MAPPING THE RIGID AND FLEXIBLE REGIONS OF A NOVEL SECRETED PROTEIN EvpP FROM THE T6SS OF EDWARDSIELLA TARDA

HU WENTAO

(B. Sc. Hons, NUS)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

NUS GRADUATE SCHOOL FOR INTERGRATIVE SCIENCES AND ENGINEERING

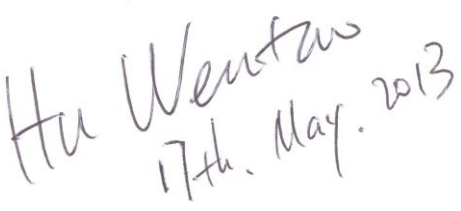
NATIONAL UNIVERSITY OF SINGAPORE

2013

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in this thesis.

This thesis has also not been submitted for any degree in any university previously.



Hu Wentao

Acknowledgements

I am cordially thankful to my supervisor, Assoc. Prof. Henry Mok, for his patient guidance and support throughout the years of my candidature. His critical thinking and advices have always encouraged me in my research work.

I would like to extend my gratitude to Prof. Yang and Assoc. Prof. Sivaraman for their generous sharing of creative ideas that have inspired me in NMR and crystallization studies.

My heartful thanks to Dr. Fan, Dr. Shiva, Dr. Lin Zhi, Dr. Chiradeep, Dr. Jana, Dr. Jobi, Dr. Tan, Dr. Lim and Dr. Smarajit for sharing invaluable experiences whenever I approached for consultation. To everyone in structural biology labs, many thanks for your companionship and helpful suggestions towards my experiments.

Special thanks to my parents, relatives and friends, you have always been supportive ever since the day I started my campus life in Singapore. Your love is the greatest treasure that I will forever cherish.

Table of Contents

Summary

Gram-negative bacteria have developed various secretion systems. These secretion systems play important roles in maintaining communication with the environment and in inflicting pathogenesis on respective hosts. Among them, Type III (T3SS) and Type VI (T6SS) secretion systems have been identified as the major components of bacterial virulence. Recently, more attention has been drawn to this research area, as proteins within these secretion systems might potentially represent a novel line of antibiotic development.

In this thesis, we aimed to gain structural insights into the novel secreted protein EvpP from the Type 6 secretion system of *Edwardsiella tarda*. EvpP is a unique protein found exclusively within the *Edwardsiella* species, which has no homologues among other genera. EvpP has been characterized as a non-stable protein which does not have a well-folded tertiary structure. Flexible regions with intrinsic disorder are well-known for their roles in mediating protein-protein interaction. Therefore we have mapped out the rigid and flexible regions of EvpP using three different approaches: limited proteolysis, nuclear magnetic resonance (NMR) backbone assignment and hydrogen-deuterium exchange mass spectrometry.

Limited proteolysis has shown a coherent pattern that the N-terminal half is more rigid than the C-terminal half. Trypsin-digested pattern further suggested another rigid region in a 4-kDa section within the C-terminal half of EvpP. We also obtained a relatively complete backbone assignment of EvpP using a stable mutant EvpP_P143T. More than 90% of the assigned peaks were located within the trypsin-digestion boundary. In addition, NMR data have suggested a third rigid region, possibly near the C-terminal end, which was supported by the hydrogen-deuterium exchange mass spectrometry (HDX-MS) experiments. Heat maps plotted using both wild-type and mutant EvpP demonstrated that the linker region between the N-terminal half and the 4-kDa section was the most flexible region throughout the entire

protein. Finally, we investigated the interaction between EvpP and EvpC, which is another secreted protein of T6SS. Glutathione S-Transferase (GST) pull-down assay using GST-EvpC and EvpP has demonstrated that the C-terminal flexible region might be involved in their interaction; while NMR titration experiments have suggested a binding stoichiometry of one to one. These results may pave a way for better understanding of the structure and function of this novel secreted protein, EvpP.

The second chapter of this thesis aimed to optimize the boundary for the crystallization of a chaperone-translocator complex, AcrH-AopB from the T3SS of *Aeromonas hydrophila*. AcrH is a Class II chaperone for translocators AopB and AopD. The structures of Class I and Class III chaperones in complex with either effector or apparatus proteins have been determined already. However, no complex structure of a Class II chaperone with translocator has been reported. Using limited proteolysis with a combination of three proteases, thrombin, thermolysin and actinase E, we obtained a much shorter boundary of the complex $AcrH_{\Delta}C$-$AopB^{61-264}$ that could still bind each other tightly. Although this construct did not yield a high quality crystal; it nevertheless suggested a shorter AcrH binding region of AopB. The N-terminal disordered region of AopB might have a novel function instead of being partially involved in binding its chaperone AcrH.

List of Table

List of Figures

LIST OF ABBREVIATIONS

Amino Acids

One letter code	Three letter code	Amino acid
A	Ala	Alanine
C	Cys	Cystein
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tyrptophan
Y	Tyr	Tyrosine

Chemicals and reagents

BSA	Bovine serum albumine
$CaCl_2$	Calcium chloride
D_2O	Deuterium Oxide
dATP	2’ deoxyadenosine 5’ triphosphate
dCTP	2’ deoxycytidine 5’ triphosphate
dGTP	2’ deoxyguanosine 5’ triphosphate
dNTP	Deoxynucleotide triphosphate
dTTP	2’ deoxythymidine 5’ triphosphate
IPTG	Isopropyl-D-thiogalactoside
KCl	Potassium chloride
$MgCl_2$	Magnesium chloride
Ni	Nickel
Ni-NTA	Nickel-Nitrilotriacetic
PBS	Phospate buffer saline
SDS	Sodium dodecyl-sulphate
TEMED	N,N,N,N’-Tetramethylethylenediamine
TFA	Trifluoroacetic acide
Tris	Tris (hydoxymethyl)-aminomenthane

Units and measurement

bp	Base pair
Da	Dalton
Hz	Hertz
K	Kelvin
kDa	Kilo Dalton
M	Molar
pH	Potential of hydrogen
ppm	Parts per million
rpm	Rotation per minute
V	Volt

Others

1D/2D/3D/4D	One dimensional/Two dimensional/3D/ Four dimensional
α	alpha
β	beta
γ	gamma
δ	delta
ε	epsilon
CD	Circular dichroism
DNA	Deoxyribonucleic acid
ESI	Electrospray Ionization
FPLC	Fast Protein Liquid Chromatography
GST	Glutathione S-transferase
HDX	Hydrogen-Deuterium exchange
HSQC	Heteronuclear Single-Quantum Correlation
LB	Luria Bertani
MALDI-TOF	Matrix-assisted laser desorption/ionization
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
OD	Optical density
PCR	Polymerase chain reaction
SDS-PAGE	Sodium Dodecyl-Sulphate Polyacrylamide Gel Electrophoresis
TOCSY	Total correlation spectroscopy

CHAPTER 1, INTRODUCTION

1.1 Secretion System

1.1.1 A general introduction to bacterial secretion systems

Secretion, the term describing the dedicated transport or translocation of molecules from the interior of a cell to its exterior, is a universal process throughout all three domains of life. Bacterial secretion systems play important roles in pathogenesis, or virulence, as well as maintaining communication among bacterial cells or between bacterial cells and their host. Through this process, bacteria could adapt to the environment and gain invasion into its respective host cell. Up to now, great breakthroughs in this research field have shed light on the structure, topology and mechanism of various bacterial secretion systems. Inhibiting secretion seems to be a novel line of antibiotic development compared to traditional antibiotics, which aim to kill bacteria directly.

1.1.2 Secretion in gram-positive bacteria

Secretion in gram-positive bacteria are more straight-forward than gram-negative bacteria because gram-positive bacteria have a single plasma membrane and a thick peptidoglycan layer in their cell walls (Shockman and Barrett, 1983). Precursor proteins destined for secretion in gram-positive bacteria usually carry an N-terminal signal peptide. Most of them are transported across the plasma membrane in an unfolded state by the general secretion pathway (Sec pathway, Karel et al, 2001). After this, the precursor gets processed and folds into its active form. Some secreted proteins carry C-terminal sorting signal peptides that will get cleaved by different sortase enzymes after crossing the plasma membrane. Sortase A cleaves the C-terminal LPXTG sorting signal and anchors proteins to the cell wall (Ton-That, H et al, 1999). Sortase B cleaves the sorting peptide NPQTN or NPKTG (by sortase B homologues) and anchors haem scavenging factors (Mazmanian et al, 2002 and

Maresso et al, 2006). Sortase C cleaves the sorting signal LPNTA and anchors proteins to the envelope of spores (Marraffini and Schneewind, 2006). Aside from the general secretion pathway, two-arginine (Tat) pathway is also found in gram-positive bacteria. The Tat machinery recognizes an N-terminal sequence motif that is rich in basic amino acid residues (S-R-R-x-F-L-K) and transports proteins in its folded state. Proteins secreted by this pathway usually have large cofactors bound to it (Muller M, 2005). In addition, Type VII Secretion system (T7SS) is found in some species of gram-positive bacteria.

1.1.3 Secretion in gram-negative bacteria

Unlike gram-positive bacteria, the secretion systems in gram-negative bacteria are more complicated because gram-negative bacteria possess not only an inner plasma membrane, but an outer membrane as well. Although the general secretion pathway and Tat pathway were also present in gram-negative bacteria; proteins secreted by the above pathways (mostly with an N-terminal signal peptide) are usually transported into the periplasmic space. Therefore additional machinery is needed to further transport the proteins across the outer membrane or proteins can be transported from the cytosol directly to the extracellular milieu. Up to seven different types of secretion systems and several subtypes have been found in gram-negative bacteria. They include chaperone-usher pathway and six other pathways, namely Type I (T1SS) to Type VI (T6SS) secretion systems. The chaperone-usher pathway is responsible for the secretion and assembly of adhesive surface structures called fimbriae (Andreas and Gabriel, 2011). T1SS to T6SS transport a wide variety of molecules ranging from nucleic acids to proteins. The process of secretion also differs from one-step (molecules are transported directly from the cytosol to the extracellular space) to two-step (molecules are transported from the cytosol to the periplasm before being transported across the outer membrane to the extracellular space). Type III and Type VI secretion systems have been discovered to be the major components of

bacterial pathogenesis towards various hosts. Therefore, these two secretion systems will be further discussed in this thesis.

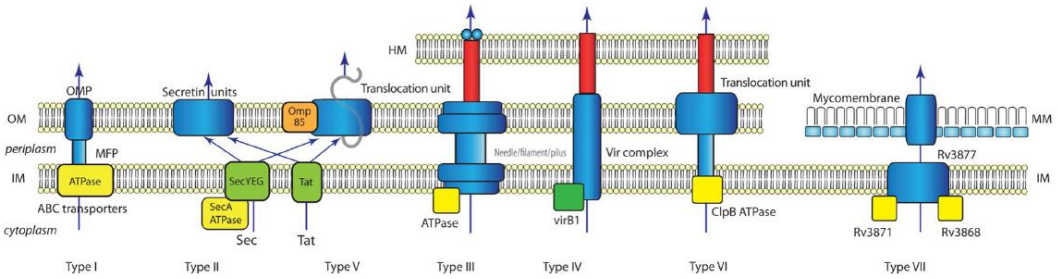


Figure 1.1, Simplified view of the seven secretion systems. HM: host membrane; OM: outer membrane; IM: inner membrane; MM: mycomembrane; OMP: outer membrane protein; MFP: membrane fusion protein. Tseng *et al. BMC Microbiology* 2009 **9**(Suppl 1):S2

1.1.4 Type III Secretion System (T3SS)

Type III secretion systems were first discovered in pathogenic *Yersinia* species, for its secretion of different Yop proteins (Michiels P et al, 1990). Since then, T3SSs have been found in both plant and animal gram-negative pathogens, some of which communicate with their host as mutualists. Similar to T1SS, T3SS is sec-pathway independent, as effectors get secreted across both inner and outer membranes in a one-step process without a periplasmic intermediate. However the assembly of T3SS, which involves the transport of T3SS subunits to the outer membrane, may partly rely on the general secretion pathway. More than 20 subunits of T3SS have to be orchestrated to form a needle complex spanning both membranes and the periplasmic space. These genes are usually localized together in a plasmid or a chromosomal locus with some species possessing more than one copy of T3SS. T3SS also has secretin-like proteins, similar to those in T2SS. Some of these proteins oligomerize to form ring-shaped structures with central pores while some stabilize the needle-like complex formed by other components of T3SS (Gauthier et al, 2003).

A typical T3SS usually consists of a basal body spanning both the inner and outer membrane of the bacteria, a needle complex that protrudes outwards from the bacterial cell surface and a translocon at the tip of the needle complex (Cornelis and

Van G, 2000). The T3SS needle complex is essentially formed by a single protein of small molecular weight. Such proteins have been identified in various organisms: PrgI in *Salmonella typhimurium* (Kimbrough and Miller, 2000), YscF in *Yersinia enterocolitica* (Hoiczyk and Blobel, 2001), MxiH in *Shigella flexneri* (Blocker et al, 2001) and PscF in *Pseudomonas aruginosa* (Pastor, et al, 2005). These small proteins are supposed to polymerize and form a conduit for the transportation of effectors. The structures of some needle complex proteins have been solved: PrgI (Wang, et al, 2007), MxiH (Deane, et al, 2006) and AscF (Yih W T, et al, 2008). They all form an anti-parallel two-helix-bundle joined by a loop turn consisting of the conserved residues [P-(S/D)-(D/N)-P].

One striking feature of T3SS is its ability to secrete effectors directly into eukaryotic host cells. This is triggered by the direct contact of the bacteria with its host. Therefore T3SS is termed as "contact dependent". At the tip of the injectisome that contacts the host cell surface, T3SS forms a "translocon" which consists of three subunits. Two of them have hydrophobic domains which are proposed to be inserted into host cell membrane, allowing effector translocation into host cell cytosol (Hakansson et al, 1996; Warawa et al, 1999; Faudry et al, 2006). The third subunit is hydrophilic and the assembly of the translocon is dependent on this hydrophilic subunit (Fields et al, 1999; Marenne et al, 2003). One well-studied translocon model involves the LcrV-YopB-YopD translocators of *Yersinia enterocolitica*; where LcrV is the hydrophilic subunit which forms a tip at the distal end of the injectisome. YopB and YopD are the hydrophobic translocators that are inserted into host cell membrane (Goure et al, 2005). The genes encoding the translocators usually localize together in a large operon. Within the same operon, there's usually one gene encoding a small protein, which functions as the chaperone for the hydrophobic translocators, sometimes with another gene encoding a regulator protein of T3SS. Some of the effectors, such as YopE and YopT as an example from *Yersinia* species, may even carry anti-pore formation activity after being injected into the host cell (Viboud and

Bliska, 2001). Although direct secretion of effectors into host cell cytosol is a fascinating feature of T3SS, it is not universal to Type 3 secretion as many proteins secreted by T3SS do not translocate into host cells.

One common feature of the T3SS translocator proteins is their possession of coiled-coil regions. A coiled-coil region usually comprises of two or more α-helices which form a bundle like structure. This structure is generally associated with increased flexibility; but important for protein-protein interactions with other members of T3SS or with itself (Anastasia D, et al, 2009). However, coiled-coil regions are not unique to translocator proteins. It has been found that HrpO from *Pseudomonas syringae*, a soluble apparatus protein possibly with ATPase activity, interacts with HrpE through the contact of the coiled-coil regions in both proteins (Anastasia D, et al, 2008).

Effectors translocated into host cell cytosol via T3SS often carry out specific functions such as repressing host cell defense response or modulating its cytoskeleton. One well-studied model organism for such effector function is *Shigella flexneri*, a gram-negative bacterium that can cause shigellosis in humans. Among the subset of effectors delivered by the T3SS of *S. flexneri*, VirA was reported to trigger host cell microtubule destabilization by directly binding to the α/β-tubulin heterodimers (Sei Y, et al, 2002). This destabilization can cause local microtubule growth, which in turn, results in the induction of Rac1, a small Rho-GTPase involved in various pathways of actin cytoskeleton remodulation (Wittmann and Waterman-Storer, 2001). Thus, the invasion of *S. flexneri* will cause local membrane ruffling that allows further bacterial endocytosis. Apart from microtubules, actin filament is another major target of T3SS effectors for host cell cytoskeleton reorganization. IpgB1, also from *S. flexneri*, has been found to be a novel effector secreted by T3SS. Upon entry into host cell, IpgB1 was able to induce membrane ruffles through the activation of Rac1 and CDC42 (Kenji O, et al, 2005). IpgB1 was later shown to activate Rac1 through the RhoG-ELMO-Dock180 pathway where IpgB1 mimics the function of Rho-GTPases by

binding to the N-terminal region of ELMO (Yutaka H, et al, 2007). Refer to Figure 1.2 for the schematic representation of *Shigella flexneri*-mediated host cell membrane ruffling.

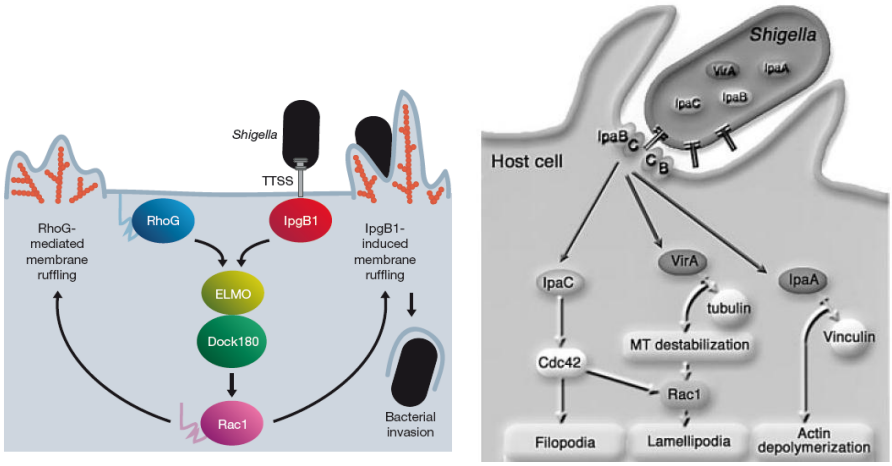


Figure 1.2, Schematic view of the proposed model for Shigella flexneri-induced membrane ruffling. (Yutaka H, et al, 2007 and Kenji O, et al, 2005)

Another effector from *S. flexneri*, IpgB2 was also shown to target Rho-GTPases in host cells. Different from IpgB1, which directly mimics the function of Rho-GTPases, IpgB2 mimics the function of guanine nucleotide exchange factors (GEFs). Additionally, its structure in complex with human RhoA has been solved. Upon binding, IpgB2 promotes the release of guanosine diphosphate (GDP) from RhoA and this process is a pre-requisite for subsequent guanosine triphosphate (GTP) loading (Bjorn U, et al, 2010).

Another important group of T3SS members is the chaperones. Many T3SS members possess extended hydrophobic regions, such as the hydrophobic translocators and many other apparatus subunits. Therefore they have to be associated with chaperones before they can interact with respective binding partners and fold into a mature state. The function of the chaperone is to prevent the pre-mature aggregation of such hydrophobic proteins and keep them in a relatively stable state until they meet their functional binding partner. So far, three classes of chaperones have been identified within the T3SSs of various gram-negative bacteria. Class I chaperones interact with effector proteins. These are the well-studied chaperones

among which, many structures have been solved, with some of them being in complex with their respective effectors. Although Class I chaperones share very little sequence homology, they all have a common fold following an α-β-β-β-α-β-β-α pattern (Nancy J, et al, 2010). When the known structures of chaperone-effector complexes were compared, a common target motif in T3SS was identified (Lilic M, et al, 2006, Figure 1.3). Firstly, the chaperones form a dimer as the core of the complex. In the case of SycN, it interacts with YscB to form a heterodimeric chaperone. The effectors are then "wrapped around" the chaperones in a non-native state. Such non-globular interaction is stabilized by two conserved β-motifs at each side of the chaperone dimer as well as other interacting regions that are specifically tailored to respective chaperone-effector complex sequences. At the two conserved β-motif sites, the effector forms an intermolecular β-sheet with its chaperone and inserts several hydrophobic residues into the hydrophobic crevice on the chaperone molecule. This was found invariably through all the known complexes. Therefore it stands as a common structural motif for chaperone-effector interaction in T3SS (Lilic M, et al, 2006).

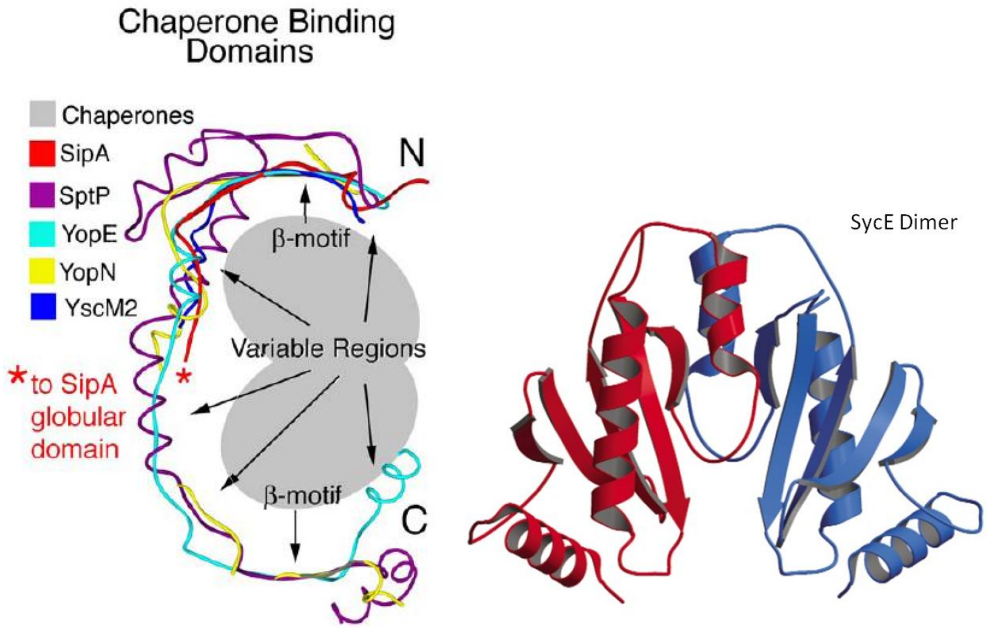


Figure 1.3, Superposition of non-globular interaction between T3SS Class I chaperones and the effectors. Left panel: The complex structures used were SipA-InvB (Lilic M, et al, 2006), SptP-SicP (Stebbins and Galan, 2001), YopE-SycE (Bertalan, et al, 2002), YopN-

SycN-YscB (Schubot F. D, et al, 2005) and YscM2-SycH (Phan J, et al, 2004). Right Panel: ribbon representation of the Class I chaperone SycE dimer.

Class II chaperones interact with translocators. The structures of many Class II chaperones have been solved. One model Class II chaperone is LcrH/SycD from *Yersinia enterocolitica* (Figure 1.4). Class II chaperones generally possess tetratricopeptide repeats (TPR), which are also found in Class III chaperones. TPR is a structural motif consisting of a degenerate 34 amino acid sequence. It is found in tandem arrays of 3 to 16, which form scaffolds to mediate protein–protein interactions and the assembly of multi-protein complexes. The binding regions of Class II chaperones on various translocators have been mapped by limited proteolysis. However, no complex structure formed by Class II chaperone and translocator has been reported.

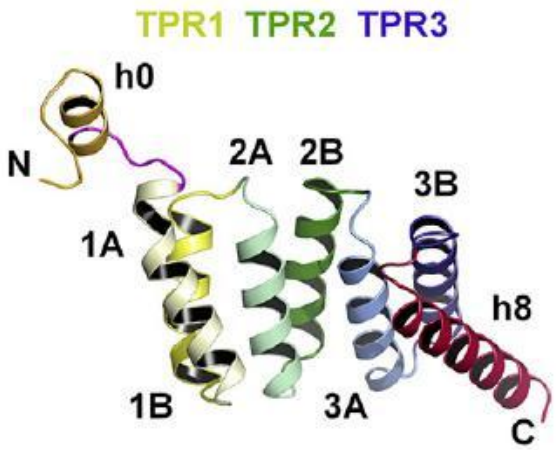


Figure 1.4, Ribbon representation of the structure LcrH/SycD. (Buttner, et al, 2007).

Class III chaperones interact with apparatus proteins. The best-studied Class III chaperone complex is the needle-complex-forming subunit PscF of *Pseudomonas aruginosa*, in complex with the heteromolecular chaperones PscE and PscG (Quinaud, et al, 2006). The orthologue complex structure of PscEGF complex in *Yersinia pestis*, YscEGF has also been solved. Similar to the PscEGF complex, YscG is a chaperone with a tetratricopeptide repeat domain. It binds to the C-terminal half of YscF while the N-terminal 49 residues of YscF are disordered. YscE interacts with the N-terminal tetratricopeptide repeat motif of YscG, but it makes little contact with YscF (Sun P, et

al, 2007). In both complexes, the apparatus proteins, PscF and YscF, bind to the concave region formed by the chaperone (Figure 1.5).

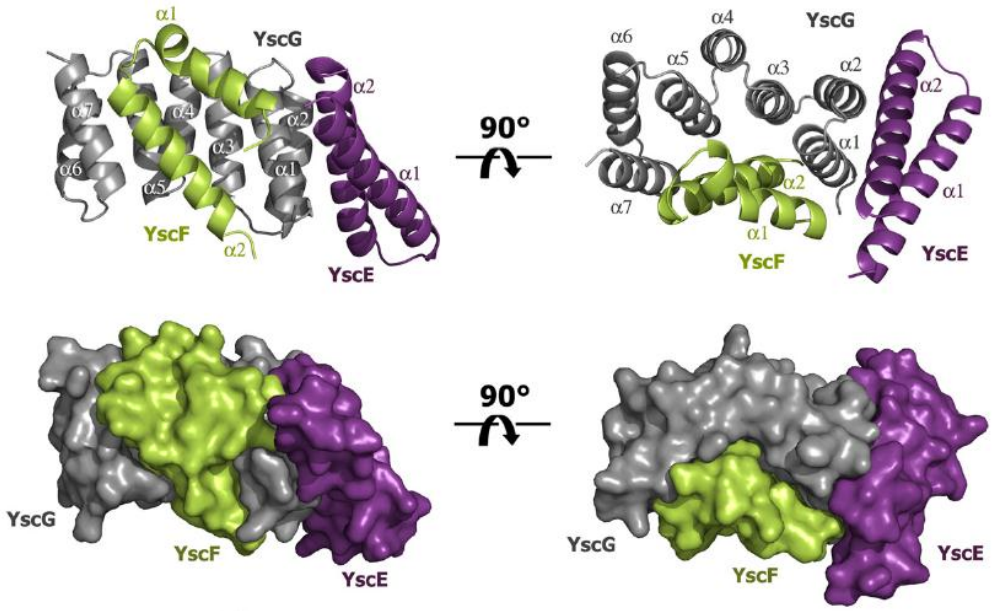


Figure 1.5, Ribbon and surface representation of the YscEGF complex. YscG was shown in grey; YscF was shown in yellow and YscE was shown in purple.

1.1.5 Type VI Secretion System (T6SS)

The Type VI secretion system was recently identified in 2006. Since then, it has been discovered in many gram-negative bacteria. It was first identified as IcmF associated homologous protein, after the discovery that one of the genes resembled icmF in Type IV secretion system (Das S et al, 2003). Similar to T3SS, T6SS is also found to be highly responsible for pathogenesis in humans and many other animals. T6SS components might also constitute a phage-tail like injectisome, which can deliver effector enzymes or toxins into the environment or directly into the host cell (Cascales E, 2008; Bingle LE et al, 2008).

Two secreted proteins, hemolysin-coregulated protein (Hcp) and valine-glycine repeat protein G (VgrG), are reliable indicators for the presence of T6SS. The crystal structure of Hcp showed a hexameric ring arrangement that polymerizes to form a tube with an inner diameter of 40 Å that can go up to 100 nm in length (Joseph

D, et al, 2006; Ballister ER et al, 2008). This suggests that Hcp may contribute to the conduit of T6SS through which effectors can be secreted into the environment or into host cell cytosol (Figure 1.6). The VgrG family of T6SS shares structural features with the T4 $gp27_3$-$gp5_3$ needle complex. Specifically, the N-terminal portion of VgrG shares similarities with gp44 protein of bacteriophage Mu. The C-terminal portion of VgrG shares similarities with gp5 protein of T4 phage. As gp44 is a structural orthologue of T4's gp27, VgrG closely resembles the $gp27_3$-$gp5_3$ needle complex with a cell-puncturing structure (Pukatzki S et al, 2007). Indeed, VgrGs form multi-meric complexes that support the hypothesis of phage-like cell-puncturing device formation. Some VgrG proteins carry functional C-terminal extensions, enabling them to function as a membrane-puncturing tip at the end of the needle complex and as an effector inside the host cell as well. VgrG1 of *Vibrio cholerae* was found to cross-link host cell actin which leads to the collapse of the host cell cytoskeleton (Sheahan KL et al, 2004).

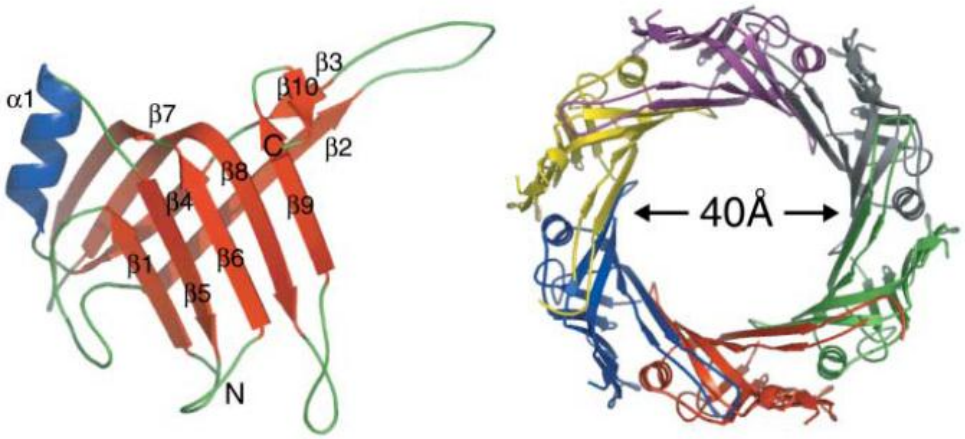


Figure 1.6, Ribbon representation of Hcp1 from *Pseudomonas aruginosa*. Left: ribbon diagram of Hcp1 monomer; Right: ribbon diagram of the hexameric ring fromed by Hcp1 where each monomer was coloured differently

A model of Type VI secretion has been proposed based on the properties of Hcp and VgrG, as Hcp was found inside the periplasmic space. VgrG, after getting into the periplasm, forms a trimeric complex similar to the T4 bacteriophage cell-puncturing device. Hcp may dock beneath the VgrG complex and start polymerizing

to form a tube. As this tube keeps growing, VgrG located at the end of the tube may puncture the outer membrane. This would allow the Hcp tube to pass through, forming a pore on the outer membrane that allows the secretion of other effectors to the extracellular space (Mougous JD et al, 2007).

A typical T6SS usually consists of 15 to 25 genes located in a single locus. Most of the T6SS members were thought to be structural components of the needle complex with a clpB homologous ATPase that provides energy for secretion. However the structure of the secretion system and function of each component still remain poorly understood.

As T6SS and T3SS are the two major systems that contribute to bacterial pathogenesis, cross-talks have been commonly found between these two secretion systems, which occur mainly at the gene transcription level. Among five species (*Salmonella enterica, Edwardsiella tarda, Aeromonas hydrophila*, *Vibrio cholerae* and *Pseudomonas aruginosa*) whose T3SS and T6SS were the major components for pathogenesis, two global regulators were found connecting them and other virulence determinants. PhoPQ, a two-component system is widely found among *Salmonella*, *Edwardsiella* and *Aeromonas* species that are able to sense the change in temperature and Mg^{2+} concentration of the environment. PhoQ is a trans-membrane histidine sensor kinase that auto-phosphorylates itself at a conserved histidine residue upon activation. It transfers the phosphoryl group to the cytoplasmic response regulator PhoP through an aspartic acid residue. This activates PhoP, which can also function as a transcription regulator (Smarajit C, et al, 2010). The other global regulator found among *Aeromonas* and *Vibrio* species is σ^{54}. σ^{54} is an alternate sigma factor required to initiate the ribonucleic acid (RNA) synthesis of virulence genes in many plant and animal pathogens. It interacts with RNA polymerase and is involved in promoter recognition at the site of transcription (Leung K Y, et al, 2011).

1.2 An introduction to *Edwardsiella tarda*

Edwardsiella tarda is a gram-negative bacillus which is facultatively anaerobic and motile. It was named after the renowned microbiologist P. R. Edward (Janda et al, 1991) and became first characterized under the *enterobacteriaceae* family in 1965 (Public Health Agency of Canada, 2001). *E. tarda* is a well-known fish pathogen with a broad range of hosts. It can live both inside and outside of its host, causing hemorrhagic septicemia (also known as edwardsiellosis). This is characterized by skin lesions and swelling which usually results in the death of the fish (Yousuf et al, 2006). It could also be an opportunistic pathogen to humans, causing gastroenteritis and diarrhea (Clarridge, 1980). Although rare, in both cases infection could lead to death if left untreated. *E. tarda* has also been shown to have a high affinity for red blood cells, due to its specific fimbriae that renders the ability of hemagglutination (Sakai et al, 2003). *E. tarda* is very susceptible to many common antibiotics like ampicillin and kanamycin. However it is quite resistant to colistin, streptogramin, rifampin and glycopeptides (Stock et al, 2001). Due to the rise of concern for antibiotic resistance (most likely because of overuse of antibiotics), it has become a health issue for both fish farm industry and human health.

Studies have been focusing on the mechanism of pathogenesis for *E. tarda* at the gene and molecular level (Rao PS et al, 2004). With the characterization of T6SS in *E. tarda* (Zheng and Leung, 2007), it was again brought to the research frontier as the secretion systems of these pathogens could lead to a new generation of antibiotic development. This is because both T3SS and T6SS were present in *E. tarda* and could contribute to pathogenesis. Cross-talks between the T3SS and the T6SS in *E. tarda* have also been partially elucidated (Leung K Y, et al, 2011). The two-component system PhoPQ acts as a global regulator whereas PhoQ senses environmental change such as temperature and low Mg^{2+} concentration. PhoQ activates PhoP by phosphorylation and PhoP, in turn, activates EsrB. EsrB is a regulator that is responsible for most of the T3SS genes and *esrC,* which encodes

another transcription regulator that controls T6SS genes and some specific T3SS genes (Figure 1.7).

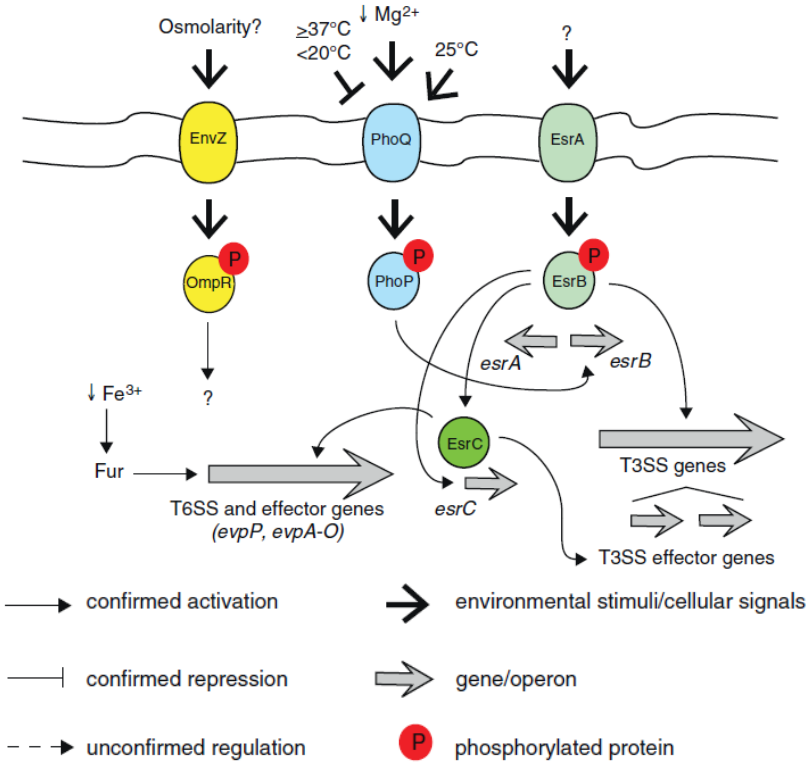


Figure 1.7, Schematic diagram representing cross communications between T3SS and T6SS in *E. tarda*. (Leung K Y, et al, 2011)

Therefore, the possibility of discovering new drugs or vaccines that can specifically target individual secretion system components or global regulators of such secretion is of great interest. These drugs or vaccines will be able to repress bacterial virulence without affecting the viability of the bacteria by itself. Thus, only by "disarming" them from doing harm to their respective hosts, this will not cause the development of drug resistance to the antibiotic.

1.2.1 T6SS in *Edwardsiella tarda* and the role of EvpP

The Type VI secretion system of *Edwardsiella tarda* was fully sequenced and characterized in 2007 (Zheng and Leung, 2007). There are at least 16 open reading frames within the gene cluster (see Figure 1.8). They were named as *E. tarda* virulent proteins (EVP) ranging from EvpA to EvpP. The transcription of EvpA to EvpO

might be polycistronic. EvpP lies upstream from EvpA where it might be utilizing a different promoter for transcription. However the transcription of other EVP members may also use the promoter for EvpP (Zheng and Leung, 2007).

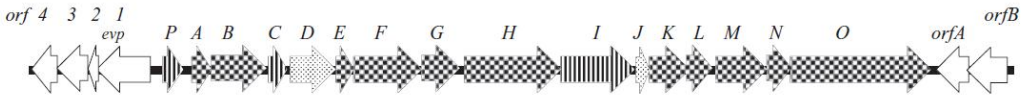


Figure 1.8, The T6SS gene cluster of Edwardsiella tarda. evp: *E. tarda* virulent protein; orf: open reading frame. (Zheng and Leung, 2007)

Within the gene cluster, EvpP, EvpC and EvpI are the three secreted proteins of T6SS. EvpC is a homologue of Hcp. The crystal structure of EvpC was solved in 2010. EvpC also forms a hexameric ring structure in high concentration like Hcp; but it will form a dimer in low concentration (Jobichen C et al, 2010). This suggests that EvpC may polymerize to form the conduit of the needle complex of T6SS. The secretion of EvpC and EvpI are mutually dependent while the secretion of EvpP requires EvpC and EvpI. Aside from the secreted proteins, most other T6SS members of *E. tarda* were thought to function as structural components. Systematic mutations of each of the EVP members have shown that most of the EVP members contribute to the virulence of *E. tarda* towards blue gourami fish except for EvpD and EvpJ. EvpO is predicted to be an ATPase protein with a walker A motif, which might be homologous to the ClpV (VCA116) protein from *Vibrio cholerae*. ClpV has been demonstrated to interact with VipA/VipB complex through its N-terminal domain where EvpA is the VipA homologue in *E. tarda* and EvpB is the homologue of VipB (Gabriele et al, 2009). Possible interactions have also been suggested between EvpO and EvpA, EvpL as well as EvpN (Zheng and Leung, 2007). However thus far, how the different components assemble into the needle complex and their respective function in T6SS have not yet been understood.

One important feature about the T6SS of *E. tarda* is that EvpP, one of its secreted proteins, is unique to this species but not found (or to have homologues) among other gram-negative pathogenic bacteria that have been confirmed to have a

T6SS. EvpP has also been found in a few strains of *Edwardsiella ictaluri* (Yang M, et al, 2012). Although a strain of *Aeromonas hydrophila* 0865 has been reported to have the *evpP* gene; this might be due to horizontal gene transfer (Xin Wang et al, 2009). EvpP has been proposed as an effector protein. This is because EvpP is a secreted soluble protein and Δ*evpP* mutants of *E. tarda* have significantly attenuated virulence towards their host (Zheng and Leung, 2007; Xin Wang et al, 2009). EvpP is regulated by *esrB* and iron, through which cross-talks between T6SS and T3SS have been proposed (Smarajit C, et al, 2010 and Leung K Y, et al, 2011). It was also shown to mediate mucus adhesion, hemolysis and host serum resistance towards Japanese flounder (Xin Wang, et al 2009). The secretion of EvpP is dependent on EvpC and EvpI, where the interaction between EvpP and EvpC has been proposed. So far, the structure and role of EvpP for the formation of T6SS needle complex still remain unknown.

1.2.2 Mapping the rigid and flexible regions of a protein

Proteins usually need to fold into a 3-dimensional structure (tertiary structure) before it can be functionally active. The information for protein folding lies within its primary structure, or amino acid sequence. A properly folded protein may be rigid or flexible to various degrees that are required for optimum function. The well-structured part of a protein is rigid, which ensures the stability of the protein. The function of flexible regions of a protein is more versatile. It has been well accepted that flexible loop regions of enzymes have critical functional roles. Sometimes the change in protein structure upon binding of its partner could be dramatic. Protein flexibility at the level of amino acid side-chains occurs universally, and is important for binding and catalysis (Huber R, 1987; Kannan and Ruth, 2007; Kaare et al, 2009). Therefore, mapping the rigid and flexible regions of a protein could give insights on its global folding as well as interacting domains.

Recent studies on T3SS have revealed the structural and functional versatility of coiled-coil domains within proteins of T3SS. Such domains consist of two or more α-helices forming a bundle structure. T3SS proteins with coiled-coil domains were often characterized by an increased structural flexibility ranging from local disordered regions to the formation of molten-globules. Such coiled-coil formation and its related structural disorder are important for their function, especially in establishing protein-protein interaction and self-polymerization (Anastasia D, et al, 2009). An example would be the interaction between T3SS chaperones with their respective binding partners. Class I chaperones form dimers in a coiled-coil fashion through the central pair of α-helices (Figure 1.3). The effectors bound to Class I chaperones are all extensively unstructured (refer to Section 1.1.4). Translocator proteins of T3SS interact with Class II chaperones, which have tetratricopeptide repeats forming a curved layer of α-helices (Pallen et al, 2003 and Buttner et al, 2008). The N-terminal region of translocator protein B (outside chaperone-binding region) is totally disordered and the C-terminal region is thought to be responsible for polymerization through the interaction of coiled-coil domains. In addition, T3SS needle complex subunit proteins bind Class III chaperone through coiled-coil interactions as well. As in PscEGF complex, the C-terminal helix of PscF is stabilized in a helical bundle formed by part of the PscG tetratricopeptide repeats (Quinaud, et al, 2007).

Since long ago, research on protein folding and dynamics have developed a wide range of methods to study protein flexibility. With the current advance in nuclear magnetic resonance (NMR) and mass spectrometry (MS), studies on protein dynamics have been rejuvenated where mapping proteins' rigid and flexible regions are allowed.

1.2.2.1 Limited protease digestion

One traditional way of mapping the rigid and flexible portions of a protein is by limited protease digestion, or limited proteolysis, using various enzymes. The underlining rationale is that: for the protease to be able to cleave the peptide bond of certain location on the substrate's amino acid chain, the substrate has to bind and adapt to the specific stereochemistry of the protease's active site (Schether and Berger, 1967). As the active sites of different proteases are not designed to fit the specific sequence of the substrate, the rigid regions of globularly folded protein are more resistant to such proteolysis. On the other hand, limited proteolysis often occurs at the sites that show a good correlation with larger crystallographic B-factor, uncertain electron density and larger dispersion of backbone angles (Fontana, 1993; 1999). Therefore limited proteolysis usually takes place at sites where the substrate displays inherent backbone flexibility (Fontana et al, 1986; 1989). Using limited protease digestion followed by sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE), we will be able to discriminate the rigid core of a protein (which will remain uncleaved, thus showing up on the SDS-PAGE as a resultant peptide) from the flexible regions (which are susceptible to limited proteolysis; therefore being "chopped up" and does not show up on the SDS-PAGE). The identity of the band on the resultant SDS-PAGE can be further determined by protein sequencing or peptide mass fingerprinting.

Limited proteolysis has been shown to be a powerful tool to map binding regions between chaperones and their respective binding partners (Yih W T, et al, 2009; Quinaud, et al, 2007). However this is not restricted to protein-protein complexes, as it can also help in determining the optimal boundaries for a single protein's core domain. A good example would be YopR, one of the least-studied members of the T3SS in *Yersinia. pestis*, its structure being determined in 2005 (Florian D, et al, 2005). While full length YopR failed to yield crystals; its core domain was elucidated (residues 38 to 149, out of a total of 165 amino acid residues)

after limited protease digestion by thermolysin, allowing the crystal structure to be determined eventually.

1.2.2.2 Hydrogen-deuterium exchange mass spectrometry (HDX-MS)

Hydrogen-deuterium exchange (HDX) is a chemical process in which a covalently bonded hydrogen atom is replaced with a deuterium atom and vice versa. This often occurs on the backbone amide proton when the protein is exposed to heavy water (deuterium oxide, D_2O) at a neutral pH level. Usually in the disordered region of a protein where there's large backbone flexibility, HD exchange takes place easily due to the lack of protection by hydrogen bonding. Well-structured regions are more protected from HDX, resulting in a slow isotope exchange which is mediated by the amino acid's local environment. With the advance in MS technology, MS-based peptide mapping is currently a widely used technique in identifying and measuring the mass shift of protein segments (Lars et al, 2010). Therefore, by using HDX in tandem with MS, we could gain structural insights into the intrinsically disordered regions of various proteins.

In HDX-MS, proteins are usually exposed to D_2O for different periods of time, depending on the experimental design. Sample aliquots are removed and quenched at low pH and low temperature so as to minimize further exchange and to "lock in" the deuterium that is already incorporated. In order to achieve amino acid sequence resolution, the sample is digested by an acid-stable enzyme (usually pepsin) before subjected to mass spectrometry (Cravello et al, 2003). In MALDI-MS (matrix-assisted laser desorption / ionization), samples are directly subjected to MS after proteolysis, while in the case of ESI-MS (electrospray ionization), one more step of liquid chromatography (usually HPLC, high-performance liquid chromatography) is recommended as it increases resolution significantly. The data collected allow us to analyze the change in the mass to charge ratio (m/z) of individual peptides as a function of time. As a result, a greater change in m/z over a designated period of time

indicates that these residues might be located in a less ordered region with higher level of solvent exposure (flexible region). Those with less HD exchange are supposed to be located in a well-structured region (Domon and Aebersold, 2006). Currently HDX-MS has become a key technique in monitoring the structure and dynamics of various proteins in solution (Figure 1.9).

Limitations of HDX-MS to probe protein intrinsically disordered regions are comprised of two major parts. First, HDX-MS usually has inadequate spatial resolution. The major advance in obtaining spatial resolution lies within proteolysis followed by liquid phase chromatography; however we are not able to control or manipulate how the enzyme digests our protein target. Although in some cases, the spatial resolution of HDX-MS can be enhanced up to a level of single residue by analyzing the HDX pattern of overlapping proteolytic fragments; it can never be generalized throughout the protein sequence (Der Mar, et al, 2005). In addition to limited spatial resolution, HDX-MS often suffers from incomplete sequence coverage as well, especially when applied to large and extensively glycosylated proteins. Proteins with multiple disulfide bonds constitute another group of targets that is difficult to attain adequate sequence coverage (Burke, et al, 2008).

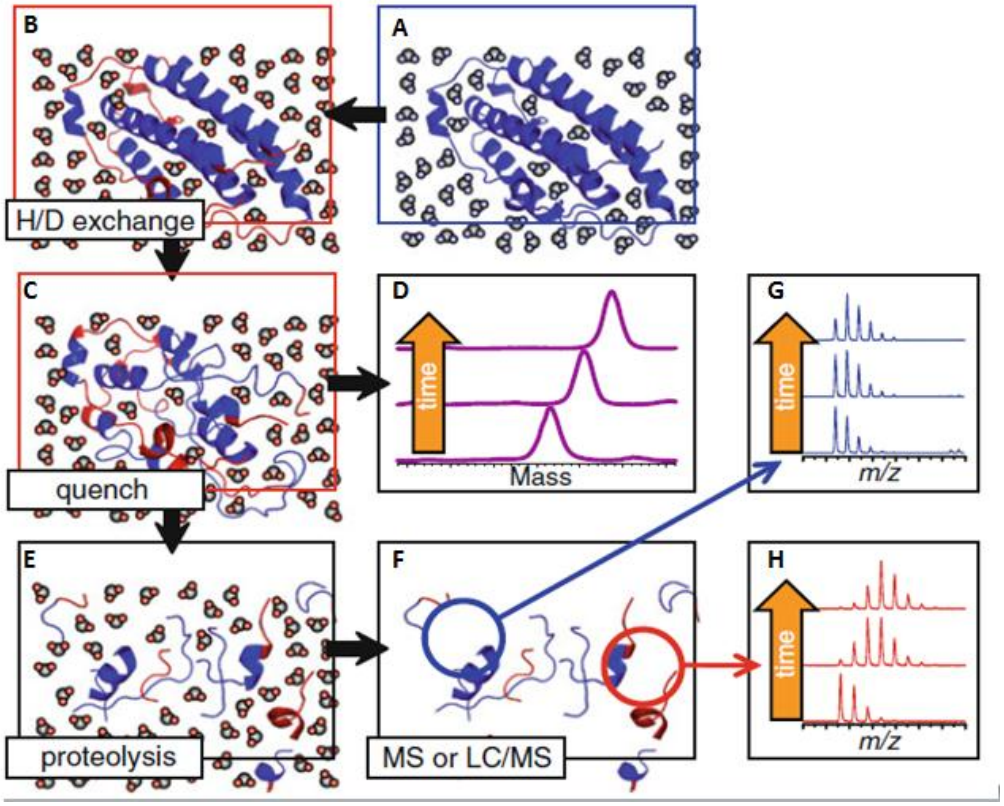


Figure 1.9, Schematic representation of HDX-MS work flow to examine protein flexibility. Panel A, protein sample exposed in aqueous buffer; Panel B, protein sample exposed in 100% deuterium oxide; Panel C, H/D exchange occurs at the backbone amide of the protein followed by quenching, incorporated deuterons are "locked"; Panel D, the change in mass through time is monitored by mass spectrometry; Panel E, to further increase spatial resolution, proteins undergo proteolysis (usually by pepsin); Panel F, protein fragments are separated by liquid chromatography followed by mass spectrometry; Panel G, a peptide fragment undergoing less H/D exchange; Panel H, a peptide fragment undergoing extensive H/D exchange. (Robst C. E. and Kaltashov I. A. *Methods Mol Biol*. 2012)

1.2.2.3 Protein dynamics by NMR

Nuclear magnetic resonance (NMR) is a widely used technique in studying protein structure and protein dynamics. Protein dynamics can occur at different time scales ranging from pico-seconds and hours, to days and even months. NMR is especially powerful in studying protein dynamics in the sense that the experiment design can be tailored to various time scales according to the mode of motion one wants to investigate (Figure 1.10).

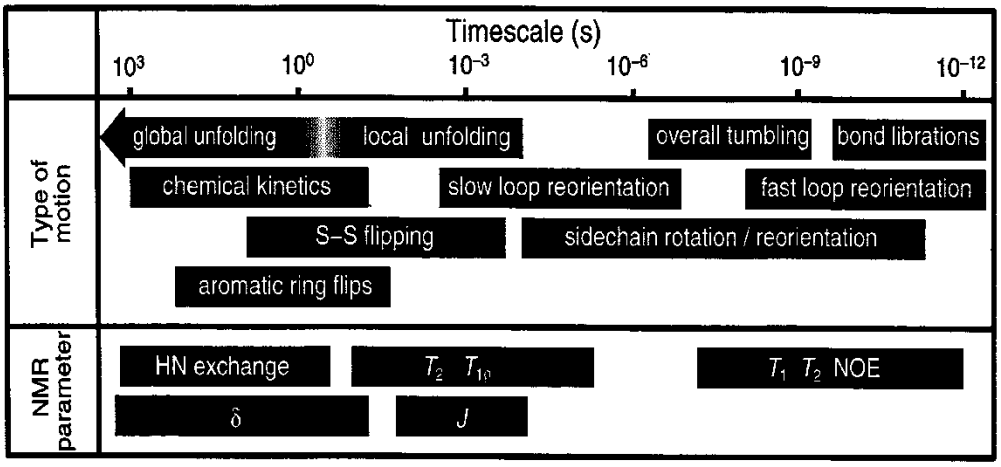
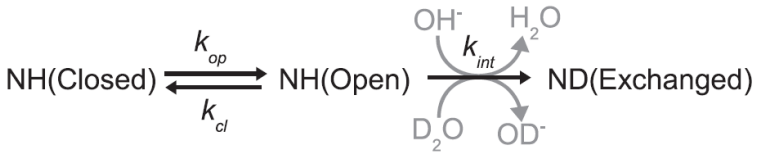


Figure 1.10, Typical time scales of molecular motion and the related NMR parameters that describes such motion.

Dynamic processes on the ~second time scale can be directly measured as a function of time. The NMR signal, or signal intensities are detected together with quantified time-correlation, the so-called "real-time" NMR (RT-NMR) (Ian R, et al, 2011). Molecular motion processes on this time scale will include protein folding, solvent hydrogen exchange and relatively slow conformational changes such as *cis-trans* isomerization of proline residues. The series of spectra acquired (in a time-dependent manner) often demonstrate a progressive weakening of an initial set of NMR signals with or without a progressive strengthening of a new set of NMR signals.

RT-NMR has been widely used in monitoring the hydrogen-deuterium exchange (HDX) of protein backbone amide protons. HDX-NMR utilizes the experiments, usually HSQC, in which ^{1}H yields visible signals, but ^{2}H can not. As the exchangeable amide protons are replaced by the solvent ^{2}H (deuterium), the observable ^{1}H signal intensity decays over time. This experiment is typically useful if provided with a complete set of backbone assignments correlated to the HSQC. The data of HDX-NMR is interpreted using the Linderstorm-Lang model (Hvidt and Nielson, 1966).



One of the experiments to study protein backbone dynamics is the measurement of backbone ^{1}H-^{15}N heteronuclear NOE. This experiment provides information about the motion of individual N-H bond vectors. Those that undergo motion faster than the overall tumbling of the molecules, in other words, has a higher level of backbone flexibility, will show a decreased NOE intensity relative to the average observed for the majority of the residues. Therefore as an example, decreased NOE intensities are usually observed at both N- and C-terminal ends of a protein where they do not form steady structures. In addition, spin-lattice relaxation (T_1), spin-spin relaxation (T_2) and spin-lattice relaxation in the rotating frame ($T_1\rho$) also provide important information on the time scale of motion for individual residues that show up on ^{1}H-^{15}N HSQC spectroscopy. Typically, a set of experiments comprising ^{15}N-edited T_1, ^{15}N-edited T_2 (usually replaced by ^{15}N-edited $T_1\rho$) and ^{1}H-^{15}N heteronuclear NOE data are usually determined by modified 2D HSQC-like NMR experiments on ^{15}N-edited proteins. In general, shorter relaxation time implies a faster motion of individual residues.

Another important parameter that can also reflect protein motion is the linewidth of the observable NMR signal. Under normal circumstances, the linewidth is defined to be the NMR signal at its half-height ($W_{1/2}$) and can be related to T_2 by the equation $W_{1/2}=1/(\pi T_2)$. Linewidth can be a general indicator for the overall tumbling time of a protein molecule. It is also highly related to the protein's molecular weight. Typically, the larger the molecule is, the broader the linewidth is expected to be. A larger-than-expected linewidth value may indicate a multimeric formation of such molecules (sometimes aggregation) or protein unfolding.

Protein side-chain dynamics occur at time scales ranging from milliseconds to picoseconds which are often related to the thermodynamics of folding, molecular interaction or conformational entropy. Side chain methyl C-H bond vectors, N-H bond vectors in the carboxyamide groups of asparagine and glutamine, as well as the secondary amines of arginine and tryptophan are generally used to probe side-chain dynamics (Nikola T, et al, 2009). Most side chain dynamics on time scales of milliseconds to microseconds can be studied using ^{13}C relaxation dispersion. Protein internal side chain motions on even faster time scales (ranging from nanoseconds to picoseconds) can be measured using auto- and cross-related relaxation as well as cross-relaxation (Yang D, 2011).

1.3 An introduction to *Aeromonas hydrophila*

Aeromonas hydrophila is a gram-negative bacterium that is oxidase-positive, anaerobic, and rod-like in shape with polar flagella. It can be found in most aquatic environments around the world (Albert M et al, 2000). *A. hydrophila* is one of the major pathogens to fish, causing motile aeromonas septicemia. It is also pathogenic to humans as well. Clinic symptoms associated with *aeromonas* infection includes gastroenteritis, wound infection and other systematic diseases (Austin B et al, 1996). *Aeromonas hydrophila* is resistant to many common antibiotics like penicillin and ampicillin. It is also able to grow at low temperatures while thriving at higher temperatures between 25℃ to 37℃. Although antibiotics like tetracycline, chloramphenicol and florenicol have been used to control aeromonas infections, concerns about drug resistance have always remained an issue.

1.3.1 Translocator operon of *Aeromonas hydrophila*

Similar to other gram-negative bacteria, the presence of a functional T3SS is essential for the pathogenesis of *Aeromonas. hydrophila*. Two kinds of proteins are secreted by the T3SS injectisome, effectors and translocators. Effector proteins are

usually delivered directly into the host cell cytosol where they will carry out specific functions such as rearrangement of host cell cytoskeleton, apoptosis of host cell, repressing host cell immune system that promote bacteria entry and survival (Zaharik ML, et al, 2002). In order for the effectors to be directly injected into the host cell, three proteins are needed to serve this purpose, two of which are hydrophobic translocators and the third is hydrophilic. A well-studied model system is the translocator operon of the *Yersinia* species (see Figure 1.11). Within the translocator operon, YopB and YopD are the two hydrophobic translocators while LcrV is the hydrophilic one. YopB and YopD were found to be inserted into host cell membrane, forming a pore that allows the transport of effector proteins directly into host cell cytosol (Neyt and Cornelis, 1999; Goure et al, 2005). The insertion of hydrophobic translocators into the host cell membrane is dependent on the hydrophilic translocator LcrV; although LcrV itself is not inserted. In the same operon, *sycD/lcrH* encodes a Class II chaperone for both YopB and YopD. Similar translocator operons were also found in *Pseudomonas* species.

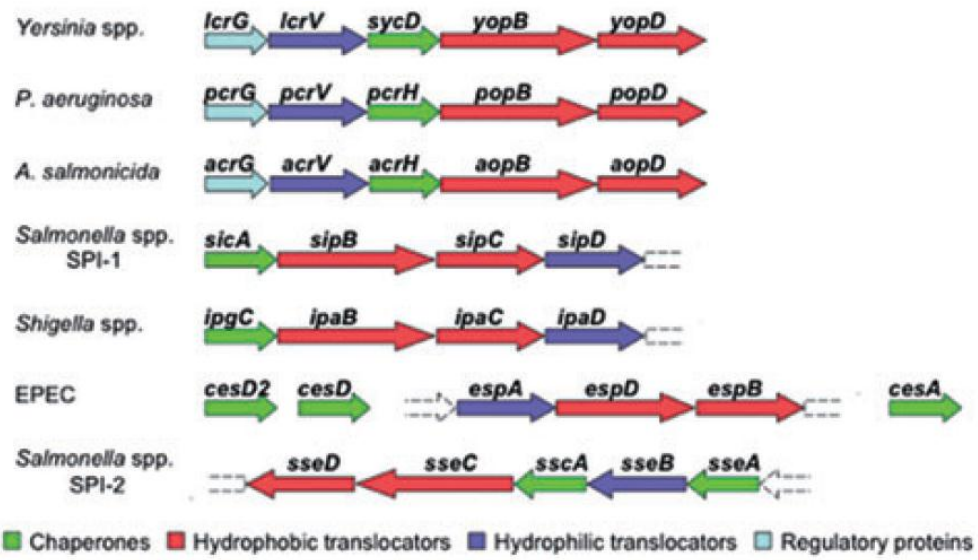


Figure 1.11, The translocator operon in different species. Translocators are coloured red; Class II chaperones are coloured green; hydrophilic protein is coloured blue. (Mueller C. A. et al, *Mol Microbiol*. 2008)

A. hydrophila has a T3SS with a translocator operon similar to the one in *Yersinia*. AopB and AopD are two hydrophobic translocators that are homologous to YopB and YopD. Inactivation of these two translocators results in decreased cytotoxicity in carp epithelial cells, increased phagocytosis and reduced virulence towards blue gourami fish (Yu HB et al, 2004). In the same operon, the gene *acrH* encodes the Class II chaperone AcrH which has been found to be able to form protein complexes with both AopB and AopD (Yih Wan Tan, et al, 2009). The crystal structure of the chaperone protein LcrH/SycD in *Yersinia* has been solved. LcrH is a pure α-helical protein with tetratricopeptide repeats (Carina B, et al, 2007). It interacts with YopB and YopD using two distinct regions. YopB binds to the concave binding groove while YopD binds to the outer convex (Edqvist PJ et al, 2006). AcrH was found to form complexes with either AopB or AopD in a one to one ratio, while the three of them can form a larger complex that is metastable (Yih Wan Tan, et al, 2009). The AcrH-binding regions of AopB and AopD have also been mapped respectively by limited protease digestion. AopD binds AcrH using two discrete regions: DF1, from Val16 to Leu147 and DF2 from Ser242 to Phe296 (digested by chymotrypsin). AopB binds AcrH through one continuous region, from Ala33 to Leu264 (digested by elastase) (Figure 1.12). The C-terminal part of AopB, a putative coiled-coil domain, is not involved in binding with its chaperone. However it is proposed to be responsible for protein complex oligomerization (Yih Wan Tan, et al, 2009). Although the structure of Class II chaperone has been solved and we have some insight into its interaction stoichiometry with the translocators; no complex structure formed between Class II chaperone and the translocator has been reported thus far. The detailed mechanism of interaction still remains unclear.

Figure 1.12, Sequence alignment of AopB from *Aeromonas hydrophila* with YopB from *Yersinia enterocolitica* and PopB from *Pseudomonas aeruginosa*. The AcrH-binding region of AcrH was boxed. Predicted trans-membrane domains were highlighted in grey. Coilded coil regions were marked in bold.

1.4 General objective and significance of this study

Type VI secretion system (T6SS) is a recently discovered secretion system which plays an important role in the pathogenesis of various gram-negative bacteria. *Edwardsiella tarda*, being one of the major fish pathogens, was also found to possess a functional T6SS. Within the gene cluster of T6SS in *E. tarda*, EvpC has been demonstrated to be an Hcp homologue and its structure was determined to be similar to Hcp1 (Jobichen C et al, 2010). Other than this, little is known about the structure or topology of the needle complex. EvpO is proposed to be an ATPase that provides energy to drive Type 6 secretion. EvpA and EvpB share homology with VipA and VipB, which are found in *Vibrio cholerae*. It has been reported that ClpV, a potential ATPase protein in *Vibrio cholerae*, interacts with VipA and VipB through its N-terminal domain while VipA/VipB has been shown to form tubule-like structures (Gabriele et al, 2009). Yet, no direct evidence of such EvpA and EvpB interaction has been reported.

More importantly, T6SS gene cluster of *E. tarda* harbours one unique gene, *evpP*, which is not found in any other gram-negative bacteria that possess a T6SS. EvpP is a novel secreted protein by the T6SS that contributes to the virulence of *E. tarda*. Recent studies at the genetic level have gained some insights to the functional potentials of EvpP (Xin Wang et al, 2009). Relatively little is understood about EvpP at the protein level: the structure as well as the interacting partners of EvpP still

remain unknown. Therefore our primary research objective was to characterize the properties of EvpP protein. We attempted to investigate the structure of EvpP in terms of mapping its rigid and flexible regions by obtaining its NMR backbone chemical shift assignment, amide proton exchange mass spectrometry as well as limited protease digestion. In addition, we attempted to examine its potential interacting partners within the T6SS gene cluster, as EvpP has been reported to be able to interact with EvpC (Zheng and Leung, 2007).

One fascinating feature about T3SS and T6SS is their ability to directly inject effectors from bacterial cytoplasm into host cell cytosol where various effectors can interact with host cell entities. Within T3SS of *Aeromonas hydrophila*, the pore-forming translocators AopB and AopD were found to be inserted into the host cell membrane where they polymerize to form a pore that is continuous from the injectisome of T3SS, allowing the direct transportation of effectors. As AopB and AopD are highly hydrophobic proteins, they have to be kept stable (perhaps in their nascent form) by the Class II chaperone AcrH before folding into mature form to become functionally active. AcrH shares a high homology with LcrH/SycD from *Yersinia* species in terms of secondary and tertiary structure. Similar to SycD, AcrH is also believed to be an α-helical protein with tetratricopeptide repeats (Pallen et al, 2003). Although AcrH has been shown to interact with either AopB or AopD in a one to one ratio, and the AcrH-binding regions on AopB and AopD have been mapped (Yih Wan Tan, et al, 2009), no structure of Class II chaperone in complex with a translocator has been reported yet. Therefore our primary research objective is to optimize the boundaries of the translocator-chaperone complex of AcrH-AopB by further limited protease digestion. The foundation of this research was based on the results of previous studies (Y W Tan, H B Yu, J Sivaraman, K Y Leung and Y K Mok, *Protein Science*, 2009). We also attempted to crystalize the complex of AcrH-AopB in order to test whether the new crystals could yield better diffraction patterns.

CHAPTER 2, MATERIALS & METHODS

2.1 Subcloning of EvpP and its mutants into expression vector

2.1.1 Bacterial host strains

Strain	Genotye
DH5-α	F^- Φ80d*lacZ*ΔM15 &(*lacZ*YA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^- ,m_k^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1
BL21-DE3	F^- *omp*T *hsd*S_B($r_B^-m_B^-$) *gal dcm*

Table 2.1, Strains used for protein subcloning and expression

2.1.2 Generation of DNA insert and Polymerase Chain Reaction

DNA insert of EvpP as well as its truncation mutants (see table 2.2 for the primer sequences used) were generated by polymerase chain reaction (PCR), using upstream oligonucleotide primer with a BamH I restriction site and downstream oligonucleotide primer with a stop codon and an EcoR I restriction site. PCR reaction mix was prepared in 50 µl mix containing 1 X Pfu buffer (20 mM Tris-HCl, pH 7.5 at 25 ℃, 8 mM $MgCl_2$, 0.5 mM DTT, 50 mg/ml BSA), 1 µl of *E. tarda* strain 130/91 genomic DNA template, 0.4 µM of each primers, 200 µM of each dATP, dCTP, dGTP and dCTP, and 1 µl of 1 U KOD Hot Start Pfu DNA Polymerase (Novagen).

PCR was carried out for 30 to 35 cycles, following the subsequent protocol: 95 ℃ of denaturation for 30 sec, 55 ℃ of primer annealing for 30 sec (usually this temperature is adjustable according to the melting temperature of the primer), 72 ℃ of extention for 30 sec (this is also adjustable according to the length of the PCR product). The PCR product was mixed with 1X DNA loading dye and analysed on 1% agarose gel electrophoresis. The desired band was recovered by Gel Extraction Kit (QIAGEN) following the standard manufacturer's protocol.

Table 2.2, Table of PCR Primers for EvpP and its Truncation Mutants. FP: forward primer; RP: reverse primer

EvpP21-endFP
5' CGG GAT CCG GCA CCG TCG TCA GTG 3'
EvpP21-endRP

5' CGG AAT TCT TAT TTC AAG GCT GAA AAG 3'
EvpP76-endFP
5' CGG GAT CCT TCC GCC ACG ATG GC 3'
EvpPFP
5' CGG GAT CCA TGT CGA TTT TGA ATT C 3'
EvpP1-74RP
5' CCC AAG CTT CTG GAC CGG GCT CCA C 3'
EvpP1-86RP
5' CCC AAG CTT CAT ACG CTG CTG CAG 3'
EvpP1-88RP
5' CCC AAG CTT GAC CAG CAT ACG CTG 3'
EvpP1-92RP
5' CCC AAG CTT GTT TAG CTT AGC GAC 3'
EvpP1-123RP
5' CCC AAG CTT TCA AAC CTT TAG GAC TAT C 3'
EvpP1-142RP
5' CCC AAG CTT TCA CAT ACC ATT ATC AAT AC 3'
EvpP1-159RP
5' CCC AAG CTT TCA GAG GTT CGA TCC CC 3'
EvpP1-168RP
5' CCC AAG CTT TCA AAT ATA GAA CTG TGT G 3'

2.1.3 Generation of mutant DNA insert for site directed mutagenesis

Mutant DNA inserts of EvpP were generated by employing the modified QuikChange™ Site-Directed Mutagenesis Kit (Stratagene) designed oligonucleotide primers (1st Base) with mismatches. Primers were designed with mismatches base pairing being flanked on both 5' and 3' ends with about 9 to 15 bases each using the software GeneRunner. The list of mismatched primers for EvpP could be seen in Table 2.3. The PCR using the high fidelity KOD Hot Start Pfu DNA Polymerase (Novagen) generated the whole plasmid consisting of the desired mismatched bases. The methylated wild-type plasmids were digested by adding 1 µl of 1 U Dpn 1 Fast Digest™ restriction enzyme at 37℃ for 1 hour. The mixture was subsequently purified using the spin columns (QIAGEN) followed by direct transformation into BL21 (DE3) competent cells. The list of primers used is shown in Table 2.2. PCR amplification was carried out with similar procedure as described in section 2.1.2 except that the elongation time was adjusted to 5 min. A flow chart of this method could be seen in Figure 2.1.

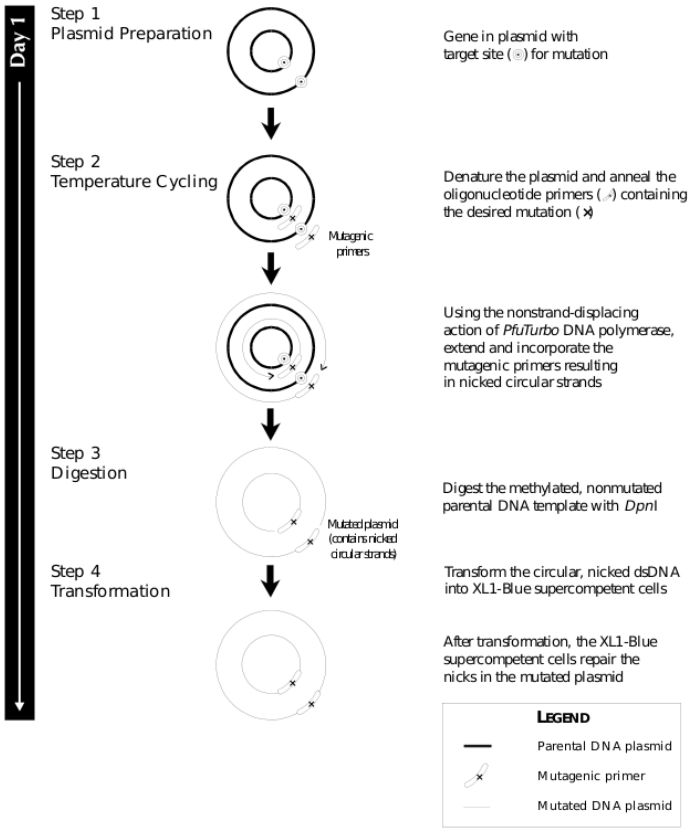


Figure 2.1, Generation of site-directed mutants. The high fidelity Hot Start KOD DNA Polymerase enabled the amplification of the DNA insert together with the vector. Dpn 1 restriction enzyme was used to digest the methylated wild-type parental DNA template. Finally, the nicked dsDNA was repaired in the BL21 (DE3) competent cells.

EvpP P20A FP
5' GT ATT TCT ACA GGC GCA GGC ACC GTC GTC AG 3'
EvpP P20A RP
5' CT GAC GAC GGT GCC TGC GCC TGT AGA AAT AC 3'

EvpP P58S FP
5' GTT CAT TGG CAA GGT CCC CGA TCA CTG ACG ATA C 3'
EvpP P58S RP
5' GTA TCG TCA GTG ATC GGG GAC CTT GCC AAT GAA C 3'

EvpP L91Q FP
5' GCT GGT CGC TAA GCA AAA CGG CGT ATC AAT G 3'
EvpP L91Q RP
5' CAT TGA TAC GCC GTT TTG CTT AGC GAC CAG C 3'

EvpP P143T FP
5' GTA TTG ATA ATG GTA TGA CCT CTA AAC GAA ATA TC 3'
EvpP P143T RP
5' GAT ATT TCG TTT AGA GGT CAT ACC ATT ATC AAT AC 3'

EvpP G156S FP
5' GAA CAA AAA GGC AGG TCA TCG AAC CTC CCC GGC 3'
EvpP G156S RP
5' GCC GGG GAG GTT CGA TGA CCT GCC TTT TTG TTC 3'

Table 2.3, List of primers used to generate mutant DNA inserts of EvpP. FP: forward primer; RP: reverse primer.

2.1.4 Preparation of competent cells

Frozen glycerol stock of DH5α or BL21 (DE3) was streaked onto plain LB agar plate and incubated overnight at 37 ℃. A single colony was inoculated into 8 ml of LB broth and incubated in a shaker overnight at 37 ℃. 1 ml of overnight culture was transferred into 100 ml of LB broth and incubated in a shaker at 37 ℃ with shaking speed of 200 rpm. Cells were grown until $O.D_{600}$ reached 0.4 and then transferred into 50 ml tubes and spun at 3,500 rpm for 10 minutes at 4 ℃. Supernatant was removed and cell pellets were resuspended gently with 40 ml of cold glycerol buffer (0.1 M $CaCl_2$, 15% glycerol) and then incubated on ice for 30 minutes. Cells were spun at 3,500 rpm for another 10 minutes at 4 ℃ and the supernatant was removed. The pellet was resuspended carefully in 4 ml of cold glycerol buffer and incubated on ice for 2 hours before aliquoted at 50 µl per tube. The aliquots were quickly freezed in liquid nitrogen before transferring to storage at -80 ℃.

2.1.5 Subcloning

For His-tagged proteins, DNA inserts of EvpP and its truncation mutants (Table 2.1) were subcloned into pET-M, a modified vector from pET-32a with S-tag and thioredoxin tag removed. For GST-tagged proteins, the inserts were subcloned into pGEX-4T-1 vector (Novagen). The DNA insert (about 50 ng/µl) was digested using 10 U of BamH I Fast Digest ™ and EcoR 1 Fast Digest ™ (both from Fermentes) in a total reaction volume of 40 µl. Two microliter of expression vector (about 50 ng/µl) was digested in a separate tube with another 10 U of the similar restriction enzyme in total reaction volume of 20 µl. Concentrated 10 X Fast Digest ™ buffer was added accordingly. Both tubes were incubated in 37 ℃ water bath for at least 1 hour.

The plasmid and DNA insert digestion products were mixed together and the mixture was purified using the spin column (QIAGEN) with a final elution volume of

35 ul. The ligation of DNA insert and plasmid was done by adding 4 µl of 10 X T4 ligase buffer and 1 µl of T4 ligase enzyme (Fermentes). Ligation was carried out for at least 1.5 hour at room temperature before transformation into DH5-α competent cell.

2.1.6 Transformation of ligation mix into DH5-α competent cell

Ligation reaction was added into a tube containing 50 µl of DH5-α competent cell and kept on ice for 30 minutes. Heat shock was carried out by incubating the cells at 42 ℃ for exactly 90 seconds, followed by 2 minutes incubation on ice. 900 µl of LB broth was added into the tube before the recovery incubation at 37 ℃ with shaking for one hour. Cells were centrifuged at 6,000 rpm for 2 minutes and 900 µl of the supernatant were removed. Cell pellet was resuspended with the remaining medium and plated on LB agar plates supplemented with 0.1 mg/ml ampicillin. Plates were incubated overnight in a 37 ℃ incubator.

2.1.7 PCR screening of transformant

The presence of DNA insert inside target plasmid was verified by PCR screening of colonies grown on LB-ampicillin agar plate prior to DNA sequencing. At least six colonies from each plate were selected for screening. Half of a colony was picked using a sterile pipette tip and inoculated into 100 µl of sterile water in an eppendorf tube. The suspension was boiled (or heated on a heat block) for five minutes and subsequently centrifuged at 13,000 rpm for a minute. One microliter of the supernatant could be used as the DNA template for PCR reaction.

The PCR reaction cocktail of 10 µl was prepared containing 1 X Taq reaction buffer (50 mM KCl, 10 mM Tris-Cl pH 9.0, 1% Triton X-100), 0.75 mM $MgCl_2$, 140 µM of each dATP, dCTP, dGTP and dCTP, 0.5 µM each of forward and reverse primer; and 1.5 U of Taq DNA Polymerase (Promega). One microliter of DNA template was added into the mixture and 30 to 35 PCR cycles were carried out with 1

minute denaturation at 94 ℃, 1 minute annealing at 52 ℃ and 1 minute elongation at 72 ℃ for each cycle. The PCR products were analyzed on a 1% agarose gel for the detection of DNA fragment of appropriate size.

2.1.8 Isolation of DNA plasmid

Colonies that showed positive PCR screening results were inoculated into LB broth (containing 100 μg/ml ampicillin) and cultured overnight at 37 ℃ with shaking. Cell pellets from 5 ml of overnight culture were recovered by centrifugation at 6,000 rpm for 15 minutes. DNA plasmid was extracted from pelleted cells using QIAGEN Miniprep Kit (QIAGEN) and eluted with 30 μl of elution buffer (QIAGEN) or water according to the supplied protocol.

2.1.9 DNA sequencing

Two microliter of plasmid DNA (about 250 ng of DNA) was added into a reaction mix containing 3 μl of water, 1 μl of T7-promoter sequencing primer (3.2 μM), 2 μl of 5X Big Dye Buffer and 2 μl of Big Dye Terminator V3.1 (Applied Biosystems). Thirty cycles of sequencing PCR reaction were carried out with denaturation step at 96 ℃ for 30 seconds, annealing at 50 ℃ for 25 seconds and extension at 60 ℃ for 4 minutes. Following the sequencing cycle, a mixture of 3 μl of 3M sodium acetate, pH 4.6, 62.5 μl of 95 % ethanol and 24.5 μl of sterile water was added into each DNA sample for precipitation. After 1 hour (or at least 15 minutes) incubation at room temperature, the mixture was centrifuged at 13,000 rpm for 20 minutes at room temperature. Supernatant was removed, and the pellet was washed twice with 500 μl of 70% ethanol and centrifuged for another 5 minutes during each wash. Supernatant was carefully removed and the DNA pellet was dried in a vacuum centrifuge (Speed-Vec) before sequencing analysis on an automated sequencer, ABI PRISM 3100 (Applied Biosystems).

2.2 EvpP protein expression and purification

2.2.1 Transformation of plasmid into BL21 (DE3) competent cell

Plasmid containing DNA insert or mutants was transformed into E. coli BL21 (DE3) cells for the expression of recombinant protein. Transformation procedure was carried out in the same way as described in section 2.1.6.

2.2.2 Protein expression

A single colony of *E. coli* BL21 cells containing the desired plasmid was inoculated into 10 ml of LB broth (containing 100 μg/ml ampicillin) and cultured at 37 ℃ overnight with shaking. Ten milliliter of overnight culture was inoculated into 1 liter of autoclaved LB broth (containing 0.1 mg/ml ampicillin). Cultures were grown at 37 ℃ with shaking at a speed of 200 rpm for approximately 3 hours or until the $O.D_{600}$ reached approximately 0.6. Subsequently, IPTG was added at a final concentration of 0.4 mM to induce protein expression. The culture was further grown at 20 ℃ to 25 ℃ overnight at 200 rpm. Cells were harvested by centrifugation at 6,000 rpm for 15 minutes at 4 ℃. Cell pellets were kept at -20 ℃ until purification.

Expression of ^{15}N or ^{13}C-labeled protein was carried out with similar conditions as mentioned above, except that the cells were grown in M9 minimal medium (Appendix I) supplemented with ^{15}N-labeled ammonium chloride and/or ^{13}C-labeled D-glucose. To obtain a single labelled sample (either ^{15}N-labeled or ^{13}C-labeled sample), cells were grown in 1 X M9 minimal medium in the presence of ^{15}N-ammonium chloride (Cambridge Isotope Laboratories) or with ^{13}C-glucose (Cambridge Isotope Laboratories) as the sole nitrogen or carbon source. For double labeled ^{15}N and ^{13}C-sample, 1 X M9 minimal medium with ^{15}N-ammonium chloride and ^{13}C-glucose (Cambridge Isotope Laboratories) was used as the sole nitrogen and carbon source respectively.

2.2.3 Protein purification using nickel-affinity chromatography

Cells from one liter culture were resuspended in 30 ml of nickel binding buffer (20 mM Tris-Cl pH 8, 0.5 M NaCl, 5 mM imidazole). The suspension was sonicated for 3 minute each time at 38% sonication amplitude for 4 to 5 times. The lysate was centrifuged at 4℃ for 30 minutes at 18,000 rpm. The supernatant was collected for purification. Around 10 ml of Ni-NTA beads (QIAGEN) were charged with 20 ml of Nickel-charge buffer (50 mM $NiSO_4$), washed with 3 colume volume of water and equilibrated with 1 X Nickel binding buffer in a chromatography column. All supernatant was slowly loaded into the column, mixed well and let flow at approximately 1 ml/min. The column was subsequently washed 15 times (or at least 10 times) with 1 X nickel washing (20 mM Tris-Cl pH 8, 0.5 M NaCl, 20 mM imidazole) buffer to remove any unbound or loosely bound proteins. His-tagged protein was eluted out using 30 ml of nickel elution buffer (20 mM Tris-Cl pH 8, 0.5 M NaCl and 500 mM imidazole). The eluted protein was dialyzed overnight against 20 mM Tris-Cl pH 7 and analyzed by SDS-PAGE.

2.2.4 Protein refolding by rapid dilution

As all the EvpP truncation mutants were insoluble and expressed in inclusion bodies; they were subjected to protein refolding in order to explore the possibility to get any soluble form of EvpP truncation mutants. His-tagged EvpP truncation mutants (Table 2.3) were purified in the same way as described in section 2.2.3 except that all the buffers used contain additional 8 M Urea (or 6 M guanidine hydrochloride, refer to Appendix II for buffer recipe). For refolding, 100 µl of eluted protein (denatured in 8 M Urea) was added into 3 ml refolding buffer (usually Tris pH7 or pH8, phosphate buffer pH6, MES buffer pH5 or Acetate buffer pH4) drop by drop with rapid stirring on a electric magnetic stirrer. The same procedure was carried out repeatedly for 100 times or until sufficient amount of denatured protein being refolded. The protein solution (in refolding buffer) was pooled together and concentrated to around 50 ml in volume and subjected to dialysis for further removal

of residual Urea or Guanidine Hydrochloride. The dialyzed protein solution was centrifuged at 6,000 rpm for 20 min and the supernatant was collected and analyzed on SDS-PAGE or CD.

2.2.5 Protein purification using GST affinity chromatography

Cells expressing EvpP (wild type as well as its truncation or site-directed mutants) from 1 liter culture were resuspended in 30 ml of PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4, pH7.4). The suspension was sonicated for 3 minute each time at 38% sonication amplitude for 4 to 5 times. The sonication lysate was centrifuged for 30 minutes at 18,000 rpm and 4℃. The supernatant was collected for purification. All supernatant was slowly loaded into a column packed with 10 ml of glutathione-sepharose 4B beads (Amersham Biosciences) equilibrated with 1 X PBS. The column was subsequently washed 15 times with a total of 1 X PBS to remove any unbound or loosely bound proteins. GST-tagged protein was eluted with 30 ml of glutathione elution buffer (50 mM Tris-Cl pH 8 and 10 mM reduced glutathione) and analyzed by SDS-PAGE. Glutathione-sepharose beads were regenerated by washing the column twice with column regeneration buffer 1 & 2 (Appendix II).

2.2.6 Gel filtration FPLC (Fast Performance Liquid Chromatography)

Gel filtration, FPLC was carried out through a HiLoad 16/60 Superdex 75 (or Superdex 200, in case of purification of AcrH-AopB complex) pg size exclusion chromatography column on AKTA FPLC System (Amersham Biosciences) using 20 mM Tris-Cl pH7 with 200 mM NaCl as running buffer. Gel filtration column was pre-equilibrated with at least three column volumes (180 ml) of running buffer before usage. Protein sample was concentrated to around 5 ml prior to loading. Chromatography was carried out at a flow rate of 1 ml/min for a total volume of 180

ml. Eluted protein peaks were detected by measuring UV absorbance at 280 nm and the respective fractions were analyzed by SDS-PAGE.

2.2.7 Ion exchange chromatography

For the purification of trypsin digested EvpP (wild type), ion exchange chromatography was carried out through a Mono Q^{TM} 10/100 GL (Amersham Biosciences) column on AKTA system (Amersham Biosciences) using 20 mM Tris-Cl pH9 as buffer A and 20 mM Tris-Cl pH9, 1 M NaCl as buffer B. The column was pre-equilibrated with at least 3 column volumes of buffer B followed by 6 column volumes of buffer A before loading of protein samples. Protein samples were concentrated to around 2 ml prior to loading. Chromatography was carried out at a flow rate of 4 ml/min until target concentration of buffer B was reached (usually 100% of buffer B was used as final concentration). Eluted protein peaks were detected by measuring UV absorbance at 280 nm and the respective fractions were analyzed by SDS-PAGE.

2.2.8 Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was carried out according to the method described by Laemmli (1970). Polyacrylamide gel was cast according to the table described in Appendix III using the Mini-PROTEAN 3 system (Bio-Rad). Protein sample was mixed with an equal volume of 2 X SDS sample buffer and boiled for 5 minutes before loading into the wells of the polyacrylamide gel. Fifteen microliter of sample was loaded into each well, and electrophoresis was carried out at 200 V for 50 min or until the dye front reached the end of the gel. The gel was stained with Coomassie blue staining solution for about 20 minutes and then destained in destaining solution (Appendix III) overnight.

2.3 Dynamic light scattering (DLS)

Dynamic light scattering was carried out for both EvpP and its mutant EvpP_P143T using DynaPro Dynamic Light Scattering (ProteinSolutions) machine linked to a Temperature Controlled MicroSampler (ProteinSolutions). Twenty microliter of protein sample was loaded into a quartz cuvette with a path length of 15 mm (ProteinSolutions). Data were recorded and analyzed using the Instrument Control Software for Molecular Research (ProteinSolutions).

2.4 Limited protease digestion of EvpP

Limited protease digestion (enzymes used were trypsin, thrombin, chymotrypsin, elastase, thermolysin and papain) was carried out for both EvpP and its mutant EvpP_P143T using protein samples in its native buffer (20 mM Tris-Cl pH7) with different enzymes at different ratio (usually 1:50, 1:100 or 1:200). Both the protein sample and the enzyme stock were prepared to be 1 mg/ml in concentration. Ten microliter of enzyme solution was added into 1 ml of protein sample to form the reaction mixture. Aliquots were removed from the reaction mixture every 10 min and mixed with equal volume of 2 X SDS sample buffer and boiled at 100 ℃ for 5 min. All the aliquots were analyzed on SDS-PAGE subsequently.

2.5 Protein cross-linking assay by glutaraldehyde

For glutaraldehyde treatment of wild type EvpP, the buffer need to be changed to 20 mM phosphate buffer pH7.0. One hundred microliter of protein sample (at a concentration of 1 mg/ml) was treated by 5 μl of freshly prepared 2.3% solution of glutaraldehyde for 10 min at 37 ℃. The reaction was terminated by the addition of 10 μl Tris-Cl pH8.0. Treated protein samples were mixed with equal volume of 2 X SDS sample buffer and boiled for 5 min before they were analyzed on SDS-PAGE.

2.6 Mass spectrometry

Trypsin digested EvpP samples were blotted onto a MALDI target plate with sinapinic acid as a matrix. The molecular weight of each component in the complex was determined using Voyager-DE STR Mass Spectrometer (GE Healthcare). A voltage of 2100V and laser intensity of 2500 was used for the experiment.

2.7 Protein sequencing (N-terminal amino acid analysis)

For N-terminal sequencing studies, EvpP proteins (after limited proteolysis by various enzymes, refer to Section 2.4 for the enzymes used) need to be immobilized onto a membrane by western blot before sequencing. Protein samples were separated on a 15% SDS-PAGE or 15% Tricine-SDS-PAGE. The proteins separated on the gel were blotted onto a PVDF membrane (Milipore) using a Mini Trans-blot Electrophoretic Tranfer Cell (Bio-Rad). Trans-blotting was performed at 4°C at a voltage of 90V for 90 min or until complete transfer of protein bands onto the PVDF membrane. Protein bands on the membrane were visualized by coomassie blue staining and excess stain on the membrane was removed by washing with a 50% methanol solution. Membrane-bound protein bands were excised and N-terminal residues were fragmented using Procise Protein sequencing system (Applied Biosystems). Data were collected and analyzed using SequencePro Data Analysis Application software v2.1.

2.8 Circular dichroism (CD) spectroscopy

Circular dichroism experiments were carried out with 20 μM EvpP protein (or more) in 20 mM Tris-Cl pH7.0 buffer. The spectra were acquired with a J-810 Spectropolarimeter (Jasco, Japan) using a quartz cuvette with 1 mm pathlength (Hellma, Germany). The spectra were recorded at wavelength region from 190 nm to 260 nm at 0.1 nm resolutions with a scan speed of 50 nm/min and a response time of 8 seconds, averaged over 10 scans.

2.9 Urea denaturation experiments by fluorescence spectroscopy

Urea denaturation experiments were carried out by mixing samples of EvpP and its mutants (P20A, P58S, L91Q, P143T, G156S) at an initial concentration of 100 µM, and denaturing buffer (20 mM Tris-Cl with 10 M Urea) at different ratios to achieve Urea concentration from 0 M to 8 M with an increment of 0.5 M each. The resultant protein concentration was 20 µM.

Protein samples were loaded into a fluorescence cuvette with a path length of 10 mm (Perkin Elmer Instruments) and the fluorescence spectra were recorded using an LS55 Luminescence Spectrometer (Perkin Elmer Instruments). Data were analyzed using FL WINLAB software. Curve fitting was performed using Kaleidagraph ver4.0.

2.10 Sequence alignment

Amino acid sequence alignment of EvpP (130/91) against other species was carried out using ClustalW (Higgins et al, 1994) software hosted at European Bioinformatics Institute server (EBI,http://www.ebi.ac.uk/Tools/). Amino acid sequences were acquired from GenBank (http://www.ncbi.nlm.nih.gov/).

2.11 Nuclear magnetic resonance and backbone assignment

3D and 4D heteronuclear NMR experiments were carried out to assign the backbone chemical shifts of EvpP and its mutant EvpP_P143T using a 2D ^{1}H-^{15}N HSQC as a reference spectrum. All NMR experiments were carried out on a Bruker AVANCE 800 MHz spectrometer equipped with a cryoprobe. Samples were loaded into a 5 mm NMR tube and all experiments were carried out at 297K. All data acquired were processed using NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed using NMRDraw (Delaglio, Grzesiek et al. 1995), SPARKY (Goddard and Kneller) and NMRSpy (Zheng, Yu; Yang, Daiwen. NMRspy, National University of Singapore). Computational processing was carried out on Dell Precision 360 workstation and SGI Origin 300 server.

2.11.1 2D ^{1}H-^{15}N HSQC spectrum

2D ^{1}H-^{15}N HSQC spectrum was used as a reference for backbone chemical shifts assignments by other 3D or 4D experiments. The spectrum was acquired with 1280 complex points over a spectral width of 11160.714 Hz on ^{1}H dimension. In the ^{15}N dimension, 256 points were collected over 2108.370 Hz. Carrier frequencies for ^{1}H and ^{15}N were at 4.782 ppm and 119.079 ppm respectively.

2.11.2 HNCACB and CBCA(CO)NH of deuterated EvpP

As His-tagged EvpP is 22 kDa and it was thought to have a dimeric form in solution; normal HNCACB and CBCA(CO)NH could not get a good signal-to-noise ratio. The experiment had to be done using deuterated EvpP protein sample. Complete deuteration of EvpP was achieved by culturing EvpP-expressing E. coli cells in M9 minimal medium that was completely prepared using deuterium oxide (D_2O) instead of hydrogen oxide (H_2O). The deuterated proteins were then purified by the standard Ni-NTA affinity chromatography using all aqueous buffers so that the backbone amide protons (but not the others) could be exchanged back to hydrogen.

Backbone assignments were carried out by HNCACB and CBCA(CO)NH experiments. The CBCA(CO)NH experiment correlates both the Cα and Cβ chemical shift with the amide from the previous residue while HNCACB provides the intra-residue correlation of its own Cα and Cβ while often, the Cα and Cβ from the preceding residue could also been seen. These two experiments provided chemical shift information for sequential assignments.

For HNCACB experiment, in the acquired dimension ^{1}H, 1280 points were collected over 11160.714 Hz of spectral width. In the dimension ^{15}N, 100 points were collected over 2107.926 Hz of spectral width while the dimension ^{13}C contained 204 points covering 12468.828 Hz of spectral width. The carrier frequencies for ^{1}H, ^{15}N and ^{13}C dimensions were at 4.782 ppm, 119.081 ppm and 41.745 ppm respectively.

3D CBCA(CO)NH experiment (Grzesiek, Vuister et al. 1993) was carried out similarly using the same data points and spectral widths as above, except that the dimension ^{13}C was recorded with 172 points.

2.11.3 ^{15}N-edited NOESY

Amide proton distance constraints were derived from 3D ^{15}N-edited Nuclear Overhauser Effect spectroscopy (NOESY) with 100 ms mixing time. Each cross-peak in a ^{15}N-NOESY amide strip represents an NOE between amide protons to a nearby proton within around 5 Å distance. 1280 complex points were recorded for the ^{1}H dimension corresponding to the directly detected ^{1}H dimension over spectral width of 11160.714 Hz. For ^{15}N dimension, 104 complex points were detected over spectral width of 2108.370 Hz, while 232 complex points were recorded for spectral width of 9601.536 Hz for F3 dimension (^{1}H). Carrier frequencies for F1, F2 and F3 were at 4.725 ppm, 119.081 ppm and 4.725 ppm respectively.

2.11.4 Four dimensional (4D) ^{13}C, ^{15}N-edited NOESY of EvpP_P143T mutant

As most triple resonance experiments do not work for uniformly ^{15}N, ^{13}C-edited large proteins expect for Nuclear Overhauser Effect spectroscopy (NOESY) and multi-quantum ^{13}C total correlation spectroscopy (MQ-CCH-TOCSY); a new strategy has been developed for backbone as well as side chain assignment by utilizing a set of spectra: HSQC, HNCA, HN(CO)CA, 4D NEOSY and HCCH-TOCSY. Clusters are formed by matching HC-NH NOE and Cα-NH (on HNCA) correlations that have identical NH chemical shifts. Spin systems are identified by sorting out intra-residue and sequential HC-NH NOE correlations from other inter-residue NOEs that can be observed on the 4D ^{13}C-^{15}N-edtited NOESY spectrum with the help of HNCA and MQ-CCH-TOCSY. Clusters are linked to form amino acid residue chains by matching the intra-residue spin system of one cluster with the

sequential spin systems of other clusters. The fragments of amino acid chains can then be fitted into the protein sequence (Yingqi Xu, et al, 2006).

1280 complex points were recorded for the directly detected ^{1}H dimension over a sweep width of 11160.714 Hz. The indirect H dimension has 104 complex points recorded over a sweep width of 9398.496 Hz. In the ^{15}N dimension, 40 complex points were recorded over a sweep width of 2108.370 Hz. In the ^{13}C dimension, 40 complex points were recorded over a sweep width of 4426.737 Hz. Carrier frequencies of the four dimensions would be 4.782 ppm, 4.782 ppm, 119.082 ppm and 56.745 ppm respectively.

2.11.5 ^{13}C-edited MQ-CCH-TOCSY of EvpP_P143T mutant

3D HCCH-TOCSY experiment was utilized to identify side-chain spin system by connecting aliphatic ^{1}H resonances with their attached ^{13}C resonances (Wang and Zuiderweg 1995). The protein sample used in this experiment was prepared by exchanging ^{13}C-labeled protein with deuterated buffer at pD 7 in 99.9% D2O. A total of 1280 complex points were collected for dimension ^{1}H over 11160.714 Hz spectral width, while in directly detected ^{13}C dimension, 180 points over 13280.212 Hz spectral width were recorded. In the indirect ^{13}C dimension, 60 complex points were recorded over 4426.737 Hz of spectral width. The carrier frequencies were at 4.725 ppm, 41.742 ppm and 41.742 ppm respectively.

2.11.6 ^{15}N-edited NOESY of EvpP_P143T mutant

This experiment was carried out under the same condition to the one carried out for EvpP wild type protein. 1280 complex points were recorded for the ^{1}H dimension corresponding to the directly detected ^{1}H dimension over spectral width of 11160.714 Hz. For ^{15}N dimension, 104 complex points were detected over spectral width of 2108.370 Hz, while 232 complex points were recorded for spectral width of

9601.536 Hz for F3 dimension (^{1}H). Carrier frequencies for F1, F2 and F3 were at 4.725 ppm, 119.081 ppm and 4.725 ppm respectively.

2.11.7 ^{1}H-^{15}N heteronuclear NOE cross relaxation experiment

For the 2D ^{1}H-^{15}N HSQC-NOE experiment, two HSQC spectra were recorded with an interlude, with and without proton saturation (NOE_{on}, NOE_{off}). And the difference in peak intensities between the two spectra will describe the relaxation of NOE. The spectrum was acquired with 1280 complex points over a spectral width of 11160.714 Hz on ^{1}H dimension. In the ^{15}N dimension, 216 points were collected over 2108.370 Hz. Carrier frequencies for ^{1}H and ^{15}N were at 4.782 ppm and 119.082 ppm respectively.

2.12 HDX-MS (hydrogen-deuterium exchange mass spectrometry)

Hydrogen-deuterium exchange mass spectrometry eperiments were carried out in order to probe flexible and rigid part of both wild type EvpP and the EvpP_P143T mutant. Sample protein was exposed to 100% $D^{2}O$ buffer for various period of time and quenched before subjected to mass spectrometry. In order to increase resolution in terms of amino acid sequence, the quenched sample was subjected to pepsin digestion. Prior to ESI mass spectrometry, the peptides were seperated by an HPLC column to further increase resolution. The whole process was carried out in Waters® nanoACQUITY UPLC® Hydrogen Deuterium Exchange (HDS) System (Waters Corporation). Data were recorded and analyzed by ProteinLynx software, ProteinLynx Global Server (Waters Corporation) and mass spectra assignment was carried out using DynamX.

2.12.1 Sample and buffer preparation

Purified wild type EvpP and EvpP_P143T mutant were each concentrated to about 50 μM before subjected to HDX experiment. Dry buffers were prepared by

drying the aqueous buffer (20 mM Tris-Cl, pH7.0) using a SpeedVac concentrator at room temperature or higher. Hundred percent deuterium oxide was added to the dry buffer powder only prior to HDX experiment. The quench buffer was made by mixing 100 µl of 2% TFA to 2.9 ml H_2O. Wash buffer was made by further topping 3 ml of quench buffer up to 5 ml by adding 2 ml of protein buffer (20 mM Tris-Cl pH7.0).

2.12.2 HDX-MS of EvpP and EvpP_P143T

For each round of experiment, 2 µl of protein sample (at a concentration of 50 µM) was mixed with 18 µl of D_2O redissolved dry buffer (to yield a final concentration of 90% D_2O) for different period of time before quenching (10 sec, 20 sec, 30 sec, 1 min, 5 min, 30 min and overnight). The sample mixture was quenched by adding 30 µl of quench buffer (to lower the pH to around 2.5) and was immediately injected into the chilled nanoACQUITY UPLC sample manager. The sample was washed through a pepsin column and the digested peptides were trapped in a C18 trap (ACQUITY BEH C18, Waters Corporation). The peptides were eluted using an 8% - 40% gradient of acetonitrile in 0.1% formic acid at 40 µl/min (nanoACQUITY Binary Solvent Manager, Waters Corporation). The peptides were detected and their respective mass was measured on a Synapt HDMS mass spectrometer (Waters Corporation).

Sequence identification was carried out using the undeuterated samples by ProteinLynx Global Server (Waters Corporation) and searched against the sequences of both EvpP and EvpP_P143T with no enzyme specified and no amino acid modifications. True identifications were revognized if they appeared in both replicated runs. The precursor ion mass tolerance was set to < 10 ppm and the fragment ion tolerance was set to < 20 ppm. The spectrum intensity was set to > 3000. Only those peptides that pass the above criteria would show up on the resultant peptide list.

The peptide spectra assignment of each identified peptide in undeuterated sample as well as in subsequent deuterated samples were carried out using the DynamX software. The output file was converted to a microsoft Excel format for further data analysis. The back-exchange constant used for calculating the exchange rate was 1.49.

2.13 Glutathione-S-Transferase pull-down assay

Fifty microliler of glutathione-sepharose 4B beads (Amersham Biosciences) was added into an eppendorf tube and was washed at least 3 times by 1 ml of 1 X PBS each time. The GST-fusion protein of GST-EvpC (around 100 μg), or GST protein alone as a control, was added into the tube and the mixture was topped up to 500 μl by 1 X PBS. The protein was incubated with the beads at 4°C for at least 1 hour. The beads were washed with 1 ml of 1 X PBS for at least 3 times. 200 μg of targeted protein, EvpP was added into the tube and the volume was topped up to 500 μl by 1 X PBS. The mixture was incubated at 4°C for at least 2 hour with mixing. The beads were washed with 1 X PBS to remove unbound or loosely bound proteins. 100 μl of 2 X SDS sample buffer was added into the beads and the suspension was boiled for 5 min followed by centrifugation at 13,000 rpm for 10 min. The supernatant was analyzed on SDS-PAGE.

2.14 Expression and purification of AcrH-AopB (1-264)

Protein expression and purification of AcrH-AopB (1-264) was carried out in the same way as described in Section 2.2 except that both the dialysis buffer as well as the FPLC buffer used were 20 mM Tris-Cl pH8.0 with 200 mM NaCl, 5% Glycerol and 5mM β-mercaptoethanol. The Buffer A for MonoQ ion exchange chromatography was 20 mM Tris-Cl pH8.0 and Buffer B was 20 mM Tris-Cl pH8.0 with 1 M NaCl.

2.15 Limited protease digestion of AcrH-AopB (1-264)

Limited protease digestion of AcrH-AopB was carried out in the same way as described in Section 2.4 except that the enzymes used were elastase, thrombin, thermolysin, actinase E, proteinase K, clostripain and bromelain.

2.16 Protein sequencing of digested AcrH-AopB complex

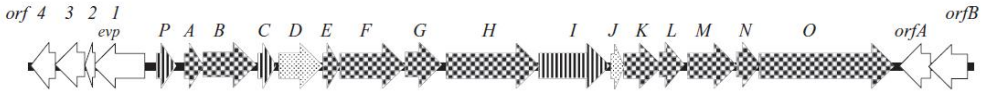
Protein sequencing of digested AcrH-AopB complex was carried out in the same way as described in Section 2.7.

2.17 Crystalization screening of digested AcrH-AopB complex

The purified AcrH-AopB complex (after partial digestion by different enzymes) was concentrated to 10 mg/ml with an Amicon filter (Milipore, 10-kDa molecular weight cut-off) before crystallization. Crystallization screening was carried out using Index Screens (Hampton Research) employing the hanging-drop vapor-diffusion technique at 302K. The drop sizes were 1 µl protein solution to 1 µl reservoir solution; 1 µl to 2 µl as well as 2 µl to 1 µl. The initial crystal was obtained within two weeks. The crystallization was further optimized by varying the pH of the reservoir buffer (in which previously crystals could form), the concentration of the protein as well as the crystallizing temperature. Preliminary diffraction pattern of crystal was collected with an in-house X-ray machine (Bruker MICROSTAR X-ray generator and PLATINUM135 CCD detector) for optimizing the condition for crystallization.

CHAPTER 3, RESULTS & DISCUSSIONS (EvpP)

3.1 The search for candidates in T6SS for structural studies



The T6SS gene cluster of Edwardsiella tarda. evp: *E. tarda* virulent protein; orf: open reading frame. (Zheng and Leung, 2007)

The gene cluster of T6SS in *E. tarda* harbours several small proteins that might be potential targets for structural study, including EvpA, EvpC, EvpE, EvpJ, EvpL, EvpN and EvpP. However, only EvpC has a known structure which is similar to Hcp1. EvpC, in high concentration, forms a hexameric ring structure that can elongate to constitute the needle conduit of T6SS. Therefore we initially examined each of the small proteins other than EvpC as potential targets for structural studies. Each protein, either full-length or a truncation mutant, was cloned into the pET-M vector and tested for expression. The insoluble ones were then tested for protein refolding by rapid dilution. All soluble protein samples were concentrated and passed through FPLC (Hiload Superdex 75) to investigate their oligomeric state in solution. The result was shown in Table 3.1 as follows.

Table 3.1, Expression of the small proteins in T6SS gene cluster of *E. tarda*

Proteins	Length	Expressed	Expression	Solubility	After Refolding
EvpA	**171a.a.**	**Full length**	**Medium**	**Inclusion body**	**Aggregation**
EvpE	**158a.a.**	**Full length**	**High**	**Inclusion body**	**Aggregation**
	158a.a.	**1-120**	**High**	**Inclusion body**	**Aggregation**
	158a.a.	**1-135**	**High**	**Inclusion body**	**Aggregation**
EvpJ	**100a.a.**	**Full length**	**Medium**	**Inclusion body**	
	100a.a.	**1-90**	**Very low**	**Inclusion body**	

EvpL	**235a.a.**	**Full length**	**No**		
	235a.a.	**54-235**	**No**		
EvpN	**216a.a.**	**1-182**	**High**	**Inclusion body**	**Aggregation**
EvpP	**185a.a.**	**185**	**Medium**	**Soluble**	

EvpA, EvpE (both full length as well as two truncation mutants) and EvpN had a high expression yield but formed inclusion bodies. They could be refolded into a soluble state by rapid dilution into normal buffers which did not contain urea or guanidine hydrochloride at neutral pH. However all of them were aggregated in solution which was shown by gel filtration as they were eluted out in void volume (data not shown). EvpJ (full length and truncation mutant $EvpJ^{1-90}$) showed little expression which also formed inclusion body. EvpL (full length and truncation mutant $EvpL^{54-235}$) had no expression at all. Only EvpP was soluble by itself that did not require protein refolding.

This result was not surprising as EvpP was found to be a secreted protein therefore it should have a higher propensity to be soluble by itself. On the other hand, the others were generally thought to be apparatus proteins that might need the presence of binding partners to reach a stable state. In addition, they may require chaperones for proper expression to prevent pre-mature association before meeting their interacting partners. As EvpP was exclusively found in *Edwardsiella* species and contributes to pathogenic virulence, it is a very interesting target for structural studies in T6SS.

3.2 Characterization of EvpP

EvpP (from *E. tarda* strain 130/91) is 185 amino acid residues in length, 20.3 kDa in molecular weight with a theoretical pI of 9.47. His-tagged EvpP (cloned in pET-M expression vector) was over-expressed in *E. coli* BL21 (DE3) as a soluble protein that could be further purified through Ni-NTA affinity column. The purified

EvpP protein was then passed through a Superdex 75 gel filtration column and it was eluted out around 72 mins at a speed of 1 ml/min (Figure 3.1). This showed that EvpP is not an aggregation and it could exist as a monomer in solution.

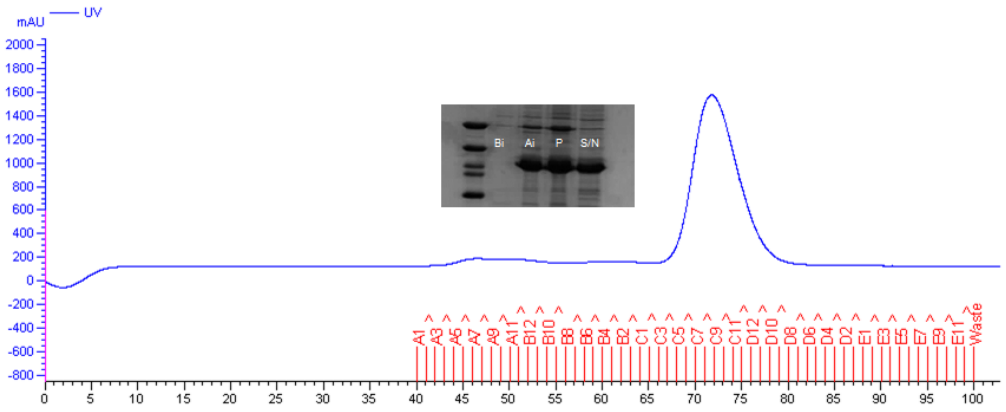


Figure 3.1, Expression and FPLC profile of EvpP (130/91). Bi: before induction; Ai: after induction; P: pellet; S/N: supernatant.

EvpP did not contain any signal sequence at the N-terminus (by SignalP 4.1 server) and no hydrophobic patch was found throughout its sequence. The secondary structure prediction showed that EvpP is not a well-structured protein; as more than half of the protein was predicted to be random coils (Figure 3.3). The circular dichroism spectra of EvpP under different pH (ranging from pH4 to pH8) also showed a pattern that is neither typical of α-helices or β-strands. However EvpP does have higher order structures as the spectra significantly differ from the pattern of a totally random coiled protein (Figure 3.2). This was further confirmed by acquiring a ^{1}H-^{15}N HSQC of EvpP (Figure 3.4).

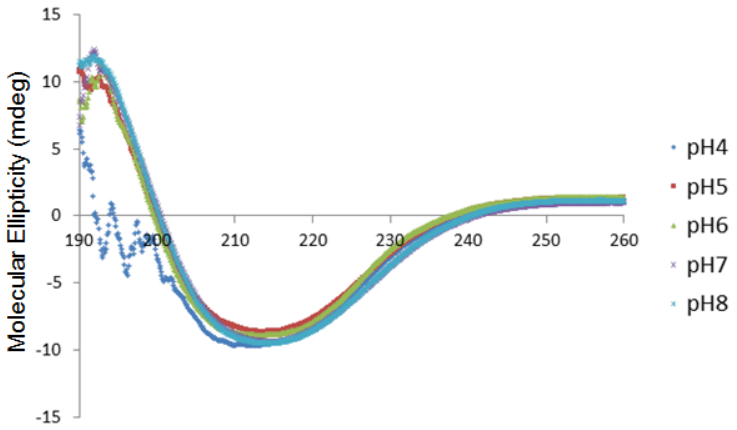


Figure 3.2, Circular dichroism spectra of EvpP under different pH conditions.

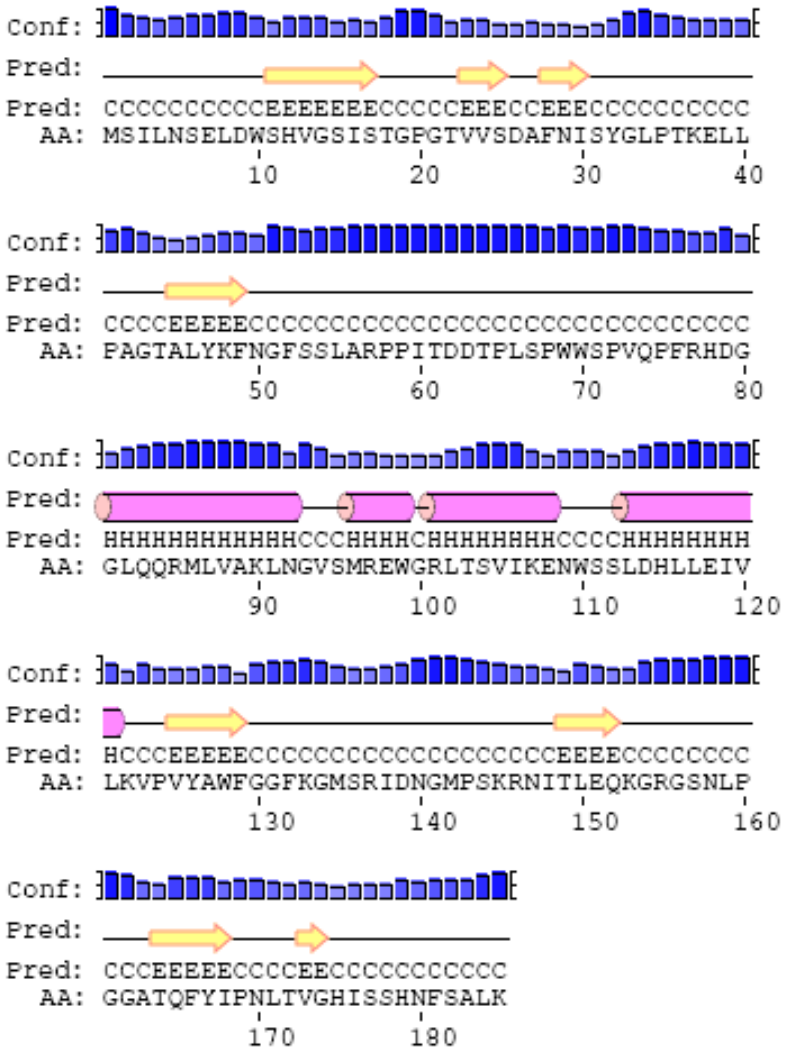


Figure 3.3, Secondary structure prediction of EvpP (by PSIpred 3.0 server, UCL Department of Computer Science)

The CD spectra showed that the secondary structure of EvpP remained unchanged across all the tested pH. The CD spectrum under pH4 partially deviated from the others, which might be due to the acidic pH where proteins started to lose their secondary structure. As the pI of EvpP was predicted to be 9.47; the CD spectra from pH5 to pH8 all converged to show a common trend. Therefore neutral pH (pH7) was used for further experiments with EvpP.

As EvpP failed to form any kind of crystal either under room temperature or lower temperatures; the structural studies on EvpP were mainly focused on NMR. The ^{1}H-^{15}N-edited HSQC of EvpP was recorded using a Bruker AVANCE 800 MHz NMR spectrometer and processed HSQC was shown in Figure 3.4. As the His-tag

was proved difficult to remove by thrombin cleavage, EvpP with His-tag was used for NMR studies. 90 separated cross-peaks could be counted from the HSQC (not counting side chain cross-peaks and very weak cross-peaks) out of a total of 203 amino acid residues of EvpP with His-tag. Since EvpP is not a well-structured protein forecast by secondary structure prediction, we hypothesized that it was the structured portion of EvpP that showed up on the HSQC spectrum while the flexible regions could not be observed. The dispersion of cross-peaks on the HSQC has further confirmed that EvpP does have a tertiary structure.

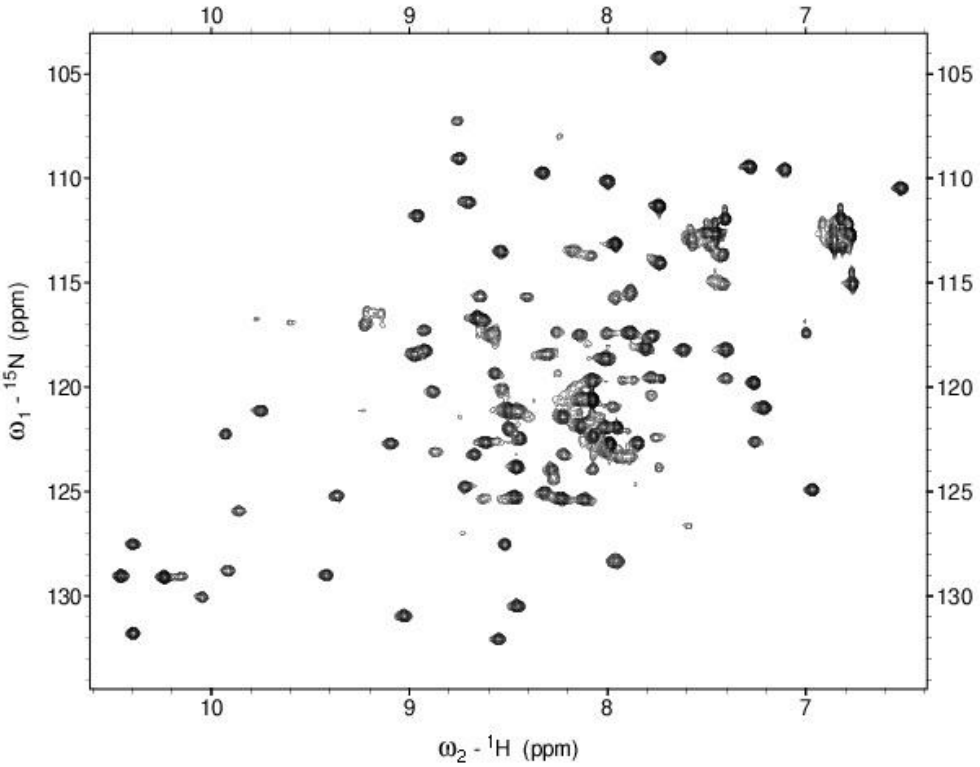


Figure 3.4, ^{1}H-^{15}N HSQC of EvpP. 90 separated cross-peaks could be observed on the HSQC (excluding side chain cross-peaks and very weak ones) out of a total of 203 amino acid residues of EvpP with His-tag. Each cross-peak could potentially represent one amino acid residue of EvpP.

EvpP was found to have a poor thermo-stability, over 90% of the protein precipitated out at 35 ℃. It could routinely precipitate at room temperature or slightly lower temperatures. Therefore it had to be stored at a low temperature such as 4 ℃. EvpP was also found to go through degradation over time. This occurred even at low temperature if given a long time period of up to a few weeks.

3.2.1 EvpP as a possible dimer in solution

Dynamic light scattering (DLS) has been a fast and reliable tool to study the oligomeric state of proteins in solution. Based on the FPLC profile alone, it is difficult to differentiate whether EvpP is a monomer or dimer in solution. At 20 ℃, EvpP showed a dimeric profile. The hydro-dynamic radius was detected to be 2.95 nm and the corresponding molecular weight was estimated to be 44.4 kDa, which is twice the molecular weight of EvpP with His-tag. The homogenicity histogram was shown in Figure 3.5. The great majority of EvpP was found to constitute a homogenous species with a hydro-dynamic radius around 2.95. Two minor species of larger entities were also present in the solution.

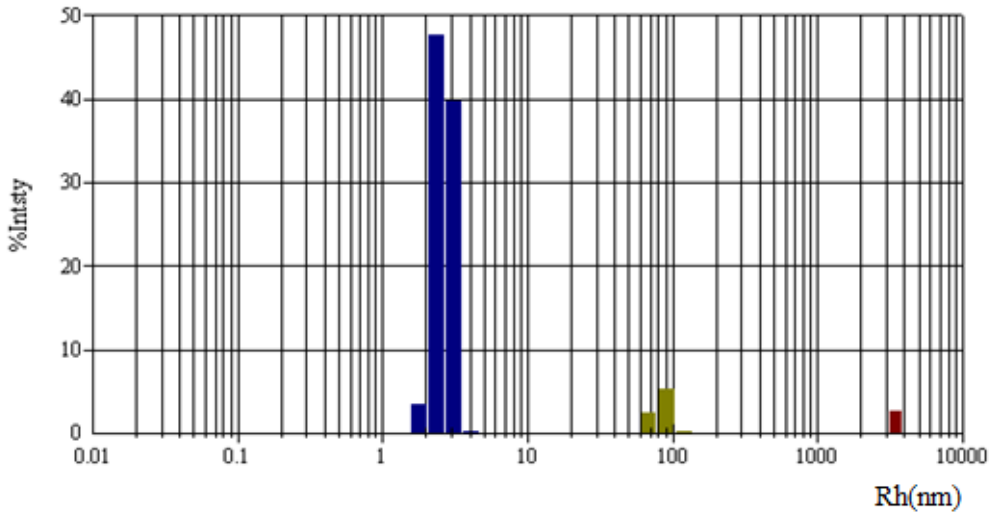


Figure 3.5, Histogram showing the homogenicity of EvpP at 20 ℃. Most of EvpP were found to be the dimer form, with a small portion adopting higher oligomeric states.

The formation of dimer was found in both low (10 μM) and high (500 μM) concentrations of EvpP. At a high concentration (500 μM) measured by NMR, the ^{15}N-edited EvpP was found to have a ^{1}H-^{15}N two-spin decay N_zH_z of around 40 ms and a $T_1\rho$ around 44 ms. Both observations demonstrated a possible dimer formation of EvpP in solution. Such dimer formation is independent of a high protein concentration.

To further verify the dimeric state of EvpP in solution, protein cross-linking assay was performed on wild-type EvpP. Freshly made 2.3% glutaraldehyde was used to cross-link EvpP at a concentration of 50 μM. The reaction was incubated at room temperature instead of 37℃, as EvpP readily precipitate at that temperature. To compensate for that, the duration of incubation was lengthened and the trend of cross-linking was monitored by SDS-PAGE (Figure 3.6).

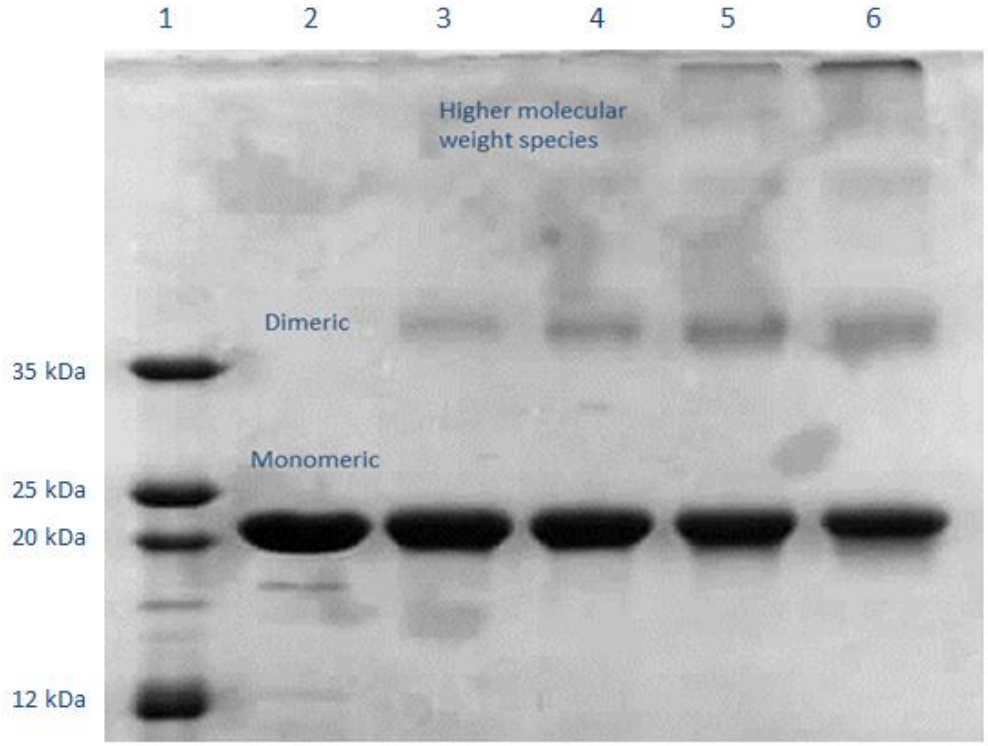


Figure 3.6, Cross-linking assay of EvpP using glutaraldehyde. From left to right, Lane 1: molecular weight marker; Lane 2, un-crosslinked EvpP; Lane 3 to Lane 6, EvpP incubated with glutaraldehyde for different period of time: Lane3, 15 min; Lane 4, 30 min; Lane 5, 45 min; Lane 6, 1 hour.

As time progressed, more EvpP was cross-linked to the dimeric form which showed up as another band of higher molecular weight than the EvpP monomers. As expected, the intensity of the band corresponding to the dimeric form increased in a progressive manner with more incubation time. Some higher oligomeric order species of EvpP started to appear after 30 min of cross-linking. One thing worth noticing was that even after one hour of reaction, the majority of EvpP still remained in the

monomeric form on SDS-PAGE. This indirectly implied that the observed dimer of EvpP might not be physiological, or the dimers formed were not strong enough to be easily cross-linked chemically.

3.3 Limited proteolysis showed a coherent protease-resistant core-region

The HSQC of EvpP as well as the secondary structure prediction both pointed to the fact that EvpP may contain a large portion of intrinsically disordered region. We speculated that EvpP has both rigid regions that are well-structured (those will show up on the HSQC) and flexible regions that may be involved in functional interactions. In order to verify this hypothesis and to further map out the rigid and flexible regions on EvpP, we performed limited protease digestion on EvpP using different enzymes. The digestion patterns of EvpP with trypsin, chymotrypsin, elastase and thermolysin were shown in Figure 3.7 and Figure 3.8. As trypsin reliably digested EvpP and generated a stable band around 10 kDa and another smeary species of lower molecular weight; it was used as a control for comparison among different proteases. More enzymes other than the four above were also tried on EvpP, such as papain, actinase E, proteinase K, etc. However they either failed to digest EvpP, or they failed to yield a convergent digestion pattern after repetitive trials.

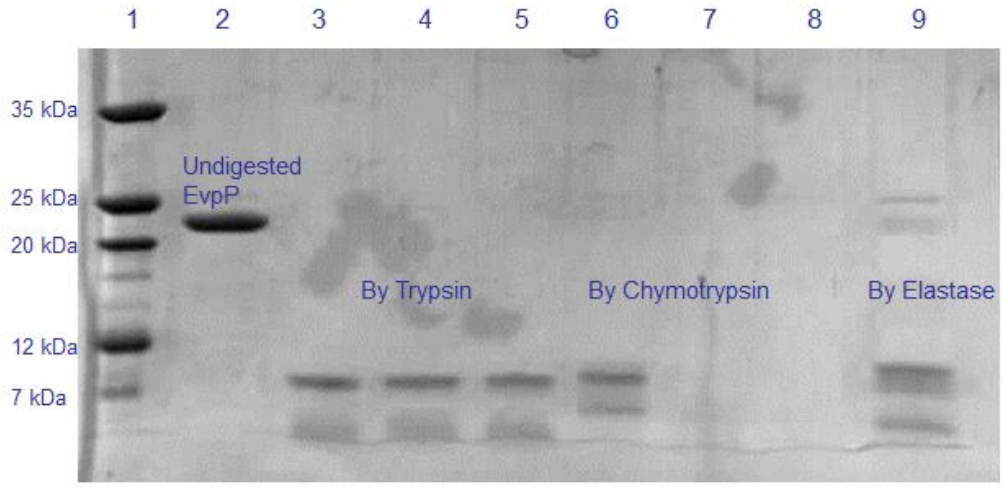


Figure 3.7, Limited protease digestion of EvpP using Trypsin, Chymotrypsin and Elastase. From left to right, Lane 1, molecular weight marker; Lane 2, undigested EvpP; Lane 3 to Lane 5, EvpP digested by trypsin (1:100) for 20 min, 40 min and 60 min; Lane 6 to Lane

8, EvpP digested by chymotrypsin (1:100) for 20 min, 40 min and 60 min; Lane 9, EvpP digested by elastase (1:100) for 20 min.

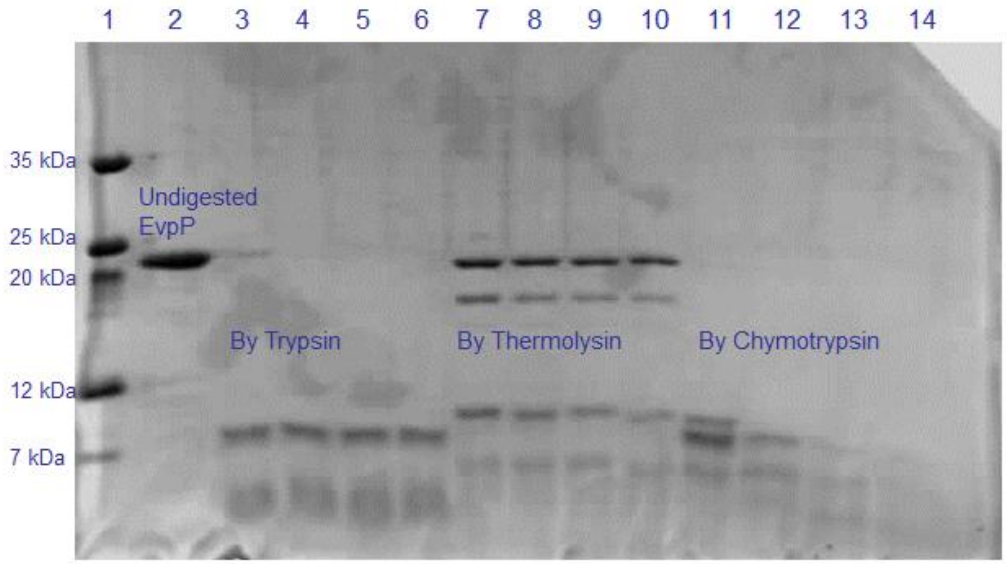


Figure 3.8, Limited protease digestion of EvpP using Trypsin, Thermolysin and Chymotrypsin. From left to right, Lane 1, molecular weight marker; Lane 2, undigested EvpP; Lane 3 to Lane 6, EvpP digested by trypsin (1:200) for 15 min, 30 min, 45 min and 60 min; Lane 7 to Lane 10, EvpP digested by thermolysin (1:200) for 15 min, 30 min, 45 min and 60 min; Lane 11 to Lane 14, EvpP digested by chymotrypsin (1:200) for 15 min, 30 min, 45 min and 60 min.

As shown in Figure 3.7 and 3.8, limited protease digestion by four enzymes generated four different patterns. Trypsin efficiently digested EvpP and left a band which was around 10 kDa in molecular weight and a smeary species of lower molecular weight. The digestion of EvpP by trypsin was not in a time-dependent or progressive manner like the other three enzymes. More importantly, the digestion pattern of EvpP by trypsin was stable over time, as shown in SDS-PAGE, there was no decay in band intensity. This was not found in all other enzymes. Through chymotrypsin and elastase, all EvpP proteins were eventually completely digested after 60 minutes of incubation at room temperature. Thermolysin was not effective in digesting EvpP. However, if the concentration of thermolysin was raised to 1:50 in molar ratio to EvpP, thermolysin was still able to completely digest EvpP within 90 minutes.

Among all four digestion patterns, there was always one band with a molecular weight of approximately 10 kDa that remained undigested. The protease-resistant 10 kDa band in trypsin-digested EvpP had sightly lower molecular weight than those digested by the other three enzymes. Although the protease-resistant 10 kDa pieces differ from each other slightly; their presence across all four digestion patterns have led to the speculation that there could be a structured core region of EvpP that is around 10 kDa.

Obviously, it is intriguing to find out the identity of the protease-resistant region of EvpP. In our attempt, we performed N-terminal sequencing for all the 10-kDa protease-resistant pieces digested by the four enzymes that were shown on SDS-PAGE. The bands on SDS-PAGE after limited proteolysis by each enzyme were first immuno-blotted onto PVDF membrane (immobilized) before being subjected to coomassie-blue staining and N-terminal sequencing. The 10 kDa piece after trypsin digestion was found to have an N-terminal sequence of GSELMSILN. The one after thermolysin digestion was found to have an N-terminal sequence of LVPRG. The ones after chymotrypsin and elastase digestion were found to have the N-terminal sequence of MHHHH. The results were summarized in the following diagram as shown in Figure 3.9.

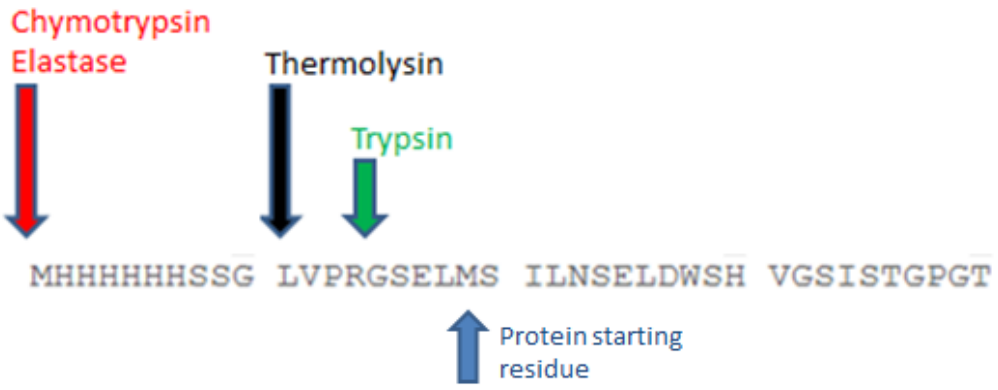


Figure 3.9, Schematic diagram showing the first 40 residues of His-tagged EvpP and the results of N-terminal sequencing of the 10 kDa bands after digestion by all four enzymes.

As shown in Figure 3.9, all the protease-resistant pieces that had a molecular weight around 10 kDa after limited proteolysis were found to start from the His-tag

region which precedes the N-terminus of EvpP. This suggested that it was the N-terminal half of EvpP that was resistant to limited protease digestion and this region was supposed to contain the structured part of EvpP. In order to map out the boundaries of the protease-resistant regions of EvpP after trypsin digestion, we determined the molecular weight of the 10-kDa band and the lower smeary species by ESI-MS. The reason for choosing trypsin was because, only trypsin digested EvpP generated a consistent and stable pattern that could be purified for molecular weight determination by mass spectrometry. To determine the molecular weight of the two components generated by trypsin limited proteolysis, trypsin-digested EvpP was first purified by FPLC (Hiload Superdex 75) to remove any residual enzymes (Figure 3.10). It was surprising that the upper band could not be separated from the smeary species by either FPLC or MonoQ ion exchange chromatography. This might imply that even after trypsin digestion, the two parts were still associated with each other so that they would always be co-purified and co-eluted.

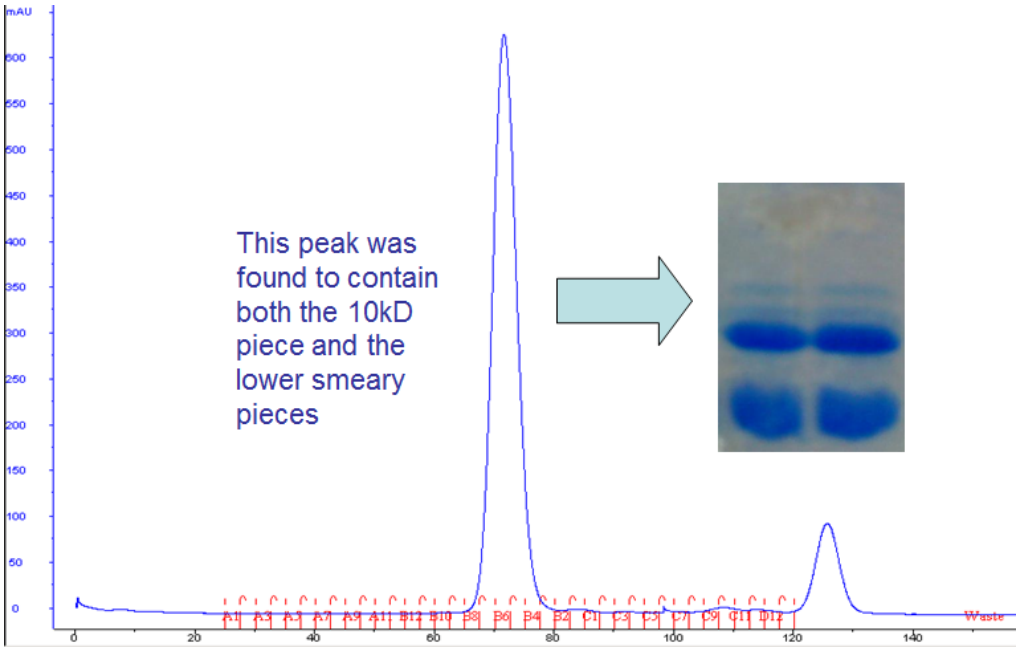


Figure 3.10, FPLC profile of trypsin-digedted EvpP. The peak eluted at 72 minutes was found to contain both the upper 10-kDa piece as well as lower molecular weight smeary species.

The upper band was found to have a molecular weight of 9,614 Da. Together with the N-terminal sequencing result, we mapped the upper band of trypsin digested

EvpP to be from the N-terminus of EvpP to residue Arg85 (Figure 3.11). The lower m.w. species after trypsin digestion was found to have a molecular weight of 4,121 Da. We also attempted to sequence this lower smeary piece, however it did not yield meaningful result through N-terminal sequencing. Instead, we performed peptide identification for the lower molecular weight species of EvpP digested by trypsin using MALDI-TOF-TOF tandem mass spectrometry. Through MALDI-TOF-TOF, two peptides were picked up and their molecular weights were found to match the sequences of EvpP: ENWSSLDHLLEIVLK and VPVYAWFGGFK. Therefore the boundaries of lower species after trypsin digestion should be around the regions containing the above peptides. Taken together with the molecular weight of the smeary species at 4,121 Da, we were able to map it from residue Ile106 to residue Gly141 (Figure 3.11).

```
        10         20         30         40         50         60
MSILNSELDW SHVGSISTGP GTVVSDAFNI SYGLPTKELL PAGTALYKFN GFSSLARPPI

        70         80         90        100        110        120
TDDTPLSPWW SPVQPFRHDG GLQQRMLVAK LNGVSMREWG RLTSVIKENW SSLDHLLEIV

       130        140        150        160        170        180
LKVPVYAWFG GFKGMSRIDN GMPSKRNITL EQKGRGSNLP GGATQFYIPN LTVGHISSHN

FSALK
```

Figure 3.11, Sequence of EvpP. The boundaries of protease-resistant regions after limited proteolysis by trypsin were boxed.

Although EvpP after limited proteolysis by chymotrypsin, elastase or thermolysin may have slightly different boundaries from the one shown for trypsin; they inevitably suggested that the N-terminal half of EvpP is more rigid, in the sense that it is more resistant to proteolysis while the C-terminal half is not as rigid.

3.3.1 The protease-resistant "core" of EvpP is not structurally stable

The trypsin-digested EvpP provided an alternate candidate for structural studies as it contains the rigid "core" of EvpP and has fewer residues. Therefore we first acquired a ^{1}H-^{15}N HSQC for trypsin digested EvpP after purification through

MonoQ ion exchange chromatography (Figure 3.12). By overlaying the HSQC of trypsin-digested EvpP against the one of full-length EvpP, we noticed that after limited proteolysis by trypsin, EvpP still retained its global structure. Although most of cross-peaks had slightly different chemical shifts, the global pattern of dispersion was unchanged according to the HSQC (Figure 3.13). However, during an attempt to acquire the 3D ^{15}N-^{13}C-edited CBCA(CO)NH and HNCACB spectra, trypsin-digested EvpP lost its structure and significant precipitation was observed in the NMR tube after a three-day experiment at 20 ℃ (Figure 3.14). Although the protease-resistant core of EvpP (after limited proteolysis by trypsin) still retained its tertiary structure, it was no longer as stable as the intact protein.

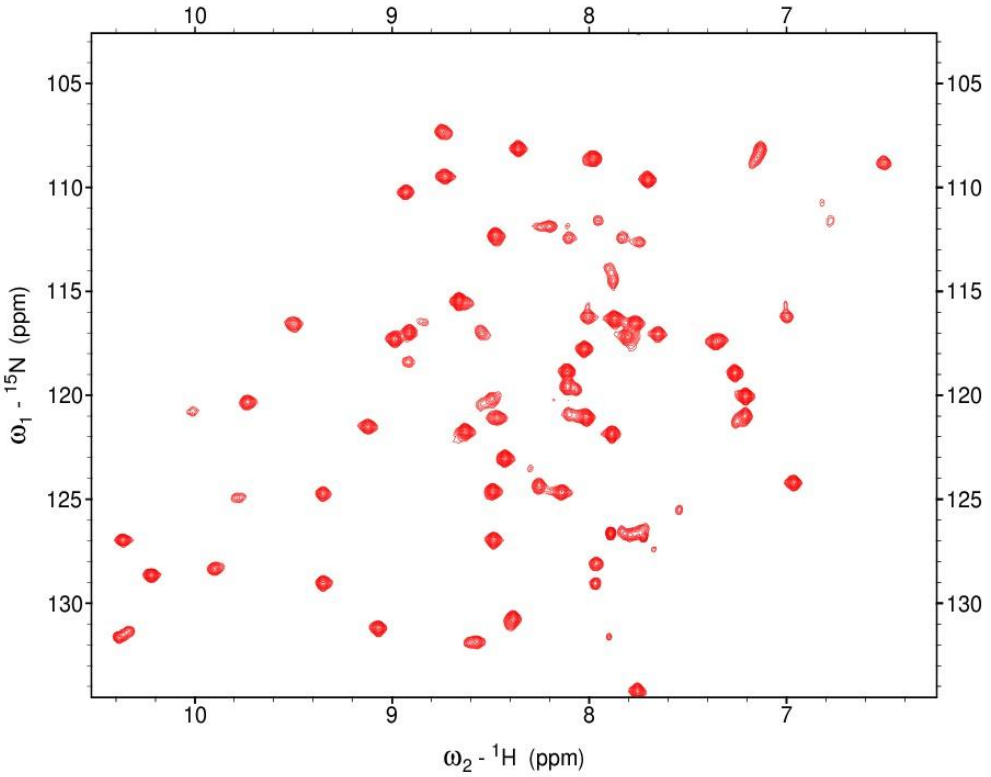


Figure 3.12, ^{15}N-^{13}C-edited HSQC of EvpP after limited proteolysis by trypsin.

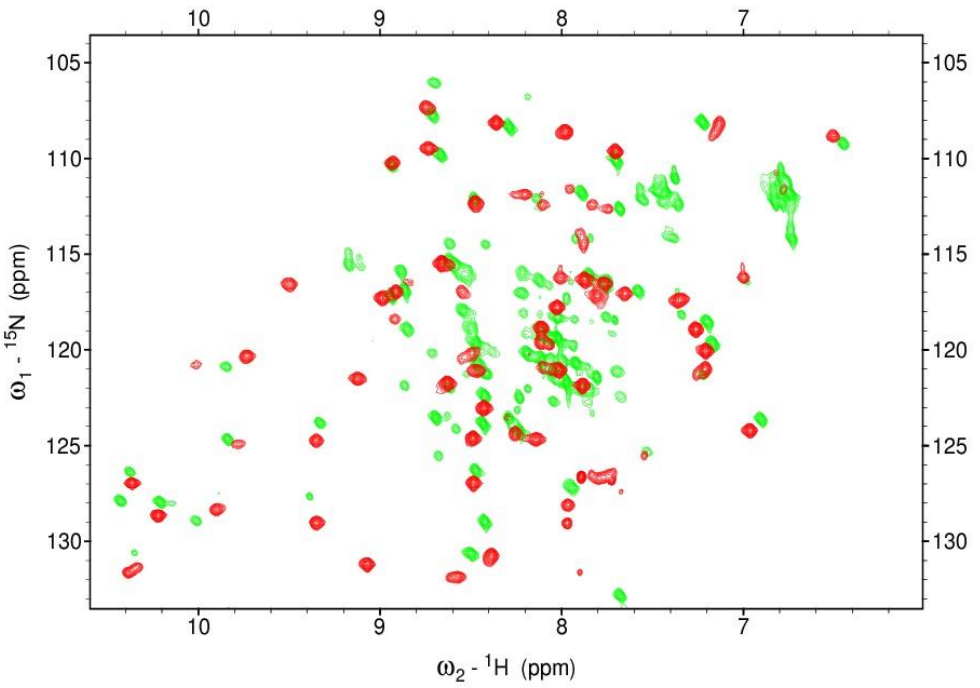


Figure 3.13, Overlay of HSQC acquired from trypsin-digested EvpP against full length EvpP. Green: Full length EvpP; Red: EvpP after limited proteolysis by trypsin

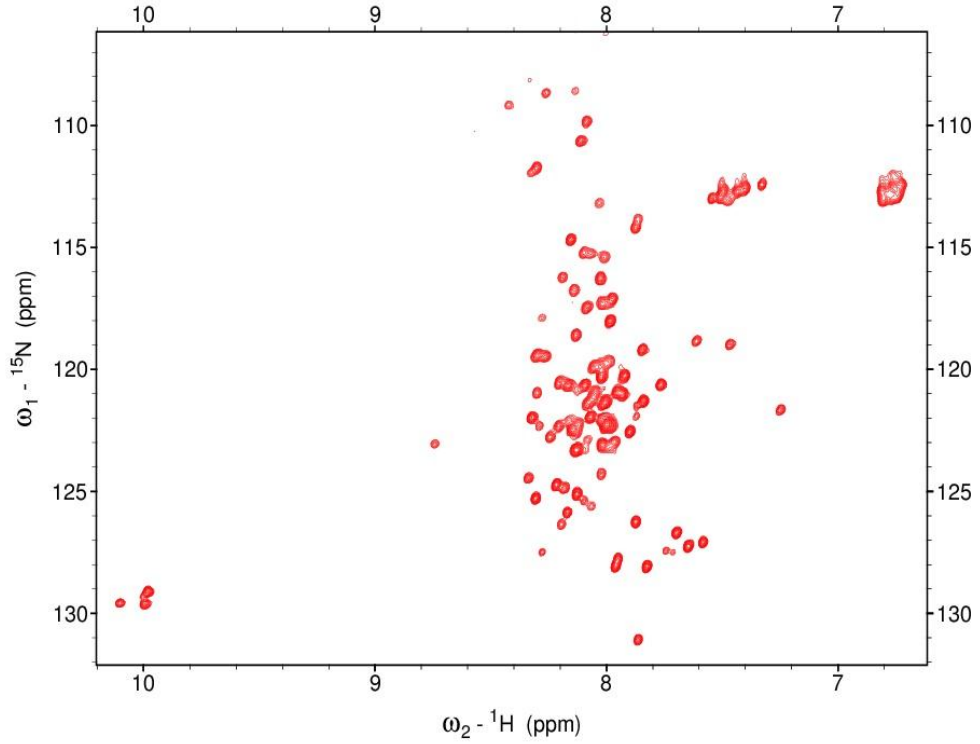


Figure 3.14, ^{15}N-^{13}C-edited HSQC of trypsin-digested EvpP after 3-day experiment at 20 ℃

As EvpP readily precipitate at temperatures of 35℃ or higher, making thermal denaturation unfeasible, we performed urea denaturation experiments measured by a fluorescence spectrometer comparing both intact as well as trypsin-digested EvpP (Figure 3.15). As expected, after limited proteolysis by trypsin, EvpP was more susceptible to unfolding by urea denaturation compared to the full-length EvpP. The estimated Gibb's free energy for wild-type EvpP was fitted at 10.619 kCal/mol, with an m value of 0.80 towards urea denaturation. Trypsin digested EvpP on the other hand, had a ΔG of 8.261 kCal/mol, which was significantly lower than wild-type EvpP. The estimated m value was 1.23, which was also larger compared to wild-type EvpP. It was worth-noticing that the full-length EvpP already showed traits of unfolding before the transition state at 5 M of urea concentration.

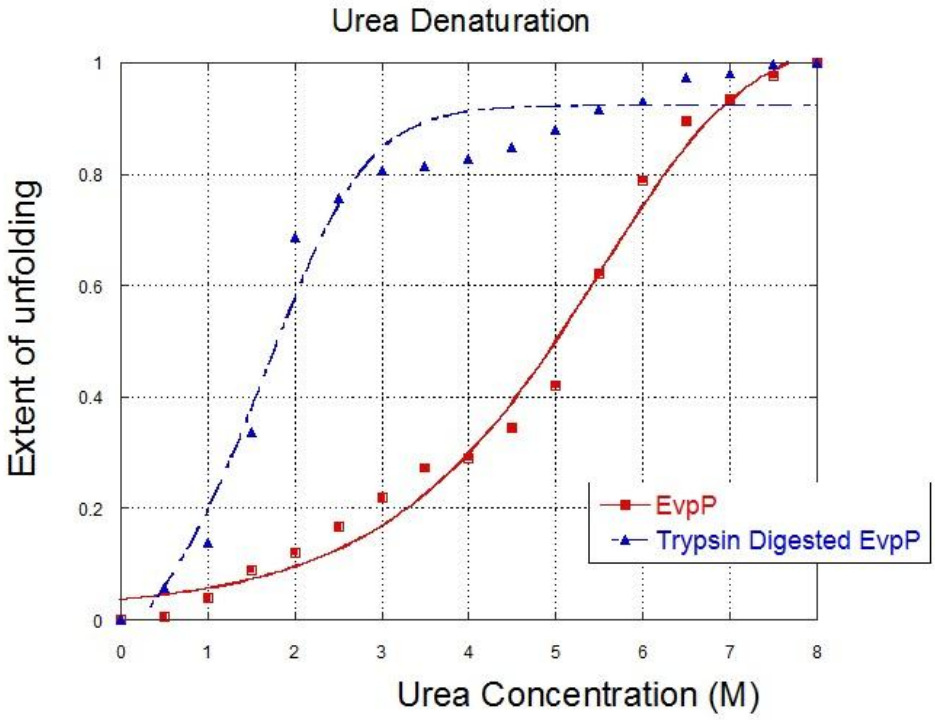


Figure 3.15, Urea denaturation measured by fluorescence spectrometry. Curve fitting was performed using Kaleidagraph ver4.0

Limited proteolysis by various enzymes has generated a relative consensus about the core region of EvpP, which is resistant to protease digestion. Among all

enzymes that were used, trypsin yielded a stable digestion pattern consisting of a 9,614-Da piece and a 4,121-Da smeary species. The band around 10 kDa was observed in all protease-digested patterns while the 4 kDa species was not. However the 4 kDa species always remained attached to the 10 kDa piece, forming a quasi-stable complex that could be co-purified and co-eluted through FPLC or ion exchange chromatography. Although this trypsin-digested EvpP is not as stable as the full-length protein; by acquiring an initial HSQC (before it starts to unfold), we learned that the structured regions of EvpP are largely located within the trypsin-digestion mapped boundaries.

3.4 NMR backbone chemical shift assignment for wild type EvpP

As the trypsin-digested EvpP is not stable for structural study; we attempted to acquire the 3D ^{15}N-^{13}C edited CBCA(CO)NH and HNCACB spectra for backbone assignment using wild-type EvpP. Due to the dimeric feature and the resultant fast relaxation time of EvpP, 3D CBCA(CO)NH and HNCACB experiments using ^{15}N-^{13}C-edited samples are not sensitive enough. Therefore deuterated EvpP samples were used to acquire these spectra. *E. coli* cells that carry plasmids expressing EvpP were allowed to grow in M9 minimal media prepared from 100% D_2O. After the cells were harvested, deuterated EvpP proteins were purified by Ni-NTA affinity column and FPLC using a normal aqueous buffer so that the backbone amide protons can be exchanged back to hydrogen. The sample was concentrated to around 700 μM for both 3D experiments. Through 3 days experiment at 20 ℃, persistent precipitation was observed in the NMR tube; however the remaining EvpP still retained its structure shown by a final HSQC acquired at the end of the 3D experiments.

Backbone chemical shifts were assigned using a pair of HNCACB and CBCA(CO)NH triple resonance experiments. First, we correlated all the amide-proton cross-peaks (on the HSQC) with the Cα and Cβ resonances of the same residue using the HNCACB spectrum. Next, CBCA(CO)NH spectrum was used to

correlate the amide-proton cross peaks with the $C\alpha_{i-1}$ and $C\beta_{i-1}$ of its preceding residue. The amide-proton cross-peaks were then linked together to form backbone fragments by matching chemical shifts of Cα and Cβ from one residue with the $C\alpha_{i-1}$ and $C\beta_{i-1}$ from another residue, or vice versa. An example of such backbone linkage was shown in Figure 3.16, exhibiting residues S2 to N5. The backbone fragments were then fitted into the sequence of EvpP aided by the signature Cα and Cβ chemical shift values of typical amino acids as starting points. For example, Ser and Thr residues have unique Cβ chemical shift values that are higher than their Cα. Another typical residue used for initial fitting is Ala, owing to its unique Cβ chemical shift value that usually appeared at ≤20 ppm, which is much lower than other residues. Finally, Gly residues were also particularly helpful as it only has a Cα cross peak appearing in the HNCACB spectrum.

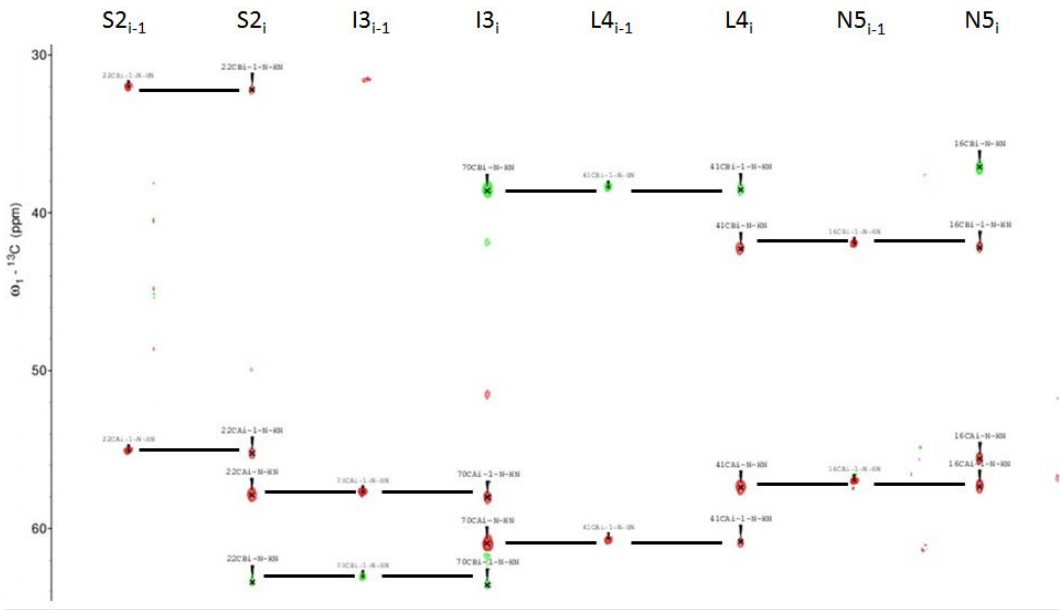


Figure 3.16, Sequential assignment of backbone chemical shifts of EvpP using a deuterated sample. Lanes marked by "i" were the corresponding HNCACB spectrum amide strips; Lanes marked by "i-1" were the corresponding CBCA(CO)NH spectrum amide strips.

Using this method, we were able to unambiguously assign 41 cross-peaks (out of 88, total number of cross-peaks) on the HSQC into the amino acid sequences. There were more backbone linkage fragments being identified. However they did not have characteristic Cα and Cβ chemical shifts like Ser, Thr or Ala residues. Therefore

we could not fit them into a particular EvpP amino acid sequence due to the lack of information. The assigned residues were shown in Figure 3.17 and they were mapped onto the sequence of EvpP in Figure 3.18.

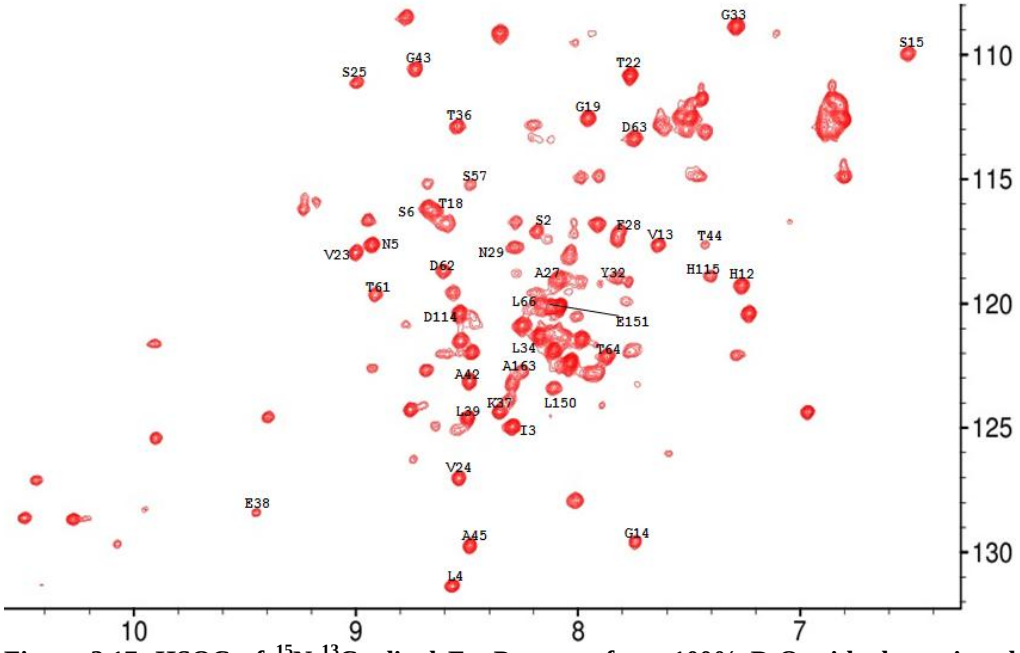


Figure 3.17, HSQC of ^{15}N-^{13}C-edited EvpP grown from 100% D_2O with the assigned residues labelled.

```
        10         20         30         40         50         60
MSILNSELDW SHVGSISTGP GTVVSDAFNI SYGLPTKELL PAGTALYKFN GFSSLARPPI

        70         80         90        100        110        120
TDDTPLSPWW SPVQPFRHDG GLQQRMLVAK LNGVSMREWG RLTSVIKENW SSLDHLLEIV

       130        140        150        160        170        180
LKVPVYAWFG GFKGMSRIDN GMPSKRNITL EQKGRGSNLP GGATQFYIPN LTVGHISSHN

FSALK
```

Figure 3.18, Sequence of EvpP with the assigned residues by HNCACB and CBCA(CO)NH mapped out in pink; the trypsin-digestion boundary was marked by red box.

As shown in Figure 3.18, almost all the assigned residues fell within the trypsin-digestion boundary, which agreed with our previous hypothesis that the trypsin-resistant core of EvpP harbours the structured regions of EvpP. Three other residues were found outside the boundary. As they no longer showed up on the HSQC of the trypsin-digested EvpP, they still agreed with our hypothesis. More importantly, most of the assigned residues were found to be at the N-terminus of

EvpP. This also supported our speculation that the N-terminal half is more rigid than the C-terminus. The assigned stretches were mainly composed of three to five residues only. This was due to the incomplete set of information gathered from the Cα and Cβ chemical shifts of wild-type EvpP.

Now we were facing one problem. We've already deuterated EvpP and thus far, only half of the cross-peaks on the HSQC could be assigned. Therefore to obtain a more complete backbone chemical shift assignment, we definitely need a better construct that gives stronger signal than wild-type EvpP.

3.5 The search for stable mutant of EvpP, P143T

Previously, several truncation mutants of EvpP were designed according to the secondary structure prediction. They were sub-cloned into pET-M expression vectors and tested for expression. In all cases, the C-terminal residues were removed in the truncation mutants as the N-terminal half accounted for most of the structure of EvpP (Figure 3.19). However, all the resultant mutants were found to be insoluble as they inevitably fell into inclusion bodies after being expressed in *E. coli* BL21 (DE3) competent cells. Although some of them could be refolded to obtain a portion of soluble protein; the yield was too low and all of the refolded mutant proteins had very poor stability.

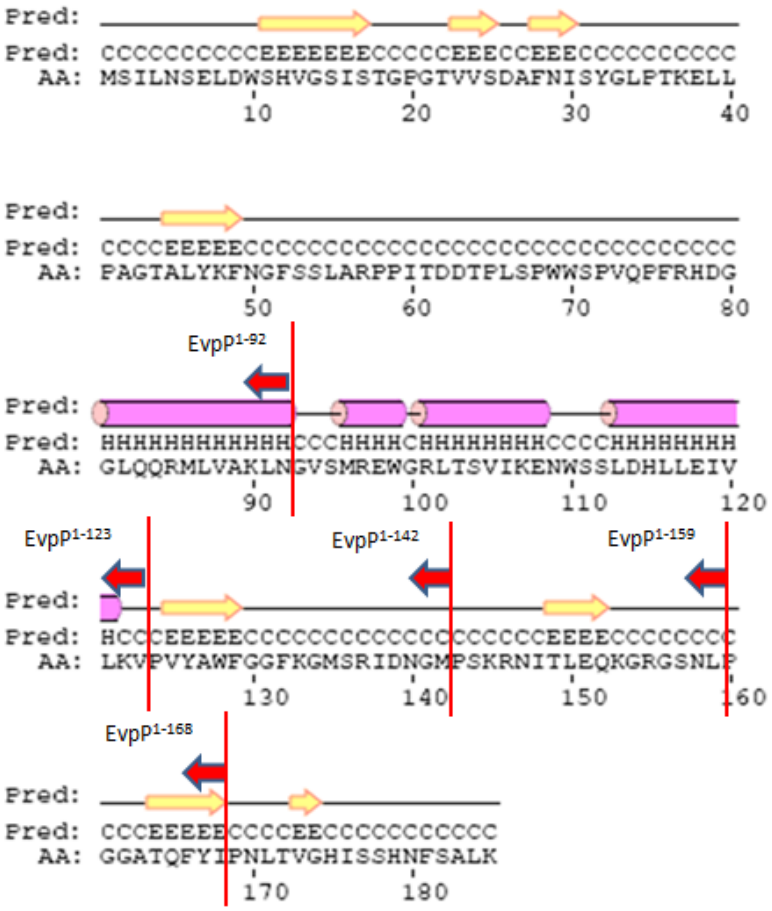


Figure 3.19, Boundaries of truncation mutants designed for EvpP marked against the sequence of EvpP with secondary structure prediction.

These truncation mutants were further cloned into vectors that contained solubility enhancement tags such as pET32a (bearing a thioredoxin tag) and pGEX-4T-1 (bearing a glutathione-S-transferase tag). However none of the recombinant truncation mutants turned out to be soluble.

We therefore explored other possibilities for making stable mutants of EvpP. As EvpP is exclusively found in *E. tarda* and some *E. ictaluri* species; we analyzed individual EvpP across different species of *Edwardsiella* (Yang M, et al, 2012). Shown in Figure 3.20, EvpP was found present in T6SS among all the five *Edwardsiella* species. The gene cluster make-up was essentially the same for all five species; but, there were subtle differences in between. For example, in the T6SS of some species, EvpP has a longer sequence (as in *E. tarda* species EIB202 and

080813), while in the others, the sequence is significantly shorter (as in *E. tarda* species FL6-60 and *E. ictaluri* species 93-146 and ATCC33202).

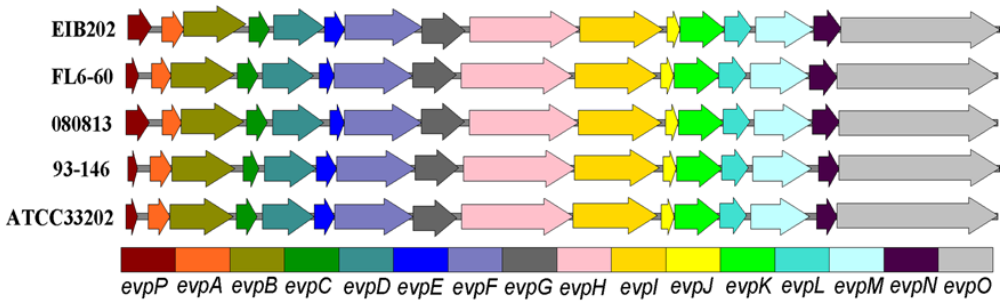


Figure 3.20, T6SS gene cluster from 5 *Edwardsiella* species. EIB202, FL6-60 and 080813 were *Edwardsiella tarda* species; 93-146 and ATCC33202 were *Edwardsiella ictaluri* species.

The *E. tarda* strain we were working with (130/91) belongs to the EdwGI type and the sequence alignment of EvpP among different species showed that our EvpP is a longer form, with particular differences if compared to other long-form EvpP proteins (Figure 3.21).

```
Aeromonas       MLTLNSELEWSDVGSICTDQKAKTYVSDAFNIAYGQPTKELLPAGTALYKFNGFSSLARS 60
EIB202          MLTLNSELEWSDVGSICTDQKAKTYVSDAFNIAYGQPTKELLPAGTALYKFNGFSSLARS 60
080813          MLTLNSELEWSGVGSICTDQKAKTYVSDAFNIAYGQPTKELLPAGTALYKFNGFNSLARS 60
FL6-60          ------------------------------------------------------------
130/91          MSILNSELDWSHVGSISTG--PGTVVSDAFNISYGLPTKELLPAGTALYKFNGFSSLARP 58

Aeromonas       PITDDSPLSPWWSPVQPFRYDGGLRQRMLIAKQNGVSMREWGRLTSVIKENWSSLDFLLE 120
EIB202          PITDDSPLSPWWSPVQPFRYDGGLRQRMLIAKQNGVSMREWGRLTSVIKENWSSLDFLLE 120
080813          PITDDSPLSPWWSPVQPFRYDGGLRQRMLIAKQNGVSMREWGRLTSVIKENWSSLEYLLE 120
FL6-60          -------------------------MLIAKQNGVSMREWGRLTSVIKENWSSLDFLLE 33
130/91          PITDDTPLSPWWSPVQPFRHDGGLQQRMLVAKLNGVSMREWGRLTSVIKENWSSLDHLLE 118

Aeromonas       IVLKIPVYAWFGGFKGMSRIDNGMTSKRNITLEQKGRSSMLPGGATQFYIPNLTVGHISS 180
EIB202          IVLKIPVYAWFGGFKGMSRIDNGMTSKRNITLEQKGRSSMLPGGATQFYIPNLTVGHISS 180
080813          ITLKVPVYAWFGGFKGMSRIDNGMTSRRNITLEQKGRSSMLPGGATQFYIPNLTVGHISS 180
FL6-60          IVLKIPVYAWFGGFKGMSRIDNGMTSKRNITLEQKGRSSMLPGGATQFYIPNLTVGHISS 93
130/91          IVLKVPVYAWFGGFKGMSRIDNGMPSKRNITLEQKGRGSNLPGGATQFYIPNLTVGHISS 178

Aeromonas       HQFSILK 187
EIB202          HQFSILK 187
080813          HQFSILK 187
FL6-60          HQFSILK 100
130/91          HNFSALK 185
```

Figure 3.21, Sequence alignment of EvpP among different *E. tarda* species. The one from *Aeromonas* species was possibly due to horizontal gene transfer from EIB202 (Xin, W, et al, 2009).

The sequence alignment clearly showed that EvpP is well conserved among all strains of *Edwardsiella tarda*. The overall sequence identity is 87%. Among them, strain FL6-60 has a shorter form of EvpP compared to other strains of *E. tarda*;

implying that there is a closer phylogenetic relationship to the *E. ictaluri* species because they tend to have shorter forms of EvpP (as shown in Figure 3.20).

EvpP from strain 130/91 is much different compared to the strains EIB202 and 080813, even though all three strains have long-form EvpP. EvpP from EIB202 and 080813 are very similar. They share more than 96% sequence identity and they both have 187 amino acid residues. Compared with these two strains, EvpP from 130/91 has a slightly shorter sequence of 185 amino acid residues. More importantly, it contains sequence variations in several well-conserved positions, namely Pro20, Leu34, Pro58, Leu91, Pro143 and Gly156 (Figure 3.21, marked red). In the above positions, all other EvpP proteins are identical while only EvpP from 130/91 has an alternative residue where these sequence variances are not conserved in terms of amino acid property (alanine to proline, glutamine to leucine and threonine to proline). Another noteworthy finding about EvpP from 130/91 is that, it contains significantly more proline residues than other strains.

It was intriguing to investigate each of the individual single residue point mutations on EvpP and the stability of resultant mutants. We mutated EvpP protein (strain 130/91) at the above positions and changed the amino acid residues to the conserved ones as in other strains of *E. tarda* to construct six single mutants, namely P20A, L34Q, P58S, L91Q, P143T and G156S. We have also attempted to sub-clone and express the wild-type EvpP protein from strain EIB202. Unexpectedly, the protein turned out to be insoluble after being over-expressed in *E. coli* BL21 (DE3) competent cells. Therefore we continued our research on EvpP (130/91) by investigating the stability of its single-residue mutants.

3.5.1 Characterization of the mutant EvpP_P143T

Six mutants (P20A, L34Q, P58S, L91Q, P143T and G156S of EvpP) were sub-cloned and tested for expression. Among them, L34Q did not show any expression. P20A, L91Q and G156S were found to have expression but they formed

inclusion bodies. P58S and P143T were found to have good expression and similar solubility compared to wild-type EvpP (Figure 3.22).

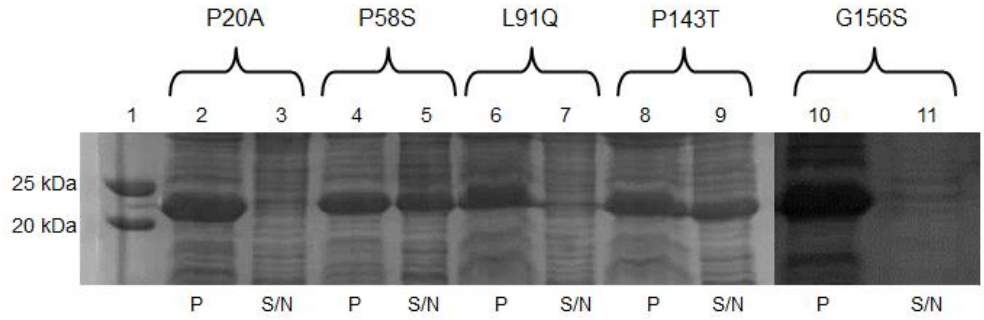


Figure 3.22, Expression profile of different single residue mutants of EvpP. P: pellet; S/N: supernatant. Among all 6 mutants, L34Q did not have expression (data not shown); P20A, L91Q and G156S were expressed but not soluble; P58S and P143T could be expressed and were soluble by themselves.

Mutant proteins of P20A, L91Q and G156S were solubilized and purified in buffers containing 8 M urea. They were later refolded back into buffers containing 20 mM Tris pH7.0 with no denaturing agent. We compared the stability of each of the individual mutants as well as the wild-type EvpP by urea denaturation experiments measured by fluorescence spectrometry (Figure 3.23).

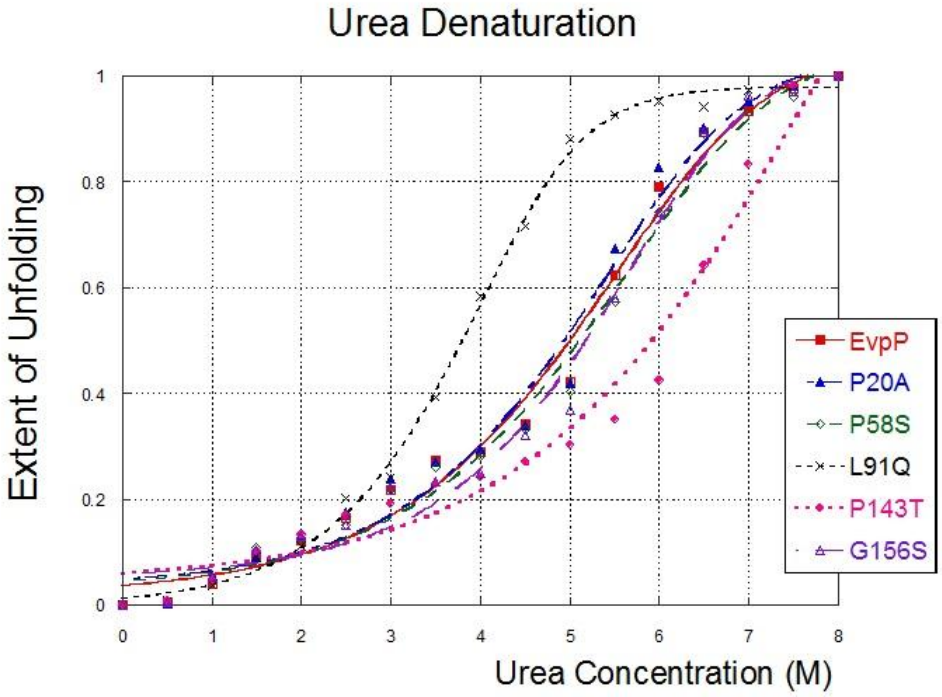


Figure 3.23, Urea denaturation of wild type of EvpP compared to 5 of its mutants. Curve fitting was performed using Kaleidagraph ver4.0

Shown in Figure 3.23, mutants P20A, P58S and G156S had similar stability towards urea denaturation like wild-type EvpP. The mutant L91Q was less stable, as the observed transition state was earlier than wild-type EvpP. Although the ΔG of L91Q was estimated to be 10.863 kCal/mol, which was comparable to wild-type EvpP (10.619 kCal/mol); the estimated m value for L91Q was much larger than wild-type EvpP (0.80) at 1.17. As leucine is a hydrophobic residue that might be buried in the structure of EvpP, it could be involved in forming hydrophobic interactions with other hydrophobic residues inside the "core" region. Once changed to a hydrophilic residue glutamine, such interaction would be disrupted, resulting in decreased stability of the protein. Interestingly, the mutant P143T showed an increased stability towards urea denaturation. It was estimated to have higher ΔG (11.15 kCal/mol) and lower m value (0.66) compared to those of wild-type EvpP. Threonine is a brunched amino acid with a polar side chain. It has a higher propensity to form interactions with the other residues compared to proline, which does not have a side chain. Proline has a few amino acid preferences for interaction, typically with aromatic residues like tryptophan or tyrosine, while threonine does not have such preference (Guilhem F, et al, 2008). Therefore new intra-molecular interactions formed by this mutation may be the reason for the improved stability of EvpP_P143T mutant protein. This has prompted us to further characterize P143T as a stable EvpP mutant.

EvpP mutant P143T has a similar molecular weight to the wild type. It was also eluted around 72 minutes through a Hiload Superdex 75 gel filtration column. Experiments by DLS showed that mutant P143T particles in solution had a hydrodynamic radius of 2.78 nm and an estimated molecular weight of 35.6 kDa. This was slightly less than the observed dimeric molecular weight of wild-type EvpP by DLS. However the estimated N_zH_z and $T_1\rho$ remained unchanged in P143T mutant, suggesting similar molecular movement compared to wild-type EvpP.

Limited protease digestion of P143T by trypsin has generated a slightly different pattern compared to wild-type EvpP (Figure 3.24). Before the digestion

came to a stable pattern (whereas wild type depicted a 10 kDa band and a lower smear of 4 kDa), there was an intermediate band that had a slightly higher molecular weight than the 10 kDa band. Although this band was soon digested and converged with the stable 10 kDa band; it nevertheless indicated a slower digestion pace of P143T compared to wild-type EvpP. This reduced pace of limited proteolysis might come from the increased intrinsic stability of mutant P143T.

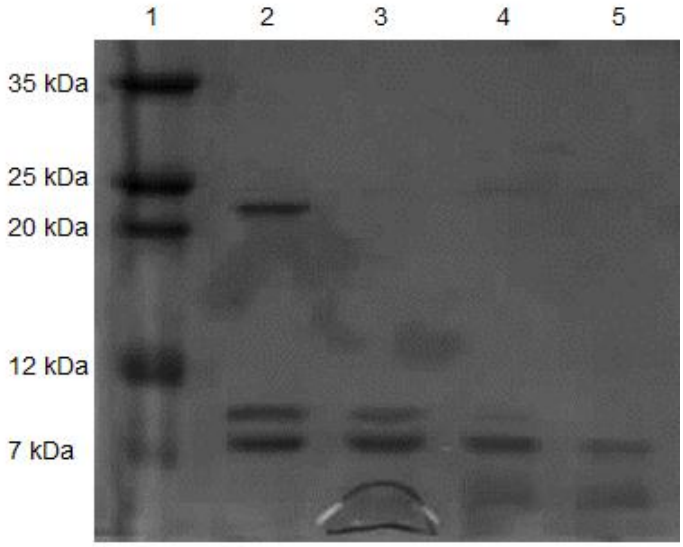
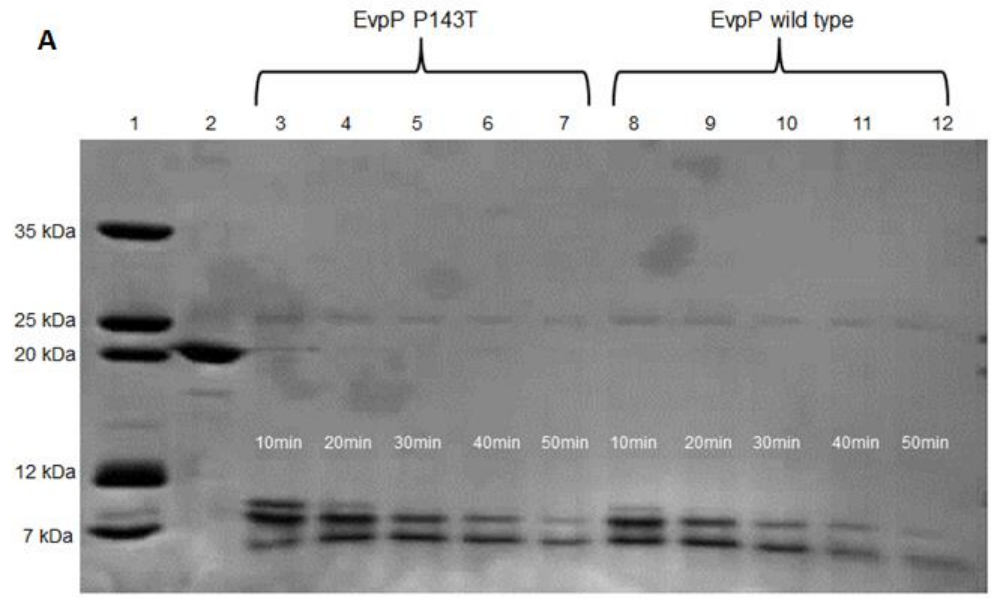
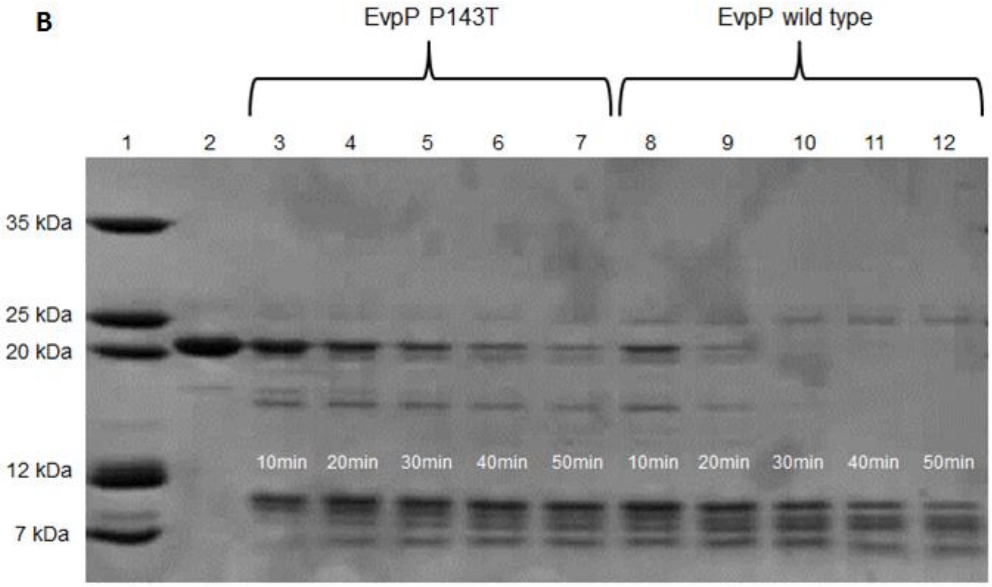


Figure 3.24, EvpP_P143T digested by trypsin (1:100). Lane 1, molecular weight marker; Lane 2 to Lane 5, EvpP mutant P143T digested by trypsin for 5 min, 10 min, 30 min and 60 min.

Mutant P143T also showed similar patterns after limited digestion by chymotrypsin, elastase and thermolysin (Figure 3.25). In all cases, it showed a slower "pace" towards proteolytic cleavage, which could be partially accounted by the increased stability of P143T.


A
EvpP P143T
EvpP wild type
1 2 3 4 5 6 7 8 9 10 11 12
35 kDa
25 kDa
20 kDa
12 kDa
7 kDa
10min 20min 30min 40min 50min 10min 20min 30min 40min 50min


B
EvpP P143T
EvpP wild type
1 2 3 4 5 6 7 8 9 10 11 12
35 kDa
25 kDa
20 kDa
12 kDa
7 kDa
10min 20min 30min 40min 50min 10min 20min 30min 40min 50min

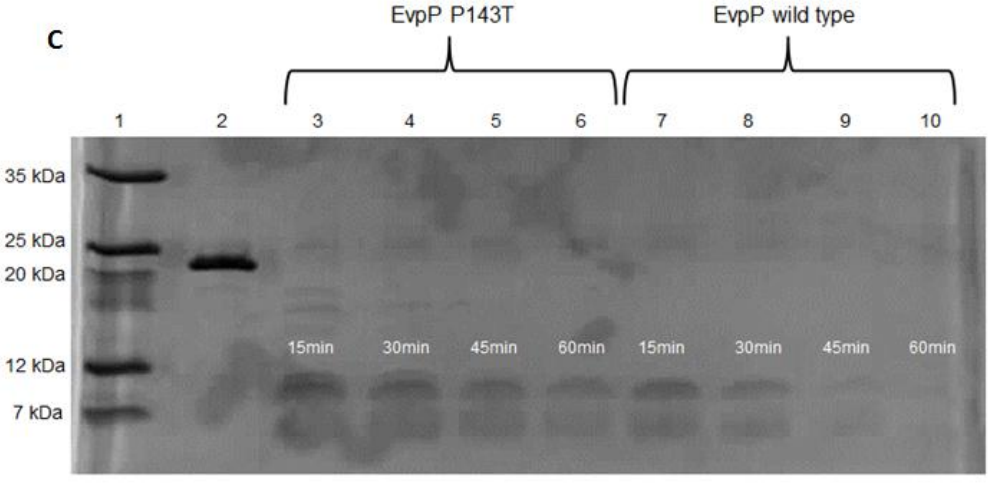

C
EvpP P143T
EvpP wild type
1 2 3 4 5 6 7 8 9 10
35 kDa
25 kDa
20 kDa
12 kDa
7 kDa
15min 30min 45min 60min 15min 30min 45min 60min

Figure 3.25, Limited protease digestion patterns of EvpP mutant P143T by chymotrypsin, elastase and thermolysin compared to wild type. Panel A, chymotrypsin digestion pattern of P143T and wild type EvpP; Panel B, elastase digestion pattern of P143T and wild type EvpP; Panel C, thermolysin digestion pattern of P143T and wild type EvpP.

3.6 Backbone assignment of EvpP_P143T

EvpP mutant P143T showed another significant improvement in protein stability, as it could be kept at 20 ℃ at a concentration of around 800 μM for more than a week before starting to precipitate. Wild-type EvpP could not last so long since previously we conducted the NMR triple resonance experiments HNCACB and CBCB(CO)NH within 3 days. Wild-type EvpP readily formed a significant amount of precipitation during the experimental process. This allowed us to re-investigate the backbone assignment of EvpP (P143T) using a 4D NMR strategy that requires a total set of five experiments: a ^{15}N-edited HSQC; ^{15}N-^{13}C-edited HNCA, HN(CO)CA, and 4D-NOESY as well as a ^{13}C-edited CCH-TOCSY.

The 4D strategy was designed for backbone and side chain assignment of large proteins without deuteration (Yingqi, X, et al, 2006). Although most of the triple resonance NMR experiments do not work for uniformly ^{15}N-^{13}C-edited large proteins (more than 25 kDa in molecular weight) due to their short relaxation time; experiments based on Nuclear Overhauser Effect (NOE) and multi-quantum ^{13}C-edited total correlation spectroscopy (MQ-CCH-TOCSY) do not rely too much on the relaxation rate. Previously owing to the dimeric characteristic of EvpP (also its mutant P143T) and the resultant short relaxation time, through-bond triple resonance experiments failed to yield a decent signal-to-noise ratio. However a through-space experiment, ^{15}N-edited NOESY, generated a moderate spectrum for EvpP. Therefore it was worth trying the 4D strategy on EvpP mutant P143T, which was able to withstand 20 ℃ for more than a week in order to acquire the 4D ^{15}N-^{13}C-edited NOESY spectrum.

Firstly, clusters were formed by grouping HSQC, HNCA, HN(CO)CA and 4D-NOESY correlated peaks according to their unique NH chemical shifts; as NH provided the amide spin pair ^{15}N and ^{1}H. In most cases, each individual cluster would represent one amino acid residue of the backbone sequence. This step was relatively non-ambiguous as both HNCA and HN(CO)CA did not contain many overlapping residues, Figure 3.26. Also, on the 4D-NOESY plane (located by the amide spin pair ^{15}N-^{1}H chemical shifts), $H_{C\alpha i}$-NH and $H_{C\alpha i-1}$-NH NOE peaks usually had strong peak intensities (Figure 3.27).

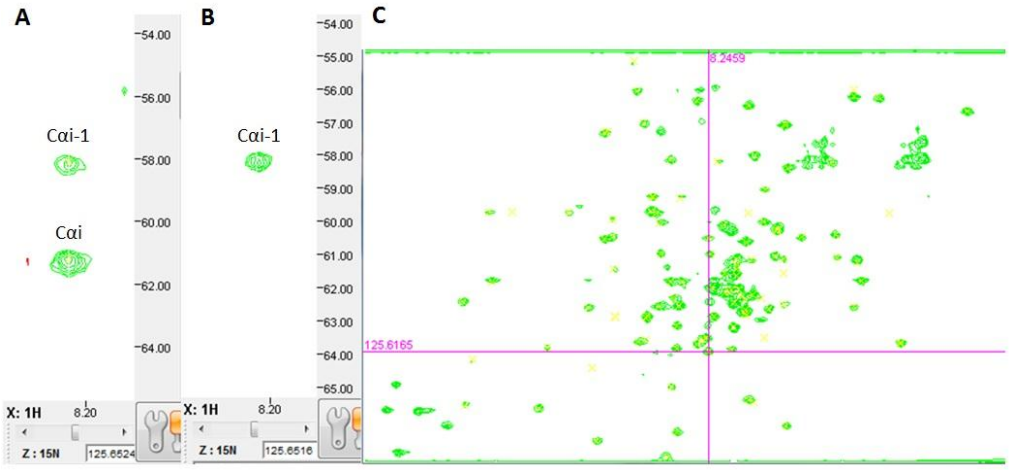


Figure 3.26, Cluster identification by correlating individual cross-peaks on the HSQC to the amide strips on both HNCA and HN(CO)CA according to the amide spin pair ^{15}N and ^{1}H. Panel A, corresponding amide strip of residue I3 on HNCA; Panel B, corresponding amide strip of residue I3 on HN(CO)CA; Panel C, HSQC of EvpP_P143T, the cross-peak position of residue I3 was located.

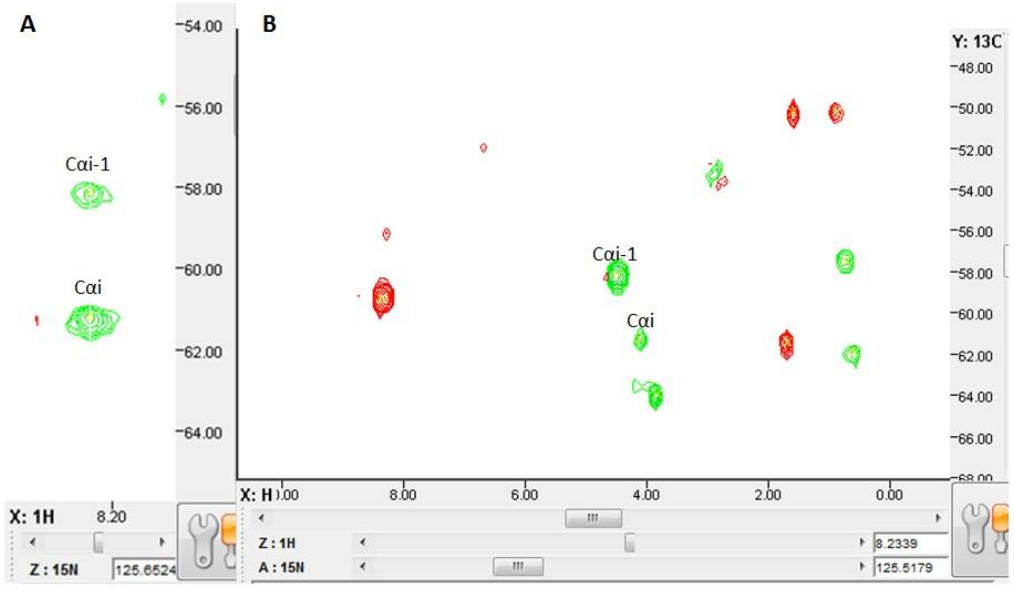


Figure 3.27, Cluster identification following grouping correlated peaks from HSQC, HNCA and HN(CO)CA; the HNCA peaks were reflected on the 4D-NOESY plane located by the amide spin pair ^{15}N-^{1}H. Panel A, corresponding amide strip of residue I3 on HNCA; Panel B, 4D-NOESY plane (XY plane) of residue I3 located by the amide NH chemical shifts in Axis A and Axis Z. Both Cαi and Cαi-1 could be found on the 4D-NOESY plane that had identical ^{13}C chemical shifts.

The second step is to identify the two spin systems within each cluster, one belongs to its own residue and the other belongs to its preceding residue. From CCH-TOCSY slices defined by the ^{13}C^{1}H spin pairs of various individual HC-NH NOEs, the CCH-TOCSY pattern of each NOE peak on the 4D-NOESY plane could be extracted. Those NOE peaks that gave rise to similar CCH-TOCSY patterns within which one of the ^{13}C chemical shift was identical to the HαCα-NH NOE or one of the HNCA peaks (previously identified clusters), were grouped into one spin system defined by that corresponding HNCA peak. Using this method, all intra-residue and sequential HC-NH NOE peaks within each cluster (representing a particular amino acid residue) could be separated from other inter-residue NOE peaks on the ^{15}N-^{13}C-edited 4D-NOESY spectrum. During the same process, an intra-residue and a sequential spin system could be established for each cluster. In addition, the amino acid type of both spin systems could be speculated based on the side chain ^{1}H and ^{13}C chemical shifts (Figure 3.28 and 3.29). Note that on the 4D-NEOSY plane, the side chain ^{13}C chemical shifts were usually folded as the observed spectral width was set to 22 ppm. Therefore the polarity of each HC-NH NOE peaks would be reversed each time the ^{13}C chemical shift was folded on the Y axis.

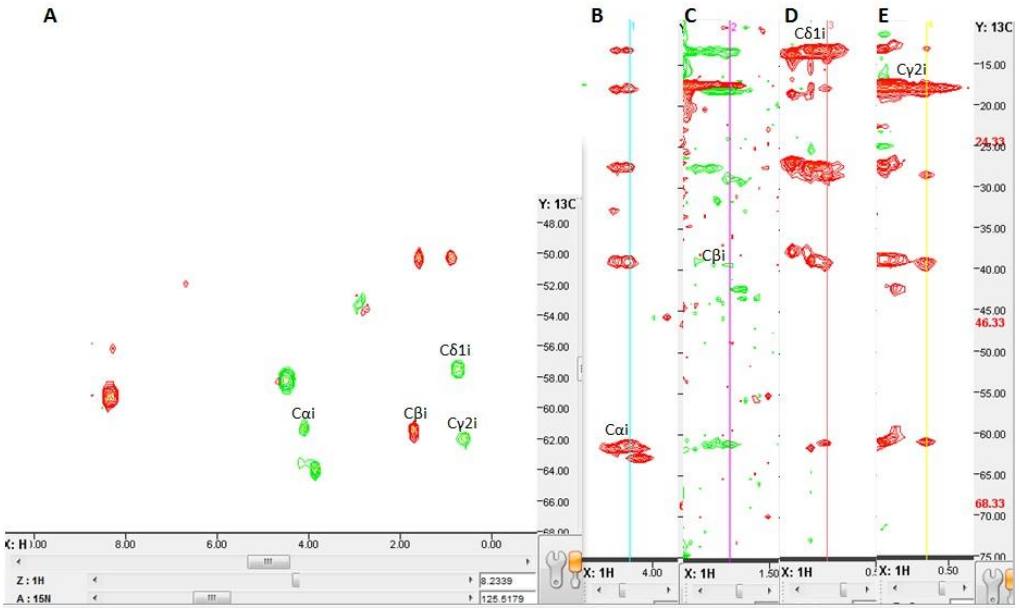


Figure 3.28, Spin system identification of residue I3 for its intra-residue NOE peaks. Panel A, 4D-NOESY plane of residue I3 located by the amide spin pair ^{15}N-^{1}H; Panel B to Panel E, corresponding CCH-TOCSY slices of individual intra-residue HC-NH NOEs on the 4D-NOESY plane: $HC_{\alpha i}$-NH, $HC_{\beta i}$-NH, $HC_{\delta 1}$-HN and $HC_{\gamma 2}$-NH.

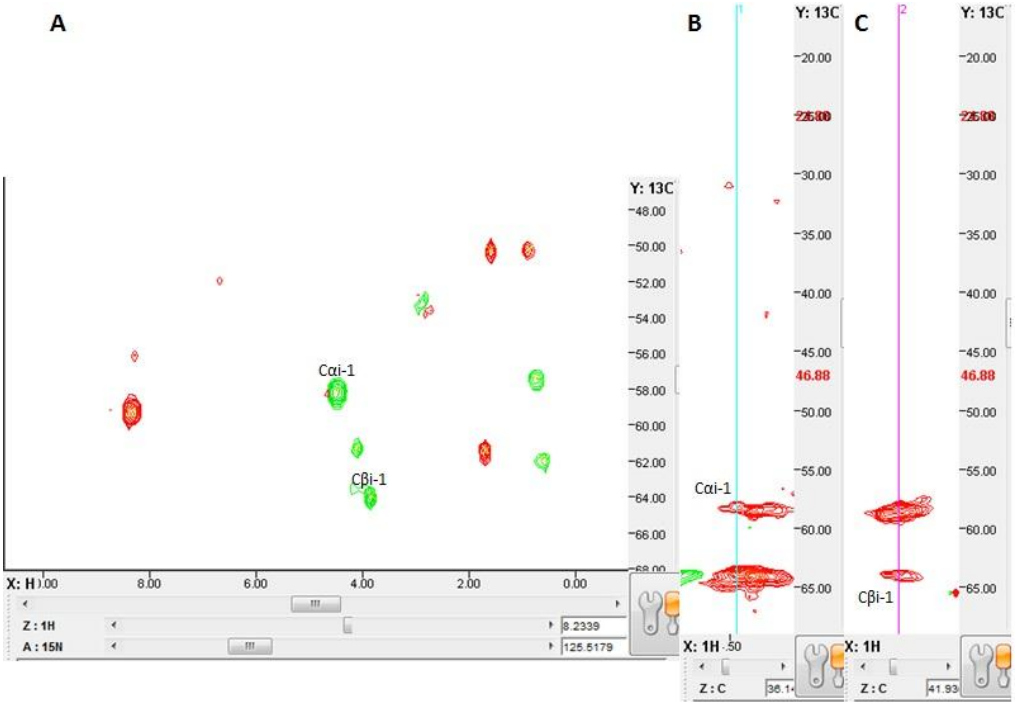


Figure 3.29, Spin system identification of residue I3 for sequential (i-1) NOE peaks. Panel A, 4D-NOESY plane of residue I3 located by the amide spin pair ^{15}N-^{1}H; Panel B and C, corresponding CCH-TOCSY slices of individual sequential HC-NH NOEs on the 4D-NOESY plane: $HC_{\alpha i-1}$-NH and $HC_{\beta i-1}$-NH.

Spin system identification was the rate-limiting step for backbone assignment. As the sole reference for grouping NOEs into a spin system, CCH-TOCSY spectrum was often crowded with overlapping problems due to the large number of amino acid

residues in protein sequence. Secondly, many clusters simply didn't have enough NOEs on the 4D-NOESY or the NOE peaks on the 4D-NOESY had very poor signal-to-noise ratio. This was further complicated by the presence of inter-residue NOEs where some of them were rather strong on the 4D-NOESY spectrum.

The next step is to link different clusters to form fragments and map the fragments to the sequence of EvpP_P143T. Intuitively, two clusters were compared with each other to examine whether they could construct a dipeptide from the two consecutive clusters. The dipeptides were further linked with other clusters to form longer fragments, which were finally mapped into protein sequence. Usually two consecutive clusters A and B would fulfill the following criteria (Zheng Yu, PhD thesis, development of NMR methods for the structural elucidation of large proteins):

1. The HNCA peak $C_{\alpha i}$ of A should have same ^{13}C chemical shift value with the $C_{\alpha i-1}$ of B, or vice versa (Figure 3.30).
2. The $H_{\alpha i}C_{\alpha i}$-NH NOE of A on 4D-NOESY should have same ^{1}H and ^{13}C chemical shift values with the $H_{\alpha i-1}C_{\alpha i-1}$-NH NOE of B, or vice versa (Figure 3.31).
3. A and B should have some common inter-residue NOEs.

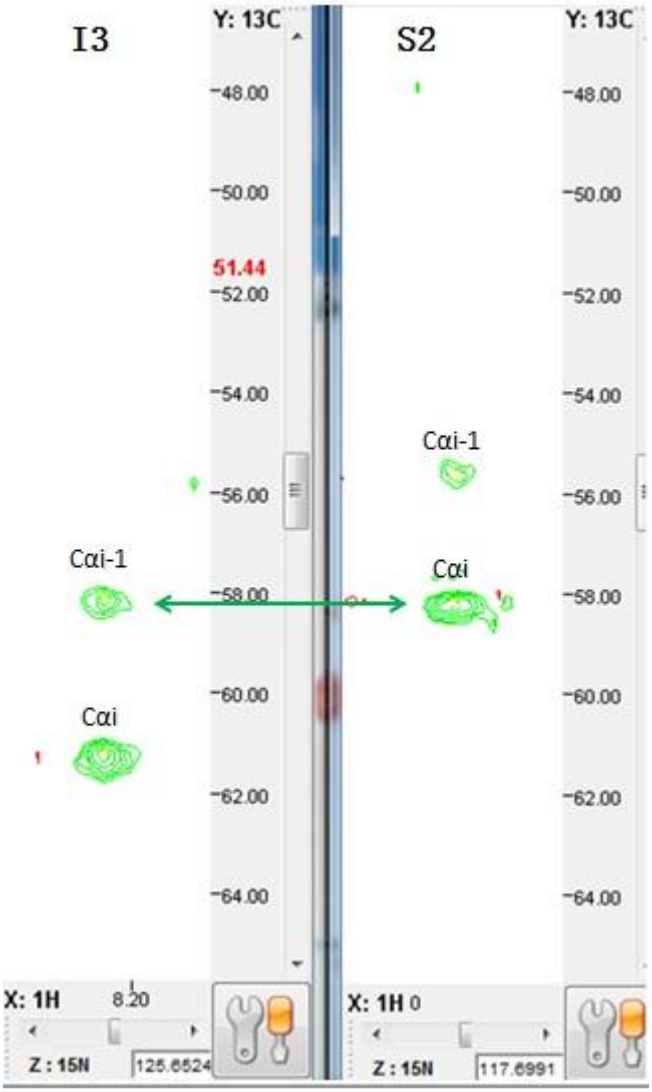


Figure 3.30, Cluster mapping: amide strips on HNCA from residue I3 and its preceding residue S2. The $C_{\alpha 1\text{-}1}$ chemical shift value of residue I3 matched the $C_{\alpha i}$ of residue S2.

Figure 3.31, Cluster mapping: the 4D-NOESY planes of residue I3 and its preceding residue S2 were compared. The $HC_{\alpha i-1}$-NH NOE peak of residue I3 matched the $HC_{\alpha i}$-NH NOE peak of residue S2. The $HC_{\beta i-1}$-NH NOE peak of residue I3 matched the $HC_{\beta i}$-NH NOE peak of residue S2. They both had sequential NOEs coming from the NH of each other.

Using this method, we were able to assign 89 cross-peaks on the ^{15}N-^{13}C-edited HSQC. Almost all the peaks on the HSQC could be assigned to EvpP_P143T sequence except for a few that had very weak signal in both HNCA and 4D-NOESY spectra. The assigned residues were labelled out on the HSQC. Some of the cross-peaks were found to come from the Histidine-tag preceding the N-terminal sequence of EvpP_P143T. These peaks were not labelled out on the HSQC (Figure 3.32).

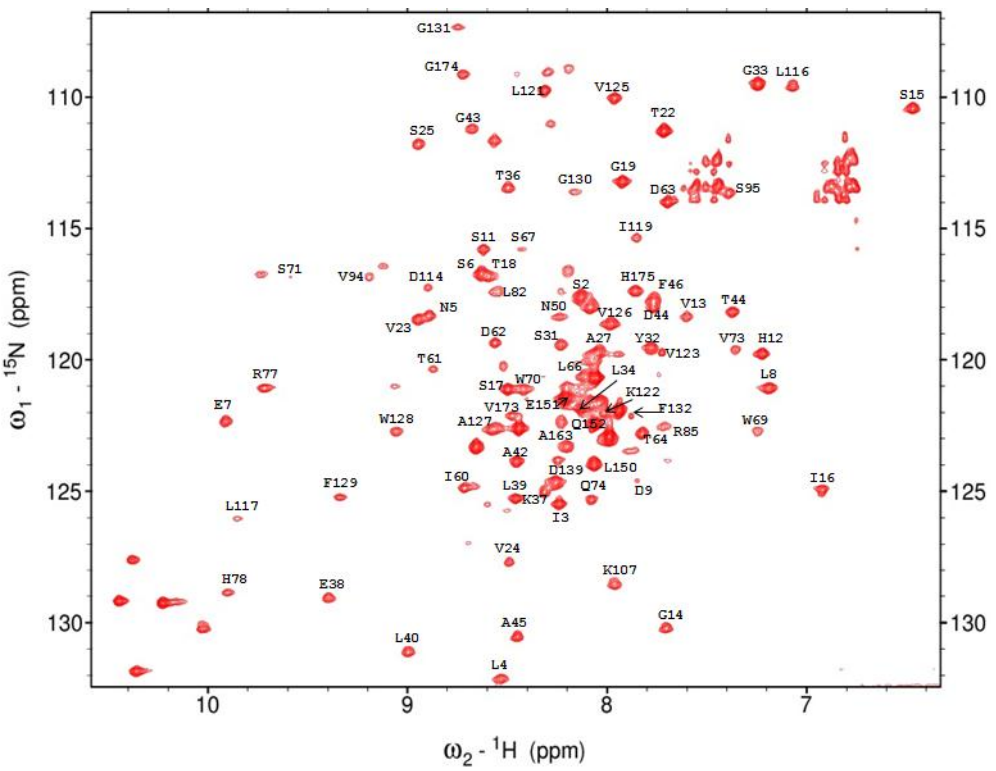


Figure 3.32, ^{1}H-^{15}N HSQC of ^{15}N-^{13}C-edited EvpP_P143T, with the assigned residues labelled out. Some of the peaks (usually with very strong signal intensity) were found to come from the His-tag preceding EvpP_P143T amino acid sequence. These peaks were not labelled in the above HSQC.

The assigned residues were mapped on the sequence of EvpP-P143T against the trypsin digestion boundary (Figure 3.33). The great majority of peaks (80 of them) were found to be within the trypsin-digestion boundary, which agrees with our previous hypothesis. The assignment was almost complete for the N-terminal half except for a few unassigned residues. Also, the presence of nine proline residues further broke the assigned amino acid chains into fragments. There was a gap of assignment from residue Tyr47 to residue Pro59, suggesting a possible internal region of flexibility within the N-terminal half of EvpP_P143T.

```
        10         20         30         40         50         60
MSILNSELDW SHVGSISTGP GTVVSDAFNI SYGLPTKELL PAGTALYKFN GFSSLARPPI

        70         80         90        100        110        120
TDDTPLSPWW SPVQPFRHDG GLQQRMLVAK LNGVSMREWG RLTSVIKENW SSLDHLLEIV

       130        140        150        160        170        180
LKVPVYAWFG GFKGMSRIDN GMTSKRNITL EQKGRGSNLP GGATQFYIPN LTVGHISSHN

FSALK
```

Figure 3.33, Sequence of EvpP_P143T with the assigned residues mapped out in pink; the trypsin-digestion boundary was marked by red box.

On the other hand, nine assigned residues were found outside the trypsin-digestion boundary. Among them, the C-terminal stretch “VGH” was still found present on the HSQC acquired for the trypsin-digested EvpP (Figure 3.12), while the others did not show up any more. In order to account for the presence of these residues on HSQC even after trypsin digestion, we examined the residues “VGH” and found that they were forming long range NOEs with residues on the N-terminal half of EvpP_P143T. Specifically, residue V173 had long-range NOEs with residue T64 (Figure 3.34) while residue G174 had long-range NOEs with residue D62 (Figure 3.35). Both residues T64 and D62 were located at the N-terminal half of EvpP_P143T within the trypsin-digestion boundary.

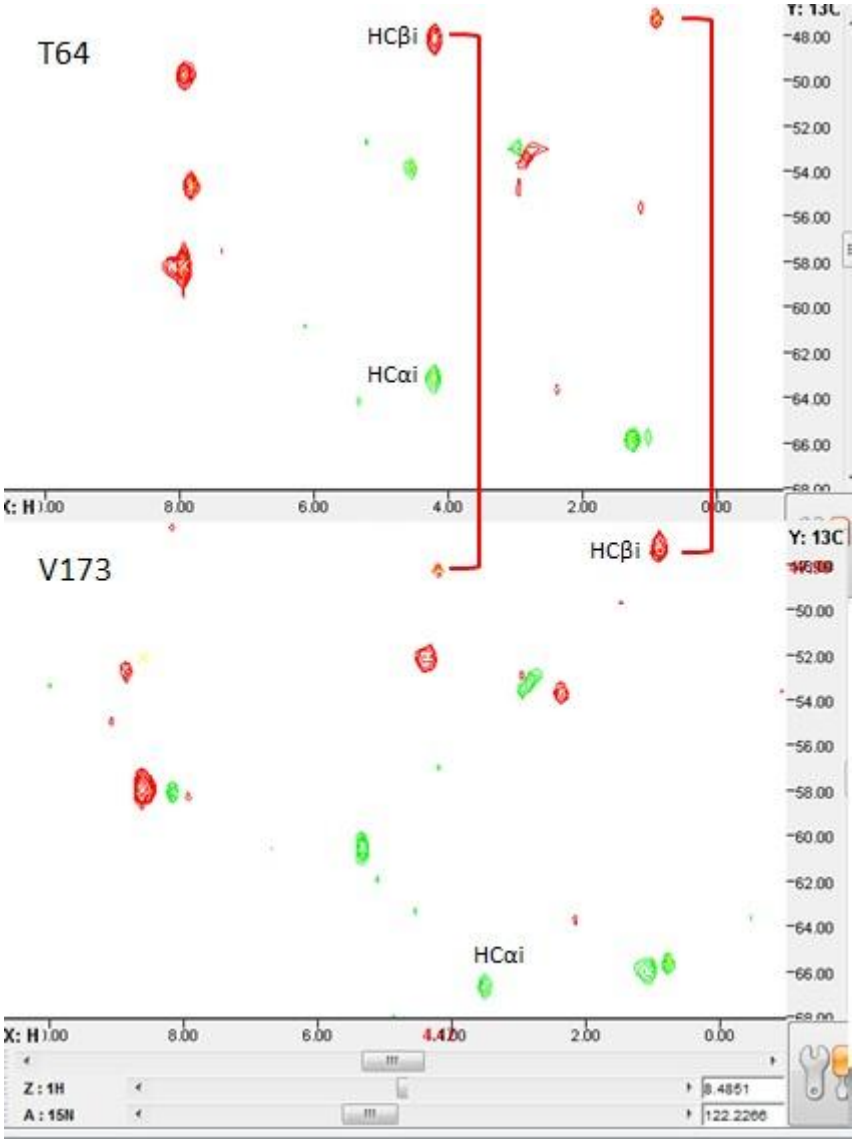


Figure 3.34, 4D-NOESY planes of residues T64 and V173 were shown next to each other with the inter-molecular long range NOEs marked out between these two residues.

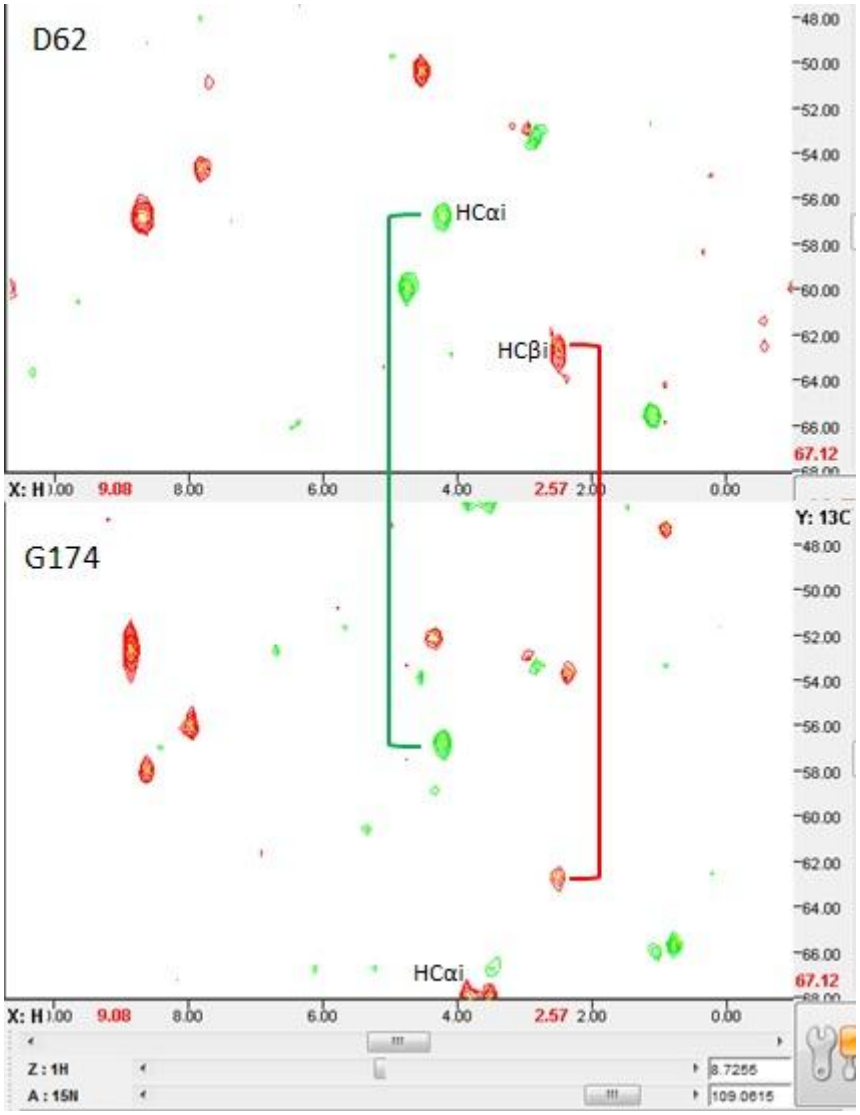


Figure 3.35, 4D-NOESY planes of residues D62 and G174 were shown next to each other with the inter-molecular long range NOEs marked out between these two residues.

Therefore it was logical to speculate that the C-terminal stretch "VGH" could be forming interactions with the N-terminal half of EvpP_P143T, probably at the region from residues D62 to T64. Such interaction would protect these residues from limited proteolysis by trypsin and after digestion, they could still be observed on the HSQC. However, they could not be detected on the SDS-PAGE of trypsin-digested EvpP. As in a denaturing condition, the interaction between these two regions would be disrupted and this C-terminal stretch was too short to be visible on the SDS gel.

3.6.1 Heteronuclear NOE experiment on EvpP mutant P143T

To further verify that residues showing up on the HSQC of EvpP account for the rigid regions, we performed heteronuclear NOE experiment using ^{15}N-edited EvpP_P143T with a relatively complete backbone assignment. The heteronuclear ^{1}H-^{15}N NOE experiment was carried out in an interleaved manner, with and without proton saturation (NOE_{on}, NOE_{off}).

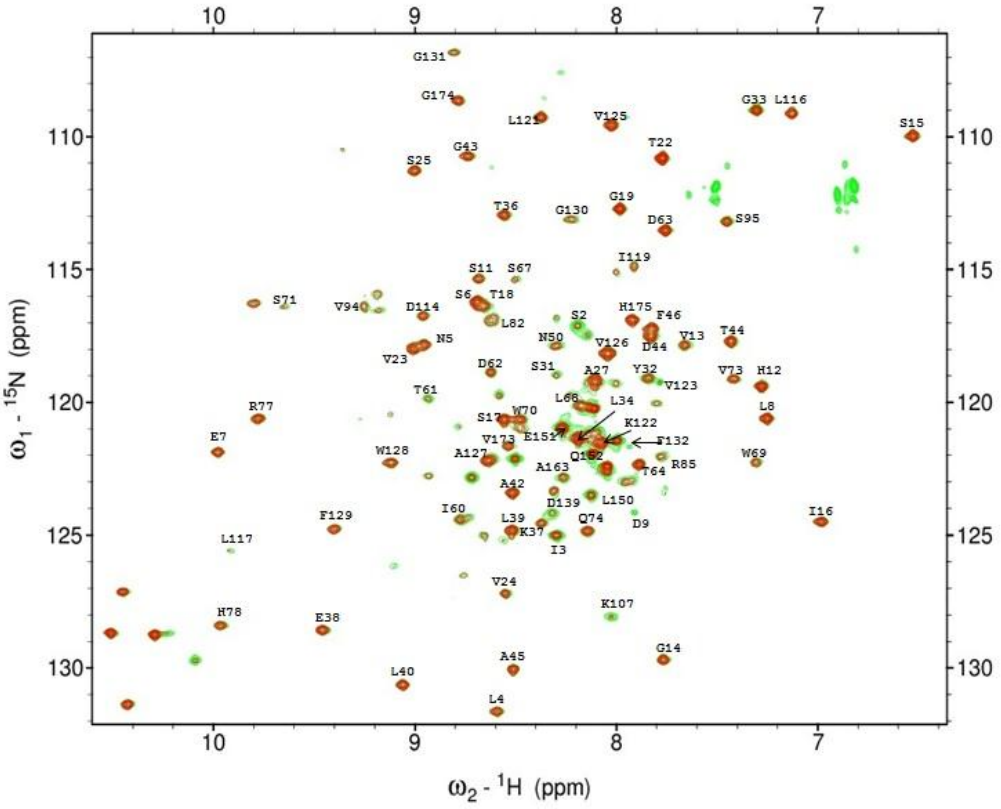


Figure 3.36, Overlay of the two ^{1}H-^{15}N heteronuclear NOE spectra. Peaks from the experiment without proton saturation were shown in green while peaks from the proton-saturated experiment were shown in red. The decrease in observed peak intensity described the flexibility of that particular residue.

The two HSQC-like spectra were overlaid for peak intensity comparison (Figure 3.36). With proton saturation, decreases in peak intensities of individual residues ranged from 20% to as much as 80%. No negative peak was observed on the proton-saturated spectrum except for the peaks from side chains of amino acids. Among the assigned peaks, residues at the very N-terminal end (such as Ser2) as well as some His-tag residues (not labelled out) were found to exhibit a significant reduction in peak intensity compared to other residues. Peaks without backbone

assignment were found to have greater reduction in intensity as well, with some of them even becoming invisible. Taken together, these findings have suggested that the backbone assignment of EvpP_P143T harbours the rigid regions of EvpP which was found to be within the trypsin-digestion boundary.

3.6.2 NOE pattern and possible secondary structure adopted by EvpP_P143T

The NOE pattern of EvpP_P143T was examined using both 3D and 4D ^{15}N-^{13}C-edited NOESY spectra. Sequential NOEs were observed for almost all assigned residues. Most $d_{\alpha N}(i, i+1)$ NOEs were observed to have comparable or stronger signal intensities than the corresponding intra-residue $d_{\alpha N}(i, i)$ NOEs. The $d_{NN}(i, i+1)$ NOEs were also observable for most assigned residues while medium range NOE could hardly be found (Figure 3.37).

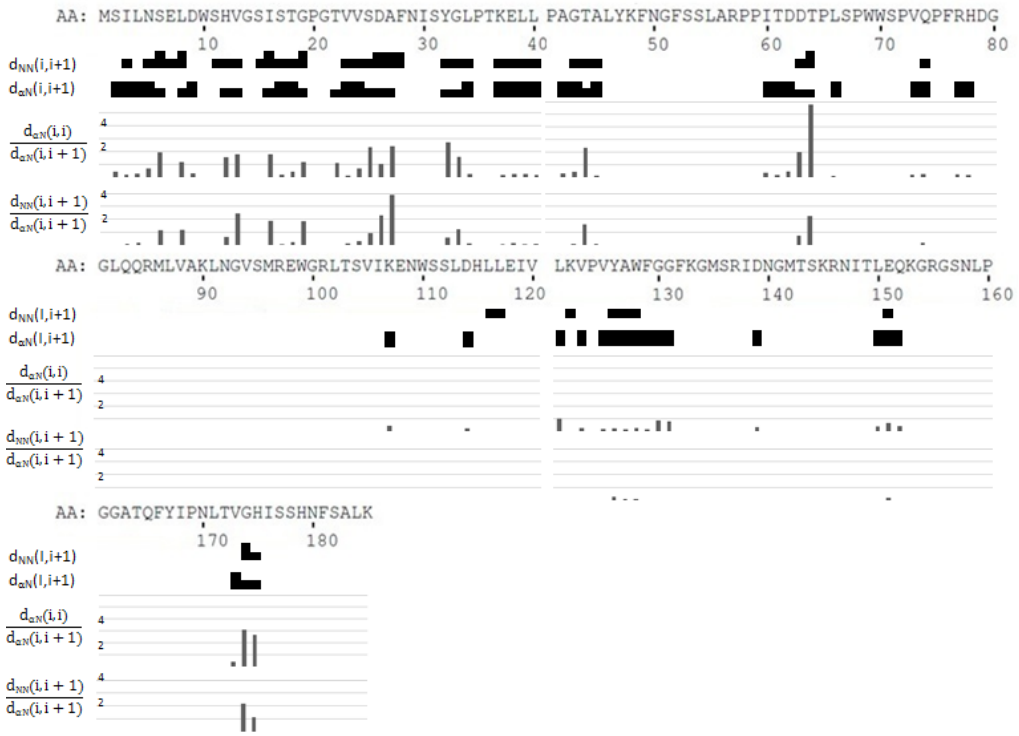


Figure 3.37, Sequential NOE pattern of EvpP_P143T. The relative signal intensities of $d_{\alpha N}(i, i+1)$ and $d_{NN}(i, i+1)$ were depicted by the thickness of the black bars. The ratio between $d_{\alpha N}(i, i)$ and $d_{\alpha N}(i, i+1)$, as well as the ratio between $d_{NN}(i, i+1)$ and $d_{\alpha N}(i, i+1)$ were also plotted against the sequence.

As shown in Figure 3.37, EvpP_P143T contains regions with extended backbone structure: S2 to N5, L34 to G43, I60 to D62, V73 to H78, V123 to F129

and L150 to Q152. In these regions, the intensity ratios of $d_{\alpha N}(i, i)/d_{\alpha N}(i, i+1)$ were significantly less than 0.44 (typical for random coil structure). The ratios of $d_{NN}(i, i+1)/d_{\alpha N}(i, i+1)$ were also much less than 0.71 (typical for random coil structure) (Ma K, et al, 2001). Whether they form β-sheets requires further experiments to prove. However the NOE data agreed with the CD spectra that EvpP adopted β-like extended structure in certain regions rather than a totally random structure. In addition, no α-helices could be predicted.

3.7 HDX-MS experiment of EvpP and its mutant P143T

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) experiment provided another approach to probe the rigid as well as flexible regions of EvpP (and its mutant P143T). Both wild-type EvpP and the mutant P143T were used to carry out the experiments. Wild-type EvpP generated a library of 117 peptide fragments with 99.5% sequence coverage while mutant P143T generated 119 of them, with sequence coverage of 98.9%. The dispersion of all the peptides among the sequence of EvpP (as well as P143T) was mapped by a shot-gun plot (Figure 3.38). A similar global pattern could be observed between EvpP and EvpP_P143T. In both proteins, peptides from the N-terminal half were usually longer while shorter peptides were found from the C-terminal half. Comparing EvpP with EvpP_P143T, it was obvious that EvpP_P143T had more longer peptides than wild-type EvpP. Particularly, at the site of mutation (residue 143), peptides from EvpP appeared to be more “fragmented” than those from EvpP_P143T. These findings have intuitively suggested that EvpP_P143T was more stable than EvpP towards complete proteolytic digestion by pepsin (the sample proteins were subjected to pepsin cleavage in the retention chamber for 3 minutes prior to the liquid chromatography followed by ESI-MS). Secondly, the N-terminal half was more protected from such digestion than the C-terminal half in both proteins.

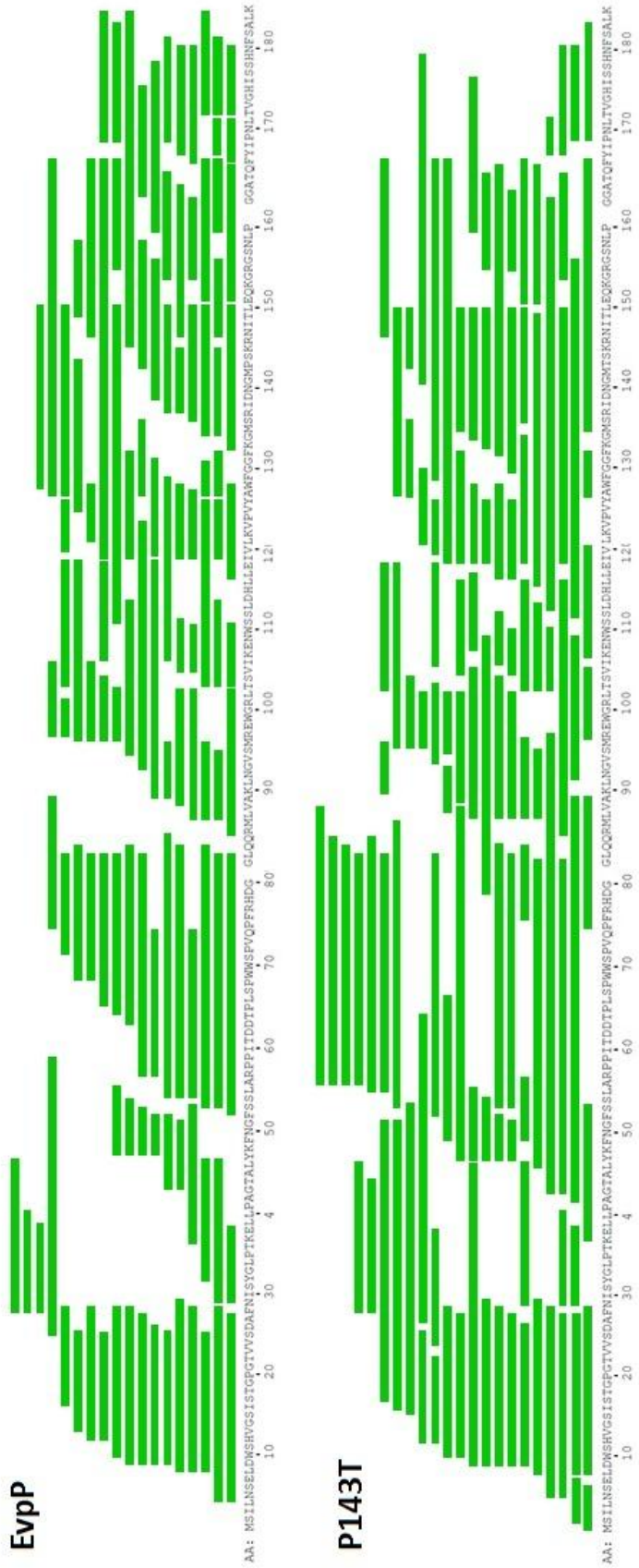
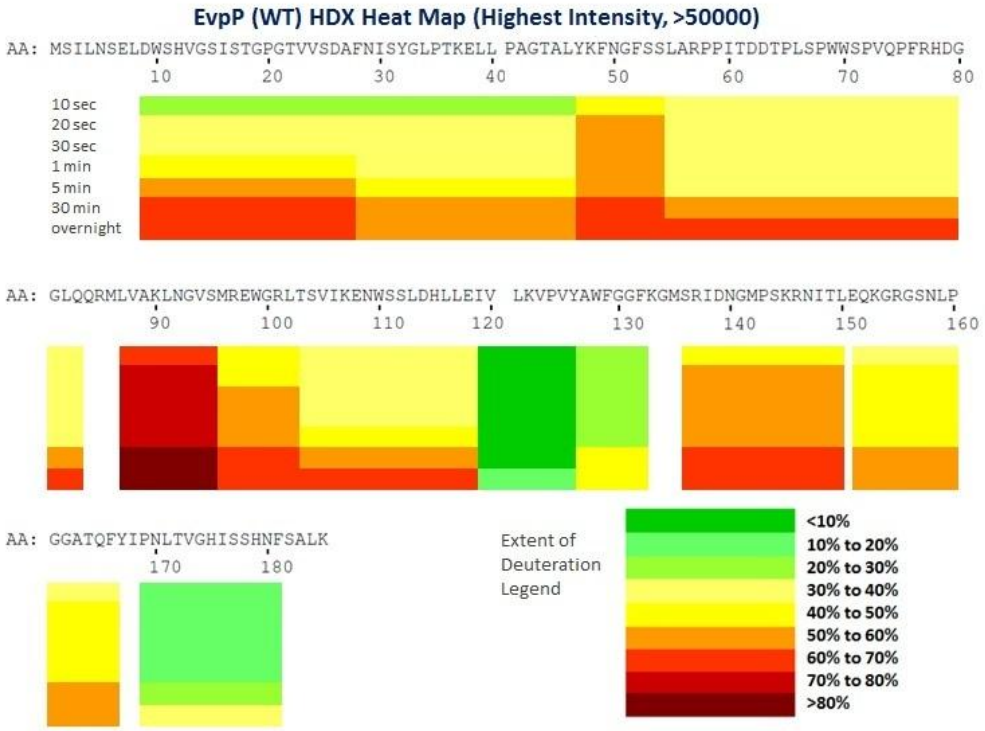


Figure 3.38, Shot gun plots of all the observed peptides of EvpP and EvpP_P143T that were found to have decent signal intensities in DynamX. The complete sequence of wild type EvpP and EvpP_P143T were shown at bottom. Peptide fragments were mapped against respective protein sequences by green bars.

To further characterize the H/D exchange patterns of EvpP and EvpP_P143T, a subset of peptides, spanning the entire protein sequence (with least overlapping residues), was chosen from each EvpP or EvpP_P143T peptide library to plot a "heat map". The extent of deuteration for individual peptides was plotted in a time-dependent manner, and the peptide was mapped onto the sequence of respective proteins to help visualize the global pattern. The criteria for choosing each subset of peptides were as follows:

1. Subset A comprised of peptides having the highest signal intensities from the mass spectrum.
2. Subset B comprised of peptides that were considerably short so that the heat map plotted by this subset would have a relatively higher "spatial resolution".
3. Subset C comprised of peptides that were chosen to fit the trypsin-digestion boundary of EvpP or Evp_P143T.

Heat maps plotted in the above 3 ways were shown in Figure 3.39 to 3.41. The heat maps plotted for EvpP and EvpP_P143T were compared side by side.



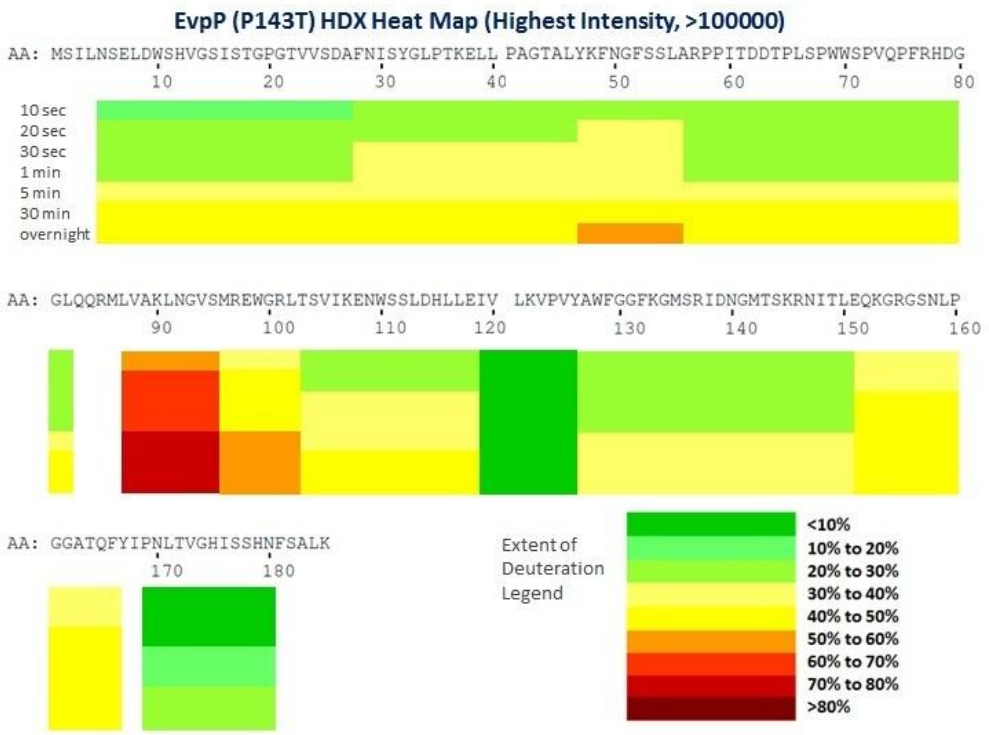
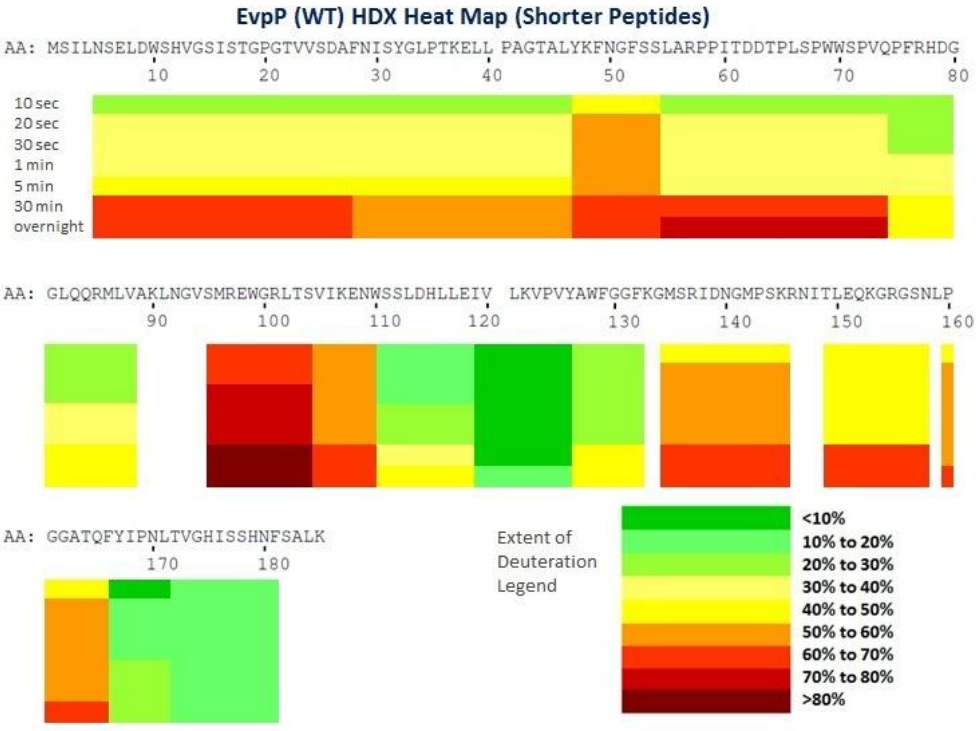


Figure 3.39, Heat maps showing H/D exchange patterns of EvpP and EvpP mutant P143T. The peptides were chosen from subset A where they had highest signal intensities from respective mass spectra. Upper panel, heat map plotted for EvpP; Lower panel, heat map plotted for EvpP mutant P143T.



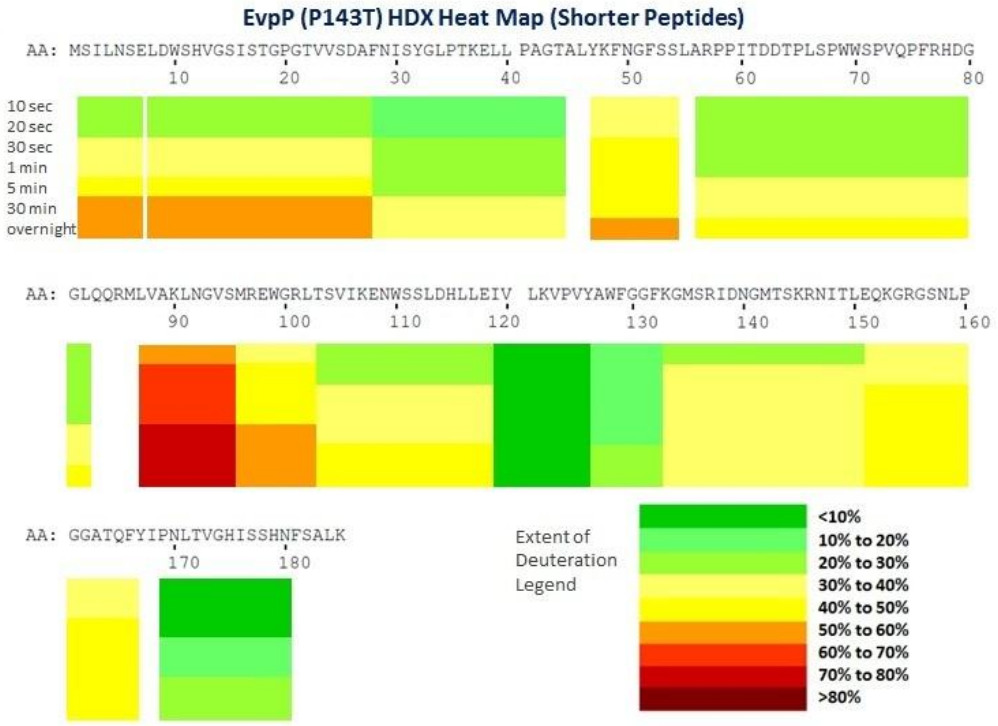
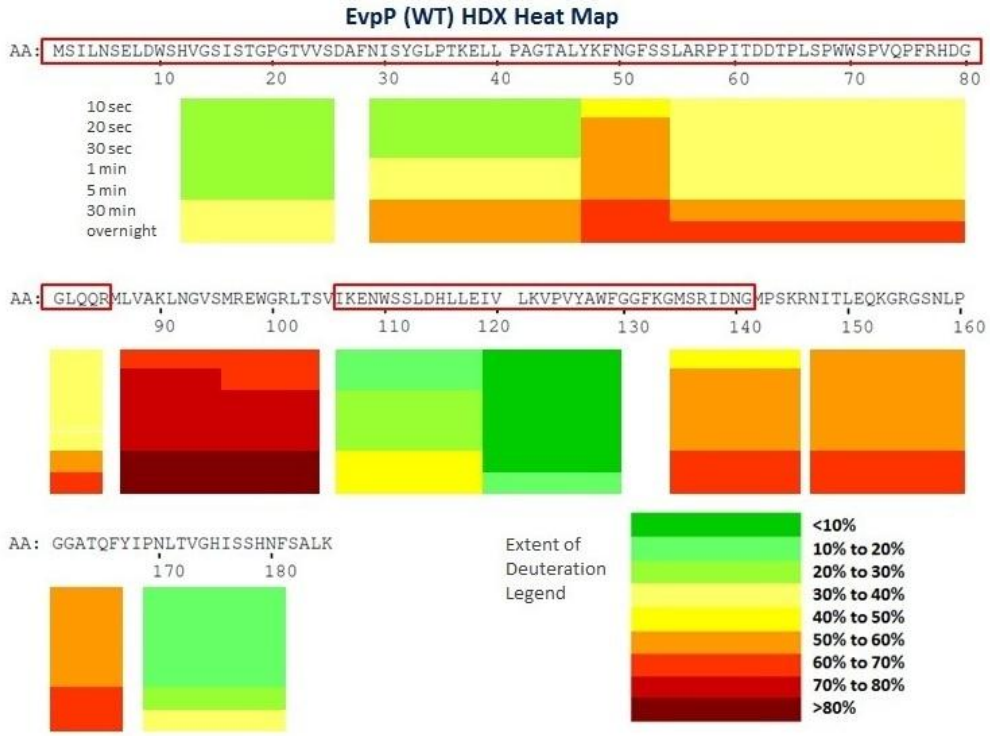


Figure 3.40, Heat maps showing H/D exchange patterns of EvpP and EvpP mutant P143T. The peptides were chosen from subset B where they were considerably short in order to construct a plot with relatively higher "spatial resolution". Upper panel, heat map plotted for EvpP; Lower panel, heat map plotted for EvpP mutant P143T.



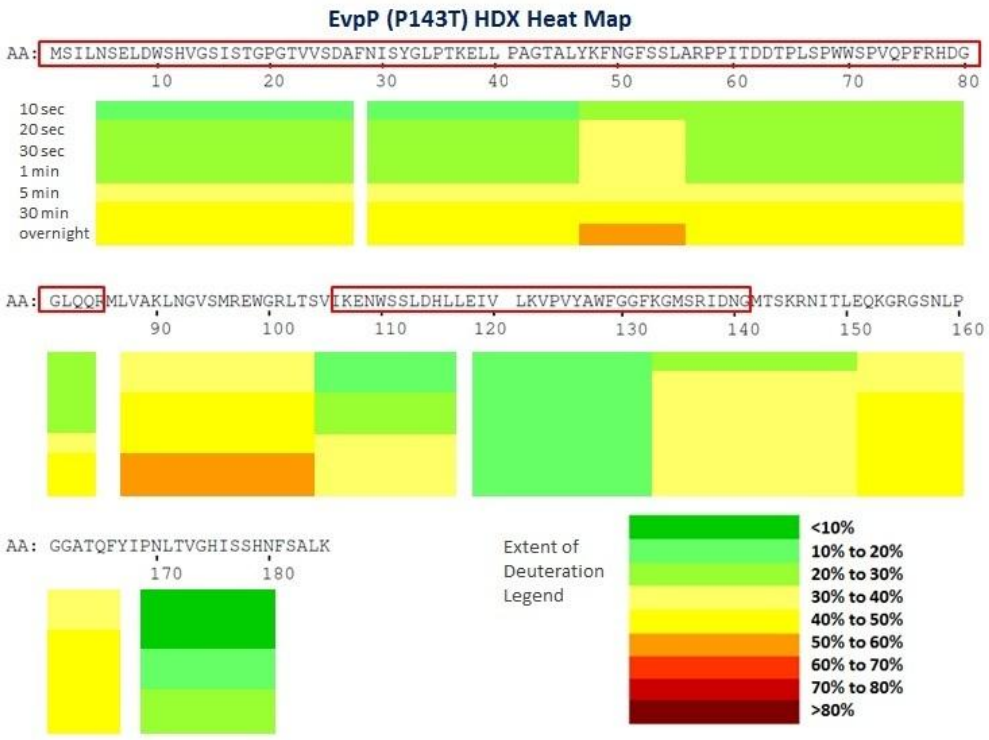


Figure 3.41, Heat maps showing H/D exchange patterns of EvpP and EvpP mutant P143T. The peptides were chosen from subset C where they were tailored to the trypsin-digestion boundary. Upper panel, heat map plotted for EvpP; Lower panel, heat map plotted for EvpP mutant P143T.

The three heat maps of EvpP demonstrated a coherent pattern of H/D exchange. Comparing these maps, the extensively-deuterated peptides all pointed to three convergent regions of EvpP, suggesting that these three regions would be considerably more flexible than their peripherals. The most flexible region was found to be from residue Leu87 to residue Ser104. This region underwent extensive H/D exchange (as much as more than 80% deuteration) in all heat maps. This is possibly because this was the linker region between the N-terminal 10-kDa piece and the C-terminal 4-kDa piece marked by the trypsin-digestion boundary. A second flexible region was observed to start from residue Met135 to residue Thr149. Although EvpP was generally found to be flexible from residue Met135 all the way to residue Phe166; the fragment from Met135 to Thr149 appeared to be more flexible than its flanking residues. The third flexible region was found to be from residue Tyr47 to residue Ser54. This short stretch was found within the N-terminal trypsin digested boundary

of EvpP. Interestingly, residues from this region were also missing from the HSQC acquired for EvpP; since the assigned backbone of EvpP_P143T did not cover this region. However this region was not digested in limited proteolysis, possibly because it is short in amino acid sequence, making it inaccessible by various proteases.

On the other hand, the heat maps also pointed to three rigid regions of EvpP. Firstly, a common pattern was observed in all heat maps plotted above. The N-terminal half of EvpP had an overall HDX extent that was significantly lower than those in the flexible regions, especially in the initial time points (10 seconds, 20 seconds and 30 seconds). A second distinctive rigid region came from residue Ser111 to residue Phe132. This region matched our previous assignment on the 4-kDa piece of EvpP after limited trypsin digestion. Unexpectedly, the most rigid region throughout the entire EvpP sequence came from residues Ile119 to Tyr126, but not in the N-terminal half. The third rigid region was found to be from residue Tyr167 to residue Phe181. This was also to our surprise that the extent of deuteration was less than most of the N-terminal regions of EvpP. The presence of such rigid region could possibly be correlated to the presence of the three assigned residues at the C-terminus of EvpP. It was noticed that the C-terminal residues Val173 and Gly174 were having long range NOEs with residues Thr64 and Asp62. Residues flanking V173 and G174 could also be interacting with residues near T64 and D62; or they could be interacting with residues from other rigid regions of EvpP, contributing to the tertiary structure. This might also explain why previously all the truncation mutants of EvpP were unstable and inevitably formed inclusion bodies. Unfortunately, only the C-terminal residues V173, G174 and H175 could be observed on the HSQC. Nevertheless, the rigid regions shown by the HDX-MS data further supported our previous hypothesis on the rigid regions of EvpP. It also helped to explain the presence of "VGH" on the HSQC even after trypsin digestion.

The heat maps of EvpP_P143T also pointed to three flexible regions coherently. Similar to EvpP, the most flexible region was found to come from

residues Leu87 to Leu102. Again this was within the linker between the two pieces coming out of limited trypsin digestion. The second flexible region was found at the N-terminal half, from residues Tyr47 to Leu55. This was the same as the third flexible region observed in wild-type EvpP. The third flexible region was found to come from residues Glu151 to Phe166. This was different from wild-type EvpP, as the same peptide in the latter was found to be less flexible than its preceding regions (from Met135 to Thr149). Therefore a significant change could be observed between wild-type and mutant EvpP. The region around the mutation point (Residue 143) has become less flexible due to the change of a single amino acid, from proline to threonine. This change could possibly be the reason why the mutant P143T was found to be more stable than wild-type EvpP.

Similar to wild-type EvpP, the rigid regions of the mutant P143T also converged to the same locations. The most rigid region of EvpP_P143T was found to come from residues Ile119 to Tyr126. The second rigid region was found at the C-terminus, from residues Pro169 to Asn180. Again this was the same as wild-type EvpP. Nevertheless, the entire N-terminal half had a lower extent of H/D exchange compared to the flexible regions.

3.7.1 Comparison between wild-type and mutant EvpP by HDX-MS revealed the major regions stabilized by the mutation P143T

By comparing the heat maps of wild-type EvpP with those of mutant P143T, it was noticed that the mutant P143T underwent less H/D exchange than wild-type EvpP. To systematically characterize the difference between EvpP and its mutant, a horizontal comparison between wild-type EvpP and EvpP_P143T was conducted using the two peptide libraries generated by HDX-MS. The two were compared against each other in search for common peptides that were present in both libraries. As a result, a total of 51 common peptides were identified. The differences in the extent of HDX for these peptides at various time-points were determined by

subtracting the HDX extent of wild-type EvpP's peptides from that of EvpP_P143T. The results were then plotted into a difference plot (Figure 3.42). A positive number indicates that this peptide undergoes less H/D exchange in mutant P143T than in wild-type EvpP, which further implies a decrease in the flexibility of the peptide's backbone in that particular region.

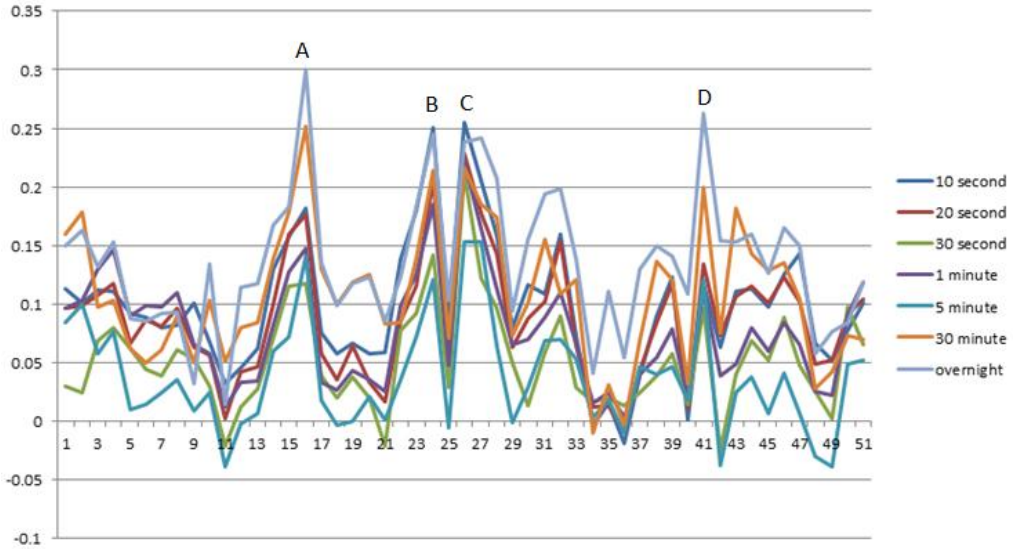


Figure 3.42, Difference plot generated for all 51 common peptides between wild type EvpP and its mutant P143T.

As shown by figure 3.42, four distinct peptides of P143T were identified to have a considerably decreased extent of H/D exchange compared to wild-type EvpP. The four peptides were mapped onto EvpP_P143T sequence in Figure 3.43.

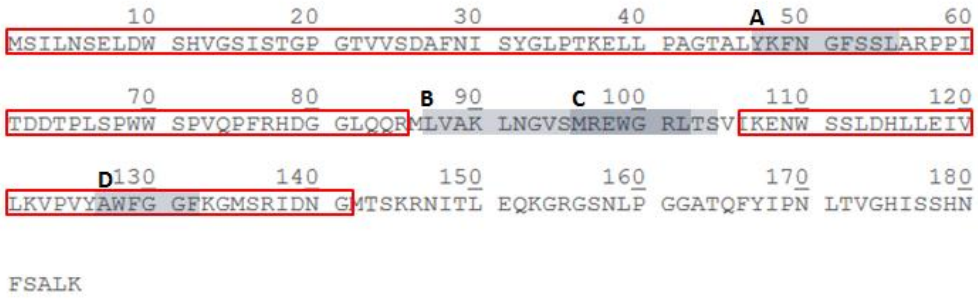


Figure 3.43, Sequence of EvpP_P143T with the two stabilized peptides mapped out in blue; the trypsin-digestion boundary was marked by red box.

Peptide A was found to come from residues Tyr47 to Leu56. This region was located within the N-terminal half of EvpP; however this stretch of residues did not

show up on the HSQC. The decrease in its flexibility might be coming from a tertiary fold of EvpP where these residues could interact with residues from the mutation point. Such a global fold of EvpP was also supposed to protect region A from limited proteolysis. Peptides B and C were found within the linker region between the N-terminal half and the 4-kDa piece. Peptide B ranged from residues Leu87 to Leu102 while peptide C was from Met96 to Ser104, with seven overlapping residues. All the above three regions were considered to be flexible in EvpP. Therefore, a decreased flexibility in these regions might suggest possible interactions among different flexible regions. The last peptide was found to be within a relatively rigid region of EvpP, ranging from residue Ala127 to Phe132, which was located within the proximity of the mutation point.

3.8 Interaction studies between EvpP and EvpC

It has been suggested by Zheng and Leung that EvpP may interact with EvpC (Zheng and Leung, 2007). To further characterize this interaction, we performed glutathione-S-transferase pull-down assay using GST-tagged EvpC with EvpP and its truncation mutants (Figure 3.44). The proteins of EvpP$^{1\text{-}168}$ and EvpP$^{1\text{-}142}$ were refolded from their inclusion bodies.

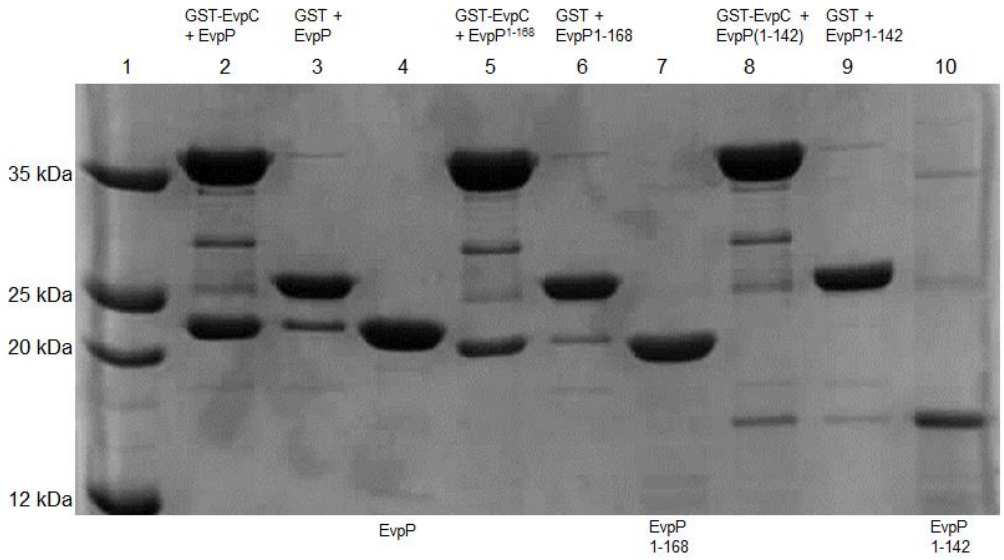
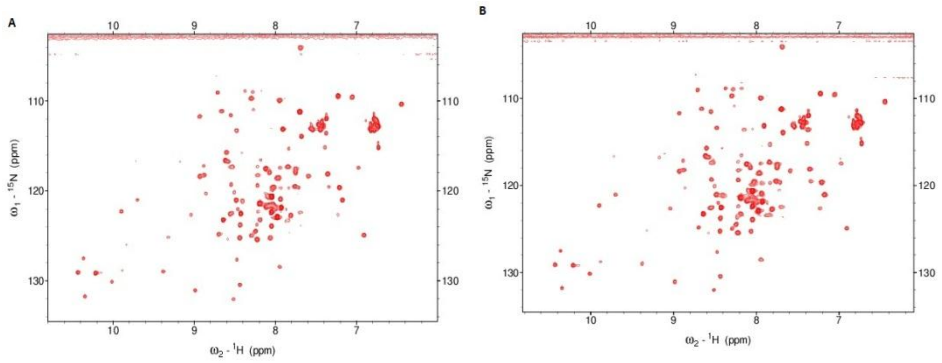


Figure 3.44, GST pull-down assay using GST-EvpC against EvpP, EvpP$^{1-168}$ and EvpP$^{1-142}$. In each of the experiment, an internal GST-control was included.

As shown in Figure 3.44, GST-tagged EvpC was able to pull down both EvpP and refolded EvpP$^{1-168}$; but not refolded EvpP$^{1-142}$. EvpP$^{1-92}$ was also unable to interact with GST-tagged EvpC (data not shown). Thus, the C-terminal flexible regions appeared to be involved in interaction with EvpC. The refolded protein EvpP$^{1-168}$, although unstable, was still able to interact with EvpC to a comparable extent as in wild-type EvpP. However further removing the flexible regions greatly abolished such interaction as shown by the inability to pull down GST-EvpC in EvpP$^{1-142}$. EvpP$^{1-92}$ could not interact with GST-EvpC at all, due to the combined effect of removing the C-terminal flexible regions and the 4-kDa rigid region of EvpP. We've also attempted to sub-clone and express GST-tagged EvpP; but this construct of EvpP formed extensive inclusion bodies (data not shown).

To further elucidate the stoichiometry of interaction between EvpC and EvpP, we performed NMR titration experiment using non-labelled GST-tagged EvpC with ^{15}N-edited EvpP. Both GST-EvpC and EvpP were concentrated to 750 µM respectively before they were mixed (in various proportions) together with NMR buffer to generated different ratios between EvpP and GST-EvpC (1: 0.2, 1: 0.4, 1: 0.6, 1: 0.8, 1: 1 and 1: 2). The concentration of ^{15}N-edited EvpP was kept constant at 150 µM in all NMR titration experiments. A series of ^{1}H-^{15}N HSQC spectra of EvpP (with different ratio to GST-EvpC) were acquired (Figure 3.45).



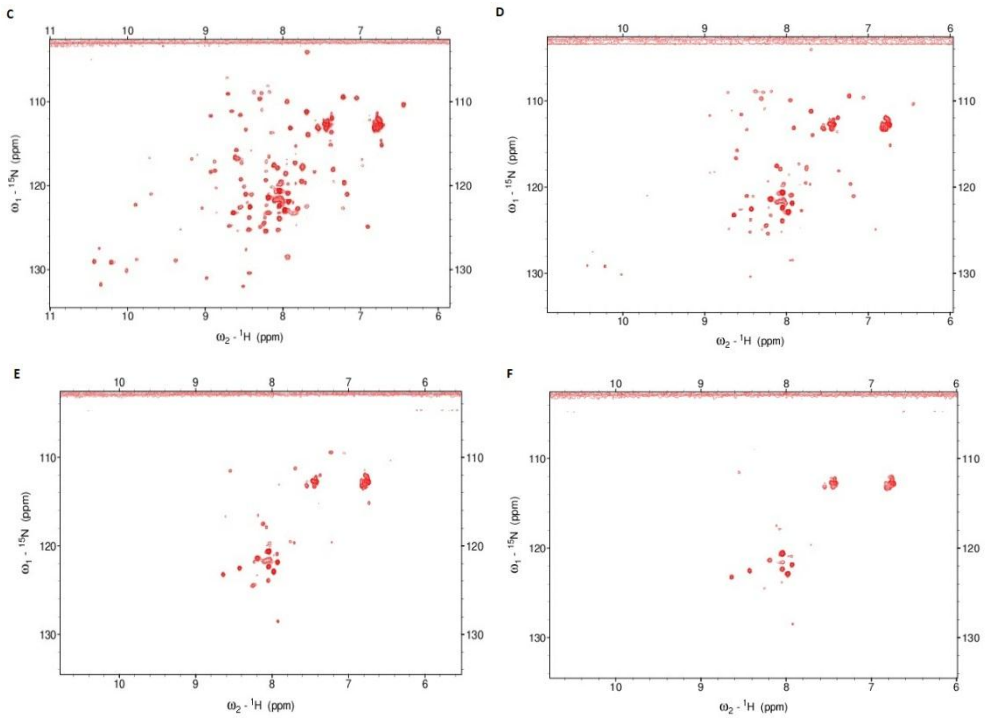


Figure 3.45, NMR titration experiments of ^{15}N-edited EvpP with different ratios of GST-EvpC. Panel A, EvpP : GST-EvpC = 1 : 0.2; Panel B, 1 : 0.4; Panel C, 1 : 0.6; Panel D, 1 : 0.8; Panel E, 1 : 1; Panel F, 1 : 2.

As shown in Figure 3.45, the interaction between EvpP and GST-EvpC was demonstrated by the progressive decrease in peak intensities in a GST-EvpC concentration dependent manner. The binding of EvpP to GST-EvpC was observed to be saturated when the concentration ratio exceeded 1: 1. However we could not specifically deduce the individual residues that mediate such interaction. The decrease in peak intensity was global and progressive, suggesting that the disappearance of observable peaks on the HSQC was due to the general increase in molecular weight by binding to GST-EvpC, instead of the binding of a small region of EvpP to EvpC. Most of the peaks that remained observable on the HSQC even after saturation of interaction were found to be peaks from the histidine tag of EvpP.

We also performed the NMR titration experiments using non-labelled GST with ^{15}N-edited EvpP as a control. The HSQC did not undergo any change even GST was titrated to a 1: 1 ratio with EvpP. Therefore, the observed interaction between GST-EvpC and EvpC was supposed to be mediated by EvpC and EvpP. Unexpectedly, his-tagged EvpC failed to interact with EvpP in both NMR titration

study as well as pull-down assays. One possible explanation could be that the oligomeric state of EvpC might be the important for the interaction to EvpP. GST-tag helps EvpC protein to adopt a dimer form, which could possibly be the pre-requisite for the interaction between EvpC and EvpP.

Lastly we performed protein cross-linking assay using EvpC with EvpP. EvpC was mixed with EvpP in a 1: 1 ratio at a concentration of 40 μM. The cross-linking was catalysed by 2.3% glutaraldehyde at room temperature (Figure 3.46).

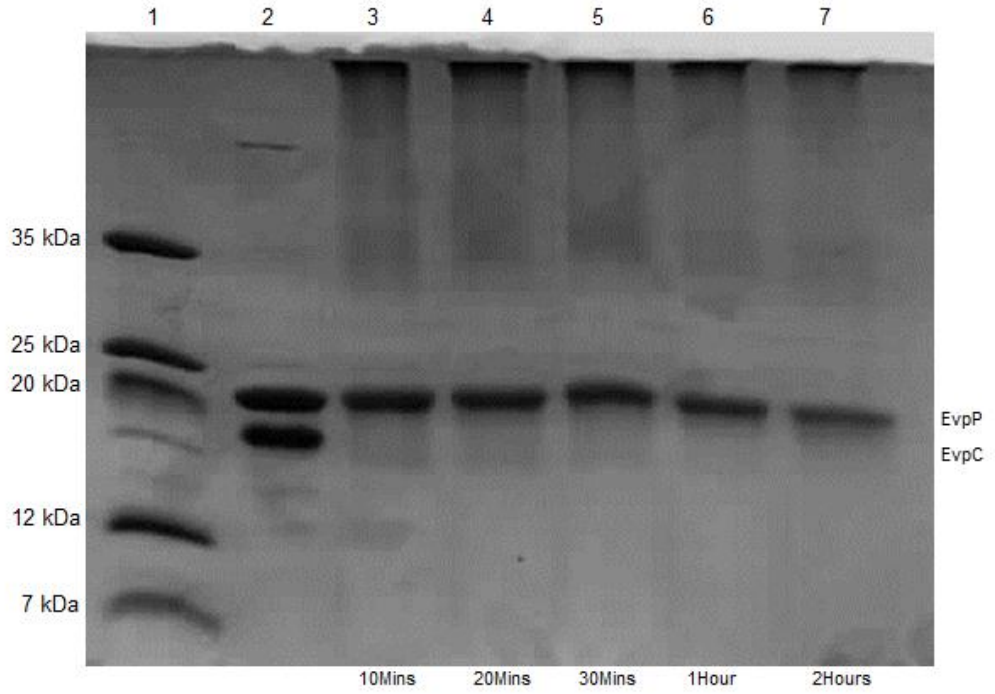


Figure 3.46, Cross-linking assay using EvpC and EvpP. Lane 1, molecular weight marker; Lane 2 to Lane 7, cross-linking of EvpP and EvpC for different period of time.

As shown in Figure 3.46, EvpC was completely cross-linked within 10 minutes of the experiment while only a small portion of EvpP was slowly cross-linked to EvpC. As there was no observable dimer formation of EvpP (refer to Figure 3.6 for the EvpP cross-linking assay), the gradual decrease in EvpP band intensity indicates possible cross-linking between EvpP and EvpC. This was not surprising as EvpC has been suggested to form hexamers at high concentration, or dimers at lower concentration. The intrinsic high propensity to form oligomers of EvpC might out-compete its cross-linking with EvpP. We've also attempted to cross-link GST-tagged

EvpC together with EvpP; however the result was not conclusive. GST-tagged EvpC was soon cross-linked to each other while the cross-linking between EvpP and GST-EvpC was not clearly observed (data not shown).

So far, only GST-tagged EvpC has been demonstrated to interact with EvpP in both NMR titration experiments and GST pull-down assay. EvpC alone could not induce significant perturbation in the HSQC of EvpP. The cross-linking assay did not show a clear pattern of interaction as well. In addition, the presence of a GST tag might further promote a dimer formation of EvpC. Whether this dimer formation has an effect on its interaction with EvpP remains unclear.

CHAPTER 4, RESULTS & DISCUSSIONS (AcrH-AopB)

4.1 Elastase digested AcrH-AopB$^{1-264}$ chaperone-translocator complex

Inside the Type III secretion system (T3SS) translocator operon of *Aeromonas hydrophila*, AopB and AopD were the two hydrophobic translocators that were found to be inserted into host cell membrane. The pore-forming ability of these proteins requires the presence of trans-membrane domains in their amino acid sequence. In order to prevent the pre-mature aggregation of AopB and AopD through their hydrophobic tans-membrane domains, a Class II chaperone AcrH binds to each of them; maintaining their stability before they become folded into a mature state. Tan Y W, et al, have laid the groundwork of the AcrH binding regions as well as their stoichiometry on AopB and AopD (Tan Y W, et al, 2009). Through limited protease digestion, the AcrH-binding regions of AopB and AopD have been mapped respectively. AopD binds AcrH using two discrete regions: DF1, from Val16 to Leu147 and DF2 from Ser242 to Phe296, as determined by digestion with chymotrypsin. AopB binds AcrH through one continuous region ranging from residues Ala33 to Leu264, as determined by digestion with elastase (Tan Y W, et al, 2009).

It has been demonstrated by the same group that AcrH could be co-expressed and co-purified with either AopB or AopD. Particularly, full length AcrH could be co-expressed with a truncated form of AopB$^{1-264}$. The purified chaperone-translocator complex went through limited protease digestion by elastase to reveal the AcrH-binding region of AopB, from residue Ala33 to Leu264. Therefore we continued this research by exploring the possibility of crystalizing this partially digested protein complex. So far, no structure of Class II chaperone in complex with translocator has been reported yet.

We purified the partially digested complex of AcrH-AopB$^{33-264}$ and set the crystallization screen using Hampton Crystal Screen I and II, Hampton Index Screen I

and II and Wizard Classic Crystallization Screen 1, 2, 3 and 4. Crystals formed in a few screening conditions (Figure 4.1), but they were all found to have very poor diffraction. Buffer optimization did not help in improving the quality of crystals formed.

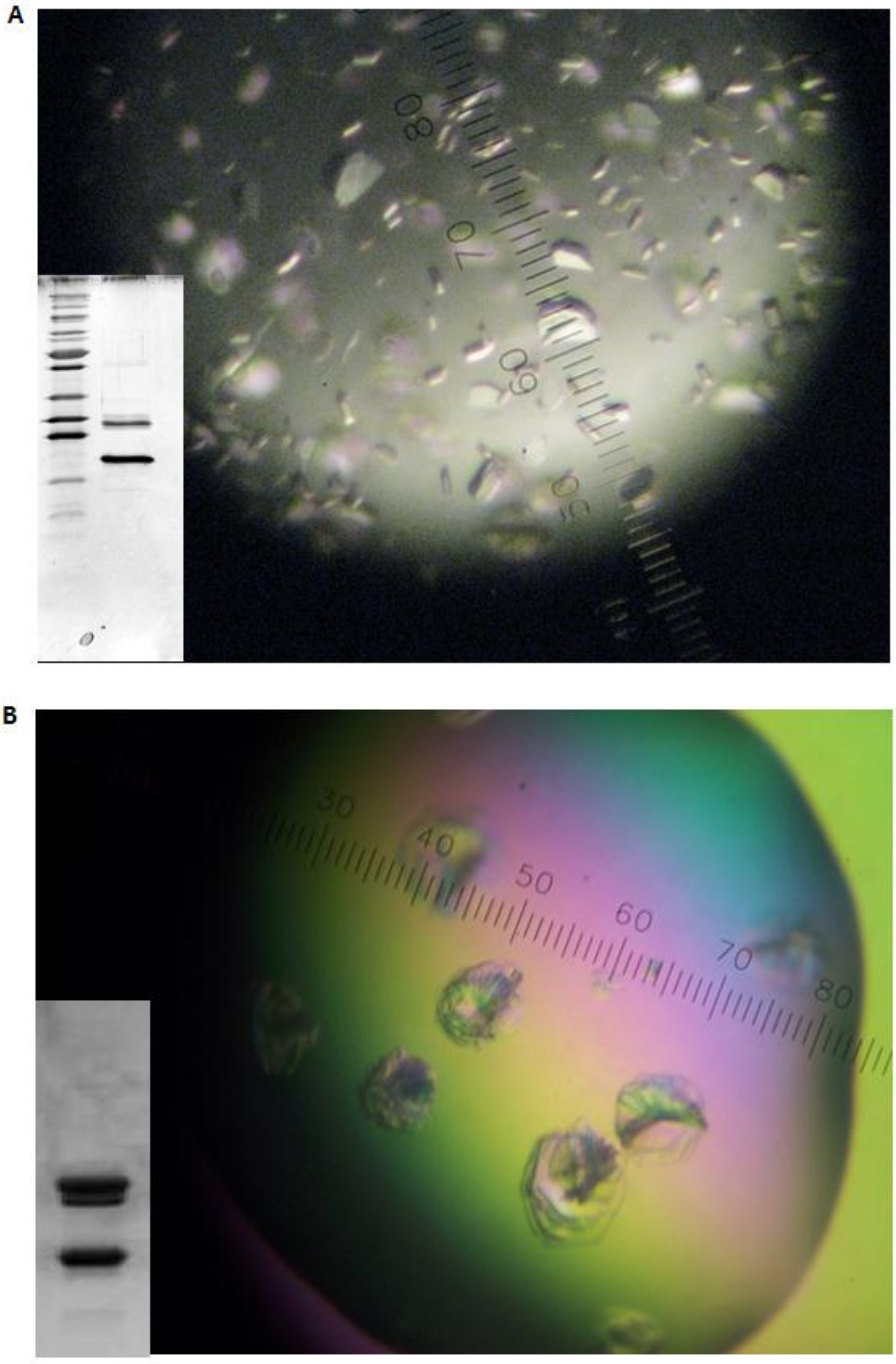


Figure 4.1, Crystals formed by the chaperone-translocator complex AcrH-AopB$^{33-264}$ in different conditions. Panel A, crystals formed by AcrH-AopB$^{33-264}$ in 1.4 M Sodium Potassium Phosphate pH6.9 at concentration of 12 mg/ml; Panel B, crystals formed by AcrH-AopB$^{33-264}$ 0.1M HEPES Sodium pH7.5 with 1.5 M Lithium Sulfate Monohydrate at concentration of 40 mg/ml.

As shown in Figure 4.1 B, another band beneath the AopB33-264 was observed from the protein crystal. This might imply that AopB33-264 has undergone degradation to a slightly shorter form. This shorter form would possibly be more stable than the one bound by elastase digestion. This, in turn, prompted us to explore other enzymes for limited proteolysis that could generate shorter boundaries on AopB, since the presence of flexible terminal residues may interfere with the quality of crystal packing.

4.2 Limited proteolysis by thrombin yielded a stable complex in which AopB was slightly shorter

In order to optimize the boundary for limited protease digestion, several enzymes were tested to partially digest AcrH-AopB$^{1-264}$. Among them, thrombin was identified to repeatedly produce a comparable digestion pattern to that of elastase. However, with thrombin, the digested AopB was found to be slightly shorter (Figure 4.2).

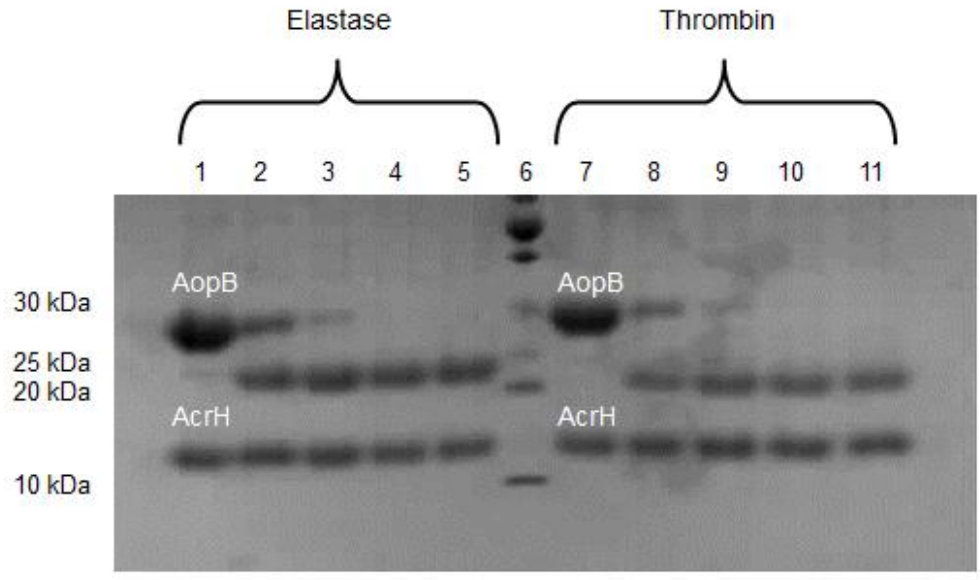


Figure 4.2, Limited protease digestion of AcrH-AopB$^{1-264}$ by elastase and thrombin. Lane 1, undigested AcrH-AopB$^{1-264}$; Lane 2 to Lane 5, AcrH-AopB$^{1-264}$ digested by elastase for 5 min, 15 min, 30 min and 60 min; Lane 6, molecular weight marker; Lane 7, undigested AcrH-AopB$^{1-264}$; Lane 8 to Lane 11, AcrH-AopB$^{1-264}$ digested by thrombin for 5 min, 15 min, 30 min and 60 min.

As shown in Figure 4.2, both elastase and thrombin were able to finish digestion within 30 minutes. In both cases, AcrH remained intact while AopB$^{1-264}$ was digested to generate shorter fragments. The patterns yielded by both enzymes were similar, whereas AopB in thrombin digestion was observed to have slightly lower molecular weight than that of elastase digestion. The thrombin digested complex was able to be co-purified through FPLC. The elution profile suggested that they were also forming monomers (Figure 4.3).

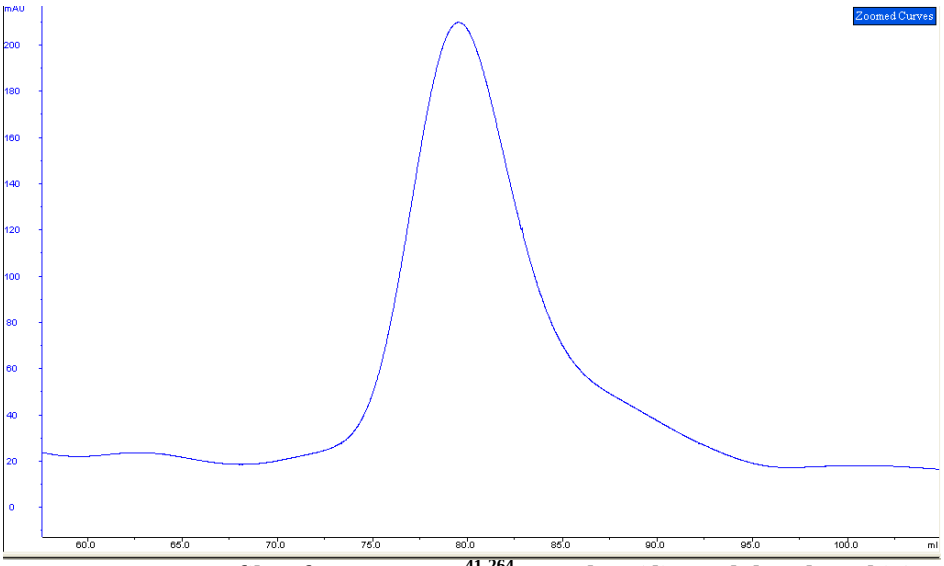


Figure 4.3, FPLC profile of AcrH-AopB$^{41-264}$ complex (digested by thrombin). The complex was eluted out at 80 minutes through a Hiload Superdex200 size exclusion column.

Therefore we performed N-terminal sequencing using thrombin digested AcrH-AopB$^{1-264}$. Thrombin digested protein bands were first immuno-blotted onto a PVDF membrane before being visualized by coomassie blue staining. The corresponding band of thrombin digested AopB was separated out and subjected to protein sequencing. As a result, an N-terminal residue stretch "VEKHGQGV" was unambiguously identified, matching the sequence of AopB from residues Val41 to Val48. This new boundary was eight residues shorter than the one generated by elastase. As the thrombin digested AcrH-AopB$^{41-264}$ complex was stable and reproducible, we continued our crystallization screen using this new complex using the same screening kits as described above. Initial crystals were subjected to buffer and additive optimization. As expected, better crystals formed in a few buffer

conditions. However the best crystal from AcrH-AopB$^{41-264}$ could only diffract to 7 Å or 8 Å (Figure 4.4).

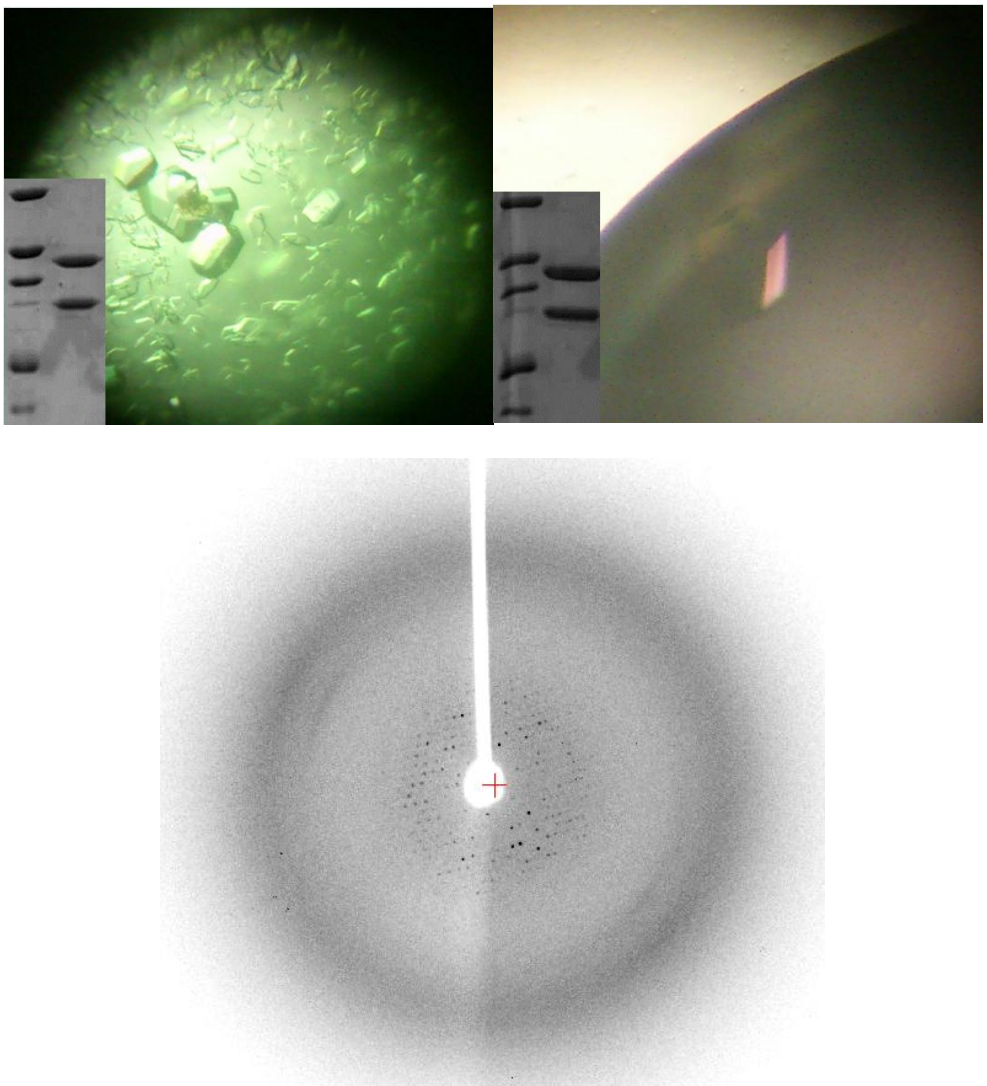


Figure 4.4, Crystals formed by AcrH-AopB41-264 (digested by thrombin). Upper left, crystals formed by AcrH-AopB$^{41-264}$ in 100 mM MES pH6.0 with 1.26 M Ammonium Sulfate; Upper right, crystal formed by AcrH-AopB$^{41-264}$ in 100 mM Cacodylate pH6.5 with 1.26 M Ammonium Sulfate; Bottom, diffraction pattern of the crystal shown in the upper right panel.

4.3 An even shorter boundary of AopB failed to yield better crystal

By shortening the digested boundary of AopB, the quality of resultant crystals could be improved. Therefore, we continued our work in seeking for even shorter boundaries on AopB digestion. Thrombin digested AcrH-AopB$^{41-264}$ was used as the baseline. Various enzymes, such as trypsin, chymotrypsin, thermolysin, actinase E, proteinase K, clostripain and bromelain, were tested to further digest this

complex. Among these enzymes, thermolysin and actinase E generated relatively stable patterns that were coherent to each other (Figure 4.5 and Figure 4.6).

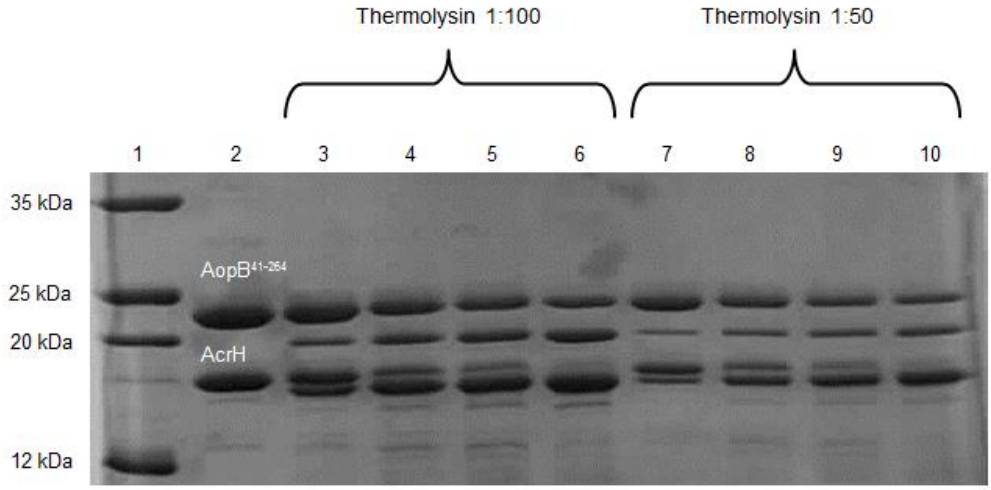


Figure 4.5, AcrH-AopB$^{41-264}$ (generated through limited proteolysis by thrombin followed by purification through FPLC) digested by thermolysin in different concentration ratios. Lane 1, molecular weight marker; Lane 2, AcrH-AopB$^{41-264}$, previously purified from thrombin digestion; Lane 3 to Lane 6, AcrH-AopB$^{41-264}$ digested by thermolysin at a ratio of 1:100 for 15 min, 30 min, 45 min and 60 min; Lane 7 to Lane 10, AcrH-AopB$^{41-264}$ digested by thermolysin at a ratio of 1:50 for 15 min, 30 min, 45 min and 60 min.

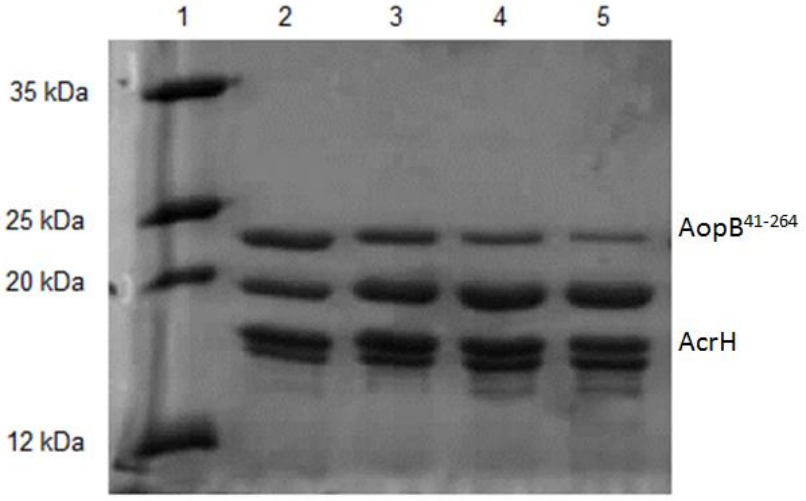


Figure 4.6, AcrH-AopB$^{41-264}$ (generated through limited proteolysis by thrombin followed by purification through FPLC) digested by actinase E at a ratio of 1:100. Lane 1, molecular weight marker; Lane 2 to Lane 5, AcrH-AopB$^{41-264}$ digested by actinase E for 15 min, 30 min, 45 min and 60 min.

Thermolysin and actinase E were able to further cleave AcrH-AopB$^{41-264}$ to generate a much shorter fragment of AopB. We subsequently performed N-terminal sequencing for the digested AopB in both cases. As expected, the digested AopB fragments were found to be the same between these two enzymes. The protein

sequencing results showed that AopB digested by either of these two enzymes had an N-terminal sequence of "MSPP" which matched the sequence of AopB from residue Met61 to Pro64 (Figure 4.7). It was intriguing to test out this new boundary in terms of crystallization. One thing we noticed was that thermolysin and actinase E were only able to cleave previously thrombin-digested AcrH-AopB$^{41\text{-}264}$ complex to generate such shorter fragments. We've attempted to use these two enzymes to digest AcrH-AopB$^{1\text{-}264}$, but they could not generate the above patterns.

As shown in Figures 4.5 and 4.6, apart from AopB, AcrH also underwent limited proteolysis by these two enzymes. We performed N-terminal sequencing for the digested AcrH and found that it had an N-terminal sequence of "MEHD", which matched residues from Met1 to Asp4. This suggested that the partial digestion on AcrH happened at the C-terminal end.

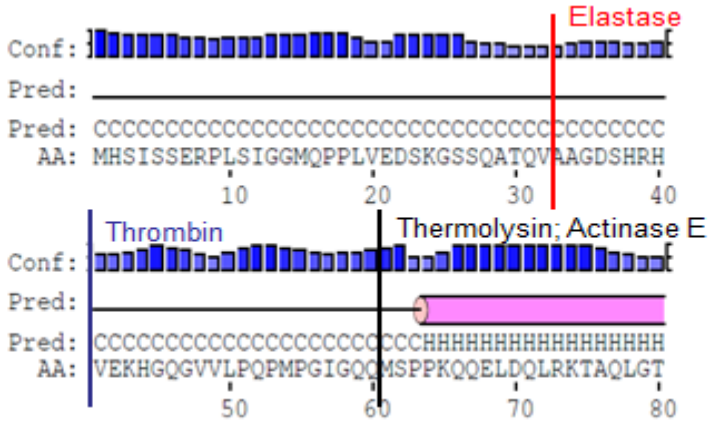


Figure 4.7, Sequence of the first 80 residues of AopB shown with the predicted secondary structure. The limited protease digestion boundaries on AopB were marked out.

Although thermolysin and actinase E were able to further digest AcrH-AopB$^{41\text{-}264}$, neither of them could achieve a complete partial digestion. In both cases, undigested AopB and AcrH were still observed in a significant amount compared to the digested ones. Apparently, thermolysin was more efficient in digesting AcrH as shown in Figure 4.5, Lane 6 and Lane 10. Complete digestion of AcrH could be achieved by thermolysin digestion in 60 minutes. On the other hand, actinase E was more efficient in digesting AopB. As shown in Figure 4.6, Lane 5, almost all AopB

have been digested to the shorter form. Therefore, to get a pure complete form of $AcrH_{\Delta}C\text{-}AopB^{61\text{-}264}$, three enzymes (thrombin, thermolysin and actinase E) were used to digest the complex in a sequential manner. First the undigested $AcrH\text{-}AopB^{1\text{-}264}$ was digested by thrombin followed by FPLC purification to yield a pure form of $AcrH\text{-}AopB^{41\text{-}264}$. This complex was then digested by thermolysin for 40 minutes. 10 mM EDTA was added to partially inhibit the activity of thermolysin before the complex was further digested by actinase E for another 20 minutes in order to minimize over-digestion (Figure 4.8).

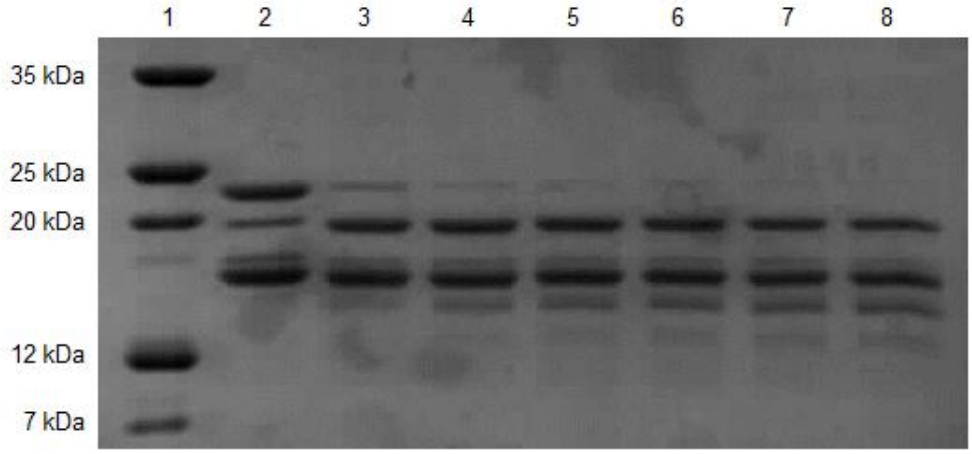


Figure 4.8, Limited protease digestion of $AcrH\text{-}AopB^{41\text{-}264}$ by a sequential digestion of thermolysin followed by actinase E. Lane 1, molecular weight marker; Lane 2, purified $AcrH\text{-}AopB^{41\text{-}264}$ digested by thermolysin for 40 minutes; Lane 3 to Lane 8, thermolysin-digested $AcrH\text{-}AopB^{41\text{-}264}$ further digested by actinase E for different period of time: 5 min, 10 min, 15 min, 20 min, 25 min and 30 min.

Using this method, we were able to get a relatively complete partial digestion for both AcrH and AopB to get the complex $AcrH_{\Delta}C\text{-}AopB^{61\text{-}264}$. However as shown in Figure 4.8, over-digestion occurred inevitably. The over-digested impurities could not be separated from size exclusion chromatography. Ion exchange chromatography helped in removing most of the over-digested products; however we still could not get a highly pure form of the $AcrH_{\Delta}C\text{-}AopB^{61\text{-}264}$ (Figure 4.9).

Nevertheless, we attempted to crystalize the triple-digested complex with the shortest form of AopB. Although crystals could form; the quality was not as good as those coming from limited proteolysis solely by thrombin (Figure 4.10). The presence

of the over-digested product as an impurity might be interfering with the crystal packing of $AcrH_{\Delta}C$-$AopB^{61-264}$.

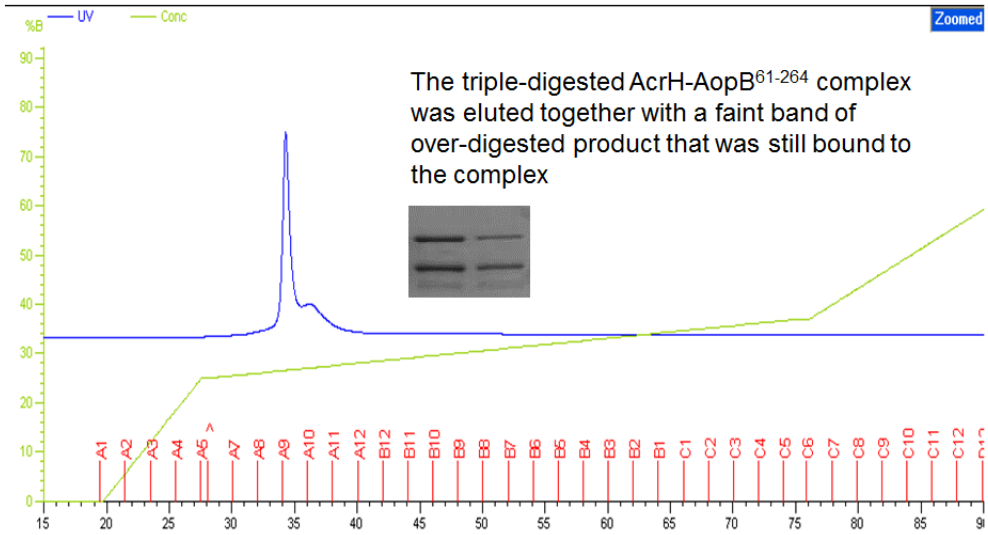


Figure 4.9, Ion exchange chromatography of triple-digested $AcrH_{\Delta}C$-$AopB^{61-264}$ complex at pH9. The elution gradient was set to be shallow however both AcrH-AopB complex as well as the over-digested product were co-eluted as a sharp peak at around 33% salt concentration.

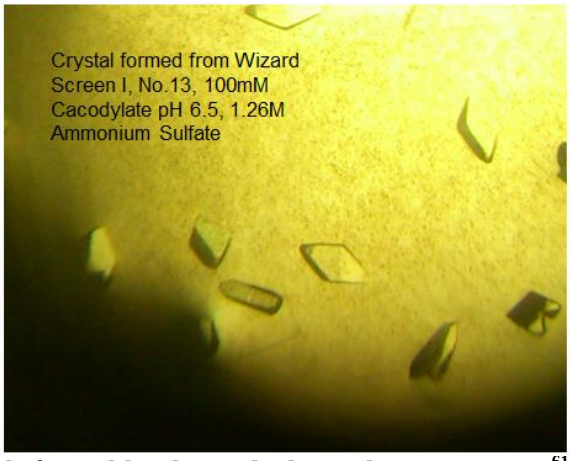


Figure 4.10, Crystals formed by the triple-digested $AcrH_{\Delta}C$-$AopB^{61-264}$. The quality was much worse than the one shown in Figure 4.4.

CHAPTER 5, CONCLUSION & FUTURE WORK

5.1 EvpP in T6SS of *Edwardsiella tarda*

Edwardsiella tarda is one of the major fish pathogens with a wide range of hosts. Both T3SS and T6SS were identified in *E. tarda* that greatly contribute to virulence. Cross-talk between T3SS and T6SS mainly occurred at the transcription level (Leung K Y, et al, 2011). While T3SS is one of the most-studied secretion systems; the mystery of T6SS remains largely unraveled. Particularly in *E. tarda*, the T6SS gene cluster harbours a novel protein EvpP, which is not found in species outside the genus *Edwardsiella*. Previous study has demonstrated that EvpP plays an important role in virulence towards various hosts. Deletion mutation in the gene *evpP* significantly attenuated its virulence towards blue gourami fish (Zheng and Leung, 2007), and was associated with retarded activity in hemolysis as well as internalization during the process of invasion (Xin W, et al, 2009). So far, little has been discovered about this unique protein EvpP, especially at the protein level.

Therefore, we attempted to gain structural insights into EvpP by mapping out its rigid and flexible regions. On the whole, EvpP is characterized to be a non-stable protein without a well-folded three-dimensional structure. More than half of the protein is predicted to be disordered. However, EvpP does have higher order structures as shown by its HSQC. Three approaches were adopted to map its rigid and flexible regions: limited protease digestion, NMR backbone assignment and HDX-MS experiments.

Limited protease digestion has shown a coherent pattern that the N-terminal half is more rigid than the C-terminal half. The trypsin-digested pattern further suggested another rigid region of a 4-kDa piece in the C-terminal half of EvpP. This is validated by the relatively complete assignment of the peaks on the HSQC using a stable mutant of EvpP_P143T. As expected, more than 90% of the assigned peaks were found to be within the trypsin-digestion boundary, while most of them were

assigned to the N-terminal half of EvpP. In addition, the backbone assignment has suggested a third rigid region, possibly near the C-terminal end, where two out of three assigned residues were found to have long-range NOEs with residues from the N-terminal half, namely Val173 with Thr64 and Gly174 with Asp62. Such interaction could be strong as the Cα and Cβ chemical shifts of the C-terminal residues deviated largely from the values observed in a random coil. The presence of a C-terminal rigid region was further supported by the HDX-MS experiments. In both experiments using wild-type EvpP as well as mutant EvpP_P143T, a C-terminal rigid region with little H/D exchange was observed. Heat maps plotted using either wild type or mutant EvpP demonstrated that the linker region between the N-terminal half and the 4-kDa piece was the most flexible region throughout entire protein. In addition, it further pointed to another short region inside the N-terminal half of EvpP, which had relatively higher flexibility than its surroundings, coming from residues Tyr47 to Ser55. Interestingly, this region also matched the only gap of backbone assignment found in the N-terminal half of EvpP. However, this flexible region was not digested by limited proteolysis, most likely because this region is not extensively flexible enough to be accessible by various proteases.

The major regions that were "stabilized" by the point mutation P143T was revealed by a comparison between wild-type EvpP and mutant EvpP_P143T in their extent of H/D exchange. The mutant P143T was found to stabilize three major regions of EvpP. The first region was the internal flexible gap in the N-terminal half that lacked backbone assignment. The second region was the linker that was found to be the most flexible region. A third region was a relatively rigid region near the mutation site. The stabilization by P143T in the flexible regions might imply potential interaction among flexible regions of EvpP. The general topologies of rigid and flexible regions of EvpP were mapped in Figure 5.1.

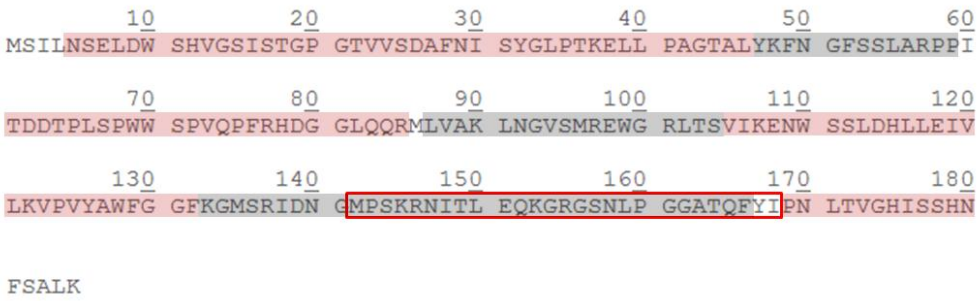


Figure 5.1, Sequence of EvpP with rigid and flexible regions mapped out. Rigid regions were highlighted in pink while flexible regions were highlighted in grey. The potential region which might be involved in interaction with EvpC was boxed.

EvpP has been suggested to bind EvpC (Zheng and Leung, 2007). We further examined this interaction by GST pull-down assay, NMR titration experiments and protein cross-linking assay. EvpC was shown to interact with EvpP in both pull down assay and NMR titration experiments. In both cases, the presence of the GST tag was required for successful interaction, while the GST protein alone failed to interact with EvpP at all. Thus, the observed interaction was supposed to be mediated between EvpC and EvpP. A C-terminal truncation mutant $EvpP^{1-168}$ was shown to be able to interact with GST-EvpC, while a further truncation mutant $EvpP^{1-142}$ largely abolished such interaction. Although both proteins were refolded from inclusion bodies with a low stability, they nevertheless led to the speculation of a role in binding EvpC for the C-terminal flexible regions (as marked in Figure 5.1). Intrinsic disordered regions are widely identified in proteins from T3SS, where they are demonstrated to play important roles in mediating protein-protein interaction (Lilic, et al, 2006 and Anastasia D, et al, 2009). The flexible regions of EvpP could also be the key to interaction activities with other members of T6SS, or even host cell entities.

5.2 Future works on EvpP

One obvious goal is to solve the tertiary structure of EvpP, although this may require other molecules to further stabilize EvpP. As EvpP failed to crystalize and only a subset of residues showed up on HSQC of NMR experiments the determination of tertiary structure using native EvpP was proven to be challenging.

This may rely on the identification of novel interacting partners. EvpC is a reasonable candidate; however the interaction between EvpP and EvpC still remains elusive. While EvpP requires an interacting partner that can largely alter the flexibility in those regions before it can be completely stabilized, finding such binding partners is also challenging. But still, as disordered regions are commonly found to be mediating interaction activities in T3SS, stabilization in flexible regions of EvpP might help us determine the tertiary structure of EvpP, possibly in complex with its interacting partner.

Another interesting line of research on EvpP is to study the localization after being secreted via T6SS. EvpP has been proposed to be an effector protein, as it is secreted and it contributes to virulence. Whether EvpP gets injected into host cells seems to be an intriguing target. If so, its subcellular location as well as its interacting host cell entities will be of great research interest, which may further lead to alteration of host cell signaling. So far, the function studies on EvpP are solely based on deletion mutation together with complementation. The mechanism of such an activity still remains unraveled.

5.3 AcrH-AopB in T3SS of *Aeromonas hydrophila*

Aeromonas hydrophila is another major pathogen to fish and sometimes humans. The presence of functional Type 3 and Type 6 secretion systems are also found to be the major components for its virulence. In *A. hydrophila*, two groups of proteins are secreted by the T3SS injectisome, effectors and translocators. In order to directly deliver effector proteins into host cell cytosol, translocator proteins AopB and AopD, with pore-forming ability, have to be inserted into the host cell membrane. They can then oligomerize to form a continuous channel connecting to the needle complex protruding out from the bacterial membrane (Neyt and Cornelis, 1999; Goure et al, 2005; Yu HB, et al, 2004). Translocator proteins are generally large and have trans-membrane domains. Therefore, genes encoding Class II chaperones are

generally located within the proximity of translocators. Class I chaperones bind to effectors while Class III chaperones bind to apparatus proteins. The structure of both Class I and Class III chaperones have been determined in complex with respective partners. However, no complex structure of a Class II chaperone with a translocator has been reported. In the translocator operon of *A. hydrophila*, AcrH is the chaperone for both translocators AopB and AopD. Tan Y W, et al, have demonstrated that AcrH forms 1:1 complex with either AopB or AopD. They also mapped the AcrH-binding regions on AopB and AopD (Tan Y W, et al, 2009).

Of our particular interest is the chaperone-translocator complex formed between AcrH and AopB, because the complex can be co-expressed and co-purified. We continued the research on AcrH-AopB complex starting from the elastase digested boundary (Tan Y W, et al, 2009). AcrH-AopB$^{1\text{-}264}$ was digested by elastase and the resultant AcrH-AopB$^{33\text{-}264}$ complex was purified for crystallization screen. Although crystals could form, they had poor diffraction. As a result, we attempted to optimize the limited proteolytic boundary of AcrH-AopB for complex crystallization. Limited proteolysis by thrombin generated a very close pattern to elastase. AopB was found to be digested to yield a shorter fragment from residues Val41 to Leu264. We subsequently subjected this new complex to crystallization screen. Better crystals formed by the complex AcrH-AopB$^{41\text{-}264}$, with the best one able to diffract to 7.5 Å. This indicated that residues between Ala33 and Val41 were not involved in binding AcrH. Therefore, by reducing the flexible residues, we were able to improve the quality of AcrH-AopB complex slightly. In an attempt to further shorten the boundary on AopB, both thermolysin and actinase E were found to generate a coherent fragment of AopB that was 20 residues shorter than the one digested by thrombin, from residues Met61 to Leu264. This even shorter fragment of AopB was still able to bind AcrH tightly as they could be co-purified through either size exclusion or ion exchange chromatography. We further optimized the digestion protocol to get all AcrH-AopB digested to the shortest form by using three enzymes in a sequential

manner. First, AcrH-AopB$^{1-264}$ complexes were digested by thrombin and purified to get a homogenous form of AcrH-AopB$^{41-264}$. The complex was then digested by thermolysin for 40 minutes before being partially inhibited by 10 mM EDTA. Actinase E was the last one to digest the remaining complex in order to get as much AopB$^{41-264}$ cleaved to AopB$^{61-264}$ as possible while keeping over-digested products at minimum. The triple-digested complex was then purified through ion exchange chromatography (MonoQ) at pH9. However we failed to attain a highly pure form of this complex AcrH$_{\Delta}$C-AopB$^{61-264}$. The presence of impurities coming from over-digested product severely interfered with the crystal formation of AcrH$_{\Delta}$C-AopB$^{61-264}$. Therefore, the best crystal still came from the one digested by thrombin.

By optimizing the boundaries for limited protease digestion alone, we could not obtain a high quality crystal from which the structure of the complex could be determined. The shorter fragments generated by different enzyme on AopB inevitably suggested a shorter binding region on AcrH. As shown in Figure 4.7, the N-terminal 63 residues of AopB were predicted to be a disordered region. With limited proteolysis using various enzymes, we've demonstrated that the N-terminal disordered region was not involved in the binding of its chaperone AcrH. AopB$^{61-264}$ was still able to bind AcrH tightly. As a result, the AcrH-binding domain of AopB might solely lie within the α-helical regions depicted in the secondary structure prediction. The N-terminal disordered region might have a completely different function other than being partially involved in binding AcrH.

5.4 Future works on AcrH-AopB

The first obvious goal is to further optimize the AcrH-AopB complex to yield a high quality crystal, from which the three-dimensional structure of a Class II chaperone in complex with a translocator could be determined for the first time. This can be performed either by limited protease digestion with more enzymes, or by engineering an internal enzymatic cleavage site on AopB that is deeper inside the

thrombin cleavage site. As thrombin was demonstrated to stably digest AopB without much over-digested products, an internal thrombin cleavage site on AopB could be one possible choice. After we manage to acquire the structure of the AcrH-AopB chaperone-translocator complex, the detailed mechanism of interaction could be investigated accordingly.

Another approach is to co-express the $AcrH_{\Delta}C$-$AopB^{1-264}$ chaperone-translocator complex and optimize from this new construct. Thermolysin and actinase E have been shown to digest a few residues at the C-terminal end of AcrH, and the remaining $AcrH_{\Delta}C$ still binds AopB strongly. We could possibly map out the remaining boundary on AcrH first and co-express this truncation mutant of AcrH together with $AopB^{1-264}$. In addition, the C-terminal residues of AcrH are not involved in the formation of tetratricopeptide repeats. Therefore, this truncation most likely will not affect the interaction between AcrH and AopB. In this manner, we no longer need to consider whether the digestion of AcrH is complete, but only focus on the digestion of AopB.

Apart from AopB, the complex formed between AcrH and AopD is also an interesting target. Tan Y W, et al, have demonstrated that AcrH, AopB and AopD were able to form a metastable complex together (Tan Y W, et al, 2009). As a result, AcrH has to interact with AopB and AopD through different regions. It has been shown that the Class III chaperone YscG binds YscF and YscE simultaneously. The apparatus protein YscF binds the concave region of the chaperone YscG while YscE binds the convex (Sun P, et al, 2007). YscG is a chaperone with tetratricopeptide repeat domain that is similar to the Class II chaperone AcrH. Therefore it is also of great interest to find out how AcrH interacts with both translocator proteins AopB and AopD.

AcrH could be co-expressed with AopB or AopD. We can purify these two chaperone-translocator complexes separately and perform HDX-MS to map out the binding regions on AcrH by comparing the HDX pattern of AcrH alone with either

AcrH-AopB or AcrH-AopD complexes. The results could be compared horizontally to find out whether there's any overlap between AopB and AopD binding regions on AcrH. Site-directly mutagenesis could be performed to further confirm the deduced binding regions.

References

Agata K, et al. The blueprint of the Type III injectisome. The Royal Society of Biological Sciences. 2012, 367, p1140-1154.

Alain F, et al. The bacterial Type VI secretion machine: yet another player for protein transport across membranes. Microbiology. 2008, 154, p1570-1583.

Alain Filloux. The Type VI secretion system: a tubular story. The EMBO Journal. 2009, 28, p309-310.

Alistair B Russel, et al. Type 6 Secretion Delivers Bacteriolytic Effectors to Target Cells. Nature. 2011, Vol.475, p.343-349

Amy T, et al. Translocation of a Vibrio cholerae Type VI secretion effector requires bacterial endocytosis by host cells. Cell Host and Microbe. 2009, 5, p234-243.

Ana T R, et al. Two-step and one-step secretion mechanisms in gram-negative bacteria: contrasting the Type IV secretion system and the chaperone-usher pathway in pilus biogenesis. Biochemical Journal. 2010, 425, p475-488.

Anastasia D, et al. Coiled-coils in Type III secretion systems: structural flexibility, disorder and biological implications. Cellular Microbiology. 2009, 11(5), p719-729.

Anastasia D, et al. Evidence for a coiled-coil interaction mode of disordered proteins from bacterial Type III secretion systems. The Journal of Biological Chemistry. 2008, 283(49), p34062-34068.

Andreas B and Gabriel W. Chaperone-usher pathways: diversity and pilus assembly mechanism. The Royal Society of Biological Sciences. 2012, 367, p1112-1122.

Angela RR, et al. The Type VI secretion system: a multipurpose delivery system with a phage-like machinery. MPMI. 2011, 24(7), p751-757.

Ariel B, et al. Structure and composition of the Shigella flexneri "needle comple", a part of its Type III secretion. Molecular Microbiology. 2001, 39(3), p652-663.

Badreddine D, et al. On the path to uncover the bacterial Type II secretion system. The Royal Society of Biological Sciences. 2012, 367, p1059-1072.

Bjorn U K, et al. Structure of Shigella IpgB2 in complex with human RhoA. The Journal of Biological Chemistry. 2010, 285(22), p17197-17208.

Carina R, et al. Structure of the Yersinia enterocolitica Type III translocator chaperone SycD. Journal of Molecular Biology. 2008, 375, p997-1012.

Cedric E and Igor A. Localizing flexible regions in proteins using hydrogen-deuterium exchange mass spectrometry. Methods in Molecular Biology. Vol896, p375-385.

Chakraborty S, et al. Two-component PhoB-PhoR regulatory system and ferric uptake regulator sense phosphate and ion to control genes in Type III and VI secretion systems of *Edwardsiella tarda*. The Journal of Biological Chemistry. 2011, 286, p.39417-39430

Chow I, et al. The N-terminal domain of Escherichia coli ClpB enhances chaperone function. FEBS Letters. 2005, 579, p4242-4248.

Chrostppher S H, et al. Bacterial contact-dependent delivery systems. Annu. Rev. Genet. 2010, 44, p71-90.

Cleveland D, et al. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. The Journal of Biological Chemistry. 1977, 252(3), p1102-1106.

Daniel F, et al. A novel sensor kinase-response regulator hybrid controls biofilm formation and Type VI secretion system activity in Berkhoderia cenocepacia. Infect. Immune. 2008.

David A Case. Molecular Dynamics and NMR spin relaxation in proteins. Accounts of Chemical Research. 2002, 35(6), p325-331.

David G T and Scott J H. Multiple pathways allow protein secretion across the bacterial outer membrane. Current Opinion in Cell Biology. 2000, 12, p420-430.

Edith N G, et al. Composition of the Type VII secretion system membrane complex. Molecular Microbiology. 2012, 86(2), p476-484.

Elisar B, et al. NMR characerization of partially folded and unfoled conformational ensambles of proteins. Biopolymers. 1999, 51, p191-207.

Ellen L Z, et al. Assembly and mechanisms of bacterial Type IV secretion machines. The Royal Society of Biological Sciences. 2012, 367, p1073-1087.

Eric C and Christian C. Structural biology of Type VI secretion systems. The Royal Society of Biological Sciences. 2012, 367, p1102-1111.

Eric C, et al. Type VI secretion toolkit. EMBO Reports. 2008, p1-7.

Felisberto-Rodrigues C, et al. Towards a Structural Comprehension of Bacterial Type VI Secretion Systems: Characterization of the TssJ-TssM Complex of an *Escherichia coli* Pathovar. PLoS Pathogens. 2011, 7(11) e1002386

Flederic B, et al. Dissecting the bacterial Type VI secretion system by a genome wide in silico analysis: what can be learned by microbial genomic resources?. BMC Genomics. 2009, 10(104).

Florian D, et al. Crystal structure of the protease-resistant core domain of Yersinia pestis virulence factor YopR. Protein Science. 2005, 14, p1679-1683.

Fontana A, et al. Probing protein structure by limited proteolysis. ABP. 2004, 51(2), p299-321.

Gabriel W. Bacterial secretion comes of age. The Royal Society of Biological

Sciences. 2012, 367, p1014-1015.

Gabriele B, et al. Remodeling of VipA/VipB tubules by ClpV-mediated threading is crucial for Type VI secretion. The EMBO Journal. 2009, p1-11.

Giovanni S, et al. Molecular characterization of a functional Type VI secretion system from a clinical isolate of Aeromonas hydrophila. Microbial Pathogenesis. 2008, 44, p344-361.

Guilhem F, et al. Protein contacts, inter-residue interactions and side-chain modelling. Biochimie. 2008, 90(4), p626-639.

Huang Z, et al. Structure insights into host GTPase isoform selection by a family of bacterial GEF mimics. Nature Structural & Molecular Biology. 2009, 16(8), p853-861.

Ian R K and Mark P F, et al. An introduction to NMR based approaches for measuring protein dynamics. Biochimica et Biophisica Acta. 2011, 1814, p942-968.

Ian R. Handerson, et al. Type V Secretion Pathway: the Autotransporter Story. Microbiology and Molecular Biology Reviews. Dec 2004, p.692-744

Igor A, et al. H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: is there a need for the top-down approach?. Analytical Chemistry. 2009, 81(19), p7892-7899.

Jack C L, et al. The Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. The Royal Society of Biological Sciences. 2012, 367, p1088-1101.

Jelger A L and Arnoid J M. The bacterial Sec-translocase, structure and mechanism. The Royal Society of Biological Sciences. 2012, 367, p1016-1028.

Jenet E Deane, et al. Molecular model of a Type III secretion needle: implications for host cell sensing. PNAS. 2006, 103(33), p12529-12533.

Jobichen C, et al. Structural Basis for the Secretion of EvpC: A Key Type VI Secretion Protein from Edawrdsiella tarda. PLoS One. 2011, 5(9), e12910

Joseph D, et al. A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. Science. 2006, 312, p1526-1532.

Julia F, et al. Twin-arginine-dependent translocation of folded proteins. The Royal Society of Biological Sciences. 2012, 367, p1029-1046.

Julie M S, et al. Structure and regulation of the Type VI secretion system. Annu. Rev. Microbiol. 2012, 66, p453-472.

Julien G, et al. Evolution of conjugation and Type IV secretion systems. Mol. Bio. Evol. 2012, p1-17.

Kaare T, et al. Functional aspects of protein flexibility. Cell and Molecular Life Sciences. 2009, 66, p2031-2047.

Kannan G and Ruth N. How different are structurally flexible and rigid binding sites?

Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. Journal of Molecular Biology. 2006, 365, p257-273.

Karel HM, et al. Translocation of proteins across the cell envelope of gram-positive bacteria. FEMS Microbiology Reviewa. 2001, 24, p437-454.

Katerina E C, et al. Breaking on through to the other side: protein export through the bacterial Sec system. Biochem J. 2013, 449, p25-37.

Kenji O, et al. IpgB1 is a novel Shigella effector protein involved in bacterial invasion of host cells. The Journal of Biological Chemistry. 2005, 280(25), p24022-24034.

Konstantin V K, et al. The Type II secretion system: biogenesis, molecular architecture and mechanism. Nature Rev. Micro. 2012, 10, p336-351.

Lars K, et al. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. The Royal Society of Chemistry. 2011, 40, p1224-1234.

Laura S, et al. Insights into enzyme structure and dynamics elucidated by H/D exchange mass spectrometry. Biochemistry and Biophysics. 2005, 433, p34-46.

Lee S, et al. The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. Cell. 2003, 115, p229-240.

Leung K Y, et al. Type VI secretion regulation: crosstalk and intracellular communication. Current Opinion in Microbiology. 2011, 14, p.9-15

Lewis B, et al. Type VI secretion: a beginner's guide. Current Opinion in Microbiology. 2008, 11, p3-8.

Li J, et al. Crystal structure of E. coli Hsp100 ClpB nucleotide binding domain 1 (NBD1) and mechanistic studies on ClpB ATPase activity. Journal of Molecular Biology. 2002, 318, p1127-1137.

Li M, et al. Structural Basis for Type VI Secretion Effector Recognition by a Cognate Immunity Protein. PLoS Pathogens. 2012, 8(4) e1002613

Lilic M, et al. A common structural motif in the binding of virulence factors to bacterial secretion chaperones. Molecular Cell. 2006, 21, p653-664.

Lisa G Pell, et al. The phage λ major tail protein structure reveals a common evolution for long-tailed phages and Type VI bacterial secretion system. PNAS. 2009, 106(11), p4160-4165.

Ma K, et al. Polyproline II helix is a key structural motif of the elastic PEVK segment of titin. Biochemistry. 2001, 40, p3427-3438.

Maria H D, et al. General secretion signal for the mycobacterial Type VII secretion pathway. PNAS. 2012, 109(28), p11342-11347.

Mark J Pallen, et al. Tetratricopeptide-like repeats in Type-III-Secretion chaperones and regulators. FEMS Microbiology Letters. 2003, 233, p53-60.

Michelle B, et al. Role of predicted transmembrane domains for Type III translocation, pore formation and signaling by the Yersinia pseudotuberculosis YopB protein. Infection and Immunity. 2005, 73(4), p2433-2443.

Mueller C A, et al. The Type III Secrertion tip complex and translocon. Molecular Microbiology. 2008, 68(5), p1085-1095.

Nancy J, et al. Analysis of the crystal structure of the ExsC-ExsE complex reveals distinctive binding interactions of the Pseudomonas aeruginosa Type III secretion chaperone ExsC with ExsE and ExsD. Biochemistry. 2010, 49, p5870-5879.

Nikola T, et al. Protein side-chain dynamics and residual conformational entropy. J.AM.CHEM.SOC. 2009, 131, p615-622.

Olaf S and Dominique MM. Protein secretion and surface display in gram-positive bacteria. The Royal Society of Biological Sciences. 2012, 367, p1123-1129.

Panagiotis F Sarris, et al. *Pseudomonas entomophila* and *Pseudomonas mendocina*: potential models for studying the bacterial type VI secretion system. Infection, Genetics and Evolution. 2011, 11, p.1352-1360

Park SB, et al. Outer Membrane Vescicles as a Candidate Vaccine against Edwardsiellosis. PLoS One. 2011, 6(3), e17629

Patrizia A, et al. Architecture of the major component of the Type III secretion system export apparatus. Nature Structural & Molecular Biology. 2012, p1-8.

Paul D, et al. Functional domains and motifs of bacterial Type III effector proteins and their roles in infection. FEMS Microbiol Rev. 2011, 35, p1100-1125.

Petr G Leiman, et al. Type VI secretion apparatus and phage-tail associated protein complexes share a common evolutionary origin. PNAS. 2009, 106(11), p4154-4159.

Petra J, et al. Tetratricopeptide repeats in Type III chaperone, LcrH: the role in substrate binding and selection. Molecular Microbiology. 2006, 59(1), p31-44.

Philip R, et al. Binding affects the tertiary and quaternary structures of the Shigella translocator protein IpaB and its chaperone IpgC. Biochemistry. 2012, 51, p4062-4071.

Pukatzki S, et al. Identification of a conserved protein secretion system in Vibrio cholerae using the Dictyostelium host model system. PNAS. 2006, 103(5), p1528-1533.

Pukatzki S, et al. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. PNAS. 2007, 104(39), p15508-15513.

Pukatzki S, et al. Type VI secretion system: translocation of effectors and effector-domains. Current Opinion in Microbiology. 2009, 12, p11-17.

Quinaud M, et al. Structure of the heterotrimeric complex that regulates Type III secretion needle formation. PNAS. 2007, 104(19), p7803-7808.

Rachel D, et al. A Type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. Cell Host and Microbe. 2010, 7, p25-37.

Radmila J, et al. Conservation of a virulence binding motif from animal to plant pathogens: structure of the HopA1(21-102)-ShcA chaperone-effector complex of Pseudomonas synringae. Journal of Bacteriology. 2012.

Rao P S, et al. Functional genomic approach to the identification of virulence genes involved in Edwardsiella tarda pathogenesis. Infection and Immunity. 2003, 71(3), p1343-1351.

Rao P S, et al. Use of proteomics to identify novel virulence determitants that are required for Edwardsiella tarda pathogenesis. Molecular Microbiology. 2004, 53(2), p573-586.

Rekha S, et al. Genome sequence of Aeromonas hydrophila ATCC 7966. Journal of Bacteriology. 2006, 188(23), p8272-8282.

Rembert P, et al. Temperature and growth phase influence the outer-membrane proteome and the expression of a Type VI secretion system in Yersinia pestis. Microbiology. 2009, 155, p498-512.

Richard K, et al. Protein structure characterization with mass spectrometry. Spectroscopy. 2004, 18, p37-47.

Rinet R, et al. Protein conformations can be probed in top-down HDX MS experiments using electron transfer dessociation of protein ions without hydrogen scrambling. American Society for Mass Spectrometry. 2009, 20, p1514-1517.

Ronald R and Wieslaw S. Yersinia pestis YopD 150-287 fragment is partially folded in the native state. Protein Expression and Purification. 2008, 58, p53-60.

Roxane S, et al. ESX/Type VII secretion system and their role in host-pathogen interaction. Current Opinion in Microbiology. 2009, 12, p4-10.

Sakshi S, et al. Identification and functional characterization of gene components of Type VI secretion system in bacterial genomes. PLoS ONE. 2008, 3(8), e2955.

Sara B and Partho G. Structure of the Yersinia Type III secretory system chaperone SycE. Nature Structure Biology. 2001, 8(11), p974-978.

Smarajit C, et al. Temperature and Mg^{2+} sensing by a novel PhoP-PhoQ two-component system for regulation of virulence in Edwardsiella tarda. The Journal of Biological Chemistry. 2010, 285(50), p38876-38888.

Sun P, et al. Structural characterization of the Yersinia pestis Type III secretion system needle protein YscF in complex with its heterodimeric chaparone YscE/YscG. Journal of Molecular Biology. 2008, 377, p819-830.

Sunny K, et al. LcrH, a Class II chaperone from the Type Three secretion system, has

a highly flexible native structure. Journal of Biological Chemistry. 2012,

Tan Y W, et al. Mapping of the chaperone AcrH binding regions of translocators AopB and AopD and characterization of oligomeric and metastable AcrH-AopB-AopD complexes in Type III secretion system of Aeromonas hydrophila. Protein Science. 2009, 18, p1724-1734.

Tan Y W, et al. Structure of AscE and induced burial regions in AscE and AscG upon formation of chaperone needle-subunit complex of Type III secretion system in Aeromonas hydrophila. Protein Science. 2008, 17(10), p1748-1760.

Tiago RD, et al. Coiled-coils in the YopD translocator family: a predicted structure unique to the YopD N-terminus contributes to full virulence of Yersinia pseudotuberculosis. Infection, Genetics and Evolution. 2012, 12, p1729-1742.

Tomoko K, et al. Supramolecular structure of the Salmonella typhimurium Type III protein secretion system. Science. 1998, 280(24), p602-605.

Tseng TT, et al. Protein secretion systems in bacterial-host associations, and their description in gene ontology. BMC Microbiology. 2009, Vol 9

Vladimir A, et al. A simple approach to analyzing protein side-chain dynamics from ^{13}C NMR relaxation data. Journal of Magnetic Resonance. 1998, 130, p329-334.

Xin H, et al. Interaction of the disordered Yersinia effector protein YopE with its cognate chaperone SycE. Biochemistry. 2009, 48, p11158-11160.

Xin W, et al. Edwardsiella tarda T6SS component evpP is regulated by esrB and ion, and plays essential roles in the invasion of fish. Fish and Shellfish Immunology. 2009, 27, p469-477.

Yahua Chen, et al. Regulation of Type VI Secretion System during *Burkholderia pseudomallei* Infection. Infection and Immunity. 2011, Vol.79, p.3064-3073

Yang D. Probing protein side chain dynamics via ^{13}C NMR relaxation. Protein and Peptide Letters. 2011, 18, p380-395.

Yang M, et al. Edwardsiella comparative phylogenomics reveal the new intra/inter-species taxonomic relationshipfs, virulence evolution and niche adaptation mechanisms. PLoS ONE. 2012, 7(5), e36987.

Yingqi Xu, et al. A new strategy for structure determination of large proteins in solution without Deuteration. Nature Methods, Vol.3 No.11 p.931-937

Yoshida S, et al. Shigella deliver an effector protein to trigger host microtubule destabilization, which promotes Rac1 activity and efficient bacterial internalization. The EMBO Journal. 2002, 21(12), p2923-2935.

Yutaka H, et al. Shigella IpgB1 promotes bacterial entry through the ELMO-Dock180 machinery. Nature Cell Biology. 2007, 9(1), p121-128.

Zhang H, et al. Crystal structure of Type VI effector Tse1 from Pseudomonas aeruginosa. FEBS Letters. 2012, 586, p3193-3199.

Zhang Z, et al. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. Protein Science. 1993, 2, p522-531.

Zheng J and Leung K Y. Dissection of a Type VI secretion system in Edwardsiella tarda. Molecular Biology. 2007, 66(5), p1192-1206.

Zheng J, et al. EscC is a chaperone for the Edwardsiella tarda Type III secretion system putative translocon components EseB and EseD. Microbiology. 2007, 153, p1953-1962.

Zheng J, et al. Genetic Analysis of Anti-Amoebae and Anti-Bacterial Activities of the Type VI Secretion System in *Vibrio cholerae*. PLoS One. 2011, 6(8) e23876

Zheng J, et al. Regulation of a Type III and a putative secretion system in Edwardsiella tarda by EsrC is under control of a two-component system, EsrA-EsrB. Infection and Immunity. 2005, 73(7), p4127-4137.

Zheng Z, et al. The transiently ordered regions in intrinsically disordered ExsE are correlated with structural elements involved in chaperone binding. Biochemical and Biophysical Research Communications. 2012, 417, p129-134.

Zhong D, et al. The Salmonella Type III secretion system inner rod protein PrgJ is partially folded. The Journal of Biological Chemistry. 2012, 287(30), p25303-25311.

Appendix I

Recipe for M9 medium

To 773 ml of sterile water, add the following:

200 ml	5X M9
2 ml	1M $MgSO_4$
5 ml	20% Glucose
0.1 ml	1M $CaCl_2$
4 ml	0.25 g/ml NH_4Cl
1 ml	0.1 g/ml Ampicillin

Composition:

1. 5X M9

   30 g Na_2HPO_4
   15 g KH_2PO_4
   2.5 g NaCl
   Add water to 1L
   (Autoclave)

2. 0.25 g/ml NH4Cl

   50 g NH_4Cl, add water to 200 ml
   (Autoclave, or filter-sterilize for ^{15}N)

3. 20% Glucose

   40 g Glucose, add water to 200 ml
   (Autoclave or filter-sterilize for ^{13}C)

4. 1M $MgSO_4$

   19.72 g $MgSO_4$, add water to 80 ml
   (Autoclave)

5 1M $CaCl_2$

   11.76 g $CaCl_2$, add water to 80 ml
   (Autoclave)

Appendix II

Buffers for Ni-NTA Affinity Chromatography

• Nickel binding buffer

5 mM	Immidazole
0.5 M	NaCl
20 mM	Tris pH 8.0

• Elution Buffer

0.5 M	Immidazole
0.5 M	NaCl
20 mM	Tris pH 8.0

• Strip Buffer

100 mM	EDTA pH 8.0
0.5 M	NaCl
20 mM	Tris pH 8.0

• Ni-Charge buffer

50 mM	$NiSO_4$

Buffers for Glutathione-Sepharose Chromatography

• Phosphate Buffer Saline (PBS)

140 mM	NaCl
2.7 mM	KCl
10 mM	Na_2HPO_4
1.8 mM	KH_2PO_4

• GST Elution Buffer

50 mM	Tris-Cl pH 8.0
10 mM	L-Gluthathione (reduced)

• Column Regeneration Buffer 1

0.1 M	Tris-Cl pH 9.0
0.5 M	NaCl

• Column Regeneration Buffer 2

0.1 M	Sodium acetate pH 4.5
0.5 M	NaCl

Appendix III

Recipe for SDS-PAGE

	15% Separating Gel	4% Stacking Gel
30% Acrylamide/0.8% Bis	3.5 ml	0.4 ml
Resolving Buffer	1.75 ml	-
Stacking Buffer	-	0.75 ml
10% SDS	70 µl	30 µl
Water	1.68 ml	1.8 ml
10% Ammonium Persulphate	42 µl	30 µl
TEMED	4.2 µl	3 µl

1. Acrylamide solution (30%, 0.8 % Bis)
2. Resolving gel buffer: 1.5 M Tris, pH 8.8

   For 200 ml, 36.3 g of Tris adjust to pH 8.8 with HCl

3. Stacking gel buffer: 0.5 M Tris, pH 6.8

   For 200 ml, 12.1 g Tris, adjust to pH 6.8 with HCl

4. 10% Ammonium persulphate: 1 g Ammonium persulphate in 10 ml water
5. 10X Tank buffer: for 2 liters

   60 g Tris base
   288g Glycine
   200 ml 10% SDS solution
   Add water to 2 liters

6. 2X SDS gel sample buffer: for 50 ml

   5 ml Glycerol
   6.25 ml 0.5 M Tris pH 6.8
   12.5 ml 10% SDS
   2.5 ml β-mercaptoethanol
   0.01 g Bromophenol blue
   Add water to 50 ml

7. Stain stock: 1% Coomassie Brilliant Blue G (Stir and filter)
8. Destaining solution: for 2 liters

   140 ml Acetic acid
   100 ml Methanol
   Add water to 2 liters