

Structural and Biochemical Characterization of the Bacterial Virulence Effector NleC

By Michelle Marian Turco

B.S., University of Wisconsin–Madison, 2006

*A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
Of the requirement for a degree of
Doctor of Philosophy
Department of Chemistry and Biochemistry
2013*

This thesis entitled
“Structural and Biochemical Characterization of the Bacterial Virulence Effector NleC”
written by Michelle Marian Turco
has been approved for the Department of Chemistry and Biochemistry

Marcelo C. Sousa

Natalie G. Ahn

Date _____

The final copy of this thesis has been examined by the signatories and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

Turco, Michelle Marian (Ph.D., Chemistry and Biochemistry)

Structural and Biochemical Characterization of the Bacterial Virulence Effector NleC

Thesis directed by Associate Professor Marcelo C. Sousa

Pathogenic bacteria have developed numerous mechanisms to subvert host cell physiology and create unique niches that facilitate colonization. Within γ -proteobacteria, a specialized molecular machine, the Type III Secretion System (T3SS), is responsible for translocating bacterial proteins directly into host cells. These virulence factors, called effectors, rewire various host cell processes for bacterial gain and are the key for a successful infection. Because many effectors share little if any sequence similarity to proteins of known function, biological knockouts, biochemical assays, and structural studies are necessary to elucidate their functions in pathogenesis.

Enterohemorrhagic *E. coli* (EHEC) encodes its T3SS on a chromosomal pathogenicity island that also contains effectors essential for its unique attaching and effacing phenotype when infecting epithelial cells. However, numerous effectors outside of this pathogenicity island are required to suppress inflammation and the host immune response, and are thus crucial for infection. Identified as a zinc protease of NF κ B transcription factors, NleC is one such effector that is essential for EHEC suppression of inflammatory cytokine secretion.

Here, I show that NleC cleaves NF κ B transcription factors with high specificity, recognizing residues proximal to the cleavage site in the NF κ B DNA-binding loop. However, elements outside of this site are also necessary for recognition by NleC making it an

unusual protease with exquisite selectivity. I determined the crystal structure of NleC revealing that despite undetectable sequence similarity to other Zn^{2+} proteases, it is a member of the Zincin zinc protease fold superfamily. Nevertheless, NleC displays distinct structural differences from currently known Zincins that may contribute to its extraordinary selectivity and the unique folding/refolding requirements implicit in translocation. Finally, I propose that NleC recognizes the $\text{NF}\kappa\text{B}$ DNA-recognition sequences by mimicking DNA in both shape and electrostatic character. This would make NleC the first example of both a DNA-mimicking T3SS effector, and a DNA-mimicking protease, representing a truly unique mechanism for subverting host signaling pathways for the benefit of the pathogen.

Dedication

To my mother, Jennifer, for fostering creativity. To my father, Gregory, for exemplifying scholarship.

Acknowledgments

I would like to thank my thesis advisor, Marcelo Sousa, for encouraging the problem solving and critical thinking that culminated in this work. I would also like to thank my other committee members, Natalie Ahn, Rob Knight, Amy Palmer, and Thomas Perkins for their helpful suggestions and thoughtful comments. The Type III Secretion System project was initiated in Marcelo Sousa's lab by Amy Dear, with whom I worked during my rotation in the lab. I would like to thank Amy Dear for helping me begin work on the project, and sharing her knowledge.

In addition, I would like to recognize David McKay for his assistance and expertise in X-ray crystallography. Theresa Nahreini assisted with preliminary cell culture experiments. William Old and Dan Gu ran and helped analyze my mass spectrometry experiments. Geoffrey Armstrong facilitated the preliminary NMR experiments. Cristina Sandoval assisted me in CD experiments. Andrea Edwards and Jeremy Trausch taught me how to use the ITC and analyze data.

I would like to further acknowledge the Sousa lab for general discussion and problem solving during my project. I would also like to thank my neighbor in the lab for the first several years, Andrea Edwards of the Batey lab, for many helpful discussions over the years. Finally, I would like to thank Clay Horiuchi for moral support and thoughtful suggestions during my final years in graduate school.

Table of Contents

Chapter I. Introduction	1
Human Host Defense Against Microbial Invasion.....	1
Gamma-proteobacteria and the Type III Secretion System.....	5
Enterohemorrhagic Escherichia coli	8
NleC Function in the Pathogenesis	12
Chapter II. Specificity of the T3SS effector NleC.....	18
Introduction.....	18
Results.....	19
Discussion.....	31
Materials and Methods	42
Chapter III. Structure of the T3SS effector NleC, a new Zincin protease	48
Introduction.....	48
Results.....	49
Discussion.....	66
Materials and Methods	80
Chapter IV. Discussion	91
The Role of NleC During Pathogenesis.....	91
Secretion as a Unique Evolutionary Pressure on Structure	112
Summary	122
Appendix A. NleC and Putative Chaperones.....	124
Appendix B. Map, Tir and CesT	130
An Exploration of Proteins on the LEE	130
Biologically relevant structure of T3SS multi-effector chaperone CesT	131
Map nucleates multimeric substrate complexes in host cells	142
References.....	147

Table of Tables

Table II-1 Summary of putative NleC cleavage sites in human p300.....	41
Table II-2 Plasmids used for the study of the specificity of NleC for NFκB and p300	43
Table III-1 NleC Crystallographic Statistics.....	55
Table III-2 Plasmids prepared for the study of the structure of NleC.....	80
Table A-1 Plasmids used in the study of NleC and putative chaperones.....	128
Table B-1 CesT Crystallography Statistics	135
Table B-2 Plasmids used for the study of CesT, Map and Tir.	142

Table of Figures

Figure I-1 NF κ B Activation Pathways.....	4
Figure I-2 Similarities between the flagella and the Type III Secretion System.....	6
Figure I-3. Non-LEE Effectors of EHEC that Impact the Inflammatory response.	12
Figure II-1 Location of NleC cleavage site in RelA.....	21
Figure II-2 Cleavage rate varies with RelA concentration.	24
Figure II-3 Comparison of cleavage rates between NF κ B subunits.....	24
Figure II-4 SDS-PAGE analysis of RelA cleavage by NleC.	25
Figure II-5 Effect on NleC catalytic efficiency of RelA mutagenesis.	28
Figure II-6 Structure of DB loop with mutagenized residues highlighted.....	29
Figure II-7 Pulldown of RelA by NleCE184A.....	30
Figure II-8 SDS-PAGE analysis of p300 interaction with NleC.....	31
Figure II-9 DNA-binding residues of RelA.	35
Figure II-10 Conserved residues among NF κ B.....	36
Figure II-11 Alignment of NF κ B subunits Rel-homology domain.	42
Figure III-1 NleC truncations during crystallization.	51
Figure III-2 Electron density map surrounding Zinc in the active site.....	53
Figure III-3 Structure of NleC.	54
Figure III-4 NleC active site.	57
Figure III-5 NleC β -sheet.....	58
Figure III-6 Conservation within the M85 NleC peptidase family.....	59
Figure III-7 Structures and topography diagrams of NleC and closest structural relatives.	61
Figure III-8 Active site tyrosine among Zincins.....	62
Figure III-9 Small angle X-ray scattering representative curves.	64
Figure III-10 NleC SAXS envelopes.....	65
Figure III-11 Zincin Subfamilies.....	67
Figure III-12 Comparison of Zincin topology.	70
Figure III-13 Structural basis for functional and non-functional truncations of NleC.	73
Figure III-14 Electrostatic surface potential of NleC.....	74
Figure III-15 Comparison of Ψ -Loop and NleC β -sheet.	76
Figure III-16 Maintenance of Zn ²⁺ coordination sphere during translocation.	79

Figure III-17 Conservation within the NleC M85 protease family.	84
Figure IV-1 Electrostatic surface potential of NleC and substrates.	95
Figure IV-2 Conserved residues among NF κ B overlap with DNA-binding residues.	98
Figure IV-3 DNA-binding residues of RelA and p50.	99
Figure IV-4 NleC active site cleft comparison with DNA showing electrostatics.	100
Figure IV-5 Model of DNA-like mode of binding for NleC bound to RelA.	101
Figure IV-6 Proposed model for the complex formation of NleC, NF κ B, and I κ B.	103
Figure IV-7 Interaction between putative NleC-NF κ B-I κ B complex.	103
Figure IV-8 Comparison of the DNA-binding domain of RelA and NFAT1.	106
Figure IV-9 Representative AFM trace of Cys-Nugx4-NleC1-280.	115
Figure IV-10 Hypothesized structure during translocation.	116
Figure IV-11 Predicted topology of T3SS NleD compared to NleC.	119
Figure A-1 NleC pulldown of EHEC proteins.	125
Figure A-2 NleC and DnaJ do not form a complex.	126
Figure B-1 Overlay of T3SS chaperone structures.	132
Figure B-2 Effector binding mode.	133
Figure B-3 CesT crystal structure.	134
Figure B-4 CesT altered helix and sixth strand.	136
Figure B-5 HSQC overlap of free N15-CesT and N15-CesT titrated with 2:1 molar ratio of Tir1-68 (pMS 932).	138
Figure B-6 HSQC overlap of free N15-CesT and N15-CesT titrated with 1:1 molar ratio of Tir1-93 (pMS 953).	139

Abbreviations

AFM: atomic force microscopy
A/E: attaching and effacing
BME: β -mercaptoethanol
CBD: chitin binding domain
DBD: DNA-binding domain
DB loop: DNA-binding loop
EDTA: ethylenediaminetetraacetic acid
EHEC: Enterohemorrhagic *Escherichia coli*
EPEC: Enteropathogenic *Escherichia coli*
I κ B: inhibitor of κ B
IL-8: interleukin 8
IPTG: isopropyl- β -D-thiogalactopyranoside
ITC: isothermal calorimetry
LEE: locus for enterocyte effacement
LPS: lipopolysaccharide
MAD: multiple wavelength anomalous dispersion
Map: mitochondrial associated protein
Nle: Non-locus of enterocyte effacement effector
NF κ B: nuclear factor κ B
PRR: pattern recognition receptor
SAXS: small angle X-ray scattering
SUMO: small ubiquitin-like modifier
T3SS: type three secretion system
TCEP: tris(2-carboxyethyl)phosphine
Tir: translocated intimin receptor
TLR: toll-like receptor
TNF- α : tumor necrosis factor α

Chapter I. Introduction

Beginning the era of modern medicine, the discovery of antibacterial agents forever changed the way doctors approached and treated communicative disease. In recent decades, however, antibiotics represent a decreasing portion of investment for pharmaceutical firms because financial returns per drug are limited as the rate of antibiotic resistance increases¹. Traditional antibiotics fight disease by preventing replication or causing death, targeting commensal and pathogenic bacteria alike. The considerable number of microbes affected by traditional antibiotics translates to a larger population searching for and evolving resistance to these drugs; thus, there is a greater propensity that pathogens will acquire resistance by horizontal gene transfer than if a smaller population was targeted^{1,2}. Hypothesizing that by targeting a smaller population, a lower evolutionary pressure for the development of resistance genes results, our lab focused on characterizing components of the pathogenic machinery³⁻⁵. This thesis concentrates on one of the virulence effector proteins translocated by Enterohemorrhagic *Escherichia coli* (EHEC) Sakai strain and its mode of function in host epithelial cells.

Human Host Defense Against Microbial Invasion

When challenged with a microbial invasion, organisms have a variety of defenses to maintain integrity. In all organisms, physical barriers often act as the first line of defense in conjunction with chemical barriers. In the gastrointestinal tract, mucus impedes the movement of pathogens, and the commensal microbial population serves as an additional

barrier, by resisting displacement physically and chemically⁶⁻⁸. If these barriers are breached, vertebrates, such as humans, utilize both an innate and the more recently evolved adaptive immune responses.

The innate immune system reacts quickly and nonspecifically, causing the responses that make humans feel sick. Bacterial pathogens are recognized by pattern recognition receptors (PRRs) on host cells, initiating a signaling cascade that leads to secretion of various cytokines^{9,10}. Upon secretion of these chemical factors, the inflammatory response is initiated, increasing blood flow and recruiting various immune cells to the site, such as leukocytes and lymphocytes. Induction of the complement system leads to protein binding to the targeted pathogen and insertion of a pore, the membrane attack complex, killing the cell. Bound complement protein recruits macrophages to ingest the targeted cell¹¹. Antimicrobial peptides are another branch of the innate immune system and are either produced constitutively or upon infection. These peptides are varied in structure and have multiple modes of action, but most often cause death by disrupting the membrane of the pathogen¹². The advantages of the innate immune system are that it is widespread and rapid, but pathogens have evolved ways to circumvent many of these general tactics, in a biological arms race¹³.

The adaptive immune response is a more recently evolved system and is antigen-specific, having “memory.”¹⁴ Many components of the innate immune system, such as complement proteins and antigen-presenting macrophages, can work in conjunction with the adaptive immune system to create a stronger response^{11,14}. The success of the adaptive immune system hinges on its ability to recognize the difference between host and pathogen and depends on the actions of two classes of lymphocytes, T cells and B cells. The adaptive

system is the basis for vaccines, as a population of both T and B cells will be converted to memory cells after clearing the infection in order to respond more rapidly upon subsequent invasion by the same pathogen¹⁴. Bacteria and viruses evade these systems by rapid evolution of surface proteins, evading identification by T and B cells. The inflammatory response is a key primary event that triggers the initiation of subsequent immunity events, including the adaptive immune response^{15,16}.

When a host cell comes in contact with an invading pathogen, PRRs are the first to recognize the pathogen and respond⁹. Of all the PRRs, Toll-like receptors (TLRs) represent the biggest class, having the most variety in ligands, and can be localized to either the plasma membrane or endosomal membranes after pathogens have been phagocytized^{9,17}. Within the cell, recognition of bacterial lipoproteins, lipopolysaccharide (LPS), and flagellin by TLR dimers activate signaling cascades through one of four adaptor proteins. The adaptor protein MyD88, is utilized most often and leads to expression and secretion of proinflammatory cytokines. The other adaptor molecules are used to trigger different combinations of transcription factors depending on the stimuli. A phosphorylation cascade eventually leads to the phosphorylation of I κ B- α , marking it for ubiquitination and destruction, releasing the sequestered NF κ B subunits (Figure I-1) ^{9,10,17,18}. The nuclear localization signal that was hidden by binding to I κ B then becomes exposed, causing the nuclear import of the NF κ Bs¹⁹. In the nucleus, NF κ B bind to the promoter region of DNA, alone or in conjunction with other transcription activators, to enhance expression of specific genes. There are two classes of NF κ B transcription factors, Class I and Class II. Class I consists of p50 and p52, which are expressed as much larger proteins that require modification to become functional. RelA, RelB, and c-Rel are in Class II, and are expressed

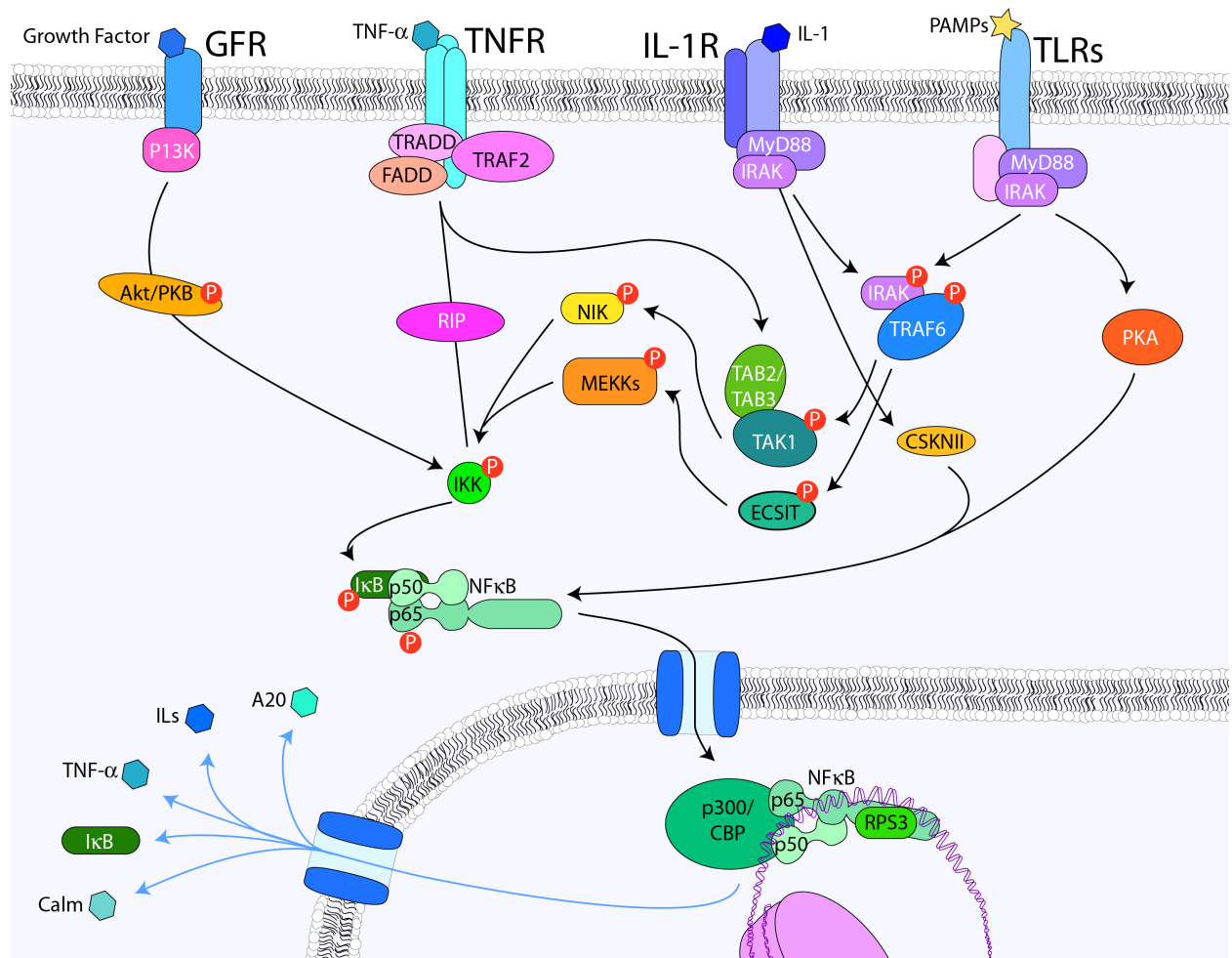


Figure I-1 NFκB Activation Pathways. NFκB-mediated transcription is initiated by a variety of external signals and receptors. The pathways converge at the point of phosphorylating the IKK kinases, which in turn phosphorylates the NFκB inhibitor, IκB. This marks IκB for proteosomal degradation and releases the NFκB dimers for import into the nucleus. There, they may associate with various transcriptional activators like CBP or p300. This promotes the expression of various proinflammatory cytokines, like TNF-α and interleukins.

in functional forms. The sequence conservation in the Rel-homology domain between subunits within a Class is >50%, whereas with between Class I and Class II is <50%. This sequence variation in part permits the subunits to recognize different DNA sequences, in both homo- and heterodimers²⁰. NFκB activation is responsible for the up-regulation of tumor necrosis factor-α, interleukin-1, interleukin-6 and interleukin-8, all of which are implicated in the acute inflammatory response^{9,21,22}. Additionally, NFκB activation inhibits apoptosis upon TNF-α signaling in nonimmune cells, and is important for signaling in adaptive immune cells, B- and T-cells²³.

Importance of NFκB in immune response

The correct functioning of the NFκB signaling cascade and the subsequent release of pro-inflammatory cytokines is essential for a rapid response to invading pathogens. Released from cells that have contacted invading microbes, these pro-inflammatory cytokines serve to act, in part, as chemoattractants to draw immunity cells to the site of infection^{9,22}. If the process is hindered at any point, microbes can begin replicating, resulting in a more serious infection. The rapidity of the inflammation onset means that it is essential for a complete immune response to occur. Because of this, the inflammatory response in general, and NFκB regulation in particular, are attractive targets for bacteria and viruses²⁴.

Gamma-proteobacteria and the Type III Secretion System

As microbes evolved, seeking and creating new survival strategies and environmental niches, some microorganisms formed dependent relationships with other organisms, either through symbiosis or parasitism. One type of bacteria, γ-proteobacteria,

developed a specialized secretion system to control its host, which appeared concurrently as the evolution of motility *via* the flagella²⁵⁻²⁸. The Type III Secretion System (T3SS) is believed to have shared a common ancestor with the flagellar system, appearing concurrently during evolution, and has numerous homologous protein components²⁵⁻²⁸. The basic machinery of both systems is highly similar, from the ATPase and pores that are embedded in the inner and outer membranes of the bacteria, to the protein subunits that make up the flagella and secretion needle (Figure I-2)²⁸.

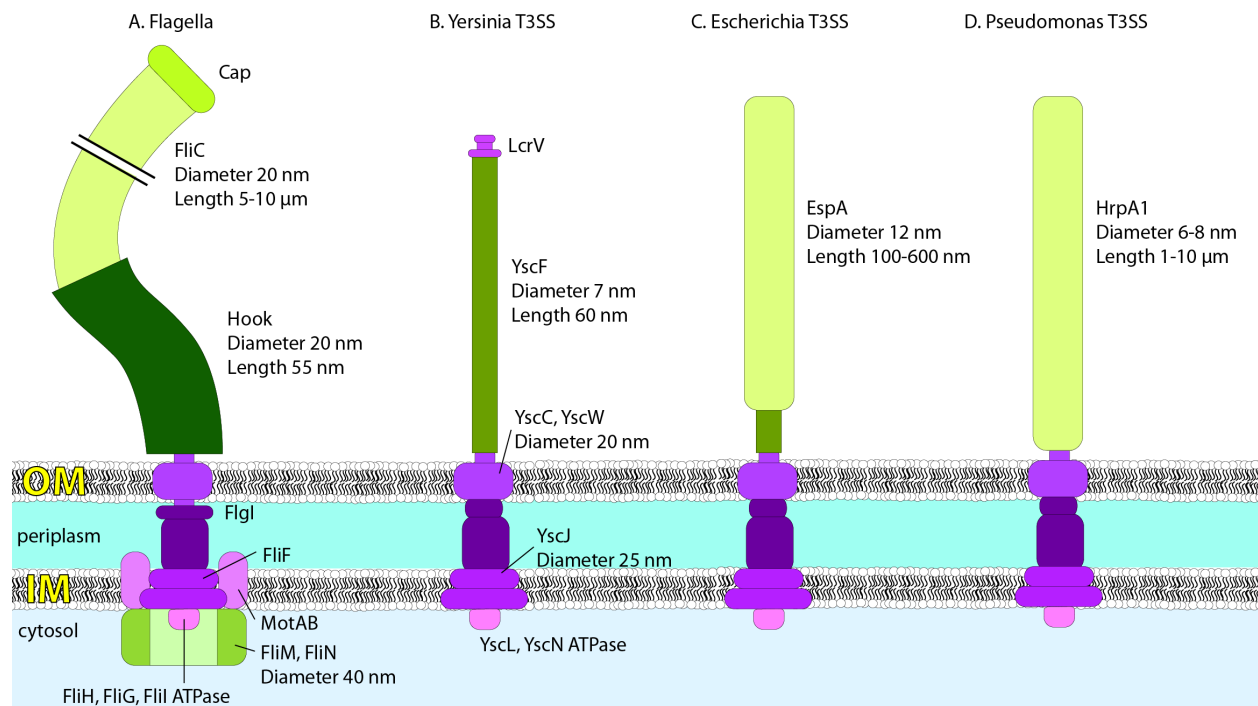


Figure I-2 Similarities between the flagella and the Type III Secretion System. The flagella and T3SS share many structural elements, mostly in the proteins that are embedded in the membrane. Other similar proteins not pictured are chaperones. The Yersinia T3SS is the most well described, and there can be variations between species in the structures of the external needle.

The T3SS itself has evolved into seven different families of related complexes. A large macromolecular complex, the *Escherichia coli* injectisome is in the same family as that

of *Salmonella typhimurium* SPI-2, *Yersinia pestis*, *Citrobacter*, and *Edwardsiella*²⁸. The injectisome consists of an inner ring complex that is embedded in the inner membrane of the bacteria and is associated with a cytosolic ATPase that provides the energy for secretion. Spanning the periplasm, a rod-like structure permits the secreted proteins passage from the inner membrane ring to the outer membrane, where a similar ring structure is embedded in the outer membrane. Through this hollow conduit, proteins are secreted from the cytosol of the bacteria to the extracellular space and assemble the outer needle structure of the injectisome. The outer needle morphology varies depending on the T3SS family, and in *E. coli*, is a filament that terminates in a pore. This pore is embedded into the host cell upon contact, opening the injectisome for direct translocation of proteins into host cells. The interior diameter of the injectisome varies along the length of the passageway, but at the narrowest point is roughly 25 nm for *E. coli*. This corridor is too constricted for most of the translocated proteins to pass through fully folded^{28,29}.

The second vital components of the translocation system, shared with flagella, are the chaperones. These chaperones bind proteins before they are translocated, whether they are components of the assembling needle, or are translocated as part of pathogenesis^{28,30}. Chaperones are also hypothesized to prevent premature assembly or aggregation within the bacterial cell, as well as target the proteins to the base of the T3SS. Many of the proteins known to associate with chaperones cannot be efficiently translocated in their absence^{31,32}. Both the chaperones that bind needle components and translocated proteins are small proteins that associate in homodimers. Whether they bind a single protein or associate with multiple proteins, the chaperones have conserved folds. However, there is no sequence conservation among the chaperone-binding domains of the different

effectors, which are all isolated at the amino-terminus^{28,30,33-35}. Studying a multi-effector chaperone will help determine how these chaperones distinguish between their varying substrates.

The proteins that are secreted by the T3SS into host cells are called effectors, because they effect changes within the host cell that make the environment more hospitable for the bacterium^{24,36}. Effectors are often shared among bacteria either *via* horizontal gene transfer, or by phage-mediated transduction³⁷⁻⁴¹. The functions of some effectors are apparent by shared sequence similarity with known eukaryotic or bacterial proteins, whereas others have unknown function because of their lack of sequence similarity. The absence of clear similarity to any other proteins besides homologous effectors has impeded efforts to study these effectors and determine their functions during pathogenesis. Besides identifying a measurable phenotype in cellular assays, the functions of these effectors can sometimes be elucidated by characterizing their structures. Through structural studies, it has been determined that some effectors are divergently derived and are homologous to known proteins of similar function⁴²⁻⁴⁴. Other effectors have convergently evolved and have the same functions as known proteins, but with a unique structure^{39,42,45-47}. Understanding how effectors function in pathogenesis is essential as they are the workhorses of T3SS virulence, with all other components only serving as the delivery system for these proteins.

Enterohemorrhagic Escherichia coli

While there are γ -proteobacteria that infect plants and other animals, a subset of bacteria are of special interest, as they are capable of infecting humans. One of the most well known modes of infection is through the gastrointestinal tract, a problem that affects

the developing and developed worlds alike⁴⁸. Attaching and effacing (A/E) pathogens are named for their phenotype of intimate attachment to intestinal epithelial cells and subsequent effacement of the microvilli^{36,49,50}. Within the A/E family of γ -proteobacteria, Enterohemorrhagic *Escherichia coli* (EHEC) is the most common subtype of Enteropathogenic *E. coli* (EPEC) to cause human disease. Infection by members of the A/E family causes diarrhea no matter the subtype; however, EHEC also releases Shiga toxins after the initial bacterial attachment⁵¹. As N-glycosidases, Shiga toxins impede host protein biosynthesis by cleaving an adenine moiety from the 28S RNA of the ribosome. This results in more severe infection with bloody diarrhea manifesting in hemorrhagic colitis, and especially among children, can cause potentially fatal hemolytic uremic syndrome⁵²⁻⁵⁴. There are 73,000 illnesses annually in the United States from EHEC O157:H7, and around 2% of these are from carbapenem-resistant strains, which are especially difficult to treat^{55,56}.

Because of the importance of understanding EHEC for human health, the course and cause of infection is well studied. In all EPEC, the genes for the T3SS machinery, chaperones, and a few important effectors are found on a large 92.7 kb plasmid within a region called the Locus for Enterocyte Effacement (LEE)⁵⁷. The genes were thought to be sufficient for virulence; however, it became apparent that there were other effectors outside of the LEE. One subtype of EHEC, the Sakai strain (EHECS), has been sequenced and mined for other putative effectors outside of the LEE³⁷. The Sakai strain became the focus of intense study because it caused the deaths of twelve schoolchildren and more than 9000 cases in an outbreak in Sakai, Japan in 1996⁵⁸. In order to understand disease caused by EHECS on a cellular level, it is essential to study all effectors, both LEE and non-LEE (Nle).

Located in the LEE, intimin and Tir were some of the first T3SS proteins characterized outside of the machinery itself. Intimin is not an effector, but is embedded in the membrane of EPEC cells and functions in creating the intimate attachment seen in the A/E phenotype^{59,60}. Originally, it was believed that intimin recognized and bound a host cell receptor in order to create the intimate attachment. However, it was later discovered that EPEC translocated the receptor itself, the effector known as Tir (translocated intimin-receptor)⁶¹. Without intimin and Tir, EPEC is easily wiped out by the immune response^{32,60}. Tir, a large 400-kDa protein, has cytosolic domains which function in cytoskeletal rearrangement, in addition to the membrane-spanning portion⁶¹. A more recently discovered function for Tir is its ability to suppress the NF κ B pathway by targeting TNF- α receptor associated factor (TRAF) and promoting its degradation⁶².

Another important LEE effector is called Map (mitochondrial associated protein), because the original assays showed this effector localized to the mitochondria⁶³. Map was also shown to have an effect on cytoskeletal rearrangements, causing initial ruffling of the host epithelial cell membrane upon association of EPEC⁶⁴. T3SS proteins found in other bacteria were found to share a conserved motif, WxxxE, with Map that was thought to be important for its virulence^{65,66}. Activating the small GTPase cdc42 caused a similar membrane ruffling phenotype, and subsequently Map was hypothesized to structurally mimic active small GTPases, without the requirement to bind ATP⁶⁵. However, complex formation between Map and cdc42 was later demonstrated and a structure was obtained showing that Map acts as a guanidine exchange factor and not as an independent GTPase mimic⁴⁷. Map also has a carboxy-terminal PDZ-binding motif that acts to localize Map to the apical membrane by associating with Ebp50, a scaffold molecule with two PDZ domains⁶⁵.

This is a common theme in effector proteins and serves to target the effector in the host cell to the substrate location²⁴.

Outside of the LEE, many effectors have been inserted by horizontal gene transfer events or transduction with phage³⁷. Several of these are important for virulence as they suppress the host inflammatory response upon infection (Figure I-3). NleH1 and NleH2 are closely related in sequence and both bind ribosomal protein S3 (RPS3). However, their activities slightly differ, with NleH1 inhibiting the nuclear trafficking of RPS3 and inhibiting its IKK β -mediated phosphorylation. This inhibits its ability to associate with other NF κ B in the nucleus and promote the expression of proinflammatory cytokines. NleH2 is not observed to have this function, and in fact, slightly increases the phosphorylation of RPS3 by phosphorylating it directly. The PDZ-binding motif of NleH1 and NleH2 also seems to influence the inhibitory or stimulatory effect on NF κ B inhibition⁶⁷.

NleE and NleB similarly work to inhibit the activation of NF κ B upstream of I κ B phosphorylation. A more potent inhibitor of the inflammatory response, NleE prohibits I κ B phosphorylation by preventing the activation of kinase TAK1, which is present in IL-1B receptor, TLR and TNFR-activated pathways. NleE functions by methylating a cysteine in the zinc finger domains of TAB2 and TAB3, inhibiting them from promoting the self-phosphorylation of TAK1. This activity prevents I κ B from being phosphorylated, and retains NF κ B in the cytosol, terminating the signaling cascade that would conclude in proinflammatory cytokine secretion. The addition of NleB supports the activity of NleE, but only inhibits NF κ B activation through the TNF- α pathway, suggesting a different mode of action^{68,69}.

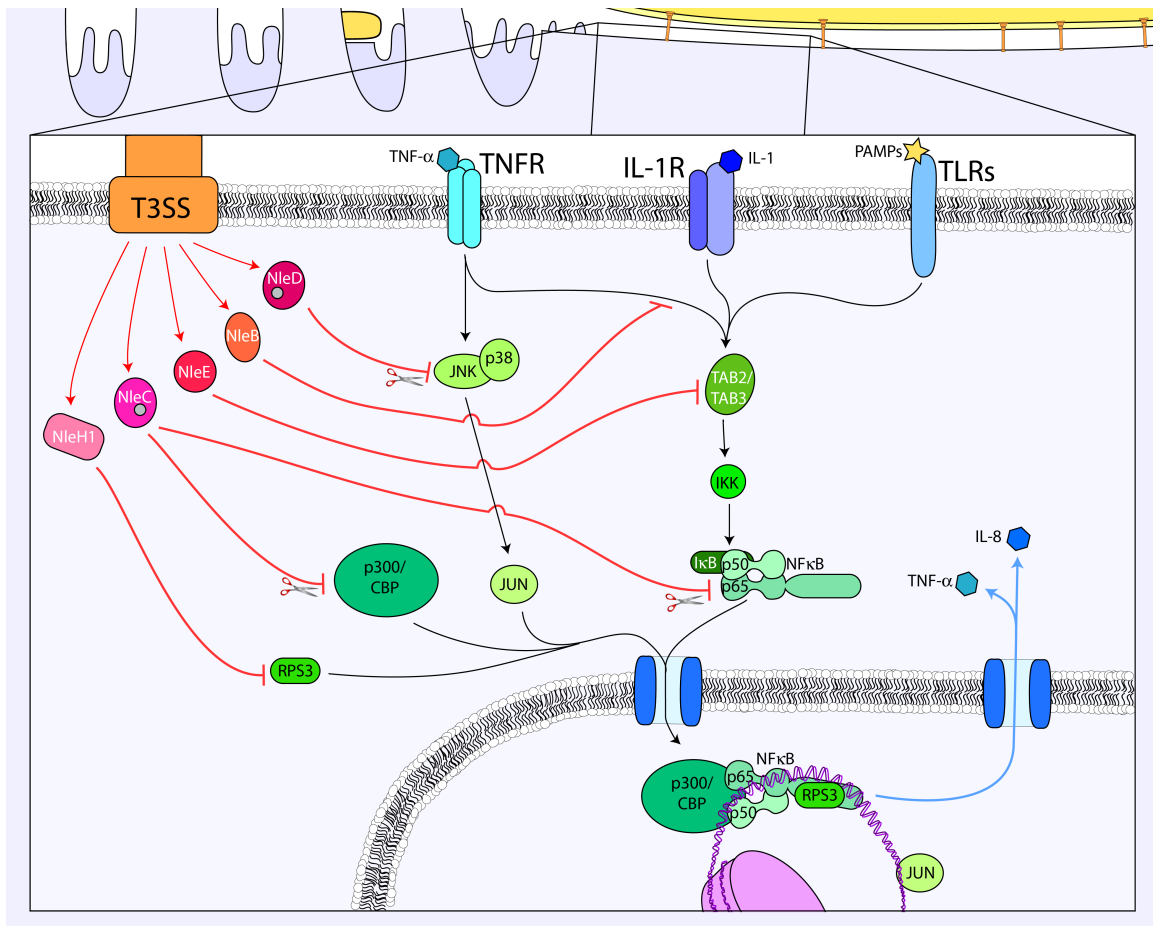


Figure I-3. Non-LEE Effectors of EHEC that Impact the Inflammatory response. EHEC secretes various effectors that are located outside of the LEE pathogenicity island, which are responsible for controlling the inflammatory response for the benefit of the bacteria. The action of these effectors results in reduction in secretion of proinflammatory cytokines, such as IL-8 and TNF- α .

NleC Function in the Pathogenesis

NleD is a zinc protease that cleaves c-Jun N-terminal kinase (JNK), preventing it from phosphorylating the AP-1 transcription factor, c-Jun. JNK is activated by a wide range of signals and is thus an important signaling mediator in inflammation, as well as proliferation and apoptosis. In the process of studying NleD and its zinc protease activity, NleC was discovered to be responsible for the abrogation of IL-8 secretion in TNF- α stimulated cells⁷⁰. Though these proteins share no sequence homology outside of a

conserved zinc-binding motif, they were both revealed to be the first zinc protease effectors identified to date.

Several groups recently published research identifying NleC as a zinc protease that abrogates the activation of NFκB and thus the secretion of IL-8, depressing the inflammatory response. Utilizing both transfection of NleC-containing plasmids directly into mammalian cells, as well as infection experiments with EPEC strains, NleC was shown to impact levels of the NFκB subunit RelA, as well as other NFκB subunits, p50 and c-Rel. Other host signaling proteins, STAT1 and Erk, were unaffected by the presence of NleC, suggesting that it is a specific protease for NFκB. The proteolytic activity of NleC was implicated in reducing NFκB levels by direct proteolysis *in vivo* through experiments that inhibited the proteasome. In addition, NFκB levels were rescued by mutagenesis of the conserved zinc-binding motif of NleC as well as with the addition of Ethylenediaminetetraacetic acid (EDTA)⁷⁰⁻⁷³. Taken together, these experiments illustrate that NleC directly proteolyzes NFκB transcription factors. Truncations of NleC to identify the minimal fragment of the protein necessary for function found that upon truncating NleC residues 1-32 and 238-330, NFκB suppression was lost. Combined with the loss-of-function results, binding experiments identified a feature of NleC within residues 33 to 66 that recognizes NFκB directly⁷².

In order to identify the cleavage site on RelA, antibody binding and N-terminal sequencing experiments were utilized. On a Western blot, only antibodies against the C-terminal portion of RelA bound to the NleC-cleaved RelA. Using Edman degradation, the cleavage site was identified between residues 38 and 39 with the sequence 38 – C|EGRS – 42⁷⁰. Another group found that a C-terminal portion of NFκB accumulated when the

proteasome was inhibited, locating the cleavage site between residues 10 and 11 with the sequence 10 – P|AEPA – 14⁷¹. Therefore, the cleavage sites identified by these two groups are inconsistent^{70,71}.

The localization of NleC in cells during infection is also in contention, and is dependent on the system used to identify the localization. When NleC was introduced to mammalian cells *via* transfection, NleC is localized to the nucleus. However, when NleC was translocated during infection by a T3SS-competent *E. coli* to mammalian cells, NleC was localized to the cytosol, and more specifically, can be seen beneath T3SS-induced pedestals. In a mouse cecal loop model, NleC similarly localized to the apical membranes at twelve hours post-infection^{70,74,75}.

The activity of NleC is complemented by other T3SS effectors in controlling the inflammatory response, but is responsible alone for cleaving NFκB. Utilizing a triple mutant of ΔNleEB/ΔNleC, the ability of EHEC to inhibit NFκB activation and the subsequent IL-8 secretion was destroyed. Upon the reintroduction of NleC, the wildtype infection phenotype was complemented for reduced nuclear RelA and the presence of cleaved RelA⁷⁰. In a time-lapse infection assay, wildtype levels of NleC did not decrease NFκB until 6 hours post infection, whereas in the absence of NleE, which prevents the degradation of IκB, NFκB levels were decreased after only four hours. This suggests that interplay of NleC and NleE during infection modulates NFκB levels in host cells^{73,74}.

In vivo studies established the effect of NleC on the success of bacterial infection. Wildtype and Δ*nleC* *C. rodentium*, an A/E pathogen, were used to infect mice. Δ*nleC*-infected mice did not differ from wild-type with respect to bacterial load during the course of infection, but did result in a higher degree of colitis in the mice, consistent with NleC's role

in suppressing inflammation. By using a cecal loop model, they also determined that $\Delta nleC$ *C. rodentium*-infected epithelial cells had a greatly increased expression of inflammatory cytokines, which would contribute to the colitis phenotype⁷⁴.

Beyond NF κ B subunits, there have been reports of additional putative targets of NleC. When studying whether NF κ B degradation was prevented by the inhibition of the proteasome, I κ B levels were found to be decreased in the presence of NleC⁷². In the normal activation of the signaling pathway, I κ B is phosphorylated, releasing NF κ B for import into the nucleus. Following this, I κ B is ubiquitinated and degraded by the proteasome. In NleC-transfected cells, NF κ B is released from I κ B, but rendered nonfunctional by the activity of NleC, leaving I κ B phosphorylated and marked for degradation. The disappearance of I κ B even when the proteasome is inhibited in these experiments suggests that NleC may be directly responsible for the degradation of I κ B.

Another putative substrate was discovered in experiments with $\Delta nlec/nleC$ EHEC-infected cells which resulted in reduced levels of phosphorylated MAPK p38. When cells were infected with EPEC $\Delta nleC$, phosphorylated MAPK was partially restored to normal levels. These results suggest that NleC may be involved in this phenotype, proteolyzing a signaling molecule upstream of MAPK phosphorylation in conjunction with another EPEC T3SS effector⁷⁴.

NleC was also found to affect NF κ B binding partners. After identifying transcriptional activator p300 in a pulldown experiment with NleC, the TAZ1 domain in the N-terminal portion of p300 was specifically found to interact with NleC. With immunofluorescence microscopy, p300 levels were shown to be reduced in EHEC-

transfected cells compared to infection with NleC-knockout EHEC, whereas *in vitro* incubation of purified NleC and p300 showed a reduction in p300 over the eight hour assay⁷⁶.

From this body of knowledge, these reports establish an important role for NleC during pathogenesis, albeit with discrepancies in the specificity of its proteolytic activity and range of substrates. Working in conjunction with other T3SS effectors NleE, NleB and NleH1, NleC helps to modulate proinflammatory cytokine production in host cells (Figure I-3). NleC clearly targets and degrades NF κ B subunits, though the location of the cleavage site is in contention. There is also disagreement about the localization of NleC during pathogenesis. P300 is reported to be another NleC substrate, but the connection between this and the capacity to proteolyze NF κ B is unknown. Additionally, the mechanism by which NleC reduces phosphorylation of p38 MAPK is undefined. Though the conserved zinc-binding motif seems to be necessary for proteolysis, the sequence of NleC bares no further resemblance to known zinc proteases. Because structural studies of other effectors have discovered novel folds, it is possible that NleC similarly has a novel zinc protease fold.

In this thesis, I elucidate the function of NleC, determining its crystal structure and biochemically characterizing its proteolytic activity for NF κ B transcription factors. Though I began the process to structurally characterize NleC first, the subsequent publication of its function in pathogenesis was instrumental in determining the X-ray crystal structure. As such, I present the work I have done on the functional characteristics of NleC first, followed by an analysis of its high-resolution structure. This also allows analysis of the structure of NleC taking into account my findings about its specificity.

The proteolysis of NF κ B subunit RelA by NleC was determined unequivocally, however there are still many inconsistencies in the literature. Described in Chapter II, I used highly purified proteins to verify *in vitro* that NleC cleaves RelA, p50 and RelB. I also utilized various fragments of RelA to determine which domains NleC recognizes. After determining the exact cleavage site in RelA, I used mutagenesis to determine which of the nearby residues are important for recognition and cleavage efficiency by NleC. I also tested cleavage and binding of p300 by NleC.

In Chapter III, I discuss the structure determination of NleC. As phasing the NleC crystals using the signal from the active site zinc enabled the structural determination, it was necessary to first know the function of NleC. After solving the structure with the Zn-phased dataset, I found that NleC has a homologous fold to the Zincin zinc protease superfamily. I identified various differences between NleC and the canonical Zincin fold, and discuss the broader NleC protein family. From the structure, I located putative binding sites on the surface that would recognize RelA. Analyzing the electrostatic surface potential of NleC, RelA and p300, I determined that an electrostatic mode of binding is very likely. Finally, I used SAXS to study the C-terminal domain of NleC, which was not resolved in the crystal structure.

In Chapter IV, I integrate the results of my research and examine the implications of the NleC structure for translocation and specificity during pathogenesis. In addition, I discuss on-going experiments and future directions which will address hypotheses developed from my research.

Chapter II. Specificity of the T3SS effector NleC

Introduction

Mediated through NF- κ B activation, host cells respond to bacterial infection with an inflammatory response, involving interleukin-8 (IL-8) and tumor necrosis factor α (TNF- α) secretion which result in activation of immunity cells in the underlying basal epithelium⁷⁷⁻⁸⁰. The T3SS effector NleC (non-LEE encoded effector C) works in concert with other effectors to repress the host inflammatory response, facilitating Enterohemorrhagic *Escherichia coli* (EHEC) colonization. NleC is a metalloprotease reported to cleave the NF- κ B subunits RelA (p65), p50 and c-Rel, thereby depressing downstream transcription events that lead to inflammation⁷⁰⁻⁷³. This activity is coordinated with that of EHEC effectors Tir, NleH1, NleE, and NleB, that respectively inhibit proteins upstream of NF κ B, TRAF, RPS3, TAB2 and TAB3, and TNF α -mediated NF κ B activation upstream of RPS3^{62,67-69,81,82}. Recent publications also implicate NleC in cleaving I κ B, preventing phosphorylation of p38, and binding of p300 and CREB-binding proteins^{72-74,76}.

Though it was proven that NleC is a highly specific protease for NF κ B transcription factors, it is unclear how proteolysis by NleC abolishes NF κ B activation. Two groups reported putative cleavage sites on RelA in the N-terminal portion of the DNA-binding domain (DBD), but disagree as to the exact sequence^{70,71}. Furthermore, NleC was shown to proteolyze multiple NF κ B subunits, but it is unknown whether a specific subunit or

subunits are preferentially proteolyzed, and are thus the principal target during pathogenesis. With recent research suggesting that NleC may target additional proteins beyond NF κ B subunits, it is possible that NleC recognizes a short sequence or small structural motif that would make it a promiscuous protease. Because of this collection of data, defining how NleC recognizes substrate is important to determine the spectrum of substrates that are important for pathogenesis.

In this chapter, I will describe how NleC recognizes NF κ B subunits using activity assays and mutagenesis. I will additionally explore NleC recognition of p300. My results suggest that NleC is a highly specific protease for NF κ B subunits that recognizes elements both proximal and distal to the DB loop. This work broaches questions about how NleC functions during pathogenesis and discusses the additional reports on NleC targets.

Results

NleC is a nonpromiscuous protease specific for NF κ B subunits

NleC cleaves NF κ B subunits RelA, RelB, and p50 in the DNA-binding loop

NleC was previously described as a zinc metalloprotease capable of cleaving NF κ B subunits RelA, p50, and c-Rel⁷⁰⁻⁷³. As mentioned in the introduction of this chapter, two groups of researchers attempted to determine the precise cleavage site on RelA, but disagree in the specific sequence. Baruch *et al.* reported cleavage between residues 38 and 39, whereas Yen *et al.* identified the cleavage site as between residues 10 and 11^{70,71}. To determine the precise cleavage site, an *in vitro* assay with highly purified proteins was required. I thus cloned and expressed full-length NleC as a fusion with the self-cleavable Intein chitin-binding domain (CBD) tag, and purified it to homogeneity. Fragments of RelA

(residues 17-291), RelB (residues 124-413) and p50 (residues 39-363), each containing the DBDs and dimerization domains, were cloned and expressed, and purified to homogeneity. To ensure thorough cleavage of the NF κ B subunits, they were incubated overnight with purified NleC. RelA was incubated in 1:1 equimolar ratio with NleC, whereas p50 and c-Rel were in excess at 1000:1 NleC. The cleavage products were analyzed with SDS-PAGE followed by electroblot transfer onto PVDF membrane and staining with Coomassie Blue. I submitted the bands corresponding to the cleavage products for Edman amino-terminal sequencing at University of Texas Medical Branch Protein Core facility. RelB was found to be proteolyzed at 144'-C|EGRSA-149 and p50 at 61'-C|EGPSH, whereas the cleavage site for RelA was confirmed as that published by Baruch, *et al.* at 38'-C|EGRSA-43. In all three subunits, the scissile bond corresponds to the highly conserved DNA-binding loop (DB loop)(Figure II-1).

Cleavage efficiency of NleC for RelA

Because the Rel-homology domain NF κ B subunits are structurally conserved and most of the published work was done with RelA, much of the further characterization was performed for RelA alone. In preparation for more exhaustive studies detailing NleC's proteolysis of RelA, I first needed to determine the basic cleavage conditions and rate. The rapid degradation in quality of purified RelA was the initial and largest barrier to defining catalysis conditions. This was apparent in the reduced capacity of NleC to cleave RelA over time. Initially it was assumed that RelA was aggregating under the storage conditions, rendering it uncleavable. However, applying the stored RelA to size exclusion chromatography resulted in a single peak corresponding to the molecular weight of RelA,

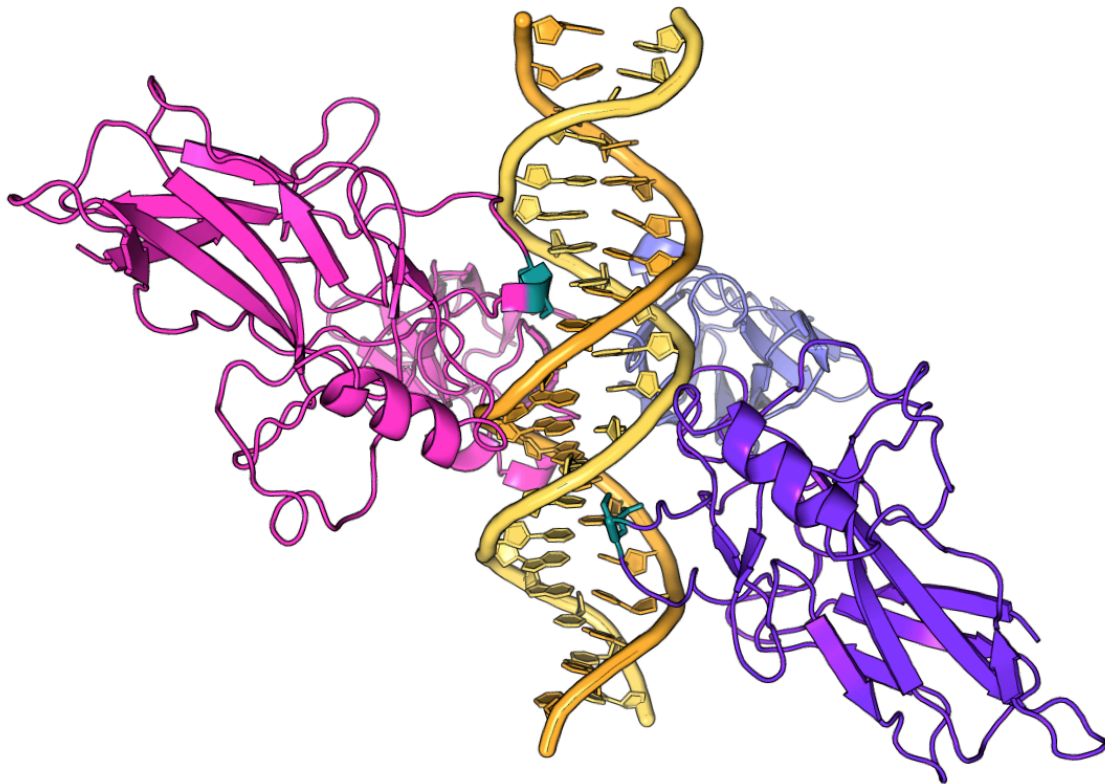


Figure II-1 Location of NleC cleavage site in RelA. The cleavage site of RelA occurs in the DNA-binding loop of RelA, here depicted in a structure of homodimeric RelA binding DNA. The two RelA subunits are in magenta and purple, whereas the DNA is in yellow. The cleavage site is colored in teal. PDB: 1RAM.

as with the freshly purified protein. These results suggest that aggregation of RelA over time was not responsible for the decrease in cleavage efficiency.

Next, because the quantity of RelA that was cleaved was similar to the amount of NleC added in these initial experiments, we hypothesized that NleC was a single turnover enzyme. Hence, because RelA is not aggregating over time, issues with the activity assay lay with NleC. To test this, I ran a series of reactions with decreasing quantities of NleC and found that significantly more RelA is cleaved in these reactions than the amount of NleC present. These experiments proved that NleC continues to turnover after the initial catalytic event.

A final hypothesis considered the cysteine present in the DB loop, which could become modified over time, either by forming a disulfide bond with another molecule of RelA, or with the reducing agent itself. To try to recover reducing conditions, activity assay conditions were altered to contain TCEP (tris(2-carboxyethyl)phosphine) instead of BME (β -mercaptoethanol), which is known to oxidize over time, forming disulfides with itself and any free cysteine residues. Newly synthesized RelA that was purified in the presence of TCEP showed a much higher level of proteolysis by NleC and was more stable during storage than preparations made with BME. All subsequent purifications of RelA or other NleC substrates were performed with TCEP as the reducing agent. Even with these precautions, however, preparations of RelA continue to degrade in quality upon long-term storage, losing the ability to be cleaved by NleC. For this reason, all activity assays are performed within one week of expression and purification of RelA.

After establishing the activity buffer, a standard reaction condition was determined. Holding NleC concentration constant at 20nM, an activity assay was performed on three concentrations of RelA at multiple time points (Figure II-2). From these data, it was determined that for 0.2 mg/mL and 1 mg/mL RelA, the product formation is linear between 0 and 5 minutes, while at 20 minutes, product formation becomes nonlinear. Because we are interested in assaying conditions that decrease the ability of NleC to proteolyze NF κ B, such as mutants, identifying a condition that produces measurable product is essential. For this reason, we chose to perform future RelA cleavage experiments at 1mg/mL concentration, quenching the reactions at 10 minutes if a full time course is not performed. Under these conditions, we are compromising between the linearity of product formation and obtaining measurable product for SDS-PAGE analysis. Unfortunately, at 5

mg/mL, the uncleaved and cleaved bands of RelA are difficult to distinguish because of their proximity, complicating the analysis.

RelA and RelB are more efficiently proteolyzed by NleC than p50

With the activity assay standardized, the rates of RelA cleavage were compared to those of other NF κ B transcription factors. RelA, RelB and p50 were expressed, purified and concentrated to 1 mg/mL into activity buffer containing 1 mM TCEP. As shown in Figure II-3, RelA and RelB have similar cleavage rates, whereas p50 is approximately three times slower by SDS-PAGE analysis. This suggests preferential cleavage of RelA and RelB.

DNA binding slows proteolysis of RelA by NleC

Because NleC cleaves RelA in the DB loop, the cleavage efficiency in the presence of DNA was tested. Palindromic DNA containing the RelA recognition site (5'-CGGCTGGAAATTTCCAGCCG-3') was incubated with RelA before proceeding with the activity assay. This is the same DNA fragment that was used to obtain the crystal structure of homodimeric RelA bound to DNA (PDB ID: 2RAM). The presence of DNA resulted in an approximate 10-fold decrease in the rate of catalysis, as analyzed by SDS-PAGE (Figure II-4A).

NleC recognizes both local and distal elements of RelA

In order to determine whether elements outside of the DB loop are necessary for recognition, RelA constructs of various sizes were tested for proteolysis by NleC. A construct originally prepared to ease the mutagenesis of residues near the active site, a fusion protein that expresses 20 residues surrounding the RelA DB loop was cloned between two well-expressing bacterial proteins: the periplasmic domain of BamA and BamD (*E. coli* β -barrel assembly machine). When NleC-catalyzed cleavage was not observed

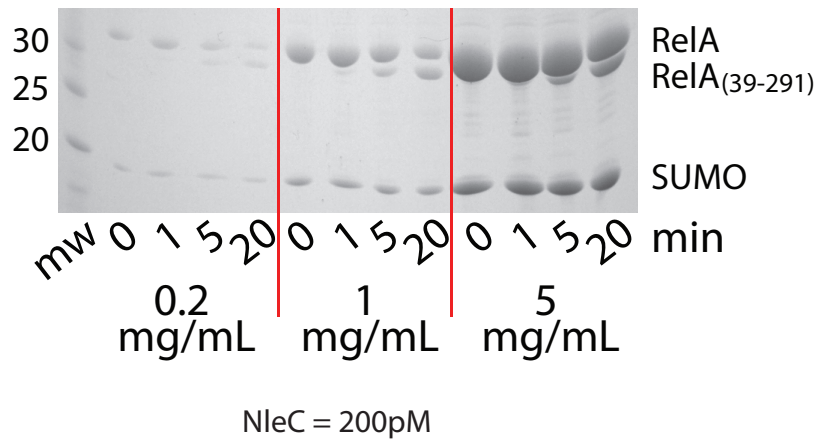


Figure II-2 Cleavage rate varies with RelA concentration. SUMO-RelA₁₇₋₂₉₁ was expressed with SUMO. During purification, the SUMO tag self-cleaved. Zero timepoints are taken before addition of NleC and take into account the small concentration change that occurs upon addition of NleC. NleC was at 200pM or 0.008mg/mL in each reaction. Reactions were quenched with addition of SDS loading buffer and followed by boiling for 5 minutes.

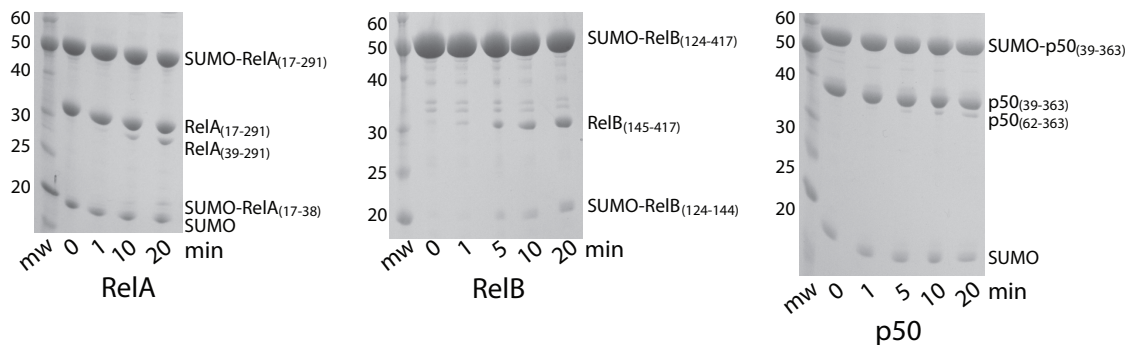


Figure II-3 Comparison of cleavage rates between NFκB subunits. All zero timepoints were taken before the addition of NleC. Reactions were stopped with the addition of SDS running buffer to the reaction, followed by boiling for five minutes. RelB and p50 reactions were run on the same day, whereas RelA was purified and tested for activity a week later. During the course of purification, the SUMO tag from RelA and p50 was partially cleaved.

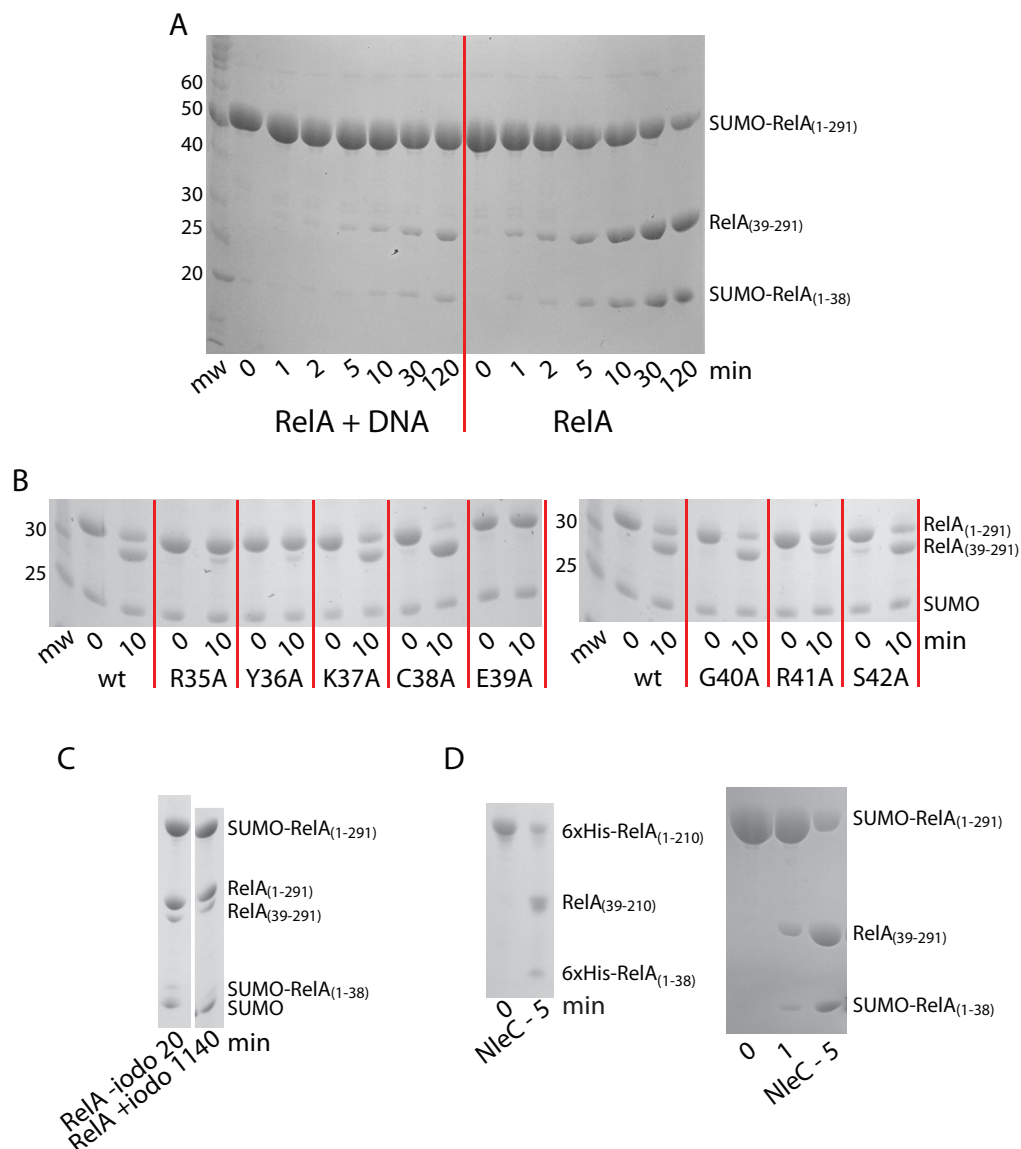


Figure II-4 SDS-PAGE analysis of RelA cleavage by NleC. A) Time course activity assay of RelA in the presence of DNA. Leftmost lanes depict a time course of RelA in the presence of palindromic DNA containing the binding site of RelA. Rightmost lanes depict the same timepoints without DNA. NleC concentration is 20 nM. B) Effect on activity of RelA alanine mutants. Single mutants were tested for cleavage by NleC at the 10 minute timepoint with 20 nM NleC. The zero timepoint is before the addition of NleC. C) Modification of scissile bond cysteine by iodoacetamide greatly hinders cleavage by NleC. The left lane shows cleavage RelA cleavage after 20 minutes by 20 nM NleC. The right lane shows RelA after incubation with and subsequent quenching of iodoacetamide after 1140 minutes with 20 nM NleC. The lanes are from the same SDS-PAGE gel and the assay was performed with the same RelA preparation. Total quantity of RelA cleaved in leftmost lane is decreased when compared to other panels because of the age of sample. D) Comparison of cleavage of RelA DBD and RelA Rel-homology domain. The Rel-homology domain contains both the DNA binding and dimerization domains. Left pane shows a construct with the DBD at 0 and 5 minutes. Right pane shows a construct with both the DNA_binding and the dimerization domain of RelA at 0, 1 and 5 minutes. The 5 minute timepoints are comparable for both. 200pM NleC was present in both assays.

with this construct, a different fusion was constructed. Because BamA and BamD may interact with each other, the RelA fragment may acquire a nonnative conformation, preventing recognition by NleC. To rule this out, BamA was replaced with a 10xHis-SUMO tag, to produce 10xHis-SUMO-RelA₂₈₋₄₈-BamD. As SUMO is a yeast protein, it is not expected to interact with BamD. However, cleavage was not observed with this fusion either. These results suggested that NleC recognizes elements outside of the immediate DB loop, though the presence of the fusion proteins may have hindered recognition by NleC.

With the inability to proteolyze the fusion proteins, small peptides containing only residues adjacent to the DB loop (RYKCEGR, YKCEGR and acetyl-YKCEGR-amide) were tested for proteolysis. These peptides would negate the possibility that NleC is unable to access the cleavage site. After overnight incubation with 20 nM NleC, at molar quantities equivalent to 1 mg/mL SUMO-RelA₁₇₋₂₉₁, the reaction mixture was subjected to mass spectrometry analysis. Using both ESI-MS and MALDI-MS, no cleavage products could be identified, as the spectra were indistinguishable from the control peptide sample which lacked NleC. In addition, isothermal titration calorimetry (ITC) was used to test binding of these peptides and the inactive NleC_{E184A}. No significant binding was observed, corroborating the lack of cleavage observed in the mass spectrometry experiments.

With the lack of cleavage and binding observed with the fusion proteins and peptides, we hypothesized that part of the binding interface was missing. Consistent with this, when the entire DBD (residues 1-193) of RelA was present, NleC was able to cleave with the same efficiency as it cleaves the whole Rel-homology domain (residues 17-291)(Figure II-4D). As this construct was less stable than that containing the whole Rel-homology domain, the remaining experiments were performed with RelA₁₇₋₂₉₁.

In order to test which residues near the scissile bond are important for cleavage, eight alanine mutants of SUMO-tagged RelA₁₇₋₂₉₁ were constructed scanning four residues on either side of the scissile bond, RYKC|EGRS (Figure II-5 and Figure II-6). After purification to homogeneity, these eight mutants were tested for cleavage in comparison to wild-type SUMO-tagged RelA (Figure II-4B and Figure II-5). All mutants were expressed and purified concurrently in order to minimize any effects of differing expression conditions as well as the time-dependent degradation in quality that occurs with RelA. Under the standardized reaction conditions of 1 mg/mL RelA incubated with 20 nM NleC for ten minutes, four of the eight residues were found to contribute most to specificity, reducing cleavage tenfold or greater. Two of the native residues appear to hinder cleavage by a relatively small amount. In decreasing order of effect on cleavage efficiency, E39A, R35A and Y36A most hindered proteolysis by NleC indicating that they are important for recognition by NleC, with R41A having an intermediate effect. Mutating C38 to an alanine resulted in 30% more cleaved product than wildtype, suggesting that a smaller residue is better accommodated at this site. To further test whether the residue at position 38 is important for recognition by NleC, iodoacetamide was incubated with wild-type SUMO-tagged RelA and tested for activity. The resulting acetamide-RelA was cleaved at 2% the rate of cleavage of untreated RelA (Figure II-4C). Together, these results suggest that the cysteine is not required for recognition, but that a smaller residue would be preferred at this site. These results corroborate my hypothesis that modification of C38 by BME was responsible for reduced cleavage efficiency by NleC. The final mutant with altered cleavage efficiency, G40A, showed a slight increase in proteolysis by NleC. This indicates that the alanine substitution results in a modest improvement in the binding interface.

A final experiment tested for stable binding to RelA by the inactive mutant, NleC_{E184A}. For this assay, 6xHis-RelA₁₋₁₉₃ was bound to Ni-NTA resin, and washed before loading NleC. Following further washing, the elutions were collected and analyzed by SDS-PAGE. Under these conditions, only a small amount of NleC_{E184A} seemed to stay associated with RelA, with most of NleC being removed in the washes. Additionally, after incubation of RelA with NleC_{E184A}, a complex was unable to be isolated by size exclusion chromatography. This suggests that the affinity of NleC for RelA is too weak for stable association.

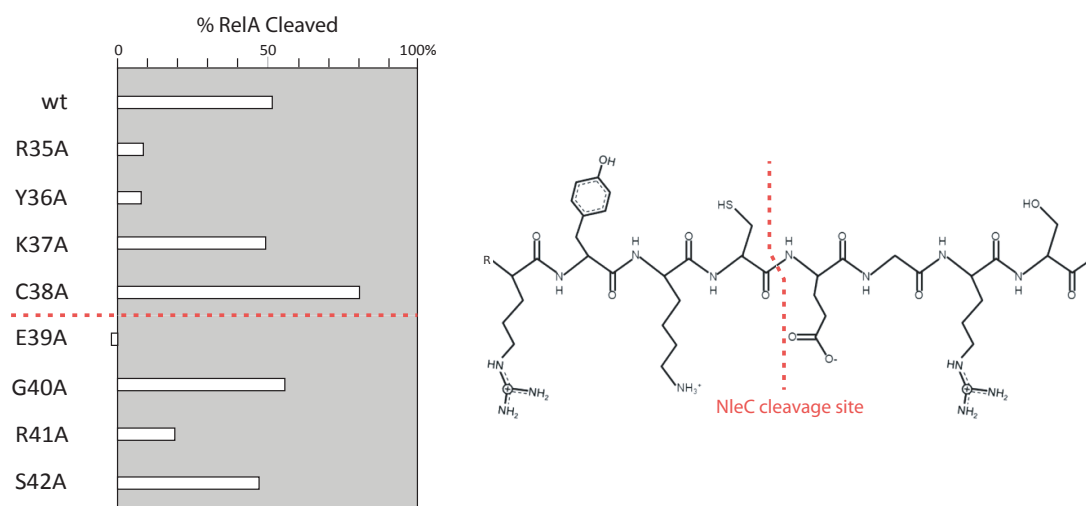


Figure II-5 Effect on NleC catalytic efficiency of RelA mutagenesis. Residues near the cleavage site of RelA₁₇₋₂₉₁ were mutated to alanine and measured for amount of NleC cleavage after 10 minutes relative to total amount of substrate. These results correspond to the gel shown in Figure II-4B. SDS-PAGE were Coomassie-stained and analyzed with ImageJ. The background from the zero timepoint was subtracted from the ten minute timepoint to normalize for staining differences. This also causes the E39A results to be negative.

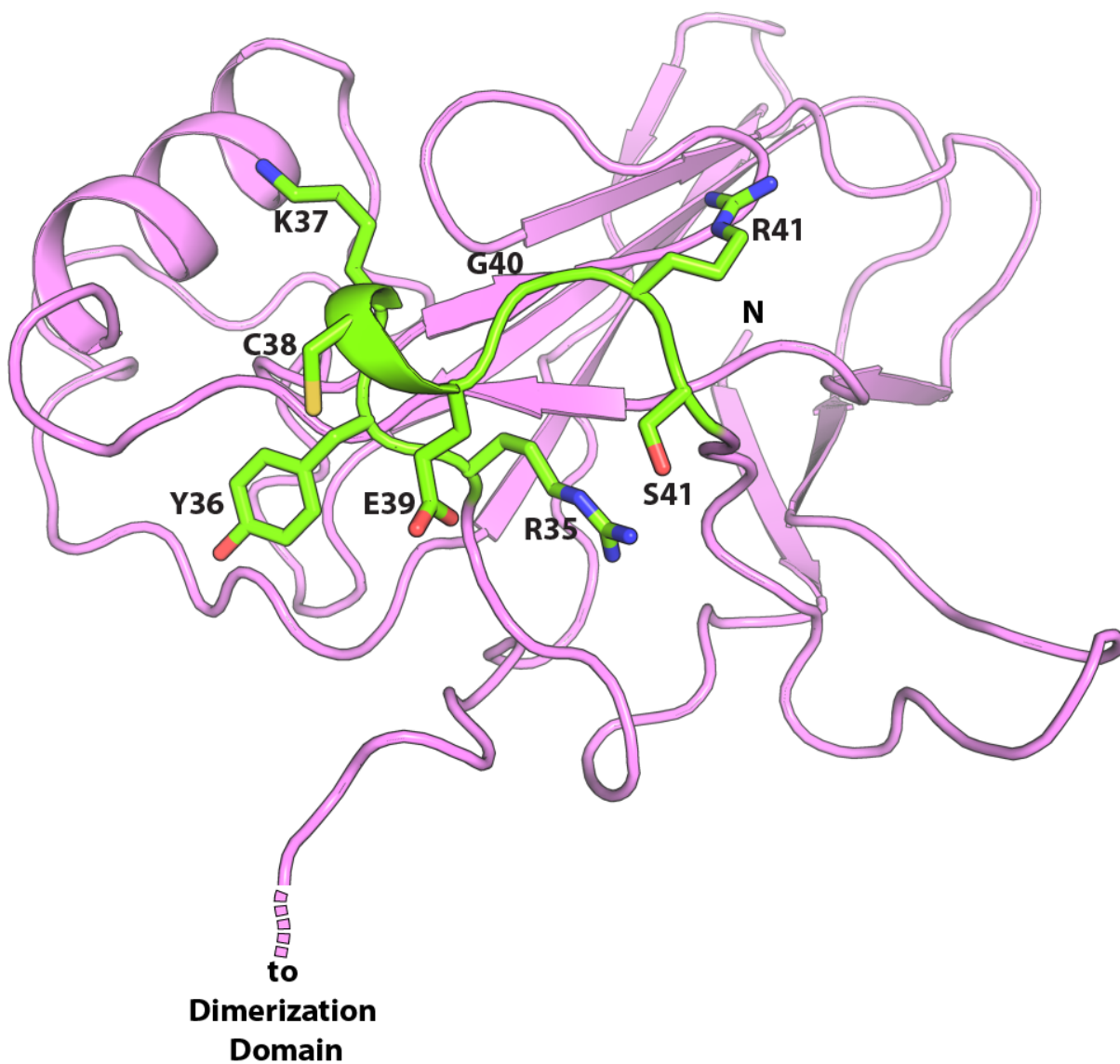


Figure II-6 Structure of DB loop with mutagenized residues highlighted. This figure highlights the DB loop of RelA within the DBD and shows its local structure. R35, Y36, E39 and R41 are the residues which most contribute to recognition by NleC, as their mutation to alanine significantly decreases cleavage efficiency.

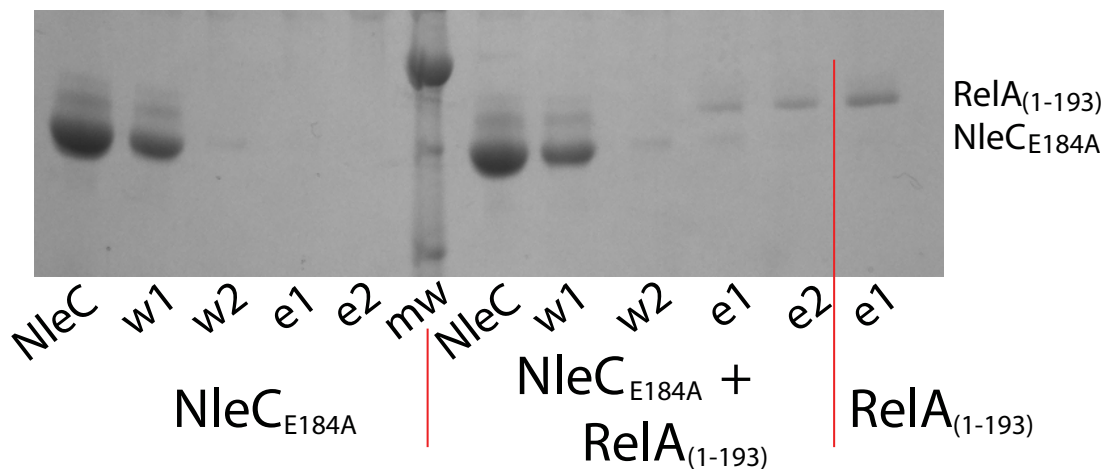


Figure II-7 Pulldown of RelA by NleC_{E184A}. Control resin, with no protein bound is shown in the leftmost lanes, whereas resin with RelA pre-bound is shown in the right lanes. The far right lane depicts the elution from resin where no NleC was added. W1 and w2 indicate washes of the resin, whereas e1 and e2 are elutions. MW stands for the molecular weight standard.

NleC interacts with, but does not cleave, p300 TAZ1 domain

Shames *et al.* recently reported binding of NleC to the N-terminal portion of p300, a transcriptional activator that has been shown to bind NF κ B in the nucleus⁷⁶. To further explore the activity of NleC, a 10xHis-SUMO tagged fusion of p300, including its TAZ1 domain, was expressed and purified (residues 1-424). After binding p300₁₋₄₂₄ to Ni-NTA and washing extensively, NleC was added to the column and further washed to remove unbound protein. NleC was observed in the imidazole elution from beads containing p300₁₋₄₂₄ but not from control resin with no p300 bound (Figure II-8A). While this binding interaction suggests a direct binding between NleC and p300, upon running an activity assay, no cleavage was observed (Figure II-8B). A very faint band did appear after overnight incubation with NleC. However, because of the extremely low amount of cleavage compared to concurrently run RelA assay, it is apparent p300 is not a main substrate. Cleavage of the TAZ1 domain alone (residues 329-424) was additionally tested and

consistent with these results, cleavage was not detected, though a Ni-NTA pulldown was not performed.

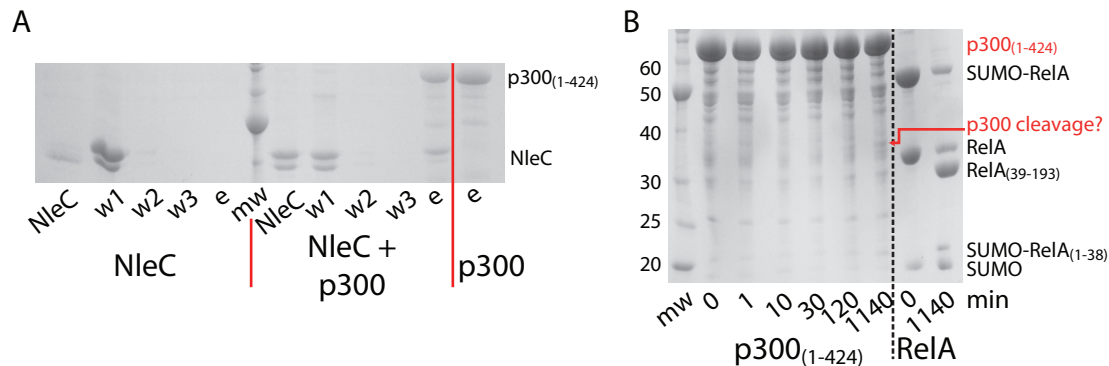


Figure II-8 SDS-PAGE analysis of p300 interaction with NleC. A) Ni-NTA pulldown of NleC by p300. Leftmost lanes depict the negative control with NleC flowed over Ni-NTA resin, washed and eluted, in the absence of any His-tagged protein. NleC+p300 lanes depict resin that has been prebound with the N-terminal domain of p300 (residues 1-424), to which NleC was added. The Ni-NTA resin was then washed and the p300 eluted. NleC is seen eluted from the column as well, signifying that NleC was pulled-down by p300. The rightmost lane shows p300 elution in the absence of NleC. NleC appears as a doublet because of C-terminal degradation that is typical of this protein. B) Activity assay of NleC with putative substrate p300 and RelA. Left lanes show time course in minutes of p300 with 200 nM NleC. RelA cleavage is shown in the rightmost lanes as a positive control with the same quantity of NleC.

Discussion

Exploring NleC Recognition of NFκB

From establishing that NleC cleaves NFκB transcription factors in their DB loop, to determining that elements both local and remote are necessary for efficient cleavage, I have shown that NleC is a NFκB-specific protease. The inability of NleC to cleave the small peptides containing the DB loop supports the hypothesis that residues or structural motifs outside of the local cleavage site are necessary for its recognition and activity. Combined with the mutagenesis results, the lack of cleavage observed with two of the DB loop peptides is corroborated, as they lack a residue that is essential for efficient cleavage by

NleC, R35. However, the longest peptide, containing R35, as well as the fusion protein with a longer stretch of residues surrounding the DB loop, should have been recognized and cleaved. The lack of cleavage of two constructs gives credence to the hypothesis that elements distal from the DB loop are necessary for recognition and efficient proteolysis.

The mutagenesis experiments additionally support the theory that a bulky group located at C38 in RelA will have a negative impact on cleavage efficiency. The mutation of C38 to an alanine increased cleavage efficiency over wildtype NleC, whereas the modification of this cysteine by disulfide bonding with BME or reacting with iodoacetamide abrogated cleavage. The curious effect of increasing the efficiency of cleavage over wildtype is likely due to a small improvement in the binding interface with an alanine at this position as opposed to cysteine. It is important to mention, additionally, that the negatively impacting modification of cysteine by forming disulfides is not physiologically relevant with the reducing environment present in cells. Furthermore, a modification here would have a negative influence on the native activity of NF κ B, as this cysteine is located in the DB loop.

Though I have established the importance of the DB loop for cleavage efficiency, there are limited experiments outside the DB loop to determine where additional binding interface occurs. Experiments in the published research did not attempt to locate a specific binding interface outside of defining where NleC cleavage occurs in RelA. Though my initial experiments suggest that only the DBD is necessary for recognition, as NleC was also able to cleave constructs that lacked the dimerization domain, I was unable to isolate a complex of NleC and RelA₁₋₁₉₃(Figure II-7). This construct of RelA contains only the DBD, and not the dimerization domain. To further characterize the specificity of NleC for RelA, it is important

to establish whether the distal elements recognized by NleC are contained within the DBD or extend to the dimerization domain, and what elements of RelA contribute to a stable binding interface.

The decrease in cleavage efficiency observed when RelA binds DNA is important when considering the physiological implications of NleC function. Depending on the time assayed, NleC either localizes in the cytoplasm near the apical membrane and site of infection, or in the nucleus. NleC that was delivered *via* the T3SS during infection was assayed after a period of only a few hours as being in the cytosol and is believed to be more relevant than the localization assays performed with transfected NleC, which localized to the nucleus. If NleC is associating with RelA while it is still in the cytoplasm, the ability of DNA to abrogate cleavage with NleC is less important for pathogenesis. My results are consistent with this, as I found that the presence of DNA drastically lowers the cleavage efficiency of NleC, a curious result if physiologically relevant. Indeed, when looking at the structure of NF κ B subunits bound to I κ B, the DB loop is exposed to solvent, suggesting it is available for binding when sequestered in the cytosol. The activity of NleE is shown to prevent the degradation of I κ B, preventing the release and activation of NF κ B, and may in fact assist in the function of NleC by retaining NF κ B in cytosol, where the DB loop is exposed to be cleaved.

By proteolyzing in the DB loop, NleC takes the most direct approach in deactivating the NF κ B transcription factors. The region in RelA that is cleaved by NleC has the highest concentration of DNA base contacts of RelA (Figure II-9)⁸³. Of the thirteen residues that are implicated in binding DNA, I have mutated five during the course of determining the binding site of NleC. Three of these residues I have found to be important for recognition by

NleC, R35, Y36, and E39. If NleC is a specific protease for NFκB subunits, the best strategy for targeting them would be to recognize additional residues near the cleavage site. The most exposed residues on the face of transcription factors that binds DNA are likely those that actually recognize the DNA. Thus, the remaining DNA-binding residues of the NFκB subunits may be good candidates for testing recognition by NleC, as the DNA-binding face of these transcription factors is essential for the function NleC is trying to abrogate.

Another way to screen which residues are important for recognition by NleC is to look at conservation between all the NFκB transcription factors. I have shown that NleC is able to cleave three of the NFκB subunits, RelA, RelB and p50 (Figure II-2 and Figure II-3). The activity of NleC for RelA, p50 and c-Rel has also been published^{72,73}. From this, it is likely that NleC can recognize the remaining NFκB subunit, p52, as well. These NFκB transcription factors have highly conserved folds, but their sequence conservation is >50% within NFκB Class I or Class II subunits, and <50% between Class I and II. Because NleC can recognize and proteolyze all NFκB subunits, it is plausible that only conserved residues between all five NFκB subunits are important for recognition. Using the Consurf server, I mapped the alignment of the human NFκB Rel-homology domain onto the structure of murine RelA (Figure II-10 and Figure II-11)⁸⁴. The Consurf server uses the alignment to score residues on the structure and colors them for visualization. The scores refer to the evolutionary rate of any given position, where slowly evolving residues are conserved and quickly evolving residues are variable. Conserved residues often have inhibitory selective pressure towards mutation, and are likely important for structural or functional reasons. Any of the highly conserved surface residues near the cleavage site of NleC may be

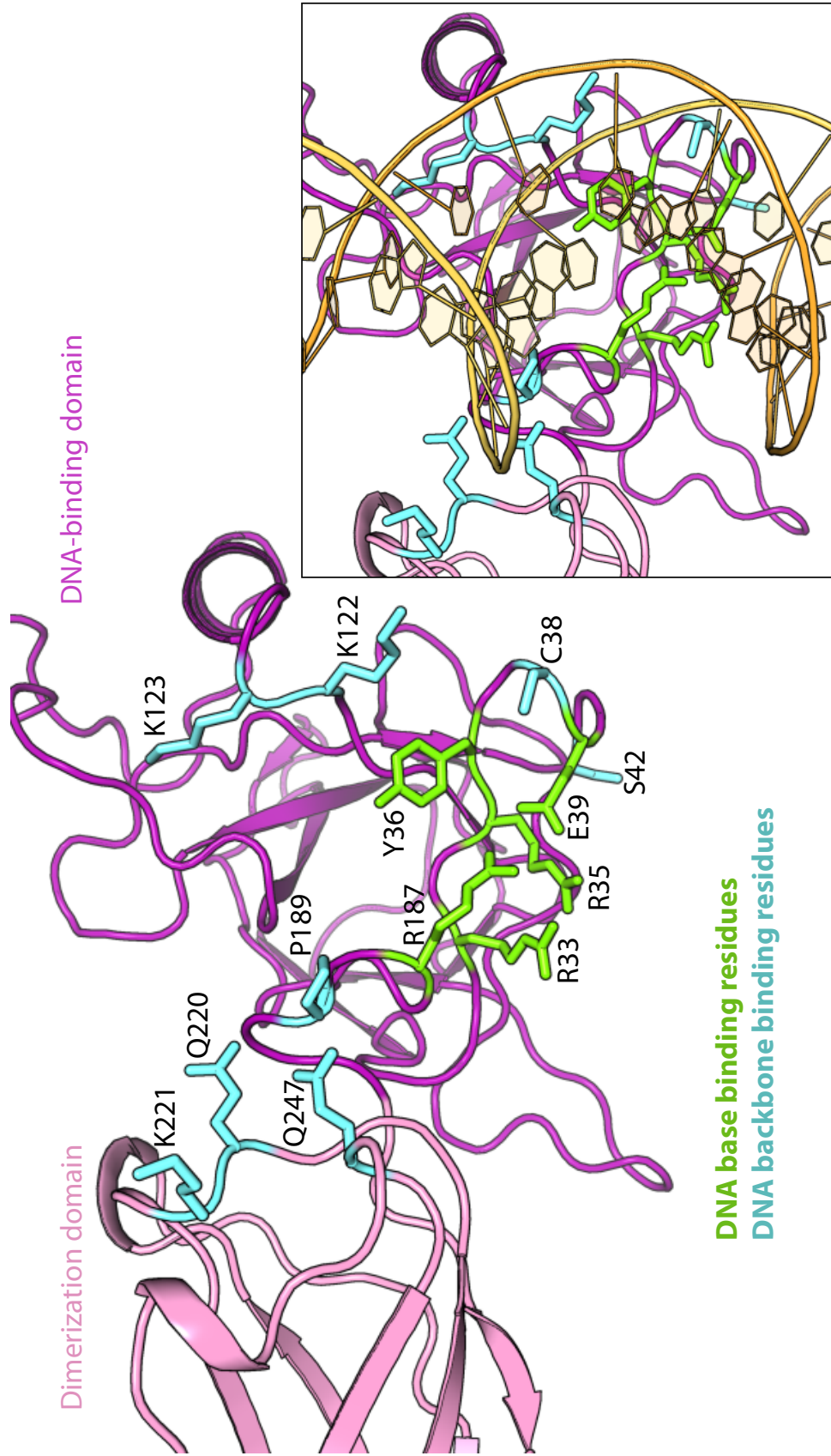


Figure II-9 DNA-binding residues of RelA. Residues within the DBD, and in the dimerization domain of RelA that are known to bind DNA are highlighted. Residues in green bind DNA bases, whereas those in cyan bind the DNA backbone. In the smaller panel, the DNA double helix is also shown for context. PDB: 2RAM.

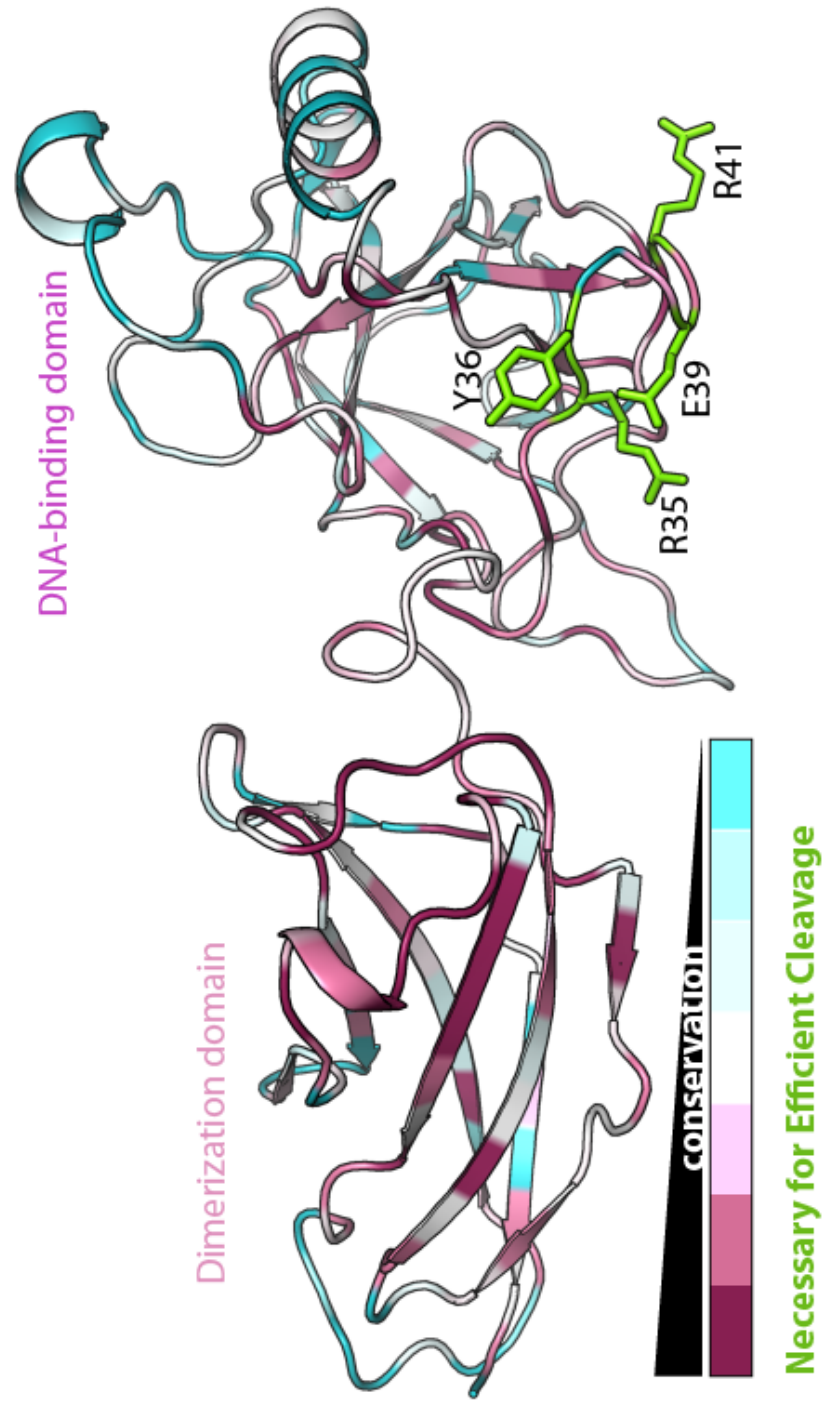


Figure II-10 Conserved residues among NFκB. In dark magenta are the most conserved residues, whereas in aqua are the most variable. Green residues are those that were found to be important for recognition by NleC. Alignment mapped onto RelA structure with the Consurf server⁸⁴.

important for recognition. In the mutagenesis assays described here, the residues that are important for cleavage by NleC, R35, Y36, and E39, are also conserved among all the subunits. The exception is R42 in RelA, which is a proline residue in p50 and p52. From the comparisons of RelA, RelB, and p50 activity, it is apparent that p50 is cleaved with less efficiency by NleC, and this may be accounted for by the substitution of arginine in the DB loop for proline (Figure II-3).

Substrate Recognition Beyond NF κ B

NleC was reported to target other host cell proteins besides NF κ B subunits. Mühlen *et al.* found that whether or not the proteasome was inhibited, NleC transfection caused I κ B level reduction⁷². I κ B functions to sequester NF κ B subunits in the cytosol, masking their nuclear localization signal and thus preventing the activation of downstream transcription events by association with DNA. Upon I κ B phosphorylation, NF κ B is released, exposing a phosphorylation site that is ubiquitinated and targets I κ B for proteasomal degradation. That the degradation of I κ B continued even when the proteasome was inhibited suggests that NleC is responsible. Because I κ B is retained when the proteasome is inhibited when NleC is absent, this suggests that NleC activity directly results in the degradation of I κ B. If verified, this activity would function in opposition of NleE, which prevents the degradation of I κ B during pathogenesis. As introduced in Chapter I, a $\Delta nleE$ strain resulted in faster degradation of NF κ B by NleC than in wildtype strains. The combination of the effects of NleE and NleC may explain the results that these two T3SS effectors work in opposition. The reduction of phosphorylated p38 MAPK in NleC-infected cells was another observed phenotype by Sham *et al.*⁷⁴. p38 MAPK is implicated in a signaling cascade that both

responds to and produces proinflammatory cytokines and in fact is in the same class of kinases as JNK, which is proteolyzed by fellow T3SS effector NleD^{70,85}. Upon the deletion of NleC from EPEC cells, this phenotype was partially restored, implying that a third effector beyond NleC and NleD may be implicated in suppressing the activation of MAPK pathways⁷⁴. In light of the experiments described here, these secondary effects of NleC are difficult to reconcile with the observations that NleC is highly specific for NFκB subunits, as I hypothesize that NleC recognizes elements outside of a small amino acid motif.

Though these secondary effects of NleC in host cells are intriguing, I chose to focus on a third putative substrate for NleC, p300, a transcriptional activator that is known to bind and increase transcription levels of RelA⁸⁶. Shames *et al.* utilized a proteomics isotope-labeling technique to attempt to identify additional substrates of NleC. Pulldowns using NleC followed by mass spectrometry identified Hsp70, Hsp90a & b, HspB, CREB-binding protein, p300, and Ubiquitin-S27a. Because heat shock proteins are often pulled down non-specifically, they chose to focus on transcriptional activators CREB-binding protein and p300, which are closely related, and interact with RelA⁸⁶. To verify this, they then pulled down NleC with p300 by co-immunoprecipitation. Performing the immunoprecipitations with truncations of p300 permitted Shames *et al.* to identify which domains of p300 are likely responsible for binding NleC. Non-specific binding was observed in the assay, complicating results for the pulldown with the C-terminal p300 domains, but the TAZ1 domain in the N-terminal portion of p300 bound to NleC more perceptibly. Immunofluorescence microscopy showed that infection with wild-type or $\Delta nleC/nleC$ EPEC depleted p300 by 20% in HeLa cells compared to $\Delta T3SS$ or $\Delta nleC$ EPEC mutants. Ectopically expressed NleC similarly reduced p300 40% in HeLa and Caco2 cells compared to empty

vector. *In vitro* analysis similarly showed that purified NleC decreased the quantity of purified p300 over the course of eight hours. Overexpression of p300 in cells causes a marked increase in IL-8 secretion. Conversely, $\Delta p300$ cells show a decrease in secretion of IL-8 upon EHEC infection. This underlines the importance of p300 in proper inflammatory signaling by RelA⁷⁶.

As described above, several lines of evidence suggest a reduction in p300 in the presence of NleC, as well as a direct interaction between these two proteins. Therefore, I targeted an N-terminal fragment of p300 containing the TAZ1 domain for verification of binding with purified protein *in vitro*. Under this assay conditions developed for the NF κ B subunits, a negligible rate of cleavage was observed, three orders of magnitude slower than that observed for RelA (Figure II-8B). These results suggest that p300 is not a primary target for NleC, or that an essential component was not present under the reaction conditions. In the published literature, an *in vitro* activity assay was performed with full-length p300, opening the possibility that the C-terminal domain of p300 is also necessary for NleC cleavage. However, similarly to the assay described here, only a very small amount of p300 appears to be degraded, and additionally, the experiment does not contain an internal loading control. In the same report, a more substantial degradation of p300 was observed when NleC was transfected into HeLa cells. This discrepancy between p300 degradation in cells and *in vitro* assays could be explained by the longer incubation times post transfection of 24-36 hours versus the eight hour incubation for the *in vitro* assay. Alternatively, the full-length p300 or an additional protein may be necessary for NleC proteolysis, such as binding partner RelA.

To determine if the lack of binding to p300 by NleC caused the reduction in cleavage from that observed in cells, a pulldown experiment was performed with NleC. When p300₁₋₄₂₄ was bound to Ni-NTA resin, NleC was pulled down in near molar ratios, compared to control resin, where NleC washed off (Figure II-8A). This result corroborates that seen in the research by Shames *et al.* where an N-terminal portion of p300 containing the TAZ1 binding domain immunoprecipitated NleC⁷⁶. However, the long incubation time required for substantial p300 degradation in the context of cells makes it difficult to distinguish whether p300 proteolysis is largely an off-target effect of NleC or additional protein components are necessary. An additional complication is the absence of the identified motif from NFκB, which only contains the DB loop. Using ScanProSite, four different motifs from specific to more general produced no hits from full-length human p300. The motifs attempted were RYxxExR, RYxxE, [RK][FYW]xx[ED]x[RK] or [RK][FYW]xx[ED], where residues in brackets could be either residue in the motif and x is any residue. However, by varying the number of nonspecific residues between the specific residues in the motif, six putative motifs were obtained. These are summarized in Table II-1. Interestingly, none of these are in the N-terminal portion of p300 that was tested for cleavage by NleC and the motifs may not occur in the proper orientation for recognition by NleC. Further discussion of p300 in the context of the structure of NleC, as well as further experiments, is discussed in Chapter IV.

Table II-1 Summary of putative NleC cleavage sites in human p300. Residues that match the general motif are in uppercase letters, whereas other residues are in lowercase. The general motif was [RK]-x(0,1)-[FYK]-x(1,3)-[ED], where residues in the brackets can be any listed, and the numbers in the parenthesis indicate that the range delineated by the numbers is acceptable.

Sequence	Residues
569-574	KqWheD
1047-1052	KiFkpE
1197-1202	RYhfcE
1203-1207	KcFnE
1342-1345	RFvD
1583-1589	KlYatmE


```

Human_NFkB_p50      -----PEVFQPPQMALPTADGPYLQILEQPKQRGFRFRYVCEGPSHG
Human_NFkB_p52      -----PKEPAP-----ETADGPYLVIVEQPKQRGFRFRYVCEGPSHG
Human_NFkB_c-Rel    -----MASGAYN-----PYIEIIEQPKQRGMRFRYKCEGRSAG
Mouse_NFkB_p65/RelA -----PYVEIIEQPKQRGMRFRYKCEGRSAG
Human_NFkB_RelB     PATPPPWGCPGLGRLVSPAPGPGPQPHLVIITEQPKQRGMRFRYECEGRSAG

Human_NFkB_p50      GLPGASSEKNNKSYPQVKICNYVGPAKVIVQ--LVTNGKNIHLHAHSLVG
Human_NFkB_p52      GLPGASSEKGRKTYPTVKICNYEGPAKIEVD--LVTHSDPPRAHAHSLVG
Human_NFkB_c-Rel    SIPGEHSTDNNRITYPSIQIMNYGKGKVRIT--LVTKNDPYKPHPHDLVG
Mouse_NFkB_p65/RelA SIPGERSTDTTKTBTPTIKINGYTGPGTVRIS--LVTKDPPHRPHPHDLVG
Human_NFkB_RelB     SILGESSTEASKTLPAIELRDCGGLREVEVTACLVWKDWPHRVHPHSLVG

Human_NFkB_p50      KHC-EDGICTVTAGFKDMVVGFANLGILHVTKKKVFETLEARMTEACIRG
Human_NFkB_p52      KQCSELGICAVSVGPKDMTAQFNNLGVLHVTKKNMMGTMIQKLQRQRLR-
Human_NFkB_c-Rel    KDCRDGYYEAEFGQERRP-LFFQNLGIRCVKKEVKEAIITRIK-----
Mouse_NFkB_p65/RelA KDCRDGYYEADLCPDRSI-HSFQNLGIQCVKKRDLEQAISQRIQ-----
Human_NFkB_RelB     KDCTDGICRVLRPHVSPRSHSFNNLGIQCVRKKEIEAAIIERKIQ-----

Human_NFkB_p50      YNPGLLVHPDLAYLQAEGGGDRQLGDREKELIRQAALQQTKEMDLSVVRL
Human_NFkB_p52      -----SRPQGLTEAEQRELEQEAKELKKVMDLSIVRL
Human_NFkB_c-Rel    -----AGINPFNVPEKQLNDIEDCDLNVRL
Mouse_NFkB_p65/RelA -----TNNPFHVPIEEQ--RGDYDLNAVRL
Human_NFkB_RelB     -----LGIDPYNAGSLKN--HQEVDMNVVRI

Human_NFkB_p50      MFAFLPDSTGSFTRRLEPVSDAIYDSKAPNASNLKIVRMDRTAGCVTVG
Human_NFkB_p52      RFAFLRADGSFSLFLKPVISQPIHDSKSPGASNLKISRMDKTAGSVRG
Human_NFkB_c-Rel    CFQVFLPDEHGNLTTALPPVSNPIYDNRAPNTAELRICRVNKNCGSVRG
Mouse_NFkB_p65/RelA CFQVTVRDPAG-RPLLLTPVLSHPIFDNRAPNTAELKICRVNRNSGSCLG
Human_NFkB_RelB     CFQASYRDQQG-QMRMDPVLSEPVYDKKSTNTSELRICRINKESGPCTG

Human_NFkB_p50      GEEIYLLCDKVQKDDIQIRFYEEEENGGVWEGFGDFSPTDVHRQFAIVFK
Human_NFkB_p52      GDEVYLLCDKVQKDDIEVRFYEDDENG--WQAFGDFSPTDVHKQYAIVFR
Human_NFkB_c-Rel    GDEIFLCDKVQKDDIEVRFVL-----NDWEAKGIFSQADVHRQVAIVFK
Mouse_NFkB_p65/RelA GDEIFLCDKVQKEDIEVYFTG-----PGWEARGSSQADVHRQVAIVFR
Human_NFkB_RelB     GEELYLLCDKVQKEDISVVFSR-----ASWEGRADSQADVHRQIAIVFK

Human_NFkB_p50      TPKYKDINITKPASVFVQLRRKSDLETSEPKPFLYYFEIKDKEEVQRKRQ
Human_NFkB_p52      TPPYHKMKIERPTVFLQLKRKRGDVSDSKQFTYYFELVEDKEEVQRKRR
Human_NFkB_c-Rel    TPPYCK-AITEPTVKMQLRRPSDQEVSESMDFRYLFDEKDTYGNKAKKQ
Mouse_NFkB_p65/RelA TPPYADPSLQAFVRVSMQLRRPSDRELSEPMEFQYLPD-----
Human_NFkB_RelB     TPPYEDLEIVEFTVNVFLQRLTDGVCSEPLPFTYLERDHDSYGVDKKRK

```

Figure II-11 Alignment of NFkB subunits Rel-homology domain. The Rel-homology domain includes the DBD and the dimerization domain. Residues highlighted with dark magenta are completely conserved among subunits, whereas highlighting in aqua signifies significant variability. Alignment produced for mapping onto murine RelA structure, so the mouse sequence was used for the alignment. Figure was modified from the output of the Consurf server⁸⁴.

Materials and Methods

Protein Cloning, Expression, and Purification

The gene encoding full length NleC (residues 1-330) was PCR amplified from EHEC O157:H7 Sakai strain genomic DNA obtained from ATCC. Primers incorporating NdeI and

XmaI sites allowed ligation into a TYB2 vector (New England Biolabs) that allows the expression of the target gene as a fusion with a self-cleavable intein and a chitin binding domain at the C-terminus. This plasmid (pMS692) was used to express full-length NleC that, upon self-cleavage of the tag, results in NleC carrying two extra amino acids (PG) at the C-terminus. All fragments of NleC used in this study were PCR amplified directly from this vector with primers incorporating NdeI and XmaI sites and cloned into the same vector backbone. Plasmids created for use in this Chapter are listed in .

Table II-2 Plasmids used for the study of the specificity of NleC for NFκB and p300

Plasmid	Gene	Residues	Organism	Gene Source	Tag	Background	Primers	Resistance
pMS692	NleC	1-330	EHEC str. Sakai	EHEC str. Sakai DNA	C-term InteIn-CBD	pTYB2	ECs0847_F1/ECs0847_R2	Amp
pMS847	NleC E184A	1-330	EHEC str. Sakai	EHEC str. Sakai DNA	C-term InteIn-CBD	pTYB2	U743/L743	Amp
pMS1014	RelA	1-551	Human	J Goodrich, Univ of CO	6x-His	pET3		Amp
pMS1013*	RelA	17-291	Human	pMS1014	10x-His-SUMO	pET24d	U919/L919	Kan
pMS845	RelA	1-210	Human	pMS1014	6x-His	pET24d	U742b/L742b	Kan
pMS933	RelA	1-193	Human	pMS1014	C-term InteIn-CBD	pET24d	U814/L814b	Amp
pMS982	RelA	1-193	Human	pMS1014	10x-His-SUMO	pET24d	U848/L848b	Kan
pMS1009	RelA	1-551	Human	pMS1014	10x-His-SUMO	pET24d	U869/L869ab	Kan
pMS1010	RelA	29-49	Human	pMS1014	6xHis-BamA-RelA-BamD	pET24d	U868/L868	Kan
pMS1020*	RelA	29-49	Human	pMS1014	10x-His-SUMO-RelA-BamD	pET24d	U927/L927/L928	Kan
pMS1032	RelA R35A	17-291	Human	pMS1013	10x-His-SUMO	pET24d	U938/L938	Kan
pMS1033	RelA Y36A	17-291	Human	pMS1013	10x-His-SUMO	pET24d	U939/L939	Kan
pMS1034	RelA K37A	17-291	Human	pMS1013	10x-His-SUMO	pET24d	U940/L940	Kan
pMS1035	RelA C38A	17-291	Human	pMS1013	10x-His-SUMO	pET24d	U941/L941	Kan
pMS1036	RelA E39A	17-291	Human	pMS1013	10x-His-SUMO	pET24d	U942/L942	Kan
pMS1037	RelA G40A	17-291	Human	pMS1013	10x-His-SUMO	pET24d	U943/L943	Kan
pMS1038	RelA R41A	17-291	Human	pMS1013	10x-His-SUMO	pET24d	U944/L944	Kan
pMS1039	RelA S42A	17-291	Human	pMS1013	10x-His-SUMO	pET24d	U945/L945	Kan
pMS1114	RelB	124-413	Human	Functional Genomics Facility - Univ of CO	10x-His-SUMO	pET24d	U1088/L1088ab	Kan
pMS1115	p50	39-363	Human	J Goodrich, Univ of CO	10x-His-SUMO	pET24d	U1089/L1089ab	Kan
pMS981	p300	1-424	Human	J Goodrich, Univ of CO	10x-His-SUMO	pET24d	U849a/L849	Kan
pMS1116	p300	329-424	Human	J Goodrich, Univ of CO	10x-His-SUMO	pET24d	U849b/L849	Kan

* Prepared by Sandra Metzner

E. coli BL21(DE3) or Rosetta(DE3) cells (Novagen) were transformed with the desired NleC-expressing plasmid and plated onto LB agar plates with 50 µg/mL ampicillin. A single colony was used to inoculate 6 mL of LB containing 50 µg/mL ampicillin and the culture was grown overnight at 37 C. This culture was used to inoculate 50 mL of LB with ampicillin, and when the optical density at 600 nm reached 0.6, it was used to inoculate 6 L of LB. Once OD₆₀₀ reached 0.6 at 37 C, the cells were incubated on ice for 30 min before expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After

induction, the cells were grown at 20 C overnight before harvesting by centrifugation at 5000 x g for 15 min at 4 C. The cells were resuspended in buffer A [25 mM Tris pH 8.0 and 150 mM NaCl] supplemented with 1 tablet protease inhibitor cocktail per 50 mL lysis buffer (Complete ethylenediaminetetraacetic acid-free; Roche), and either frozen until later use or immediately lysed on ice by sonication. The soluble fraction was separated by centrifugation at 20,000 x g for 20 min at 4 C, and added to a chitin column pre-equilibrated with Buffer B [25 mM Tris pH 8.0 and 500 mM NaCl]. The soluble fraction was incubated on the column for 30 min at 4 C before the column was washed with at least 10 column volumes of Buffer B. The column was then flushed with one column volume of cleavage buffer [Buffer C: Buffer B + 10 mM _{DL}-dithiothreitol] and incubated overnight at 4 C to allow for cleavage of NleC from the chitin tag. The eluted fractions were concentrated and loaded onto a size-exclusion column (HiLoad 26/60 Superdex 75; Amersham Pharmacia Biotech) pre-equilibrated with Buffer A. The NleC containing fractions were pooled and concentrated. NleC was concentrated to a maximum of 30 mg/mL and stored at 4 C. NleC may also be stored at -80 C. Long term storage will result in degradation from both the amino- and carboxy-terminus, but was not found to affect activity significantly (as discussed in depth in Chapter III).

Plasmids encoding human RelA (p65), p50 and p300 were obtained from Dr. Jim Goodrich (University of Colorado at Boulder). A plasmid encoding RelB was obtained from the Functional Genomics Facility at the University of Colorado at Boulder. The gene fragments encoding RelA DBD (residues 17-291), RelB Rel homology domain (residues 124-413), and p50 (residues 39-363) were PCR amplified and ligated into pMS984, a modified pET24d plasmid obtained from Dr. Deborah Wuttke (University of Colorado at Boulder).

This vector encodes a 6-His-tag and small ubiquitin-like modifier (SUMO) at the N-terminus of the target gene. The tag is removed *via* a SUMO protease that results in a native N-terminus. The gene fragments encoding RelA residues 1-210 and residues 1-323 were also cloned into a modified pMS173, described above, that adds an N-terminal His-tag and TEV protease site. The gene fragments containing the N-terminus of p300 including the TAZ1 domain (residues 1-424) and the TAZ1 domain alone (residues 329-424) were PCR-amplified and ligated into pMS984.

All the substrate constructs were prepared in the same general way. *E. coli* Rosetta (DE3) cells (Novagen) were transformed with the plasmids and plated onto LB agar plates with 50 µg/mL kanamycin. Single colonies were used to inoculate LB containing 50 µg/mL kanamycin, and cells were grown, induced and harvested, as with NleC above. After sonication and centrifugation, the soluble fractions of cells were affinity purified on Ni-NTA beads (Qiagen). After 1 hour incubation of the soluble fraction on beads, Buffer E (Buffer A + 1 mM tris(2-carboxyethyl)-phosphine) was used to wash the beads with at least ten column volumes. Elution of the proteins was achieved with Buffer F (Buffer E + 250 mM imidazole (Gold Bio Inc.)), incubated with the beads for 1 hour to overnight. The beads were then washed with additional Buffer F, and these elutions were concentrated. The concentrated elutions were loaded onto the Superdex 75 for the RelA fragments and p300-TAZ1 domain, and the Superdex 200 for p300 N-terminus (Amersham Pharmacia Biotech). The protein fractions were utilized as is or concentrated as was necessary.

The RelA fragments were especially sensitive to oxidizing conditions, as we discovered after purifying and storing them in buffer containing β-mercaptoethanol (BME). The BME oxidized over time and formed an adjunct on the DNA-binding-loop cysteine,

preventing the recognition and cleavage by NleC. Beyond this cysteine modification, RelA degrades as a NleC substrate with some rapidity and all activity assays with the protein should be completed within a week of expressing the protein. Storage at -80 C slows the degradation but does not halt it. Though no proteolytic degradation is apparent, nor does RelA precipitate, the samples become “uncleavable” by NleC. Initial activity assays with RelA and the p300 constructs were performed in the presence of BME but were repeated with TCEP to verify that findings were not due to cysteine-related modifications.

NleC Biochemical Assays

The proteolytic activity of NleC on RelA, RelB, p50, and the p300 fragments was tested in Buffer E (25 mM Tris pH 8, 150 mM NaCl, 1 mM TCEP) under various protein concentrations. NleC also exhibits activity with other buffers, including MOPS pH 7. Cleavage was visualized using SDS-PAGE. In order to define the cleavage sites of RelA, RelB and p50, the reaction was allowed to proceed overnight and a SDS-PAGE gel overloaded with the products. This gel was electroblotted onto PVDF Mini Problott membrane (Applied Biosystems) and stained with Coomassie Blue according to standard procedure. The bands corresponding to the C-terminal pieces of RelA, RelB or p50 were sent for N-terminal sequencing by Edman degradation at the UTMB Protein Chemistry Laboratory. The N-terminus of a fragment of NleC after incubation with aminopeptidase was also determined with N-terminal sequencing with the same methods described above. The addition of metals in the reaction buffer was not necessary to obtain cleavage, though in other works testing the *in vitro* activity of NleC, ZnCl₂, CaCl₂ and MgCl₂ are often added. To test the cleavage efficiency in the presence of DNA, a palindromic 20-bp DNA oligomer (5'-

CGGCTGGAAATTTCCAGCCG-3') containing the NF κ B consensus sequence, GGRRNNYYCC, was incubated with RelA for 1 hour prior to the addition of NleC for the activity assay.

Cleavage of both fragments of p300 was tested under the same conditions as RelA above. For the pulldown assay, the 6xHis-SUMO-tagged N-terminal fragment (residues 1-424) was bound to Ni-NTA resin. Inactive NleC (E184A) was added in excess and washed with Buffer E for four column volumes, followed by washing with Buffer G (Buffer E + 50 mM imidazole) for four column volumes. Buffer F was then used to elute the contents of the Ni-NTA columns. Negative controls were NleC on Ni-NTA resin with no His-tagged protein bound and p300 bound Ni-NTA resin with no NleC added. The various fractions were visualized using SDS-PAGE.

SDS-PAGE Analysis

After activity assays are performed, the results were visualized with SDS-PAGE and Coomassie stained. To quantify the resulting gels, they were imaged and then analyzed with ImageJ. The lanes were selected and pixels across the length of the gel were readout into a histogram. From this, the peaks on the histogram that represent the bands of interest were selected. The peaks were integrated with a standard width at the base. The same area in control lanes was selected as well. To account for differences in staining, the same area in the control zero timepoint lane was subtracted from the experimental lanes. For the mutagenesis activity assay, this resulting number was then normalized against the wild-type activity to get a percentage cleavage efficiency.

Chapter III. Structure of the T3SS effector NleC, a new Zincin protease

Introduction

Structural studies of Type III Secretion System (T3SS) effectors are instrumental in understanding mechanism and for designing inhibitors, in part because many effectors do not have detectable sequence similarity to characterized proteins with known function. Upon structural characterization of these effectors, they exhibit a degree of structural mimicry when compared to well-known proteins of the same function, without being homologous. This means that structural characterization of effectors often reveals a novel, divergent structural solution for a functional, chemical problem. Characterizing the structure of an effector can lead to hypotheses about function, which otherwise would be difficult to unravel. Often, knockout experiments of individual effectors do not result in a visible phenotypical difference, and without clues into the function, molecular characterization is nearly impossible. With this in mind, we sought to determine the structures of effectors with unknown function and structure with the hope of shedding light on these proteins' places in pathogenicity.

When our lab initiated study on T3SS effectors, we pursued a “high-throughput” approach to identify effectors of unknown function that were amiable to crystallization. We targeted the recently described putative effectors from EHEC Sakai strain, detailed by Tobe *et al.* for cloning into plasmids for expression of His-tagged fusions³⁷. NleC was one of the

effectors that expressed solubly, determined *via* a novel method, the colony filtration blot or CoFi blot. This method screens colonies directly for expression of soluble proteins His-tagged proteins⁸⁷.

After optimizing the expression platforms and screening for the crystallizable core fragment of NleC, I obtained a crystal of an NleC fragment. With the publication of reports disclosing the function of NleC as a zinc protease, we were able to obtain a full dataset using the anomalous signal from the active site Zn²⁺ ion.

Here I report the structure of NleC from EHEC Sakai strain refined to 1.8Å resolution with coverage from residues 22-280. Through comparison with other zinc proteases, NleC is shown to be a member of the Zincin zinc protease superfamily and represents a new subfamily, with a unique active site motif and deviant fold. The structure also provides the basis for hypotheses into how NleC recognizes its substrates.

Results

Crystallization and structure determination of NleC

Full length NleC (amino acids 1-330) was expressed and purified to homogeneity as described in Materials and Methods. Initial crystallization trials in sitting drops yielded 50 µm crystals from a precipitant containing 20% PEG-3350 and 0.2M Mg(CHO₂)₂ at 2.5 mg/mL final protein concentration, but only after eight months of incubation at 16 °C. A native dataset to 2.6Å resolution was collected from one of these crystals, which belonged to space group P2₁2₁2₁ with a 45.02 by 67.55 by 81.95Å asymmetric unit, and likely can accommodate one molecule. This estimate is made by calculating the Matthews coefficient, which is the crystal volume per protein molecular weight (kDa), and comparing this to a

database of known values⁸⁸. Requiring a long incubation time to obtain the crystals suggests that proteolysis occurring during incubation may have yielded the crystallizable fragment. SDS-PAGE of the crystallization drops confirmed the degradation of the full-length protein and accumulation of a lower molecular weight bands (Figure III-1A). In-gel trypsin digestion of these bands followed by mass spectrometry analysis showed significant sequence coverage from amino acid 8 to 287. The molecular weight of this fragment, at 30 kDa compared to 37 kDa for full-length protein, is consistent with the ~5 kDa gel shift of the NleC bands from the crystallization trays. There were, however, less represented C-terminal fragments present in the mass spectrometry experiments, leading us to question the confidence of the mass spectrometry derived C-terminus. By digesting full-length NleC with aminopeptidase followed by Edman N-terminal sequencing at UTMB, the aminopeptidase resistant core was further refined. Though many residues were visible in the Edman degradation results, the best represented fragment began at residue 19. Based on these data, a variety of N-terminally and C-terminally truncated NleC constructs were cloned and expressed to facilitate more rapid crystal growth and phase determination (Figure III-1B).

One of the N-terminal truncations of EHEC NleC, NleC₁₉₋₃₃₀ (residues 19-330), yielded crystals after several months of incubation in similar crystallization conditions to the full-length protein crystal. Once the crystal was mounted at the Advanced Light Source Synchrotron at Lawrence Berkeley National Labs, a fluorescence scan of the crystals can determine the presence of any heavy atoms with absorption peaks at wavelengths between 1 and 2 Å, which is the wavelength range accessible at the Synchrotron beamline. An excitation scan was used to determine that zinc was present in a native NleC crystal. Upon

excitation with X-rays, zinc fluoresces at with a characteristic K-edge at 1.44 Å. Additionally, this signal is, in favorable cases, strong enough to be used as an anomalous scatterer in phase determination. The NleC crystal containing zinc diffracted very well, and full datasets were collected, at the zinc peak 1.2831 Å, at the inflection point 1.2835 Å, and at a remote wavelength 1.2574 Å. I had also prepared a number of NleC crystals with different heavy metal soaks to increase the likelihood of collecting data on a derivative that would allow phase determination. Unfortunately, the soaked crystals did not diffract as well, their lattices seemingly distorted by the extra manipulation or heavy metal soak.

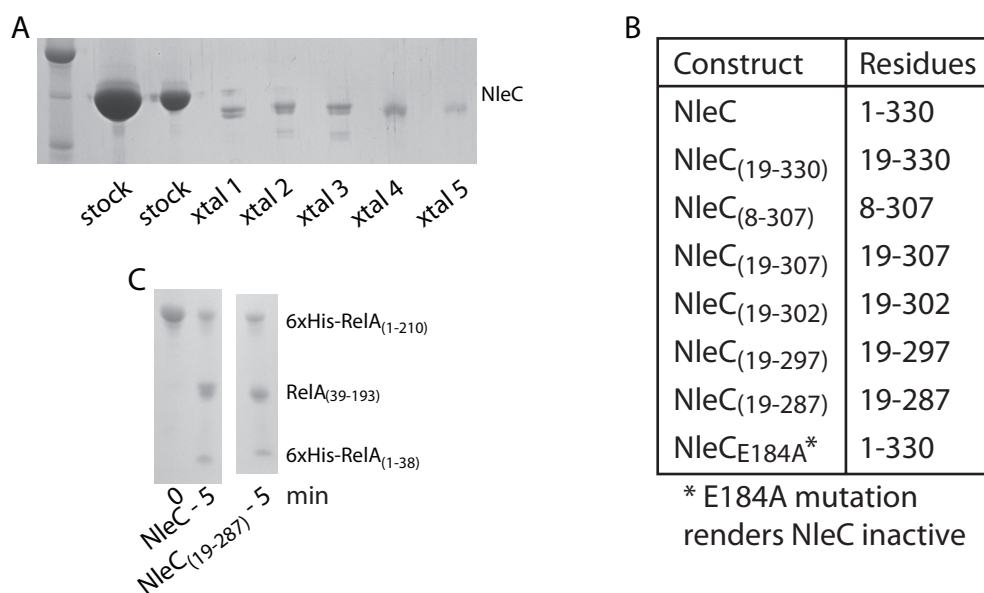


Figure III-1 NleC truncations during crystallization. A) Degradation during incubation in crystallization conditions as shown with SDS-PAGE. The two stock lanes are the same stock at different concentrations. Xtal 1-5 lanes contain different crystallization conditions showcasing the lower molecular weight protein present after eight months incubation. B) Different NleC constructs were constructed to identify the crystallizable core and obtain diffraction-quality crystals with shorter incubation times. An inactive full-length mutant was also constructed. C) Comparison of RelA cleavage by full-length NleC and NleC₁₉₋₂₈₇, which approximates the fragment seen in the crystal structure. Similar quantities of RelA are cleaved in these two 5 minute assays, which were not done at the same time.

The structure of NleC was determined from the native crystal using Multiwavelength Anomalous Dispersion methods and the three-wavelength dataset described above (see Materials and Methods for details). After determining from analysis of the initial images that the crystal lattice was orthorhombic, a full dataset at the three wavelengths was collected to ensure completeness. First the data sets were indexed and scaled, and the space group identified as $P2_12_12_1$. to identify peaks and combine the numerous raw images for further processing. To phase the data, the peak and inflection datasets were compared to the remote wavelength to determine the intensity differences structure factors that arose from the single Zn^{2+} anomalous signal in the unit cell. This was enough to unambiguously determine the location of the Zn^{2+} atom and use it to calculate the phases. From this, an initial electron density map is obtained from the experimental Zn-derived phases and the preliminary model was automatically built. Through a series of manual rebuilding of the model to better fit the electron density data, alternated with automated refinement, the original model was refined, improving the R_{work} and R_{free} to an R_{work}/R_{free} of 22.7/18.3. This measurement represents how well the model fits the data compared to an unaltered control dataset. The final model, refined to 1.8-Å resolution, contains residues 22-280, a Zn^{2+} ion in the active site, and a Mg^{2+} ion distal from the active site (RCSB PDB=undeposited) (Figure III-3, Figure III-2). Data collection, phasing and refinement statistics are summarized in Table III-1.

Using molecular replacement, the final NleC model was used to phase the dataset obtained from the crystallization trial of full-length NleC. No electron density for additional residues was observed in this structure and the model, refined to 2.6Å resolution, is identical to the higher-resolution model. An NleC fragment (residues 19-287)

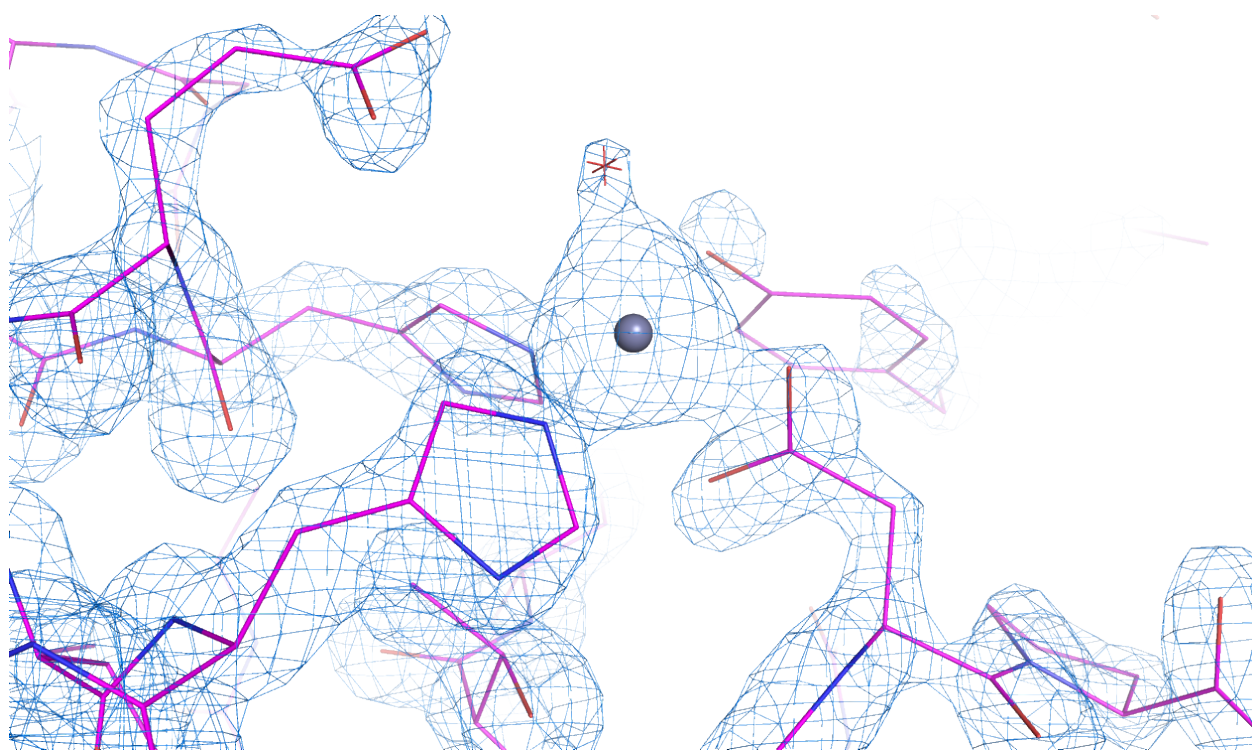


Figure III-2 Electron density map surrounding Zinc in the active site. Electron density contoured to 2σ .

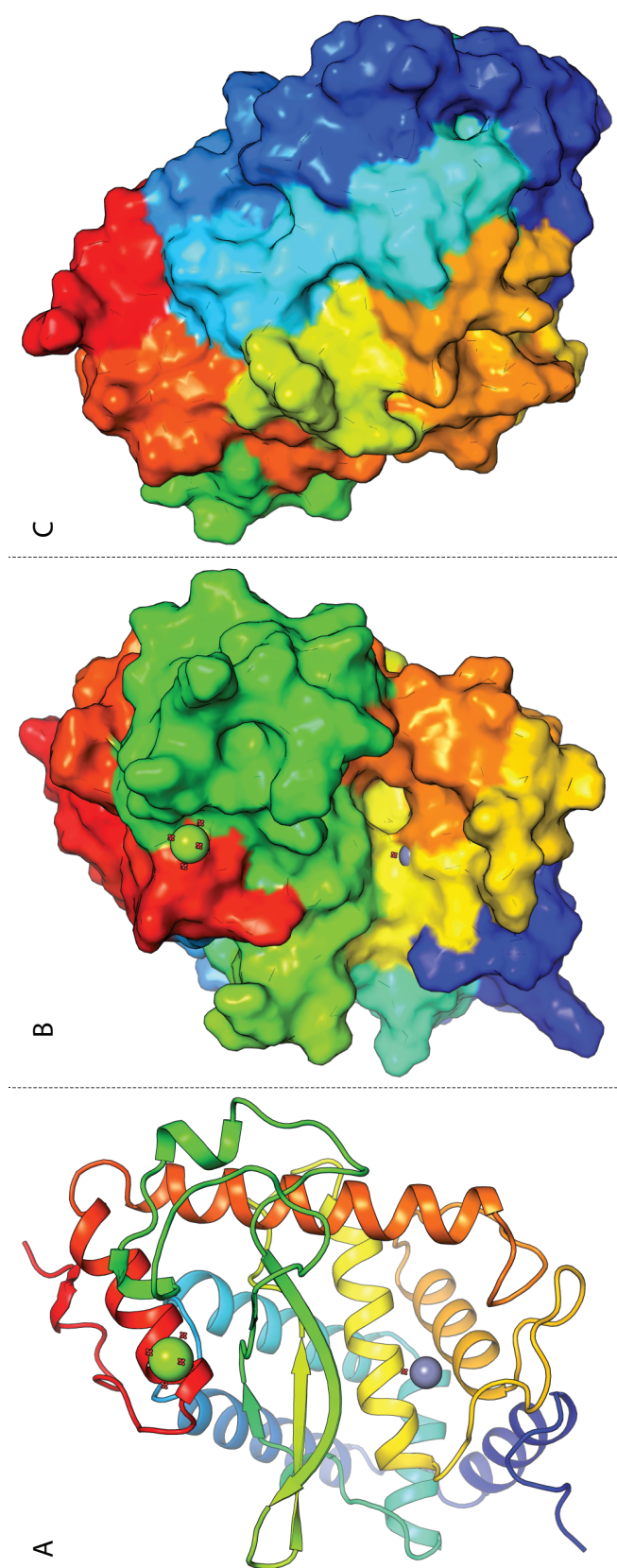


Figure III-3 Structure of NleC. A) Cartoon representation of NleC in where blue is the amino-terminus and red is the carboxy-terminus. Active site zinc ion is in grey and calcium ion is in pink. B & C) Surface representation of NleC, produced with added hydrogens. The zinc ion is mostly buried by the coordinating residues, while the calcium ion is completely exposed, illustrating the lack of coordination by NleC directly.

Table III-1 NleC Crystallographic Statistics

Data Collection			
Space group	P212121		
Cell dimensions			
a	45.21		
b	67.48		
c	81.69		
$\alpha=\beta=\gamma$	90		
	<i>Remote</i>	<i>Peak</i>	<i>Inflection</i>
Wavelength	1.2574	1.2831	1.2835
Resolution (Å)	50-1.9	50-1.9	50-1.9
<i>R</i> _{linear}	8.1 (48.1)	7.7 (47.0)	7.7 (44.7)
<i>I</i> / σ <i>I</i>	13.3 (1.8)	15.8 (1.8)	14.7 (1.5)
Completeness (%)	96.4 (97.0)	96.0 (88.2)	96.4 (96.5)
Redundancy	2.3 (1.8)	2.4 (1.7)	2.3 (1.8)
Refinement			
Resolution (Å)	34.9-1.9		
No. reflections	19840		
R _{work} /R _{free}	18.3/22.7		
No. atoms			
Protein	2204		
Ligand/ion	2		
Water	137		
B-factors			
Protein	27.5		
Zn (Mg)	36.8 (47.0)		
Water	34.1		
R.M.S. deviations			
Bond lengths (Å)	1.008		
Bond angles (°)	0.014		

approximating the crystallographic model is similar to full-length NleC in its ability to cleave RelA, indicating that the model represents the catalytic core of NleC (Figure III-1C).

NleC structure

The structure of NleC consists of eight extended α -helices and three β -strands, cradling the active site zinc (Figure III-3). A hexahydrated magnesium ion is also present in the structure, but it not coordinated to the protein directly, and is not expected to be necessary for function. NleC was suspected to represent a novel family of the Zincin zinc protease fold superfamily because it shares the highly conserved zinc-binding motif, HExxH, while outside of this motif, displays undetectable sequence similarity with other

Zincins (SCOP FSF d.92.1, PFAM CL0126 Peptidase MA, MEROPS clan MA)⁸⁹⁻⁹². A search on the Dali server confirms this hypothesis and is further discussed in the next section.

The zinc-ligating residues His183, His187, Asp194 and Tyr227 reside in the motif, HEIIH + D + Y (Figure III-4). The presence of a zinc-coordinating aspartate places NleC in the Aspzincin subfamily. Aspzincins as a family overlap with Metzincins, which often have aspartates at the same position coordinating zinc, but additionally have a structural element called the Met-turn⁹³⁻⁹⁶. The fifth coordination site to the zinc ion is provided by a water molecule thought to be the nucleophile that attacks the carbonyl of the substrate scissile bond (Figure III-4). The coordination geometry in NleC is penta-coordinated, with the glutamate exhibiting bidentate binding geometry. By custom, a bidentate glutamate is considered one coordination site⁹⁷⁻⁹⁹.

While most Zincins have a five-stranded β -sheet, NleC has a three-stranded sheet. The three-strand mixed β -sheet of NleC is similar to the structural element known as the Ψ -loop motif, but is unique (Figure III-5). The interior strand exits the sheet midway, allowing the two exterior strands to interact directly. This allows the last strand to interact with both the middle and first strands. We are unaware of another occurrence of this fold in known structures and a Dali server search with only this portion did not return any results.

In addition to NleC from EHEC *E. coli*, BLAST reveals that NleC homologs are found in other enterobacteria, in the fish pathogens *P. damsela* and *Vibrio* species, and the insect pathogen *Arsenophonus nasoniae* (Figure III-17). The NleC family consensus active site motif, LIHExxHxxTxxxD + Y, was identified through a sequence alignment of these family members (MEROPS family M85)⁸⁹. The ConSurf server was used to map the conservation

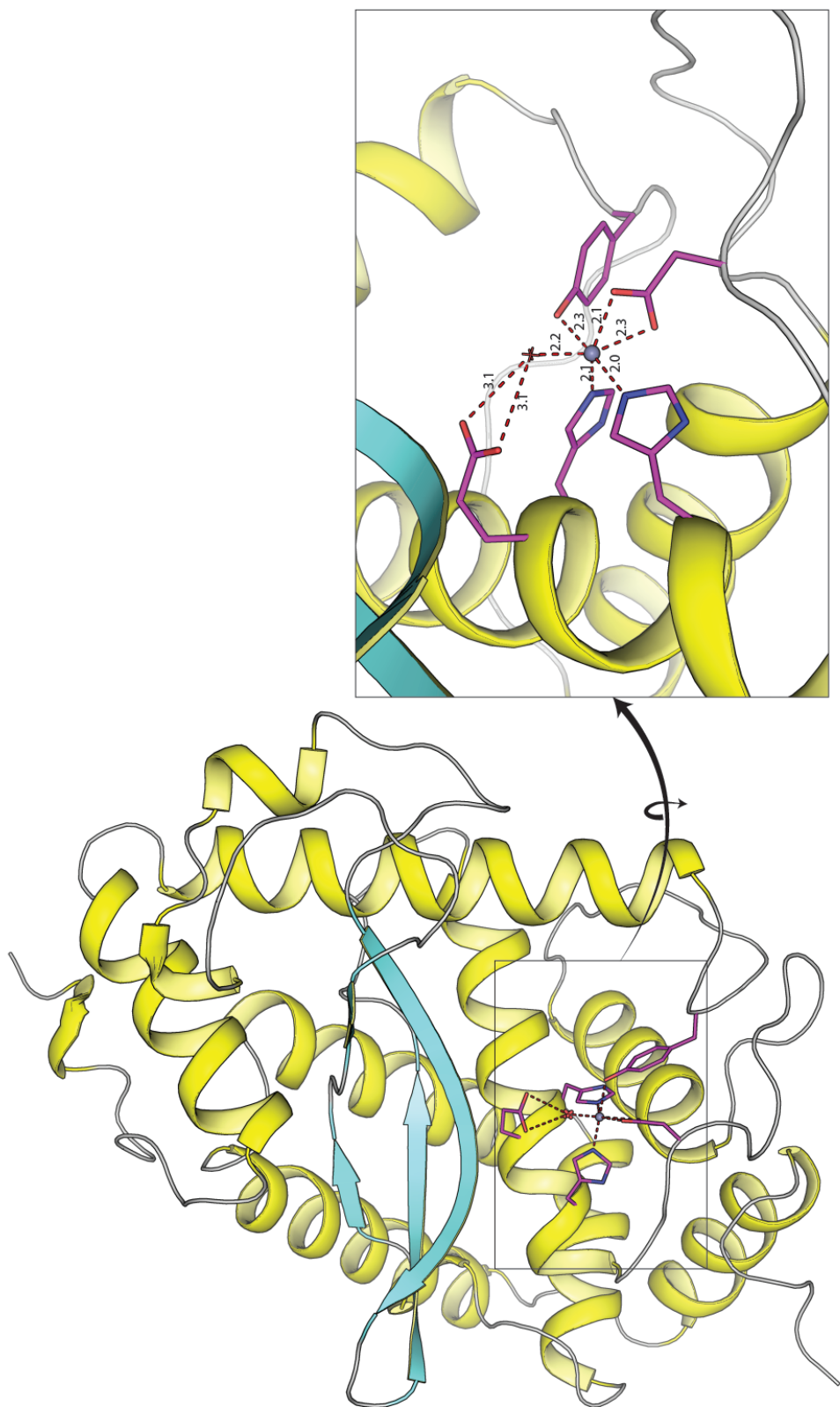


Figure III-4 NleC active site. NleC is shown in the standard Zincin orientation as described by Gomis-Rüth⁹⁶. The Zn²⁺-coordinating residues are shown in magenta, whereas secondary structure elements helices are shown in yellow and strands in aqua. Label units are Angstroms.

onto the structure of *E. coli* EHEC NleC (Figure III-6)^{84,100}. Within the NleC family, the active site is strictly conserved, as are many residues that make up the hydrophobic core of the protein. However, the surface of the protein is highly variable, and there are insertions into many loops, likely due to the differences in targeting and pathogenic environments encountered.

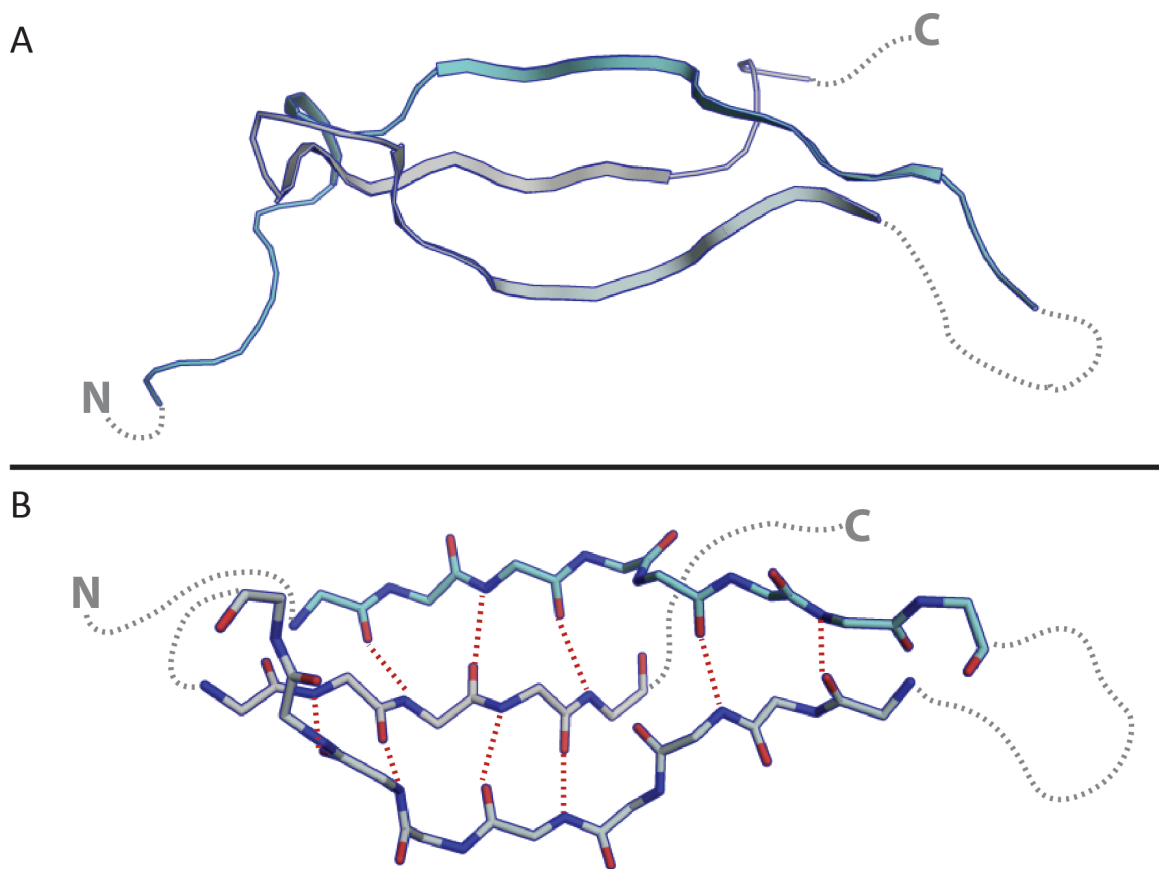


Figure III-5 NleC β -sheet. The β -sheet of NleC has a unique structure, similar to a ψ -loop motif. A) The outer strands of the motif are amino-terminal to the central strand, which in this case, exits the β -sheet and allows the two outer strands to interact directly. This is a very unique structure and a Dali search using just the β -sheet did not provide any similar structures. B) Hydrogen bonding within the NleC β -sheet. Red dotted lines delineate the hydrogen bonds which are representative of β -sheet formation.

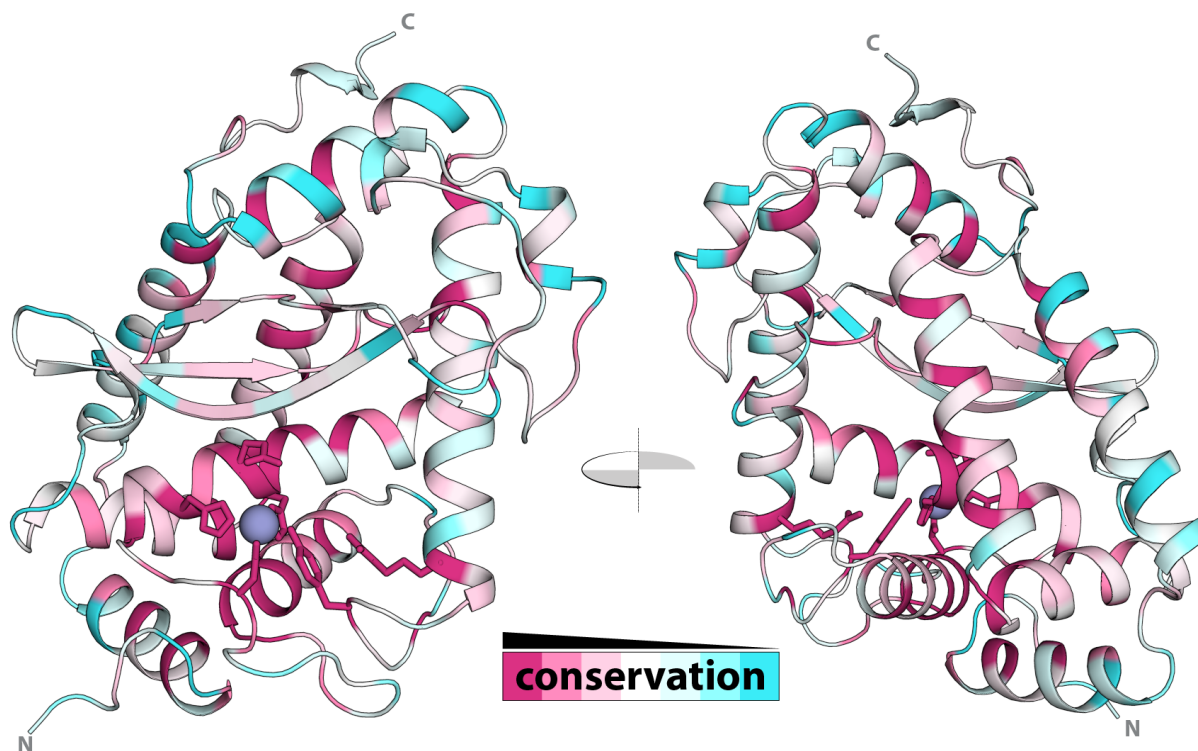


Figure III-6 Conservation within the M85 NleC peptidase family. A multiple sequence alignment was mapped on to the structure of NleC with the Consurf server. Magenta indicates complete conservation among all family members, whereas aqua indicates the most variability. The Zn^{2+} -coordination and active site is the most conserved region of the NleC family, followed by many interior residues. The sequence alignment that was used to construct this is found in Figure III-17 at the end of the chapter.

Structural Comparison of NleC to Other Zincins

When I compared the NleC structure with other Zincins, I found that NleC maintains the basic Zincin fold of a three-helix bundle and mixed β -sheet around the active site but has several unique features. Utilizing the Dali Server, the closest structural relatives are aminopeptidase tricorn (Z score = 6.3, RMSD = 3.3) and botulinum toxin (Z score = 5.7, RMSD = 3.3) (Figure III-7)⁹². The Dali Z-score is calculated by the comparison of intramolecular distances between two structures, and a score over 2 represents significant similarity, often culminating in similar folds. The unique three-stranded β -sheet present in

NleC is a deviation from the five-strand canonical Zincin fold which is represented by both closest structural relatives as well as snapalysin, the minimal structure for family MA of the MEROPS peptidase database^{89,101}. The closest structural neighbors, while sharing a similar catalytic domain, also contain insertions and have significant additional sequence N and C-terminal to the catalytic domain. Finally, the presence of a zinc-coordinated tyrosine is only found in one other Zincin family, the astacin family (Figure III-8A), though many Zincins have a tyrosine in a similar location, hydrogen bonding a water molecule that then directly coordinates the zinc (Figure III-8B)⁸⁹.

Distinct C-terminal domain

The crystal structure I solved for NleC shows electron density from residues 22-280, with the final 50 residues of NleC not visible in the electron density. This lack of density either means that the C-terminal portion is unstructured and not in a discrete, crystallizable state, or that it was actually missing from the protein, having been proteolyzed during the course of incubation. Because we did observe a truncated NleC in mass spectrometry analysis of crystal drops, the latter case is more likely. NleC₂₈₀₋₃₃₀ could exist in one of two ways: either it wraps around the catalytic domain and is not a separate domain itself, or it is a discrete domain, whether folded or unstructured in structure. To differentiate between these possibilities, we performed small angle X-ray scattering (SAXS) on the full length NleC and on a C-terminal truncation, NleC₁₉₋₂₈₇. Using this technique, it is possible to calculate a molecular envelope of the molecule in solution. However, it would be difficult to differentiate the presence of the C-terminus if it wraps around the crystallographic core domain or is unstructured, as the results may be similar to the C-terminal NleC truncation, NleC₁₉₋₂₈₇.

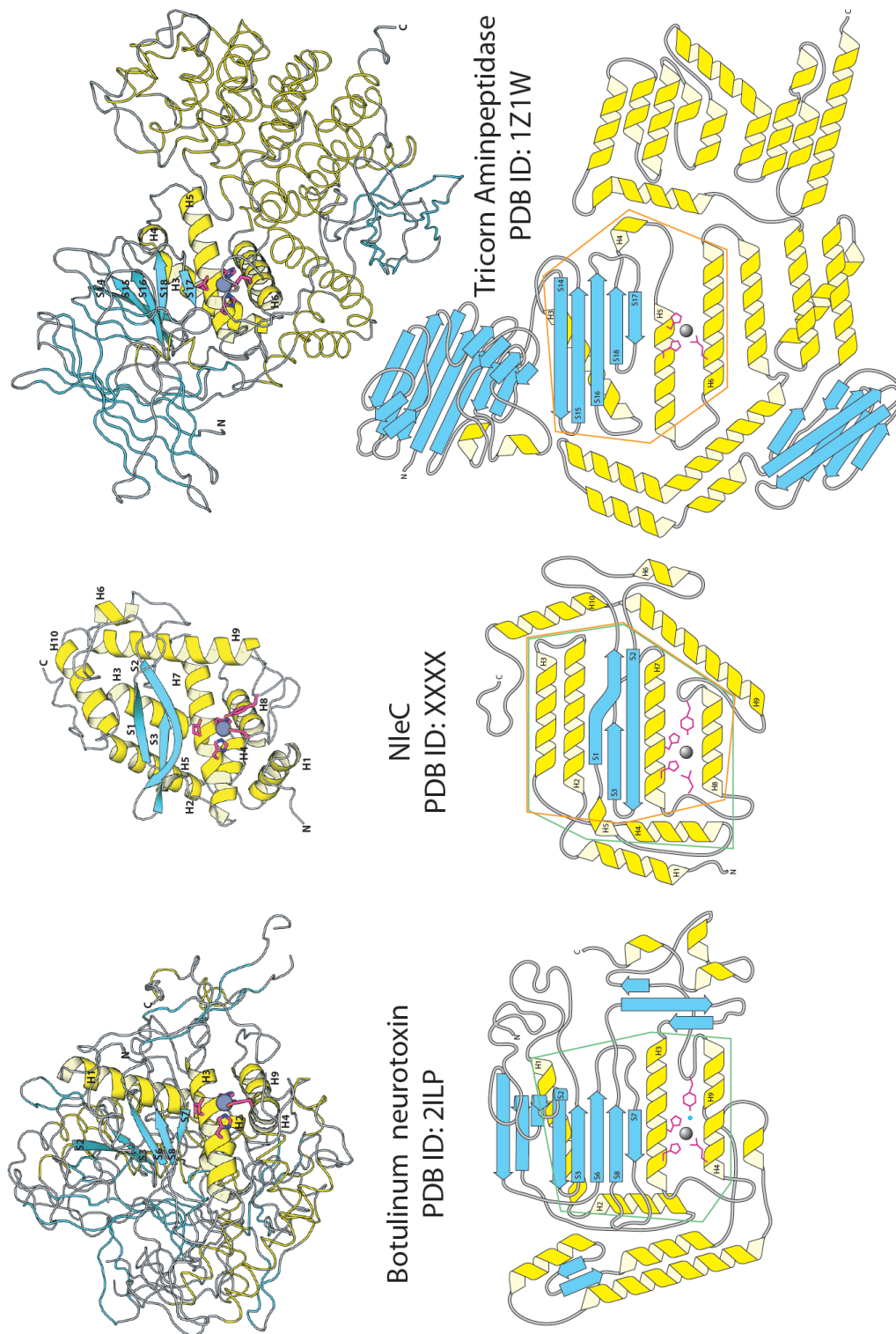


Figure III-7 Structures and topography diagrams of NleC and closest structural relatives. Searching with the Dali server, when compared to NleC, tricorn aminopeptidase has a Z-score of 6.3, and botulinum neurotoxin has a Z-score of 5.7. In the depictions of the structures, only the features that are shared with NleC are shown as cartoon representations, whereas the rest of the structure is shown in the ribbon format. In the topography diagrams, the features that are shared with NleC are encircled with green for botulinum toxin, and orange for tricorn aminopeptidase.

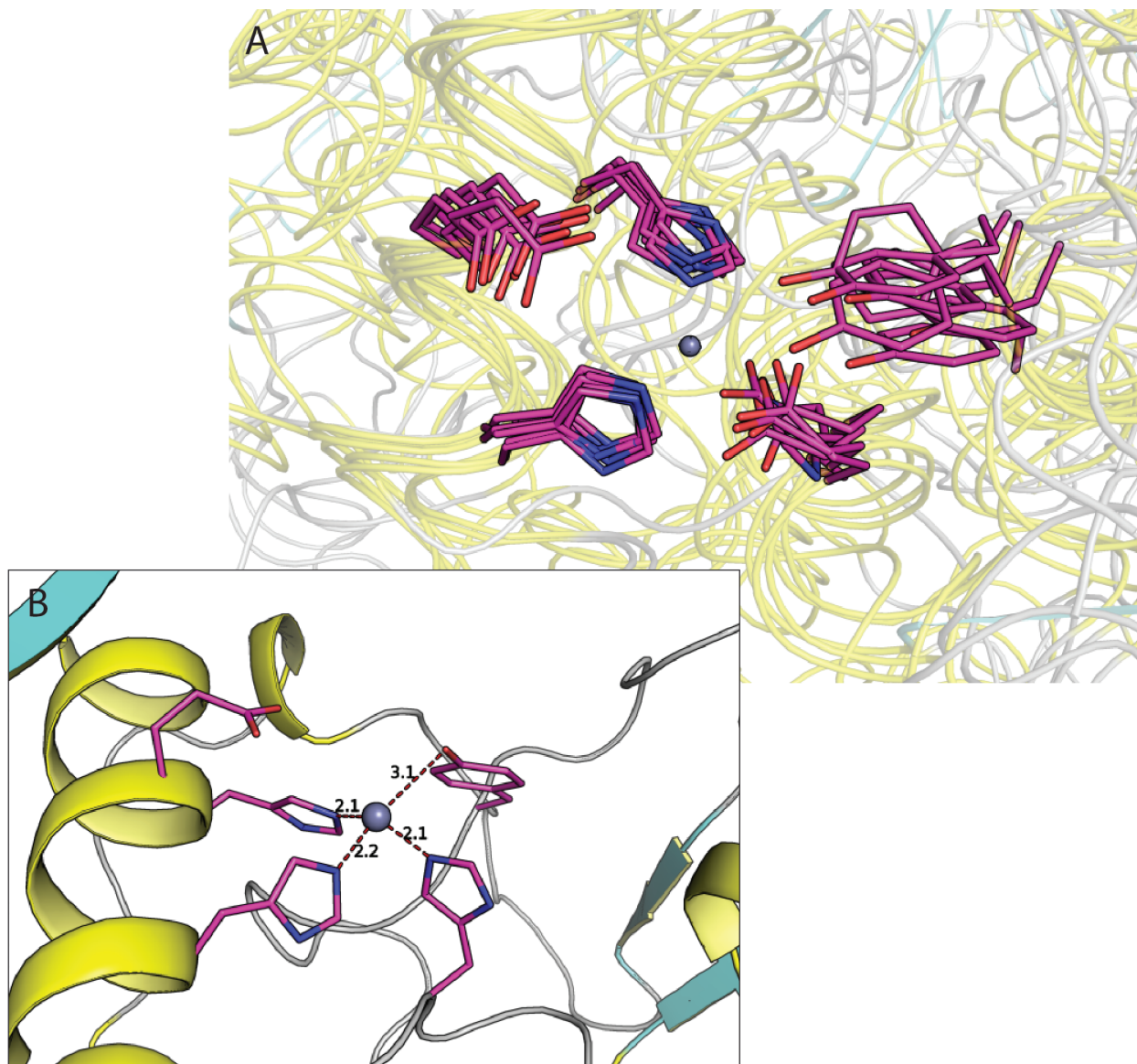


Figure III-8 Active site tyrosine among Zincins. A) Numerous Zincin proteases have a tyrosine residue in their active sites that do not coordinate the Zn^{2+} ion. It is unknown what the function of this tyrosine is. PDB IDs = 2ILP, 1SRP, 2C6F, 3O0Y, 3FU6, 3BOK, 1YVG, 3EBI. B) The Astacin M12 zinc protease family is the only other known family where the tyrosine in the active site is within range to coordinate the Zn^{2+} ion directly, though it is not as close as the tyrosine residue in NleC. PDB = 1QJJ.

After purification and expression, NleC and NleC₁₉₋₂₈₇ were concentrated and SAXS data was collected at the ALS Synchrotron at Lawrence Berkeley National Labs. For the full-length NleC protein, the longest exposures and highest concentration of protein resulted in aggregation and altered signal, however the lower concentrations and shorter exposures

were better behaved (Figure III-9, top panel). Unfortunately, the truncated protein, NleC₁₉₋₂₈₇ appeared to be aggregated under most conditions, preventing the calculation of an envelope (Figure III-9, bottom panel). Using the ATSAS program suite, I was able to calculate an envelope for full-length NleC with both the GASBOR and DAMMIF methodologies (Figure III-10). Both programs use the SAXS data and user-input on the size of the molecule to calculate an *ab initio* model of the molecule in solution, but use different approaches. GASBOR creates an envelope using a beads-on-a-string model and iteratively melts and refolds this model by calculating a theoretical SAXS dataset based on the model and matching this to the experimental data. DAMMIF starts with a cube and stepwise compares the resulting theoretical SAXS data to the experimental data as it removes portions of the cube. When an iteration does not change the model substantially from the previous step, the program terminates, creating a final best-fit molecular envelope for the given experimental data. For these data, I calculated 20 different envelopes for both methodologies, and used the DAMMIN analysis suite to weigh and average these into a final model for each.

The SAXS envelopes obtained from the full-length NleC data clearly show two separate domains, the largest of which can accommodate the zinc protease domain (Figure III-10). The smaller domain envelope similarly is large enough to fit the final fifty residues. There are also twenty residues from the amino-terminal end missing from the structure. It is unknown where these residues would fit in the envelope, whether wrapped around the main protease domain, incorporated with the carboxy-terminal fragment in the accessory domain, or in an unstructured random coil.

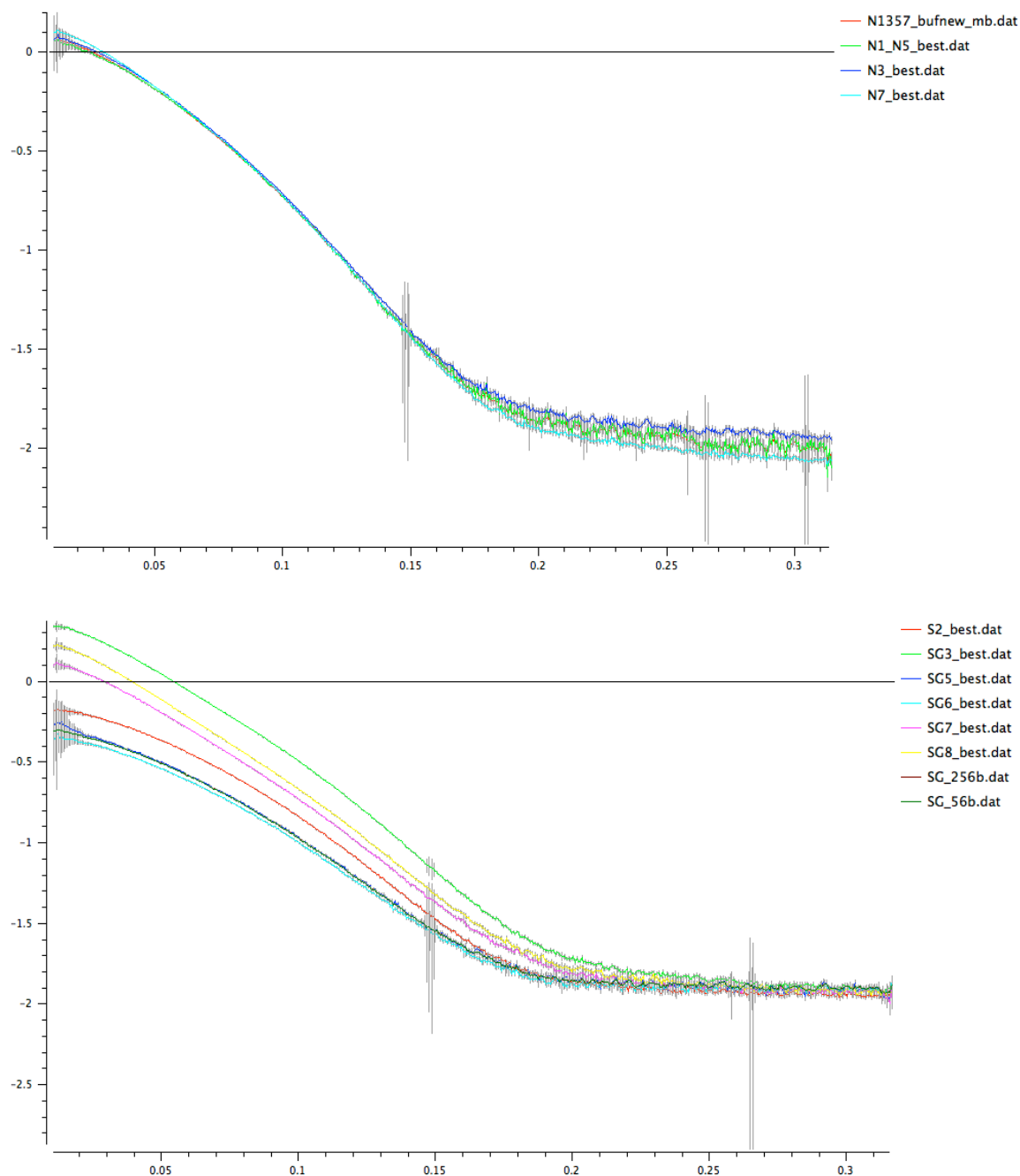


Figure III-9 Small angle X-ray scattering representative curves. Top panel: Full-length NleC curves. Low concentration samples and shortest exposure times produced curves that overlapped well and could be combined to be analyzed further. Bottom panel: NleC₁₉₋₂₈₇ truncation curves. None of the curves overlay, regardless of exposure time or protein concentration, suggesting that the data is too poor to analyze with any confidence. If the SAXS curves overlap, this shows that the protein is maintained in a similar molecular envelope. When the curves change between samples or exposure times, this suggests that the protein envelope is different, most likely from aggregation, and combining these samples together for analysis would give an averaged, or false molecular envelope.

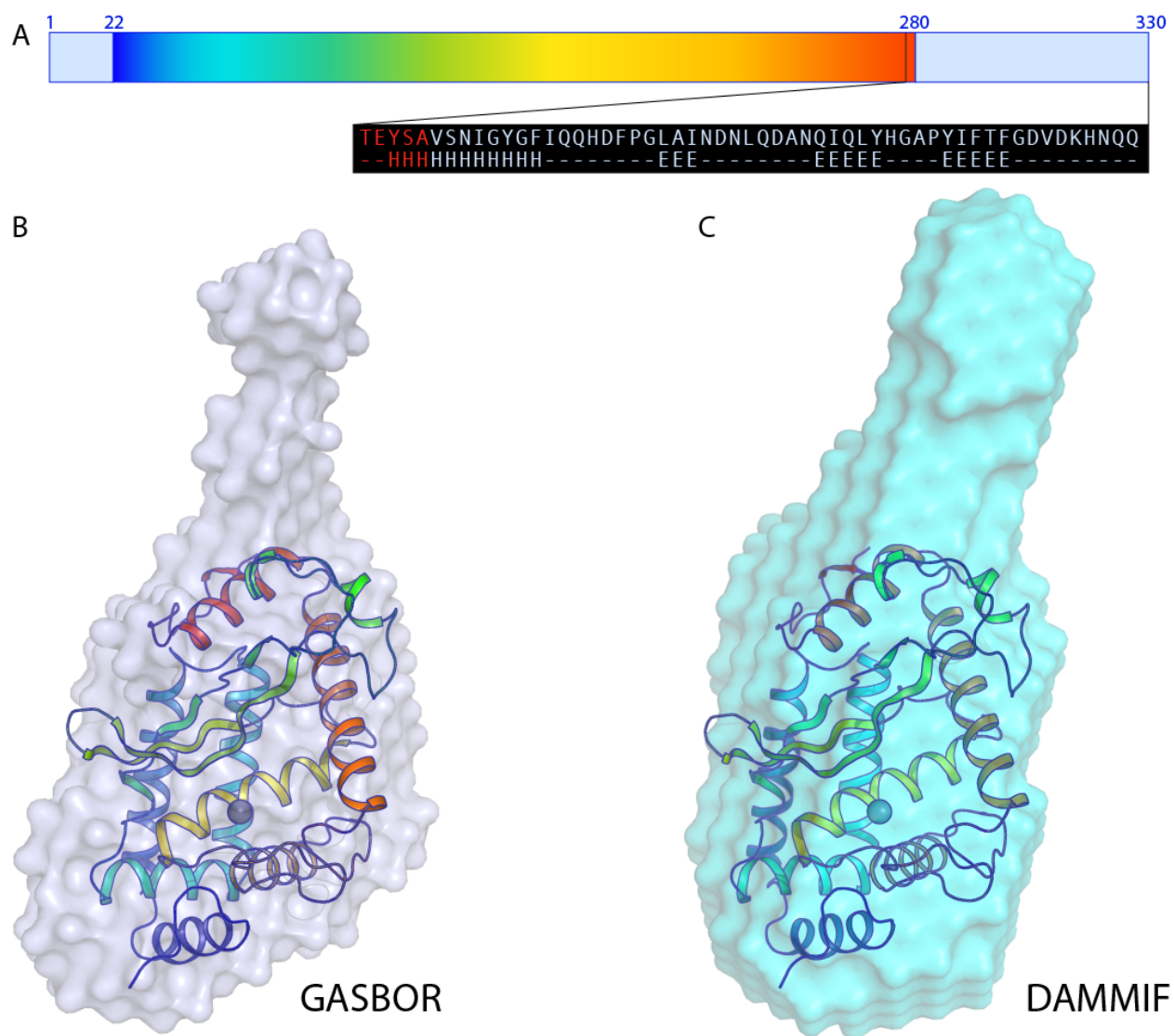


Figure III-10 NleC SAXS envelopes. A) Depiction of NleC crystallographic fragment and C-terminal domain. The portions of the NleC sequence that are not present in the crystal structure are shown in light blue, whereas the crystallographic fragment itself is colored in blue at the N-terminus and red at the C-terminus. The second bar depicts the sequence and Jpred secondary structure prediction for the C-terminal fifty residues, which are missing from the crystal structure. B) and C) are the GASBOR and DAMMIF-calculated molecular envelopes for full-length NleC. The crystal structure of NleC was fitted into the molecular envelopes and the remaining envelope analyzed. Both methods show an extended portion of the envelope which could account for the truncated portions, especially the C-terminal fragment.

Discussion

NleC and Zincin Conservation

Early in evolution, proteins utilized metals for various purposes. Most early metalloprotein folds were able to bind different metal ions by one site or had multiple metal-binding sites per protein. The metals that were most prevalent in the early ocean were iron, manganese, cobalt and nickel and these were some of the first specific-metal folds to develop¹⁰². However, the affinity of the most commonly used metals in coordination environments is described by the Irving-Williams series, and is $Mg^{+2}/Ca^{+2} < Mn^{+2} < Fe^{+2} < Co^{+2} < Ni^{+2} < Cu^{+2} \sim Zn^{+2}$ ¹⁰²⁻¹⁰⁴. This strong binding of zinc by ligands is perhaps why zinc-specific protein folds were in fact the first metal-specific protein fold to develop and helps account for the prevalence of zinc-binding proteins in the genome^{102,105,106}.

Among zinc metalloproteins, the zinc acts in either a structural or a catalytic role. Many transcription factors, for example, contain zinc fingers, where zinc plays a structural role. Zinc metalloproteases, such as the large Zincin family, have a water molecule coordinated in the active site, which nucleophilically attacks the carbonyl of the substrate peptide bond. The prototype for this mechanism is thermolysin¹⁰⁷. In Zincins, defined by the conserved motif HExxH and a similar fold, the zinc is coordinated by the two histidines of the motif and the glutamate serves to hydrogen bond the catalytic water^{96,107}. A third coordinating residue defines the subfamilies of Zincins, with Glut zincins as the most common with an additional glutamate. In another subfamily, the Asp zincins, an aspartate makes up the third residue, whereas the third family, Metzincins, is characterized by a

conserved methionine near the active site, and has either a histidine or aspartate as the third zinc-coordinating residue (Figure III-11).

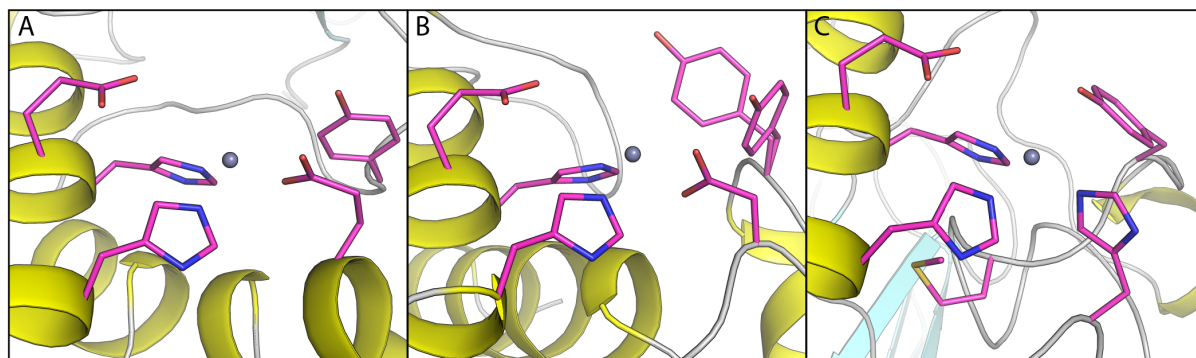


Figure III-11 Zincin Subfamilies. A) Gluzincins, or Zincins with a Zn²⁺-coordinating glutamate are represented here by botulinum neurotoxin from *Clostridium botulinum*. MEROPS family M27. PDB: 2ILP. B) Aspzincins were originally a part of the Metzincin family, but lacking the conserved methionine-turn, are now a separate subfamily. Deuterolysin from *Aspergillus oryzae*, a widely used in the food industry, represents this family. This structure also shows two orientations for the active site, a mechanism called the tyrosine switch, which is expected to occur to accommodate substrate binding. MEROPS family M35. PDB: 1EB6. C) Metzincins contain a conserved methionine residue located near the active site that is in a Met-turn motif. The third zinc-coordinating residue is either a histidine or aspartate in this family. In this example, serralyisin from *Serratia marcescens* sp. E-15, a third histidine coordinates the zinc. MEROPS family M10. PDB: 1SRP.

Appearing before Eukarya split from the last common ancestor, the Zincin zinc protease fold superfamily (SCOP FSF d.92.1, PFAM CL0126 Peptidase MA, MEROPS clan MA) is one of the most ancient metal-binding protein folds^{89-91,102,106}. Though sequence similarity is normally necessary to establish evolutionary relationships, bioinformaticians who work with structures have developed techniques to suggest ancestry between proteins^{102,108-111}. These techniques depend on the deduction that folds are maintained throughout evolution, whereas the frequency of change in sequence is much more rapid. The key that determines whether structures are homologous versus the result of convergent evolution is the topology of the fold, and more specifically, how the secondary

structure elements appear in sequence. It is in using these techniques that we are hypothesize that NleC is a member of the Zincin fold superfamily. While the sequence similarity of NleC to other Zincins is negligible, the maintenance of the fold, activity, and motif shared by other Zincins structure suggests that NleC is evolutionarily related (Figure III-7). However, this hypothesis is impossible to test without sequences linking the evolution of NleC directly from another Zincin family.

Zincins are divided into three major groups depending on what residue coordinates the zinc downstream of the conserved HExxH motif (Figure III-11). NleC is a member of the Aspzincin subfamily. The coordinating aspartate in the NleC family occurs in a motif unique among Aspzincins, LIHExxHxxTxxxD + Y and contains an active site tyrosine, Tyr227, which occurs C-terminally to the other ligating residues. This tyrosine is unique because the structure shows it directly coordinating the zinc. Astacin zincins (MEROPS Family M12), similarly have a Zn²⁺-coordinating tyrosine but are in the Metzincin subfamily (Figure III-8B). Numerous other families have tyrosine residues in a similar location in the active site, though these are not observed coordinating the zinc directly. One of these families, deuterolysins and peptidyl-Lys endopeptidases (MEROPS Family M35), has multiple structures where the tyrosine is present in two conformations, a mechanism called the “tyrosine switch” (Figure III-11B). The tyrosine flips to accommodate substrate binding and acts as a proton donor in catalysis^{94,112}. NleC Tyr227 does not interact with any other residues, and so may be available for a similar action, though in this case, another ligand would have to replace tyrosine in the zinc coordination sphere. NleC is penta-coordinated in a trigonal bipyramidal geometry, as are some other Zincins, although most are in

tetrahedral geometry. Penta-coordination is only present in 31% of zinc-containing enzymes, whereas a bidentate-coordinated carboxylate is rarer, at 17%^{99,113}.

From the solution of the thermolysin structure in 1988 by Matthews *et al.*, researchers describe Zincins as separated into an N- and C-terminal subdomain by the active site cleft (Figure III-12)^{95,96,114-116}. In NleC the C-terminus wraps around the N-terminal portion. Upon review of many Zincin structures, however, this subdomain separation seems tenuous. Though there are Zincins that are structured in a strictly sequential manner, such as thermolysin, many Zincins have structures with mixed elements from both the amino and carboxy-terminus. Additionally, most Zincins have significant structure backing the active site helix, making the distinction between the subdomains questionable. Because thermolysin is a well-known structure that was solved early in the exploration of zinc proteases, the nomenclature developed to describe it persists despite waning evidence. It would be more accurate to say that the catalytic core of Zincin proteases is constant in size and orientation throughout known structures and most essential elements are N-terminal to the catalytic helix and Zn²⁺-coordinating residues. However, carboxy-terminal elements may also be necessary for catalysis for any given Zincin protease.

Within the NleC family, the complete conservation around the active site and the high conservation in the interior of the protein suggest that all family members share conserved folds (Figure III-6). With many of the internal residues conserved between NleC family members, internal contacts will be maintained, making it highly likely that the overall fold is very similar. With the recent finding that NleC family member Aip56 from *Photobacterium damsela* also cleaves p65, it is additionally expected that all NleC

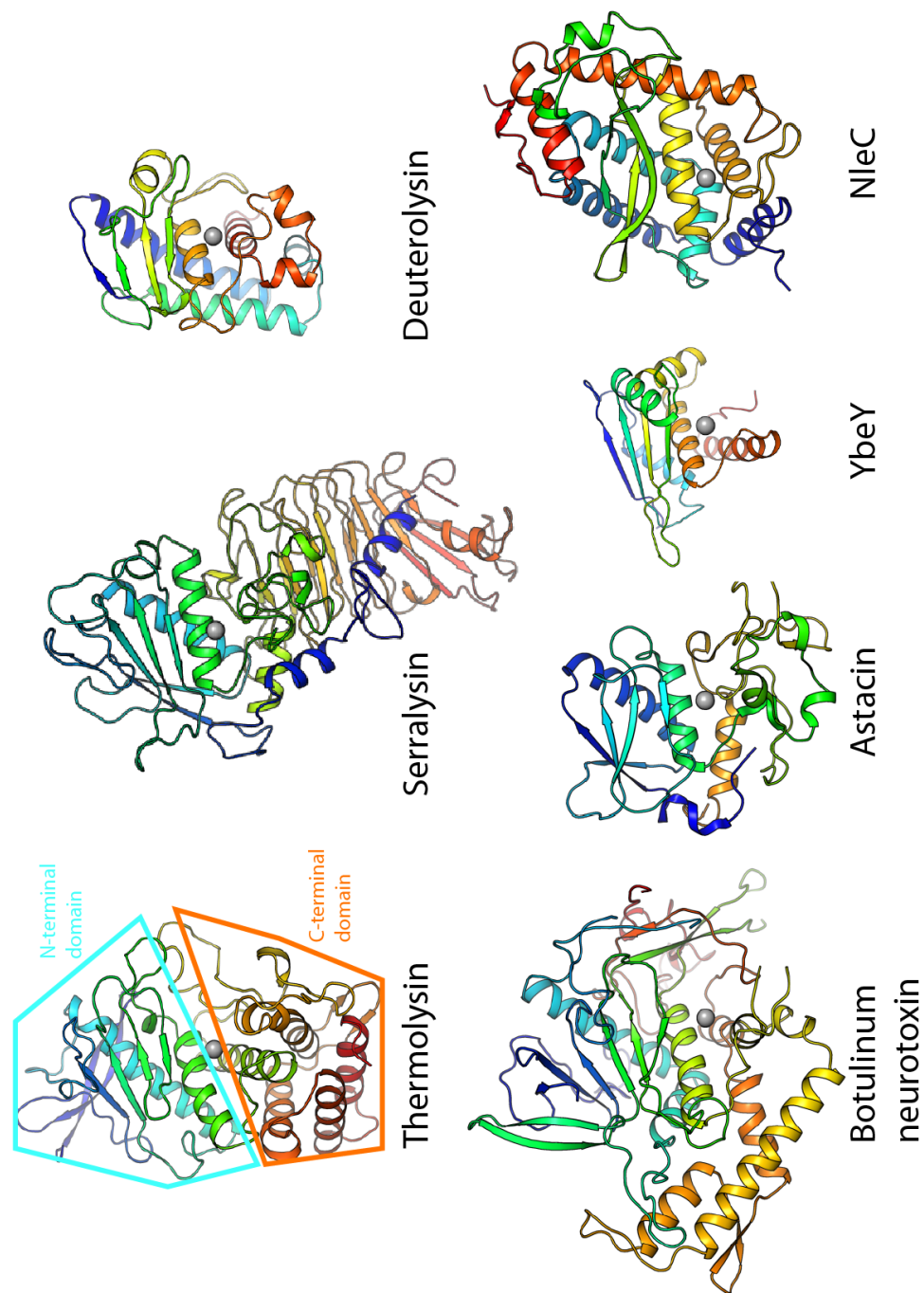


Figure III-12 Comparison of Zincin topology. Various Zincins are shown here with the N-terminus in blue and the C-terminus in red. Conventional subdomain separation is shown for thermolysin with aqua and orange lines. Though it is conventional to describe thermolysin in terms of two subdomains, in many other Zincins, it is difficult to separate the structures into sub-domains.

proteins share specificity for NF κ B subunits; this is despite the lack of sequence similarity among surface residues between the mammalian, insect and fish NleC proteases¹¹⁷. Though there are large insertions to the loops present in some fish and insect NleC proteins, this would be unlikely to affect the overall fold as they are at the periphery of the protein and may be attributed to a differing delivery system beyond the T3SS and differing host environments (Figure III-17). To infect host fish cells, *Photobacterium* does not require direct contact to deliver the NleC homologue Aip56, although multiple copies of the T3SS machinery appear in its genome¹¹⁸. While the N-terminal domain of Aip56 is the NleC zinc protease domain, the C-terminal domain facilitates entry into the host cytosol¹¹⁷. These observations suggest that Aip56 is an exotoxin and not directly translocated like *E. coli* NleC. Insect pathogen *Arsenophonus* has two T3SSs in the genome but the function and delivery of NleC during pathogenesis has not been studied in depth¹¹⁹.

Structural Considerations for Function

One of the most important aspects about discovering a new Zincin protease family member is learning how it compares to presently known Zincins in substrate recognition and specificity. Even with the prevalence of Zincins, ancient and distributed across all domains of life, they maintain a similar mode of binding. Zincins minimally recognized one residue with a large S1' pocket, recognizing the P1' residue C-terminal to the scissile bond in all cases studied. Beyond this, the family varies in its specificity, acting as a very general peptidase, such as tricorn aminopeptidase, to highly specific, like botulism neurotoxin⁹⁶. With this knowledge, I sought to understand how NleC recognizes the substrates that had been identified, the NF κ B transcription factors, and whether there were other substrates of NleC that would be important for pathogenesis.

The structure of NleC provides a new framework in which to discuss the NFκB proteolytic function. One important consideration is whether the X-ray structure described here contains all elements necessary for recognition and catalysis of substrates. As discussed above, a fragment approximating the catalytic core, NleC₁₋₂₈₇ was able to cleave RelA with near wild-type efficiency. Additionally, Mühlen *et al.* showed an NleC fragment, residues 1-266, was able to abrogate RelA activation in cells as completely as wildtype. They also tested a smaller fragment, residues 1-237, which was not able to impede RelA signaling. This truncation would remove a portion of a central helix and would likely destabilize the structure of NleC, destroying its activity, though there are no catalytic elements contained within the truncated portion (Figure III-13)⁷².

The surface features of an enzyme are the main source of information for how it might bind substrate. Electrostatic surface potential analysis of NleC, calculated with APBS in Pymol, reveals that the face containing the active site of NleC is extremely negative, which may provide a mode of binding (Figure III-14)¹²⁰. The surface of NleC can also be analyzed for pockets that may be involved in binding. Like mentioned above, thermolysin and other Zincins are known to recognize the P1' position of the substrate, the residue C-terminal to the scissile bond, with a large S1' pocket^{96,114,116,121}. This obvious subsite is easily recognized on the surface of NleC as well and with strong probability binds the conserved glutamate of the DNA-binding loop, C|EGRSA (Figure III-14). Models of how NleC binds RelA may help determine where other residues of RelA bind NleC.

From the SAXS data, the C-terminal portion of NleC that was not visible in the crystal structure is likely folded into a distinct domain from the catalytic domain. Because the crystallizable core, NleC₁₇₋₂₈₀ was necessary and sufficient for cleavage of RelA, the missing

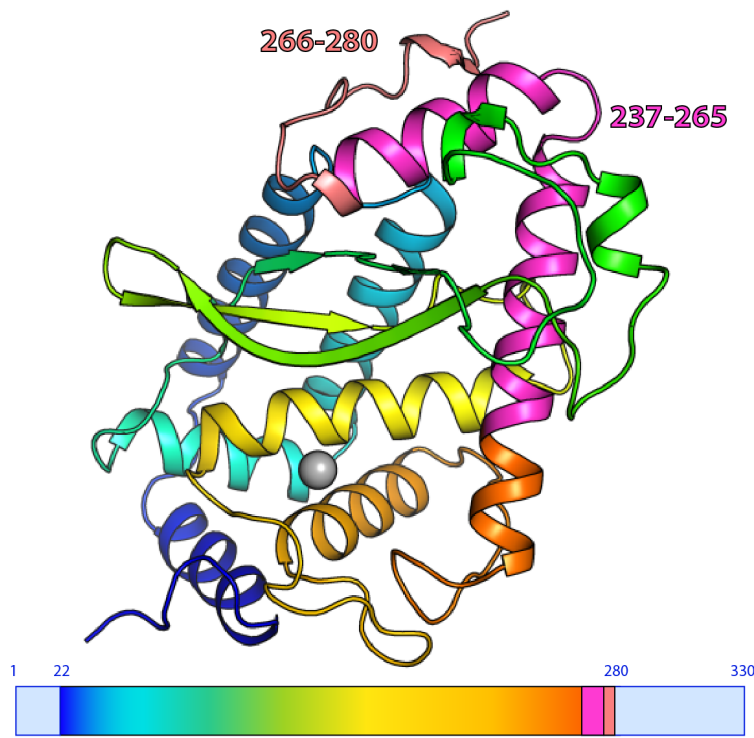


Figure III-13 Structural basis for functional and non-functional truncations of NleC. Mühlen *et al.* truncated NleC in two places C-terminally and suggested the shorter of these, NleC₁₋₂₃₆ was non-functional because a catalytically-important region had been ablated. This is in comparison to another truncation they tested, NleC₁₋₂₆₅, which retained its ability to cleave NFκB. Upon analysis of the structure, however, it is apparent that truncating the region between residues 237 and 265 did not affect the active site directly, but likely destabilized the whole protein, rendering it incapable of catalysis.

carboxy-terminal 50 residues could be extraneous, or alternatively, have a distinct function from the majority of the protein. Because effector proteins are highly evolved and are likely minimized for the necessity of being translocated for pathogenesis, I hypothesized that the C-terminal portion of the protein had a function outside of proteolysis, instead of being expendable. Two possible functional hypotheses are that the C-terminal domain binds an

unknown chaperone and is necessary for efficient secretion, or that it is used during pathogenesis to target NleC to a certain area of the cell.

In Chapter IV, I discuss further implications the structure has for the function of NleC in relation to pathogenesis and binding. Furthermore, I discuss experiments we are pursuing to explore these hypotheses.

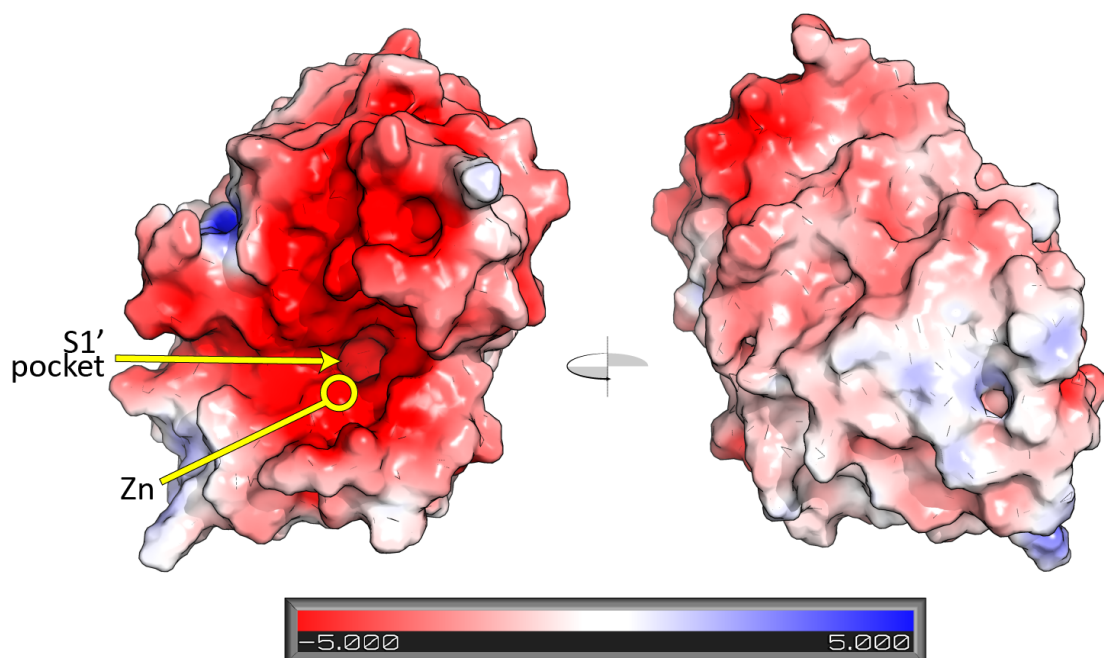


Figure III-14 Electrostatic surface potential of NleC. The face of NleC is highly electronegative, with little positive electrostatic character. The active site Zinc and S1' pocket are indicated. Electrostatic surface potential calculated with the APBS module in Pymol, where the coloring for the solvent accessible surface is shown on the van der Waals calculated surface.

Structural Considerations for Translocation

The structure of NleC contains the core elements of the Zincin scaffold where a three-helix bundle and a mixed β -sheet surround the active site. However, several variations from the canonical fold may allow NleC to accommodate its unique substrate

specificity and maintain competence for secretion. The highly negative surface character of NleC may facilitate translocation, as suggested by Rathinavelan *et al.*¹²²(Figure III-14). Deviating from the most similar protein structures (botulinum toxin and tricorn aminopeptidase), the structure of NleC substitutes two beta strands with the H2 α -helix (Figure III-7). In botulinum neurotoxin, this corresponds to S2 and S3, and in tricorn aminopeptidase, S14 and S15. Although several MA clan families have members with only three β -strands (Families M2, M13, and M48), in NleC the H2 α -helix replaces the missing strands in both sequence and space, suggesting that this change may be functionally beneficial⁸⁹. This modification increases the overall alpha helical content of NleC, which may be advantageous for secretion, as expounded further below.

The three-strand β -sheet of NleC modifies upon the Ψ -loop seen in the lower three strands of other Zincins, S6, S7 and S8 in botulinum neurotoxin, and S16, S17, and S18 in tricorn aminopeptidase (Figure III-7). Ψ -loop motifs are rare and believed to be difficult to fold because the interior strand must be inserted between the two outer strands, which are N-terminal to the middle strand (Figure III-15A)^{123,124} With this modification, the interior β -strand exits the sheet halfway, permitting the two longer exterior strands to interact directly in a two-stranded antiparallel sheet (Figure III-15B). Because the unique pathogenic trajectory of T3SS effectors requires that they are partially unfolded for translocation and refolded upon entry to the cell, I hypothesize that the modification to the NleC Ψ -loop, which zippers the two exterior strands together, aids in this function. Analysis of the residues in the NleC β -sheet reveals glutamate and proline residues in each of the exterior strands, which are shown to be rare and contribute to destabilization of β -sheets^{125,126}. This may cause the initial interaction between the exterior strands to be

unstable upon translation or translocation, permitting the more C-terminal interior strand to insert in between the two N-terminal strands after it emerges. This interplay between stabilizing and destabilizing elements in the NleC β -sheet may be necessary for the unique secretion course.

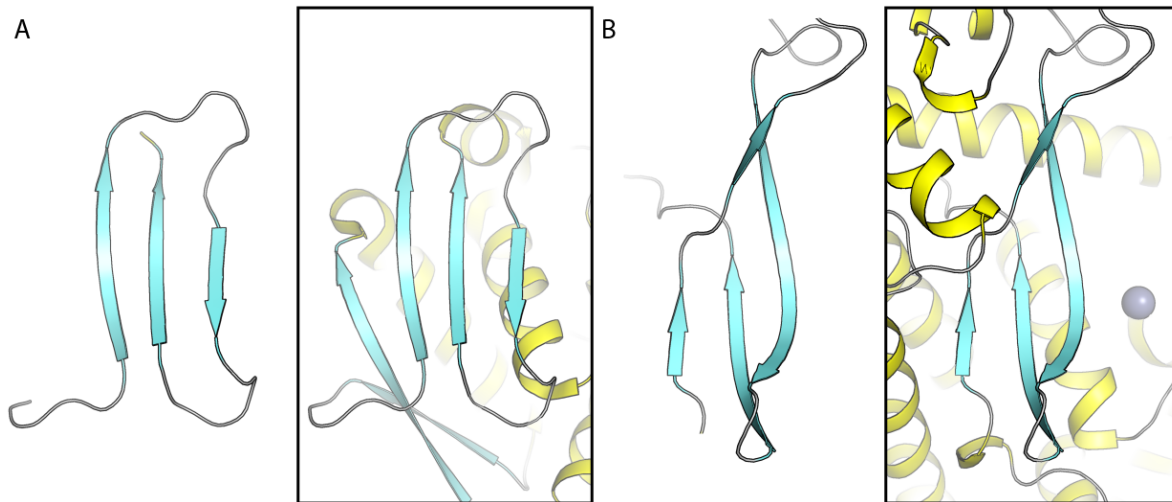


Figure III-15 Comparison of Ψ -Loop and NleC β -sheet. A) A classic Ψ -Loop found in Zincin snapalysin, isolated and within context of the molecule. MEROPS Family M7. PDB: 1C7K. B) The modified Ψ -Loop found in NleC, where the interior strand that does not maintain contact with the two outer strands, but rather exists the sheet halfway, allowing the two outer strands to interact directly.

The unique structure of NleC with a combination of a highly α -helical fold and the unique β -sheet may provide the balance necessary in stability for the unfolding and refolding function. The prevalence of a highly alpha-helical fold is common among effectors, many of which lack or only have short β -strands present^{127,128}. Likewise, the T3SS machinery components that must be translocated to build the needle are highly alpha helical, predominantly containing helix-turn-helix motifs¹²². This trend lends the hypothesis that a highly alpha-helical structure may allow for easier translocation through

the 25 Å T3SS needle interior, and the evolution of the altered Ψ -loop motif in NleC was necessary for efficient secretion of refolding upon translocation. Additionally, the highly electronegative surface potential of NleC may assist in translocation, as the interior of the T3SS needle is electronegative as well and is predicted to propel effectors towards secretion utilizing repulsive electrostatics¹²².

Can NleC be translocated in complex with the active site Zn^{2+} ?

The variability presented by the Irving-Williams series allows biology to utilize the metals for different functions and is the basis for why metals are highly regulated in cells¹⁰⁵. For instance, zinc is very rarely found as a free ion because of its high affinity for various ligands, and so procuring zinc from the environment to use in various cellular functions is an essential process, with growth-limiting potential¹²⁹.

In mammalian cells, zinc homeostasis is tightly regulated. Bacteria which form dependent relationships with hosts have strategies to obtain zinc in this difficult environment. *E. coli* has two zinc transporters, one with lower affinity when zinc is more widely available and a high-affinity transporter for use in environments when zinc is scarce^{130,131}. In pathogens especially, the ability to acquire zinc can directly affect success in host invasion¹³⁰. Vertebrates in turn are found to sequester zinc upon infection, a innate immune system mechanism that is still being researched¹³².

With this knowledge, it is apparent that it would be advantageous for NleC to be translocated with the active site zinc in place. Many of the structural elements in NleC may be evolved to reduce disruption of the coordination sphere for this to occur. By measuring the distance between those structural elements that maintain the coordination sphere, I can assess whether translocation with the Zn^{2+} ion in site is possible. Knowing that the

interior diameter of the EHEC T3SS needle is $\sim 25\text{\AA}$, I measured the width of helix bundles in the structure of NleC (Figure III-16). The Zn^{2+} -coordinating residues are contained in the catalytic helix, the loop following this helix, and the loop following the next helix. The catalytic helix with the following helix and loop would have to translocate together to maintain the integrity of the zinc coordination sphere. Measurements made in Pymol between the most distal side chains in these helices show that this is possible, but only if dynamic remodeling of the loops occurs during translocation. Possible experiments to test this hypothesis are discussed in Chapter IV.

An additional observation from these measurements is the spiraling orientation that these helices adopt (Figure III-16B). Molecular dynamics analysis of the interior of the T3SS needle reveal that helix-turn-helix substrates move through the needle in a spiraling motion. This is facilitated by a tryptophan groove that is aligned in a right-handed spiral in the interior of the needle¹²². In NleC, the placement of the structural elements around the Zn-coordination sphere also appears in a right-handed helix orientation and may expedite this secretion mechanism.

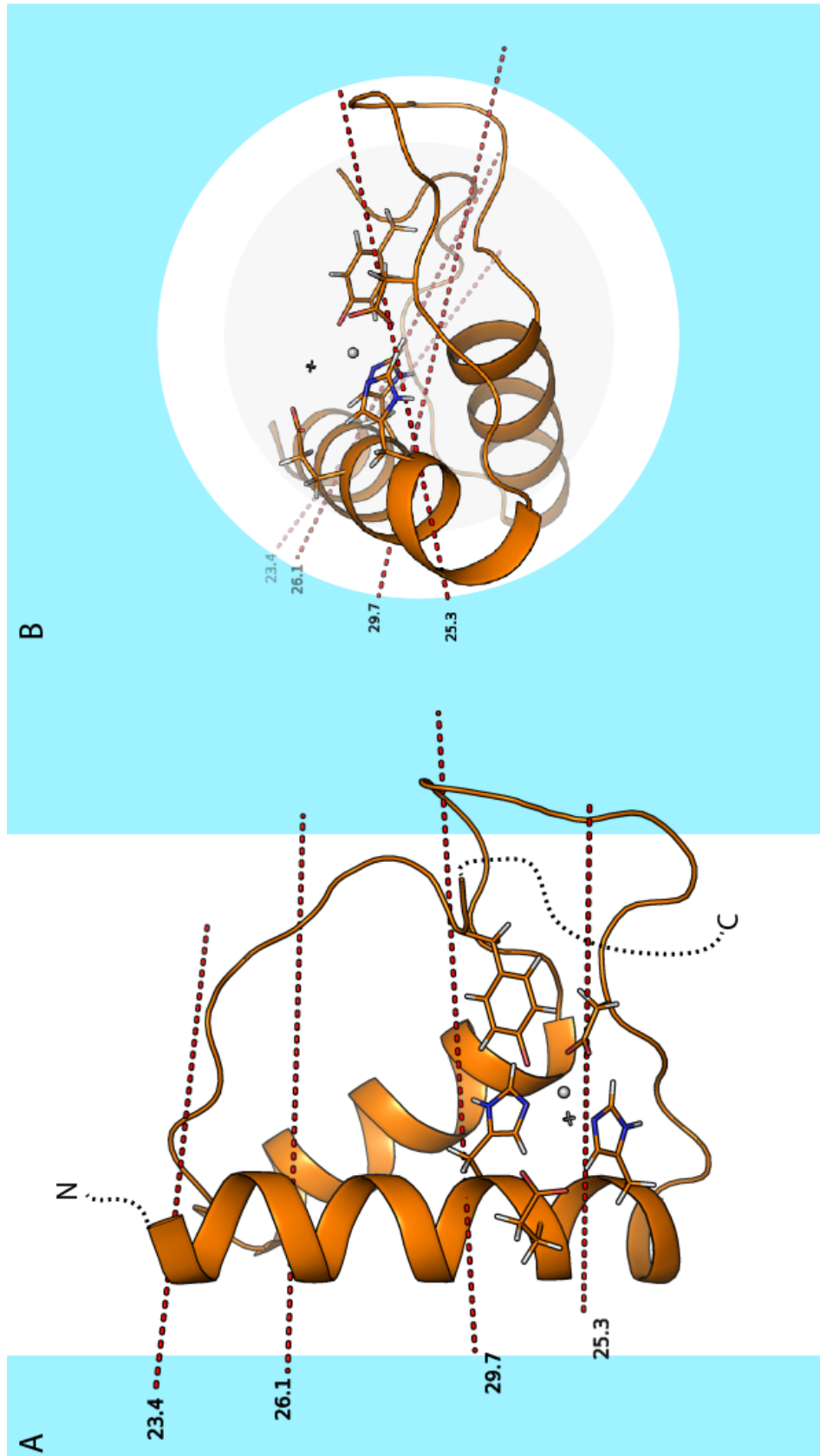


Figure III-16 Maintenance of Zn^{2+} coordination sphere during translocation. A) Measurements across residues in catalytic helix, following helix, and loops (residues not visualized explicitly). In order to be translocated with all Zn^{2+} -coordinating residues intact, the catalytic helix, the backing helix, and the loops containing the tyrosine would have to be translocated through the T3SS together. This would only be possible if the loops containing the coordinating aspartate and tyrosine are dynamic, as the interior diameter of the *E. coli* T3SS needle is estimated to be 25Å. B) Head-on view of catalytic helix, backing helix, and loops. From this angle, the spiraling

Materials and Methods

Protein Cloning, Expression, and Purification

NleC for crystallography trials was cloned into pTYB2 by PCR amplifying the *nleC* gene from *E. coli* str. Sakai with NdeI and XmaI-encoding primers. The parent plasmid and NleC insert were digested with NdeI and XmaI restriction enzymes before ligation. The resultant plasmid is pMS 692. Truncations of NleC were cloned by PCR amplifying the target sequence out of pMS692 with primers encoding NdeI and XmaI sites. The inactive mutant NleC_{E184A} was cloned from pMS692 with primers encoding for the mutation. Plasmids used in this work are listed in Table III-2.

Table III-2 Plasmids prepared for the study of the structure of NleC

Plasmid	Gene	Residues	Organism	Gene Source	Tag	Background	Primers	Resistance
pMS513*	NleC	1-330	EHEC str. Sakai	EHEC str. Sakai DNA	6xHis	pET24d	ECs0847_F1/ECs0847_R1	Kan
pMS692	NleC	1-330	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	ECs0847_F1/ECs0847_R2	Amp
pMS847	NleC E184A	1-330	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	U743/L743	Amp
pMS693	NleC	8-307	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	U637/L637	Amp
pMS934	NleC	19-287	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	U755/L637d	Amp
pMS850	NleC	19-302	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	U755/L637a	Amp
pMS849	NleC	19-307	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	U755/L637	Amp
pMS846	NleC	19-330	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	U755/ECs0847_R2	Amp
pMS935	NleC E184A	19-287	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	U743/L743	Amp
pMS1023**	NleC E184A	22-279	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U932/L932/L933	Kan
pMS1024**	NleC E184A	22-279	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	U934/L934	Kan
pMS1023**	NleC E184A	22-270	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U932/L936/L937	Kan
pMS1075**	NleC E184A	279-330	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U1061/L1062ab	Kan
pMS1076**	NleC E184A	290-330	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U1060/L1062ab	Kan

* Prepared by Amy Dear

** Prepared by Sandra Metzner

Plasmids was transformed and expressed in *E. coli* Rosetta (DE3) cells in LB broth with 50 mg/mL ampicillin. From a freshly transformed plate, one colony was used to inoculate 1L of LB, for up to 6L per expression. After cultures reach an OD₆₀₀ of 0.6, the cultures were placed in ice baths for 10-30 minutes and then induced with 1mM IPTG. They then grew overnight at 20C for 16-19 hours. The cells were harvested by centrifugation at 5000 x g for 15 minutes at 4C. Then the cells were resuspended with 7.5mL per 1L culture

in Buffer A [25 mM Tris pH 8.0 and 150 mM NaCl] with 1 tablet protease inhibitor cocktail (Roche Complete EDTA-free) per 50 mL lysis buffer. This protease inhibitor cocktail does not inhibit zinc proteases. At this point, the resuspended cells could be frozen for later use if necessary. When ready to purify, the cells were lysed by sonication and the cell debris separated from the soluble fraction by centrifugation at 20,000 x g for 20 minutes at 4°C. The soluble fraction was applied to a chitin column, with 5 mL of beads per liter of culture grown. Alternatively, the chitin purification can be repeated with the same flow-through and a smaller quantity of beads after the first fraction is eluted off and the beads are washed with 0.3M NaOH and water. After incubation of the soluble fraction for 30 minutes on the column, the column was washed with ten times the column volume with Buffer B [25 mM Tris pH 8.0 and 500 mM NaCl]. The column was then flushed with one column volume of Buffer C [Buffer B + 10 mM DL-dithiothreitol] and incubated overnight at 4°C. The column was then eluted and the fractions concentrated for size exclusion chromatography (HiLoad 26/60 Superdex 75; Amersham Pharmacia Biotech). The column is preequilibrated in Buffer D [Buffer A + 5 mM BME]. The NleC-containing fractions were then pooled and used as is or concentrated. Another portion was dialyzed into Buffer E [20 mM MOPS pH 7 and 20 mM NaCl] before concentration. The expression and purification procedures were the same for all NleC variants. For crystallization trials, NleC can be concentrated up to 30 mg/mL, but most trials were done at 15 mg/mL maximum concentration. Purified NleC can be stored at 4°C or -80°C, however degradation does occur over time. A 6xHis-tagged NleC fusion was also produced for this project, but it was more difficult to obtain a clean purification. The Intein-tagged NleC expresses very well, is soluble, and results in a very clean purification off of the chitin column.

Protein Crystallization

Full-length NleC crystals were grown with the sitting-drop vapor diffusion method (protein:precipitant, 0.2:0.2 μL) after eight month incubation. The precipitant was 20% PEG-3350 and 0.2M $\text{Mg}(\text{CHO}_2)_2$ at 2.5 mg/mL final protein concentration. The crystals were transferred to 35% PEG-3350 and 0.2M $\text{Mg}(\text{CHO}_2)_2$ for 1 min before flash-freezing in liquid N_2 . An N-terminal truncation of NleC, NleC-SAP (residues 17-330), gave the best crystals in similar conditions to the full-length NleC using the same method, in 10% PEG-3350, 50 mM $\text{Mg}(\text{CHO}_2)_2$ and 0.1 M MES pH 6.0 at 6 mg/mL final protein concentration.

Structure Determination

A fluorescence scan for Zn^{2+} confirmed the presence of zinc in the native crystal at the Advanced Light Source at the Lawrence Berkeley National Laboratory. A dataset to 1.8- $\text{\AA}$ resolution was collected at the peak 1.2831 $\text{\AA}$, at the inflection point 1.2835 $\text{\AA}$, and at a remote wavelength 1.2574 $\text{\AA}$ for zinc. After indexing and scaling the data according to the space group to identify peaks with HKL-2000, the differences in the peaks intensities between the three datasets was measured with AutoSol in PHENIX. AutoSol requires this scaled data, a sequence file containing the protein sequence of interest, the identity of the anomalous scatterer and how many are expected per unit cell, and which wavelengths the data was collected at their derivatives. Autosol then identified where the anomalously scattering zinc was located in the unit cell and phased the data. With the phasing dataset, AutoSol modified the electron density map and, with the user-input protein sequence, built an initial model of NleC. After this, AutoBuild from the PHENIX suite was launched to rebuild and improve upon the initial model. This program used the experimental structure factor amplitudes from the remote wavelength and the experimental phases determined by

AutoSol. AutoBuild then iteratively used the model as phase data to recalculate the electron density map, and subsequently refined the model using this new map. With this AutoBuild-generated model, I manually rebuilt missing portions of NleC that were obvious from the density. The only major refinement of the AutoBuild structure was a helix that was clearly visible in the density but not modeled. Manual refining of the NleC model with Coot was alternated with refinement using Refine on PHENIX until R_{free} and R_{work} could no longer be improved. R_{work} describes how the structure factor amplitudes that were used from model building fit in the model, whereas R_{free} describes how a control set of structure factor amplitudes that was not used to build the model fits into the built model. The final NleC model has residues 22-280 modeled and refined, and the statistics are shown in Table III-1. A hexahydrated Mg^{2+} ion is also in the structure between two unit cells.

A dataset to 2.6 Å resolution was collected for the crystal grown with full-length NleC. This structure was solved by molecular replacement with the NleC-SAP structure described above using PHENIX AutoMR. This structure has residues 22-280 as well, and is identical.

Figure III-17 Conservation within the NleC M85 protease family. N termini truncations: lcl|24: Residues 1-550. First panel shows which sequences were aligned using the BLAST server.

lcl	FASTA Header	Max score	E-value
1	>ref NP_286533.1 hypothetical protein Z0986 [Escherichia coli O157:H7 str. EDL933]	684	0
2	>ref YP_002328603.1 T3SS secreted effector NleC homolog [Escherichia coli O127:H6 str. E2348/69]	682	0
3	>gb EFX07422.1 hypothetical protein ECO5101_22535 [Escherichia coli O157:H7 str. G5101]	668	0
4	>gb EFX27229.1 T3SS secreted effector NleC-like protein [Escherichia coli O55:H7 str. USDA 5905]	657	0
5	>gb EFZ76165.1 non-LEE encoded type III effector C [Escherichia coli RN587/1]	649	0
6	>gb EFZ40131.1 non-LEE encoded type III effector C [Escherichia coli EPECa14]	642	0
7	>ref YP_003229213.1 T3SS effector NleC [Escherichia coli O26:H11 str. 11368]	641	0
8	>ref ZP_03043710.1 non-LEE encoded type III effector C [Escherichia coli E22]	641	0
9	>ref YP_003234807.1 T3SS secreted effector NleC-like protein [Escherichia coli O111:H- str. 11128]	640	0
10	>ref YP_003234967.1 T3SS secreted effector NleC-like protein [Escherichia coli O111:H- str. 11128]	629	0
11	>ref YP_003365223.1 T3SS effector protein NleC [Citrobacter rodentium ICC168]	626	0
12	>gb EFX31295.1 Non-LEE encoded type III effector C [Escherichia coli O157:H7 str. LSU-61]	601	0
13	>gb EGX10984.1 non-LEE encoded type III effector C [Escherichia coli STEC_H.1.8]	583	0
14	>ref ZP_04620510.1 Non-LEE encoded type III effector C [Yersinia aldovae ATCC 35236]	579	0
15	>ref ZP_02576318.1 non-LEE encoded type III effector C [Salmonella enterica subsp. enterica serovar 4,[5],12:i:- str. CVM23701]	473	1.E-166
16	>gb EFX16868.1 Non-LEE encoded type III effector C [Escherichia coli O157:H- str. H 2687]	432	3.E-152
17	>dbj BAF99004.1 apoptosis inducing protein [Photobacterium damsela subsp. piscicida]	182	1.E-50
18	>ref ZP_02194626.1 hypothetical protein 1103602000593_AND4_00648 [Vibrio sp. AND4]	181	2.E-50
19	>ref YP_003422532.1 Aip56 [Photobacterium damsela subsp. piscicida]	177	7.E-49
20	>ref ZP_07742130.1 Aip56 [Vibrio caribbenthicus ATCC BAA-2122]	175	7.E-48
21	>emb CBA72068.1 non-LEE encoded type III effector C [Arsenophonus nasoniae]	169	8.E-46
22	>emb CBA76058.1 non-LEE encoded type III effector C [Arsenophonus nasoniae]	131	9.E-32
23	>ref ZP_08731022.1 Aip56 [Vibrio nigrapulchritudo ATCC 27043]	127	6.E-30
24	>emb CBA72300.1 apoptosis inducing protein [Arsenophonus nasoniae]	109	6.E-24
25	>emb CBA74519.1 non-LEE encoded type III effector C [Arsenophonus nasoniae]	94.4	1.E-18
26	>gi 15830101 ref NP_308874.1 hypothetical protein ECs0847 [Escherichia coli O157:H7 str. Sakai] *	*	*

1

lcl 23	-----M Q N N Q D Q V N Q-----	-	-
lcl 11	-----	-	M K I
lcl 19	-----M I K K V L V I S-----L F L A M N T K F V L A N N-----	-	-
lcl 4	-----	-	M K I
lcl 9	-----	-	M K I
lcl 24	P N G E Y A R F D G V R H K F V R V M H Q G E V F I F P A R E I S V A G R P K L E M N N N P G Y E F	-	-
lcl 10	-----	-	M K I
lcl 25	-----M I H N Q P E V K Q L Q A N Q L E I I R N L I S Q G I I Y Y P Y P Q F E E E H R P V Y P I	-	-
lcl 22	-----M K R K N-----M R L N Q Q Q P N P A-----	-	-
lcl 13	-----	-	M K I
lcl 1	-----	-	M K I
lcl 3	-----	-	M K I
lcl 5	-----	-	M K I
lcl 18	-----M T A I F S L A I N S N F V L A N N-----	-	-
lcl 8	-----	-	M K I
lcl 15	-----	-	M K I
lcl 14	-----	-	-
lcl 7	-----	-	M K I
lcl 16	-----	-	-
lcl 2	-----	-	M K I
lcl 20	-----M T A I F S L A I N S N F V L A N N-----	-	-
lcl 21	-----M-----	-	-
lcl 12	-----M Y K I I L G I N-----	-	M K I
lcl 26	-----	-	M K I
lcl 17	-----M I G K S L S V A I L L V L A M N T K F V L A N N-----	-	-
lcl 6	-----	-	M K I

1c1 23	-- -- -- -- -- -- -- -- -- -- -- NEIFR	KMD	DYAEYVVTNG	GARA	--	ELSSRQIE		
1c1 11	PVLQPSFNF	FAPAGYSA	AAVAPNR	ADNAYADYVLD	IGKQT	--	PLSAADLSN	
1c1 19	-- -- -- -- -- -- -- -- -- -- -- GAFDR	DS	DEYADYVVRNG	GANG	--	QLTYAQMVA		
1c1 4	PVLQPGFNF	FAPAGYSA	AAVAPNR	AENAYADYVLD	IGKRI	--	PLSAADLSN	
1c1 9	PVLQPGFNF	FAPAGYSA	AAVAPNR	AENAYADYVLD	IGKRI	--	PLSAADLS	
1c1 24	FPLDIEGEFF	GITRHGL	XGGMNK	FLDSYVEHVT	AMGSQD	--	ALDYPSRVQ	
1c1 10	PVLQSGFNF	FAPAGYSA	AAVAPNR	ADNAYADYVLD	IGKQI	--	PLSAADLSN	
1c1 25	QGLDLQLEGI	QLDLA	IRALNSL	FAQSNQYAEFVCTE	GQRN	IDQ	IDDNFISN	
1c1 22	-- -- -- -- -- -- -- -- -- -- -- HETFR	RS	DGYAAHVIT	IGERN	--	RLSEAQLYD		
1c1 13	PSLQSNFNF	SAPAGYS	APIAPNR	AENAYADYVLD	IGKRI	--	PLSAADLSN	
1c1 1	PSLQSNFNF	SAPAGYS	APIAPNR	AENAYADYVLD	IGKRI	--	PLSAADLSN	
1c1 3	PVLQPGFNF	FAPAGYSA	AAVAPNR	AENAYADYVLD	IGKRI	--	PLSAADLSN	
1c1 5	SVLQPGFNF	FAPAGYSA	AAVAPNR	AENAYADYVLD	IGKRI	--	PLSAADLS	
1c1 18	-- -- -- -- -- -- -- -- -- -- -- DKPDA	SD	KYADYVVR	LGSEH	--	PLNHTQIE		
1c1 8	PSLQSNFNF	SAPAGYS	AAVAPNR	AENAYADYVLD	IGKRI	--	PLSAADLSN	
1c1 15	SSSTPC	LN	FAPQKEYS	AAVVPHP	SKNAYADYVLE	TGKRI	--	PFSAADLSN
1c1 14	-- -- -- -- -- -- -- -- -- -- -- MD	IGKRI	--	PLSAADLSN				
1c1 7	PSLQSNFNF	FAPAGYSA	AAVAPNR	AENAYADYVLD	IGKRI	--	PLSAADLSN	
1c1 16	-- -- -- -- -- -- -- -- -- -- --							
1c1 2	PSLQSNFNF	SAPAGYS	APIAPNR	AENAYADYVLD	IGKRI	--	PLSAADLSN	
1c1 20	-- -- -- -- -- -- -- -- -- -- -- DKPDA	SD	KYADYVVR	LGSEH	--	PLNHTQIE		
1c1 21	-- -- -- -- -- -- -- -- -- -- -- KNR	NED	AYASYVC	GN	HEF	--	SLSDDVLTR	
1c1 12	TSLQPSFNF	SAQEGYS	AAVAPNH	SDNAYADYVLD	IGKRI	--	PFSTADLGN	
1c1 26	PSLQSNFNF	SAPAGYS	APIAPNR	AENAYADYVLD	IGKRI	--	PLSAADLSN	
1c1 17	-- -- -- -- -- -- -- -- -- -- -- DRPDQ	HS	DEYADYVVT	GADQ	--	QLTYSQLVA		
1c1 6	PSLQPSFNF	FAPAGYSA	AAVAPNR	SDNAYADYVLD	IGKRI	--	PLSAEDLG	

lcl 23	MII	DVRRTINDRASFLVDRHTAIAIEDTIEHALLSQOTFRQAVVFSLREE
lcl 11	VYE	NVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNK
lcl 19	LSS	AISRSVSVSYGSIIVDRHTAAAIEYTLIDALFNSPTFQHAVSFGILNND
lcl 4	VYES	VISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 9	VYES	SVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 24	IED	NVSLGIRDSYRRLIDRYTADSIERTVHEALHLSETFQHAISFSKQRN
lcl 10	VYE	NVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNK
lcl 25	YMIQ	IRWAIOLSGAGYIIDEEATAFNIEETIEHAFTSEPTFRLAMAFSIGES
lcl 22	IIV	RVRQSIQSICYFLAEPYTAFADIEDTIRTAIHLSTTFRNALAFSGSRE
lcl 13	VYES	SVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 1	VYES	SVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 3	VYES	SVISRAVDHSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 5	VYES	SVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 18	LSS	AVSRAVLLSYPNIIDRYTAAATEYTVIDLALFHSPTFRRHIVSFGILHNQ
lcl 8	VYES	SVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 15	LYQ	SVIYAVHSRSRLIDQHTANMIGNTVLDALSRSQOTFRDAVIYGIHNK
lcl 14	VYES	SVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 7	VYES	SVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 16	- - -	- MR
lcl 2	VYES	SVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 20	LSS	AVSRAVLLSYPNIIDRYTAAATEYTVIDLALFHSPTFRRHIVSFGILHNQ
lcl 21	LTI	ELRTAVRSYHNLLDRHTAYAIEDTIREALLMSQAQFQRIIDFSIREDD
lcl 12	LYEN	NVISRAVHDSLRLIDQHTADMIGNTVLDALSRSQOTFREAVSYGIHNH
lcl 26	VYES	SVISRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQOTFRDAVSYGIHNE
lcl 17	LSS	AISRSVSVSYGSIADRHTSAEIEYTLIDALFNSPTFQHSVAFGLHNH
lcl 6	LYEN	NVISRAVRDHSRKLIHQHTVDMIGNTILDALSRSQOTFRDAVSYGIHNN

4

lc1 23	RNNLNFSHNSVLPN	LYYMN	LYELR	-	-	D	-	R	H	R	F	Q	D	I	R	F	E	E	I	T	R	-	S	T	V	D	T
lc1 11	EVYIGS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 19	IDNVGN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 4	EVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 9	EVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 24	YDPLES	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 10	EVYIGS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 25	RSNLFNFQNPVLSE	VRYRN	YYLRNP	NSNVRL	ENLAAYD	LD	TAL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 22	RRDAGVLRSDITR	LHYRN	LYVMRQ	XTD	-	R	-	R	R	L	D	Q	L	S	L	T	E	I	L	T	-	-	-	-	-	-	-
lc1 13	KVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 1	KVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 3	EVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 5	EVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 18	QENLGH	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 8	KVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 15	EVQLGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 14	EVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 7	EVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 16	RYYTLVA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 2	KVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 20	QENLGH	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 21	NEQLGF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 12	EVHLGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 26	KVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 17	RENVGQ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
lc1 6	EVHIGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

5

lc1 23	VPI	I	G	I	S	S	N	S	N	L	T	D	N	I	Y	Q	D	-	-	-	I	I	T	I	S	V	A	P	N	Y	L	S	E	H	Y	P	I	W	Q	Y	T	L	I	H	E	I	I	H						
lc1 11	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	E	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 19	VPI	F	A	-	-	-	-	-	-	-	-	N	H	E	A	K	E	D	N	N	-	G	T	P	I	V	N	V	G	A	A	P	H	I	N	T	S	E	Y	F	W	W	Q	E	S	L	I	H	E	V	I	H		
lc1 4	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	D	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 9	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	E	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 24	SPI	Y	V	H	M	-	-	-	-	-	-	Q	F	A	E	T	P	D	-	G	T	P	L	I	K	V	G	A	A	P	A	P	D	T	P	R	Y	E	H	W	Q	L	G	L	I	H	E	I	M	H				
lc1 10	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	D	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 25	PV	L	S	N	I	Y	F	S	L	P	E	G	N	I	N	H	-	-	-	-	P	Y	I	A	I	P	V	V	P	N	P	E	S	Y	F	Y	P	L	W	Q	A	C	L	I	H	E	I	I	H					
lc1 22	VPI	V	-	-	-	-	-	-	-	-	-	P	E	E	I	L	T	F	E	D	A	R	D	P	D	L	F	V	S	I	S	I	S	P	N	H	N	S	V	H	Y	P	F	W	Q	T	G	L	I	H	E	I	I	H
lc1 13	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	D	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 1	EPI	F	P	I	C	-	-	-	-	-	-	E	A	G	E	N	D	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 3	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	E	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 5	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	E	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 18	VPL	I	A	-	-	-	-	-	-	-	-	F	N	E	A	R	E	D	R	A	-	G	T	P	I	V	N	M	G	V	A	P	S	L	F	S	G	R	Y	S	W	Q	E	A	L	I	H	E	I	V	H			
lc1 8	EPI	F	P	I	C	-	-	-	-	-	-	E	A	G	E	N	D	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 15	EPI	L	P	I	C	-	-	-	-	-	-	E	A	R	K	D	E	H	-	E	E	A	Y	I	S	F	S	A	A	P	D	T	D	S	C	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 14	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	E	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 7	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	E	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 16	EPI	F	P	I	C	-	-	-	-	-	-	E	A	G	E	N	E	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 2	EPI	F	P	I	C	-	-	-	-	-	-	E	A	G	E	N	D	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 20	VPL	V	A	-	-	-	-	-	-	-	-	F	Y	E	A	R	E	D	R	A	T	G	T	P	I	V	N	M	G	V	A	P	S	L	F	S	G	R	Y	S	W	Q	E	A	L	I	H	E	I	V	H			
lc1 21	MPI	I	H	G	-	-	-	-	-	-	-	E	A	G	E	D	A	N	-	E	H	P	Y	V	N	I	S	P	A	P	S	X	T	S	R	D	Y	N	H	W	Q	S	S	L	I	H	E	L	I	H				
lc1 12	EPV	F	P	I	C	-	-	-	-	-	-	E	A	G	E	D	D	H	-	E	E	P	Y	V	S	F	S	V	A	P	D	A	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 26	EPI	F	P	I	C	-	-	-	-	-	-	E	A	G	E	N	D	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				
lc1 17	VPL	I	A	-	-	-	-	-	-	-	-	S	H	E	A	R	E	D	D	Y	-	G	N	P	V	N	I	G	A	A	P	R	I	D	S	S	E	Y	S	W	Q	E	A	L	I	H	E	V	I	H				
lc1 6	EPI	L	P	I	C	-	-	-	-	-	-	E	A	G	E	N	D	N	-	E	E	P	Y	V	S	F	S	V	A	P	D	T	D	S	Y	E	M	P	S	W	Q	E	G	L	I	H	E	I	I	H				

6

lcl 23	AITNAGDPP	EGEE	--	RIGPTDIL	AYRIAS	EMD---	LR-IQR	-	FNNY	YSL
lcl 11	HVTGASDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 19	QITASQD	DELDRH	--	GPTEIL	AQRIAN	ELN---	WP-IP	T	FKGY	ADP
lcl 4	HVTGASDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 9	HVTGASDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 24	VLTDAKDPS	SASEN	-RL	GPTEIL	ARKVMA	EMG---	KNNIP	D	FANY	FQP
lcl 10	HVTGASDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 25	AITNAMDPS	LDGPR	--	GPTEIL	TQYVVT	EMNVVIR	YPQSR	VM	FNSY	DDP
lcl 22	AITDAGDPP	ENEQTAR	L	GPTEIL	ANQIAR	EIS---	WN-IP	V	FTAY	DSV
lcl 13	HVTGSSDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 1	HVTGSSDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 3	HVTGASDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	ADP
lcl 5	HVTGASDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 18	HVTGSSDTH	ENKQ	--	GPTEIL	AQMVA	ELH---	WT-IP	T	FKGY	SDP
lcl 8	HVTGSSDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 15	HVTGAGDPL	EDGNI	-EP	GPTEIL	ARRIAQ	ELG---	WS-IP	E	FTGY	ASP
lcl 14	HVTGASDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 7	HVTGSSDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 16	HVTGSSDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 2	HVTGSSDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 20	HVTGSSDTH	ENKQ	--	GPTEIL	AQMVA	ELH---	WA-IP	T	FKGY	SDP
lcl 21	HLTGADPE	SNED	--	RLGPTEIL	TQRVAI	DLN---	WH-IP	Q	FRSY	NSP
lcl 12	HVTGASDPS	RDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WP-IP	D	FIGY	ADP
lcl 26	HVTGSSDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WS-VPD	-	FKGY	AEP
lcl 17	HITGSSDTH	EVDRH	--	GPTEIL	AQRIAN	ELH---	WP-IP	I	FKGY	ADP
lcl 6	HVTGASDPS	GDSNI	-EL	GPTEIL	ARRVAQ	ELG---	WT-VPD	-	FIGY	AEP

7

lcl 23	ER	RAIY	EHDF	AC	LC	ETI	YRHHE	H	PTEVIN	R	L	F	A	I	N	F	F	V	E	N	R	P	E	D	L	N	N	H	P																				
lcl 11	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	K	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 19	ER	TRAI	A	E	R	G	F	R	A	L	L	Q	T	I	D	R	H	P	S	E	T	V	A	L	L	E	R	I	A	T	I	S	R	G	L	K	A	S	A	-	S	F	S	L	L	T			
lcl 4	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	E	N	E	R	A	F	F	E	R	L	G	M	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 9	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	K	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 24	ER	E	A	W	I	K	Q	R	D	H	L	A	V	D	L	I	E	R	V	P	D	R	-	A	A	L	M	K	R	I	D	K	V	A	K	Q	R	G	T	A	K	-	-	-	-	-			
lcl 10	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	E	N	E	R	A	F	F	E	R	L	G	M	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 25	DR	E	R	A	A	Q	Q	L	E	W	V	A	L	L	H	C	L	F	R	-	S	E	N	L	S	E	R	V	N	R	L	L	E	I	G	D	R	M	V	A	E	I	-	A	A	Q	P	Y	A
lcl 22	DR	N	E	F	I	S	E	Y	E	F	Q	S	L	R	Q	G	I	Y	R	H	H	Q	R	G	Y	E	L	L	E	R	L	C	D	I	N	H	L	T	S	S	P	-	N	F	A	G	I	D	
lcl 13	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	E	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 1	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	E	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 3	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	K	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 5	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	K	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 18	AR	VE	A	I	Q	E	R	D	F	H	S	L	E	M	F	Q	R	H	G	S	E	L	G	F	L	F	T	R	L	A	T	I	A	K	G	K	K	A	S	P	-	D	F	G	T	L	T		
lcl 8	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	K	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 15	DR	V	A	H	L	R	T	R	N	L	N	A	L	R	Q	T	A	T	R	H	E	D	N	E	E	A	F	F	E	R	L	D	V	I	S	E	G	Y	E	A	S	A	-	D	F	T	E	Y	P
lcl 14	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	K	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 7	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	E	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 16	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	E	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 2	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	E	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 20	AR	VE	A	I	Q	E	R	D	F	H	S	L	N	M	F	Q	R	H	G	S	E	L	G	F	L	F	T	R	L	A	T	I	A	K	G	K	K	A	S	P	-	D	F	G	T	L	T		
lcl 21	ER	L	Q	A	I	R	A	R	N	F	R	S	L	E	S	I	D	R	H	P	R	E	N	E	L	L	E	R	L	I	S	I	S	E	N	V	G	A	S	S	-	D	F	S	S	L	S		
lcl 12	DR	EAHL	R	A	R	N	L	N	A	L	R	Q	A	A	T	R	H	E	S	N	E	G	A	F	F	E	R	L	D	M	I	S	D	K	Y	E	V	S	P	-	D	F	T	E	Y	A			
lcl 26	ER	EAHL	R	L	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	E	N	E	R	A	F	F	E	R	L	G	T	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			
lcl 17	ER	T	Q	A	L	K	E	R	N	F	Q	A	L	L	H	T	I	H	R	P	Y	E	A	E	A	L	N	R	L	A	T	I	S	S	G	L	K	A	S	P	-	R	F	S	L	L	T		
lcl 6	DR	EAHL	R	G	R	N	L	N	A	L	R	Q	A	A	M	R	H	E	D	N	E	R	T	F	F	E	R	L	G	M	I	S	D	R	Y	E	A	S	P	-	D	F	T	E	Y	S			

8

```

lc1|23 P Q Y F R Y L L P D - - - - H L F C D D N D V Q L G S - - A F F M G A S - - - - A S H C K
lc1|11 A V T N I G Y G F I Q Q H D F P G V A I D E N L Q D V N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|19 S F C S S G I I E L P - - - K L P D F D D D D F S M G A - - A F F T G A S - - - - A S N L G
lc1|4 A V P N I G Y G F I Q Q H D F P G L A I N D N L Q D A N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|9 A V S N I G Y G F I Q Q H D F L G V A I D E N L Q D V N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|24 - - - - - L I Q T D Q Y P P A P L L Q P - - - - - - - - - - - - - - - - - - - - - - - - - -
lc1|10 A V S N I G Y G F I Q Q H D F P G L A I N D N L Q D A N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|25 H E P N A E L L S I C P - - - - - F D D D Q E H L R F - - F F F S G A H - - - - - G I S V A
lc1|22 T L S D E C L I P T - - - - - H S S D E D T P A H W G A - - F L L S G A H - - - - - A T H Q R
lc1|13 A V S N I G Y G F I Q Q H D F P W I G Y R R E F T G C K - - S D P T V S W C S L Y L Y I W G C G Q T
lc1|1 A V S N I G Y G F I Q Q H D F P G L A I N D N L Q D A N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|3 A V S N I G Y G F I Q Q H D F P G V A I D E N L Q D V N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|5 A V S N I G Y G F I Q Q H D F P G V A I D E N L Q D V N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|18 S F C S E G I R S F P - - - K Y P D H D D D F N G G G A - - F F L P S A S - - - - - A D S S V
lc1|8 A V S N I G Y G F I Q Q H D F P G V A I D E N L Q D V N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|15 V M S D M V K E L N K P H D F P G L V I N D N T M D A D P D Q I Q L Y H G Q P - Y I F T F V - - - -
lc1|14 A V S N I G Y G F I Q Q H D F P G V A I D E N L Q D V N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|7 A V S N I G Y G F I Q Q H D F P G V A I D E N L Q D V N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|16 A V S N I G Y G F I Q Q H D F P G L A I N E N L Q D A N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|2 A V S N I G Y G F I Q Q H D F P G L A I N D N L Q D A N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|20 S F C S E G I S S F P - - - K Y P D H D D D F N G G G A - - F F L P S A S - - - - - A D S S V
lc1|21 E S S N D S S V S T A A C R F E H G D S D D D F T - - G S - - T F F R G A K - - - - - A A - - -
lc1|12 V V S N I E Y G L I Q Q H D F P G L A I D D N L Q D A N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|26 A V S N I G Y G F I Q Q H D F P G L A I N D N L Q D A N - - Q I Q L Y H G A P - Y I F T F G - - - -
lc1|17 S F C S S G I I D L P - - - K L P D F D D D D F S M G A - - A F F T G A S - - - - - A S S S G
lc1|6 A V S N I E Y G F I Q Q H D F P G L A I D D N L Q D A N - - Q I Q L Y H G A P - Y I F T F G - - - -

```

9

```

lc1|23 N N L F A A K Y E - - - - - N I I F S - E N L P L T Q F H L E K Y N L V I A A M A V N R A E Q G
lc1|11 - - - - - - - - - - - D V D R H N - - - - -
lc1|19 E C S L D A A G R - V E P V S S I T F E G G Q L - L I Q R D L K N L N L L V A K L A F L R A K N N
lc1|4 - - - - - - - - - - - D V D R H N - - - - -
lc1|9 - - - - - - - - - - - D V D R H N - - - - -
lc1|24 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
lc1|10 - - - - - - - - - - - D V N K H N Q R - - - - -
lc1|25 K S F F C I N H S - - - - - D I I F N D G - L K L S N L H L I K Y D L A A A T V A V N L V K N G
lc1|22 K D V F A T K Y K - - - - - N I V F N K E E L P L T Q W H L E K Y G L V I A E K A V N R V E H G
lc1|13 Q S A M I R L C S D I G L L L I D I N G V D D G - - - - -
lc1|1 - - - - - - - - - - - D V D K H N Q R - - - - -
lc1|3 - - - - - - - - - - - D V D R H N - - - - -
lc1|5 - - - - - - - - - - - D V D R H N - - - - -
lc1|18 E C T F D V L N R - I E P V D D S I K F E G G N L - L I K N D F K N L N L R V A Q L S F L N A K K G
lc1|8 - - - - - - - - - - - D V D R H N - - - - -
lc1|15 - - - - - - - - - - - D K H N Q R - - - - -
lc1|14 - - - - - - - - - - - D V D R H N - - - - -
lc1|7 - - - - - - - - - - - D V D K H N Q - - - - -
lc1|16 - - - - - - - - - - - D V D R H N - - - - -
lc1|2 - - - - - - - - - - - D V D K H N Q Q - - - - -
lc1|20 E C T F D V L N R - I E P V D D S I K F E G G N L - L I K N D F K N L N L R V A Q L S F L N A K K G
lc1|21 - N Y L C H K K V - - - - - K I I F E N D - L S L E Y W H L K K Y N L I I A E L A V S R V E N G
lc1|12 - - - - - - - - - - - G A D R H N Q R - - - - -
lc1|26 - - - - - - - - - - - D V D K H N Q Q - - - - -
lc1|17 S C S L D A A G R - V D P V S S I T F E G G Q L - L I Q R D L K N L N L L V A E L A F M R A K N K
lc1|6 - - - - - - - - - - - D V D K H N Q R - - - - -

```

10

```

lcl|23      SGFYKEECESWKKEWYKSKEWRKYLFGKGIFEHKGKYEAKGSFIYGPYGLLF
lcl|11      -----
lcl|19      GGFYSKNWESWKHWYKDSGW-KNLFGFGIYGYGLDEAEGNDIYSPYGLTF
lcl|4        -----
lcl|9        -----
lcl|24      -----
lcl|10      -----
lcl|25      SGFNKSEYLLWKDWDYHDKEW-RGIFKSGIYDYGK-----NLFNNECGLVF
lcl|22      SGFYKKKYKSWKEWYQSLAW-KHVFGYGINGYQGAEAEGYFIYSPYGEIF
lcl|13      -----
lcl|1        -----
lcl|3        -----
lcl|5        -----
lcl|18      SGFYRKNWDSWKSWYQASSW-KNGLNSGLYGYGHDESEGNLIYSPYGITF
lcl|8        -----
lcl|15      -----
lcl|14      -----
lcl|7        -----
lcl|16      -----
lcl|2        -----
lcl|20      SGFYRKNWDSWKSWYQASSW-KNGLNSGLYGYGHDESEGNLIYSPYGITF
lcl|21      SGFYNNKKYKSWKEWYHSSAW-KTILGTGIYDYGLSEIGKNASSKPYGFIF
lcl|12      -----
lcl|26      -----
lcl|17      GGFYSKKSKSWKHWHYKDSAW-KHLFGYGIYGYGLEEANGNSIYDPYGLTF
lcl|6        -----

```

11

```

lcl|23      DDGSFAVG VKG---KD-IEEYGYNDNWTTL SGKNWY PLEYAGQIYFDKNG
lcl|11      -----
lcl|19      NDGSFSIGVVG---QD-VKESTKSDNFTKLSGTNWHTIKYAGQMFFDMNG
lcl|4        -----
lcl|9        -----
lcl|24      -----
lcl|10      -----
lcl|25      KDGSFLVGISGMMLNSLKKNF GHEDKWVK-ENKNLINNSSFGVMFFDKNG
lcl|22      EDCSFAVGITG---TD-IKKWGHNDNWTNLAGENWSSAS-AGQMYFDKNG
lcl|13      -----
lcl|1        -----
lcl|3        -----
lcl|5        -----
lcl|18      NDGSFSIGFSS---RKHINDNTKDDNFVKLNNANWSSFYAGQMFFDKNK
lcl|8        -----
lcl|15      -----
lcl|14      -----
lcl|7        -----
lcl|16      -----
lcl|2        -----
lcl|20      NDGSFSIGFSS---RKHINDNTKDDNFVKLNNANWSSFYAGQMFFDKNK
lcl|21      EDGSFSVGVTs---DD-AKKYGYNDTWTNYFKDDKETIY-AGQMFFDKNG
lcl|12      -----
lcl|26      -----
lcl|17      NDGSFSIGVVG---QD-VNKNTISDNFTRLAGTNWHTIKYAGQMFFDKNG
lcl|6        -----

```


12

```

lcl|23      RPIAIVMTNPPIGLLGDGWSFIYNKGKWEYESKDDWDERRFAGR-TESLD
lcl|11      -----
lcl|19      RPVALVITDVLTTGGFGSGWSFVYNNGKWEYESKDDWDDRRFQGSSTLSLD
lcl|4        -----
lcl|9        -----
lcl|24      -----
lcl|10      -----
lcl|25      RPVAVVFTVADKGFSGSGYSFIYNKKKWEYVLDENWDMNKFFGN-TKSLD
lcl|22      RPVAIVMTNLITGFFGAGWSFIYNEGKWEYESKDDWDKHRFSEQ-NESLD
lcl|13      -----
lcl|1        -----
lcl|3        -----
lcl|5        -----
lcl|18      RPVALVITEPLNAAFAGAGWSYIYKDGKWHYEAQDDWDQRLFKDS-TLSLD
lcl|8        -----
lcl|15      -----
lcl|14      -----
lcl|7        -----
lcl|16      -----
lcl|2        -----
lcl|20      RPVALVITEPLNAAFAGAGWSYIYKDGKWHYEAQDDWDQRLFKDS-TLSLD
lcl|21      RPIALTITNKISGYLGKGLSSLYIKGGWTWNIKDTWDIHNFKDN-YESLD
lcl|12      -----
lcl|26      -----
lcl|17      RPVVVLVITDALTTGGFGSGWSFIYNNGKWQYESKDDWDDRFFKGS-TLSLD
lcl|6        -----

```

13

```

lcl|23      PYAPRFLIKNI-----
lcl|11      -----
lcl|19      SNAPKFSTH-----
lcl|4        -----
lcl|9        -----
lcl|24      -----
lcl|10      -----
lcl|25      SHAPRFLINKYIVKSIDSEHESGGL
lcl|22      RYAARFLIKNI-----
lcl|13      -----
lcl|1        -----
lcl|3        -----
lcl|5        -----
lcl|18      PHAPQFIN-----
lcl|8        -----
lcl|15      -----
lcl|14      -----
lcl|7        -----
lcl|16      -----
lcl|2        -----
lcl|20      PHAPQFIN-----
lcl|21      RYAPRFLIKNKL-----
lcl|12      -----
lcl|26      -----
lcl|17      PHAPKFKL-----
lcl|6        -----

```

Chapter IV. Discussion

T3SS pathogens often employ ingenious and unexpected tactics for subverting the host immune response to create a unique biological niche. Secreted directly into host cells, T3SS effectors are responsible for the majority of pathogen-induced changes in infection^{24,49,50}. Therefore, a central goal of studying T3SS effectors is to understand how they affect virulence and whether developing inhibitors against them would decrease human disease and mortality.

When our lab first initiated research with the T3SS, NleC was a putative effector with unknown function during EHEC infection. In the Introduction, I discussed the published advances in knowledge regarding EHEC infection and NleC. I established that NleC proteolyzes NF κ B transcription factors in the DB loop in Chapter II and discussed how my results support the conclusion that NleC is highly specific for NF κ B, defining NF κ B residues crucial for substrate recognition by NleC. I also determined the crystal structures of NleC to high resolution and described how NleC varies from the many members of the Zincin zinc protease family and the potential implications of its unique structural features for translocation and specificity. In this final chapter, I will integrate these discussions, introducing a novel hypothesis for substrate recognition by NleC. Furthermore, I describe ongoing and future experiments to test the hypotheses arising from my research.

The Role of NleC During Pathogenesis

First and foremost, NleC is a T3SS effector with a significant role in pathogenesis. NleC knockouts of *Citrobacter* trigger significantly more inflammatory response during

mouse infection, and across A/E pathogens, is responsible for increased pro-inflammatory cytokine secretion⁷⁰⁻⁷⁴. NleC serves to inhibit the inflammatory response, limiting exposure of the bacteria to immunity cells and prolonging its adhesion to the epithelia of infected hosts. In Chapter II, I described how NleC achieves this anti-inflammatory effect by specifically targeting NF κ B subunits in their DB loop and preventing the activation of downstream transcription events. Here, I will combine this knowledge with that gained from the structural studies of NleC in Chapter III and discuss the structure of NleC is a new light.

Recognition of NF κ B Transcription Factors

While altering the DB loop of RelA with mutagenesis rendered it uncleavable by NleC, this fragment alone is not sufficient to promote cleavage. Experiments utilizing a minimized RelA substrate, either as small peptides containing only the DB loop, or a slightly larger fragment cloned in between two unrelated proteins, failed to show cleavage by NleC. Because cleavage was unobserved using these two systems, but was attained in the context of RelA₁₋₁₉₃, NleC likely recognizes distal elements of RelA beyond the local sequence of the cleavage site. This would indicate that NleC is a highly specific Zincin protease and that it is unlikely that proteins are targeted for proteolysis with any physiologically relevant efficiency. Insight into such specificity is facilitated when the structures of enzyme and substrate are both known, as is now the case.

A dominant feature from the structure of NleC is its strongly negative electrostatic face cradling the active site zinc (Figure III-14A). Electrostatic interactions are integral to binding, thus the negative character of NleC may contribute to binding of substrates by electrostatic interactions. This hypothesis is supported by surface analysis of the known

NleC substrates. RelA has a positive electrostatic surface potential due to its interaction with negatively charged DNA (Figure III-14B). However, though the surface of RelA is positively charged and expected to interact with the negative face of NleC, I was unable to isolate a stable complex of RelA₁₋₁₉₃ bound to the inactive mutant of NleC_{E184A}. This inability may be due to the absence of the dimerization domain, which may be important for stable binding.

If general electrostatic association by NleC is not the chief mode of binding NFκB transcription factors, recognition of specific residues must be considered. Detailed in Chapter II, several residues of NFκB subunits near the cleavage site are important for recognition by NleC. Mutagenesis of these residues in RelA, R35, Y36, E39, and R41, to alanine reduces cleavage efficiency significantly. Essential for recognition by NleC, the E39A mutant completely abrogates proteolysis. In other Zincins, the residue in this position determines whether cleavage occurs, binding in a large cavity on the surface near the active site Zn²⁺. Relative to the NleC-targeted scissile bond of NFκB subunits, this glutamate is directly C-terminal, in the P1' position. An obvious pocket on the surface of NleC correlates to the S1' subsite of Zincins, and would be responsible for specifically recognizing this residue.

As discussed in Chapter II, the P1' glutamate is conserved among all the NFκB subunits, as are R35 and Y36, in positions P4 and P3, respectively. However, in the Class I NFκB subunits p50 and p52, position P3' corresponds to a proline instead of arginine as in RelA, RelB and c-Rel. Because the mutagenesis experiments indicate that R41, the P3' residue, is important for efficient cleavage, the reduced cleavage observed for p50 may be due to this substitution. Additionally, mutagenesis of the P1 and P2' positions to alanine,

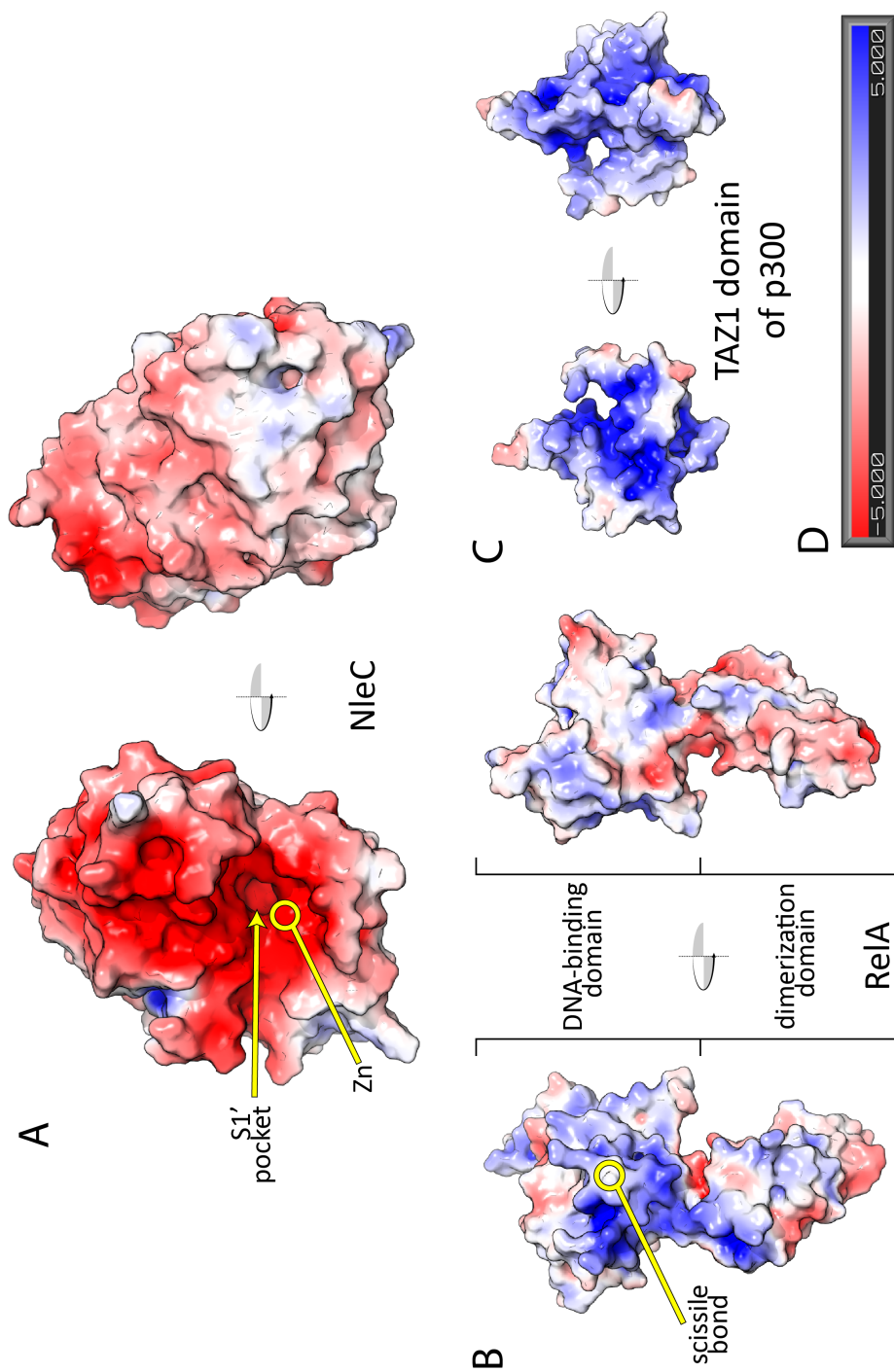


Figure IV-1 Electrostatic surface potential of NleC and substrates. A) NleC has a highly electronegative face with little positive character, though the positive charge of the Zn^{2+} ion was not taken into consideration with this software. A large S1' binding pocket is oriented directly behind the active site zinc. B) RelA is largely positive in character on the face of the DBD that comes in contact with DNA. The scissile bond targeted by NleC is centrally located in this positive electrostatic patch. The dimerization domain has more mixed character. C) The TAZ1 domain of p300 is a zinc finger domain that binds DNA and as such is also highly positive electrostatically. D) The key for the electrostatic coloring. All images were created with the APBS Pymol module and are colored with the electrostatic surface potential at solvent accessible distance and mapped on the Van der Waals calculated surface.

corresponding to C38 and G40 of RelA, slightly increase cleavage efficiency. To explain the structural basis of these results, it is important to obtain information about how NleC binds NFκB, either from an experimentally derived complex or from a computational model.

Because NleC family members in other bacteria were discovered to also target RelA, knowledge about the sequence conservation on the surface of NleC may lead to a more robust model of how NleC recognizes NFκB subunits. However, because the sequence conservation on the surface of NleC family members is very limited, structural elements that do not depend on conservation would have to be considered, such as pocket formation, general electrostatic binding, and hydrophobic binding. Such a model may lend additional insight into the NleC-NFκB binding interaction.

DNA mimicry: A putative NleC-substrate binding model

To further determine how NleC recognizes the NFκB transcription factors with such precision, we can look at the NFκB surface conservation and residues that are responsible for binding DNA, as discussed in Chapter II. While the NFκB transcription factors overlay structurally, their sequence conservation between Class I and Class II subunits is less than 60%. Utilizing the sequence alignment of the five NFκB subunits and the Consurf server, I visualized the conserved residues shared by all subunits onto the structure for RelA (Figure II-9)⁸⁴. Additionally, because NleC cleaves NFκB in the DB loop, we can restrict our search for a putative NleC binding site to the face that recognizes DNA (Figure II-8)^{83,133}. Interestingly, three of the same residues that contribute to DNA-binding were also shown to be important for binding by NleC in the mutagenesis experiments, R35, Y36 and E39. It is important to note that these three residues are also conserved among all NFκB subunits. In

fact, all of the residues that are important for DNA-binding by RelA are also conserved among the NFκB subunits, except for three (Figure IV-2)⁸³. NFκB subunit p50 uses these same conserved DNA-binding residues to recognize DNA, in addition to four others that RelA does not use¹³³(Figure IV-3). This enables the NFκB subunits to bind to different DNA sequences, and may also contribute to the preferential cleavage of RelA over p50 by NleC.

The overlap in RelA residues that bind both NleC and DNA raises the possibility that NleC and DNA occupy similar binding surfaces on RelA. In fact, when NleC is manually overlaid with DNA from the RelA-bound structure, striking similarities are apparent in the size and shape of the NleC active site cleft. Indeed, the major groove of the DNA helix where RelA binds superimposes with the large cleft on the surface of NleC that cradles the active site Zn²⁺(Figure IV-4). The strong electronegative character of the NleC cleft also mimics the surface electrostatics of DNA.

Furthermore, upon the addition of RelA from the DNA-bound structure, almost all residues avoid steric clashes with the surface of NleC, without rearrangement of either crystal structure (Figure IV-5). The few steric hindrances that occur are likely negligible as minor local rearrangements could accommodate binding. For proteolysis to occur, changes do need to be made in the DB loop of RelA, as the carboxyl of the scissile bond is 1.9Å from the coordination sphere of the active site Zn²⁺, where it would be located for proteolytic attack. The general binding mode of the model need not change for this rearrangement. The lip of the DB loop where the scissile bond is located does not have any regular secondary structure and would therefore be free to move into the NleC active site pockets.

The model derived from the DNA mimicry hypothesis of NleC further permits modeling of how NFκB, IκB, and NleC may interact in the cytosol upon EHEC infection.

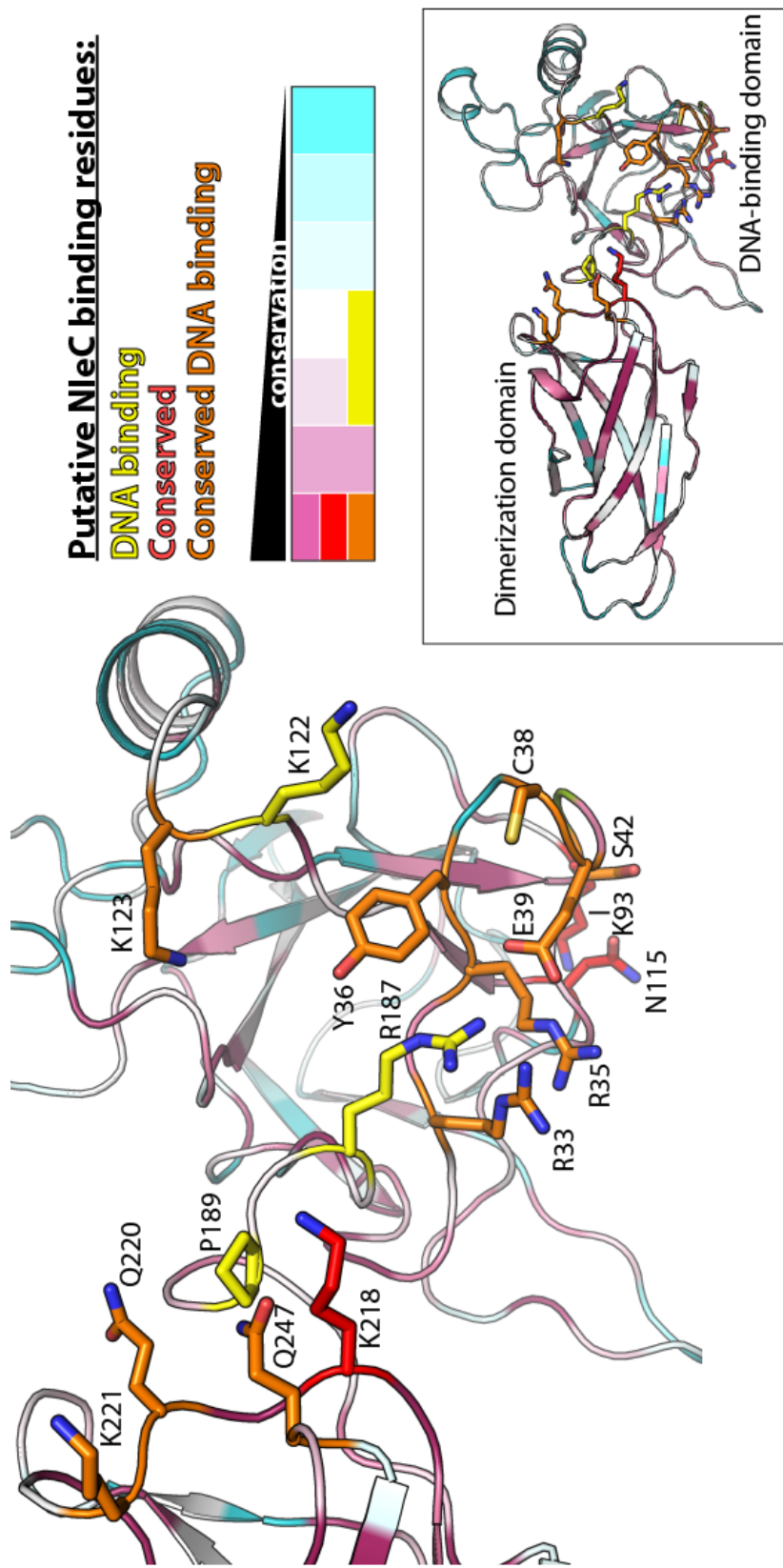


Figure IV-2 Conserved residues among NFκB overlap with DNA-binding residues. Many of the residues responsible for binding DNA in RelA are also conserved between all NFκB transcription factors (orange). Only three residues that important for DNA-binding are not conserved between the NFκB subunits (yellow). Three other residues that are conserved between NFκB subunits are within the area suspected to be important for NleC binding (red). Alignment mapped onto RelA structure with the Consurf server⁸². PDB ID: 2RAM

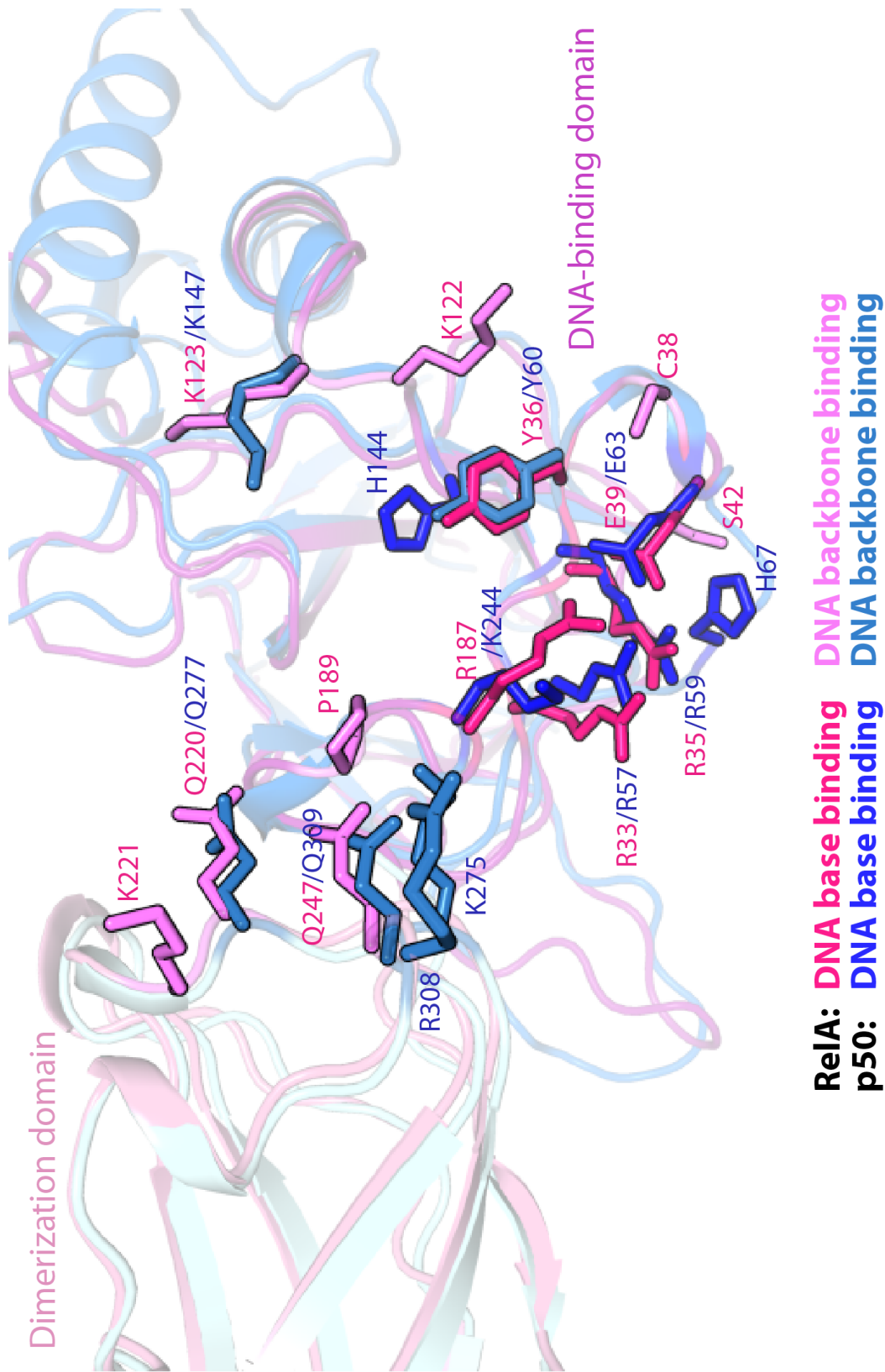


Figure IV-3 DNA-binding residues of RelA and p50. The majority of the DNA-binding residues are conserved between RelA and p50. Those that are not conserved are responsible for the specificity of the NFκB subunit to a particular DNA sequence. RelA is shown in pink, whereas p50 is shown in blue^{81,133}. PDB IDs: 2RAM, 1SVC

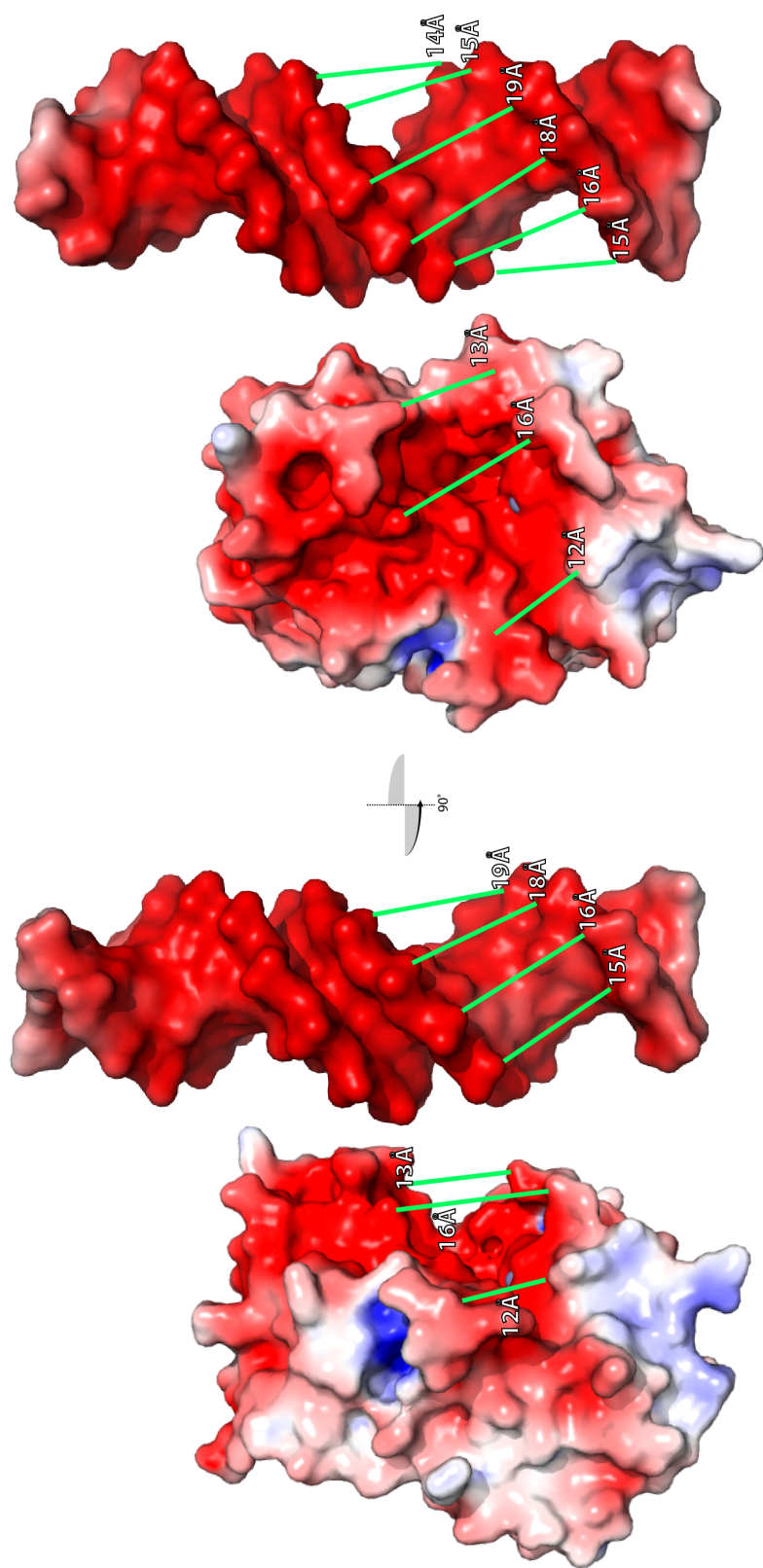


Figure IV-4 NleC active site cleft comparison with DNA showing electrostatics. Three views of NleC and the binding site of RelA on DNA. Both the highly negative character and the active site cleft of NleC are remarkably similar to that of the major groove of DNA. Measurements made in Pymol. Electrostatic surface calculations made with APBS in Pymol. PDB IDs: 2RAM.

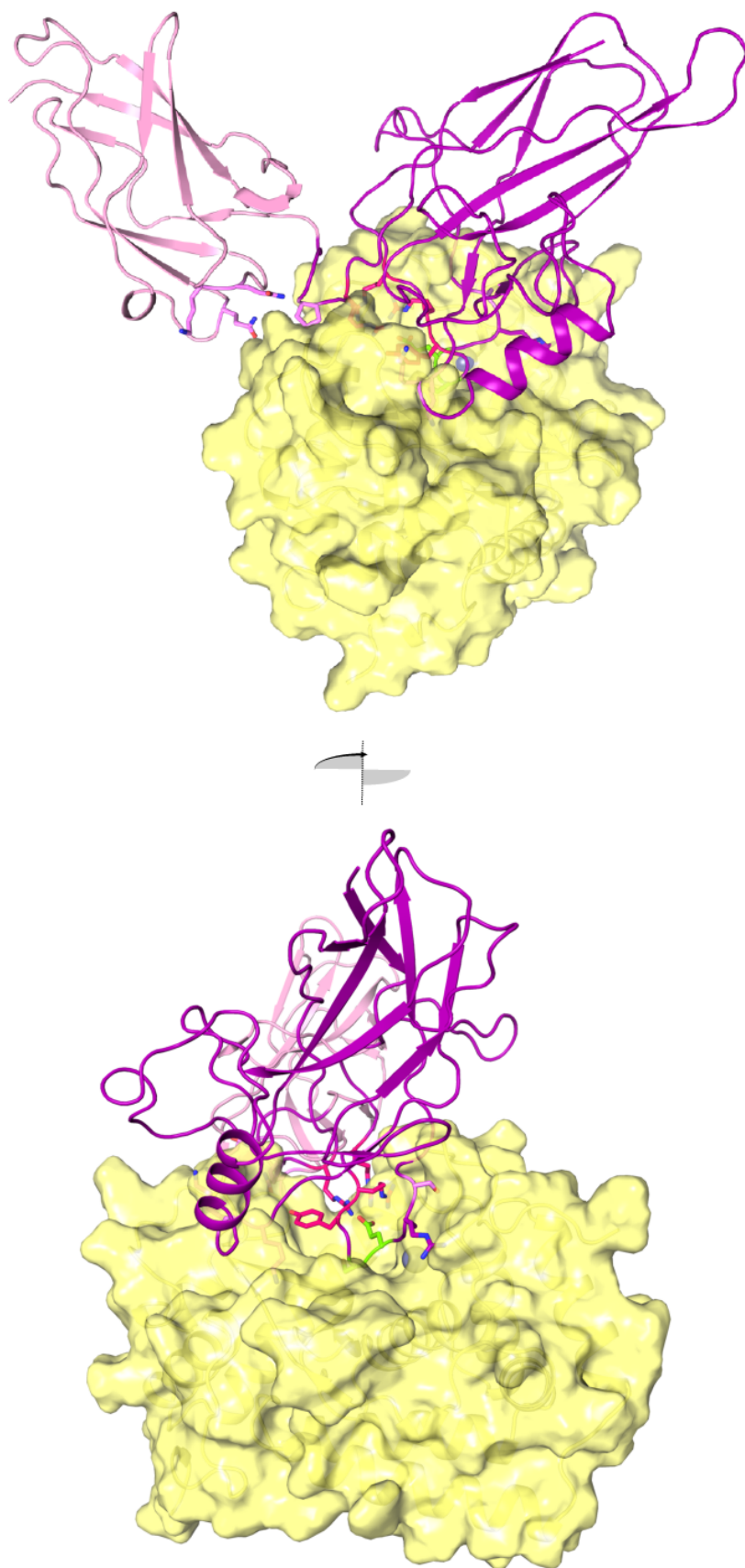


Figure IV-5 Model of DNA-like mode of binding for NleC bound to RelA. NleC is shown as a surface in yellow. The DBD of RelA is shown in magenta, and the dimerization domain of RelA is pink. Residues shown in stick-format are the DNA-binding residues. PDB IDs: 2RAM.

Between the structures of RelA in complex with DNA and in complex with I κ B, the orientation of the Rel-homology domain in respect to the dimerization domain is twisted. Because I κ B recognizes and binds only the dimerization domain when sequestering NF κ B, the dimerization domains of the I κ B-bound and DNA-bound RelA can be superimposed. This will retain the orientation of RelA to NleC while also modeling in I κ B to permit insight into how the complex forms *in vivo* (Figure IV-6). This model illustrates that the DNA mimicry binding mode that I propose is fully consistent with NleC recognition of RelA while being sequestered by I κ B. Analysis of this model reveals slight steric hindrances between NleC and the C-termini of I κ B that on further inspection would not be expected to occur in solution (Figure IV-7A). Because there are no favorable interactions between the C-termini of I κ B and the dimerization domain of RelA, the orientation of the I κ B C-termini in the structure is likely the result of crystal contacts and is probably flexible in solution (Figure IV-7B). As such, a very minor rearrangement of I κ B would completely avoid the few clashes observed in the model. In fact, this model is consistent with the hypothesis that I κ B and NleC interact during pathogenesis and may provide a mode for NleC to bind and cleave I κ B after proteolysis of NF κ B, given the close proximity and increased local concentration⁷². From this model, the interaction of the rear face of NleC with the other dimer in the NF κ B complex is also possible.

The overlap of a DNA helix with the active site groove of NleC, combined with the highly electronegative character of NleC suggests that NleC structurally mimics DNA. The fact that NleC recognizes and proteolyzes the DB loop of RelA, combined with the data that suggests NleC recognizes elements outside of this loop, is consistent with the exciting

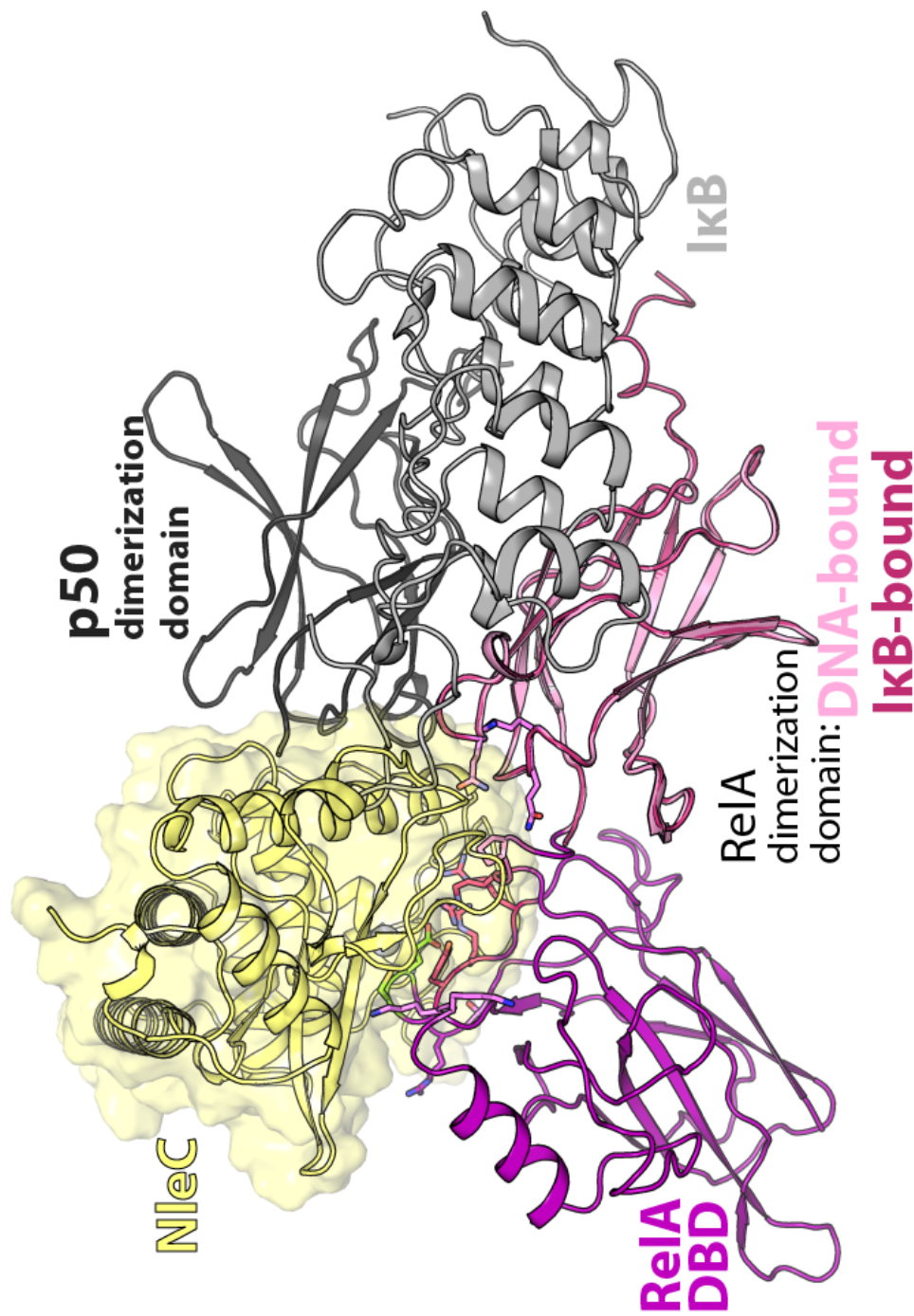


Figure IV-6 Proposed model for the complex formation of NleC, NFκB, and IκB. The model of NleC (yellow) and RelA (fuchsia) is the same as depicted in Figure IV-5, where NleC is overlaid with DNA. Here, the dimerization domain from the DNA-bound RelA (pink) is overlaid with the dimerization domain from the IκB-bound RelA (magenta) to illustrate how the complex may form *in vivo*. In this case, the DBDs of RelA do not overlay, as the orientation of the DBD to the dimerization domains are different in the structures. The DNA-bound structure has a more rigidly associated dimerization domain, as both the DBD and the dimerization domain are bound to the DNA through specific interactions. Here, IκB in grey is able to interact with the side of NleC. The p50 dimerization domain is also shown in this model in black and may interact with NleC on the face opposite its active site. PDB IDs: 2RAM, 1IKN.

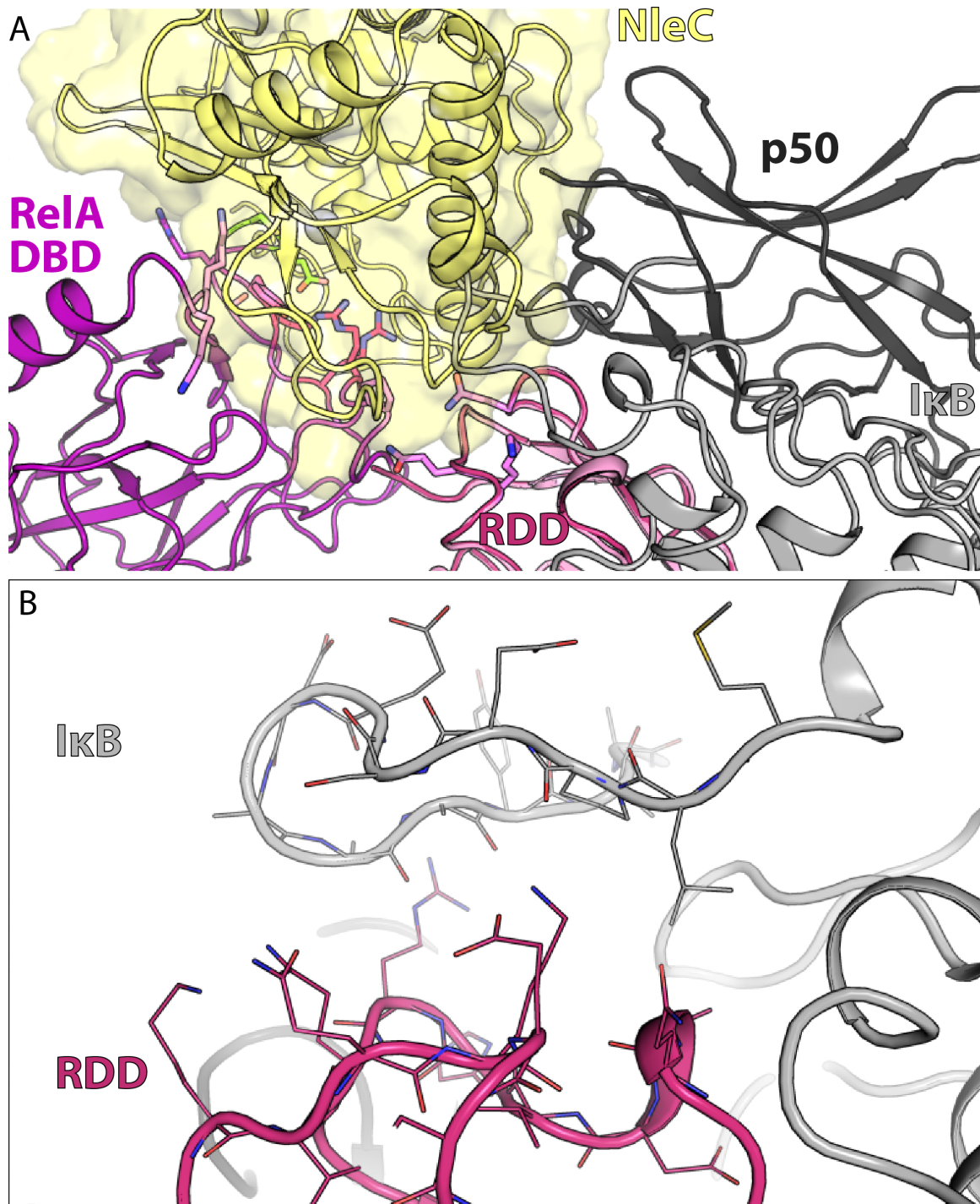


Figure IV-7 Interaction between putative NleC-NFκB-IκB complex. A) A closeup view of the interactions between NleC (yellow), RelA (pink), p50 (black) and IκB (gray). The interaction of NleC with RelA is shown more clearly in Figure IV-5. Here, the C-termini of IκB is sterically hindered by NleC. B) RDD refers to the RelA dimerization domain. From examination of the crystal structure of RelA in complex with IκB, it is apparent that the C-termini is not specifically associated with RelA and so is likely to be free to move in solution. This would alleviate the steric hindrance in model. PDB ID: 2RAM, 1IKN

hypothesis that NleC may resemble the binding partner that normally occupies this site, DNA. This model can thus be used to formulate hypotheses about which distal elements of RelA are necessary for recognition by NleC. Additionally, the structural mimicry of DNA by NleC may explain why other DNA-binding proteins, such as p300, are bound by NleC, whether specifically or not. If proven, this DNA-mimicking model would be novel and especially intriguing, because even though T3SS effectors have been shown to mimic host-cell proteins, none are yet implicated in mimicking the binding mode of DNA^{43,45}. Proteins that exhibit DNA mimicry are still rare, with less than 20 examples published^{134,135}. The function of the described DNA mimicking proteins to date is to sequester or inhibit DNA-binding proteins from interacting with their cognate DNA, and this is certainly the function of NleC^{135,136}. A DNA-mimic which additionally proteolyzes its substrate is especially nefarious.

Putative Substrate, p300

A second putative NleC substrate is the transcriptional activator p300. As stated in Chapter II, I successfully pulled down NleC with the N-terminal domain of p300, the TAZ1 domain of which has a highly positive electrostatic surface (Figure III-14C). Because I was unable to observe significant cleavage of p300 but could detect binding, it is conceivable that NleC requires full-length p300 or an unknown binding partner to cleave efficiently. Additional possibilities are that the cleavage of p300 is a secondary effect, and that binding to p300 may serve as a localization signal for NleC to target RelA in the nucleus. Given the strong complementary character of the electrostatic surface potentials, the binding of p300 may also be an artifact resulting from a nonspecific interaction with NleC acting as a DNA mimic.

The question of whether p300 binding is physiologically relevant can be teased apart by considering NleC localization during infection. However, there are conflicting reports as to whether NleC is predominantly localized to the cytosol or the nucleus^{70,71,73,74,76}. I noted two major differences between whether a nuclear or cytosolic localization was reported in the literature. Nuclear localization dominated when tagged NleC was introduced to the cell *via* transfection of a plasmid and imaging occurred an average of 16 hours post-transfection. When NleC was found in the cytoplasm, tagged NleC was translocated *via* T3SS infection and cells were imaged after only a few hours. It is also important to note that the NleC cleavage site of the NFκB subunits is exposed when bound to IκB in the cytosol, where it would be easily accessible, while in the nucleus, DNA would prevent association. These observations suggest that the more physiologically significant experiments delivering NleC by T3SS have greater relevancy. I hypothesize that NleC functions chiefly in the cytosol to cleave IκB-bound NFκB directly after NleC is translocated, whereas after longer incubation times, NleC may accumulate in the nucleus where there are many positively charged transcription factors to bind nonspecifically. In the nucleus, NleC would also be able to complete cleavage of any nuclear-localized NFκB factors, though their association with DNA would slow this activity.

NleC and other transcription factors

Given the DNA-mimicry model of NleC, it is possible that NleC will be able to recognize and cleave other transcription factors beyond NFκB. An intriguing possibility is the NFAT transcription factors, which also contain a Rel-homology domain that includes both the DBD and the dimerization domain¹³⁷. However, there is little sequence similarity

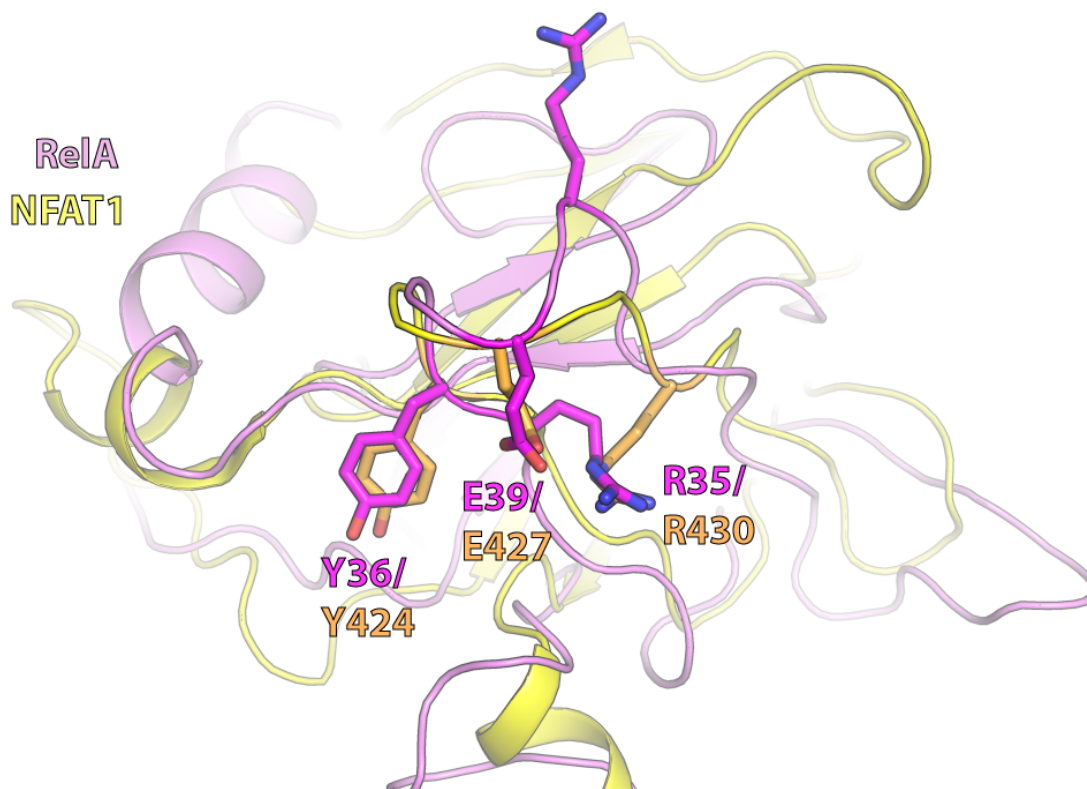


Figure IV-8 Comparison of the DNA-binding domain of RelA and NFAT1. The three residues in the DNA-binding loop of RelA that are most important for cleavage by NleC are maintained structurally, though not sequentially, in NFAT1. This opens the possibility that NleC is capable of cleaving NFAT transcription factors, which also contain a Rel-homology domain. PDB ID: 2RAM, 2AS5

to the Rel-homology domain of NF κ B. The DB loop of NF κ B subunits that I characterized as important for efficient cleavage, **RYKCEGRS**, where the residues recognized by NleC are bold, is partially maintained in the NFAT transcription factors, which contain the sequence *HYETEGSR*, where the italicized residues correlate with the emboldened residues in the NF κ B DB loop, with an insertion in between EG and R at the end of the sequence¹³⁸. Upon viewing the structures of RelA and NFAT1, it is apparent that the arginine that is essential for efficient proteolysis in RelA, R35, is actually overlaid with the final arginine in the NFAT sequence, R430 (Figure IV-8). This would also mean that the final arginine that has an intermediate effect on NleC recognition of RelA, R41, is not present in NFAT. However, the

maintenance of the three most important residues for efficient cleavage is intriguing. The structural similarity of this DB loop between NF κ B and NFAT, as well as that of the general Rel-homology fold, opens the possibility the NleC is capable of recognition and cleavage of these transcription factors. These NFAT proteins are additionally implicated in the immune response, largely in B- and T-cells, but in nonimmune cell types as well, regulating the expression of interleukins and TNF^{137,139}. If NleC is capable of proteolyzing NFAT proteins, it would have further consequences in the proper immune response of host cells upon infection by EPEC beyond the disrupting the NF κ B signaling cascade discussed in this work.

The C-terminal Domain of NleC in Function

In the crystal structure of NleC, the N-terminal and C-terminal portions were not present and are not required for cleavage of NF κ B subunits *in vitro*. Using SAXS, I determined that the C-terminal fifty residues that were truncated from the crystal structure fold into a discreet domain separate from the Zincin domain. This domain likely has a function during the course of translocation and infection, as it is maintained in copies of NleC from other enterobacteria and effectors often have a minimized structure to facilitate translocation. A possible hypothesis for this C-terminal domain is that it is necessary for pathogenesis independently of proteolysis, such as for targeting within the host cell.

In order to evade the immune response, bacteria utilizing T3SSs need to invade as rapidly and efficiently as possible. Because of this, many effectors contain targeting motifs or domains in order to facilitate locating substrates. Tir, one of the first translocated effectors during infection, incorporates into the host apical membrane, where it serves as a receptor for bacterial adhesion and is ideally located to modify actin dynamics. The

effectors EspF and EspG additionally target actin and microtubules, modifying the cytoskeleton and disrupting endosomal trafficking. The EHEC effectors Map, NleH, and EspI all contain PDZ-binding motifs that target them to the scaffolding protein NHERF2/Ebp50, located at the apical membrane²⁴.

With this precedent, we can look for additional binding interactions for NleC besides NF κ B subunits, such as p300, and specifically focus on the C-terminal domain. To explore this possibility, future experiments utilizing the p300 pulldown experiment described in Chapter II can be repeated with full-length NleC and a truncation that lacks the C-terminus, NleC₁₋₂₈₀. Pulldown experiments with only the C-terminal domain, NleC₂₈₀₋₃₃₀, as bait in whole cell mammalian lysate may isolate other putative binding partners.

An experiment that tests the function of the NleC C-terminal domain during infection was conducted by collaborators at Osaka University, Hilo Yen and Toru Tobe. These experiments examined whether the C-terminal domain was necessary for function in pathogenesis by infecting THP-1 cells and measuring the output of NleC infection, TNF- α . Unfortunately the effect on TNF- α during infection was unchanged when NleC lacked the C-terminal fifty residues, suggesting that this domain does not affect targeting to NF κ B transcription factors under these circumstances. However, because TNF- α was measured only at the end of the assay, seven hours after infection began, it is still possible that the absence of the NleC C-terminal domain had an influence on pathogenesis during the course of infection. This effect may be lost with long incubation times, perhaps related to NleC localization at different time-points post-infection.

NleC as a Tool

As we know, NleC is a specific protease for NF κ B transcription factors and may even acts as a DNA mimic. In studying how NF κ B transcription factors function both *in vivo* and *in vitro*, NleC would be a useful tool for inhibiting their activation both by proteolyzing and sequestering the DNA-binding site. If transformed into cells with an inducible plasmid, resultant wildtype or inactive NleC can be controlled in a temporal manner. This would allow inhibition of NF κ B on a quicker timescale than that obtained by knockdown of the NF κ B gene and may also be less deleterious than knockout experiments that are in some cases nonviable¹⁴⁰⁻¹⁴². With *in vitro* studies, active NF κ B can be inhibited with the addition of NleC. Derived from my experience working in RelA, iodoacetamide may also be used to prevent the association of DNA and RelA by disrupting the cysteine in the DB loop.

Synergy among T3SS effectors

As explained in Chapter I, there are many EHEC T3SS effectors that have an impact on the inflammatory response in host cells. These effectors work in tandem to tightly regulate host cells, beneficially elongating infection for the invading bacteria. Five EHEC T3SS effectors have been identified that directly impact the NF κ B signaling cascade. Tir inhibits the effect of TNF- α signaling by promoting the degradation of TRAF adaptor proteins, halting the signaling pathway at the receptor complex⁶². Effective only in the TNF- α induced pathway, NleB seemingly also functions upstream of NF κ B, before the TNFR, IL-receptor, and TLR pathways converge⁶⁸. NleE functions downstream of Tir and NleB to methylate a zinc-binding cysteine in TAK1, preventing it from phosphorylating I κ B. This activity has an effect on signaling cascades initiated by TNF- α , interleukins, and the direct

recognition of the bacteria⁸¹. NleC works at the level of NF κ B, as I've presented in this work, cleaving in the DB loop and thus preventing NF κ B from interacting with DNA once imported into the nucleus. NleH1 and NleH2 work together to tightly regulate the nuclear importation and phosphorylation state of NF κ B co-transcriptional factor RSP3^{67,82}.

Given this information, hypotheses can be made about how NleC may work in conjunction with the other effectors implicated in impacting the NF κ B pathway. The effectors that work upstream of NleC in the signaling cascade are expected to support the activity of NleC, if NleC functions chiefly in the cytosol while the NF κ B subunits are bound to I κ B. This is because these effectors, Tir, NleE and NleB, are not expected to directly interact with NleC, and would result in the extended sequestration of NF κ B in the cytosol, where DNA would not competitively bind the DB loop. However, experimental data from EHEC infections in the absence and presence of NleE show that the activity of NleE actually slows the efficient cleavage by NleC^{73,74}. This may be because having I κ B bound to NF κ B prevents the association with NleC, disproving the model I presented above. Alternatively, because the signaling molecule affected by NleE, TAK1, is important for activation of multiple transcription factors, the interplay between these other signaling cascades and NF κ B may affect NleC's ability to proteolyze it. For instance, the MAPK JNK and FKBP51, a Hsp90 co-chaperone, are both shown to interact with I κ B and may affect the ability of NleC to target NF κ B for degradation¹⁴³. Further experiments testing the effect of NleB and Tir knockouts on NleC activity are necessary to determine whether I κ B degradation is necessary for efficient cleavage of NF κ B by NleC or if this is an NleE-specific result.

Downstream, once in the nucleus, the action of NleH1 and NleH2 to affect RPS3 would support the ability of NleC to suppress the inflammatory response, as RPS3 is a NF κ B co-transcriptional activator that specifically promotes the expression of pro-inflammatory cytokines^{67,144}. By regulating the nuclear import of RPS3, NleH effectors essentially alter the affinity of NF κ B away from pro-inflammatory genes, the transcription of which are positively correlated with the presence of RPS3⁶⁷. This would support the activity of NleC in inhibiting the inflammation response of host cells through NF κ B.

Synthesizing these data together to establish how exactly EHEC controls the inflammation response, and NF κ B specifically, during pathogenesis is necessary. Preliminary experiments suggest that NleE and NleB work to support the suppression of NF κ B activation⁶⁹. NleH1 and NleH2 knockouts in live animal models have resulted in conflicting evidence, allowing easier clearance of the mutant *Citrobacter* strain in mouse model, while resulting in a hypervirulent strain in infected piglets. This difference also seems dependent on whether one or both NleH effectors are removed⁶⁷. Deleting NleC from *Citrobacter* similarly resulted in an infection with higher levels of colitis and permitted the bacteria to colonize further up the intestines⁷⁴. The activity of NleC, which would be predicted to be supported by NleE inhibiting the phosphorylation of I κ B upstream, is actually more effective when NleE is absent during infection, as discussed above^{73,74}. These results foremost illustrate that the regulation of the NF κ B pathway by T3SS effectors is complicated, and furthermore suggest that the assemblage of effectors present in A/E pathogens is the result of careful evolution. The appearance of more virulent phenotypes upon deletion of some of these effectors further inform that a successful infection with regard to the invading bacterium is thus not necessarily the most virulent. NleC has an

important role in this interplay, being one of many effectors that directly impact the NF κ B pathway.

Secretion as a Unique Evolutionary Pressure on Structure

As discussed in Chapter III, NleC has a highly α -helical structure with a modified, zippered Ψ -loop. These alterations from the canonical Zincin structure may be necessary to permit NleC to unfold for translocation and, once in the host cell, refold in order to fulfill its pathogenic fate. Furthermore, the question of whether the C-terminal domain serves a purpose in translocation is unanswered. Maintaining the Zn²⁺ ion in the active site throughout translocation is a further difficulty, though it would be advantageous given the highly controlled nature of zinc homeostasis in the host. To address these various questions, which would also yield insight into how other effectors are evolved for translocation, I discuss experiments already initiated, as well as hypotheses I have developed.

T3SS as an Engineered Protein Delivery System

A long-term goal of understanding how proteins are translocated lies in the possibility of using the T3SS as a protein delivery system. Because these enterobacteria are evolved to prosper in the intestinal milieu, minimally viable bacteria could be developed that deliver engineered proteins to intestinal epithelial cells and later cleared by the immune system. By understanding how proteins are altered by evolution to be translocated, these proteins can be better engineered to be efficiently delivered by the T3SS. The requirement for a chaperone or targeting signal for secretion, as well as the

kinetic requirements for physical translocation need to be defined before this ambitious engineering tactic can be successful.

Mechanical Process of Unfolding

To study translocation in more depth, we have initiated collaboration with Tom Perkins at University of Colorado at Boulder. The Perkins group is both highly skilled and innovative in single molecule techniques, atomic force microscopy (AFM) and optical tweezers. The unique secretion trajectory of T3SS effectors means that while the thermal or chemical denaturation of these proteins may not show unusual instability, the more physiologically relevant experiment is mechanical unfolding. This is because effectors need to be at least partially unfolded by an ATPase that pulls the protein through a narrow orifice and propels it through the secretion needle. We hypothesize that NleC has evolved to be easier to mechanically unfold and refold than a Zincin of similar size.

For these experiments, we have created constructs of NleC with two copies of NUG2 on either side. NUG2 is a synthetic protein designed by David Baker at University of Washington Seattle and has been studied in depth by the Perkins group with AFM. This provides an internal control for the experiment. Though many proteins are sticky and adhere to the AFM base and tip without special preparation, using different chemistries at either the base or the tip ensures directionality and specificity. Initial experiments were done without special treatment of the amino or carboxy terminus of the protein, however since then an N-terminal cysteine has been introduced to allow covalent attachment utilizing a maleimide-activated crosslinker. These constructs created further difficulties, however, as the preparation of the base that reacts with the N-terminal cysteine may also react with three internal cysteines in NleC. The preparation of a construct that utilizes a

specialized modification on both termini will help remedy this. Because the chemistry used to adhere the proteins to the base requires incubation for an hour, by switching the cysteine-based chemistry to the AFM tip, we will reduce the propensity of internal cysteines from reacting. If this tactic fails, the cysteines in NleC can also be mutated out. On the N-terminus of the protein we are using a genetically encoded tag that is modified by a bacterial protein using a novel method. This small tag, containing a cysteine, is specifically recognized by a co-expressed formyl-glycine generating enzyme and modified into an aldehyde in cells. This tag will then be covalently bonded to the AFM base utilizing hydrazide chemistry.

Initial results from the AFM experiments are promising. Constructs were designed with both full length NleC, as well as the C-terminal truncation NleC₁₋₂₈₀, to distinguish whether the C-terminal domain altered the way NleC unfolded. The Perkins group has been able to distinguish between the NUG2 and the NleC traces (Figure IV-9). With the improved cysteine and aldehyde-based chemistries, accurate data will be obtained for the force required to unfold NleC versus NleC₁₋₂₈₀. With this data, comparisons to Zincins of similar sizes will be initiated. Additionally, experiments unfolding NleC in the presence of binding partners or in the absence of Zn²⁺ in the active site will contribute new information to the stability of a Zincin effector.

Using the structure of NleC, I also have hypothesized about structural elements that would be translocated intact. In Chapter III, I discussed the possibility of translocating the zinc coordination sphere by maintaining contacts between the catalytic helix, backing helix, and following loop. Applying this same logic to the remainder of the protein, I developed hypotheses about how the remainder of the protein might unfold during translocation

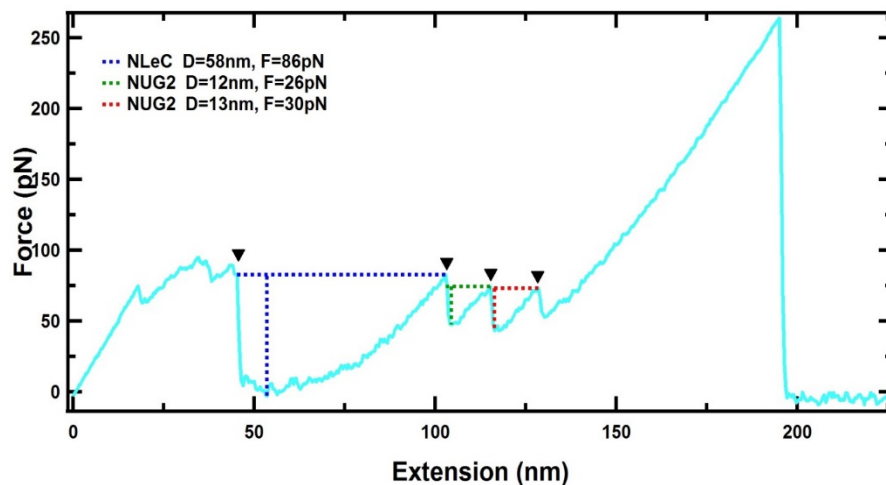


Figure IV-9 Representative AFM trace of Cys-Nugx4-NleC₁₋₂₈₀. The sample is bound to a glass slide with maleimide chemistry before loading onto the Cypher AFM. For this trace, the AFM base moved at 800nm/s towards the tip after an initial 200 pN force of the tip into the glass slide to adhere the protein. The initial force seen in this trace is likely adhesion forces from the protein sticking intramolecularly to the base. After 50 nm, NleC is unfolded, resulting in a drop in force, and extends for 58nm. The next two peaks are Nug2 proteins unfolding, while the last increase in force is the polypeptide being extended with the large drop resulting from the protein detaching from the base or tip.

(Figure IV-10). The majority of the helices might be translocated individually, as discrete structural elements are not apparent from the helices orientation to neighboring helices. The unique β -sheet of NleC with the zippered outer strands may permit translocation of the whole sheet including the extended loop. This β -sheet structural element is directly before the catalytic helix and may function to protect the zinc-coordination helices so that their integrity is maintained during translocation. The final C-terminal helices and the C-terminal domain described with SAXS are then likely translocated individually. These hypotheses can be tested explicitly using AFM, by dividing NleC into these smaller structural elements and comparing the force required to unfold them versus the full-length protein. If either of these elements, the β -sheet or the active site helices, require similar forces to unfold as the

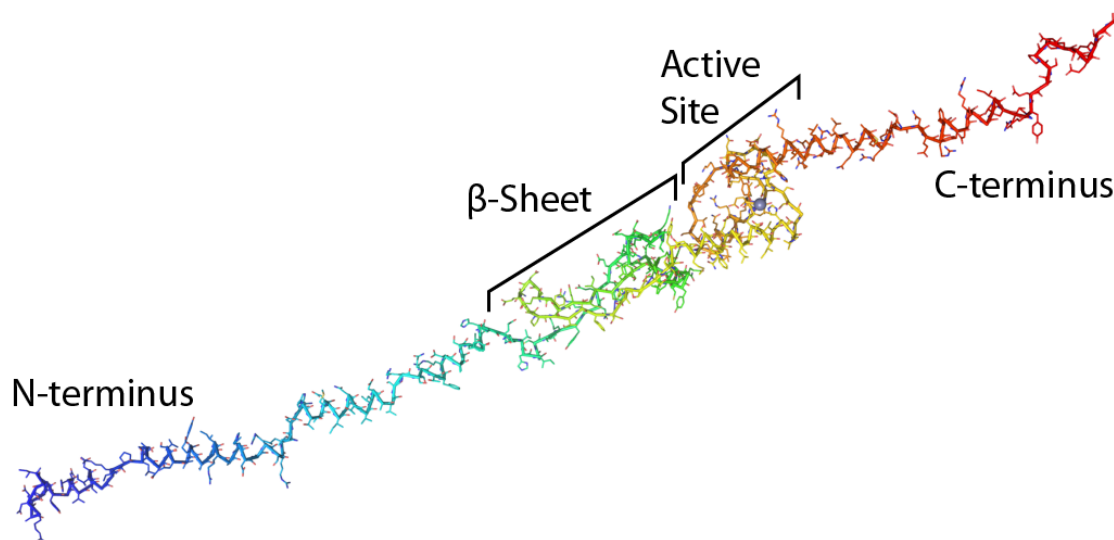


Figure IV-10 Hypothesized structure during translocation. From this hypothesized structure, NleC can be fragmented into discrete structural elements for unfolding by AFM. This will elucidate whether the β -sheet motif and the active site helices, especially, are responsible for the majority of the force required to completely unfold NleC by AFM.

full-length NleC, it can be deduced that these elements contribute significantly to the stability of NleC. Additionally, these studies will help to develop hypotheses about how much force is applied to proteins during translocation by T3SS, as well as about the stability, size, and shape of translocatable structural elements.

Can the Zinc be Translocated as well?

Highly regulated in both bacterial and eukaryotic cells, zinc is rarely found uncoordinated. Bacterial species, like pathogens, that live inside host organisms have elaborate systems for scavenging zinc from the environment^{129,132}. Because T3SS effectors are partially unfolded to be translocated for pathogenesis, an intriguing question is whether Zincin effectors are secreted in complex with the active Zn^{2+} . This would be preferential given the limited availability of zinc in the host cell. In Chapter III, I discussed

how the helices cradling the active site Zn^{2+} are in close enough proximity to allow the translocation of the whole coordination sphere, especially if the adjacent loops are dynamic. In order to confirm this hypothesis, whole cell assays specifically examining secretion or infection must be conducted.

To look at whether Zn^{2+} is translocated explicitly, secretion experiments must be performed using T3SS-competent cells. From our collaborators at Osaka University, Hilo Yen and Toru Tobe, we are obtaining wild type EHEC cells and the reconstituted minimal T3SS strain (TOB02) that they developed. Using zinc-free media, we can probe for secreted zinc using fluorescent zinc probes upon stimulation of T3SS secretion. Plasmids containing NleC will have to be transformed into both EHEC and TOB02 cells to ensure a measurable quantity is secreted. As a control, the level of protein secretion will be measured by SDS-PAGE or Western Blot and compared to measured zinc levels. These experiments can be coupled with AFM of NleC in the presence and absence of Zn^{2+} , removed by EDTA. From these experiments, I expect to observe that NleC translocates *via* the T3SS with the Zn^{2+} -coordination sphere intact, and that the holoenzyme is more stable than the apoenzyme. Entering into the host cells with Zn^{2+} in place ensures that NleC is stable enough to refold around the active site and can perform its function of proteolyzing NF κ B subunits immediately.

Translocation as an Evolutionary Pressure

When considering whether evolutionary pressure from the act of translocation results in a greater α -helical content than the typical Zincin, while maintaining a high level of specificity, the ideal situation would be to study other zinc protease effectors to determine if there is a trend. Conveniently, EHEC has another Zincin protease in its effector

arsenal, NleD, that has no sequence similarity to NleC or any other Zincin outside of the conserved HExxH motif. Structural studies of NleD would be most helpful in exploring these hypotheses. However, I have performed expression and solubility tests on NleD, but with limited success. The two tags that were utilized, an N-terminal 6xHis tag and the C-terminal Intein-CBD tag that was so successful with NleC, rendered NleD insoluble. Expression with the 10xHis-SUMO tag used for expressing NF κ B subunits may help both expression and solubility of NleD, at levels necessary for crystallographic determination. Alternatively, because the host-binding target for NleD has been identified as JNK, a co-expression platform could be used to drive the solubility of NleD and obtain the enzyme-substrate complex at the same time.

In lieu of a crystal structure of NleD, secondary structure predictions of NleD, combined with what we know of the topology of the catalytic domain in Zincins, can lead to a robust prediction of NleD's overall structure (Figure IV-11). This model topological map can be compared to that of NleC and its closest structural relatives from Chapter III (Figure III-8). Because NleD is a smaller protein than NleC, at 232 residues compared to 330 residues for NleC, features outside of the catalytic domain can be expected to be minimal. From these predictions, NleD is expected to have more β -strands than NleC, but these β -strands are shorter in length. In determining whether effector Zincins have reduced β -strand content, secondary structural features of several Zincins were tallied. NleD has 12% β -strand content and 36% helical content, whereas NleC has 12% β -strand and 35% helical content. For the closest structural neighbors of NleC, botulinum neurotoxin has 13% β -strand and 35% helical content, whereas tricorn aminopeptidase has 24% β -strand and 45% helical content, where only secondary structure elements that are shared with NleC

were counted (delineated by the colored outline in Figure III-8). The wide range in β -strand character between botulinum neurotoxin and tricorn aminopeptidase illustrates that the catalytic core of Zincins is highly variable. Though NleC and NleD have slightly less β -strand character than the Zincins that were measured, this small difference might not be significant for translocation.

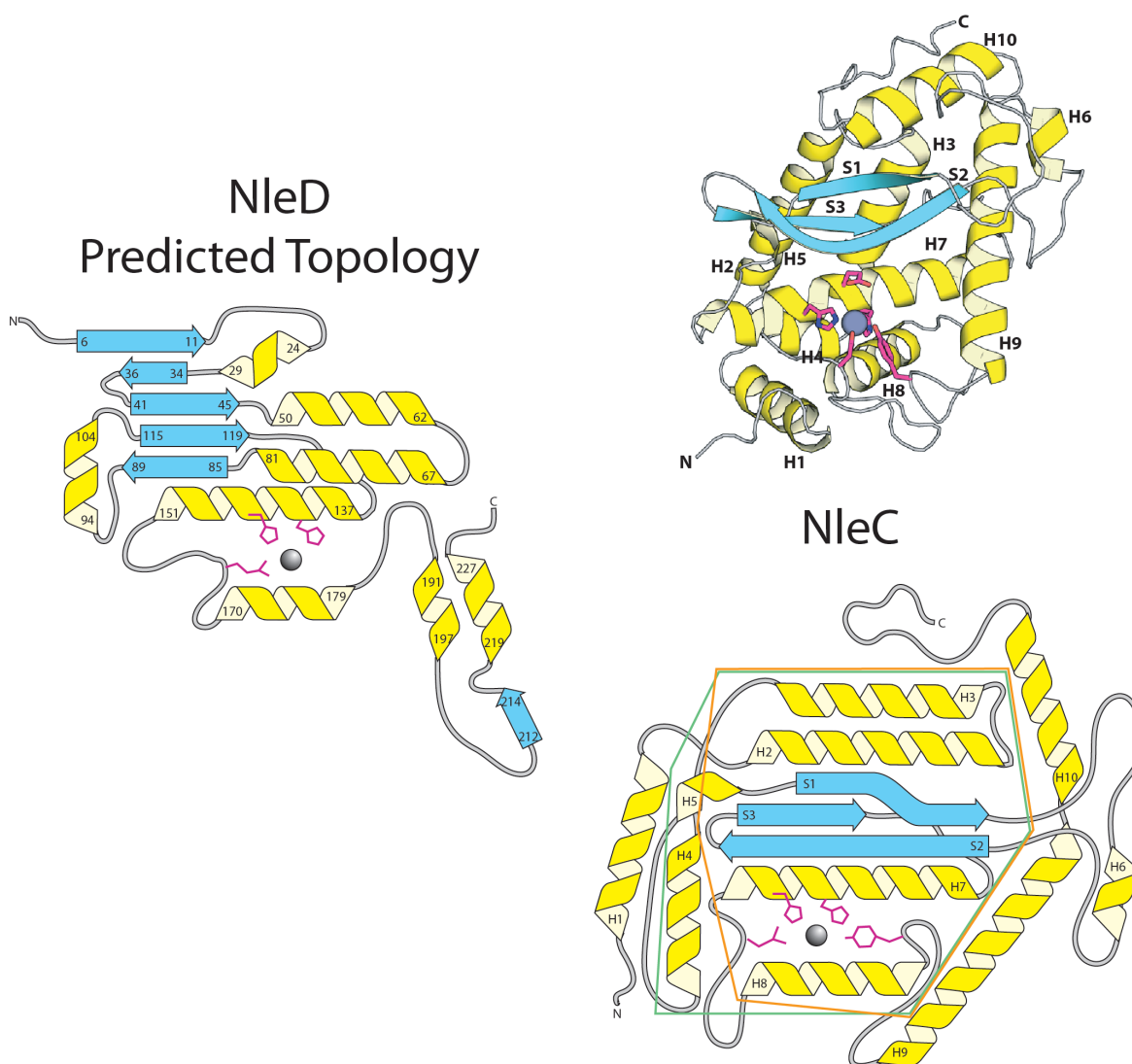


Figure IV-11 Predicted topology of T3SS NleD compared to NleC. Secondary structure analysis using the Jpred server combined with comparisons to Zincins with known structure permitted the construction of model topographic map of NleD.

Function in Translocation for the Amino and Carboxy Termini of NleC

From the SAXS data presented in Chapter III, I have shown that the C-terminus of NleC is folded into a discreet domain separate from the crystallographic core that functions in proteolysis. I also discussed the hypotheses about a role for the NleC C-termini during pathogenesis as a targeting domain and am exploring its effect on the mechanical unfolding of NleC. An alternative hypothesis for the C-terminal domain of NleC concerns preparation for translocation by potentially binding a chaperone. Most T3SS effectors are known to bind a specific T3SS-encoded chaperone that is encoded nearby in the virulence locus^{35,145}. However, in EHEC, many of the effectors found outside of the LEE were acquired by horizontal gene transfer or by phage, as evidenced by the surrounding sequence³⁷. In these cases, where a cognate chaperone does not exist, a more general chaperone encoded across many bacterial species may be preferential. Discussed in Appendix I in more detail, we searched for a chaperone for NleC in conjunction with collaborators at Osaka University, Hilo Yen and Toru Tobe. NleC constructs that I developed and expressed were sent to Yen and Tobe for pulldown experiments with EHEC lysate. After being specifically pulled down by full-length NleC compared to control resin and tag, the general chaperone DnaJ was identified by mass spectrometry. Though subsequent experiments failed to confirm whether binding occurs between NleC and DnaJ *in vitro*, the findings ushered the possibility of a putative role for general chaperones in T3SS pathogenesis. The work by Shames *et al.* resulted in pull downs of multiple Heat Shock proteins, including DnaK, the co-chaperone to DnaJ⁷⁶. The lack of binding *in vitro* between NleC and DnaJ may have resulted from the absence of an additional binding partner, such as DnaK.

The two portions of NleC that have unknown structure may be responsible for binding a chaperone, whether T3SS-specific or general, prior to translocation. The first eighteen residues on the N-terminus, as well as the final fifty residues on the C-terminus were not present in the crystal structure. If either of these fragments are important for efficient translocation by the T3SS, it is expected that truncating NleC will result to less NleC secreted by EHEC, and less virulence in infection assays. The same experiment discussed above, which was performed by our collaborators at Osaka University to determine whether the C-terminal domain affects TNF- α secretion, can be used to draw conclusions about its effect on translocation. A less efficient infection by NleC would result in increased pro-inflammatory cytokine secretion. Because no change in TNF- α secretion was seen between cells infected with full-length NleC and the C-terminal truncation, NleC₂₈₀₋₃₃₀, it is likely that the efficiency of translocation was unaffected.

This same experiment can give insight into the function of the N-terminus of NleC, which may function like the N-termini of other effectors, hypothesized to target the effector to the T3SS and initiate the translocation process. The N-terminus of many effectors has been shown to be essential for translocation by the T3SS and contains unusual amino acid composition and secondary structure characteristics from the remainder of the protein¹⁴⁶⁻¹⁴⁸. These observations have in fact been used to identify putative effectors using computational methods^{147,148}. In analyzing the NleC N-terminus, this putative translocation signal can be identified. NleC is enriched for alanine, proline, and serine residues, with 16% content of each, in the first 25 amino acids compared to the remainder of the protein, that contains 7.5%, 4.3%, and 7.9% respectively. The emergence of this N-terminal secretion signal that is rich in polar residues is suggested to have convergently evolved given the

selective pressure of translocation¹⁴⁷. These polar residues are suggested to facilitate the threading of the protein into the T3SS needle, thus propelling the translocation of the whole protein¹⁴⁶⁻¹⁴⁸.

With this in mind, investigation of the NleC family member conservation reveals that several members do not contain this polar rich region and thus are not likely to be translocated by T3SSs. These sequences are those from the insect and fish pathogens, *Arsenophonus*, *Vibrio*, and *Photobacterium* and also include two mammalian enteropathogenic NleC family members from *E. coli* and *Yersinia*. The fish pathogen NleC members are suggested to be exotoxins, which is consistent with this proposal. It is unknown whether *Arsenophonus* NleC is translocated *via* the T3SS, as it has not been studied in depth. Finally, the only two enteropathogenic NleC proteins in my multiple sequence alignment that lack the N-terminal putative signal sequence are truncated further into the protein, making it likely they are incompletely transferred genes and may not be functional in during pathogenesis. From this analysis, it is apparent that the T3SS effector N-terminal translocation signal can be utilized in NleC to determine whether the proteins are functional in T3SS-dependent pathogenesis. Any secretion or infection assays that disrupt this initial 25-residue region in Sakai strain NleC are expected to significantly hinder efficient translocation and would be a future experiment to validate these hypotheses.

Summary

The structure of NleC demonstrates how bacteria can utilize and alter ancient, phylogenetically-distributed protein architecture to accomplish their unique goals¹⁰². This is a complementary approach to the convergent strategy seen in many T3SS effectors, such

as the *Salmonella* effector SifA^{39,45,149}. Utilizing a well-known architecture, I propose that NleC evolved additional function in acting as a DNA mimic, while maintaining the ability to proteolyze protein. This especially nefarious and functional combination enables NleC to bind NFκB transcription factors and not only inhibit them from binding DNA physically, but also by cleaving the DNA-binding loop. This would be the first example of a DNA mimic among T3SS effectors. The DNA-mimicry ability may also give NleC the ability to bind other DNA-binding proteins, such as p300, though it is unknown whether this has a function during pathogenesis. While the NleC family of proteins continues to be studied, it will especially interesting to determine whether the fish and insect pathogenic forms also exhibit DNA mimicry, as their surfaces are variable compared to enteropathogenic NleC. A variety of structural, biochemical and microbiology assays are required to fully characterize this interesting protein family. Given the diversity of T3SS effectors, with each functional and structural characterization, we gain further insight in how these pathogenic bacteria disable and hijack proteins, cells and hosts.

Appendix A. NleC and Putative Chaperones

Chaperones are required for many effectors to be efficiently secreted, both providing specificity and directionality to the T3SS, as well as preventing aggregation in the bacterial cell^{28,29}. Because all known effectors that bind chaperones do so with their amino-terminal tails, we hypothesized that NleC may similarly bind an effector with the N-terminal portion that is not present in the crystal structure. The catalytic activity of effectors is unaffected by chaperone binding, and structures solved of effectors with and without their chaperones confirm this, showing that the chaperone binding occurs in a domain separate from the catalytic domain³⁴.

To uncover a putative chaperone for NleC, we worked with collaborators at Osaka University, Toru Tobe and Hilo Yen, who have executed infection and secretion experiments with wild type EHEC Sakai strain, as well as engineered EPEC and a reconstituted T3SS-competent lab strain *E. coli*⁷¹. I prepared full-length NleC with a C-terminal Intein-CBD (chitin-binding domain) tag for use as bait in pulldown experiments. After expression, the NleC-Intein-CBD was loaded onto a chitin-column and washed extensively. An unrelated protein (GCAP1) that had the same carboxy-terminal Intein-CBD tag, as well as the Intein-CBD tag alone, were also expressed for use as controls. The resins with proteins bound were shipped to the collaborators in Japan. They lysed wild type EHEC Sakai strain cells and applied this lysate to the NleC-loaded resin. Through three replicates of this pulldown with full-length NleC and the unrelated protein, they identified DnaJ and Skp as putative chaperones but no T3SS-specific chaperone was identified (Figure A-1).

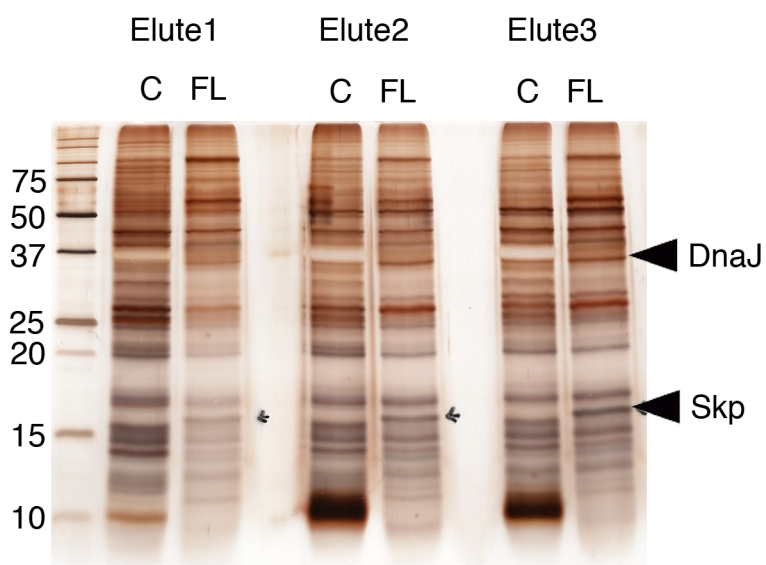


Figure A-1 NleC pulldown of EHEC proteins. Performed by Hilo Yen and Toru Tobe of Osaka University. C designates elution from control resin, whereas FL designates full-length NleC-bound resin. NleC-Intein can be seen around 60 kDa. The indicated bands were analyzed by mass spectrometry and returned DnaJ and Skp peptides to high confidence.

Upon learning of the results of the pulldown from Yen and Tobe, we hypothesized that perhaps a general chaperone like DnaJ or DnaK could be utilized for T3SS in the absence of a specific chaperone for effectors located outside of the LEE. Because many of the non-LEE effectors are shared *via* horizontal gene transfer and phage infection, utilizing a general chaperone would negate the requirement that a cognate chaperone be transferred as well³⁷.

To test for binding to DnaJ by NleC, a pulldown experiment was performed. DnaJ was cloned into a plasmid that expresses a 10xHis-SUMO tag at the N-terminus of the protein. This protein was expressed, loaded onto a Ni-NTA column, and washed extensively. Purified NleC was then added to the column. The control conditions used were

resin with only DnaJ, and resin with NleC added alone. After addition of NleC, the resin was further washed and then eluted. Unfortunately, NleC was not retained on the column containing DnaJ, illustrating that DnaJ does not strongly associate with NleC *in vitro* under these conditions. Isolation of the complex was also attempted by size exclusion chromatography (SEC). NleC was incubated with DnaJ in equimolar ratios before loading onto the size exclusion column. The resultant peaks were compared to chromatograms of NleC and DnaJ alone on the same column. This also illustrated that NleC and DnaJ do not bind strongly under the conditions tested (Figure A-2).

These results do not rule out the possibility that NleC and DnaJ are indeed binding partners. Additional proteins may be necessary to mediate the interaction, such as DnaK, which is known to be involved in secretion processes^{150,151}. A plasmid containing both DnaJ

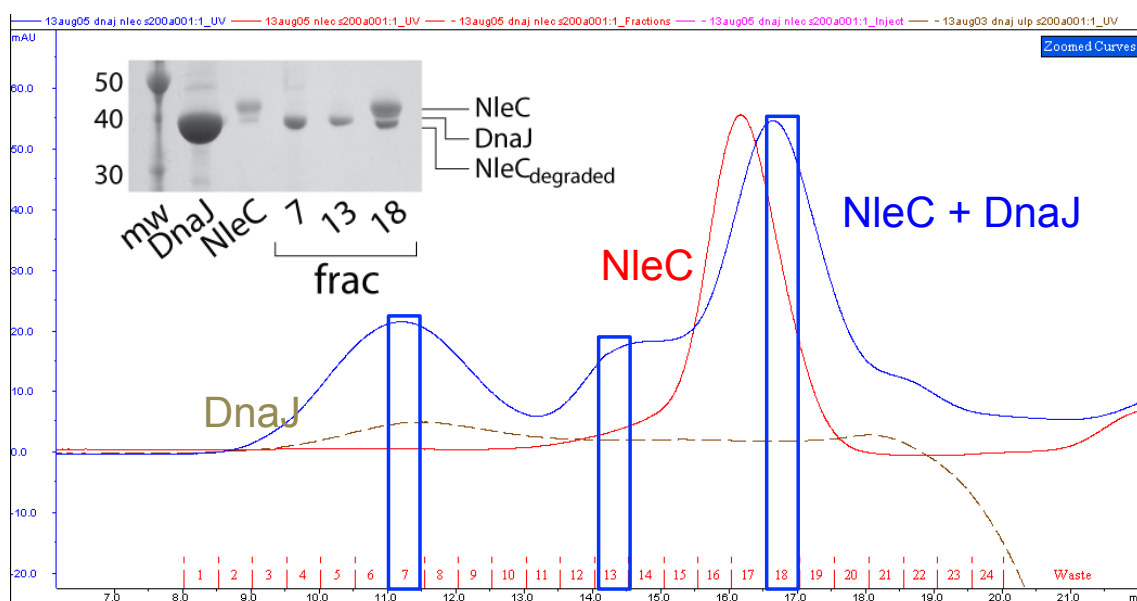


Figure A-2 NleC and DnaJ do not form a complex. SEC was used to try to isolate a complex between DnaJ and NleC. DnaJ elutes in two peaks because it self-associates as a homodimer. Both DnaJ and NleC overlay with the peaks from each protein alone. Inset shows SDS-PAGE depicting the fractions from the NleC and DnaJ injection. NleC runs in two bands from time-dependent degradation.

and DnaK has been obtained and the next step will be to attempt *in vitro* complex formation with both DnaJ and DnaK, and NleC.

There is also the possibility that the pulldown of DnaJ in the original results were a result of misfolded NleC on the resin, as generalized chaperones are often obtained as false positives. This would mean that the result has no physiological relevance.

Methods

All plasmids used for this study are listed in Table A-1. NleC truncations were developed using the crystal structure of NleC and hydrophobicity plots to minimize hydrophobic tail exposure. NleC₁₋₂₉, NleC₁₋₄₃, and NleC₁₋₇₀ were PCR amplified from pMS692, the wildtype NleC plasmid with primers encoding NdeI and XmaI sites. pMS692 was then digested with NdeI and XmaI, and ligated with the NleC inserts. All NleC plasmids were expressed and purified as described in Chapter II. For preparation of NleC-bound chitin resin, the purification was halted after washing the resin extensively. The resin was then sent to Osaka University for pulldown experiments with EHEC lysate. EHEC O157 VT strain which possesses multi copies of *pchA* was grown for the lysate. This strain produces high levels of T3SS-related genes and so if a T3SS-specific chaperone binds NleC, it would be detectable. The pulldown was performed in triplicate and the proteins eluted with 25 mM Tris pH 8 and 600 mM NaCl. Elutions were visualized by SDS-PAGE and silver stained. Bands that appear from NleC-bound resin but not from control resin were excised, trypsin digested and analyzed by TOF-MS/MS. The other proteins identified from the DnaJ band by mass spectrometry were Lpxd, Ptnab, OmpC, RpoA, Dpo3b, Ftsz, OmpF, Adhp, Asna, Phol, Cfa, Yfgl, Hflk, and Ruvb. Most of these are not cytosolic and would not be physiologically relevant. Skp was the only protein identified from its band.

Table A-1 Plasmids used in the study of NleC and putative chaperones.

Plasmid	Gene	Residues	Organism	Gene Source	Tag	Background	Primers	Resistance
pMS692	NleC	1-330	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	ECs0847_F1/ECs0847_R2	Amp
pMS1017	NleC	1-29	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	ECs0847_F1/L913a	Amp
pMS1018	NleC	1-43	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	ECs0847_F1/L913b	Amp
pMS1019	NleC	1-70	EHEC str. Sakai	EHEC str. Sakai DNA	C-term Intein-CBD	pTYB2	ECs0847_F1/L913c	Amp
pMS1073	DnaJ	1-376	Human	Dave McKay - CU Boi	6xHis	pET		Amp
pMS1117	DnaJ	1-376	Human	pMS1073	10x-His-SUMO	pET24d	U1064/L1064ab	Kan

The plasmid for DnaJ was obtained from Dave McKay at University of Colorado at Boulder. DnaJ was PCR amplified out of this plasmid with primers encoding SfoI and KpnI restriction sites. The plasmid pMS984 that encodes an N-terminal 10xHis-SUMO tag was then digested with SfoI and KpnI, and ligated overnight with T4 ligase. The resulting plasmid was transformed into BL21 (DE3) cells. A single colony from a LB-Kan was used to inoculate 1 L of LB with 50 µg/mL Kanamycin. The cells were grown until an OD of 0.6 at 37 °C then incubated on ice for ten minutes. The cells were then induced with 1mM IPTG. This was then grown overnight at 20 °C for 16 hours. The cells were harvested by centrifugation at 4,000xG and resuspended in Buffer A [25 mM Tris pH 8, 150 mM NaCl, 1 mM TCEP] supplemented with a protease inhibitor cocktail (Complete EDTA-free, Roche). This was sonicated to lyse the cells and centrifuged at 15,000G to pellet cell debris. The soluble cellular fraction was applied to 2 mL of Ni-NTA resin and washed with Buffer A. After ten column volumes of washing, the resin was split into two equal aliquots. One had 2 mg of NleC added to it, whereas the other was used as a control. Clean resin with no DnaJ added was used as a control of nonspecific NleC binding. Ten column volumes were used to wash the resins after this addition. The three samples were eluted from the resin with Buffer B [Buffer A + 250 mM imidazole]. The washes and elutions were analyzed by SDS-PAGE.

The concentration of DnaJ from the DnaJ-only resin was determined and incubated with equimolar amount of NleC. After an hour, this was injected onto a size exclusion column (Super 75 26/60, GE Healthcare) and compared to runs with DnaJ and NleC alone. The fractions were compared by SDS-PAGE as well.

Appendix B. Map, Tir and CstT

An Exploration of Proteins on the LEE

A handful of EHEC T3SS effectors are especially well known and studied. These effectors are found in the most evident genomic signature of Type III Secretion for A/E pathogens, the locus of enterocyte effacement (LEE), which contains the machinery of the T3SS, as well as these important effectors and chaperones. Two of these effectors are called Tir (translocated intimin receptor) and Map (mitochondrial associated protein).

The function of Tir was known before its identity was discovered. It was recognized that a bacterial cell-surface protein bound a receptor on the host cell membrane in order to establish the tight and intimate relationship between bacterial cell and host cell, but the identity of the receptor was unknown. Upon inspection, it was determined that instead of using a eukaryotic receptor, the A/E pathogens translocate a special receptor that inserts into the host membrane. This receptor is called Tir⁶¹.

With Map, however, the function was only discovered after the identity of the protein was known, and much postulation. Map is a member of a group of proteins called WxxxE effectors, which are named for a conserved sequence found in all members. These proteins were discovered to stimulate small GTPase pathways by an unknown mechanism. The predominant hypothesis at the time was that these proteins are able to mimic active small GTPases themselves, but lacking the ability to bind ATP, thus causing a constitutively active pathway downstream^{65,66}. The true nature of the WxxxE proteins was discovered only with the structural determination of a Map-cdc42 complex. This structure showed

unequivocally that Map was a guanidine nucleotide exchange factor (GEF) of the SopE T3SS effector family, though with no sequence similarity to either eukaryotic GEF proteins or SopE⁴⁷.

Biologically relevant structure of T3SS multi-effector chaperone CesT

What these effector proteins share, besides being in the LEE and being important for infection, is a bacterial chaperone that is essential for efficient secretion by the T3SS. This chaperone is called CesT, and it is known to bind and escort the effectors to the T3SS^{31,152}. For such an important function, we know very little about how these small, homo-dimeric chaperones work. Among T3SS chaperones, some bind only one effector, whereas others, like CesT, bind multiple effectors^{28,34}. Though the sequence similarity is quite low between different T3SS chaperones, they have a shared structure and overlay well (Figure B-1). From structures of these chaperones with their binding partners, we know that they all share a similar binding mode, with the N-terminus of the binding partner wrapping around the homodimeric chaperone (Figure B-2)³⁴. However, sequence analysis of the binding partners is unable to locate a conserved chaperone-binding domain, even among proteins that bind to the same chaperone. The apparent lack of consensus sequence led researchers to search for a structural basis for recognition by the chaperones. The only structural feature shared by all the chaperone-bound proteins is a β -strand, extending the chaperones β -sheet. This observation makes it likely that chaperones bind their cognate proteins by β -augmentation with sequences of residues that have β -strand character³⁴. Studying a multi-effector chaperone, in complex with two different effectors, will help elucidate how these chaperones recognize diverse effector targets. It is also important to note that the N-

terminal signal sequence of T3SS effectors discussed in Chapter IV does not appear to be necessary for binding to chaperones, and the two actually reside in distinct regions in sequence^{35,153,154}.

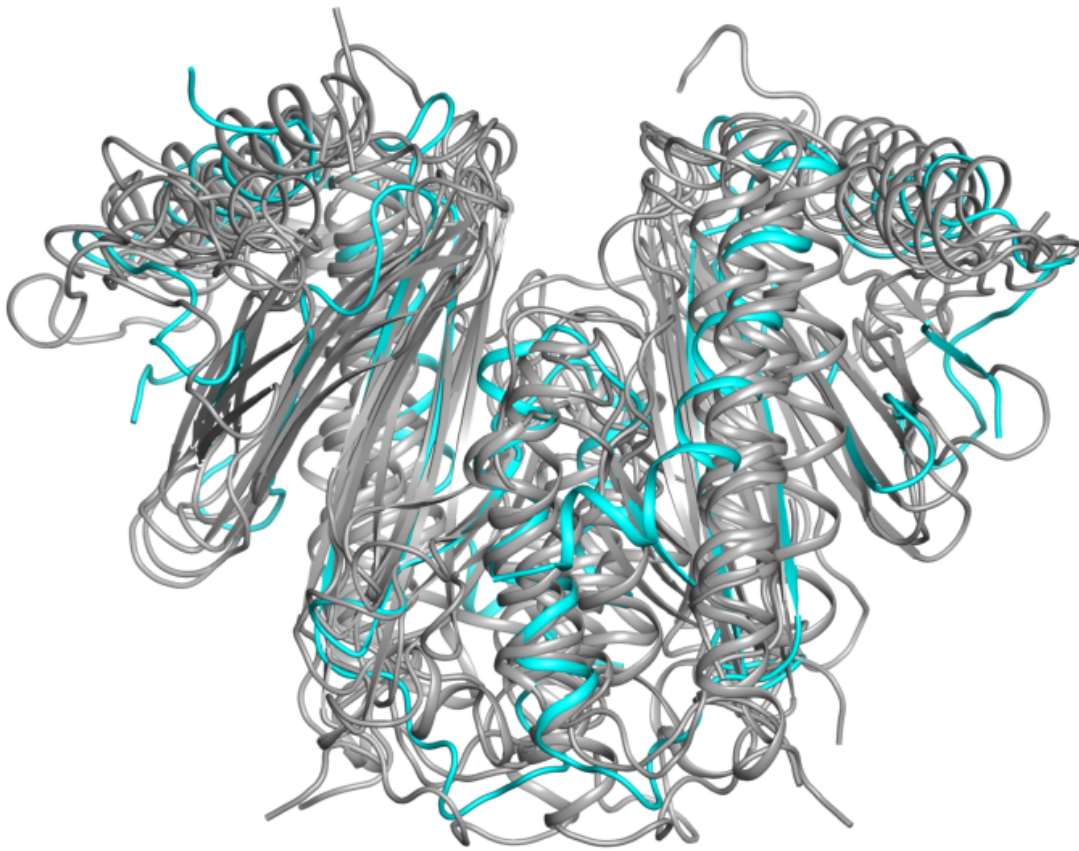


Figure B-1 Overlay of T3SS chaperone structures. Shown in aqua in the structure I solved of CesT. Other chaperone PDB IDs:3KXY, 2FM8, 1XKP, 1TTW, 1L2W, 1JY0

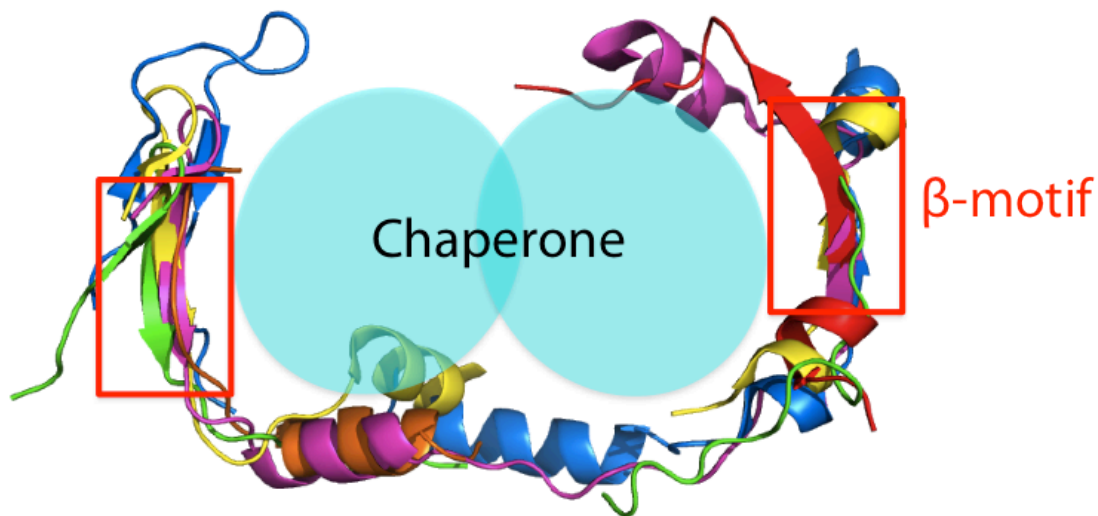


Figure B-2 Effector binding mode. Of the structures that have been solved between T3SS effectors and their cognate chaperones in complex, all of them form an extended conformation, wrapping around the exterior of the chaperone dimer. All of the effector structures in these complexes take a β -strand character, forming β -motifs which bind by β -augmentation.

To this end, I pursued the structural characterization of the CesT-Map complex. Map was expressed as 6xHis-MBP-Map fusion protein. The cleavage of the MBP tag resulted in unstable Map, which was prone to aggregating, especially at higher concentrations. To obtain the CesT-Map complex, 6xHis-CesT was purified on Ni-NTA resin, after which the tag was removed. Upon expression of 6xHis-MBP-Map, it was loaded on a Ni-NTA column and washed. To this resin-bound Map fusion protein, purified tagless CesT was added, which bound to the column in 2:1 CesT-Map ratio. The column was further washed, and then eluted with imidazole. TEV (Tobacco Etch Virus) protease was added to the eluted fraction to cleave the 6xHis-MBP tag and this was dialyzed overnight to remove the imidazole. After a final SEC step, the purified CesT-Map complex could be used for crystallography studies.

After incubation under various crystallography conditions, several crystals were obtained from the CesT-Map complex screens. Upon screening these crystals at the home source and at the synchrotron, however, it was apparent that the unit cell was too small to contain both the CesT and Map. A domain swapped structure of CesT had been solved previously and was used to construct a model of a non-domain swapped monomer. Using this monomer, I solved the crystal structure using molecular replacement, proving that the crystal contained only CesT monomer per asymmetric unit, and no Map (Figure B-3)(Table B-1). SDS-PAGE analysis of crystal drops versus crystals themselves validated this result. This structure reveals a non-domain swapped, physiologically relevant CesT dimer, whose existence had been supported by previously published NMR data. All other known structures of T3SS dimers are non-domain swapped and obtaining this structure verifies that CesT does not exist in a different form from others T3SS chaperones.

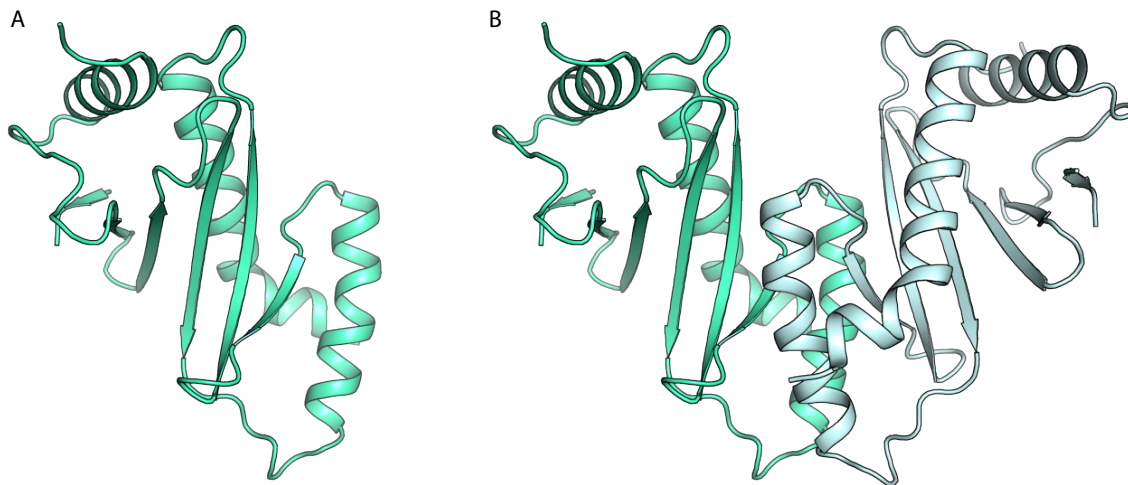


Figure B-3 CesT crystal structure. A) Because the CesT homodimer is symmetrical, the asymmetric unit contained only a CesT monomer, with the dimerization interaction appearing across a crystallographic symmetry axis. B) The CesT dimer with a crystallographic symmetry mate making up the other monomer.

Data Collection			
Space group	P3221		
Cell dimensions			
a (Å)	66.058	α (°)	90
b (Å)	66.058	β (°)	90
c (Å)	70.692	γ (°)	120
Refinement			
Resolution (Å)	44-1.9		
No. reflections			
Rwork/Rfree	19.7/23.7		
No. atoms			
Protein	1064		
Ligand/ion	0		
Water	44		
R.M.S. deviations			
Bond lengths (Å)	0.007		
Bond angles (°)	1.079		

Table B-1 CesT Crystallography Statistics

Using the non-domain swapped structure of CesT, I hypothesized how CesT may identify the different effectors it is responsible for binding. As mentioned above, Lilic *et al.* identified a β -motif that is shared between all effectors binding chaperones in known structures³⁴. This involves β -augmentation of the chaperone-binding domain of the effector by the chaperone extended β -sheet, as demonstrated earlier with experimentally determined structures of chaperone-effector complexes (Figure B-2). CesT has an interesting β -sheet in the sense that it appears to have six strands per monomer, whereas most T3SS chaperones have five. Additionally, this sixth, outermost strand appears to be quite variable. In my CesT structure, this sixth strand appears to belong to a neighboring monomer in the lattice, bending the H5 helix from that obtained in the previous domain swapped structure reported by Luo *et al.*(Figure B-4)¹⁵⁵. It is possible that this final sixth strand can directly bind effector proteins, or can move to accommodate effectors binding

the fifth strand. This would provide some variability in the binding mode of CesT and may allow it to bind multiple effectors.

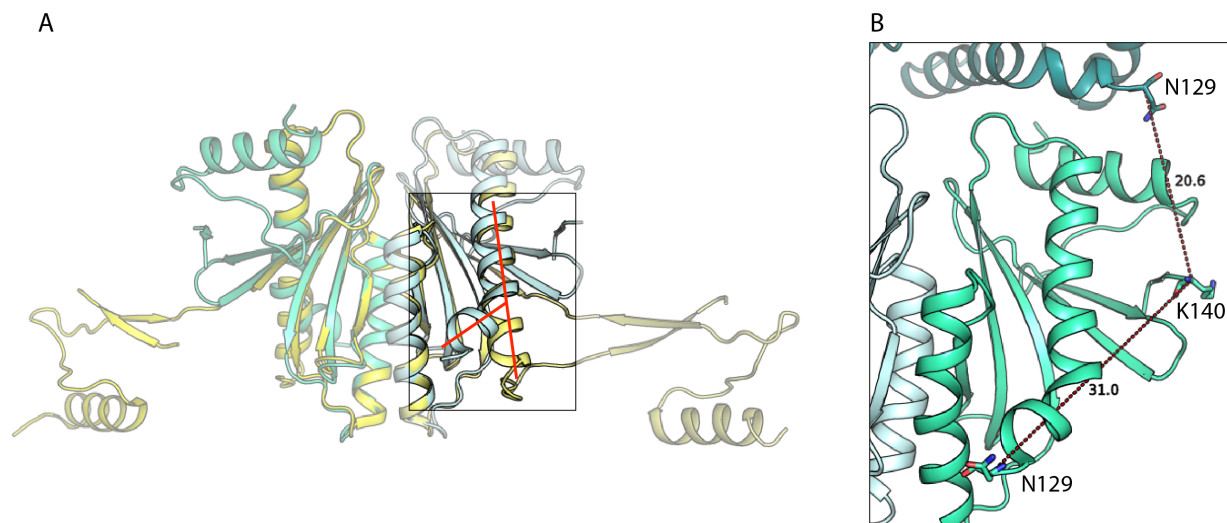


Figure B-4 CesT altered helix and sixth strand. A) Overlays of previous domain-swapped CesT structure and the structure solved here. Aqua: Our CesT structure. Yellow: 1KE3. The C-terminal helix is bent from the conformation in the previous structure. The portions of the previous structure that do not align with our structure are domain swapped with neighbor monomers in the crystal lattice. **B)** The sixth strand is fairly distal from the end of the C-terminal helix, especially as it is bent from the previous crystal structure. This makes it more probable that the sixth strand is donated by the neighboring monomer in the crystal lattice.

The structures of the amino terminal domains of Tir and Map have not been determined. Therefore, I used secondary structure prediction approaches to identify putative β -strands that may be responsible for binding to CesT. These fragments were cloned into 10xHis-SUMO N-terminally tagged plasmids. The Map fragments did not express well and were not soluble, but the Tir fragments expressed well and solubly. These fragments were shown to bind CesT by a Ni-NTA pulldown, like that described for Map above. Nuclear magnetic resonance (NMR) can be used to study the interactions between CesT and its effectors, and the studies would be facilitated by the published resonance

assignments for CesT^{156,157}. N¹⁵-labeled CesT was expressed and purified into the buffer described in the manuscript reporting the resonance assignments. Initial Nitrogen Hydrogen correlation spectra of CesT showed a substantial amount of overlap between peaks we obtained and those that were published. With the expectation that binding of an effector fragment would change the chemical shifts for the residues involved in the interaction, I titrated a non-labeled N-terminal Tir fragment into the labeled CesT sample. However, the peaks began broadening and disappearing instead of shifting (Figure B-5 and Figure B-6). This suggests that instead of a few residues being affected by the binding of the Tir fragments, the entire molecule was altered, possibly because of aggregation. To test this possibility, SEC was used to assess whether the CesT-Tir NMR sample was aggregated. CesT and Tir eluted at the same retention time as each protein individually, suggesting that irreversible aggregation was not an issue. However, we cannot rule out that reversible aggregation occurs at the protein concentrations required to collect NMR data.

The future directions for this project are well defined. Because of the strong expression and solubility of CesT and the Tir fragments, complexes can be made and used to set up crystallization trays. If a CesT-Map complex cannot be obtained according to methods for nonlabeled CesT, crosslinkers can be used to lock the complex together before setting up crystal trays. Whereas the affinity was high enough to survive purification by SEC, the complex dissociated under crystallization conditions, preventing formation of a crystal of the complex. Trapping the complex may permit its crystallization. Other effectors that are known to bind to CesT may also be used for NMR or crystallization attempts. These effectors are EspF, EspH, NleA, EspG, NleG, NleH, NleH2, and EspZ.

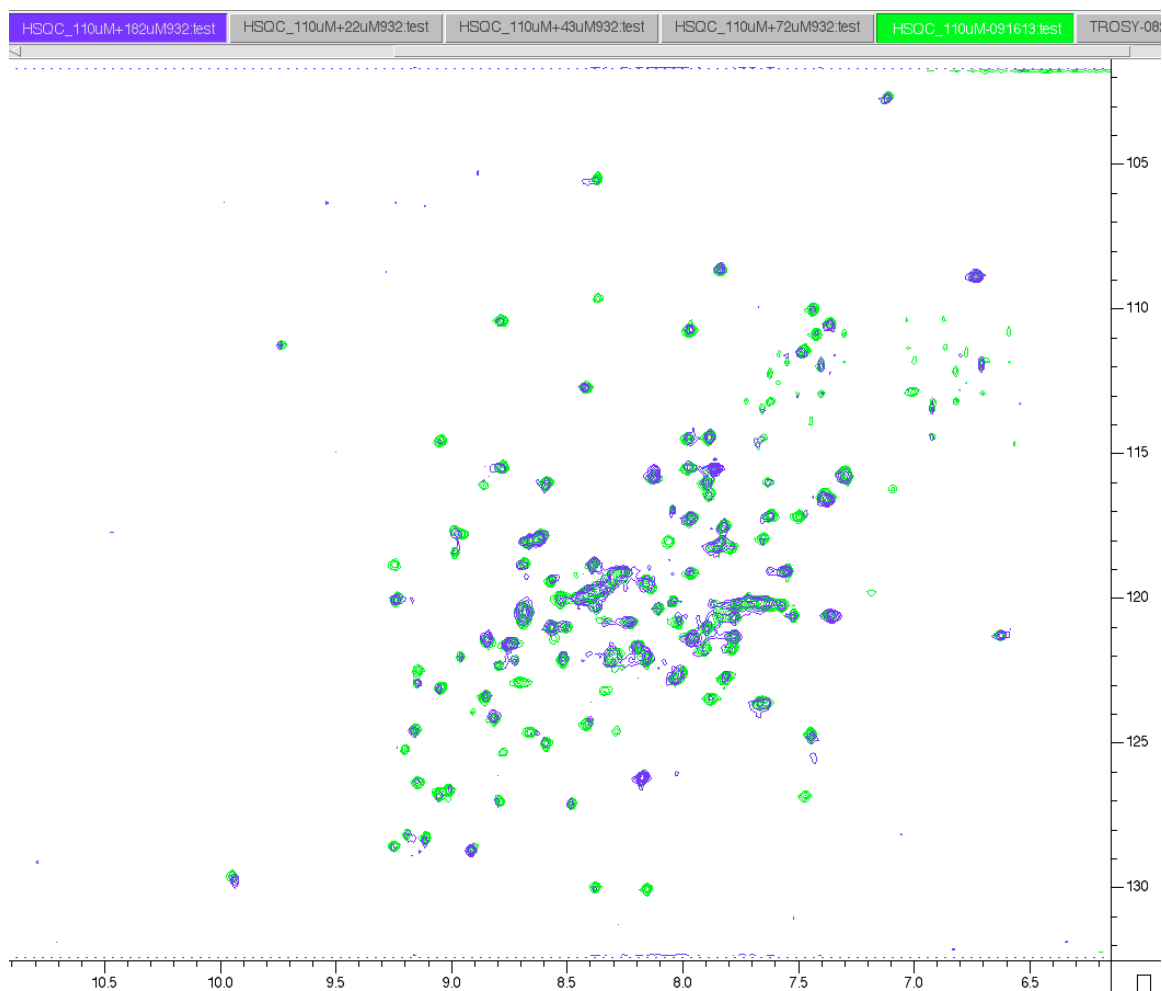


Figure B-5 HSQC overlap of free N^{15} -CesT and N^{15} -CesT titrated with 2:1 molar ratio of Tir₁₋₆₈ (pMS 932). Obtained at the 900 MHz NMR maintained by University of Colorado at Anschutz Medical Campus, visualized with CCPN Analysis.

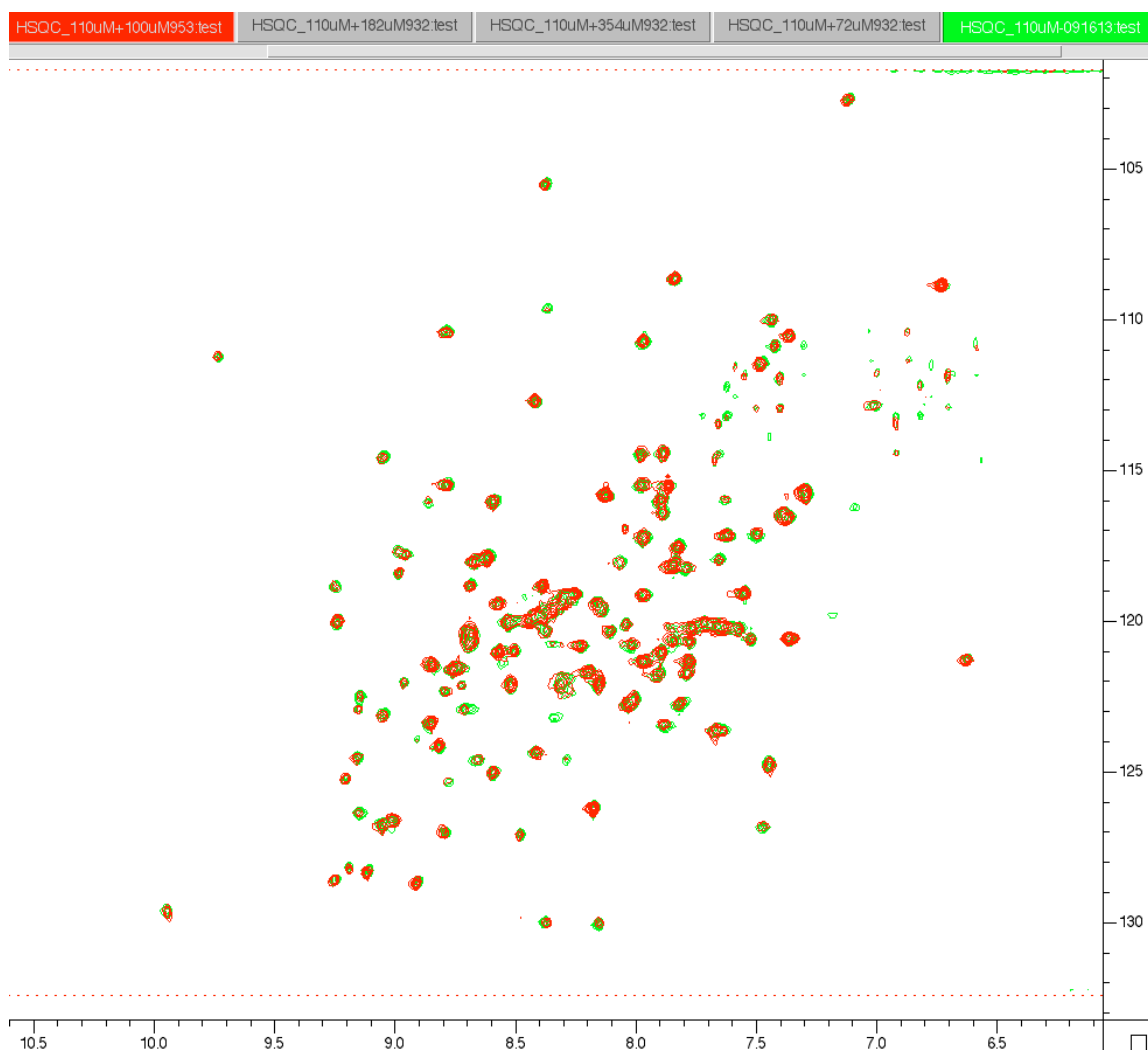


Figure B-6 HSQC overlap of free N^{15} -CesT and N^{15} -CesT titrated with 1:1 molar ratio of Tir₁₋₉₃ (pMS 953). Obtained at the 900 MHz NMR maintained by University of Colorado at Anschutz Medical Campus, visualized with CCPN Analysis.

Methods

All plasmids used for this study are listed in Table B-2. CesT was amplified from EHEC Sakai strain genomic DNA with primers containing NcoI and XmaI restriction sites. These sites allowed the ligation into pMS173, a pET28a+ vector that was modified to add an XmaI site plus stop codon in the multiple cloning site as well as to remove an internal XmaI site in the kanamycin resistance gene. pMS173 was also digested with NcoI and XmaI,

encoding a fusion protein of CesT with an amino-terminal 6xHis tag. This plasmid was sequenced and transformed into BL21 (DE3) *E. coli* cells, induced after cells reached $OD_{600}=0.6$, and grown for three hours at 37 C. Expression and solubility tests showed that this CesT construct was highly expressed at 37 C, compared to low expression levels at 20 C. After expression, the cells were harvested by centrifugation at 4000xG for 20 minutes, and resuspended in Buffer A (25 mM Tris pH 8, 300 mM NaCl, 10% glycerol, 10 mM imidazole, 5 mM β -mercaptoethanol) with protease inhibitors (Complete EDTA-free, Roche). This slurry was sonicated at 40% intensity for 48 cycles of {10 s on: 50 s off} to break the cells. Cell debris was then pelleted by centrifugation at 15,000xG for 20 minutes. The soluble fraction was loaded onto 1/4 volume Ni-NTA resin. After the incubation, the resin is washed in column with 24x the resin volume with Buffer B (25 mM Tris pH 8, 300 mM NaCl, 10% glycerol, 25 mM imidazole, 5 mM β -mercaptoethanol). CesT was then eluted with Buffer C (25 mM Tris pH 8, 300 mM NaCl, 10% glycerol, 250 mM imidazole, 5 mM β -mercaptoethanol). After elution, 1/500 w/w TEV was added to the sample plus 5 mM DTT. Because the CesT 6xHis tag is difficult to cleave, after three days, more TEV (1/500 w/w) and DTT (5mM) was added and allowed to incubate for a further three days. The CesT was then concentrated and purified by SEC (SEC) (Superdex 75 26/60, GE Healthcare).

CesT also had to be prepared for analysis by NMR. The assignments have been deposited into BioMagResBank (<http://www.bmrb.wisc.edu>) under accession number 6451. We prepared a N^{15} -labeled, deuterated CesT by growing the above 6xHis-CesT plasmid in BL21 (DE3) cells in minimal media supplemented with N^{15} -glucose. The Ni-NTA purification, cleavage of the His-tag and SEC purification were the same as described above for the unlabeled CesT, except the SEC was done in Buffer C (50 mM potassium phosphate

pH 6.8, 100 mM NaCl, 5 mM DTT, and 0.5 mM EDTA). After concentration, 10% D₂O was added to the sample.

The Tir fragments (residues 1-68 and 1-93) were similarly amplified from EHEC Sakai strain genomic DNA with primers containing KpnI and SfoI restriction sites. This allowed ligation into a plasmid encoding 10xHis-SUMO tag at the amion-terminus. These constructs were expressed in BL21 (DE3) *E. coli* cells after transformation by ice shocking and inducing after the cells reached OD₆₀₀= 0.4. These cells were then grown at 20 C overnight (~16 hours). The protein was then harvested and purified exactly as above with CesT. After purification from the Ni-NTA column, however, a fraction was concentrated with the 10xHis-SUMO tag still attached to the N-terminus, and another fraction had the tag removed by incubation with ULP, the SUMO-specific protease. This fraction, while bound to the Ni-NTA bead, had 6xHis-CesT added to it in equimolar quantities. This was then eluted. ULP and TEV protease were added at 1/100 V/V with 10 mM DTT. This reaction went for three days and the cleaved SUMO, His tag and proteases (both His-tagged) were removed by flowing the sample over Ni-NTA beads. Unfortunately, there was too little of this complexed sample for analysis by NMR. The 10xHis-SUMO tagged Tir fragments, however, were purified by SEC, as done with CesT above. These samples were concentrated to ~1mM.

A NMR spectra of CesT was obtained with 1mM N¹⁵-labeled CesT in 10% D₂O. Initially, the His tag was not completely cleaved off the CesT, and so this contributed to many peaks centrally located on the HSQC (Heteronuclear Single Quantum Coherence). After complete cleavage of the His tag was obtained, the HSQC was repeated and many of the peaks could be identified from the published assignments. We then diluted the sample

to a working concentration of $\sim 100\mu\text{M}$, verifying that all peaks were still present at this lower concentration. Then we titrated one of the Tir fragments, Tir₍₁₋₉₃₎, into the CesT sample at ratios of CesT:Tir at 10:1, 5:1, 5:2, 10:7, 1:1, 2:1 and 5:1. Tir₍₁₋₆₈₎ fragment was only titrated in at 1:1 and 1:2 CesT:Tir. After analysis by NMR, these CesT-Tir samples were run on SEC to check for aggregation.

Table B-2 Plasmids used for the study of CesT, Map and Tir.

Plasmid	Gene	Residues	Organism	Gene Source	Tag	Background	Primers	Resistance
pMS439*	CesT	1-156	EHEC str. Sakai	EHEC str. Sakai DNA	6xHis	pET24d	ECs4560_F1/ECs4560_R1	Kan
pMS354*	Map	1-203	EHEC str. Sakai	EHEC str. Sakai DNA	6xHis	pET24d	U199/L199	Kan
pMS357*	Map	1-203	EHEC str. Sakai	EHEC str. Sakai DNA	6xHis-MBP	pMAL	U199/L199	Amp
pMS498*	Map	1-203	EHEC str. Sakai	EHEC str. Sakai DNA	6xHis-MBP	pMAL	U505/L505	Amp
pMS594	Map	1-203	EHEC str. Sakai	pMS354	6xHis	pET24d	U199/L199	Kan
	Ebp50	1-358	Human				U522b/L522	
pMS595	Map	1-203	EHEC str. Sakai	pMS354	6xHis	pET24d	U199/L199	Kan
	Ebp50 (PDZ1)	1-138	Human				U522b/L523	
pMS596	Map	1-203	EHEC str. Sakai	pMS354	6xHis	pET24d	U199/L199	Kan
	Ebp50 (PDZ2)	139-248	Human				U524/L524	
pMS691	Map	1-203	EHEC str. Sakai	pMS354	6xHis	pET24d	U199/L199	Kan
	CesD	1-151	EHEC str. Sakai	EHEC str. Sakai DNA				
pMS818**	Map	1-52	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U820/L820	Kan
pMS819**	Map	1-52	EHEC str. Sakai	pMS818	10xHis-SUMO-Map-Intein	pET24d	U822/L822	Kan
pMS820**	Map	1-81	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U820/L821	Kan
pMS819**	Map	1-81	EHEC str. Sakai	pMS818	10xHis-SUMO-Map-Intein	pET24d	U820/L821	Kan
pMS1007	Map	1-73	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U870/L870ab	Kan
pMS1008	Map	1-96	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U870/L871ab	Kan
pMS449	Ebp50	1-138	Human	Anthony Bretscher, Cornell	3x-GST	pGEX		Amp
pMS450	Ebp50	139-248	Human	Anthony Bretscher, Cornell	3x-GST	pGEX		Amp
pMS451	Ebp50	1-358	Human	Anthony Bretscher, Cornell	3x-GST	pGEX		Amp
pMS463*	Ebp50	1-138	Human	pMS449	6xHis	pET24d	PDZ1_F/PDZ1_R	Kan
pMS464*	Ebp50	139-248	Human	pMS450	6xHis	pET24d	PDZ2_F/PDZ2_R	Kan
pMS465*	Ebp50	1-138	Human	pMS449	no tag	pET24d	PDZ1_F/PDZ1_R	Kan
pMS466*	Ebp50	139-248	Human	pMS450	no tag	pET24d	PDZ2_F/PDZ2_R	Kan
pMS485*	Ebp50	1-358	Human	pMS451	6xHis	pET24d	U481/L502	Kan
pMS844	cdc42	1-210	Human	Natalie Ahn, CU-Boulder	6xHis	pET24d	U589/L589	Kan
pMS848	Tir	1-559	EHEC str. Sakai	EHEC str. Sakai DNA	6xHis	pET24d	U776b/L776	Kan
pMS932**	Tir	1-93	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U831/L831/L832	Amp
pMS953**	Tir	1-68	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U851/L851	Amp
pMS954**	Tir	24-68	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U852/L852	Amp
pMS955**	Tir	24-68	EHEC str. Sakai	EHEC str. Sakai DNA	10xHis-SUMO	pET24d	U853/L853	Amp

* Prepared by Amy Dear

** Prepared by Sandra Metzner

Map nucleates multimeric substrate complexes in host cells

When the mechanism of the LEE-effector Map was still unknown, it was a very attractive target for structural study. The WxxxE effectors were known to activate the small

GTPase pathways, but as stated above, it was unknown whether they achieved this by directly binding the small GTPases or by acting on their downstream partners^{65,66}. Though the mechanism of Map was still unknown, a binding partner had been identified. A targeting sequence on the C-terminus of Map was found to be recognized by the PDZ1 domain of NHERF1/Ebp50. This would target Map to apical membrane of epithelial cells, beneath where EHEC cells have formed their intimate connection with T3SS and pili⁶⁵.

A post-doctoral researcher in our lab, Amy Dear, set out to determine the structure of Map. Map was initially cloned into pMS173 that encodes a 6xHis N-terminal tail, though upon expression, it was found to be insoluble. Map was then cloned into a pMAL plasmid that expressed an N-terminal MBP (maltose binding protein) tag. This tag rendered the fusion protein soluble and purified Map could be obtained after cleavage of MBP by TEV. Early crystallization screens showed excessive precipitation, however. Because many effectors are more stable upon the addition of a binding partner, Map binding partners CesT, the PDZ1 domain and Ebp50 His-tagged constructs were also cloned. However, binding between either the PDZ1 domain alone or to Ebp50 with Map could not be obtained. The nonnative C-terminal tail with an additional -PG, a result of cloning, was deemed to be responsible for abrogating binding. When Map was recloned into the pMAL plasmid maintaining the native C-terminus, binding was achieved with both the PDZ1 domain and Ebp50. At this point, I took over the project. Ebp50 was found to be unstable upon expression and easily degraded, resulting in a non-homogenous purified product. The PDZ1 domain was more easily purified and was used to create MBP-Map + PDZ complexes. A protocol for purifying Map with its chaperone CesT was also developed as described

above. From this complex, the only crystals obtained were that of CesT alone, as described above.

After the publication of the structure of Map in complex with cdc42, I hypothesized that Map could bind both cdc42 and its targeting substrate, the PDZ1 domain of Ebp50, concurrently. Cdc42 was cloned into a plasmid encoding a N-terminal 6xHis tag and purified cdc42. A complex containing all three proteins, Map, cdc42 and the PDZ1 domain of Ebp50 was successfully obtained through Ni-NTA purification and SEC. When run on SEC, the complex appeared in two distinct peaks, but did not appear to be aggregated. These peaks were large sub-megadalton complexes that could account for multimers of the complex bound together. Because Map only has one carboxy-terminus, only one PDZ1 domain can bind, suggesting that Map and PDZ1 always bind in a 1:1 ratio. Cdc42 is able to form homodimers at high concentrations, so it is possible that a dimer of two cdc42 molecules can bind two Map molecules. Based on this, a likely multimeric complex would involve 2:2:2 cdc42:Map:PDZ1. Corresponding to the smaller of the two peaks from the SEC, this complex would have a weight of 117kDa with Map (22.9 kDa), PDZ1 (14.7 kDa) and cdc42 (19.7 kDa). The larger of the two peaks appears to correspond to a weight of 543 kDa and could possibly be accounted for by a 12mer of the cdc42:Map complex with a smaller ratio of PDZ1 bound. This estimate is based on SDS-PAGE analysis of fractions of this larger peak, as well as on the molecular weights of the components. These two multimeric complexes were purified separately and used to set up crystallization trials at 7 mg/mL for the 543 kDa complex and 2 mg/mL for the 117 kDa complex. Though phase separation was apparent in many of the crystallization conditions, none of these developed into diffracting crystals.

Though diffraction-quality were not obtained in the initial crystallization screening, these large macromolecular complexes may have physiological relevance and deserve further examination. Studying the envelope of the complex, either by SAXS or by EM, may result in determination of how the complex associates, if the particles all have similar shapes and if individual components can be identified. Further crystallization experiments may also be useful in determining the stoichiometry of the complex.

Methods

All plasmids used for this study are listed in Table B-2. The Map gene was amplified from EHEC Sakai DNA using primers that inserted NdeI and XmaI restriction sites at the ends. These sites were used to ligate the Map gene into pMS173 that encodes a N-terminal 6xHis tag. A TEV protease site is located between the His tag and the gene to allow the removal of the tag. Ebp50 and the PDZ1 domain of Ebp50 (residues 1-138) were amplified from pGEX plasmids obtained from the Evans lab at the Imperial College of Science (UK) and put into the same pET28c vector as Map. A plasmid containing cdc42 was obtained from the Ahn lab at the University of Colorado at Boulder and amplified out with primers encoding NdeI and EcoRI, as above for Map. This was also ligated into the pMS173 vector. Additionally, Map was ligated into a pMAL vector that encoded an N-terminal 6xHis-MBP tag with TEV site allowing the removal of the tag. Map was amplified with NdeI and XmaI restriction sites, and both Map and the pMAL vector were then digested with NdeI and XmaI. The pMAL vector was then dephosphorylated with CIP before PCR cleanup, whereas the Map insert was gel purified. These were incubated at a 1:1 molar ratio with T4 ligase. After ligation, constructs were transformed into XL-10 Gold cells, miniprepmed and verified by sequencing. A Map construct with a native C-terminus was amplified with NdeI and

EcoRI. All other steps were as above. These constructs were all tested for expression and solubility during a 37C and a 20C growth in LB with 50 µg/mL kanamycin. MBP-Map fusion protein was more soluble when expressed at 20C for overnight (~17 hours). All constructs expressed as soluble proteins and at high levels at 37 C growth and induction temperatures. The expression was induced by addition of IPTG to 1mM when cells reached an $OD_{600} = 0.6$ and were harvested after 3 hours by centrifugation at 4000xG. Cells were resuspended into Buffer A and frozen or processed immediately. The cells were lysed by sonication for 3 minutes in pulses of 10 seconds followed by 50 seconds of rest at 40% amplitude. The cell pellets were separated from the soluble fractions by centrifugation at 14,000xG.

References

1. Anomaly, J. Collective action and individual choice: rethinking how we regulate narcotics and antibiotics. *J Med Ethics* **39**, 752–756 (2013).
2. Thaller, M. C. *et al.* Tracking Acquired Antibiotic Resistance in Commensal Bacteria of Galápagos Land Iguanas: No Man, No Resistance. *PLoS ONE* **5**, e8989 (2010).
3. Kimura, K. *et al.* A small-molecule inhibitor of the bacterial type III secretion system protects against in vivo infection with *Citrobacter rodentium*. *J Antibiot* **64**, 197–203 (2011).
4. Keyser, P., Elofsson, M., Rosell, S. & Wolf-Watz, H. Virulence blockers as alternatives to antibiotics: type III secretion inhibitors against Gram-negative bacteria. *J Intern Med* **264**, 17–29 (2008).
5. Veenendaal, A. K. J., Sundin, C. & Blocker, A. J. Small-Molecule Type III Secretion System Inhibitors Block Assembly of the Shigella Type III Secretion. *J Bacteriology* **191**, 563–570 (2009).
6. Martinez, J. L. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proc R Soc B* **276**, 2521–2530 (2009).
7. Kelly, D., Conway, S. & Aminov, R. Commensal gut bacteria: mechanisms of immune modulation. *Trends Immun* **26**, 326–333 (2005).
8. Packey, C. D. & Sartor, R. B. Commensal bacteria, traditional and opportunistic pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases. *Curr Opin Infectious Diseases* **22**, 292–301 (2009).
9. Takeuchi, O. & Akira, S. Sterile inflammation: sensing and reacting to damage. *Cell* **140**, 805–820 (2010).
10. Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol* **11**, 373–384 (2010).
11. Carroll, M. C. The complement system in regulation of adaptive immunity. *Nat Immunol* **5**, 981–986 (2004).
12. Ganz, T. The role of antimicrobial peptides in innate immunity. *Integr Comp Biol* **43**, 300–304 (2003).
13. Gunn, J. S. Bacterial modification of LPS and resistance to antimicrobial peptides. *J Endotoxin Res* **7**, 57–62 (2001).
14. Dempsey, P. W., Vaidya, S. A. & Cheng, G. The art of war: Innate and adaptive immune responses. *Cell Mol Life Sci* **60**, 2604–2621 (2003).
15. Joffre, O., Nolte, M. A., Spörri, R. & Sousa, C. R. E. Inflammatory signals in dendritic cell activation and the induction of adaptive immunity. *Immunol Rev* **227**, 234–247 (2009).
16. Iwasaki, A. & Medzhitov, R. Regulation of adaptive immunity by the innate immune system. *Science* **327**, 291–295 (2010).
17. Kumar, H., Kawai, T. & Akira, S. Pathogen recognition by the innate immune system. *Int Rev Immunol* **30**, 16–34 (2011).
18. Kawai, T. & Akira, S. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* **34**, 637–650 (2011).

19. Beg, A. A. *et al.* I kappa B interacts with the nuclear localization sequences of the subunits of NF-kappa B: a mechanism for cytoplasmic retention. *Genes Dev* **6**, 1899–1913 (1992).
20. Wong, D. *et al.* Extensive characterization of NF-κB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. *Genome biology* **12**, R70 (2011).
21. Blackwell, T. S. & Christman, J. W. The role of nuclear factor-kappa B in cytokine gene regulation. *Am J Respir Cell Mol Biol* **17**, 3–9 (1997).
22. Harada, A. *et al.* Essential involvement of interleukin-8 (IL-8) in acute inflammation. *J Leukocyte Biol* **56**, 559–564 (1994).
23. Hayden, M. S. & Ghosh, S. Signaling to NF-kappaB. *Genes Dev* **18**, 2195–2224 (2004).
24. Wong, A. R. C. *et al.* Enteropathogenic and enterohaemorrhagic Escherichia coli: even more subversive elements. *Mol Microbiol* **80**, 1420–1438 (2011).
25. Gophna, U., Ron, E. Z. & Graur, D. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* **312**, 151–163 (2003).
26. Young, G. M., Schmiel, D. H. & Miller, V. L. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. *PNAS* **96**, 6456–6461 (1999).
27. Young, B. M. & Young, G. M. YplA Is Exported by the Ysc, Ysa, and Flagellar Type III Secretion Systems of Yersinia enterocolitica. *J Bacteriology* **184**, 1324–1334 (2002).
28. Cornelis, G. R. The type III secretion injectisome. *Nat Rev Micro* **4**, 811–825 (2006).
29. Galán, J. E. & Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444**, 567–573 (2006).
30. Feldman, M. F. & Cornelis, G. R. The multitasking type III chaperones: all you can do with 15 kDa. *FEMS Microbiol Letters* **219**, 151–158 (2003).
31. Creasey, E. A. *et al.* CesT is a bivalent enteropathogenic Escherichia coli chaperone required for translocation of both Tir and Map. *Mol Microbiol* **47**, 209–221 (2003).
32. Abe, A. *et al.* Enteropathogenic Escherichia coli translocated intimin receptor, Tir, requires a specific chaperone for stable secretion. *Mol Microbiol* **33**, 1162–1175 (1999).
33. Cooper, C. A. *et al.* Structural and biochemical characterization of SrcA, a multi-cargo type III secretion chaperone in Salmonella required for pathogenic association with a host. *PLoS Pathog* **6**, e1000751 (2010).
34. Lilic, M., Vujanac, M. & Stebbins, C. E. A Common Structural Motif in the Binding of Virulence Factors to Bacterial Secretion Chaperones. *Mol Cell* **21**, 653–664 (2006).
35. Parsot, C. The various and varying roles of specific chaperones in type III secretion systems. *Curr Opin Microbiology* **6**, 7–14 (2003).
36. Frankel, G. *et al.* Enteropathogenic and enterohaemorrhagic Escherichia coli: more subversive elements. *Mol Microbiol* **30**, 911–921 (1998).
37. Tobe, T. *et al.* An extensive repertoire of type III secretion effectors in Escherichia coli O157 and the role of lambdoid phages in their dissemination. *PNAS* **103**, 14941–14946 (2006).
38. Galán, J. E. & Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322–1328 (1999).
39. Stebbins, C. E. & Galán, J. E. Structural mimicry in bacterial virulence. *Nature* **412**,

- 701–705 (2001).
40. Hacker, J. & Kaper, J. B. Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**, 641–679 (2000).
 41. Ohnishi, M. *et al.* Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. *PNAS* **99**, 17043–17048 (2002).
 42. Stebbins, C. E. & Galán, J. E. Modulation of host signaling by a bacterial mimic: structure of the *Salmonella* effector SptP bound to Rac1. *Mol Cell* **6**, 1449–1460 (2000).
 43. Stebbins, C. E. & Galán, J. E. Structural mimicry in bacterial virulence. *Nature* **412**, 701–705 (2001).
 44. Fu, Y. & Galán, J. E. A salmonella protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. *Nature* **401**, 293–297 (1999).
 45. Hume, P. J. & Vassilis, K. Mimicry Is the Sincerest Form of Flattery? *Cell Host and Microbe* **4**, 410–412 (2008).
 46. Stender, S. *et al.* Identification of SopE2 from *Salmonella typhimurium*, a conserved guanine nucleotide exchange factor for Cdc42 of the host cell. *Mol Microbiol* **36**, 1206–1221 (2002).
 47. Huang, Z. *et al.* Structural insights into host GTPase isoform selection by a family of bacterial GEF mimics. *Nat Struct Mol Biol* **16**, 853–860 (2009).
 48. Thapar, N. & Sanderson, I. R. Diarrhoea in children: an interface between developing and developed countries. *Lancet* **363**, 641–653 (2004).
 49. Deborah Chen, H. & Frankel, G. Enteropathogenic *Escherichia coli*: unravelling pathogenesis. *FEMS Microbiol Reviews* **29**, 83–98 (2005).
 50. Dean, P., Maresca, M. & Kenny, B. EPEC's weapons of mass subversion. *Curr Opin Microbiology* **8**, 28–34 (2005).
 51. Frankel, G. & Phillips, A. D. Attaching effacing *Escherichia coli* and paradigms of Tir-triggered actin polymerization: getting off the pedestal. *Cellular Microbiology* **10**, 549–556 (2008).
 52. Sandvig, K. Shiga toxins. *Toxicon* **39**, 1629–1635 (2001).
 53. Endo, Y. *et al.* Site of action of a Vero toxin (VT2) from *Escherichia coli* O157:H7 and of Shiga toxin on eukaryotic ribosomes. RNA N-glycosidase activity of the toxins. *Eur J Biochem* **171**, 45–50 (1988).
 54. Guttman, J. A. *et al.* Attaching and effacing pathogen-induced tight junction disruption in vivo. *Cellular Microbiology* **8**, 634–645 (2006).
 55. Center for Disease Control and Prevention (CDC). Antibiotic resistance threats in the United States, 2013. www.cdc.gov (2013).
 56. Rangel, J. M., Sparling, P., Crowe, C., Griffin, P. M. & Swerdlow, D. L. Epidemiology of *Escherichia coli* O157:H7 Outbreaks, United States, 1982–2002. *Emerging Infect Dis* **11**, 603–609 (2005).
 57. Hayashi, T. Complete Genome Sequence of Enterohemorrhagic *Escherichia coli* O157:H7 and Genomic Comparison with a Laboratory Strain K-12. *DNA Res* **8**, 11–22 (2001).
 58. Michino, H. *et al.* Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. *Am J Epidemiol* **150**, 787–796 (1999).
 59. Donnenberg, M. S. & Kaper, J. B. Enteropathogenic *Escherichia coli*. *Infect Immun*

- 60**, 3953 (1992).
60. Donnenberg, M. S. *et al.* The role of the eae gene of enterohemorrhagic *Escherichia coli* in intimate attachment in vitro and in a porcine model. *J Clin Invest* **92**, 1418–1424 (1993).
 61. Kenny, B. *et al.* Enteropathogenic *E. coli* (EPEC) Transfers Its Receptor for Intimate Adherence into Mammalian Cells. *Cell* **91**, 511–520 (1997).
 62. Ruchaud-Sparagano, M.-H., Mühlen, S., Dean, P. & Kenny, B. The Enteropathogenic *E. coli* (EPEC) Tir Effector Inhibits NF- κ B Activity by Targeting TNF α Receptor-Associated Factors. *PLoS Pathog* **7**, e1002414 (2011).
 63. Kenny, B. & Jepson, M. Targeting of an enteropathogenic *Escherichia coli* (EPEC) effector protein to host mitochondria. *Cellular Microbiology* **2**, 579–590 (2000).
 64. Jepson, M. A. *et al.* Synergistic roles for the Map and Tir effector molecules in mediating uptake of enteropathogenic *Escherichia coli* (EPEC) into non-phagocytic cells. *Cellular Microbiology* **5**, 773–783 (2003).
 65. Alto, N. M. *et al.* Identification of a Bacterial Type III Effector Family with G Protein Mimicry Functions. *Cell* **124**, 133–145 (2006).
 66. Alto, N. M. Mimicking small G-proteins: an emerging theme from the bacterial virulence arsenal. *Cellular Microbiology* **10**, 566–575 (2008).
 67. Pham, T. H. *et al.* Functional Differences and Interactions between the *Escherichia coli* Type III Secretion System Effectors NleH1 and NleH2. *Infect Immun* **80**, 2133–2140 (2012).
 68. Newton, H. J. *et al.* The Type III Effectors NleE and NleB from Enteropathogenic *E. coli* and OspZ from *Shigella* Block Nuclear Translocation of NF- κ B p65. *PLoS Pathog* **6**, e1000898 (2010).
 69. Nadler, C. *et al.* The Type III Secretion Effector NleE Inhibits NF- κ B Activation. *PLoS Pathog* **6**, e1000743 (2010).
 70. Baruch, K. *et al.* Metalloprotease type III effectors that specifically cleave JNK and NF- κ B. *EMBO J* 1–11 (2010).
 71. Yen, H. *et al.* NleC, a Type III Secretion Protease, Compromises NF- κ B Activation by Targeting p65/RelA. *PLoS Pathog* **6**, e1001231 (2010).
 72. Mühlen, S., Ruchaud-Sparagano, M.-H. & Kenny, B. Proteasome-independent degradation of canonical NF- κ B complex components by the NleC protein of pathogenic *Escherichia coli*. *J Biol Chem* **286**, 5100–5107 (2011).
 73. Pearson, J. S., Riedmaier, P., Marches, O., Frankel, G. & Hartland, E. L. A type III effector protease NleC from enteropathogenic *Escherichia coli* targets NF- κ B for degradation. *Mol Microbiol* **80**, 219–230 (2011).
 74. Sham, H. P. *et al.* Attaching and effacing bacterial effector NleC suppresses epithelial inflammatory responses by inhibiting NF- κ B and p38 mitogen-activated protein kinase activation. *Infect Immun* **79**, 3552–3562 (2011).
 75. Yen, Y. T., Bhattacharya, M. & Stathopoulos, C. Genome-wide in silicomapping of the secretome in pathogenic *Yersinia pestis* KIM. *FEMS Microbiol Letters* **279**, 56–63 (2008).
 76. Shames, S. R. *et al.* The pathogenic *Escherichia coli* type III secreted protease NleC degrades the host acetyltransferase p300. *Cellular Microbiology* **13**, 1542–1557 (2011).
 77. Hecht, G. & Savkovic, S. D. Review article: Effector role of epithelia in inflammation -

- interaction with bacteria. **11**, 64–69 (1997).
78. MacDonald, T. T. & Pettersson, S. Bacterial regulation of intestinal immune responses. *Inflamm Bowel Dis* **6**, 116–122 (2000).
 79. Ruchaud-Sparagano, M.-H., Maresca, M. & Kenny, B. Enteropathogenic *Escherichia coli* (EPEC) inactivate innate immune responses prior to compromising epithelial barrier function. *Cellular Microbiology* **9**, 1909–1921 (2007).
 80. Savkovic, S. D., Koutsouris, A. & Hecht, G. Attachment of a noninvasive enteric pathogen, enteropathogenic *Escherichia coli*, to cultured human intestinal epithelial monolayers induces transmigration of neutrophils. *Infect Immun* **64**, 4480–4487 (1996).
 81. Zhang, L. *et al.* Cysteine methylation disrupts ubiquitin-chain sensing in NF- κ B activation. *Nature* **481**, 204–208 (2011).
 82. Royan, S. V. *et al.* Enteropathogenic *E. coli* non-LEE encoded effectors NleH1 and NleH2 attenuate NF- κ B activation. *Mol Microbiol* **78**, 1232–1245 (2010).
 83. Chen, Y. Q., Ghosh, S. & Ghosh, G. A novel DNA recognition mode by the NF-kappa B p65 homodimer. *Nat Struct Biol* **5**, 67–73 (1998).
 84. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529–33 (2010).
 85. Johnson, G. L. & Lapadat, R. Mitogen-Activated Protein Kinase Pathways Mediated by ERK, JNK, and p38 Protein Kinases. *Science* **298**, 1911–1912 (2002).
 86. Gerritsen, M. E. *et al.* CREB-binding protein/p300 are transcriptional coactivators of p65. *PNAS* **94**, 2927–2932 (1997).
 87. Dahlroth, S.-L., Nordlund, P. & Cornvik, T. Colony filtration blotting for screening soluble expression in *Escherichia coli*. *Nat Protoc* **1**, 253–258 (2006).
 88. Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865–1871 (2003).
 89. Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Res* **38**, D227–33 (2010).
 90. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**, 536–540 (1995).
 91. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res* **40**, D290–D301 (2011).
 92. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545–W549 (2010).
 93. Fushimi, N., Ee, C. E., Nakajima, T. & Ichishima, E. Aspzincin, a family of metalloendopeptidases with a new zinc-binding motif. *J Biol Chem* **274**, 24195 (1999).
 94. Hori, T. *et al.* Structure of a new ‘aspzincin’ metalloendopeptidase from *Grifola frondosa*: implications for the catalytic mechanism and substrate specificity based on several different crystal forms. *Acta Crystallogr D Biol Crystallogr* **57**, 361–368 (2001).
 95. Gomis-Rüth, F. X. Structural Aspects of the Metzincin Clan of Metalloendopeptidases. *Mol Biotech* **24**, 157–202 (2003).

96. Gomis-Rüth, F. X., Botelho, T. O. & Bode, W. A standard orientation for metallopeptidases. *BBA Prot Proteomics* **1824**, 157–163 (2012).
97. Ryde, U. Carboxylate binding modes in zinc proteins: a theoretical study. *Biophys J* **77**, 2777–2787 (1999).
98. Dudev, T. & Lim, C. Effect of Carboxylate-Binding Mode on Metal Binding/Selectivity and Function in Proteins. *Acc Chem Res* **40**, 85–93 (2007).
99. Harding, M. M. The architecture of metal coordination groups in proteins. *Acta Crystallogr D Biol Crystallogr* **60**, 849–859 (2004).
100. Celniker, G. *et al.* ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. **53**, 199–206 (2013).
101. Kurisu, G. *et al.* Structure of the zinc endoprotease from *Streptomyces caespitosus*. *J Biochem* **121**, 304–308 (1997).
102. Dupont, C. L., Butcher, A., Valas, R. E., Bourne, P. E. & Caetano-Anollés, G. History of biological metal utilization inferred through phylogenomic analysis of protein structures. *PNAS* **107**, 10567–10572 (2010).
103. Irving, H. & Williams, R. J. P. 637. The stability of transition-metal complexes. *J Chem Soc* 3192–3210 (1953).
104. Gorelsky, S. I. *et al.* Spectroscopic and DFT Investigation of [M{HB(3,5- iPr 2pz) 3}{SC 6F 5)] (M = Mn, Fe, Co, Ni, Cu, and Zn) Model Complexes: Periodic Trends in Metal–Thiolate Bonding. *Inorg Chem* **44**, 4947–4960 (2005).
105. Andreini, C., Banci, L., Bertini, I. & Rosato, A. Counting the Zinc-Proteins Encoded in the Human Genome. *J Proteome Res* **5**, 196–201 (2006).
106. Caetano-Anollés, G. & Wang, M. Personal communication via email. (2013).
107. Coleman, J. E. Zinc enzymes. *Curr Opin Chem Biol* **2**, 222–234 (1998).
108. Caetano-Anollés, G. & Caetano-Anollés, D. An evolutionarily structured universe of protein architecture. *Genome Research* **13**, 1563–1571 (2003).
109. Caetano-Anollés, G., Wang, M., Caetano-Anollés, D. & Mittenthal, J. E. The origin, evolution and structure of the protein world. *Biochem J* **417**, 621 (2009).
110. Wang, M., Yafremava, L. S., Caetano-Anolles, D., Mittenthal, J. E. & Caetano-Anolles, G. Reductive evolution of architectural repertoires in proteomes and the birth of the tripartite world. *Genome Research* **17**, 1572–1585 (2007).
111. Ingles-Prieto, A. *et al.* Conservation of Protein Structure over Four Billion Years. *Structure* **21**, 1690–1697 (2013).
112. Gomis-Ruth, F. X. Catalytic Domain Architecture of Metzincin Metalloproteases. *J Biol Chem* **284**, 15353–15357 (2009).
113. Patel, K., Kumar, A. & Durani, S. Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures. *BBA Prot Proteomics* **1774**, 1247–1253 (2007).
114. Adekoya, O. A. & Sylte, I. The thermolysin family (M4) of enzymes: therapeutic and biotechnological potential. *Chem Biol Drug Des* **73**, 7–16 (2009).
115. Matthews, B. W. Structural basis of the action of thermolysin and related zinc peptidases. *Acc Chem Res* **21**, 333–340 (1988).
116. Bode, W. Structural basis of matrix metalloproteinase function. *Biochem Soc Symp* 1–14 (2003).
117. Silva, D. S. *et al.* The Apoptogenic Toxin AIP56 Is a Metalloprotease A-B Toxin that Cleaves NF- κ B P65. *PLoS Pathog* **9**, e1003128 (2013).

118. Do Vale, A. *et al.* AIP56, a novel plasmid-encoded virulence factor of *Photobacterium damsela* subsp. *piscicida* with apoptogenic activity against sea bass macrophages and neutrophils. *Mol Microbiol* **58**, 1025–1038 (2005).
119. Wilkes, T. E. *et al.* The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis. *Insect Mol Biol* **19**, 59–73 (2010).
120. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *PNAS* **98**, 10037–10041 (2001).
121. Ragusa, S., Mouchet, P., Lazennec, C., Dive, V. & Meinnel, T. Substrate recognition and selectivity of peptide deformylase. Similarities and differences with metzincins and thermolysin. *J Mol Biol* **289**, 1445–1457 (1999).
122. Rathinavelan, T. *et al.* A repulsive electrostatic mechanism for protein export through the type III secretion apparatus. *Biophys J* **98**, 452–461 (2010).
123. Hutchinson, E. G. & Thornton, J. M. HERA--a program to draw schematic diagrams of protein secondary structures. *Proteins* **8**, 203–212 (1990).
124. E G Hutchinson, J. M. T. PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Sci* **5**, 212 (1996).
125. Fasman, G. D. *Prediction of protein structure and the principles of protein conformation*. (Springer Us, 1989).
126. Richardson, J. S. & Richardson, D. C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *PNAS* **99**, 2754–2759 (2002).
127. Pallen, M. J., Dougan, G. & Frankel, G. Coiled-coil domains in proteins secreted by type III secretion systems. *Mol Microbiol* **25**, 423–425 (1997).
128. Delahay, R. M. & Frankel, G. Coiled-coil proteins associated with type III secretion systems: a versatile domain revisited. *Mol Microbiol* **45**, 905–916 (2002).
129. Jackson, K. A., Valentine, R. A., Coneyworth, L. J., Mathers, J. C. & Ford, D. Mechanisms of mammalian zinc-regulated gene expression. *Biochem Soc Trans* **36**, 1262 (2008).
130. Gabbianelli, R. *et al.* Role of ZnuABC and ZinT in *Escherichia coli* O157:H7 zinc acquisition and interaction with epithelial cells. *BMC Microbiol* **11**, 36 (2011).
131. Hantke, K. Bacterial zinc uptake and regulators. *Curr Opin Microbiology* **8**, 196–202 (2005).
132. Kehl-Fie, T. E. & Skaar, E. P. Nutritional immunity beyond iron: a role for manganese and zinc. *Curr Opin Chem Biol* **14**, 218–224 (2010).
133. Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. Structure of the NF- κ B p50 homodimer bound to DNA. *Nature* **373**, 311–317 (1995).
134. Wang, H.-C., Wu, M.-L., Ko, T.-P. & Wang, A. H. J. *Neisseria* conserved hypothetical protein DMP12 is a DNA mimic that binds to histone-like HU protein. *Nucleic Acids Res* **41**, 5127–5138 (2013).
135. Dryden, D. & Tock, M. R. DNA mimicry by proteins. *Biochem Soc Trans* **34**, 317–319 (2006).
136. Roberts, G. A. *et al.* Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic Acids Res* **40**, 8129–8143 (2012).
137. Macian, F. NFAT proteins: key regulators of T-cell development and function. *Nat Rev Immunol* **5**, 472–484 (2005).

138. Jain, J., Burgeon, E., Badalian, T. M., Hogan, P. G. & Rao, A. A similar DNA-binding motif in NFAT family proteins and the Rel homology region. *J Biol Chem* **270**, 4138–4145 (1995).
139. Müller, M. R. & Rao, A. NFAT, immunity and cancer: a transcription factor comes of age. *Nat Rev Immunol* **10**, 645–656 (2010).
140. Beg, A. A. & Baltimore, D. An essential role for NF-kappaB in preventing TNF-alpha-induced cell death. *Science* **274**, 782–784 (1996).
141. Burkly, L. *et al.* Expression of relB is required for the development of thymic medulla and dendritic cells. *Nature* **373**, 531–536 (1995).
142. Sha, W. C., Liou, H.-C., Tuomanen, E. I. & Baltimore, D. Targeted disruption of the p50 subunit of NF-κB leads to multifocal defects in immune responses. *Cell* **80**, 321–330 (1995).
143. Staibano, S., Mascolo, M., Siano, M., Ilardi, G. & De, G. in *intechopen.com* (Research on Melanoma - A Glimpse into Current Directoins and Future Trends, 2011). doi:10.5772/19966
144. Gao, X. *et al.* Bacterial Effector Binding to Ribosomal Protein S3 Subverts NF-κB Function. *PLoS Pathog* **5**, e1000708 (2009).
145. Thomas, N. A., Ma, I., Prasad, M. E. & Rafuse, C. Expanded roles for multicargo and class 1B effector chaperones in type III secretion. *J Bacteriology* **194**, 3767–3773 (2012).
146. Harrington, A. T. *et al.* Structural Characterization of the N Terminus of IpaC from *Shigella flexneri*. *Infect Immun* **71**, 1255–1264 (2003).
147. Arnold, R. *et al.* Sequence-Based Prediction of Type III Secreted Proteins. *PLoS Pathog* **5**, e1000376 (2009).
148. Samudrala, R., Heffron, F. & McDermott, J. E. Accurate Prediction of Secreted Substrates and Identification of a Conserved Putative Secretion Signal for Type III Secretion Systems. *PLoS Pathog* **5**, e1000375 (2009).
149. Ohlson, M. B. *et al.* Structure and Function of Salmonella SifA Indicate that Its Interactions with SKIP, SseJ, and RhoA Family GTPases Induce Endosomal Tubulation. *Cell Host and Microbe* **4**, 434–446 (2008).
150. Wild, J., Altman, E., Yura, T. & Gross, C. A. DnaK and DnaJ heat shock proteins participate in protein export in *Escherichia coli*. *Genes Dev* **6**, 1165–1172 (1992).
151. Wild, J., Rossmeyssl, P., Walter, W. A. & Gross, C. A. Involvement of the DnaK-DnaJ-GrpE chaperone team in protein secretion in *Escherichia coli*. *J Bacteriology* **178**, 3608–3613 (1996).
152. Thomas, N. A. *et al.* CesT is a multi-effector chaperone and recruitment factor required for the efficient type III secretion of both LEE- and non-LEE-encoded effectors of enteropathogenic *Escherichia coli*. *Mol Microbiol* **57**, 1762–1779 (2005).
153. Lloyd, S. A., Norman, M., Rosqvist, R. & Wolf-Watz, H. *Yersinia* YopE is targeted for type III secretion by N-terminal, not mRNA, signals. *Mol Microbiol* **39**, 520–531 (2001).
154. Birtalan, S. C., Phillips, R. M. & Ghosh, P. Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens. *Mol Cell* **9**, 971–980 (2002).
155. Luo, Y. *et al.* Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. *Nat Struct Biol* **8**, 1031–1036 (2001).

156. Rumpel, S., Lakshmi, R., Becker, S. & Zweckstetter, M. Assignment-free solution NMR method reveals CesT as an unswapped homodimer. *Protein Sci* **17**, 2015–2019 (2008).
157. Rumpel, S., Kim, H.-Y., Vijayan, V., Becker, S. & Zweckstetter, M. Letter to the Editor: Backbone resonance assignment of the homodimeric, 35 kDa chaperone CesT from enteropathogenic *Escherichia coli*. *J Biomol NMR* **31**, 377–378 (2005).