ELUCIDATION OF PROTEIN-PROTEIN AND PROTEIN-SMALL MOLECULE INTERACTIONS IN THE TYPE III SECRETION SYSTEM

By

Amritangshu Chakravarty

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

_____
Chairperson: Roberto N. De Guzman, Ph.D.

_____
Krzysztof Kuczera, Ph.D.

_____
Joanna Slusky, Ph.D.

_____
Erik Holmstrom, Ph.D.

_____
Prajna Dhar, Ph.D.

Date Defended: 05/03/2021

The Dissertation Committee for Amritangshu Chakravarty

certifies that this is the approved version of the following dissertation:

ELUCIDATION OF PROTEIN-PROTEIN AND PROTEIN-SMALL MOLECULE INTERACTIONS IN THE TYPE III SECRETION SYSTEM

_____

Chairperson: Roberto N. De Guzman, Ph.D.

Date Approved: 05/14/2021

Abstract

Many biological processes are characterized by transient and weak interactions. Nuclear Magnetic Resonance (NMR) is a powerful tool to characterize these interactions on a per-residue basis. In this dissertation, I have utilized NMR spectroscopy to characterize protein-protein and protein-small molecule interactions of different proteins of the type III secretion system (T3SS) in pathogenic Gram-negative bacteria.

Pathogenic Gram-negative bacteria such as *Salmonella, Shigella, Burkholderia, Pseudomonas and Yersinia* assemble the T3SS to inject virulence proteins into eukaryotic cells, causing millions of deaths worldwide. A major public health concern is the development of antibiotic resistance among the Gram-negative pathogens. This problem is exacerbated by the low number of new antibiotics in the pipeline. Because of its essential role in virulence, understanding the assembly of the T3SS can lead to the development of new antimicrobials.

The structural component of the T3SS is the needle complex, which consists of a base, an extracellular needle, a tip, and a translocon. The translocon forms a pore on the host cell membrane to allow the injection of virulence proteins into the host cell. The two membrane proteins that assemble the translocon – the major translocon protein and the minor translocon protein – are designated based on their relative sizes. The minor translocon proteins from various bacteria are predicted to contain one transmembrane domain. IpaC is the minor translocon protein of *Shigella*, and plays an essential role in pathogenesis. The atomic structure for any minor translocon protein and its conformational changes in membrane mimetics and micelles are currently unknown. The protein-protein interactions of IpaC are also not well understood. In this dissertation, I show by NMR that full-length IpaC, as well as its N-terminal domain (NTD) and

C-terminal domain (CTD), are intrinsically disordered. NMR titrations show that IpaC and its NTD and CTD bind to the chaperone protein IpgC and the tip protein IpaD. Further, the IpaC CTD interacts with the N-terminal domain of the major translocon protein IpaB. IpaC is known to oligomerize, and NMR detected homotypic protein-protein interactions of the IpaC CTD. My results show the first NMR spectrum for any minor translocon protein, identify previously unknown binding partners of IpaC, and provide possible insights into the mechanism of oligomerization of IpaC. Circular dichroism (CD) spectroscopy and NMR also show that full-length IpaC and its domains undergo conformational changes in DPC, LMPG and SDS micelles. This knowledge is important in the biophysical characterization of IpaC.

The tip protein in *Salmonella* is SipD, which is exposed on the cell surface and is essential in virulence. This makes SipD an attractive target for developing new antimicrobials. Currently, the only known small molecules that bind to SipD are bile salts and small molecule compounds based on hydroxyindole, morpholinoaniline and indole-acetic acid scaffolds. Results of the computational screening of compound libraries for binding to tip proteins is also currently unknown. In this dissertation, I report the results of our computational screening of a compound library containing ~ 8.33 million compounds using Rosetta and ROCS/FastROCS for binding to SipD. The screening predicted the binding of three compounds to SipD, based on piperidine, piperazine and phenyl scaffolds. Through a collaboration, we also used surface plasmon resonance (SPR) to screen one small molecule library containing 104 compounds with structures similar to known antibiotics, and another library consisting of 440 compounds with structures similar to intermediates in the synthesis of larger drug-like candidates. We identified the binding of three small molecules based on quinoline and phenyl scaffolds using the SPR screen. I then used NMR to identify the residues and surfaces of SipD that interact with these small molecules. This knowledge increases the number of known fragments that can bind to SipD, and can be used

to screen other virulence proteins from the T3SS, aiding in designing novel therapeutics to combat the growing threat posed by multidrug resistance in bacteria.

My dissertation illustrates the strength of NMR spectroscopy to characterize weak and transient protein-ligand interactions in the T3SS.

Acknowledgements

My journey as a doctoral student would have been impossible without the support of my advisor, Dr. Roberto N. De Guzman. He constantly encouraged and motivated me in my tenure as a graduate student and introduced me to the world of NMR spectroscopy. His constructive criticisms have enabled me to become a better scientist, oral presenter and science writer. Following his advice, I have come to realize that scientific writing can be really enjoyable. His encouragement guided me throughout this journey, through the plethora of hardships and disappointments in terms of failed experiments. He taught me to think independently in terms of designing experiments, as well as analyzing and presenting data. I will forever be grateful to him.

I would like to extend a special note of acknowledgement to all my current and former committee members – Dr. Krzysztof Kuczera, Dr. Joanna Slusky, Dr. Erik Holmstrom, Dr. Prajna Dhar and Dr. Eric Deeds – for their suggestions and guidance that have helped me in my doctoral research. I especially thank Dr. Mark Richter, my former committee member and the chair of my oral proposal defense committee, for his valuable suggestions on my research and proposal. He will be dearly missed.

I am really thankful to the members of the NMR Core Facility at KU - Dr. Justin Douglas and Dr. Minli Xing (former) – for their assistance and training on setting up NMR experiments. Our collaborators at KU – Dr. David Johnson (Director, Computational Chemical Biology Core Facility), Dr. Jeff Aube (former Distinguished Professor at the Department of Medicinal Chemistry) and Anne Cooper (Protein Production Group) – helped me complete the projects on protein-small molecule interactions, and were thus invaluable to my research.

A loud shoutout goes out for my labmate, Pallavi, for making the De Guzman lab a fun place to work in. She also briefly mentored me initially during my rotation period in my lab, that factored into my decision to finally join the lab. I also thank the former De Guzman lab members – Dr. Supratim Dey, Dr. Andrew McShan and Dr. Kawaljit Kaur – whose advice and guidance were very important during my initial days in the lab. I whole-heartedly thank former undergraduates, Nick and Helen, for their immense help in protein expression and purification. Nick also assisted me in performing the CD experiments of my dissertation and in preparing some figures for my manuscripts.

I thank the graduate selection committee in the Department of Molecular Biosciences for selecting me as one of their potential graduate students in 2015 and awarding me with the Robert Weaver Graduate Student Fellowship that supported me in my first year of study. The excellently designed core courses of the Biochemistry and Biophysics program served as brilliant refreshers to the important concepts. The knowledge has proved to be invaluable in my research. John Connolly, graduate advisor, and Cynthia Cowan, administrative associate at the department, took care of all the official formalities in a timely manner which helped me concentrate on my graduate work.

In conclusion, I would like to thank my parents for instilling in me the drive to pursue doctoral research and encouraging me to move ~ 5000 miles away from home to USA. Daily conversations with them in the evenings provided me the much-needed respite from the stress of graduate life. A huge shoutout for my wife Mou, who came to the USA to live with me in 2017, and has since provided me constant support in this journey.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANS – 8-anilino-1-naphthalenesulfonic acid
Bip – *Burkholderia* invasion protein
CD – Circular Dichroism
CM – Carboxymethyl
CSD – Chemical Shift Deviation
CTD – C-terminal Domain
DMSO – Dimethyl Sulfoxide
DOPC – 1,2-dioleyl-sn-glycero-3-phosphocholine
DOPG – 1,2-dioleyl-sn-glycero-3-phospho-(1'-rac-glycerol)
DPC – n-dodecylphosphocholine
DPPC – Dipalmitoylphosphatidylcholine
DTT – Dithiothreitol
EPR – Electron Paramagnetic Resonance
HSQC – Heteronuclear Single Quantum Coherence
IDP – Intrinsically disordered proteins
ILV – Isoleucine, Leucine and Valine
Ipa – Invasion plasmid antigen
IPTG – Isopropyl-β-D-thiogalactopyranoside
LMPG – Lyso-myristoylphosphatidylglycerol
NMR – Nuclear Magnetic Resonance
NTD – N-terminal Domain
OD – Optical Density
PAGE – Polyacrylamide Gel Electrophoresis
PBS – Phosphate Buffered Saline
PEI – Polyethyleneimine
PMSF – Phenylmethylsulfonyl fluoride
RMSD – Root Mean Square Deviation
ROCS – Rapid Overlay of Chemical Structures
SDS – Sodium Dodecyl Sulfate
Sip – *Salmonella* invasion protein
SPR – Surface Plasmon Resonance
STD – Saturation Transfer Difference
T3SS – Type III Secretion System
TEV – Tobacco Etch Virus
TROSY – Transverse Relaxation Optimized Spectroscopy
UV – Ultraviolet
Yop – *Yersinia* outer protein
TXI-RT – Inverse Triple Resonance-Room Temperature
GB1 – β1 domain of streptococcal protein G
LB – Luria-Bertani
TB – Terrific Broth
T_m – Melting temperature
Lcr – Low calcium response
Prg – PhoP-repressed gene
FRET – Förster Resonance Energy Transfer
K_d – Dissociation constant

Chapter 1: Introduction

1.1. The Type III Secretion System (T3SS)

Pathogenic Gram-negative bacteria such as *Salmonella, Shigella, Burkholderia, Pseudomonas and Yersinia* use the Type III Secretion System (T3SS) to inject virulence proteins into host eukaryotic cells, causing millions of deaths worldwide [1-5]. A major public health concern is the development of antibiotic resistance in *the Gram-negative pathogens* [6, 7]. This problem is exacerbated by the low number of new antibiotics in the pipeline [8]. Because of its essential role in virulence, understanding the assembly of the T3SS can lead to the development of new antimicrobials [8-10].

1.1.1. Components of the T3SS

The structural component of the T3SS consists of the sorting platform in the bacterial cytoplasm, the export apparatus and the needle complex [11-15] **(Figure 1-1**, adapted from Park *et al* [16]**)**. The export apparatus is formed by proteins embedded in the bacterial inner membrane and forms a channel that shuttles effectors across the bacterial membrane [17] (not shown in Figure 1-1). Effectors are bacterial proteins that are secreted through the T3SS. The sorting platform is a multiprotein complex in the bacterial cytoplasm. It selects and delivers the effectors to the export apparatus in a temporal manner [18]. The needle complex is made up of a base, an extracellular needle, a tip complex and a translocon **(Figure 1-1,** adapted from Park *et al* [16]**)** [19, 20]. The tip complex, formed by the tip protein, sits on top of the needle, and serves as a platform for the assembly of the translocon. The translocon forms a pore in the host cell membrane and completes the channel that allows the passage of effectors from the bacterial cytoplasm into the host cell. Inside the host cells, the effectors can modulate cytoskeletal rearrangements and cell signaling of the host cells, allowing the survival and multiplication of the

pathogens inside the host [21, 22]. The bacterial cytoplasm contains a class of proteins called the chaperones that bind to the effectors such as the tip protein and the translocon proteins to prevent their premature secretion [3, 11].

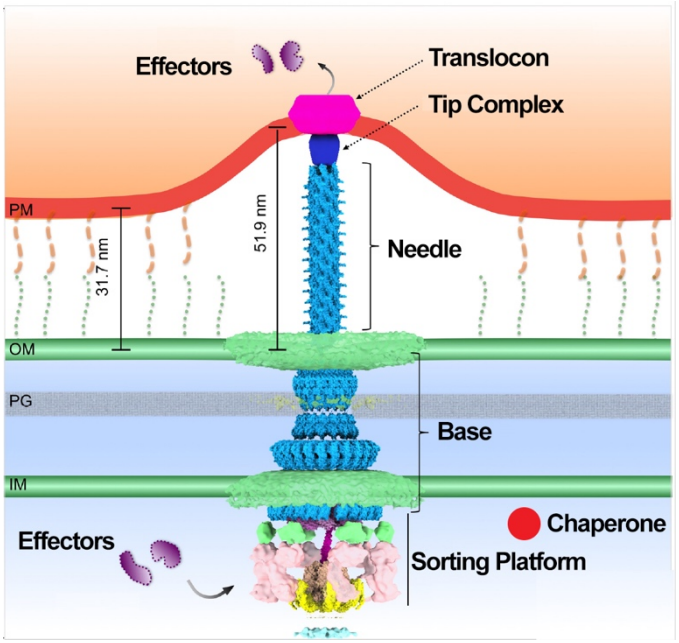

Figure 1-1. The bacterial type III secretion system. Molecular model of the T3SS needle complex (Figure adapted from Park *et al* [16]. The needle complex consists of a base, an extracellular needle, a tip complex and a translocon. The chaperone proteins (red) present in the bacterial cytoplasm prevent the premature secretion of the effector proteins. The export apparatus is embedded in the inner membrane (IM) and is not shown here.

1.1.2. The tip protein

The tip protein is positioned on top of the needle protein **(Figure 1-1)**, and functions as a platform for the assembly of the translocon. There is structural homology among the reported

crystal structures of the tip proteins – *Salmonella* SipD [20, 21], *Shigella* IpaD [23] and *Burkholderia* BipD [23, 24] **(Figure 1-2,** adapted from [25]**)**. The Cα RMSD of SipD with IpaD is 1.3 Å, while being 2.2 Å with BipD [25]. Thus, SipD is structurally more similar to IpaD than to BipD. There are three common structural features that are observed in SipD, IpaD and BipD – an N-terminal α-helical hairpin, a central coiled-coil and a mixed α/β domain **(Figure 1-2)**. The N-terminal α-helical hairpin and the mixed α/β domain are present on the two ends of the central coiled-coil. A major difference among the structures of the tip proteins is the presence of an N-terminal globular domain in the *Yersinia* tip protein LcrV [26, 27] instead of the α-helical hairpin found in SipD, IpaD and BipD **(Figure 1-2)**. The globular domain of LcrV packs at one end of the central coiled-coil. Further, in LcrV, a second globular domain is connected to the other end of the coiled-coil, similar to the mixed α/β domain in SipD, IpaD and BipD **(Figure 1-2)**.

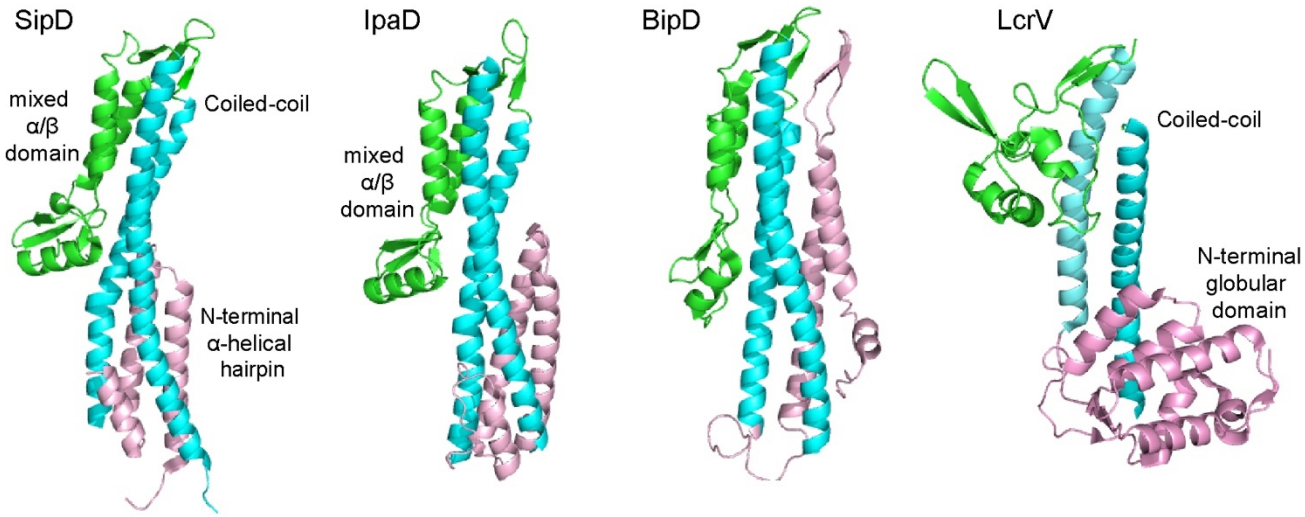

Figure 1-2. Crystal structures of the tip proteins SipD, IpaD, BipD and LcrV, colored by their domains: N-terminal α-helical hairpin or globular domain (pink), central coiled-coil (cyan) and mixed α/β domain (green). The figure is adapted from Dey *et al* [25].

1.1.3. The tip complex

Multiple subunits of the tip protein assemble to form the tip complex at the distal end of the needle. For example, pentamers of IpaD form the tip complex at the tip of the needle protein MxiH in *Shigella* [23, 28-30]. Similarly, the *Salmonella* tip complex is formed by pentameric SipD at the tip of the needle protein PrgI [20]. The proximal part of SipD interacts with the needle protein PrgI, while the distal part interacts with the N-terminal domain of the major translocon protein SipB [31] **(Figure 1-3,** adapted from [25, 31]**)**.

An important step in developing novel drug candidates against the T3SS is the identification of novel small molecules that can bind to different protein components of the T3SS. SipD is essential in virulence and is exposed on the bacterial surface. This makes it an excellent target for developing novel antimicrobials. The only small molecules known to bind the T3SS tip proteins are bile salts, which bind to *Salmonella* SipD [21, 32] and the *Shigella* IpaD [33, 34], and small molecules screened from the Zenobia 2.0 library for SipD [35] and IpaD [36]. In my dissertation, I have reported the screening of a new library of small molecules by Surface Plasmon Resonance (SPR), and another library of ~ 8.33 million drug-like compounds by computational methods, for binding to SipD. I have also determined the residues and surfaces of SipD interacting with the small molecules and compounds. The data showing the interaction of SipD with the small molecules and compounds are presented in Chapters 4 and 5.

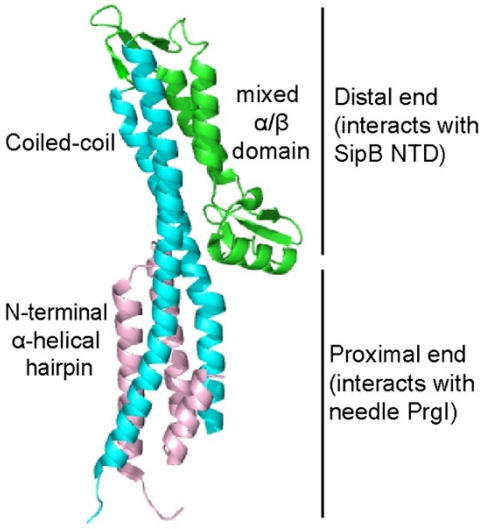

Figure 1-3. SipD interacts with PrgI and SipB-NTD. The proximal part of the *Salmonella* tip protein SipD is oriented towards the needle and interacts with the needle protein PrgI. The distal part is oriented towards the host cell membrane and interacts with the N-terminal domain of the major translocon protein SipB. The figure is adapted from McShan *et al* [31] and Dey *et al* [25].

1.1.4. The translocon

The formation of translocon pore in the host cell membrane is the final step in the assembly of the T3SS, allowing the injection of effector proteins into the host cell cytoplasm. Two hydrophobic membrane proteins assemble the translocon pore – the major and the minor translocon proteins, named on the basis of their relative sizes [37-40]. The major translocon proteins have two transmembrane domains, while the minor translocon proteins have one transmembrane domain **(Figure 1-4,** adapted from [25]**)**. These proteins have putative coiled-coil domains [41] and disordered regions [42, 43]. The crystal structures of the N-terminal domains of the major translocon proteins IpaB (in *Shigella*), SipB (in *Salmonella*) [44], and AopB (in *Aeromonas*) in complex with its chaperone [45] have been reported. These are currently the only

reported structures of the translocon proteins at atomic resolution. The atomic structure of any minor translocon protein remains unknown.

In my dissertation, I have reported the first NMR spectra of the full-length *Shigella* minor translocon protein IpaC along with those of its N-terminal and C-terminal domains. Since IpaC is a hydrophobic membrane protein, I hypothesized that IpaC and its domains undergo conformational changes in presence of membrane mimetics and detergents. I used CD and NMR spectroscopy to study the conformational changes of these proteins in DPC, LMPG and SDS micelles. The NMR spectra of full-length IpaC and its domains, along with their conformational changes in micelles, are reported in Chapter 2.

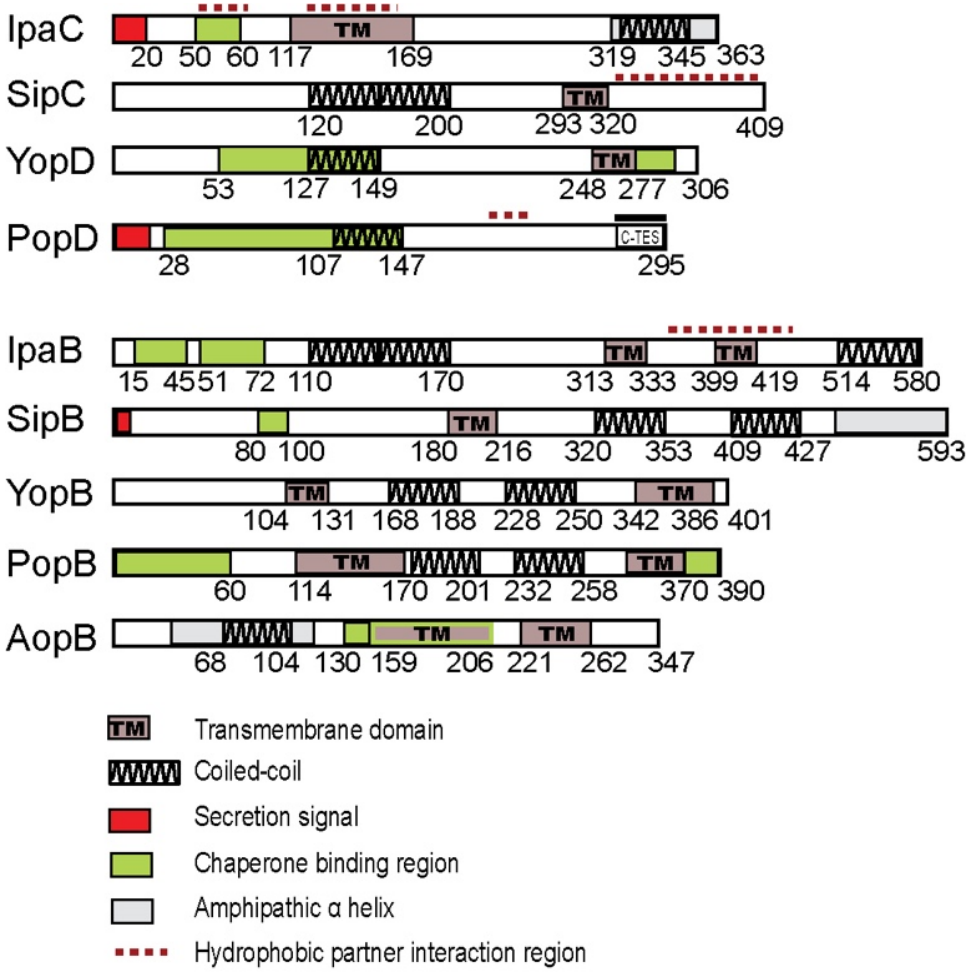

Figure 1-4. Domain maps of the translocon proteins. Domain maps are shown for the minor (top) and major (bottom) translocon proteins. The minor translocon proteins (IpaC, SipC, YopD, PopD) have one transmembrane domain, whereas the major translocon proteins (IpaB, SipB,YopB, PopB, AopB) have two transmembrane domains. The figure is adapted from Dey *et al* [25].

IpaC has 33% sequence identity and 52% similarity with the *Salmonella* minor translocon protein SipC [46]. The C-terminal domain (CTD) of both IpaC and SipC are expected to be inside the host cells, while the N-terminal domain (NTD) are extracellular **(Figure 1-5**, adapted from Russo *et al* [46]**).** Interestingly, the C-terminal 15 residues of IpaC loop back within the pore lumen and is accessible to the extracellular side, presenting possibilities of protein-protein interactions with other components of the T3SS. The protein-protein interactions of IpaC are not completely understood [47-49]. In my dissertation, I have used NMR to identify previously unknown binding partners of IpaC and to determine the domains of IpaC responsible for these interactions. I have also reported that the oligomerization of IpaC occurs via the interaction of neighboring C-terminal domains. The data of the protein-protein interactions of IpaC are presented in Chapter 3.

The key findings of my dissertation and the future directions are summarized in Chapter 6.

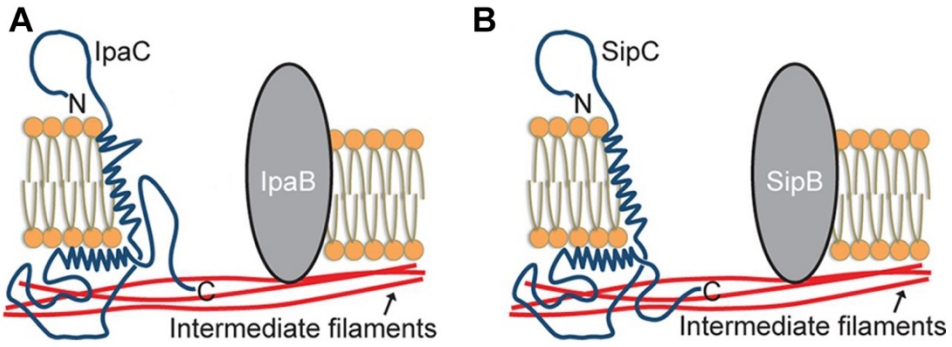

Figure 1-5. Topology of IpaC and SipC. The C-terminal region of IpaC (A) and SipC (B) are located inside the host cells, while the N-terminal region is extracellular. The C-terminal 15 residues of IpaC loop back into the pore lumen. The C-terminal regions of both proteins interact with the intermediate filaments of the host. The figure is adapted from Russo *et al* [46].

1.2. Use of NMR to study protein-ligand interactions

Many biological processes are characterized by weak interactions. Nuclear Magnetic Resonance (NMR) is a powerful tool to characterize these interactions on a per-residue basis [50-54]. The chemical shift of each atom is very sensitive to its surrounding environment, making NMR chemical shift mapping a widely used method to study protein-ligand interactions [55]. An isotopically labeled protein is titrated with unlabeled ligand, and changes in chemical shift are monitored. The NMR-active nuclei at the binding interface undergo the largest changes in chemical shift upon ligand binding [54-56], making this technique especially useful for the determination of interaction interfaces.

1.2.1. NMR exchange regimes

An excellent technique to study protein-ligand interactions is NMR titration. The binding affinity of the ligand (proteins, small molecules, nuclei acids, carbohydrates) dictates the appearance of the NMR spectra [57], pertaining to three exchange regimes **(Figure 1-6,** adapted from [57]**)** in the NMR timescale.

- Tight-binding interactions yield slow-dissociation complexes. This is called the **slow exchange regime**, when two sets of NMR peaks are observed, corresponding to the free and the bound states of the protein. As the ligand concentration is increased, the peaks from the free protein gradually reduce in intensity and disappear, while intensities of peaks from the bound state gradually increase **(**boxes, **Figure 1-6A)** (adapted from [58]) [53, 55, 57]. The slow exchange regime is associated with dissociation constants in the sub-micromolar range.

- In the **intermediate exchange regime**, significant interconversion occurs between the free and the ligand-bound states of the protein. The affected peaks reduce in intensity (peak broadening) with increasing ligand concentration, and can disappear **(Figure 1-6B** and **Figure 1-6D)**. Peak broadening may be associated with a shift in the position of the peak. The intermediate exchange regime is associated with micromolar dissociation constant [53, 55, 57].

- **Fast exchange regime** corresponds to weak binding, with high micromolar to millimolar dissociation constant. Very fast interconversion occurs between free and bound states. The observed chemical shift is the weighted average of the chemical shifts of the free and bound forms. There is a continuous change in peak position with increasing ligand concentration **(Figure 1-6C** and **Figure 1-6E)** [53-55, 57].

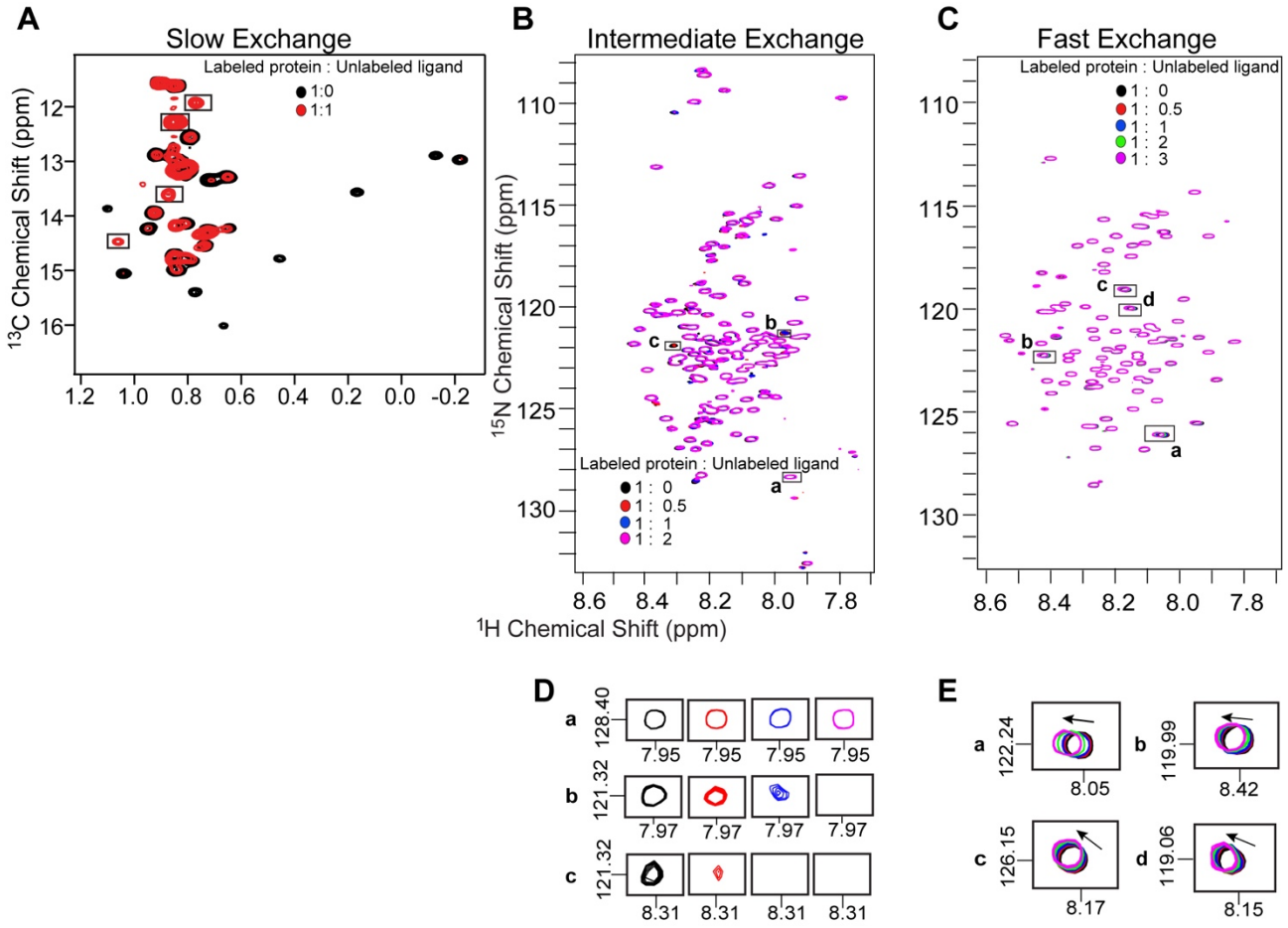

Figure 1-6. NMR exchange regimes. The binding affinity of the ligand dictates the type of NMR spectra. There are distinct NMR peaks observed for free and ligand-bound states of the protein. Dissociation constants are in (A) submicromolar (figure adapted from [58], (B) micromolar and (C) high micromolar to low millimolar ranges. The affected slow exchange peaks are shown in boxes in panel A. Selected affected peaks are shown for (D) intermediate exchange and (E) fast exchange.

1.2.2. Isotopic labeling to study protein-ligand interactions

The naturally occurring isotopes of nitrogen (^{14}N) and carbon (^{12}C) are NMR-inactive, whereas the most abundant hydrogen isotope (^{1}H) is NMR-active. Proteins are routinely labeled with stable, NMR-active isotopes (^{15}N, ^{13}C) to study protein-ligand interactions using NMR. This is achieved by expressing the protein in *E. coli* grown in minimal media supplemented with labeled ^{15}N and/or ^{13}C [52-54]. In my dissertation, I used three different NMR labeling schemes.

- **Uniform ^{15}N-labeling.** *E. coli* is grown in M9 minimal media supplemented with ^{15}N-ammonium chloride. This uniformly labels the backbone amide groups and sidechain amides of asparagine and glutamine [50, 53, 59].

- **Specific amino acid labeling.** The NMR spectra of large (MW ~ 25KDa or more) and intrinsically disordered proteins often show poor resolution and significant signal overlap. In such cases, the backbone amides of specific amino acids of the protein are labeled with ^{15}N by growing cells in media containing the specific ^{15}N-labeled amino acid. The remaining amino acids are supplied in the unlabeled form. This reduces signal complexity by minimizing interference from other, often poorly resolved signals [53, 56, 57].

- **Sidechain ^{13}C methyl (ILV) labeling.** The sidechain methyl groups of isoleucine ($^{13}C\delta1$; I), leucine ($^{13}C\delta$; L) and valine ($^{13}C\gamma$; V) are labeled using ^{13}C-labeled keto acids **(Figure 1-7)**. The ILV residues are present in the hydrophobic core of proteins, and many protein interactions involve the hydrophobic interface. This makes the ^{13}C-ILV methyl groups remarkable and sensitive probes to study protein-ligand interactions [53, 56, 60, 61]. The ILV-labeling can also significantly reduce signal complexity.

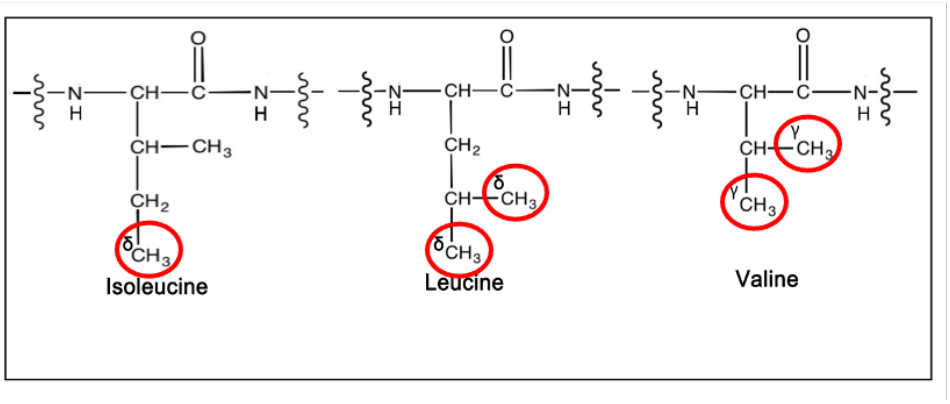

Figure 1-7. ILV-labeling of proteins. The sidechain methyl groups of isoleucine (Cδ1), leucine (Cδ) and valine (Cγ) are labeled. The labeled atoms are circled in red.

1.2.3. 2D NMR spectra and folding status of proteins

The 2D NMR spectrum of a protein provides an idea about the folding status of the protein. For a **fully folded** protein, the NMR peaks are well-dispersed [31], with peak dispersions greater than 1 ppm in the proton dimension **(Figure 1-8** and **Figure 1-9).** A **partially folded** protein will typically show peak dispersions less than 1 ppm in the proton dimension **(Figure 1-10** and **Figure 1-11)** [62]. A partially folded protein often has regions that are dynamic in nature. In such situations, the NMR spectrum shows less than the expected number of peaks [62]. In my dissertation, I used NMR to study the proteins SipD (Chapters 4 and 5) and IpaC (Chapters 2 and 3). SipD is a fully folded protein, while IpaC is partially folded.

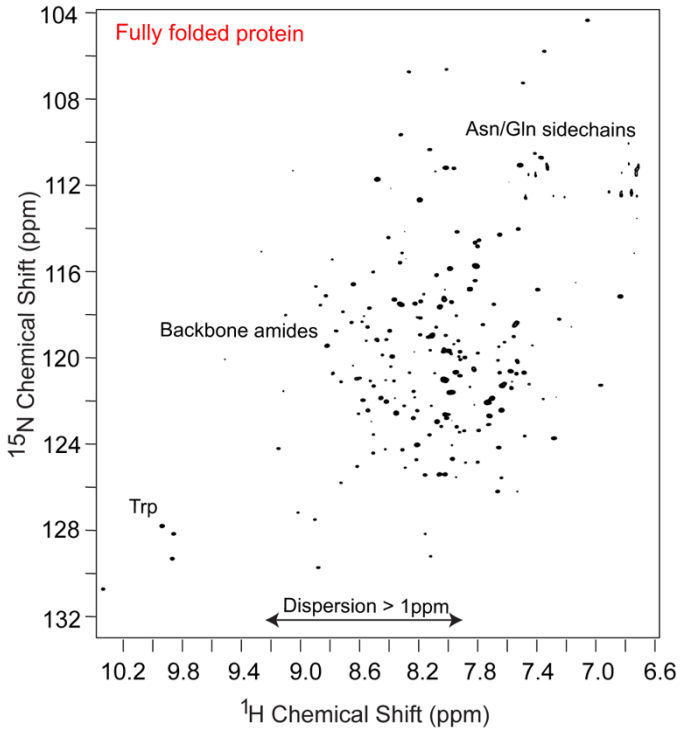

Figure 1-8. ^{1}H-^{15}N TROSY spectrum of a fully folded protein showing a peak dispersion greater than 1 ppm in the proton dimension. The backbone and sidechain amides of a protein are represented in this spectrum at specific regions.

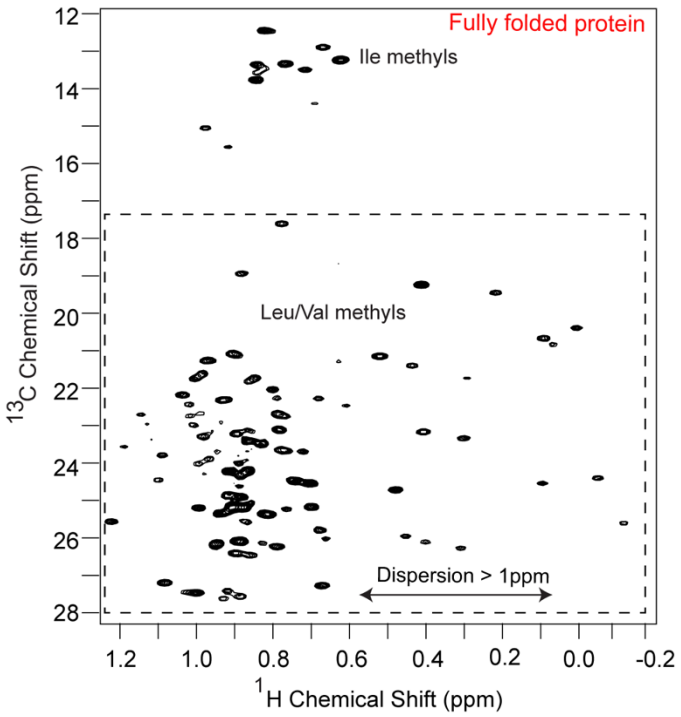

Figure 1-9. ^{1}H-^{13}C HSQC spectrum of a fully folded ILV-labeled protein showing a peak dispersion greater than 1 ppm in the proton dimension. The sidechain methyl groups of isoleucine, leucine and valine of a protein are represented in this spectrum at specific regions.

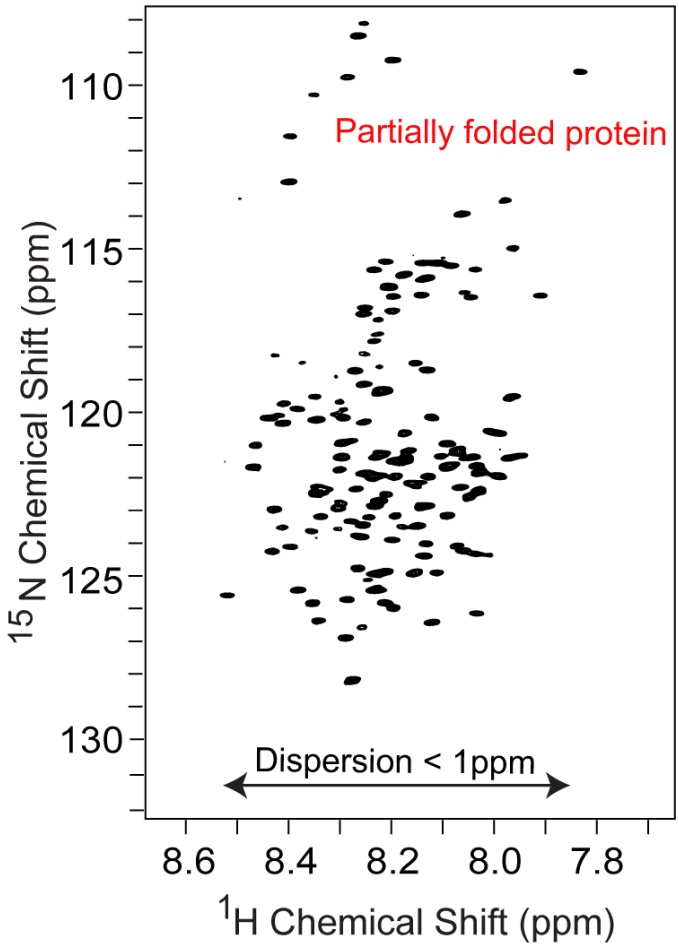

Figure 1-10. ^{1}H-^{15}N TROSY spectrum of a partially folded protein showing a peak dispersion less than 1 ppm in the proton dimension.

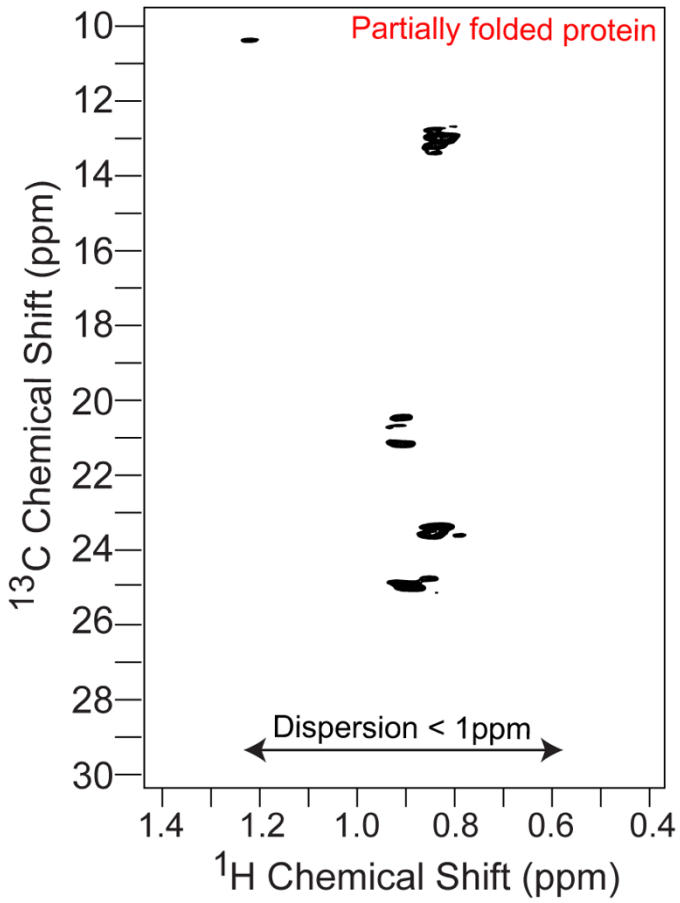

Figure 1-11. **^{1}H-^{13}C HSQC spectrum of a partially folded ILV-labeled protein** showing a peak dispersion less than 1 ppm in the proton dimension.

1.3. References

1. Hueck, C.J., *Type III protein secretion systems in bacterial pathogens of animals and plants.* Microbiol. Mol. Biol. Rev., 1998. **62**(2): p. 379-433.
2. Kubori, T., et al., *Supramolecular structure of the Salmonella typhimurium type III protein secretion system.* Science, 1998. **280**(5363): p. 602-605.
3. Cornelis, G.R., *The type III secretion injectisome.* Nat. Rev. Microbiol., 2006. **4**(11): p. 811-825.
4. Coburn, B., I. Sekirov, and B.B. Finlay, *Type III Secretion Systems and Disease*, in *Clin. Microbiol. Rev.* 2007.
5. CDC, *Antibiotic Resistance Threats in the United States, 2013.* 2013.
6. Wellington, E.M., et al., *The role of the natural environment in the emergence of antibiotic resistance in gram-negative bacteria.* Lancet Infect Dis, 2013. **13**(2): p. 155-65.
7. Davies, J. and D. Davies, *Origins and evolution of antibiotic resistance.* Microbiol Mol Biol Rev, 2010. **74**(3): p. 417-33.
8. McShan, A.C. and R.N. De Guzman, *The bacterial type III secretion system as a target for developing new antibiotics.* Chem. Biol. Drug. Des., 2015. **85**(1): p. 30-42.
9. Keyser, P., et al., *Virulence blockers as alternatives to antibiotics: type III secretion inhibitors against Gram-negative bacteria.* J Intern Med, 2008. **264**(1): p. 17-29.
10. Duncan, M.C., R.G. Linington, and V. Auerbuch, *Chemical inhibitors of the type three secretion system: disarming bacterial pathogens.* Antimicrob Agents Chemother, 2012. **56**(11): p. 5433-41.
11. Galan, J.E., et al., *Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells.* Annu. Rev. Microbiol., 2014. **68**: p. 415-38.
12. Loquet, A., et al., *Atomic model of the type III secretion system needle.* Nature, 2012. **486**(7402): p. 276-9.
13. Schraidt, O. and T.C. Marlovits, *Three-dimensional model of Salmonella's needle complex at subnanometer resolution.* Science, 2011. **331**(6021): p. 1192-5.
14. Worrall, L.J., et al., *Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body.* Nature, 2016. **540**(7634): p. 597-601.
15. Hu, B., et al., *In Situ Molecular Architecture of the Salmonella Type III Secretion Machine.* Cell, 2017. **168**(6): p. 1065-1074 e10.
16. Park, D., et al., *Visualization of the type III secretion mediated Salmonella-host cell interface using cryo-electron tomography.* Elife, 2018. **7**.
17. Dietsche, T., et al., *Structural and Functional Characterization of the Bacterial Type III Secretion Export Apparatus.* PLoS Pathog, 2016. **12**(12): p. e1006071.
18. Lara-Tejero, M., et al., *A sorting platform determines the order of protein secretion in bacterial type III systems.* Science, 2011. **331**(6021): p. 1188-91.
19. Chatterjee, S., et al., *Structure and biophysics of type III secretion in bacteria.* Biochemistry, 2013. **52**(15): p. 2508-17.
20. Lunelli, M., et al., *Crystal structure of PrgI-SipD: insight into a secretion competent state of the type three secretion system needle tip and its interaction with host ligands.* PLoS Pathog., 2011. **7**(8): p. e1002163.
21. Chatterjee, S., et al., *The crystal structures of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate.* Protein Sci, 2011. **20**(1): p. 75-86.
22. Russo, B.C., et al., *Intermediate filaments enable pathogen docking to trigger type 3 effector translocation.* Nat Microbiol, 2016. **1**: p. 16025.
23. Johnson, S., et al., *Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD.* J. Biol. Chem., 2007. **282**(6): p. 4035-4044.

24. Erskine, P.T., et al., *High Resolution Structure of BipD: An Invasion Protein Associated with the Type III Secretion System of Burkholderia pseudomallei.* J. Mol. Biol., 2006. **363**(1): p. 125-136.
25. Dey, S., et al., *The type III secretion system needle, tip, and translocon.* Protein Sci, 2019. **28**(9): p. 1582-1593.
26. Derewenda, U., et al., *The structure of Yersinia pestis V-antigen, an essential virulence factor and mediator of immunity against plague.* Structure, 2004. **12**(2): p. 301-306.
27. Chaudhury, S., et al., *Structure of the Yersinia pestis tip protein LcrV refined to 1.65 Å resolution.* Acta Crystallogr. Sect. F, 2013. **69**(Pt 5): p. 477-81.
28. Sani, M., et al., *IpaD is localized at the tip of the Shigella flexneri type III secretion apparatus.* Biochim. Biophys. Acta, 2007. **1770**(2): p. 307-11.
29. Espina, M., et al., *IpaD localizes to the tip of the type III secretion system needle of Shigella flexneri.* Infect. Immun., 2006. **74**(8): p. 4391-4400.
30. Veenendaal, A.K., et al., *The type III secretion system needle tip complex mediates host cell sensing and translocon insertion.* Mol. Microbiol., 2007. **63**(6): p. 1719-30.
31. McShan, A.C., et al., *NMR Identification of the Binding Surfaces Involved in the Salmonella and Shigella Type III Secretion Tip-Translocon Protein-Protein Interactions.* Proteins, 2016. **84**: p. 1097-1107.
32. Wang, Y., et al., *NMR characterization of the interaction of the Salmonella type III secretion system protein SipD and bile salts.* Biochemistry, 2010. **49**(19): p. 4220-6.
33. Barta, M.L., et al., *Identification of the bile salt binding site on IpaD from Shigella flexneri and the influence of ligand binding on IpaD structure.* Proteins, 2012. **80**(3): p. 935-45.
34. Dickenson, N.E., et al., *Conformational Changes in IpaD from Shigella flexneri upon Binding Bile Salts Provide Insight into the Second Step of Type III Secretion.* Biochemistry, 2011. **50**: p. 172-180.
35. McShan, A.C., et al., *Characterization of the Binding of Hydroxyindole, Indoleacetic acid, and Morpholinoaniline to the Salmonella Type III Secretion System Proteins SipD and SipB.* ChemMedChem, 2016. **11**(9): p. 963-971.
36. Dey, S., et al., *Characterization of Small-Molecule Scaffolds That Bind to the Shigella Type III Secretion System Protein IpaD.* ChemMedChem, 2017. **12**(18): p. 1534-1541.
37. Blocker, A., et al., *The tripartite type III secreton of Shigella flexneri inserts IpaB and IpaC into host membranes.* J.Cell Biol., 1999. **147**(3): p. 683-693.
38. Neyt, C. and G.R. Cornelis, *Insertion of a Yop translocation pore into the macrophage plasma membrane by Yersinia enterocolitica: requirement for translocators YopB and YopD, but not LcrG.* Mol. Microbiol., 1999. **33**(5): p. 971-81.
39. Schoehn, G., et al., *Oligomerization of type III secretion proteins PopB and PopD precedes pore formation in Pseudomonas.* EMBO J., 2003. **22**(19): p. 4957-67.
40. Ide, T., et al., *Characterization of translocation pores inserted into plasma membranes by type III-secreted Esp proteins of enteropathogenic Escherichia coli.* Cell. Microbiol., 2001. **3**(10): p. 669-79.
41. Delahay, R.M. and G. Frankel, *Coiled-coil proteins associated with type III secretion systems: a versatile domain revisited.* Mol Microbiol, 2002. **45**(4): p. 905-16.
42. Gazi, A.D., et al., *Coiled-coils in type III secretion systems: structural flexibility, disorder and biological implications.* Cell Microbiol, 2009. **11**(5): p. 719-29.
43. Dey, S., A. Basu, and S. Datta, *Characterization of molten globule PopB in absence and presence of its chaperone PcrH.* Protein J, 2012. **31**(5): p. 401-16.
44. Barta, M.L., et al., *The structures of coiled-coil domains from type III secretion system translocators reveal homology to pore-forming toxins.* J. Mol. Biol., 2012. **417**(5): p. 395-405.

45. Nguyen, V.S., et al., *Structure of AcrH-AopB Chaperone-Translocator Complex Reveals a Role for Membrane Hairpins in Type III Secretion System Translocon Assembly.* Structure, 2015. **23**(11): p. 2022-31.
46. Russo, B.C., J.K. Duncan, and M.B. Goldberg, *Topological Analysis of the Type 3 Secretion System Translocon Pore Protein IpaC following Its Native Delivery to the Plasma Membrane during Infection.* MBio, 2019. **10**(3).
47. Harrington, A.T., et al., *Structural characterization of the N terminus of IpaC from Shigella flexneri* Infect. Immun., 2003. **71**(3): p. 1255-1264.
48. Page, A.-L., et al., *Characterization of the interaction partners of secreted proteins and chaperones of Shigella flexneri.* Mol. Microbiol., 2001. **42**(4): p. 1133-1145.
49. Davis, R., et al., *Protein-protein interactions in the assembly of Shigella flexneri invasion plasmid antigens IpaB and IpaC into protein complexes.* Biochim.Biophys.Acta, 1998. **1429**(1): p. 45-56.
50. Vaynberg, J. and J. Qin, *Weak protein-protein interactions as probed by NMR spectroscopy.* Trends Biotechnol, 2006. **24**(1): p. 22-7.
51. Perkins, J.R., et al., *Transient protein-protein interactions: structural, functional, and network properties.* Structure, 2010. **18**(10): p. 1233-43.
52. Takeuchi, K. and G. Wagner, *NMR studies of protein interactions.* Curr Opin Struct Biol, 2006. **16**(1): p. 109-17.
53. Marintchev, A., D. Frueh, and G. Wagner, *NMR methods for studying protein-protein interactions involved in translation initiation.* Methods Enzymol, 2007. **430**: p. 283-331.
54. Bieri, M., et al., *Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination.* FEBS J, 2011. **278**(5): p. 704-15.
55. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Prog. Nucl. Magn. Reson. Spectrosc., 2013. **73**: p. 1-16.
56. Barrett, P.J., et al., *The quiet renaissance of protein nuclear magnetic resonance.* Biochemistry, 2013. **52**(8): p. 1303-20.
57. Kleckner, I.R. and M.P. Foster, *An introduction to NMR-based approaches for measuring protein dynamics.* Biochim Biophys Acta, 2011. **1814**(8): p. 942-68.
58. Chaudhury, S., et al., *The LcrG tip chaperone protein of the Yersinia pestis type III secretion system is partially folded.* J. Mol. Biol., 2015. **427**: p. 3096-3109.
59. Fernandez, C. and G. Wider, *TROSY in NMR studies of the structure and function of large biological macromolecules.* Curr Opin Struct Biol, 2003. **13**(5): p. 570-80.
60. Tugarinov, V. and L.E. Kay, *Methyl Groups as Probes of Structure and Dynamics in NMR Studies of High-Molecular-Weight Proteins.* ChemBioChem, 2005. **6**(9): p. 1567-1577.
61. Hajduk, P., D. Augeri, and J. Mack, *NMR-based screening of proteins containing ^{13}C-labeled methyl groups.* J. Am. Chem. Soc., 2000. **122**: p. 7898-7904.
62. Konrat, R., *NMR contributions to structural dynamics studies of intrinsically disordered proteins.* J Magn Reson, 2014. **241**: p. 74-85.

Chapter 2: The *Shigella* Minor Translocon Protein IpaC and its Domains Are Intrinsically Disordered and Undergo Conformational Changes in Micelles

2.1. Abstract

IpaC is the minor translocon protein of *Shigella*, and plays a critical role in virulence. Data regarding the detailed biophysical characterization of IpaC in atomic resolution, as well as studies on its conformational changes in lipid environments remain sparse. In this chapter, I show by NMR that full-length IpaC, as well as its N-terminal domain (NTD) and C-terminal domain (CTD), lack overall tertiary structures. I have also used circular dichroism (CD) spectroscopy and NMR to show that full-length IpaC and its domains undergo conformational changes in n-dodecylphosphocholine (DPC), lyso-myristoylphosphatidylglycerol (LMPG) and sodium dodecyl sulfate (SDS) micelles. This knowledge is important in understanding the biophysical characteristics of the IpaC protein family and can be used in future for detailed structural characterization of the protein.

2.2. Introduction

Shigella is a Gram-negative pathogen that causes bacillary dysentery [1] and uses the type III secretion system (T3SS) to gain entry into host epithelial cells by injecting virulence effector proteins into the host cytoplasm [2]. The development of antibiotic resistance in *Shigella* and other T3SS pathogens such as *Salmonella*, *Yersinia*, *Chlamydia* and *Pseudomonas* [3, 4] is a global health concern. This problem is intensified by the low number of new antibiotics in the pipeline [5, 6]. The T3SS plays an essential role in virulence, and understanding its assembly could lead to the development of new antimicrobials.

The structural component of the T3SS is the needle complex, which consists of a base, an extracellular needle, a tip complex, and a translocon [7, 8]. In *Shigella*, several copies of the

tip protein IpaD interact to form a tip complex, which serves as a platform for the assembly of the translocon (Kawaljit Kaur, University of Kansas 2016; Dissertation). IpaC is the minor translocon protein in *Shigella*. The N-terminal and C-terminal domains of IpaC have been shown to interact with other proteins of the T3SS as well as host cell components [9-16]. Previous results also showed that IpaC interacts with liposomes [17] and phospholipid membranes [18]. These suggest that IpaC plays a central role in the protein-protein interaction network responsible for pathogenesis.

Wright and Dyson reported [19] that proteins participating in diverse protein-protein interactions in the cell are often intrinsically disordered. An intrinsically disordered protein lacks a well-defined three-dimensional structure. An intrinsically disordered protein (IDP) contains disordered sequences [19] and can exist in multiple conformational states ranging from completely extended, or regions with transient or partial secondary structures, to a more compact globule [20]. The folding pathway of proteins often proceeds through an intermediate state called the molten globule [20] (**Figure 2-1**), that has defined secondary structures without an overall tertiary structure [21]. The atomic structure for any minor translocon protein and its conformational changes in membrane mimetics and micelles are currently unknown. In this chapter, I report using circular dichroism (CD) spectroscopy and NMR that full-length IpaC, as well as its N-terminal domain (NTD) and C-terminal domain (CTD) are IDPs. I have also used CD spectroscopy and NMR to show that full-length IpaC and its domains undergo conformational changes in DPC, LMPG and SDS micelles. This knowledge is important for the detailed biophysical characterization of IpaC.

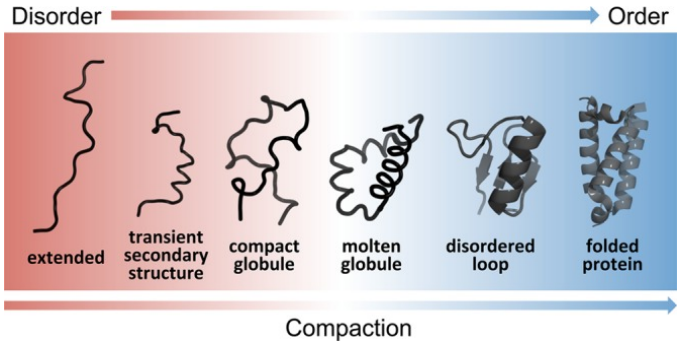

Figure 2-1. Models of protein structures. Proteins can typically exist in multiple conformations between the disordered (red) and ordered (blue) states. The molten globule is an intermediate in the pathway, having well defined secondary structure without a tertiary structure. The figure is adapted from van der lee *et al* [20].

2.3. Methods

2.3.1. Expression and purification of proteins

Full-length IpaC was over-expressed and purified as described previously [13] with modifications. Full-length IpaC was subcloned in the NdeI/XhoI sites of pET 22b (Yu Wang, University of Kansas 2009; Dissertation), which appended a non-cleavable 8-residue tag (LEH_6) at the C-terminus. The plasmid expressing IpaC was transformed into *E. coli* BL21(DE3) Codon Plus competent cells (Agilent) and grown in culture media containing carbenicillin (100 μg/mL) (Kawaljit Kaur, University of Kansas 2016; Dissertation) and chloramphenicol (25 μg/mL). Unlabeled IpaC was obtained by expression in 4 L of TB (Terrific Broth) medium. ^{15}N-labeled IpaC for NMR was obtained by growing cells in M9 minimal medium supplemented with 1 g/L of ^{15}N-ammonium chloride (Sigma). Cells were grown at 37°C and induced with 1 mM

isopropyl-β-D-thiogalactopyranoside (IPTG) at OD_{600} ~ 0.8 (Kawaljit Kaur, University of Kansas 2016; Dissertation). Cell growth was continued at 25°C overnight until OD_{600} ~2.9 for TB medium or ~2.3 for M9 medium. Cells were harvested by centrifugation at 4,000 rpm (2,392×g) for 10 min, resuspended in binding buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 1 mM phenylmethylsulfonyl fluoride) and lysed by sonication. The cell lysate was centrifuged at 13,000 rpm (13,865×g) for 10 min, and 700 μL of 5% polyethyleneimine (PEI) was added to the supernatant to precipitate the nucleic acids. The mixture was centrifuged at 13,000 rpm (13,865×g) for 10 min. The supernatant was discarded and the pellet, which contained inclusion bodies of IpaC, was resuspended in binding buffer containing 6 M urea to solubilize the protein. The solubilized IpaC was incubated by rocking at room temperature for 30 min, and centrifuged at 18,000 rpm (39,000×g) for 20 min. The supernatant was incubated with 25 mL of Ni^{2+}-affinity resin (Qiagen, # 30230) for 30 min. The resin was washed 3× with 150 mL of binding buffer containing 6 M urea. IpaC was eluted with 50 mL of elution buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM imidazole) containing 6 M urea. The denatured IpaC was refolded by stepwise dialysis in buffer without urea (20 mM Tris-HCl pH 8.0, 500 mM NaCl) and then in buffer (20 mM sodium phosphate, pH 6.8, 50 mM NaCl).

$IpaC^{1-100}$, the N-terminal domain (NTD), was expressed and purified as described above for full-length IpaC. $IpaC^{1-100}$ also formed inclusion bodies upon expression.

$IpaC^{173-363}$, the C-terminal domain (CTD), was expressed under native conditions as a fusion protein. IpaC residues 173-363 were subcloned in the NdeI/BamHI site of the expression plasmid pDZ1[22], forming a fusion protein with an N-terminal His_6 tag for purification, a GB1 domain as a solubility enhancer, a tobacco etch virus (TEV) protease cleavage site, and $IpaC^{173-363}$. The expression plasmid for $IpaC^{173-363}$ was transformed into *E. coli* BL21(DE3) Codon

Plus competent cells and grown in culture media containing carbenicillin (100 μg/mL) and chloramphenicol (25 μg/mL). Unlabeled IpaC$^{173-363}$ was obtained by cell growth in TB medium, and ^{15}N-labeled IpaC$^{173-363}$ was obtained by cell growth in M9 medium, as above. Cells were grown at 37°C, induced with 1 mM IPTG at an OD_{600} ~ 0.8, and cell growth was continued at 25°C overnight until OD_{600} ~ 2.9 for LB ~ 2.6 for M9. Cells were harvested by centrifugation (4,000 rpm, 2392×g, 10 min), resuspended in binding buffer, and sonicated to lyse the cells. The cell lysate was centrifuged (13,000 rpm, 13,865xg, 10 min) to remove cellular debris, and to the supernatant was added 700 μL of 5% PEI to precipitate the nucleic acids, followed by centrifugation (13,000 rpm, 13,865×g, 10 min). The supernatant was passed through a 25 mL Ni^{2+}-affinity resin (Qiagen, # 30230), followed by 150 mL of binding buffer. IpaC$^{173-363}$ was eluted with 50 mL of elution buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM imidazole). To cleave IpaC$^{173-363}$ from the fusion tag, the elution fractions were incubated in buffer (50 mM Tris-HCl pH 8.0, 20 mM NaCl, 0.5 mM EDTA, 1 mM DTT) at room temperature in 250 μL of 0.04 mM recombinant TEV protease [23]. The digest was passed through a 25 mL Ni^{2+}-affinity resin. The His_6-GB1 tag bound to the resin, while IpaC$^{173-363}$ was collected and dialyzed in buffer (20 mM sodium phosphate pH 6.8, 50 mM NaCl). All the proteins were concentrated using Amicon Ultra 3K filter (Millipore) and protein concentration was determined by A_{280}.

The sidechain methyl groups of isoleucine, leucine and valine of full-length IpaC, IpaC$^{1-100}$ and IpaC$^{173-363}$ were labeled (ILV-labeling) according to protocols described previously [24, 25] for *Salmonella* tip protein SipD with minor modifications. Briefly, cells were grown at 37°C, and at OD_{600} ~ 0.4, the culture medium was supplemented with 60 mg/L of 2-ketobutyric acid-4-^{13}C acid sodium salt hydrate (Sigma, #571342) to label the Cδ1 methyl group of isoleucine with ^{13}C and 100 mg/L of 2-keto-3-(methyl-^{13}C)-butyric-4-^{13}C acid sodium salt (Sigma, #571334) to label the leucine Cδ and valine Cγ methyl groups with ^{13}C. Cells were induced with 1 mM

isopropyl-β-D-thiogalactopyranoside (IPTG) at an OD_{600} ~ 0.8, and cell growth was continued at 25°C overnight until OD_{600} ~ 2.6.

Two other IpaC constructs, $IpaC^{1-173}$ and $IpaC^{101-363}$, were also successfully subcloned as described above for full-length IpaC, $IpaC^{1-100}$ and $IpaC^{173-363}$. These constructs were designed to contain the central transmembrane domain. However, protein expression was unsuccessful from both of these constructs.

2.3.2. Preparation of detergent stock solutions

The detergents DPC (Avanti Polar Lipids # 850336), LMPG (Avanti Polar Lipids # 858120) and SDS (Midsci #M107) were dissolved in water (for CD) or in buffer for NMR (20 mM sodium phosphate pH 6.8, 50 mM NaCl) to prepare stock solutions of 200 mM. The DPC and LMPG stocks were stored at 4^0C and the SDS stock was stored at room temperature for future use.

2.3.3. Circular Dichroism (CD) Spectroscopy

A Jasco J-815 spectropolarimeter was used to collect far-UV CD data. Protein samples were diluted to a final concentration of 0.1-0.2 μM (final volume of 3 mL) with water. CD spectra were acquired in triplicate at 20^0C with a scan speed of 50 nm/min (Kawaljit Kaur, University of Kansas 2016; Dissertation). Thermal denaturation was monitored by the change in the molar ellipticity at 222 nm over a temperature range of 20^0C to 75^0C using an increasing temperature gradient of 5^0C/min.

2.3.4. NMR spectroscopy

NMR data were acquired using a Bruker Avance 800 MHz NMR spectrometer equipped with a cryogenic triple resonance probe, processed with NMRPipe [26], and analyzed using NMRView [27]. Two-dimensional ^{1}H-^{15}N TROSY and ^{1}H-^{13}C HSQC spectra were acquired at 20ºC using 0.1 mM of ^{15}N-labeled IpaC, IpaC$^{1-100}$ and IpaC$^{173-363}$ in NMR buffer (20 mM sodium phosphate pH 6.8, 50 mM NaCl containing 10% D_2O). The acquisition parameters for ^{1}H-^{13}C HSQC were 32 scans, a 22 ppm ^{13}C sweep width with the center at 15 ppm and ^{1}H sweep width of 12 ppm centered at 4.7 ppm, whereas typical ^{1}H-^{15}N acquisition parameters were 48 scans, with a ^{15}N sweep width of 30 ppm centered at 118 ppm; and ^{1}H sweep width of 18 ppm centered at 4.7 ppm.

NMR data were also acquired using 0.1 mM of each of ^{15}N-labeled IpaC, IpaC$^{1-100}$ and IpaC$^{173-363}$ in presence of 25 mM DPC, 3 mM LMPG and 20 mM SDS. The acquisition parameters were identical to those described above for the proteins without the detergents.

2.4. Results

2.4.1. Expression and purification of full-length IpaC, IpaC$^{1-100}$ and IpaC$^{173-363}$

The N-terminal and C-terminal domains of IpaC were previously shown to interact with other proteins of the *Shigella* T3SS as well as host cell components [9-16]. To characterize IpaC and its domains, I subcloned full-length IpaC and four different truncated IpaC constructs – IpaC$^{1-100}$, IpaC$^{101-363}$, IpaC$^{1-173}$ and IpaC$^{173-363}$ **(Figure 2-2)**. IpaC$^{1-173}$ and IpaC$^{101-363}$ did not express. I overexpressed and purified full length IpaC **(Figure 2-3A)**, IpaC$^{1-100}$ (NTD) **(Figure**

2-3B) and IpaC$^{173-363}$ (CTD) **(Figure 2-4)**. IpaC$^{173-363}$ expressed as a soluble protein. However, IpaC and IpaC$^{1-100}$ expressed as inclusion bodies, which required denaturation and solubilization in 6 M urea followed by stepwise dialysis to remove urea. Full-length IpaC precipitated above a concentration of 0.2 mM. The precipitation observed in full-length IpaC may be due to aggregation caused by interactions among neighboring C-terminal domains due to the presence putative coiled-coils [28] involving residues 319-345 **(Figure 2-2)**.

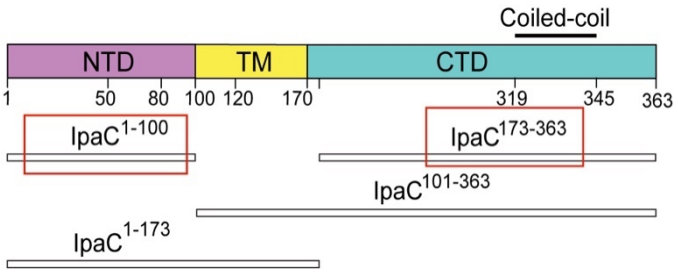

Figure 2-2. Domain map of IpaC. IpaC$^{1-100}$ (NTD), and IpaC$^{173-363}$ (CTD) (red boxes) along with full-length IpaC were overexpressed and purified. IpaC$^{1-173}$ and IpaC$^{101-363}$ did not express.

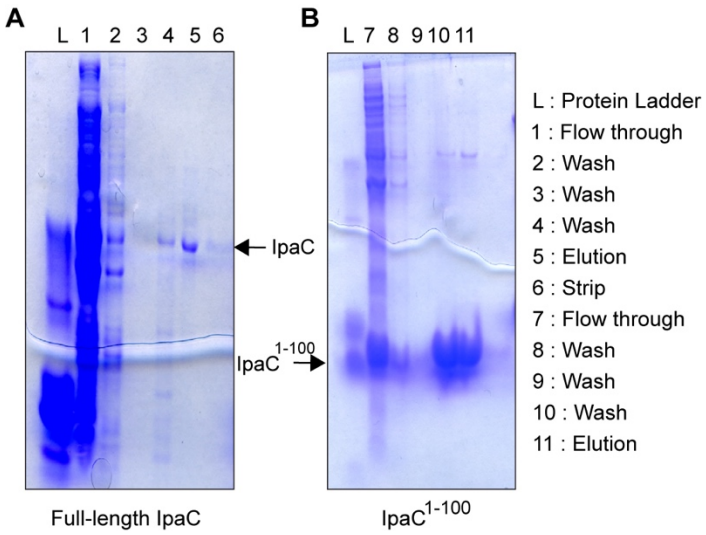

Fig. 2-3: SDS-PAGE of full-length IpaC and IpaC$^{1-100}$ purification. ^{15}N/ILV-labeled full-length IpaC and IpaC$^{1-100}$ were purified by Ni^{2+}-affinity chromatography. (A) Lane 5 shows the purified full-length IpaC (40.8 KDa) and (B) Lanes 10 and 11 show the purified IpaC$^{1-100}$ (12.3 KDa).

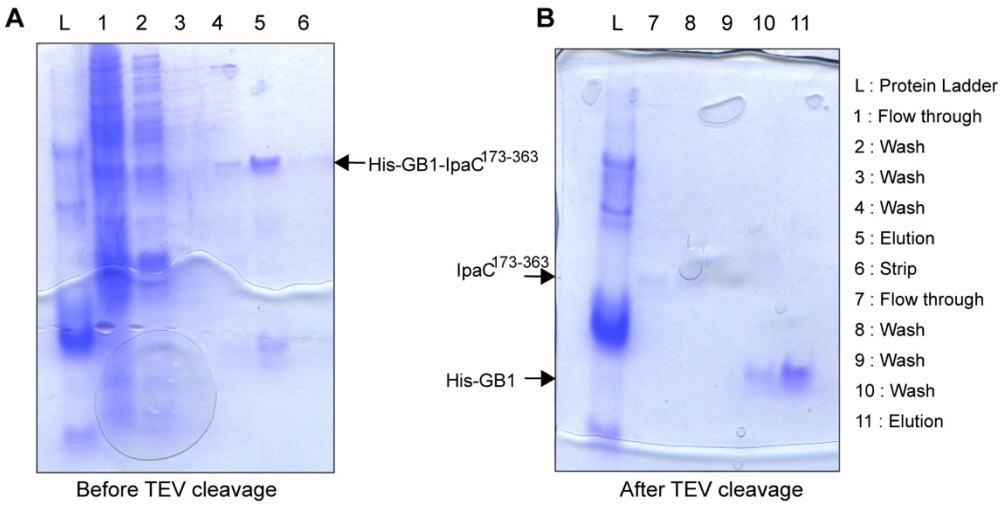

Figure 2-4. SDS-PAGE of IpaC$^{173-363}$ purification. ^{15}N/ILV-labeled IpaC$^{173-363}$ was purified by Ni^{2+}-affinity chromatography. (A) Lane 5 shows the eluted His-GB1-IpaC$^{173-363}$ (29.9 KDa) with a contaminant. **(B)** Lane 7 shows the purified IpaC$^{173-363}$ (20.9 KDa) after cleavage with TEV protease and another round of Ni^{2+}-affinity chromatography. Lanes 10 and 11 show the cleaved His-GB1.

2.4.2. Far-UV CD spectroscopy of full-length IpaC and its domains

PSIPRED [29] is a server to analyse the secondary structure of a protein. PSIPRED [29] analysis showed that full-length IpaC contains α-helical and disordered regions (data not shown). The far-UV CD spectrum of freshly prepared full-length IpaC **(Figure 2-5A)** showed two minima at 208 and 222 nm, which are characteristic of α-helical proteins. The ratio of the molar ellipticities at 222 nm and 208 nm (θ_{222}/ θ_{208}) was 1.07, suggesting the presence of interacting α-helices [30, 31] in contrast to the ratio of ~ 0.9 found in proteins with non-interacting or single-stranded α-helices [32, 33]. However, IpaC was previously reported to exist predominantly in a random coil conformation [17, 34]. Those spectra were acquired under different buffer conditions [17, 34] and the samples often contained 200 mM urea [34], indicating that change in buffer composition alters the overall secondary structure of IpaC. The thermal denaturation profile **(Figure 2-5B)** of full-length IpaC showed that the protein lacks an overall tertiary structure because of the absence of a sharp transition with no well-defined melting temperature (T_m), as seen in well-folded proteins [31].

In contrast, the far-UV CD spectra of $IpaC^{1-100}$ and $IpaC^{173-363}$ **(Figure 2-6)** exhibited prominent minima at 200 nm (for $IpaC^{1-100}$) and 202 nm ($IpaC^{173-363}$). $IpaC^{173-363}$ also showed slight negative dip at 222 nm. This indicates that both the NTD and CTD of IpaC exist predominantly as random coils in solution and $IpaC^{173-363}$ also contains some helical character. Overall, far-UV CD spectra (**Figure 2-5** and **Figure 2-6**) suggest that separating the NTD and CTD from full-length IpaC disrupts the overall secondary structure of IpaC.

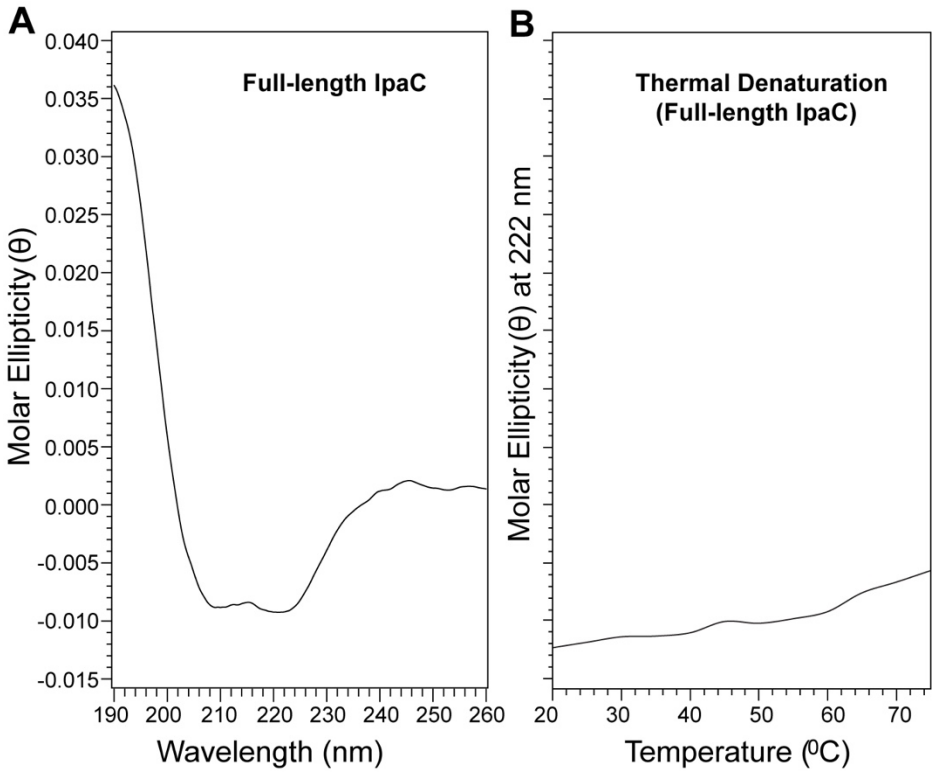

Figure 2-5. Far-UV CD spectrum of full-length IpaC. (A) The far-UV spectrum of full-length IpaC shows that the protein exists as an α-helix in solution. (B) CD thermal denaturation plot of full-length IpaC.

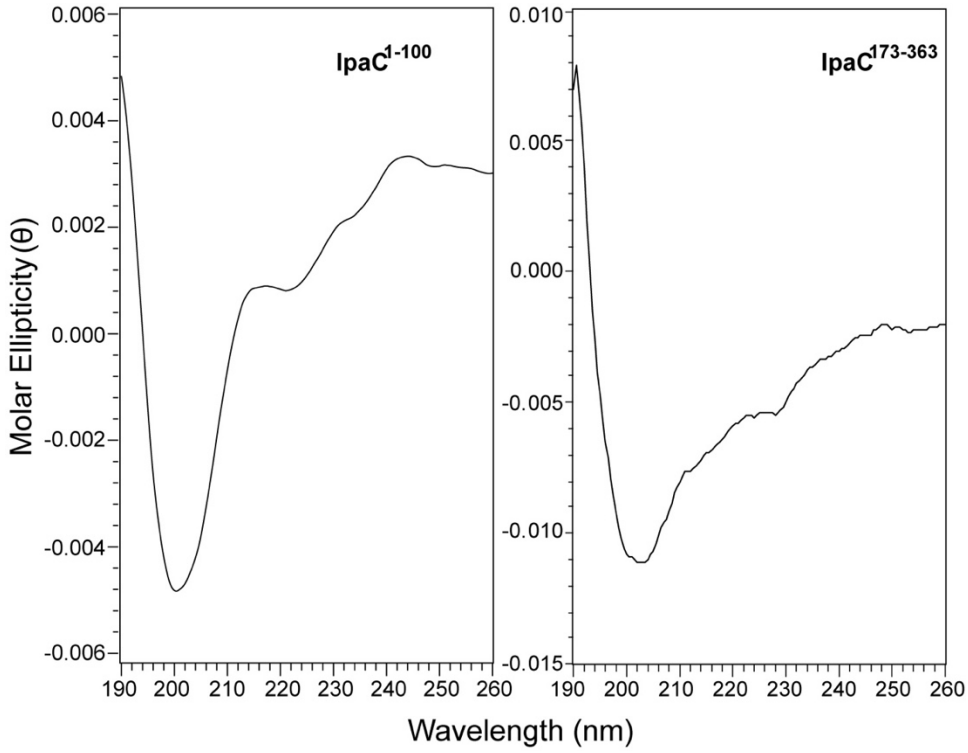

Figure 2-6. Far-UV CD spectra of IpaC domains show that IpaC$^{1-100}$ (NTD) and IpaC$^{173-363}$ (CTD) exist predominantly as random coils in solution. IpaC$^{173-363}$ also contains some helical characters.

2.4.3. NMR spectroscopy of full-length IpaC and its domains

The 2D ^{1}H-^{15}N TROSY spectrum of full length IpaC **(Figure 2-7A)** showed about 181 backbone amide peaks, representing only about 50% of the expected backbone amide peaks for the 363-residue IpaC. Half of the possible IpaC backbone amides were therefore missing in the 2D ^{1}H-^{15}N TROSY of IpaC. This suggests differential backbone dynamics for different segments or regions of the protein. Aggregation was ruled out as a possibility for the missing backbone amide peaks as the remaining 181 peaks were sharp and distinct. The N-terminal construct, $IpaC^{1-100}$, consisting of 108 residues (including the 8 residues as cloning artefact) showed 106 backbone amide peaks in the 2D ^{1}H-^{15}N TROSY spectrum **(Figure 2-7B)**. Thus, nearly all the backbone amide peaks of $IpaC^{1-100}$ were visible in the 2D ^{1}H-^{15}N TROSY spectrum **(Figure 2-7B)**. For the C-terminal domain, 148 peaks, representing 79% of the total number of expected backbone amide peaks, were present in the 2D ^{1}H-^{15}N TROSY spectrum of $IpaC^{173-363}$ **(Figure 2-7C)**. The 2D ^{1}H-^{15}N TROSY spectra of IpaC, $IpaC^{1-100}$, and $IpaC^{173-363}$ **(Figure 2-7)** showed peak dispersions of less than 1 ppm in the proton dimension, indicating that IpaC and its two domains are intrinsically disordered. This finding is supported by the 2D ^{1}H-^{13}C NMR spectra of the three proteins **(Figure 2-8)** showing only ^{13}C peaks from the sidechain methyl groups of isoleucine, leucine and valine (ILV). Very few, poorly resolved ILV peaks were visible **(Figure 2-8)**. These are currently the only available 2D NMR spectra for a minor translocon protein of the type III secretion system, and they are important because they can be used to study the molecular interactions of IpaC. The amino acid sequence of full-length IpaC (shown below) indicates that 194 residues out of 363 are disorder-promoting (shown in red) [35], consisting of 9 Arg, 20 Glu, 7 Pro, 35 Ala, 17 Gly, 29 Gln, 50 Ser and 27 Lys. Only 85 order-promoting residues (shown in green) [35] are present, consisting of 4 Tyr, 2 Phe, 39 Leu, 11 Val and 29 Ile.

Amino acid sequence of full-length IpaC

MEIQNTKPTQTLYTDISTKQTQSSSETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQ
VQHDSEIIARLTDKKMKDLSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLS
LIAFDATKSAAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEK
ELAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDKIDTQRRT
YELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASSKQAEEASQVSKEA
SQATNQLIQKLLNIIDSINQSKNSAASQIAGNIRA

Green - Order-promoting residues [35] (85 out of 363)
Red – Disorder-promoting residues [35] (194 out of 363)

On comparing the 2D ^{1}H-^{15}N TROSY spectra of the three proteins **(Figure 2-7)**, I found that 39 NMR peaks from the spectrum of IpaC$^{1-100}$ **(Figure 2-7B)** can be superimposed on the spectrum of full-length IpaC **(Figure 2-7A)**. Likewise, 14 NMR peaks from the spectrum of IpaC$^{173-363}$ **(Figure 2-7C)** can be superimposed on the spectrum of full-length IpaC **(Figure 2-7A)**. This finding supports the far-UV CD data (**Figure 2-5A** and **Figure 2-6**), indicating that in solution, the NTD and CTD of IpaC individually adopt tertiary structures that are different from that of full-length IpaC.

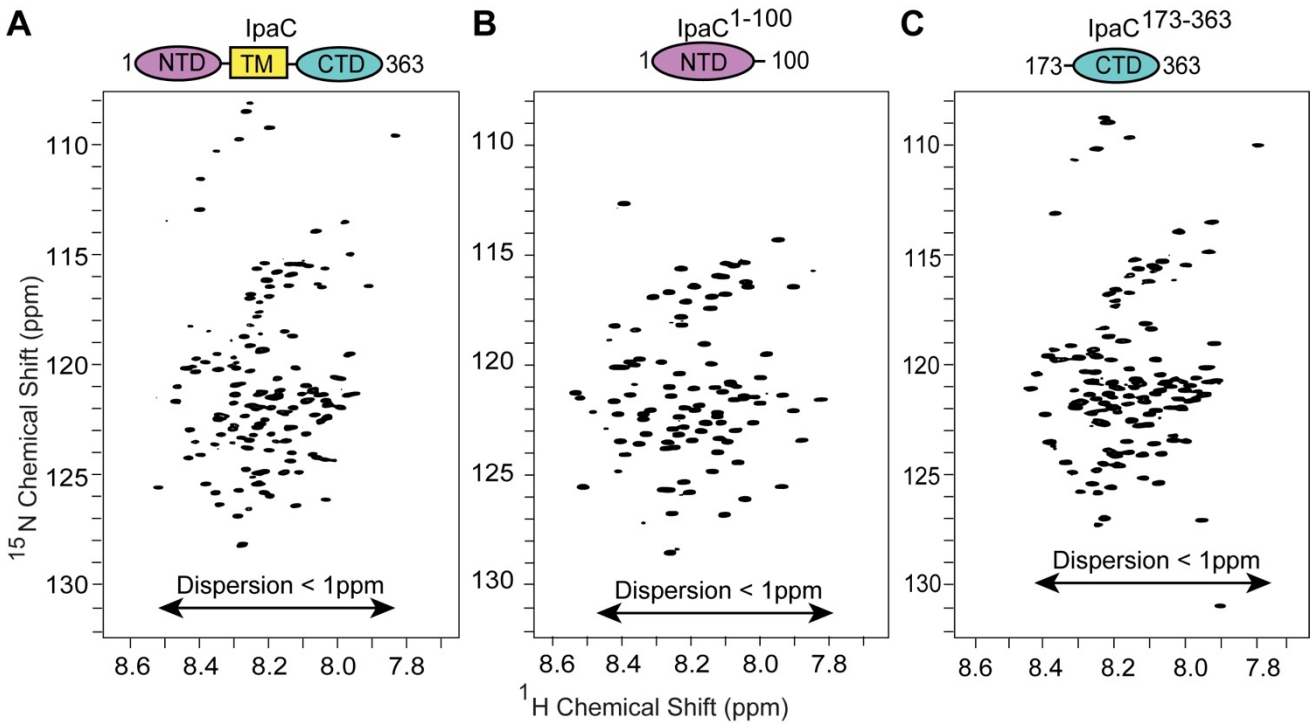

Figure 2-7. 2D ^{1}H-^{15}N TROSY spectra of IpaC. (A) Full-length IpaC, (B) IpaC$^{1-100}$ and (C) IpaC$^{173-363}$ are intrinsically disordered because the peak dispersions are less than 1 ppm.

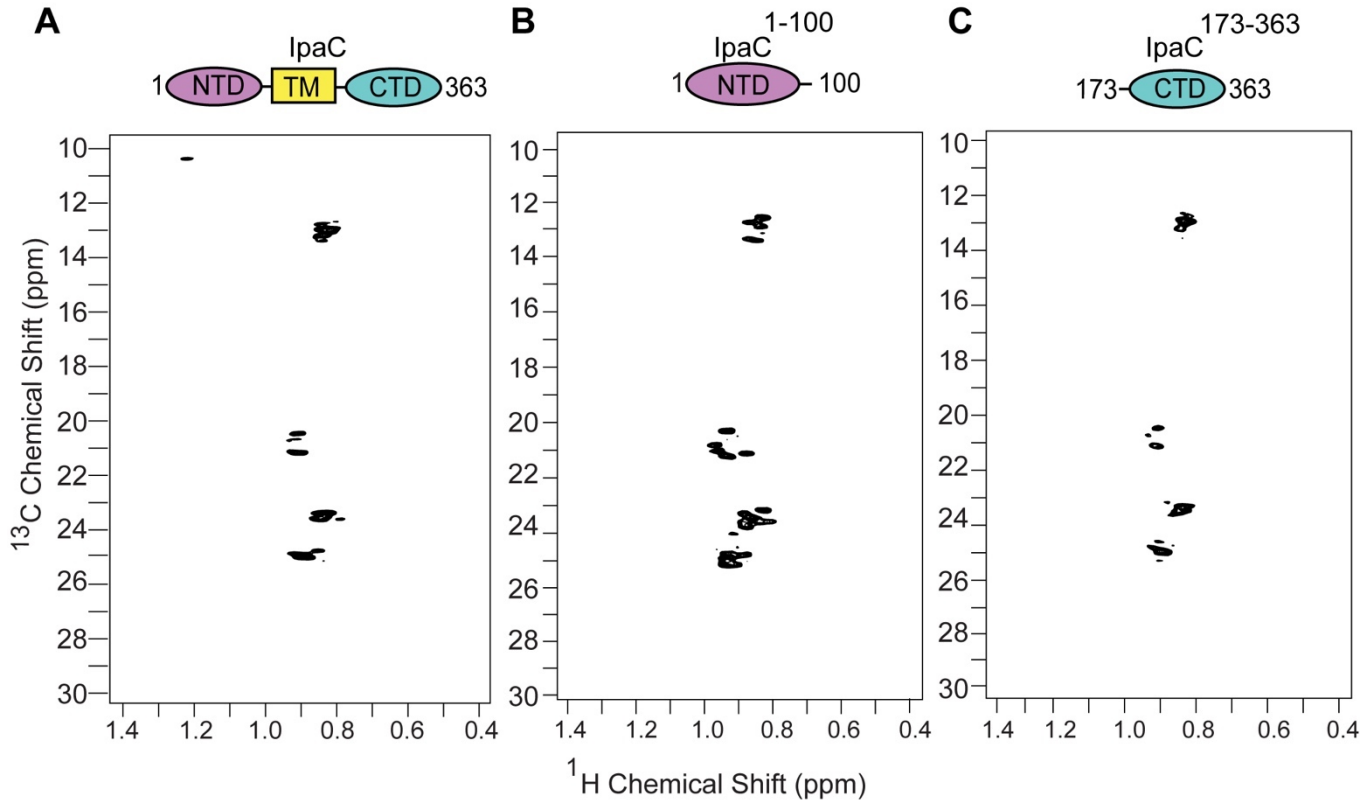

Figure 2-8. 2D ^{1}H-^{13}C HSQC spectra of IpaC. (A) Full-length IpaC, (B) IpaC$^{1-100}$ and (C) IpaC$^{173-363}$ are intrinsically disordered because the peak dispersions are less than 1 ppm. The peaks are very few in number and are poorly resolved.

2.4.4. IpaC undergoes conformational change in micelles

Previous results showed that IpaC interacts with liposomes [17] and phospholipid membranes [18]. However, there are no published data on the conformational changes of IpaC or its domains in the presence of surfactants or membrane mimetics. I used DPC, LMPG and SDS **(Figure 2-9)** to determine how the conformation of IpaC changes in presence of micelles. DPC, LMPG and SDS form micelles in solution, and can solubilize membrane proteins at concentrations above their respective critical micellar concentration (CMC). The CMC is the concentration of a detergent at which micelles form. The CMC of the detergents used here as follows: DPC, 1.1 mM; LMPG, 0.3 mM; SDS, 8 mM. I recorded the ^{1}H-^{15}N TROSY spectra of full-length IpaC, IpaC$^{1-100}$ and IpaC$^{173-363}$ in absence and presence of 25 mM DPC, 3 mM LMPG and 20 mM SDS. All these detergent concentrations were above their respective CMC. I also used far-UV CD spectroscopy to determine the secondary structural changes in these proteins in presence of increasing concentrations of DPC, LMPG and SDS. Since SDS is strongly ionic and is known to induce α-helix formation in proteins [36-38], I also used the zwitterionic detergent DPC and the mildly ionic detergent LMPG.

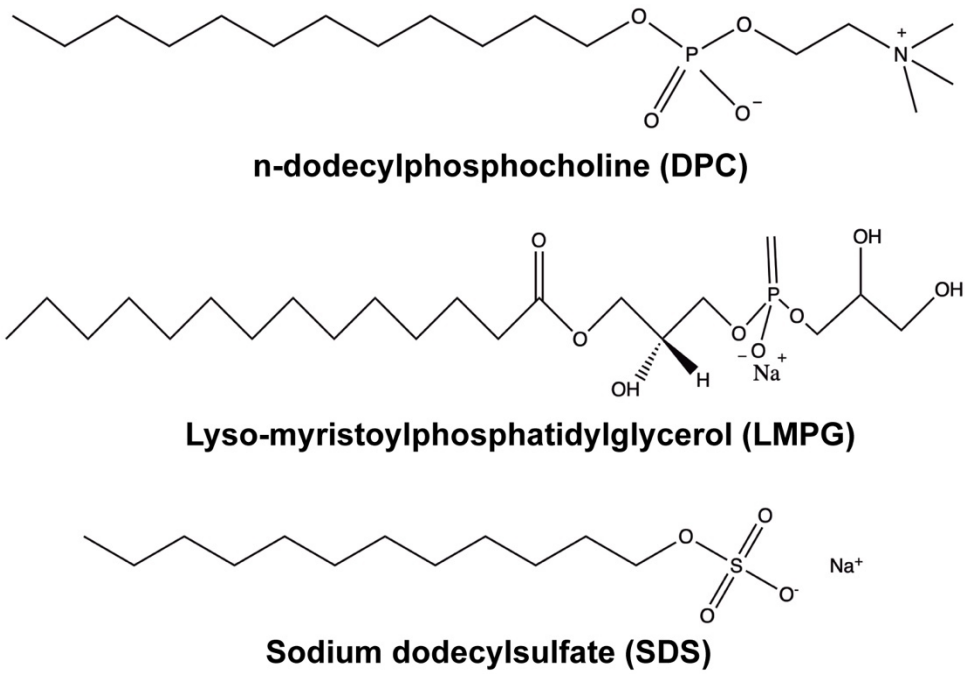

Figure 2-9. Molecular structures of the detergents used in this study. DPC and LMPG are neutral, while SDS is an anionic surfactant.

The ^{1}H-^{15}N TROSY data showed decrease in peak intensity and changes in peak positions for many residues of full-length IpaC (shown by arrowheads) **(Figure 2-10A)** in presence of 25 mM DPC. This suggests that full length IpaC undergoes conformational changes and interacts with DPC micelles. However, the peak dispersion along the proton dimension remained unchanged (less than 1 ppm), indicating that full-length IpaC remained intrinsically disordered in presence of DPC **(Figure 2-10A)**. I also recorded the far-UV CD spectra of full-length IpaC in presence of DPC **(Figure 2-10B)** to understand the changes in IpaC secondary structure. Titration with increasing concentrations of DPC showed an increase in α-helical content of full-length IpaC, characterized by increased negative dips at 208 nm and 222 nm **(Figure 2-10B)**. Similarly, ^{1}H-^{15}N TROSY spectra of full-length IpaC in presence of 3 mM LMPG **(Figure 2-11A)** and 20 mM SDS **(Figure 2-11B)** showed peak broadening and peak shifts for specific IpaC peaks (shown by arrowheads), suggesting conformational changes of IpaC and interaction with these micelles. The $IpaC^{1-100}$ and $IpaC^{173-363}$ constructs of IpaC also showed similar effects in their ^{1}H-^{15}N TROSY spectra in presence of DPC **(Figure 2-12)**, LMPG **(Figure 2-13)** and SDS **(Figure 2-14)**. These truncated constructs did not contain the central hydrophobic transmembrane domain. However, $IpaC^{1-100}$ has 31 hydrophobic residues while $IpaC^{173-363}$ has 60 hydrophobic residues, which possibly contributed to their observed conformational changes in presence of micelles.

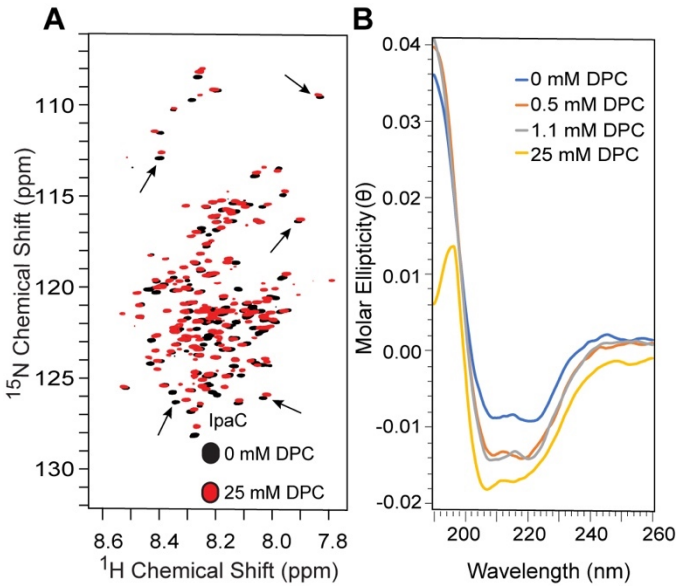

Figure 2-10. Conformational change of full-length IpaC in DPC micelles, shown by (A) ^{1}H-^{15}N TROSY spectroscopy (representative affected peaks shown by arrowheads) and (B) Far-UV CD spectroscopy.

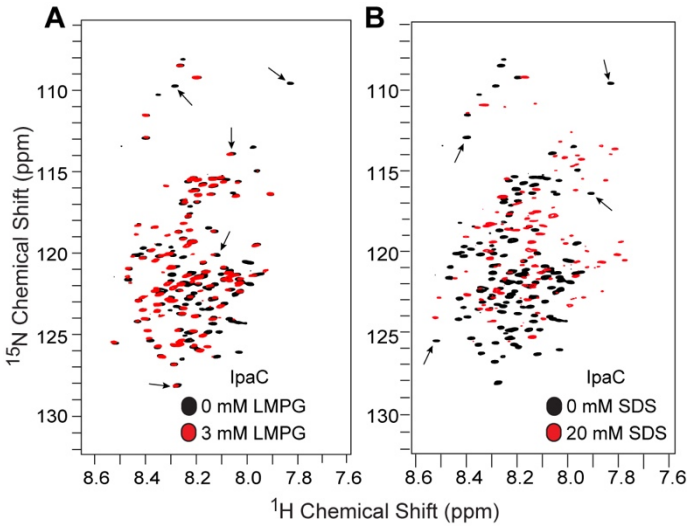

Figure 2-11. ^{1}H-^{15}N TROSY spectra of full-length IpaC in LMPG and SDS micelles. NMR spectra were acquired in presence and absence of (A) 3 mM LMPG and (B) 20 mM SDS. Representative affected peaks are shown in black arrows.

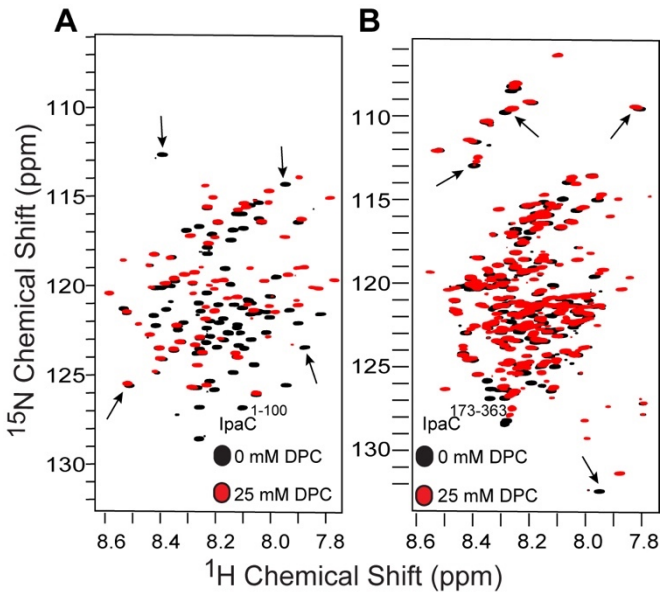

Figure 2-12. ^{1}H-^{15}N TROSY spectra of IpaC domains in DPC micelles show conformational changes for (A) IpaC$^{1-100}$ and (B) IpaC$^{173-363}$. Representative affected peaks are shown in black arrows.

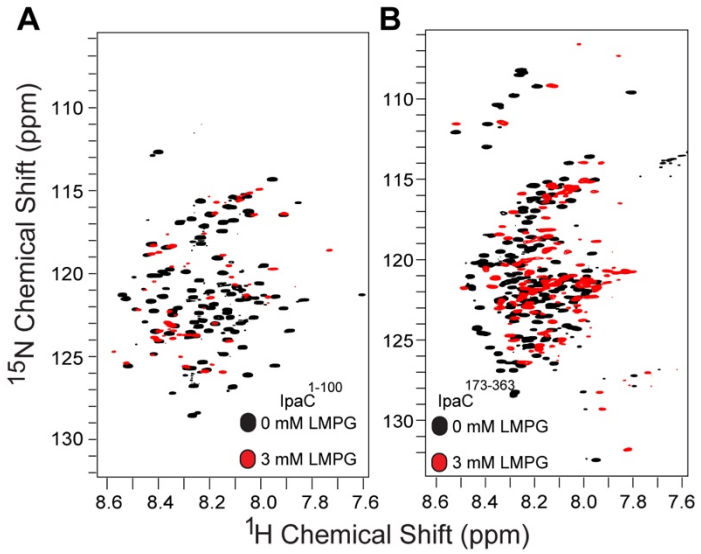

Figure 2-13. ^{1}H-^{15}N TROSY spectra of IpaC domains in LMPG micelles, showing conformational changes for (A) IpaC$^{1-100}$ and (B) IpaC$^{173-363}$. The spectra in absence of LMPG are shown in black, while the spectra in presence of 3 mM LMPG are shown in red.

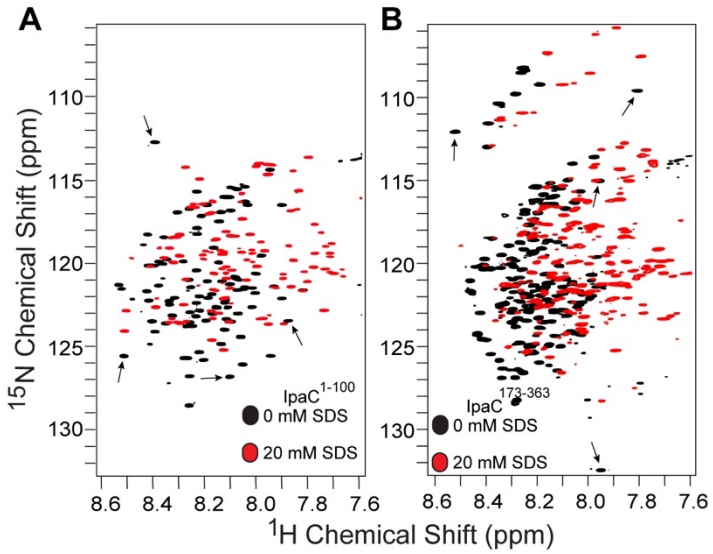

Figure 2-14. ^{1}H-^{15}N TROSY spectra of IpaC domains in SDS micelles, showing conformational changes for (A) IpaC$^{1-100}$ and (B) IpaC$^{173-363}$. Representative affected peaks are shown in black arrows.

The far-UV CD spectra of full-length IpaC, IpaC$^{1-100}$ and IpaC$^{173-363}$ titrated with increasing concentrations of SDS **(Figure 2-15)** showed that all three proteins undergo conformational changes in SDS micelles. An increase in α-helical character was observed for full-length IpaC **(Figure 2-15A)**, shown by increased negative dips at 208 nm and 222 nm in the CD spectrum. The NTD (IpaC$^{1-100}$) of IpaC underwent a change from random coil to α-helix, characterized by a shift in the dip from 200 nm to 208 and 222 nm **(Figure 2-15B)** in the CD spectrum. The CTD (IpaC$^{173-363}$) of IpaC also showed a shift in the dip from 200 nm to 208 nm **(Figure 2-15C)**, suggesting a coil-to-helix transition. Only IpaC$^{1-100}$ showed well-behaved far-UV CD spectra in presence of LMPG **(Figure 2-16)**. I found that this construct undergoes a transition from a random coil to an α-helix in presence of LMPG **(Figure 2-16)**. The full-length IpaC and IpaC$^{173-363}$ did not show good quality CD spectra in presence of LMPG (data not shown). The CD spectra of IpaC$^{173-363}$ in presence of DPC were also of poor quality (data not shown). This is possibly because these spectra were collected 4 days after the initial purification of the proteins, leading to their aggregation. In presence of DPC, IpaC$^{1-100}$ transitioned to an α-helical conformation from a random coil **(Figure 2-17)**. The far-UV CD data collectively suggest that full-length IpaC and its domains show distinct behaviors to the three detergents used.

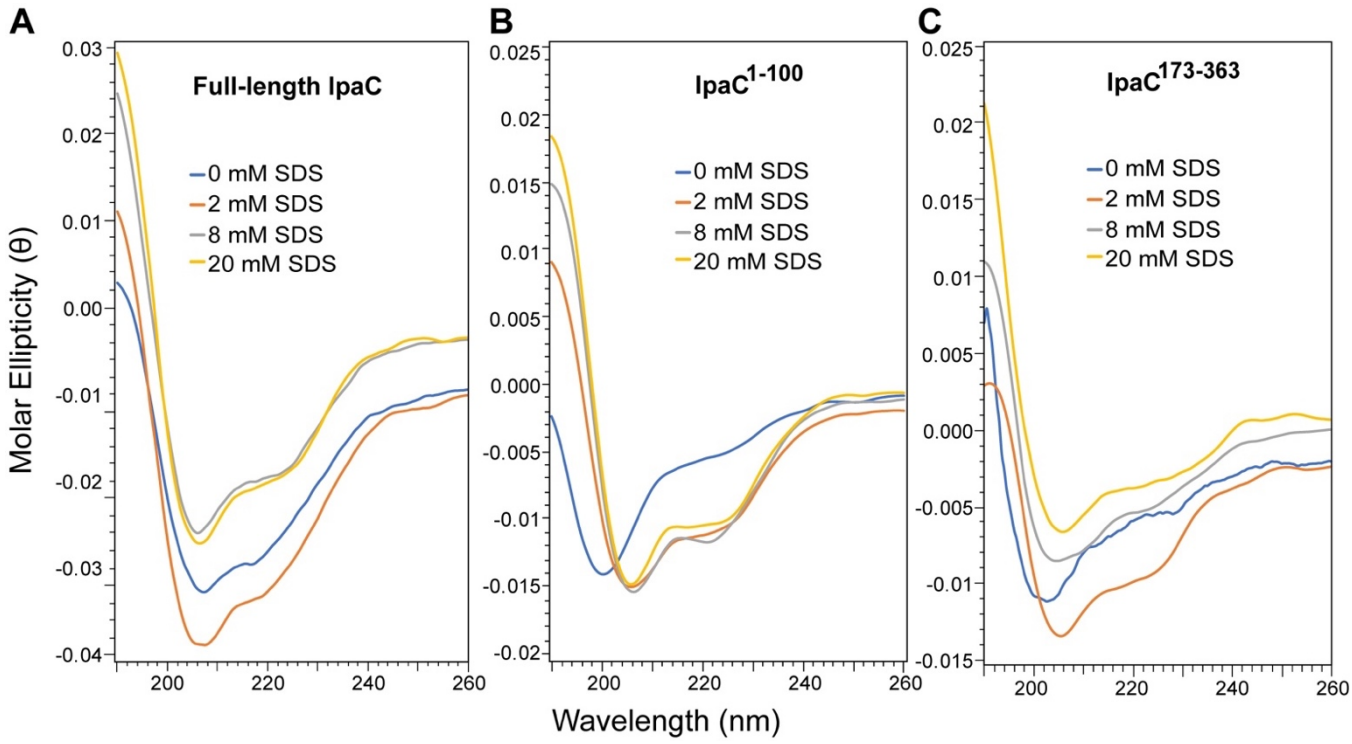

Figure 2-15. Secondary structure changes of IpaC in SDS micelles. Conformational changes induced by SDS micelles are shown for (A) full-length IpaC, (B) IpaC$^{1-100}$ and (C) IpaC$^{173-363}$

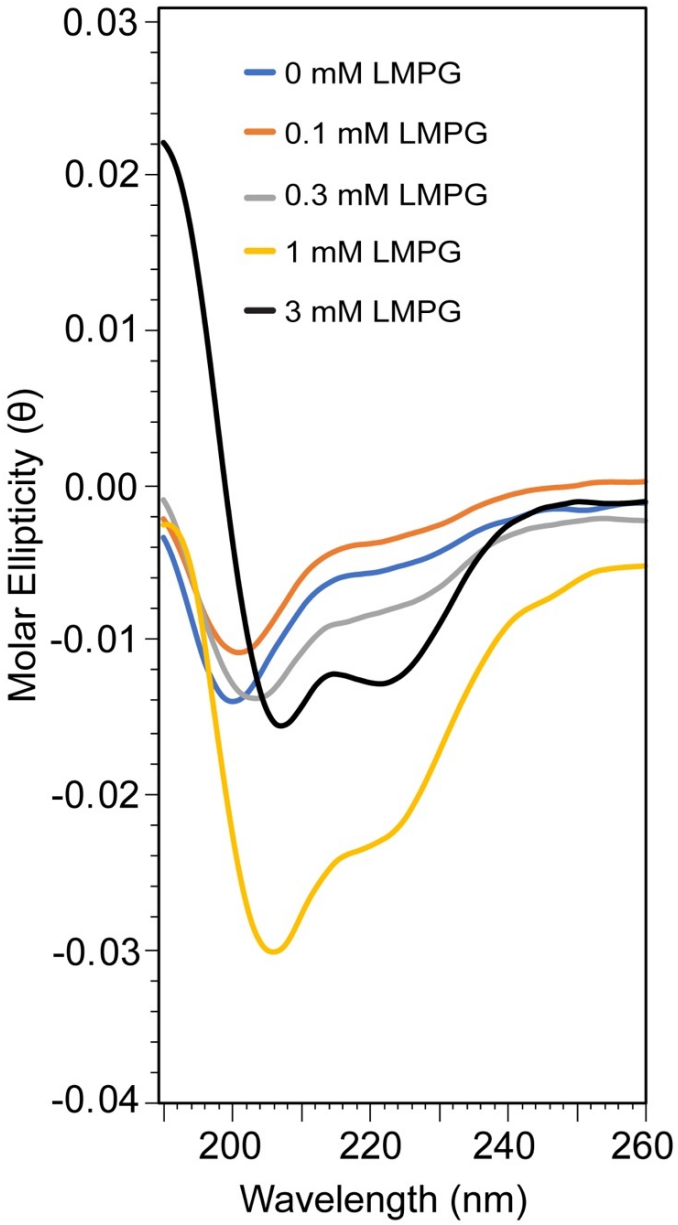

Figure 2-16. Secondary structure changes of IpaC$^{1-100}$ in LMPG micelles. IpaC$^{1-100}$ undergoes a coil (blue) to helix (black) transition with increasing concentration of LMPG.

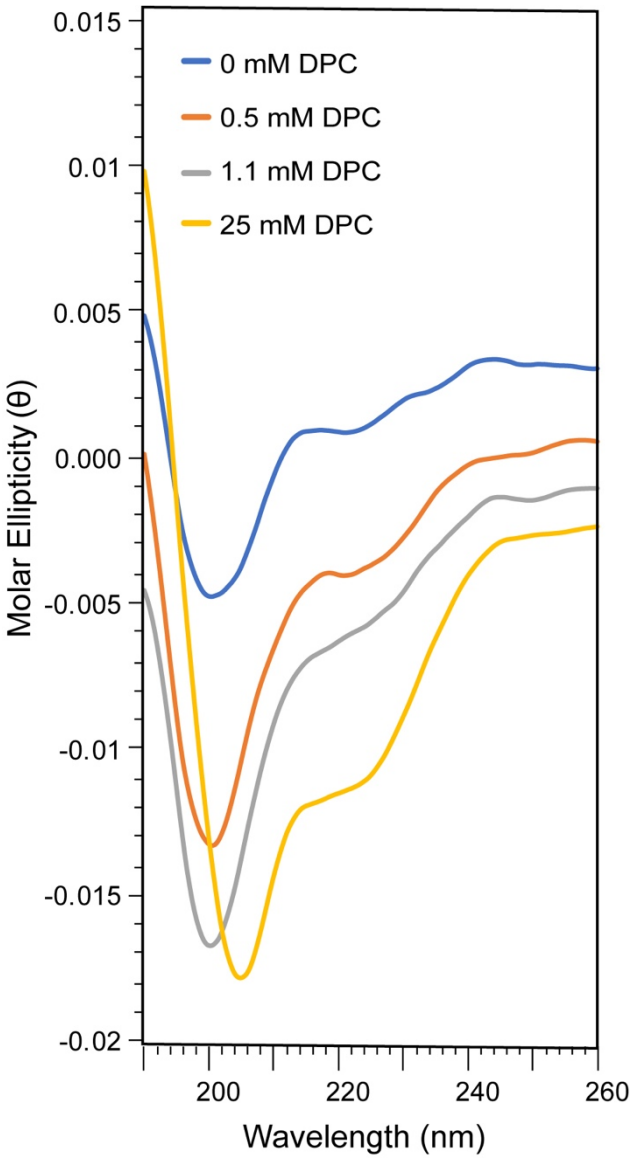

Figure 2-17. Secondary structure changes of IpaC$^{1-100}$ in DPC micelles. The data show that IpaC$^{1-100}$ undergoes coil (blue) to helix (yellow) transition in presence of increasing DPC concentration.

2.5. Discussion

The atomic structure of any minor translocon protein is currently unknown (Kawaljit Kaur, University of Kansas 2016; Dissertation). The far-UV CD data **(Figure 2-5A** and **Figure 2-6)** suggest that the overall secondary structure of IpaC is disrupted on separating the NTD and CTD from the full-length protein. The ratio of the molar ellipticities at 222 and 208 nm ($\theta_{222}/\ \theta_{208}$) for full-length IpaC was 1.07, suggesting the presence of interacting α-helices [30, 31] in contrast to the ratio of ~0.9 found in proteins with non-interacting or single-stranded α-helices [32, 33]. This finding supports previously published reports that IpaC contains putative coiled-coils [39] and can potentially oligomerize [28, 40, 41]. The ability to oligomerize probably led to protein aggregation and contributed to the precipitation of full-length IpaC above a concentration of 0.2 mM in the experiments. Konrat *et al* reported [42] that the ^{1}H-^{15}N HSQC spectra of intrinsically disordered proteins (IDP) show severe spectral overlap and peak dispersions of less than 1 ppm in the proton dimension, which were not resolved even at low temperature and/or low pH. The amino acid sequence and the ^{1}H-^{15}N TROSY spectra of full-length IpaC and its domains (**Figure 2-7**) indicate that all the proteins are intrinsically disordered. Full-length IpaC and its CTD contain regions that are conformationally dynamic, because less than the expected number of NMR peaks were observed for these proteins (**Figure 2-7** and **Figure 2-8**). This inherent flexibility of IpaC probably makes it a difficult candidate for crystallization. These are currently the only available NMR spectra of a minor translocon protein. The intrinsically disordered nature of IpaC is possibly necessary for its secretion through the channel formed by the needle protein. Other proteins in the T3SS, like the inner rod protein PrgJ in *Salmonella* are also partially folded [43]. Thus, IpaC adds to the growing list of intrinsically disordered proteins essential for the function of the T3SS [43, 44].

Collectively, the far-UV CD and NMR spectra suggest that full-length IpaC may exist as a molten globule in solution, which is an intermediate between the native and unfolded states in the folding pathway of many proteins [20]. Molten globules are often formed under mildly denaturing conditions such as low/high pH, temperature and low concentrations of denaturing agents like urea [45], and have defined secondary structures without any well-formed tertiary structures with exposed hydrophobic patches [46]. The CD thermal denaturation plot **(Figure 2-5B)** suggests that full-length IpaC lacks an overall tertiary structure, and this is supported by the NMR data (**Figure 2-7A** and **Figure 2-8A**). The major translocon protein PopB in *Pseudomonas* [47] and the minor translocon protein YopD [48] in *Yersinia* were reported to exist as molten globules. Further experiments, like ANS fluorescence studies, will need to be performed with IpaC to test the validity of this hypothesis. The molten globule state of IpaC in solution might be important in its transport through the T3SS channel in a partially unfolded state, and might also prevent its proper folding and oligomerization prior to host membrane insertion.

Both DPC and LMPG micelles have been used previously [49-52] in the structural and functional studies of membrane proteins. The NMR assignment of any minor translocon protein is currently unknown. Since IpaC is a membrane protein, assigning its NMR spectrum acquired in presence of a membrane mimetic or a surfactant will be more biologically relevant. IpaC was previously shown to interact with liposomes [17] and phospholipid membranes [18]. However, conformational changes of the protein in presence of varying concentration of detergents, liposomes or phospholipids have not been reported. The NMR data showed that full-length IpaC undergoes conformational change and interacts with DPC, LMPG and SDS micelles at concentrations above their respective CMC **(Figure 2-10A** and **Figure 2-11)**. Interaction with micelles probably leads to lower tumbling rate of the IpaC, leading to peak broadening and disappearance. The corresponding far UV CD spectra showed that the α-helical characters of full-length IpaC increased in presence of DPC **(Figure 2-10B)** and SDS **(Figure 2-15A)**. This

conformational change might be essential for IpaC to perform its effector functions in the host cell [10-15] under conditions that activate the T3SS, like host cell contact and a temperature ~ 37^0C [53]**.** The NMR data also showed conformational changes of the NTD and CTD of IpaC in presence of DPC **(Figure 2-12)**, LMPG **(Figure 2-13)** and SDS **(Figure 2-14)** micelles. Far-UV CD spectroscopy showed that the NTD undergoes a coil-to-helix transition in presence SDS **(Figure 2-15B)**, LMPG **(Figure 2-16)** and DPC **(Figure 2-17)**. But the CTD undergoes a coil-to-helix transition in SDS **(Figure 2-15C)**. This transition of the CTD might be due to the ability of SDS to induce helix formation in proteins [36-38]. Further experiments needed to understand the effects of LMPG on the secondary structures of full-length IpaC and its CTD. From the data presented in this chapter, DPC is a good micelle system to further study IpaC.

My findings are essential in studying the protein-protein interactions of IpaC and its domains (Chapter 3) with other components of the T3SS. The data will be a good starting point for the future use of detergents and other membrane mimetics to assign the NMR spectra of IpaC.

2.6. References

1. Coburn, B., I. Sekirov, and B.B. Finlay, *Type III Secretion Systems and Disease*, in *Clin. Microbiol. Rev.* 2007.
2. Blocker, A., et al., *Structure and composition of the Shigella flexneri "needle complex", a part of its type III secretion.* Mol. Microbiol., 2001. **39**(3): p. 652-663.
3. Wellington, E.M., et al., *The role of the natural environment in the emergence of antibiotic resistance in gram-negative bacteria.* Lancet Infect Dis, 2013. **13**(2): p. 155-65.
4. Davies, J. and D. Davies, *Origins and evolution of antibiotic resistance.* Microbiol Mol Biol Rev, 2010. **74**(3): p. 417-33.
5. Boucher, H.W., et al., *Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America.* Clin. Infect. Dis., 2009. **48**(1): p. 1-12.
6. Spellberg, B., et al., *The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America.* Clin. Infect. Dis., 2008. **46**(2): p. 155-64.
7. Chatterjee, S., et al., *Structure and biophysics of type III secretion in bacteria.* Biochemistry, 2013. **52**(15): p. 2508-17.

8. Lunelli, M., et al., *Crystal structure of PrgI-SipD: insight into a secretion competent state of the type three secretion system needle tip and its interaction with host ligands.* PLoS Pathog., 2011. **7**(8): p. e1002163.
9. Romano, F.B., et al., *Type 3 secretion translocators spontaneously assemble a hexadecameric transmembrane complex.* J Biol Chem, 2016. **291**(12): p. 6304-15.
10. Tran Van Nhieu, G., et al., *IpaC induces actin polymerization and filopodia formation during Shigella entry into epithelial cells.* EMBO J, 1999. **18**(12): p. 3249-62.
11. Veenendaal, A.K., et al., *The type III secretion system needle tip complex mediates host cell sensing and translocon insertion.* Mol. Microbiol., 2007. **63**(6): p. 1719-30.
12. Blocker, A., et al., *The tripartite type III secreton of Shigella flexneri inserts IpaB and IpaC into host membranes.* J.Cell Biol., 1999. **147**(3): p. 683-693.
13. Marquart, M.E., W.L. Picking, and W.D. Picking, *Soluble invasion plasmid antigen C (IpaC) from Shigella flexneri elicits epithelial cell responses related to pathogen invasion.* Infect. Immun., 1996. **64**(10): p. 4182-4187.
14. Menard, R., et al., *The secreted Ipa complex of Shigella flexneri promotes entry into mammalian cells.* Proc. Natl. Acad. Sci. U.S.A., 1996. **93**(3): p. 1254-8.
15. Russo, B.C., et al., *Intermediate filaments enable pathogen docking to trigger type 3 effector translocation.* Nat Microbiol, 2016. **1**: p. 16025.
16. Harrington, A.T., et al., *Structural characterization of the N terminus of IpaC from Shigella flexneri* Infect. Immun., 2003. **71**(3): p. 1255-1264.
17. Kueltzo, L.A., et al., *Structure-function analysis of invasion plasmid antigen C (IpaC) from Shigella flexneri* J. Biol. Chem., 2003. **278**(5): p. 2792-2798.
18. Osiecki, J.C., et al., *IpaC from Shigella and SipC from Salmonella possess similar biochemical properties but are functionally distinct.* Mol. Microbiol., 2001. **42**(2): p. 469-481.
19. Wright, P.E. and H.J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation.* Nat Rev Mol Cell Biol, 2015. **16**(1): p. 18-29.
20. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins.* Chem Rev, 2014. **114**(13): p. 6589-631.
21. Uversky, V.N., S. Winter, and G. Lober, *Use of fluorescence decay times of 8-ANS-protein complexes to study the conformational transitions in proteins which unfold through the molten globule state.* Biophys Chem, 1996. **60**(3): p. 79-88.
22. Chatterjee, S., et al., *The crystal structures of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate.* Protein Sci, 2011. **20**(1): p. 75-86.
23. Geisbrecht, B.V., S. Bouyain, and M. Pop, *An optimized system for expression and purification of secreted bacterial proteins.* Protein Expr. Purif., 2006. **46**(1): p. 23-32.
24. Chatterjee, S., et al., *The crystal structure of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate.* Protein Sci., 2011. **20**: p. 75-86.
25. McShan, A.C., et al., *Characterization of the Binding of Hydroxyindole, Indoleacetic acid, and Morpholinoaniline to the Salmonella Type III Secretion System Proteins SipD and SipB.* ChemMedChem, 2016. **11**(9): p. 963-971.
26. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J. Biomol. NMR, 1995. **6**(3): p. 277-293.
27. Johnson, B.A., *Using NMRView to visualize and analyze the NMR spectra of macromolecules.* Methods Mol. Biol., 2004. **278**: p. 313-352.
28. Terry, C.M., et al., *The C-terminus of IpaC is required for effector activities related to Shigella invasion of host cells.* Microb. Pathog., 2008. **45**(4): p. 282-9.
29. McGuffin, L.J., K. Bryson, and D.T. Jones, *The PSIPRED protein structure prediction server.* Bioinformatics, 2000. **16**(4): p. 404-5.

30. Choy, N., V. Raussens, and V. Narayanaswami, *Inter-molecular coiled-coil formation in human apolipoprotein E C-terminal domain.* J. Mol. Biol., 2003. **334**(3): p. 527-39.
31. Wang, Y., et al., *NMR structure of the N-terminal coiled-coil domain of the Andes hantavirus nucleocapsid protein.* J. Biol. Chem., 2008. **283**(42): p. 28297-28304.
32. Kiss, R.S., C.M. Kay, and R.O. Ryan, *Amphipathic alpha-helix bundle organization of lipid-free chicken apolipoprotein A-I.* Biochemistry, 1999. **38**(14): p. 4327-34.
33. Lau, S.Y., A.K. Taneja, and R.S. Hodges, *Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils.* J Biol Chem, 1984. **259**(21): p. 13253-61.
34. Bernard, A.R., et al., *Detergent Isolation Stabilizes and Activates the Shigella Type III Secretion System Translocator Protein IpaC.* J Pharm Sci, 2016. **105**(7): p. 2240-8.
35. Jorda, J., et al., *Protein tandem repeats - the more perfect, the less structured.* FEBS J, 2010. **277**(12): p. 2673-82.
36. Otzen, D.E. and M. Oliveberg, *Burst-phase expansion of native protein prior to global unfolding in SDS.* J Mol Biol, 2002. **315**(5): p. 1231-40.
37. Jirgensons, B., *Effects of n-propyl alcohol and detergents on the optical rotatory dispersion of alpha-chymotrypsinogen, beta-casein, histone fraction F1, and soybean trypsin inhibitor.* J Biol Chem, 1967. **242**(5): p. 912-8.
38. Mattice, W.L., J.M. Riser, and D.S. Clark, *Conformational properties of the complexes formed by proteins and sodium dodecyl sulfate.* Biochemistry, 1976. **15**(19): p. 4264-72.
39. Gazi, A.D., et al., *Coiled-coils in type III secretion systems: structural flexibility, disorder and biological implications.* Cell Microbiol, 2009. **11**(5): p. 719-29.
40. Pallen, M.J., G. Dougan, and G. Frankel, *Coiled-coil domains in proteins secreted by type III secretion systems.* Mol. Microbiol., 1997. **25**(2): p. 423-5.
41. Davis, R., et al., *Protein-protein interactions in the assembly of Shigella flexneri invasion plasmid antigens IpaB and IpaC into protein complexes.* Biochim.Biophys.Acta, 1998. **1429**(1): p. 45-56.
42. Konrat, R., *NMR contributions to structural dynamics studies of intrinsically disordered proteins.* J Magn Reson, 2014. **241**: p. 74-85.
43. Zhong, D., et al., *The Salmonella type III secretion system inner rod protein PrgJ is partially folded.* J. Biol. Chem., 2012. **287**(30): p. 25303-11.
44. Chaudhury, S., et al., *The LcrG tip chaperone protein of the Yersinia pestis type III secretion system is partially folded.* J. Mol. Biol., 2015. **427**: p. 3096-3109.
45. Vassilenko, K.S. and V.N. Uversky, *Native-like secondary structure of molten globules.* Biochim Biophys Acta, 2002. **1594**(1): p. 168-77.
46. Ptitsyn, O.B., *Molten globule and protein folding.* Adv Protein Chem, 1995. **47**: p. 83-229.
47. Dey, S., A. Basu, and S. Datta, *Characterization of molten globule PopB in absence and presence of its chaperone PcrH.* Protein J, 2012. **31**(5): p. 401-16.
48. Faudry, E., et al., *Type III secretion system translocator has a molten globule conformation both in its free and chaperone-bound forms.* FEBS J., 2007. **274**(14): p. 3601-10.
49. Kang, C., et al., *Structure of KCNE1 and implications for how it modulates the KCNQ1 potassium channel.* Biochemistry, 2008. **47**(31): p. 7999-8006.
50. Kang, C. and Q. Li, *Solution NMR study of integral membrane proteins.* Curr Opin Chem Biol, 2011. **15**(4): p. 560-9.
51. Sanders, C.R. and F. Sonnichsen, *Solution NMR of membrane proteins: practice and challenges.* Magn Reson Chem, 2006. **44 Spec No**: p. S24-40.
52. Page, R.C., et al., *Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins.* J Struct Funct Genomics, 2006. **7**(1): p. 51-64.

53. Dey, S., et al., *The type III secretion system needle, tip, and translocon.* Protein Sci, 2019. **28**(9): p. 1582-1593.

Chapter 3: Protein-Protein Interactions of the *Shigella* Minor Translocon Protein IpaC With Other Components of the Type III Secretion System

3.1. Abstract

IpaC is the minor translocon protein of *Shigella*. In the previous chapter, I showed that full-length IpaC and its N-terminal domain (NTD) and C-terminal domain (CTD) lack tertiary structures. Here I used NMR titrations to show that IpaC and its NTD and CTD bind to the chaperone IpgC and the tip protein IpaD. Further, the IpaC CTD interacts with the N-terminal domain of the major translocon protein IpaB. IpaC is known to oligomerize, and NMR detected interactions among the CTD of IpaC in solution. My results identify previously unknown binding partners of IpaC and provide insights into the mechanism of oligomerization of IpaC.

3.2. Introduction

IpaC is the minor translocon protein in *Shigella* and forms a hetero-oligomeric complex with the major translocon protein IpaB [1]. IpaC was also found to function as an effector within the host cell [2-6] by interacting with the intermediate filaments - vimentin and keratin [7]. This interaction helps *Shigella flexneri* to dock to host cells in a stable manner [8]. The premature secretion and complex formation of IpaC and IpaB are prevented by the interaction with their common chaperone IpgC [9]. The N-terminal region of IpaC is involved in binding IpgC and IpaB, and IpaB competes with IpaC for binding IpgC [10]. The central hydrophobic region of IpaC also interacts with IpaB [10]. Results from truncations and yeast two-hybrid assays showed that residues 367-458 of IpaB harbor the IpaC-binding site [11]. Residues 50-80 of IpaC are important for binding to IpgC and IpaB [10]. IpaC is predicted to contain one putative transmembrane helix from residues 100-120 and the N-terminal domain is outside the host cells, while the C-terminal domain is inside [7]. Residues 319-345 in the C-terminal domain of IpaC forms a putative

coiled-coil [12]. IpaC has been reported to interact with itself [13], indicating that IpaC can potentially form oligomers.

The protein-protein interactions of IpaC are not well understood. Here I used NMR spectroscopy to test my hypothesis that IpaC binds to the tip protein IpaD. Additionally, I determined that the C-terminal domain of IpaC also harbors an IpgC binding site and interacts with the N-terminal domain of IpaB. IpaC oligomerization possibly occurs due to interactions among neighboring C-terminal domains and not due to an interaction between the N-terminal and C-terminal domains. My results improve our understanding of the protein-protein interactions of IpaC. This knowledge is needed to elucidate how the translocon is assembled, and will aid in developing new antimicrobials targeted at disrupting the protein-protein interactions involved in the assembly of the T3SS needle complex.

3.3. Methods

3.3.1. Expression and purification of proteins

Unlabeled as well as uniformly ^{15}N-labeled full-length IpaC, IpaC$^{1-100}$ (NTD) and IpaC$^{173-363}$ (CTD) were overexpressed and purified as described in the previous chapter.

^{15}N-leucine labeled IpaC. Full-length IpaC was specifically labeled with ^{15}N-leucine by growing cells in M9 minimal media supplemented with 20 individually weighed amino acids. Into a 2.8 L Flask A, was added 700 mL water, 200 mg each of unlabeled valine and isoleucine, 150 mg of each of the remaining 17 amino acids except leucine and M9 salts sufficient for a 1 L culture media. Into Flask B was added the following: 150 mL water, 125 mg ^{15}N-leucine, 450 mg each

of valine and isoleucine, and 300 mg each of the remaining 17 amino acids. The flask was placed on a heating plate to dissolve the contents, and the volume was made up to 300 mL. Flask A and Flask B were autoclaved to sterilize the contents. After cooling, the following were added to Flask A: 5 ml 40% D-glucose and 500 μL of each of 1 M $MgSO_4$, 0.1 M $CaCl_2$, BME vitamin solution (Sigma # B6891), and trace mineral supplement (ATCC MD-TMS). A starter culture in 50 mL LB overnight growth was centrifuged (4,000 rpm, 2392×g, 10 min) and the pellet was used to inoculate the culture media in Flask A. Cells were grown at 37°C until OD_{600} ~ 0.7, and the contents of Flask B were added to Flask A. At OD_{600} ~ 0.8, cells were induced with 1 mM IPTG. Cell growth was continued overnight at 25°C until OD_{600} ~ 2.7. Cells were harvested by centrifugation and the ^{15}N-Leu IpaC was purified following the protocol described for full-length IpaC in the previous chapter.

IpaD, IpaB, and IpgC. IpaD (residues 38-332, C322S) and $IpaB^{9-226}$ were overexpressed and purified as described previously [14]. IpgC was expressed in *E. coli* BL21(DE3) DNAY cells and grown in culture media containing carbenicillin and kanamycin (30 μg/mL). The expression and purification of IpgC followed the protocol described for $IpaC^{173-363}$ in the previous chapter. Unlabeled IpaD, $IpaB^{9-226}$, and IpgC were expressed and purified from cell growth in 2 L LB culture media. All the three proteins were purified from the soluble fraction.

3.3.2. NMR spectroscopy

Two-dimensional ^{1}H-^{15}N TROSY spectra were acquired for ^{15}N-labeled full-length IpaC, $IpaC^{1-100}$ and $IpaC^{173-363}$ using a Bruker Avance 800 MHz NMR spectrometer equipped with a cryogenic triple resonance probe or a Bruker Avance 600 MHz NMR spectrometer equipped with a TXI-RT probe as described in the previous chapter. Two-dimensional ^{1}H-^{15}N TROSY spectra

were acquired at 20°C using 0.1 mM of ^{15}N-labeled IpaD in the NMR buffer. The ^{1}H-^{15}N acquisition parameters for IpaD were 32 scans with a ^{15}N sweep width of 30 ppm with the center at 118 ppm and ^{1}H sweep width of 18 ppm centered at 4.7 ppm. The ^{1}H-^{15}N acquisition parameters for full-length IpaC and IpaC$^{1-100}$ were identical to those reported in the previous chapter. For ^{15}N-Leu IpaC, the typical ^{1}H-^{15}N acquisition parameters were 96 scans, with a ^{15}N sweep width of 30 ppm centered at 118 ppm; and ^{1}H sweep width of 18 ppm centered at 4.7 ppm. For IpaC$^{173-363}$, the acquisition parameters were identical to those reported in the previous chapter except the use of 96 scans on the 600 MHz spectrometer.

3.3.3. NMR titrations

^{15}N-IpaC, ^{15}N-IpaC$^{1-100}$ or ^{15}N-IpaC$^{173-363}$ was titrated with unlabeled IpgC, IpaD or IpaB$^{9-226}$. ^{15}N-IpaD was also titrated with unlabeled IpaC, IpaC$^{1-100}$ or IpaC$^{173-363}$. Three to five titration points were recorded using increasing amounts of protein-protein molar ratios by acquiring 2D ^{1}H-^{15}N TROSY spectra. For titrations using labeled IpaD, the peak intensity ratio ($I_{1:1}/I_{1:0}$) for non-overlapped peaks were calculated and subtracted from 1, which is the intensity ratio in absence of any interaction. The (1- Intensity Ratio) of the strongly affected IpaD residues were mapped onto the IpaD crystal structure (PDB ID: 2J0O) [15].

3.4. Results

3.4.1. Full-length IpaC and its domains bind to IpgC

In the previous chapter, I used ^{1}H-^{15}N TROSY and ^{1}H-^{13}C HSQC spectra to show that full-length IpaC and its domains contain regions that are dynamic in nature **(Figures 2-6** and **2-7)**. This observation was confirmed by ^{1}H-^{15}N TROSY spectrum of ^{15}N-Leu IpaC **(Figure 3-1)**. This spectrum showed only the leucine peaks from full-length IpaC. About 36 peaks out of the expected 40 (corresponding to 39 leucine residues of IpaC and 1 from the C-terminal tag) **(Figure 3-1)** were observed, indicating that there are regions in IpaC that are conformationally dynamic.

IpaC and its NTD ($IpaC^{1-100}$) are known to interact with its chaperone protein IpgC [10, 16], and this interaction was used to determine that purified, recombinant IpaC and $IpaC^{1-100}$ are functional. I titrated ^{15}N-Leu IpaC with increasing concentrations of unlabeled IpgC and found concentration-dependent peak shifts for specific Leu peak a, while concentration-dependent decrease in intensity for peaks c, d and e **(Figure 3-2A and Figure 3-2C)**. This indicates that full-length refolded IpaC is functional, as it interacts with IpgC. Titration of unlabeled IpgC with labeled $IpaC^{1-100}$ showed changes in peak positions for peaks a, b, c and d with increasing IpgC concentration **(Figure 3-2B** and **Figure 3-2D)**, showing that refolded $IpaC^{1-100}$ is also functional. $IpaC^{173-363}$ does not contain the IpgC-binding regions [10, 16]. I hypothesized that this construct does not bind to IpgC. But ^{1}H-^{15}N TROSY spectrum of $IpaC^{173-363}$ titrated with increasing molar concentrations of IpgC showed concentration-dependent shifts of specific peaks a, b, c and d **(Figure 3-3)**. This indicates that IpgC binds to $IpaC^{173-363}$. This is the first published report of the presence of an IpgC-binding region in the C-terminal domain of IpaC. This newly reported interaction occurs in the fast exchange regime in the NMR timescale, with dissociation constants

in the millimolar range. My NMR data suggest that the chaperone IpgC interacts with both the NTD and the CTD of IpaC **(Figure 3-2B, Figure 3-2D and Figure 3-3)**.

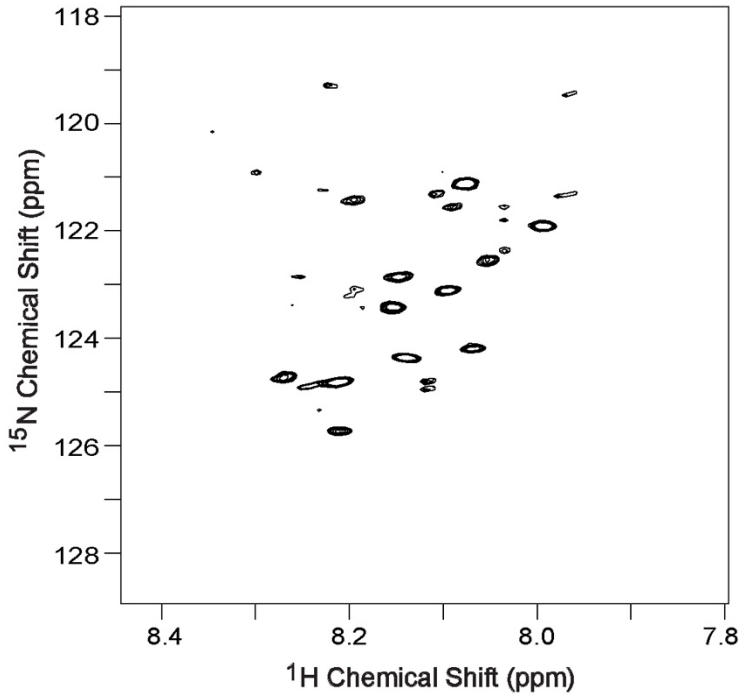

Figure 3-1. ^{1}H-^{15}N TROSY spectrum of ^{15}N-Leu IpaC, showing only peaks from the leucine residues of IpaC.

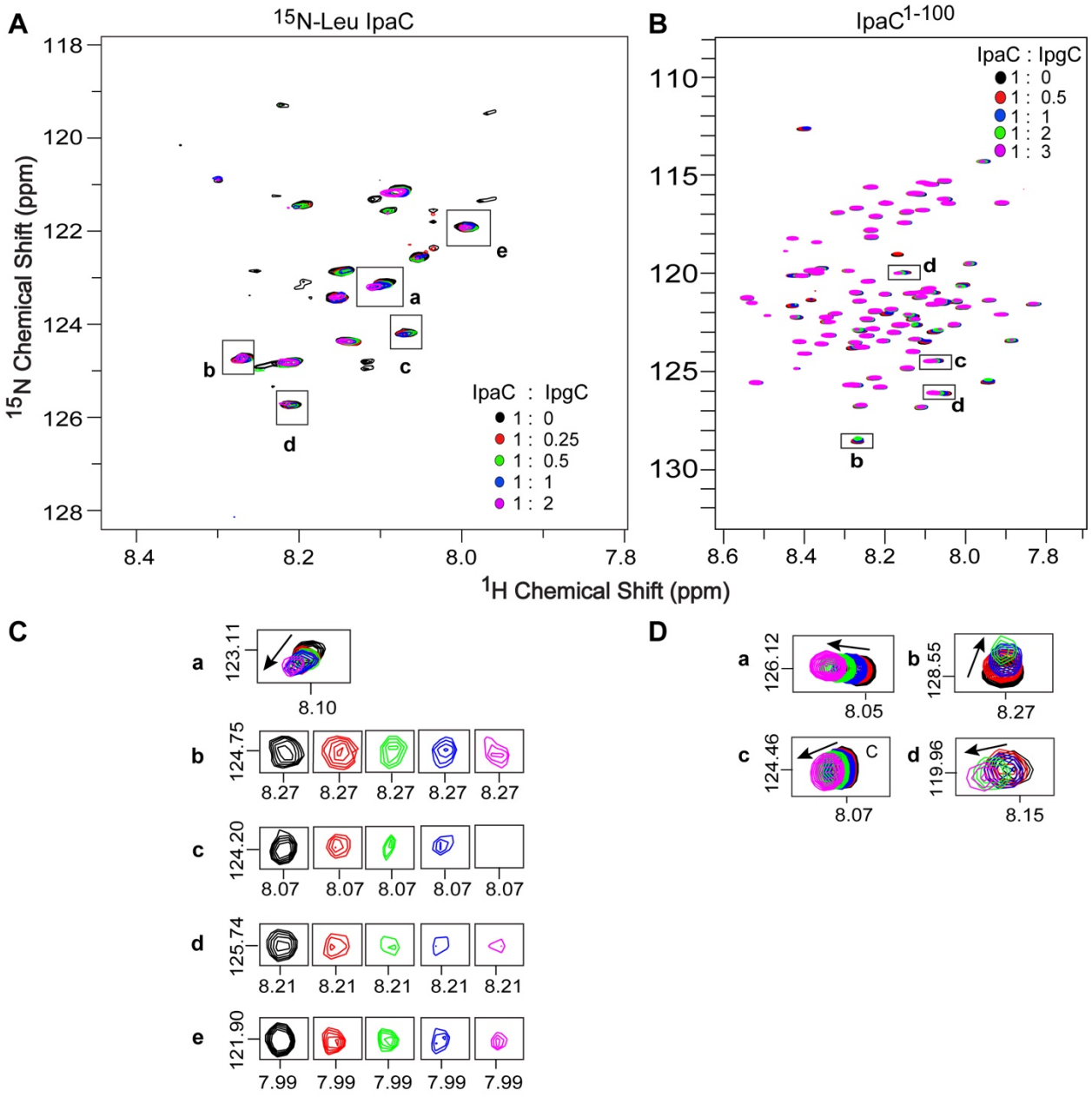

Figure 3-2. Full-length IpaC and IpaC$^{1-100}$ are functional. ^{1}H-^{15}N TROSY spectra of the titrations of (A) ^{15}N-Leu IpaC and (B) IpaC$^{1-100}$ with IpgC. Selected affected peaks are shown for (C) ^{15}N-Leu IpaC and (D) IpaC$^{1-100}$. The data indicate that refolded full-length IpaC and IpaC$^{1-100}$ are functional.

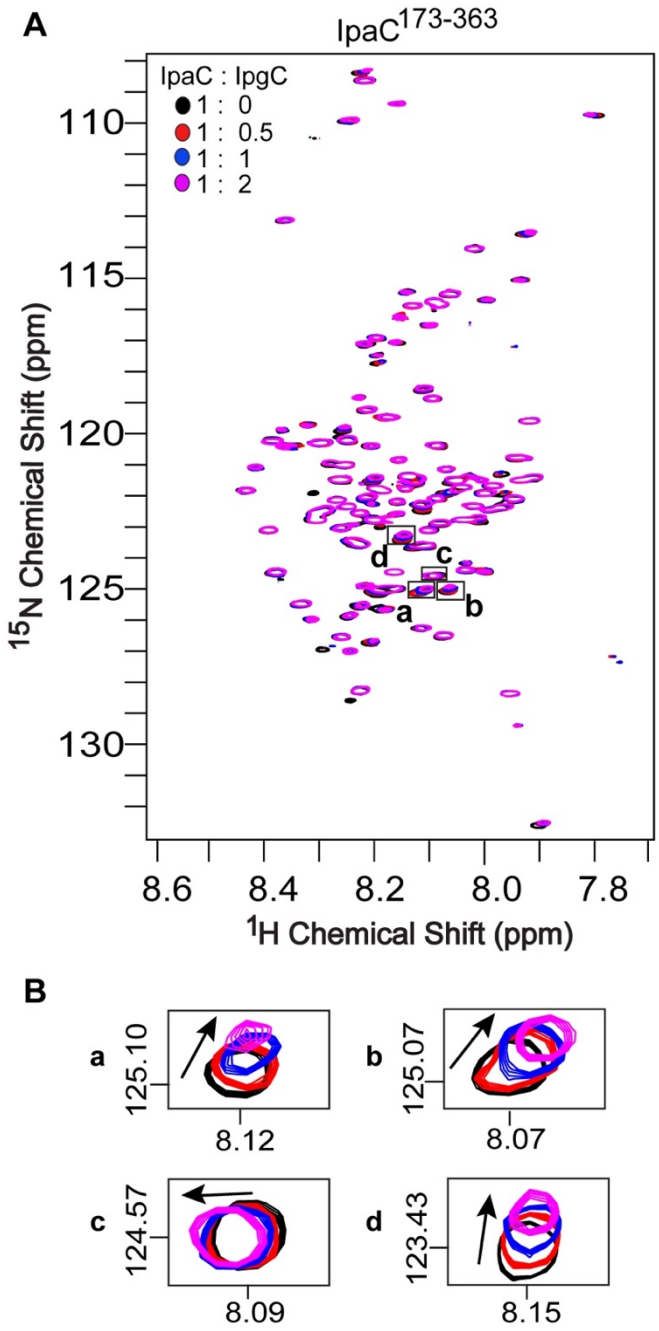

Figure 3-3. IpaC$^{173-363}$ binds to IpgC. (A) ^{1}H-^{15}N TROSY spectrum of the titration of IpaC$^{173-363}$ with IpgC. (B) Selected affected peaks a, b, c and d are shown.

3.4.2. The C-terminal domain of IpaC interacts with itself

IpaC contains putative coiled-coils [17] and can potentially oligomerize [12, 13, 18]. Fluorescence and cross-linking experiments showed [13] that full-length IpaC interacts with itself, forming dimers, trimers and tetramers. Deletion of residues 261-363 in the C-terminal domain (CTD) of IpaC abrogated the IpaC-IpaC interactions [19]. To better understand the role of the CTD in IpaC oligomerization, I titrated ^{15}N-labeled IpaC$^{173-363}$ (CTD) with unlabeled IpaC$^{173-363}$ and observed concentration-dependent changes in peak positions of two specific peaks a and b (**Figure 3-4B**), indicating interaction in the fast exchange NMR timescale, with millimolar dissociation constant. A similar titration of labeled IpaC$^{1-100}$ (NTD) with unlabeled IpaC$^{173-363}$ showed neither any change in chemical shift nor a decrease in peak intensity **(Figure 3-4A)**. This indicates that IpaC$^{173-363}$ interacts with itself, but does not interact with IpaC$^{1-100}$. The NMR data suggest that IpaC oligomerization occurs due to interactions among the neighboring CTD of IpaC and not due to an interaction between the NTD and the CTD. My findings serve as the first direct biophysical evidence of the possible mechanism of IpaC oligomerization.

3.4.3. The C-terminal domain of IpaC interacts with the N-terminal domain of IpaB

Residues 367-458 of the major translocon protein IpaB harbor the IpaC-binding site [11]. IpaB$^{9-226}$, a construct containing the N-terminal domain of IpaB, was reported [14] to bind to the tip protein IpaD. I wanted to determine if this N-terminal IpaB construct also interacts with IpaC. Titration of uniformly ^{15}N-labeled full-length IpaC, IpaC$^{1-100}$ and IpaC$^{173-363}$ with unlabeled IpaB$^{9-226}$ showed **(Figure 3-5)** that IpaC binds to IpaB$^{9-226}$ using its C-terminal domain (IpaC$^{173-363}$).

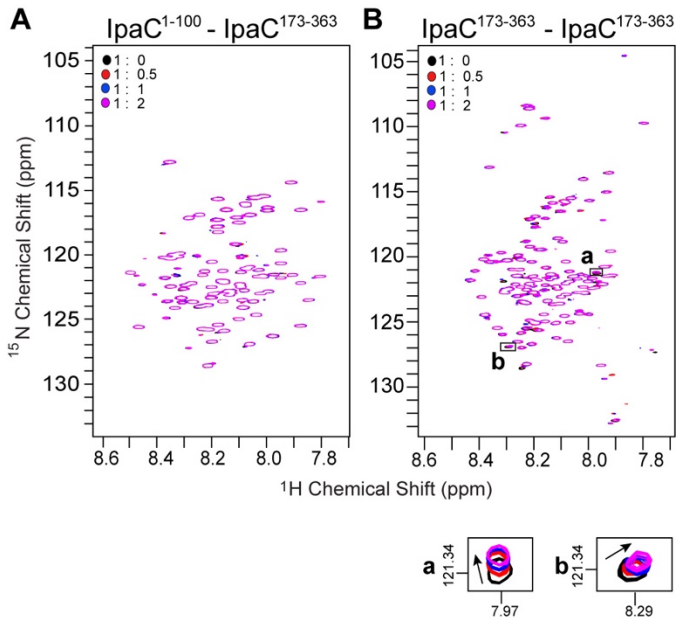

Figure 3-4. IpaC CTD interacts with itself. ^{1}H-^{15}N TROSY spectra show that (A) $IpaC^{173-363}$ does not interact with $IpaC^{1-100}$ and (B) $IpaC^{173-363}$ interacts with itself

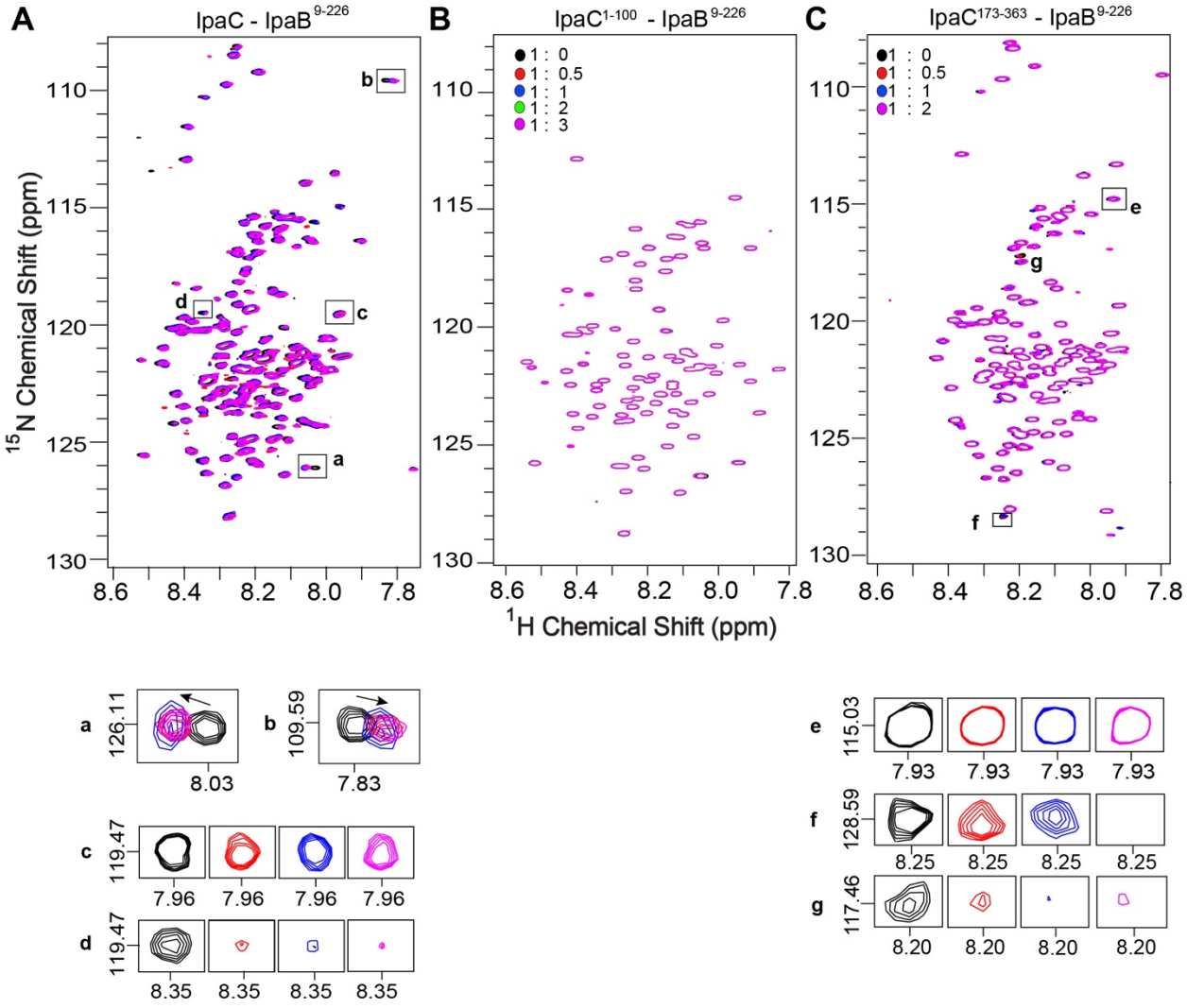

Figure 3-5. IpaC-CTD interacts with IpaB-NTD. ^{1}H-^{15}N TROSY spectra for the titration with IpaB$^{9-226}$ shown for (A) full-length IpaC, (B) IpaC$^{1-100}$ and (C) IpaC$^{173-363}$. Selected affected peaks are also shown for the respective titrations.

3.4.4. IpaC interacts with the tip protein IpaD

Fluorescence studies [13] could not detect any interaction of IpaC with the tip protein IpaD. Here I used NMR to detect similar interactions between *Shigella* IpaC and IpaD. I titrated ^{15}N-labeled IpaC with unlabeled IpaD which resulted in a decrease in peak intensity for specific peaks b, c and d in a concentration-dependent manner **(Figure 3-6A)**, indicating interaction in the intermediate exchange regime with micromolar dissociation constant. The intensity of many peaks (for example, peak a) **(Figure 3-6A)** remained unchanged at the highest concentration of IpaD, indicating that the observed intensity decrease was not due to protein aggregation. Titration of labeled IpaC$^{1\text{-}100}$ with unlabeled IpaD showed changes in positions of specific peaks a,b,c and d with increasing IpaD concentrations **(Figure 3-6B)**, indicating interaction in the fast exchange regime in the NMR timescale with millimolar dissociation constant. IpaC$^{173\text{-}363}$, titrated with increasing molar concentrations of IpaD, resulted in a concentration-dependent decrease in peak intensity **(Figure 3-6C)**. My NMR data indicate that full-length IpaC and its NTD and CTD interact with tip protein IpaD. The data further suggest that the CTD of IpaC (IpaC$^{173\text{-}363}$) interacts with IpaD **(Figure 3-6C)** with a stronger affinity than the NTD (IpaC$^{1\text{-}100}$) **(Figure 3-6B)**.

The NMR assignments of IpaC and its domains are currently unknown. However, the NMR spectra of IpaD were assigned previously [20]. I used the ^{15}N-IpaD assignments to identify the residues and surfaces of IpaD interacting with full-length IpaC and its domains **(Figure 3-7)**. The intensity ratios ($I_{1:1}/I_{1:0}$) were calculated from the NMR titration data using molar concentration ratios of 1:1 and 1:0, and subtracted from 1, which is the intensity ratio in absence of any interaction, to calculate (1-Intensity Ratio) for all IpaD residues. Plots of (1-Intensity Ratio) of the residues revealed the strongly affected IpaD peaks **(Figure 3-8)**. Further, the (1-Intensity Ratio) of the strongly affected peaks were mapped onto the crystal structure of IpaD (PDB ID: 2J0O)

[15] to identify the surfaces of IpaD interacting with IpaC and its domains **(Figure 3-9)**. All three NMR titrations show **(Figure 3-7, Figure 3-8** and **Figure 3-9)** that the N-terminal α-helical hairpin of IpaD contains many strongly affected residues and is the major interacting region.

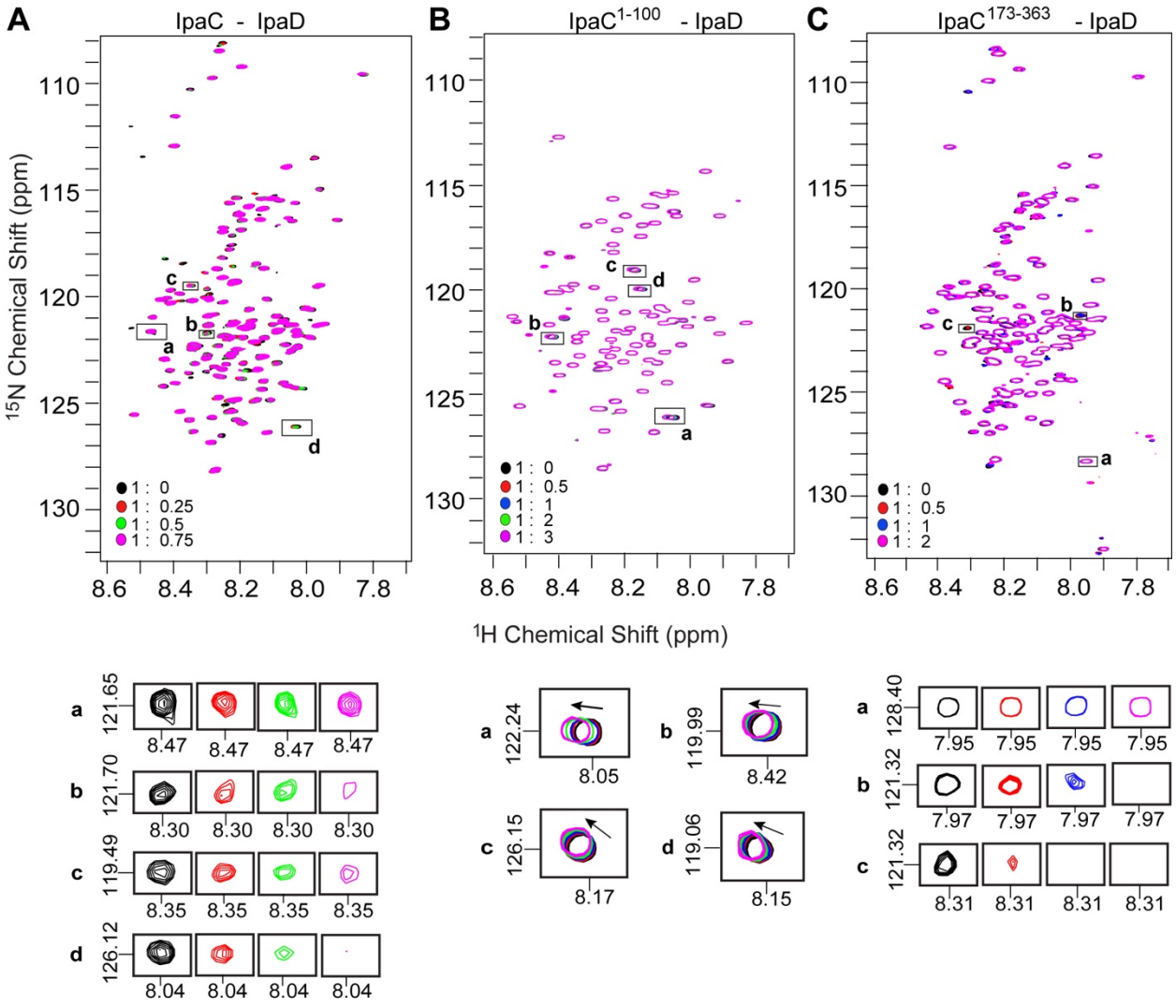

Figure 3-6. Both the NTD and CTD of IpaC interact with IpaD. ^{1}H-^{15}N TROSY spectra for the titration with unlabeled IpaD shown for (A) full-length IpaC, (B) IpaC$^{1-100}$ and (C) IpaC$^{173-363}$. Selected affected peaks are also shown for the respective titrations.

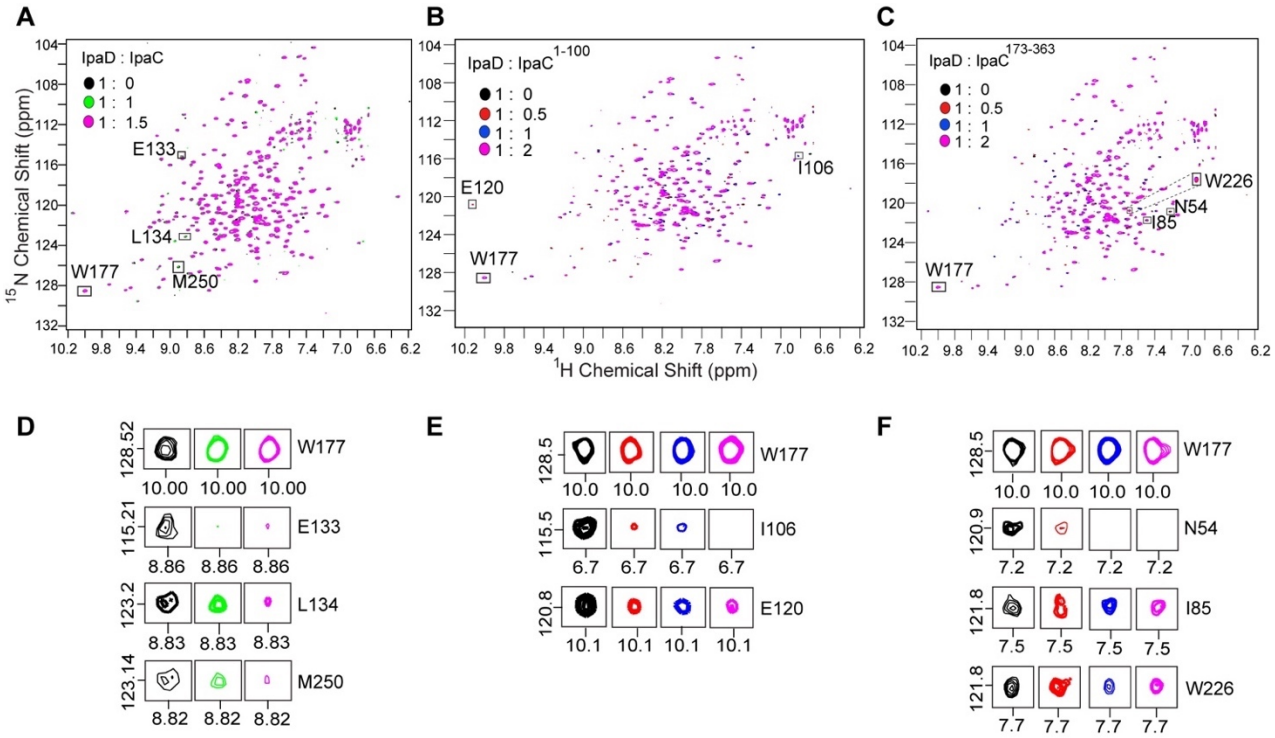

Figure 3-7. ^{15}N-IpaD interacts with IpaC. ^{1}H-^{15}N TROSY spectra for the titration of IpaD with (A) full-length IpaC, (B) IpaC$^{1-100}$ and (C) IpaC$^{173-363}$. Selected affected peaks are shown for titration with (D) full-length IpaC, (E) IpaC$^{1-100}$ and (F) IpaC$^{173-363}$.

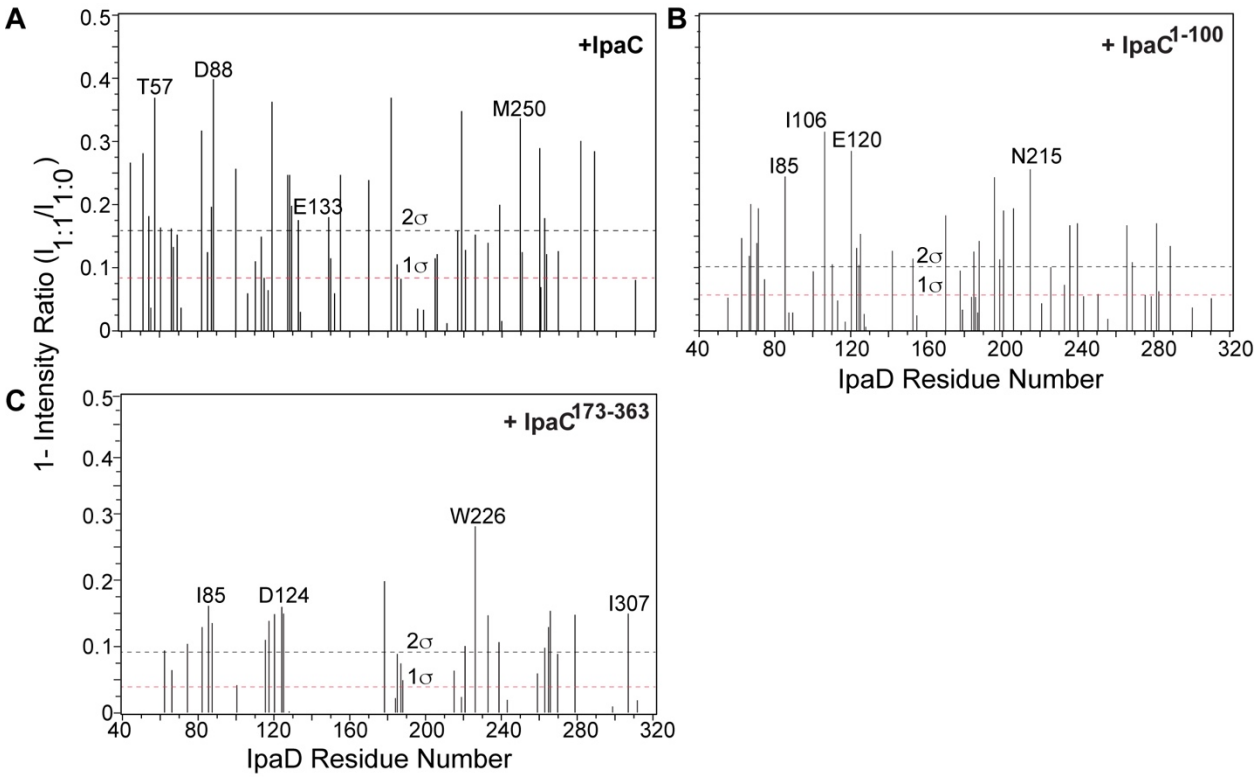

Figure 3-8. Plots of (1-Intensity Ratio) of IpaD residues. The (1-Intensity Ratio) for all IpaD residues were plotted for ^{15}N titrations with (A) full-length IpaC, (B) $IpaC^{1-100}$ and (C) $IpaC^{173-363}$. The peaks showing (1-Intensity Ratio) values greater than 2σ were plotted on the IpaD crystal structure.

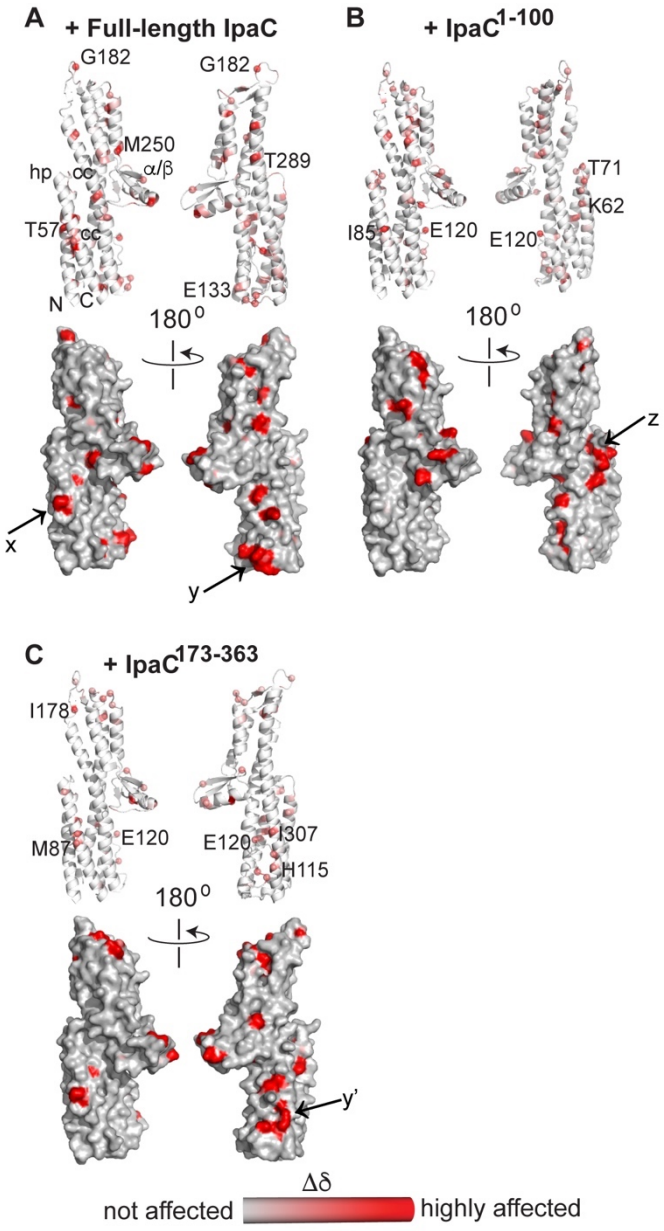

Figure 3-9. IpaC interaction regions on IpaD. IpaC and its domains interact with the N-terminal α-helical hairpin of IpaD. Strongly affected IpaD residues were plotted on the IpaD crystal structure. The residues, colored red (strongly affected) to grey (not affected), are shown for ^{15}N titrations with (A) full-length IpaC, (B) $IpaC^{1-100}$ and (C) $IpaC^{173-363}$. The different domains of IpaD are labeled as *hp* (hairpin), *cc* (coiled-coil) and *α/β* (mixed αβ domain). Pockets x, y and z are the major interacting surfaces on IpaD.

Titration of IpaD with full-length IpaC showed decrease in peak intensities of specific ^{15}N peaks of IpaD, such as E133, L134 and M250 **(Figure 3-7D)**. The majority of the strongly affected residues, such as T57, D88, K119 and E133, were clustered in the N-terminal α-helical hairpin **(Figure 3-8A** and **Figure 3-9A)**. The affected residues formed two binding surfaces on the IpaD hairpin (**Figure 3-9A)**, designated as pockets x and y. Pocket x is formed by residues L51, I54, T57, S82, M87 and D88, while pocket y is formed at the bottom of the central coiled-coil domain by residues Q127, M128 and I129 located in the loop connecting the hairpin to the coiled-coil (**Figure 3-9A)**. Pockets x and y are separated by ~ 28 Å, and are possibly brought closer due to conformational changes upon interacting with IpaC. Titration of IpaD with $IpaC^{1-100}$ and $IpaC^{173-363}$ (**Figure 3-7B** and **Figure 3-7C)** showed that the truncated IpaC constructs retained their ability to bind to the hairpin, even though the interacting surfaces were altered (**Figure 3-9B** and **Figure 3-9C**). For example, $IpaC^{1-100}$ binds to pocket z, formed by strongly affected residues located in the hairpin, such as K62, S66, Q67, L70, T71 and I85 **(Figure 3-8B** and **Figure 3-9B)**. $IpaC^{173-363}$ binds to IpaD predominantly at pocket y' **(Figure 3-9C)** formed by strongly affected residues H115, A117, D124 and G125 at the bottom of the coiled-coil. Another interaction surface is formed by I307 in coiled-coil and E120 in the hairpin **(Figure 3-9C)**. The NMR titrations further show that other residues located in the mixed $\alpha\beta$ domain and coiled-coil of IpaD, such as G182, M250 and T289 **(Figure 3-9A)**, N215 **(Figure 3-8B)** and W226 **(Figure 3-8C)**, are strongly affected. But these interactions are most likely non-specific. Overall, NMR data suggest that in solution, the *Shigella* tip protein IpaD interacts with IpaC predominantly through its N-terminal α-helical hairpin and the lower half (proximal region) of the coiled-coil. I reported in the previous chapter that less than 20 NMR peaks of each of $IpaC^{1-100}$ and $IpaC^{173-363}$ can be superimposed on the NMR spectrum of full-length IpaC, indicating that the tertiary structures of the IpaC domains are different from that of full-length IpaC in solution. This possibly contributes to the use of different binding surfaces by IpaD to interact with full-length IpaC, $IpaC^{1-100}$ and $IpaC^{173-363}$.

3.5. Discussion

The final step in the T3SS assembly is the formation of a translocon pore in the host cell membrane, allowing the passage of virulence effector proteins to the host cytoplasm (Sukanya Chaudhury, University of Kansas 2013; Dissertation). In *Shigella*, the minor translocon protein IpaC performs this critical step together with the major translocon protein IpaB, making IpaC an essential component in the bacterial virulence. Determining the protein-protein interactions of IpaC is thus critical to understand the assembly of the T3SS.

In this chapter, I used NMR to show that refolded full-length IpaC and $IpaC^{1-100}$ are functional because it binds to its chaperone IpgC **(Figure 3-2)**. NMR data identified previously unknown binding partners of IpaC in solution. $IpaC^{173-363}$, the C-terminal domain of IpaC, can also bind IpgC **(Figure 3-3)**. Additionally, the C-terminal domain interacts with itself in solution, suggesting its involvement in IpaC oligomerization **(Figure 3-4)**. This IpaC domain is also involved in interacting with the N-terminal domain of the major translocon protein IpaB **(Figure 3-5)**. Further, I have reported for the first time that in solution, both the N-terminal and the C-terminal domains of IpaC harbor binding sites for the *Shigella* tip protein IpaD **(Figure 3-6)**. NMR data also show that the N-terminal α-helical hairpin of IpaD is involved in interacting with IpaC and its domains (**Figure 3-7**, **Figure 3-8** and **Figure 3-9**). All protein-protein interactions reported in this chapter occur either in the fast exchange regime **(Figures 3-2A, 3-2B, 3-3, 3-4A, 3-5A** and **3-6B)** with millimolar dissociation constants, or in the intermediate exchange regime (**Figures 3-2A, 3-5A, 3-5C, 3-6A, 3-6C** and **3-7)** with high micromolar dissociation constants. Aggregation was ruled out as a reason for the peak intensity decreases observed in case of intermediate exchange interactions, because peak intensites of many peaks remained unchanged at the highest molar

concentration ratios, for example, W177 **(Figure 3-7A)**. My findings are summarized in **Figure 3-10**.

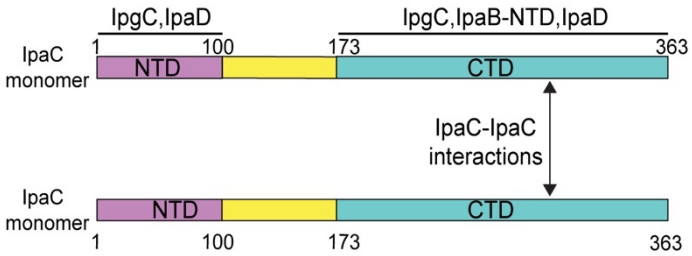

Figure 3-10. Summary of IpaC protein-protein interactions in solution. My NMR data show that the IpaC N-terminal domain (NTD) interacts with chaperone IpgC and tip protein IpaD. The C-terminal domain (CTD) of IpaC interacts with IpgC, IpaD and the N-terminal domain of the major translocon protein IpaB. The IpaC CTD also interacts with itself.

The N-terminal domain and parts of the central hydrophobic transmembrane domain of IpaC were previously [10, 16, 21] found to constitute the IpgC-binding site. Francis *et al* [22] showed that the N-terminal domain as well as an amphipathic domain in the C-terminal region of the *Yersinia* minor translocon protein YopD interacts with its chaperone LcrH predominantly through hydrophobic interactions. IpaC$^{1-100}$ has 31 hydrophobic residues, while IpaC$^{173-363}$ has 60 hydrophobic residues. IpaC has ~ 19% sequence identity and ~ 38% similarity with YopD, and possibly employs similar mode of hydrophobic interactions with IpgC using both the N-terminal and the C-terminal domains. The involvement of both these domains in binding IpgC possibly aids IpgC to efficiently sequester IpaC in the bacterial cytosol, preventing its premature secretion. IpaC-IpaC interactions were reported previously [13, 19]. My data (**Figure 3-4A**) provide the first direct evidence using NMR that the C-terminal domain is involved in IpaC oligomerization via homotypic interactions with neighboring C-terminal domains. This finding is in agreement with the observation that the C-terminal domain of the *Yersinia* homolog YopD can oligomerize [22]. The ability of the IpaC C-terminal domain to oligomerize is possibly important for the incorporation of IpaC into large multiprotein complexes upon host cell contact [23].

Residues 367-458 in the C-terminal domain of the major translocon protein IpaB was previously reported [24] to constitute the IpaC-binding site. My NMR data show that in solution, the C-terminal domain of the minor translocon protein IpaC additionally interacts with IpaB$^{9-226}$ (**Figure 3-5C**), the same IpaB N-terminal domain that was previously found to interact with the tip protein IpaD [14]. Myeni *et al* [25] reported that C-terminal residues 340-409 in the *Salmonella* minor translocon protein SipC contains the binding site for the major translocon protein SipB, and this interaction is essential for the membrane incorporation of SipC. IpaC shares ~31% sequence identity and ~49% similarity with SipC, thus possibly employing a similar mode of interaction with IpaB. Complex formation between 8 subunits of the *Pseudomonas* minor translocon protein PopD with 8 subunits of the major translocon protein PopB in membranes is known [1]. NMR titration

experiments should be performed in presence of membrane mimetics such as liposomes and phospholipids, or in micelles such as DPC, to better understand the protein-protein interactions between the *Shigella* major and minor translocon proteins in a lipid environment.

The interaction between IpaC and tip protein IpaD has not been reported previously. In fact, fluorescence anisotropy studies [13] failed to detect IpaC-IpaD interactions. The discrepancy between fluorescence and my NMR data may be due to the difference in the protein concentrations used in the two techniques. Fluorescence polarization [13] used 30 nM labeled IpaC with up to 100 nM unlabeled IpaD. I used 0.1 mM of ^{15}N-labeled IpaC in my NMR titrations (**Figure 3-6A**). NMR data (**Figure 3-6**) suggest that in solution, the C-terminal domain of IpaC interacts with IpaD with a higher affinity than the N-terminal domain. The N-terminal α-helical hairpin of IpaD is the domain that interacts with IpaC (**Figure 3-9A**) and the interaction involves pockets x and y. This domain is important for anchoring IpaD to the tip of the needle protein MxiH [26]. This presents the possibility that small molecule inhibitors can bind to N-terminal hairpin and disrupt the IpaD-IpaC interactions. NMR suggests (**Figure 3-9B** and **Figure 3-9C**) that IpaC truncations do not alter the IpaD-IpaC interaction significantly, as IpaC$^{1\text{-}100}$ and IpaC$^{173\text{-}363}$ also interact predominantly with the α-helical hairpin of IpaD. Interestingly, Yadava *et al* showed (Yadava, Sanjay; Thesis 2018, University of Kansas) that NMR titration of IpaD with the *Burkholderia* BipC resulted in concentration-dependent intensity decrease of I106 and L134 located in the α-helical hairpin of IpaD, indicating that this domain is involved in the interaction. This suggests that the tip-minor translocon interaction region is conserved across the different bacterial species. Further experiments using *Salmonella* SipC can confirm this proposition.

In conclusion, I have identified previously unknown binding partners of IpaC. My findings have established the importance of the C-terminal domain of IpaC in IpaC oligomerization, as well

as in binding the chaperone protein IpgC and the N-terminal domain of IpaB. NMR titrations showed that IpaC binds predominantly to two binding pockets x and y located on the N-terminal α-helical hairpin of IpaD. My findings extend our knowledge of the protein-protein interactions of IpaC, and are important for the future design of novel antimicrobials that can disrupt these interactions to reduce drug resistance in the pathogenic bacteria.

3.6. References

1. Romano, F.B., et al., *Type 3 secretion translocators spontaneously assemble a hexadecameric transmembrane complex.* J Biol Chem, 2016. **291**(12): p. 6304-15.
2. Tran Van Nhieu, G., et al., *IpaC induces actin polymerization and filopodia formation during Shigella entry into epithelial cells.* EMBO J, 1999. **18**(12): p. 3249-62.
3. Veenendaal, A.K., et al., *The type III secretion system needle tip complex mediates host cell sensing and translocon insertion.* Mol. Microbiol., 2007. **63**(6): p. 1719-30.
4. Blocker, A., et al., *The tripartite type III secreton of Shigella flexneri inserts IpaB and IpaC into host membranes.* J.Cell Biol., 1999. **147**(3): p. 683-693.
5. Marquart, M.E., W.L. Picking, and W.D. Picking, *Soluble invasion plasmid antigen C (IpaC) from Shigella flexneri elicits epithelial cell responses related to pathogen invasion.* Infect. Immun., 1996. **64**(10): p. 4182-4187.
6. Menard, R., et al., *The secreted Ipa complex of Shigella flexneri promotes entry into mammalian cells.* Proc. Natl. Acad. Sci. U.S.A., 1996. **93**(3): p. 1254-8.
7. Russo, B.C., J.K. Duncan, and M.B. Goldberg, *Topological Analysis of the Type 3 Secretion System Translocon Pore Protein IpaC following Its Native Delivery to the Plasma Membrane during Infection.* MBio, 2019. **10**(3).
8. Russo, B.C., et al., *Intermediate filaments enable pathogen docking to trigger type 3 effector translocation.* Nat Microbiol, 2016. **1**: p. 16025.
9. Menard, R., et al., *Extracellular association and cytoplasmic partitioning of the IpaB and IpaC invasins of S. flexneri.* Cell, 1994. **79**(3): p. 515-25.
10. Harrington, A.T., et al., *Structural characterization of the N terminus of IpaC from Shigella flexneri* Infect. Immun., 2003. **71**(3): p. 1255-1264.
11. Page, A.-L., et al., *Characterization of the interaction partners of secreted proteins and chaperones of Shigella flexneri.* Mol. Microbiol., 2001. **42**(4): p. 1133-1145.
12. Terry, C.M., et al., *The C-terminus of IpaC is required for effector activities related to Shigella invasion of host cells.* Microb. Pathog., 2008. **45**(4): p. 282-9.
13. Davis, R., et al., *Protein-protein interactions in the assembly of Shigella flexneri invasion plasmid antigens IpaB and IpaC into protein complexes.* Biochim.Biophys.Acta, 1998. **1429**(1): p. 45-56.
14. McShan, A.C., et al., *NMR Identification of the Binding Surfaces Involved in the Salmonella and Shigella Type III Secretion Tip-Translocon Protein-Protein Interactions.* Proteins, 2016. **84**: p. 1097-1107.

15. Johnson, S., et al., *Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD.* J. Biol. Chem., 2007. **282**(6): p. 4035-4044.
16. Birket, S.E., et al., *Preparation and characterization of translocator/chaperone complexes and their component proteins from Shigella flexneri.* Biochemistry, 2007. **46**(27): p. 8128-37.
17. Gazi, A.D., et al., *Coiled-coils in type III secretion systems: structural flexibility, disorder and biological implications.* Cell Microbiol, 2009. **11**(5): p. 719-29.
18. Pallen, M.J., G. Dougan, and G. Frankel, *Coiled-coil domains in proteins secreted by type III secretion systems.* Mol. Microbiol., 1997. **25**(2): p. 423-5.
19. Picking, W.L., et al., *Identification of functional regions within invasion plasmid antigen C (IpaC) of Shigella flexneri.* Mol Microbiol, 2001. **39**(1): p. 100-11.
20. Dickenson, N.E., et al., *Conformational Changes in IpaD from Shigella flexneri upon Binding Bile Salts Provide Insight into the Second Step of Type III Secretion.* Biochemistry, 2011. **50**: p. 172-180.
21. Lunelli, M., et al., *IpaB-IpgC interaction defines binding motif for type III secretion translocator.* Proc. Natl. Acad. Sci. U.S.A., 2009. **106**(24): p. 9661-6.
22. Francis, M.S., et al., *A study of the YopD–LcrH interaction from Yersinia pseudotuberculosis reveals a role for hydrophobic residues within the amphipathic domain of YopD.* Mol. Microbiol., 2000. **38**(1): p. 85-102.
23. Marquart, M.E., W.L. Picking, and W.D. Picking, *Structural analysis of invasion plasmid antigen D (IpaD) from Shigella flexneri.* Biochem.Biophys.Res.Commun., 1995. **214**(3): p. 963-970.
24. Yang, S.C., et al., *The roles of the virulence factor IpaB in Shigella spp. in the escape from immune cells and invasion of epithelial cells.* Microbiol Res, 2015. **181**: p. 43-51.
25. Myeni, S.K., L. Wang, and D. Zhou, *SipB-SipC complex is essential for translocon formation.* PLoS One, 2013. **8**(3): p. e60499.
26. Zhang, L., et al., *Identification of the MxiH needle protein residues responsible for anchoring invasion plasmid antigen D to the type III secretion needle tip.* J. Biol. Chem., 2007. **282**(44): p. 32144-51.

Chapter 4: Characterization of Novel Small Molecules Identified Using Surface Plasmon Resonance For Binding to the *Salmonella* Tip Protein SipD

4.1. Abstract

The emergence of antibiotic resistance in bacterial pathogens is a global health concern. The number of small molecules known to bind to the tip proteins of the T3SS is limited. The tip protein of the T3SS is surface exposed and plays an essential role in virulence. Small molecules based on hydroxyindole, morpholinoaniline and indole-acetic acid scaffolds have been previously identified to bind to the *Salmonella* tip protein SipD, a key virulence protein of the *Salmomella* T3SS. Through a collaboration, surface plasmon resonance (SPR) was used to screen a library of 104 fragments with structures similar to known antibiotics, and another library consisting of 440 compounds with structures similar to intermediates in the synthesis of larger drug-like candidates. The binding of quinolin-8-amine, 4-methoxybenzene-1-carboximidamide and 2-(3,4-dimethoxyphenyl)-ethan-1-amine to SipD were identified. I then used NMR to identify the residues and surfaces of SipD that interact with these small molecules. My results increase the known number of fragments that can bind to SipD, and this knowledge will aid in designing new molecules that target a key virulence protein in pathogenesis.

4.2. Introduction

Pathogenic Gram-negative bacteria like *Salmonella* cause gastroenteritis and death due to food-borne illness [1]. Antibiotic resistance among pathogens that require the T3SS for virulence [2, 3] and the dearth of new antibiotics in the pipeline [4, 5] are major global health concerns that necessitate the development new antimicrobials. The T3SS is an attractive target for developing new antimicrobials because of it is essential in virulence and its presence only among pathogens [6]. SipD is the tip protein in *Salmonella* and is exposed on the cell surface

prior to host cell contact [7]. Upon contact with the host cells, the translocon assembles on the tip protein complex and forms a pore in the host cell membrane to allow the passage of bacterial effector proteins from the bacteria to the host cytosol [7]. The crystal structure of SipD [8, 9] revealed that it contains an N-terminal α-helical hairpin, a central coiled-coil and a mixed $\alpha\beta$ domain [8].

An important step in developing new drug candidates against the T3SS is the identification of novel small molecules that can bind to different protein components of the T3SS [15]. SipD is an attractive target for developing novel antimicrobials because it is important in pathogenesis and is exposed on the bacterial surface. Currently, limited number of small molecules are known to bind the T3SS tip proteins, such as bile salts, which bind to SipD [10, 11] and the *Shigella* tip protein IpaD [12, 13], and small molecule compounds screened from the Zenobia library 2.0 for SipD [14] and IpaD [15]. These small molecules were based on hydroxyindole, morpholinoaniline, indole, methylquinoline and pyrrolidine-aniline scaffolds. The Zenobia library contains molecules having drug-like properties and structures presenting possibilities of elaboration, modification and linking. In collaboration with the Center of Biomedical Research Excellence in Protein Structure and Function at the University of Kansas, we used SPR-based screening to identify new small molecule scaffolds binding to SipD. Two small molecule libraries were screened - a targeted library whose members were substructures of known antibiotics, and a non-targeted library whose members were intermediates in the synthesis of larger drug candidates. I used NMR to determine the residues and surfaces of SipD that interact with these small molecules and to identify potential druggable hotspots in SipD. My findings can be used in future to design new antimicrobials against the multidrug resistant bacteria.

4.3. Materials and Methods

4.3.1. Small molecule libraries

Two different small molecule libraries were synthesized by Jeff Aube's Center for Chemical Methodologies and Library Development at the University of Kansas. One was a targeted library containing 134 compounds (Library 1) whose structures were similar to known antibiotics. The other was a non-targeted library of 489 compounds (Library 2) whose structures resembled those of intermediates found in the synthesis of drug-like candidates. The compound libraries were all synthesized satisfying "Rule of Three" [16] to make them druglike.

4.3.2. Expression and Purification of SipD

The expression and purification of SipD (residues 39-343, C244S) were done as described previously [8, 14] with minor modifications. Briefly, the SipD construct contained an N-terminal His_6-tag, a soluble GB1 domain and a tobacco etch virus (TEV) protease cleavage site. The SipD expression plasmid was transformed into *E. coli* BL21(DE3) DNAY competent cells and grown in culture media containing carbenicillin (100 μg/mL) and kanamycin (30 μg/mL). 1 L of LB medium was inoculated with 10 mL of LB starter culture [15] and used to obtain unlabeled proteins. Cells were grown at 37°C, and 1 mM IPTG was used to induce protein expression at OD_{600} ~ 0.8. Cell growth was continued overnight at 15°C until OD_{600} ~ 2.9, harvested and SipD was purified by nickel-affinity chromatography followed by digestion with tobacco etch virus protease as described previously [8, 14]. For NMR, the backbone amides of SipD were labeled with ^{15}N, and the sidechain methyl groups of isoleucine, leucine and valine residues were labeled with ^{13}C [15]. To label SipD simultaneously with ^{15}N and ILV, cells were grown in M9 minimal media

supplemented with 1g/L of ^{15}N-ammonium chloride (Sigma). Cells were grown at 37°C, and at OD_{600} ~ 0.4 [15], the culture medium was supplemented with 60 mg/L of 2-ketobutyric acid-4-^{13}C acid sodium salt hydrate (Sigma, #571342) to label the Cδ1 methyl group of isoleucine with ^{13}C and 100 mg/L of 2-keto-3-(methyl-^{13}C)-butyric-4-^{13}C acid sodium salt (Sigma, #571334) to label the leucine Cδ and valine Cγ methyl groups with ^{13}C [8, 14]. 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was used to induce protein expression at an OD_{600} ~ 0.8, and cell growth was continued at 15°C overnight until OD_{600} ~ 2.6. Cells were harvested by centrifugation and ^{15}N/ILV-labeled SipD was purified as described [8, 14].

4.3.3. SPR screening

The surface plasmon resonance (SPR)-based screening of the two small molecule libraries was done in collaboration with the Center of Biomedical Research Excellence in Protein Structure and Function at the University of Kansas as described before [14]. Briefly, SipD was covalently immobilized to the CM5 sensor chip surface (GE Healthcare, #BR100399) by standard amine coupling chemistry using the amine coupling kit (GE Healthcare #BR-1106-33). The running buffer was 1X PBS. The small molecule fragments were injected for 60 sec at a flow rate of 60 µL/min followed by a dissociation period of an additional 60 sec. The unbound fragments were washed with a 1:1 solution of DMSO-water.

4.3.4. NMR Spectroscopy

NMR spectra were acquired as described previously [14, 15] using a Bruker Avance 800 MHz spectrometer equipped with a cryogenic triple resonance or a Bruker 600 MHz spectrometer

equipped with a TXI-RT probe. The NMR data were processed with NMRPipe [17] and analyzed using NMRView [18]. NMR samples containing 0.3 mM of ^{15}N/ILV-labeled SipD in NMR buffer (10 mM sodium phosphate pH 7.4, 20 mM NaCl and 10% D_2O) were used to acquire 2D ^{1}H-^{13}C HSQC and ^{1}H-^{15}N TROSY spectra at 30°C using the 800 MHz NMR. The acquisition parameters for ^{1}H-^{13}C HSQC for ILV-labeled SipD were 8 scans, an 18 ppm ^{13}C sweep width with the center at 18 ppm and ^{1}H sweep width of 10 ppm centered at 4.69 ppm, whereas typical ^{1}H-^{15}N acquisition parameters were 16 scans at 30 ppm with an ^{15}N sweep width with the center at 118 ppm.

4.3.5. NMR titrations

The NMR titrations were performed as described previously [14, 15]. Briefly, the small molecules were dissolved in ~250 μL of 100% [D_6] DMSO (Cambridge Isotope Laboratories, Inc., Andover, MA) to make stock solutions of 2 M. The compounds were titrated into 0.3 mM ^{15}N/ILV-labeled SipD at various molar ratios of SipD to compounds. Five titration points were obtained for each of the three compounds. All the NMR titration samples contained final [D_6] DMSO concentration of 1.5%. The previous ^{15}N and ILV NMR assignments of SipD [11, 19] were used to identify the residues of SipD that were affected upon binding to the compounds. The weighted chemical shift deviation ($\Delta\delta$) were calculated using the equation $\Delta\delta_{HN} = ½ [(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2]$ [20] for the backbone amides and $\Delta\delta_{ILV} = ½ [(\Delta\delta_H)^2 + (\Delta\delta_C/2)^2]$ for the ILV spectra [14]. The weighted chemical shift deviations of specific SipD residues were plotted vs. compound concentrations and dissociation constants were estimated. A single-site binding model was used to perform curve fittings using the GraphPad Prism Version 9 (GraphPad Software, Inc., USA). The R^2 values of the fits were ~ 0.97 – 0.99.

4.3.6. Saturation Transfer Difference (STD) NMR

One-dimensional ^{1}H STD NMR [21] data were acquired as described previously [14] with minor modifications. Briefly, STD-NMR data were acquired at 30°C using a Bruker 600 MHz NMR equipped with a TXI-RT probe using the pulse program stddiffesgp.3 [21]. The STD-NMR data were processed using MestReNova (v14.1.1-24571, Mestrelab Research, Spain) [22]. Each sample contained protein and compound at 1:100 molar ratio (40 μM protein, 4000 μM compound, 10% D_2O, 0.4% [D_6] DMSO). The protein was saturated with a Gaussian pulse of 50 ms for a period of 2 s. The off-resonance pulse was fixed at 40 ppm, while the on-resonance pulse was varied from -0.2 to 0.1 ppm for an optimal STD signal. The acquisition parameters used were 128 scans, 16 ppm ^{1}H sweep width centered at 4.701 ppm, with a 2 s recycle delay.

4.3.7. Druggable sites on SipD

The servers DoGSiteScorer [23] and PockDrug-Server [24] were used to predict the druggable sites in SipD (PDB ID : 3NZZ) [8].

4.4. Results

4.4.1. Small molecule libraries

The properties of the two small molecule libraries are summarized in **Table 4-1**. The properties of all the library members conformed to the "Rule of Three" for orally active drugs [16].

Table 4-1. Small molecule libraries used for SPR screening

Name of library and "Rule of Three"	Median MW (Daltons)	ClogP	Median # H-bond Acceptors	Median # H-bond Donors
Library 1	143.41	0.19	2.69	1.67
Library 2	216.86	1.61	2.82	1.08
"Rule of Three"	< 300	≤ 3	≤ 3	≤ 3

For example, the ClogP values of all the compounds of the two libraries are less than 3. The ClogP of a compound is a measure of its bioavailability, and is the calculated partition coefficient between water (aqueous phase) and octanol (lipid phase) [25]. All the compounds are less than 300 daltons in size, which can possibly lead to better intracellular diffusion [26]. It has been reported previously [27] that high numbers of hydrogen bond donors and acceptors can decrease the permeability across lipid bilayers. The conformity of the properties of the library

members with those predicted by the "Rule of Three" **(Table 4-1)** [16] suggest that all the compounds can potentially serve as starting lead candidates for future drug development.

4.4.2. SPR screening

SPR screening of the two small molecule libraries identified three compounds that can bind to SipD **(Figure 4-2)**. These compounds are based on **quinoline** [quinolin-8-amine, Compound **1**] and **phenyl** [4-methoxybenzene-1-carboximidamide, Compound **2,** and 2-(3,4-dimethoxyphenyl)-ethan-1-amine, Compound **3**] **(Figure 4-1)** scaffolds. Compounds **1** and **2** are members of the targeted library, having structures similar to known antibiotics; compound **3** belonged to the non-targeted library having structure similar to intermediates in the synthesis of drug-like molecules. Analogs of these compounds were present in the libraries (Compounds **4** and **5**, **Figure 4-1**) and did not bind to SipD. This led to the identification of the important functional groups for binding SipD. For example, in compound **1**, the amino group is important for binding SipD, because replacing it with the hydroxyl group (compound **4**, Library 1) made the molecule bind to SipD, the *Shigella* homolog IpaD and *Yersinia* LcrV (data not shown). Replacement of the quinoline moiety by a purine ring (compound **5**, targeted library) abrogated the binding to SipD (data not shown). This knowledge will be helpful in future to further derivatize these molecules to bind SipD optimally.

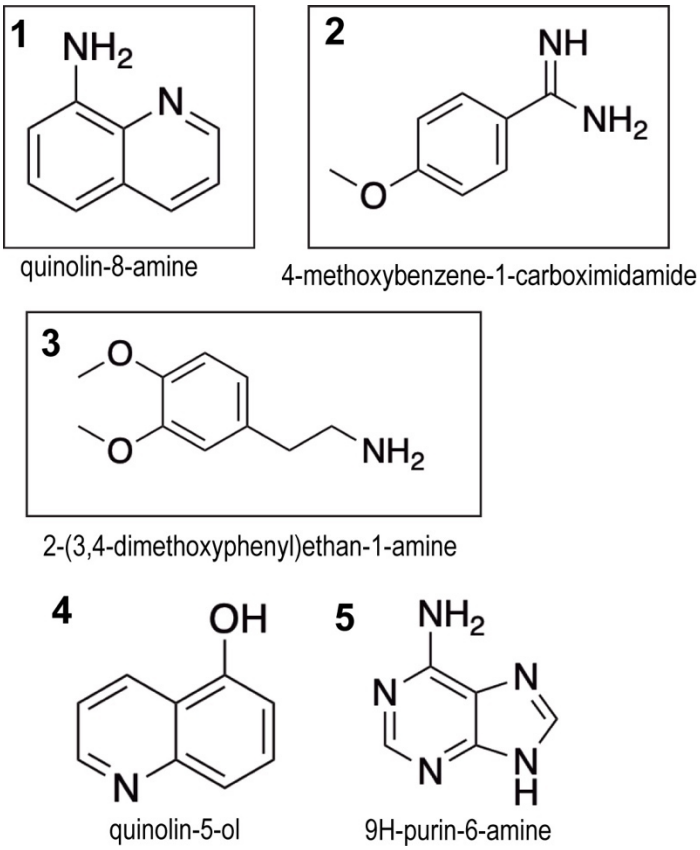

Figure 4-1. The three small molecules that bind to SipD are Compounds **1**, **2** and **3** (boxed). Compounds **4** and **5** are analogues of Compound **1,** and they did not bind to SipD. Compounds **1**, **2**, **4**, and **5** are from the Library 1 consisting of fragments found in antibiotics, whereas Compound **3** is part of the Library 2 with structures similar to intermediates in the synthesis of larger drug-like candidates.

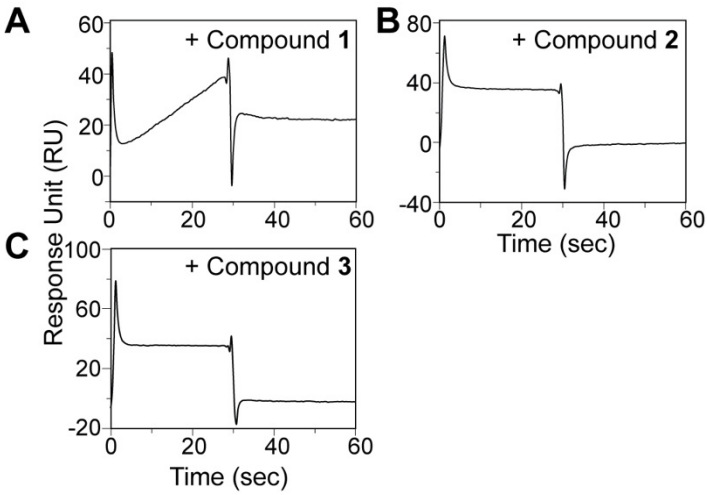

Figure 4-2. Surface Plasmon Resonance (SPR) sensorgrams of SipD with (A) Compound **1**, (B) Compound **2**, and (C) Compound **3**.

4.4.3. STD NMR

I used saturation transfer difference (STD) NMR [21] to identify which groups of the small molecules interact with SipD [15]. In STD NMR, two spectra are recorded by selectively saturating the proton resonances of the protein with a Gaussian pulse that is turned on or off [15]. If a compound binds to the protein, then magnetization is transferred from the protons of the protein to the protons of the compounds, and detected in the 'on-resonance' spectrum. The 'off-resonance' spectra (top panels, **Figure 4-3**) contain peaks from the protein and the small molecules. The STD spectra (bottom panels, **Figure 4-3**), obtained by subtracting the on-resonance spectra (data not shown) from the off-resonance spectra [15], show peaks from only the protons of the compounds that contact the protein. Some protons may not interact with the protein, and their peaks are absent from the STD spectra. My STD NMR data show that for compound **1 (Figure 4-3A)**, all the protons of the quinoline moiety, labeled as a,b,c,d,e and f interact with SipD. Similarly, for compound **2 (Figure 4-3B)**, all the protons of the phenyl ring, marked as d/e and f/g, as well as the protons of the methoxy group, labeled as a/b/c, interact with SipD. Likewise, all the protons of compound **3**, labeled as a/b, c/d, e/f/g, h/i/j and k/l, except those of the terminal amino group, interact with SipD **(Figure 4-3C)**. My STD NMR data further suggest that the compounds **1**, **2** and **3** all bind to SipD on the surface.

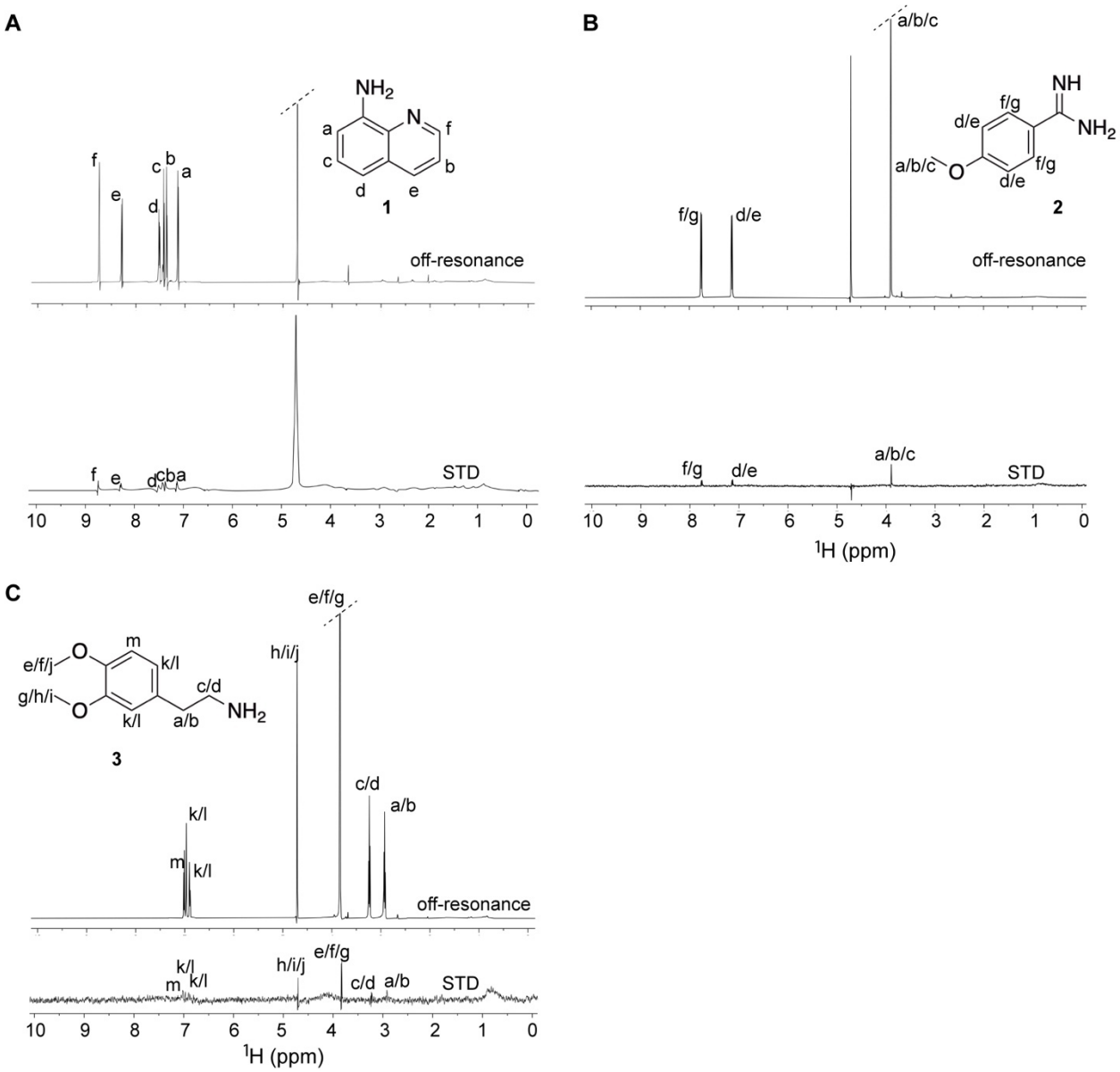

Figure 4-3. STD NMR of SipD with compounds identified by SPR, shown for (A) Compound **1**, (B) Compound **2**, and (C) Compound **3**. The off-resonance spectra (top panels) show the proton assignments of each molecule. The on-resonance spectra (not shown) are subtracted from the off-resonance spectra [15] to obtain the STD spectra (bottom panels). The protons in contact with SipD show reduced peak intensities in the STD spectra.

4.4.4. NMR titrations of ^{15}N/ILV-SipD with the compounds

NMR spectroscopy was used to identify the residues and surfaces of SipD interacting with the compounds. ^{15}N/ILV-labeled SipD was titrated with increasing concentrations of compounds **1**, **2** and **3**, and 2D ^{1}H-^{15}N TROSY and 2D ^{1}H-^{13}C HSQC spectra were recorded. The previous ^{15}N and ILV NMR assignments of SipD [11, 19] were used to identify the residues of SipD that were affected upon binding the small molecules.

SipD titrated with compounds **1**, **2** and **3** showed concentration-dependent chemical shift perturbation of specific ^{15}N-amide **(Figures 4-4A, 4-5A, 4-6A**; **Figures 4-7A, B, C)** and ILV peaks **(Figures 4-4C, 4-5B, 4-6C**; **Figures 4-7D, E, F)**. This indicates that all the three small molecules bind to SipD, showing interaction in the fast exchange regime in the NMR timescale **(Figure 4-7)**. The dissociation constants (K_d) were estimated to be ~ 79 mM (for compound **1**) and ~ 30 mM (for compound **3**) **(Figure 4-4B** and **Figure 4-6B)**, suggesting weak interactions of SipD with compound **1** and **3**. The weighted chemical shift deviations were plotted for the compounds **(Figure 4-8)** to identify the strongly affected ^{15}N and ILV peaks. The weighted chemical shift deviations were mapped onto the structure of SipD [8, 28], showing the surfaces and binding pockets for the compounds on SipD **(Figure 4-9)**.

The most strongly affected ^{15}N residues in the titration with compound **1** (F117, A111 and E133) are located in the 23-residue loop connecting the N-terminal α-helical hairpin to the central coiled-coil. Many strongly affected residues are clustered in the α-helical hairpin – E71, R76, K92 (missing electron density in the SipD crystal structure) and S107 (in the ^{15}N titration), as well as L83, L87 and L90 in the ILV titration **(Figure 4-8A** and **Figure 4-8B)**. The residues L90, K92 and E133 cluster on the same surface of SipD, forming pocket *x* **(Figure 4-9A)**. For titration with

compound **2**, residue V213, located in the mixed αβ-domain, was the most strongly affected ^{15}N peak. The mixed αβ-domain contains many strongly affected residues – T212, G220 (missing electron density in the SipD crystal structure) and K224 (in the ^{15}N spectrum); I208 and V256 (in the ILV spectrum) **(Figures 4-8C,4-8D)**. All these residues form a common binding surface, designated as pocket *y* **(Figure 4-9B)**. For ^{15}N-titration with compound **3**, the most strongly affected residue was G220 located in a loop in the mixed αβ-domain, but with no electron density in the SipD crystal structure [28]. Majority of the affected residues were present in the N-terminal α-helical hairpin and the loop connecting the hairpin to the coiled-coil – I68, R76, T84, S96, A111, F117 and E133 (in the ^{15}N titration); L48 and L90 (in the ILV titration). However, the weighted chemical shift deviation of the ILV peaks were ~ 3 times lower than those of the ^{15}N peaks **(Figures 4-7C**, **7F**, **Figure 4-8E**, **4-8F)**, indicating that compound **3** mainly interacts with the backbone amide groups of SipD. Residues K92, S96 and E133 cluster on the same surface at the bottom of the coiled-coil and form pocket *x* **(Figure 4-9C)**, indicating that compound **3** also binds to pocket *x*. Thus, results from the ILV titrations complemented the results from the ^{15}N titrations. My NMR data collectively suggest that SipD uses polar contacts mediated by electrostatic interactions and hydrogen bonding as well as hydrophobic contacts to interact with the small molecules.

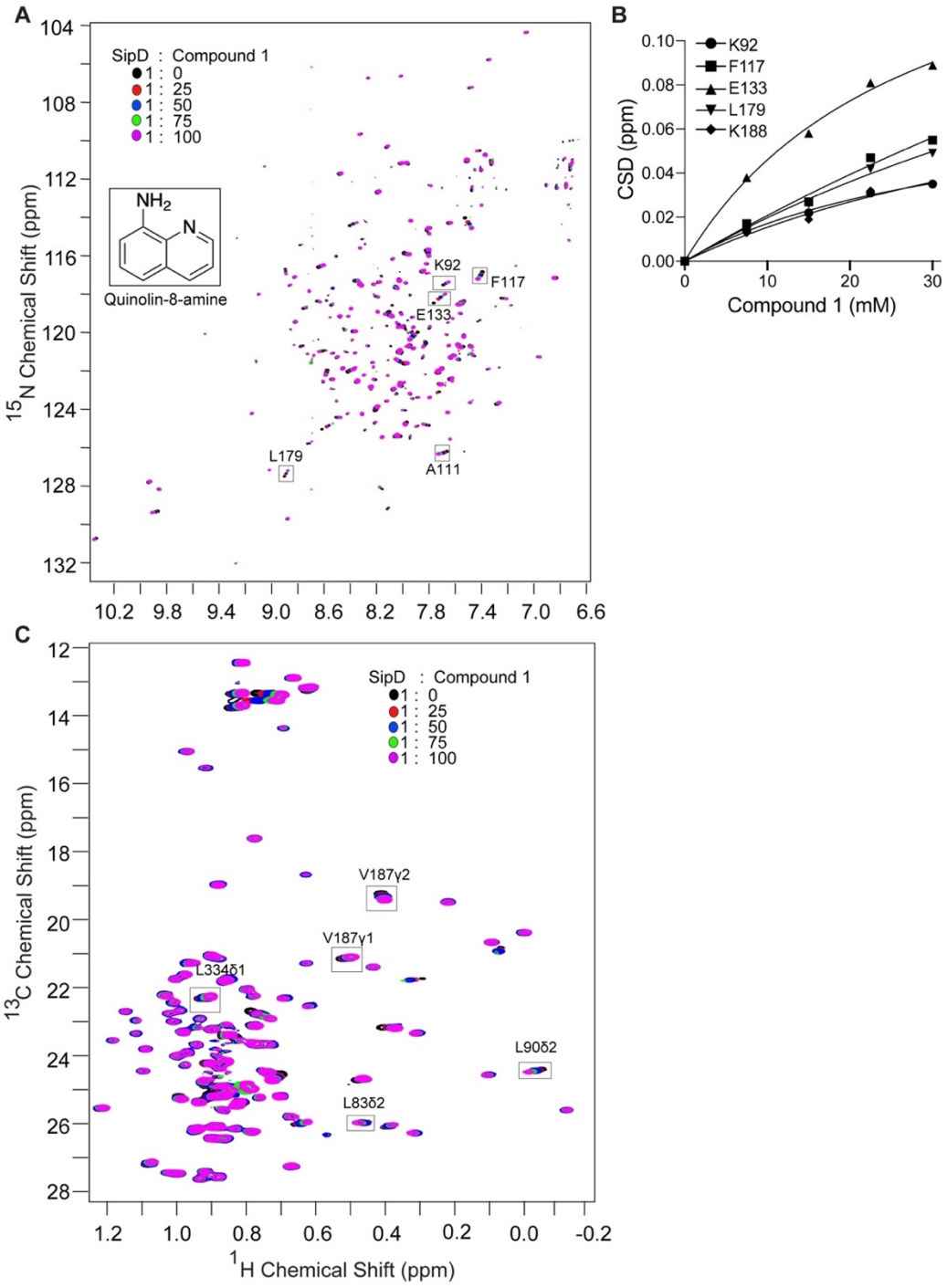

Figure 4-4. NMR titration of SipD with Compound 1. (A) ^{15}N-titration of SipD with Compound **1**. (B) Weighted chemical shift deviation (CSD) vs. concentration of Compound **1** for selected ^{15}N peaks. (C) ILV-titration of SipD with Compound **1**.

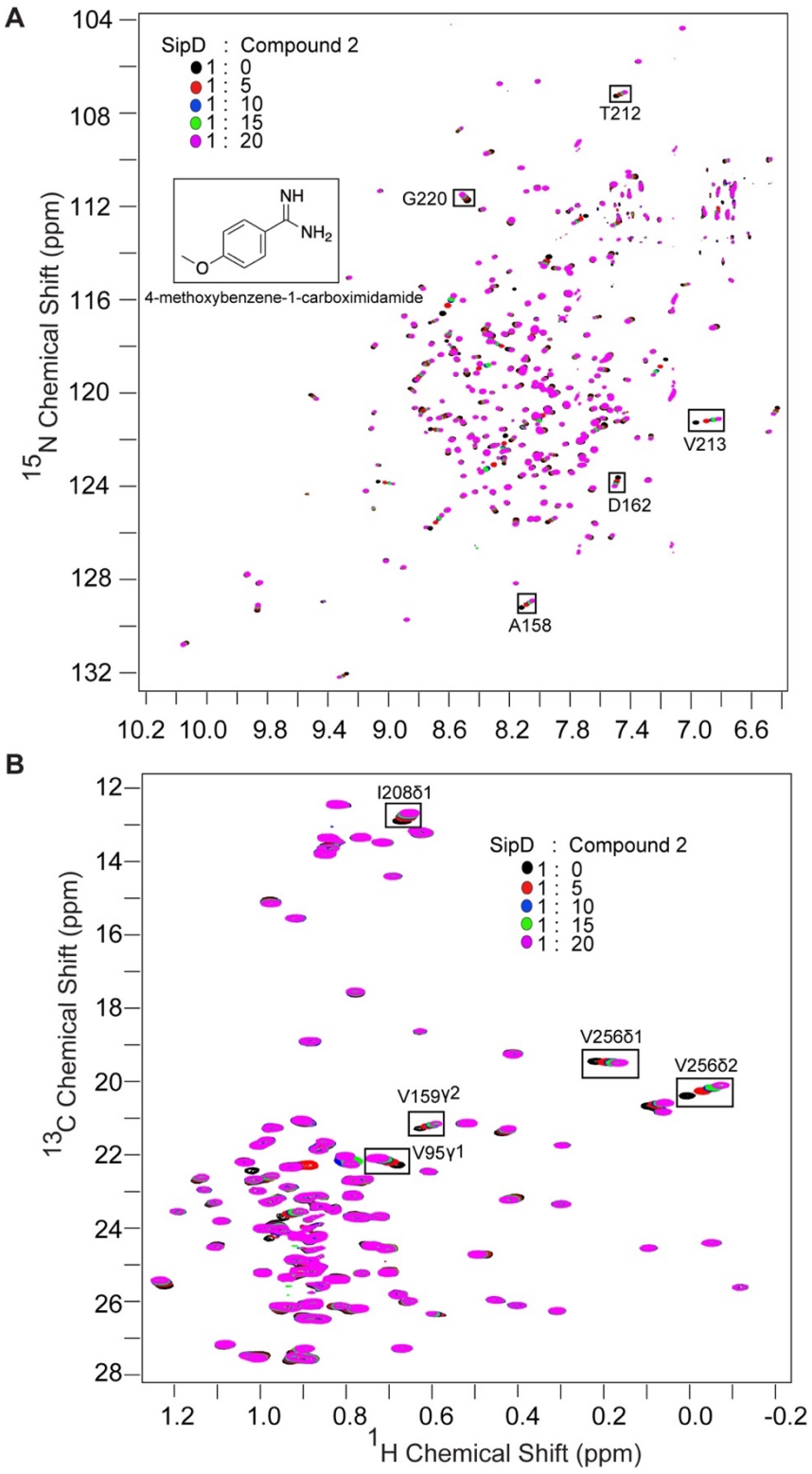

Figure 4-5. NMR titration of SipD with Compound 2. (A) ^{15}N and (B) ILV titrations of SipD with Compound **2**.

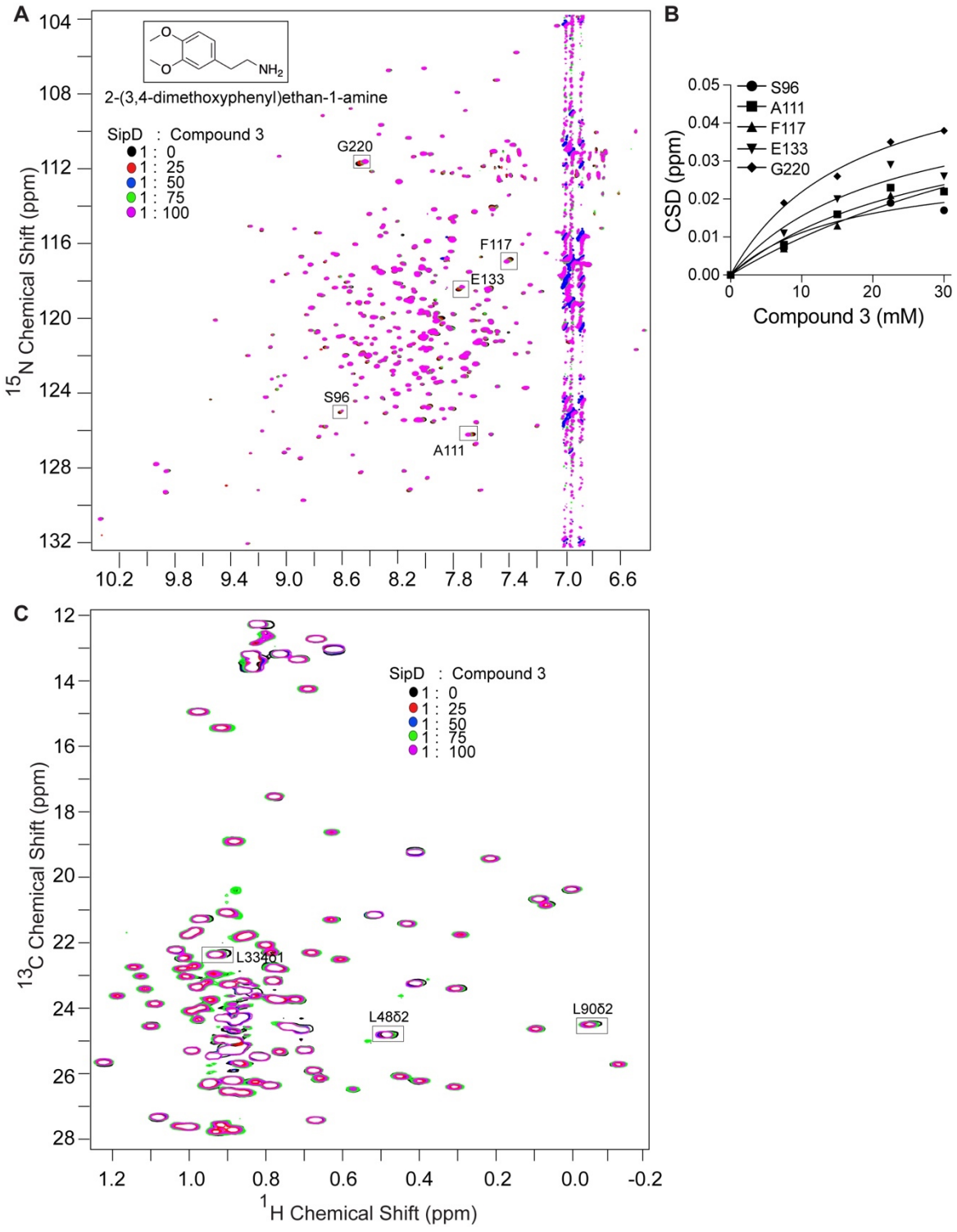

Figure 4-6. NMR titration of SipD with Compound 3. (A) ^{15}N-titration of SipD with Compound **3**. (B) Weighted chemical shift deviation (CSD) vs. concentration of Compound **3** for selected ^{15}N peaks. (C) ILV-titration of SipD with Compound **3**.

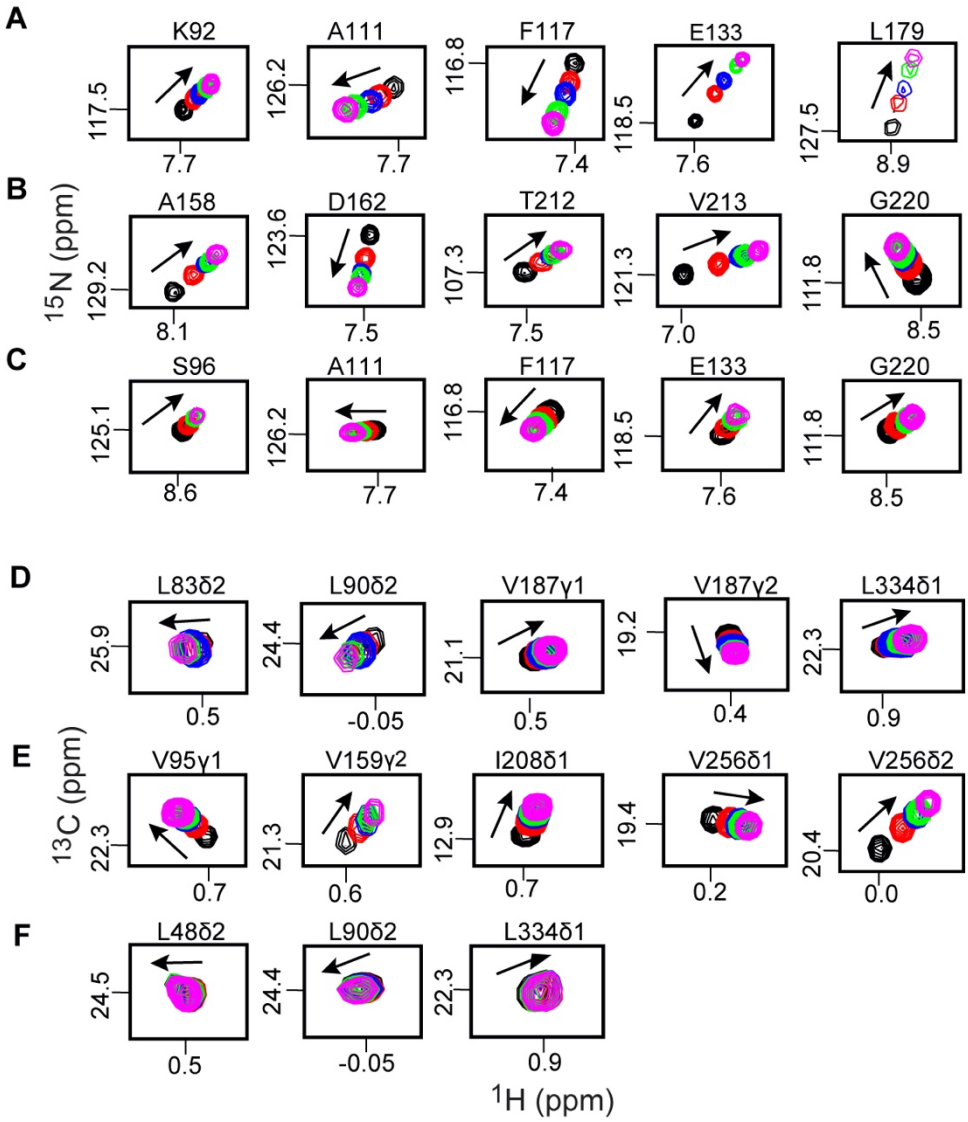

Figure 4-7. Selected affected peaks from the NMR titrations of SipD with the compounds. Specific regions were expanded from 2D ^{1}H-^{15}N TROSY (A, B, C) and 2D ^{1}H-^{13}C HSQC (D, E, F) spectra of SipD with Compound **1** (A, D), Compound **2** (B, E) and Compound **3** (C, F). The direction of the change in the NMR peak positions is indicated by arrowheads. Peaks are colored black, red, blue, green and magenta according to the increasing molar ratios [15] of SipD with each compound.

For titration with all the compounds, residues in other regions - L179, V187 and K188 in the mixed αβ-domain for compound **1**; L48 and V95 in the α-helical hairpin for compound **2**; K204, G220 and A231 in the mixed αβ-domain for compound **3 (Figure 4-8** and **Figure 4-9)** were strongly affected. But these interactions are probably transient and non-specific. For titrations with compound **1** and **3**, most of the affected residues were clustered in the proximal region of SipD **(Figure 4-8** and **Figure 4-9)** that binds to the needle protein PrgI [19]. For titrations with compound **2**, the mixed αβ-domain of SipD was the major interacting region, containing many of the affected residues **(Figure 4-8** and **Figure 4-9)**. The mixed αβ-domain is a part of the distal region of SipD, and was previously shown to interact with the N-terminal domain of the major translocon protein SipB [19].

The bottom of the coiled-coil, designated as pocket *x* in my study, was previously identified as a hotspot for binding drug-like fragments from the Zenobia library 2.0 [14] and bile salts [10]. Pockets *x* and *y* are separated by ~ 37 Å, leading to the future possibility of linking compounds **1**/**3** and **2**, which bind to these pockets. This can generate novel drug-like molecules which can bind optimally to SipD. The pocket *x* identified by NMR **(Figure 4-9)** was also predicted by the servers DoGSiteScorer [23] **(Figure 4-10A)** and PockDrug [24] **(Figure 4-10B)**.

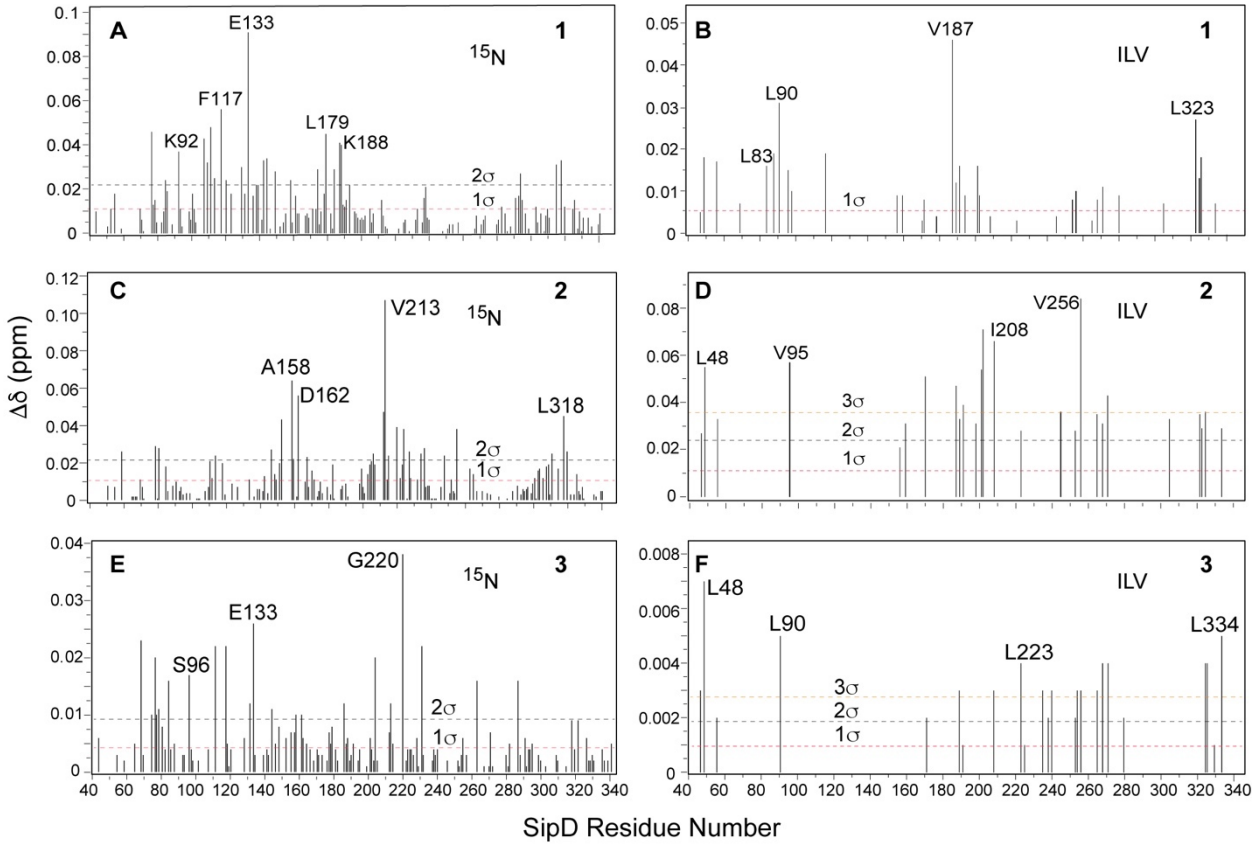

Figure 4-8. Plots of weighted chemical shift deviation of SipD residues. The weighted chemical shift deviation ($\Delta\delta$) of SipD residues upon titration with Compound **1** (A, B), Compound **2** (C, D) and Compound **3** (E, F). Data from the ^{15}N titrations are shown on the left panels (A, C, E) and data from the ILV titrations are shown on the right panels (B, D, F).

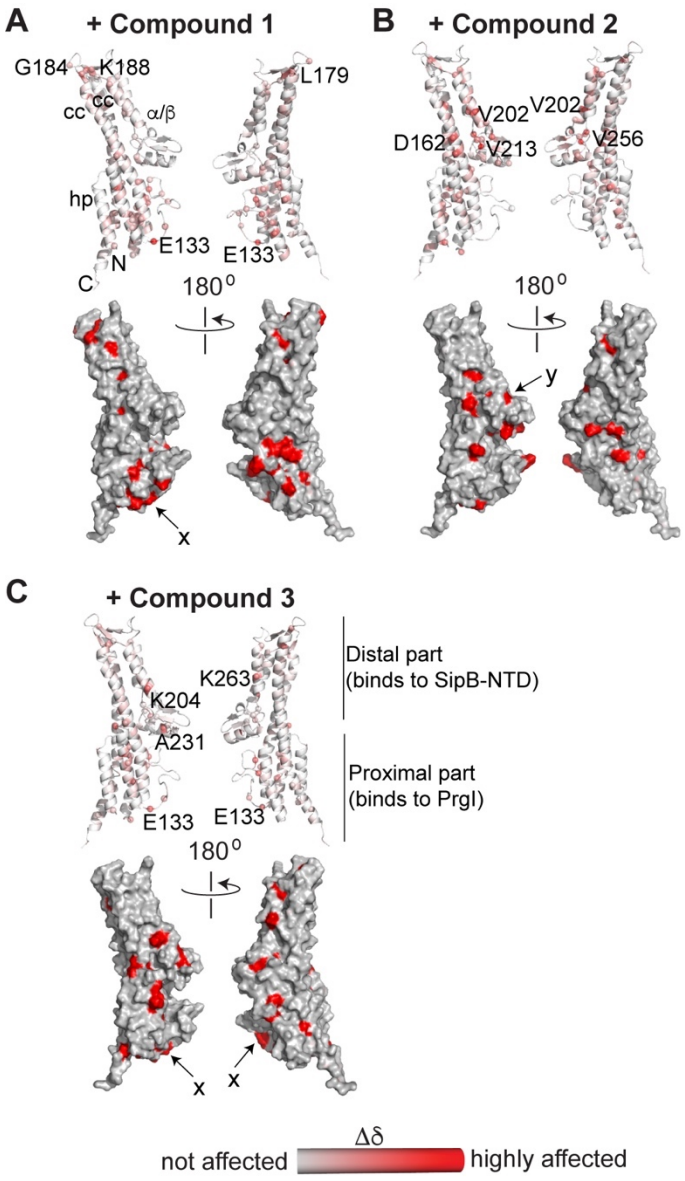

Figure 4-9. Cartoon and surface representations of affected SipD regions. The weighted chemical shift deviation of (A) Compound **1**, (B) Compound **2** and (C) Compound **3** are shown on the cartoon and surface structures of SipD. The residues are colored according to the value of ($\Delta\delta$), with the least affected residues colored gray and the most strongly affected residues colored red. Compounds **1** and **3** bind to pocket *x*, while Compound **2** binds to pocket *y*. The different parts of SipD are labeled as the amino *(N)* and carboxy *(C)* termini, hairpin *(hp)*, coiled-coils *(cc)* and the mixed $\alpha\beta$-domain *($\alpha\beta$)*.

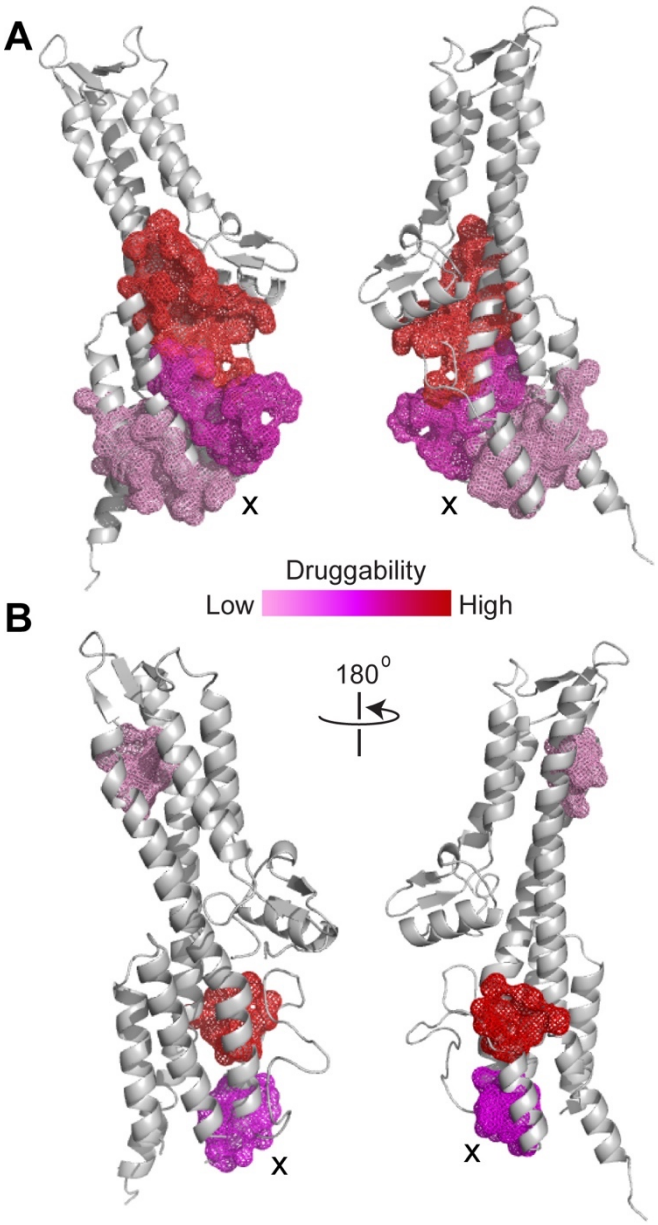

Figure 4-10. The druggable sites on SipD were predicted by (A, B) DoGSiteScorer [23] and (C, D) PockDrug [24] servers. The predicted sites are shown as mesh and colored according to their probability of druggability. The binding pocket *x*, identified by NMR, was also predicted by the servers.

4.5. Discussion

A major global health concern is the rise in antibiotic resistance combined with the low number of new antibiotics in the pipeline. This makes the development of new antimicrobials necessary. The tip protein of the T3SS is an excellent target for the development of novel anti-infectives because it is exposed on the bacterial cell surface and plays an essential role in virulence [15]. The only small molecules known to interact with the tip proteins include the bile salts, such as deoxycholate [9, 10] to SipD and its *Shigella* homologue IpaD, and fragment-based small molecules screened from the Zenobia library 2.0 for binding to SipD [14] and IpaD [15]. Here three novel small molecules based on quinoline and phenyl scaffolds were identified to bind to SipD, thus increasing the number of known scaffolds that can bind to SipD. Two different libraries containing 134 and 489 small molecules were screened using SPR for binding to SipD **(Figure 4-2)**. All the compounds of the two libraries can potentially serve as leads for the future development of novel anti-infectives, because their properties conformed to the "Rule of Three" **(Table 4-1)** [16]. I validated the SPR results using NMR to characterize the interactions of the three identified small molecules **(Figure 4-1)** to SipD (**Figures 4-3, 4-4, 4-5, 4-6, 4-7, 4-8**). My results have identified two possible binding pockets, *x* and *y*, for binding the molecules **(Figure 4-9)**.

My NMR data identified two druggable hotspots, designated as pockets *x* and *y*, for binding to quinoline (Compound **1**) and phenyl (Compound **2** and Compound **3**) scaffolds. These scaffolds have not been identified previously to bind to SipD. It was previously showed [15] that a quinoline-based scaffold interacts with the *Shigella* homologue IpaD. But the scaffold showed binding at the interface between the central coiled-coil and the mixed $\alpha\beta$-domain at a pocket different from my identified pocket *x* [15]. Thus, even though SipD and IpaD have structural

homology [8, 9, 29] and share 38% sequence identity with 56% similarity, they differ in their ability to bind specific chemical scaffolds. My NMR data showed that the most strongly affected residues are clustered in the proximal region of SipD **(Figure 4-8** and **Figure 4-9)** for Compound **1** and Compound **3**, as well as in the distal region of SipD involving the mixed $\alpha\beta$-domain **(Figure 4-8** and **Figure 4-9)** (for Compound **2**). The proximal part of SipD was previously shown to interact with the needle protein PrgI [19, 28] and the distal part was found to interact with the N-terminal domain of the major translocon protein SipB [19]. This presents the possibility that the bound compounds can disrupt these interactions. The presence of additional interaction surfaces on SipD **(Figure 4-9)** for the small molecules is possibly due to their weak binding affinities (in the millimolar range), making them non-selective for one specific surface.

Previous studies [30] have shown that modifying the piperazine moiety in the 8-hydroxyquinoline derivative INP1750 improved the inhibition of the T3SS of *Yersinia pseudotuberculosis*. INP1750 was reported to inhibit the activity of the YscN ATPase, flagellar motility and toxin secretion in *Yersinia pseudotuberculosis* [31]. Further experiments should be carried out to investigate whether Compound **1**, Compound **2** and Compound **3** can block the SipD-PrgI and SipD-SipB NTD interactions and inhibit the activity of the T3SS without killing the bacteria. Further chemical modifications of the molecules such as linking and introducing chirality can also be carried out. This will lead to the generation of novel drug candidates that can reduce the problem of antimicrobial resistance.

4.6. References

1. Majowicz, S.E., et al., *The global burden of nontyphoidal Salmonella gastroenteritis.* Clin. Infect. Dis., 2010. **50**(6): p. 882-9.
2. Wellington, E.M., et al., *The role of the natural environment in the emergence of antibiotic resistance in gram-negative bacteria.* Lancet Infect Dis, 2013. **13**(2): p. 155-65.
3. Davies, J. and D. Davies, *Origins and evolution of antibiotic resistance.* Microbiol Mol Biol Rev, 2010. **74**(3): p. 417-33.

4. Boucher, H.W., et al., *Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America.* Clin. Infect. Dis., 2009. **48**(1): p. 1-12.
5. Spellberg, B., et al., *The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America.* Clin. Infect. Dis., 2008. **46**(2): p. 155-64.
6. McShan, A.C. and R.N. De Guzman, *The bacterial type III secretion system as a target for developing new antibiotics.* Chem. Biol. Drug. Des., 2015. **85**(1): p. 30-42.
7. Lara-Tejero, M. and J.E. Galan, *Salmonella enterica serovar typhimurium pathogenicity island 1-encoded type III secretion system translocases mediate intimate attachment to nonphagocytic cells.* Infect. Immun., 2009. **77**(7): p. 2635-42.
8. Chatterjee, S., et al., *The crystal structure of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate.* Protein Sci., 2011. **20**: p. 75-86.
9. Lunelli, M., et al., *Crystal structure of PrgI-SipD: insight into a secretion competent state of the type three secretion system needle tip and its interaction with host ligands.* PLoS Pathog., 2011. **7**(8): p. e1002163.
10. Chatterjee, S., et al., *The crystal structures of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate.* Protein Sci, 2011. **20**(1): p. 75-86.
11. Wang, Y., et al., *NMR characterization of the interaction of the Salmonella type III secretion system protein SipD and bile salts.* Biochemistry, 2010. **49**(19): p. 4220-6.
12. Barta, M.L., et al., *Identification of the bile salt binding site on IpaD from Shigella flexneri and the influence of ligand binding on IpaD structure.* Proteins, 2012. **80**(3): p. 935-45.
13. Dickenson, N.E., et al., *Conformational Changes in IpaD from Shigella flexneri upon Binding Bile Salts Provide Insight into the Second Step of Type III Secretion.* Biochemistry, 2011. **50**: p. 172-180.
14. McShan, A.C., et al., *Characterization of the Binding of Hydroxyindole, Indoleacetic acid, and Morpholinoaniline to the Salmonella Type III Secretion System Proteins SipD and SipB.* ChemMedChem, 2016. **11**(9): p. 963-971.
15. Dey, S., et al., *Characterization of Small-Molecule Scaffolds That Bind to the Shigella Type III Secretion System Protein IpaD.* ChemMedChem, 2017. **12**(18): p. 1534-1541.
16. Congreve, M., et al., *A 'rule of three' for fragment-based lead discovery?* Drug Discov Today, 2003. **8**(19): p. 876-7.
17. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J. Biomol. NMR, 1995. **6**(3): p. 277-293.
18. Johnson, B.A., *Using NMRView to visualize and analyze the NMR spectra of macromolecules.* Methods Mol. Biol., 2004. **278**: p. 313-352.
19. McShan, A.C., et al., *NMR Identification of the Binding Surfaces Involved in the Salmonella and Shigella Type III Secretion Tip-Translocon Protein-Protein Interactions.* Proteins, 2016. **84**: p. 1097-1107.
20. Grzesiek, S., et al., *The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR.* Biochemistry, 1996. **35**(32): p. 10256-61.
21. Mayer, M. and B. Meyer, *Characterization of ligand binding by saturation transfer difference NMR spectroscopy.* Angew. Chem. Int. Ed., 1999. **38**(12): p. 1784-1788.
22. Bernstein, M.A., et al., *Optimization and automation of quantitative NMR data extraction.* Anal Chem, 2013. **85**(12): p. 5778-86.
23. Volkamer, A., et al., *Combining global and local measures for structure-based druggability predictions.* J Chem Inf Model, 2012. **52**(2): p. 360-72.
24. Hussein, H.A., et al., *PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins.* Nucleic Acids Res, 2015. **43**(W1): p. W436-42.

25. Testa, B., et al., *Lipophilicity in molecular modeling.* Pharm Res, 1996. **13**(3): p. 335-43.
26. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Adv Drug Deliv Rev, 2001. **46**(1-3): p. 3-26.
27. Abraham, M.H., et al., *Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the delta log P parameter of seiler.* J Pharm Sci, 1994. **83**(8): p. 1085-100.
28. Rathinavelan, T., et al., *NMR model of PrgI-SipD interaction and its implications in the needle-tip assembly of the Salmonella type III secretion system.* J. Mol. Biol., 2014. **426**(16): p. 2958-69.
29. Johnson, S., et al., *Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD.* J. Biol. Chem., 2007. **282**(6): p. 4035-4044.
30. Enquist, P.A., et al., *Derivatives of 8-hydroxyquinoline-antibacterial agents that target intra- and extracellular Gram-negative pathogens.* Bioorg. Med. Chem. Lett., 2012. **22**(10): p. 3550-3553.
31. Anantharajah, A., et al., *Salicylidene Acylhydrazides and Hydroxyquinolines Act as Inhibitors of Type Three Secretion Systems in Pseudomonas aeruginosa by Distinct Mechanisms.* Antimicrob Agents Chemother, 2017. **61**(6).

Chapter 5: Characterization of Novel Compounds Identified Using Computational Screening For Binding to the *Salmonella* Tip Protein SipD

5.1. Abstract

In the previous chapter, I reported the SPR-based screening and characterization of three novel small molecules based on quinoline and phenyl scaffolds for binding to the *Salmonella* tip protein SipD. The computational screening of compound libraries for binding to tip proteins is currently unknown. Here, through a collaboration, I present the results of the computational screening of a compound library containing ~ 8.33 million compounds using Rosetta and ROCS/FastROCS for binding to SipD. The compounds were screened on the basis of their similarity in shape and chemical features with pocket-containing conformations of SipD. The screening predicted the binding of three compounds to SipD, based on piperidine, piperazine and phenyl scaffolds. I also used NMR to identify the residues and surfaces of SipD that interact with these compounds. This knowledge further increases the known number of fragments that can bind to SipD, and can be used to screen other virulence proteins from the T3SS. This knowledge will help in designing novel therapeutics to combat the growing threat posed by multidrug resistance in bacteria.

5.2. Introduction

SipD is the tip protein in *Salmonella*. Computational screening of compounds for binding to the tip proteins has not been reported previously. Here I screened a library containing ~ 8.33 million compounds using Rosetta and ROCS/FastROCS [1-6]. This approach involved predicting low-energy pocket-containing conformations in SipD by biased simulations, followed by using 'exemplars' derived from these conformations as templates to screen the compound library [1]. The technique is based on matching the overall shape and chemical features between SipD and

the compounds [1] and is much faster over traditional docking approaches [1]. Rosetta and ROCS/FastROCS predicted three compounds that can bind to SipD, based on piperidine, piperazine and phenyl scaffolds. I used NMR to determine the residues and surfaces of SipD that interact with these small molecules and to identify potential druggable hotspots in SipD. The findings can be used in future to design new antimicrobials against the multidrug resistant bacteria.

5.3. Methods

5.3.1. Computational screening of compound library

The computational screening of ~8.33 million drug-like compounds belonging to the ZINC database [7] was done as described previously [1] in collaboration with Dr. David Johnson (Director, COBRE CBID Computational Chemical Biology Core Facility, University of Kansas). Briefly, the *Salmonella* tip protein SipD (PDB ID: 3NZZ) [8], lacking a 23-residue loop spanning residues 110-132, was docked to trimers of the *Salmonella* needle protein PrgI. Docking to the entire needle ring was computationally unfeasible. Techniques of conformational sampling of proteins were employed, leading to the generation of exemplars, using the Rosetta suite [1, 9]. Ensembles of pocket-containing conformations were generated by a biased [2] Monte Carlo sampling method in the Rosetta suite [9], taking into account both backbone and sidechain moves. This methodology favoured the conformations of SipD containing a surface pocket, simultaneously decreasing the energetic cost to open other pockets [2, 3]. An ensemble of exemplars was generated from each pocket by identifying the shape of the pocket and the location of ideal hydrogen bond donors and acceptors as described previously [3].

The compounds of the ZINC database [7] were used as decoys in an ultra-high-throughput virtual screening experiment using the exemplars. The compounds were ranked by traditional virtual screening and Rosetta-optimization of shape-matched pose. The compounds that had shape-complementarity with the pockets were ranked higher in induced-fit docking. Virtual screening was performed using the ROCS [4, 6] and FastROCS [4, 5] softwares. Steric clashes were not taken into account for Rosetta optimization. Further refinement of the structure is necessary in order to accommodate clashes.

5.3.2. Expression and purification of SipD

SipD was overexpressed and purified as described in the previous chapter.

5.3.3. NMR spectroscopy

NMR data were acquired as described in the previous chapter, using identical acquisition parameters.

5.3.4. NMR titrations

The compounds **1** (Molport # 009-563-322), **2** (Molport # 009-420-836) and **3** (Molport # 016-589-411) were dissolved in ~250 uL of 100% [D_6] DMSO (Cambridge Isotope Laboratories, Inc., Andover, MA) to make stock solutions of 2 M. The compounds were titrated into 0.3 mM ^{15}N/ILV-labeled SipD at various molar ratios of SipD to compounds. Five titration points were obtained for each of the three compounds. All the NMR titration samples contained final [D_6] DMSO concentration of 1.5%. The previous ^{15}N and ILV NMR assignments of SipD [10, 11] were

used to identify the residues of SipD that were affected upon binding to the compounds. The weighted chemical shift deviation ($\Delta\delta$) were calculated using the equation $\Delta\delta_{HN} = ½ [(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2]$ [12] for the backbone amides and $\Delta\delta_{ILV} = ½ [(\Delta\delta_H)^2 + (\Delta\delta_C/2)^2]$ for the ILV spectra [13]. The weighted chemical shift deviations of selected SipD residues were plotted vs. compound concentrations to estimate the dissociation constants using GraphPad Prism Version 9 (GraphPad Software, Inc., USA).

5.3.5. Saturation Transfer Difference (STD) NMR

STD NMR data were acquired using acquisition parameters identical to those described in the previous chapter.

5.4. Results

5.4.1. Computational screening of ZINC database

The computational screening of the ZINC database [7] identified three compounds that can bind to SipD. These compounds were based on **piperidine** [(1-{1-[4-amino-6-(phenylamino)-1,3,5-triazin-2-yl]-ethyl}piperidin-4-yl)methanol, ZINC52819972, Compound **1**], **piperazine** [2-(4-{1-[4-amino-6-(phenylamino)-1,3,5-triazin-2-yl]ethyl}piperazin-1-yl)ethan-1-ol, ZINC32850916, Compound **2**] and **phenyl** [3-(5-chloro-2,4-dimethoxyphenyl)-1-{4-[(6-methylpyridazin-3-yl)amino]phenyl}urea, ZINC65363539, Compound **3**] **(Figure 5-1)** scaffolds. All the compounds were modeled to bind to pocket *x* **(Figure 5-2)**, at the interface between a helix of the central coiled-coil and a loop in the mixed $\alpha\beta$-domain of SipD [11]. Compounds **1** and **2** are very similar

in structure, and provide interactions bridging the helix and the loop. The extra nitrogen in Compound **2** changes the orientation of the ethanol substituent, providing additional interactions between the helix and the loop. However, its docking and complex energies were lower than that of compound **1**.

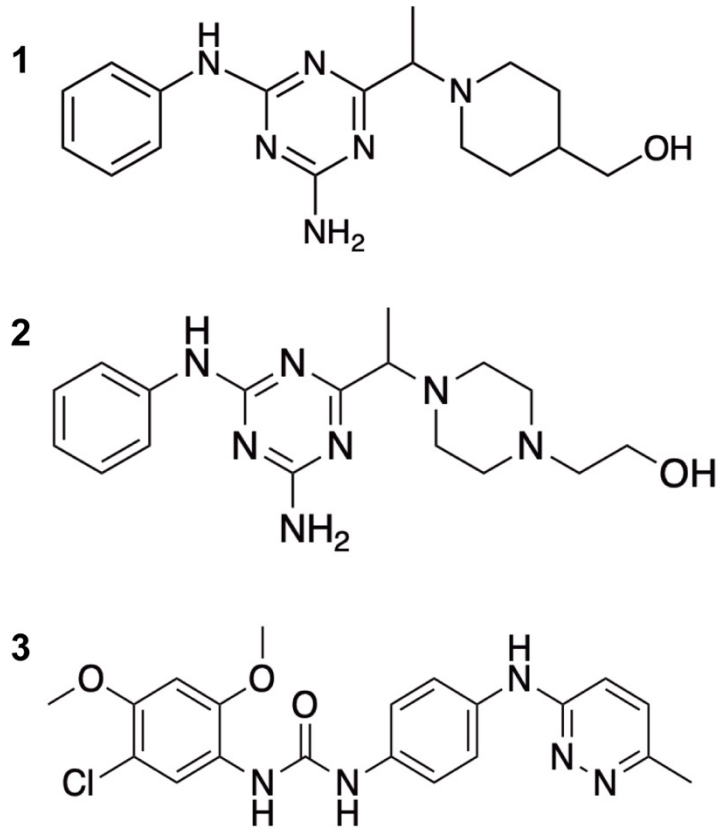

Figure 5-1. Structures of compounds identified by computational screening. The three scaffolds predicted to bind to SipD were identified by induced fit docking (Compounds **1** and **2**) and Rosetta optimization (Compound **3**). The compounds are based on piperidine (**1**), piperazine (**2**) and phenyl **(3)** scaffolds.

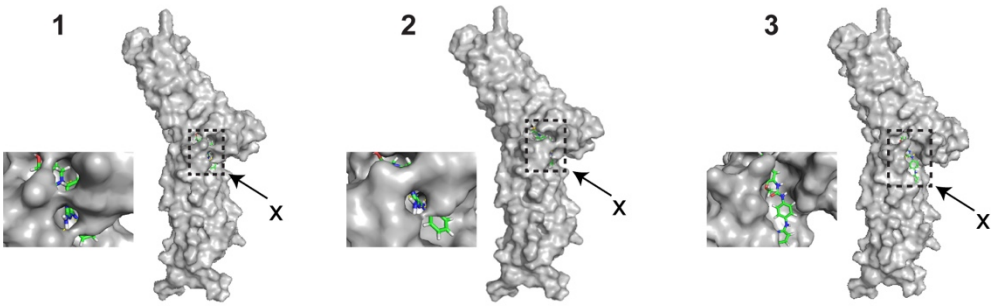

Figure 5-2. Pockets on SipD identified by computational screening. Compounds **1**, **2** and **3** were all identified by computational screening to bind to SipD at pocket x.

5.4.2. STD NMR

Saturation transfer difference (STD) NMR [14] was used to identify the atoms and parts of the compounds interacting with SipD [18]. In STD NMR, two spectra are recorded by selectively saturating the proton resonances of the protein with a Gaussian pulse that is turned on or off [18]. If a compound binds to the protein, then magnetization is transferred from the protons of the protein to the protons of the compounds, and detected in the 'on-resonance' spectrum. The 'off-resonance' spectra (top panels, **Figure 5-3**) contain peaks from the protein and the small molecules. The STD spectra (bottom panels, **Figure 5-3**), obtained by subtracting the 'on-resonance spectra' (data not shown) from the off-resonance spectra [18], show peaks from only the protons of the compounds that contact the protein. Some protons may not interact with the protein, and their peaks are absent from the STD spectra. The STD NMR data indicate that for compounds **1 (Figure 5-3A)** and **2 (Figure 5-3B)**, all the atoms of the phenyl ring, marked as p, q/r and s/t, are in contact with SipD. The data further suggest that the phenyl rings of these compounds are embedded in a pocket on SipD. Compound **3** did not show binding **(Figure 5-3C)** to SipD. This is possibly because of steric clashes between the compound and SipD. The Rosetta optimization protocol that was employed for compound **3** did not take clashes into account. Further refinement of the structure is necessary to accommodate steric clashes.

Figure 5-3. STD NMR of SipD with the compounds. STD NMR data are shown for Compounds (A) 1, (B) 2 and (C) 3. Top panels show off-resonance spectra and the assignment of the protons of each compound. Lower panels show the STD spectra, obtained by subtracting the on-resonance spectra (not shown) from the corresponding off-resonance spectra [18]. The STD spectra show the protons of the compounds in contact with SipD.

5.4.3. NMR titrations of ^{15}N/ILV-SipD with the compounds

NMR spectroscopy was used to identify the residues and surfaces of SipD interacting with the compounds. ^{15}N/ILV-labeled SipD was titrated with increasing concentrations of compounds **1**, **2** and **3**, and 2D ^{1}H-^{15}N TROSY and 2D ^{1}H-^{13}C HSQC spectra were recorded. The previous ^{15}N and ILV NMR assignments of SipD [10, 11] were used to identify the residues of SipD that were affected upon binding the compounds. Compound **3** did not show binding to SipD in the NMR titrations **(Figure 5-4)**, complementing the STD NMR data **(Figure 5-3C)**.

Titration of SipD with compounds **1** and **2** showed concentration-dependent chemical shift perturbation of specific ^{15}N-amide **(Figure 5-5A, Figure 5-66A)** and ILV peaks **(Figure 5-5D, Figure 5-6D)**. This indicates that both the compounds bind to SipD, showing interaction in the fast exchange regime in the NMR timescale **(Figures 5-5B, 5-5E, 5-6B and 5-6E)**. The dissociation constants (K_d) were calculated to be 1.5 mM (for compound **1**) and ~ 1 mM (for compound **2**) **(Figure 5-5C, Figure 5-6C)**. The lower K_d value for compound **2** supports the finding from induced-fit docking. The extra nitrogen in the ring in compound **2** possibly favours the interaction by changing the orientation of the ethanol substituent. The weighted chemical shift deviations were plotted for both the compounds **(Figure 5-7)** to identify the strongly affected ^{15}N and ILV peaks. The chemical shift deviations were mapped onto the structure of SipD [15], showing the surfaces and binding pockets for the compounds on SipD **(Figure 5-8)**.

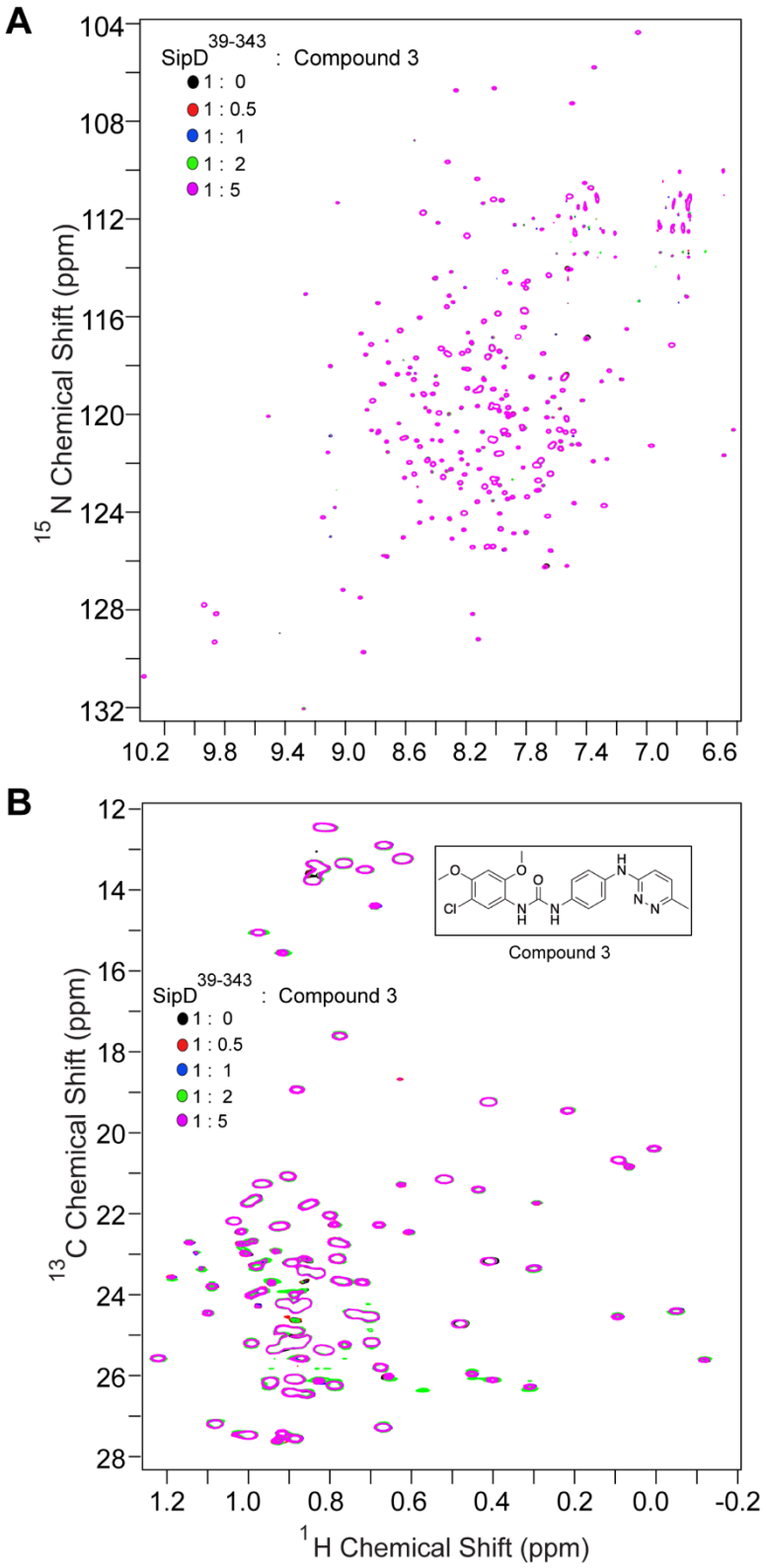

Figure 5-4. NMR titrations of SipD with Compound 3. (A) ^{15}N and (B) ILV titrations of SipD with Compound **3** show that SipD does not bind to Compound **3**.

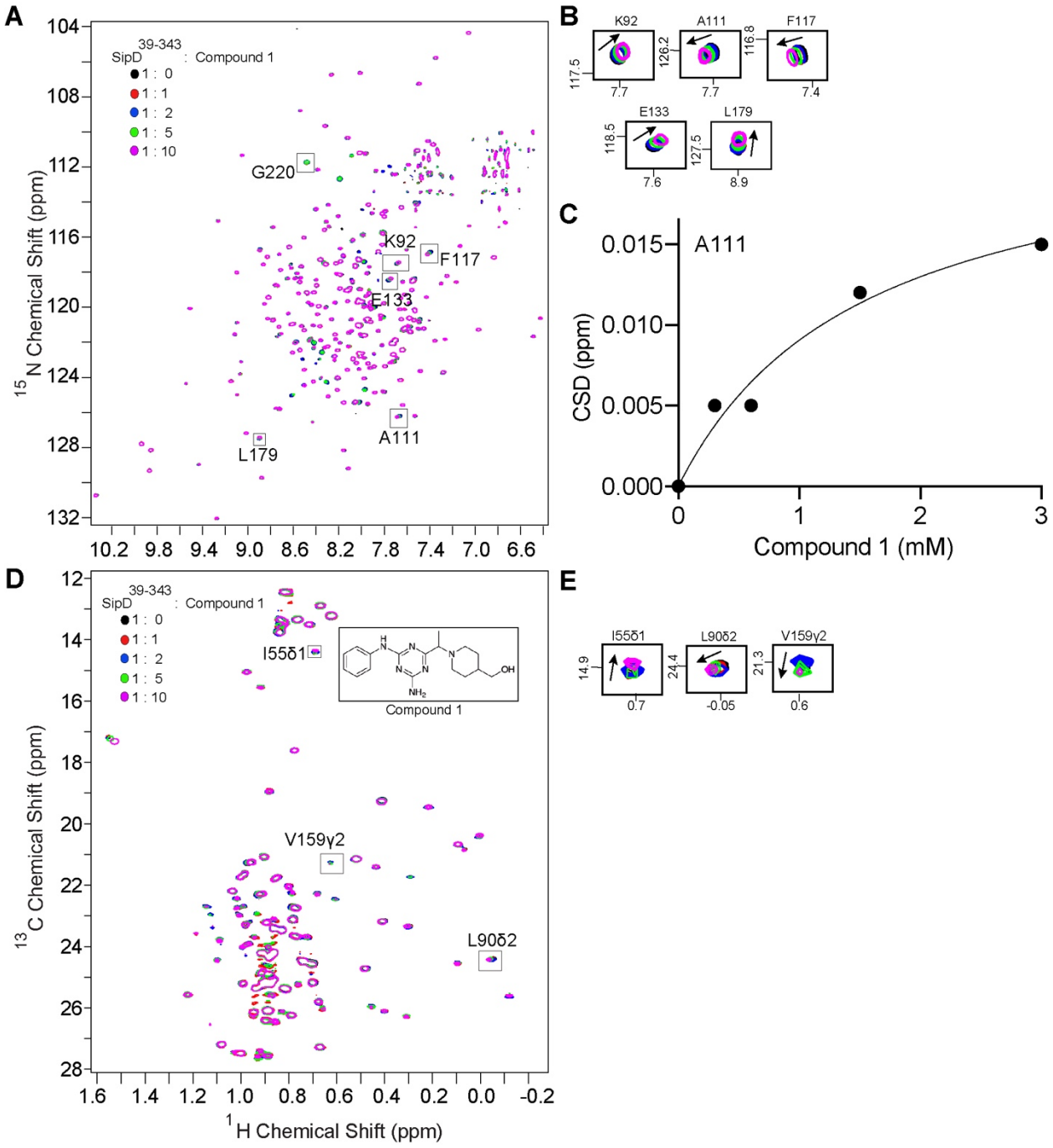

Figure 5-5. NMR titrations of SipD with Compound 1. (A) ^{15}N titration of SipD with Compound **1**. (B) Selected peaks from the 2D ^{1}H-^{15}N TROSY spectrum of SipD titrated with Compound **1**. (C) Plot of the weighted chemical shift deviation (CSD) vs. concentration of Compound **1** for A111. (D) ILV titration of SipD with Compound **1**. (E) Selected peaks from the 2D ^{1}H-^{13}C HSQC spectrum of SipD titrated with Compound **1**.

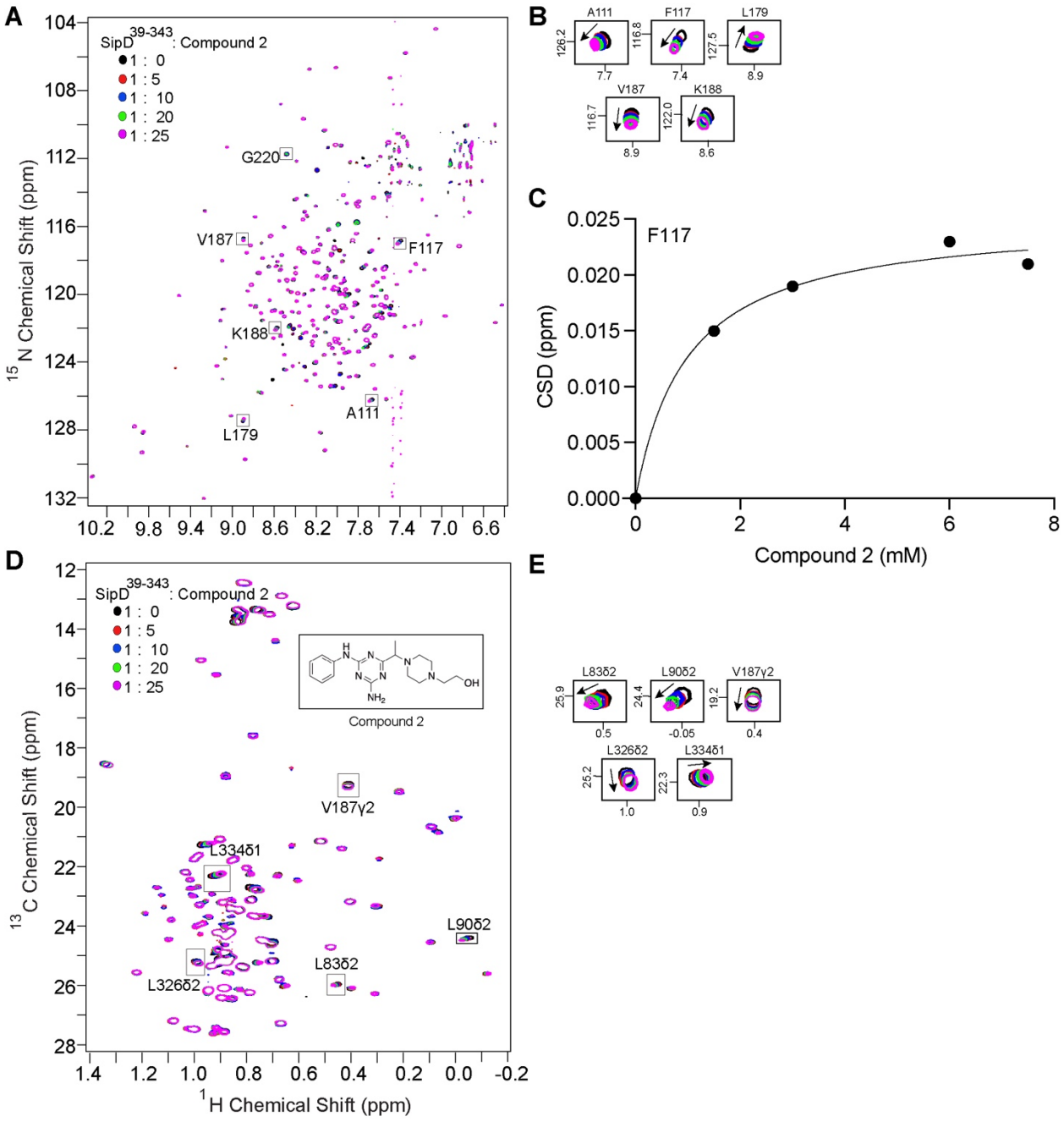

Figure 5-6. NMR titrations of SipD with Compound 2. (A) ^{15}N titration of SipD with compound **2**. (B) Selected peaks from the 2D ^{1}H-^{15}N TROSY spectrum of SipD titrated with Compound **2**. (C) Plot of the weighted chemical shift deviation (CSD) vs. concentration of Compound **2** for F117. (D) ILV titration of SipD with Compound **2**. (E) Selected peaks from the 2D ^{1}H-^{13}C HSQC spectrum of SipD titrated with Compound **2**.

The most strongly affected ^{15}N residues in the titration with Compound **1** (F117, A111 and E133) are located in the 23-residue loop connecting the N-terminal α-helical hairpin to the central coiled-coil. Many affected residues are clustered in the α-helical hairpin – E71 and K92 (in the ^{15}N titration), while I55 and L90 in the ILV titration **(Figures 5-7A, 5-7B)**. The residues L90, K92 and E133 cluster on the same surface of SipD, forming pocket *y* **(Figure 5-8A)**. Similarly, compound **2**, having a very similar structure to compound **1**, also binds to pocket *y* **(Figure 5-8B)**. For titration with compound **2**, residue E133 was the most strongly affected ^{15}N peak. The α-helical hairpin contained many strongly affected residues – L83, L90 (in the ILV spectrum) and L87, L90, K92, L97, A111 and F117 (in the ^{15}N spectrum) **(Figures 5-7C, 5-7D)**. Thus, the results of the ILV titration complemented the results from the ^{15}N titration. For titration with both the compounds, residues in other regions (L179, V187 and K188 in the mixed $\alpha\beta$-domain) **(Figure 5-7)** were strongly affected. But these interactions are probably transient and non-specific. Most of the affected residues are located in the proximal region of SipD that interacts with the needle protein PrgI [11]. The NMR data collectively suggest that SipD uses predominantly hydrophobic interactions as well as some polar contacts mediated by electrostatic interactions to interact with the compounds.

Some SipD residues in pocket *x* such as D162 for compound **1** and E70, D74 and D162 for compound **2**, were predicted by induced-fit docking to be involved in the interaction. But these peaks showed ~ 100 times lesser chemical shift perturbations in NMR titrations than the strongest ones **(Figures 5-7A, 5-7C)**. The residue G220 located in pocket *x* was modeled by induced fit docking to bind to both compounds **1** and **2**. This residue showed concentration-dependent intensity decrease in the ^{15}N-titrations with both the compounds **(Figures 5-5A, 5-6A)**, indicating interaction in the intermediate exchange regime in the NMR timescale with micromolar dissociation constants, suggesting that the binding affinities of both the compounds to pocket *x*

are in general lower than those to pocket *y* identified in the NMR titrations. The pocket *y* is possibly inaccessible to the compounds in the beginning. After the initial binding to pocket *x* using residue G220, conformational changes in SipD might open pocket *y*, allowing the subsequent binding of the compounds. Other interacting residues predicted by induced-fit docking, such as N211 and L219, are located in pocket *x*, but were assigned ambiguously in NMR [10, 11] to prevent their analysis in the titrations. Thus, I have identified a potential hotspot (pocket *y*) **(Figure 5-8)** for piperidine (compound **1**) and piperazine (compound **2**)-based scaffolds at the bottom of the coiled-coil. This region was previously identified as a hotspot for binding drug-like fragments from the Zenobia library 2.0 [13] and bile salts [16].

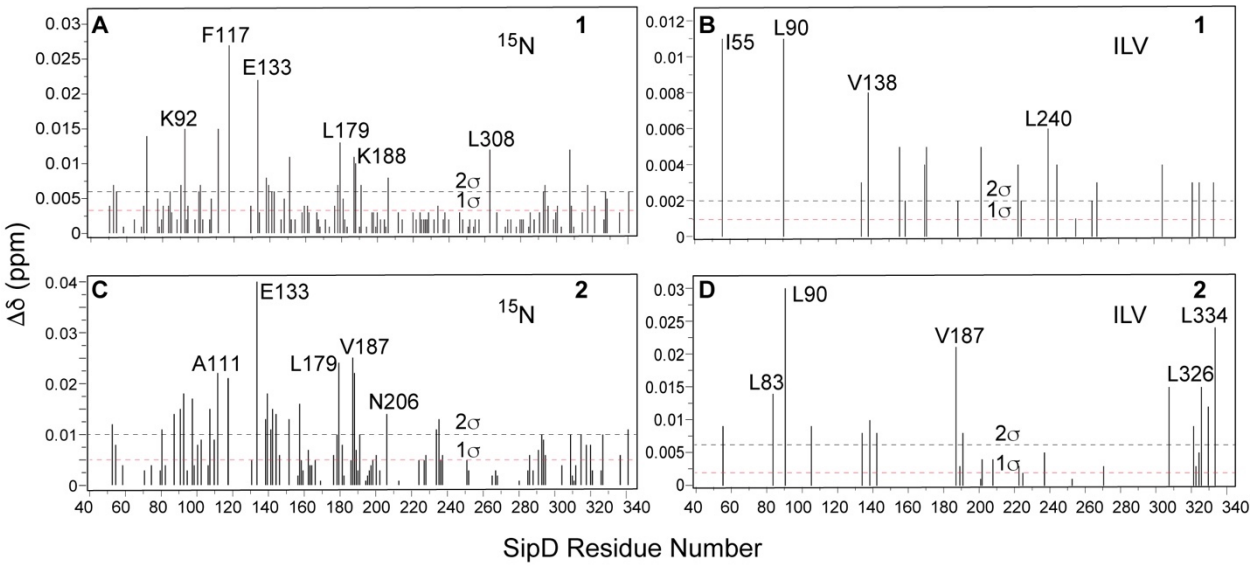

Figure 5-7. Plots of the weighted chemical shift deviation of SipD. The weighted chemical shift deviations ($\Delta\delta$) were plotted for Compounds (A, B) **1** and (C, D) **2**. The plots show the SipD peaks that are strongly affected relative to the others. The data from the ^{15}N titrations are shown on the left panels (A, C), and the results from the ILV titrations are on the right panels (B, D).

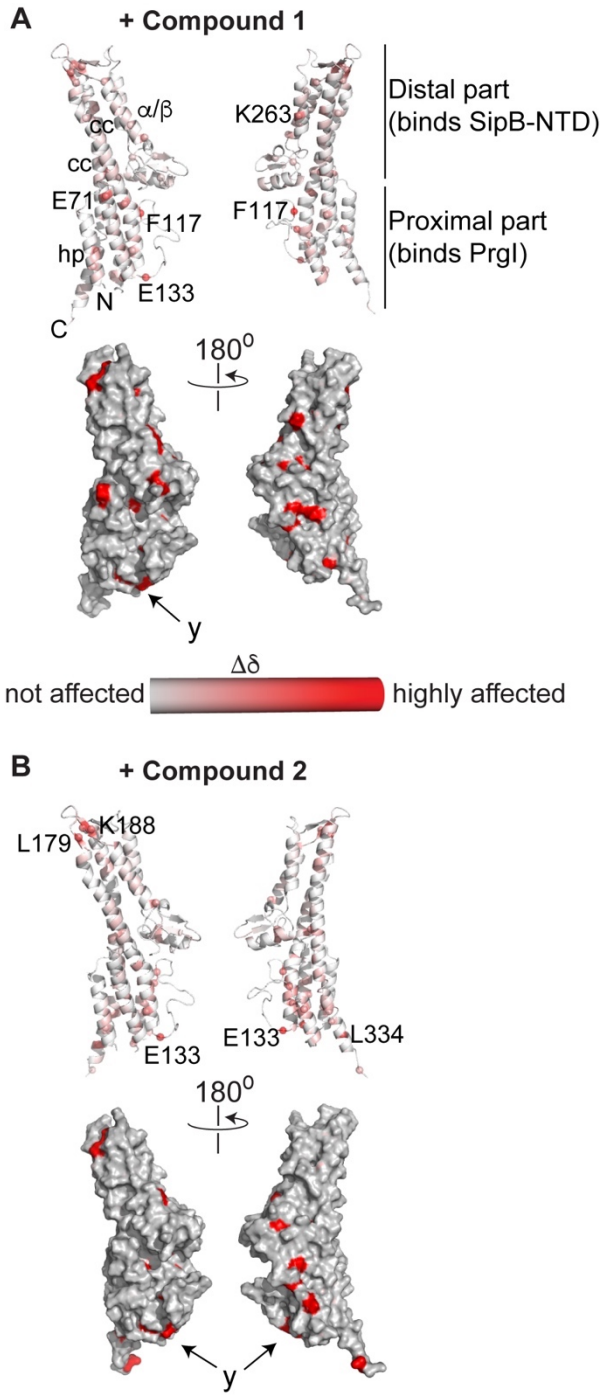

Figure 5-8. Cartoon and surface representations of affected SipD regions. The weighted chemical shift deviation ($\Delta\delta$) of Compounds (A) **1** and (B) **2** are shown on the cartoon and surface representations of SipD. The residues are colored according to the value of $\Delta\delta$, with the least affected residues colored gray, and the most strongly affected residues colored red. Compounds **1** and **2** both bind to the pocket *y*. The different parts of SipD are labeled as the amino *(N)* and carboxy *(C)* termini, hairpin *(hp)*, coiled-coils *(cc)* and the mixed $\alpha\beta$-domain *($\alpha\beta$)*.

5.5. Discussion

A major global health concern is the rise in antibiotic resistance combined with the low number of new antibiotics in the pipeline. This makes the development of new antimicrobials necessary. The tip protein of the T3SS is an attractive target for developing new anti-infectives because it is exposed on the bacterial surface and plays an essential role in virulence. Currently, the only known chemical scaffolds that are known to bind to the tip proteins include the bile salts, such as deoxycholate [16, 17] to SipD and its *Shigella* homologue IpaD, and fragment-based small molecules screened from the Zenobia library 2.0 for binding to SipD [13] and IpaD [18]. Here computational screening identified two novel compounds based on piperidine and piperazine scaffolds for binding to SipD (**Figure 1** and **Figure 2**), thus increasing the number of known scaffolds that can bind to SipD.

The NMR data identified a druggable hotspot, designated as pocket *y*, for binding to piperidine (compound **1**) and piperazine (compound **2**) scaffolds. These scaffolds have not been identified previously to bind to SipD. In fact, it was previously showed [18] that piperidine and piperazine scaffolds did not bind to the *Shigella* homologue IpaD. Thus, even though SipD and IpaD have structural homology [8, 17, 19] and share 38% sequence identity with 56% similarity, they differ in their ability to bind specific chemical scaffolds. My NMR data showed that the most strongly affected residues are clustered in the proximal region of SipD **(Figure 8)**, the region previously shown to interact with the needle protein PrgI [11, 15]. This presents the possibility that the bound compounds can disrupt this interaction. The binding affinity of compound **2**, based on the piperazine scaffold, to this surface is ~ 1.5 times higher than that of compound **1 (Figures 5C, 6C)**, presumably because the extra nitrogen in the piperazine ring changes the orientation of the ethanol substituent to allow additional interactions. This finding supports the induced-fit

docking prediction. The presence of additional interaction surfaces on SipD **(Figure 8)** for the two compounds is possibly due to their weak binding affinities (in the millimolar range), making them non-selective for one specific surface.

Previous studies [20] have shown that modifying the piperazine moiety in the 8-hydroxyquinoline derivative INP1750 improved the inhibition of the T3SS of *Yersinia pseudotuberculosis*. INP1750 was reported to inhibit the activity of the YscN ATPase, flagellar motility and toxin secretion in *Yersinia pseudotuberculosis* [21]. My data present the first available report of the binding of piperidine and piperazine-based scaffolds to any tip protein of the T3SS, identified by computational screening. Further experiments should be carried out to investigate whether compounds **1** and **2** can block the SipD-PrgI interaction and inhibit the activity of the T3SS without killing the bacteria. This will lead to the generation of novel drug candidates that can reduce the problem of antimicrobial resistance.

5.6. References

1. Johnson, D.K. and J. Karanicolas, *Ultra-High-Throughput Structure-Based Virtual Screening for Small-Molecule Inhibitors of Protein-Protein Interactions.* J Chem Inf Model, 2016. **56**(2): p. 399-411.
2. Johnson, D.K. and J. Karanicolas, *Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface.* PLoS Comput Biol, 2013. **9**(3): p. e1002951.
3. Johnson, D.K. and J. Karanicolas, *Selectivity by small-molecule inhibitors of protein interactions can be driven by protein surface fluctuations.* PLoS Comput Biol, 2015. **11**(2): p. e1004081.
4. Hawkins, P.C., A.G. Skillman, and A. Nicholls, *Comparison of shape-matching and docking as virtual screening tools.* J Med Chem, 2007. **50**(1): p. 74-82.
5. Osguthorpe, D.J., W. Sherman, and A.T. Hagler, *Exploring protein flexibility: incorporating structural ensembles from crystal structures and simulation into virtual screening protocols.* J Phys Chem B, 2012. **116**(23): p. 6952-9.
6. Rush, T.S., 3rd, et al., *A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction.* J Med Chem, 2005. **48**(5): p. 1489-95.

7. Irwin, J.J., et al., *ZINC: a free tool to discover chemistry for biology.* J Chem Inf Model, 2012. **52**(7): p. 1757-68.
8. Chatterjee, S., et al., *The crystal structure of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate.* Protein Sci., 2011. **20**: p. 75-86.
9. Leaver-Fay, A., et al., *ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules.* Meth. Enzymol., 2011. **487**: p. 545-574.
10. Wang, Y., et al., *NMR characterization of the interaction of the Salmonella type III secretion system protein SipD and bile salts.* Biochemistry, 2010. **49**(19): p. 4220-6.
11. McShan, A.C., et al., *NMR Identification of the Binding Surfaces Involved in the Salmonella and Shigella Type III Secretion Tip-Translocon Protein-Protein Interactions.* Proteins, 2016. **84**: p. 1097-1107.
12. Grzesiek, S., et al., *The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR.* Biochemistry, 1996. **35**(32): p. 10256-61.
13. McShan, A.C., et al., *Characterization of the Binding of Hydroxyindole, Indoleacetic acid, and Morpholinoaniline to the Salmonella Type III Secretion System Proteins SipD and SipB.* ChemMedChem, 2016. **11**(9): p. 963-971.
14. Mayer, M. and B. Meyer, *Characterization of ligand binding by saturation transfer difference NMR spectroscopy.* Angew. Chem. Int. Ed., 1999. **38**(12): p. 1784-1788.
15. Rathinavelan, T., et al., *NMR model of PrgI-SipD interaction and its implications in the needle-tip assembly of the Salmonella type III secretion system.* J. Mol. Biol., 2014. **426**(16): p. 2958-69.
16. Chatterjee, S., et al., *The crystal structures of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate.* Protein Sci, 2011. **20**(1): p. 75-86.
17. Lunelli, M., et al., *Crystal structure of PrgI-SipD: insight into a secretion competent state of the type three secretion system needle tip and its interaction with host ligands.* PLoS Pathog., 2011. **7**(8): p. e1002163.
18. Dey, S., et al., *Characterization of Small-Molecule Scaffolds That Bind to the Shigella Type III Secretion System Protein IpaD.* ChemMedChem, 2017. **12**(18): p. 1534-1541.
19. Johnson, S., et al., *Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD.* J. Biol. Chem., 2007. **282**(6): p. 4035-4044.
20. Enquist, P.A., et al., *Derivatives of 8-hydroxyquinoline-antibacterial agents that target intra- and extracellular Gram-negative pathogens.* Bioorg. Med. Chem. Lett., 2012. **22**(10): p. 3550-3553.
21. Anantharajah, A., et al., *Salicylidene Acylhydrazides and Hydroxyquinolines Act as Inhibitors of Type Three Secretion Systems in Pseudomonas aeruginosa by Distinct Mechanisms.* Antimicrob Agents Chemother, 2017. **61**(6).

Chapter 6: Concluding Remarks and Future Directions

6.1. Overview

The focus of my dissertation was to understand the biophysical properties and protein-protein interactions of the *Shigella* minor translocon protein IpaC and characterizing novel compounds that bind to the *Salmonella* tip protein SipD. IpaC makes a pore in the host cell membrane [1] and serves as an effector inside the host cells by interacting with intermediate filaments [1-5]. SipD is exposed on the bacterial cell surface and provides a platform for the assembly of the translocon [6]. The functions of IpaC and SipD make them essential for virulence. My findings advance our understanding of the assembly of the Type III Secretion System (T3SS) and increase the repertoire of known chemical scaffolds that can target the tip protein. This knowledge will be useful for the future design of next-generation anti-infectives targeting the essential protein components of the T3SS.

6.2. Biophysical characterization of IpaC and its domains

6.2.1. Key findings

The atomic structure of any minor translocon protein is unknown. The biophysical studies of IpaC and its domains are detailed in Chapter 2. I reported that full-length IpaC contains α-helical regions in solution, but the protein overall lacks a tertiary structure. My ^{1}H-^{15}N TROSY spectra of IpaC and its domains are the first available NMR spectra of any minor translocon protein, and suggest that other proteins belonging to the IpaC family are intrinsically disordered in solution. Far-UV CD and NMR data suggest that IpaC may exist as a molten globule similar to the *Pseudomonas* major translocon protein PopB [7] and minor translocon protein PopD [8].

Further, CD and NMR spectroscopy showed that IpaC and its domains undergo conformational changes in micelles formed by the detergents n-dodecylphosphocholine (DPC), lyso-myristoylphosphatidylglycerol (LMPG) and sodium dodecyl sulfate (SDS). The results presented in this dissertation have identified DPC as a good micelle system for further biophysical characterization of IpaC.

6.2.2. Future directions to characterize IpaC

The far-UV CD and NMR data reported in Chapter 2 were recorded at 20^0C. In *Shigella* and *Salmonella*, the tip proteins bind to bile salts, which serve as environmental sensors [9-14]. Host cell contact, a shift in temperature to 37^0C and change in oxygen tension, among other factors, can activate the T3SS and assemble the translocon [15] (**Figure 6-1**, adapted from Dey *et al* [16]. Far-UV CD and NMR spectra of IpaC should be recorded at 37^0C to better understand the changes in its secondary and tertiary structures under conditions of infection. Fluorescence studies using 8-anilino-1-naphthalenesulfonic acid (ANS) and IpaC at 20^0C and 37^0C can test my proposition that IpaC exists as a molten globule. IpaC has previously been shown to interact with membrane mimetics such as liposomes [17] and phospholipids [18]. Titration of IpaC with phospholipids such as dipalmitoylphosphatidylcholine (DPPC) or liposomes such as 1,2-dioleyl-sn-glycero-3-phosphocholine (DOPC) and 1,2-dioleyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DOPG), and monitoring the secondary and tertiary structure changes using CD and NMR spectroscopy can aid in understanding the conformational changes in IpaC during infection.

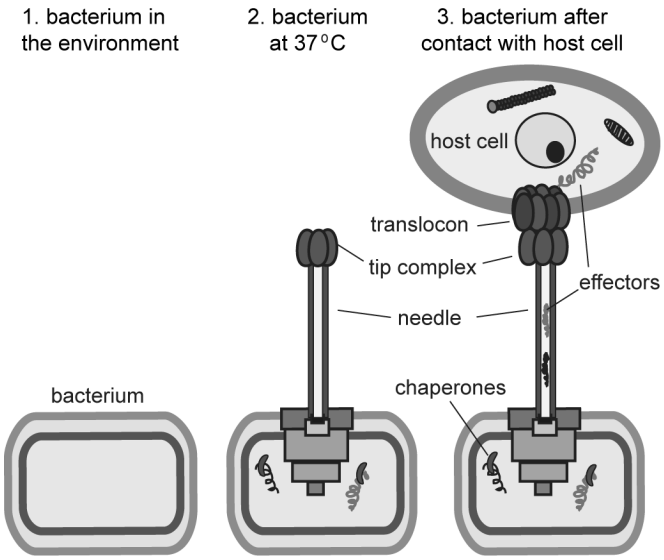

Figure 6-1. T3SS assembly. A temperature shift to 37^0C and contact with the host cell trigger the sequential assembly of the needle, tip complex and the translocon. The figure is adapted from Dey *et al* [16].

In my dissertation (Chapter 2), I overexpressed and purified IpaC$^{1\text{-}100}$ and IpaC$^{173\text{-}363}$. These constructs do not contain the predicted transmembrane domain (residues 100-120) [19]. I was able to successfully subclone the two constructs – IpaC$^{1\text{-}173}$ and IpaC$^{101\text{-}363}$, but these did not express proteins, preventing further characterization. My findings suggest that the transmembrane domain is important in the structural stability of IpaC, because IpaC$^{1\text{-}100}$ and IpaC$^{173\text{-}363}$ were found to exist as random coils in solution (Chapter 2). Alternative IpaC constructs, such as IpaC$^{1\text{-}121}$ and IpaC$^{121\text{-}363}$, containing the transmembrane region, can be designed to better understand the role of the transmembrane domain in IpaC conformation.

About 50% of the NMR peaks are missing in the ^{1}H-^{15}N TROSY spectrum of IpaC. Additionally, the peak dispersion in the proton dimension is less than 1 ppm (Chapter 2). This makes NMR assignment of IpaC challenging. An alternative approach to assign the NMR

spectrum of intrinsically disordered proteins (IDP) such as IpaC is by ^{15}N detection (**Figure 6-2**, adapted from [20]). This technique enhances the resolution of the NMR spectrum of intrinsically disordered proteins **(Figure 6-2)** [20]. Through a collaboration, the NMR spectrum of IpaC can be assigned by ^{15}N detection in absence and presence of detergents such as DPC or membrane mimetics such as liposomes. This knowledge will be important in further biophysical characterization of the IpaC protein family.

6.3. Protein-protein interactions of IpaC

6.3.1. Key findings

The protein-protein interactions of IpaC are not well understood [1-5, 17, 21-23]. Previously, the involvement of the C-terminal domain (CTD) of IpaC in interacting with the chaperone protein IpgC and the N-terminal domain of the major translocon protein IpaB was unknown. In my dissertation (Chapter 3), I have reported that in solution, the CTD of IpaC interacts with IpgC and the N-terminal domain of IpaB. Further, the CTD is responsible for IpaC oligomerization by interacting with the CTD of neighboring IpaC molecules. Interactions between IpaC and the tip protein IpaD was previously not detected [24]. My NMR studies (Chapter 3) show that in solution, both the N-terminal domain (NTD) and the CTD of IpaC bind to the tip protein IpaD predominantly at its N-terminal α-helical hairpin. My findings increase the knowledge of the protein-protein interactions of IpaC, and to better understand the role of these interactions in pathogenesis. My findings are summarized in **Figure 6-3.**

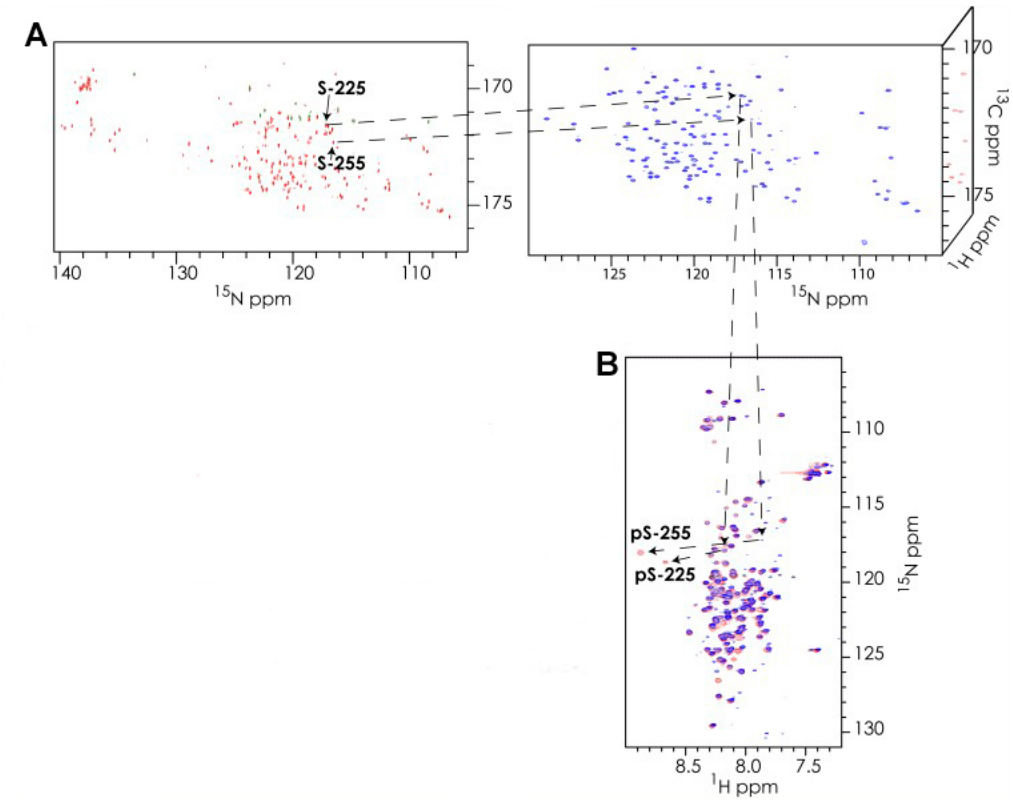

Figure 6-2. ^{15}N detection enhances the resolution of the NMR spectrum of IDPs. Assignments are transferred from A) ^{15}N-detected to B) ^{1}H NMR experiments used to assign the IDP, $NFAT_{131-294}$ [20]. Residues S225 and S255 are shown as examples. The resolution of the NMR spectrum from the ^{15}N-detected experiment (panel A) is higher than the ^{1}H-detected experiment (panel B). The figure is adapted from Chhabra *et al* [20].

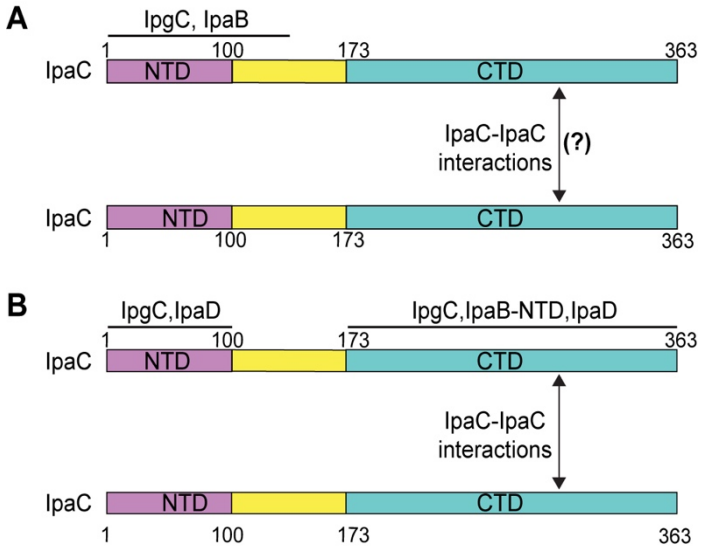

Figure 6-3. Key findings of IpaC protein-protein interactions. (A) Chaperone protein IpgC and the major translocon protein IpaB was known to bind to the N-terminal domain and parts of the central transmembrane domain of IpaC [23]. The precise involvement of the C-terminal domain in IpaC oligomerization was unknown. (B) My findings show that in solution, the C-terminal domain also interacts with IpgC, the N-terminal domain of IpaB and itself. Both the N-terminal and the C-terminal domains of IpaC interact with the tip protein IpaD. The IpaC oligomerization possibly occurs due to interactions among the CTD of neighboring IpaC molecules.

6.3.2. Future directions to study IpaC protein-protein interactions

The protein-protein interactions of IpaC discussed in Chapter 3 were studied in solution. Since IpaC is a membrane protein [1], studying its protein-protein interactions in membrane mimetics such as phospholipids and liposomes will help in understanding the biological relevance of these interactions. Assigning the NMR spectrum of IpaC by ^{15}N detection [20] can identify the residues and surfaces of IpaC involved in these interactions. Alternate techniques such Electron Paramagnetic Resonance (EPR) spectroscopy and fluorescence can also provide this information. If protein expression is successful using alternative IpaC constructs - $IpaC^{1-121}$ and $IpaC^{121-363}$ - then these constructs can also be used to study protein-protein interactions of the IpaC domains. These constructs contain the transmembrane domain of IpaC, and will provide a more complete understanding of the function of the IpaC domains in interactions with their binding partners. The IpaC protein-protein interactions should also be studied at 37^0C, which is the temperature at which IpaC interacts with intermediate filaments [1-5] inside host cells. Novel small molecule libraries can be screened by computational as well as surface plasmon resonance-based screening to identify compounds that can bind to IpaC. The residues and surfaces of IpaC that are involved in its interactions with the compounds can be determined using NMR spectroscopy. Compounds that were identified to bound to IpaC can then be tested for their ability to disrupt the protein-protein interactions of IpaC. This knowledge will be important in the future design of novel anti-infectives targeted against the *Shigella* minor translocon protein.

6.4. Characterization of the binding of novel compounds to SipD

6.4.1. Key findings

SPR-based screening identified three novel small molecules based on quinoline and phenyl scaffolds that bind to the *Salmonella* tip protein SipD. NMR data suggest two hotspots for binding the compounds (Chapter 4). Computational screening identified two new chemical scaffolds, based on the piperidine and piperazine, that bind to SipD. NMR data suggest a common hotspot for both the compounds (Chapter 5). Interestingly, the N-terminal α-helical hairpin and a portion of the coiled-coil of SipD referred to as the proximal region [25], was found to be important in binding all the identified scaffolds (Chapter 4 and Chapter 5). Non-specific interactions were also detected in another part of the coiled-coil, referred to as the distal region [25]. The proximal region of SipD interacts with the needle protein PrgI, while the distal region interacts with the N-terminal domain of the major translocon protein SipB [25]. The new scaffolds expand the repertoire of novel compounds that can target the *Salmonella* tip protein. This knowledge will aid the design of novel inhibitors of the protein-protein interactions of SipD and will contribute in the development of new anti-infectives against multidrug-resistant bacteria that rely on the T3SS for pathogenesis.

6.4.2. Future directions to design the next-generation anti-infectives

The design of next-generation inhibitors of the T3SS should be aimed at designing novel anti-infectives - compounds that do not necessarily kill the pathogenic bacteria but blocks the T3SS [26, 27], and thereby prevent the bacteria from infecting hosts. This approach will make the compounds selective for the pathogens without harming the host [28]. The novel chemical

scaffolds identified in my study (Chapter 4 and Chapter 5) should be expanded and tested for their ability to disrupt the SipD-PrgI and SipD-SipB NTD interactions using fluorescence methods such as Förster Resonance Energy Transfer (FRET). All the scaffolds show weak binding, with dissociation constants in the millimolar range. Future collaborations with medicinal chemists can lead to modifications of the scaffolds, with possibilities of elaboration, linking and introducing chirality. Biologically, proteins often interact with their binding partners in a stereoselective manner [29]. Introducing chiral centers in the chemical scaffolds identified in my study can increase their potency while simultaneously improving their specificity, solubility and bioavailability [29, 30]. These approaches will aid in the future design of lead compounds targeted against the tip protein of the T3SS.

6.5. References

1. Marquart, M.E., W.L. Picking, and W.D. Picking, *Soluble invasion plasmid antigen C (IpaC) from Shigella flexneri elicits epithelial cell responses related to pathogen invasion.* Infect. Immun., 1996. **64**(10): p. 4182-4187.
2. Tran Van Nhieu, G., et al., *IpaC induces actin polymerization and filopodia formation during Shigella entry into epithelial cells.* EMBO J, 1999. **18**(12): p. 3249-62.
3. Veenendaal, A.K., et al., *The type III secretion system needle tip complex mediates host cell sensing and translocon insertion.* Mol. Microbiol., 2007. **63**(6): p. 1719-30.
4. Blocker, A., et al., *The tripartite type III secreton of Shigella flexneri inserts IpaB and IpaC into host membranes.* J.Cell Biol., 1999. **147**(3): p. 683-693.
5. Menard, R., et al., *The secreted Ipa complex of Shigella flexneri promotes entry into mammalian cells.* Proc. Natl. Acad. Sci. U.S.A., 1996. **93**(3): p. 1254-8.
6. Lara-Tejero, M. and J.E. Galan, *Salmonella enterica serovar typhimurium pathogenicity island 1-encoded type III secretion system translocases mediate intimate attachment to nonphagocytic cells.* Infect. Immun., 2009. **77**(7): p. 2635-42.
7. Dey, S., A. Basu, and S. Datta, *Characterization of molten globule PopB in absence and presence of its chaperone PcrH.* Protein J, 2012. **31**(5): p. 401-16.
8. Faudry, E., et al., *Type III secretion system translocator has a molten globule conformation both in its free and chaperone-bound forms.* FEBS J., 2007. **274**(14): p. 3601-10.
9. Lunelli, M., et al., *Crystal structure of PrgI-SipD: insight into a secretion competent state of the type three secretion system needle tip and its interaction with host ligands.* PLoS Pathog., 2011. **7**(8): p. e1002163.
10. Chatterjee, S., et al., *The crystal structure of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate.* Protein Sci., 2011. **20**: p. 75-86.

11. Barta, M.L., et al., *Identification of the bile salt binding site on IpaD from Shigella flexneri and the influence of ligand binding on IpaD structure.* Proteins, 2012. **80**(3): p. 935-45.
12. Dickenson, N.E., et al., *Conformational Changes in IpaD from Shigella flexneri upon Binding Bile Salts Provide Insight into the Second Step of Type III Secretion.* Biochemistry, 2011. **50**: p. 172-180.
13. Wang, Y., et al., *NMR characterization of the interaction of the Salmonella type III secretion system protein SipD and bile salts.* Biochemistry, 2010. **49**(19): p. 4220-6.
14. Stensrud, K.F., et al., *Deoxycholate interacts with IpaD of Shigella flexneri in inducing the recruitment of IpaB to the type III secretion apparatus needle tip.* J. Biol. Chem., 2008. **283**(27): p. 18646-54.
15. Puhar, A. and P.J. Sansonetti, *Type III secretion system.* Curr Biol, 2014. **24**(17): p. R784-91.
16. Dey, S., et al., *The type III secretion system needle, tip, and translocon.* Protein Sci, 2019. **28**(9): p. 1582-1593.
17. Kueltzo, L.A., et al., *Structure-function analysis of invasion plasmid antigen C (IpaC) from Shigella flexneri* J. Biol. Chem., 2003. **278**(5): p. 2792-2798.
18. Osiecki, J.C., et al., *IpaC from Shigella and SipC from Salmonella possess similar biochemical properties but are functionally distinct.* Mol. Microbiol., 2001. **42**(2): p. 469-481.
19. Russo, B.C., J.K. Duncan, and M.B. Goldberg, *Topological Analysis of the Type 3 Secretion System Translocon Pore Protein IpaC following Its Native Delivery to the Plasma Membrane during Infection.* MBio, 2019. **10**(3).
20. Chhabra, S., et al., *(15)N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins.* Proc Natl Acad Sci U S A, 2018. **115**(8): p. E1710-E1719.
21. Romano, F.B., et al., *Type 3 secretion translocators spontaneously assemble a hexadecameric transmembrane complex.* J Biol Chem, 2016. **291**(12): p. 6304-15.
22. Russo, B.C., et al., *Intermediate filaments enable pathogen docking to trigger type 3 effector translocation.* Nat Microbiol, 2016. **1**: p. 16025.
23. Harrington, A.T., et al., *Structural characterization of the N terminus of IpaC from Shigella flexneri* Infect. Immun., 2003. **71**(3): p. 1255-1264.
24. Davis, R., et al., *Protein-protein interactions in the assembly of Shigella flexneri invasion plasmid antigens IpaB and IpaC into protein complexes.* Biochim.Biophys.Acta, 1998. **1429**(1): p. 45-56.
25. McShan, A.C., et al., *NMR Identification of the Binding Surfaces Involved in the Salmonella and Shigella Type III Secretion Tip-Translocon Protein-Protein Interactions.* Proteins, 2016. **84**: p. 1097-1107.
26. Allen, R.C., et al., *Targeting virulence: can we make evolution-proof drugs?* Nat Rev Microbiol, 2014. **12**(4): p. 300-8.
27. Clatworthy, A.E., E. Pierson, and D.T. Hung, *Targeting virulence: a new paradigm for antimicrobial therapy.* Nat. Chem. Biol., 2007. **3**(9): p. 541-8.
28. Duncan, M.C., R.G. Linington, and V. Auerbuch, *Chemical inhibitors of the type three secretion system: disarming bacterial pathogens.* Antimicrob Agents Chemother, 2012. **56**(11): p. 5433-41.
29. Brooks, W.H., W.C. Guida, and K.G. Daniel, *The significance of chirality in drug design and development.* Curr Top Med Chem, 2011. **11**(7): p. 760-70.
30. Villoutreix, B.O., et al., *A leap into the chemical space of protein-protein interaction inhibitors.* Curr Pharm Des, 2012. **18**(30): p. 4648-67.