PART I: CRYSTALLIZATION OF A TYPE IV PILIN FROM *PSEUDOMONAS AERUGINOSA*

PART II: CHARACTERIZATION OF A PEPTIDYL-PROLYL *CIS,TRANS*-ISOMERASE THROUGH X-RAY CRYSTALLOGARPHY

AYAT YASEEN

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

GRADUATE PROGRAM IN CHEMISTRY

YORK UNIVERSITY

TORONTO, ONTARIO

JANUARY 2018

`

Abstract

Within a host, pathogenic bacteria employ several mechanisms that enhance their survival and motility. These mechanisms include their ability to adhere, replicate (in order to avoid eradication), and secretion of virulent proteins. The study of proteins involved in bacterial pathogenesis provides us with a more thorough understanding of their mechanism and function, which can lead to the development of more effective therapeutics. In this study proteins secreted by gram-negative bacteria are explored, specifically those from *Pseudomonas aeruginosa* and *Helicobacter pylori*. *P. aeruginosa* is a common opportunistic pathogen associated with 10% of hospital infections, mainly owed to their ability to bind to biotic and abiotic surfaces. Type IV pili secreted by *P. aeruginosa* are associated with adhesion, motility and DNA transfer. Part I details the cloning to crystallization of KB7, a pilin secreted by *P. aeruginosa*. The second bacteria studied in Part II is *H. pylori,* associated with gastric ulcers and gastric inflammation compromising 50% of the global population, where severity of infection is highly dependent on the strain of *H. pylori,* and the individual infected. The secreted protein HP0175 from *H. pylori* binds to Toll-Like Receptor 4 and activates a cascade of mechanisms leading to apoptosis, as well as triggering the innate immune response. HP0175 is also classified as a peptidyl-prolyl *cis,trans*-isomerase involved in the isomerization of proline peptide bonds preceding the N-terminal. Here, apo-HP0175 was crystallized to 2.09 Å in space group $P3_221$ with one monomer in the asymmetric unit; the dimer is generated through symmetry mates. A comparison to indole-2-carboxylic acid bound HP0175 shows N- and C- terminal helix extensions upon interaction of the catalytic residues in the binding pocket. Helix extension supports other parvulin findings that N- and C- terminal helices stabilize proteins undergoing catalysis by protein-protein interaction.

`

Acknowledgments

I would first like to thank my supervisor Dr. Gerald Audette for giving me this incredible opportunity to pursue my passion in research. Thank you for always believing in me and providing all the tools I needed to succeed in the field of protein crystallography. I would like to thank my committee members Dr. Derek Wilson and Dr. Vivian Saridakis who have always been accessible and welcoming in providing constructive feedback, guidance and assistance. As well as the chemistry department for all their assistance with regards to organization, support and GSA; Magy Baket, Mary Mamais, and Natasha Bissoon.

I thank my colleagues in the Life Sciences Building who have made my time here very memorable with an amazing support system. Cristina Lento, for welcoming me into the Audette lab as her student, a dear friend who I could always count on to guide me in the right direction and demonstrate an admirable work ethic. Shaolong Zhu, who had always been supportive and welcomed me as a part of Dr. Wilson's lab group, thank you for always believing a protein was bound to crystallize. Kerene Brown who was always accessible for mass spectrometry related feedback and help. Irina Organesyan and Xiaojing Haung who were very helpful with mass spectrometry. John Van Nostrand for always providing his wisdom about instrumentation, work, and anything anyone needed help with, he was always happy to assist. Dayana D'Amora who was has always showed a relentless work ethic a great friend and colleague who kept me in great company on long days and holidays. Jyotsna Vinayak for providing help with cloning or any biology related assays and also providing a great environment to work in. Ana Vakiloroayaei for being supportive. From the Bayfield lab, Rawaa Hussain and Stefano Marrella for creating such a nice and welcoming work environment. As well as their feedback into molecular biology was always well appreciated. My lab member Nicholas Bragagnolo for all his assistance and friendship.

I would also like to thank my family, my mother Huda Yaqub, my sister Wid Yaseen, and my brother Ahmed Yaseen for their constant support throughout this process. I could not have done this without them, and would like to dedicate my thesis to my mother who has always demonstrated to us that hard work and passion will always guide us.

`

Table of Contents

`

`

List of Tables

`

List of Figures

`

`

List of Abbreviations

GC	Gonococcal
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
MBP	Maltose binding protein
MBP	Membrane binding partner
DNA	Deoxyribonucleic acid
T4PM	Type IV pili membrane
sfGFP	Superfolded green fluorescent protein
Cryo-ET	Cryo electron tomography
T4P	Type IV pili
SEC	Size exclusion chromatography
PNT	Protein nanotubes
IL	Interleukin
Th	T helper cell
TGF	Transforming growth factor
TLR	Toll like receptor
VEGF	Vascular endothelial growth factor
MMP	Metalloproteinase
PAMPs	Pathogen-associated molecular patterns
LPS	Lipopolysaccharide
NF-κB	Nuclear factor kappa B
CsA	Cyclosporine
PPIase	Peptidyl prolyl isomerase
PRRs	Pathogen recognition receptors
CLR	C-lectin receptors
NOD	Nucleotide binding oligomerization domain
AGS	Human Caucasian gastric adenocarcinoma
ASK1	Apoptosis signal regulating kinase 1
MAPK	Mitogen-activated protein kinase
TNF	Tumor necrosis factor
IRAK	Interleukin 1 receptor associated kinase
MAP	Mitogen activated protein
IFN	Interferon

`

ssRNA	Single stranded ribonucleic acid
CMCF	Canadian Macromolecular Crystallography Facility
CLS	Canadian Light Source
OD	Optical density
LB	Luria Bertani
TB	Terrific broth
EMBL	European Molecular Biology Laboratory
NEB	New England Biolabs
CIP	Calf intestinal phosphate
TCAG	The Centre for Applied Genomics
Km	Kanamycin
GST	Glutathione-S-transferase
PEG	Polyethylene glycol
MMS	Microseed matrix screening
RMMS	Random matrix microseed screening
Cryo-EM	Cryo-electron microscopy
NMR	Nuclear magnetic resonance
MS	Mass spectrometry
TEM	Transmission electron microscopy
RF	Rotation function
TF	Translational function
MR	Molecular replacement
Phenix	Python-based Hierarchiacal ENvironment for Integrated Xtallography
CCP4	Collaborative Computational Project No. 4
ESI	Electrospray Ionization
m/z	Mass to charge ratio
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MWCO	Molecular weight cut-off
Tris	Tris(hydroxymethyl)aminomethane
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
SAD	Single-wavelength anomalous diffraction
MAD	Multi-wavelength anomalous dispersion

Chapter 1 : Introduction

1.1. Infections Caused by Pathogenic Bacteria

The continuous study of microorganisms is often driven due to their pathogenesis, which can affect the health of humans, animals, and plants. This complex phenomenon carried by bacterial system proteins, begins with adhesion to mucosal surfaces and is commonly observed in respiratory, gastrointestinal, and genitourinary tracts.[1] For successful colonization of the host, bacteria must overcome both non-specific and specific host defense mechanisms such as sneezing, coughing, cellular apoptosis, as well as avoid recognition by the immune system.[1] Bacteria are capable of colonization through distinct mechanisms, including adhesion to epithelial cells using various fimbrial proteins, replication (to avoid eradication), and secretion of various virulent proteins that can cause disease.

1.2. Bacterial Pili

Bacterial membranes harbor interesting "nano-machines" that aid in movement and biofilm formation.[2] Flagella are involved in movement and motility, while pili affect micro-colonization,[3] biofilm formation,[4] cell signalling,[5] cell-host adhesion,[6] DNA uptake by natural transformation,[7] phage attachment,[8,9] and twitching motility.[10] Pili are a diverse class of proteins often found in gram-negative bacteria (such as *Pseudomonas aeruginosa*), as well as several gram-positive bacteria (such as *Bacillus subtilis*).[11] Structural and functional studies of pili provide insight on the formation of infection at the molecular level.[11] Extensive studies of pili has led to their classification based on their assembly in gram-negative bacteria (Figure 1.1). Examples of highly studied pili classes include: the chaperone-usher pathway (Type I pili), nucleation-dependent polymerization (curli),[12] and the type II secretion system (Type IV pili).[11–13]

Type I pili are linear, unbranched polymers, composed of pilin subunits that range from 12 kDa to 20 kDa in size. Their assembly is mediated by the chaperone-usher pathway of pilus biogenesis, where the chaperone directs pilus assembly to be secreted through the usher pore.[14] Several type I pili are also involved with specific host attachment and evasion.[15] Curli contain a coiled-coil structure that can be viewed by electron microscopy. Involved in bacterial aggregation,

`

Curli are able to withstand harsh environments such as 90% formic acid and high temperatures.[16,6] Type IV pili contain similar structural characteristics, with a mode of secretion related to the type II system. However, in comparison to other systems, structural and functional studies are highly sought after in type IV pili, due to their DNA transformation capability.[12]

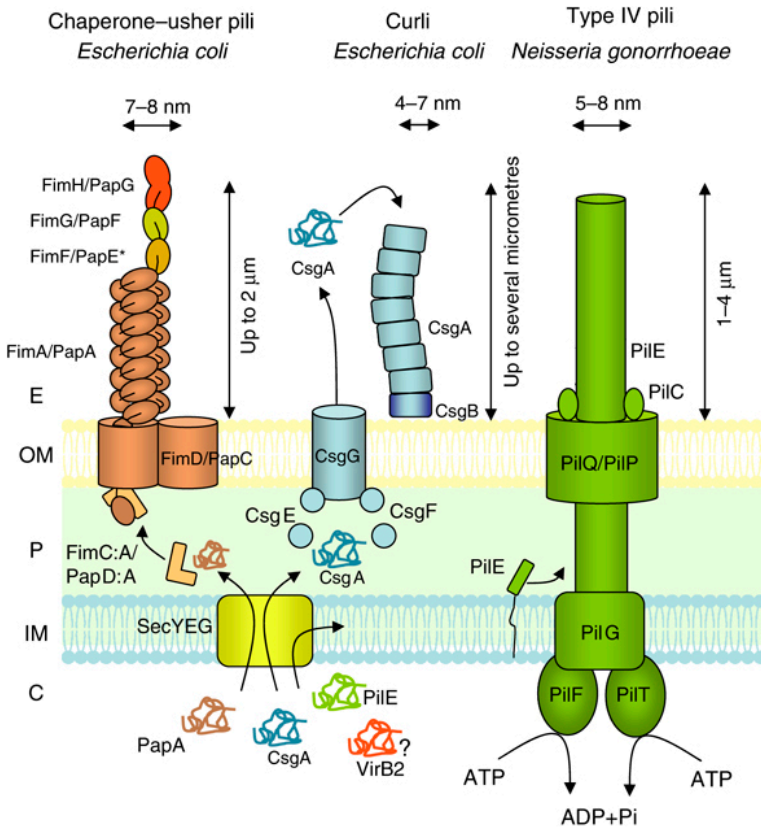


Figure 1.1. A structural comparison of gram-negative bacterial pili. The chaperone-usher pili (type I) and Curli contain simple systems that span the outer membrane. Chaperone-usher pili denoted by the PapA polymer is assembled as a right-handed "one-start" helix. The tip known as the PapG pilin contains adhesive properties that is linked to the homopolymer via PapF and PapE. Curli filaments are assembled through edge-to-edge aggregation of CsgA and CsgB, secreted by the CsgG outer membrane pore. Type IV pili is comprised of a complex system that spans from the cytoplasm to the outer membrane. Assembly and disassembly of pilin monomers is driven by an ATPase pump that extends from the inner membrane (PilF/PilT). PilQ/PilP are involved in the transportation and stabilization of PilE monomers through the outer membrane. Figure adapted from Fronzes et al., 2008.[6]

`

1.2.1. Type IV Pili Structural Characteristics

Type IV pili are polymers composed of monomeric pilin subunits that share a distinctive architecture (Figure 1.2) with an N-terminal α-helix, C-terminal globular head domain, $\alpha\beta$-loop, and a disulfide bounded region (D-region). Moreover, these pili are further classified as type IVa and type IVb, which is determined primarily by their amino acid sequence. Type IVb pili remain commonly associated with bacteria that colonize human intestines such as *Escherichia coli* and *Salmonella enterica serovar typhi*. Their composition includes a leader peptide sequence of 15-30 amino acids, 190 amino acids of mature protein, a 55-amino acid D-region, and a variable N-terminal methylated residue of methionine, leucine or valine.[9] By contrast, type IVa pili possess a broad range of targets including mammals, plants, fungi, as well as other bacteria such as *P. aeruginosa*, *Shewanella spp.,* and *Geobacter spp.*[17] Overall, type IVa pili possess shorter peptide sequences with a leader sequence of 5-6 amino acids, mature proteins composed of 150 amino acids, a D-region that on average spans 22 residues, and interestingly the N-methylated N-terminal residue is phenylalanine.

Crystallization of a native pilin is challenging due to the hydrophobic nature of the α-helical N-terminal domain, as well their tendency to form fibers exceeds the rate of crystallization.[18] To date, only two full-length pilin proteins have been characterized by X-ray crystallography. Comparison of the N-terminal α-helix was achieved through crystallization of full-length gonococcal (GC) pilin from *Neisseria gonorrhoeae* (MS11 strain)[19] and the *P. aeruginosa* strain K (PAK) pilin.[20] Comparison of these two structures lead to the observation that the N-terminal α-helical domains behave similarly.[18] Truncation of the N-terminal helix aids in pilin solubilization, thus full-length and truncated PAK structures were compared. Removal of 28 amino acids (Δ1-28) had no effect on the folding of the globular head domain with a root-mean-square deviation (RMSD) of 0.69 Å for all atoms.[19] The N-terminal α-helix domain (residues 1-52) is generally 80 Å length and contains a slight curvature (kink), caused by a proline residue at position 22 both PAK and GC pilins. A second kink is observed in residue 42, occupied by a proline in PAK pilin and a glycine in the GC pilin.[18,19] The kinks are thought to contribute to fiber flexibility, and may reduce inter-subunit packing, which subsequently plays a role in fiber assembly and disassembly.[17] In comparison to the α-helix in type IVa pili, the α-helix of type IVb pili does not contain the same kinks generated by Pro22 and Gly42. The structure of type IVb pili

`

is hypothesized to be less curved, and therefore more flexible due to the lack of this constraint. In addition, the N-terminal α-helix is not required for protein oligomerization, observed in studies of truncated *P. aeruginosa* strain K122-4 (Δ1-28 a.a) pili.[21–23]

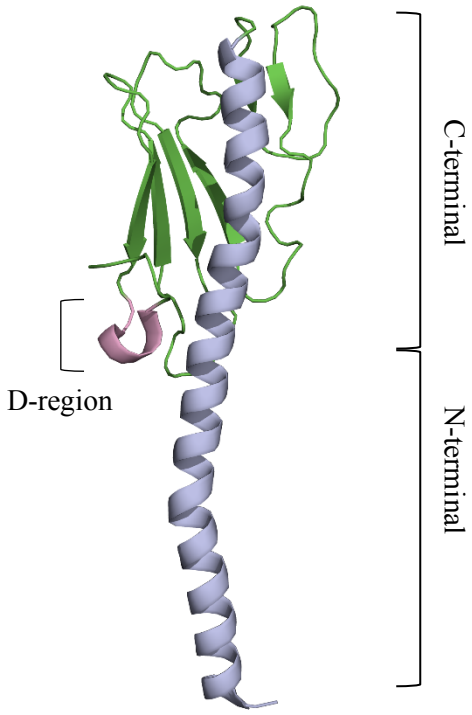


Figure 1.2. Schematic representation of a type IV pilin. The full-length PAK pilin structure highlights the N-terminal *α*-helix that extends to the C-terminal region (green). The D-region (pink) binds to steel and cell surfaces for motility. Upon packing, the D-region is generally conserved except at the tip of the pilus. The helical polymer contains 4-5 subunits per turn, 40 Å pitch, and is 6 nm in length. Figure adapted from PDB entry 1OQW.[18]

1.2.2. Type IV Pili Fibers

Type IV pili fibers are assembled at a rate of 1000-2000 subunits/s and are able to withstand forces greater than 100 pN.[24,25] Stabilization of pilus assembly is achieved through the *α*-helical N-terminal region, through hydrophobic and subunit-subunit interactions.[17] Assembly of the pilus can be characterized by the axial rise of the fiber, as well as the progression of protein addition along the filament.[26] Based on axial direction, this assembly can be described in three ways: 1) a

`

right handed 1-start, 2) left handed 3-start or 3) a right handed 4-start helical assembly. The smallest axial rise represents a right-handed 1-start symmetry, where n represents the first protein, n+1 the second, and n+2 the third. In the left-handed 3-start helical symmetry, subunits along the axial position are n, n+3, n+6 and so forth (Figure 1.3A and 1.3B). The last representation is the right-handed 4-start assembly of proteins n, n+4, n+8, etc. The kinks provided by Pro22 and Gly42 in the α-helix influences the formation of an "S shape" that aids in the close packing of the polymer.[27]

Interactions of type IVa pili within the C-terminal globular head domain are facilitated by subunit-subunit interactions, which are primarily found in the loop regions. Further analysis into interactions within the C-terminal region displays polar interactions between the $\alpha\beta$-loop and the D-region of the next subunit, and interactions proceed between loop regions connecting β-strands in the 3-start and 4-start helical views, respectively.[17] Type IVa pili possess a diameter of 60 Å in comparison to type IVb pili that have larger diameters of up to 90 Å. The larger diameter in type IVb pili is associated with decreased interactions within the C-terminal region, which is only facilitated by polar and hydrophobic interactions between $\alpha\beta$-loop, $\alpha 3$, and parts of the N-terminal domain.[19,28] The looser packing has been hypothesized to decrease fiber resilience within temperature, harsh environments, and proteases when compared to type IVa pili (Figure 1.3C).[17]

The pathogenic gram-negative *P. aeruginosa* has been implicated in infections such as ventilator associated pneumonia.[29,30] Type IV pili secreted by *P. aeruginosa* can bind to biotic and abiotic surfaces, allowing it to be successful for targeting immune-compromised hospital patients.[31] Specifically, these infections can occur with cancer,[32] cystic fibrosis,[33] and human immunodeficiency virus (HIV) patients.[34] The primary mode of contact for infection begins with type IV pili, which has been observed to enhance the pathogen virulence. Further analysis into type IV pili binding to cell surface receptors demonstrated that pilus interaction is only achieved at the surface monomer through the D-region. The D-region recognizes asialo-gangliotetraosylceramide; GM_1 and GM_2 cell surface receptors, specifically the disaccharide β-D-GalNAc(1-4)-β-D-Gal moiety.[35]

`

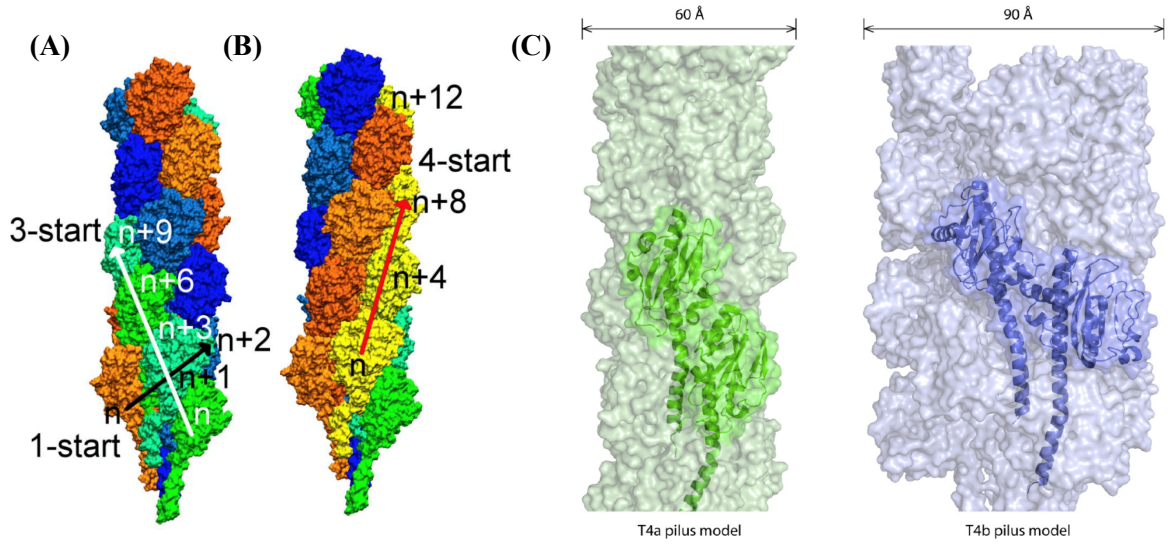


Figure 1.3. X-ray and cryo-electron microscopy representation of a T4P assembly. **(A)** Assembly of T4P in a right handed 1-start helical view (outlined by a black arrow). Left handed 3-start helical view (outlined by a white arrow). **(B)** Right handed 4-start helical view (outlined by a red arrow). Adapted from Baker et al.,2013.[26] **(C)** The T4a pilus is held together through tight interactions between N-terminal α-helix within the core. The T4b pilus model shows a bulkier assembly in comparison to T4aP, as C-termini head groups are more loosely packed. The closer packing of the T4a pilus (60 Å) can be attributed to its apparent resistance to heat, chaotropic agents and proteases. Adapted from Giltner et al., 2012.[17]

1.2.3. *In vivo* Assembly of Type IV Pili Fibers

Interestingly, bacteria containing type IV pili are able to alter their sequence while maintaining their characteristic pili structure.[22] Presence of the same overall structure of the pilin monomer allows for the assembly and disassembly of pili through a type II secretion system, regardless of sequence diversity. The working model for assembly and disassembly is based on studies with *N. gonorrhoeae* bacteria associated with sexually transmitted infections.[28] A general mechanism, suggested by Craig et al. in 2006,[28] demonstrated the addition of pilin subunits within the inner membrane (IM). Monomer assembly is facilitated by diffusion and electrostatic interactions between the incoming Glu5 (monomer n) to the pre-existing Phe1 subunit (n+1). For the helix to be assembled, adenosine triphosphate (ATP) bound to the hexameric ATPase must be hydrolyzed from the cytosol, inducing a conformational change. The new conformation allows for piston-like movement in the membrane binding partner (MBP) protein. Once the fiber filament has extended 10.5 Å from the outer membrane (OM), ATP replaces bound ADP and MBP returns

`

to its initial position. The addition of pili monomers within the IM occurs in three active sites surrounding the filament diameter. Figure 1.4 represents the addition at one active site.[28]

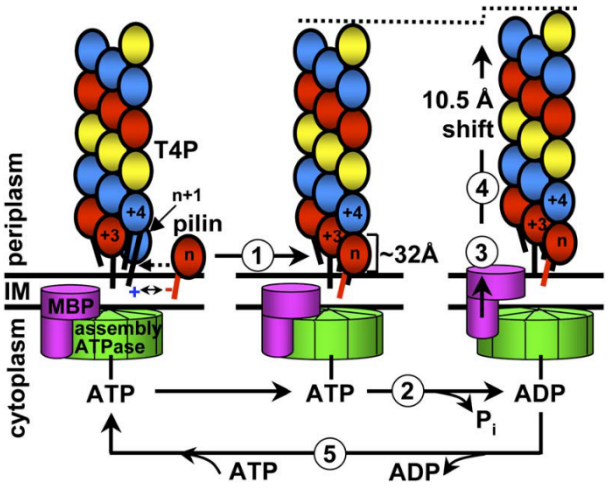


Figure 1.4. The proposed mechanism of pilus assembly. **(1)** Addition of subunit n to the growing pilus filament. **(2)** Hydrolysis of ATP. **(3)** Hydrolyzed ATP induces piston-like motion of the membrane bound protein (MBP). **(4)** Pilus filament extended 10.5 Å into the periplasm. **(5)** Exchange to ATP, back to resting position. Adapted from Craig et al., 2006.[28]

Recently, Chang et al.[36] further explored type IVa pili assembly and disassembly with intact *Myxococcus xanthus* cells, using X-ray crystallography and cryo-electron tomography (Cryo-ET). This nano-machine spans both the inner and outer membrane and is comprised of following: the outer membrane pore, periplasmic stem, cytoplasmic disc and dome, three-connected periplasmic rings, and a ring within the cytoplasm.[37] A total of 10 proteins constitute the type IV pili machinery (T4PM): PilQ, PilP, TsaP, PilN, PilO, PilA, PilC, PilB, PilT, and PilM. The major pilin protein, PilA, is found in the periplasmic space as single monomers until the assembly into a protein fiber is facilitated by PilC. Peptidoglycan-binding protein (Tsap) forms the periplasmic ring around PilQ, while the mid-periplasmic ring is composed of PilP. The lower periplasmic ring is constructed of PilO and PilN, connected by coiled-coils that extend to the IM until it reaches the PilM in the cytosolic ring. Hydrolysis of the triple ATPase PilB induces rotation of PilC in order to drive the PilA fiber through the OM. Once PilA adheres to a surface, it relays a

`

signal back to the T4PM to induce a conformational change that releases PilB. PilT can then bind, which allows PilC to rotate in a manner that disassembles PilA monomers from the fiber.[36]

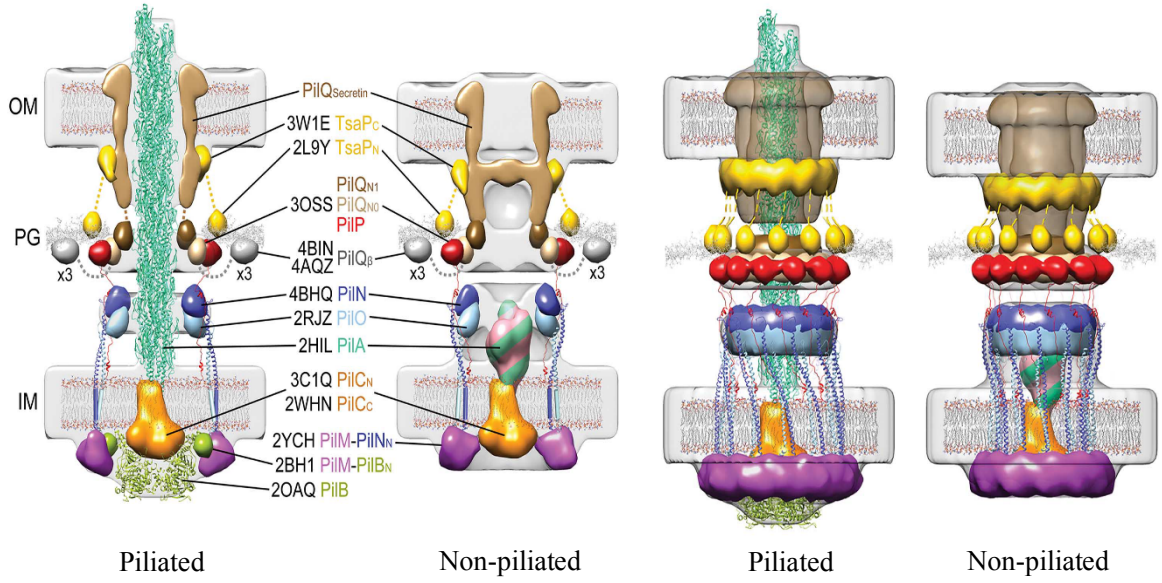


Figure 1.5. Type IV pilus machinery. A representation between piliated and non-piliated T4P machinery containing 10 different proteins. PilB binds at the basal body of PilM with PilC and is involved in the extraction of PilA and PilT in the retraction of PilA. PilB is absent in the non-piliated structure, therefore not involved in pilus retraction. PilQ extends from the outer membrane into the periplasm, the periplasmic ring encompassing PilQ is composed of Tsap and the mid-periplasmic ring is composed of PilP. Adapted from Chang et al., 2016.[36]

1.2.4. Pili Secreted by *P. aeruginosa*

P. aeruginosa causes respiratory infections with patients whose immune systems are compromised when suffering from diseases such as cancer, cystic fibrosis, as well as patients recuperating in intensive care units.[38] Infection by *P. aeruginosa* can become chronic, seen with biofilm formation in cystic fibrosis patients.[39] Pili secreted on the surface of *P. aeruginosa* have been rigorously studied due to their ability to adhere to host cells.[40] The different clinical strains of *P. aeruginosa* retain a consistent pilin structure, variability is observed within the C-terminal and D-regions. The D-region is able to bind to the cell surface receptor, asialo-GM_1 and GM_2, and is a potential target for cross-reactive vaccines. Development of cross-reactive vaccines for these targets requires the detailed study of structural differences, the effects of targeted binding on the

`

pili, and factors that contribute to immunogenicity.[41] To generate a cross-reactive vaccine, the goal is to synthesize an antibody that will undergo a conformational change and fit into the binding regions of *P. aeruginosa* pili strains PAK, KB7, P1, and PAO. The antibody PAK-13 was selected for pili binding and was found to block adherence by interacting with the receptor binding site. The receptor binding domain, which was characterized by nuclear magnetic resonance (NMR) with the antibody, confirmed the presence of two β-turns. It was also found that the sequence elements required for binding was in the form of: Asp134-X-X-Phe137 and Pro139-X-Gly-Cys143 (where X is any residue), demonstrating that cross-reactive antibody recognition is possible.[41,42]

Site-directed mutagenesis of specific residues in the binding region mapped the binding affinity of KB7 and PAK to A549 cells.[43] Important residues for this affinity were determined to be: Ser131, Gln136, Ile138, Pro139, Gly141, and Lys144. The disulfide bridge of cysteines 129 and 142 was also important for receptor recognition. In KB7, the residues of importance were found to be: Ala130, Thr131, Thr132, Val133, Asp134, Ala135, Lys136, Arg138, and Pro139. Receptor binding motif (Adhesintope) mapping confirmed that the backbone conformations of PAK and KB7 are very similar within the two β-turn regions. While changing several residues does not affect receptor interactions, in PAK and KB7,[40] the exact positions of basic or charged residues differs. The receptor binding site for Ile138 in PAK is replaced by Arg138 in KB7, with Val133 maintaining hydrophobicity, while the positive charge from the main residue Lys144 in PAK is replaced by Arg138 or Lys136 in KB7.

1.2.5. *In vitro* Assembly of T4P-like Nanotubes

Type IV pili form robust nanofibers, and interestingly, their truncated monomers have been used to synthesize protein nanotubes (PNTs) *in vitro*.[44] It was observed that pilin monomers can be used to form PNTs using a size-exclusion (SEC) Superdex column, indicating that hydrophobic probes induce oligomerization. Previously, it was hypothesized that filament formation must occur with T4P machinery to drive assembly, and yet recent work has determined otherwise.[44] Soluble protein nanotubes generated what resembled T4P filaments *in vitro*, and PNTs generated had a total length of 100 μm, an outer diameter of 6 nm, and an inner diameter of 2 nm.[44] The increased length is thought to be a consequence of a lack of T4P machinery, however the overall PNTs generated resembled the structure and function of T4P.[22,23,45] Particularly, it was surprising to

`

observe PNTs with the truncated *P. aeruginosa* monomer ΔK122-4, since the N-terminal α-helical domain was hypothesized to play a major role in fiber formation.[38] Oligomerization of ΔK122-4 was induced by a hydrophobic probe[44] (C_{11}-SH), and was not incorporated into the nanofiber as a replacement for the N-terminal α-helix. This was observed with a constrained hydrophobe covalently attached to maleimide-activated plates.[38]

1.3. Peptidyl-Prolyl *cis,trans*-Isomerase from *Helicobacter pylori*

1.3.1. *H. pylori*

The gram-negative *H. pylori* infects 50%[46] of the global population, about 15-20%[47] progress to develop peptic ulcers,[48] cell death,[49] and adenocarcinomas.[50] Severity of infection is dependent on the strain of *H. pylori*, as well as the innate immunity of the host. As with any pathogenic bacterium, *H. pylori* secretes several virulent proteins that include, but are not limited to: Vacoulating cytotoxin (VacA),[51] urease,[52] γ-glutamyl transpeptidase,[53] Cytotoxin associated gene A (CagA),[53] and HP0175.[54] VacA increases the formation of vacuoles in the cytoplasm of gastric cells and disrupts mitochondrial function.[51] CagA utilizes a type IV secretion system to translocate into the cytoplasm where it becomes phosphorylated, and activates interleukin-8 (IL-8).[53] Regulation of apoptosis by *H. pylori* has been linked to multiple proteins, however the peptidyl-prolyl *cis, trans*-isomerase HP0175 has been shown to increase the rate of apoptosis.[50]

Patients infected with *H. pylori* elicit an up-regulated pro-inflammatory response mediated by macrophages, chemokines, and cytokines.[55,56] HP0175 increases the activation of human effector T-cells (Th17), through its interaction with the Toll-like Receptor 4 (TLR4) pathway.[57] Key secreted cytokines associated with Th17 involved in activating other pro-inflammatory pathways include: IL-β1, IL-6, IL-17,[58] IL-21, and transforming growth factor-β (TGF-β).[59] The HP0175 activated cytokine IL-β1 can directly affect epithelial cells, and has been implicated in the progression of gastric cancer.[60,61] TGF-β upregulates vascular endothelial growth factor (VEGF), which increases the production of blood vessels.[62,63] IL-21 can stimulate the overproduction of matrix metalloproteinase (MMP)-2 and MMP-9, leading to matrix degradation to further contribute to tumour malignancy and invasion.[64] Patients infected with *H. pylori* and

`

diagnosed with gastric cancer were found to have increased levels of IL-23 and IL-17 within gastric tissue.[65]

Inflammation of gastric tissue is a distinguishing feature of *H. pylori* infection.[66] Most *H. pylori* infections progress into chronic gastritis, associated with an inadequate immune response against the bacterium.[66] The adaptive immune response activates a signalling cascade facilitated by TLRs expressed on mucosal epithelium in order to identify pathogen-associated molecular patterns (PAMPs).[67] Since *H. pylori* is a gram-negative bacterium, it is directly recognizable by TLR4 through its lipopolysaccharide (LPS).[68] It was also observed that *H. pylori* increased the expression levels of TLR4 at the apical cell surface.[67,69] Activation of the TLR4 complex triggers activation of the nuclear factor Kappa B (NF-κB) pathway, which further activates pro-inflammatory cytokines and the IL-8 pathway.[66–68]

1.3.2. Peptidyl-Prolyl *cis,trans*-Isomerases

Peptidyl prolyl *cis, trans*- isomerases (PPIases) are involved in the correct folding of newly synthesized proteins within the cell. Folding occurs with the isomerization of imidic peptide bonds N-terminal to protein residues, from *cis* to *trans*.[70] The *trans* conformation of amino acid side chains are favoured over the *cis* conformation, since it is associated with steric hindrance and releases a free energy difference of 2.6 kcal/mol (Figure 1.6A). The free energy difference for a peptidyl proline side chain is 0.5 kcal/mol between *cis* and *trans* conformations (Figure 1.6B).[71,72] Characterization of PPIases further subdivides them based on drug specificity: 1) cyclosporine A (CsA)-binding cyclophilins,[73] 2) FK506-binding proteins (FKBPs), and 3) parvulin-like[74] PPIases that do not bind to immunosuppressants (Figure 1.7).[75] PPIases are classified as globular proteins which primarily consist of β-folds in the central region (Figure 1.7).[76] FKBP proteins are characterized by the structure of FK506 (Figure 1.7A), which contains four anti-parallel β-strands and one α-helix. The overall structure of cyclophilin PpiB is a β-barrel, composed of eight anti-parallel β-strands surrounding a hydrophobic core while two short 3_{10} helices lie above and below the β-sheet (Figure 1.7B).[77] The parvulin subfamily of PPIases represented by human peptidyl-prolyl *cis,trans*-isomerase NIMA-interacting 1, encoded by gene Pin1 (Figure 1.7C), is composed of two anti-parallel β-strands surrounded by four α-helices. Although PPIases are known for

`

catalyzing peptide bonds preceding proline residues, studies have demonstrated that protein-protein interactions still occur even with PPIase domain removal.[76,78] This has now expanded the role of PPIases to include an association with signal transduction, as well as gene regulation on both transcriptional and translational levels.[78]

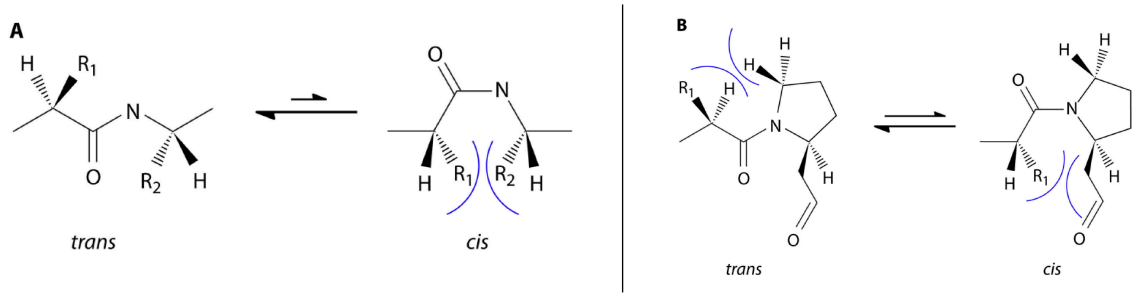


Figure 1.6. Cis- and trans- isomers of the peptide bond. (A) Denotes a general structure of a protein amide backbone. In the trans-configuration, the side chains R1 and R2 are not within close proximity. In the cis-configuration, side chains R1 and R2 could clash. **(B)** The peptide bond of proline. The trans- and cis-configurations both demonstrate rigidity, however the cis-conformation is favoured with an energy difference of 0.5 kcal/mol. Figure adapted from Ünal and Steinert, 2014.[72]

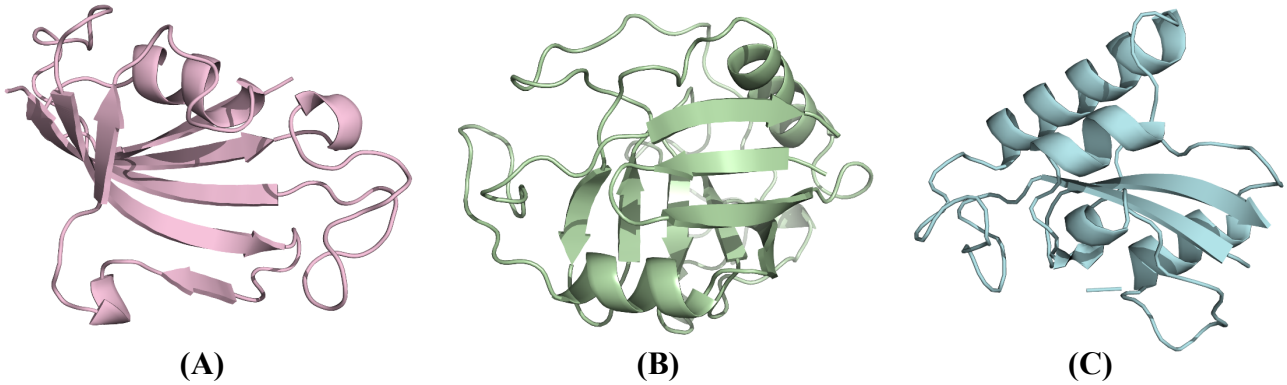


Figure 1.7. Ribbon representation of three peptidyl prolyl isomerase families. (A) The FKB12 protein regulates cell growth and metabolism in humans. Its classification is based on its ability to bind to the immunosuppressant FK506. Crystallographic representation of its PPIase: a four-stranded anti-parallel β-sheet below two α-helices (PDB ID: 1FKJ).[79] **(B)** *E. coli* cyclophilin PpiB, found in the cytoplasm binds to the immunosuppressant CsA. Overall structure of this PPIase domain is a β-barrel, with eight anti-parallel β-strands located between two α-helices (PDB ID: 2NUL).[77] **(C)** Parvulin type PPIase; represented by human Pin1 involved in the regulation of G2/M transition of eukaryotic cells. The PPIase domain (residues 43-163) are depicted with two anti-parallel β-sheets and four α-helices (PDB ID: 1PIN). Partially adapted from Ünal and Steinert, 2014.[72]

`

1.3.3. HP0175 Induced Cell Death

Cellular recognition of invaders such as pathogenic bacteria are usually performed by pathogen recognition receptors (PRRs). These PRRs recognize specific invariant structures through different sub categories of sensors; pathogen associated molecular patterns (PAMPs), TLRs, nucleotide binding oligomerization domain (NOD)-Like Receptors (NLR), and C-lectin receptors (CLR).[80,81] HP0175 has been shown to interact with TLR4 and induce apoptosis.[54] Pre-incubation of HP0175 in human Caucasian Gastric Adenocarcinoma (AGS) cells with addition of anti-TLR4 resulted in the inhibition of apoptosis. The signaling pathway for cell death by HP0175 functions through initial interaction with TLR4. TLR4 activates Apoptosis Signal regulating Kinase 1 (ASK1), which proceeds to activate p38 MAPK and caspase 8. Bid is cleaved and translocated to the mitochondria, cytochrome c is released from the mitochondria due to the membrane potential loss which subsequently releases caspase 9 and 3, and cell death occurs (Figure 1.8).[54] Additionally, cell lysates containing TLR4 can be pulled down by immobilized HP0175 on Ni^{2+}-NTA agarose, confirming an interaction between TLR4.[82] Inhibited p38 MAPK decreased apoptosis induced cell death by HP0175, while ERK1/2 MAPK had no effect. Since it was determined that p38 MAPK was involved in the apoptosis pathway, as a default from the known cascade, so is ASK1, a member of the kinase kinase family is also involved.[83–85]

1.3.4. Overview of Caspases

Apoptosis is often referred to as programmed cell death which is an important mechanism for central nervous system development, maintaining homeostasis and early brain development.[86] Caspases play an important role in cell death activation linked to their cysteine aspartate-specific protease, their primary function is to cleave peptide bonds of aspartic acid residues.[87] Caspases are present as inactive pre-cursors (procaspases) in the cell until activated by specific cell death signals. Two distinct pathways highlighted are the initiation of death receptors tumor necrosis factor (TNF) or Fas, requiring Caspase 8 or 10 in the receptor complex. Activated caspase 8 is able to further cleave downstream caspases such as caspase 7 or 3 that lead to apoptosis.[88] While an interesting pathway includes the activation through a stimuli that causes hypoxic stress, chemotherapeutic agents, growth factor withdrawal or irradiation that induces mitochondrial damage allowing for the release of cytochrome c. Cytochrome c complexed with dATP interacts

`

with Apaf-1, induces a conformational change and activates Caspase 9, which further cleaves downstream caspases.[89]

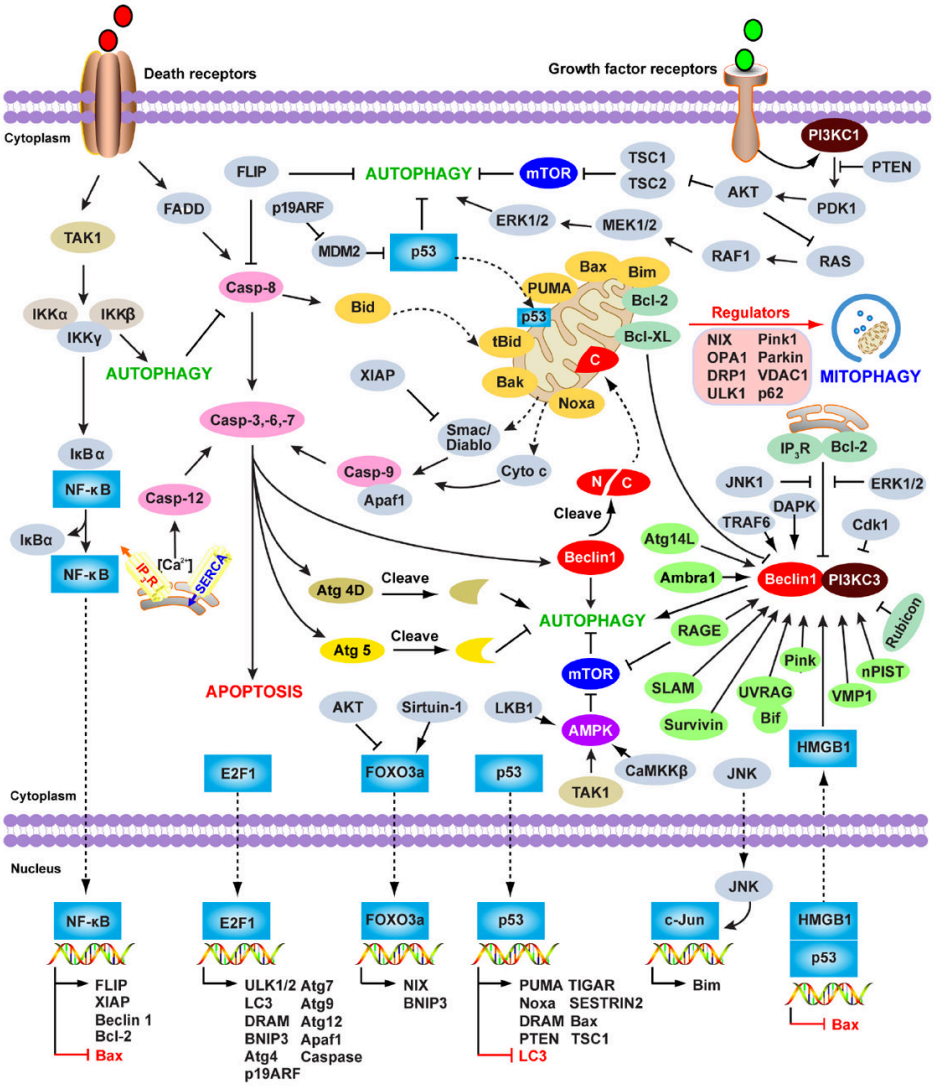


Figure 1.8. Schematic representation of the multiple pathways associated with cell death and autophagy. Apoptosis pathways outlined represent mediation by extrinsic pathway involving cell surface receptors, intrinsic pathway utilizing mitochondria and endoplasmic reticulum. Additional pathway delivers porforin by cytolytic T and NK cells. The pathways of interest for the *H.* pylori induced cell death focuses on nuclear factor Kappa B (NF-kB), caspase-8, -9, and -3, Bid, cytochrome c, ERK1/2. Adapted from Tang et al., 2012.[90]

`

1.3.5. Toll-Like Receptors

Toll-like receptors (TLRs) are membrane bound proteins involved in triggering immune responses due to their ability to recognize different pathogenic compounds, and are classified as pathogen recognition receptors (PRRs).[91] There are ten TLR paralogs found in humans that respond to different PAMPs including: lipopeptides (TLR2 associated with TLR1 or TLR6), viral dsRNA (TLR3), viral or bacterial ssRNA (TLRs 7 and 8), lipopolysaccharide (TLR4), bacterial flagellin (TLR5), CpG-rich unmethylated DNA (TLR9), and several others. TLR paralogs are located within the membrane of endosomes; their function is to defend against pathogenic organisms and compounds. TLRs are type I integral membrane receptors and are composed of N-terminal ligand recognition domain, single transmembrane helix and a C-terminal cytoplasmic signaling domain. The TLR signaling domain is referred to as Toll IL-1 receptor due to its homology to the IL-1R family. In most TLRs these signaling domains are the ones that interact with the pathogens and can initiate the signaling cascade.[92] A cascade of signalling occurs when a ligand or molecule bind to the extracellular domain of TLR and induces a conformational change within the intracellular domain.[93]

When TLRs bind to their specific target ligand, they transmit signals through myeloid differentiation primary response protein 88 (MyD88) followed by TIR domain and tumor necrosis receptor associated with factor 6 (TRAF6).[94] MyD88 also transmits signals that result in inflammatory chemokines. Association of TLRs and MyD88 activates interleukin 1 receptor associated kinase (IRAK) family. Phosphorylation of IRAK1 and IRAK4 is observed and dissociated from MyD88 results in activation of (TRAF6), that subsequently goes on to activate transforming growth factor B-activated protein kinase 1 (TAK1). Another pathway of signaling from TLRs and MyD88 is the phosphorylation of inhibitor Kappa B kinase (IKK) enzyme complex, and mitogen activated protein (MAP) kinase.[95] Eventually, activated NF-κB translocates to the nucleus for the transcription of pro-inflammatory genes such as inflammatory cytokines, Type I interferon (IFN) and chemokines in inflammation.[94]

`

1.3.5.1. TLR4

TLR4 is well known for activating an immune response through its ability to recognize lipopolysaccharides found on the surface of gram-negative bacteria.[96] The recognition of pathogens involves hetero-dimerization with myeloid differentiation factor 2 (MD-2) as represented by the crystallographic structure in Figure 1.9.[96] Previous data by Park et al. (2009)[96] demonstrated that TLR4 behaves in complex with MD-2, where binding of the pathogen is recognized by the hydrophobic pocket of MD-2. With the lipopolysaccharide tested, five of the six chains of LPS interact with MD-2 buried in the hydrophobic pocket, induces a slight conformational change that allows for TLR4 and MD-2 to interact through hydrophilic residues. Once LPS binds to MD-2, it bridges the two MD-2's and forms a multimeric complex in the shape of an M.[96]

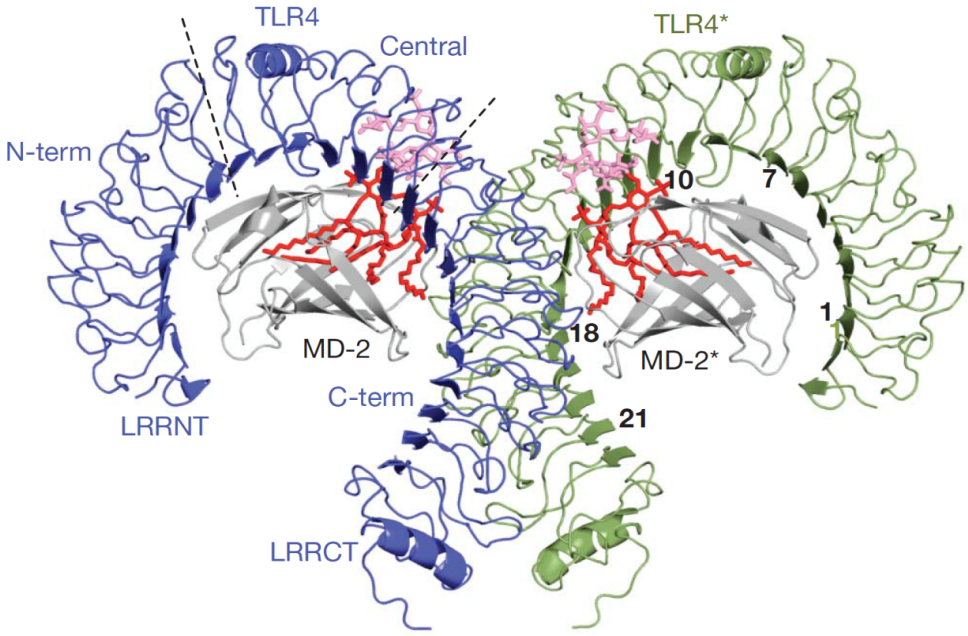


Figure 1.9. Crystal structure of TLR4-MD-2 complex bound to LPS. TLR4 and MD-2 interact to form a heterodimer that can recognize a certain pattern of lipopolysaccharides secreted by gram-negative bacteria. TLR4 is represented in blue and green (dimer), MD-2 is represented in grey and the lipopolysaccharide is divided in red (LPS) and pink (Lipid A). Adapted from Park et al., 2009 (PDB ID: 3FXI).[96]

`

1.3.6. Crystal Structure of I2CA bound HP0175

The crystal structure of HP0175 or referred to as *H. pylori* cell-binding factor 2 (*Hp*CBF2) has been recently reported by Naveen et al. (2016).[97] The protein was crystallized bound to indole 2-carboxylic acid (I2CA), a recognised inhibitor of parvulin-type PPIases. Data was acquired to 2.4 Å using the National Synchrotron Radiation Research Center of Taiwan (Hsinchu, Taiwan), and Se-Methionine derivative crystals were used to obtain initial phases. It was crystallized in the space group $P3_121$ with unit cell parameters a=b= 62.0 Å and c=366.4 Å, and Mathews coefficient of 3.16 $Å^3$, corresponding to 61% solvent content comprising a homodimer in the asymmetric unit.[97] Analytical ultracentrifugation identified the molecular weight of HP0175 to be 60.4 kDa, suggesting a dimer in solution. The structure represented 2 α-helices in the N-(54-153) and 1 α-helix in the C-(261-296) terminal, with two short antiparallel β-strands (Figure 1.10). The PPIase domain is composed of 4-antiparallel β-sheets surrounded by four α-helices. Binding of the inhibitor ligand is found near α6 and β4 - α7 and β5 - β6.[97] Previously, no structural data on the native un-liganded HP0175 was available. Therefore, comparative analysis of I2CA-HP0175 was performed with other proteins within the parvulin PPIase family, in particular comparison against *Campylobacter jejuni* PEB4 (CjPEB4),[98] *Bacillus subtilis* PrsA (*Bs*PrsA)[99] and *E. coli* SurA (*Ec*SurA).[100]

Construction of the apo and I2CA-bound HP0175 structures utilizing *ab initio* calculation through their small angle X-ray scattering (SAXS) profiles resulted in different overall shapes. Examination of the generated shapes from the top view, I2CA-HP0175 resembles a V and apo-HP0175 resembles that of a W-shape (Figure 1.11). This indicates a slight difference in conformation between the apo- and I2CA-bound proteins. The crystal structure of HP0175 has a strong resemblance to other PPIases within the Parvulin type family. The comparison obtained a root mean square deviation of 0.4 Å for 75 of 98 Cα atoms to CjPEB4, 0.6 Å for 61 of 89 Cα atoms in *Bs*PrsA and 1.8 Å for 77 of 105 Cα atoms in *Ec*SurA. The CjPEB4 also contained a domain-swapped manner that interlocked the two homodimers together, similar to the one observed by HP0175.[97]

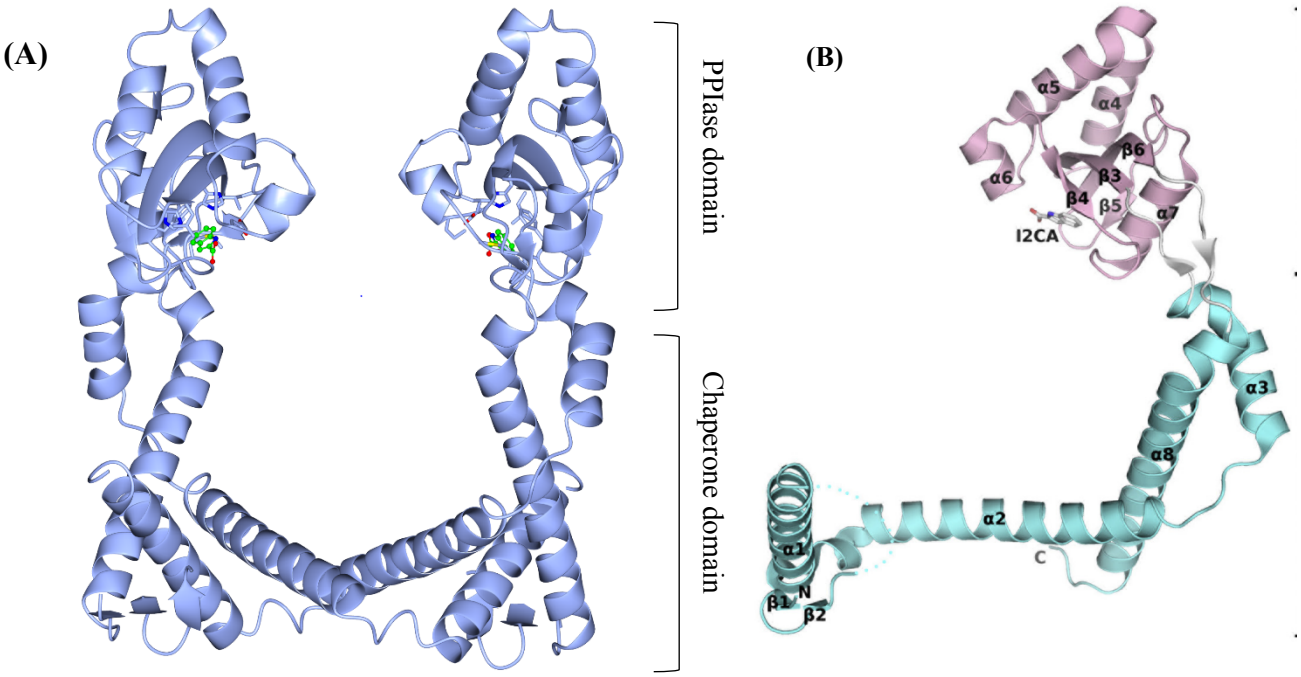


Figure 1.10. Crystal structure of HP0175 complexed with I2CA. (A) Dimeric representation of I2CA-HP0175, crystallized in space group $P3_121$, resolution range 30-2.4 Å with unit cell parameters of a=b=62.0, c=366.4 Å. **(B)** Monomeric representation of I2CA-HP0175. The binding pocket of I2CA domain is involved with residues K165, H161, D199, L211, M219, F223, F244 and H247. Labelling of helices and β sheets is depicted in (PDB entry 5EZ1) **(B).** Adapted from Naveen et al., 2016.[97]

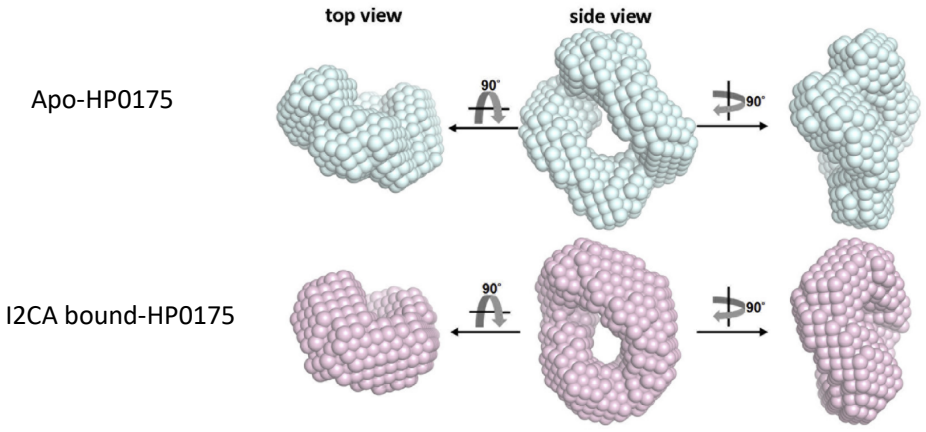


Figure 1.11. SAXS analysis of apo-HP0175 and I2CA-bound HP0175. *Ab initio* reconstruction of native HP0175 (blue) and I2CA-bound HP0175 (pink) based on merging 20 independent models. Representing a smaller crevice for the I2CA-bound HP0175 in comparison to the native. Top view of native HP0175 demonstrates a W, while the I2CA-bound reveals a V shape. Adapted from Naveen et al., 2016.[97]

`

1.4. A Brief Primer on Structural Biology by X-ray Crystallography

1.4.1. Initial Crystallization Trials

Protein crystals are composed of an ordered and repeating array of protein in a thermodynamically stable environment. Crystallization of macromolecules has often been referred to as a 'black box', with the majority of the experiment left for trial and error. While crystallization of macromolecules has been predominantly an empirical process, methods have been introduced to favor crystal growth. Controllable factors such as protein concentration, purity, pH, salts and cryoprotectants have been utilized as parameters that can be altered for crystal growth. Increasing protein concentration will increase the likelihood of protein-protein interaction and decrease protein-solution interaction, favouring crystallization. Adjustment of the pH of a solution has an impact on protein solubility, the closer the pH of a solution is to the isoelectric point of a protein solubility is decreased, less solvation and favouring crystallization. The addition of salts or cryoprotectants decreases protein-solution interaction by increasing salt-solution interactions, thereby increasing protein-protein interactions. The effect of multiple variables influencing crystallization can be represented in a 2D schematic diagram (Figure 1.12), the phase of the protein is characterized with respect to environmental influences. The phase diagram is composed of multiple regions; undersaturation, saturation, metastable and precipitation. Undersaturation represents a single liquid phase, saturation denotes a crystal will remain in equilibrium within the solution. Metastable, or a supersaturated solution would continue growth on the surface of a crystal. Precipitation is observed as solidification from a solution.[101,102]

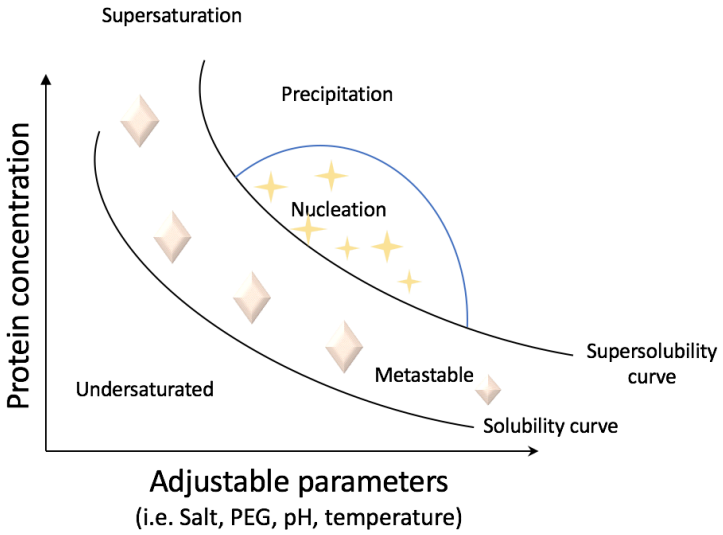


Figure 1.12. Protein solubility curve. Adjustable parameters are plotted with protein concentration and salts, PEG (cryoprotectant) and temperature. Undersaturated; low concentration of protein, would not crystallize. Metastable; within the supersaturation region, where the protein crystal is in equilibrium with the environment/solution it is in. Nucleation; growth of crystal nuclei where protein crystals can grow and move to the metastable region. Precipitation occurs with a high concentration of protein, protein crashes out of solution. Adapted from Luft et al., 2011.[101]

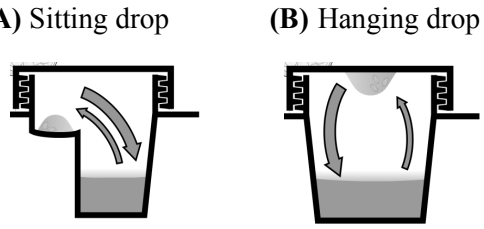


Figure 1.13. Representation of vapour diffusion for crystallization. Initial screening was performed with sitting drop vapour diffusion **(A)** sealed with tape. While optimization was performed with hanging drop **(B)** sealed with vacuum grease.[103]

1.4.2. Optimization with Microseeding

Many crystals obtained through initial screening do not diffract to high resolution and therefore require further refinement. A common method to enhance crystal quality is through utilizing poorly diffracted crystals for nucleation, a process known as seeding, and has become a powerful tool for generating higher resolution crystals for diffraction. Seeding can be done using

`

larger crystals, a process known as macroseeding, or by crushing seed crystals to make microcrystals for seeding, and accordingly known as microseeding. Seeding involves the transfer of crystalline material as a nucleation site in a new crystallization experiment with similar conditions.[104] To further elaborate, de-novo nucleation is bypassed by the addition of a crystal that acts as a nucleation site where crystal growth can proceed. Macroseeding focuses on the transfer of one crystal into a newly equilibrated drop, the chosen crystal is rinsed to remove any unwanted nucleation sites prior to transferring. In microseeding, a crystal is crushed into many smaller crystals and transferred to a newly equilibrated protein and buffer drop.[104] In addition, microseeding can be administered into a new drop by serial dilution and pipetting or by using cat whiskers or horse hair for streaking. Whiskers and hair contain rough edges allowing crystals to adhere, thereby enabling transfer to the new drop. Crystals can also be added to an optimization screen known as microseed matrix screening (MMS; Figure 1.14),[105] or they can be added in re-screening new conditions, a process known as random matrix microseed screening (RMMS).[106]

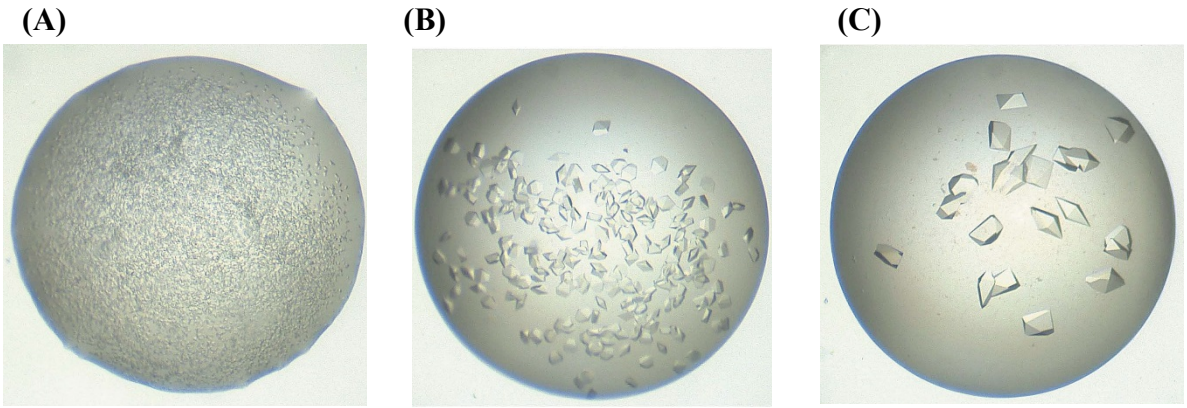


Figure 1.14. Adjustment of seed stock concentration on crystallization. Sitting drop vapour diffusion was used to represent the effect of seed stock concentration on the growth of tyrosine kinase. The seed stock was composed of crystals that were grown in 10 mM HEPES (pH 7.0), 8% (w/v) PEG800, and 7% (v/v) ethylene glycol. MMS crystallization was performed and identified a condition with 0.1 M MES pH 6.5, 25% (w/v) PEG 550 monomethyl ether (MME). In the above, drops represent seeding with **(A)** undiluted microseed stock, **(B)** a 1:100 dilution and **(C)** 1:1000 dilution of the microseed stock, which provided the best results. Adapted from D'Arcy et al. [105]

1.4.3. Diffraction, Structure Solution and Refinement

Structural characterization of proteins in three-dimensions is an important aspect to understanding their function and in the development of therapeutic drugs.[107,108] There are several

`

methods that can characterize the structure of biological systems such as Cryo-electron microscopy (Cryo-EM), transmission electron microscopy (TEM), nuclear magnetic resonance (NMR), X-ray crystallography, and mass spectrometry. Each of these methods contain their own distinct advantages and limitations, nonetheless, when utilized together can provide powerful insight and validation on the systems studied. In respect to atomic resolution, X-ray crystallography currently remains the only method that can characterize a range of molecules, from large proteins to small organic and inorganic molecules to an angstrom metric.[109]

Viewing the structure of proteins requires a brief understanding of wavelengths and optics in re-generating an image. Objects are not viewed in terms of wavelengths but rather as images attributed to the lenses in our retinas focusing the light diffracted by objects. The wavelength of visible light (electromagnetic radiation) is between 400-700 nm, these wavelengths are also used by cameras, microscopes and retinas. In order to view smaller sizes such as those for molecules and proteins, we require a wavelength small enough to match bonding of atoms calculated around 0.15 nm or 1.5 Å. Positions of hydrogen atoms for protein molecules are often generated from bond angles and conformation assumptions based on small molecules.[109] Determination of a crystal structure is not as straightforward as the extraction of an image from a lens, where the construction of wavelength interference are collected as intensities.

Monochromatic X-rays are diffracted by electrons within the crystal, the magnitude of the diffracted beam is dependent on the number of electrons in each atom. The diffraction pattern generated by a crystal is described as an array of intensities observed in a regular net or reciprocal lattice. Sir W.H. Bragg and Sir W.L. Bragg explained this phenomenon in what is known as the Bragg Law (Equation 1):

Equation 1: $$\mathbf{2\,d\sin\theta = n\lambda}$$

where d represents the spacing between atomic layers within the crystal planes, lambda (λ) represents wavelength of the incident beam, theta (θ) is the angle of reflection, and n is an integer.[109]

The reflections observed in a diffraction pattern represent a summation of waves that travel from the electron positions in the lattice. Since waves can add constructively and destructively,

`

amplitude can be calculated from the intensity of the reflection, but the phase (∅) is lost; this is largely referred to as the crystallographic phase problem. The following is required for the generation of the electron density map exemplified by the Fourier transform (Equation 2). Over the years several methods have been developed for solving the phase problem and generating electron density, such as direct methods, single wavelength anomalous dispersion (SAD) and molecular replacement.

Equation 2: $$\boldsymbol{\rho\,(xyz) = \left(\frac{1}{V}\right)|\sum_{hkl} F(hkl)|\{\exp(i\emptyset)\exp(-2\,\pi\,i\,(hx+ky+lz)\}}$$

Electron density Amplitude (Structure factor)

Molecular replacement utilizes the coordinates of a previously determined structure to generate the phases for the unknown protein crystal, assuming amino acids applied in a similar sequence will contain similar folds. Relating the known protein's phase for the unknown crystal requires search for rotation and translation. Rotation function (RF) is defined as the spatial orientation of the known and unknown structure with respect to each other for determining the spatial orientation of the unknown crystal. The second is translational function (TF) for determining the exact position in the unit cell for the superposition of structure in the new unit cell with the correct orientation. Packing analysis is performed following RF and TF to determine any clashes between atoms of symmetry related molecules. In the last stage, rigid body refinement is performed on the final model and the phase is calculated. A problem that arises from this method is structural bias imposed from the previously determined protein structure.[110] Selection of the model used for molecular replacement must be similar in structure or sequence to the unknown crystal, with an RMSD variation less than 1.5 Å, or based on sequence identity greater than 20-30%.[110]

1.5. Thesis Research Objective and Overview

Bacterial infections are common in our society and continue to evolve. There is much we do not know from their behaviour, function and adaptation strategies. Further investigation on different strains and various machineries encompassing their behaviour would pose a beneficial strategy to generate a novel mode of disruption. Infections obtained from strains such as

`

Pseudomonas aeruginosa are commonly found in individuals with a comprised immune system such as burn victims, cystic fibrosis, acute leukaemia and HIV patients.[111,112] An important characteristic of this bacteria is through the adhesion by pili which utilize complex machinery as seen with T4P that bind to biotic and abiotic surfaces.[38] In addition, the bacterium *H. pylori*, associated with peptic ulcers[48] and adenocarcinomas,[50] is associated with multiple proteins that contribute to the following through activation of the adaptive immune system. HP0175 is involved with activating cell death, causing inflammation at a more significant rate, making the study of the protein important.

The overall objective of the research outlined within this thesis is to examine proteins secreted by pathogenic bacteria through X-ray crystallography, specifically the KB7 and P1 pilins secreted by *Pseudomonas aeruginosa* and HP0175 of *Helicobacter pylori*. Characterization of a protein's structure will provide information on how it behaves and how it may differ from other proteins within the family. As in *P. aeruginosa,* the overall structure of the pilin secreted remains similar with variability of the D-region. Crystallographic determination of the pilins would highlight how bacteria evades eradication through sequence alterations, which could lead to the synthesis of cross-reactive vaccines. Examination of the structure of apo-HP0175 from *H. pylori* will provide insight into how the protein functions and its conformational changes upon substrate binding. Previously, parvulin peptidyl prolyl isomerases were only observed structurally in one configuration; either in a bounded or un-bounded state. Here we show the changes of HP0175 undergone upon targeted interaction with indole-2-carboxylic acid and infer on the changes when bound to a peptide or a protein.

`

Chapter 2: Experimental Methods

2.1. Materials and Methods

Chemicals and antibiotics used in the following study were purchased from Sigma-Aldrich, Alfa Aesar, VWR, Hampton Research, Fischer Scientific, Bio Basic, J.T. Baker, Anachemia, Amresco and EMD. Restriction enzymes, DNA ladders and DNA staining were purchased from New England Biolabs (NEB). Plasmid vectors pGEX-4T2, pET28a:HP0175 and *E. coli* BL21 (DE3) and DH5α were obtained from cryogenic lab stocks. Final concentrations of antibiotics utilized were 50 μg/mL Kanamycin (Km) and 100 μg/mL Ampicillin (Amp). A list of all primers for the cloning of pilin proteins are represented in Table 2.1.

Liquid chromatography was performed on an ÄKTA Purifier 10 (GE Healthcare); all columns were purchased from GE Healthcare, Phenomenex and Wyatt. Resins for the purification of proteins were purchased from Fisher scientific. Polymerase chain reactions were performed using a TECHNE PCR thermocycler. Displacement of proteins from the cytoplasm was performed via sonication on a Sonic Dismembrator 500 (Fisher Scientific). Electrospray ionization mass spectrometry was performed on a Waters Synapt G1 instrument. Concentrating proteins to the appropriate volumes were completed using 10,000 and 30,000 Da molecular weight cut off (MWCO) centrifugal concentrators (Fisher Scientific, GE Healthcare). Observation of crystallization experiments were conducted using a Nikon SMZ1500-Fibre Lite MI 150 stereoscopic microscope. Diffraction studies were carried utilizing Rigaku 007HF MicroMax Rotating Anode with Saturn 944+ CCD detector. Data of protein crystals were collected with beamline 08ID-1 and 08B1-1 of the Canadian Macromolecular Crystallography Facility (CMCF) at the Canadian Light Source (CLS) in Saskatoon, Saskatchewan.

Table 2.1. Primer design for cloning ΔKB7 and ΔP1

Construct	Restriction Enzyme	Tm (°C)	Primer (5'-3')
KB7_Forward	EcoRI	57.5	ATGAGAATTCTATCTCGCTCTCAGGTC
KB7_Reverse	XhoI	59.6	CTTACTCGAGTTATCAGTCCGACGTTGTAAAC
P1_Forward	EcoRI	63.2	TATTCGAATTCTCGCCCGTACCCAGGT G
P1_Reverse	XhoI	65.5	TTGATTCTCGAGTCATCA CATTAGCCGGAGCGTAGT TGG

*Underlined are the restriction enzyme cut sites *EcoRI* and *XhoI*. *Δ refers to the truncated KB7and P1 pilin which consists the removal of amino acids 1-28.*

`

2.2. Generation of Competent DH5α and BL21 (DE3) Cells

Overnight cultures of DH5α cells were prepared in 20 mL LB. From the overnight culture, 1 mL was transferred to a total of 100 mL fresh LB and inoculated at 37 °C/200 rpm until mid-log phase growth (OD_{600}= 0.4-0.7). Pipette tips, falcon tubes, microfuge tubes, ddH_2O and 100 mM $CaCl_2$ were pre-chilled at 4 °C. The cells were placed on ice for 10 minutes followed by centrifugation (2 x 50 mL), 10 minutes at 2,700 x g and 4 °C. Cell cultures were re-suspended with 1.6 mL ice cold 100 mM $CaCl_2$ and incubated on ice for 30 minutes. Cells were centrifuged, re-suspended in 100 mM $CaCl_2$ and incubated on ice for 20 minutes. Ice-cold glycerol (0.5 mL, 80%) was added to 3.2 mL of DH5α cells. Competent cells were stored for a period of 6 months in -80 °C. The same procedure was followed for the generation of competent BL21 (DE3) cells.

2.3. Cloning *P. aeruginosa* pili ΔKB7 and ΔP1

2.3.1. Plasmid Extraction of ΔKB7 and ΔP1 from pMAL p2 [tm]

Overnight cultures of pMAL:ΔKB7, pMAL:ΔP1 and empty pGEX-4T2 in DH5α were prepared with 20 mL LB and 20 µL Ampicillin (100 µg/mL) from laboratory stocks. Plasmid extraction was performed with GeneJet plasmid mini-prep kit following manufacturers protocol from 6-8 mL of overnight cultures. Concentrations of each plasmid were determined using a NanoDrop 2000 Spectrophotometer, and verified on a 1.2% agarose gel (1.2 g agarose with 1X TAE buffer: 40 mM Tris:HCl pH 8.5, 20 mM acetic acid and 1 mM EDTA were incubated at 65 °C to dissolve and added with 0.2 mg/mL ethidium bromide).

2.3.2. PCR Amplification of ΔKB7 and ΔP1

Primers utilized in this study followed design guidelines presented by European Molecular Biology Laboratory (EMBL).[113] The following guidelines included appropriate GC content (40-60%), melting temperature for the forward and reverse primer being within 3-5°C (Table 2.2), the temperature to generate hairpins being lower than the annealing temperature, and the free energy of homo- and hetero- dimerization were below -10 kJ and 30 °C.[114] PCR amplification parameters

`

were: initial denaturation for 5 minutes at 98 °C, annealing temperatures were 5 °C below the T_m of the primers, 72 °C extension step, with a final extension at 72 °C for 7 minutes, and held at 4 °C. PCR amplicons of ΔKB7 and ΔP1 were PCR purified for restriction enzyme digestion.

2.3.3. Restriction Enzyme Double-Digestion

Restriction enzyme digestion on the PCR amplicon of ΔP1 and ΔKB7 were performed following parameters of NEB kit outlined in Table 2.2. The target plasmid pGEX-4T2 was also double digested to create complementary sticky ends with XhoI and EcoRI restriction enzyme sites. Plasmid vectors and PCR amplicons were gel extracted following double digestion to ensure removal of the enzymes.

Table 2.2. Restriction enzyme double digestion

Components	Volume (μL)	Final Concentration
ddH_2O	To 50	
PCR amplicon		1 μg
10X NEB Buffer 2.1	5	1 X
XhoI	0.5	20, 000 units
EcoRI-HF	0.5	20, 000 units

Double digestion was carried at 37 °C for 1-2-hour incubation.

2.3.4. Dephosphorylation of pGEX-4T2

Calf Intestinal phosphatase (CIP) was used for de-phosphorylation of the 5' and 3' ends of the pGEX-4T2 plasmid. This ensured that plasmid used would not self-ligate, and therefore increase ligation efficiency with the PCR amplicons. CIP was incubated with pGEX-4T2 at 37 °C for 2 hours (2000 ng plasmid, 4 μL smart buffer and 1μL CIP).

2.3.5. Ligation and Heat-Shock Transformation into DH5α Cells

The mass of amplicon to plasmid were performed utilizing 1:3, 1:5 and 1:7 as well as self-ligation trials. Calculations of masses (PCR amplicon to plasmid) were designed using NEB's ligation calculator. The following was prepared with 2 μL ligase buffer, vector DNA, insert DNA,

`

ddH2O and 1 μL DNA ligase for a reaction total of 20 μL. Ligations were incubated at room temperature for 30 minutes followed by heat inactivation at 65 °C for 10 minutes. The samples were chilled on ice for 5 minutes prior to transforming 5 μL of the reaction to 20 μL DH5α competent cells. The transformed cells were incubated on ice for 20 minutes before incubation at 42 °C for 90 seconds, placed on ice again for 5 minutes and transferred into 300 μL LB for 1 hour at 37 °C and inoculated on agar plates containing Amp (100 μg/ml).

2.3.6. Gene Extraction and Sequencing

In order to verify the correct placement of the gene, several colonies were inoculated overnight at 37 °C followed by gene extraction the following day. Restriction enzyme double digest with XhoI and EcoRI-HF were performed as per section 2.3.3. The plasmids containing the correct insert size as per 1.2% agarose gel after double digestion were sent for sequencing at The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children (Toronto, ON).

2.3.7. Heat-Shock Transformation into BL21 (DE3) Cells

Plasmids containing the correct genes after sequencing were transformed into BL21 (DE3) competent cells as per section 2.3.5.

2.4. Expression and Purification of GST-ΔKB7 and HP0175

2.4.1. Expression Trials for GST-ΔKB7

E. coli BL21 (DE3) cells containing pGEX-4T2:ΔKB7 were inoculated overnight at 37 °C in 20 mL LB, 1 mM dextrose and 50 μL Amp (100 μg/μL). Cell cultures were transferred to fresh 50 mL LB and Terrific Broth (TB). The cells were expressed to mid log-phase (OD_{600} 0.4-0.7) at 37 °C/200 rpm. Only half of the cultures were induced while half were left to grow without induction, with a total expression of 4 hours. Cell pellets were harvested by centrifugation: 6000 x g, 20 min, 4 °C, and stored at -20 °C.

`

2.4.2. Large Scale Expression

E. coli BL21 (DE3) cells harbouring pGEX:ΔKB7 were grown overnight in LB at 37 °C. Aliquots of these cells were allowed to reach mid-log phase (OD_{600}= 0.4-0.7) in 500 mL fresh LB containing Amp 100 μg/μL Amp and 1 mM dextrose. Protein expression was induced with the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM final for a total of 4 hours at 32 °C and 200 rpm.

Glycerol stocks of HP0175 were originally a gift from our collaborator, Professor M. Kundu of the Bose Institute in Calcutta, India. HP0175 was N-terminally hexa-histidine tagged and expressed in *E. coli* BL21 (DE3) cells harboring the pET28a expression vector. From the overnight expression, cell cultures were grown to mid-log phase in terrific broth media (TB: 12g Tryptone, 24 g yeast extract, 4 mL 100% glycerol in 900 mL ddH_2O, 100 mL salt containing 0.17 M KH_2PO_4 and 0.72 M K_2HPO_4). Protein expression was also achieved by adding 1 mM IPTG and incubated for 4 hours in 32 °C/200 rpm. Cell pellets were harvested by centrifugation 6,000 x g, 30 min, 4 °C and stored for up to 6 months in -20 °C.

2.5. Purification of GST- ΔKB7 and HP0175

2.5.1. GST Affinity Purification of ΔKB7

Cell pellets containing GST-ΔKB7 were re-suspended in lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA), incubated with lysozyme (1 mg/ml final concentration) for 1 hour with shaking at 4 °C. The sample was incubated with lysozyme to increase cell lysis efficiency by hydrolyzing the peptidoglycan layer, and therefore allowing ease of disruption through sonication. Release of GST-ΔKB7 was performed via sonication on ice for a total of 5 minutes with pulses (40 % amplitude, 15 s pulse on, 30 s pulse off). Cells were centrifuged at 45,000 x g, 30 mins at 4 °C. Cell debris was removed by syringe filtration (0.8 μm) prior to loading onto a glutathione column. The use of GST as a tag is beneficial for solubilizing proteins that would otherwise become inclusion bodies. It is also easily removed from the sample through enzymatic digestion and re-purification on a glutathione column, size exclusion or ion exchange chromatography.[115] The elution buffer was composed of 20 mM Tris pH 8.0, 10 mM reduced glutathione and 150 mM

`

NaCl. Fractions containing GST-ΔKB7 were pooled, concentrated (30,000 MWCO; Millipore) and buffer exchanged for 36 hours via dialysis into 10 mM Tris:HCl pH 7.4.

2.5.2. Immobilized Metal Affinity Purification of HP0175

Cell pellets of pET28a:HP0175 were re-suspended in 1:10 (w/v) loading buffer (20 mM HEPES, 500 mM NaCl pH 7.4). Protein was released via sonication over a total of 5 minutes at 30% amplitude with brief pulses (15 s on, 30 s off), following which samples were centrifuged at 45,000 x g, 20 minutes at 4 °C. The supernatant was syringe filtered (0.8 μm) for removal of cell debris and purified with immobilized metal affinity chromatography (IMAC). Similar to the affinity purification used for GST-ΔKB7, the polyhistidine peptide is used as a tag and can bind to transition metals such as Ni^{2+}, Co^{2+}, Cu^{2+} and Zn^{2+}. Immobilized Nickel (II) to nitrotriacetic acid (NTA) resin was used for the purification of HP0175.[116] Fractions containing HP0175 (36 kDa) were pooled and buffer exchanged overnight via dialysis into 10 mM HEPES pH 7.4 before concentrating with 10 K MWCO concentrators. Concentration and purity were verified with UV-VIS and 12.5% sodium dodecyl sulfate protein agarose gel electrophoresis (SDS-PAGE).

2.6. Crystallization of Macromolecules

Initial crystallization trials for GST-ΔKB7 were prepared with sitting drop (Figure 1.13A) vapour diffusion (96-well Art Robins Instruments plates), buffers used for screening included MCSGII, MCSGIV, TOP96 and PurePEGs, purchased from Anatrace. Sitting drops were prepared with 1:1 protein-to-buffer ratio (0.6 μL) with a reservoir of 90 μL. The same plates were also arranged utilizing a reservoir comprising only 1.5 M NaCl to increase the rate of evaporation through high salt concentration and sealed with tape. Hanging drop vapour diffusion (Figure 1.13B) was prepared for optimization of certain conditions, plastic slides containing protein and buffer drop (0.8 μL; 1:1 ratio) were sealed with grease above 1 mL reservoir in a 24-well plate. GST-ΔKB7 was screened within the range of 5-15 mg/ml.

Initial crystallization trials for HP0175 were performed in sitting drop vapour diffusion 96-well plates (ARI) with Microlytic MCSG II crystallization screen (Anatrace). The protein to buffer

`

ratio was 1:1 with protein density ranging from 5-20 mg/ml, 0.75 μL protein: 0.75 μL buffer. The following screen condition yielded a protein crystal in 0.1 M Tris:HCl pH 8.5 and 2.5 M ammonium sulfate; these conditions were optimized to generate spherulites in a 24-well plate hanging drop vapour diffusion (0.8 μL protein: 0.8 μL buffer content). The spherulites were crushed and used as a microseed stock following Hampton Microlytic seed bead kit; dilutions of the crushed crystal were made into 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}. The seed stock was then added to the hanging drop optimization screen, after 48 hours to allow for equilibration. This optimization screen yielded a 7.8 Å diffracting crystal from reservoirs containing 0.1 M Tris:HCl pH 8.3 and 47.4% ammonium acetate. The same concentrations of protein and microseed stock that provided the diffracting crystal were re-screened with fresh MCSG I-III buffers (Anatrace). The re-screened HP0175 (16.5 mg/ml) yielded a crystal that diffracted to 2.09 Å resolution from a reservoir containing 0.1 M HEPES:NaOH (pH 7.0) and 30% (v/v) Jeffamine (O,O'-Bis(2-aminopropyl)polypropylene glycol-block-polypropylene glycol). All crystallizations were conducted at room temperature (21 ºC).

2.7. Data Collection & Processing of Apo-HP0175

Initial diffraction screening of crystals was performed on a Rigaku MicroMax-007 HF rotating anode X-ray generator with Saturn 944+ CCD detector. Data collected here consisted of 100 images at a crystal-to-detector distance of 60.0 mm, 20 s exposure time with angles 285-346°. Once diffraction confirmed the presence of a protein crystal, the samples were sent to the Canadian Macromolecular Crystallography Facility (CMCF) located at the Canadian Light Source (CLS) in Saskatoon, Canada. At the CMCF, the 08ID-1 beamline was used with a Pilatus 6M detector, 200 mm distance, exposure time set to 0.3 s, λ = 0.979 Å, covering an angular range of 0-180° to a total of 900 frames (0.2° per image). Images collected were indexed, refined and integrated operating XDS.[117,118] Space group analysis and data reduction was performed by Aimless/Pointless[119,120] through the CCP4[121] suite of programs.

`

2.7.1. Structure Solution & Refinement

The structure of apo-HP0175 was solved by molecular replacement using PHASER[122] through PHENIX.[123] Inhibited HP0175 bound to indole 2-carbocylic acid (I2CA) was used as the search model for molecular replacement. The search model was fragmented into three regions; PPIase domain (residues 151-257) and the N-terminal helices (residues 54-98 and 105-130) represented in Figure 2.1. Autobuild (rebuild in space)[124] was performed in PHENIX[123] for positioning of the model based on electron density. The structure was further manually re-built in COOT[125] using 2FOFCWT and PH2FOFCWT maps and refined with REFMAC in the CCP4i[121] program suite interface. The final refined structure contains amino acid residues 43-60 and residues 69-299, with only one break in the structure for the missing residues 61-68; no electron density was observed prior to residue 43 and therefore these amino acid residues are not included in the final refined structure. The final refined structure and structure factors were deposited in the Protein Data Bank with accession code 6BHF.

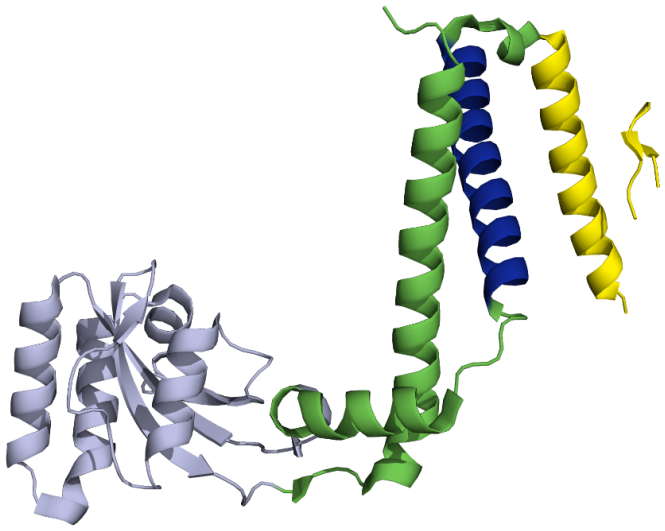


Figure 2.1. Molecular replacement model. The PPIase domain is represented in mauve (151-257 aa), helix (105-130aa) is colored blue, and yellow represents helix with residues 54-98aa. Figure adapted from PDB ID entry 5EZ1.[97] All figures of protein structures are generated in Pymol.[126]

`

2.8. Electrospray Ionization Mass Spectrometry

Mass spectrometry has become a sensitive bioanalytical method for studying biological samples.[127] Mass spectrometry separates analytes in the gas phase through a mass to charge ratio. Solution phase can be directly coupled to the electrospray ionization (ESI) source, where ions are produced upon passing through a capillary with electric potential between 2-6 kV. This high electric potential generated at the tip of the capillary produces a dispersion in the sample, creating a Taylor cone.[128] Due to the electrically conductive liquid, shape altercation is through surface tension and the electric field becomes prominent upon increasing capillary voltage.[128] Once the solvent is evaporated and sprayed, the ions travel to the mass analyzer that separates based on mass-to-charge (m/z) ratio. Both positive ion mode and negative ion mode can be used to asses proteins;[128] the positive ion mode was used for analysis of apo-HP0175. Instrument settings used for the analysis of HP0175 with ESI-MS includes; capillary voltage 2.8 kV, sampling cone 80.0, extraction cone 2.0, source temperature 120 ˚C, desolvation temperature 150 ˚C, cone gas flow 30 L/Hr, nanoflow 1.0 Bar, desolvation gas flow 500.0 L/Hr, trap collision energy 15.0, transfer collision energy 10.0, trap gas flow 2.0 ml/min and source gas flow 15.0 ml/min.

`

Chapter 3: Results and Discussion

3.1. Cloning ΔKB7 and ΔP1 into pGEX-4T2

The pilins explored in this study were truncated (Δ1-28) KB7 and P1 extracted from the pMAL-p2 plasmid vector (lab stocks). Expression in pMAL-p2 contained the periplasmic expression of maltose binding protein (MBP) tagged to ΔKB7 and ΔP1. Column cleavage of MBP is not recommended due to the large amount of Factor Xa required, as well as the duration associated, estimating 5% cleavage for 24-48 hours.[129] Secondary MBP purification with an amylose column is not amenable because MBP after elution contains maltose, not removable with dialysis and therefore cannot bind to the amylose column again.[129] Ion exchange is not ideal associated with the close isoelectric points of MBP and KB7 (5.2 and 4.5, respectively), while high resolution size exclusion chromatography would induce oligomerization of the pilin (GF Audette; personal communication). Therefore, a different expression system is required for ΔKB7 and ΔP1. The truncated pilins of KB7 and P1 were PCR amplified from pMAL-p2 plasmid vector and were successfully cloned into pGEX-4T2 vector (Figure 3.1 and 3.2). Amino acid sequences of ΔKB7 and ΔP1 are found in Appendix A.

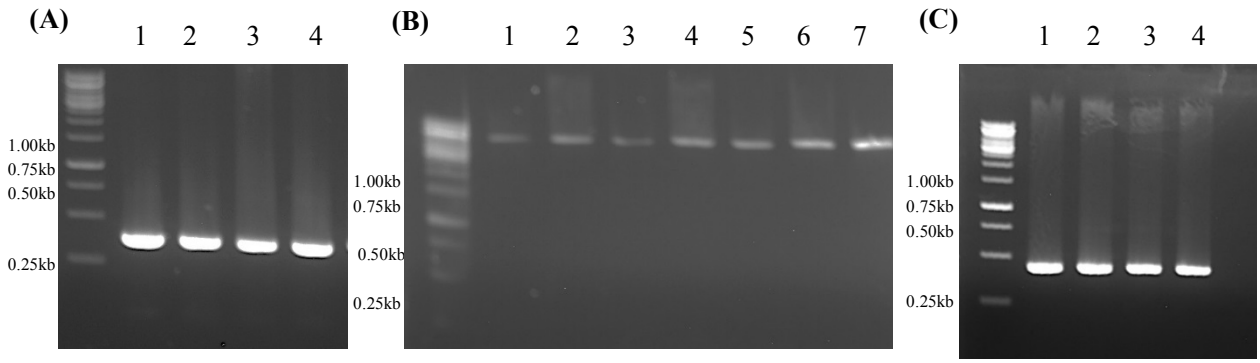
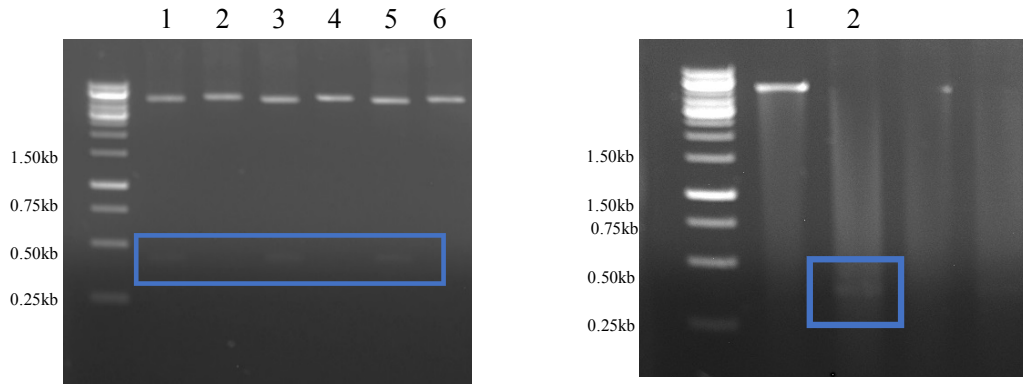


Figure 3.1. PCR products of ΔP1 and ΔKB7. The products were introduced onto a 2% agarose gel and viewed with UV light. **(A)** ΔP1 was PCR amplified with primers P1_forward and P1_reverse lanes 1-4. **(B)** Representation of pGEX-4T2 plasmid vector double digested with XhoI and EcoRI-HF lanes 1-7. **(C)** PCR amplicon of ΔKB7 achieved with primers KB7_Forward and KB7_Reverse lanes 1-4.

`

3.2. GST-ΔKB7 and -ΔP1 Sequence Analysis

Prior to sequencing the cloned constructs of ΔKB7 and ΔP1 in the pGEX plasmid, they were gene extracted and double digested with EcoRI and XhoI. Digestion of the plasmid with the required restriction enzymes verified presence of ΔKB7 and ΔP1 genes within the 500-base pair region. Figure 3.2 represents successful placement of the ΔKB7 and ΔP1 genes and was subsequently sent for sequencing at The Centre for Applied Genomics at SickKids, Toronto.



(A) **(B)**

Figure 3.2. Verification of cloning through restriction enzyme double digest. (A) 1,3 and 5: pGEX plasmid double digested with XhoI and EcoRI-HF contained ΔKB7 gene. 2; pGEX EcoRI-HF digestion. 4 and 6: pGEX XhoI digestion. **(B)** 1: pGEX plasmid with ΔP1gene. 2: pGEX plasmid with ΔP1 gene, double digested with XhoI and EcoRI-HF.

3.3. GST-ΔKB7 and ΔP1 Expression Trials

Expression trials were performed to determine the optimal temperature and broth at which GST-ΔKB7 growth was favoured. It was observed that GST tagged proteins contained a leaky expression, which should be further optimized. From Figure 3.3.1; the best conditions for expression are that of 3.5 hours/37 °C in LB with induction, as well as the expression in TB/37 °C with induction. However, since TB contains more components, the most economical choice remains in expressing with LB media. Expression trials for GST-ΔP1 in LB are represented in Figure 3.3.2 and demonstrates a wider range of temperature to minimize GST expression.

`

Expression trails against TB media are not represented here. While optimal expression of GST-ΔP1 is observed at 37°C/2 hours with IPTG induction.

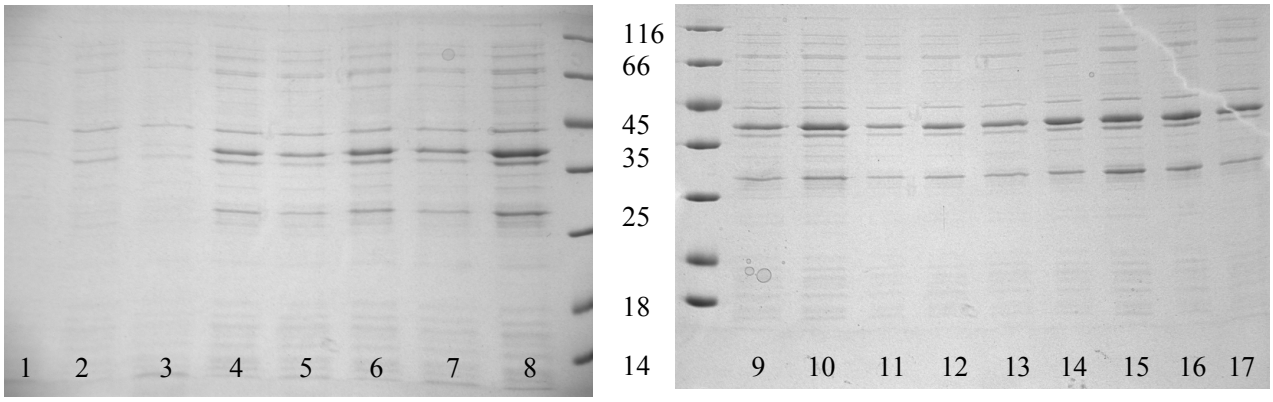


Figure 3.3.1. Expression trials of ΔKB7 expressed with GST tag in *E. coli* BL21 (DE3). a) 1) 32°C, 0 hrs in LB. 2) 32°C, 0 hr in TB. 3) 37°C, 0 hr in LB. 4) 37°C, 0 hr, in TB. 5) 32°C, 1 hr, uninduced in LB. 6) 32°C, 1 hr in TB, induced. 7) 37°C, 1 hr, uninduced in LB. 8) 37 °C, 1 hr, induced in LB. b) 9) 37°C, 2 hr, induced in TB. 10) 37°C, 2 hr, induced in LB. 11) 32°C, 2 hr, induced in TB. 12) 32°C, 3 hr, induced in TB. 13) 37°C, 3 hr, induced in TB. 14) 37°C, 3 hr, induced in LB. 15) 32°C, 3.5 hr, induced in LB. 16) 37°C, 3.5 hr, induced in LB. 17) 37°C, 3.5 hr, induced in TB.

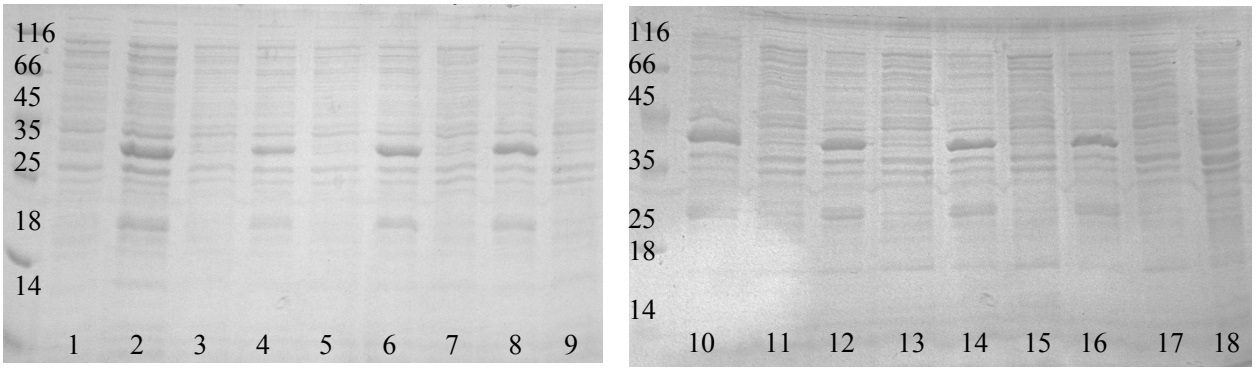


Figure 3.3.2. **Expression trials of ΔP1 expressed with GST tag in *E. coli* BL21 (DE3)**. Different temperatures were used in LB media. 1) 0 hr. 2) 32°C, induced at 1 hr. 3) 32°C, un-induced, 1 hr. 4) 37°C, induced, 1 hr. 5) 37°C, un-induced, 1 hr. 6) 32°C, induced, 2 hr. 7) 32°C, un-induced, 2 hr. 8) 37°C, induced, 2 hr. 9) 37°C, un-induced, 2 hr. 10) 32°C, induced, 3 hr. 11) 32°C, un-induced, 3 hr. 12) 37°C, induced, 3 hr. 13) 37°C, un-induced, 3 hr. 14) 32°C, induced, 4 hr. 15) 32°C, un-induced, 4 hr. 16) 37°C, induced, 4 hr. 17) 37°C, un-induced, 4 hr. 18)16°C, induced, 16 hr.

`

3.4. Purification of GST-ΔKB7 and GST-ΔP1

Isolation of GST-ΔKB7 was achieved using batch purification utilizing a glutathione resin, which allowed an additional period for binding. This was the optimal method in comparison to the ÄKTA purifier, as it resulted in half of the protein eluting in the flow through. The supernatant (Figure 3.4A) was syringe filtered and incubated on column for a total of 1 hour before elution (loading buffer: 50 mM Tris pH 8.0, 150 mM NaCl and 1mM PMSF). Fractions 1-10 (Figure 3.4B) were pooled and concentrated utilizing a 30 MWCO Millipore concentrator (Figure 3.4C). Large amounts of the GST tag passed through the 30 MWCO membrane, thereby providing for a cleaner sample in comparison to the MBP tagged protein. Purification of GST-ΔP1 was not successful in producing a clean and concentrated protein of interest. Purification of ΔKB7 with a GST tag (25 kDa) in comparison to the MBP (45 kDa) indicated that a larger tag is required to solubilize the protein of interest. This was observed with the amount of protein lost in the pellet after sonication (Figure 3.4A).

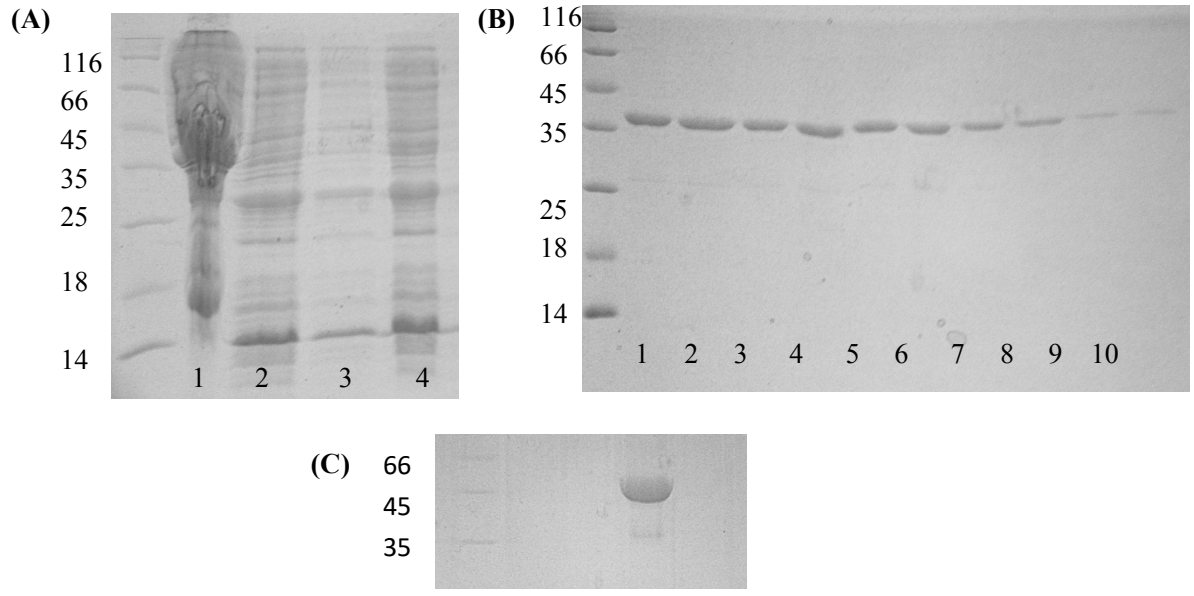
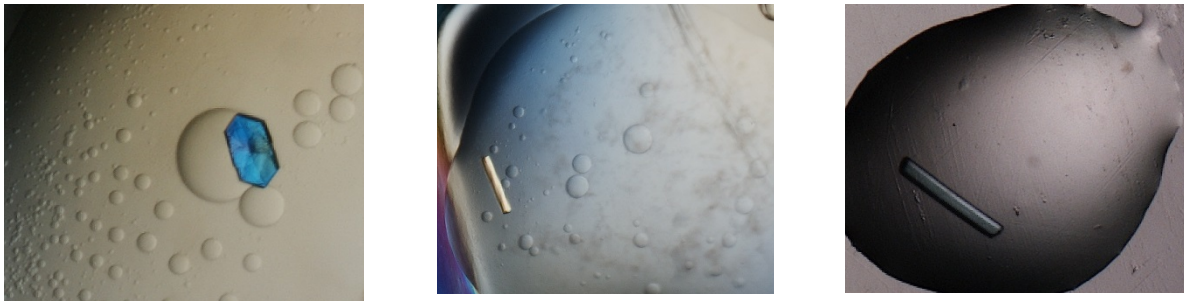


Figure 3.4. Purification of ΔKB7. Fractions demonstrated on a 12.5% SDS PAGE. **(A)** Pellet (1), supernatant (2), Flow through from purification (3), cell pellet after sonication (4). **(B)** Fractions 1-10 represent eluted GST-ΔKB7 in 10 mM Tris and 10 mM glutathione (pH 8.2). **(C)** Concentrated GST-ΔKB7 prior to crystallization.

`

3.5. Crystallization of GST-ΔKB7

Crystallization of GST-ΔKB7 was arranged using sitting drop vapour diffusion with 5-15 mg/ml protein densities. The crystallization suites used for initial screening were JCSG + core (Qiagen), MCSG II, IV and Top96 (Anatrace). Several successful conditions were achieved (Figure 3.5), however, only one crystal was diffracted (Figure 3.5A and 3.6). The conditions achieved did not contain cryoprotection, and therefore further optimization is required. With the tools available, 0.2% bromothymol blue dye was added to the droplet containing the crystal to observe colour change. Packing of salt crystals are very constricted and does not allow dye to penetrate. Since protein crystals contain roughly 20-50% solvent, they allow dye to diffuse and can be used as a technique to distinguish protein crystal from non-protein.



(A) **(B)** **(C)**

Figure 3.5. Crystallization conditions achieved for GST-ΔKB7. Buffer conditions; **(A)** 2.0 M Ammonium sulfate, 0.2 M Sodium chloride, 0.1 M Sodium cacodylate pH 6.5. **(B)** 3.2 M Ammonium sulfate, 0.1 M Citric acid pH 4.0, final pH of 5.0. Crystal represented in (B) was stained in **(C)** with 0.02% bromothymol blue.

3.5.1. Crystallization Screening

One of the crystals achieved from Figure 3.5 (A) was screened for diffraction. The following crystal was cryoprotected through soaking with increasing concentrations of cryoprotectant (glycerol: 10, 15 and 35%). The following did not provide for a protein diffraction pattern as only ice rings and small molecule diffraction was observed (Figure 3.5.1).

`

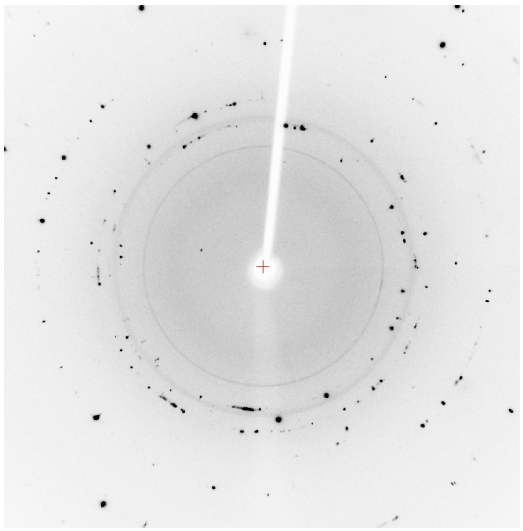


Figure 3.5.1. Diffraction analysis of GST-ΔKB7. The crystal was soaked with increasing concentrations of glycerol (5, 10 and 35%). Ice rings and small molecule reflections were observed instead of protein diffraction. The detector distance was set to 60 mm with exposure time to 10 s, and 0˚ image frame.

3.6. Expression and Purification of HP0175

HP0175 was provided in a hexahistidine-tagged construct in the pET28a plasmid by our collaborator, M. Kundu of the Bose Institute (Calcutta, India), and was transformed into *E. coli* (BL21) DE3 by a previous lab member. Glycerol stocks of the plasmid in BL21 (DE3) cells were stored in 50% glycerol at -80 °C. Optimal expression of HP0175 was determined to be in terrific broth for a total of 4 hours with IPTG (1 mM) induction at 32 °C and 200 rpm. Following expression, cells were harvested by centrifugation (6,000 x g, 4 °C, 20 min), cell pellets were stored at -20 °C for later use. Release of protein (36393.73 Da) from the cell was performed via sonication in loading buffer (20 mM HEPES:NaOH pH 7.4 and 500 mM NaCl) and purified on ÄKTA Purifier 10S (GE Healthcare) using Nickel Immobilized Affinity Chromatography (IMAC). HP0175 was eluted with 35% (7 mM HEPES:NaOH pH 7.4, 175 mM NaCl and 35 mM Imidazole) elution of linear gradient (Figure 3.6). Fractions containing HP0175 were pooled and buffer exchanged into 10 mM HEPES:NaOH (pH 7.4) with dialysis overnight. Since purification of HP0175 produced a large amount of protein, excess was stored at -80 °C with 50% glycerol. The protein was then free to use once dialyzed (2 days) into 10 mM HEPES:NaOH pH 7.4.

`

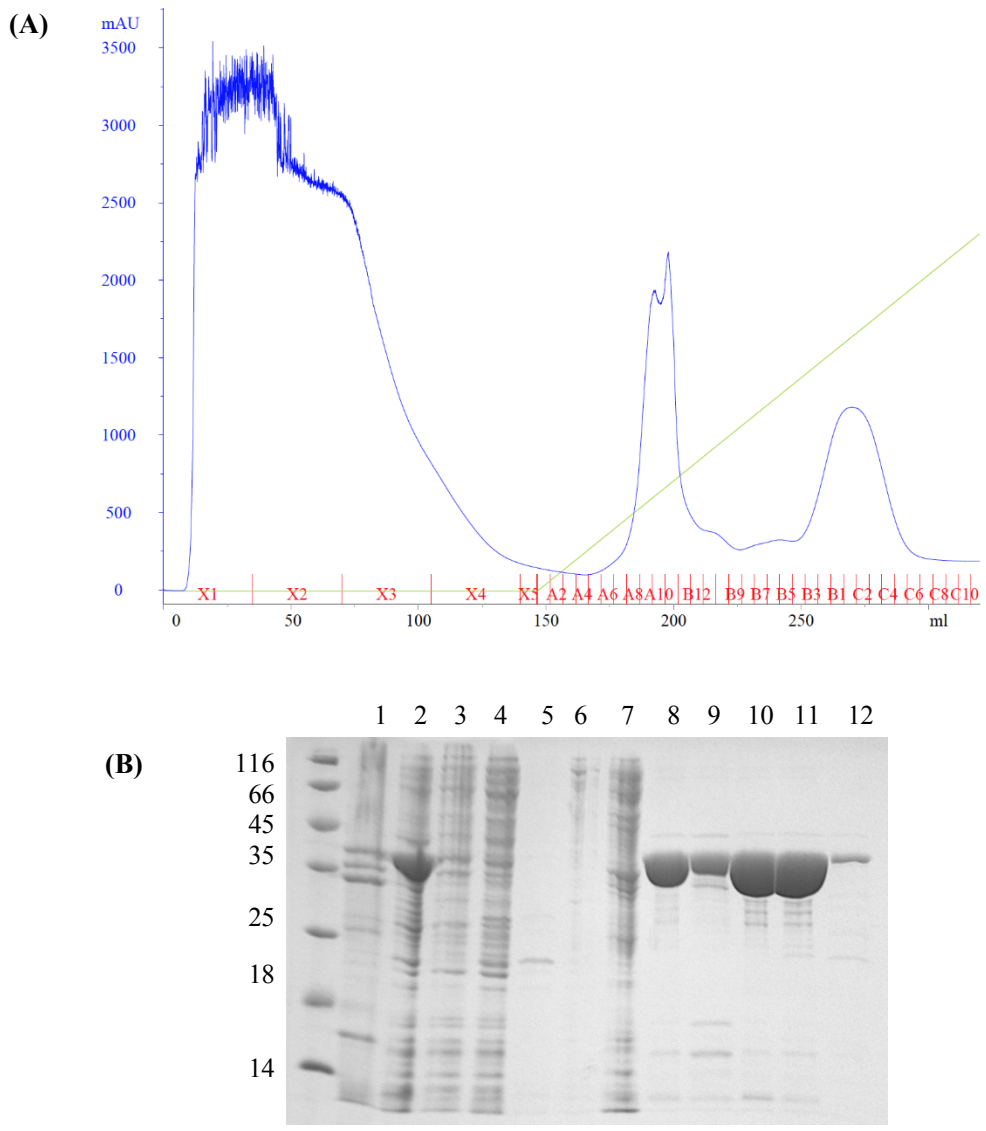


Figure 3.6. IMAC purification of HP0175. (A) IMAC chromatographic profile for the purification of HP0175. Blue line represents the absorbance of proteins at 280 nm, while the linear green line represents the elution profile of the buffer. The first peak shows different proteins expressed by the bacteria binding to the column between 0-100 ml. HP0175 elutes at 35% which represents fractions 8-12 in 7 mM HEPES pH 7.4, 175 mM NaCl and 350 mM imidazole. **(B)** 12.5% SDS PAGE; pellet after sonication (lane 1), supernatant for purification (lane 2), flow through fractions (lanes 3-5), elution of peak 1 from purification chromatogram in (A) is demonstrated in fractions 6 and 7, fractions from peak 2 beginning from 250 ml are in lanes 8-12, representing the elution of HP0175.

`

3.7. Crystallization of HP0175

Crystallization trials were prepared as per Section 2.6 and 2.7 using vapour diffusion sitting drop and hanging drop geometries, for screening and optimization, respectively. Initial crystallization plates were arranged with densities varying between 5-25 mg/ml (0.75 μL protein and 0.75 μL buffer content), using the MCSG I-III crystallization suites. A promising crystal was obtained from reservoirs of 0.1 M Tris:HCl pH 8.5 and 2.5 M ammonium sulfate (Figure 3.7A); however this crystal was not cryoprotected, therefore further optimization was obligatory. Upon optimization, spherulites (Figure 3.7B) were observed which presented the opportunity for microseeding. Microseeding trials were performed with the spherulites of dilutions 1:10, 1:100, 1:1000 and 1:10000, the seeds were applied at different stages of protein and buffer incubations, plates were equilibrated and allowed to reach saturation over the course of 24 and 48 hours, this also ensured the seed added would not dissolve. A crystal seed diluted by 1:1000 was added after 48 hours of equilibration, which provided a large crystal (Figure 3.7C) in 0.1 M Tris:HCl, 47.4 % ammonium sulfate and 7 % (w/v) PEG 3350. Conversely, this crystal did not provide for a solvable diffraction pattern at 7.8 Å (Figure 3.7.1A). The condition in Tris:HCl (Figure 3.7C) yielded a poorly diffracting crystal, but provided a thorough estimate of crystallization concentration. HP0175 was re-screened against fresh crystallization buffers MCSGI-III suite buffers; a diffraction quality crystal was obtained. This crystal was observed to diffract to a resolution of 2.3 Å (Figure 3.7.1B) on the home source, and to 2.09 Å at the synchrotron (Figure 3.7.1C).

`

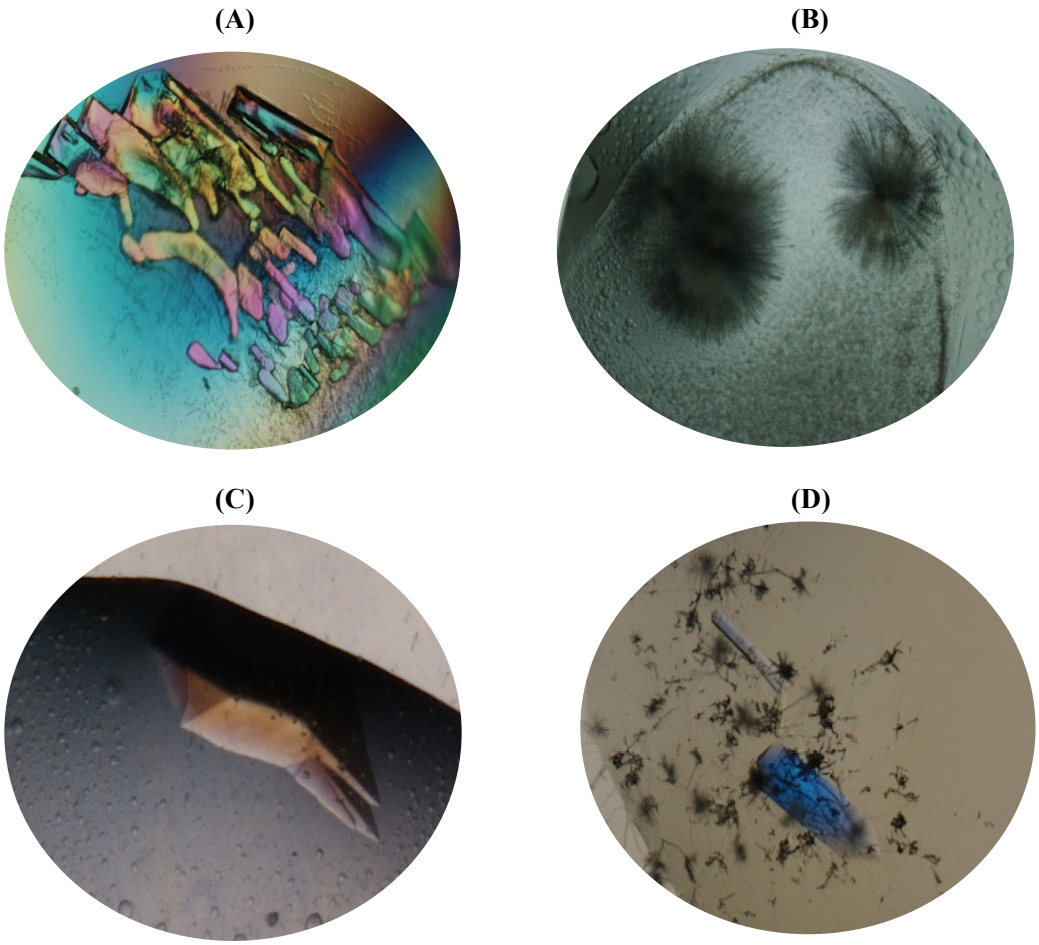


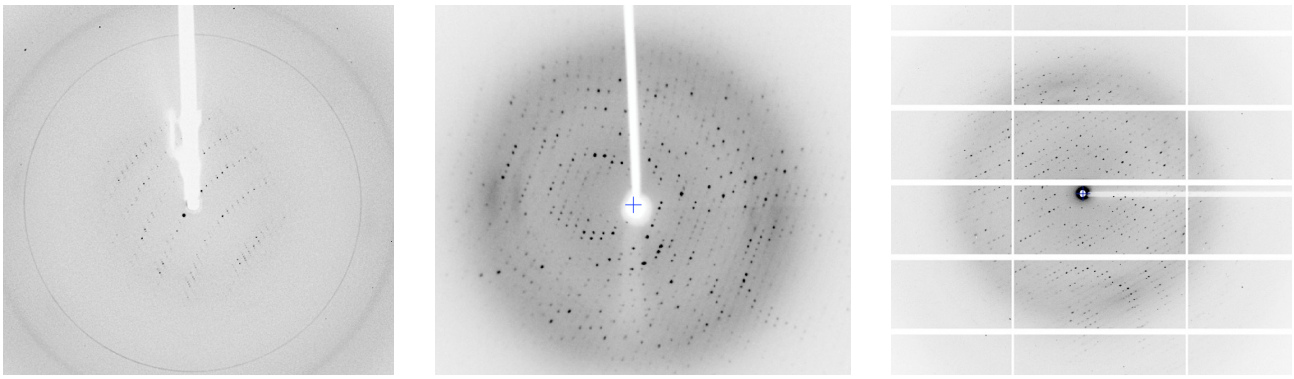
Figure 3.7. Crystallization of HP0175. (A) Initial crystal achieved with 14.5 mg/ml protein sample in MCSG-II (condition H11) comprising of 0.1 M Tris:HCl pH 8.5 and 2.5 M ammonium sulfate, discovered 9 days after plating. **(B)** Spherulite achieved with optimization plate in hanging drop vapour diffusion, condition consisted of 0.1 M Tris-HCl pH 8.4, 43.1 % ammonium sulfate and 7.5% (w/v) PEG 3350. The following condition was obtained based on the crystal in (A). **(C)** Hanging drop vapour diffusion obtained from microseeding with (B); 0.1 M Tris:HCl pH 8.3, 47.4% ammonium sulfate and 7.5% (w/v) PEG3350. **(D)** Re-screening of HP0175 with 16.7 mg/ml in fresh MCSGII crystallization suite, condition: 0.1 M HEPES:NaOH (pH 7) and 30% (v/v) Jeffamine, also known as O,O'-Bis(2-aminoprolyl)polypropylene glycol-block-polypropylene glycol.

3.7.1. Diffraction Analysis

The HP0175 crystal obtained from microseeding (Figure 3.7C) diffracted to 7.8 Å (Figure 3.7.1A) at the synchrotron. While further screening achieved with MCSG-II suite diffracted to 2.3 Å (Figure 3.7.1B) on the home source (Micromax007HF) and to 1.8 Å at the synchrotron. Analysis

`

of the data indicated the highest diffraction resolution was not 1.8 Å, but rather 2.09 Å due to poor resultant $CC_{1/2}$ (0.09) and I/ σ(I) (0.003) values beyond 2.09 Å.



(A) **(B)** **(C)**

Figure 3.7.1. Diffraction images for HP0175 crystals. (A) CMCF 08B1-1, detector distance set to 400 mm, 3 s exposure time, 90°, indicated a resolution of 7.8 Å. **(B)** Diffraction acquired on Rigaku Micro-Max 007 HF, resolution range 29.96-2.3 Å, exposure time 20 s at 286°. **(C)** CMCF 08ID-1 beamline, pilatus 6M detector, 200.00 mm distance, 0.3 s exposure, 4°, comprising a resolution range of 48.48 - 2.09 Å.

The initial space group identified from the diffraction data (both on the home source and beamline data) was determined to be $P3_121$. Upon phasing via molecular replacement, the enantiomeric space group $P3_221$ was determined as the appropriate space group, therefore the diffraction data was re-processed accordingly using $P3_221$. Images were indexed utilizing XDS-GUI,[117] scaled and merged by the Pointless/Aimless/cTruncate routine[119] (Table 3.1) in the CCP4 program suite.[121] Molecular replacement was performed with Phaser[122] followed by Autobuild routine in PHENIX;[123] the model was rebuilt with Coot[125] and refined by Refmac. Merging statistics are found in Table 3.1 while refinement statistics are outlined in Table 3.2.

`

Table 3.1. Merging statistics for diffraction data

	081D-1 Beamline, CLS
Space group	P 3_2 2 1
Wavelength (Å)	0.979
Resolution Range (Å)	48.48 - 2.09 (2.14 - 2.09)
Unit cell parameters (Å, °)	a=b= 90.95, c = 67.05 $\alpha = \beta = 90$ and $\gamma = 120$
Total number of observations	190,760 (14742)
No. of unique reflections	19446 (1459)
Completeness (%)	99.9 (99.5)
Average I/σ(I)	15.2 (1.5)
R_{meas} (all I+ & I-)	0.088 (1.690)
R_{merge}[a]	0.084 (1.604)
$CC_{1/2}$	0.999 (0.786)
B Wilson (Å^2)	42.5
Multiplicity	9.8 (10.1)
Mosaicity (°)	0.33

*Brackets contain highest resolution shell (2.14-2.09 Å)

[a]$R_{merge} = \sum_h \sum_i (|< I_h > - I_{h,i}| / \sum_h \sum_i I_{h,i})$, where h represents unique reflections, i depicts their symmetry equivalent

Table 3.2. Summary of refinement statistics

Resolution range (Å)	45.48-2.09
Reflections used	18461
Completeness for range (%)	99.9
R_{work}[b]	0.217
R_{free}[c]	0.275
Average B-factors (Å^2)	61.78
No. of atoms	2061
Water	57
Ramachandran Plot (%)	
Favoured	98.0
Allowed	2.0

[b]$R_{work} = \sum_{hkl} \| F^{obs} hkl| - |F^{calc} hkl\| / \sum_{hkl} |F^{obs} hkl|$

[c]R_{free} = is the cross validation R factor using 5% of reflections

`

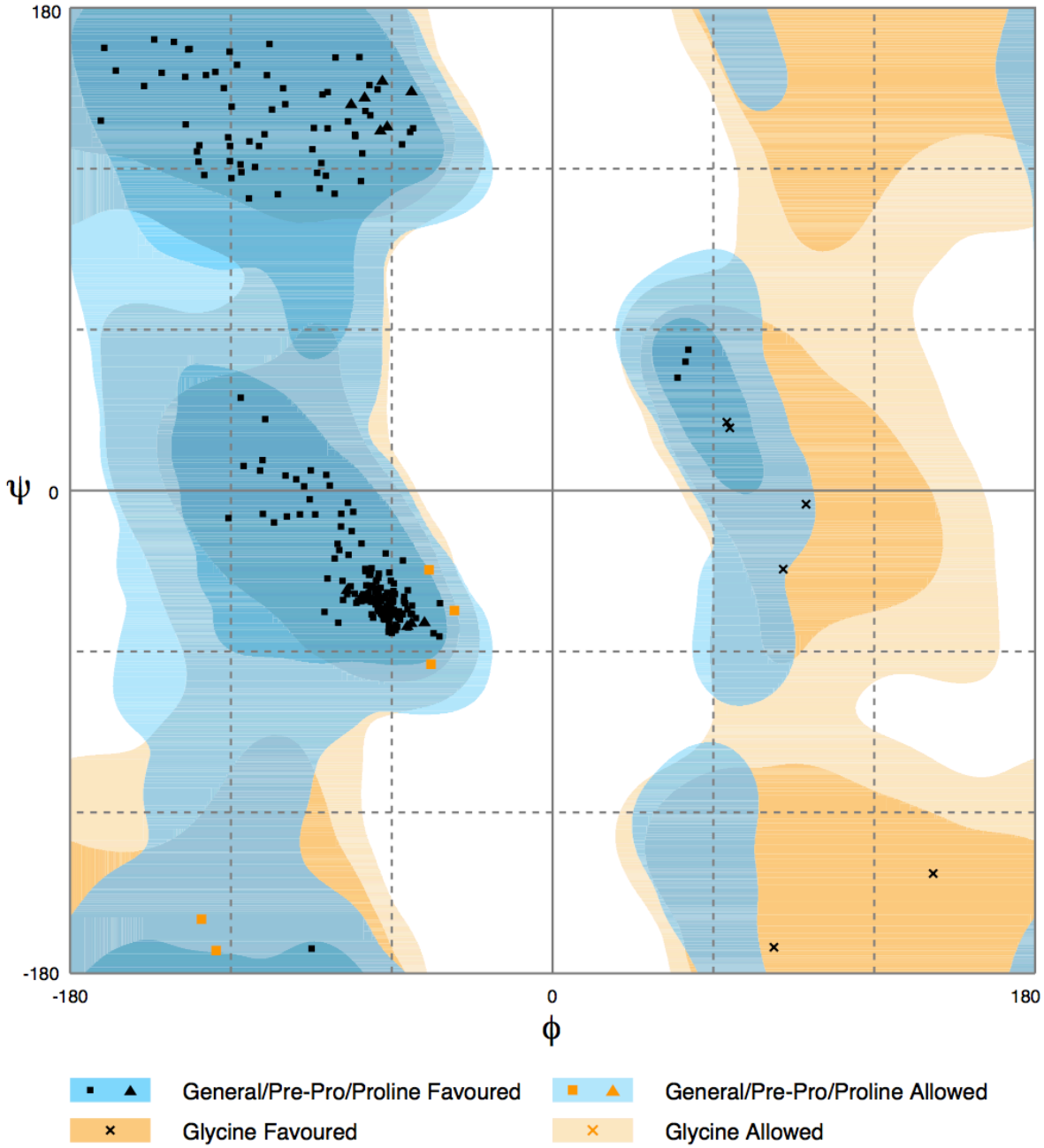


Figure 3.7.2. Ramachandran plot for HP0175 model building. Number of residues in favoured region (~ 98.0% expected): 240 (98.0%). Number of residues in allowed region (~2.0% expected): 5 (2.0%), and there are no residues in outlier region. Figure generated by RAMPAGE.[130]

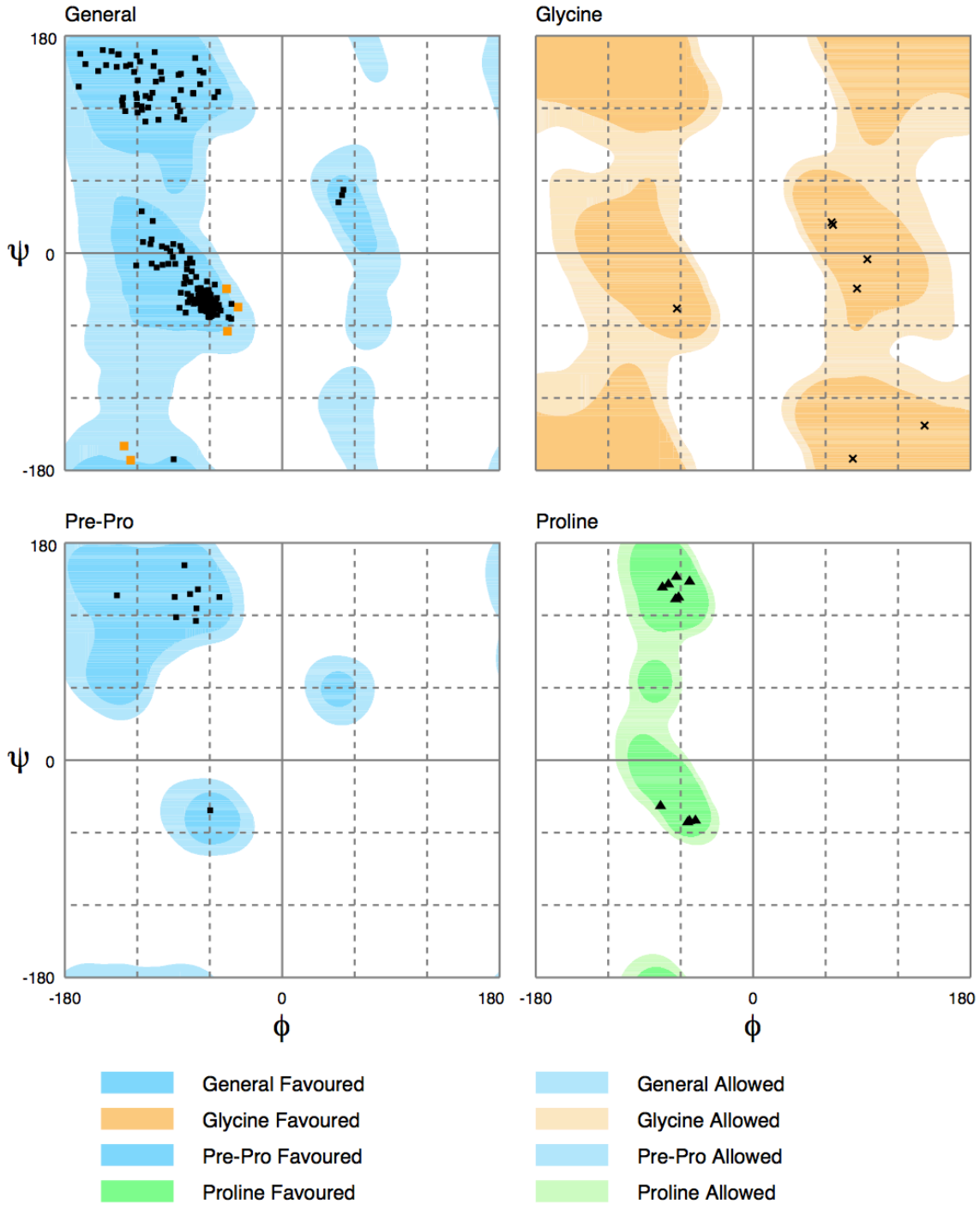


Figure 3.7.3. Ramachandran plot for specific residues in HP0175. Number of residues in favoured region (~ 98.0% expected): 240 (98.0%). Number of residues in allowed region (~2.0% expected): 5 (2.0%), and there are no residues in outlier region. Figure generated by RAMPAGE.[130]

`

3.7.2. Overall Structure of Apo-HP0175

We have crystallized the apo form of the peptidyl-prolyl *cis,trans*-isomerase HP0175 to a resolution of 2.09 Å in the space group $P3_221$, with one monomer in the asymmetric unit. The overall monomeric structure of apo-HP0175 comprises three domains, the PPIase domain and the N- and C-terminal helical chaperone domains (Figure 3.7.5A). The N- and C-terminal α-helical chaperone domains are between residues 75-97 ($\alpha1$), 104-132 ($\alpha2$) and 261-292 ($\alpha8$), respectively. The conserved PPIase domain is located between residues 151-257 and is represented by β-strands ($\beta3$, $\beta4$ and $\beta5$) enclosed within four helices ($\alpha4$, $\alpha5$, $\alpha6$ and $\alpha7$). Respectively, the overall monomeric shape of HP0175 resembles that of an asymmetric dumbbell. In addition to the helices of the N- and C-terminal domains, β-strands are located in residues 45-49, 51-53 and 295-297 (Figure 3.7.5A). The linker region Loop 1 (98-103 aa) connects helices $\alpha1$ (75-97 aa) and $\alpha2$ (104-132 aa) facilitating stabilization of the helices. Within the first turn, hydrogen bonding is observed between the carbonyl of the main chain of Ala95 ($\alpha1$) and main chain amide of Leu100 (Loop1). Within Loop 1, stabilizing H-bonds are observed between the main chain carbonyl of Lys99 and amide backbone of Asp101 (3.18 Å). Thr103 contains two important interactions, the first interface is within Leu100 of loop 1 (main chain; carbonyl) and the second with Lys107 of $\alpha2$ (main chain; amide), H-bond interactions measured as 2.95 Å and 2.89 Å, respectively.

Parvulin-type PPIases are classified as domain-swapped dimeric proteins, observed via the structures of *cj*PEB4,[98] *Bs*PrsA,[99] and *Ec*SurA.[100] Apo-HP0175 was crystallized in the space group $P3_221$ with one monomer in the asymmetric unit. We therefore sought to understand if HP0175 was a monomer or dimer prior to binding a target molecule (such as I2CA), initially through size exclusion chromatography (SEC). SEC analysis of HP0175 was however inconclusive in determining monomeric or dimeric state of the protein (not shown) with multiple overlapping species being present, which is likely attributable to the extended N- and C-terminal chaperone domain helices of HP0175 increasing interactions with the SEC matrix versus that of a more globular protein. A similar abnormal SEC separation has previously been observed in PEB4, a parvulin-like PPIase with 34.5% sequence identity to HP0175,[98] where the protein was observed to elute abnormally through interaction with the SEC matrix.[98] We therefore explored electrospray ionization mass spectrometry (ESI-MS) in order to provide further insight into the potential monomeric or dimeric state of apo-HP0175 in solution. Interestingly, the ESI-MS spectrum

`

observed for apo-HP0175 in 10 mM ammonium acetate (Figure 3.7.4) indicated both monomeric and dimeric species with differing mass to charge (m/z) distributions being present. Indeed, three (3) species are observed: monomeric protein between m/z 2000-3000 and +12 to +20 charge, dimeric HP0175 at m/z 3700-4700 and +16 to +20 charge, and unfolded protein observed between m/z 1000-5000 and +21 to +31 charge. Under these conditions, the monomeric form of the protein ionizes at a much more significant rate in comparison to the dimeric form of the protein. The average molecular masses observed for the monomeric and dimeric species of HP0175 were 36,421.85 Da and 72,945.27 Da, respectively; a molecular mass of 35,252.25 Da was also observed in ESI-MS experiments, which can be associated to an HP0175 monomer following the loss of the hexahistidine purification tag.

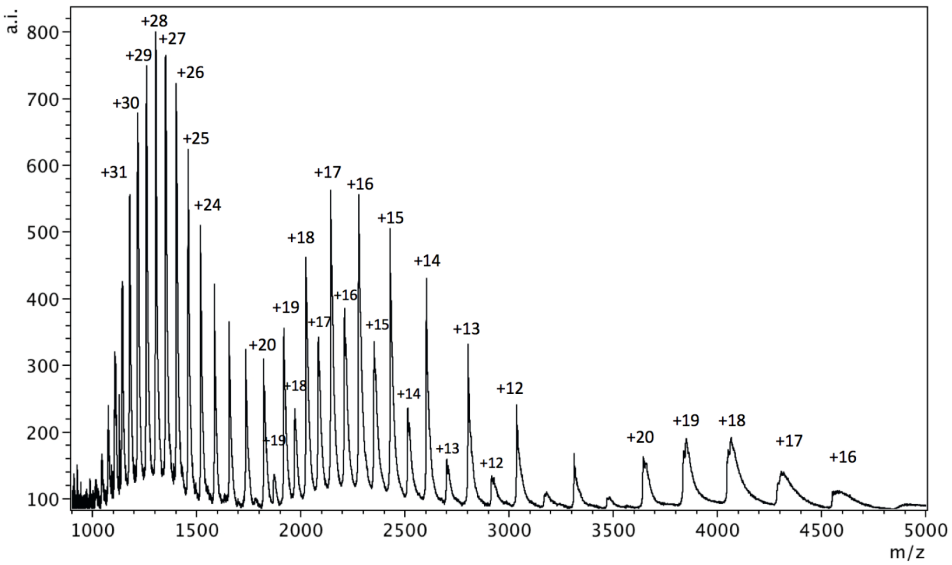


Figure 3.7.4. Electrospray ionization mass spectrum of apo-HP0175. ESI-Mass spectrometry chromatogram on native HP0175 in 10 μM Ammonium acetate pH 8.1 (NH_4:Cl). A high range of charge distribution is observed for the monomeric HP0175, ranging between +11 and +34 charge states (1000-3000 m/z), providing an average molecular weight of 36,421.85 Da. A lower molecular weight was also observed, calculated to be 35,252.25 Da (2000-3000 m/z). The dimer representation shows a structured protein in comparison to the monomer, charge states between +16 to +20, averaging the dimeric molecular weight to 72,945.27 Da.

These observed masses from ESI-MS are different from what was observed in the final refined structure of the protein. While the mass observed in ESI-MS indicated the presence of all

`

residues at the start of crystallization, in the final refined structure, only residues 43-60 and 69-299 could be modeled within electron density maps. The original pET28a expression vector expressed a protein of 322 residues, as observed in ESI-MS, and consists of 10 amino acids for the purification tag and 299 residues of the complete HP0175 gene; of these 299 residues is a predicted 21 aa N-terminal signal peptide (as predicted by SingalP).[131] It is therefore likely that these 31 residues were auto-catalytically removed during the time course of crystallization, resulting in a crystallized protein that spanned residues 22-299, of which the noted residues were observed in the electron density. It is interesting to note that the HP0175 construct used in the determination of the I2CA-bound protein (PDB ID 5EZ1; Naveen et al., 2016)[97] was cloned to span residues 33 to 299 of the protein with a C-terminal hexahistidine purification tag. The authors indicated that their construct represented the protein following the removal of the signal peptide,[97] although the predicted signal peptide cleavage is between residues 21-22, their final model consists of residues 54-61 and 74-296.[97]

The dimer representation of apo-HP0175 was generated with a symmetry related monomer (X-Y, -Y, -Z+1/3), in reference to the first monomer in coordinates (X, Y, Z) represented in Figure 3.7.5B. Stabilization of the homodimer is through intertwining of the NC α-helices and the hydrogen bonding between N-terminal β-strand 1 (aa 45-49) of the (X, Y, Z) monomer and the C-terminal β-strand 6 (aa 295-297) of the symmetry mate. Specifically, hydrogen bonding is observed between the backbones of Asp49 and Lys294, Thr47 and Val296, Leu45 and Ile297, Val44 and Asn298. Examination of the thermal factors (B_{av}) for β-strands 1, 2 and 6 indicate that B_{av} for β-strand 1 and 6 were 59.2 and 53.8 Å^2 respectively, while β-strand 2 was determined as 70.4 Å^2. The differences in thermal B_{av} calculated, further support that homodimer is stabilized by the interaction between β-strand 1 and 6; β2 is not similarly stabilized, rendering it slightly more variable in position.

`

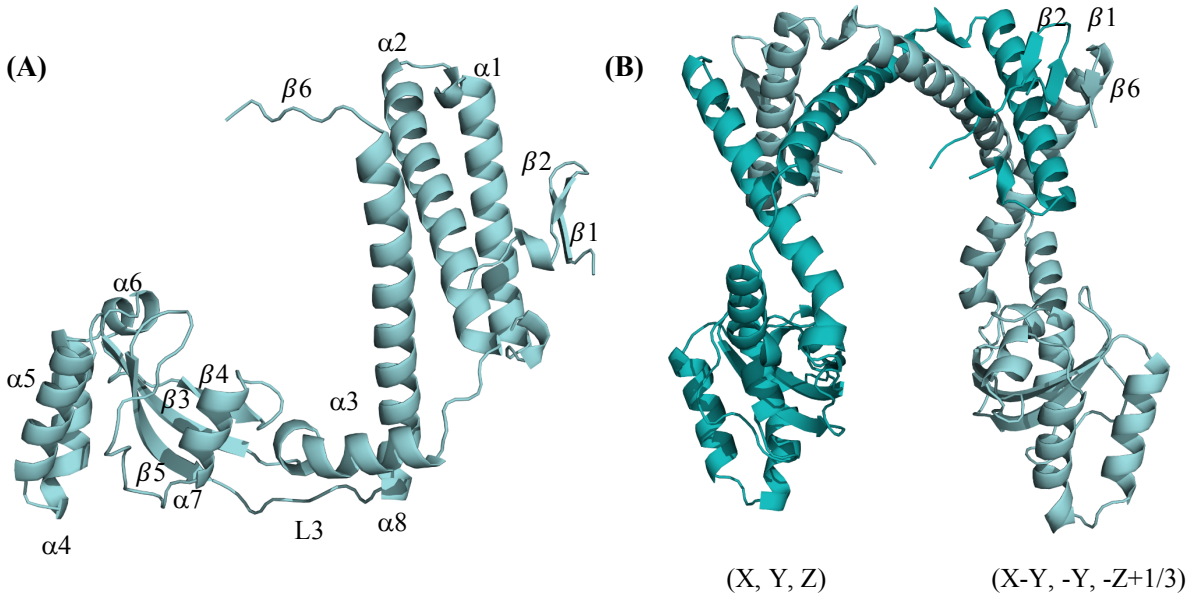


Figure 3.7.5. Monomeric and dimeric forms of apo-HP0175. (A) The apo-HP0175 monomer observed in the crystallographic asymmetric unit. The monomer consists of three domains; the N-terminal chaperone domain, composed of β1 (aa 45-49), β2 (aa 51-53), α1 (aa75-97), and α2 (aa 104-132), precedes the PPIase domain (aa 151-257), comprised of four β-strands (β3, β4 and β5) enclosed within four helices (α4, α5, α6 and α7), and ends with the C-terminal chaperone domain of α8 (aa 261-292) and β6 (aa 295-297). Connecting loops are shown as L1 (aa 98-103), L2 (aa 133-136), and L3 (aa 255-261). **(B)** The HP0175 dimer is generated the symmetry (X-Y, -Y, -Z+1/3) related monomer (cyan) to that of the (x, y, z) monomer (teal). Stabilization of the dimer occurs through intertwining of the N- and C-terminal helices, as well as hydrogen bonding between β-strands 1 and 6, and specific interactions occurring between main chain atoms of Val44 and Asn298, Leu 45 and Ile297, Asp49 and Lys294, Thr47 and Val296.

3.7.3. Structural Variations Between Apo-HP0175 and I2CA-HP0175

Target molecule binding in the PPIase domain results in only modest changes (Figure 3.7.6). The catalytic residues in the PPIase domain (aa 151-257) are His161, Asp199, Met219, His161 and His247.[97] Superposition of apo-HP0175 and I2CA-HP0175 PPIase domains obtained an r.m.s.d of 0.384 Å for 91 C_α atoms (Figure 3.7.6A). Structural alterations directly associated with residues interacting with the indole-2-carboxylic acid ligand are Asp199, Met219, Leu211, Phe223, Phe244 and His247 were observed (Figure 3.7.6B).[97] Specifically, addition of the ligand weakened the stabilization amongst His161, His247 and Asp199 (Figure 3.7.6C). The histidine residues 161 and 247 had contact distances of 2.9 Å in apo-HP0175 which increased to 3.9 Å in

`

I2CA-HP0175. In relation, His161 and Asp199 increased to 3.4 Å from 2.7 Å in I2CA-HP0175. While the sulfur atom of Met219 is positioned closer to the I2CA-phenyl ring with a distance of 3.3 Å (S···C), resulted in decreasing clashes between the terminal methyl in Met219 and the I2CA.[97]

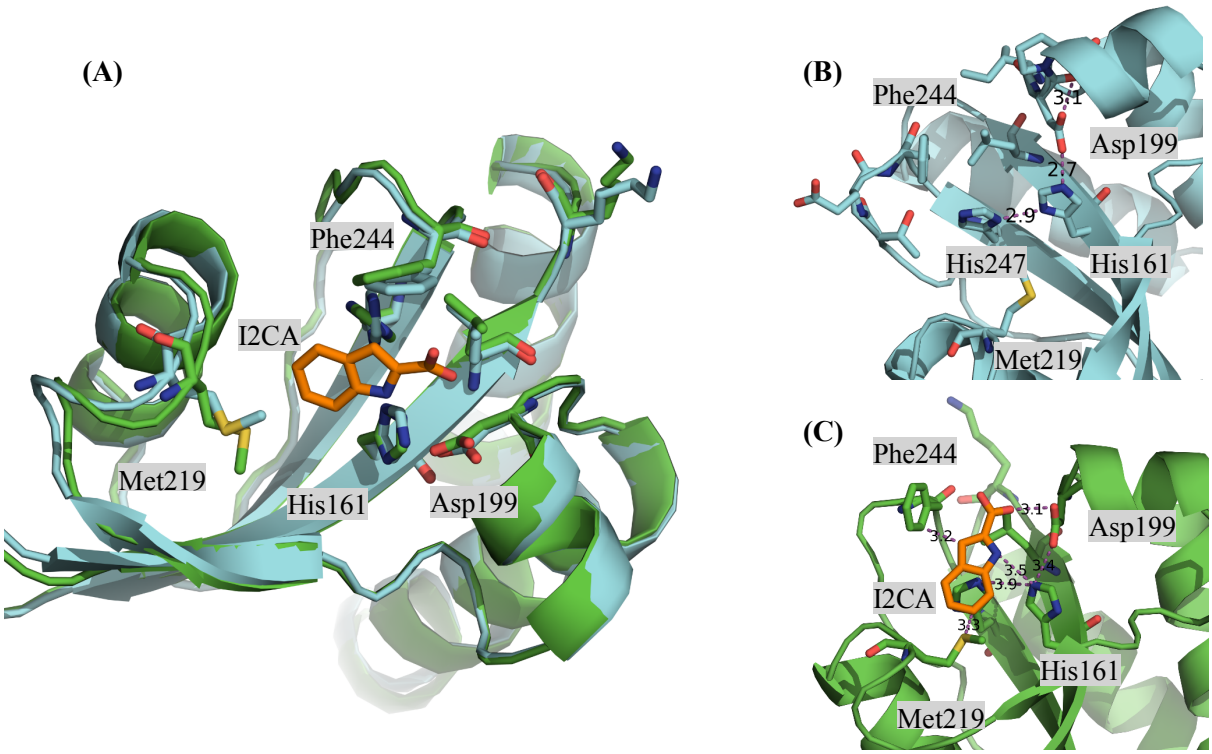


Figure 3.7.6. Superposition of the catalytic site in apo and I2CA bound HP0175. (A) Superposition of the PPIase domain in apo-HP0175 (cyan) and I2CA-HP0175 (green) is demonstrated by a cartoon figure, where the residues involved in the catalytic pocket is demonstrated in sticks. Indole-2-carboxylic acid (inhibitor) is coloured in orange for a clearer representation. **(B)** Apo-HP0175 shows interactions between His 247, His161 and Asp199. **(C)** I2CA-HP0175 demonstrates that interaction between His161 and Asp199 are weakened and disrupted between His247 and His161 due to the addition of I2CA, while the sulfur atom is closer to the indole ring and methyl group is directed away. (PDB ID I2CA-HP0175: 5EZ1 and 6BHF of APO-HP0175).

Previous studies had used cjPEB4 as the structural reference for apo-HP0175 conformation, elaborating on targeted interaction; PEB4 was a prime candidate for representing apo-HP0175 based on 34.5% sequence identity and superposition yielding an r.m.s.d of 3.3 Å for 223 C_α atoms to I2CA-HP0175.[97] With the determination of the structure of apo-HP0175, we are

`

able to compare the dimer conformation of HP0175 in the apo and I2CA bound to further extrapolate on structural changes; one noted observation is the opening of the central cavity within the dimer. Distances within the central cavity were measured between two Phe143 (α3) and two Gln279 (α8) residues, while the distance separating the PPIase domain was measured between Gln206 residues between monomers. In the apo-HP0175 structure, the distances were calculated as 53.0 Å (α3) and 40.5 Å (α8) for the central cavity, and 47.3 Å for the PPIase domain, respectively (Figure 3.7.7A). Binding of I2Ca in the PPIase domain increases the opening in the central cavity to 68.1 Å (α3) and 47.8 Å (α8), respectively; binding of I2CA results in a closing of the distance between the PPIase domains, halving the distance between each monomeric PPIase domain (22.8 Å) (Figure 3.7.7B).

The structure of native HP0175 and I2CA-HP0175 were determined in enantiomeric space groups ($P3_221$ and $P3_121$, respectively) from slightly different crystallization conditions, though they diffract to similar resolutions and have similar solvent contents. The present similarities and differences in the two structures allow for the comparison of thermal factors in certain domains (Table 3.3) in support of observed structural differences between the two structures. Thermal factors (B_{av} or B-value) are based on the static dynamic motion of atoms within the crystal and respectively determine contouring of the electron density map.[132] For the purpose of the current comparison, each domain was arranged from low to high B_{av} values found in one crystal. The domains are then compared to the same domains within the second crystal and/or second monomer to identify any significant changes. This is performed as an indirect way to standardize the measured values in each crystal. In addition, crystal packing of the two structures resembled similar contacts, therefore the thermal factors analyzed are not a result of variances in packing interactions (Table 3.4 – Table 3.6), which will be discussed shortly.

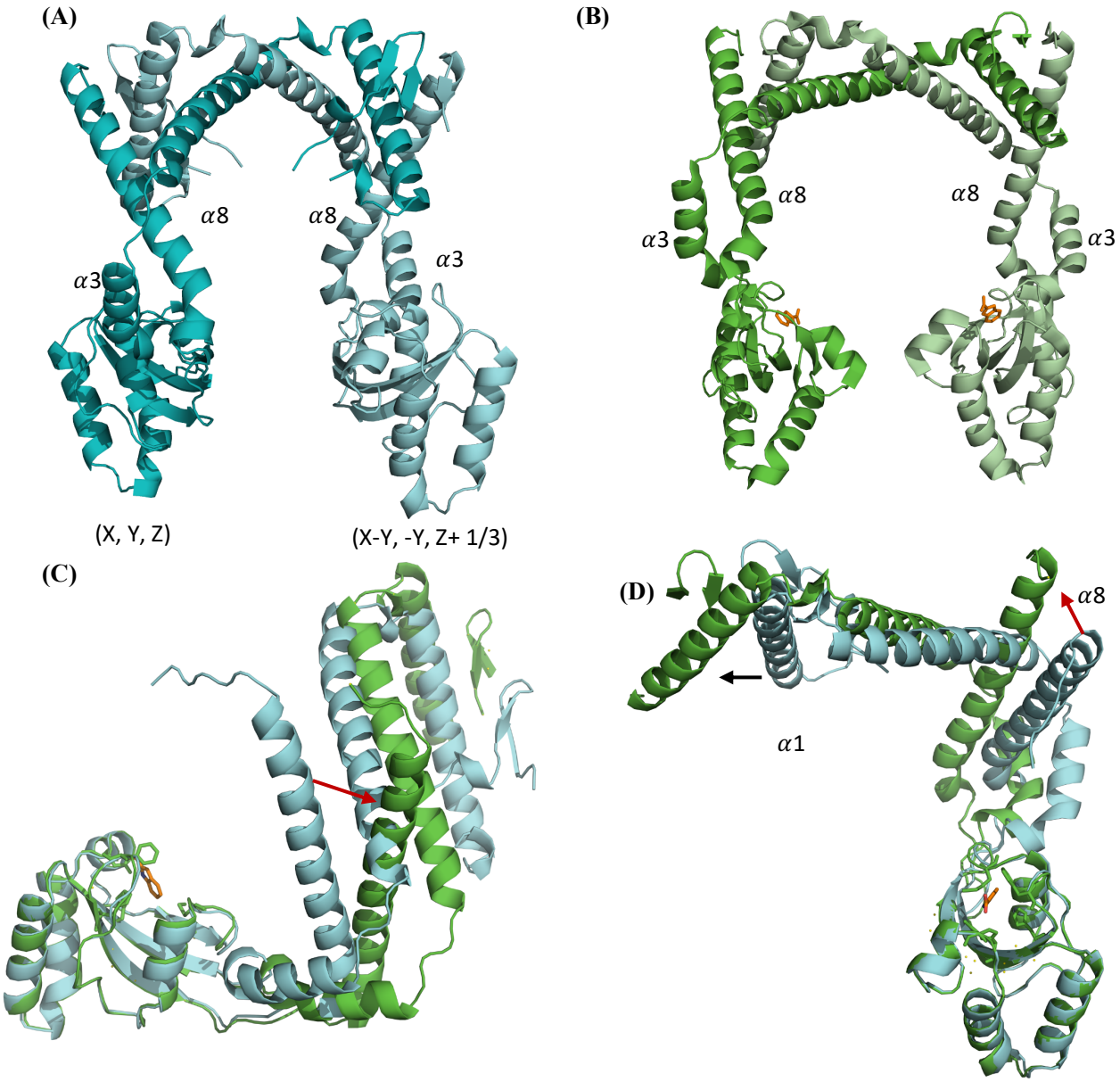


Figure 3.7.7. Comparison of apo-HP0175 and I2CA-HP0175. **(A)** Symmetry generated dimer of apo-HP0175 in comparison to dimeric I2CA-HP0175 **(B)**. **(C)** Superposition of monomeric apo and I2CA-HP0175; anchored at the PPIase domain to demonstrate the extension of helix α8 from the apo (cyan) to I2CA (orange)-bound HP0175, red arrow to show the path of extension of helix α8 from the apo (cyan) to I2CA (orange)-bound HP0175, red arrow to show the path of extension. **(D)** Top-view representation demonstrating the extension of the helices, black arrow shows the extension of α1 from the apo to ligand bound HP0175. (PDB ID I2CA-HP0175: 5EZ1[97] and 6BHF of apo-HP0175).

`

Table 3.3. Comparison of B_{av} values[a] for apo-HP0175 and I2CA-HP0175

Protomer	Label	Residues	B_{av} MC	B_{av} Range	B_{av}
Apo-HP0175			59.0	50.08-92.7	62.0
Chain A: I2CA-HP0175	α1	75-97	57.0	26.7-117.7	61.0
Chain B: I2CA-HP0175			64.0	32.8-120.4	67.7
Apo-HP0175			59.1	52.2-80.4	63.5
Chain A: I2CA-HP0175	α2	104-132	41.4	31.3-86.7	50.0
Chain B: I2CA-HP0175			47.1	34.0-86.1	55.5
Apo-HP0175			55.5	48.7-71.3	58.6
Chain A: I2CA-HP0175	L3	255-262	61.3	35.6-69.0	59.4
Chain B: I2CA-HP0175			82.1	69.8-104.1	82.0

[a]B_{av} values are provided in $Å^2$, Main Chain (MC)

The comparison of B-values between apo- and I2CA-bound in the current study is based on a similar report by Edwards and colleagues;[77] in this study, thermal factors from the structures of *E. coli* cytoplasmic peptidyl-prolyl *cis,trans*-isomerase (EcoCyPA) bound with a tripeptide substrate succinyl-Ala-Pro-Ala-*p*-nitroanilide and in its unliganded form were compared. The study compared protein crystals grown in different crystallization conditions and space groups; $C222_1$ and $P2_12_12_1$ for the uncomplexed and complexed EcoCyPA, respectively. The authors stated that comparison of crystal structures in two different space groups provides more insight on how a structure may behave in solution, and comparison of the two structures determined similar overall conformations, the regions compared were those involved in ligand binding.[77] Within a crystal structure, an analysis of the B-values is a means to assess the regional thermal motion; this analysis can only be applied for thermal motion within particular regions of the protein, rather than as a metric for the protein as a whole, as crystal contacts may affect thermal motion of loops, helices or β-strands. Additionally, thermal factors can be compared across different protein-DNA structures through normalization of B-factors by unity-based scaling or z-score normalization, as demonstrated by Schneider et al., 2014.[132] Since apo- and I2CA-bound HP0175 were of similar overall structure, and crystallized in enantiomeric space groups to similar diffraction resolutions, it was unnecessary to scale the B-factors. A more in-depth analysis of the protein's dynamics in the apo- and ligand-bound states to compliment the thermal factor analysis presented herein could be undertaken through Hydrogen Deuterium Exchange Mass Spectrometry or heteronuclear magnetic resonance NMR.

`

Analysis of the N-terminal helices demonstrated $\alpha 1$ helix is extended by 6.6° from apo to the I2CA-bound conformation (Figure 3.7.7D), measured between Lys74 ($\alpha 1$), Thr97 ($\alpha 1$) and Asp101 (Loop 1). In the apo-HP0175 structure, helices $\alpha 1$ and $\alpha 2$ contained similar B_{av} values of 62 Å^2 and 63.5 Å^2, indicating helix $\alpha 1$ contained a higher rigidity in comparison to $\alpha 2$. However, the small difference between the two B_{av} values is not significant and can therefore be concluded that both $\alpha 1$ and $\alpha 2$ are equally variable. The reverse is observed with I2CA-HP0175, where the $\alpha 1$ helix contained thermal factors of 61 Å^2 in Chain A and 67.7 Å^2 in Chain B, while the $\alpha 2$ helix contained a B_{av} of 50 Å^2 (Chain A) and 55.5 Å^2 (Chain B). Increased rigidity of the $\alpha 2$ helix can be explained by $\alpha 1$ extension, inferring once helices are extended, $\alpha 2$ is forced into a locked configuration.

The differences observed in/near the PPIase domain show increased thermal factors in Loop 3 with 59.4 and 82.02 Å^2 for chains A and B of I2CA-HP0175, respectively, whereas in apo-HP0175 the B_{av} in loop 3 is 58.6 Å^2. From this comparison, we can infer upon the effect of ligand interaction through the changes noted with residues His161, His247, Asp199, Met219 and Phe244 subsequently increasing thermal motion in Loop 3 (aa 255-261).[97] Increased thermal motion can be hypothesized to induce the extension of the C-terminal $\alpha 8$ helix by 7.4° from apo to I2CA-HP0175, measured between Glu287 ($\alpha 8$), Tyr262 and Lys254 (PPIase) (Figure 3.7.7C). Extension of the N- and C- terminal helices in ligand interaction of HP0175, supports the hypothesis of stabilizing protein undergoing catalysis through protein-protein interactions.

The thermal factors between helices $\alpha 1$ and $\alpha 2$ and Loop L3 of apo-HP0175 and I2CA-HP0175 were also compared. In order to assess if differences in structures were attributable to crystal contacts, we reviewed the symmetry mates associated with these regions (Tables 3.4-3.6). Within the N-terminal helices in apo-HP0175, only one symmetry contact was observed within $\alpha 2$ helix; Lys131 and symmetry related Thr166 of the PPIase domain. The thermal factors associated with $\alpha 1$ (62.0 Å^2) were only slightly lower than $\alpha 2$ (63.5 Å^2), even though $\alpha 1$ did not contain any symmetry related crystal contacts. B_{av} values of the N-terminal helices $\alpha 1$ and $\alpha 2$ in the I2CA bound HP0175 shows a similar pattern, in Chain A; $\alpha 1$ and $\alpha 2$ are calculated as 61.0 Å^2 and 50.0 Å^2. In Chain B; the thermal factors were similar and measured as 67.7 Å^2 and 55.52 Å^2 for $\alpha 1$ and $\alpha 2$. In both chains A and B, crystal symmetry contacts are observed in Thr97, Chain B contains

`

an additional symmetry contact involving $\alpha 2$ (Glu111 — Lys294), however this does not change the outcome of both Chain A and B thermal values.

Apo-HP0175 loop L3 (aa 255-261) shows one symmetry related interaction with Ser256 to Asp83 (-X, -X+Y, -Z+2/3) with a B average of 58.6 Å^2 (range: 52-70 Å^2). While the I2CA-HP0175 chain A contained 2 symmetry related crystal contacts between Ser256 to symmetry contact Thr231 and Pro232, the thermal factor associated with loop 3 was 59.4 Å^2, while Chain B did not contain any symmetry related crystal contacts, and a B_{av} of 82 Å^2. Although chain B did not contain any symmetry related crystal contacts, chain A contained 2 symmetry related crystal contacts, however, the thermal factor remained relatively higher based on the range of 21-94 Å^2 (Chain B).

Table 3.4. Apo-HP0175 symmetry crystal contacts

Contact residue to symmetry related contact	Symmetry Operator	Distance (Å)
Val44 (O) — Asn298 (ND2)	X-Y, -Y, -Z+1/3	2.99
Leu45 (O) — Asn298 (N)	X-Y, -Y, -Z+1/3	2.92
Thr47 (N) — Val296 (O)	X-Y, -Y, -Z+1/3	3.02
Thr47 (O) — Val296 (N)	X-Y, -Y, -Z+1/3	3.19
Asp49 (N) — Lys294 (O)	X-Y, -Y, -Z+1/3	2.80
Asp49 (OD1) — Lys294 (N)	X-Y, -Y, -Z+1/3	2.68
Lys131 (O) — Thr166 (OG1)	-Y, X-Y, Z+ 2/3	2.83
Lys154 (N) — Glu287 (OE1)	-X+Y, -X, Z+1/3	2.66
Thr166 (OG1) — Lys131 (O)	-X+Y, -X, Z+1/3	2.83
Glu243 (OE2) — Lys148 (NZ)	-Y, X-Y, Z+2/3	2.54
Lys254 (O) — Ile297 (N)	-X+Y, -X, Z+1/3	3.03
Ser256 (OG) — Asp83 (OD1)	-X, -X+Y, -Z+2/3	2.43
Lys270 (NZ) — Asp168 (OD2)	-Y, X-Y, Z+2/3	3.10
Gln274 (NE2) — Asp168 (OD2)	-Y, X-Y, Z+2/3	3.07
Lys294 (N) — Asp49 (OD1)	X-Y, -Y, -Z+1/3	2.68
Lys294 (O) — Asp49 (N)	X-Y, -Y, -Z+1/3	2.80
Val296 (O) — Thr47 (N)	X-Y, -Y, -Z+1/3	3.02
Val296 (N) — Thr47 (O)	X-Y, -Y, -Z+1/3	3.19
Ile297 (N) — Lys254 (O)	-Y, X-Y, Z+2/3	3.03
Asn298 (N) — Leu45 (O)	X-Y, -Y, -Z+1/3	2.92
Asn298 (ND2) — Val44 (O)	X-Y, -Y, -Z+1/3	2.99
Lys299 (O) — Lys254 (NZ)	-Y, X-Y, Z+2/3	2.81

`

Table 3.5. I2CA-HP0175 symmetry crystal contacts for Chain A

Contact residue to symmetry related contact	Symmetry Operator	Distance (Å)
Thr97 (O) — Thr97 (OG1) (B)	-X, -X+Y, -Z+1/3	2.77
Gln150 (NE2) — Glu167 (OE1) (B)	X, Y, Z	2.22
Gln150 (NE2) — Tyr246 (OH) (B)	X, Y, Z	3.10
Asp178 (O) — Tyr262 (N) (A)	Y, X, -Z	2.87
Lys182 (N) — Glu263 (OE2) (A)	Y, X, -Z	3.01
Asn194 (ND2) — Gln206 (OE2) (A)	Y, X, -Z	2.99
Arg195 (O) — Arg195 (NH1) (A)	Y, X, -Z	2.98
Gln206 (OE1) — Asn194 (ND2) (A)	Y, X, -Z	2.99
Ser256 (OG) — Thr231 (OG1) (B)	Y, X, -Z	2.91
Ser256 (N) — Pro232 (O) (B)	Y, X, -Z	2.95
Glu263 (OE2) — Lys182 (N) (A)	Y, X, -Z	3.01
Lys294 (NZ) — Glu111 (OE2) (B)	-X, -X+Y, -Z+1/3	3.07

The letters A and B in brackets refer to the symmetry related to Chain A (A) and Chain B (B)

Table 3.6. I2CA-HP0175 symmetry crystal contacts for Chain B

Contact residue to symmetry related contact	Symmetry Operator	Distance (Å)
Thr97 (OG) — Thr97 (O)	-X, -X+Y, -Z+1/3	2.77
Glu111 (OE2) — Lys294 (NZ)	-X, -X+Y, -Z+1/3	3.07
Glu167 (OE1) — Gln150 (NE2)	X, Y, Z	2.22
Thr231 (OG1) — Ser256 (OG)	Y, X, -Z	2.91
Pro232 (O) — Ser256 (N)	Y, X, -Z	2.95
Tyr246 (OH) — Gln150 (NE2)	X, Y, Z	3.10

The symmetry contacts relate to Chain A

Coupling the analysis of the structural changes differences in B-values in the regions of motion between apo- and I2CA-HP0175 provides insight on the possible function of HP0175 upon binding to a substrate. In order to understand the structural differences, both spatially and in terms of local thermal motion, we examined the crystallographic contacts made by HP0175 in both structures (Tables 3.4-3.6). Although it is natural for proteins to be stabilized through crystal packing, no overly influential crystal contacts were observed in with the apo- or I2CA-bound HP0175 structures, thereby indicating that the differences observed are not a result of the crystalline condition. Therefore, the current analysis of the bound and unbound HP0175 provides a better understanding into the possible motions of the protein upon interacting to an inhibitor (I2CA-HP0175)[97] or other target molecule. It has been suggested that HP0175 binds to TLR4

`

through an immunoassay in Basak et al.[50] TLR4 is known to detect bacterial infections through the lipopolysaccharide complexed with MD-2 that interacts with the extracellular domain of TLR4.[133] This interaction may prove to be beneficial in combating *H. pylori* infections, as HP0175 could in theory be used as an adjuvant in vaccine design for eliciting an innate immune response.[134] However, more research must be performed into the interaction between HP0175 and TLR4 and its effects *in vitro* prior to development of such approaches.

3.7.4. Comparative Analysis of HP0175 to Other Parvulins

The parvulin subgroup of PPIases contain similar overall structures as observed in PEB4,[98] PrsA,[99] SurA[100] and HP0175 (Figure 3.7.8). SurA is a chaperone protein found in *Esherichia coli*, involved in the correct folding of outer membrane proteins (OMPs) and crystallizes with 4 monomers in the asymmetric unit.[100] PEB4, found in *Campylobacter jejuni*, is a periplasmic chaperone that also crystallizes with 1 monomer in the asymmetric unit and the dimer is generated through a symmetry-related monomer.[98] PrsA is from the gram-positive *Bacillus subtilis*, and is associated with the correct folding of periplasmic proteins requiring export; PrsA crystallizes as a dimer in the asymmetric unit.[99] These proteins all contain N- and C- terminal α-helices and the catalytic PPIase domain (4 α-helices and 3 β-strands). Dimerization is stabilized by intertwining of the N- and C- helices, as well as hydrogen bonding between β-strands of the N- and C-terminals; HP0175, PEB4 and PrsA (Figure 3.7.8). Although the overall structure of these proteins is relatively consistent covering a sequence identity between 10-34%, we also see that molecular replacement fails for obtaining phases, possibly attributed to the large structural alterations in the helices that give rise to an r.m.s.d over 7 Å.

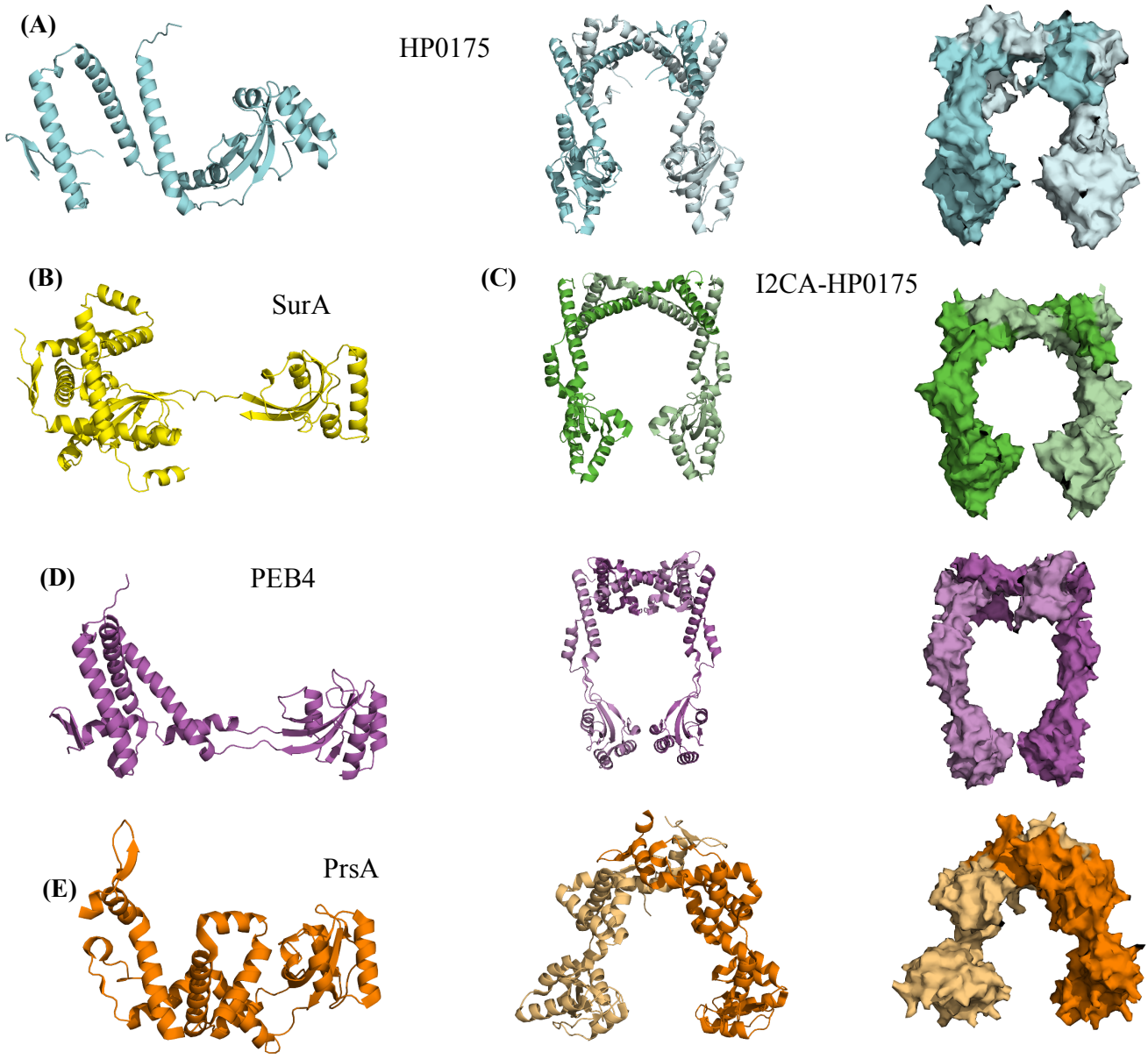


Figure 3.7.8. The parvulins share a common fold. (A) Apo-HP0175 contains a relatively compact overall structure, with crevice dimensions ~22 X 42 Å. **(B)** SurA the only parvulin represented with 2 PPIase domains and is demonstrated as a monomeric species. (PDB ID: 1M5Y)[100] **(C)** I2CA-HP0175 contains a crevice similar to apo-HP0175, however it is slightly larger with dimensions of 55 X 50.7 Å. (PDB ID: 5EZ1)[97] **(D)** PEB4, contains an elongated linker connecting to the PPIase domain, when observed as a dimer a large cavity is generated calculated to be ~55 X 50 Å. The inner region of crevice is more hydrophobic in comparison to the exposed regions. (PDB ID: 3RFW)[98] **(E)** PrsA, slightly more compact between PPIase and NC-domains. Inner region of crevice is comparably hydrophobic to the exterior, dimensions were determined as 20 X 30 X 35 Å. (PDB ID: 4WO7).[99]

`

The common dimeric conformation of the parvulins contains a crevice that has been demonstrated for protein-protein interaction.[98,100] Although SurA does not dimerize as seen with other parvulins, it maintains a small crevice through the N- and C- helices of the monomer. Another feature of parvulins was noted in PEB4 where it inhibited protein aggregation through isolation of unfolded proteins; however, this was not observed in PrsA and has not been heavily studied amongst all PPIases.[98,99] PrsA revealed NC-helix interactions with propeptide AmyE from α-amylase *B. subtilis* through NMR [^{15}N, ^{1}H]-TROSY spectra.[99] The different proteins assisted by PrsA through correct folding are involved in cell wall metabolism, motility and membrane bioenergetics. This feature of parvulins makes them interesting to study especially when chemical energy is not required for their function, making them a fascinating target for optimizing protein export in biotechnological applications.[99] In relation to the proteins presented here, HP0175 is the only parvulin crystallized in the apo and liganded form (I2CA-HP0175).[97] The two crystal structures demonstrate NC domain extensions upon targeted interaction, which infers NC domains may play a role in peptide stabilization, both in HP0175 and other parvulin PPIases.

`

Chapter 4: Conclusions and Future Work

4.1. Conclusions

4.1.1. Peptidyl-prolyl *cis,trans*-Isomerase HP0175

The secreted peptidyl-prolyl *cis,trans*-isomerase HP0175 from *Helicobacter pylori* (strain 26695) has been crystallized, and the structure solved to 2.09 Å resolution. Crystallization of HP0175 was optimized from 7.8 Å to 2.09 Å resolution, in 0.1 M HEPES: NaOH (pH 7.0) and 30% (v/v) Jeffamine ED 2001. Native/apo (ligand free) HP0175 crystallizes as one monomer in the asymmetric unit of the space group $P3_221$, unit cell dimensions a=b=90.95 Å, c=67.05 Å and angles 90˚ and 120˚. The structure was refined to R_{work}/R_{free} of 21.4 and 27.5 % to a completeness of 99.9% for the resolution range 45.48-2.09 Å. Apo-HP0175 was demonstrated as a dimer through symmetry generation (X-Y, -Y, -Z+1/3) in reference to (X, Y, Z). Mass spectrometry analysis for the molecular weight concluded HP0175 is relatively stable as a dimer in comparison to a wider mass-to-charge distribution for the monomer form. The overall structure of the HP0175 monomer resembles an asymmetric dumbbell, with N- and C- terminal helices and an extended PPIase domain. The dimer shows intertwining of NC domains for stabilization as well as H-bond stabilization through β-strands 1 (N-terminal) and 6 C-terminal of the symmetry generated monomer.

The current report shows the first crystallographic based interpretation of a parvulin peptidyl prolyl isomerase in its bound (to an inhibitor) in comparison to the un-bound 'native' conformation. Previously, interpretations for bound vs. unbound states were demonstrated using different tools such as NMR in PrsA with propeptide AmyE,[99] SAXS for I2CA-HP0175[97] and UV-VIS in PEB4.[98] The only documented crystallographic representation of bound and native form of peptidyl-prolyl isomerases was demonstrated in *E. coli* cytoplasmic cyclophilin A (CypA) in the native and bound (tri-peptide) conformations.[77] Structural comparison of apo-HP0175 and I2CA-HP0175 PPIase domain shows structural alterations directly associated with residues interacting with indole-2-carboxylic acid; Asp199, Met219, Leu211, Phe223, Phe244 and His247. The addition of I2CA ligand weakened the interaction between His161, His247 and Asp199, as Asp199 has an interaction to I2CA indole ring, and the sulfur group of Met219 interacts with the phenyl group of I2CA. The global implications of I2CA binding is found with N and C-terminal helices.

`

The N-terminal helix of α1 was measured to be extended by 6.6° and C-terminal helix (α8) was extended by 7.4° from apo to I2CA bound HP0175. The effect of these extensions has a direct effect on the crevice within the dimer. The crevice has extended from 53.0 Å (two; α3) and 40.5 Å (two; α8) to 68.1 Å (two; α3) and 47.8 Å (two; α8). While the PPIase domain becomes occluded upon ligand binding, 47.3 Å to 22.8 Å, which can be extended to include activity of peptide binding.

Since the structure of apo-HP0175 and I2CA-HP0175 were determined in enantiomeric space groups and similar resolutions, we analyzed thermal factors (B_{av}) associated with each domain in the two crystals in compliment to the observed structural differences. Only three domains demonstrated differences between the two crystal structures, those of α1 (aa 75-97), α2 (aa 104-132) and Loop L3 (aa 255-262). The N-terminal helices α1 and α2 support the extension of helices forming a rigid and locked configuration upon targeted interaction of the PPIase domain. The N-terminal and C-terminal helices are hypothesized to stabilize peptide-protein or protein-protein interaction, increasing the rigidity of the N-terminal helices supports this theory. While Loop L3, the linker between PPIase domain and C-terminal helix α8, shows a slight increase in B_{av} thermal factors, the increased thermal motion can be hypothesized to induce C-terminal α8 helix extension by 7.4° from apo- to I2CA-bound HP0175. This indicates that increased variability in Loop 3 may extend the C-terminal helix, and concurrently extend N-terminal helices as well. We also notice the occlusion of PPIase domains for further interaction. These structural rearrangements in HP0175 can facilitate protein-protein interaction for binding to TLR4 receptor, although this remains a focus of on-going study. Thermal factor analysis was performed with respect to crystal symmetry contacts present in both crystal structures, acknowledging that a fulsome understanding of the motion of a protein on a whole is not achieved with analyzing thermal factors alone, rather use of other methods such as NMR or HDX-MS can provide more detailed insights in this regard.

4.1.2. *P. aeruginosa* ΔKB7 and ΔP1 Pilins

The type IV pilins secreted by *P. aeruginosa* were successfully cloned into pGEX-4T2 plasmid. Expression and purification was optimized for GST-ΔKB7, however, the expression of

`

GST-ΔP1 did not provide pure protein. The expression of GST-ΔP1 produced for a leaky expression that was only later observed in the purification gel. Crystallization trials for GST-ΔKB7 provided for initial conditions in 2.0 M ammonium sulfate, 0.2 M sodium chloride, 0.1 M sodium cacodylate, pH 6.5, and in 3.2 M ammonium sulfate, 0.1 M citric acid, pH 4.0, final pH of 5.0. Diffraction analysis for one of the crystals (GST-ΔKB7) did not provide for a protein diffraction, instead represented ice-ring and small molecule diffraction.

4.2. Future Work

4.2.1. Peptidyl-Prolyl *cis,trans*-Isomerase: HP0175

The biological function of the chaperone (N- and C-) domains of parvulin PPIases have not been extensively studied, and therefore, it would be interesting to explore specific behavior of the following domains. Several ways to account for the behavior of the following domains is to crystallize HP0175 with extracellular domain of toll like receptor 4 and full-length, as this interaction is very important for eliciting immunogenic responses. The following can also be further explored utilizing time resolved hydrogen deuterium exchange mass spectrometry. This method can highlight the domains within TLR4 and HP0175 interaction. Further understanding of the protein can also include interaction studies of HP0175 with a protein that it would catalyze for proper folding. This interaction can be studied with mass spectrometry or can mutate several residues for suicide inhibition interaction and subsequently crystallize the protein with a peptide undergoing isomerization. The following interaction can also be used as a new mode for vaccine design, where domains of HP0175 is used for eliciting the innate immunity.

4.2.2. *P. aeruginosa*: ΔKB7 and ΔP1

Since cloning of GST-ΔKB7 was successful, there remained a reasonable amount of the protein found in the pellet after sonication. This could be a result of solubility issues, and therefore it would be best to clone the following pilin into a different construct. The ΔKB7 pilin could instead be cloned into a cytosolic expression of MBP, with a TEV site for ease of release of the tag, and a histidine tag for separation of MBP from ΔKB7 (ΔKB7-polyHis-TEV-MBP). As for the purification of ΔP1, it should be re-cloned into pGEX-4T2 plasmid, to ensure for a better construct

`

so that it is possible to crystallize with the tag. The ΔP1 pilin has previously been crystallized to 1.2 Å by Prof. Audette, but structure solution by molecular replacement has proven unsuccessful to date (GFA, personal communication). A recent approach has been taken by Üson et al., for solving the phase problem through *ab initio*, based on the tertiary structure of the protein and previously solved proteins available.[135,136] The following approach was not successful, and thus, was not represented in this thesis. Another method in which the phase can be solved for ΔP1 is through the mutation of all Leucine residues to Methionine residues.[137] The protein lacks methionine residues, and therefore, mutating the 6 Leucines in the sequence can provide for a phase. Once methionine is present in P1, selenomethionine derivative can be expressed and would provide for single/multi anomalous dispersion (SAD/MAD) phasing.

`

Chapter 5: References

(1) Krogfelt, K. . Bacterial Adhesion : Genetics , Biogenesis , and Role in Pathogenesis of Fimbrial Adhesins of Escherichia Coli Author. *Rev Infect Dis.* **1991**, *13* (4), 721–735.

(2) Chevance, F. F.; Hughes, K. T. Coordinating Assembly of a Bacterial Macromolecular Machine. *Nat Rev Microbiol.* **2008**, *6* (6), 455–465.

(3) Kirn, T. J.; Lafferty, M. J.; Sandoe, C. M.; Taylor, R. K. Delineation of Pilin Domains Required for Bacterial Association into Microcolonies and Intestinal Colonization by Vibrio Cholerae. *Mol Microbiol.* **2000**, *35* (4), 896–910.

(4) O'Toole, G.; Kaplan, H. B.; Kolter, R. Biofilm Formation as Microbial Development. *Annu Rev Microbiol.* **2000**, *54*, 49–79.

(5) Davies, D. G.; Parsek, M. R.; Pearson, J. P.; Iglewski, B. H.; Costerton, J. W.; Greenberg, E. P. The Involvement of Cell-to-Cell Signals in the Development of a Bacterial Biofilm. *Science.* **1998**, *280* (5361), 295–298.

(6) Fronzes, R.; Remaut, H.; Waksman, G. Architectures and Biogenesis of Non-Flagellar Protein Appendages in Gram-Negative Bacteria. *EMBO J.* **2008**, *27* (17), 2271–2280.

(7) Chen, I.; Dubnau, D. DNA Uptake During Bacterial Transformation. *Nat Rev Micro.* **2004**, *2* (3), 241–249.

(8) Craig, L.; Pique, M. E.; Tainer, J. A. Type IV Pilus Structure and Bacterial Pathogenicity. *Nat Rev Microbiol.* **2004**, *2* (5), 363–378.

(9) Craig, L.; Li, J. Type IV Pili: Paradoxes in Form and Function. *Curr Opin Struct Biol.* **2008**, *18* (2), 267–277.

(10) Bradley, D. E. A Function of Pseudomonas Aeruginosa PAO Polar Pili: Twitching Motility. *Can J Microbiol.* **1980**, *26* (2), 146–154.

(11) Sauer, F. G.; Mulvey, M. A.; Schilling, J. D.; Martinez, J. J.; Hultgren, S. J. Bacterial Pili: Molecular Mechanisms of Pathogenesis. *Curr Opin Struct Biol.* **2000**, *3* (1), 65–72.

(12) Telford, J. L.; Barocchi, M. a; Margarit, I.; Rappuoli, R.; Grandi, G. Pili in Gram-Positive Pathogens. *Nat Rev Microbiol.* **2006**, *4* (7), 509–519.

(13) Soto, G. E.; Hultgren, S. J. Bacterial Adhesins : Common Themes and Variations in Architecture and Assembly. *J Bacteriol.* **1999**, *181* (4), 1059–1071.

(14) Waksman, G.; Hultgren, S. J. Structural Biology of the Chaperone-Usher Pathway of Pilus Biogenesis. *Nat Rev Microbiol.* **2009**, *7* (11), 765–774.

(15) Busch, A.; Phan, G.; Waksman, G. Molecular Mechanism of Bacterial Type 1 and P Pili Assembly. *Philos Trans A Math Phys Eng Sci.* **2015**, *373* (2036).

(16) Wu, H.; Fives-taylor, P. M. Molecular Strategies for Fimbrial Expression and Assembly. *Crit Rev Oral Biol Med.* **2001**, *12* (2), 101-115.

(17) Giltner, C. L.; Nguyen, Y.; Burrows, L. L. Type IV Pilin Proteins : Versatile Molecular Modules. *Microbiol Mol Biol Rev.* **2012**, *76* (4), 740–772.

(18) Craig, L.; Taylor, R. K.; Pique, M. E.; Adair, B. D.; Arvai, A. S.; Singh, M.; Lloyd, S. J.; Shin, D. S.; Getzoff, E. D.; Yeager, M.; et al. Type IV Pilin Structure and Assembly: X-Ray and EM Analyses of Vibrio Cholerae Toxin-Coregulated Pilus and Pseudomonas Aeruginosa PAK Pilin. *Mol Cell.* **2003**. *11*(5),1139-1150.

(19) Parge, H. E.; Forest, K. T.; Hickey, M. J.; Christensen, D. A.; Getzoff, E. D.; Tainer, J. A. Structure of the Fibre-Forming Protein Pilin at 2.6 A Resolution. *Nature.* **1995**. *378* (6552),32-38.

`

(20) Hazes, B.; Sastry, P. A.; Hayakawa, K.; Read, R. J.; Irvin, R. T. Crystal Structure of Pseudomonas Aeruginosa PAK Pilin Suggests a Main-Chain-Dominated Mode of Receptor Binding. *J Mol Biol.* **2000**, *299* (4), 1005–1017.

(21) Audette, G. F.; Irvin, R. T.; Hazes, B. Crystallographic Analysis of the Pseudomonas Aeruginosa Strain K122-4 Monomeric Pilin Reveals a Conserved Receptor-Binding Architecture. *Biochemistry.* **2004**, *43* (36), 11427–11435.

(22) Schaik, E. J. van; Giltner, C. L.; Audette, G. F.; Keizer, D. W.; Bautista, D. L.; Slupsky, C. M.; Sykes, B. D.; Irvin, R. T. DNA Binding: A Novel Function of Pseudomonas Aeruginosa Type IV Pili. *J Bacteriol.* **2005**, *187* (4), 1455–1464.

(23) Yu, B.; Giltner, C. L.; Van Schaik, E. J.; Bautista, D. L.; Hodges, R. S.; Audette, G. F.; Li, D. Y.; Irvin, R. T. A Novel Biometallic Interface: High Affinity Tip-AssociatedBinding by Pilin-Derived Protein Nanotubes. *J Bionanosci.* **2007**, *1* (2), 73–83.

(24) Merz, A. J.; So, M.; Sheetz, M. P. Pilus Retraction Powers Bacterial Twitching Motility. *Nature.* **2000**, *407* (6800), 98–102.

(25) Gold, V.; Kudryashev, M. Recent Progress in Structure and Dynamics of Dual-Membrane-Spanning Bacterial Nanomachines. *Curr Opin Struct Biol.* **2016**, *39*, 1–7.

(26) Baker, J. L.; Biais, N.; Tama, F. Steered Molecular Dynamics Simulations of a Type IV Pilus Probe Initial Stages of a Force-Induced Conformational Transition. *PLoS Comput Biol.* **2013**, *9* (4), e1003032.

(27) Craig, L.; Pique, M. E.; Tainer, J. A. Type IV Pilus Structure and Bacterial Pathogenicity. *Nat Rev Microbiol.* **2004**, *2* (5), 363–378.

(28) Craig, L.; Volkmann, N.; Arvai, A. S.; Pique, M. E.; Yeager, M.; Egelman, E. H.; Tainer, J. A. Type IV Pilus Structure by Cryo-Electron Microscopy and Crystallography: Implications for Pilus Assembly and Functions. *Mol Cell.* **2006**, *23* (5), 651–662.

(29) Todd, T. R.; Franklin, A.; Mankinen-Irvin, P.; Gurman, G.; Irvin, R. T. Augmented Bacterial Adherence to Tracheal Epithelial Cells Is Associated with Gram-Negative Pneumonia in an Intensive Care Unit Population. *Am Rev Respir Dis.* **1989**, *140* (6), 1585–1589.

(30) Höffken, G.; Niederman, M. S. Nosocomial Pneumonia: The Importance of a De-Escalating Strategy for Antibiotic Treatment of Pneumonia in the ICU. *CHEST.* **2002**, *122* (6), 2183–2196.

(31) Giltner, C. L.; Van Schaik, E. J.; Audette, G. F.; Kao, D.; Hodges, R. S.; Hassett, D. J.; Irvin, R. T. The Pseudomonas Aeruginosa Type IV Pilin Receptor Binding Domain Functions as an Adhesin for Both Biotic and Abiotic Surfaces. *Mol Microbiol.* **2006**, *59* (4), 1083–1096.

(32) Maschmeyer, G.; Braveny, I. Review of the Incidence and Prognosis of Pseudomonas Aeruginosa Infections in Cancer Patients in the 1990s. *Eur J Clin Microbiol Infect Dis.* **2000**, *19* (12), 915–925.

(33) Da Silva, L. V. R. F.; Ferreira, F. de A.; Reis, F. J. C.; de Britto, M. C. A.; Levy, C. E.; Clark, O.; Ribeiro, J. D. Pseudomonas Aeruginosa Infection in Patients with Cystic Fibrosis: Scientific Evidence Regarding Clinical Impact, Diagnosis, and Treatment. *J Bras Pneumol.* **2013**, *39* (4), 495–512.

(34) Manfredi, R.; Nanetti, A.; Ferri, M.; Chiodo, F. Pseudomonas Spp. Complications in Patients with HIV Disease: An Eight-Year Clinical and Microbiological Survey. *Eur J Epidemiol.* **2000**, *16* (2), 111–118.

`

(35) Lee, K. K.; Sheth, H. B.; Wong, W. Y.; Sherburne, R.; Paranchych, W.; Hodges, R. S.; Lingwood, C. A.; Krivan, H.; Irvin, R. T. The Binding of Pseudomonas Aeruginosa Pili to Glycosphingolipids Is a Tip-Associated Event Involving the C-Terminal Region of the Structural Pilin Subunit. *Mol Microbiol.* **1994**, *11* (4), 705–713.

(36) Chang, Y.-W.; Rettberg, L. A.; Treuner-Lange, A.; Iwasa, J.; Søgaard-Andersen, L.; Jensen, G. J. Architecture of the Type IVa Pilus Machine. *Science.* **2016**, *351* (6278), aad2001.

(37) Wolfgang, M.; van Putten, J. P.; Hayes, S. F.; Dorward, D.; Koomey, M. Components and Dynamics of Fiber Formation Define a Ubiquitous Biogenesis Pathway for Bacterial Pili. *EMBO J.* **2000**, *19* (23), 6408–6418.

(38) Giltner, C. L.; van Schaik, E. J.; Audette, G. F.; Kao, D.; Hodges, R. S.; Hassett, D. J.; Irvin, R. T. The Pseudomonas Aeruginosa Type IV Pilin Receptor Binding Domain Functions as an Adhesin for Both Biotic and Abiotic Surfaces. *Mol Microbiol.* **2006**, *59* (4), 1083–1096.

(39) O'Toole, G. A; Kolter, R. Flagellar and Twitching Motility Are Necessary for *Pseudomonas Aeruginosa* Biofilm Development. *Mol Microbiol.* **1998**, *30* (2), 295–304.

(40) Campbell, A. P.; Wong, W. Y.; Houston Jr, M.; Schweizer, F.; Cachia, P. J.; Irvin, R. T.; Hindsgaul, O.; Hodges, R. S.; Sykes, B. D. Interaction of the Receptor Binding Domains of Pseudomonas Aeruginosa Pili Strains PAK, PAO, KB7 and P1 to a Cross-Reactive Antibody and Receptor Analog: Implications for Synthetic Vaccine Design. *J Mol Biol.* **1997**, *267*, 382–402.

(41) Campbell, A. P.; McInnes, C.; Hodges, R. S.; Sykes, B. D. Comparison of NMR Solution Structures of the Receptor Binding Domains of Pseudomonas Aeruginosa Pili Strains PAO, KB7, and PAK: Implications for Receptor Binding and Synthetic Vaccine Design. *Biochemistry.* **1995**, *34* (50), 16255–16268.

(42) McInnes, C.; Sönnichsen, F. D.; Kay, C. M.; Hodges, R. S.; Sykes, B. D. NMR Solution Structure and Flexibility of a Peptide Antigen Representing the Receptor Binding Domain of Pseudomonas Aeruginosa. *Biochemistry.* **1993**, *32* (49), 13432–13440.

(43) Wong, W. Y.; Campbell, A. P.; McInnes, C.; Sykes, B. D.; Paranchych, W.; Irvin, R. T.; Hodges, R. S. Structure-Function Analysis of the Adherence-Binding Domain on the Pilin of Pseudomonas Aeruginosa Strains PAK and KB7. *Biochemistry.* **1995**, *34* (40), 12963–12972.

(44) Audette, G. F.; van Schaik, E. J.; Hazes, B.; Irvin, R. T. DNA-Binding Protein Nanotubes: Learning from Nature's Nanotech Examples. *Nano Lett.* **2004**, *4* (10), 1897–1902.

(45) Petrov, A.; Lombardo, S.; Audette, G. F. Fibril-Mediated Oligomerization of Pilin-Derived Protein Nanotubes. *J Nanobiotechnology.* **2013**, *11*, 24.

(46) Fiorentino, M.; Ding, H.; Blanchard, T. G.; Czinn, S. J.; Sztein, M. B.; Fasano, A. Helicobacter Pylori-Induced Disruption of Monolayer Permeability and Proinflammatory Cytokine Secretion in Polarized Human Gastric Epithelial Cells. *Infect Immun.* **2013**, *81* (3), 876–883.

(47) Graham, D. Y. History of Helicobacter Pylori, Duodenal Ulcer, Gastric Ulcer and Gastric Cancer. *World J Gastroenterol.* **2014**, *20* (18), 5191–5204.

(48) Kohda, K.; Tanaka, K.; Aiba, Y.; Yasuda, M.; Miwa, T.; Koga, Y. Role of Apoptosis Induced by Helicobacter Pylori Infection in the Development of Duodenal Ulcer. *Gut* **1999**, *44* (4), 456–462.

`

(49) Rudi, J.; Kuck, D.; Strand, S.; von Herbay, A.; Mariani, S. M.; Krammer, P. H.; Galle, P. R.; Stremmel, W. Involvement of the CD95 (APO-1/Fas) Receptor and Ligand System in Helicobacter Pylori-Induced Gastric Epithelial Apoptosis. *J Clin Invest.* **1998**, *102* (8), 1506–1514.

(50) Basak, C.; Pathak, S. K.; Bhattacharyya, A.; Pathak, S.; Basu, J.; Kundu, M. The Secreted Peptidyl Prolyl Cis,Trans-Isomerase HP0175 of Helicobacter Pylori Induces Apoptosis of Gastric Epithelial Cells in a TLR4- and Apoptosis Signal-Regulating Kinase 1-Dependent Manner. *J Immunol.* **2005**, *174* (9), 5672–5680.

(51) Kuck, D.; Kolmerer, B.; Iking-Konert, C.; Krammer, P. H.; Stremmel, W.; Rudi, J. Vacuolating Cytotoxin of Helicobacter Pylori Induces Apoptosis in the Human Gastric Epithelial Cell Line AGS. *Infect Immun.* **2001**, *69* (8), 5080–5087.

(52) NIH Consensus Conference. Helicobacter Pylori in Peptic Ulcer Disease. NIH Consensus Development Panel on Helicobacter Pylori in Peptic Ulcer Disease. *JAMA*. **1994**, *272* (1), 65–69.

(53) Odenbreit, S.; Püls, J.; Sedlmaier, B.; Gerland, E.; Fischer, W.; Haas, R. Translocation of Helicobacter Pylori CagA into Gastric Epithelial Cells by Type IV Secretion. *Science.* **2000**, *287* (5457), 1497–1500.

(54) Basak, C.; Pathak, S. K.; Bhattacharyya, A.; Pathak, S.; Basu, J.; Kundu, M. The Secreted Peptidyl Prolyl Cis,Trans-Isomerase HP0175 of Helicobacter Pylori Induces Apoptosis of Gastric Epithelial Cells in a TLR4- and Apoptosis Signal-Regulating Kinase 1-Dependent Manner. *J Immunol.* **2005**, *174* (9), 5672–80.

(55) Sebkova, L.; Pellicanò, A.; Monteleone, G.; Grazioli, B.; Guarnieri, G.; Imeneo, M.; Pallone, F.; Luzza, F. Extracellular Signal-Regulated Protein Kinase Mediates Interleukin 17 (IL-17)-Induced IL-8 Secretion in Helicobacter Pylori-Infected Human Gastric Epithelial Cells. *Infect Immun.* **2004**, *72* (9), 5019–5026.

(56) Amedei, A.; Cappon, A.; Codolo, G.; Cabrelle, A.; Polenghi, A.; Benagiano, M.; Tasca, E.; Azzurri, A.; D'Elios, M. M.; Del Prete, G.; et al. The Neutrophil-Activating Protein of Helicobacter Pylori Promotes Th1 Immune Responses. *J Clin Invest.* **2006**, *116* (4), 1092–1101.

(57) Amedei, A.; Munari, F.; Della Bella, C.; Niccolai, E.; Benagiano, M.; Bencini, L.; Cianchi, F.; Silvestri, E.; D'Elios, S.; Farsi, M.; et al. Helicobacter Pylori HP0175 Promotes the Production of IL-23, IL-6, IL-1β and TGF-β. *Eur J Inflamm.* **2013**, *11* (1), 261–268.

(58) Miossec, P.; Korn, T.; Kuchroo, V. K. Interleukin-17 and Type 17 Helper T Cells. *N Engl J Med.* **2009**, *361* (9), 888–898.

(59) Amedei, A.; Munari, F.; Bella, C. D.; Niccolai, E.; Benagiano, M.; Bencini, L.; Cianchi, F.; Farsi, M.; Emmi, G.; Zanotti, G.; et al. Helicobacter Pylori Secreted Peptidyl Prolyl Cis, Trans-Isomerase Drives Th17 Inflammation in Gastric Adenocarcinoma. *Intern Emerg Med.* **2014**, *9* (3), 303–309.

(60) Drici, A. E.-M.; Moulessehoul, S.; Tifrit, A.; Diaf, M.; Turki, D. K.; Bachir, M.; Tou, A. Effect of IL-1β and IL-1RN Polymorphisms in Carcinogenesis of the Gastric Mucosa in Patients Infected with Helicobacter Pylori in Algeria. *Libyan J Med.* **2016**, *11*.

(61) El-Omar, E. M.; Carrington, M.; Chow, W. H.; McColl, K. E.; Bream, J. H.; Young, H. A.; Herrera, J.; Lissowska, J.; Yuan, C. C.; Rothman, N.; et al. Interleukin-1 Polymorphisms Associated with Increased Risk of Gastric Cancer. *Nature.* **2000**, *404* (6776), 398–402.

`

(62) Folkman, J. Angiogenesis. *Annu Rev Med.* **2006**, *57* (1), 1–18.
(63) Ferrara, N.; Davis-Smyth, T. The Biology of Vascular Endothelial Growth Factor. *Endocr Rev.* **1997**, *18* (1), 4–25.
(64) Caruso, R.; Pallone, F.; Monteleone, G. Emerging Role of IL-23/IL-17 Axis in H Pylori-Associated Pathology. *World J Gastroenterol.* **2007**, *13* (42), 5547–5551.
(65) Zhang, B.; Rong, G.; Wei, H.; Zhang, M.; Bi, J.; Ma, L.; Xue, X.; Wei, G.; Liu, X.; Fang, G. The Prevalence of Th17 Cells in Patients with Gastric Cancer. *Biochem Biophys Res Commun.* **2008**, *374* (3), 533–537.
(66) Wilson, K. T.; Crabtree, J. E. Immunology of Helicobacter Pylori: Insights into the Failure of the Immune Response and Perspectives on Vaccine Studies. *Gastroenterology.* **2007**, *133* (1), 288–308.
(67) Iwasaki, A.; Medzhitov, R. Toll-like Receptor Control of the Adaptive Immune Responses. *Nat Immunol.* **2004**, *5* (10), 987–995.
(68) Schmauber, B.; Andrulis, M.; Endrich, S.; Lee, S. K.; Josenhans, C.; Müller-Hermelink, H.-K.; Eck, M. Expression and Subcellular Distribution of Toll-like Receptors TLR4, TLR5 and TLR9 on the Gastric Epithelium in Helicobacter Pylori Infection. *Clin Exp Immunol.* **2004**, *136* (3), 521–526.
(69) Käbisch, R.; Mejías-Luque, R.; Gerhard, M.; Prinz, C. Involvement of Toll-Like Receptors on Helicobacter Pylori-Induced Immunity. *PLoS ONE.* **2014**, *9* (8), e104804.
(70) Schmid, F. X. Protein Folding. Prolyl Isomerases Join the Fold. *Curr Biol.* **1995**, *5* (9), 993–994.
(71) Weiss, M. S.; Jabs, A.; Hilgenfeld, R. Peptide Bonds Revisited. *Nat Struct Mol Biol.* **1998**, *5* (8), 676–676.
(72) Ünal, C. M.; Steinert, M. Microbial Peptidyl-Prolyl Cis/Trans Isomerases (PPIases): Virulence Factors and Potential Alternative Drug Targets. *Microbiol Mol Biol Rev.* **2014**, *78* (3), 544–571.
(73) Handschumacher, R. E.; Harding, M. W.; Rice, J.; Drugge, R. J.; Speicher, D. W. Cyclophilin: A Specific Cytosolic Binding Protein for Cyclosporin A. *Science.* **1984**, *226* (4674), 544–547.
(74) Rahfeld, J. U.; Rücknagel, K. P.; Schelbert, B.; Ludwig, B.; Hacker, J.; Mann, K.; Fischer, G. Confirmation of the Existence of a Third Family among Peptidyl-Prolyl Cis/Trans Isomerases. Amino Acid Sequence and Recombinant Production of Parvulin. *FEBS Lett.* **1994**, *352* (2), 180–184.
(75) Shaw, P. E. Peptidyl-Prolyl Isomerases: A New Twist to Transcription. *EMBO Rep.* **2002**, *3* (6), 521–526.
(76) Göthel, S. F.; Marahiel, M. A. Peptidyl-Prolyl Cis-Trans Isomerases, a Superfamily of Ubiquitous Folding Catalysts. *Cell Mol Life Sci.* **1999**, *55* (3), 423–436.
(77) Edwards, K. J.; Ollis, D. L.; Dixon, N. E. Crystal Structure of Cytoplasmic Escherichia Coli Peptidyl-Prolyl Isomerase: Evidence for Decreased Mobility of Loops upon Complexation11Edited by R. Huber. *J Molec Biol.* **1997**, *271* (2), 258–265.
(78) Gerard, M.; Deleersnijder, A.; Demeulemeester, J.; Debyser, Z.; Baekelandt, V. Unraveling the Role of Peptidyl-Prolyl Isomerases in Neurodegeneration. *Mol Neurobiol.* **2011**, *44* (1), 13–27.
(79) Wilson, K. P.; Yamashita, M. M.; Sintchak, M. D.; Rotstein, S. H.; Murcko, M. A.; Boger, J.; Thomson, J. A.; Fitzgibbon, M. J.; Black, J. R.; Navia, M. A. Comparative X-Ray Structures of the Major Binding Protein for the Immunosuppressant FK506 (Tacrolimus)

in Unliganded Form and in Complex with FK506 and Rapamycin. *Acta Crystallogr D Biol Crystallogr.* **1995**, *51* (4), 511–521.
(80) Castaño-Rodríguez, N.; Kaakoush, N. O.; Mitchell, H. M. Pattern-Recognition Receptors and Gastric Cancer. *Front Immunol.* **2014**, *5*.
(81) White, J. R.; Winter, J. A.; Robinson, K. Differential Inflammatory Response to Helicobacter Pylori Infection: Etiology and Clinical Outcomes. *J Inflamm Res.* **2015**, *8*, 137–147.
(82) Choi, I. J.; Kim, J. S.; Kim, J. M.; Jung, H. C.; Song, I. S. Effect of Inhibition of Extracellular Signal-Regulated Kinase 1 and 2 Pathway on Apoptosis and Bcl-2 Expression in Helicobacter Pylori-Infected AGS Cells. *Infect Immun.* **2003**, *71* (2), 830–837.
(83) Ichijo, H.; Nishida, E.; Irie, K.; Dijke, P. ten; Saitoh, M.; Moriguchi, T.; Takagi, M.; Matsumoto, K.; Miyazono, K.; Gotoh, Y. Induction of Apoptosis by ASK1, a Mammalian MAPKKK That Activates SAPK/JNK and P38 Signaling Pathways. *Science.* **1997**, *275* (5296), 90–94.
(84) Tobiume, K.; Matsuzawa, A.; Takahashi, T.; Nishitoh, H.; Morita, K.; Takeda, K.; Minowa, O.; Miyazono, K.; Noda, T.; Ichijo, H. ASK1 Is Required for Sustained Activations of JNK/P38 MAP Kinases and Apoptosis. *EMBO Rep.* **2001**, *2* (3), 222–228.
(85) Nishitoh, H.; Saitoh, M.; Mochida, Y.; Takeda, K.; Nakano, H.; Rothe, M.; Miyazono, K.; Ichijo, H. ASK1 Is Essential for JNK/SAPK Activation by TRAF2. *Mol Cell.* **1998**, *2* (3), 389–395.
(86) Jacobson, M. D.; Weil, M.; Raff, M. C. Programmed Cell Death in Animal Development. *Cell.* **1997**, *88* (3), 347–354.
(87) Chang, H. Y.; Yang, X. Proteases for Cell Suicide: Functions and Regulation of Caspases. *Microbiol Mol Biol Rev.* **2000**, *64* (4), 821–846.
(88) Thornberry, N. A.; Lazebnik, Y. Caspases: Enemies Within. *Science.* **1998**, *281* (5381), 1312–1316.
(89) Kuida, K. Caspase-9. *Int J Biochem Cell Biol.* **2000**, *32* (2), 121–124.
(90) Tang, D.; Kang, R.; Coyne, C. B.; Zeh, H. J.; Lotze, M. T. PAMPs and DAMPs: Signal 0s That Spur Autophagy and Immunity. *Immunol Rev.* **2012**, *249* (1), 158–175.
(91) Kim, H. M.; Park, B. S.; Kim, J. I.; Kim, S. E.; Lee, J.; Oh, S. C.; Enkhbayar, P.; Matsushima, N.; Lee, H.; Yoo, O. J.; et al. Crystal Structure of the TLR4-MD-2 Complex with Bound Endotoxin Antagonist Eritoran. *Cell.* **2007**, *130* (5), 906–917.
(92) Botos, I.; Segal, D.; Davies, D. The Structural Biology of Toll-like Receptors. *Structure.* **2011**, *19* (4), 447–459.
(93) Kim, H. M.; Park, B. S.; Kim, J. I.; Kim, S. E.; Lee, J.; Oh, S. C.; Enkhbayar, P.; Matsushima, N.; Lee, H.; Yoo, O. J.; et al. Crystal Structure of the TLR4-MD-2 Complex with Bound Endotoxin Antagonist Eritoran. *Cell.* **2007**, *130* (5), 906–917.
(94) Uno, K.; Kato, K.; Shimosegawa, T. Novel Role of Toll-like Receptors in Helicobacter Pylori - Induced Gastric Malignancy. *World J Gastroenterol.* **2014**, *20* (18), 5244–5251.
(95) Smith, S. M. Role of Toll-like Receptors in Helicobacter Pylori Infection and Immunity. *World J Gastrointest Pathophysiol.* **2014**, *5* (3), 133–146.
(96) Park, B. S.; Song, D. H.; Kim, H. M.; Choi, B.-S.; Lee, H.; Lee, J.-O. The Structural Basis of Lipopolysaccharide Recognition by the TLR4–MD-2 Complex. *Nature.* **2009**, *458* (7242), 1191–1195.

`

(97) Naveen, V.; Chu, C.; Chen, B.; Tsai, Y.; Hsiao, C.; Sun, Y. Helicobacter Pylori Cell Binding Factor 2 : Insights into Domain Motion. *J Struct Biol.* **2016**, *194* (1), 90–101.

(98) Kale, A.; Phansopa, C.; Suwannachart, C.; Craven, C. J.; Rafferty, J. B.; Kelly, D. J. The Virulence Factor PEB4 (Cj0596) and the Periplasmic Protein Cj1289 Are Two Structurally Related SurA-like Chaperones in the Human Pathogen Campylobacter Jejuni. *J Biol Chem.* **2011**, *286* (24), 21254–21265.

(99) Jakob, R. P.; Koch, J. R.; Burmann, B. M.; Schmidpeter, P. A. M.; Hunkeler, M.; Hiller, S.; Schmid, F. X.; Maier, T. Dimeric Structure of the Bacterial Extracellular Foldase PrsA. *J Biol Chem.* **2015**, *290* (6), 3278–3292.

(100) Bitto, E.; McKay, D. B. Crystallographic Structure of SurA, a Molecular Chaperone That Facilitates Folding of Outer Membrane Porins. *Structure.* **2002**, *10* (11), 1489–1498.

(101) Luft, J. R.; Wolfley, J. R.; Snell, E. H. What's in a Drop? Correlating Observations and Outcomes to Guide Macromolecular Crystallization Experiments. *Cryst Growth Des.* **2011**, *11* (3), 651–663.

(102) Hickman, A. B.; Davies, D. R. Principles of Macromolecular X-Ray Crystallography. *Current Protocols in Protein Science*; John Wiley & Sons, Inc. **2001**.

(103) Protein Crystallization. *Wikipedia*; https://en.wikipedia.org/wiki/Protein_crystallization#/media/File:CrystalDrops.svg (Accessed Jul 15, 2017)

(104) Oswald, C.; Smits, S. H. J.; Bremer, E.; Schmitt, L. Microseeding – A Powerful Tool for Crystallizing Proteins Complexed with Hydrolyzable Substrates. *Int J Mol Sci.* **2008**, *9* (7), 1131–1141.

(105) D'Arcy, A.; Bergfors, T.; Cowan-Jacob, S. W.; Marsh, M. Microseed Matrix Screening for Optimization in Protein Crystallization: What Have We Learned? *Acta Crystallogr F Struct Biol Commun.* **2014**, *70* (Pt 9), 1117–1126.

(106) Till, M.; Robson, A.; Byrne, M. J.; Nair, A. V.; Kolek, S. A.; Shaw Stewart, P. D.; Race, P. R. Improving the Success Rate of Protein Crystallization by Random Microseed Matrix Screening. *J Vis Exp.* **2013**, No. 78.

(107) Kuhn, P.; Wilson, K.; Patch, M. G.; Stevens, R. C. The Genesis of High-Throughput Structure-Based Drug Discovery Using Protein Crystallography. *Curr Opin Chem Biol.* **2002**, *6* (5), 704–710.

(108) Congreve, M.; Murray, C. W.; Blundell, T. L. Structural Biology and Drug Discovery. *Drug Discov Today.* **2005**, *10* (13), 895–907.

(109) Berman, H. M. Introduction to Macromolecular Crystallography: McPherson, A. *Biochem Mol Biol Educ.* **2003**, *31* (2), 155a–156.

(110) Evans, P.; McCoy, A. An Introduction to Molecular Replacement. *Acta Crystallogr D Biol Crystallogr.* **2008**, *64* (Pt 1), 1–10.

(111) Lyczak, J. B.; Cannon, C. L.; Pier, G. B. Establishment of Pseudomonas Aeruginosa Infection: Lessons from a Versatile Opportunist. *Microbes Infect.* **2000**, *2* (9), 1051–1060.

(112) Hancock, R. E. W.; Speert, D. P. Antibiotic Resistance in Pseudomonas Aeruginosa: Mechanisms and Impact on Treatment. *Drug Resist Updat.* **2000**, *3* (4), 247–255.

(113) Cloning - PCR Strategy - Primer Design - EMBL https://www.embl.de/pepcore/pepcore_services/cloning/pcr_strategy/primer_design/ (Accessed Nov 8, 2017).

`

(114) Erdogan, F.; Lento, C.; Yaseen, A.; Nowroozi-Dayeni, R.; Kheyson, S.; Audette, G. F. Conjugative Mating Assays for Sequence-Specific Analysis of Transfer Proteins Involved in Bacterial Conjugation. *J Vis Exp.* **2017**, No. 119.

(115) Harper, S.; Speicher, D. W. Purification of Proteins Fused to Glutathione S-Tranferase. *Methods Mol Biol.* **2011**, *681*, 259–280.

(116) Bornhorst, J.A. Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzymol.* **2000**, *324*,245-254.

(117) Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr.* **2010**, *66* (Pt 2), 125–132.

(118) Kabsch, W. Integration, Scaling, Space-Group Assignment and Post-Refinement. *Acta Crystallogr Sect D Biol Crystallogr.* **2010**, *66* (2), 133–144.

(119) Evans, P. Scaling and Assessment of Data Quality. *Acta Crystallogr Sect D Biol Crystallogr.* **2006**, *62* (1), 72–82.

(120) Evans, P. R. An Introduction to Data Reduction: Space-Group Determination, Scaling and Intensity Statistics. *Acta Crystallogr Sect D Biol Crystallogr.* **2011**, *67* (4), 282–292.

(121) Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E. A Graphical User Interface to the CCP4 Program Suite. *Acta Crystallogr D Biol Crystallogr.* **2003**, *59* (Pt 7), 1131–1137.

(122) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J Appl Crystallogr.* **2007**, *40* (Pt 4), 658–674.

(123) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; et al. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr Sect D Biol Crystallogr.* **2010**, *66* (2), 213–221.

(124) Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L.-W.; Read, R. J.; Adams, P. D. Iterative Model Building, Structure Refinement and Density Modification with the PHENIX AutoBuild Wizard. *Acta Crystallogr Sect D Biol Crystallogr.* **2008**, *64* (1), 61–69.

(125) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr Sect D Biol Crystallogr.* **2010**, *66* (4), 486–501.

(126) Schrödinger, L. *The PyMOL Molecular Graphics System*.

(127) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J Phys Chem.* **1984**, *88* (20), 4451–4459.

(128) Taylor, G. Disintegration of Water Drops in an Electric Field. *Philos Trans A Math Phys Eng Sci.* **1964**, *280* (1382), 383–397.

(129) Maltose Binding Protein Expression | NEB https://www.neb.com/applications/protein-expression-and-purification/coupled-protein-expression-and-purification/maltose-binding-protein-expression (Accessed Jan 29, 2018).

(130) Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by Calpha Geometry: Phi,Psi and Cbeta Deviation. *Proteins.* **2003**, *50* (3), 437–450.

(131) Petersen, T. N.; Brunak, S.; Heijne, G. von; Nielsen, H. SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nat Methods.* **2011**, *8* (10), 785–786.

(132) Schneider, B.; Gelly, J. C.; de Brevern, A. G.; Černý, J. Local Dynamics of Proteins and DNA Evaluated from Crystallographic B Factors. *Acta Crystallogr D Biol Crystallogr.* **2014**, *70* (Pt 9), 2413–2419.

`

(133) Ohto, U.; Yamakawa, N.; Akashi-Takamura, S.; Miyake, K.; Shimizu, T. Structural Analyses of Human Toll-like Receptor 4 Polymorphisms D299G and T399I. *J Biol Chem.* **2012**, *287* (48), 40611–40617.
(134) Vogel, F. R. Improving Vaccine Performance with Adjuvants. *Clin Infect Dis.* **2000**, *30* (Supplement_3), S266–S270.
(135) Rodríguez, D.; Sammito, M.; Meindl, K.; de Ilarduya, I. M.; Potratz, M.; Sheldrick, G. M.; Usón, I. Practical Structure Solution with ARCIMBOLDO. *Acta Crystallogr D Biol Crystallogr* **2012**, *68* (Pt 4), 336–343.
(136) Sammito, M.; Millán, C.; Rodríguez, D. D.; de Ilarduya, I. M.; Meindl, K.; De Marino, I.; Petrillo, G.; Buey, R. M.; de Pereda, J. M.; Zeth, K.; et al. Exploiting Tertiary Structure through Local Folds for Crystallographic Phasing. *Nature Methods* **2013**, *10* (11), 1099–1101.
(137) Heavy Atom Choices http://xray0.princeton.edu/~phil/Facility/heavyatompick.html (Accessed Jul 24, 2017).

Appendices

Appendix A

Relevant Amino Acid Sequences for Proteins in this Thesis

Truncated P1

```
        10         20         30         40         50         60
ISEFARTQVT RAVSEVSALK TAAESAILEG KEIVSSATPK DTQYDIGFTE STLLDGSGKS

        70         80         90        100        110        120
QIQVTDNKDG TVELVATLGK SSGSAIKGAV ITVSRKNDGV WNCKITKTPT AWKPNYAPAN
```

CPKS

Residues 124
Molecular weight= 13034.66 Da
Theoretical pI 8.59

Truncated KB7

```
        10         20         30         40         50         60
ISEFSRSQVS RVMAEAGSLK TAVEACLQDG RTAVGTAAGQ CDPGATGSSL LTGASQTSQT

        70         80         90        100        110        120
LPTNTGVPQV LDPLTTQTTI IVTFGNGASA AISGQTLTWT RDVNGGWSCA TTVDAKFRPN
```

GCTD

Residues 124
Molecular weight 12568.90 Da
Theoretical pI 4.77

`

Helicobacter pylori (HP0175) with N-terminal Histidine tag

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH MASMKKNILN LALVGALSTS FLMAKPAHNA NNATHNTKKT

        70         80         90        100        110        120
TDSSAGVLAT VDGRPITKSD FDMIKQRNPN FDFDKLKEKE KETLIDQAIR TALVENEAKT

       130        140        150        160        170        180
EKLDSTPEFK AMMEAVKKQA LVEFWAKKQA EEVKKVQIPE KEMQDFYNAN KDQLFVKQEA

       190        200        210        220        230        240
HARHILVKTE DEAKRIISEI DKQPKAKKEA KFIELANRDT IDPNSKNAQK WRDLGKFQKN

       250        260        270        280        290        300
QMAPDFSKAA FALTPGDYTK TPVKTEFGYH IIYLISKDSP VTYTYEQAKP TIKGMLQEKL

       310        320
FQERMNQRIE ELRKHAKIVI NK
```

Residues 322
Molecular weight 36756.13
Theoretical pI 9.45

Reference:

Gasteigner E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005), pp. 571-60.

`

Appendix B

Table A. Temperature factors for apo-HP0175

Label	Residue range	B_{av} (MC)	B_{av} (SC)	B_{av}
α1	75-97	59.01	64.31	61.95
α2	104-132	59.11	67.17	63.46
α3	137-151	64.92	73.28	69.44
α4	167-179	57.52	63.97	61.00
α5	182-196	56.84	62.84	60.27
α6	200-205	56.58	63.57	59.91
α7	221-229	59.71	63.47	61.54
α8	261-292	56.69	62.73	61.11
β1	45-48	58.33	59.84	59.18
β2	51-53	67.51	71.95	70.38
β3	156-164	52.07	54.26	53.23
β4	211-215	57.12	49.26	59.58
β5	246-254	52.02	52.98	52.56
β6	295-297	52.74	55.10	53.83
L1	99-103	65.73	69.13	67.64
L2	133-136	64.52	68.90	66.75
L3	152-155	56.40	58.37	57.40
L4	216-220	69.50	77.44	73.11
L5	230-245	59.81	57.59	62.68
L6	255-261	55.55	61.17	58.63

`

Table B. Temperature factors of I2CA-HP0175 chain A

Label	Residue range	B_{av} (MC)	B_{av} (SC)	B_{av}
α1	75-97	57.06	62.01	61.03
α2	104-132	41.36	55.65	50.06
α3	137-151	32.40	50.00	42.00
α4	167-179	34.96	46.94	41.78
α5	182-196	18.00	23.97	21.83
α6	200-205	19.75	36.53	28.00
α7	221-229	50.53	48.13	50.45
α8	261-292	37.36	48.11	44.25
β1	55-56	65.40	92.14	85.08
β2	59-60	82.19	107.05	94.62
β3	156-164	23.52	30.00	27.74
β4	211-215	30.26	36.08	35.95
β5	246-254	30.01	35.19	33.80
β6	295-296	84.59	48.16	67.50
L1	99-103	43.23	45.50	44.30
L2	133-136	54.90	60.46	58.34
L3	152-155	45.20	61.77	53.18
L4	216-220	58.73	67.12	64.21
L5	230-245	47.68	41.91	47.33
L6	255-261	61.31	55.97	59.41

`

Table C. Temperature factors I2CA-HP0175 chain B

Label	Residue range	B_{av} (MC)	B_{av} (SC)	B_{av}
α1	75-97	64.07	68.37	67.74
α2	104-132	47.15	61.62	55.52
α3	137-151	95.75	94.53	96.15
α4	167-179	85.22	88.12	87.73
α5	182-196	84.25	80.53	84.91
α6	200-205	87.37	70.76	80.93
α7	221-229	74.02	78.82	79.22
α8	261-292	64.93	73.46	71.75
β1	55-56	75.20	82.45	81.27
β2	59-60	88.48	128.12	108.30
β3	156-164	75.81	70.34	75.46
β4	211-215	69.90	58.27	72.37
β5	246-254	64.35	60.44	62.77
β6	295-297	74.94	63.68	69.40
L1	99-103	48.61	47.75	49.10
L2	133-136	100.11	85.33	93.32
L3	152-155	91.55	76.55	85.26
L4	216-220	89.01	80.53	85.64
L5	230-245	76.28	70.41	78.87
L6	255-261	82.14	80.56	82.08