that the primary sequence of CdiI^{YK} is 68.9% and 51.5% identical to CdiI^{EC869} and CdiI^{YP413}, respectively. The extended L1 loop observed in CdiI^{YP413} is notably absent from CdiI^{YK}. Of note, the structure CdiI^{YK} in the absence of toxin retains the architecture of the toxin β-hairpin binding site. Furthermore, a structural search using the structure of CdiI^{EC869} revealed an additional homolog from *Neisseria meningitis* (NM) present in the protein data bank (PDB ID 2GKP). The CdiI^{NM} structure superimposes well with the three other CdiI structures with rmsds of less than 0.7 Å (Figure 3.6). These CdiI structures show that the tertiary structure is unperturbed in the absence of toxin and is not the result of interactions due to complex formation. Because of this, the β-hairpin binding pocket of these CdiI proteins could potentially function as a scaffold for structure-based inhibitor design to disrupt complex formation.

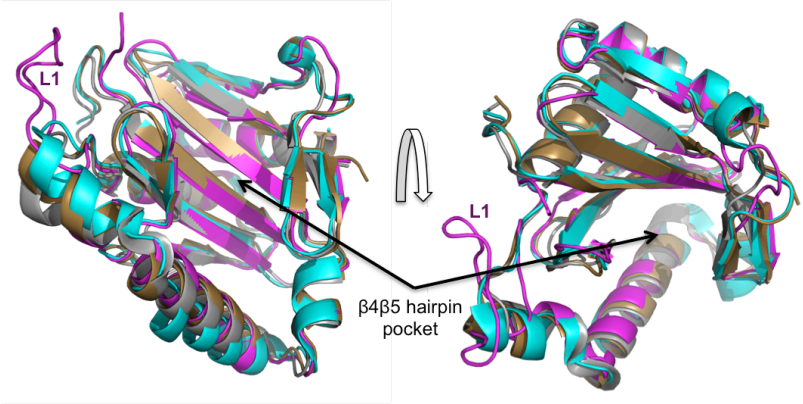

Figure 3.6. Structural superimposition of CdiI^{YP413}, CdiI^{EC869}, CdiI^{YK} and CdiI^{NM} (colored in magenta, cyan, gold, and gray respectively). We determined the structure of CdiI^{YK} in absence of bound toxin. The location of the conserved β-hairpin binding pocket is indicated. Extended loop L1 of CdiI^{YP413} is labeled.

Structure of CdiI^{EC869} in complex with a cyclic β-hairpin mimic

As the immunity proteins of the "EC869-like" family of proteins is structurally conserved in the absence of toxin, the CdiI^{EC869} β-hairpin binding pocket could be a starting-point for small molecules or peptides design to inhibit formation of the "EC869-like" toxin/immunity complexes. As a preliminary investigation, a small macrocyclic peptide (CT-MAC) that mimics the CdiA-CT^{EC869} β-hairpin was designed and synthesized by our collaborators, James Nowick and Mandy Zheng, at UCI. CT-MAC contains CdiA-CT^{EC869} residues 242-253, with K242 and S253 connected by a δ-linked ornithine turn unit (19, 20). Unfortunately, *in vitro* assays to test whether CT-MAC inhibits CdiA-CT^{EC869}/CdiI^{EC869} complex formation were inconclusive. We therefore attempted to solve the crystal structure of CdiI^{EC869} in complex with CT-MAC (CdiI^{CT-MAC}) to demonstrate that the macrocycle can stably bind to CdiI^{EC869}. For crystallization trials, a 10:1 ratio of macrocycle:protein ratio was used. We solved the structure of CdiI^{CT-MAC} to 2.0-Å resolution by MR, using the structure of CdiI^{EC869} alone as a search model (Figure 3.7a). CdiI^{CT-MAC} crystallized in space group was $P2_1$ and contains two complexes per asymmetric unit. The

final CdiI$^{CT\text{-}MAC}$ model contains two CdiI^{EC869} proteins containing residues 1-168, two macrocyclic peptides, and 132 waters, resulting in an R_{work}/R_{free} of 18.4/23.1.

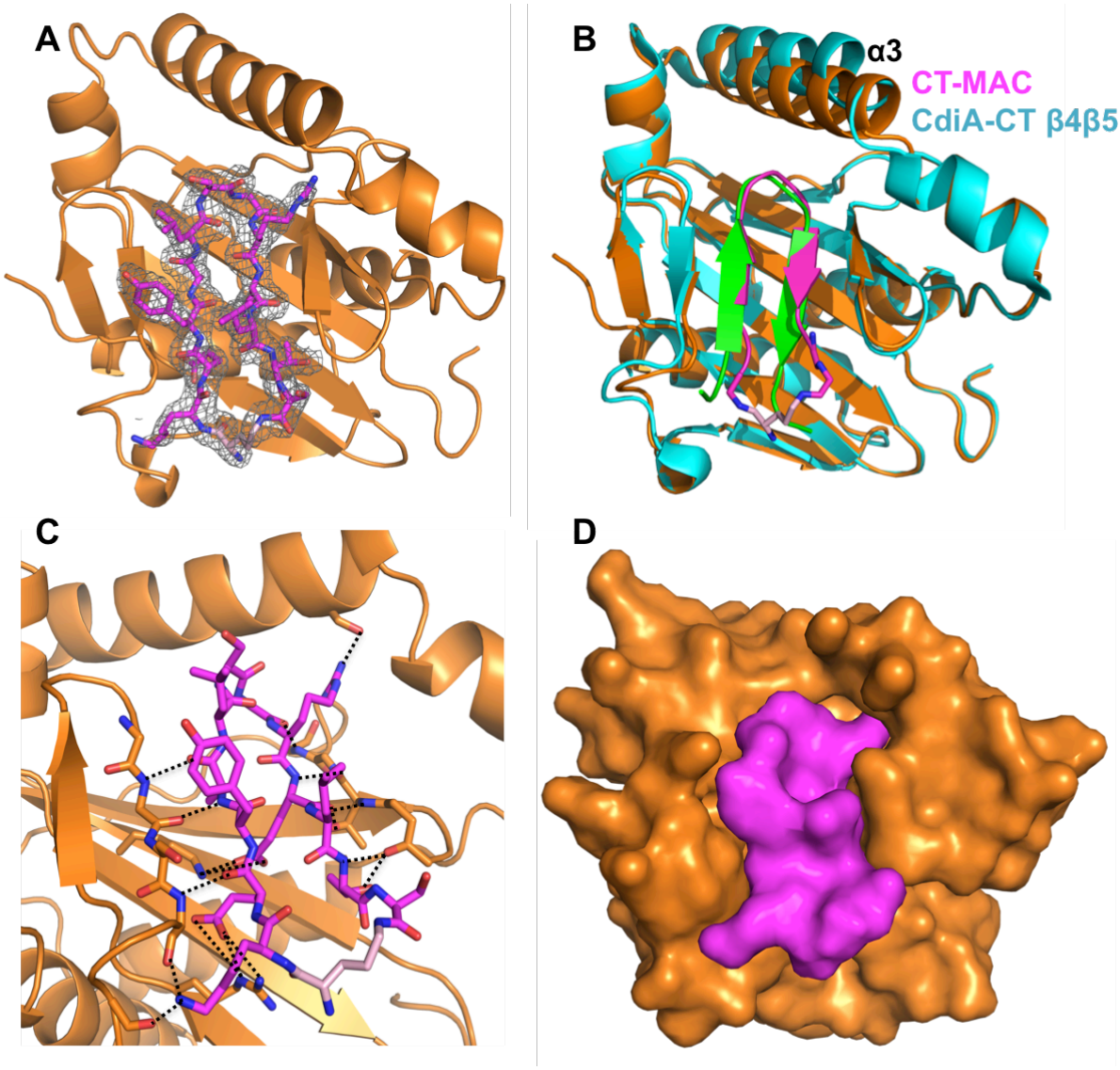

Figure 3.7. Structure of CdiI$^{CT\text{-}MAC}$. (A) Crystal structure of CdiI$^{CT\text{-}MAC}$ with CdiI^{EC869} depicted as orange ribbons, and CT-MAC displayed as sticks, carbon, nitrogen, oxygen, and ornithine carbons colored mageneta, blue, red, and pink, respectively. The CT-MAC $2F_o$-F_c electron density map is shown in grey mesh, contoured at 1.0σ. (B) Structural superimposition of CdiI$^{CT\text{-}MAC}$ (colored same as panel A) and the CdiA-CT^{EC869}/CdiI^{EC869} complex. Only the CdiA-CT^{EC869} β-hairpin (residues 242-253) are shown (green). CdiI^{EC869} is colored teal. The CdiI α3 helix is labeled. (C). CT-MAC interactions with CdiI^{EC869} via a network of hydrogen bonds and ion pairs. Interacting bonds are shown as black dotted lines. CdiI^{EC869} β-strands that H-bond with CT-MAC (β-strands β3a and β7) are shown as sticks. (D) Surface representation of the CdiI$^{CT\text{-}MAC}$ structure, oriented as in panels A and B, depicting the complementarity of CdiI^{EC869} and CT-MAC .

The structure of CdiI$^{CT\text{-}MAC}$ overlays with CdiI^{EC869} from the complex structure with rmsd of 0.437 Å over all α-carbons. However, there is an observed movement of one structural element between the two complexes (Figure 3.7b). In the CdiI$^{CT\text{-}MAC}$ structure, the CdiI^{EC869} α3 helix is displaced 3.0 Å, creating a slightly altered β-hairpin binding pocket. Despite this, in the CdiI$^{CT\text{-}MAC}$ structure CT-MAC is observed as a two-stranded cyclic β-sheet that forms a β-augmentation interaction with CdiI, resulting in an extensive network of backbone H-bond interactions (Figure 3.7c). H-bond or ion pair interactions are also observed via the side chains of five CT-MAC residues, in contrast to the six interacting residues from the β-hairpin of the CdiA-CT^{EC869}/CdiI^{EC869} complex structure. In the CdiA-CT^{EC869}/CdiI^{EC869} complex, the sidechains of β-hairpin residues Tyr244 and Ser247 interact with CdiI^{EC869}, however the corresponding CT-MAC residues (Tyr4 and Ser7) do not interact with CdiI^{EC869}. An additional interacting residue is CT-MAC Thr13 (corresponding to CdiA-CT^{EC869} Thr 252), which H-bonds with the backbone carbonyl of CdiI^{EC869} Gly29. Additionally, the orientation of Arg10 in CdiI$^{CT\text{-}MAC}$ (Arg249 in CdiA-CT^{EC869}) is different compared to that observed in the CdiA-CT^{EC869}/CdiI^{EC869} structure, resulting in it interacting with a different CdiI^{EC869} residue. Finally, the β-hairpin formed by CT-MAC only contains four cross-strand H-bonds, less than the five observed in the CdiA-CT^{EC869} β-hairpin. CT-MAC contains a bulge adjacent to the ornithine turn, preventing the formation of a fifth H-bond (Figure 3.7b). This suggests that the CT-MAC β-hairpin confirmation observed in the structure may be less stable than the CdiA-CT^{EC869} β-hairpin. Despite this, the CdiI$^{CT\text{-}MAC}$ structure establishes that structure-based designed macrocyclic peptides can bind in the CdiI^{EC869} β-hairpin binding pocket, forming contacts reminiscent of what is observed in EC869 family CdiA-CT/CdiI complexes.

Discussion

This work builds upon our previous characterization of the EC869 family of toxin and immunity proteins by demonstrating the significance of the toxin β-hairpin. Not only is the presence of the β-hairpin required for toxin binding to its cognate immunity protein, but the specific amino acid sequence of the β-hairpin contributes to the stability of the toxin/immunity interaction. It has been suggested that despite CDI toxin and immunity proteins from the same family share varying levels of sequence homology (and structural similarity), evolutionary pressures cause key residues involved at the complex interface to mutate, diversifying the protein family and improving binding specificity of cognate pairs (14). In general, CdiI sequence diversification seems to be rapid compared to CdiA-CTs, and this is likely the mechanism for ensuring continued high levels of specificity for cognate toxin/immunity pairs. CdiA-CT evolution is slowed by the need to maintain toxin activity. However in the case of EC869 family CdiA-CTs, the β-hairpin sequence is extremely diverse (Figure 3.1), likely due to the fact that it is involved in cognate CdiI binding. It is unclear whether the β-hairpin contributes to toxin activity. The CdiA-CT^{YP413}/CdiI^{YP413} structure reported herein unsurprisingly demonstrates conservation of the β-augmentation interaction observed in our previous EC869 complex structure. However, aspects of this structure allow us to explore the diversification of EC869 family toxin/immunity complexes. The CdiA-CT^{YP413} β4β5-hairpin consists mostly of hydrophobic residues, resulting in only two side chain interactions (via E242 and K243) with CdiI^{YP413} (Figure 3.8). Compared to CdiA-CT^{EC869}/CdiI^{EC869} in which six hairpin side chains form interactions, this would suggest

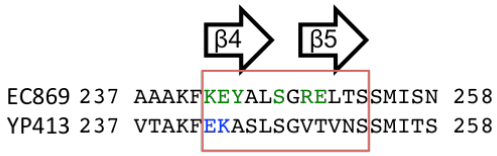

Figure 3.8. Alignment of the CdiA-CT^{EC869} and CdiA-CT^{YP413} β-hairpin sequences. Side-chains that interact with cognate CdiI are shown in green (EC869) or blue (YP413).

that the YP413 toxin and immunity proteins may have reduced binding specificity. Interestingly, the charges of the two interacting CdiA-CT^{YP413} residues are switched as compared to CdiA-CT^{EC869}. Also, compared to EC869, the YP413 complex forms more salt bridges outside of the β-hairpin region. This includes interaction via CdiI^{413} loop L1, which is a loop not present in any of the other immunity proteins studied (Figures 3.5 and 3.6). While side chain interactions via the CdiA-CT^{YP413} β4β5-hairpin are few, these additional strong interactions may account for the high binding specificity and affinity. Based on our studies, one could speculate that formation of the EC869 family complexes may proceed first by sampling of the CdiI "β-hairpin pocket" by the CdiA-CT β-hairpin backbone, followed by formation of the stable CdiA-CT/CdiI complex through interactions via the protein side chains, with some of these interactions occurring away from the β-hairpin binding region. This proposed interaction sequence is also in agreement with our binding data with CdiA-CT$^{EC869\text{-}NL}$ and CdiA-CT$^{EC869\text{-}YK}$, in which the presence of cognate β-hairpins offer a certain level of binding affinity, but the lack of specific interacting residues outside of this region prevents high affinity complex formation.

The β-hairpin binding pocket in EC869 family CdiI proteins offers the possibility to function as a scaffold for small molecules or peptides designed to inhibit formation of the CdiA-CT/CdiI complex. Small cyclic peptides that fold into β-hairpins have been shown to be useful in studying protein-protein and protein-DNA interactions in the past (21), with some even displaying our desired effect of specifically inhibiting formation of a protein-protein complex (22, 23). As showcased by our CdiI$^{CT\text{-}MAC}$ structure, macrocyclic peptides can be designed to bind and fit into the CdiI β-hairpin binding pocket by mimicking the cognate toxin's β-hairpin. Notably, this structure represents the first structural example of a macrocyclic peptide cyclized by an ornithine turn unit in complex with a protein target. However, our design could be improved to enhance binding affinity and possibly allow for function as a protein-protein complex inhibitor. Previous ornithine cyclized macrocyclic β-sheets used for crystallography have contained longer β-strands than in the CT-MAC used for our study (24). Perhaps including

additional residues in our β4β5 macrocyclic mimic, while still preserving the residues important for CdiI interaction, could possibly increase H-bond interaction between immunity protein β-strands. This could potentially improve mimicry of the CdiA-CT^{EC869} β-hairpin, as the toxin β-hairpin contains an additional cross-strand H-bond compared to the macrocycle. In addition, to improve the binding ability of the macrocycle to CdiI^{EC869}, modifications to the amino acid sequence can be introduced to replace the Ser-Gly β-turn currently present in the peptide. The Ser-Gly turn mimics the CdiA-CT^{EC869} β-hairpin, but neither of these residues interacts with CdiI^{EC869}. Replacing the Ser with a Gln could allow for an ion pair to form between the Gln and a nearby CdiI^{EC869} Glu. Additional residues in the macrocycle can also be changed to achieve additional macrocycle-protein interactions. For example, CT-MAC residues corresponding to CdiA-CT^{EC869} residues Ala245 and Leu246 could potentially be changed to acidic amino acids to allow for interactions with a pair of nearby CdiI^{EC869} Lys residues. Generation of this high affinity macrocycle could be a useful tool by allowing for selective disruption of the CdiA-CT^{EC869}/CdiI^{EC869} complex, thus causing bacterial death.

References

1. Aoki SK*, et al.* (2005) Contact-dependent inhibition of growth in Escherichia coli. *Science* 309(5738):1245-1248.
2. Aoki SK*, et al.* (2010) A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* 468(7322):439-442.
3. Aoki SK*, et al.* (2008) Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Molecular microbiology* 70(2):323-340.
4. Ruhe ZC, Wallace AB, Low DA, & Hayes CS (2013) Receptor polymorphism restricts contact-dependent growth inhibition to members of the same species. *mBio* 4(4).
5. Webb JS*, et al.* (2013) Delivery of CdiA nuclease toxins into target cells during contact-dependent growth inhibition. *PLoS One* 8(2):e57609.
6. Morse RP*, et al.* (2012) Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems. *Proceedings of the National Academy of Sciences of the United States of America* 109(52):21480-21485.
7. Nikolakakis K*, et al.* (2012) The toxin/immunity network of Burkholderia pseudomallei contact-dependent growth inhibition (CDI) systems. *Molecular microbiology* 84(3):516-529.
8. Diner EJ, Beck CM, Webb JS, Low DA, & Hayes CS (2012) Identification of a target cell permissive factor required for contact-dependent growth inhibition (CDI). *Genes Dev* 26(5):515-525.
9. Beck CM*, et al.* (2014) CdiA from Enterobacter cloacae Delivers a Toxic Ribosomal RNase into Target Bacteria. *Structure* 22.
10. Aoki SK, Webb JS, Braaten BA, & Low DA (2009) Contact-dependent growth inhibition causes reversible metabolic downregulation in Escherichia coli. *Journal of bacteriology* 191(6):1777-1786.
11. Poole SJ*, et al.* (2011) Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. *PLoS genetics* 7(8):e1002217.
12. Blower TR, Salmond GP, & Luisi BF (2011) Balancing at survival's edge: the structure and adaptive benefits of prokaryotic toxin-antitoxin partners. *Current opinion in structural biology* 21(1):109-118.
13. Cascales E*, et al.* (2007) Colicin biology. *Microbiology and molecular biology reviews : MMBR* 71(1):158-229.
14. Ruhe ZC, Low DA, & Hayes CS (2013) Bacterial contact-dependent growth inhibition. *Trends in microbiology* 21(5):230-237.
15. Battye TG, Kontogiannis L, Johnson O, Powell HR, & Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography* 67(Pt 4):271-281.
16. Otwinowski Z & Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* 276:307-326.
17. Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.0.
18. Bergfors T (2003) Seeds to crystals. *Journal of structural biology* 142(1):66-76.
19. Woods RJ*, et al.* (2007) Cyclic modular beta-sheets. *Journal of the American Chemical Society* 129(9):2548-2558.
20. Nowick JS & Brower JO (2003) A new turn structure for the formation of beta-hairpins in peptides. *Journal of the American Chemical Society* 125(4):876-877.
21. Loughlin WA, Tyndall JD, Glenn MP, & Fairlie DP (2004) Beta-strand mimetics. *Chemical reviews* 104(12):6085-6117.

22. Dias RL*, et al.* (2006) Protein ligand design: from phage display to synthetic protein epitope mimetics in human antibody Fc-binding peptidomimetics. *Journal of the American Chemical Society* 128(8):2726-2732.
23. Fasan R*, et al.* (2004) Using a beta-hairpin to mimic an alpha-helix: cyclic peptidomimetic inhibitors of the p53-HDM2 protein-protein interaction. *Angewandte Chemie* 43(16):2109-2112.
24. Pham JD, Chim N, Goulding CW, & Nowick JS (2013) Structures of oligomers of a peptide from beta-amyloid. *Journal of the American Chemical Society* 135(33):12460-12467.

CHAPTER 4

CdiA from *Enterobacter cloacae* delivers a toxic ribosomal RNase into target bacteria

Summary

Contact-dependent growth inhibition (CDI) is one mechanism of inter-bacterial competition. CDI^+ cells export large CdiA effector proteins, which carry a variety of C-terminal toxin domains (CdiA-CT). CdiA-CT toxins are specifically neutralized by cognate CdiI immunity proteins to protect toxin-producing cells from auto-inhibition. Here, we use structure determination to elucidate the activity of a unique CDI toxin from *Enterobacter cloacae* (ECL). The CdiA-CT^{ECL} structure is similar to the C-terminal nuclease domain of colicin E3, which cleaves 16S ribosomal RNA to disrupt protein synthesis. In accord with this structural homology, we show that CdiA-CT^{ECL} uses the same nuclease activity to inhibit bacterial growth. Surprisingly, although colicin E3 and $CdiA^{ECL}$ carry equivalent toxin domains, the corresponding immunity proteins are unrelated in sequence, structure and toxin-binding site. Together, these findings reveal unexpected diversity amongst 16S rRNases and suggest that these nucleases are robust and versatile payloads for a variety of toxin-delivery platforms.

Introduction

Bacterial genomes and plasmids encode a variety of peptide and protein toxins that mediate inter-bacterial competition. Colicins were the first of such toxins to be identified and characterized from strains of *Escherichia coli*. Subsequently, it was discovered that other bacteria release similar toxins, which are now collectively termed bacteriocins (1). Bacteriocins are diffusible proteins that parasitize cell-envelope proteins to enter and kill bacteria. These toxins are composed of three domains, each responsible for a distinct step in the cell-killing pathway. The central domain binds specific receptors on the surface of susceptible bacteria. The N-terminal domain mediates translocation across the cell envelope, and the C-terminal domain carries the bacteriocidal activity. This modular structure allows for delivery of diverse C-terminal toxins using conserved translocation and receptor-binding domains. For example, colicins E2 through E9 share virtually identical N-terminal domains but carry different C-terminal toxins with DNase (2), ribosomal RNase (3, 4) or tRNA anticodon nuclease activities (5). Bacteriocin genes are always closely linked to immunity genes, which encode small proteins that bind the toxin domain and neutralize its toxicity. Thus, cells that harbor bacteriocinogenic plasmids are protected from toxin activity, but they may still be susceptible to the bacteriocins produced from other plasmids. Many different bacteriocin/immunity types are typically present in a given environment (6, 7), and these plasmids are predicted to have a significant impact on bacterial population structures (8, 9).

Research over the past decade has uncovered additional bacterial competition systems that require direct cell-to-cell contact for toxin delivery (10-14). There are at least two pathways – mediated by type V and type VI secretion systems – for contact-dependent toxin delivery between Gram-negative bacteria (15, 16). The type V mechanism was the first identified and this phenomenon was termed "CDI" for contact-dependent growth inhibition (10). CDI is mediated by the CdiB/CdiA family of two-partner secretion proteins. CdiB is a predicted b-barrel

protein that resides in the outer membrane and is required for export of CdiA effectors. CdiA proteins are very large (250 – 600 kDa) and are thought to extend from the inhibitor cell to interact with neighboring target bacteria. Although CdiA and bacteriocins are unrelated, these effector proteins share a number of general features. Like bacteriocins, CdiA proteins bind to specific receptors on the surface of target bacteria and these interactions determine the target-cell range (17, 18). Additionally, CDI toxin activity is carried at the extreme C-terminus of CdiA, and some portion of this CdiA-CT region is translocated into target bacteria (11, 19, 20). CDI loci also encode CdiI immunity proteins, which bind and inactivate CdiA-CTs to protect toxin-producing cells from auto-inhibition. Finally, CDI systems deploy a variety of toxin domains with distinct biochemical activities. Remarkably, chimeric CDI effectors can be produced by fusing different toxins onto CdiA at the conserved VENN peptide motif that demarcates the CdiA-CT region (11). There is also evidence that bacteria exchange *cdiA-CT/cdiI* genes through horizontal transfer (21), suggesting that effector modularity is exploited to switch toxin/immunity type. In fact, bacteria collectively contain a large repository of toxin/immunity genes that are shared by a variety of toxin-delivery systems (21-24). For example, at least two CdiA proteins carry toxins that resemble bacteriocin nucleases. $CdiA^{Dd3937}$ from *Dickeya dadantii* 3937 carries a CT domain with 35% identity to the pyocin S3 DNase domain (11), and the C-terminal region of $CdiA^{K96243}$ from *Burkholderia pseudomallei* K96243 is 49% identical to the anticodon tRNase domain of colicin E5. Biochemical analyses have confirmed that each CDI toxin has the same nuclease activity as the corresponding bacteriocin (11, 25). Together, these observations suggest that CDI loci integrate toxin/immunity gene pairs from diverse sources and that this diversity contributes to inter-strain competition.

In an effort to understand CDI toxin/immunity diversity and uncover new toxin activities, we have initiated structural studies of CdiA-CT/CdiI pairs from various bacteria. Here, we describe the structure and function of the CDI toxin/immunity protein pair from *Enterobacter*

cloacae ATCC 13047 (ECL). The CdiA-CT^{ECL} toxin shares no significant sequence identity with proteins of known function, but the three-dimensional structure of CdiA-CT^{ECL} reveals similarity to the C-terminal nuclease domain of colicin E3. In accord with the structural homology, CdiA-CT^{ECL} cleaves 16S rRNA at the same site as colicin E3 and this nuclease activity is responsible for growth inhibition. By contrast, CdiI^{ECL} does not resemble the colicin E3 immunity protein (ImE3), and the two immunity proteins bind to different sites on their respective cognate toxin domains. Inspection of other CdiA proteins from *Erwinia chrysanthemi* EC16 (Uniprot: P94772), *Enterobacter hormaechei* ATCC 49162 (F5S237) and *Pseudomonas viridiflava* UASWS0038 (K6CF79) has revealed that their toxin domains share a common nuclease motif with colicin E3 (26). Analysis of CdiA-CT^{EC16} from *Erwinia chrysanthemi* EC16 confirms that this toxin has 16S rRNase activity and demonstrates that the associated CdiI^{EC16} immunity protein is specific to CdiA-CT^{EC16} and does not provide protection against the CdiA-CT^{ECL} nuclease. Together, these observations indicate that 16S rRNase toxins are more diverse and widespread than previously recognized.

Materials and Methods

Bacterial strains, plasmids and growth conditions

All bacterial strains and plasmids used in this study are listed in Table S2 of Beck *et al* (27). Bacteria were grown in LB media or LB-agar with the appropriate antibiotics as described in the supplement. *E. cloacae* genes were deleted using the same protocol as described for *E. coli* (28). DNA sequences located upstream and downstream of target genes were amplified and cloned into plasmid pKAN or pSPM (29) to flank kanamycin- or spectinomycin-resistance cassettes, respectively. The resulting plasmids were linearized by restriction endonuclease digestion and electroporated into *E. cloacae* cells expressing the phage l Red proteins from

plasmid pKOBEG (30). The details of all strain and plasmid constructions are provided in the supplement (27).

Protein purification and crystallography

CdiA-CT/CdiI-His_6 complexes were purified and the toxin and immunity proteins isolated from one another as described (25, 31). Protein crystallization of the ECL complex was as described (32). Crystals were grown at room temperature by hanging drop-vapor diffusion with a reservoir containing 1.5 M $(NH_4)_2SO_4$, 0.1 M Bis Tris (pH 5.1) and 1% (wt/vol) PEG 3350. The structural model was determined as described (19). All crystallography and refinement statistics are presented in Table 4.1. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (www.pdb.org) as PDB ID code 4NTQ.

Nuclease assays (Experiments performed by Hayes lab, UCSB)

Ribosomes were isolated from S30 lysates of *E. coli* as described (33) and incubated with purified CdiA-CT toxins and CdiI immunity proteins as described in the supplement. All reactions were analyzed by northern blot using a probe complementary to the 3´-end of *E. coli* 16S rRNA. CdiA-CT cleavage sites were determined using ribosomes from *E. coli* Δ*rsmE* cells. Reactions were quenched with guanidinium isothiocyanate-phenol and rRNA extracted for primer extension analysis as described (33)

Growth competitions (Experiments performed by Hayes lab, UCSB)

E. cloacae inhibitor cells were co-cultured with *E. cloacae* Δ*cdiAI* target cells on LB-agar supplemented with 0.2% L-arabinose. Cells were harvested and enumerated as colony forming units (CFU). Immunity function was evaluated through expression of *cdiI* genes in target cells from plasmid constructs as described in the supplement. Cross-species competitions were performed under the same conditions using *E. coli* target cells. Chimeric EC93-ECL CDI

systems were expressed from cosmids in *E. coli* EPI100. Inhibitor cells were co-cultured with target cells in LB media. Samples were taken for enumeration of viable target cells. *E. coli* EPI100 cells carrying cosmid pWEB-TNC were used as mock (CDI–) inhibitors.

Table 4.1. X-ray diffraction data and atomic refinement for the CdiA-CT^{ECL}/CdiI^{ECL} complex

	Peak	**Remote**	**Inflection**	**Native**
Space Group	$P4_122$	$P4_122$	$P4_122$	$P4_122$
Unit cell dimensions (Å)	85.64 × 85.64 × 75.17	85.64 × 85.64 × 75.17	85.64 × 85.64 × 75.17	85.25 × 85.25 × 74.91
pH of crystallization condition	5.1	5.1	5.1	5.1
Protein concentration (mg/mL)	9	9	9	9
Data set				
Wavelength (Å)	0.9759	1.377	0.9794	1
Resolution range	50-2.85	50-3.0	50-2.9	50-2.4
Unique reflections (total)	5486 (191798)	4645 (163018)	5179 (181470)	11315 (324387)
Completeness (%)*	100.0 (100.0)	100.0 (100)	100.0 (100.0)	100 (100)
Redundancy *	27.8 (28.8)	27.6 (28.5)	27.8 (28.7)	28.7 (29.3)
R_{merge}*,†	0.114 (0.47)	0.129 (0.501)	0.109 (0.445)	0.088 (0.455)
I/σ*	31.1 (10.42)	27.8 (8.7)	34.8 (10.8)	44.7 (11)
NCS copies	1			1
No. of Selenium sites/a.u	6			
FOM	0.49			
Model refinement				
Resolution range (Å)				38.125-2.400
No. of reflections				11291/538
No. of protein atoms				1760
No. of water molecules				62
Missing residues				CdiA-CT 1-160
R_{work}/R_{free}‡, %				18.3/23.7
R.m.s deviations				
Bond lengths (Å)				0.008
Bond angles (degrees)				1.15
Ramachandran Plot				
Most favorable region (%)				93.61
Additional allowed region (%)				6.39
Disallowed region				0.0
PDB ID Code				4NTQ

*Statistics for the highest resolution shell are given in (brackets)

†$R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma I$

‡$R_{work}=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$ R_{free} was computed identically except where all reflections belong to a test set of 10% randomly selected data.

Results

Crystallization and structure of the CdiA-CT^{ECL}/CdiI^{ECL} complex

In a previous study, we used structural analysis to elucidate the activities of CDI toxins from *E. coli* EC869 and *Burkholderia pseudomallei* 1026b (19). Because the CDI toxin/immunity pair from *E. cloacae* ATCC 13047 shares no sequence homology with proteins of known function, we followed a similar structure-based approach to characterize this system. The CdiA-CT^{ECL} region is demarcated by the AENN peptide motif and corresponds to residues Ala3087 to Asp3321 of full-length CdiA^{ECL}. We co-expressed CdiA-CT^{ECL} with His$_6$-tagged CdiI^{ECL} and purified the complex to near homogeneity. The N-terminal region of CdiA-CT^{ECL} underwent significant degradation during crystallization, presumably because this region is disordered. Similar N-terminal degradation has been observed with other CdiA-CTs (19). The CdiA-CT^{ECL}/CdiI^{ECL} complex crystallized in space group $P4_122$ with one heterodimeric complex per asymmetric unit. The structure was solved by selenium multiple wavelength anomalous dispersion (Se-MAD) phasing to 2.4 Å resolution. The final refined model contains CdiA-CT^{ECL} residues 160 – 235 (numbered from Ala1 of the AENN motif) and CdiI^{ECL} residues 1 – 145. In addition, 62 well-resolved water molecules are included in the final model resulting in R_{work}/R_{free} of 18.3/23.7 (Table 4.1).

The resolved C-terminal domain of CdiA-CT^{ECL} consists of an N-terminal α-helix followed by a twisted five-stranded antiparallel β-sheet (Figure 4.1a). The domain contains two long loops, L2 and L4, which connect β1 to β2 and β3 to β4, respectively (Figure 4.1a). Weak electron density was observed for loop L4, likely due to its flexibility, and thus Ser206 – Asn211 were modeled as alanine residues. The CdiI^{ECL} immunity protein comprises three- and four-stranded antiparallel β-sheets, forming a β-sandwich that is decorated with three α-helices

(Figure 4.1a). The toxin and immunity protein interface is elaborate and mediated by a series of hydrogen-bonds (H-bond), electrostatic and hydrophobic interactions (Figure 4.1b, Table 4.2). CdiA-CT^{ECL} residues within loops L2 – L6 form H-bonds and ion-pair interactions with CdiI^{ECL} residues in loops L1′, L2′ and L3′ and the edge of the β-sandwich (β3′, β5′ and β6′) (Figure 4.1b). A water-mediated network of H-bonds also contributes to the interface, resulting in more than 20 ion-pair/H-bond interactions between toxin and immunity proteins (Figure 4.1b, Table 4.2). In addition, there is a hydrophobic interface of approximately 300 Å^2 consisting of Ile178, Val192, Tyr199 and Phe216 from CdiA-CT^{ECL}, and Phe76, Phe78, Val95 and Phe97 from CdiI^{ECL} (Figure 4.1c). Overall, the CdiA-CT^{ECL}/CdiI^{ECL} complex has an interface of 1399 Å^2, burying 27.6 and 17.1% of the solvent-accessible surface areas of the toxin and immunity proteins, respectively.

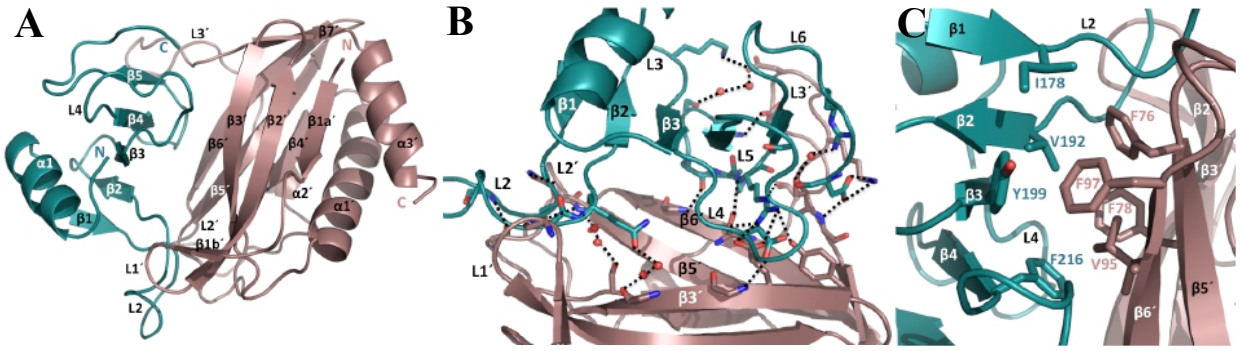

Figure 4.1. Structure of the CdiA-CT/CdiI^{ECL} complex. (A) The CdiA-CT^{ECL} toxin (teal) and CdiI^{ECL} immunity protein (salmon pink) are depicted in ribbon representation with secondary structure elements. The amino and carboxyl termini are indicated by N and C, respectively. CdiI^{ECL} elements are denoted with a prime symbol (′) to differentiate them from the toxin secondary structure elements. (B) The CdiA-CT/CdiI^{ECL} interface is mediated by an extensive network of ion-pair and hydrogen-bond interactions. Water molecules are depicted as red spheres and interacting bonds as black dotted lines. (C) The CdiA-CT/CdiI^{ECL} interface also contains hydrophobic interactions mediated by the indicated residues. The view in panel B represents a 90° clockwise rotation of panels A & C.

Table 4.2. Hydrogen bonds and ion pairs between CdiA-CT^{ECL} and CdiI^{ECL}

CdiA-CT toxin	CdiI immunity	Distance (Å)
Pro179 O	Asn75 ND2	2.99
Pro180 O	Asn75 N	2.81
Pro180 O	Phe76 N	3.89
Lys182 N	*Ala73* O[a]	2.70
Gly189 N	Glu14 O	3.07
Lys197 NZ	Asp122 OD1, OD2	3.75, 2.82
Asp205 OD2	Glu31 N	3.15
Asn217 ND2	Asp120 OD1	2.82
Gly220 O	Val95 N	2.68
Lys221 NZ	Thr117 OG1	3.19
Arg222 NH2	Tyr28 OH	3.39
Arg222 NH1, NH1, NH2	Glu30 OE1, OE2, OE1	2.86, 3.49, 3.36
Arg222 N, NH1	Asn93 O, OD1	3.27, 2.96
Gly224 N	Ser127 OG	3.34
Asp235 OD1, OD2	Lys125 NZ, NZ	3.40, 3.51
Asp235 OD2	Glu128 N	3.20

CdiA-CT^{ECL} is structurally homologous to the nuclease domain of colicin E3

CdiA-CT^{ECL} shares no structural homology with previously characterized CDI toxins from *E. coli* EC869 and *B. pseudomallei* 1026b (19, 25). Searches for structural homologues using the DALI server (34) revealed that CdiA-CT^{ECL} is similar to the C-terminal nuclease domain of colicin E3 (ColE3-CT). Colicin E3 is a plasmid-encoded bacteriocin found in some *E. coli* strains, and its nuclease domain cleaves 16S rRNA between residues A1493 and G1494 (*E. coli* numbering) to interfere with protein synthesis (35, 36). The CdiA-CT^{ECL} and ColE3-CT domains share a twisted antiparallel b-sheet and superimpose with an rmsd of 2.1 Å over 76 α-carbons, corresponding to a Z-score of 4.8, whereas the sequence identity between the two domains is approximately 18% (Figures 4.2a & 4.3a). Residues Asp510, His513 and Glu517 of colicin E3 are thought to function directly in catalysis (26, 36, 37), and CdiA-CT^{ECL} residues Asp203, Asp205 and Lys214 superimpose upon the colicin E3 active site residues (Figures 4.2b & 4.3a). Together, these structural similarities suggest that CdiA-CT^{ECL} may share 16S rRNA nuclease activity with colicin E3.

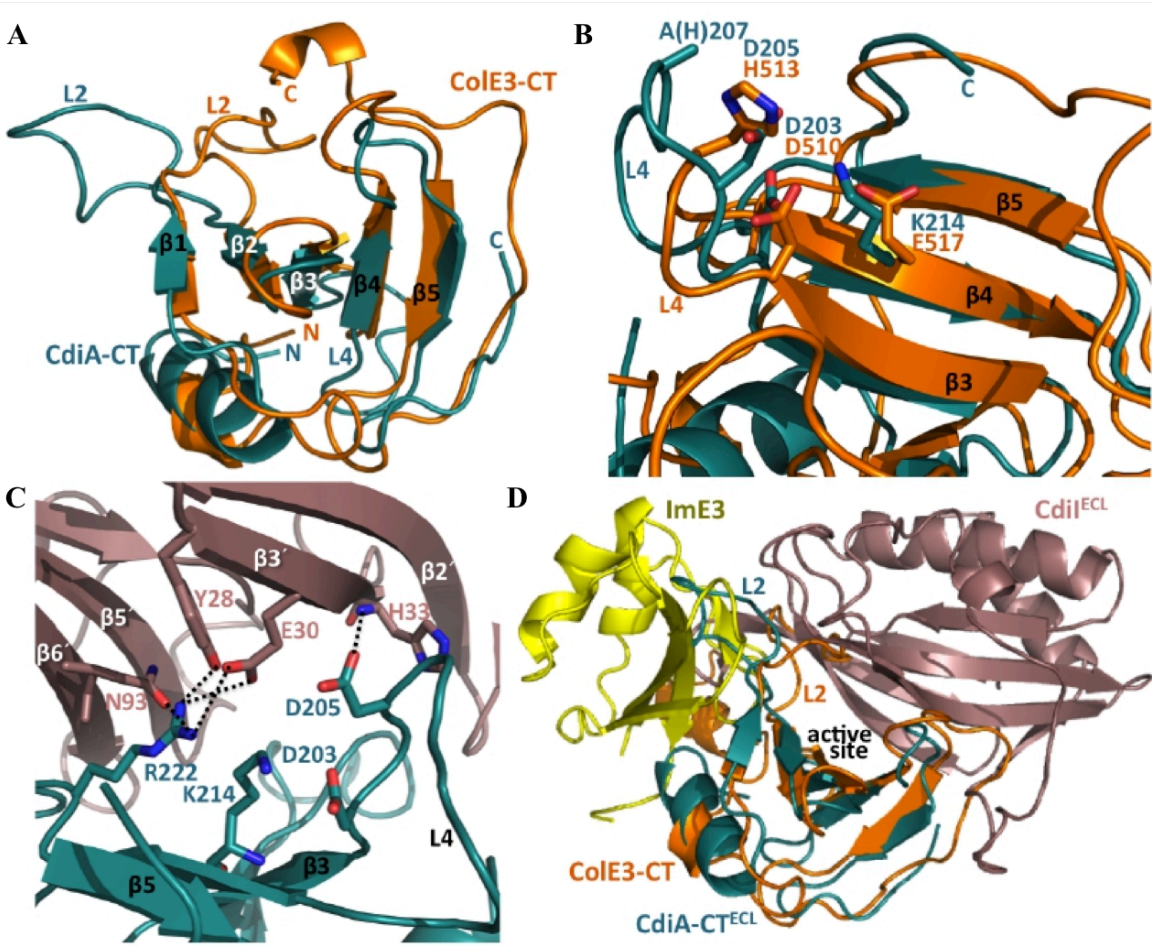

Figure 4.2. CdiA-CT^{ECL} share structural similarities with the nuclease domain of colicin E3. (A) Superimposition of CdiA-CT^{ECL} (teal) and the C-terminal nuclease domain of colicin E3 (ColE3-CT, orange) (PDB ID: 2B5U). The toxin domains superimpose with an rmsd of 2.1 Å. (B) Colicin E3 residues Asp510, His513 and E517 are involved in catalysis and superimpose with residues Asp203, Asp205 and Lys214 of CdiA-CT^{ECL}. His207 of CdiA-CT^{ECL} is located within disordered loop L4 and is modeled as an alanine residue. Residues are indicated in one-letter code and rendered as stick representations. (C) The predicted CdiA-CT^{ECL} active site is occluded by bound CdiI^{ECL}. Interacting bonds are represented by black dotted lines. D) Superimposition of CdiA-CT/CdiI^{ECL} with the ColE3-CT/ImE3 complex. Ribbon representations of CdiA-CT^{ECL} (teal), CdiI^{ECL} (salmon pink), ColE3-CT (orange) and ImE3 (yellow) are depicted.

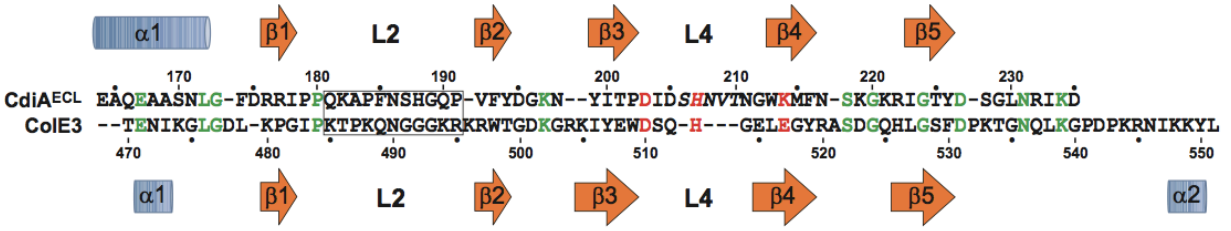

Figure 4.3. Alignment of CdiA-CT^{ECL} with the nuclease domain of colicin E3**.** The sequences of CdiA-CT^{ECL} and ColE3-CT were aligned based on the superimposed structures of the nuclease domains. Secondary structure elements are indicated as blue α-helices and orange β-sheets. Boxed residues correspond to loop L2 regions, which adopt different positions in the two toxins. Italicized residues in CdiA-CT^{ECL} correspond to the unresolved loop L4 region. Identical residues are rendered in green and the predicted active-site residues are shown in red.

Although the CdiA-CT^{ECL} and ColE3-CT toxin domains are structurally similar, the corresponding immunity proteins are not related to one another in either primary or tertiary structure (Figures 4.4a & 4.4b). The colicin E3 immunity protein (ImE3) is significantly smaller than CdiI^{ECL} (~9.9 versus 16.9 kDa), and the two proteins have different folds (Figure 4.4b). A DALI search reveals that CdiI^{ECL} is most similar to the Whirly family of single-stranded DNA binding proteins (38). The closest structural homologues are two proteins of unknown function from cyanobacteria (PDB ID codes: 2IT9 and 2NVN), which superimpose onto CdiI^{ECL} with rmsd of 3.6 – 4.0 Å over 120 – 122 α-carbons (Figure 4.4c). CdiI^{ECL} and ImE3 also bind their cognate toxins differently. ImE3 binds to an 'exosite' that leaves the colicin E3 active site exposed (39), whereas CdiI^{ECL} binds directly over the predicted active site (Figure 4.2c). Structural alignment of the complexes shows that immunity protein binding occurs at distinct non-overlapping positions (Figure 4.2d). Interestingly, ColE3-CT contains a C-terminal extension not found in CdiA-CT^{ECL} (Figure 4.2a). This C-terminal tail forms a short α-helix in one ColE3-CT structure (37), and this element would likely interfere with CdiI^{ECL} binding were it present in CdiA-CT^{ECL}. Similarly, the orientation of loop L2 differs considerably between the toxins (Figure 4.2a), and these loops could block the binding of non-cognate immunity proteins (Figure 4.2d).

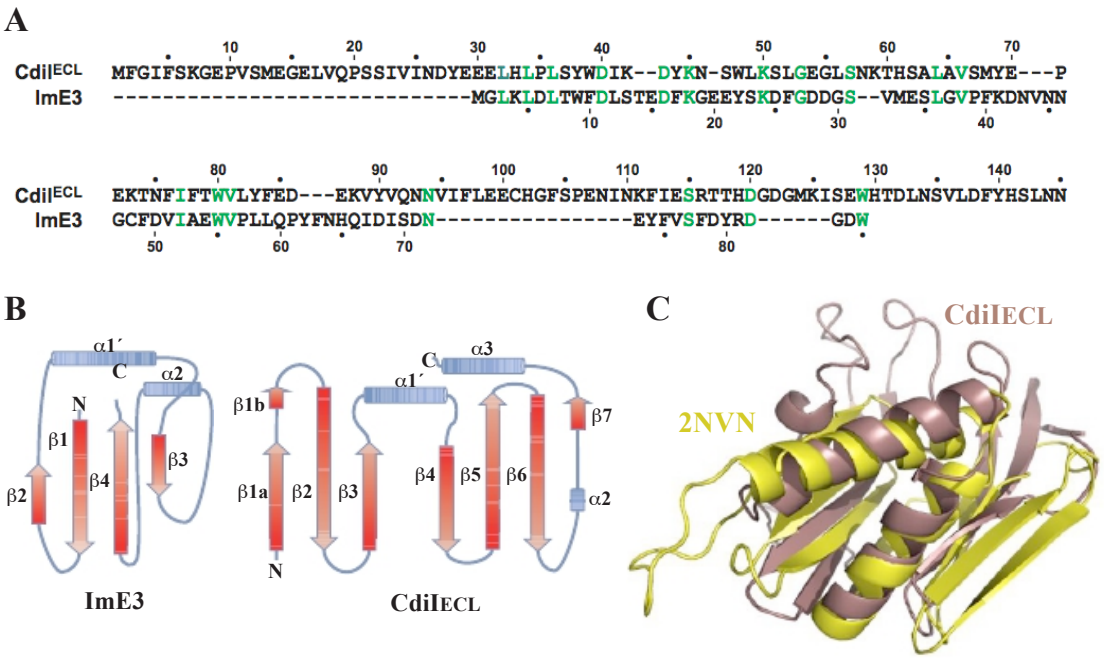

Figure 4.4. CdiI^{ECL} and ImE3 immunity proteins are unrelated. (A) Alignment of CdiI^{ECL} and ImE3 immunity proteins sequences using Clustal-W. Identical residues are rendered in green and the two proteins share 11.7% identity. (B) Topologies of CdiI^{ECL} and ImE3 immunity proteins. The N- and C- termini are indicated for each immunity protein, as are the secondary structure elements. (C) Superimposition of CdiI^{ECL} and Uniprot entry Q31MH7 from *Synechococcus elongatus* PCC 7942 (PDB:2NVN). The proteins overlay with rmsd 3.6 Å over 121 α-carbons and a Z-score of 4.1.

CdiA-CT^{ECL} cleaves 16S rRNA in vivo to inhibit cell growth (Experiments performed by Hayes lab, UCSB)

The structural resemblance of CdiA-CT^{ECL} to ColE3-CT suggests that the CDI toxin also cleaves 16S rRNA. To test this prediction, we cloned *cdiA-CT*^{ECL} under the control of an arabinose-inducible promoter and asked whether 16S rRNA is cleaved upon induction with L-arabinose. *E. coli* cells carrying the *cdiA-CT*^{ECL} construct do not grow when the media is supplemented with L-arabinose (data not shown), confirming that CdiA-CT^{ECL} is an inhibitory toxin. We isolated total RNA from the inhibited cells and analyzed 16S rRNA by northern blot. This analysis revealed that 16S rRNA is cleaved in cells expressing *cdiA-CT*^{ECL}, but remains intact in control cells that carry the vector plasmid alone (data not shown). We next tested

CdiI^{ECL} function to determine whether it neutralizes the growth inhibition and nuclease activities of CdiA-CT^{ECL}, and demonstrated that indeed CdiI^{ECL} prevents CdiA-CT^{ECL}-mediated growth inhibition (data not shown). Together, these results show that CdiA-CT^{ECL} and CdiI^{ECL} constitute a cognate toxin/immunity pair that targets the ribosome.

CdiI^{ECL} immunity function is specific for its cognate toxin (Experiments performed by Hayes lab, UCSB)

At least one other CdiA protein is predicted to possess 16S rRNase activity. Kleanthous and colleagues discovered that HecA from *Erwinia chrysanthemi* EC16 contains the same catalytic motif as colicin E3 (26, 27). HecA was originally identified as an adhesin that promotes bacterial colonization of plant hosts (40, 41), but this protein shares 68% sequence identity with a known CdiA effector (Uniprot: E0SCQ6) from *Dickeya dadantii* 3937 (11). Together, these observations suggest that HecA actually functions in CDI, and therefore we refer to this protein as CdiA^{EC16}. The Hayes lab, then went on to show that CdiA-CT^{ECL} and CdiA-CT^{EC16} share a common growth inhibition activity, and the associated immunity proteins only provide protection against their cognate toxins.

Purified CdiA-CT^{ECL} and CdiA-CT^{EC16} cleave 16S rRNA in vitro (Experiments performed by Hayes lab, UCSB)

In principle, CdiA-CT^{ECL} and CdiA-CT^{EC16} could induce an endogenous nuclease activity that actually catalyzes 16S rRNA cleavage. The Hayes then showed that each toxin is directly responsible for 16S rRNA cleavage.

We next used primer extension analysis to determine whether the CDI toxins cleave 16S rRNA at the same position as colicin E3. We generated an oligonucleotide that hybridizes to residues C1501 – C1521 of *E. coli* 16S rRNA (Figure 4.5a) and used it as a primer in reverse transcription reactions to screen for cleavage sites. Residue U1498 of 16S rRNA is methylated

at the N3 position (Figure 4.5a) and this modified base is predicted to interfere with reverse transcription. Therefore, we repeated the *in vitro* nuclease reactions using ribosomes isolated from an *E. coli* Δ*rsmE::kan* mutant, which lacks the U1498 methyltransferase (42). Analysis of these nuclease reactions shows a strong primer-extension arrest corresponding to residue G1494 (Figures. 4.5a & 4.5b). This primer extension product is not observed when ribosomes are mock-treated with buffer, nor when the reactions contain equimolar cognate CdiI protein (Figure 4.5b). These data are consistent with CdiA-CT-mediated cleavage of the phosphodiester bond linking residues A1493 and G1494 (Figure 4.5a). Thus, CdiA-CT^{ECL} and CdiA-CT^{EC16} both appear to cleave 16S rRNA at the same site as colicin E3.

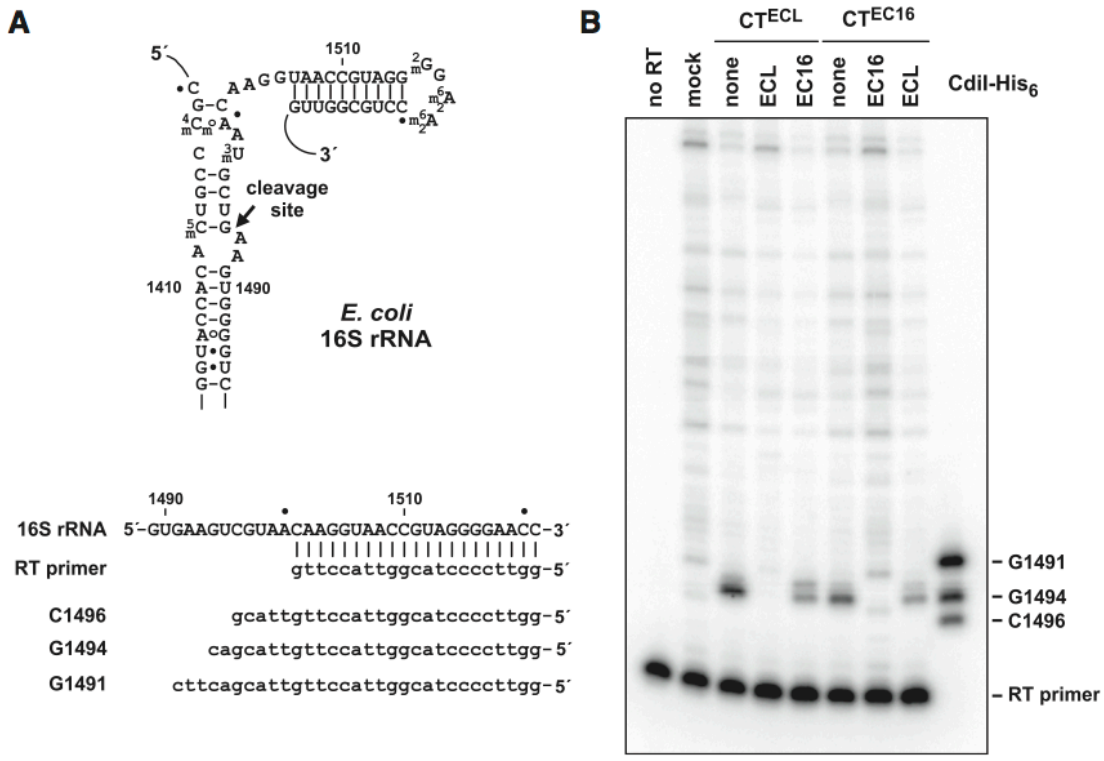

Figure 4.5. CdiA-CT^{ECL} and CdiA-CT^{EC16} cleave 16S rRNA between A1393 and G1394. (A) Nucleotide sequence and secondary structure of the 30S subunit decoding center. The sequence of the reverse transcription (RT) primer is shown in heteroduplex with its complementary sequence in 16S rRNA. Oligonucleotides C1496, G1494 and G1491 were used as gel-migration standards. The 16S rRNA cleavage site is indicated by an arrow. (B) Ribosomal RNA was extracted from the indicated *in vitro* nuclease reactions and hybridized to radiolabeled RT primer for primer extension analysis using reverse transcriptase. Reactions were resolved on a denaturing gel and visualized by phosphorimaging. The migration positions of oligonucleotides C1496, G1494 and G1491 are indicated.

Mutational analysis confirms active site residues identified in the CdiA-CT^{ECL} structure (Experiments performed by Hayes lab, UCSB)

The side-chains of CdiA-CT^{ECL} Asp203 and Lys214 overlay with active site residues Asp510 and Glu517 (respectively) of colicin E3 (Figures 4.2b & 4.3). However, because loop L4 is not well resolved in the CdiA-CT^{ECL} structure, it is difficult to unambiguously identify a catalytic residue corresponding to His513 of colicin E3. Therefore, we mutated CdiA-CT^{ECL} residues Asp203, Asp205, His207 and Lys214 individually to alanine and tested the resulting proteins for toxicity *in vivo* and 16S rRNase activity *in vitro*. CdiA-CT^{ECL} variants containing Asp203Ala, His207Ala or Lys214Ala mutations have no effect on *E. coli* cell growth (data not shown), suggesting that nuclease activity is disrupted. The Asp205Ala variant shows a delayed inhibition phenotype, in which cell growth is arrested ~90 min after toxin expression is induced (data not shown). Comparable results were obtained with *in vitro* reactions using purified toxin variants. CdiA-CT^{ECL} carrying the Asp203Ala, His207Ala and Lys214Ala mutations have no detectable rRNase activities *in vitro*, whereas the Asp205Ala variant exhibits lower activity than the wild-type enzyme (data not shown). Together, these experiments indicate that Asp203, His207 and Lys214 are required for toxin activity and could function in catalysis, whereas Asp205 plays an important yet non-essential role.

CdiA-CT^{ECL} is delivered into target bacteria during CDI (Experiments performed by Hayes lab, UCSB)

We next asked whether the *E. cloacae* CDI system is expressed and deployed for competition. We reasoned that *E. cloacae* mutants lacking the immunity gene should be susceptible to inhibition. When the CDI^{ECL} system is induced, the growth of target cells is suppressed approximately 20-fold compared to co-cultures with *E. cloacae* cells carrying the wild-type *cdi* locus (data not shown). Moreover, target cell growth is restored if they carry a plasmid-borne copy of the *cdiI*^{ECL} immunity gene, but the non-cognate *cdiI*^{EC16} gene provides no

protection (data not shown). We also tested the inducible *E. cloacae* inhibitor cells in co-cultures with *E. coli* target cells. Remarkably, *E. coli* cells are more sensitive to growth inhibition, with viable target cell counts reduced ~100-fold after four hrs of co-culture (data not shown). Moreover, *E. coli* targets are protected by plasmid-borne *cdiI*^{ECL}, but not *cdiI*^{EC16} (data not shown), confirming that growth inhibition is due to CdiA-CT^{ECL} toxin activity.

Many CdiA-CT toxins are modular and can be exchanged between different CdiA proteins to generate functional effector molecules (11, 19, 20, 25). To test whether the CdiA-CT/CdiI^{ECL} toxin/immunity protein complex is functional in the context of another CDI system, we replaced the *cdiA-CT/cdiI*^{EC93} region of the *E. coli* EC93 CDI system with the *E. cloacae* toxin/immunity coding sequences. This fusion produces a chimeric CdiA protein with CdiA-CT^{ECL} grafted onto CdiA^{EC93} at the VENN peptide motif. *E. coli* cells expressing the *cdiA*^{EC93}*-CT*^{ECL} chimera are potent inhibitors, capable of reducing viable *E. coli* target cells ~10^4-fold after three hrs of co-culture (data not shown). Again, target cells that carry the *cdiI*^{ECL} immunity gene are not inhibited and grow to the same level as cells cultured with mock-inhibitor cells that lack a CDI system (data not shown). We isolated total RNA from each competition co-culture and performed northern blot analysis to assay for RNase activity. Cleaved 16S rRNA is readily detected when the target cells lack immunity or express non-cognate *cdiI*^{EC16} immunity (data not shown). We also generated and tested inhibitor cells that express chimeric *cdiA*^{EC93}*-CT*^{ECL} that contains the His207Ala active-site mutation. Cells expressing the mutant effector do not inhibit *E. coli* targets, and no 16S rRNA cleavage is detected in the competition co-culture (data not shown). Together, these results demonstrate that the CdiA-CT^{ECL} toxin is delivered into target bacteria during CDI and that 16S RNase activity is solely responsible for growth inhibition.

Discussion

CdiA proteins carry a variety of sequence-diverse C-terminal domains, which represent a collection of distinct toxins. Determining the biochemical activities of so many different toxins remains an important problem in the field (11, 15). Aravind and colleagues have successfully used comparative sequence analyses to predict that many CdiA-CT toxins have nuclease activities (24). However, these predictions often do not identify specific nucleic acid substrates and may be inaccurate in some instances. In fact, the current annotation for CdiA-CT^{ECL} (Pfam PF15526, http://pfam.sanger.ac.uk/family/Toxin_46) suggests that this toxin adopts a BECR (Barnase-EndoU-colicin E5/RelE) protein fold and targets tRNA molecules for cleavage. The work presented here demonstrates that CdiA-CT^{ECL} is actually most similar to the C-terminal nuclease domain of colicin E3. Consistent with this structural homology, CdiA-CT^{ECL} is a site-specific 16S rRNase rather than a tRNase. Furthermore, we note that even accurate protein-fold predictions can lead to erroneous assignments of biochemical activity. For example, CdiA-CT^{Bp1026b} from *Burkholderia pseudomallei* 1026b has the same fold as type IIS restriction endonucleases, yet this toxin is a specific tRNase and has no detectable DNase activity (19, 25). These discrepancies between prediction and experimental characterization underscore the need for careful biochemical analysis to test sequence-based hypotheses.

The activity of colicin E3 was first described over 40 years ago (3, 4), yet a catalytic mechanism has only recently been proposed based on the structure of the enzyme bound to the ribosome (36). The mechanistic model postulates that Glu517 of colicin E3 acts as a general base to abstract a proton from the 2´-OH of 16S rRNA residue A1493. The resulting alkoxide subsequently attacks the phosphodiester linking A1493 and G1494 to cleave the 16S rRNA chain. The side-chain of His513 is thought to stabilize the transition state as well as donate a proton to the 5´-OH leaving group after cleavage. Colicin E3 residues Asp510 and Glu515 are within H-bonding distance of His513 and may promote protonation of its imidazole ring (36).

Comparative structure analysis suggests that residues Asp203, Asp205 and Lys214 of CdiA-CT^{ECL} are involved in catalysis because they are in the same relative positions as Asp510, His513 and Glu517 (respectively) of colicin E3. The superimposition of these residues is remarkable given that loop L4 of CdiA-CT^{ECL}, which contains the predicted active-site residues, is significantly longer and more flexible than the corresponding region in colicin E3 (Figure 4.2a). Mutagenesis experiments confirm that the CdiA-CT^{ECL} residues are important for nuclease activity, but it is not clear that the two enzymes share the same catalytic mechanism. For example, Lys214 in CdiA-CT^{ECL} is unlikely to function as a general base as proposed for Glu517 of ColE3-CT, especially as nearby residues Asp203 and Asp205 within CdiA-CT^{ECL} should favor protonation of Lys214. Lysine residues are often found in the active sites of nucleases and typically function to position the scissile phosphodiester or stabilize pentavalent transition states (43-45). Therefore, it seems likely that Lys214 serves one of the aforementioned functions, leaving His207 to act as the general base that initiates 16S rRNA cleavage. Though we have no structural information for CdiA-CT^{ECL} bound to the ribosome, the available data suggest that colicin E3 and CdiA-CT^{ECL} probably utilize distinct catalytic strategies.

The lack of sequence identity between CdiA-CT^{ECL} and ColE3-CT also raises questions about how the CDI toxin binds to the ribosome. Ramakrishnan, Kleanthous and their colleagues have shown that ColE3-CT loop L2 makes a number of specific contacts with the ribosome A site (36). Residues Arg495 and Gln489 bind the nucleobase and phosphate of 16S rRNA residue A1493, Lys496 interacts with C518, and Lys494 holds G530 in the *syn* conformation through a bridging water molecule (36). These toxin residues are highly conserved between colicins E3, E4, E6 and cloacin DF, suggesting that these enzymes all bind the ribosome in the same manner. By contrast, none of these loop L2 residues are shared with CdiA-CT^{ECL} (Figure 4.3). In fact, loop L2 of CdiA-CT^{ECL} is significantly displaced compared to ColE3-CT. This

displacement may result from the binding of CdiI^{ECL}, which would clash with loop L2 if it were in the ColE3-CT conformation. In the absence of CdiI^{ECL}, it is possible that loop L2 of CdiA-CT^{ECL} adopts the same conformation seen in ColE3-CT, but the sequence divergence suggests that each loop makes distinct contacts with the ribosome. ColE3-CT makes additional contacts with ribosomal protein S12 within the A site. Residues Tyr460 – Tyr464 form an intriguing pseudo-b-sheet interaction with the side-chains from His462, Asp463 and Tyr464 making specific H-bond contacts with S12 (36). Unfortunately, the corresponding region of CdiA-CT^{ECL} was degraded during crystallization, precluding a direct comparison of these structures. But again, the primary sequences in this region share no obvious homology, indicating that the two toxins probably interact with ribosomal protein S12 in distinct manners.

Finally, we note that there are several fundamental differences between CdiI^{ECL} and ImE3 immunity proteins. CdiI^{ECL} and ImE3 differ significantly in molecular mass, share less than 12% sequence identity, and bind to non-overlapping sites on their cognate nuclease domains. Moreover, each immunity protein has a distinct tertiary structure and fold. Despite these differences, each immunity protein is predicted to prevent its cognate toxin from entering the ribosome A site (Figure 4.6) (36), and therefore toxin inactivation is fundamentally the same for both systems. Structural homology searches reveal that CdiI^{ECL} is most similar to the Whirly family of single-stranded DNA-binding proteins. Although this homology is relatively weak (Z-scores 4.1 to 4.3 and rmsd ~4.0 Å), CdiI^{ECL} shares a characteristic topology with all Whirly proteins. The fact that the immunity proteins for colicin E3 and CdiA-CT^{ECL} toxins are unrelated in both primary sequence and tertiary structure suggests that these toxin-immunity pairs have independent origins. Because cognate toxin/immunity gene pairs are closely linked, they must presumably co-evolve as a unit. This process is thought to involve initial changes in the immunity protein, followed by compensatory mutations in the toxin that restore high-binding affinity between the two proteins (46, 47). In general, there are few constraints to impede the

drift of immunity genes, because they need only encode proteins that bind toxins. By contrast, toxins are often enzymes and must retain the ability to bind substrates and catalyze reactions. This model is largely supported by analyses showing that immunity proteins diverge more rapidly than toxins (15, 47). Thus, although it is formally possible that ImE3 and CdiI^{ECL} arose from a common ancestor, the differences in immunity protein folds make this model much less likely. Based on this reasoning, we speculate that CdiA-CT^{ECL}/CdiI^{ECL} and ColE3-CT/ImE3 evolved from different lineages, and that the structural and enzymatic similarities between the toxins reflect convergent evolution.

Acknowledgments

This work was supported by the National Institutes of Health through grants R21 AI099687 (C.S.H. and C.W.G.) and U01 GM102318 (D.A.L., C.S.H. and C.W.G). We thank the staff at both the Advanced Light Source (U.S. Department of Energy under Contract No. DE-AC02-05CH11231) at Berkeley National Laboratories and the Stanford Synchrotron Radiation Lightsource (supported in part by P41 GM103393) for their invaluable help in data collection. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

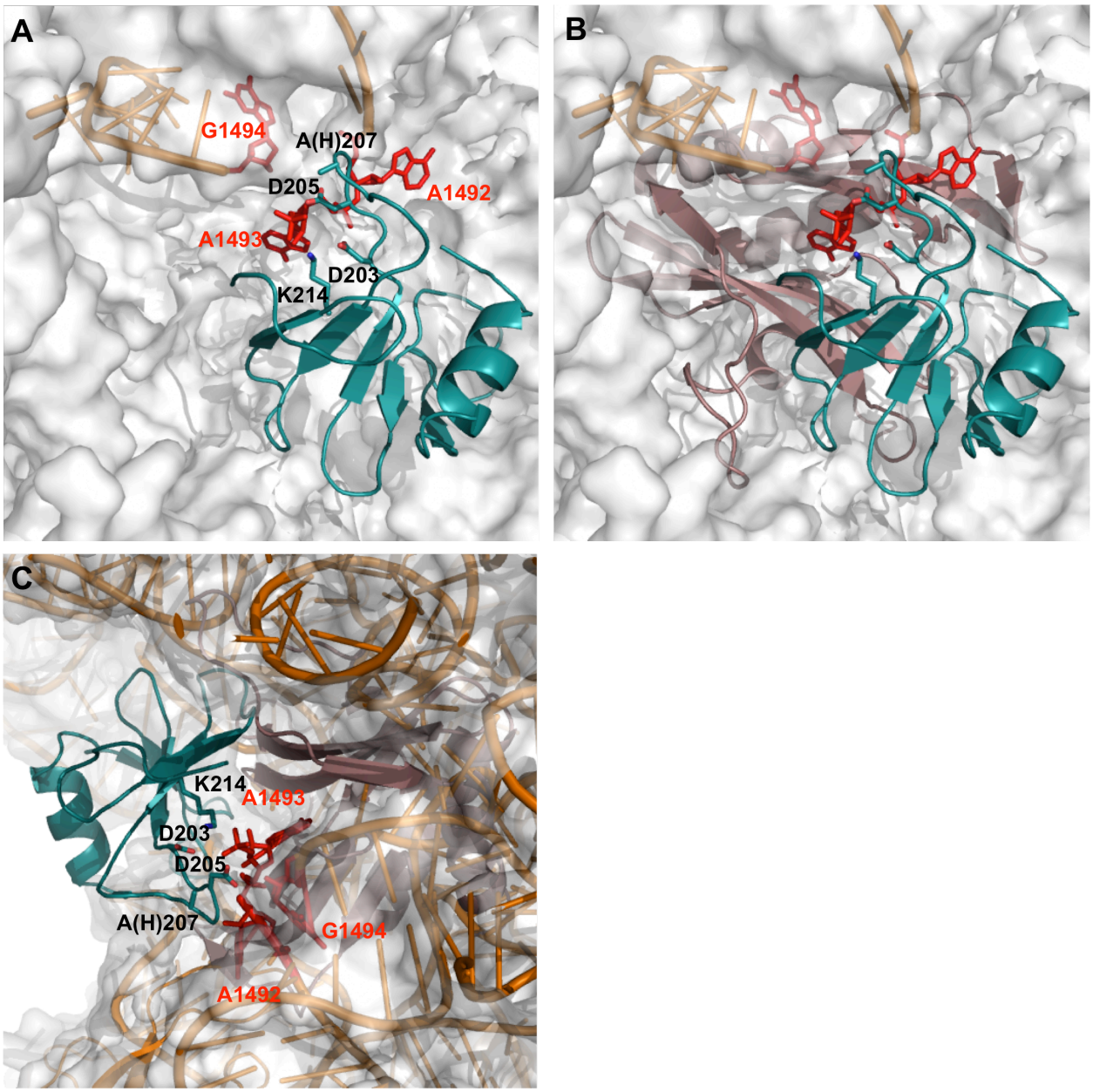

Figure 4.6. Modeling of CdiA-CT^{ECL} on the ribosome. (A) Superimposition of CdiA-CTECL (teal ribbon) onto the structure of ColE3-CT bound to the ribosome in the post-cleavage state (PDB 2XFZ). The surface of the ribosome is rendered in grey, and the cleaved 16S rRNA strand is shown in orange ladder representation, with residues A1492 – G1494 rendered as red sticks. The ColE3-CT domain has been removed for clarity. Putative CdiA-CTECL active site residues are shown in stick representation with nitrogen, oxygen and phosphate atoms colored blue, red and orange, respectively. (B) CdiI^{ECL} (salmon pink) is superimposed upon the model to illustrate steric clashes that should prevent the complex from binding the ribosome A site. (C) Side view of CdiA-CT/CdiI^{ECL} as depicted in B showing CdiI^{ECL} buried within the rRNA.

References

1. Cascales E*, et al.* (2007) Colicin biology. *Microbiology and molecular biology reviews : MMBR* 71(1):158-229.
2. Schaller K & Nomura M (1976) Colicin E2 is DNA endonuclease. *Proc Natl Acad Sci U S A* 73(11):3989-3993.
3. Bowman CM, Dahlberg JE, Ikemura T, Konisky J, & Nomura M (1971) Specific inactivation of 16S ribosomal RNA induced by colicin E3 in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 68(5):964-968.
4. Senior BW & Holland IB (1971) Effect of colicin E3 upon the 30S ribosomal subunit of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 68(5):959-963.
5. Ogawa T*, et al.* (1999) A cytotoxic ribonuclease targeting specific transfer RNA anticodons. *Science* 283(5410):2097-2100.
6. Gordon DM, Riley MA, & Pinou T (1998) Temporal changes in the frequency of colicinogeny in *Escherichia coli* from house mice. *Microbiology* 144 (Pt 8):2233-2240.
7. Riley MA & Gordon DM (1992) A survey of Col plasmids in natural isolates of *Escherichia coli* and an investigation into the stability of Col-plasmid lineages. *J Gen Microbiol* 138(7):1345-1352.
8. Chao L & Levin BR (1981) Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proc Natl Acad Sci U S A* 78(10):6324-6328.
9. Czaran TL, Hoekstra RF, & Pagie L (2002) Chemical warfare between microbes promotes biodiversity. *Proc Natl Acad Sci U S A* 99(2):786-790.
10. Aoki SK*, et al.* (2005) Contact-dependent inhibition of growth in Escherichia coli. *Science* 309(5738):1245-1248.
11. Aoki SK*, et al.* (2010) A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* 468(7322):439-442.
12. Hood RD*, et al.* (2010) A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* 7(1):25-37.
13. MacIntyre DL, Miyata ST, Kitaoka M, & Pukatzki S (2010) The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. *Proc Natl Acad Sci U S A* 107(45):19520-19524.
14. Zheng J, Ho B, & Mekalanos JJ (2011) Genetic analysis of anti-amoebae and anti-bacterial activities of the type VI secretion system in *Vibrio cholerae*. *PLoS ONE* 6(8):e23876.
15. Ruhe ZC, Low DA, & Hayes CS (2013) Bacterial contact-dependent growth inhibition. *Trends in microbiology* 21(5):230-237.
16. Silverman JM, Brunet YR, Cascales E, & Mougous JD (2012) Structure and regulation of the type VI secretion system. *Annu Rev Microbiol* 66:453-472.
17. Aoki SK*, et al.* (2008) Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Mol Microbiol* 70(2):323-340.
18. Ruhe ZC, Wallace AB, Low DA, & Hayes CS (2013) Receptor polymorphism restricts contact-dependent growth inhibition to members of the same species. *mBio* 4(4):e00480-00413.
19. Morse RP*, et al.* (2012) Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems. *Proceedings of the National Academy of Sciences of the United States of America* 109(52):21480-21485.
20. Webb JS*, et al.* (2013) Delivery of CdiA nuclease toxins into target cells during contact-dependent growth inhibition. *PLoS One* 8(2):e57609.

21. Poole SJ*, et al.* (2011) Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. *PLoS Genet* 7(8):e1002217.
22. Zhang D, Iyer LM, & Aravind L (2011) A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic acids research* 39(11):4532-4552.
23. Holberger LE, Garza-Sanchez F, Lamoureux J, Low DA, & Hayes CS (2012) A novel family of toxin/antitoxin proteins in Bacillus species. *FEBS Lett* 586(2):132-136.
24. Zhang D, de Souza RF, Anantharaman V, Iyer LM, & Aravind L (2012) Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct* 7:18.
25. Nikolakakis K*, et al.* (2012) The toxin/immunity network of Burkholderia pseudomallei contact-dependent growth inhibition (CDI) systems. *Molecular microbiology* 84(3):516-529.
26. Walker D, Lancaster L, James R, & Kleanthous C (2004) Identification of the catalytic motif of the microbial ribosome inactivating cytotoxin colicin E3. *Protein Sci* 13(6):1603-1611.
27. Beck CM*, et al.* (2014) CdiA from Enterobacter cloacae Delivers a Toxic Ribosomal RNase into Target Bacteria. *Structure* 22.
28. Hayes CS, Bose B, & Sauer RT (2002) Proline residues at the C terminus of nascent chains induce SsrA tagging during translation termination. *The Journal of biological chemistry* 277(37):33825-33832.
29. Koskiniemi S*, et al.* (2013) Rhs proteins from diverse bacteria mediate intercellular competition. *Proc Natl Acad Sci U S A* 110(17):7032-7037.
30. Perez A*, et al.* (2007) Cloning, nucleotide sequencing, and analysis of the AcrAB-TolC efflux pump of *Enterobacter cloacae* and determination of its involvement in antibiotic resistance in a clinical isolate. *Antimicrob Agents Chemother* 51(9):3247-3253.
31. Diner EJ, Beck CM, Webb JS, Low DA, & Hayes CS (2012) Identification of a target cell permissive factor required for contact-dependent growth inhibition (CDI). *Genes Dev* 26(5):515-525.
32. Goulding CW & Perry LJ (2003) Protein production in Escherichia coli for structural studies by X-ray crystallography. *Journal of structural biology* 142(1):133-143.
33. Diner EJ & Hayes CS (2009) Recombineering reveals a diverse collection of ribosomal proteins L4 and L22 that confer resistance to macrolide antibiotics. *J Mol Biol* 386(2):300-315.
34. Holm L & Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic acids research* 38(Web Server issue):W545-549.
35. Lancaster LE, Savelsbergh A, Kleanthous C, Wintermeyer W, & Rodnina MV (2008) Colicin E3 cleavage of 16S rRNA impairs decoding and accelerates tRNA translocation on *Escherichia coli* ribosomes. *Mol Microbiol* 69(2):390-401.
36. Ng CL*, et al.* (2010) Structural basis for 16S ribosomal RNA cleavage by the cytotoxic domain of colicin E3. *Nat Struct Mol Biol* 17(10):1241-1246.
37. Soelaiman S, Jakes K, Wu N, Li C, & Shoham M (2001) Crystal structure of colicin E3: implications for cell entry and ribosome inactivation. *Molecular cell* 8(5):1053-1062.
38. Desveaux D, Marechal A, & Brisson N (2005) Whirly transcription factors: defense gene regulation and beyond. *Trends Plant Sci* 10(2):95-102.
39. Carr S, Walker D, James R, Kleanthous C, & Hemmings AM (2000) Inhibition of a ribosome-inactivating ribonuclease: the crystal structure of the cytotoxic domain of colicin E3 in complex with its immunity protein. *Structure* 8(9):949-960.
40. Rojas CM, Ham JH, Deng WL, Doyle JJ, & Collmer A (2002) HecA, a member of a class of adhesins produced by diverse pathogenic bacteria, contributes to the attachment,

aggregation, epidermal cell killing, and virulence phenotypes of *Erwinia chrysanthemi* EC16 on *Nicotiana clevelandii* seedlings. *Proc Natl Acad Sci U S A* 99(20):13142-13147.
41. Rojas CM*, et al.* (2004) The *Erwinia chrysanthemi* EC16 *hrp/hrc* gene cluster encodes an active Hrp type III secretion system that is flanked by virulence genes functionally unrelated to the Hrp system. *Mol Plant Microbe Interact* 17(6):644-653.
42. Basturea GN, Rudd KE, & Deutscher MP (2006) Identification and characterization of RsmE, the founding member of a new RNA base methyltransferase family. *RNA* 12(3):426-434.
43. Richardson RM, Pares X, & Cuchillo CM (1990) Chemical modification by pyridoxal 5'-phosphate and cyclohexane-1,2-dione indicates that Lys-7 and Arg-10 are involved in the p2 phosphate-binding subsite of bovine pancreatic ribonuclease A. *Biochem J* 267(3):593-599.
44. Gite S, Reddy G, & Shankar V (1992) Active-site characterization of S1 nuclease. I. Affinity purification and influence of amino-group modification. *Biochem J* 285 (Pt 2):489-494.
45. Pingoud A & Jeltsch A (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res* 29(18):3705-3727.
46. Riley MA (1993) Positive selection for colicin diversity in bacteria. *Mol Biol Evol* 10(5):1048-1059.
47. Tan Y & Riley MA (1997) Positive selection and recombination: major molecular mechanisms in colicin diversification. *Trends Ecol Evol* 12(9):348-351.

CHAPTER 5

Crystal structures of uropathogenic *E. coli* CdiA-CT in complex with CysK, and an *E. coli* CdiA domain

Summary

Contact-dependent growth inhibition (CDI) is a newly discovered mechanism of interbacterial competition. In CDI, large cell-surface effector proteins (CdiA) contact specific membrane receptors on target cells, initiating cleavage and translocation of the toxic C-terminal domain (CdiA-CT). Previous structural characterization has focused on complexes formed between CdiA-CTs and their cognate CdiI immunity proteins, which protect target cells from growth inhibition. Here we report the crystal structure of the CdiA-CT toxin from uropathogenic *E. coli* strain 536 in complex with the biosynthetic enzyme CysK, which activates the toxin's tRNase activity. The complex formed mimics the interaction found in the cysteine synthase complex. Additionally, we report the structure of a domain from the large, non-toxin region of *E. coli* CdiA. This represents the first structural characterization of this region of CdiA proteins. Together, these structures offer a starting point towards further understanding of CDI.

Introduction

Contact-dependent growth inhibition (CDI) is a newly discovered method of inter-bacterial competition that is widely distributed in gram-negative bacteria, including important human pathogens (1, 2). CDI utilizes a CdiA/CdiB two-partner secretion system by which CdiB, the bacterial outer membrane, facilitates export of CdiA proteins to the cell surface. CdiA proteins are large (250-600 kDa) hemagglutinin (HA) repeat containing proteins, and are thought to form extended "stick-like" structures that protrude from the cell, facilitating delivery of the CdiA-derived toxin to a target cell. (3). In the current model, toxic C-terminal CdiA domains (CdiA-CT, 25-50kDa) are cleaved from CdiA upon contact with a specific target cell membrane receptor, and then translocated into the cytosol of the target bacterium (4-6). To prevent autoinhibition upon CdiA-CT delivery to "self" cells, CDI systems encode small immunity proteins, CdiI, which bind tightly to their cognate CdiA-CT domains (1). While the process of CdiA-CT cleavage and translocation is currently not well understood, it is clear that the non-toxin region of CdiA plays a role in cell-to-cell contact and target specificity. Recent studies utilizing the CDI system from intestinal isolate *E. coli* EC93 revealed that $CdiA^{EC93}$ obtains target specificity by recognizing variable loops on the highly conserved membrane protein BamA (5). Additionally, $CdiA^{EC93}$ is capable of delivering heterologous CdiA-CTs to target cells, suggesting that $CdiA^{EC93}$ recognition of BamA is independent of its toxin domain (1). BamA does not appear to be a ubiquitous membrane receptor for all CdiA proteins, as CdiA from uropathogenic *E. coli* strain 536 (UPEC536) recognize a different membrane protein (unpublished data from Hayes lab, UCSB). The precise mechanism of receptor recognition by CdiA proteins is currently unknown.

The toxic functions of CdiA-CT toxins are diverse and include dissipating the proton motive force, degrading DNA, and cleaving various forms of RNA (1, 7-9). Toxin diversity is especially intriguing in the unique case of CdiA-CT from uropathogenic *E. coli* strain 536

(UPEC536). The CdiA-CT^{UPEC536} toxin is a tRNase capable of cleaving the anticodon loop of tRNA^{Arg}, however it is inactive unless bound to the biosynthetic enzyme CysK (*O*-acetylserine sulfhydrylase A) (10). CysK is one of two sulfhydrolase enzymes (the other being CysM, or *O*-acetylserine sulfhydrylase B) that catalyze the final step of the synthesis of cysteine from serine. Expression of CysK in the target cell cytosol to is required to activate the CdiA-CT^{UPEC536} toxin. Remarkably, CdiA-CT^{UPEC536} and CysK form a stable complex, and the interaction appears to mimic the interaction observed in the cysteine synthase complex (CSC), formed between CysK and CysE (serine *O*-acetyltransferase) (11-15). The CSC is formed by the insertion of the C-terminal tail, GDGI, of CysE into the active site cleft of CysK (16-19). CdiA-CT^{UPEC536} contains a similar C-terminal tail, GYGI, which has been shown to be required for CdiA-CT^{UPEC536} binding to CysK, however it is unclear how this interaction activates CdiA-CT^{UPEC536} (10). Furthermore, there are currently no structures of the CSC in the protein data bank (PDB). The evolutionary benefit of CysK-mediated toxin activation is not immediately obvious, however it is clear that CDI systems have evolved interesting mechanisms to promote toxic activity.

Our previous structural studies of CDI have focused on cognate CdiA-CT/CdiI complexes. These studies offered significant insight into both CdiA-CT activity and the mechanism of immunity offered by cognate CdiI. In an effort to gain additional structural understanding into the CDI system, and in particular, the "stick region" of CdiA, the aims of this chapter is to explore further aspects of the CDI system and structurally characterize: **1)** the complex of the CdiA-CT^{UPEC536} toxin and its permissive factor, CysK, and **2)** a segment of the large, non-toxin, stick region of CdiA. Herein we report the crystal structure of CdiA-CT^{UPEC536} in complex with CysK. As predicted, the C-terminal tail of CdiA-CT^{UPEC536} mediates the complex interface. Further, we have solved the crystal structure of a segment of the *E. coli* EC93 CdiA stick domain encompassing residues 2681-2909 (CdiA$^{2681-2909}$). The sequence of this domain ends with the VENN motif that marks the start of the C-terminal CdiA-CT^{EC93} toxin domain. This

structure offers the first glimpse of a portion of the CdiA stick domain. Taken together, these new structures offer a starting point towards additional research and understanding of the CDI system.

Materials and Methods

Expression vectors for His_6-CdiA-$CT^{UPEC536}$/CysK and EC93 His_6-$CdiA^{2681-2909}$

The pET21S plasmids for over-expression of His_6-CdiA-$CT^{UPEC536}$ containing a H178A missense mutation and His_6-CysK have been previously described (10). To produce non-tagged CysK, the above His_6-CysK expressing plasmid was modified by site-directed mutagenesis (SDM) to generate a stop codon prior to the C-terminal His_6-tag (Table 5.1). Following the PCR, the SDM product was digested with *Dpn*1 overnight, and then transformed into TOP10 *E. coli* cells. Generation of the stop codon was confirmed by DNA sequencing (Laguna). The pET21 EC93 His_6-$CdiA^{Q2681-A2909}$ over-expression plasmid was designed and constructed in the laboratory of Christopher Hayes (UC Santa Barbara). This plasmid over-expresses an N-terminally His_6-tagged *E. coli* EC93 CdiA soluble domain spanning amino acids Q2681-A2909.

Table 5.1. Primers for creation of stop codon in CysK expression plasmid

Primer	Sequence
CysK-stop For	GAATTGCAACAGACTAGTTGAGAGCACCACCACCACCAC
CysK-stop Rev	GTGGTGGTGGTGGTGCTCTCAACTAGTCTGTTGCAATTC

Protein expression and purification for crystallography

E. coli CysK (no His_6 tag) and His_6-CdiA-$CT^{UPEC536}$ and were over-expressed separately in *E. coli* BL21 (DE3) Gold cells (Strategene). Cells were grown aerobically at 37 °C in LB medium containing 50 µg/mL ampicillin. Expression of each protein was induced by the addition of 1 mM isopropyl-β-D-thiogalactosidase at an $OD_{600} \approx 0.8$ and grown for a further 3-4 h before harvesting. Cells were collected by centrifugation at 5,500 × g for 25 min and then washed with resuspension buffer (20 mM Sodium Phosphate, pH 7.0, 150 mM NaCl). Pellets of

cells containing overexpressed His_6-CdiA-$CT^{UPEC536}$ and CysK were combined, and this cell mixture was lysed on ice by sonication in resuspension buffer containing 10 mg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride (PMSF). Unbroken cells and debris were removed by centrifugation at 18,000 × g for 30 min followed by filtration through a 1.0-μm filter. Clarified lysates were loaded onto a Ni^{2+}-charged HiTrap column (5 mL; GE Healthcare) and washed with resuspension buffer supplemented with 10 mM imidazole. Proteins were eluted with a linear gradient of imidazole (10-250 mM) in resuspension buffer. Fractions were collected, combined, and concentrated to a volume of ~500 μL using a 10-kDa centrifugal concentrator (Centricon; Millipore). Proteins were further purified by gel filtration on a Superdex 200 column equilibrated with 20 mM Sodium Phosphate, pH 7.0 and 150 mM NaCl using an AKTA FPLC. The purified CdiA-$CT^{UPEC536}$/CysK complex was concentrated to 20 mg/mL for crystallization trials.

The selenomethionine-derivatized His_6-$CdiA^{2681-2909}$ domain ($CdiA^{2681-2909}$) was grown in M9 minimal medium supplemented with amino acid supplements (leucine, isoleucine, and valine at 50 mg/L; phenylalanine, lysine, and threonine at 100 mg/L; and selenomethionine at 75 mg/L) as described (20). Following over-expression in *E. coli* BL21 (DE3) Gold cells (Strategene), purification proceeded similar to as described for CdiA-$CT^{UPEC536}$/CysK complex with a few changes. First, all buffers contained 20 mM Tris, pH 7.4, instead of sodium phosphate buffer. During the Ni^{2+}-affinity purification, an imidazole linear gradient of 10-500 mM was used. Lastly, gel filtration was performed on a Superdex 75 column. $CdiA^{2681-2909}$ was concentrated to 85 mg/mL for crystallization trials.

Crystallization, Data Collection, Structure Determination, and Refinement

CdiA-$CT^{UPEC536}$/CysK complex crystals were grown over the period of 3 weeks by hanging drop vapor-diffusion with a reservoir containing 0.2M Na Sulfate, 0.1M Bis Tris

Propane, pH 7.9, and 20% (wt/vol) PEG 3350, and a 20 mg/mL protein solution. The hanging drop contained a 1:1 ratio of protein:reservoir. The protein complex crystallized in space group $P4_1$ with unit cell dimensions 64.01 Å × 64.01 × Å 365.37 Å, and two CdiA-$CT^{UPEC536}$/CysK complexes per asymmetric unit. Crystals were mounted and collected under cryoconditions with the addition of 40% (vol/vol) glycerol as cryoprotectant to the reservoir condition. A data set was collected at 70K with wavelength 1.0 Å on SSRL beamline 7-1. Images were indexed, integrated, and reduced using iMOSFLM (21). The initial phases were determined by molecular replacement by autoMR in PHENIX using the structure of CysK as a search model (pdb ID 1OAS) (22). The initial model built by Autobuild in PHENIX only contained the two CysK molecules; therefore the CdiA-$CT^{UPEC536}$ molecules were built through iterative manual building in Coot, followed by subsequent Autobuild cycles (23, 24). The final model was refined with phenix.refine. The final model contains CdiA-$CT^{UPEC536}$ and CysK residues 127-227 and 2-314, respectively, and 110 water molecules resulting in an R_{work}/R_{free} (%) of 20.0/22.4, with 97.1% and 2.9% of residues in the favorable allowed, and allowed regions, respectively. Additionally, each CysK molecule in the asymmetric unit contains a pyridoxal 5′-phosphate (PLP) bound in a Schiff base linkage to Lys42. Due to lack of observable electron density, CdiA-$CT^{UPEC536}$ loop L1 residues 168, 170-172, and 175 were modeled as alanines. Data collection and refinement statistics are presented in table 5.2. All molecular graphics were prepared with PyMOL (25).

The selenomethionine-derivatized $CdiA^{2681-2909}$ crystals were grown over the period of two months by hanging-drop vapor diffusion with a reservoir containing 0.1M MIB buffer, pH 6.7, 25% PEG 1500, and a 85 mg/mL protein solution. The hanging drop contained a 1:2 ratio of protein:reservoir. $CdiA^{2681-2909}$ crystallized in space group $P3_121$ with unit cell dimensions 90.14 Å × 90.14 × Å 107.80 Å, and two molecules per asymmetric unit. Crystals were mounted and collected under cryoconditions with the addition of 40% (vol/vol) glycerol as cryoprotectant to the reservoir condition. A Se-single wavelength anomalous dispersion (SAD) dataset was collected at 70K at the Se-absorption edge (.979 Å) on beamline 7-1 at SSRL. Data reduction

Table 5.2. X-ray diffraction data and refinement statistics

	CdiA-CT^{UPEC536}/CysK	**CdiA$^{2681-2909}$**
Space Group	$P4_1$	$P3_121$
Unit cell dimensions (Å)	64.01 × 64.01 × 365.37	90.14 × 90.14 × 107.79
pH of crystallization condition	7.9	6.7
Protein concentration (mg/mL)	20	85
Data set		
Wavelength (Å)	1.0	0.979
Resolution range	64.01-2.70	53.90-2.7
Unique reflections (total)	39863 (303136)	14362 (140476)
Completeness (%)*	99.4 (99.2)	99.9 (100.0)
Redundancy *	12.8 (13.4)	9.8 (10.1)
R_{merge}*,†	0.163 (0.759)	0.077 (0.521)
I/σ*	9.9 (3.0)	20.8 (4.6)
NCS copies	2	2
No. of Selenium sites/a.u	-	18
FOM	-	0.408
Other ions	-	2 BO_3^{3-}
Model refinement		
Resolution range (Å)	44.92-2.70	44.35-2.7
No. of reflections (working/free)	39794 (3974)	14334 (1428)
No. of protein atoms	6216	2315
No. of water molecules	110	90
Missing residues	CdiA-CT: 1-126 CysK: 1, 316-326	1-31, 109-110, 154-163 210-231
R_{work}/R_{free}‡, %	20.0/22.4	20.6/24.8
R.m.s deviations		
Bond lengths (Å)	0.01	0.009
Bond angles (degrees)	1.20	1.30
Ramachandran Plot		
Most favorable region (%)	97.1	98.38
Additional allowed region (%)	2.9	1.62
Disallowed region	0	0
PDB ID Code		

*Statistics for the highest resolution shell are given in (brackets)
†$R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma I$
‡$R_{work}=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$ R_{free} was computed identically except where all reflections belong to a test set of 10% randomly selected data.

was carried out in iMOSFLM, resulting in significant anomalous differences up to 3.5 Å. Experimental phasing and initial model building was performed using AutoSol and Autobuild in PHENIX, in which 18 Se atoms were located per asymmetric unit (26). The final model was built through iterative manual building in Coot and refined with phenix.refine. The final model contains CdiA$^{2681-2909}$ residues 33-108, 111-153, and 164-209 (numbered from residue 2681), 97 water molecules and two borate molecules resulting in an R_{work}/R_{free} (%) of 21.0/23.6, with 98.38% and 1.62% of residues in the favorable allowed, and allowed regions, respectively.

Residues 1-32, 109-110, 154-163, and 210-233 were omitted due to lack of observable electron density. Residues 111-118 were also omitted in chain A only. Additionally, residues Q36, K84, N107, N119, and E136 of chain A, and residues 32-35, N107, Q113, V114, N119, D133, and Q189 of chain B were modeled as alanines due to lack of observable electron density. Data collection and refinement statistics are presented in Table 5.2.

Results

Structure of CdiA-CT^{UPEC536}/CysK

To characterize the interaction between CdiA-CT^{UPEC536} and CysK, we solved the X-ray crystal structure of the CdiA-CT^{UPEC536}/CysK complex to 2.7 Å resolution by molecular replacement (MR), using the structure of *Salmonella typhimurium* CysK as a search model (pdb ID 1OAS) (22). The crystal space group was $P4_1$, with one heterotetramer comprised of two CdiA-CT ^{UPEC536}/CysK complexes per asymmetric unit. The final model contains CdiA-CT^{UPEC536} residues 127-227 (numbered from Val1 of the VENN motif) and CysK residues 2-315. CdiA-CT^{UPEC536} contains an H178A missense mutation (His3193 in full length CdiA), which has previously been shown to abolish the tRNase activity of this toxin (10). Similar to what we observed in our previous structures of CdiA-CT/CdiI complexes, the structural model of CdiA-CT^{UPEC536} only contains the putative C-terminal catalytic domain, as the N-terminal regions of CdiA-CTs are predicted to have little structure and we propose are susceptible to degradation (8, 27). The active site of each CysK molecule contains pyridoxal 5'-phosphate (PLP) bound in a Schiff base linkage to Lys42. In addition, 110 waters are included in the final model resulting in an R_{work}/R_{free} (%) of 20.0/22.4 (Table 5.2).

The structural model of CdiA-CT^{UPEC536}/CysK is a heterotetramer, with the interaction of two CdiA-CT^{UPEC536}/CysK complexes mediated entirely by a CysK-CysK dimer interface (Figure 5.1a), which is consistent with previous studies that have show CysK is a dimer (28). Multiple

CysK structures have been previously reported in the literature, but our model represents the first *E.coli* CysK structure. The first reported CysK structure, from *Salmonella typhimurium* (PDB ID: 1OAS), superimposes upon *E. coli* CysK model with an rmsd of 0.477 Å over all α-carbons, illustrating that there is little structural difference between the homologs (22). The structure of CdiA-CT^{UPEC536} consists of a single domain containing four α-helices (α1-α4) (Figure 5.1b). Notably, there is a long flexible loop connecting helices α2 and α3 (L1) consisting of residues 165-177. This loop is directly adjacent to the known catalytic residue H178 (A178 in the structure), suggesting that it may be important in CdiA-CT^{UPEC536} activity (Figure 5.1b). As expected based on the homology between the C-terminal residues of CysE and CdiA-CT^{UPEC536}, the CdiA- CT^{UPEC536}/CysK complex interface is mediated by the interaction of the GYGI tail of CdiA-CT^{UPEC536} with the CysK active site (Figure 5.1c). This interaction consists of a network of four hydrogen bonds (H-bonds) between CdiA-CT^{UPEC536} GYGI backbone atoms and CysK, as well as distinct shape complementarity between the two proteins (Figures 5.1d & 5.1e). There are seven additional H-bonds between CdiA-CT^{UPEC536} and CysK in regions outside of the GYGI tail that presumably aid in complex stabilization (Table 5.3). CdiA-CT^{UPEC536} residues within helices α2- α4 interact with CysK, including an ion pair between CdiA-CT^{UPEC536} Gln183 and CysK Asp207. The CdiA-CT^{UPEC536}/CysK complex has an interface 1280 Å^2, burying 9.2 and 19.4% of the solvent accessible surface areas of CdiA-CT^{UPEC536} and CysK, respectively.

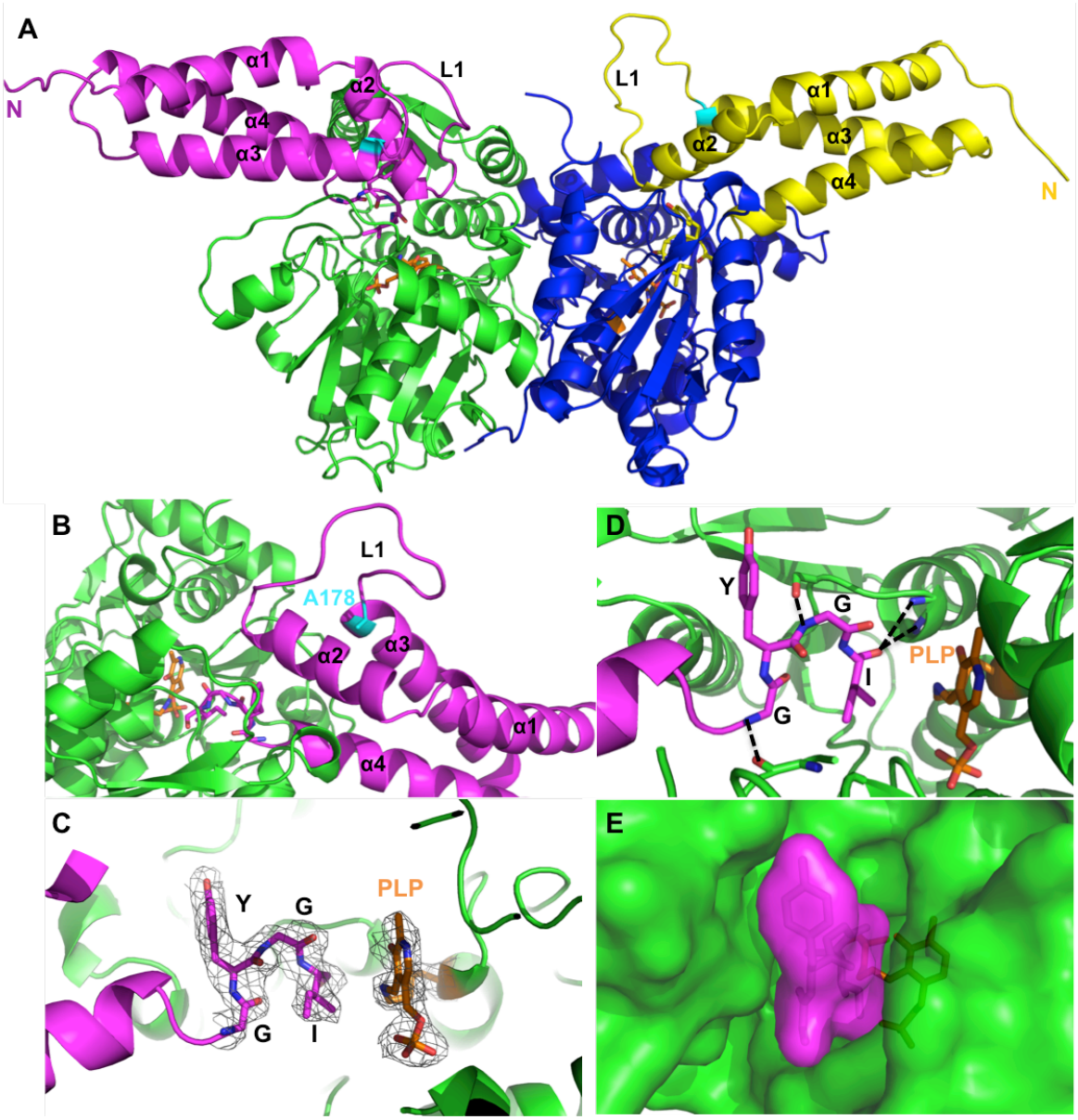

Figure 5.1. The structure of the CdiA-CT^{UPEC536}/CysK complex. (A) Ribbon representation of the heterotetromeric CdiA-CT^{UPEC536}/CysK complex, with CysK molecules colored green and blue, and CdiA-CT^{UPEC536} molecules colored magenta and yellow. CdiA-CT^{UPEC536} secondary structure elements and proposed active site loops (L1) are labeled, and the H178A active site mutation is colored cyan. GYGI C-terminal peptides and PLP are represented as sticks, with oxygen, nitrogen, and PLP carbons colored red, blue, and orange, respectively. (B) Ribbon representation of a single complex of CdiA-CT^{UPEC536} (magenta) and CysK (green) highlighting the H178A active site mutation and the proposed active site loop. (C) The CdiA-CT^{UPEC536} C-terminal GYGI peptide inserts into the CysK active site, adjacent to PLP. $2F_o$-F_c electron density map of the GYGI peptide and PLP is shown in grey mesh, contoured at 1.2σ (D) H-bonds between the GYGI peptide and CysK are indicated with dashed black lines. (E) Surface representation of the GYGI peptide and CysK. The transparent surface reveals stick representation of the GYGI peptide and PLP.

Table 5.3. Hydrogen bond and ion pairs between CdiA-CT^{UPEC536} and CysK

CdiA-CT^{UPEC536}	CysK	Distance (Å)
LYS 161[NZ]	THR 95[O]	3.09
GLN 183[NE2]	PRO 224[O]	3.20
GLN 183[NE2]	ASP 207[OD2]	3.35
ARG 190[NH2]	PRO 222[O]	2.95
SER 220[O]	LYS 121[N]	3.17
ALA 221[O]	MET 120[N]	2.90
GLY 224[N]	ALA 232[O]	2.77
GLY 226[N]	SER 70[OG]	2.64
ILE 227[O]	THR 73[N]	3.42
ILE 227[O]	ASN 72[N]	3.71

Presently, it still remains unclear how the interaction of CdiA-CT^{UPEC536} with CysK activates the toxin. CdiA-CT^{UPEC536} shares no sequence or structural homology with previously solved CdiA-CT structures, including CdiA-CT^{1026b} (from *Burkholderia pseudomallei* 1026b), which also has tRNase activity (8). A search utilizing DALI for structural homologs of CdiA-CT^{UPEC536} revealed no hits to endonucleases (29), however, a wide range of proteins with functions unrelated to endonucleases have similar α-helical folds to the small toxin. It is conceivable that CdiA-CT^{UPEC536} could represent a novel fold for endonucleases, or that some residues of CysK may contribute to tRNase activity. Both possibilities warrant further investigation to identify the role of CysK in the activation of CdiA-CT^{UPEC536} tRNase activity.

Structure of CdiA$^{Q2681\text{-}A2909}$

In CDI, the large, non-toxin region of CdiA proteins likely plays a role in cell-to-cell contact and receptor recognition. However, outside of the N-terminal HA repeat region, the N-terminal stick-like regions of CdiA proteins share no obvious sequence homology with any previously characterized proteins, thus both their complex structure and perhaps function have largely remained a mystery. We have undertaken a structural-piecewise approach to glean

information about the structure and role of the stick region of CdiA. To this end, we have solved the structure of *E.coli* EC93 CdiA$^{2681\text{-}2909}$, a predicted domain directly preceding CdiA-CT^{EC93}. CdiA$^{2681\text{-}2909}$ was solved to 2.7 Å by selenium single wavelength anomalous dispersion (Se-SAD). CdiA$^{2681\text{-}2909}$ crystallized in space group $P3_121$ and contained two molecules per asymmetric unit. The final model contains CdiA$^{2681\text{-}2909}$ residues 33-108, 111-153, and 164-209 (numbered from residue 2681). Residues 1-32, 109-110, 154-163, and 210-231 were excluded from the final model due to lack of observable electron density. In addition, 97 waters and two borate molecules are included in the final model resulting in an R_{work}/R_{free} (%) of 20.6/24.8 (Table 5.2).

The structural model of each CdiA$^{2681\text{-}2909}$ monomer comprises a single, all α-helical domain (α1-α7). However, due to lack of observable electron density in loop regions, CdiA$^{2681\text{-}2909}$ is modeled as three non-continuous polypeptide chains (α1-α3, α4-α5, α6-α7) (Figure 5.2). The structure consists of three long α-helices (α1, α3, and α7), which are decorated by four shorter α-helices (α2, α4-6). Within the α3 helix, there is a small four-residue loop of sequence GTGS, which results in a discontinuous α3 helix (α3a and α3b). Contrary to what was expected, the N and C termini of CdiA$^{2681\text{-}2909}$ lie on the same face of this domain. This suggests that this particular CdiA domain may form a “bulge” along the predicted extended structure of CdiA, as depicted in Figure 5.5. Since this domain directly precedes the toxic CdiA-CT^{EC93} domain, it could also play a role in cleavage of the toxin from CdiA. Unfortunately, structural homology searches using DALI revealed no strong hits, with all alignments over less than 50% of the CdiA$^{2681\text{-}2909}$ structure. Based solely on the structure reported, the function of the CdiA$^{2681\text{-}2909}$ domain in the CDI system remains unclear.

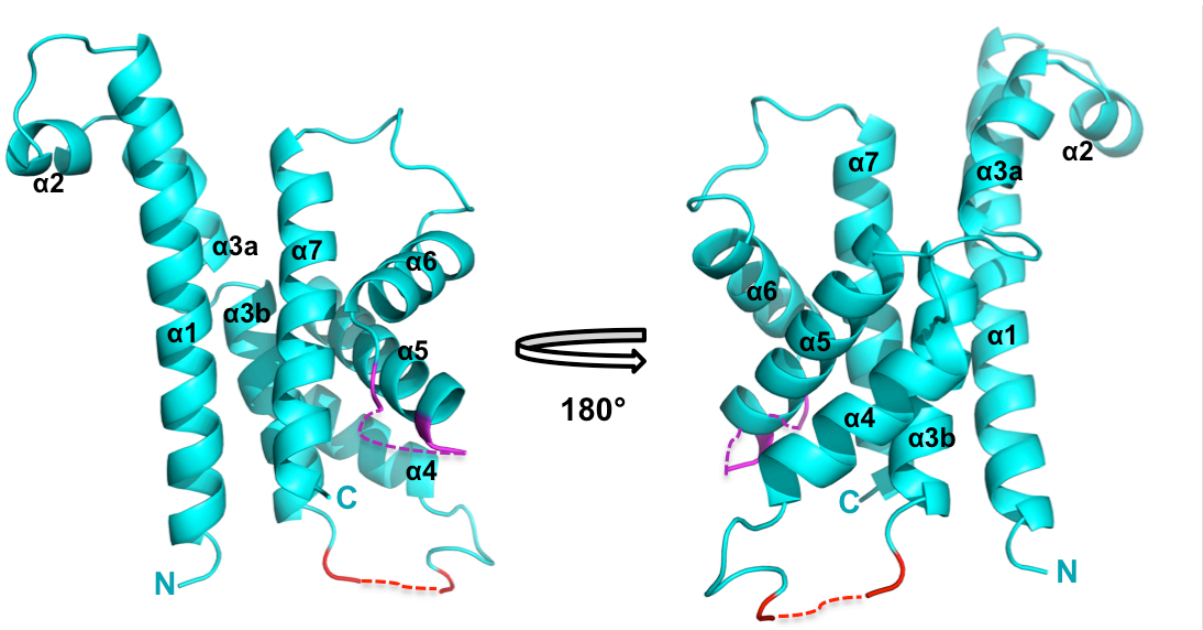

Figure 5.2. Ribbon representation of the CdiA$^{2681\text{-}2909}$ domain. CdiA$^{2681\text{-}2909}$ secondary structure elements, as well as N and C termini are labeled. Red and magenta labels represent the locations of flexible loops not observed in the crystal structure (shown as dashed lines).

Discussion

The CdiA-CT^{UPEC536}/CysK complex structure reported here provides an important glimpse into the unique uropathogenic *E. coli* CDI system. Remarkably, the C-terminal peptide of the cytotoxic UPEC536 toxin inserts into the active site of the biosynthetic enzyme CysK, mimicking the proposed interaction occurring in the CSC complex (between CysK and CysE). While multiple CysK and CysE structures have been reported, the structure of the CSC has not, thus our CdiA-CT^{UPEC536}/CysK structure represents the first structural model of CysK in complex with another protein. Additionally, the CysE C-terminal residues are not reported in the CysE structures from both *E. coli* and *Haemophilis influenza*, likely due to flexibility of this C-terminal region (30, 31). However, previous CysK structures have included small, five amino acid peptides that mimick the CysE C-terminal residues (32). Importantly, the structure of CysE is distinct from that of CdiA-CT^{UPEC536}. The carboxy-terminal domain of CysE, the region of the protein that would interact with CysK, is a left-handed β-helix, while CdiA-CT^{UPEC536} is all α-

helical (Figure 5.3). The evolutionary significance is unclear, but it is intriguing that two proteins with distinct folds would bind to a common protein partner in a similar manner. Conceivably, CdiA-CT^{UPEC536} has hijacked the well-conserved CysE/CysK interaction to allow for its activation in an array of bacterial species.

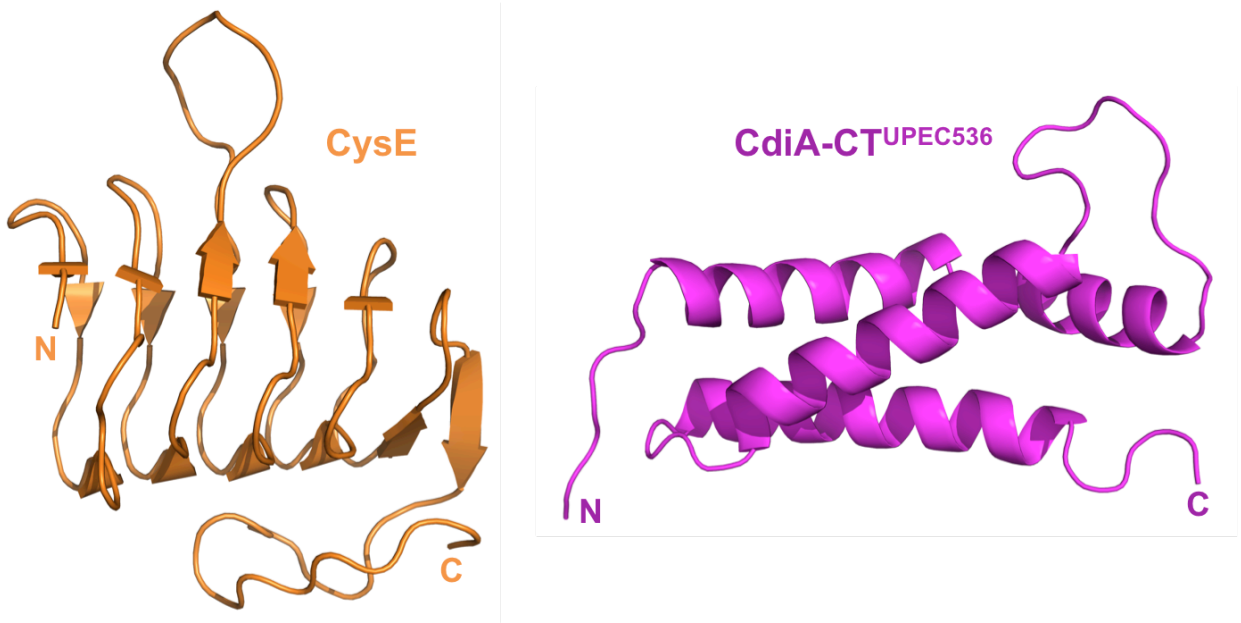

Figure 5.3. Comparison of CdiA-CT^{UPEC536} (magenta) and the carboxy-terminal domain of *E. coli* CysE (PDB ID: 1T3D) (30). CysE amino acids 142-262 are shown. The N and C termini of both proteins are labeled.

It was earlier established that CysK binding is required to activate the tRNase activity of CdiA-CT^{UPEC536}. Despite the CysK/CdiA-CT^{UPEC536} structure reported herein, the mechanism of activation is still not clear. In our structure, loop L1 is disordered, and this loop likely plays a role in activity as it is in close proximity to the known catalytic residue H178A (Figure 5.1b). Additionally, CdiA-CT^{UPEC536} shares no structural homology with previously reported endonucleases. Attempts at crystallizing CdiA-CT^{UPEC536} in the absence of CysK have thus far been unsuccessful, but this model could prove important to identify any CdiA-CT^{UPEC536} conformational changes upon CysK binding. Previously, we were able to infer activity of a CdiA-CT from *Enterobacter cloacae* by comparing it to the cytotoxic domain of colicin E3, which

has a similar fold (27). Similar to CdiA-CT^{UPEC536}, the C-terminal cytotoxic domain of colicin D (ColD-CT) cleaves the anticodon loop of tRNA^{Arg} (33). CdiA-CT^{UPEC536} and ColD-CT are not structurally similar, but both contain an active site histidine critical for tRNase activity (34). The ColD-CT active site also contains a network of positively charged lysine residues (Lys 607, Lys608, and Lys610) adjacent to the active site histidine (His 611). Mutating any of these lysine residues significantly reduces the activity of the toxin. These lysine residues are thought to play a role in either substrate recognition, by interacting with the negatively charged tRNA molecule, or catalysis. CdiA-CT^{UPEC536} contains no analogous patch of positively charged residues, suggesting that these two toxins may achieve cleavage of tRNA^{Arg} by differing mechanisms. CysK could conceivably play an active role in tRNA cleavage either by enhancing tRNA binding specificity or by involvement in the catalytic mechanism. Further research is needed to explore the CdiA-CT^{UPEC536} catalytic mechanism and the role of CysK.

In *E. coli* EC93, recent research has shown that the N-terminal stick region of CdiA^{EC93} specifically recognizes extracellular loops 6 and 7 on the highly conserved outer-membrane protein, BamA (5). BamA sequence variability is concentrated within these extracellular loops, providing a potential mechanism for bacteria to distinguish one another during CDI. Indeed, altering the sequence of these loops abolishes the efficacy of the EC93 CDI system. Recent research suggests that a CdiA domain near the middle of the protein (spanning residues R1358-F1646) could be responsible for BamA recognition (unpublished data from Hayes lab, UCSB). Despite this, the mechanism of target cell recognition by CdiA proteins, and the subsequent cleavage of the CdiA-CT toxin is still largely unknown. Here we reported the first structure of a segment of the CdiA non-toxin, stick-like region, CdiA$^{2681-2909}$, in attempt to shed light on toxin release following cell-to-cell contact. CdiA$^{2681-2909}$ is an entirely α-helical domain that contains multiple disordered loops that were not revolved in the structural model. The protein sequence of this domain is well conserved in CdiA proteins across bacterial species, signifying that it may

have a common function in different CDI systems (Figure 5.4). Recent work suggests that this CdiA domain may interact directly with CdiA-CT, as constructs that contain the CdiA$^{2681-2909}$ domain and CdiA-CT are inactive, and toxin activation is achieved by cleavage of the CdiA$^{2681-2909}$ domain (unpublished data from Hayes lab, UCSB). Finally, the N and C termini of CdiA$^{2681-2909}$ are located on the same side of the structure, and approximately 20-30 amino acids on each terminus are not resolved in the crystal structure. All of the CdiA-CT structures we have solved also lack their N-terminal regions, likely due to flexibility. This suggests that between the CdiA$^{2681-2909}$ domain and the catalytic CdiA-CT domain, there could be ~100 unstructured amino acids. Taken together, the CdiA$^{2681-2909}$ domain could form a "bulge" on the elongated CdiA structure, which may allow it to interact with target cells or the CdiA-CT domain inactivating it (Figure 5.5). Further research is required to determine the precise role the CdiA$^{2681-2909}$ domain plays in toxin delivery and perhaps inactivation, as well as other segments of the stick-like region of CdiA.

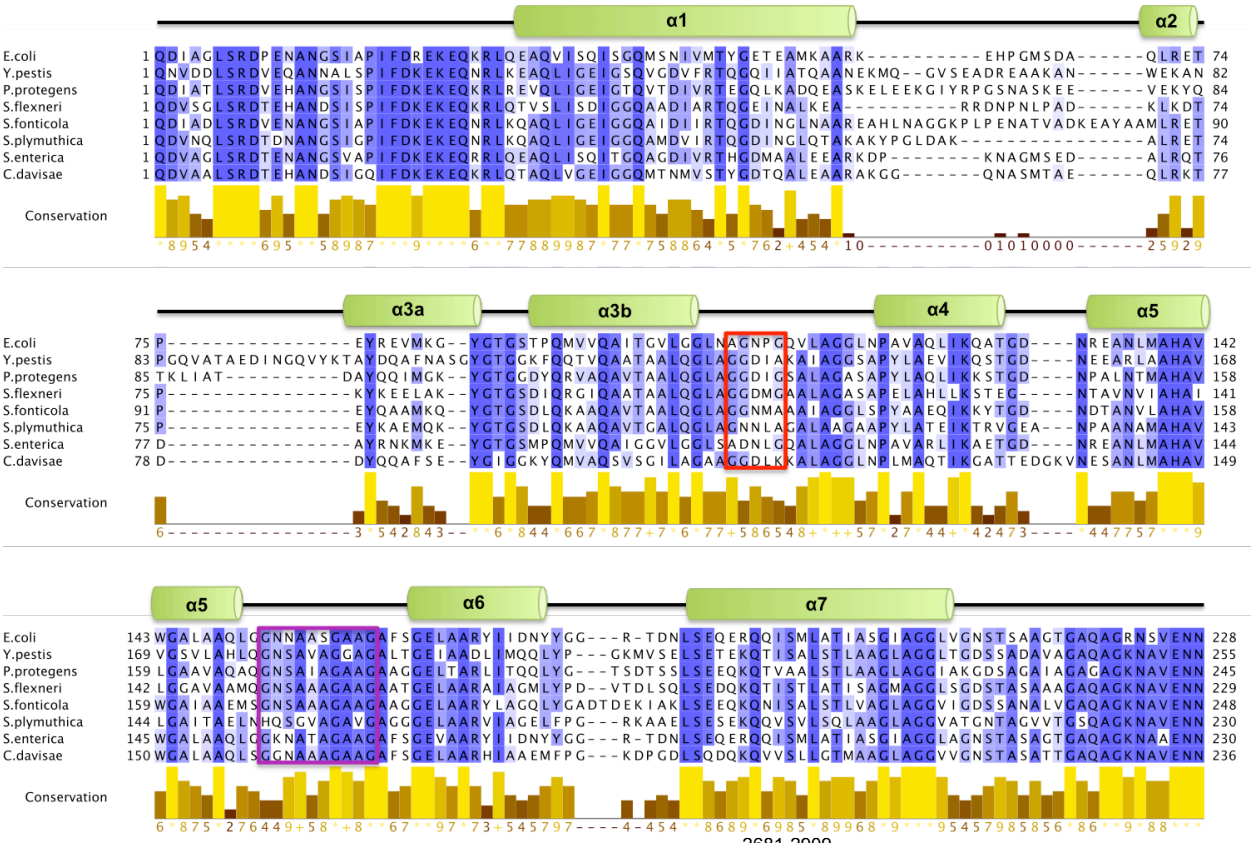

Figure 5.4. Protein sequence alignment of the CdiA$^{2681\text{-}2909}$ domain and homologs. The alignment was prepared using Jalview, with progressively darker shades of purple indicating greater residue conservation. Secondary structure elements shown correspond to the CdiA$^{2681\text{-}2909}$ structure. Disordered loops are outlined in boxes, colored as in Figure 5.2.

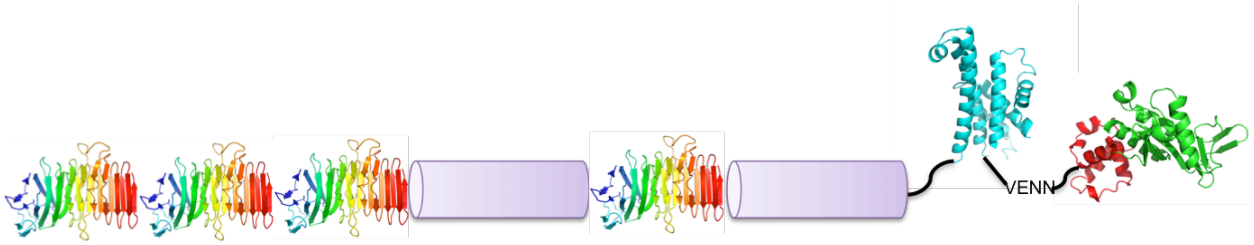

Figure 5.5. Proposed model of a full length CdiA protein. CdiA proteins are predicted to form an elongated "stick-like" structure on the surface CDI^{+} cells. HA repeats are shown in rainbow ribbon representation (PDB ID 1RWR (35)), CdiA$^{2681\text{-}2909}$ in cyan, and CdiA-CT^{EC869} (PDB ID 4G6U (8)) in red and green. CdiA-CT^{EC869} is separated from CdiA$^{2681\text{-}2909}$ by a proposed unstructured region (black lines), containing the conserved VENN peptide motif. The structure of a large portion of CdiA proteins in currently unknown, and this is represented by purple cylinders. Based on presence of flexible regions in our CdiA$^{2681\text{-}2909}$ structure, as well as the N-terminal region of all CdiA-CTs we have solved, CdiA$^{2681\text{-}2909}$ could form a "bulge" on the elongated CdiA structure.

References

1. Aoki SK*, et al.* (2010) A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* 468(7322):439-442.
2. Aoki SK*, et al.* (2005) Contact-dependent inhibition of growth in Escherichia coli. *Science* 309(5738):1245-1248.
3. Kajava AV*, et al.* (2001) Beta-helix model for the filamentous haemagglutinin adhesin of Bordetella pertussis and related bacterial secretory proteins. *Molecular microbiology* 42(2):279-292.
4. Webb JS*, et al.* (2013) Delivery of CdiA nuclease toxins into target cells during contact-dependent growth inhibition. *PLoS One* 8(2):e57609.
5. Ruhe ZC, Wallace AB, Low DA, & Hayes CS (2013) Receptor polymorphism restricts contact-dependent growth inhibition to members of the same species. *mBio* 4(4).
6. Aoki SK*, et al.* (2008) Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Molecular microbiology* 70(2):323-340.
7. Nikolakakis K*, et al.* (2012) The toxin/immunity network of Burkholderia pseudomallei contact-dependent growth inhibition (CDI) systems. *Molecular microbiology* 84(3):516-529.
8. Morse RP*, et al.* (2012) Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems. *Proceedings of the National Academy of Sciences of the United States of America* 109(52):21480-21485.
9. Aoki SK, Webb JS, Braaten BA, & Low DA (2009) Contact-dependent growth inhibition causes reversible metabolic downregulation in Escherichia coli. *Journal of bacteriology* 191(6):1777-1786.
10. Diner EJ, Beck CM, Webb JS, Low DA, & Hayes CS (2012) Identification of a target cell permissive factor required for contact-dependent growth inhibition (CDI). *Genes Dev* 26(5):515-525.
11. Kumaran S, Yi H, Krishnan HB, & Jez JM (2009) Assembly of the cysteine synthase complex and the regulatory role of protein-protein interactions. *The Journal of biological chemistry* 284(15):10268-10275.
12. Mino K*, et al.* (2000) Effects of bienzyme complex formation of cysteine synthetase from escherichia coli on some properties and kinetics. *Bioscience, biotechnology, and biochemistry* 64(8):1628-1640.
13. Zhu X, Yamaguchi T, & Masada M (1998) Complexes of serine acetyltransferase and isozymes of cysteine synthase in spinach leaves. *Bioscience, biotechnology, and biochemistry* 62(5):947-952.
14. Droux M, Ruffet ML, Douce R, & Job D (1998) Interactions between serine acetyltransferase and O-acetylserine (thiol) lyase in higher plants--structural and kinetic properties of the free and bound enzymes. *European journal of biochemistry / FEBS* 255(1):235-245.
15. Kredich NM, Becker MA, & Tomkins GM (1969) Purification and characterization of cysteine synthetase, a bifunctional protein complex, from Salmonella typhimurium. *The Journal of biological chemistry* 244(9):2428-2439.
16. Kumaran S & Jez JM (2007) Thermodynamics of the interaction between O-acetylserine sulfhydrylase and the C-terminus of serine acetyltransferase. *Biochemistry* 46(18):5586-5594.
17. Zhao C*, et al.* (2006) On the interaction site of serine acetyltransferase in the cysteine synthase complex from Escherichia coli. *Biochemical and biophysical research communications* 341(4):911-916.

18. Mino K*, et al.* (2000) Characteristics of serine acetyltransferase from Escherichia coli deleting different lengths of amino acid residues from the C-terminus. *Bioscience, biotechnology, and biochemistry* 64(9):1874-1880.
19. Mino K*, et al.* (1999) Purification and characterization of serine acetyltransferase from Escherichia coli partially truncated at the C-terminal region. *Bioscience, biotechnology, and biochemistry* 63(1):168-179.
20. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, & Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *Journal of molecular biology* 229(1):105-124.
21. Battye TG, Kontogiannis L, Johnson O, Powell HR, & Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography* 67(Pt 4):271-281.
22. Burkhard P*, et al.* (1998) Three-dimensional structure of O-acetylserine sulfhydrylase from Salmonella typhimurium. *Journal of molecular biology* 283(1):121-133.
23. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 4):486-501.
24. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2):213-221.
25. Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.0.
26. Terwilliger TC*, et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta crystallographica. Section D, Biological crystallography* 65(Pt 6):582-601.
27. Beck CM*, et al.* (2014) CdiA from Enterobacter cloacae Delivers a Toxic Ribosomal RNase into Target Bacteria. *Structure* 22.
28. Rabeh WM & Cook PF (2004) Structure and mechanism of O-acetylserine sulfhydrylase. *The Journal of biological chemistry* 279(26):26803-26806.
29. Holm L & Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic acids research* 38(Web Server issue):W545-549.
30. Pye VE, Tingey AP, Robson RL, & Moody PC (2004) The structure and mechanism of serine acetyltransferase from Escherichia coli. *The Journal of biological chemistry* 279(39):40729-40736.
31. Gorman J & Shapiro L (2004) Structure of serine acetyltransferase from Haemophilus influenzae Rd. *Acta crystallographica. Section D, Biological crystallography* 60(Pt 9):1600-1605.
32. Salsi E*, et al.* (2010) Design of O-acetylserine sulfhydrylase inhibitors by mimicking nature. *Journal of medicinal chemistry* 53(1):345-356.
33. Tomita K, Ogawa T, Uozumi T, Watanabe K, & Masaki H (2000) A cytotoxic ribonuclease which specifically cleaves four isoaccepting arginine tRNAs at their anticodon loops. *Proceedings of the National Academy of Sciences of the United States of America* 97(15):8278-8283.
34. Graille M, Mora L, Buckingham RH, van Tilbeurgh H, & de Zamaroczy M (2004) Structural inhibition of the colicin D tRNase by the tRNA-mimicking immunity protein. *The EMBO journal* 23(7):1474-1482.
35. Clantin B*, et al.* (2004) The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway. *Proceedings of the National Academy of Sciences of the United States of America* 101(16):6194-6199.

CHAPTER 6

Conclusions and Future Directions: Understanding contact-dependent growth inhibition protein diversity

Summary

Contact-dependent growth inhibition (CDI) is a unique bacterial competition mechanism by which polymorphic toxin (CdiA-CT) and immunity (CdiI) proteins mediate the growth of neighboring cells. CdiA-CT toxins bind tightly to cognate CdiI immunity proteins, which prevents toxin producing "inhibitor-cells" from killing "self-cells." Here we presented the initial structural characterization of proteins in the CDI system, including the structures of CdiA-CT/CdiI complexes from three families. These structures revealed three distinct CdiA-CT/CdiI interaction interfaces, and were essential to elucidate each CdiA-CT toxin activity. We further explored the unique *E. coli* EC869 β-augmentation CdiA-CT/CdiI complex interface through additional structural and biochemical characterization. Among the members of this "EC869-like" CDI protein family, the complex interface is conserved, but heterologous EC869 family CdiA-CT and CdiI proteins cannot form stable complexes. Additionally, we described the structure of the CdiA-CT tRNase toxin from uropathogenic *E. coli* strain 536 in complex with the biosynthetic enzyme CysK, which is required to activate the toxin. For future research, we are especially interested in the CDI cell-to-cell contact and translocation mechanisms mediated by the non-toxin "stick" region of CdiA. We solved the structure of a non-toxin CdiA domain, providing a starting point for future characterization of the CdiA "stick" region.

CdiI-CT/CdiI structural characterization reveals diverse protein interfaces

Our initial structural characterization of CDI systems focused on proteins from *Escherichia coli* EC869 (EC869) and *Burkholderia pseudomallei* 1026b (1026b) (1). The CDI systems in these two bacteria are distinct in both gene organization and the conserved peptide that demarcates the CdiA-CT toxins (2). The crystal structures of both EC869 and 1026b CdiA-CT/CdiI complexes were solved, which surprisingly revealed similar CdiA-CT endonuclease folds, despite these domains sharing only ~15% sequence identity. However, the toxins interact with their cognate immunity proteins through unique protein-protein interfaces, thereby reducing the possibility of heterologous toxin/immunity interactions. A network of electrostatic interactions mediates the 1026b complex interface, while the EC869 complex is formed by a unique β-augmentation interaction. Importantly, the immunity proteins share no structural similarity, and bound to completely different locations on the surface of their cognate toxins. These results support the hypothesis that CdiI sequence and structural diversification is predicted to be the dominant driving force for continued binding specificity between cognate CdiA-CT/CdiI pairs (3).

In a separate study, we solved the crystal structure of a CdiA-CT/CdiI protein complex from *Enterobacter cloacae* ATCC 13047 (ECL) (4). CdiA-CT^{ECL} shares no sequence or structural homology with CdiA-CT^{EC869} or CdiA-CT^{1026b}, yet is structurally similar to the C-terminal nuclease domain of colicin E3 (ColE3-CT). The ECL complex interaction is distinct from both 1026b and EC869, and is mediated by a small hydrophobic interface (area 300 Å^2) surrounded by a network of over 20 ion-par and hydrogen bonds between the two proteins. Similar to what we observed in the EC869 and 1026b complexes, the cognate immunity proteins to CdiA-CT^{ECL} and ColE3-CT differ structurally, and bind different locations on their toxins. Thus far, we have solved structures of CdiA-CT/CdiI complexes from three different CDI families, all

of which exhibited unique interaction interfaces, providing evidence as to why toxins and immunities from different CDI protein families do not interact.

Heterologous CdiI-CT and CdiI proteins from the same family cannot form stable complexes

Over 60 CdiA-CT/CdiI families have been identified, which suggests a wide variety of toxin activities and CdiA-CT/CdiI interaction interfaces (2). We sought to further investigate the "EC869-like" family CdiA-CT toxins, all of which are predicted to interact with cognate immunity proteins by β-augmentation. The crystal structure of CdiI from *Yersinia kristensenii* (YK), solved in absence of a bound toxin, revealed that the immunity proteins retain their fold and the β-hairpin binding pocket is accessible. Additionally, the structure of the CdiA-CT/CdiI complex from *Yersinia pseudotuberculosis* (YP413) showed high structural homology to the EC869 complex, including the β-augmentation interaction. These two structures suggest that there is structural conservation across the "EC869-like" CDI family. Despite their similar folds, $CdiI^{EC869}$ only interacts weakly with $CdiA\text{-}CT^{YP413}$ (the same is true for $CdiI^{YP413}$ and $CdiA\text{-}CT^{EC869}$). Recent research has revealed a similar phenomenon in the *B. pseudomallei* 1026b toxin/immunity family, in which structural conservation is observed across the family, yet heterologous toxins and immunities do not form stable complexes (unpublished data, Parker Johnson, Goulding lab). Sequence alignments of CdiI homologs within both of these families reveal that toxin interacting residues are poorly conserved. This stipulates that CdiI sequence diversity contributes to CdiA-CT/CdiI binding specificity within CDI families.

In general, the sequence diversity of CdiA-CTs within families is much more limited than CdiIs, because the toxins need to retain their catalytic functions. EC869-like CdiA-CTs are a special case, as the β-hairpin they donate as part of the β-augmentation interaction is not part of the toxin's active site, and is very diverse among family members. One could speculate that

other CdiA-CT families could have evolved similar CdiI-binding motifs that allow for diversification of the toxin protein sequence, and thus improved binding specificity to cognate CdiI proteins. One such type of potential binding motif is observed in the Doc/Phd toxin/antitoxin (TA) complex in *E. coli*. In this complex, binding of the Phd (prevents host death) antitoxin to Doc (death on curing) is mediated by fold complementation, in which a Phd α-helix inserts into a groove on the Doc surface (5). Similar to EC869 CdiA-CT/CdiI complexes, Phd binds a Doc "exosite", and Doc residues that interact with Phd are not well conserved. Further structural characterization of additional CDI families is needed to uncover additional unique protein-protein interfaces that allow for diversification of both the toxin and immunity protein sequences.

CdiA-CT activation by target cell machinery

We have determined the crystal structure of CdiA-CT^{UPEC536} (from uropathogenic *E. coli* 536) in complex with its permissive factor, CysK (*O*-acetylserine sulfhydrylase A). As expected, the C-terminal peptide of CdiA-CT^{UPEC536} mediates the complex interface, mimicking the interaction observed in the cysteine synthase complex (CSC) formed between CysK and CysE (6). The CdiA-CT^{UPEC536} toxin has a long flexible loop directly adjacent to a known catalytic residue, His178, leading us to hypothesize that this loop may be important for the toxin's tRNase activity. A recently solved crystal structure of the CysK/CdiA-CT^{UPEC536}/CdiI^{UPEC536} ternary complex by another member of the Goulding lab has shown that CdiI^{UPEC536} binds to CdiA-CT^{UPEC536} in such a way as to bury both the catalytic His178 residue and the flexible loop (unpublished data, Parker Johnson, Goulding lab). However, despite our wealth of structural data it remains unclear as to how binding to CysK activates CdiA-CT^{UPEC536}. Further studies are currently underway in our lab to determine the role of CysK in CdiA-CT^{UPEC536} activation and substrate recognition.

This phenomenon of toxin activation by target cells has also been observed in other bacterial competition systems. At least one colicin, colicin M, requires processing into its active conformation (7, 8). Following translocation into the target cell, the cell-killing ability of colicin M is dependent on conversion of a specific Phe-Pro *trans* bond into a *cis* bond by a periplasmic peptidyl prolyl *cis-trans* isomerase. Given the diversity of the CDI system and the similarities it shares with the colicin system, it is likely that additional CdiA-CT toxins may require components within the target cell to activate their functions, either by binding a permissive factor or processing.

CdiA-CT modularity

Here we presented crystal structures of four distinct CdiA-CT toxins: CdiA-CT^{EC869}, CdiA-CT^{1026b}, CdiA-CT^{ECL}, and CdiA-CT^{UPEC536}. The CdiA-CT structures all lacked a significant number of amino terminal residues (ranging from 70-173 amino acids) either because they were unresolved in the final model (EC869), degraded during crystallization (ECL and UPEC536), or the toxin was expressed as a truncated domain (1026b) (1, 4). We have cloned, expressed, and purified the EC869 and 1026b CdiA-CT N-terminal domains, however crystallization of these isolated domains has thus far been unsuccessful. The N-terminal regions of CdiA-CTs are predicted to be flexible, while the C-terminal domains are responsible for toxin activity and CdiI interactions. These CdiA-CT N-terminal regions are hypothesized to be important for translocation of the toxin across the target cell membrane into the cytosol. This function and modularity would be similar to what is observed in the colicin toxin/immunity system (9). Colicin proteins contain three domains: A N-terminal domain that mediates transport across the target cell membrane, a central domain responsible for receptor binding, and a C-terminal toxin domain. The N-terminal region of CdiA-CTs could function in a similar manner to the N-terminal domain of colicins. Structures of complete colicins often contain long α-helical regions separating the receptor binding and translocation domains from the C-terminal toxin (10-12).

Additionally, many colicin structures lack their N-terminal regions due to flexibility. For example, the translocation domain of colicin A has been shown to lack secondary structure by both circular dichroism and the poor resolution of its heteronuclear nuclear magnetic resonance spectrum (13). If the CdiA-CT N-terminal region does indeed function in toxin translocation, structural characterization of these regions could greatly improve our understanding of the CDI system.

We recently solved the crystal structure of an *E. coli* EC93 CdiA domain encompassing residues 2681-2909 ($CdiA^{2681-2909}$). This is the final predicted domain in the non-toxin "stick" region of the CdiA protein, and directly precedes the CdiA-CT^{EC93} domain. *In vitro* assays have shown that various CdiA-CTs exhibit less activity when fused to this domain, suggesting that the $CdiA^{2681-2909}$ domain may interact with the CdiA-CTs prior to toxin cleavage (unpublished data from Hayes lab, UCSB). These $CdiA^{2681-2909}$-CdiA-CT fusion proteins are attractive candidates for future structural studies as they may prevent flexibility in the toxin N terminal regions. We are currently pursuing a number of promising crystallization leads with $CdiA^{2681-2909}$ fused to EC869, UPEC536, and ECL CdiA-CTs.

Target cell receptor recognition

One of the least understood aspects of the CDI system is the mechanism of target cell receptor binding. CdiA-CT^{EC93} is known to interact with two BamA extracellular loops, however the precise mechanism by which BamA is recognized is unknown (14). Preliminary results suggest an interaction mediated by a $CdiA^{EC93}$ domain (spanning residues R1358-F1646), which shows significant sequence variation among CdiA homologs (unpublished data from Hayes lab, UCSB). We plan on performing $CdiA^{EC93}$-BamA co-crystallization trials in attempt to characterize receptor binding. The crystal structures of BamA from *Neisseria gonorrhoeae* and *Haemophilus ducreyi* were recently described, which provide us with a general protocol for BamA crystallization (15). Additionally, the receptor-binding domains of colicins and their

membrane protein receptors bind tightly *in vitro*, suggesting that perhaps under the right conditions we could create a stable CdiA^{EC93}-BamA complex (16, 17). Alternatively, cross-linkers could be employed to stabilize the interaction. This characterization would provide important insights into the CDI cell-to-cell contact mechanism, one of the least understood aspects of the CDI system.

Concluding remarks

This dissertation presented a structural and functional analysis of a diverse array of proteins within the CDI system. We elucidated specific CdiA-CT toxicity mechanisms and offered important insights regarding CdiA-CT neutralization by cognate CdiI proteins, including why toxin/immunity interactions are highly specific for cognate pairs. However, much work still remains to be done to characterize the CDI system. The discovery of the EC93 CDI system suggested that bacteria utilize CDI to achieve a competitive advantage, and could potentially be used by bacteria outcompete the normal flora of a host. Additionally, CDI systems are present in many important human pathogens, suggesting that CDI may be important in pathogenesis. In a recent study, a strain of *E. coli* was engineered to specifically recognize, seek, and eradicate both dispersed and biofilm-incased *Pseudomonas aeruginosa* cells within a host (18). Upon detection of quorum sensing molecules secreted by *P. aeruginosa*, these *E. coli* secreted an antimicrobial peptide and nuclease to kill the target cells. One could speculate that the specificity of CdiA-receptor interactions could allow for bacteria to be engineered in a similar way, allowing the CDI system to be utilized for targeting specific pathogens. Furthering our understanding of CDI would not only increase our understanding of a unique bacterial communication mechanism, but could allow for the development of new antimicrobial therapies.

References

1. Morse RP*, et al.* (2012) Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems. *Proceedings of the National Academy of Sciences of the United States of America* 109(52):21480-21485.
2. Aoki SK*, et al.* (2010) A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* 468(7322):439-442.
3. Ruhe ZC, Low DA, & Hayes CS (2013) Bacterial contact-dependent growth inhibition. *Trends in microbiology* 21(5):230-237.
4. Beck CM*, et al.* (2014) CdiA from Enterobacter cloacae Delivers a Toxic Ribosomal RNase into Target Bacteria. *Structure* 22.
5. Garcia-Pino A*, et al.* (2008) Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. *The Journal of biological chemistry* 283(45):30821-30827.
6. Diner EJ, Beck CM, Webb JS, Low DA, & Hayes CS (2012) Identification of a target cell permissive factor required for contact-dependent growth inhibition (CDI). *Genes Dev* 26(5):515-525.
7. Helbig S, Patzer SI, Schiene-Fischer C, Zeth K, & Braun V (2011) Activation of colicin M by the FkpA prolyl cis-trans isomerase/chaperone. *The Journal of biological chemistry* 286(8):6280-6290.
8. Hullmann J, Patzer SI, Romer C, Hantke K, & Braun V (2008) Periplasmic chaperone FkpA is essential for imported colicin M toxicity. *Molecular microbiology* 69(4):926-937.
9. Cascales E*, et al.* (2007) Colicin biology. *Microbiology and molecular biology reviews : MMBR* 71(1):158-229.
10. Hilsenbeck JL*, et al.* (2004) Crystal structure of the cytotoxic bacterial protein colicin B at 2.5 A resolution. *Molecular microbiology* 51(3):711-720.
11. Soelaiman S, Jakes K, Wu N, Li C, & Shoham M (2001) Crystal structure of colicin E3: implications for cell entry and ribosome inactivation. *Molecular cell* 8(5):1053-1062.
12. Wiener M, Freymann D, Ghosh P, & Stroud RM (1997) Crystal structure of colicin Ia. *Nature* 385(6615):461-464.
13. Deprez C*, et al.* (2002) Macromolecular import into Escherichia coli: the TolA C-terminal domain changes conformation when interacting with the colicin A toxin. *Biochemistry* 41(8):2589-2598.
14. Ruhe ZC, Wallace AB, Low DA, & Hayes CS (2013) Receptor polymorphism restricts contact-dependent growth inhibition to members of the same species. *mBio* 4(4).
15. Noinaj N*, et al.* (2013) Structural insight into the biogenesis of beta-barrel membrane proteins. *Nature* 501(7467):385-390.
16. Baboolal TG*, et al.* (2008) Colicin N binds to the periphery of its receptor and translocator, outer membrane protein F. *Structure* 16(3):371-379.
17. Smarda J & Macholan L (2000) Binding domains of colicins E1, E2 and E3 in the receptor protein BtuB of Escherichia coli. *Folia microbiologica* 45(5):379-385.
18. Hwang IY*, et al.* (2013) Reprogramming Microbes to Be Pathogen-Seeking Killers. *ACS synthetic biology*.

APPENDIX A

Structural and biochemical insights into the *Mycobacterium tuberculosis* heme-degrading protein, MhuD

Summary

Mycobacterium heme utilization degrader (MhuD) is a heme degrading protein from *Mycobacterium tuberculosis* responsible for extracting the essential nutrient iron from host-derived heme. MhuD has been previously shown to produce unique organic products as compared to those of human heme oxygenases (hHOs), as well as those of the structurally homologous IsdG and IsdI heme degrading enzymes from *Staphylococcus aureus*. Here we report results of a mutagenic analysis of the MhuD active site, showing that mutations to either the proximal or distal heme ligands abolish the catalytic capability of MhuD. Additionally we report the X-ray crystal structure of cyanide-inhibited MhuD (MhuD–heme–CN), which shows MhuD in an active, monoheme confirmation. The degree of heme ruffling in MhuD-heme-CN is greater than that observed for hHO and less than that observed for IsdI. The structure provides insight into the α-meso regioselectivity of MhuD-catalyzed heme degradation.

Introduction

Heme degradation is performed by a diverse range of organisms, fulfilling various physiological functions. Heme oxygenases (HO) are ubiquitous enzymes in nature, and their mechanism of heme catabolism has been widely studied. Initially discovered in eukaryotes about 40 years ago (1), many bacterial HOs have now been characterized, with most having significant structural homology to human HO-1 (2-5). Importantly, heme degradation plays a role in the availability of the essential nutrient iron for pathogenic bacteria. Heme degradation across most HOs results in the production of α-biliverdin and carbon monoxide (CO) (Figure 1). In eukaryotes, this is coupled with the conversion of biliverdin to bilirubin by biliverdin reductase (6). However, the ultimate fate of biliverdin in bacteria is still unclear.

The IsdG-family of HOs is structurally distinct from the HO-1 family (7-9). HO-1 family proteins are monomeric α-helical proteins with a single active site, while IsdG-family HOs form homodimeric β-barrel structures, with two separate active sites. Each active site contains a trio of catalytic residues (Asn, Trp, His), that for IsdG and IsdI have been shown to be required for activity (8). Recently it was discovered that IsdG and IsdI degrade heme into the novel product "staphylobilin" (oxo-bilirubin) and formaldehyde (10, 11) (Figure 1). Heme degradation by IsdG and IsdI yields a mixture of 2 different staphylobilins, 5-oxo-δ-bilirubin and 15-oxo-β-bilirubin, resulting from cleavage at the δ and β carbon positions on the tetrapyrole ring, respectively.

Our lab previously solved the crystal structure of MhuD, an IsdG-family HO from *Mycobacterium tuberculosis*, which revealed two stacked heme molecules in the MhuD active site (MhuD-diheme) (12). However, no heme-degradation is observed by MhuD-diheme, only the monoheme-protein complex is capable of degrading heme. Recent work has shown that the organic by-products of MhuD-catalyzed heme degradation are unique (Figure 1). MhuD cleaves the porphyrin ring of heme at the α-meso carbon, and this carbon is retained as an aldehyde group in the final mycobilin product (13). Despite high structural homology, MhuD-diheme contains significant differences to IsdG and IsdI (Figure 2) (12). Importantly, MhuD is capable of

binding two heme molecules per active site (MhuD-diheme), while the Isd proteins can only bind one. The ability to bind multiple heme molecules may be attributed to the fact that in IsdG and IsdI, there is a phenylalanine in the active site (Phe73 in IsdG) that may sterically hinder binding of a second heme molecule, whereas in MhuD this phenylalanine is an alanine (Ala71). Additionally, in the structure of MhuD, α-helix 2 is extended, whereas the corresponding residues in IsdG and IsdI consist of two α-helices, bent at a 45° angle. The residues in MhuD α-helix 2 and the succeeding loop are conserved in other mycobacterial MhuD homologs (12). Finally, when comparing propionate groups it is apparent that the solvent-protected heme in the MhuD diheme structure is rotated ~90° when compared to the heme in the IsdG and IsdI active sites (Figure 2) (11).

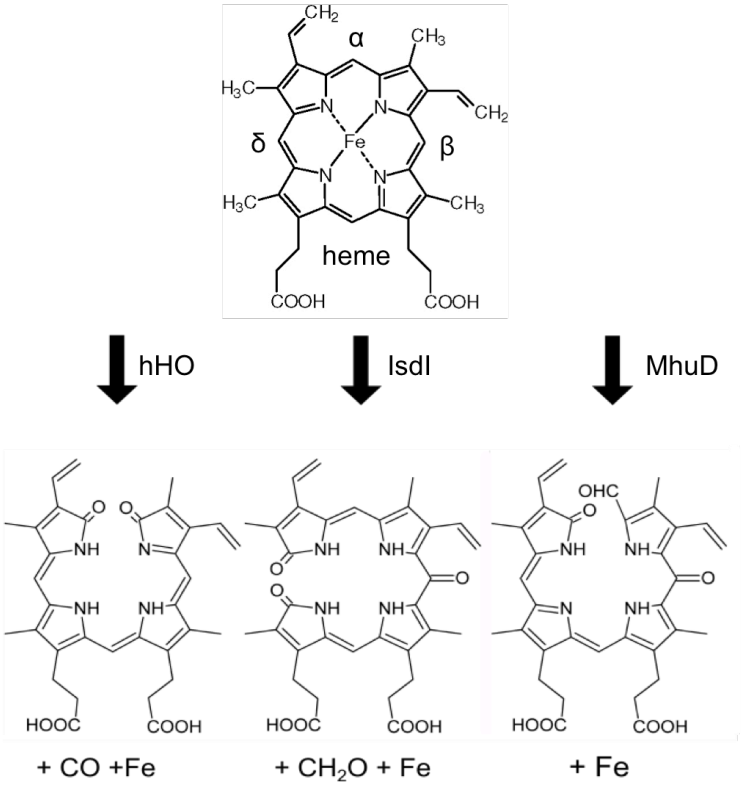

Figure 1. The distinct products of heme degradation by hHO, IsdI, and MhuD. hHO produces biliverdin, CO, and iron as products. *S. aureus* IsdI (and IsdG) prodcuce staphylobilin (5-oxo-δ-bilirubin shown) and formaldehyde. MhuD degrades heme to mycobilin and iron.

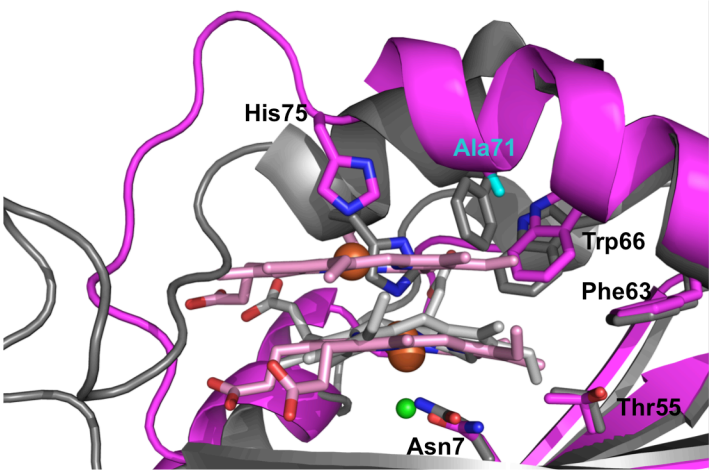

Figure 2. Comparison of the heme binding sites of MhuD-diheme (PDB ID 3HX9, magenta) and IsdI-heme (PDB ID 3LGN, grey). Heme molecules are in stick representation, with carbon atoms depicted in pink and light grey for MhuD-diheme and IsdI, respectively, and nitrogen and oxygen atoms in blue, red respectively. Iron atoms, and ordered Cl^- atom (MhuD–diheme structure) depicted as orange, green spheres, respectively. Labeled residues correspond to MhuD residue numbers. Most of the residues in the active sites are conserved, except for Ala71 (cyan).

In this study we further characterize the MhuD active site through mutational analysis. In addition, we report the first crystal structure of a monoheme form of MhuD, using cyanide as an inhibitor. This structure offers the first glimpse of MhuD in an active conformation, and reveals a His-ligated heme substrate in a similar position as the solvent-protected heme in the MhuD-diheme structure. This structure offers insights into MhuD-catalyzed heme degradation, and the unique products formed compared to IsdG and IsdI.

Materials and Methods

Plasmids, Protein Expression, and Purification

The preparation of a pET-22b (Amp^r) plasmid encoding wild type (WT) *M. tuberculosis* MhuD along with a C-terminal His_6 tag has been previously described (12). Site-directed mutagenesis (SDM) was utilized to generate active site point mutations (Table 1). Following PCR, the SDM product was digested overnight with *Dpn*1, and transformed the following day

into TOP10 *E. coli* cells. Generation of the mutations was confirmed by DNA sequencing (Laguna). For recombinant protein expression, the pET-22b plasmid encoding MhuD was transformed into BL21-GOLD (DE3) cells (Stratagene) as previously described. For all studies presented herein, expression and purification of apo-MhuD was carried out as previously described (12). Expression and purification of IsdG (cloned into a pET-15b plasmid (7)) was carried out the same as apo-MhuD.

Table 1. Primers for creation of MhuD point mutations

Primer	Sequence
N7A For	CCAGTGGTGAAGATCGCGGCAATCGAGGTGCCC
N7A Rev	GGGCACCTCGATTGCCGCGATCTTCACCACTGG
A71F For	GGCAAACGGGCCCTTCATCGCAGCCCATGC
A71F Rev	GCATGGGCTGCGATGAAGGGCCCGTTTGCC
H75A For	GCCATCGCAGCCGCGGCCGGACACCGG
H75A Rev	CCGGTGTCCGGCCGCGGCTGCGATGGC

Heme Binding Assay

Heme titration experiments were performed as described (12). Briefly, aliquots of hemin (1 μM) were added into ~5 μM apo-MhuD (or apo-IsdG), 50 mM Tris, pH 8.0, 150 mM NaCl at 25 °C, and heme binding was monitored by difference absorption spectroscopy (DU800, Beckman Coulter) between 300 and 700 nM. The change in absorption at 410 nm was calculated by subtracting heme-into-buffer control spectra from heme-into-protein spectra.

Heme Degradation Assay

Reconstitution of apo-MhuD with heme was performed as previously described (12). Protein concentrations were determined by a Bradford assay utilizing bovine serum albumin as a standard (14). The single turnover degradation reaction of MhuD-heme was performed as previously described. Briefly, 10 mM sodium ascorbate was added to 5 μM MhuD-heme, 50

mM Tris, pH 7.4, and 150 mM NaCl. Spectral changes were monitored between 300 and 700 nM every 5 min.

Crystallization and Data Collection of MhuD-heme-CN

Purified apo-MhuD was exchanged into 50 mM sodium phosphate (NaPi) pH 7.4, 150 mM NaCl by concentrating and resuspending three times (Amicon). A 0.5 mM hemin solution was prepared by dissolving 3.3 mg of hemin chloride in 300 µL of 1 M NaOH, to which 50 mM NaPi pH 7.4, 150 mM NaCl was added to a final volume of 10 mL. The pH of the hemin solution was adjusted to 7.4 using HCl. Several crystals of potassium cyanide (KCN) were added to the hemin solution. Purified apo-MhuD (concentrated to 100 µM) and cyanide-bound heme were combined in a 1:1.05 ratio (apo-MhuD:cyanide-bound heme). MhuD–heme–CN was concentrated to 1 mL (Amicon), and was exchanged into 50 mM NaPi pH 7.4, 150 mM NaCl using a PD-10 desalting column (GE Healthcare). MhuD–heme–CN was concentrated to 10 mg/mL in 50 mM NaPi pH 7.4, 150 mM NaCl for crystallization trials. MhuD–heme–CN crystallized in 1.9 M ammonium sulfate, 0.1 M sodium acetate pH 6.0, 0.2 M NaCl. Crystals were mounted and data was collected under cryo-conditions with the addition of 40% (*v*/*v*) glycerol as cryoprotectant to the reservoir condition. The native diffraction dataset was collected at a wavelength of 1.0 Å at 70 K, with crystal diffracting to 1.9 Å with unit cell dimensions 40.97 Å x 60.40 Å x 78.46 Å and one dimer per asymmetric unit in space group $P2_12_12_1$. Images were indexed, integrated, and reduced using iMOSFLM (15). The initial phases were determined by molecular replacement by autoMR in PHENIX using the MhuD–diheme structure without the heme molecules present as a search model (PDB ID 3HX9). The initial model building was performed by Autobuild in PHENIX (16). The final model was built through iterative manual building in Coot and refined with phenix.refine.(17) During refinement, minimal restraints were placed on the heme molecule to allow for greater torsional rotation of bonds within the porphyrin ring. Minimal distance and no angle restraints were used for the iron-cyanide bond. Chains A

and B of the MhuD dimeric model contain residues 2-102 and 2-101, respectively. The final dimeric model contains two heme substrates, two cyanide ligands, two acetate molecules, and 94 waters (PDB ID 4NL5). Programs from Phenix (16), Coot (17), and Pymol (18), were used to analyze the stereochemistry and geometry of the models and were found to be acceptable. Data and refinement statistics are presented in Table 1. All molecular graphics were prepared with PyMOL (18).

Table 2. X-ray data collection and refinement statistics for the structural determination of MhuD–heme–CN

	MhuD–heme–CN
Space Group	$P2_12_12_1$
Unit cell dimensions (Å)	40.97 × 60.40 × 78.46
pH of crystallization condition	6.0
Protein concentration (mg/mL)	10
Data set	
Wavelength (Å)	1.0
Resolution range	39.23-1.90
Unique reflections (total)	15989 (103418)
Completeness (%)*	99.6 (99.9)
Redundancy *	6.5 (6.7)
R_{merge}*,†	6.4 (35.5)
I/σ*	17.3 (4.9)
NCS copies	2
Model refinement	
Resolution range (Å)	39.23-1.90
No. of reflections (working/free)	15935 (1595)
No. of protein atoms	1505
No. of water molecules	94
No. of heme/dimer	2
No. of cyanide.dimer	2
Missing residues	103-105
R_{work}/R_{free} (%)‡	17.3/22.4
Average B-factor (Å^2)	
Protein	31.9
Heme & CN	27.0
Water	37.6
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (degrees)	1.098
Ramachandran Plot	
Most favorable region (%)	95.43
Additional allowed region (%)	4.06
Disallowed region	0.51
PDB ID Code	4NL5

*Statistics for the highest resolution shell are given in (brackets)

†$R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma I$

‡$R_{work}=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$ R_{free} was computed identically except where all reflections belong to a test set of 10% randomly selected data.

Results

Effect of Ala71 on MhuD heme binding

We propose that the ability for MhuD to bind two heme molecules is due to the absence of a conserved phenylalanine in the heme-binding pocket, which is substituted with an alanine (A71) in MhuD. The lack of the larger phenylalanine residue may provide the necessary space required for a second molecule of heme to bind. To investigate this, an A71F point mutation was engineered in MhuD by site-directed mutagenesis to mimic the heme-binding pocket of IsdG and IsdI. To test the heme binding capacity of this mutant, hemin was incrementally titrated into MhuD-A71F and was monitored using absorption spectroscopy (Figure 3A). Heme binding was monitored by the appearance of a Soret peak at 410 nm, and difference spectra were generated by subtracting the free heme spectra from the heme-titrated MhuD-A71F spectra. Plotting the absorbance difference at 410 nm against heme concentration revealed the MhuD-A71F mutant can still bind two molecules of heme, thus A71 is not solely responsible for the multiple heme-binding capability of MhuD. A similar experiment with IsdG confirmed its single-heme capacity (Figure 3B).

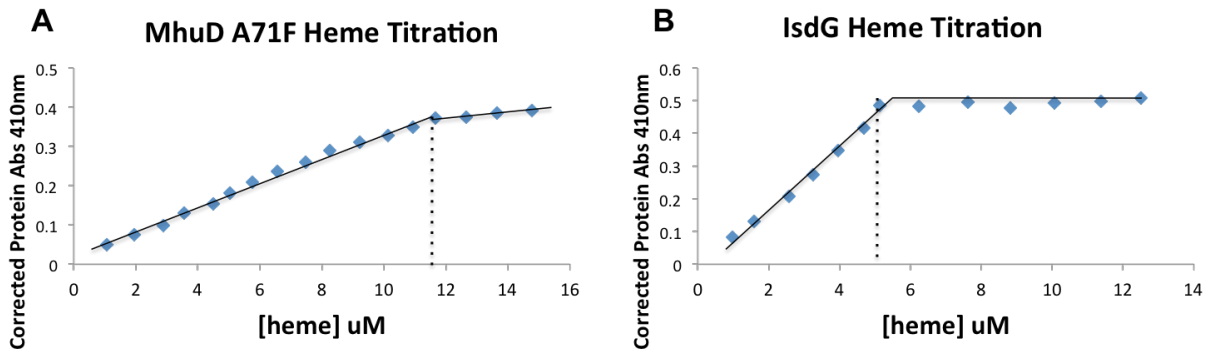

Figure 3. Effects of the MhuD-A71F mutation on heme binding. Heme was incrementally titrated into 5.4 μM MhuD (A) and 5 μM IsdG (B), and the change in absorbance at 410 nm versus heme concentration shows the saturation point for each protein (dotted line). MhuD-A71F retains the ability to bind two heme molecules.

Conserved catalytic residues are required for heme degradation

IsdG-family heme degrading enzymes share a conserved trio of catalytic residues, which in MhuD consist of Asn7, Trp66, and His75. In the MhuD-diheme crystal structure, His75 coordinates the solvent-exposed heme, while Asn7 coordinates the solvent-protected heme through a chloride axial ligand. To investigate the importance of these residues in the heme degradation reaction, N7A and H75A mutations were made by side-directed mutagenesis. Heme-degradation assays were performed using these two mutants, as well as the aforementioned A71F mutant as a positive control (Figure 4). Purified mutant MhuD was incubated with heme overnight at a ratio of 1:1, purified by size- exclusion chromatography, and the concentration of protein and heme were approximately 1:1 as determined by Bradford and pyridine hemochrome assays, respectively (14, 19). Incubation of heme bound MhuD-A71F with ascorbate caused a disappearance of the soret peak over time, indicating that MhuD- A71F is degrading heme (Figure 4A). In contrast, both the N7A and H75A mutants showed no heme degradation activity, which is consistent for homologous mutations in IsdG and IsdI (Figure 4B-4C) (8). This could suggest a similar degradation mechanism between MhuD, IsdG, and IsdI.

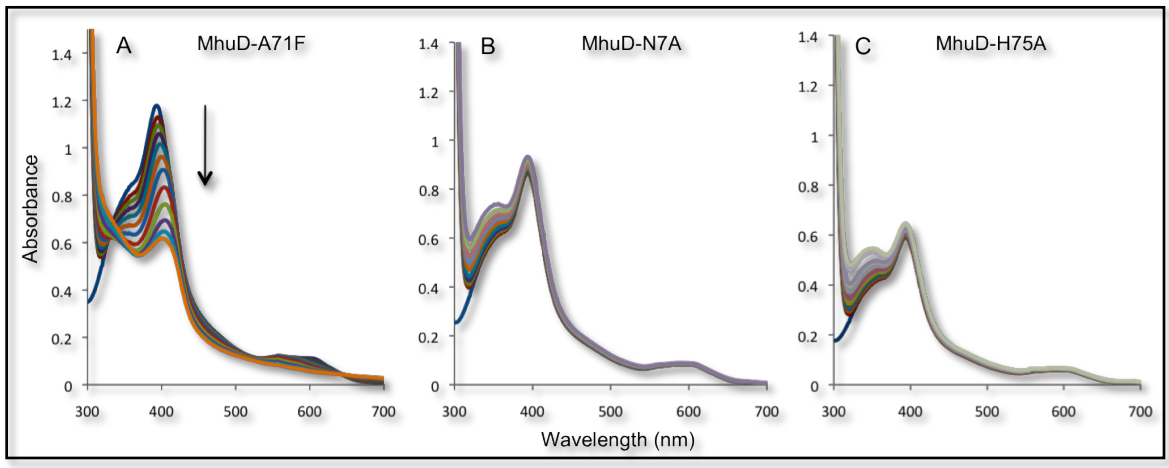

Figure 4. The degradation of heme by MhuD point mutants was observed by the decrease of the Soret peak (arrow) by monitoring the spectral changes from 300-700 nm every 5 minutes. Ascorbate (10 mM) was added to MhuD-heme (5 μM, 1:1). Degradation was observed by the A71F mutant (A), but not for the N7A and H75A mutants (B, C).

Structure of MhuD-heme-CN

We have previously determined the crystal structure of MhuD–diheme, which revealed two stacked heme molecules in the MhuD active site (12). However, no heme degradation activity was observed by MhuD–diheme and only the monoheme complex was capable of degrading heme. To observe MhuD in its monoheme active conformation, we determined the structure of MhuD–heme–CN to 1.9 Å resolution, with one homodimer in the asymmetric unit. MhuD–heme–CN retains its ferrodoxin-like α + β-barrel fold as observed for MhuD–diheme; however only one bound molecule of heme is observed in each active site (Figure 5A). In accord with previously reported spectroscopic data (13), His75 coordinates the iron of MhuD–heme–CN on its proximal side (2.1 Å, Figure 5B), and the His75 imidazole ring is hydrogen (H)-bonded to the backbone carbonyl of Ala71. A fully occupied cyano group was modeled into the electron density observed on the distal side of the heme iron (2.1 Å). The bound CN atoms refine with B-factors of approximately 22 $Å^2$, similar to those of the heme irons, and fit the electron density well (Figure 6). The CN ligands are observed in a bent coordinating mode, with Fe–C–N angles of 118 and 120° for chains A and B, respectively, whereas the Fe–C–N bonds are nearly perpendicular to the porphyrin plane in the IsdI–heme–CN structure with Fe–C–N angles of 171° and 158° (20). The Fe–C–N observed for MhuD–heme–CN is more similar to the 139° angle in cyanide-inhibited rat heme oxygenase (rHO–heme–CN) (21). In the MhuD–heme–CN active site, the CN ligand forms an H-bond with Asn7 NH1. Furthermore, the CN-inhibited heme substrate is stabilized by hydrophobic interactions with Ile9, Phe23, Phe39, Val53, Thr55, Phe63, and Trp66; H-bonds between propionate 6 and Arg22 NH1, Arg26 NH2 and the Val83 backbone amide; and H-bonds between the bent propionate 7 and a water molecule (W1), which in turn H-bonds to Arg26 NH1.

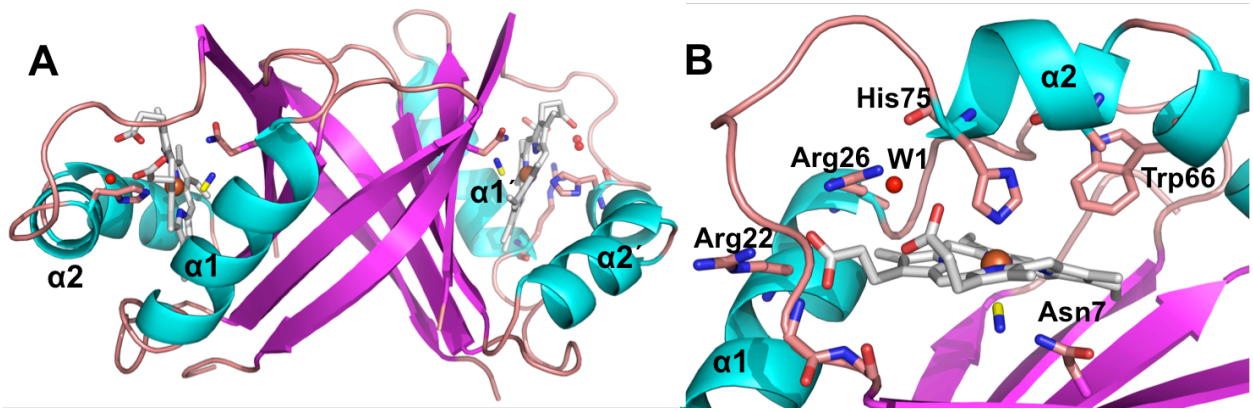

Figure 5. X-ray crystal structure of MhuD–heme–CN (PDB ID 4NL5). Panel A: Ribbon representation of the dimeric MhuD–heme–CN complex. Panel B: Ribbon representation of the MhuD–heme–CN heme-binding pocket. α-helices and β-strands are depicted in cyan and magenta, respectively. Loops and sidechain carbons are shown in salmon. All α-helices are labeled, with the second polypeptide chain differentiated by a prime symbol ('). Heme–CN, one per active site, is represented as a stick model where nitrogen, oxygen, heme carbon, and cyano carbon atoms are in blue, red, white, and yellow, respectively. Iron atoms and ordered water molecules are depicted as orange and red spheres, respectively

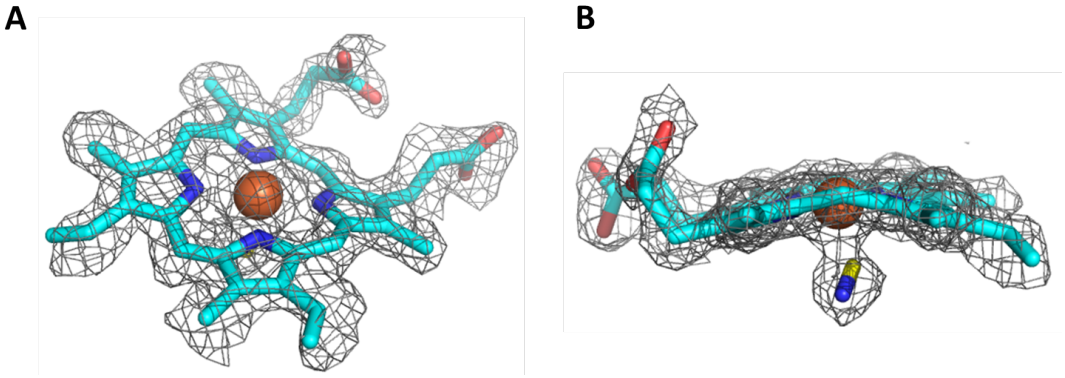

Figure 6. $2mF_o$-ΔF_c composite omit electron density map for bound heme–CN calculated with the heme and cyanide ligands omitted. The electron density mesh contoured at 1.0σ is colored in grey. Heme–CN is represented as a stick model, where nitrogen, oxygen, heme carbon, and cyano carbon atoms are in blue, red, cyan, and yellow, respectively. Iron is depicted as on orange sphere. Heme is shown in two confirmations to demonstrate that the (A) vinyl groups fit the density well and that the (B) porphyrin ring is distorted.

The MhuD–diheme and MhuD–heme–CN structures superimpose with a root-mean-square deviation (rmsd) of 0.29 Å over all Cα atoms (12). Within the active site pocket, the monoheme substrate overlays with the solvent-protected heme from the MhuD–diheme structure, however the modeled heme is rotated 180° about the α–γ axis (Figure 7A). Additionally, there is an increase in heme out-of-plane distortion. The distortions of the hemes in

MhuD–heme–CN (1.4 and 1.5 Å) are more severe than those in the solvent-protected heme from MhuD–diheme (0.7 Å), as analyzed by normal-coordinate structural decomposition (22). It was suggested that Phe22 contributes to heme ruffling in IsdG by contacting the γ-meso carbon.(23, 24) In MhuD–diheme and MhuD–heme–CN structures the sidechains of the corresponding residue, Phe23, overlay suggesting that Phe23 does not contribute to ruffling as the MhuD–diheme solvent-protected heme is planar as compared to the distorted MhuD–heme–CN porphyrin ring (Figure 7A). Although, the MhuD–diheme second solvent-exposed heme may play a role in the planar nature of the solvent-protected heme. The most notable structural differences are within the α2 helix and the subsequent loop region surrounding the active site (Figure 7A). In MhuD–heme–CN, the α2 helix is kinked after residue Asn68, while in the MhuD–diheme structure this helix (α2) is extended. This kink results in the movement of His75 (4.5 Å) so that it may coordinate with heme iron in the MhuD–heme–CN structure (Figure 7B). Furthermore, there is one notable altered sidechain conformation, Arg26, between the MhuD–heme–CN and MhuD–diheme structures. In the MhuD–diheme structure, Arg26 forms an H-bond with both propionate 6 (P6') and propionate 7 (P7') from the solvent-exposed heme, but not with the solvent-protected heme molecule (Figure 7C). However, in MhuD–heme–CN Arg26 is flipped into the reduced volume heme active site where it H-bonds to an active site water molecule (W1), which in turn H-bonds to the carbonyl oxygen of His75 and heme propionate 7 (P7), whose orientation is rotated toward the active site water (W1, Figure 7B). Further, Arg26 is also in H-bonding distance to heme propionate 6 (P6). One could speculate that the ordered water and alternative conformation of the Arg26 sidechain may stabilize the otherwise flexible loop to form a stable monoheme active site. The position of this water molecule is conserved in both active sites of the MhuD dimer, however in one of two active sties of the dimer there is a second water molecule that also H-bonds with Arg26 (not shown).

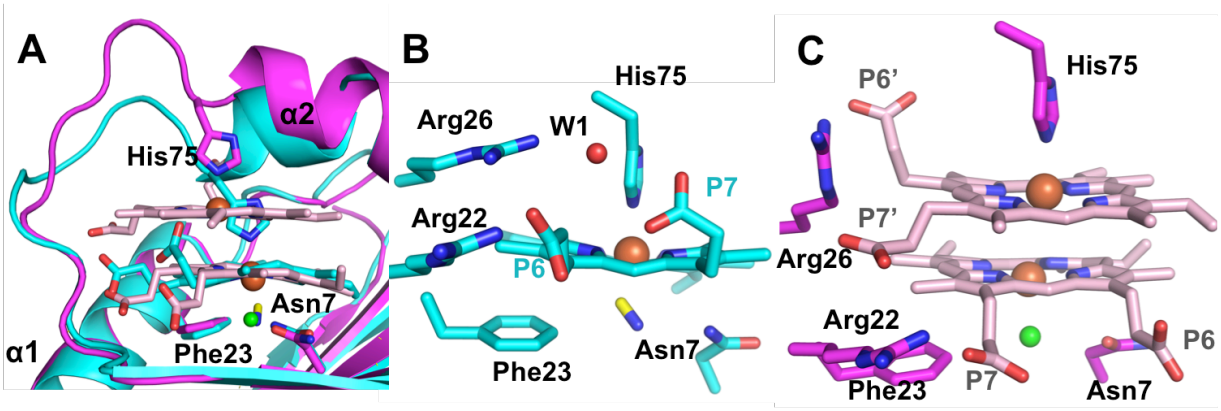

Figure 7. Structural comparison of the active sites of MhuD–heme–CN and MhuD–diheme. Panel A: Superposition of MhuD–heme–CN (PDB ID 4NL5, cyan) with MhuD–diheme (PDB ID 3HX9, pink) shows that α2 in MhuD–heme–CN is kinked while it is extended in the MhuD–diheme structure. This kink results in MhuD–heme–CN His75 moving 4.5 Å to coordinate heme iron (orange spheres). Panels B and C: The orientations of most residues within the MhuD–heme–CN (B) and MhuD–diheme (C) active sites are unchanged, but Arg26 in the MhuD–heme–CN structure flips into the reduced volume active site. Heme propionates 6 and 7 are denoted by P6 and P7, respectively, and the solvent-exposed heme propionates (MhuD–diheme) are differentiated with a prime symbol ('). Heme molecules are in stick representation, with carbon atoms depicted in light cyan and light pink for MhuD–heme–CN and MhuD–diheme, respectively, and nitrogen, oxygen and cyano carbon atoms in blue, red and yellow, respectively. Iron atoms, an ordered Cl^- atom (MhuD–diheme structure) and water molecule (W1, MhuD–heme–CN structure) are depicted as orange, green and red spheres, respectively.

Discussion

It has been suggested that the heme ruffling observed in structures of IsdG-like heme degrading proteins is required for their activity. The Cα atoms of the MhuD–heme–CN structure superimpose with those of N7A IsdG (PDB ID 2ZDO) and IsdI–heme–CN (PDB ID 3QGP) with rmsds of 1.59 and 1.24 Å, respectively (Figure 8A, IsdG is not shown for clarity) (20, 23). The 1.5 Å distortion of heme from planarity in the MhuD–heme–CN structure is less than that observed for N7A IsdG and IsdI–heme–CN, which show heme distortions of 1.9 and 2.3 Å, respectively (Figure 9). In contrast, the overall out-of-plane distortion (1.5 Å) and the degree of heme ruffling (1.4 Å) in MhuD–heme–CN is more than that observed in rHO-heme-CN (PDB ID 2E7E), which displays only 0.6 Å distortion from planarity and 0.5 Å ruffling (Figure 9). Finally, while the bound hemes of MhuD–heme–CN, IsdG, and IsdI occupy similar positions within their

respective structures, the entire heme molecule, and thus including the propionate groups, for both IsdG and IsdI are rotated approximately 90° about the axis normal to the tetrapyrrole ring compared to the MhuD-bound heme (Figure 8B). The positional difference within the heme molecules is dictated by the α1 helix and the loop region directly following the α2 helix. Within MhuD–heme–CN, the C-terminus of the α1 helix has an additional turn as compared to that of heme-bound IsdG/I, which enables Arg26 to form a H-bonding network with a water molecule, His75, and heme propionates 6 and 7. The last turn of this α1 helix is a loop region in both the IsdG and IsdI structures and, thus, the IsdG/I Arg26 Cα is displaced ~6 Å from the heme molecule, with its sidechain solvent accessible instead of participating within the active site, as observed for Arg26 in the MhuD–heme–CN structure (Figures 7B & 8B). Furthermore, the structural variance in the loop region directly following the α2 helix combined with that of the α1 helix results in the MhuD heme propionates pointing towards this loop region, whereas the corresponding propionates in IsdG/I point towards the α1 helix, reducing their solvent accessibility compared to those of MhuD (Figure 8). First, the kink observed in the α2 helix of MhuD–heme–CN is reminiscent of the corresponding helix in the *S. aureus* IsdG and IsdI heme-degrading proteins (23). Second and most importantly, the observed heme rotation may play a role in the variant location of tetrapyrrole ring cleavage during the heme degradation reaction of MhuD and Isd proteins, whereby the two products of IsdI suggest cleavage at the β- and δ-meso carbons and the products of MhuD suggest cleavage at the α-meso carbon (13, 24).

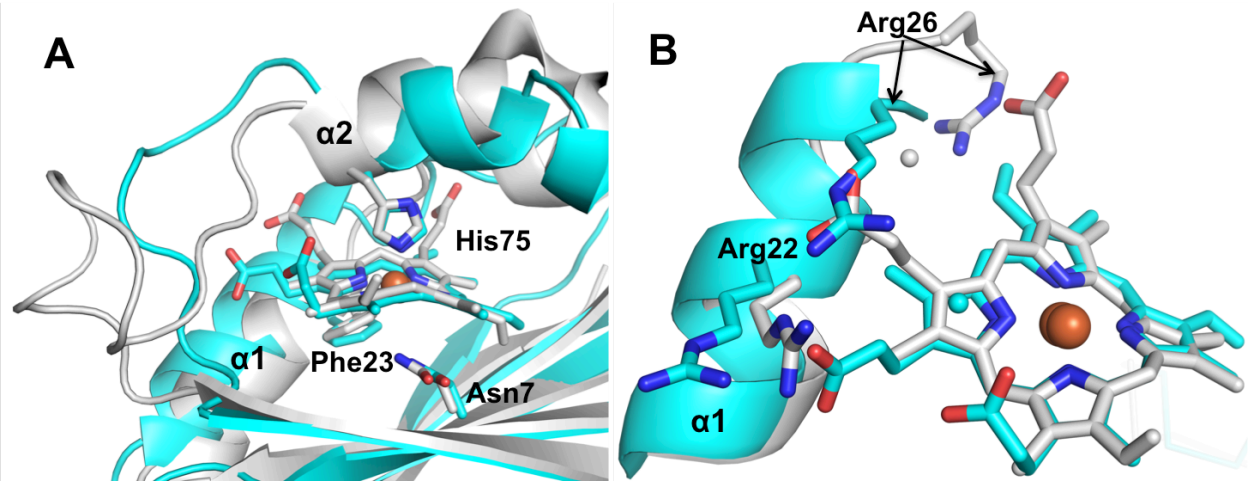

Figure 8. Structural comparison of MhuD–heme–CN with IsdI. Panel A: Superposition of MhuD–heme–CN (cyan, PDB ID 4NL5) with IsdI–heme–CN (white, PDB ID 3QGP) shows that the orientation of heme within the active site of MhuD is different compared to that of IsdG and IsdI, whereby the heme propionates in MhuD are rotated 90° around the axis normal to the heme plane. Panel B: The final turn of the α1 helix of MhuD (contains Arg26) is a loop region in the IsdI structure, enabling IsdI Arg26 to flip from within the active site as observed in MhuD to be surface exposed. Residue sidechains and heme molecules are represented as sticks, with oxygen, nitrogen, and iron atoms colored red, blue, and orange, respectively. Ordered water molecules are represented as spheres.

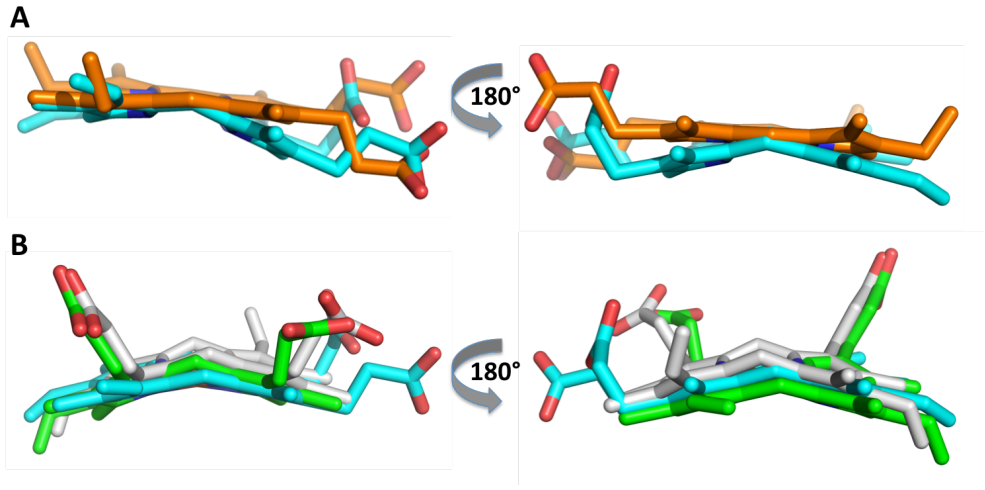

Figure 9. Degree of heme ruffling within heme degrading enzymes. Panel A: Overlay of MhuD–heme–CN (cyan) and rHO-heme-CN (orange, PDB ID 2E7E). Both panels show rHO heme being mostly planar. Panel B: Overlay of hemes from MhuD–heme–CN (cyan, PDB ID 4NL5), N7A IsdG (white, PDB ID 2ZDO) and IsdI–heme–CN (green, PDB ID 3QGP) following structural overlay.

Biological Relevence

M. tuberculosis, the causative agent of the human disease tuberculosis, acquires the vital nutrient iron from its host. Recently, it has been shown that *M. tuberculosis* can uptake host-derived heme as an iron source (25, 26). To date, four members of this pathway have been identified and characterized: Rv0203 (27), an excreted protein that binds extracellular heme; MmpL3 and MmpL11 (28), two membrane proteins that receive heme from Rv0203; and MhuD, which catalyzes the final step of *M. tuberculosis* heme acquisition and degradation (12). Recent studies strongly suggest that *M. tuberculosis* can use exogenous heme as an iron source in the absence of non-heme iron (25, 26, 29). *In vitro*, growth of an iron siderophore-deficient strain of *M. tuberculosis* can be recovered in the presence of heme or human hemoglobin (25). An iron siderophore-deficient strain of a similar species, *Mycobacterium bovis*, was able to survive in mice, suggesting that heme acquisition is also utilized by mycobacteria *in vivo* (29). These discoveries have led researchers to suggest that proteins from the *M. tuberculosis* heme acquisition and degradation pathway, including MhuD, represent promising new anti-mycobacterial targets (30)

References

1. Tenhunen R, Marver HS, & Schmid R (1969) Microsomal heme oxygenase. Characterization of the enzyme. *The Journal of biological chemistry* 244(23):6388-6394.
2. Ratliff M, Zhu W, Deshmukh R, Wilks A, & Stojiljkovic I (2001) Homologues of neisserial heme oxygenase in gram-negative bacteria: degradation of heme by the product of the pigA gene of Pseudomonas aeruginosa. *Journal of bacteriology* 183(21):6394-6403.
3. Zhu W, Wilks A, & Stojiljkovic I (2000) Degradation of heme in gram-negative bacteria: the product of the hemO gene of Neisseriae is a heme oxygenase. *Journal of bacteriology* 182(23):6783-6790.
4. Schuller DJ, Wilks A, Ortiz de Montellano PR, & Poulos TL (1999) Crystal structure of human heme oxygenase-1. *Nat Struct Biol* 6(9):860-867.
5. Schmitt MP (1997) Utilization of host iron sources by Corynebacterium diphtheriae: identification of a gene whose product is homologous to eukaryotic heme oxygenases and is required for acquisition of iron from heme and hemoglobin. *Journal of bacteriology* 179(3):838-845.
6. Drummond GS & Kappas A (1981) Prevention of neonatal hyperbilirubinemia by tin protoporphyrin IX, a potent competitive inhibitor of heme oxidation. *Proceedings of the National Academy of Sciences of the United States of America* 78(10):6466-6470.
7. Skaar EP, Gaspar AH, & Schneewind O (2004) IsdG and IsdI, heme-degrading enzymes in the cytoplasm of Staphylococcus aureus. *The Journal of biological chemistry* 279(1):436-443.
8. Wu R*, et al.* (2005) Staphylococcus aureus IsdG and IsdI, heme-degrading enzymes with structural similarity to monooxygenases. *The Journal of biological chemistry* 280(4):2840-2846.
9. Lee WC, Reniere ML, Skaar EP, & Murphy ME (2008) Ruffling of metalloporphyrins bound to IsdG and IsdI, two heme-degrading enzymes in Staphylococcus aureus. *The Journal of biological chemistry* 283(45):30957-30963.
10. Matsui T*, et al.* (2013) Heme degradation by Staphylococcus aureus IsdG and IsdI liberates formaldehyde rather than carbon monoxide. *Biochemistry* 52(18):3025-3027.
11. Reniere ML*, et al.* (2010) The IsdG-family of haem oxygenases degrades haem to a novel chromophore. *Molecular microbiology* 75(6):1529-1538.
12. Chim N, Iniguez A, Nguyen TQ, & Goulding CW (2010) Unusual diheme conformation of the heme-degrading protein from Mycobacterium tuberculosis. *Journal of molecular biology* 395(3):595-608.
13. Nambu S, Matsui T, Goulding CW, Takahashi S, & Ikeda-Saito M (2013) A New Way to Degrade Heme: The *Mycobacterium tuberculosis* Enzyme MhuD Catalyzes Heme Degradation without Generating CO. *J. Biol. Chem.* 288(14):10101-10109.
14. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248-254.
15. Battye TGG, Kontogiannis L, Johnson O, Powell HR, & Leslie AGW (2011) *iMOSFLM*: A New Graphical Interface for Diffraction-image Processing with *MOSFLM*. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 67(4):271-281.
16. Adams PD*, et al.* (2010) *PHENIX*: A Comprehensive Python-based System for Macromolecular Structure Solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66(2):213-221.
17. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of *Coot*. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66(4):486-501.
18. Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.0.

19. Falk JE (1963) *Chemistry and Biochemistry of Porphyrins and Metalloporphyrins. Comprehensive Bio- chemistry* (Elsevier Publishing Company, Amsterdam).
20. Takayama SJ, Ukpabi G, Murphy ME, & Mauk AG (2011) Electronic Properties of the Highly Ruffled Heme Bound to the Heme Degrading Enzyme IsdI. *Proc. Natl. Acad. Sci. U.S.A.* 108(32):13071-13076.
21. Sugishima M*, et al.* (2007) Alternate cyanide-binding modes to the haem iron in haem oxygenase. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* 63(6):471-474.
22. Jentzen W, Song X-Z, & Shelnutt JA (1997) Structural Characterization of Synthetic and Protein-bound Porphyrins in Terms of the Lowest-frequency Normal Coordinates of the Macrocycle. *J. Phys. Chem. B* 101(9):1684-1699.
23. Lee WC, Reniere ML, Skaar EP, & Murphy ME (2008) Ruffling of Metalloporphyrins Bound to IsdG and IsdI, Two Heme-degrading Enzymes in *Staphylococcus aureus*. *J. Biol. Chem.* 283(45):30957-30963.
24. Reniere ML*, et al.* (2010) The IsdG-family of Haem Oxygenases Degrades Haem to a Novel Chromophore. *Mol. Microbiol.* 75(6):1529-1538.
25. Tullius MV*, et al.* (2011) Discovery and Characterization of a Unique Mycobacterial Heme Acquisition System. *Proc. Natl. Acad. Sci. U.S.A.* 108(12):5051-5056.
26. Jones CM & Niederweis M (2011) *Mycobacterium tuberculosis* Can Utilize Heme as an Iron Source. *J. Bacteriol.* 193(7):1767-1770.
27. Owens CP, Du J, Dawson JH, & Goulding CW (2012) Heme Binding Protein Involved in *Mycobacterium tuberculosis* Heme Uptake. *Biochemistry* 51(7):1518-1531.
28. Owens CP*, et al.* (2013) The *Mycobacterium tuberculosis* Secreted Protein Rv0203 Transfers Heme to Membrane Proteins MmpL3 and MmpL11. *J. Biol. Chem.* 288(30):21714-21728.
29. Tullius MV, Harth G, Maslesa-Galic S, Dillon BJ, & Horwitz MA (2008) A Replication-Limited Recombinant *Mycobacterium bovis* BCG Vaccine Against Tuberculosis Designed for Human Immunodeficiency Virus-Positive Persons is Safer and More Efficacious than BCG. *Infect. Immun.* 76(11):5200-5214.
30. Owens CP, Chim N, & Goulding CW (2013) Insights on how the *Mycobacterium tuberculosis* heme uptake pathway can be used as a drug target. *Future Med. Chem.* 5(12):1391-1403.