A University of Sussex DPhil thesis

Available online via Sussex Research Online:

http://sro.sussex.ac.uk/

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details
Biophysical Characterisation of LcrH, a Class II Chaperone of the Type III Secretion System

Sunny Kumar Singh

A thesis submitted for the degree of Doctor of Philosophy (Chemistry) at the University of Sussex

September 2015
Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signed:

Sunny Kumar Singh, 18th September 2015
In memory of my Grandfather, Grandmother and Father
“It does not matter how slowly you go as long as you do not stop.”

- Confucius
Acknowledgements

First and foremost I thank my supervisor Prof Louise Serpell. I could not have come this far without your support and kindness. I am grateful to the University of Sussex for funding this project. I do not think I can thank Dr Ewan Main enough. Not because of the standard polite gratitude that needs to be given to PhD supervisors in acknowledgements but because I truly am indebted to you. You provided me with the opportunity to work, helped me get through the tough times of recurring illnesses and had the patience of letting me stick around for what now seems like centuries. I also acknowledge my co-supervisor Dr Darren Thompson for his support, laboratory equipment and amusing anecdotes.

I thank Prof Andrew Smith and his group in former lab 13, Chemistry, University of Sussex especially Dr Wendy Doyle for her advice on molecular biology which was instrumental in kick-starting my project. Additionally, I thank Dr Yalda Javadi, Dr Maria Karatsa-Dodgson and Dr Sarah Pannell for education on instrumentation, techniques and inspiration.

I have cherished the time spent both inside the lab and out with my dear friends Dr Jonathan (JJ) Phillips and Dr. Charlotte Millership both in Sussex and in London. I will never forget our culinary gatherings.

An enormous thanks to Dr Ruth-Sarah Rose for her support during my time at QMUL. She trained me for using various instruments and provided invaluable help in the composition of my thesis. Without her guidance I would not have accomplished this. I am grateful to Kelly Clark largely for her hospitality but also for the initiation into British culture.

I thank the members past and present of office 4.36 in G.E. Fogg building at QMUL especially Dr Erica Belgio, Anna Yates, Dr Helen Marks, Dr Tatiana Novoselova and Dr Christopher Duffy. There are many others, too numerous to mention but your contribution and support is no less significant. Also, the members of G35 in the Joseph Priestly Building at QMUL, in particular Dr Nadine Younan for her expertise regarding CD.
I am deeply indebted to my friends Inderjeet, Souvik, Anirban, Srijit, Mahesh and Tina. You made me feel at home away from home. Your companionship gave me the courage to fight against all odds and crawl out of deep despair. I would not have survived without you all.

Without my late Grandfather, this journey would not have been possible. It was his encouragement and sacrifice that allowed me to come to the UK and pursue my passion. I am overwhelmed by the selfless love from my mother who has taken care of me single-handedly in the absence of my father. I have no words to thank my little brother who has taken care of my family in Mumbai in my absence and has helped me pay all of my debts. The love from my cousins and extended family has continued to feed my enthusiasm for my career in science. No one has ever pursued science in my family so I thank you for your support, even if it does make it so much harder for me to explain what I’ve done. I hope I have made you proud.
Summary

The type three-secretion system (T3SS) is a large and complex protein nano-machine that many gram-negative pathogens employ to infect host cells. A key structure of this machine is a proteinaceous pore that inserts into the target membrane and forms a channel for bacterial toxins to flow from bacteria into the host cell. The pore is mainly formed from two large membrane proteins called “translocators”. Importantly, effective secretion and thus pore formation of the translocators depends on their binding to and being transported by small specialized chaperones after synthesis in the bacterial cytosol. Recent crystal structures have shown these chaperones are formed from modular tetratricopeptide repeats (TPRs). However, each crystal structure produced different homodimeric structures, suggesting flexibility in their topology that may be of importance to function.

Given the crucial role of the translocator chaperones, we investigated the conformational stability of the chaperone LcrH (*Yersinia pestis*). Mutational analysis coupled with analytical ultra-centrifugation and equilibrium chemical denaturations showed that LcrH is a weak and thermodynamically unstable dimer (\(K_\text{D} \approx 15 \mu\text{M}, \Delta G_{\text{H}_2\text{O}} = 7.4 \text{ kcal/mol}\)). The modular TPR structure of the dimer allows it to readily unfold in a non-cooperative manner to a one-third unfolded dimeric intermediate (\(\Delta G_{\text{H}_2\text{O}} = 1.7 \text{ kcal/mol}\)), before cooperatively unfolding to a monomeric denatured state (\(\Delta G_{\text{H}_2\text{O}} = 5.7 \text{ kcal/mol}\)). Thus under physiological conditions the chaperone is able to populate C-terminally unravelled partially folded states, whilst being held together by its dimeric interface. Such ability suggests a “fly-casting” mechanism as a route to binding their far larger translocator cargo.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>ATP binding cassette</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Ampicillin resistance</td>
</tr>
<tr>
<td>A. tumefaciens</td>
<td>Agrobacterium tumefaciens</td>
</tr>
<tr>
<td>AUC</td>
<td>Analytical ultracentrifugation</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>CBD</td>
<td>Chaperone binding domain</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>cm</td>
<td>Centimetre</td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic light scattering</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>D&lt;sup&gt;-&lt;/sup&gt;L-Dithiothreitol</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetate</td>
</tr>
<tr>
<td>fl-LcrH</td>
<td>Full length LcrH</td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast purification liquid chromatography</td>
</tr>
<tr>
<td>g</td>
<td>Grams</td>
</tr>
<tr>
<td>Gdm</td>
<td>Guanidinium</td>
</tr>
<tr>
<td>GPS</td>
<td>General secretory pathway</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>Hcp</td>
<td>Haemolysin coregulated protein</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IM</td>
<td>Inner membrane</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calorimetry</td>
</tr>
<tr>
<td>kb</td>
<td>Kilo base</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani broth</td>
</tr>
<tr>
<td>lb/in²</td>
<td>Pounds per square inch</td>
</tr>
<tr>
<td>LcrH</td>
<td>Low calcium response protein H</td>
</tr>
<tr>
<td>L65E tr-LcrH</td>
<td>Truncated LcrH carrying the L65E mutation</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>mdeg</td>
<td>Millidegree</td>
</tr>
<tr>
<td>MFP</td>
<td>Membrane fusion protein</td>
</tr>
<tr>
<td>MS</td>
<td>Inner membrane ring (T3SS)</td>
</tr>
<tr>
<td>OM</td>
<td>Outer membrane</td>
</tr>
<tr>
<td>OMP</td>
<td>Outer membrane pore forming protein</td>
</tr>
<tr>
<td><em>P. aeruginosa</em></td>
<td><em>Pseudomonas aeruginosa</em></td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PP</td>
<td>Periplasm</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>SRP</td>
<td>Signal recognition particle</td>
</tr>
<tr>
<td>T1SS</td>
<td>Type I Secretion System</td>
</tr>
<tr>
<td>T2SS</td>
<td>Type II Secretion System</td>
</tr>
<tr>
<td>T3SS</td>
<td>Type III Secretion System</td>
</tr>
<tr>
<td>T4SS</td>
<td>Type IV Secretion System</td>
</tr>
<tr>
<td>T5SS</td>
<td>Type V Secretion System</td>
</tr>
<tr>
<td>T6SS</td>
<td>Type VI Secretion System</td>
</tr>
<tr>
<td>Tat</td>
<td>Twin-arginine translocation</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>TPR</td>
<td>Tetratricopeptide repeat</td>
</tr>
<tr>
<td>Tris</td>
<td>Trizma base</td>
</tr>
<tr>
<td>tr-LcrH</td>
<td>truncated LcrH</td>
</tr>
<tr>
<td>T&lt;sub&gt;m&lt;/sub&gt;</td>
<td>Melting temperature</td>
</tr>
<tr>
<td>UV-Vis</td>
<td>Ultraviolet-visible</td>
</tr>
<tr>
<td>V. cholera</td>
<td>Vibrio cholera</td>
</tr>
<tr>
<td>Y. enterocolitica</td>
<td>Yersinia enterocolitica</td>
</tr>
<tr>
<td>Y. pestis</td>
<td>Yersinia pestis</td>
</tr>
</tbody>
</table>
Table of Contents

CHAPTER 1  INTRODUCTION........................................................................................................16

1.1 SYMBIOSIS, PATHOGENESIS & COMMENSALISM ............................................................... 16
1.2 GRAM POSITIVE AND GRAM NEGATIVE BACTERIA ......................................................... 17
1.3 PROTEIN SECRETION BY BACTERIA.................................................................................... 18
  1.3.1 SECRETION THROUGH THE INNER MEMBRANE....................................................... 19
    1.3.1.1 The Sec Pathway ........................................................................................................ 19
    1.3.1.2 Signal Recognition Particle (SRP) ........................................................................... 20
    1.3.1.3 The Tat System ......................................................................................................... 20
  1.3.2 GRAM-NEGATIVE BACTERIAL SECRETION SYSTEMS ................................................ 20
    1.3.2.1 Type I Secretion System (T1SS) .......................................................................... 21
    1.3.2.2 Type II Secretion System (T2SS) .......................................................................... 22
    1.3.2.3 Type III Secretion System (T3SS) .......................................................................... 24
    1.3.2.4 Type IV Secretion System (T4SS) .......................................................................... 25
    1.3.2.5 Type V Secretion System (T5SS) .......................................................................... 26
    1.3.2.6 Type VI Secretion System ..................................................................................... 28
  1.3.3 TYPE THREE SECRETION SYSTEM IN DETAIL .......................................................... 31
    1.3.3.1 The role of Chaperones in the T3SS .................................................................... 40
    1.3.3.2 Class I Chaperones ............................................................................................... 40
    1.3.3.3 Class II & III Chaperones ..................................................................................... 44
    1.3.3.4 Structure and biophysical characterisation of Class II Chaperones ...................... 48
    1.3.3.5 Structure and biophysical characterisation of Class III Chaperones .................... 53

1.4 THESIS AIMS ..................................................................................................................... 56

CHAPTER 2  MATERIALS AND METHODS ........................................................................... 57

2.1 MOLECULAR BIOLOGY METHODS ...................................................................................... 57
  2.1.1 FL-LCRH GENE SYNTHESIS ....................................................................................... 57
    2.1.1.1 Recursive PCR ........................................................................................................ 57
    2.1.1.1.1 Agarose Gel Electrophoresis .............................................................................. 58
  2.1.2 POLYMERASE CHAIN REACTION (PCR) ..................................................................... 58
    2.1.2.1 Cloning of l-LcrH gene into TOPO TA cloning vector .......................................... 59
    2.1.2.1.1 Poly-A tailing with Taq polymerase ..................................................................... 59
    2.1.2.1.2 Ligation for TOPO TA cloning ......................................................................... 60
2.1.2.1.3 Transformation of Ligated fl-LcrH in chemically competent cells............................60
2.1.2.1.4 DNA extraction and purification ........................................................................61
2.1.2.1.5 Estimation of DNA concentration .......................................................................61
2.1.2.1.6 Analytical Restriction Digestion with EcoRI ......................................................62
2.1.2.1.7 DNA sequencing ...............................................................................................62
2.1.3 INSERTION OF SYNTHESISED fl-LCRH GENE INTO AN EXPRESSION VECTOR ........62
2.1.3.1 Restriction digest of fl-LcrH gene .........................................................................63
2.1.3.2 Purification of DNA from an Agarose Gel ..............................................................64
2.1.3.3 ligation ................................................................................................................64
2.1.3.4 Transformation using Electro-competent cells ......................................................64
2.1.3.5 Confirmation of DNA Clones ................................................................................65
2.1.4 SYNTHESIS & INSERTION OF TRUNCATED LCRH (TR-LCRH) GENE IN INTO pTRC-GST VECTOR .......65
2.1.5 SYNTHESIS OF LE65 MUTANT OF TR-LCRH ......................................................66
2.1.5.1 Site Directed Mutagenesis ..................................................................................66
2.2 PROTEIN PRODUCTION AND PURIFICATION.........................................................67
2.2.1 PROTEIN PRODUCTION .......................................................................................67
2.2.1.1 Cell Lysis for Protein Extraction .......................................................................68
2.2.1.1.1 Cell Lysis by Sonication ................................................................................68
2.2.1.1.2 Cell Lysis by Homogenisation ......................................................................68
2.2.1.2 Purification by Affinity Chromatography .........................................................68
2.2.1.3 Buffer Solutions ................................................................................................69

NAME OF BUFFER ........................................................................................................69

HIS-TAG WASH BUFFER ..........................................................................................69

HIS-TAG ELUTION BUFFER .......................................................................................69

GST WASH BUFFER ..................................................................................................69

GST ELUTION BUFFER .............................................................................................69

2.2.1.4 Ni²⁺ Affinity chromatography and cleavage of His-tagged proteins ....................69
2.2.1.5 GST Affinity purification and cleavage of GST-tagged proteins ........................70
2.2.1.6 Purification using Size Exclusion Chromatography by AKTA FPLC ..................70
2.2.2 PROTEIN PURITY ...............................................................................................70
2.2.2.1 Mass Spectrometry .........................................................................................71
SAMPLE AND MATRIX PREPARATION

2.2.2.2 SDS-Polyacrylamide Gel Electrophoresis ......................................................... 71
2.2.3 DETERMINATION OF PROTEIN CONCENTRATION.................................................. 71
2.2.4 STORAGE OF PURIFIED PROTEIN ........................................................................ 72

2.3 BIOPHYSICAL CHARACTERIZATION OF LCRH PROTEINS ........................................... 72

2.3.1 MULTIMERIC STATE OF LCRH PROTEINS USING ANALYTICAL SIZE EXCLUSION CHROMATOGRAPHY (SEC) ................................................................. 72

2.3.2 ANALYTICAL ULTRACENTRIFUGATION (AUC) ......................................................... 73
2.3.3 DYNAMIC LIGHT SCATTERING (DLS) ...................................................................... 75

2.4 SECONDARY AND TERTIARY STRUCTURE CHARACTERIZATION ....................................... 76

2.4.1 CIRCULAR DICHROISM (CD) SPECTROSCOPY .......................................................... 76
2.4.2 FLUORESCENCE SPECTROSCOPY ........................................................................ 77

2.5 EQUILIBRIUM STUDIES FOR MEASURING THERMODYNAMIC STABILITY ...................... 78

2.5.1 GENERAL BACKGROUND ....................................................................................... 78
2.5.2 EQUIPMENT AND GENERAL PROCEDURES ............................................................ 79
2.5.3 PREPARATION OF SAMPLES FOR EQUILIBRIUM CHEMICAL DENATURATION .................. 79

2.5.1.1 Sample preparation for LcrH protein concentrations 1 to 12 µM ................................. 80
2.5.1.2 Sample preparation for LcrH protein concentrations 25 to 80 µM .............................. 80

CHAPTER 3 GENE CONSTRUCTION, PURIFICATION STRATEGY AND INITIAL BIOPHYSICAL CHARACTERISATION OF LCRH PROTEINS ........................................................................ 81

3.1 INTRODUCTION ............................................................................................................ 81
3.2 EXPERIMENTAL APPROACHES .................................................................................. 81

3.3 RESULTS ...................................................................................................................... 82

3.3.1 FULL LENGTH LCRH (FL-LCRH) GENE CONSTRUCTION ........................................... 82

3.3.2 PRODUCTION AND PURIFICATION OF FULL LENGTH LCRH (FL-LCRH) ...................... 83

3.3.3 TRUNCATED LCRH (TR-LCRH) ............................................................................... 88

3.3.4 EXPRESSION & PURIFICATION OF TRUNCATED LCRH (TR-LCRH) ............................. 91

3.3.5 CONSTRUCTION OF L65E TR-LCRH ....................................................................... 91

3.3.6 EXPRESSION AND PURIFICATION OF L65E TR-LCRH ................................................... 93

3.3.7 INITIAL CHARACTERIZATION OF LCRH CONSTRUCTS ............................................ 94

3.3.7.1 Defining buffer & protein concentrations to maintain “native” oligomeric states of LcrH constructs ......................................................................................... 94
3.3.7.2 Far-UV circular dichroism of native LcrH constructs.................................................................97
3.3.8 CONCLUSION ..................................................................................................................................98

CHAPTER 4  OLIGOMERIC STATE, FLEXIBILITY AND THERMODYNAMIC STABILITY OF LCRH
PROTEIN CONSTRUCTS..........................................................................................................................99
4.1 INTRODUCTION ......................................................................................................................................99
4.2 EXPERIMENTAL APPROACHES .........................................................................................................99
4.3 RESULTS ................................................................................................................................................99
4.3.1 OLIGOMERIC STATES OF LCRH USING ANALYTICAL SIZE EXCLUSION CHROMATOGRAPHY (SEC) ....99
4.3.2 DETERMINATION OF DIMERIZATION DISSOCIATION CONSTANT (Kd) USING ANALYTICAL
ULTRACENTRIFUGATION (AUC) ..............................................................................................................102
4.3.2.1 Calculating percentage of dimer at a given protein concentration .................................................103
4.3.3 LOSS OF LCRH’S A-HELICAL STRUCTURE ON N-TERMINAL DELETION & CONCENTRATION REDUCTION
.................................................................................................................................................................106
4.3.4 EQUILIBRIUM UNFOLDING AND THERMODYNAMIC STABILITY OF LCRH ....................................107
4.3.4.1 The Best Structural Probe to Monitor Chemical Denaturation ......................................................107
4.3.4.2 Determining the most suitable chemical denaturant .................................................................111
4.3.4.3 Minimum Equilibration Time for Equilibrium Chemical Denaturation .........................................114
4.3.4.4 The Reversibility of Equilibrium Unfolding ..................................................................................115
4.3.5 EQUILIBRIUM CHEMICAL DENATURATION OF LCRH’S PROTEIN CONSTRUCTS .........................116
4.3.5.1 Biphasic Denaturations Show LcrH Unfolds via Partially Folded Dimeric Intermediate .............118
4.3.5.2 Deletion of N-terminus Causes No Change in Equilibrium Unfolding .......................................122
4.3.6 THERMODYNAMIC STABILITY OF LCRH ......................................................................................122
4.3.6.1 Data analysis .................................................................................................................................123
4.3.6.2 Proposed equilibrium unfolding pathway: ......................................................................................126
4.3.6.3 Thermodynamic Stability of Monomeric vs Dimeric LcrH ...........................................................127
4.3.7 CONCLUSION ...................................................................................................................................129

CHAPTER 5  CONCLUSIONS .......................................................................................................................130
5.1 FUTURE WORK ....................................................................................................................................133

CHAPTER 6  BIBLIOGRAPHY ....................................................................................................................134
Chapter 1 Introduction

Life on earth can be categorised in numerous ways. One method classifies the diverse array of organisms on our planet through their biochemical characteristics. This method divides organisms into three distinct groups called domains:

(i) **Eukarya** (eukaryotes) – Eukarya consist of all macroscopic organisms including human beings as well as many microscopic, unicellular organisms such as yeast. The most important characteristic of *eukaryotes* is the presence of a well-defined nucleus within each cell.

(ii) **Bacteria** – These are unicellular organisms lacking a nucleus, and are referred to as *prokaryotes*. They constitute a large domain of prokaryotic microorganisms. Bacteria were among the first life forms to appear on Earth, and are present in most of its habitats.

(iii) **Archaea** – Carl Woese first discovered these organisms in 1977 and classified them as prokaryotes as they were bacteria-like. However, they are biochemically quite distinct from bacteria. These distinctions include: archaea have cell membrane containing ether linkages, cell walls that lack peptidoglycan and have genes and enzymes that behave more like eukaryotes. Thus they were reclassified as a domain of their own [1-3].

This thesis is concerned with the biochemical methods employed by certain bacteria to survive in and around higher multicellular organisms (humans).

1.1 Symbiosis, Pathogenesis & Commensalism

There are three distinct types of relationship between bacteria and higher multicellular organisms. They can live in symbiosis, as commensals or as pathogens. Symbiosis occurs when two organisms obtain a mutually beneficial relationship with one another [4-6]. For example, many bacteria are vital for the maintenance of homeostasis in higher organisms. The human gut hosts an estimated 500-1000 species of bacteria [1,3,7,8]. Here, the human intestine provides nutrients to the resident bacteria, whereas bacteria help in the digestion of food, absorption of nutrients, production of vitamins such as biotin and vitamin K, regulation of the immune system, and prevention of colonization of pathogenic microorganisms [4,6,9].
In contrast, pathogenesis occurs when one organism gains advantage to the detriment of the other. Bacteria are another excellent example of this interaction, with many attacking higher organisms by damaging cell tissues in the search for nutrients. For example, *Yersinia pestis* is the causative agent of bubonic plague and many *Shigella* species cause dysentery [7,8,10].

In between the two extremes of symbiosis and pathogenesis lie commensals. Commensalism is an association between two different species where one enjoys benefit and the other is not significantly affected [9,11]. Commensal bacteria are found on all animal body surfaces as part of normal microbial flora. However, under specific conditions such commensals can turn pathogenic and exert pathologic effects on the host organism [10,12]. For example, *Staphylococcus aureus* normally colonises human skin without causing any harm. However if the epidermis of the skin is broken they can cause skin infection and more serious complaints. Recently, there has been an emergence of antibiotic resistant forms of *S. aureus* (MRSA: Methicillin resistant *S. aureus*), making it a significant health risk [11,13].

### 1.2 Gram Positive and Gram Negative Bacteria

Bacteria can be separated into two groups on the basis of their Gram staining: Gram-positive and Gram-negative (Figure 1.1). This simple staining process has lasted into our high-technology era as a useful tool for identifying bacteria. The Danish physician Hans Christian Gram developed it in 1884. A smear of bacterial cells on a microscope slide is soaked in a violet dye and treated with iodine; it is then washed with alcohol and counterstained with saffranine (a red dye). Gram-positive bacteria retain the violet dye and appear blue to purple. The alcohol washes the violet stain out of Gram-negative cells; these cells then pick up the saffranine counterstain and appear pink to red. For many bacteria, the Gram-staining results compare roughly with the cell wall structure. In Gram-positive bacteria, peptidoglycan forms a thick layer outside the plasma membrane. Whereas, the Gram-negative cell wall has only one-fifth as much peptidoglycan. In order to provide additional protection for Gram-negative bacteria, outside the peptidoglycan layer, a second, outer membrane quite distinct in chemical
makeup from the plasma membrane surrounds the cell. Between the inner and outer membranes of Gram-negative bacteria is an additional periplasmic space. This space contains enzymes important in digesting specific molecules, transporting others, and detecting chemical gradients in the environment [2,12].

Interestingly, common causes of infections are Gram-negative bacterial pathogens. The prevalence and the rate of resistance among these pathogens to existing antimicrobial agents is on the rise [12,13]. The pathogenicity of gram-negative bacteria is often associated with the lipopolysaccharide (LPS) layer present in the outer membrane. The chemical structure of the LPS layer is often unique to particular bacterial strains (i.e. sub-species) and is responsible for many of the antigenic properties of these strains [2,14].

![Figure 1.1 Gram staining and the bacterial cell wall: when treated with Gram stain, the cell wall components of different bacteria react in one of two ways. (a) Gram positive bacteria have a thick peptidoglycan cell wall that retains the violet dye and appears blue or purple. (b) Gram negative bacteria have a thin peptidoglycan layer that does not retain the violet dye, but picks up the counterstain and appears pink-red. Figure was taken from [12,15].](image)

### 1.3 Protein secretion by bacteria

The cytoplasm of the bacterial cell is isolated from the external environment by a phospholipid bilayer termed as the plasma membrane. In order to interact with the environment and for its sustenance it becomes essential for the bacterium to export
elements from its cytosol to the external milieu and \textit{vice versa}. Until the 1960s, protein secretion across the bacterial membrane was thought to be a rare phenomenon. It was believed that secretion occurred in a protein specific manner [14,16]. This view was dismissed following the discovery of numerous systems dedicated to the transport of proteins across the plasma membrane. Furthermore, in the case of Gram-negative bacteria, proteins are transported through the periplasm and the outer membrane. These secretory systems export proteins either to be attached outside the bacterial cell surface, released into the external environment, or in some cases even directly into the cytoplasm of other cells.

1.3.1 \textbf{Secretion through the inner membrane}

The first challenge faced by a bacterial cell is the transportation of protein molecules across the inner plasma membrane, irrespective of whether it a Gram-positive or Gram-negative bacteria. Protein molecules travel across this barrier with the help of membrane embedded protein transport systems. These proteins then enter the periplasm or they get inserted into the membrane. There is a remarkable array of protein transport systems found in bacteria, of which three are present in most bacterial species: The Sec (general secretory pathway or GSP), signal recognition particle (SRP) and twin-arginine translocation (Tat) pathways [15,17].

1.3.1.1 \textbf{The Sec Pathway}

The Sec pathway is formed by the interaction of a numerous proteins, which are conserved both in prokaryotes and eukaryotes (known as the Sec 61 complex, where it transports proteins across the endoplasmic reticulum [16,18]. In particular, the Sec system is involved in the secretion of unfolded proteins across the cytoplasmic membrane and the insertion of membrane proteins into the cytoplasmic membrane [17,19]. To perform the transport, three proteins SecY, SecE & SecG (SecYEG) form a heterotrimeric integral membrane complex. SecA is then recruited to the complex, by binding to SecY on the bacterial cytoplasmic side. SecA has ATPase activity and acts as a motor to drive secretion [18,20]. SecB binds the protein requiring transport and delivers it to SecA. Once bound to SecA, ATP is used to release the cargo of SecB to SecA and translocate the cargo through the membrane [19,21-23].
1.3.1.2 Signal Recognition Particle (SRP)

The SRP is very similar to the Sec pathway, in as much as they use the same complex SecY, SecE, SecG (SecYEG). However it does not use SecA and SecB to drive secretion or recruit cargo. Instead the SRP molecule is recruited to the ribosome to bind newly formed proteins that need to be membrane bound by ribosome-nascent chain (RNC) complex [20,24]. The SRP+RNC complex then targets the protein SRP receptor FtsY which in turn directs it to the SecYEG machinery [21-23,25].

1.3.1.3 The Tat System

The Sec pathway is the universally conserved pathway in prokaryotes. However, there also exists the Tat pathway. This system is found in some bacteria and has been identified as essential in only a few organisms. Nevertheless, the Tat system plays a significant role in some cellular processes such as iron and phosphate acquisition in *E. coli* and *Salmonella spp.* The Sec pathway is employed for the transportation of unstructured proteins across the bacterial plasma membrane. In contrast, the Tat pathway is used for the transportation of folded proteins. Proteins are identified for transport through the Tat system via a signal peptide containing a twin arginine repeat motif [24,26].

The Tat system comprises of two inner membrane integrated subunits *i.e.* TatA and TatC. In some bacterial systems, there are multiple copies of components of the Tat system. For example, *Bacillus subtilis* has two copies of *tatC* and three copies of *tatA* [25,27] whilst a few organisms such as *E. coli* employ an additional protein TatB.

1.3.2 Gram-negative Bacterial Secretion Systems

Secretion across the outer membrane is a process specifically undertaken by Gram-negative bacteria. In order to achieve this task Gram-negative bacteria have evolved a series of mechanisms which allow them to either export proteins as a two-step process, exploiting one of the methods mentioned above to export the protein across the inner membrane, or as a one step process, where the protein is exported from the cytosol to outside of the cell without any intermediate steps. The mechanisms available to accomplish this task are named, in a simple numerical manner. There are
at present six major systems that are well described: the type I-VI secretion systems. The following sections briefly describe each of those systems.

1.3.2.1 Type I Secretion System (T1SS)

Type-I secretions systems (T1SS) allow for the movement of proteins from the cytoplasm to outside of the cell in a one-step manner, utilising a simple system of just three proteins. These three proteins are an ATP-binding cassette (ABC) transporter, a membrane fusion protein (MFP) and an outer membrane pore forming protein (OMP) [26-28] (see Figure 1.2). The ABC protein consists of a cytoplasmically located nucleotide binding domain, and a transmembrane domain produced from six α-helices [27,29]. ABC proteins characteristically function as homodimers or trimers in producing a functional pore through which the secreted protein can pass through [27,28,30]. Additionally, ABC protein provides substrate specificity to the secretion machinery [29,31]. The MFPs interact in a trimeric fashion with the ABC proteins in order to create a periplasmic channel through which the secreted protein can traverse [30,32].

It has been suggested that the binding of substrates to the ABC protein leads to a conformational change in the MFP, such that it interacts with the OMP to complete the channel to the external environment [31,33]. However, the exact mechanism of this interaction remains unclear [32,34]. The OMP, as characterised by the TolC protein from *E. coli* exists as a trimer anchored in the outer membrane by a β-barrel structure [26,33]. It has been shown that the ABC and OM proteins can interact together directly, although this interaction *in vivo* requires the presence of the MFP [34-37].

There are a wide variety of proteins exported by T1SS machinery, from enzymes to toxins and adhesins [26]. Proteins secreted by T1SSs contain a C-terminal secretion system, most commonly in the terminal 15-30 amino acids [35-39], implying that the molecule must be secreted in a post-translational fashion. There is no specific consensus for this signal sequence, although it would seem that there is a preference towards certain amino acids (LDAVTSIF) [26,40,41].
1.3.2.2 Type II Secretion System (T2SS)

Compared to the T1SSs, type II secretion systems (T2SSs) are considerably more complex and are commonly employed in Gram-negative bacteria such as *Yersinia* spp and *Klebsiella* spp. In order to export proteins through the plasma membrane the Sec or Tat pathways is employed. Once in the periplasm, the protein is trafficked across the outer membrane using the T2SS. T2SSs employ 12 to 16 proteins for enabling secretion of folded proteins through the outer membrane. Perhaps surprisingly though only a couple of these are actually located in the outer membrane [38,39,42,43]. The majority of the proteins are located in the plasma membrane or in the periplasmic space (see Figure 1.3). For the T2SS, the general secretory pathway (Gsp) nomenclature is used, followed by a capital letter.
The cytoplasmic protein GspE is a hexameric ATPase, providing energy to drive the export of proteins [38,40,41]. GspE interacts with the inner membrane platform. This is composed of 4 core proteins (GspC, F, L and M) which anchor the ATPase to the pseudopilus apparatus [39,42-45]. The pseudopilus is composed of 5 proteins termed pseudopilins. The pseudopilins themselves are a group of proteins (GspG-K) that come together to form a large multimeric structure called the pseudopilus. The base of the pseudopilus is in the periplasm and interacts with the plasma membrane. Based on evidence from several experiments it has been hypothesised that the pilus may grow in order to push secreted molecules through the outer membrane complex, or alternatively as a cork to close off the outer membrane channel when not required [38,46].

The outer membrane complex of T2SS consists of two components, (i) the pore forming protein GspD, which exists as a multimer of 12-14 copies forming a pore in the outer membrane, and (ii) a lipoprotein GspS known as the “pilotin”, stabilises the GspD multimer in the outer membrane [44,45,47-49]. The pore formed by GspD is about 95 Å in diameter, a size large enough for proteins to pass through the T2SS assembly in a folded state [46,50,51].
Figure 1.3 Schematic representation of the Type 2 Secretion System (T2SS). The position of the inner membrane (IM), periplasm (PP), outer membrane (OM) and the major components of the T2SS are shown. The cytoplasmic hexameric ATPase GspE (shown in blue) provides energy for protein export and interacts with the inner membrane platform composed of GspF, GspL, GsM and GspC (shown in red, light blue, purple and dark blue respectively). The inner membrane platform anchors the GspE hexamer to the pseudopilin (shown in yellow) comprising of GspG-K which links to the pore forming multimeric protein GspD (shown in light orange) in the outer membrane. GspS (shown in orange) stabilises the GspD multimer in the OM. The protein (shown as orange, rounded rectangle) is first transported from bacterial cytoplasm to periplasm using the Sec pathway (shown in green) and trafficked across the OM using the T2SS assembly.

1.3.2.3 Type III Secretion System (T3SS)

The type III secretion system (T3SS) is found in Gram-negative bacteria that interact with both plants and animal hosts, either as pathogens or symbionts [47-49,52]. The primary function of the T3SS machinery termed as the “injectisome” is the delivery of effector proteins across the bacterial and host membranes into the cytosol of host cells to hijack host cell functions, including immune and defence responses [50-52]. This thesis is based on the T3SS of pathogenic Yersinia pestis and is described in detail in section 1.3.3.
1.3.2.4 Type IV Secretion System (T4SS)

In comparison to other secretion systems, the T4SS (Figure 1.4) is unique in its ability to transport a wide range of substrates such as single peptide molecules, protein-protein complexes and protein-DNA complexes. Furthermore, the T4SS is employed in the infection of a wide range of hosts ranging from higher eukaryotes such as animal and plant cells, to yeast and prokaryotes [52]. Many bacteria have homologous type IV secretion systems, including the pathogens *Agrobacterium tumefaciens* (VirB/D4), *Helicobacter pylori* (CAG/ComB) and *Pseudomonas aeruginosa* (TraS/TraB) [52,53]. Although T4SSs have garnered attention because of their significance in pathogenesis, it is important to point out that not all pathogenic bacteria have a T4SS.

There are several subclasses of T4SS, however the majority of the work to date, has focused on *Agrobacterium tumefaciens*. The VirB system from A. tumefaciens is capable of exporting DNA-protein complex from its Ti plasmid into the host [52,54]. Here, VirB4, VirB11 and VirD4 are cytoplasmic ATPases that generate the energy for the system, driving complex assembly followed by substrate translocation. VirB1 is a lytic transglycosylase responsible for making holes in the peptidoglycan layer thus allowing the formation of the protein channel [53,55]. VirB6-10 form a scaffold creating a translocation channel, connecting the plasma membrane to the outer membrane. This assembly enables the transport of the substrate directly from the cytoplasm to the extracellular milieu. The channel is a cylindrical structure with a diameter and length of 185 Å to allow for the exportation of folded protein molecules [54,56].
1.3.2.5 Type V Secretion System (T5SS)

The type five secretion system (T5SS) (Figure 1.5) is typically referred to as the autotransporter system and is characterised by the requirement of SEC machinery and a β-barrel domain [14,55,57-59]. The process of transport from the plasma membrane to the outer membrane can be divided into three sub classes, monomeric autotransporter system (T5aSS), two partner secretion pathway (T5bSS) and trimeric autotransporters (T5cSS) [14,56-59]. The T5SS serves as a pathway for secretion of a
large number of proteins exceeding the number of proteins secreted by the T2SS. Over 500 proteins are secreted through the T5aSS class alone. The majority of the T5SS secreted proteins that have been characterised to date are responsible for the virulence of animals or human pathogenic bacteria [14,57-59].

The T5SS proteins are produced with an N-terminal signal peptide, which exports them into the periplasm using the Sec pathway. The bacterial proteins secreted via the T5aSS typically comprise an N-terminal passenger domain from 40 - 400 kDa in size and a conserved C-terminal domain which forms a β-barrel [14,57-60]. The β-barrel is able to insert into the outer membrane and is needed for translocation of the passenger domain into the extracellular milieu. In some cases, such as adhesins, the passenger domain remains attached to the β-barrel and the protein remains attached in the outer membrane. In other cases, the passenger domain is cleaved from the β-barrel to form a toxin or hydrolytic enzyme. Recently, detailed structural studies have suggested that the barrel is not capable of transporting the passenger domain by itself. It has been hypothesised that a helper protein, Omp85/YaeT enables the transportation across the outer membrane [57,59]. Proteins secreted via the T5aSS include IgA protease from Neisseria meningitidis and Pertactin from Bordetella pertussis [14,60].

A second class of proteins secreted via the T5SS is called T5bSS proteins. In contrast to T5aSS, the passenger and translocator functions in T5bSS are located on separate polypeptide chains. T5bSS contains a pair of proteins in which one of the proteins carries the β-barrel domain, and the other is the secreted protein (passenger polypeptide). This process is called two partner secretion (TPS) [57,61]. The mechanism of transport of the passenger polypeptide of the T5bSS is similar to the passenger domain transport in T5aSS. T5bSS (TPS) secreted proteins include adhesins such as HecA/HecB of the plant pathogen Dickeya dadantii (Erwinia chrysanthemii) and cytolysins such as ShlA/ShlB of Serratia marcescens [14,62].

The third sub-class of T5SS proteins, called T5cSS are trimeric autotransporters, in which a single β-barrel is formed by contributions from each of the three monomeric subunits. In contrast to the monomeric T5aSS, they are usually adhesins and don’t exhibit enzymatic function. However, the method of transportation of the passenger
domain is identical to that in T5aSS. An example of T5cSS protein is YadA, the Yersinia adhesin A of Yersinia enterocolitica [61,63-68].

Figure 1.5 Schematic representation of the Type 5 Secretion System (T5SS). Inner membrane (IM), periplasm (PP), outer membrane (OM) and the major components of the T5SS are shown. The secretion pathway of the monomeric autotransporter system (T5aSS) is depicted at the bottom left, the two-partner system (T5bSS) is displayed in the centre, and the trimeric autotransporters (T5cSS) is shown on the right. The four functional domains of the proteins are shown: the signal sequence, the passenger domain, the linker region, and the β-domain. The autotransporter proteins are synthesised and exported through the cytoplasm across the IM via the Sec machinery. Once, through the IM, the signal sequence is cleaved and the β-domain inserts into the OM in a β-barrel structure forming a pore in the OM. The passenger domain then inserts into the pore and is translocated across the OM to the extracellular milieu.

1.3.2.6 Type VI Secretion System
The type VI secretion system (T6SS) is the most recently identified Gram-negative bacterial secretion system. The name T6SS was first coined in 2006, when it was identified in Vibrio cholera [62,69-71]. T6SS is highly conserved and is found in nearly 25% of all Gram-negative bacteria such as Pseudomonas, Escherichia, Agrobacterium
and *Helicobacter spp.* [63-68,72]. It is a large complex composed of 13 core components that span from the cytoplasm, to the extracellular milieu and the target cell. These proteins collectively assemble a structure reminiscent of an upside-down bacteriophage-like structure anchored to the bacterial cell envelope [69-71,73]. T6SS machinery is capable of delivering effector proteins directly into the host cell cytoplasm in a single-step, independent of the Sec and Tat pathways [72,74].

The T6SS core components can be grouped into the base plate complex, sheath, tube and spike. In both *V. cholerea* and *P. aeruginosa*, the base plate complex comprises of TssE (type six subunit E), TssF, TssG, TssJ, TssK, TssL, TssM and TssA. TssEFG are located in the cytosolic side of the bacterial inner membrane and in close proximity to the tube complex [73,75-77]. TssL is anchored in the inner bacterial membrane, TssM is bound to the inner membrane and extends outwards interacting with TssJ, an outer membrane lipoprotein [74,78]. The roles of TssF, TssG and TssA are unclear although they are universally conserved [73,75-77]. The formation of the base plate assembly is followed by the haemolysin coregulated protein (Hcp) tube formation. The Hcp tube undergoes polymerisation and emerges from the base plate complex spanning the bacterial cytosol to the extracellular milieu [78,79]. Then the VipA/VipB sheath polymerises around the growing Hcp tube. TssK is an essential cytoplasmic protein, which interacts with membrane bound TssJ-TssL-TssM complex as well as the tube and sheath components of the T6SS [73]. Analogous to phage, a conformational change in the sheath structure causes contraction. This propels the Hcp tube out of the cell and across the target cell membrane. This contraction event enables the delivery of effectors loaded onto the VgrG (valine-glycine repeat G)/PAAR (proline-alanine-alanine-arginine motif containing proteins) spike complex or the distal end of the Hcp tube [79,80]. The contraction process disassembles the unsheathed Hcp tube; parts of the tube that are not expelled from the cell are recycled within the cytosol. The contracted sheath is remodelled by the ATPase activity of ClpV, restoring the pool of available sheath subunits. Sheath components VipA/VipB interact with ClpV that is associated with the baseplate components in the bacterial cytosol. The naked baseplate complex is then either recycled or disassembled, depending on the
requirement and activation state of the T6SS [7,73]. The T6SS assembly is shown in Figure 1.6.

Figure 1.6 Schematic representation of the Type 6 Secretion System (T6SS). The inner membrane (IM), periplasm (PP), outer membrane (OM) and the major components of the T6SS are shown. The components of the base plate complex: TssEFG are located on the cytosolic side of the IM, TssL is anchored in the IM, TssM is bound to the IM and extends to the OM and interacts with TssJ, TssK interacts with membrane bound TssJLM complex, and sheath and tube components. The tube is formed of Hcp polymer, emerging from the base plate spanning the cytoplasm to the extracellular milieu. The sheath is composed of VipA/VipB which polymerises around the growing Hcp tube. Conformation change in sheath components causes contraction of Hcp tube causing the propulsion of the spike complex of VgrG/PAAR located at the tip of the Hcp tube along with the loaded effectors into the target cell in the extracellular milieu. ATPase ClpV remodels the sheath, recycles the tube and drives the translocation of effectors and VgrG/PAAR spike complex through the T6SS machinery.
1.3.3 Type Three Secretion System in Detail

Pathogenic *Yersinia* species cause human diseases ranging from relatively mild intestinal disease for *Yersinia enterocolitica* [80-84] to bubonic plague for *Yersinia pestis* [7,85]. Despite the differences in disease, virulence of the *Yersinia* species is caused by a conserved type three secretion system (T3SS), which is now a well-established model system for this form of protein secretion. Although the T3SS was first discovered in *Yersinia pestis*, it is a conserved virulence factor amongst other human pathogens such as enteropathogenic *Escherichia coli* (EPEC), *Salmonella* spp., *Pseudomonas aeruginosa*, *Shigella flexneri*, and *Chlamydia* spp. [81-84,86]. These pathogenic bacteria collectively cause numerous fatalities and significant healthcare costs annually [85,87-90].

T3SS uses a molecular syringe for the delivery of cytotoxic effectors into eukaryotic host cells. These effectors influence and manipulate diverse cellular pathways in the eukaryotic host for the survival and proliferation of the pathogen by either repressing the inflammatory response or altering phagocytosis [83,86,90]. Interfering with the delivery of effectors could have substantial consequences on disease pathology. Since this virulence mechanism is conserved in numerous pathogenic bacteria, it makes an attractive target for novel therapeutics.

In *Yersinia* spp., the genes of T3SS are located on a 70 kb virulence plasmid, and the expression of genes *in vitro* is dependent on temperature and calcium concentration, a phenomenon called the low calcium response (LCR) [87-92]. T3SS genes are not expressed at ambient temperature. However, upon increasing the temperature of *Yersinia* cultures from 26°C to 37°C in the presence of millimolar concentrations of calcium, conditions representing the mammalian host, T3SS genes are expressed at low levels and the molecular machinery known as the “injectisome” is assembled [90,92-94].

The “injectisome” comprises several components that can be broken down into the various units: (i) the basal body which is composed of the inner membrane ring, the scaffold proteins, the outer membrane ring, the export apparatus, the ATPase complex and the C ring; (ii) the needle which is an extension from the basal body into the
extracellular milieu and (iii) the pore complex also known as the “translocon” which is formed in the membrane of the target (Figure 1.7).

Figure 1.7 Schematic representation of the Type 3 Secretion System (T3SS). The inner membrane (IM), periplasm (PP), outer membrane (OM), the host cell membrane and the major components of the T3SS are shown: basal body, needle and pore complex. The basal body comprises of YscC, YscD, YscJ which form a platform for the assembly of YscQ, YscRSTUV and the complex YscNLK of which YscN is the ATPase powering the secretion of substrates. The needle comprises of YscI which forms a rod through which YscF is secreted and forms the needle. The pore complex is formed by the polymerisation of LcrV at the distal end of YscF forming a pentameric tip and complexes with translocators YopB and YopD which then insert into host cell membrane creating a pore or “translocon”. The class II chaperone LcrH (shown in yellow) is depicted as binding to YopB and YopD before being channeled across the T3SS assembly. The cargo of YopB/D is then released which gets inserted into the host cell membrane forming a pore.
THE BASAL BODY

Assembly of the T3SS is driven by the formation of the basal body which is initiated by the oligomerisation of YscC on the periplasmic side of the outer membrane and is termed the outer membrane ring (OM) [91,92,95,96] Next, a ring of YscD is assembled in the inner membrane and is thought to connect the inner and outer membrane rings [92-94,97]. YscD then binds to YscJ, which oligomerises to complete the inner membrane (MS) ring [92,93,95,96,98-104]. The assembly of these structures forms a basic channel through the bacterial envelope, which serves as a platform for the assembly of the remaining injectisome components.

On the cytosolic face of the basal body is the C-ring. This is composed of YscQ, which colocalises with YscC in the outer membrane stabilising the whole assembly. Additionally, a complex composed of YscN, YscL and YscK is formed and interacts with the C-ring. YscN is the ATPase that powers the secretion of substrates by the T3SS, YscL is a negative regulator of ATPase activity, while the function of YscK is unknown [97,105].

The next to follow is the assembly of the export apparatus composed of integral membrane proteins YscRSTUV [93,98-104,106-110] within the inner membrane, independently of the scaffold proteins. The pathways for assembly converge at this point with the recruitment of the export apparatus to YscJ in the MS ring of the scaffold [105,111,112]. The formation of the fully functional basal body is achieved with the amalgamation of the scaffold, ATPase complex and export apparatus. The basal body is now capable of exporting secretion substrates.

THE NEEDLE AND THE “EARLY STAGE”

On completion of the basal body, YscI integrates into the inner membrane and forms the rod that allows substrate transport across the inner membrane. This stage is referred to as the “early” stage because only “early” substrates are translocated. These include the first proteins to be secreted: YscIFPXO and YopR [106-110,113-116]. The needle protein YscF is secreted through the YscI rod into the periplasmic space
allowing substrate transport across the inner membrane [111,112,117,118]. YscF then undergoes polymerisation to form the needle, which extends ~41 nm from the bacterium Y. pestis or ~58 nm from Y. enterocolitica. The needle has an outer diameter of ~6-7 nm and an inner diameter of ~2-3 nm [97,113-116].

YscPXO and YopR are not shown in the diagram but have a regulatory role. Indeed, YscP is required to control the needle length while YopR is important in the early stages of needle assembly, though its exact role is not clear [117,118]. Likewise, YscX and YscO are essential for the secretion of early substrates but are not secreted from the bacterium until needle assembly is complete [97,119]. YscY is thought to be the chaperone for YscX secretion [2,106] and YscY has also been shown to bind to SycD/LcrH, which is the chaperone for facilitating secretion of the middle substrates YopB and YopD. The role of the chaperones such as LcrH is covered more extensively below.

**THE MIDDLE STAGE AND TRANSLOCON ASSEMBLY**

The secretion of translocators LcrV, YopB and D into the extracellular milieu initiates the middle stage. LcrV polymerises at the distal end of YscF forming a pentameric tip on the needle complex [119,120]. YopB and YopD complex with LcrV at the needle tip and are inserted into the host cell membrane creating a pore. This large assembly is termed the pore complex or "translocon". This is the key step in the creation of the T3SS as organisms in which the translocon formation is impeded are rendered non-lethal *in vivo* [120,121].

Once the translocon is formed, Yops are delivered through the complex and the late stage, involving the delivery of effector proteins such as YopN and TyeA begins [121,122]. Here, immune supressing proteins and molecules targeting macrophages are released and the process of infection begins [122,123].
Table 1.1 Secretion Systems of Gram-negative bacteria describing their subunit composition, functional characteristics and citations.

<table>
<thead>
<tr>
<th>Secretion System</th>
<th>Composition of subunits</th>
<th>Functional Characteristics</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1SS</td>
<td>ATP Binding Cassette (ABC) transporters also known as Inner Membrane Component (IMC) located in the inner membrane. e.g. AcrB in <em>E. coli</em></td>
<td>ABC is an ATPase. It supplies energy for transport of variety of substrates such as ions and macromolecules.</td>
<td>AcrB [123,124], PDB: 4CDI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MFP interacts with the OMP to complete the channel to the external milieu for export of various compounds.</td>
<td>AcrA [124,125], PDB: 2F1M</td>
</tr>
<tr>
<td></td>
<td>Membrane Fusion Protein (MFP) located in the periplasm. e.g. AcrA in <em>E. coli</em>.</td>
<td>OMP forms an exit duct for diverse molecules like protein toxins and antibacterial drugs.</td>
<td>TolC [125,126], PDB: 2XMN</td>
</tr>
<tr>
<td>Secretion System</td>
<td>Composition</td>
<td>Function</td>
<td>Citations</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>T2SS</td>
<td>Inner Membrane (IM) platform comprising four core proteins: GspC, GspF, GspL and GspM tightly associated with cytoplasmic hexameric ATPase GspE.</td>
<td>GspE is an ATPase, provides energy to drive export of proteins and interacts with IM platform. GspC, F, L and M anchor GspE to the pseudopilus apparatus.</td>
<td>GspE - [126,127], PDB: 4KSR. GspC - [127,128], PDB: 3OSS. GspF - [128-130], PDB: 3C1Q. GspL - [129-131], PDB: 2W7V, 2BH1. GspM - [131,132], PDB: 1UV7.</td>
</tr>
<tr>
<td>Periplasmic pseudopilus, a large multimeric structure composed of the major pseudopilin subunit GspG and minor pseudopilin subunits GspL, GspJ, GspK and GspH.</td>
<td>The base of pseudopilus is in the periplasm and interacts with the plasma membrane. Pilus may grow in order to push secreted molecules through the OM complex or block the OM channel when not required.</td>
<td>GspG - [132,133], PDB: 3G20. GspH - [133,134], PDB: 4DQ9. Gsp I/J/K - [134,135], PDB: 3CI0.</td>
<td></td>
</tr>
<tr>
<td>Outer Membrane (OM) complex consists of a dodecameric secretin GspD and a lipoprotein GspS (pilotin).</td>
<td>GspD forms a large pore in the OM for passage of folded proteins and GspS stabilizes the GspD multimer in the OM.</td>
<td>GspD - [135,136], PDB: 3EZJ. GspS - [136,137], PDB: 3SOL.</td>
<td></td>
</tr>
<tr>
<td>Secretion System</td>
<td>Composition</td>
<td>Function</td>
<td>Citations</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>T3SS</strong></td>
<td>Basal body</td>
<td>YscC forms a ring in the OM, acting as scaffold. YscD &amp; YscJ form a ring in the IM creating a scaffold. YscRST promote polymerization of YscV. YscU is essential for substrate specificity. YscN is the ATPase for substrate secretion by T3SS. YscN is a negative regulator of ATPase activity.</td>
<td>YscD - [137,138], PDB: 4D9V YscU - [138,139], PD: 2V5G</td>
</tr>
<tr>
<td></td>
<td>Scaffold proteins: YscC, YscD, YscJ Export apparatus proteins: YscR, YscS, YscT, YscU, YscV. Cytoplasmic components: YscQ (C-ring) and YscN, YscL, YscK (ATPase complex).</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>The needle</strong></td>
<td>YscI, YscF, YscP, YscX, YscO and YopR</td>
<td>YscI forms the rod to allow substrate passage across the IM. YscF undergoes polymerization to form the needle for translocation of substrates. YscP regulates the length of the needle. YscX and YscO are required for early substrate secretion.</td>
<td>YscF - [139,140], PDB: 2P58 YscO - PDB: 4MH6 YopR - [140,141], PDB: 1Z21</td>
</tr>
<tr>
<td></td>
<td>YscI forms the rod, followed by secretion of YscF which forms the needle</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pore complex</strong></td>
<td>LcrV (needle tip complex) YopB/YopD (translocon pore)</td>
<td>LcrV assists YopB/YopD for insertion in the host cell membrane. YopB/YopD create a pore into the host cell membrane and form the translocon for delivery of virulent proteins.</td>
<td>LcrV - [141,142], PDB: 1R6F YopD - [53,142], PDB: 4AM9</td>
</tr>
<tr>
<td>Secretion System</td>
<td>Composition</td>
<td>Function</td>
<td>Citations</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>T4SS</td>
<td>Inner Membrane (IM) complex is composed of three hexameric ATPases (VirD4, VirB4 and VirB11) located at the cytoplasmic side of the IM, together with VirB3, VirB6 and VirB8.</td>
<td>VirD4, VirB4 and VirB11 ATPases provide energy for substrate transport and T4SS biogenesis. VirB11 assists VirB4 in pilus biogenesis and polymerization. VirB8 helps in the assembly of the pili.</td>
<td>VirB4 - [53,143], PDB: 4AG5. VirB11 - [143,144], PDB: 2GZA. VirB8 - [144,145], PDB: 4AKZ.</td>
</tr>
<tr>
<td></td>
<td>Outer membrane (OM) complex comprises of VirB7, VirB9 and VirB10 with VirB10 extending from the IM to the OM.</td>
<td>VirB7 and VirB9 together from a pore in the OM for substrate translocation. VirB10 forms part of a scaffold around the translocation channel for substrate.</td>
<td>VirB7/VirB9/VirB10 - [145,146], PDB: 3JQO. VirB9/VirB10 - [74,146], PDB: 3ZBJ</td>
</tr>
<tr>
<td></td>
<td>Conjugative pilus is composed of VirB2 and VirB5.</td>
<td>VirB2 and VirB5 form the major and minor subunit of the pilus respectively, connecting the IM to the OM.</td>
<td></td>
</tr>
<tr>
<td>Secretion System</td>
<td>Composition</td>
<td>Function</td>
<td>Citations</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>T6SS</td>
<td><strong>Baseplate complex comprising TssE (type six subunit E), TssF, TssG, TssJ, TssK, TssL, TssM, TssA and ClpV (ATPase). TssEFG, TssK, TssL, ClpV are located on the cytosolic side of bacterial IM, TssM is bound to the IM and extends onto the OM where it connects to TssJ.</strong></td>
<td>TssM is bound to the IM and extends outwards interacting with TssJ. TssK interacts with membrane bound TssJLM complex, the tube and sheath components. ClpV is an ATPase, provides energy for assembly/disassembly of the sheath and baseplate components.</td>
<td>TssJ/TssM - [74,147], PDB: 3RX9 TssL - [147,148], PDB: 3U66 ClpV - [148,149], PDB: 3ZRJ</td>
</tr>
<tr>
<td></td>
<td><strong>Sheath component: VipA/VipB (TssB/TssC)</strong></td>
<td>Interacts with ClpV in bacterial cytosol, causes contraction of the growing Hcp tube and propels the spike complex loaded with effectors.</td>
<td>VipA/VipB - [79,149], PDB: 3J9G.</td>
</tr>
<tr>
<td></td>
<td><strong>Tube component is haemolysin coregulated protein (Hcp) hexamer (Hcp1), connecting the IM with the OM.</strong></td>
<td>Virulence factor, transporter of effectors and chaperone.</td>
<td>Hcp1 - [79,150], PDB: 4W64.</td>
</tr>
<tr>
<td></td>
<td><strong>Spike complex is made of PAAR-VgrG protein complex and in the OM.</strong></td>
<td>Transportation of effectors into the extracellular milieu, killing of target cell.</td>
<td>PAAR-VgrG - [150,151], PDB: 4JIV</td>
</tr>
</tbody>
</table>
1.3.3.1 The role of Chaperones in the T3SS

For efficient export of both hydrophobic translocators and effector proteins specific chaperones are required. They act as “bodyguards” preventing their substrates from generating premature or non-specific interactions with other proteins in the bacterial cytosol prior to export. Moreover, they can maintain their substrates in a partially folded state in order to enable their passage through the narrow channel of the injectisome and may also play a role in secretion process [151-154]. Interestingly, the T3SS chaperones lack the ability to bind and hydrolyse ATP unlike classic chaperones like DnaK or GroEL. T3SS chaperones are divided into three sub-groups based on their binding partners and their modes of interaction: class I chaperones interacting with effector proteins, class II chaperones interacting with the translocators and class III chaperones interacting with T3SS needle components [152-155].

1.3.3.2 Class I Chaperones

Class I chaperones are relatively small (~10-15 kDa), acidic (pI 4-5) and generally display a homodimeric structure with a conserved α-β fold (Figure 1.8). They interact with their cognate effector protein through a chaperone-binding domain (CBD), which is located within the first 100 amino acids of the effector protein. This is downstream from the short amino terminal export signal. They are classified into two categories based on their interaction with their effectors: class IA chaperone bind a single effector, whereas class IB chaperones bind multiple effectors.

Class 1A chaperones: The majority of the structures solved to date fall into class IA. The examples for class IA chaperones are SycE [155,156], SycH [156,157], SycN-YscB (see Figure 1.8a) [157-159], SycT (see Figure 1.8b) [158-160] from Yersinia spp. and CesT [160,161] from enteropathogenic E. coli (EPEC).
The class I chaperones show low sequence similarity, but possess a conserved 3D structure and a common mode of binding. Each of these dimeric chaperones use a similar fold, where the monomer comprises of three α-helices and a five-stranded antiparallel β sheet arranged in the order of α-β-α-β-α, where the helices are positioned on one face of the β sheet. The central α helix and the loop and β strand, which immediately follows, are responsible for dimerization. This covers a hydrophobic surface area of generally around 2500 Å² [156,161-163]. The CBD of effector proteins wrap around chaperone homodimers in an extended, horseshoe like conformation with some α-helical secondary structure organisation [156,160,162-164]. The presence of hydrophobic surfaces on the chaperone homodimers are responsible for the binding of the cognate effector protein [156,160,162,164]. A good example of this interaction is SycH and its effector YscM2. Here the stable YscM2[33-50] fragment wraps around the homodimeric SycH in an extended fashion. YscM2[33-50] interacts with SycH via a structured α helix and β strand that dock with the β sheet and helix of the SycH [156,157] (see Figure 1.9).
Interestingly there are deviations from the homodimeric structure described above. For example, the crystal structure of class IA chaperone SycN-YscB is unique amongst the class I chaperones because it is a heterodimer \[157,161\] (see Figure 1.11a). Here, although the topology of each monomer is conserved, the heterodimeric nature of the assembly leads to an asymmetry that does not exist in homodimeric class IA chaperones. Even though the \(\alpha2\) helix of YscB lies parallel to the dimer interface, as observed in the homodimers, the corresponding helix in SycN is tilted and induces a shift of \(\beta4\) and \(\beta5\) of YscB. This asymmetry in the dimer interface has a direct influence on the specific binding of YscB to its effector, YopN \[152,161\] (see Figure 1.10). YopN contains an N-terminal secretion signal followed by a chaperone-binding domain (CBD) where its two chaperones, SycN and YscB, bind. The YopN CBD wraps around SycN-YscB chaperone keeping YopN partially unfolded and secretion competent \[152,165\].

Figure 1.9 Ribbon representation of the crystal structure of *Yersinia pestis* T3SS Class IA chaperone SycH in complex with a stable fragment of its cognate effector protein YscM\(_2\)\(_{33-50}\) (PDB code 1TTW). The monomeric units of SycH are shown in red and magenta respectively. The effector YscM\(_2\)\(_{33-50}\) fragment is shown in green. Figure was prepared using PyMOL.
Class 1B Chaperones: As stated previously, class 1B chaperones contrast to the Class 1A by binding to multiple diverse effectors. For example, Spa15 from *Shigella flexneri* and InvB from *Salmonella enterica* (see Figure 1.11a and Figure 1.11b, respectively). Interestingly, in spite of possessing the ability to bind to various effectors, the monomers of each chaperone share the same overall fold as the class IA chaperones. However, the main difference lies with the orientation of the dimers. In particular, the α2 helices that form the dimeric interface have a different orientation, leading to a dimer in which the subunits are rotated about 30° with respect to each other. The residues that cause this differing shape and orientation of the α2 helix are conserved among other putative class 1B chaperones and thus distinguish class 1B from class IA chaperones. The change in dimeric shape also changes the orientation of the hydrophobic pockets that bind the effectors. The effectors still wrap around chaperone homodimers in the extended, horseshoe-like conformation, but the rotation of the assembly causes a longer wrapping linker. Moreover, the residues that form the hydrophobic binding regions are more highly conserved, enabling differing effectors to bind. This suggests that the chaperones of the 1B class are more general chaperones than the more specific class 1A [142,165-169].

**Figure 1.10** Ribbon representation of the crystal structure of *Yersinia pestis* T3SS macromolecular assembly of effector YopN in complex with its heterodimeric class IA chaperone SycN-YscB (PDB code 1XKP). SycN, YscB and YopN are shown in red, magenta and green respectively. Figure was prepared using PyMOL.
1.3.3.3 Class II & III Chaperones

Both class II & class III chaperones are small, acidic and slightly larger than class I chaperones (~15-20 kDa). Both bind to parts of the secretion apparatus that is formed outside the bacteria. The class II chaperones bind to and stabilise the translocon forming translocator proteins (for example, YopB and YopD in Yersinia spp., PopB and PopD in Pseudomonas aeruginosa and IpaB and IpaC in Shigella spp., respectively). Similarly, the class III T3S chaperones bind to the needle component preventing their premature association or aggregation within the bacterial cytoplasm (for example YscF in Yersinia spp. and PscF in Pseudomonas aeruginosa). Importantly, the translocon pore and needle complexes can only form if their respective chaperones are present in the bacterial cytosol [142,166-171]. If the chaperones are knocked out, no needle or pore formation occurs, thus rendering the bacteria non-infective.

Interestingly, class II and class III chaperones, although binding to molecules that perform completely different functions, have remarkably similar structures. These are formed from multiple copies of a motif called the tetratricopeptide repeat (TPR). TPRs are short 34 amino acid motifs that adopt a α-helix-turn-α-helix conformation and stack onto each other to form elongated structures [170-172]. The X-ray structure of a domain containing three TPRs from protein phosphatase 5 discovered the helix-turn-helix arrangement, with adjacent TPR motifs packing in a parallel fashion, resulting in a
spiral of repeating anti-parallel alpha-helices. The two helices are denoted as A Helix and B Helix (Figure 1.12). The packing angle between the two helices is \( \sim 24^\circ \) within a single TPR and produces a right-handed super helical shape. Two protein surfaces are generated: mainly residues on helices A contribute to the inner concave surface, and the other surface presents residues from both helices A and B. There can be between 3 and 16 direct repeats in a given protein domain [170,172]. Individual TPR motifs are not stable and must be arrayed in tandem to fold correctly. TPR containing proteins are widespread in prokaryotic and eukaryotic organisms and are involved in a variety of biological processes, such as cell cycle regulation, transcriptional control, mitochondrial and peroxisomal protein transport, neurogenesis and protein folding [142,166-170].

Figure 1.12 Consensus sequence of TPR. The consensus TPR motif (yellow) and solvating helix (green) sequence are displayed with the corresponding sequence underneath, ‘n’ denotes the number of consensus TPRs. The highly conserved residues are highlighted in yellow (>50% conservation of an identical residue) and pink (50% of similar residue type). The consensus sequence was taken from SMART (a Simple Modular Architecture Research Tool) database with accession no. SM00028.

Excitingly, crystal structures for both Class II and Class III chaperones were recently solved as follows:
1. **Class II Chaperones:** SycD from *Yersinia enterocolitica*, IpgC from *Shigella flexneri* and PcrH from *Pseudomonas aeruginosa* [139,142,166-169,173] (Figure 1.13).

Figure 1.13 Ribbon representation of crystal structures of homodimeric T3SS class II chaperones: (a) SycD$_{21-163}$ from *Yersinia enterocolitica* (PDB code 2VGX), (b) IpgC from *Shigella flexneri* (PDB code 3GY2) and (c) PcrH$_{21-160}$ from *Pseudomonas aeruginosa* (PDB code 2XCC). The monomers are shown in red and magenta respectively in each of the above structures. Figure was prepared using PyMOL.
2. **Class III Chaperones:** YscG from *Yersinia* enterocolitica and PscG from *Pseudomonas aeruginosa* [139, 167, 168, 173] (Figure 1.14).

![Ribbon representation of the crystal structure of T3SS class III chaperones](image)

Figure 1.14 Ribbon representation of the crystal structure of T3SS class III chaperones: (a) YscG (coloured green) in complex with YscE (coloured blue) from *Yersinia enterocolitica* (PDB code 2P58) and (b) PscG (coloured green) in complex with PscE (coloured blue) from *Pseudomonas aeruginosa* (PDB code 2UWJ). Figure was prepared using PyMOL.

The structures reveal that each class of chaperone possess a TPR domain of 3.5 TPR motifs. These form a basket-like molecule that has concave and convex sides. In each case their secretion apparatus cargo is bound to the concave face of the molecule. However, the mode of binding within this concave face and oligomeric structure of each chaperone is very different for both classes. Class II chaperones form homodimers that bind to extended peptides of the translocator proteins. Whereas Class III chaperones form heterodimers with a further protein and bind a structured helical segment of their needle cargo. This might be due to the class II chaperones being required to bind two types of cargo protein and the class III chaperones only being required to bind one cargo protein. To dissect the structures in more detail each class is discussed separately in detail below.
1.3.3.4 Structure and biophysical characterisation of Class II Chaperones

**Oligomeric Structure:** Although class II chaperones are generally homodimers their exact mode of dimerization is somewhat ambiguous. The crystal structures provide a number of possible conformations depending on the protein and construct used (Figure 1.15 and Figure 1.16). For N-terminally shortened SycD (21–163) and IpgC (10–151), further mutational analysis suggested that the head-to-head dimer was the relevant unit (Figure 1.15). In the case of SycD, crystal structures show two possible head-to-head conformations (Figure 1.15a&b). In contrast, the relevant unit of full-length IpgC (1–151) seems to be a back-to-head conformation (Figure 1.15d). N-terminally shortened PcrH (21–160) crystal structures show two dimer structures that might be stable in solution, an asymmetric back-to-back dimer and a symmetric head-to-head dimer (Figure 1.15 (e) & (f)).

To investigate the physiological role of class II chaperone dimerization, *in vivo* experiments were carried out in Yersinia and in Shigella, respectively. Here, the mutated monomeric mutant showed characteristics typical of a null mutant. These consisted of a lack of translocator secretion and an inability to invade mammalian cells [166-168,174]. Yet studies on AcrH and PcrH have shown that translocator binding disrupts their dimeric structures and produces 1:1 chaperone-translocator complexes [166,174-176].

Such structural ambiguity suggests certain flexibility in the topology of the translocator chaperones, which would seem to be of importance to their function. This appears logical given the modular TPR units that form their structure. However, there is little information and no biophysical studies to have specifically investigated whether this is the case. These questions form the basis of this thesis and will be elaborated in Section 1.4 – Thesis Aims.
Figure 1.15 Ribbon representation of the crystal structures of class II chaperones: (a) & (b) head-to-head confirmations of dimeric obtained without YopD peptide SycD$_{21-163}$. Chains A and B of SycD$_{21-163}$ are coloured blue and red, respectively. Structures (a) and (b) were produced from 2VGX.pdb and 2VGY.pdb, respectively. Leucine 65 is represented by green spheres and was mutated to glutamic acid to obtain monomeric protein. LcrH differs from SycD in only two positions (N136D and P138T). These amino acids are shown as space fill spheres in each of the monomeric units. (c) & (d) Two differing conformations of dimeric IpgC$_{10-151}$ (c) and dimeric IpgC$_{1-151}$ (d). Chains A and B of IpgC are coloured cyan and orange, respectively. Structures (c) and (d) were produced from 3GZ1.pdb and 3KS2.pdb, respectively. (e) & (f) Two differing conformations of dimeric PcrH$_{21-160}$. Chains A and B of PcrH$_{21-160}$ are coloured dark and green respectively. Structures (e) and (f) were both produced from 2XCC.pdb. Figures (a) to (f) were prepared using PyMOL.
**Cargo Binding:** Although biochemical studies have implicated more than one interaction between chaperone and translocators, structures of IpgC with IpaB (51–72), PcrH (21–160) with PopD (47–56), and SycD (21-163) with YopD (56–65) show a common main interaction where an extended peptide from each translocator binds to the same concave face of its cognate TPR domain (1:1 ratio) [142,166,175,176] (Figure 1.16). Thus, the structures suggest that the chaperone can only bind one translocator at any given time.
Figure 1.16 Crystal structures of class II chaperones bound to their cognate translocator protein peptides. (a) Ribbon representation of the asymmetric dimer IpgC\textsubscript{1-151} with bound peptides of IpaB\textsubscript{51-72} (major translocator protein) in back-to-head confirmation (PDB code 3GZ1). Chain A and chain B of IpgC\textsubscript{1-151} are coloured cyan and orange, respectively. The bound peptides of IpaB are coloured dark green. Residues in chain A of IpgC that interact with IpaB are shown as sticks and coloured purple. These are located in TPR1-3. (b) Ribbon representation of PcrH\textsubscript{21-160} with the bound peptide of PopD\textsubscript{47-56} (minor translocator protein) (PDB code 2XCB). Chain A and chain B of PcrH are coloured dark grey and green, respectively. The bound peptides of PopD are coloured dark green. Residues in PcrH that interact with PcrH are shown as sticks and coloured purple. These are located in TPR1-3. Figures were prepared using PyMOL. (c) Ribbon representation of the dimer SycD\textsubscript{21-163} with bound peptides of YopD\textsubscript{56-65} (minor translocator protein) in back-to-head confirmation (PDB code 4AM9). Chain A and chain B of SycD\textsubscript{21-163} are coloured blue and red respectively. The bound peptides of YopD are colored dark green. Residues in chain A of SycD that interact with YopD are shown as sticks and coloured purple. These are located in TPR1-3. Leucine 65 is represented by green spheres and was mutated to glutamic acid to obtain monomeric protein. SycD differs from LcrH in only two positions (N136D and P138T respectively). Figures were prepared using PyMOL.
A good example to illustrate the chaperone-translocator binding is the crystal structure of SycD in complex with a peptide of the N-terminal region of the minor translocator protein YopD (Figure 1.16 c). The N-terminal YopD hexapeptide PElik^P^1 has three hydrophobic residues (PElik^P^63) that dock into three specific conserved hydrophobic clefts found along the concave face of the chaperone. The clefts are lined by aromatic and aliphatic residues mainly from the A helices of the first and second TPR motifs [142,167]. The same type of interaction and conserved residues were found in Shigella and Pseudomonas where the major translocator IpaB (PElik^P^63) and the minor translocator PopD (VELNAP^P^54) bind to IpgC and PcrH, respectively [167]. The binding mode is shown in Figure 1.16 and the sequence conservation of the chaperones is shown in Figure 1.17. Interestingly, Lunelli et al. performed Isothermal Titration Calorimetry (ITC) experiments to obtain a K_d for the interaction of PELKAP peptide with IpgC and various constructs of IpgC in which the peptide-binding site was disrupted. PELKAP bound to IpgC with a K_d of 625 ± 11 µM, whereas the IpaB^P^51-72 peptide, used in the co-crystallization bound to IpgC with a K_d of 72 ± 8 µM. The mutated IpgC showed no binding to the peptide, was unable to complement the deletion of ipgC in epithelial cell invasion and failed to allow secretion of translocator IpaB. Additionally, they assayed the K_d of peptides encoding the putative chaperone-binding motif of YopB (VQLPAP) and YopD (PElikP) to SycD. These still showed an interaction, but with a very weak K_d of 455 ± 19 µM and 1,370 ± 38 µM respectively [167,173]. Thus although the peptide binding interfaces shown in the crystal structures are important, the ITC results suggest a further interaction site on the chaperone is probable.
53

Figure 1.17 Structure based sequence alignment of T3SS Class II chaperones. IpgC from *Shigella flexneri*, PcrH from *Pseudomonas aeruginosa*, SycD from *Yersinia enterocolitica*, fl-LcrH and tr-LcrH from *Yersinia pestis* (residues are shown in red). Residues forming the TPR helices are highlighted in yellow. C-terminal solvating helix forming residues are highlighted in green. Residues involved in hydrophobic and van der Waal’s interaction with the class II chaperone’s corresponding translocator extended peptides are highlighted in red. Amino acids involved in salt bridge formation by interaction with side chain residues of corresponding translocator peptides are highlighted in light blue. The difference between SycD and LcrH (fl-LcrH and tr-LcrH) is in only two positions: N136D coloured dark blue and P138T coloured purple.

1.3.3.5 Structure and biophysical characterisation of Class III Chaperones

As stated earlier, although the class III chaperones possess a TPR domain like the class II chaperones, they don’t form homodimers and don’t bind their cargo as an extended peptide. Instead they interact with an additional protein to form a heterodimer and interact with their cargo through a structured α helix-turn-loop arrangement.

Oligomeric State: In both reported structures, the class III TPR chaperone domain binds a small 2 α-helical protein at its N-terminus. In *Pseudomonas* the TPR chaperone PscG binds PscE and in *Yersinia* the TPR chaperone YscG binds YscE (Figure 1.19). In both structures the interaction between these proteins is very similar, unlike the numerous different homodimers of the Class II chaperones. The two helices of
PscE/YscE interact through hydrophobic interactions with the first two N-terminal TPR containing helices of PscG/YscG. Protein sequence alignment of YscG, PscG and other related proteins showed that these hydrophobic residues are conserved (Figure 1.18). In the case of YscE/G, the total buried surface area of the interaction is 1548 Å². Moreover, the positioning of PscE/YscE in the complex means that it makes very little direct contact with the chaperones cargo, PscF/YscF (Figure 1.19). The total buried surface area between YscF and YscE in the complex is merely 276 Å². Thus it would seem that the role of PscE/YscE is most likely to be that of stabilising PscG/YscG in a form that can readily accept its cargo. The fact that YscE null mutant strains of Yersinia do not accumulate YscG and that PscE null mutant strains of Pseudomonas are non-cytotoxic support this hypothesis [139,173,177].

Figure 1.18 Structure based sequence alignment of T3SS class III chaperones sharing homology from diverse pathogens. PscG from Pseudomonas aeruginosa (nosocomial infections), YscG from Yersinia pestis (bubonic plague), YscG from Vibrio parahaemolyticus (gastrointestinal illness), LscG from Photorhabdus luminescens (insect pathogen), and YscG from Aeromonas hydrophila (respiratory illness in humans, reptiles, and birds). Residues showing 100 % conservation are highlighted in red. PscG homologs share high conservation within residues that interact with PscE (purple) and PscF (orange).
**Cargo Binding:** The main binding site of the heterodimer is located on the hydrophobic concave groove of the TPR domains of PscG/YscG. PscG/YscG acts as a scaffold to organise the assembly of PscE/YscE and PscF/YscF into the heterotrimeric complex. In both cases, the concave surface of PscG/YscG formed by TPR containing α-helices (H1, H3 and H5) and a C-terminal α-helix (H7), display a highly hydrophobic platform, which grasps the C-terminal helix of PscF/YscF. [139,173,177-180]. In YscEFG complex, the C-terminal residues of YscF$_{50-87}$ adopt a α-helical hairpin conformation within the TPR groove. The total buried surface area between YscF and YscG in the complex is 2724 Å$^2$. Thus, the complex that YscG makes with YscF is far more extensive than those made by the class II chaperones with their cargo. In synergy with the structure of YscEFG, the interaction of PscF mirrors that of YscF, but differs at the N-terminus. Here instead of the hairpin-helix conformation, the PscF shows an extended peptide. One explanation for the difference is the PscF construct used was truncated whereas the YscF was full length. Interestingly, further experimentation concluded that the N-terminal section of YscF (1-49) could not be resolved crystallographically as it was unstructured.

![Figure 1.19](image_url)  
*Figure 1.19 Crystal structures of the heterotrimeric complex comprising the class III chaperones: (a) YscEFG complex comprising the heterodimer chaperone of YscE/YscG (coloured blue/red) and needle protein YscF$_{50-87}$ (coloured green). (b) PscEFG complex showing the heterodimer chaperone of PscE/PscG (coloured blue/red) and needle protein PscF$_{54-84}$ (coloured green). Figures were prepared using PyMOL.*
1.4 Thesis Aims

Given the crucial role of the class II translocator chaperones, this study aims to investigate their conformational stability and the thermodynamic importance of the homodimeric interface. To accomplish these aims the class II chaperone LcrH from *Yersinia pestis* was chosen as a model system. LcrH is identical to SycD from *Y. enterocolitica*, except for two minor amino acid substitutions, (Figure 1.15 and Figure 1.17).

Initially we will investigate the oligomeric states of different constructs of LcrH and their secondary structure characteristics. These will be achieved by: (i) constructing the LcrH genes - full length LcrH (*fl*-LcrH), truncated LcrH (*tr*-LcrH) and monomeric L65E LcrH (*tr*-LcrH L65E), (ii) carrying out molecular cloning of constructed LcrH genes into suitable expression vectors for protein expression followed by purification strategies and (iii) initial biophysical characterisation of LcrH protein constructs. The oligomeric states of these proteins will be studied by analytical size exclusion chromatography (SEC) and their secondary structure characterised through far-UV circular dichroism (CD).

Furthermore we will examine the monomer-dimer equilibrium and oligomeric state of various LcrH protein constructs which will be undertaken by the use of analytical ultracentrifugation (AUC), analytical size exclusion chromatography (SEC) and dynamic light scattering (DLS). In addition I intend to explore the thermodynamic stability and flexibility of LcrH proteins through equilibrium chemical denaturations (monitored by far-UV CD and fluorescence spectroscopy).
Chapter 2 Materials and Methods

2.1 Molecular Biology Methods

In this study, three different constructs of LcrH – full length LcrH (fl-LcrH, 168 amino acids), truncated LcrH (tr-LcrH, amino acids 21-163, 143 residues) and mutant L65E LcrH (sequence identical to tr-LcrH, except at amino acid position 65 where leucine was mutated to glutamate) was produced. The techniques used for their gene construction and cloning are described as follows.

2.1.1 fl-LcrH gene synthesis

The fl-LcrH gene was synthesized using the steps outlined below:

2.1.1.1 Recursive PCR

The full length LcrH (fl-LcrH) gene was initially synthesized using a recursive PCR method. This required the design of eight coding and non-coding overlapping oligonucleotides that spanned the entire sequence of the LcrH gene (Figure 2.1). The overlapping regions in the oligonucleotides were 20 base pairs in length and engineered to give melting temperatures ($T_m$) in the range 54 – 64 °C. Computer based sequence analysis using ClustalW multiple sequence alignment of the entire gene sequence with the sequence of the overlap regions was performed. The sequence alignment showed them to be unique and unlikely to give non-specific priming. The designed sequences were synthesised by VH Bio Limited, UK and supplied in lyophilised form. They were dissolved in ddH$_2$O to give a stock concentration of 50 µM.

Figure 2.1: Schematic showing the recursive oligonucleotides required for synthesis of the LcrH gene. The gene is represented as a blue box, the coding oligonucleotides represented as black lines and the non-coding oligonucleotides represented as red lines. Note that each oligonucleotide overlapped with each of its neighbours by 20 bases.
To perform the recursive PCR the overlapping oligonucleotides (0.5 µM final concentration) were added to 10 x reaction buffer (5 µL, Stratagene), dNTPs (0.2 mM of each final concentration), sterile deionised distilled water and *Pfu Turbo* DNA polymerase (1 µL of 2.5 U µL⁻¹ –Stratagene) to give a total reaction volume of 50 µL in a sterile PCR tube. The recursive PCR was carried out starting with 95 °C of initial denaturation for 2 minutes, followed by a cycle repeated 25 times: denaturation at 95 °C for 30 seconds, annealing temperature of 52 °C for 30 seconds, elongation at 72 °C for 30 seconds, and finally 10 minutes of extension at 72 °C in a thermal cycler.

2.1.1.1 Agarose Gel Electrophoresis

0.75 g of Molecular Biology grade agarose (Fisher, UK) was added to 50 ml of 1 x TBE buffer (Tris-Borate-EDTA buffer - 90 mM Boric acid, 2 mM EDTA, 90 mM Tris-HCl, pH 8.0) in a 100 ml conical flask. The mixture in the conical flask was then heated in the laboratory’s microwave oven to prepare a 1.5 % w/v agarose gel. Ethidium bromide was then added to the molten agarose (final concentration 0.5 µg ml⁻¹). The complete mix was added to the gel mould and left to set. DNA samples were prepared by adding 1 µL of 6x gel loading buffer (0.25% Bromophenol Blue, 0.25% Xylene cyanol FF, 15% Ficoll) to 5 µL of DNA preparation. The gel loading buffer contains bromophenol blue dye, which helps increase the DNA sample’s density allowing it to fall into the well of the gel and provides a visible marker to monitor the progress of electrophoresis. Once the gel was set, it was submerged in 1 x TBE buffer and loaded with DNA samples and 100 bp DNA ladder (0.5 µg of 100 bp DNA ladder diluted in loading buffer – NEB) for detecting the band of 500 bp *LcrH* gene. The 100 bp DNA ladder contains standard molecular size marker ranging from 100-1,517 bp. Agarose gel electrophoresis was performed in an electrophoresis tank (BIORAD, UK) by applying a current of 35 mA and voltage of 70 volts. DNA is negatively charged at pH 8.0, hence the application of an electric current caused migration of DNA towards the anode. DNA fragments migrate through the matrix of agarose gel at rates that are inversely proportional to the log₁₀ of the molecular weight of DNA.

2.1.1.2 Polymerase Chain Reaction (PCR)

In order to extract a full length *LcrH* gene (*fl-LcrH*) with restriction sites for further cloning, a conventional PCR was performed using the recursive PCR product as template. Oligonucleotides were designed to match the termini of *fl-LcrH* (coding and non-coding
strands) with melting temperatures ($T_m$) of $\approx 70^\circ$C. The coding oligonucleotide incorporated a $BamHI$ and the non-coding oligonucleotide incorporated $HindIII$ restriction site at their ends. The reverse oligonucleotide also incorporated a stop codon after the gene sequence, but before the $HindIII$ restriction site. The designed sequences were synthesized by Integrated DNA Technologies, Belgium and supplied in lyophilised form. They were dissolved in ddH$_2$O to give a stock concentration of 50 µM.

To perform the PCR the oligonucleotides (0.5 µM final concentration of each) were added to the recursive PCR product (2 µL), 10 x reaction buffer (5 µL, Stratagene), dNTPs (0.2 mM of each final concentration), sterile deionised distilled water and $Pfu$ Turbo DNA polymerase (1 µL of 2.5 U µL$^{-1}$ –Stratagene) to give a total reaction volume of 50 µL in a sterile PCR tube. The PCR was carried out starting with 95 °C of initial denaturation for 2 minutes, followed by a cycle repeated 25 times: denaturation at 95 °C for 30 seconds, annealing temperature of 52 °C for 30 seconds, elongation at 72 °C for 30 seconds, and finally 10 minutes of extension at 72 °C in a thermal cycler. As before, on completion of the PCR, the entire PCR product was run on a 1.5 % agarose gel to confirm the success of the reaction (Section 2.1.1.1.1) showed a clear band of DNA at $\sim$ 500 base pairs. The correct sized DNA was excised from the gel and purified (Section 2.1.3.2).

### 2.1.2 Insertion of synthesised $LcrH$ gene into a cloning vector

The newly synthesized $fl-LcrH$ gene was first incorporated into a cloning vector using the following strategy.

#### 2.1.2.1 Cloning of $l-LcrH$ gene into TOPO TA cloning vector

2.1.2.1.1 Poly-A tailing with Taq polymerase

1 unit of $Taq$ polymerase was added to the PCR reaction from Section 2.1.1.2 and was incubated for 8-10 minutes (no cycle) at 72°C in a thermal cycler. This introduced a poly-A tail (‘A’ overhangs) of the 3’ end of DNA which complements the ‘T’-overhangs already present in the pCR®2.1-TOPO® vector (Invitrogen™). This enables the incorporation of the PCR product into the cloning vector for which the map is shown in Figure 2.2.
Figure 2.2 pCR2.1-TOPO cloning vector contains a number of features for efficient subcloning. The vector is supplied as a linearized vector containing T overhangs for direct ligation of Taq-amplified PCR products. The T7 promoter and M13 forward and reverse primer sites are present for in vitro RNA transcription and sequencing. EcoR I sites flank the PCR product insertion site for easy excision of inserts. The presence of lac promoter and operator allows for easy blue/white colony screening on X-Gal-IPTG agar plates. Kanamycin and ampicillin resistant sites in the vector allow for the choice of selection in E. coli. The vector map was produced using SnapGene Viewer 1.5.3.

2.1.2.1.2 Ligation for TOPO TA cloning
The ligation reaction was performed using 3 µL of the poly-A extended fl-LcrH PCR product, 2 µL of pCR®2.1-TOPO® vector, 1 µL of T4 DNA ligase (NEB, UK), 1 µL of T4 DNA ligase 10 x reaction buffer, 3 µL of sterile deionised distilled water to give total reaction volume of 10 µL. The reaction mixture was incubated at 14 °C for 4 hours.

2.1.2.1.3 Transformation of Ligated fl-LcrH in chemically competent cells
The chemically competent cells used for the transformation of the fl-LcrH gene cloned in the TOPO TA cloning vector was One Shot® TOP10 E. coli cells (Invitrogen™).

One Shot® TOP10 Genotype:

F- mcrA Δ( mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG

2 µL of the ligation mixture was added to 50 µL of One Shot® TOP10 E. coli cells in a sterile microcentrifuge tube and allowed to incubate on ice for 30 minutes. The mixture was
incubated at 42 °C in a water bath for 30 seconds and placed on ice for 2 minutes. 250 µL of LB media was added to the transformed cells and incubated on a shaker (250 RPM) at 37 °C for one hour. After incubation, the cells were centrifuged at 6,000 RPM for 2 minutes; the supernatant was discarded leaving approximately 100 µL of LB media. The cell pellet was resuspended and plated on LB agar plate containing 25 µg/mL of Kanamycin, 20 µg/ml of X-Gal and 0.2 mg/ml of IPTG for blue/white colony screening. The plate was left to incubate at 37 °C overnight and checked the next morning. White colonies represented the E. coli cells that contained the fl-LcrH gene incorporated into the TOPO TA cloning vector.

2.1.2.1.4 DNA extraction and purification
Representative white colonies were picked using sterile toothpicks, and used to inoculate 5 ml of sterile LB broth containing 50 µg/ml of Kanamycin. This was incubated at 37 °C for 14-16 hours on a shaker (250 RPM). Small-scale isolation of plasmid DNA from these bacterial colonies was then performed (known as minipreparation of plasmid DNA (plasmid miniprep)). This process is based on the alkaline lysis method for extraction of plasmid DNA from a relatively small volume of bacteria and was performed using a QIAprep Spin Miniprep Kit (QIAGEN Ltd.). The plasmid miniprep was carried out using the manufacturer’s instructions.

2.1.2.1.5 Estimation of DNA concentration
Plasmid DNA samples obtained from the plasmid miniprep method were diluted using ddH2O for measurement of DNA concentration. A Shimadzu UV-1601 UV-VIS Scanning spectrophotometer was baselined using ddH2O as blank sample. 100 µL of the diluted DNA sample was added to a quartz cuvette and a spectrum was measured from 350 to 220 nm. Three differing dilutions for a given DNA sample were prepared and an average of three measured spectra was taken for ensuring accurate DNA concentration measurement. DNA absorbs at 260 nm, hence this wavelength was used to determine the amount of DNA. The following Equation 2.1 was used to measure the concentration of DNA:

Equation 2.1

\[ DNA \text{ concentration} = A_{260 \text{ nm}} \times \text{dilution factor} \times \text{conversion factor} \]

where, \( A_{260\text{nm}} \) denotes the absorbance at 260 nm; conversion factor is 50 since an absorbance at 260 nm of 1.0 corresponds to 50 ng ml\(^{-1}\) of double stranded DNA, the
resultant DNA concentration is measured in ng ml\(^{-1}\). A\(_{260nm}\) to A\(_{280nm}\) ratio of 1.6 or above indicated pure, contamination free DNA sample.

2.1.2.1.6 Analytical Restriction Digestion with EcoRI

The fl-LcrH gene would be incorporated between EcoRI sites in the pCR\(^\circ\)2.1-TOPO\(^\circ\) vector on successful ligation (the EcoRI sites flank the PCR product insertion site in the vector). Hence, the purified plasmid miniprep DNA samples were digested with EcoRI restriction enzyme to confirm positive clones. The digestion was carried out using 8 \(\mu\)L of plasmid DNA, 0.5 \(\mu\)L of EcoRI restriction enzyme (NEB, UK), 0.5 \(\mu\)L of sterile deionised distilled water, and 1 \(\mu\)L of NEB buffer 2 to give a total reaction volume of 10 \(\mu\)L. The reaction mixture was incubated at 37 °C for 3 hours. The digested bands were visualized on a 1.5 % agarose gel (Section 2.1.1.1.1). Presumptive positive clones were sent for DNA sequencing (Section 2.1.2.1.7) and sequenced using M13 reverse primer.

2.1.2.1.7 DNA sequencing

DNA samples each with a volume of 15 \(\mu\)L of miniprep DNA at a concentration of 100 ng \(\mu\)L\(^{-1}\) were prepared in labelled microcentrifuge tube and sent with the correct sequencing primer. DNA sequencing was carried out by Lark Technologies, Essex, UK (now Beckman Coulter Genomics Ltd). The resultant DNA chromatograms were viewed and analyzed using ChromasPro software version 1.34 (Technelysium Pty Ltd.).

2.1.3 Insertion of synthesised fl-LcrH gene into an expression vector

Two expression vectors were used for expressing fl-LcrH. These were pProEX-HTb (Invitrogen) and pTrc-GST (designed vector obtained from Prof. A.R. Fersht and based on the pRSETA vector, Life Technologies). These are shown in Figure 2.3. The fl-LcrH gene was inserted into these vectors between the restriction sites BamHI and HindIII.
Figure 2.3: pProEX-HTb (a) and pTrc-GST (b) vectors. The vectors contain a number of features including, an ampicillin resistance gene (shown in green colour) that permits the selection of cells carrying the expression vector and multiple cloning site (MCS) shown in orange. The Trc promoter sequence contains part of the tryptophan promoter, part of the lactose promoter and the lactose repressor gene, thus allowing high level of expression when induced with isopropyl β-D-1-thiogalactopyranoside (IPTG). Vector maps were produced using SnapGene viewer version 1.5.2.

2.1.3.1 Restriction digest of fl-LcrH gene

Fl-LcrH in the TOPO vector and each expression vector were digested separately using the restriction enzymes BamHI and HindIII. Due to the star activity of BamH1 a sequential restriction digestion was carried out. 16 µL of gel purified DNA (PCR product or plasmid vector DNA), 2 µL of NEB buffer 2 (NEB), 1 µL of HindIII restriction enzyme (NEB), and incubated for 3 hours at 37 °C. The HindIII was then heat deactivation by incubating at 65 °C for 20 minutes and subsequently placed on ice. 0.4 µL of NaCl (5mM stock solution NEB), 0.2 µL of BSA (1 x final concentration from 100 X stock solution, NEB), 1 µL of BamHI restriction enzyme (NEB), 0.4 µL of sterile ddH2O was added to the reaction mixture to give a total volume of 20 µL and incubated for 3 hours at 37 °C. BamHI was deactivated by heating the reaction mixture at 65 °C for 20 minutes. The digested DNA samples were visualized on 1.5 % agarose gel (Section 2.1.1.1.1) and DNA bands of correct size were excised and gel purified as below.
2.1.3.2 Purification of DNA from an Agarose Gel

The correct sized DNA band visualized on agarose gel was excised and placed in a PCR tube (the PCR tube was punctured with a hole at the bottom and filled with glass wool), and was placed into a microcentrifuge tube. The lid of the microcentrifuge tube was punctured and the PCR tube containing the excised gel was inserted to fit tightly at the top. This was centrifuged for 10 minutes at 13,000 RPM to separate the DNA from the gel. The gel was held on the glass wool with the DNA in solution collected at the bottom of the microcentrifuge tube. This DNA was then purified from this solution via DNA precipitation using the Pellet Paint Co-Precipitant (Novagen) as per the manufacturer’s instructions. The DNA was then re-solubilised in 20 µL of sterile ddH₂O.

2.1.3.3 Ligation

The gel purified digested insert and vector were ligated in a ratio of 3:1 (insert: vector) in terms of volume. The ligation reaction was performed by adding 6 µL of the gel purified insert, 2 µL of the digested vector DNA, 0.5 µL of sterile ddH₂O, 1 µL of T4 DNA ligase 10 x reaction buffer (NEB, UK), 0.5 µL of T4 DNA ligase (NEB, UK) to give a total reaction volume of 10 µL. The reaction mixture was incubated at 4 °C for 14-16 hours. The ligase was deactivated by incubating the reaction mixture at 65 °C for 20 minutes

2.1.3.4 Transformation using Electro-competent cells

The ligation mixture was transformed in electrocompetent XL1-Blue E. coli cells (Agilent Technologies). The XL1-Blue strain of E. coli allows blue-white colour screening for recombinant plasmids, is an excellent host strain for routine cloning applications using plasmid or lambda vectors and helps in preparation of high quality plasmid DNA.

XL1-Blue Genotype:

\n
recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F’ proAB lacIq ΔM15 Tn10 (Tet’)]

(Genes listed signify mutant alleles. Genes on the F’ episome, however, are wild type unless indicated otherwise).

In order to perform transformation, 1 µL of the ligation mixture was added to 30 µL of XL1-Blue cells (in 10 % glycerol), mixed well and added to 0.2 cm electroporation cuvette (BIORAD). The cuvette was placed into the chamber slide and electroporated using the Bio-Rad Micropulser™. A pre-programmed electroporation setting called EC2 (voltage of 2.5
kV) was used, as this was the manufacturer’s recommended setting. A voltage delivery time of approximately 5 milliseconds after pulsing was an indicator of efficient electroporation. Once pulsed, 1 ml of LB broth was immediately added to the sample and the mixture incubated at 37 °C for 1 hr. The culture produced from the transformation was plated out (=50 µL) onto LB-Agar plates containing ampicillin (100 µg ml⁻¹ final concentration) and incubated overnight at 37 °C. Single bacterial colonies were picked the following day, inoculated in 5 ml LB broth containing 100 µg ml⁻¹ ampicillin and incubated at 37 °C on a shaker (250 RPM) for 14-16 hours.

2.1.3.5 Confirmation of DNA Clones
To determine whether the ligation had been successful, the isolated DNA was restriction digested (Section 2.1.3.1, except the reaction volume was 10 µL) and run on a 1.5 % agarose gel (Section 2.1.1.1.1). A positive result showed a DNA band that corresponded to the size of the LcrH gene (500 bases). The DNA that produced positive restriction digests were then sent for DNA sequencing (Section 2.1.2.1.7).

2.1.4 Synthesis & insertion of truncated LcrH (tr-LcrH) gene into pTrc-GST vector
The shortened version, tr-LcrH was synthesized by performing a conventional PCR using fl-LcrH in pProEXHTb vector as a template. Oligonucleotides were designed to match the termini of tr-LcrH (coding and non-coding strands) with melting temperatures (T_m) of 67–74 °C. The coding oligonucleotide incorporated a BamHI and the non-coding oligonucleotide incorporated HindIII restriction site at their ends. The reverse oligonucleotide also incorporated a stop codon after the gene sequence, but before the HindIII restriction site. The designed sequences were synthesized by Integrated DNA Technologies Inc. and supplied in lyophilised form. They were dissolved in ddH₂O to give a stock concentration of 50 µM. The PCR reaction was performed and success confirmed as described in Section 2.1.1.2.

The correctly sized tr-LcrH PCR product was then incorporated into the pTrc-GST vector by a simple restriction digest of vector and gene, followed by DNA extraction, ligation and transformation (Section 2.1.2.1.3). The success was confirmed with analytical restriction digest and DNA sequencing (Section 2.1.3.5).
2.1.5 Synthesis of LE65 mutant of tr-LcrH

In order to create the monomeric L65E LcrH construct, site directed mutagenesis was performed on the pTrc-GST expression vector containing the tr-LcrH gene. A point mutation was carried out in the tr-LcrH gene, where the 65th amino acid Leucine was mutated to Glutamic acid. Oligonucleotides for the coding and non-coding strand spanning the point mutation site in tr-LcrH were synthesized (Integrated DNA Technologies, Inc.) with Tm of 78 °C. Site directed mutagenesis PCR was performed using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene). The formula commonly used for estimating the Tm for the oligonucleotides is shown in Equation 2.2 (QuikChange® Site-Directed Mutagenesis kit instruction manual)

\[
T_m = 81.5 + 0.41(\% GC) - \frac{675}{N} - \% mismatch
\]

For calculating Tm: N is the primer length in bases, values for % GC and % mismatch are whole numbers.

2.1.5.1 Site Directed Mutagenesis

Site directed mutagenesis was carried out using 1.25 µL each of oligonucleotides for the coding and non-coding strand (125 ng each), 2 µL of template (10 ng of tr-LcrH-pTrc-GST plasmid vector), 5 µL of 10 x reaction buffer (Stratagene), 1 µL of dNTPs (10mM mix), 38.5 µL of ddH2O and 1 µL of Pfu turbo DNA polymerase (2.5 U/µL –Stratagene) to give a total reaction volume of 50 µL in a sterile PCR tube.

The mutagenesis PCR was performed starting with 95 °C of initial denaturation for 30 seconds, followed by a cycle repeated 18 times: denaturation at 95 °C for 30 seconds, annealing temperature of 55 °C for 30 seconds, elongation at 68 °C for 5 minutes 30 seconds (1 minute per kb of plasmid length) and finally 5 minutes of extension at 68 °C in a thermal cycler. 1 µL of DpnI restriction enzyme (10 U/µL) was added to the reaction mixture after the completion of PCR, thoroughly mixed, spun down in a microcentrifuge tube for 1 minute and incubated at 37 °C for 1 hour. This DNA was used to transform XL1-Blue electrocompetent cells (Section 2.1.3.4). Colonies were picked for growing bacterial cultures overnight, followed by isolation and purification of plasmid DNA (Section 2.1.3.2). The
plasmid DNA samples were subjected to analytical restriction digestion with *BamH*I and *HindIII* (Section 2.1.3.1) except that the reaction volume was 10 µL. The restriction digestion was analysed on a 1.5% agarose gel. DNA samples showing the correct sized bands were selected, sent for DNA sequencing (Section 2.1.2.1.7) and sequencing was carried out using pTrc-His reverse primer.

### 2.2 Protein Production and Purification

LcrH proteins fl-LcrH, tr-LcrH and monomeric L65E LcrH were produced using the protocol described below.

#### 2.2.1 Protein Production

BL21 C41 (DE3) Electrocompeent strain of *E. coli* cells were transformed (Section 2.1.2.1.3) with expression plasmid vectors pTrc-GST and pProEX-HTb, both carrying the *LcrH* gene respectively. BL21 C41(DE3) cells are *E. coli* strains that are effective in expressing toxic proteins from all classes of organisms, including bacteria, yeast, plant, viruses, and mammals. The C41 strains contain genetic mutations phenotypically selected for conferring tolerance to toxic proteins. The strain C41 (DE3) was derived from BL21 (DE3). This strain has a mutation that reduces the level of T7 RNAP activity, thereby preventing cell death associated with overexpression of many recombinant toxic proteins [178-181].

**BL21 C41 (DE3) Genotype:** F⁻ *ompT hsdSB (rB⁻ mB⁻) gal dcm* (DE3)

Single colonies containing the *LcrH* gene were picked and grown in 10 ml of LB medium containing final concentration of 50 µg/ml ampicillin overnight at 37 °C on a shaker. The 10 ml overnight grown bacterial culture was used to inoculate 1L of 2 x YT + ampicillin in a 2L flask and was grown at 37 °C with constant shaking at 200 RPM until the absorption of the cells at 600 nm (A₆₀₀) reached 0.6 – 1.0 (= 3 – 4 hours). The absorption was monitored using a UV-2401PC UV-Vis recording spectrophotometer. Isopropyl-thio-β-D-galactopyranoside (IPTG) (Melford Laboratories Ltd., UK) was added to a final concentration of 0.1 mM to induce protein expression. The temperature was lowered to 18 °C for improving the stability and solubility of the protein undergoing expression and placed on the shaker overnight (200 RPM). The cells were harvested after 16-18 hours post induction in a
Beckman Avanti J-30I centrifuge and using the JLA 9.1000 rotor for 10 minutes at 4,000 x g at 4 °C.

2.2.1.1 Cell Lysis for Protein Extraction

2.2.1.1.1 Cell Lysis by Sonication
The harvested bacterial cell pellet was resuspended in 40 ml of ice-cold 1x PBS containing 10 % glycerol. The cells were lysed by sonication using a Sonics Vibracell Ultrasonic Processor (Sonics & Materials Inc.). The mixture was retained on ice, whilst 6 cycles of 15 seconds on, 45 seconds off, with an output of 30 volts were applied, until the cell mixture became non-viscous and darker in colour. The cell debris was then removed by centrifugation at 33,000 x g in Beckman Avanti J-30I centrifuge, in a JA-30.50 rotor for 1 hour at 4 °C.

2.2.1.1.2 Cell Lysis by Homogenisation
The suspension of harvested bacterial cells in ice cold 1X PBS with 10 % glycerol, 5 mM DTT was filtered through a sieve and poured into the inlet of high pressure homogenizer Emulsiflex-C3 (Avestin Inc., Canada). The cell lysis was carried out under pressure of 20,000 lb/in² with nitrogen gas according to the manufacturer’s instructions. The cells were lysed by two passages through the Emulsiflex-C3 until the lysate became non-viscous and homogenous. The cell lysate was then centrifuged at 33,000 x g for 1 hour at 4 °C using Beckman-Coulter JA-30.50 rotor.

2.2.1.2 Purification by Affinity Chromatography
The supernatant protein extracted after cell lysis and subsequent centrifugation was filtered through a 0.2 µm syringe filter (Sartorius Stedim Biotech GmbH, Germany) and subjected to affinity purification. Affinity purification of LcrH proteins was carried out by two methods depending on the expression plasmid used for protein production i.e. pProEX HTb plasmid containing the *fl-LcrH* gene produced a hexahistidine tagged protein whereas pTrc-GST plasmid produced GST tagged proteins:

1) Ni²⁺ affinity chromatography – pProEX HTb plasmid containing the *fl-LcrH* gene produced a hexahistidine-tagged fl-LcrH protein. His-tagged fl-LcrH protein was purified by using Ni²⁺ charged column packed with Iminodiacetic acid agarose resin (Sigma, UK). The resin was
charged with Ni\(^{2+}\) by passing 2 column volumes of 100 mM NiSO\(_4\) and the unbound NiSO\(_4\) was removed with a column volume of 1x PBS.

2) GST-affinity chromatography was used to purify fl-LcrH, tr-LcrH and monomeric L65E LcrH proteins produced from pTrc-GST vector containing the respective genes. GST-tagged proteins were purified by using a column packed with glutathione agarose resin called Super Glu – Glutathione Superflow (Generon, UK).

### 2.2.1.3 Buffer Solutions

All buffer solutions were made with deionised distilled H\(_2\)O (resistance of 18.2 \(\Omega\)) and vacuum filtered using 0.2 \(\mu\)m membrane filter (Whatman) prior to usage. The names of buffers and their respective compositions are shown in Table 2.1.

**Table 2.1: Buffer solutions and their compositions**

<table>
<thead>
<tr>
<th>Name of buffer</th>
<th>Buffer composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>His-tag Wash buffer</td>
<td>1x PBS (137 mM NaCl, 10 mM phosphate, 2.7 mM KCl), pH 7.4</td>
</tr>
<tr>
<td>His-tag Elution buffer</td>
<td>50 mM phosphate, 150 mM NaCl, 250 mM Imidazole, pH 7.0</td>
</tr>
<tr>
<td>GST Wash buffer</td>
<td>20 mM phosphate, 150 mM NaCl, pH 7.2, 5 mM DTT</td>
</tr>
<tr>
<td>GST Elution buffer</td>
<td>20 mM phosphate, 150 mM NaCl, 10 mM reduced glutathione, pH 7.2, 5mM DTT</td>
</tr>
</tbody>
</table>

### 2.2.1.4 Ni\(^{2+}\) Affinity chromatography and cleavage of His-tagged proteins

The filtered supernatant was loaded onto a Ni\(^{2+}\) affinity column pre-equilibrated with his-tag wash buffer. After passing the supernatant protein through the Ni\(^{2+}\) affinity column (column volume = 10 ml), it was washed with 20 column volumes of his-tag wash buffer. The protein was eluted with his-tag elution buffer. After elution, the column was washed with 2 column volumes of 6 M GdmHCl to unfold any residual protein bound to the column. The column
was then washed with 5 column volumes of ddH₂O, followed by 2 column volumes of 20 % ethanol and stored at 4 °C. The eluted protein was cleaved overnight at 25 °C by addition of TEV protease to remove the fused N-terminal His-tag. The affinity purification and cleavage of LcrH was monitored by visualisation on SDS-PAGE (Section 2.2.2.2).

2.2.1.5 GST Affinity purification and cleavage of GST-tagged proteins
The filtered supernatant was loaded onto the GST affinity column (column volume = 10 ml) and washed with 20 column volumes of GST wash buffer. LcrH – GST fusion protein was cleaved on the resin overnight at 25 °C by addition of thrombin protease (GE Healthcare, UK). The cleaved LcrH was eluted as a flow through in the GST wash buffer. The bound GST-tag was removed from the column by the addition of GST elution buffer. The column was subsequently washed with two column volumes of 6 M GdmHCl to unfold and remove any residual proteins bound to the column. The column was immediately washed with 5 column volumes of ddH₂O, followed by 2 column volumes of 20 % ethanol and stored at 4 °C.

2.2.1.6 Purification using Size Exclusion Chromatography by AKTA FPLC
The final step in the purification used size exclusion chromatography using AKTA FPLC system (GE Healthcare) by loading the cleaved LcrH protein fractions onto a HiLoad™ 16/60 Superdex™ 200 gel filtration column (GE Healthcare, UK) with a bed volume of 120 ml. The column was equilibrated with 180 ml of 50 mM phosphate buffer, 5 mM DTT, pH 7.0. The cleaved LcrH protein fractions were concentrated using Amicon® Ultra Centrifugal Filter units (Millipore) and filtered through 0.2 µm syringe filter (Sartorius Stedim Biotech GmbH, Germany). The protein samples (under 10 ml volume) were loaded onto the gel filtration column with a 50 ml superloop (GE Healthcare). 1 column volume of 50 mM phosphate buffer, 5 mM DTT, pH 7.0 was applied to the column at a flow rate of 1 ml/min. Fractions were collected in 2.5 ml aliquots, in 30 mm tubes.

2.2.2 Protein Purity
Mass spectrometry (Section 2.2.2.1) and SDS-PAGE (Section 2.2.2.2) of the proteins confirmed both the size and checked the purity, respectively. In all cases the purity of LcrH proteins was greater than 95 %. The nucleic acid contamination of the protein was checked by the 280/260 nm ratio. A ratio of 1.6 or above indicated no contamination.
2.2.2.1 Mass Spectrometry

Sample and Matrix preparation

A fresh saturated matrix solution for MALDI-ToF was prepared in a microcentrifuge tube by dissolving Sinapinic acid (Formula weight = 224.21) in 50 % acetonitrile, 0.1 % TFA in ddH₂O. The mixture was vortexed for 1 minute and centrifuged for 20 seconds to remove any undissolved matrix from the solution and the clean top matrix was used. Protein samples of 30 µM concentration were given ZipTip® (Millipore) treatment as per manufacturer’s instructions. Sample and matrix solutions were mixed in the ratio of 1: 4 in a microcentrifuge tube. 0.5 µL of the mixture was immediately loaded onto a MALDI plate and allowed to dry at room temperature prior to insertion into the mass spectrometer.

Mass spectra were recorded on a Micromass TofSpec mass spectrometer (Micromass, Manchester, UK), equipped with nitrogen UV laser (337 nm with a 4 nm pulse width).

2.2.2.2 SDS-Polyacrylamide Gel Electrophoresis

All solutions were made as described in the laboratory manual [181,182]. SDS-PAGE was carried out according to standard protocols using the Laemmli system [182,183]. LcrH protein samples were resuspended in SDS loading buffer (50 mM Tris-HCl, pH 6.8, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol), boiled for 10 minutes and loaded on an 18% SDS-PAGE gel for analysis.

2.2.3 Determination of Protein Concentration

Protein concentration was determined spectrophotometrically, using the UV-2401PC-UV-Vis recording spectrophotometer, according to Equation 2.3.

Equation 2.3

\[ C = \left( \frac{A_{280} - A_{320}}{\varepsilon l} \right) \times \text{dilution factor} \]

Where \( C \) is the protein concentration in M. \( A_x \) is the absorbance at x wavelength. \( \varepsilon \) is the molar extinction coefficient, and \( l \) is the cuvette path length in cm.
The molar extinction coefficient at 280 nm wavelength, $\varepsilon_{280 \text{ nm}}$, is calculated by Equation 2.4 [183,184]:

Equation 2.4

$$\varepsilon_{280 \text{ nm}} = (\# \text{ Trp}) (5500) + (\# \text{ Tyr}) (1490) + (\# \text{ Cys}) (125)$$

Where ‘# Trp’ refers to the number of tryptophan residues, ‘# Tyr’ refers to the number of tyrosine residues and ‘# Cys’ refers to the number of cysteine residues in the protein. The molar extinction coefficient ($\varepsilon_{280 \text{ nm}}$) for the LcrH proteins are as follows; for fl-LcrH, 12,420 M$^{-1}$ cm$^{-1}$, tr-LcrH and mutant L65E LcrH, 10,805 M$^{-1}$ cm$^{-1}$.

### 2.2.4 Storage of Purified Protein

The LcrH proteins were dialysed into 50 mM Tris, pH 8.0, 5 mM DTT overnight at 4 °C. They were then concentrated using centrifugal concentrators (Millipore) to ≈ 100-200 µM and 1 ml aliquots in micro-centrifuge tubes were rapidly frozen using liquid nitrogen. The samples were then stored at -80 °C.

### 2.3 Biophysical Characterization of LcrH proteins

In all experiments, 50 mM phosphate, pH 7.0, 5 mM DTT buffer was used. The molarity of the stock solutions of urea was determined using a refractometer (model NAR-1T, Atago, Tokyo).

#### 2.3.1 Multimeric state of LcrH proteins using Analytical Size Exclusion Chromatography (SEC)

In order to determine the multimeric state of LcrH proteins, analytical SEC was performed on an AKTA FPLC system (GE Healthcare). This was done by injecting 100 µL of protein at concentrations from 3 µM to 100 µM in 50 mM Phosphate buffer, pH 7.0, 5 mM DTT onto a Superdex$^\text{TM}$ 75 10/30 column and Superdex$^\text{TM}$ 200 10/30 column (GE Healthcare). The column was pre-equilibrated with LcrH buffer before injection of protein samples. 1 column volume of 50 mM Phosphate buffer, pH 7.0, 5 mM DTT, was applied at a flow rate of 0.5 ml/min. For the SEC experiments involving urea denatured protein samples, the column
and protein sample was pre-equilibrated for 4 hours and run with the required concentration of urea.

2.3.2 Analytical Ultracentrifugation (AUC)

Two types of sedimentation experiment can be performed using an analytical ultracentrifuge; sedimentation equilibrium and sedimentation velocity. Sedimentation velocity experiments involve spinning molecules at a high speed and looking at the rate at which the molecules move in response to the centrifugal force that is generated. From the rate at which the molecules move, a sedimentation coefficient can be determined (as the centrifugal force is known). Information regarding the shape and sometimes aggregation/oligomerisation state can be determined, although the latter parameter is more easily determined by sedimentation equilibrium experiments. In contrast, sedimentation equilibrium experiments involve spinning molecules at moderate speeds so that centrifugal force causes the protein to sediment gradually, creating a concentration gradient across the cell, whilst diffusive forces act to oppose the sedimentation forces. Eventually these two forces are in equilibrium and so no net movement of the concentration gradient is observed. This gradient can then be measured and from knowing how this concentration gradient is affected by changing rotor speed, molecular masses (and hence oligomerization states) can be determined. Masses can be determined accurately as, at equilibrium, the concentration distribution is only dependent on the mass of the sedimenting species and not on its shape (unlike for sedimentation velocity experiments). If a species is self-associating, dissociation constants can often be obtained if they are in the range measurable by the instrument (usually between 1 μM – 0.5 mM, depending on the absorbance of the molecule in question). All experiments performed on this system are sedimentation equilibrium experiments as this is the method used when an accurate, shape-independent molecular weight is required.

Aimee Boyle from School of Chemistry, University of Bristol performed sedimentation equilibrium experiments on the analytical ultracentrifuge. The experiments were conducted using a Beckman-Optima XL-I analytical centrifuge fitted with an An-60 Ti rotor. Solutions were prepared at 40 μM protein concentration in 50 mM Phosphate buffer, pH 7.0, 5 mM DTT. Data was recorded at 25 °C using speeds in the range 17,000 – 30,000 rpm. Datasets
were initially fitted to a single, ideal species model using the program Ultrascan [184,185] which employed the Equation 2.5.

Equation 2.5

\[ C(x) = \exp \left\{ \frac{\ln(A)+M \omega^2 (1-\nu \cdot D) (x^2-x_r^2)}{2 R T} \right\} + B \]

Where \( C(x) \) is the concentration of peptide as a function of radial position (M), A is the measured absorbance of the sample, M is the molar mass of the species (Da), \( \omega \) is the angular velocity of the rotor (rad s\(^{-1}\)), \( \nu \) is the partial specific volume of the protein in question (ml g\(^{-1}\)), D is the solvent density (g ml\(^{-1}\)), \( x^2 \) is the radial position (cm\(^2\)), \( x_r^2 \) is the reference radial position (cm\(^2\)), R is the gas constant (J mol\(^{-1}\) K\(^{-1}\)), T is the temperature (K) and B is a baseline offset which compensates for any non-ideality in the data acquisition or the cells.

The partial specific volume of the peptide (0.7341 cm\(^3\) g\(^{-1}\)) and the density of the solvent (1.0042 g cm\(^{-3}\)) were calculated using Sednterp. A monomer-dimer equilibrium model (Equation 2.6) was used to determine the \( K_D \) of the species. A fit was deemed appropriate if the residuals from the fit were randomly scattered and had an absorbance close to zero. Any trends in the residuals indicated a different model was required.

The monomer-dimer equilibrium model was used for determining the dimerization dissociation constant \( K_D \) after performing equilibrium sedimentation is shown in Equation 2.6.

Equation 2.6

\[ C_T = C_M(R_0) \exp \left[ M \varphi (R^2 - R_0^2) + \left( \frac{2}{K_D \cdot M} \right) [C_M(R_0)]^2 \exp \left[ 2M \varphi (r^2 - r_0^2) \right] \right] \]

Where \( C_T \) is the total protein concentration, \( R_0 \) is the reference radius (cm), \( R \) is the radius (cm) and \( C_M(R_0) \) is the monomer concentration at \( R_0 \), \( M \) is the molecular weight of monomer, \( \varphi = \frac{\omega^2 (1-\nu \cdot D)}{2RT} \), \( \omega \) is the angular velocity of the rotor (rad s\(^{-1}\)), \( \nu \) is the partial specific volume of the protein in question (ml g\(^{-1}\)), D is the solvent density (g ml\(^{-1}\)), R is the gas constant (J mol\(^{-1}\) K\(^{-1}\)) and T is the absolute temperature (K).
The protein concentration \( C_T \) versus radius \( r' \) plots were analysed to determine \( K_D \) of the species according to the above theoretical Equation 2.6 for monomer-dimer equilibrium. Least squares curve fitting was performed using the Ultrascan software. Errors were obtained by Monto Carlo analysis and given with 5 % confidence limit.

### 2.3.3 Dynamic Light Scattering (DLS)

Dynamic light scattering (DLS), also known as quasi-elastic light scattering (QELS) and photon correlation spectroscopy (PCS), is a non-invasive, well-established technique in the field of protein crystallography. DLS is used to measure hydrodynamic sizes, polydispersities and aggregation effects of protein samples in solution [185,186]. Apart from proteins, its applications include emulsions, micelles, polymers, nanoparticles or colloids. It measures the size and size distribution of molecules in the submicron region, and with the latest technology lower than 1nm. In DLS, a laser beam illuminates the sample and the Brownian motion of particles or molecules in suspension scatters the laser light at different intensities. Analysis of these intensity fluctuations yield the velocity of the Brownian motion and hence the particle size using the Stokes-Einstein relationship [186,187] (Equation 2.7).

**Equation 2.7**

\[
R_h = \frac{k_B T}{6\pi\eta D}
\]

Where, \( R_h \) is the hydrodynamic radius, \( k_B \) is Boltzmann constant, \( T \) is temperature, \( \eta \) is the solvent viscosity and \( D \) is the translational diffusion coefficient (obtained through the velocity of Brownian motion by DLS).

DLS measurements were carried out using Zetasizer Nano S instrument (Malvern Instruments Ltd). Protein samples were prepared at concentrations of 50 µM in 50 mM phosphate buffer, pH 7.0, 5 mM DTT. For experiments involving urea denatured protein samples, the protein sample was pre-equilibrated for 4 hours and run with the required concentration of urea. All samples were spun at 10,000 rpm for 10 minutes before taking measurements. The results were analyzed by Zetasizer software version 6.12 (Malvern Instruments Ltd).
2.4 Secondary and Tertiary Structure Characterization

2.4.1 Circular Dichroism (CD) Spectroscopy

Circular Dichroism (CD) Spectroscopy is a spectroscopic technique where the CD of molecules is measured over a range of wavelengths. CD is the difference in the absorption of left and right handed circularly polarised light in chiral (structurally asymmetrical) molecules and occurs when a molecule contains one or more chiral chromophores (light-absorbing groups). Depending on the CD instrument, the data are recorded as either as the difference in absorbance of right- and left-handed circularly polarised light, \( \Delta A(\lambda) = A_L(\lambda) - A_R(\lambda) \), or as ellipticity, \( \theta_{obs} \) in degrees or millidegrees. Data in both formats can be converted to molar values, that is, to the differential molar circular dichroic extinction coefficient, \( \Delta \varepsilon = \varepsilon_L - \varepsilon_R \) and to the molar ellipticity, [\( \theta \)]. \( \Delta \varepsilon \) and [\( \theta \)] are related by Equation 2.8.

Equation 2.8

\[
[\theta] = 3300 \ \Delta \varepsilon
\]

It should be noted that the concentration standards are different for [\( \theta \)] and \( \Delta \varepsilon \). \( \Delta \varepsilon \) is the differential absorbance of a 1 mol/l solution in a 1 cm cell, whereas [\( \theta \)] is the rotation in degrees of a 1 dmol/cm\(^3\) solution and a path length of 1 cm.

The principle of CD involves a monochromatic beam of light passing through a Photo Elastic Modulator (PEM) that converts the linear polarised light into alternating left and right handed polarised light. The two polarisations are differently absorbed by the chiral sample, and the detector called Photo Multiplier Tube (PMT) [187-189], detects this difference in absorption and conveys the CD signal. Along with the CD signal, the CD spectropolarimeter also records the High Tension (HT) voltage of the detector. This HT voltage is a measure of the amount of voltage the detector is using and is dependent on the amount of light striking the detector (more light – lower HT voltage). Furthermore, HT voltage is roughly proportional to the absorbance of the sample. If the HT voltage goes above ~800 Volts the noise will become disproportionate to the sample signal and the output will be completely unreliable (Jasco J-715 Spectropolarimeter Operating Manual).

CD spectroscopy is widely used to study a range of chiral molecules, but its most important application is in the study of large biological molecules. Largely, it is used for analysis of the
secondary structure or conformation of proteins, as secondary structure is sensitive to environment, temperature or pH. CD can be used to monitor secondary structural changes with environmental conditions or interaction with other molecules. Structural, kinetic and thermodynamic information about proteins can be derived from CD spectroscopy [188-190].

Far-UV (spectral region 190-250 nm) CD was used to probe the secondary structure of the LcrH protein at different protein concentrations and denaturant concentrations. Spectra were acquired on either a Jasco J-715 CD Spectropolarimeter (JASCO, UK) or a Chirascan™ CD Spectrometer (Applied Photophysics Ltd, UK). All far-UV CD spectra were acquired in LcrH buffer, at 25 °C. The following pathlengths were used: 10 or 5 mm pathlengths for protein concentrations at 3 µM; 1, 2 or 5 mm path lengths for proteins concentrations at 6, 12 and 25 µM and 1 mm path length for protein concentrations at 50 µM and 80 µM. Spectra were recorded from 200 to 300 nm. To compare data, and thus correct for pathlength and concentration, all ellipticities were converted to molar ellipticity (degrees cm² dmol⁻¹), shown in Equation 2.9.

Equation 2.9

\[ [\theta]_{\text{molar}} = \left( \frac{100 \times \theta_{\text{obs}}}{d \times m} \right) \]

Where, \([\theta]_{\text{molar}}\) is molar ellipticity in deg cm² dmol⁻¹, \(\theta_{\text{obs}}\) is the observed CD signal in millideg, \(d\) is the cuvette path length in cm and \(m\) is molar concentration of protein.

### 2.4.2 Fluorescence Spectroscopy

Fluorescence is a phenomenon where a molecular system absorbs, then emits light. During absorption, high energy (short wavelength) light excites the system, promoting electron within the molecule to move from the ground state, to the excited state [190,191]. Emission is observed when the excited electron returns from the excited state back to the ground state. As some energy is always lost by non-radiative processes, such as vibrational transitions, the energy of the emitted light is always lower than that of the absorbed light [191,192]. The energies and relative intensities of the fluorescence signals give an indication about the structure and environment of the fluorophores (substances that exhibit fluorescence).
Fluorescence spectrophotometer also called Spectrofluorimeter or Fluorimeter is a device for measurement of fluorescence of proteins that contain intrinsic fluorophores (naturally occurring fluorophores) such as tryptophan, tyrosine and phenylalanines. A typical fluorescence spectrophotometer contains a sample holder, incident photon source (a xenon lamp), monochromators for selecting particular incident wavelengths, focussing optics, photon-collecting detector and a control software unit. The detector is usually set at 90° to the light source. A fluorescence emission spectrum is recorded when the excitation wavelength is held constant and the emission beam is scanned as a function of wavelength. A fluorescence spectrophotometer detects the changes in fluorescence intensity, which enables the study of the tertiary structural changes of proteins or binding of two molecules. For example, the wavelength of tryptophan fluorescence can be used to determine whether a tryptophan is in an aqueous environment (longer wavelength) or buried deep within the protein (shorter wavelength) [192,193].

All measurements were performed in a thermostatted cuvetter holder at 25 °C using a Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies). Tyrosine fluorescence was employed due the lack of Tryptophan residues in LcrH. The excitation wavelength was 274 nm and band passes for excitation and emission was 5 nm. The fluorescent emission of the LcrH protein was measured from 294 to 400 nm. The λ\text{max} of 303 nm (emission wavelength for tyrosine) was used in analysis unless otherwise stated.

### 2.5 Equilibrium Studies for measuring Thermodynamic Stability

#### 2.5.1 General Background

The thermodynamic stability of proteins can be investigated by performing unfolding studies under equilibrium conditions. These experiments involve unfolding the protein by varying one parameter (concentration of chemical denaturant, temperature, pH and pressure) and utilising a probe (fluorescence and circular dichroism) to measure the equilibrium produced. In this study, chemical denaturations were performed using urea and circular dichroism (Section 2.4.1) was used to monitor the changes in secondary structure. Fluorescence spectroscopy (Section 2.4.2) was used to measure the changes in tertiary structure. The far UV spectra from 200 to 300 nm monitored the secondary structure of the
protein. LcrH proteins are alpha helical containing three TPRs in each monomer, thus depicted in a negative peak at 222 nm.

2.5.2 Equipment and General Procedures

All experiments, unless otherwise stated, were carried out at 25 °C in 50 mM Phosphate buffer, pH 7.0, 5 mM DTT. All stock solutions of urea were made using volumetric flasks, and the urea solutions were flash frozen and stored at -20 °C to prevent degradation. All protein solutions were filtered 0.2µm syringe filter prior to addition to denaturant solutions. GdmHCl and urea solutions were checked by a refractometer (model NAR-1T, Atago, Tokyo), using Equation 2.10 and Equation 2.11 respectively [168,193].

Equation 2.10

\[
Molarity = 57.1 (\Delta N) + 38.7 (\Delta N)^2 - 91.6 (\Delta N)^3
\]

Equation 2.11

\[
Molarity = 117.7 (\Delta N) + 29.7 (\Delta N)^2 + 185.5 (\Delta N)^3
\]

Where, \( \Delta N \) is the difference in refractive index between the denaturant solution and 50 mM Phosphate buffer, pH 7.0, 5 mM DTT.

2.5.3 Preparation of Samples for Equilibrium Chemical Denaturation

A stock solution of GdmHCl (≈ 8 M) or urea (≈ 10 M) was diluted with ddH2O (resistance of 18.2 Ω) to obtain a large range of denaturant concentrations using a Hamilton Microlab dispenser (Hamilton Company). The protein and denaturant stock solutions were filtered through 0.2 µm syringe filter. Cuvettes of differing path lengths were used depending on the protein concentration for the denaturation experiment. Smaller path length cuvette was used to scan higher protein concentration samples to prevent saturation of the detector. The preparation of samples for different protein concentrations and the use of cuvettes of suitable path lengths have been described as follows.
2.5.3.1 Sample preparation for LcrH protein concentrations 1 to 12 µM
Experiments involving denaturation of protein samples from 1 to 12 µM, a 9 x stock solution of protein (9 to 108 µM) and buffer (450 mM Phosphate buffer, pH 7.0, 45 mM DTT) was prepared. 100 µL of this 9 x stock solution was added to 800 µL of denaturant sample resulting in a 1 x concentration of proteins ranging from 1 to 12 µM. Cuvettes of path lengths 10 mm (for 1 µM LcrH), 5 mm (for 3 µM and 6 µM LcrH) and 2 mm (for 12 µM) were used for performing far-UV CD scans.

2.5.3.2 Sample preparation for LcrH protein concentrations 25 to 80 µM
A 6x stock solution of protein (150 to 480 µM) and buffer (300 mM Phosphate buffer, pH 7.0, 30 mM DTT) was prepared for conducting experiments involving protein samples from 25 to 80 µM. 50 µL of this 6 x stock solution was added to 250 µL of denaturant sample resulting in a 1x concentration of proteins ranging from 25 to 80 µM. The final conditions were 50 mM Phosphate buffer pH 7.0, 5 mM DTT. 1 mm path length cuvette was used to perform far-UV CD scans for LcrH protein with concentrations 25 to 80 µM.

The protein/denaturant solutions were pre-equilibrated at 25 °C for a minimum of 4 hours prior to performing experiments on the CD. Scans were taken in a thermostatted cuvette holder at 25 °C ± 0.2 °C and between 210 nm and 250 nm at a scan rate of 1 nm s⁻¹. The largest difference in signal between native and denatured LcrH was observed at 222 nm, and this wavelength was used to monitor unfolding. For easy comparison of differing data sets, the CD signal at 222 nm (λ_{obs}) was converted to relative spectroscopic signal (λ_{rel}) using Equation 2.12.

Equation 2.12

\[ \lambda_{rel} = \frac{(\lambda_{obs} - \lambda_N)}{(\lambda_D - \lambda_N)} \]

Where \( \lambda_{obs} \) is the observed CD signal, \( \lambda_N \) is the signal of the native state and \( \lambda_D \) is the signal of the denatured state.
Chapter 3 Gene Construction, Purification Strategy and Initial Biophysical Characterisation of LcrH proteins

3.1 Introduction

This chapter describes the following: (i) gene construction of full length LcrH (fl-LcrH), truncated LcrH (tr-LcrH) and monomeric L65E LcrH, (ii) purification strategies of these three constructs and (iii) initial biophysical characterisation - oligomeric state, conditions under which these oligomeric states were maintained (i.e. abrogation of any aggregation) and secondary structure characterisation.

3.2 Experimental approaches

Genes were constructed using PCR and site directed mutagenesis. Protein purification was carried out using affinity and size exclusion chromatography. SDS-PAGE gels and Mass Spectrometry followed protein purification. Once purified, analytical size exclusion chromatography was used to determine oligomeric state and conditions under which these oligomeric states were maintained. Finally, protein secondary structure was analysed using far-UV circular dichroism (CD).
3.3 Results

3.3.1 Full length LcrH (fl-LcrH) gene construction

The fl-LcrH gene sequence was obtained from NCBI Reference Sequence: NC_009595.1 showing 507 bp region from 55032 to 55538. The translated protein sequence of full length LcrH, comprising of 168 amino acids, is shown in Figure 3.1.

(a)

1 ATG CAA CAA GAG AGC AGA ACT CAA GAA TAC CAG CTG GCA ATG
46 GAA TCC TCT CTA AAA GGA GGG GGA ACT ATG GCC ATG CTC AAC GAA
91 ATT TCA AGT GAC ACT TTA GAG CAA CTC TAC TCT CTT GCA TTT AAC
136 CAA TAC CAG TCA GGA AAA TAC GAG GTT GCT CAC AAC AAG GTC TTT CAA
181 GCT CTC TGT GTG CTA GAC CAC TAT GAT TCA CGT TTC TTT TTA GGG
226 CTA GCC GCT TGT CTT CAA GCC ATG GGG CAA TAC GAC TTA GCG ATT
271 CAT AGC TAC AGC TAT GCC GCC ATA ATG GAT ATG AAA GAA CCT GGT
316 TTT CCG TTT CAT GCG GCC GAA TGT TTA CTG CAA AAG GGA GAG CTT
361 GCT GAA GCA GAA ATG GSC TGG TTC TCT TGG GTT CAA GAG CTT ATC GCA
406 GAC AAA ACT GAG TTT AAG GAG CTT TCC ACC CGA GTT AGC TCA ATG
451 TTA GAA GCA ATT AAA TTG AAA AAG GAG ATG GAA CAT GAG TGC GTT
496 GAT AAC CCA TGA

(b)

1........10 ........20 ........30 ........40 ........50 ........60
MQQETDDQEQ YQLAMESFLK GGTTIAMNLE ISSDTLEQLY SLaFQYQSG KYEDAHKVQF

........70 ........80 ........90 ........100 ........110 ........120
ALCVLDHYDS RFFLGLGACR QAMGQYDLAI HSYSGAIMD IEKPRFFFAE AECLLQKGEL

........130 ........140 ........150 ........160 ........168
AEAESGLFIA QELIADKTEF KELSTRVSSM LEAIKLKKEM EHECVDNP

Figure 3.1: Full length LcrH (fl-LcrH) (a) gene NCBI Reference Sequence: NC_009595.1 showing 507 bp region from a Yersinia pestis CA88-4125 plasmid pCD1 and (b) fl-LcrH protein sequence with accession no. AAA27642 in FASTA format. Sequences are numbered in red and amino acids are shown in black.

The fl-LcrH gene was constructed using recursive PCR (2.1.1.1) where eight coding and non-coding oligonucleotides spanning the entire length of the gene were used. Each oligonucleotide overlapped with its neighbour by 20 bases (Figure 2.1). The resultant recursive PCR product was run on 1% agarose gel (2.1.1.1.1) and showed a range of DNA sizes from ~100 – 500 bp (Figure 3.2a). This is expected and is caused by recombination and extension of oligonucleotides by different amounts. In order to obtain a full length LcrH
gene from the recursive PCR mixture, a conventional PCR was performed using the recursive PCR product as the template and PCR primers encoding the 5’ and 3’ of the gene with additional restriction sites (BamHI and HindIII at the 5’ and 3’ end, respectively) (Figure 3.3). Figure 3.2b showed that the PCR yielded DNA bands of ~500 bp which corresponds to the correct size of fl-LcrH gene (504 bp). The PCR product was then gel purified (section 2.1.3.2) and inserted into pProEX HTb (section 2.1.3) and pTrc-GST (section 2.1.4) vector for protein expression. DNA sequencing of the genes within the pProEX HTb and pTrc-GST vectors confirmed correct gene sequence had been synthesised.

Figure 3.2 (a) 1% agarose gel of recursive PCR of fl-LcrH gene. Lane 1 - recursive PCR product of fl-LcrH gene showing a range of DNA sizes ranging from ~100 to 500 bp, lane 2 - 1 Kb DNA ladder. (b) 1.5% agarose gel showing PCR product of fl-LcrH gene. Lanes 1&2 contain the PCR product of fl-LcrH gene showing DNA bands of ~500 bp. Lane 3: 100 bp DNA ladder.

Figure 3.3 Schematic showing the incorporated restriction sites of BamHI and HindIII at the 5’ and 3’ termini respectively of the full length LcrH gene using conventional PCR.

3.3.2 Production and purification of full length LcrH (fl-LcrH)
The pProEX HTb vector (section 2.1.3) expressed the fl-LcrH as a fusion protein with an N-terminal hexahistidine tagged protein. Protein expression was carried out in the E. coli
expression strain BL21 C41 (DE3). The conditions that gave the greatest soluble protein were found to be growth at 37 °C until an OD₆₀₀nm of 0.6 was reached and then induction with 0.1 mM IPTG at 27 °C for 14-16 hours. The fl-LcrH protein expression at different time points is shown in Figure 3.4.

![Figure 3.4: 15% SDS-PAGE gel showing the production of his-tagged fl-LcrH protein at different time points. Lanes 1, 2, 3 and 4 are lysed bacterial cell pellets after induction with 1mM IPTG at time t = 0 hrs, t = 2 hrs, t = 4 hrs and t = 16 hrs respectively. Lane 5 – protein ladder. The protein production was induced with IPTG at 16 °C and bacterial cells were grown overnight before harvesting for proteins.](image)

**Native Purification:** Fl-LcrH was extracted from the bacterial cells (Section 2.2.1.1) and subjected to affinity chromatography using Ni²⁺ column (Section 2.2.1.4). The his-tagged fl-LcrH protein was eluted (Figure 3.5a) and the his-tag cleaved using TEV protease (Figure 3.5b). Fl-LcrH has a molecular weight of 19 KDa and on the SDS-PAGE gel in Figure 3.5b, the cleaved fl-LcrH is observed as band of ≈ 20 KDa. However, there was a significant proportion of fl-LcrH found to be present in the insoluble pellet after sonication (inclusion bodies).

**Denatured Purification:** The insoluble pellet (I) was re-suspended in a denaturing lysis buffer (1x PBS, pH 7.0, 6 M GdmHCl, 10 % glycerol, 5 mM DTT), sonicated on ice, centrifuged and the resultant supernatant was subjected to affinity chromatography using a Ni²⁺ column pre-equilibrated with 5 column volumes of lysis buffer. The Ni²⁺ column was then washed with reducing concentrations of GdmHCl to gradually fold the bound proteins on the column before finally washing with 1 x PBS, pH 7.0, 10 % glycerol to completely remove the any trace of GdmHCl. The his-tagged fl-LcrH was then eluted using the his-tag elution buffer (section 2.2.1.2) and the his-tag from the protein was cleaved by TEV protease (Figure 3.5b).
**Purification of fl-LcrH**: Both natively purified and denatured purified cleaved fl-LcrH was then subjected to size exclusion chromatography to remove residual TEV protease and other protein impurities (section 2.2.1.6). Figure 3.5(c) shows the size exclusion chromatography (SEC) profile of fl-LcrH purification. The inset in Figure 3.5(c) shows the SDS-PAGE gel of the protein fractions eluted from the SEC of fl-LcrH. In the SDS-PAGE gel, lane 2 contains the protein fraction loaded onto the SEC, lanes 3-5 show higher molecular weight impurities and lane 6 onwards fl-LcrH band ~19 kDa can be observed. Lanes 6 & 7 showed considerable amount of higher molecular weight proteins co-eluting with fl-LcrH. The SDS-PAGE gel lanes 8-11 contained relatively higher purity of fl-LcrH protein. Thus, it became clear that the purified fl-LcrH samples obtained after the final SEC contained higher molecular weight impurities. Therefore, to increase purity to 95 % an additional wash step was introduced in the nickel column step. Here, the columns were washed with 1x PBS containing low concentration of imidazole (5 mM) before elution with his-tag elution buffer. However, the co-elution of higher molecular weight impurities could not be completely avoided. Both purification strategies are summarised in Figure 3.5 and achieved more than 90 % purity. Protein fractions corresponding to the lanes 8-11 in the inset Figure 3.5 (c) were pooled and concentrated as more than 90 % purity was achieved.
Figure 3.5 15% SDS PAGE gel showing (a) elution profile of his-tagged fl-LcrH. Lanes from left to right, lane 1 – supernatant containing soluble protein (s), lane 2 – supernatant from re-solubilized inclusion bodies (I), lane 3 – eluted fl-LcrH from soluble supernatant (s), lane 4 – eluted fl-LcrH from re-solubilized inclusion bodies (I), lane 5 – flow through (f/t) from (s), lane 6 – flow through (f/t) from (I), lane 7 – protein ladder. (b) Cleavage of his-tag from fl-LcrH (combined elutions from soluble supernatant and re-solubilised inclusion bodies) using TEV protease, lane 1 – protein ladder, lane 2 – uncleaved his-tagged fl-LcrH, lanes 3 and 4 – cleaved fl-LcrH. (c) Size exclusion chromatography profile of cleaved fl-LcrH, the chromatogram showing the elution of fl-LcrH and the 15 % SDS-PAGE gel (inset) showing the eluted protein fractions. Lanes from left to right, lane 1 & 12 – protein ladder, lane 2 – cleaved fl-LcrH sample loaded on the FPLC, lane 3-11: eluted protein fraction from SEC. The protein fractions loaded onto the SDS-PAGE gel are shown as the corresponding lane numbers on the SEC chromatogram.
In order to attain higher levels of protein purity, the fl-LcrH gene was inserted in pTrc-GST expression vector (section 2.1.3). The pTrc-GST vector produces the fl-LcrH as an N-terminally GST tagged fusion protein (section 2.2.1.5). As before the vector was transformed in BL21 C41(DE3) expression strain of E. coli and grown in an identical manner to that used for producing his-tagged fl-LcrH mentioned i.e. bacterial growth at 37 °C until reaching an O.D$_{600}$ nm of 0.6. However, expression was induced with 0.1 mM IPTG at 16 °C for 14-16 hours. The fl-LcrH protein was then extracted from bacterial cells (section 2.2.1.1) and subjected to affinity purification by passing it through a column containing glutathione agarose resin pre-equilibrated with GST wash buffer (section 2.2.1.3). Thrombin protease was added to the resin bound with the GST-fl-LcrH fusion protein for cleavage on column in order to separate the GST tag from fl-LcrH protein. The fl-LcrH eluted from the column as a flow through while the GST tag remained bound to the glutathione agarose resin. The protein production and purification of GST-tagged fl-LcrH followed by thrombin cleavage is shown in Figure 3.6.

(a) 18 % SDS-PAGE showing the protein production and elution profile of GST-tagged fl-LcrH protein. Lanes from left to right: lane 1 - lysed cells, lane 2 – supernatant protein, lanes 3 & 4 – GST-fusion fl-LcrH protein bound to the glutathione agarose resin, lanes 5 & 6 – cleaved GST tag bound to the glutathione agarose resin, lanes 7 & 8 – fl-LcrH protein, lane 9 – protein marker. The GST tag was cleaved using thrombin protease to achieve fl-LcrH protein of ≈ 19 kDa. (b) Chromatogram showing size exclusion profile of fl-LcrH using S200 16/60 column. The fl-LcrH protein fractions from 60 – 110 ml were pooled and concentrated further.
The fl-LcrH protein bands of ≈ 19 kDa can be clearly observed in the 18 % SDS-PAGE gel (Figure 3.6a) after cleavage with thrombin protease. The affinity purified fl-LcrH protein was further concentrated and subjected to size exclusion chromatography for removal of thrombin (Figure 3.6b). The fl-LcrH protein obtained using the GST-tagged fl-LcrH fusion method was easier to purify, provided a greater yield and was of higher purity in comparison to his-tagged fl-LcrH (> 95 % vs. < 90 %, respectively).

3.3.3 Truncated LcrH (tr-LcrH)

After the successful expression and purification of fl-LcrH, it proved to be relatively unstable and prone to aggregation on concentrating (discussed more fully later in this chapter). Moreover, soon after we had purified fl-LcrH the crystal structure of SycD was published. This is the same protein as LcrH but found in Yersina enterocolitica. The sequence identity of both proteins is extremely high with only two variant amino acid positions (position 136 and 138 – Figure 3.7 a). Importantly, the paper showed that the construct used to obtain crystals was a shortened version of SycD (tr-SycD) that was more stable and less prone to aggregation. This shortened construct spanned from position 21 to 163. The N-terminal deletion removed a mostly unstructured region, whereas the C-terminal deletion removed a solvent exposed cysteine. It was therefore decided that we should also obtain a truncated LcrH (tr-LcrH).

(a)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>1</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>SycD</td>
<td>MQQETTDQYQLAMESFLKGGTIAMLNEISSDTLEQLYSFLNQYSGKYEDAHKVFQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr-SycD</td>
<td>---------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>fl-LcrH</td>
<td>MQQETTDQYQLAMESFLKGGTIAMLNEISSDTLEQLYSFLNQYSGKYEDAHKVFQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr-LcrH</td>
<td>---------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>SycD</td>
<td>ALCVDHYDSRFFGLGACRQMGQYDIAHSYGAIMDIEPRFFHAAECLLQKGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr-SycD</td>
<td>---------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>fl-LcrH</td>
<td>ALCVDHYDSRFFGLGACRQMGQYDIAHSYGAIMDIEPRFFHAAECLLQKGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr-LcrH</td>
<td>---------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>SycD</td>
<td>AEAESGLFLAQELIARKPEFKEKSTRVSSMLEAIKLKKEHCEVDPN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr-SycD</td>
<td>---------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>fl-LcrH</td>
<td>AEAESGLFLAQELIARKPEFKEKSTRVSSMLEAIKLKKEHCEVDPN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr-LcrH</td>
<td>---------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
</tr>
</tbody>
</table>
Figure 3.7: (a) Multiple sequence alignment of protein sequences of SycD and tr-SycD from *Yersinia enterocolitica* (both shown in black) with fl-LcrH and tr-LcrH (both shown in dark red) from *Yersinia pestis*. Asparagine ‘N’ in SycD and tr-SycD is aspartic acid ‘D’ both in fl-LcrH and tr-LcrH and is highlighted in yellow. Proline ‘P’ in SycD and tr-SycD is Threonine ‘T’ both in fl-LcrH and tr-LcrH and is highlighted in turquoise. The 65th Leucine residue is highlighted in green and was changed to Glutamic acid to obtain monomeric tr-SycD and L65E tr-LcrH. Amino acid sequence numbers are shown in red. This figure was generated by ClustalW multiple alignment software. (b) Ribbon representation of head-to-head conformation of dimeric SycD<sub>21-163</sub> (tr-SycD) from *Yersinia enterocolitica* which is 99 % identical to tr-LcrH from *Yersinia pestis*. tr-LcrH differs from SycD in only two positions (N136D and P138T). Asparagine and Proline at amino acid positions 136 and 138 are shown in yellow and blue spheres respectively. The 65th leucine residue is shown in green spheres and was mutated to glutamic acid to obtain the monomeric L65E LcrH. Chains A and B of SycD<sub>21-163</sub> are coloured cyan and red respectively. Figure was prepared using PyMOL.

The tr-LcrH gene was constructed by PCR with the fl-LcrH gene as template and a forward primer starting at the DNA codons of the 21st amino acid and a reverse primer starting from DNA codons of the 163rd amino acid (Figure 3.8). Both primers possessed a BamHI and HindIII restriction sites, respectively for insertion into the pTrc-GST expression vector (Section 2.1.4).
Figure 3.8 Schematic representing the construction of tr-LcrH through PCR using fl-LcrH as template, forward primer spanning the DNA encoding 21st amino acid with BamHI restriction site, reverse primer spanning the DNA encoding the 163rd amino acid which also included a stop codon and HindIII restriction site. The resultant PCR product obtained is shown as the shortened tr-LcrH spanning amino acids from position 21-163.

The pTrc-GST vector inserted with tr-LcrH gene was transformed into XL-1 Blue strain of *E. coli* (section 2.1.3.4) followed by isolation of DNA from individual colonies. This was followed by restriction digestion of the isolated DNA using BamHI and HindIII (section 2.1.4) in order to screen the colonies that yielded DNA containing correctly sized insert. Those with a DNA size of ~430 base pairs (Figure 3.9) were sent for DNA sequencing. DNA sequencing confirmed that tr-LcrH gene was successfully cloned in pTrc-GST vector (section 2.1.2.1.7).
3.3.4 Expression & Purification of Truncated LcrH (tr-LcrH)

As before the pTrc-GST vector containing the tr-LcrH gene was transformed into BL21 C41 (DE3) expression strain of E. coli to produce GST-tagged tr-LcrH protein. The protocol for GST-tagged tr-LcrH protein production was identical to that used for the production of GST-tagged fl-LcrH production (Section 2.1.4). The protein production and elution profile of GST-tagged tr-LcrH is shown in Figure 3.10(a). The GST-tag was cleaved off the tr-LcrH protein by using thrombin protease giving a band of ≈ 16 kDa shown in Figure 3.10b.
The tr-LcrH protein was concentrated and further subjected to purification by size exclusion chromatography (Section 2.2.1.6), which ensured the removal of thrombin protease and enabled to achieve more than 95 % purity. The protein samples eluted from the size exclusion chromatography we loaded on an 18 % SDS-PAGE gel (Section 2.2.2.2) to check for the level of purity, shown in Figure 3.11. The tr-LcrH protein samples eluted from the size exclusion chromatography column showed bands of ≈ 16 kDa more than 95 % purity. Thus, tr-LcrH was successfully produced and purified.

![Figure 3.11: (a) Size exclusion chromatography profile of tr-LcrH using S200 16/60 column. (b) 18 % SDS-PAGE gel showing tr-LcrH protein samples purified by size exclusion chromatography. Lanes 1 to 8 are tr-LcrH SEC fractions from 70 ml until 110 ml with each fraction of 5 ml volume.](image-url)

3.3.5 Construction of L65E tr-LcrH

The study that described the crystal structure of SycD (and thus obviously LcrH) also showed that a single amino acid mutation could render the SycD dimer monomeric [168,194]. This mutation was L65E (Figure 3.7). It was, therefore, important that we also obtained this monomeric version of LcrH. To obtain LE65 tr-LcrH site directed mutagenesis was performed on the pTrc-GST expression plasmid containing the tr-LcrH gene (Section 2.1.5). Confirmation of the mutation was obtained through DNA sequencing by Beckman Coulter Genomics Ltd. (Section 2.1.2.1.7).
3.3.6 Expression and Purification of L65E tr-LcrH

The protein was produced using the protocol identical to that used for the production of GST-tagged fl-LcrH (Section 3.3.2) and GST-tagged tr-LcrH (Section 3.3.3) proteins. The protein samples were loaded onto an 18% SDS-PAGE for analysis and Figure 3.12 shows the protein production, affinity purification and cleavage of the GST tag using thrombin protease.

![Image](image.png)

Figure 3.12: (a) Size exclusion chromatography elution profile of L65E tr-LcrH using Superdex 200 16/60 column. Eluted protein fractions from 70-115 ml were pooled and concentrated further. (b) 18% SDS-PAGE gel showing the GST-tagged L65E tr-LcrH production, its affinity purification profile and cleavage of GST tag using thrombin protease. Lane 1 – protein ladder, lane 2 – lysed cells, lane 3 – supernatant protein, lane 4 – flow through, lanes 5 & 6 – resin sample showing the bound GST-tagged L65E tr-LcrH, lanes 7 & 8 – resin sample showing the thrombin cleavage of GST-tagged L65E tr-LcrH, lanes 9 & 10 – eluted L65E tr-LcrH showing a band ≈ 16 kDa, identical to tr-LcrH protein.

The eluted L65E tr-LcrH after cleavage of GST-tag showed a pure L65E tr-LcrH band of ≈ 16 kDa, identical to tr-LcrH (Figure 3.11 (b)). L65E tr-LcrH was then subjected to size exclusion chromatography for the removal of thrombin protease and achieve more than 95% purity. Importantly, the size exclusion chromatography showed that the LE65 mutant eluted at a larger volume than tr-LcrH (~95 ml vs ~80 ml). This is consistent with LE65 tr-LcrH being monomeric and tr-LcrH being dimeric.
3.3.7 Initial characterization of LcrH constructs

3.3.7.1 Defining buffer & protein concentrations to maintain “native” oligomeric states of LcrH constructs

During the purification of all LcrH constructs it was obvious that they were prone to aggregation. This manifested itself mainly during concentration steps where: (i) protein concentration was lost and (ii) UV-Vis absorbance showed high signal at 320 nm. Therefore it was important to define suitable buffer and protein concentrations conditions were they would remain in “native” oligomeric states.

Buffer conditions were chosen by: (i) calculating the isoelectric point (pl) of the proteins using the ExPASy – ProtParam tool [166-168,194] and (ii) choosing a pH that was 2 pH units away from the pl. The theoretical pl of the proteins is as follows: 4.70 for fl-LcrH, 4.95 for tr-LcrH and 4.87 for L65E tr-LcrH. Thus, 50 mM phosphate buffer at pH 7.0 was chosen (this, fortuitously, also emulated physiological conditions). Phosphate buffer was also chosen as it can specifically stabilise certain proteins. In addition 5 mM DTT was added to prevent disulphide bond formation.

To determine what protein concentrations these buffer conditions enabled (i.e. no aggregation was observed) the proteins were firstly concentrated until protein loss was observed through gain of UV absorbance at 320 nm and loss at 280 nm. The highest non-aggregating concentrations for fl-LcrH, tr-LcrH and L65E tr-LcrH were 300 µM, 500 µM and 150 µM, respectively. To confirm whether these conditions were stable over 1 week, each protein was concentrated to the aforementioned value and subjected to analytical SEC (Section 2.3.1). They were then stored at 4 °C for one week and then analytical SEC was repeated. Figure 3.13 shows the analytical SEC traces for each LcrH protein before and after 1 week at 4°C.
Figure 3.13 Analytical size exclusion chromatography profile of LcrH proteins using Superdex™ 75 10/30 column before and after 1 week: (a) fl-LcrH at 300 µM, (b) tr-LcrH at 500 µM and (c) L65E tr-LcrH at 150 µM. The elution profile in black denotes the first analytical SEC run for the respective LcrH protein samples at the aforementioned protein concentrations. The elution profile in red denotes the analytical SEC of the same protein samples stored for 1 week at 4 °C. TPR standards: CTPR3 (blue) and CTPRa6 (green). Each protein sample at the respective concentrations was loaded using 100 µl sample loop in 50 mM Phosphate, pH 7.0, 5 mM DTT at 25 °C. Figure was prepared using KALEIDAGRAPH 4.0.

Figure 3.13(a) did not show any evidence of aggregation in the protein samples. The elution peak for both samples i.e. before and after storage at 4 °C was ≈ 9.8 ml. However, a tailing peak at ≈ 11-13 ml was observed in both sample runs, which can be attributed to protein degradation. Figure 3.13 (b) shows the analytical SEC profile for tr-LcrH at 500 µM before and after storage at 4 °C. The elution profile for both sample runs was identical. The
difference in the absorbance peaks between the two samples could be attributed to the variability in sample runs. The analytical SEC profile for L65E tr-LcrH at 150 µM was indistinguishable for both sample runs. The elution volume for both sample runs was ≈ 10.8 ml. Additionally, the elution volumes of both tr-LcrH and fl-LcrH are close to that of the dimeric TPR standard CTPRa6 (10.55 ml). Thus, confirming the dimeric oligomeric state of tr-LcrH and fl-LcrH. Figure 3.13 (c) represents the analytical SEC profile of monomeric mutant L65E tr-LcrH. Both elution profiles for L65E tr-LcrH are nearly identical. The elution profiles gives an indication of aggregation (small elution peaks before the void volume (V_v) = 8.5 ml) and some degradation products (tailing peaks) in both sample runs.

On comparing the analytical SEC profile of LcrH proteins at their respective protein concentrations it was observed that over the course of one week the proteins did not undergo significant changes. This confirmed that the highest non-aggregating protein concentrations determined for each of LcrH protein constructs and the buffer conditions used were stable over time. As can be seen all proteins were stable over this time course at their respective concentrations with little loss of protein. This was confirmed by obtaining the concentration of each protein by UV-Vis absorbance at 280 nm before and after one week at 4 °C.

The elution profiles of the protein samples before and after one week storage at 4 °C were identical. For fl-LcrH, tr-LcrH and L65E tr-LcrH this equated to peak elution volumes of ~9.6 ml, ~9.8 ml and ~10.8 ml respectively. The tailing peak in the fl-LcrH elution profile is consistent with slight degradation (similar in both runs). The slight difference in the size of the elution peaks is simply due to the variability of the UV-Vis detector of the AKTA FPLC. The protein concentration was confirmed by obtaining UV-Vis absorbance at 280 nm of each protein before and after one week prior to each analytical SEC run. Interestingly, a difference in size of the elution peaks of LcrH proteins was observed. The peak sizes for LcrH proteins were in the following order: fl-LcrH>tr-LcrH>L65E tr-LcrH. This is suggestive of fl-LcrH and tr-LcrH being dimeric and L65E tr-LcrH monomeric (discussed in Chapter 4).
3.3.7.2 Far-UV circular dichroism of native LcrH constructs

To confirm that the conditions used above produced correctly folded LcrH constructs far-UV CD spectra were obtained (section 2.4.1). Figure 3.14 shows the native and urea denatured far-UV CD spectra of fl-LcrH, tr-LcrH and L65E tr-LcrH at 3 μM (spectra measured in 5 mm path length cuvette) and 50 μM (spectra measured in 1 mm path length cuvette) respectively. The native spectra clearly show correctly alpha-helically folded protein, as they possess minima at 222 nm and 208 nm. Interestingly the fl-LcrH has approximately 10 % greater signal at 222 nm than both tr-LcrH and its LE65 mutant (which are identical). When 8 M urea is added to LcrH proteins there is a complete loss of secondary structure as shown by the disappearance of the minima at 208 & 222 nm.

Figure 3.14 Far-UV CD spectra comparing native fl-LcrH (black), tr-LcrH (blue), L65E tr-LcrH (green) and denatured L65E tr-LcrH in 8 M urea (red) for concentrations: (a) 3μM (spectra measured using 5 mm path length cuvette) and (b) 50μM (spectra measured using 1 mm path length cuvette) The detector on the CD spectrophotometer became saturated with HT voltage exceeding 850 volts after 212 nm for samples in 5 mm cuvette and after 208 nm for samples in 1 mm path length cuvette. Figure was prepared using KALEIDAGRAPH 4.0.

In the above Figure 3.14, far-UV CD spectra for LcrH protein samples at 3 μM and 50 μM, were masked beyond wavelengths at 212 nm and 208 nm respectively because the HT voltage of the detector of the CD spectropolarimeter increased to above 800 volts. This is because if the HT voltage goes above ~800 volts the noise becomes disproportionate to the sample signal and the output is completely unreliable (Section 2.4.1 and Jasco J-715 Spectropolarimeter Operating Manual).
3.3.8 Conclusion
The three protein constructs of LcrH – fl-LcrH, tr-LcrH and L65E tr-LcrH were successfully produced and purified. We were able to define suitable buffer and protein concentration conditions for maintenance of native oligomeric states of the respective proteins through analytical size exclusion chromatography. Far-UV CD spectroscopy enabled to confirm that the conditions used produced correctly folded proteins. In the following chapter we explore the oligomeric states and thermodynamic stability of LcrH protein constructs.
Chapter 4 Oligomeric State, Flexibility and Thermodynamic Stability of LcrH protein constructs

4.1 Introduction
In this chapter we explore the oligomeric state, flexibility and thermodynamic stability of each of our constructs of LcrH. These investigations enable the dissection of the role of oligomeric state and stability on LcrH’s biological function.

4.2 Experimental Approaches
The monomer-dimer equilibrium and the oligomeric state of the different constructs of LcrH was determined by analytical size exclusion chromatography, dynamic light scattering and analytical ultracentrifugation. We then examine the thermodynamic stability and flexibility of our different LcrH constructs by performing equilibrium unfolding through chemical denaturation. The unfolding was monitored by far-UV circular dichroism and fluorescence spectroscopy as probes for secondary and tertiary structures respectively.

4.3 Results
4.3.1 Oligomeric States of LcrH using Analytical Size Exclusion Chromatography (SEC)
Class II chaperones of the type three secretion system (TTSS) in gram negative bacteria Yersinia, Shigella and Pseudomonas species have been shown to be homodimeric in solution [166-168,195]. In Chapter 3 we obtained conditions where fl-LcrH, tr-LcrH and L65E tr-LcrH were “stable” (i.e. maintained a “native” oligomeric state). These experiments were carried out at high concentrations and gave elution profiles in analytical SEC of dimeric size for fl-LcrH and tr-LcrH but monomeric for L65E tr-LcrH. To qualitatively probe the monomer-dimer equilibrium of fl-LcrH and tr-LcrH, analytical
SEC was performed over a range of protein concentrations. LcrH is a non-globular protein with each monomeric unit comprising of 3.5 modular tetratricopeptide repeats (TPRs). Thus, two non-globular protein standards – CTPR3 (3.5 continuously arrayed TPR motifs) and CTPRa6 (6.5 continuously arrayed TPR motifs) were chosen as protein standards for comparing the molecular sizes. The crystal structure of SycD (99% identical to LcrH) shows that the dimeric form of the protein produces a “head-to-head” dimer. Thus, dimeric LcrH should have a similar size to that of CTPRa6. Whereas the monomeric LcrH (L65E tr-LcrH) should elute with size similar to that of CTPR3 (Figure 4.1).

![Figure 4.1](image)

**Figure 4.1** Ribbon representation of the crystal structures of (a) head-to-head confirmation of dimeric SycD$_{21-163}$ [Protein Data Bank (PDB) entry: 2VGX], chains A and B are shown in cyan and red respectively, (b) CTPR3 (PDB entry: 1NA0), (c) CTPRa6. The figure was prepared using PYMOL.

Figure 4.2 (a) and Figure 4.2 (b) show the analytical SEC profile for fl-LcrH and tr-LcrH from 3 – 100 µM respectively. These show that at higher concentrations (100 µM), fl-LcrH and tr-LcrH eluted with volumes of 9.7 ml and 9.6 ml respectively, and were consistent with a homodimeric structure (in close agreement with an elution peak of 9.6 ml for the designed TPR protein CTPRa6 containing 6.5 stacked repeats).
Interestingly, when the protein concentrations were lowered to more physiological concentrations i.e. from 100 µM to 3 µM, the elution volumes for fl-LcrH and tr-LcrH were observed to shift to that of a monomeric sized protein. The lowest concentration of 3 µM for tr-LcrH eluted at 10.7 ml, which was identical to the elution volume of 10.7 ml for monomeric L65E tr-LcrH and similar to elution volume of 10.55 ml for CTPR3 protein (containing 3.5 stacked repeats).

Figure 4.2 Analytical size exclusion chromatography profile of fl-LcrH (a) and tr-LcrH (b) using Superdex 75 HR 10/30 column at protein concentrations of 100 µM (red), 50 µM (blue), 25 µM (brown), 12 µM (green), 6 µM (purple) and 3 µM (orange). For calibration, two protein standards – CTPRa6 and CTPR3 were also run. CTPRa6 is a designed TPR protein that contains 6.5 continuously arrayed TPR motifs. CTPR3 is also a designed TPR protein that contains 3.5 continuously arrayed TPR motifs. Both the protein standards are coloured black; with cTPRa6 having elution volume of 9.6 ml and cTPR3 shows an elution volume of 10.55 ml. Monomeric L65E tr-LcrH is shown in grey and has an elution volume of 10.7 ml. In the respective inset figures, the elution volumes of LcrH proteins are plotted against their respective protein concentrations. Figure was prepared using KALEIDAGRAPH 4.0.
The change in elution profile shows that LcrH is present in monomer-dimer equilibrium at physiological protein concentrations and therefore has a relatively weak dimerisation interaction. It also suggests that its $K_D$ value should be in the low micromolar range.

4.3.2 Determination of Dimerization Dissociation Constant ($K_D$) using Analytical Ultracentrifugation (AUC)

The previous section suggested that the dimerization dissociation constant ($K_D$) should be in the low micromolar ($\mu$M) concentration range for tr-LcrH and fl-LcrH. In order to accurately determine the $K_D$ value, equilibrium sedimentation by analytical ultracentrifugation for tr-LcrH was performed (section 2.3.2, Figure 4.3).

Figure 4.3 Equilibrium Analytical Ultracentrifugation (AUC) for tr-LcrH, the figure shows (a) the representative curve obtained from equilibrium sedimentation using analytical ultracentrifugation denoted by black dots and the line of best fit (red) to a monomer-dimer equilibrium model using the program Ultrascan, (b) fitted residuals shown by black filled circles.
From Figure 4.3 (b), it can be observed that the residuals from the fit appear to be scattered randomly and have an absorbance close to zero. This deemed the fit appropriate. No other trends can be observed in the residuals so does not need to be fitted to another model except the monomer-dimer equilibrium model (Equation 2.6).

Multiple curves from the equilibrium sedimentation experiment were fitted and calculations were carried out using the Equation 2.6. This gave a dimerization dissociation constant $K_D$ of $15 \pm 5 \ \mu M$. The errors and confidence limits were calculated as mentioned in section 2.3.2.

4.3.2.1 Calculating percentage of dimer at a given protein concentration

After determining the dimerization dissociation constant $K_D$ for tr-LcrH, the percentage of the protein population that is dimeric can be calculated as follows:

The monomer-dimer chemical equilibrium at a given protein concentration can be illustrated by the Equation 4.1 below.

Equation 4.1

$$M + M \rightleftharpoons D$$

Where $D$ is the dimer and $M$ is the monomer. Thus, the dimerization dissociation constant $K_D$ of this equilibrium is:

Equation 4.2

$$K_D = \frac{[M][M]}{[D]} = \frac{[M]^2}{[D]}$$

The total protein concentration ($C_T$) for the dimeric system can be written as:

Equation 4.3

$$[C_T] = [M] + 2[D]$$

Rearranging Equation 4.3 and solving for $M$ gives

Equation 4.4

$$[M] = [C_T] - 2[D]$$
Substituting Equation 4.4 into the $K_D$ expression (Equation 4.2) gives

Equation 4.5

$$K_D = \frac{(C_T - 2D)(C_T - 2D)}{D}$$

Solving the above Equation 4.5 into quadratic form gives:

Equation 4.6

$$[D] = \frac{4C_T + K_D \pm \sqrt{(8C_T * K_D + K_D^2)}}{8}$$

Fraction of dimeric species ($F_D$) at any given protein concentration can be defined by:

Equation 4.7

$$F_D = \frac{2[D]}{[C_T]}$$

‘$F_D$’ is the fraction of total ‘D’ (dimer), which can vary from 0 to 1 and ‘$C_T$’ is the total protein concentration.

Thus, by substituting Equation 4.6 in Equation 4.7 we obtain:

Equation 4.8

$$F_D = \frac{4C_T + K_D \pm \sqrt{(8C_T * K_D + K_D^2)}}{4[C_T]}$$

A saturation-binding curve of $F_D$ (fraction of total dimer) vs $C_T$ (total protein concentration) using Equation 4.8 with the dimerization dissociation constant $K_D$ value of 15 µM was plotted (Figure 4.4).
Figure 4.4 Saturation binding curve for dimerization of tr-LcrH showing a graph of $Y$ vs $C_T$, where $Y$ is the fraction of total possible dimer which varies between 0 to 1, $C_T$ is the total protein concentration ($\mu$M) and the dimerization dissociation constant $K_D = 15 \pm 5 \mu$M.

The curve in Figure 4.4 is hyperbolic with half-maximal dimer formation occurring at a total protein concentration of $C_T = K_D$. It can also be noted that even at a total protein concentration of 600 $\mu$M, which is 40-fold greater than $K_D$, only 90 % of the total possible dimer (D) is formed ($F_D = 0.90$). When $C_T = 90 \mu$M, 75 % of the protein is dimeric ($Y = 0.75$). The percentages of dimer at protein concentrations we will use later are shown in Table 4.1. The table shows that even at lower concentrations of tr-LcrH there is some dimer present.

Table 4.1: Dimer percentages at respective protein concentrations of tr-LcrH

<table>
<thead>
<tr>
<th>Total Protein Concentration ($\mu$M)</th>
<th>Percentage of Dimer (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>34</td>
</tr>
<tr>
<td>12</td>
<td>46</td>
</tr>
<tr>
<td>25</td>
<td>58</td>
</tr>
<tr>
<td>50</td>
<td>68</td>
</tr>
<tr>
<td>80</td>
<td>74</td>
</tr>
<tr>
<td>100</td>
<td>76</td>
</tr>
</tbody>
</table>

The table shows percentages of dimer population of tr-LcrH at the given total protein concentrations ($\mu$M). The dimer percentages have been rounded up to the next highest number.
4.3.3 Loss of LcrH’s α-Helical Structure on N-terminal Deletion and Concentration Reduction

As LcrH is a completely α-helical protein, far-UV CD provides an excellent probe for monitoring its secondary structure in solution. In the previous chapter we observed that under our conditions all LcrH constructs exhibit strong α-helical signals with indicative minima at 208 and 222 nm. Thus, to probe whether any structure is lost between fl-LcrH and tr-LcrH or transition from dimer to monomer, far-UV CD spectra were recorded for protein concentrations of 3 – 100 µM. The CD ellipticity at each concentration was converted to molar ellipticity (Equation 2.9) for ease of data comparison and to correct for path length and protein concentration (Figure 4.5).

![Graph showing far-UV CD spectra](image)

Figure 4.5 far-UV CD spectra showing molar ellipticity from 200 to 300 nm for native fl-LcrH (a) and tr-LcrH (b) at concentrations of 100 µM (red), 50 µM (blue), 25 µM (brown), 12 µM (green), 6µM (purple), and 3 µM (orange). Figure was prepared using KALEIDAGRAPH 4.0.

Interestingly, we observed a reduction in the signal (molar ellipticity at 222 nm) when tr-LcrH was compared with fl-LcrH and when the concentration of each protein was decreased. In each case the change was small as follows: at 50 µM, fl-LcrH and tr-LcrH
have molar ellipticities at 222 nm of $-3.1 \times 10^6$ and $-2.7 \times 10^6$ degrees cm$^2$ dmol$^{-1}$, respectively, and at 3.1 µM, fl-LcrH and tr-LcrH have molar ellipticities of $-2.8 \times 10^6$ and $-2.5 \times 10^6$ degrees cm$^2$ dmol$^{-1}$, respectively. In both comparisons this equates to a maximal change of $\approx 10\%$. The reduction of $\alpha$-helical signal from fl-LcrH to tr-LcrH suggested that the N-terminal 20 amino acid deletion of LcrH removes some $\alpha$-helical structure. The loss of $\alpha$-helical signal with a decrease in protein concentration is consistent with the weak monomer/dimer equilibrium i.e. the reduction in concentration of LcrH results in a monomeric state, which is less $\alpha$-helically structured than the dimer.

### 4.3.4 Equilibrium Unfolding and Thermodynamic Stability of LcrH

To determine the flexibility and thermodynamic stability of fl-LcrH, tr-LcrH and L65E tr-LcrH equilibrium chemical denaturation experiments were performed (section 2.5.3). However, to achieve analysable denaturation experiments that were reproducible, we first needed to determine the following:

i. The best structural probe to monitor the chemical denaturation

ii. The most suitable chemical denaturant

iii. The minimum equilibration time that produced a proper equilibrium at every denaturant concentration

iv. The reversibility of equilibrium unfolding

#### 4.3.4.1 The Best Structural Probe to Monitor Chemical Denaturation

In general, two highly used biophysical probes for protein structural change are far-UV circular dichroism (far-UV CD) and fluorescence. This is because both these techniques require small concentrations of protein sample, low sample volumes and exhibit high sensitivity. Far UV CD is an absorbance-based technique that monitors the secondary structure of a protein while fluorescence monitors the tertiary structural changes. LcrH should be amenable to both techniques, as firstly it has a high alpha-helical native structure thereby making it easy to gain information about the secondary structure and folded state of proteins using far UV CD. Secondly, fl-LcrH and tr-LcrH contain 8 and 7 tyrosine residues respectively. Hence, tyrosine fluorescence can be used as a
tertiary structure probe to monitor changes in LcrH’s protein structure when subjected to unfolding.

To observe the change in signal of both these probes on LcrH’s unfolding, each protein construct was incubated with 8 M urea. Figure 4.6 (a-c) shows the far-UV CD spectra of native and denatured LcrH proteins (fl-LcrH, tr-LcrH and L65E tr-LcrH) at 3 µM in 50 mM Phosphate, pH 7.0, 5 mM DTT. The native and 8 M urea denatured LcrH protein samples were incubated at 25 °C for 6 hours prior to recording far-UV CD spectra. The native protein samples showed strong α-helical signal with minima at 222 nm and 208 nm (minima at 208 nm not shown in figure). This is consistent with their high alpha helical native secondary structure. When a high molarity of chemical denaturant (8 M urea) was introduced these indicative minima disappeared, displaying that all native structure was unfolded. Importantly, the change in signal between folded and unfolded protein samples was suitably large.

**Figure 4.6** Far-UV CD spectra at 3 µM protein concentration of (a) fl-LcrH, (b) tr-LcrH, (c) L65E tr-LcrH showing native (black filled circles) and 8 M urea denatured (red filled circles). The CD spectra of the samples was taken using 5 mm path length quartz cuvette at 25 ± 0.2 °C. The data was masked for wavelengths below 212 nm as the CD detector became saturated with voltage exceeding 850 HT volts. (d) Tyrosine fluorescence of 3 µM tr-LcrH showing native (black) and 8 M urea denatured (blue). The excitation wavelength was 274 nm, the fluorescence spectra was measured from 294 nm to 400 nm in 1 cm pathlength quartz cuvette at 25 ± 5 °C. Figure was prepared using KALEIDAGRAPH 4.0.
After monitoring the secondary structure of LcrH proteins using far-UV CD, it was then decided to use fluorescence spectroscopy to compare the native and denatured tertiary structure of LcrH. Figure 4.6 (d) shows the comparison between the fluorescence intensities of native and 8 M urea denatured 3 µM tr-LcrH protein samples. Here, the sample was excited at 274 nm and emission spectrum from 294 nm to 400 nm was recorded. A peak due to tyrosine emission can be seen at 303 nm, which decreases on addition of urea. However, it is also obvious that the change in signal in fluorescence on unfolding was not as great a change as that observed for the far-UV CD. Figure 4.6 showed that the largest difference in the far-UV CD signal between the native and denatured LcrH samples was observed at 222 nm, which was then used to monitor equilibrium unfolding.

![Figure 4.7: (a) Tyrosine fluorescence monitored at 303 nm (black filled circles) for urea induced equilibrium unfolding of 3 µM tr-LcrH. The excitation wavelength was 274 nm and slit width for excitation and emission was 5 nm. Fluorescence emission was recorded using 1 cm path length quartz cuvette at 25 ± 5 °C. (b) Equilibrium unfolding profile of 3 µM tr-LcrH monitored by far-UV CD at 222 nm wavelength. Measurements were taken using 0.5 cm path length quartz cuvette at 25 ± 5 °C in 50 mM Phosphate, pH 7.0 and 5 mM DTT. The samples were incubated at 25 °C for 6 hours prior to far-UV CD and fluorescence measurements. Figure was prepared using KALEIDAGRAPH 4.0.](image)

Given that both probes showed a clear difference between native and denatured protein, it was decided to obtain a full equilibrium chemical denaturation. Here samples containing 3 µM tr-LcrH were prepared with a series of urea concentrations from 0 to 8 M urea (section 2.5.3.1). The samples were incubated for 6 hours at 25 °C
before taking far-UV CD and fluorescence measurements. The equilibrium unfolding profile of 3 μM tr-LcrH for each probe is shown in Figure 4.7. When the two profiles are compared, a number of striking differences become obvious. The tyrosine fluorescence monitored equilibrium unfolding showed that tr-LcrH possesses no sigmoidal co-operative transition. In contrast, although far-UV CD monitored equilibrium unfolding (Figure 4.6 (b)) also showed an initial gradual unfolding profile, there is a distinct sigmoidal transition between the initial (folded) and final (unfolded) states. The fluorescence monitored unfolding data did not bear any resemblance to the sigmoidal unfolding pattern observed in far-UV CD monitored unfolding data. One explanation of the differing curves can be suggested by the positioning of the tyrosine residues in tr-LcrH (Figure 4.8). The surface analysis of the tr-LcrH protein structure showed that 50% of the tyrosine residues, which are located in the N-terminus of the protein were found to be solvent exposed in the native state of the protein. Consequently, as the protein undergoes unfolding, the remaining 50% of tyrosine residues too become solvent exposed. However, resultant change in the fluorescence intensity between native and unfolded LcrH protein in comparison to the secondary structural change observed from far-UV CD data (Figure 4.6) was much lower. This is because fluorescence is based on the environment of fluorophores i.e. tryptophans and tyrosines that are solvent exposed fluoresce at a different wavelength compared to those in a non-solvent exposed environment (the hydrophobic core of the protein). In order to observe a difference in fluorescence between a native and unfolded protein, there must be an overall change in the interaction of the solvent exposed tyrosines and tryptophans with the solution. Tryptophan has stronger fluorescence and higher quantum yield in comparison to tyrosine. But, LcrH does not contain any tryptophan residues. Although, tyrosine is a weaker emitter than tryptophan, it may still contribute significantly to protein fluorescence if present in great numbers. tr-LcrH and fl-LcrH contain only 7 and 8 tyrosine residues respectively, which is not a large number. Furthermore, tyrosine is frequently quenched due to loss of proton on the aromatic hydroxyl group by undergoing ionization in the excited state or due to its interaction with the peptide chain.
Thus, the equilibrium unfolding monitored by tyrosine fluorescence was not found to be suitable. Far-UV CD proved to be the best structural probe for monitoring the equilibrium unfolding of LcrH’s protein constructs.

4.3.4.2 Determining the most suitable chemical denaturant

It was previously shown that LcrH underwent complete loss of secondary structure on denaturation with high molar concentration (8 M) of the chemical denaturant urea (Figure 4.6). When 6 M Guanidine hydrochloride (GdmHCl) was used to unfold 3- μM tr-LcrH, same result was achieved (Figure 4.9).
After observing the effect of GdmHCl, it was decided to perform an equilibrium chemical denaturation experiment using GdmHCl as a chemical denaturant and 3 µM tr-LcrH protein as the protein sample (Section 2.5.3). The samples were incubated at 25 ± 0.2°C for 6 hours followed by far-UV CD measurements. The GdmHCl induced equilibrium unfolding of tr-LcrH was compared to the urea induced equilibrium unfolding (Figure 4.10). GdmHCl is a stronger chemical denaturant than urea. Figure 4.10 shows that tr-LcrH undergoes unfolding at lower concentration of GdmHCl compared to that of urea. Importantly, the data comparison showed that there is a lack of folded (native) baseline for GdmHCl induced unfolding compared to that of urea. This lack of native baseline caused difficulties in fitting equilibrium unfolding curves to 2-state or multi-state denaturation models.
Figure 4.10 Equilibrium denaturation of 3 µM tr-LcrH using urea (black filled circles) and GdmHCl (black filled diamonds) monitored by changes in far-UV CD at 222 nm plotted against the denaturant concentrations (M) of urea and GdmHCl respectively. The samples were equilibrated for 6 hours prior to taking far-UV CD scans, in 5 mm path length quartz cuvette at 25 °C ± 0.2 °C, in 50 mM Phosphate pH 7.0, 5 mM DTT. Figure was prepared using KALEIDAGRAPH 4.0.

Therefore, it was decided that urea was the most suitable chemical denaturant. Henceforth, all the equilibrium-unfolding experiments of LcrH protein constructs were carried out using urea.
4.3.4.3 Minimum Equilibration Time for Equilibrium Chemical Denaturation

It is important to confirm that all protein samples at each denaturant concentration reach equilibrium. If insufficient time is allowed for equilibration, an incorrect stability and/or non-reproducible experiments may occur [195-197]. In order to determine the minimum equilibration time for LcrH, 3 μM tr-LcrH was chosen as a test case. A mixture of protein and urea in buffer were prepared (Section 2.5.3) and the samples were incubated for differing lengths of time (4, 12 and 24 hours). This was followed by taking far-UV CD measurements at 222 nm in a thermostatted cuvette holder at 25 ± 0.2 °C in a 5 mm path length quartz cuvette. Figure 4.11 shows the overlaid results.

![Figure 4.11 Urea induced equilibrium denaturation of tr-LcrH monitored by changes in far-UV CD signal (ellipticity in millidegrees) at 222 nm plotted against the concentration of urea (M). The superimposed equilibrium unfolding profiles of 3 μM tr-LcrH corresponds to differing equilibration times: 4 hours (filled circles), 12 hours (open circles) and 24 hours (closed squares) respectively. The inset figure shows the same unfolding profiles converted to relative spectroscopic signal (fraction unfolded) for easy comparison of different data sets. Figure was prepared using KALEIDAGRAPH 4.0.](image)

Importantly, all three differing equilibration times were observed to be superimposable. It could thus be inferred that 4 hours was the minimum time required for achieving chemical equilibrium. Therefore, all equilibrium-unfolding experiments
were left for a minimum equilibration time of 4 hours prior to taking measurements on the CD spectrophotometer.

4.3.4.4 The Reversibility of Equilibrium Unfolding

The final control required was to confirm that the chemical denaturations were reversible. This is requirement, as otherwise thermodynamic parameters observed from fitting the curves will not report on the unfolding process, but instead on aggregation. For this experiment, tr-LcrH and fl-LcrH protein samples at 27 µM were unfolded with 8 M urea, equilibrated for 4 hours at 25 ± 0.2 °C respectively. The unfolded protein samples of both protein constructs were then refolded by aliquoting into buffer with differing concentrations of urea. This gave a final protein concentration of 3 µM in each differing urea concentration. The diluted/refolded samples were re-equilibrated for 4 hours followed by far-UV CD measurement at 222 nm. Figure 4.12 shows the far-UV CD monitored refolded equilibrium titrations superimposed onto the unfolded equilibrium titrations at 3 µM concentration of tr-LcrH and fl-LcrH respectively.

![Figure 4.12](image)

Figure 4.12 Reversibility of urea induced equilibrium unfolding of (a) tr-LcrH and (b) fl-LcrH at 3 µM protein concentration showing the unfolded titrations (black closed diamonds) and refolded titrations (green closed circles) respectively. The unfolding and refolding equilibrium titrations were monitored by changes in far-UV CD at 222 nm, in 5 mm path length quartz cuvette at 25 °C ± 0.2 °C, in 50 mM Phosphate, pH 7.0, 5 mM DTT. Figure was prepared using KALEIDAGRAPH 4.0.

The unfolded and subsequently refolded sample retained nearly 100% of their native far-UV CD signal. This confirmed that urea induced equilibrium unfolding of both tr-LcrH and fl-LcrH is completely reversible under the buffer conditions used.
4.3.5 Equilibrium Chemical Denaturation of LcrH’s Protein Constructs

Having determined conditions where reversible, aggregation-free chemical denaturations of LcrH can be performed and monitored, we conducted such denaturations at a number of different protein concentrations. For fl-LcrH, these were 3 and 50 μM (Figure 4.13), for tr-LcrH these ranged from 1 to 80 μM (Figure 4.14) and for monomeric L65E tr-LcrH these were 1.5 and 3 μM (Figure 4.15). Reproducibility of each denaturation was confirmed by repeating each experiment at least once (shown in the aforementioned figures). For both fl-LcrH and tr-LcrH the 50 and 80 μM chemical denaturations were the highest protein concentration that could be prepared. This was due to the methodology that chemical denaturations require, i.e. the protein stock solution is diluted to the same amount into a series of increasing urea concentrations.

Figure 4.13 Urea induced equilibrium chemical denaturation of fl-LcrH at protein concentrations of (a) 3 μM shown by blue filled circles, far-UV CD measurements were taken using a 5 mm path length cuvette and (b) 50 μM shown by open and filled blue triangles for reproducibility, far-UV CD measurements were taken using a 1 mm path length cuvette. Figure was prepared using KALEIDAGRAPH 4.0.
Figure 4.14 Urea induced unfolding of tr-LcrH monitored by change in far-UV CD at 222 nm at protein concentrations: (a) 1 µM (1 cm path length), (b) 3 µM (5 mm path length), (c) 6 µM (5 mm path length), (d) 12 µM (2 mm path length), (e) 25 µM (1 mm path length), (f) 50 µM (1 mm path length) and (g) 80 µM (1 mm path length). The inset figures are the denaturation profiles at respective concentrations with the CD signal changed to relative spectroscopic signal (fraction unfolded) for ease of comparison. Each experiment was repeated at least once to confirm reproducibility and shown by black open and filled circles. Figure was prepared using KALEIDAGRAPH 4.0.
Figure 4.15 Urea induced equilibrium chemical denaturation of monomeric mutant L65E tr-LcrH at protein concentrations of (a) 1.5 µM shown by open blue circles and (b) 3 µM shown by closed blue circles. Far-UV CD measurements were taken using a 5 mm path length quartz cuvette at 25 ± 0.2°C. Figure was prepared using KALEIDAGRAPH 4.0.

For ease of comparison, each denaturation was converted from ellipticity at 222 nm vs urea to relative spectroscopy signal vs urea (Equation 2.9, Materials & Methods). Figure 4.16 shows fl-LcrH overlaid with tr-LcrH and the overlaid plots tr-LcrH from 1 µM to 80 µM, respectively. From these graphs the following features and trends were immediately apparent.

4.3.5.1 Biphasic Denaturations Show LcrH Unfolds via Partially Folded Dimeric Intermediate

The profile of each equilibrium denaturation and its protein concentration dependence delineates the mechanism by which oligomeric proteins unfold. Such concentration dependence is a unique characteristic of oligomeric protein systems and is a direct consequence of protein unfolding being coupled with oligomeric dissociation. Once determined, the data can be fit to the correct dimer denaturation model and thus obtain thermodynamic stability of the protein. Figure 4.16 shows that for both forms of LcrH the denaturation profile of each protein concentration consists of two transitions which becomes clearer with the increase in protein concentrations.
The initial unfolding produces a first sloping transition that accounts for the loss of approximately one third of the native α-helical signal. This ends at a saddle point at ≈ 3 M urea. At higher denaturant concentrations, a second sharper transition was observed. This results in the complete unfolding of the protein at urea concentrations of greater than 6 M. The biphasic unfolding pattern showed that a stable, partially structured intermediate is populated during the unfolding of LcrH. Significantly, the first transition is concentration-independent, and the second transition is
concentration-dependent (the midpoint ‘[D]_{50}’ shifts to higher urea concentration with an increase in protein concentration). The observed pattern of unfolding (protein concentration-independent first transition to a stable partially folded intermediate followed by a protein concentration-dependent second transition) is consistent with a scheme where native dimer (N$_2$) partially unfolds to a dimeric intermediate (I$_2$), which then unfolds to a denatured monomer (2D), 

$$N_2 \rightleftharpoons I_2 \rightleftharpoons 2D$$

(Figure 4.17). Moreover, it is important to note that the shallow nature of the first unfolding transition, native dimer to partially unfolded intermediate, shows a lack of co-operative unfolding. This indicates the structure of LcrH is particularly flexible and can easily adopt partially unfolded confirmations.

![Figure 4.17 Schematic illustration of the proposed equilibrium unfolding pathway of LcrH. A head-to-head dimeric structure of SycD (PDB entry: 2VGX) is shown next to a topological map of the fold for reference (helices are shown as circles with the individual protein chains coloured red and blue). During unfolding, the native state (N$_2$) non-cooperatively frays through unfolding of C-terminal helices. This leaves an N-terminal dimeric intermediate (I$_2$). We note that various partially unfolded dimeric intermediate conformations could be populated during the unfolding transition. These are presented by showing the two extreme possible examples (one monomer of the dimer fully folded and the other mostly unfolding) and a third that is the median (both subunits partially unfolded to the same extent). These possible intermediate structures are shown in the grey box. At higher concentrations of urea the intermediate dissociates and denatures to a monomeric denatured state (2D, where D is the denatured state).](image)

To confirm the dimeric intermediate unfolding mechanism, analytical SEC and dynamic light scattering of tr-LcrH at 50 µM were performed in urea concentrations of 0 – 6 M urea. Both methods showed that there was an increase in the size of tr-LcrH on increasing urea concentrations (Figure 4.18). In particular, the size of tr-LcrH was
similar from 0 to 2 M urea but then significantly increased in 3 M urea as observed from the analytical SEC profile of tr-LcrH (Figure 4.18 (b)). In order to investigate LcrH’s increase in molecular size, the hydrodynamic radii at respective denaturant concentrations were plotted against the urea concentrations used for unfolding (Figure 4.18 (a) inset). It was revealed that tr-LcrH increased in size until 5 M urea and a reduction in hydrodynamic radius was observed in 6 M urea. It can be inferred that the first unfolding transition of LcrH involves partial unfolding to an expanded dimeric intermediate. By 6 M urea, tr-LcrH undergoes complete denaturation, with the shape of the peak by SEC becoming more symmetric. At, 0 M urea, tr-LcrH elutes at ~ 10.2 ml near the elution volume of 10.55 ml of the monomeric sized TPR standard CTPR3 which elutes at 10.55 ml (Figure 4.18 (b) and section 4.3.1). At 6 M urea, tr-LcrH elutes at ~ 8.6 ml in comparison to the dimeric TPR standard CTPRa6’s elution volume of 9.6 ml (Figure 4.18 (b) and section 4.3.1). Thus, from the Figure 4.18 below, it can be seen that further unfolding of tr-LcrH produced an expanded monomeric species, which was greater in size than the dimeric TPR standard CTPRa6.

Figure 4.18 (a) Dynamic light scattering [volume of protein (%) against the hydrodynamic radius (nm)] of 50 µM tr-LcrH when in 0 M urea (black), 1 M urea (blue), 2 M urea (green), 3 M urea (pink), 4 M urea (orange), 5 M urea (purple), 6 M urea (red) and inset: Hydrodynamic radius (nm) against Urea concentration (M), smooth curve plotted in black. (b) Analytical SEC using S75 HR 10/30 column: 50 µM tr-LcrH in 0 M urea (black), 3 M urea (pink), 6 M urea (red), TPR standards - CTPR3 (blue) and CTPRa6 (green). Figure was prepared using KALEIDAGRAPH 4.0.
4.3.5.2 Deletion of N-terminus Causes No Change in Equilibrium Unfolding

When the equilibrium denaturations of fl-LcrH and tr-LcrH are compared at 3 and 50 μM, their profiles are indistinguishable (Figure 4.16). It is therefore clear that fl-LcrH and tr-LcrH have identical thermodynamic equilibrium unfolding and thus must infold via the same mechanism and have the same stability. This shows that although, the N-terminus of LcrH may possess some α-helical content, it does not alter the conformational stability or the nature of equilibrium unfolding of the protein.

4.3.6 Thermodynamic stability of LcrH

Having delineated the precise mechanism of LcrH’s unfolding, the thermodynamic stabilities of the intermediate and native states were obtained by globally fitting the chemical denaturations performed at 50 and 80 μM for tr-LcrH and 50 μM for fl-LcrH to a three-state unfolding model with a dimeric intermediate (Figure 4.19). The derivation of the three-state unfolding model and Equations used to fit the data are described in detail in the following data analysis section. Lower protein concentration denaturations were excluded, as the intermediate state was not sufficiently populated to enable explicit fitting.

Figure 4.19 Urea denaturation of tr-LcrH monitored by changes in far-UV CD at 222 nm showing the global fit of the 50 and 80 μM data sets to a three-state model with a dimeric intermediate (Equation 4.4). For clarity, only one denaturation is shown per concentration: 6 μM (filled squares), 12 μM (open squares), 25 μM (filled triangles), 50 μM (open triangle) and 80 μM (filled diamonds). Figure was produced using GraphPad Prism 6.0.
4.3.6.1 Data analysis

In this study the unfolding of LcrH was fitted to a three-state model, which involves a dimeric intermediate (scheme 1)

\[
\begin{align*}
K_1 & \quad K_2 \\
N_2 & \rightleftharpoons I_2 \rightleftharpoons 2D
\end{align*}
\]

Scheme 1

where \(N_2\) is the native dimer, \(I_2\) is the intermediate dimer and \(D\) is the denatured state. In this model the equilibrium constants for the first \((K_1)\) and second \((K_2)\) transitions can be defined, respectively, as \(K_1 = [I_2]/[N_2]\) and \(K_2 = [D]^2/[I_2]\). The total protein concentration is \(P_t = 2[N_2] + 2[I_2] + [D]\), and the sum of the fractions of individual species is equal to 1: \(f_N + f_I + f_D = 1\), where \(f_i\) represents the fraction of monomeric subunits involved in the intermediate state. Combining these relationships gives:

Equation 4.9

\[
K_1 = \frac{f_I}{f_N}
\]

and

Equation 4.10

\[
K_2 = \frac{2P_tD^2}{f_I}
\]

Rearranging and substitution of the above Equations in terms of \(f_D\) (fraction of denatured protein) allows \(f_D\) to be expressed in terms of \(K_1, K_2\) and \(P_t\):

Equation 4.11

\[
f_D = \frac{-K_1K_2 + \sqrt{(K_1K_2)^2 + 8(1+K_1)(K_1K_2)P_t}}{4P_t(1+K_1)}
\]

In this model the spectroscopic signal obtained becomes:

Equation 4.12

\[
\lambda_{obs} = \lambda_N f_N + \lambda_I f_I + \lambda_D f_D
\]
Where $\lambda_{\text{obs}}$ is the spectroscopic signal, $\lambda_N$ is the signal of the native state, $\lambda_i$ is the signal of the intermediate state and $\lambda_D$ is the signal of the native state. Substituting Equation 4.9 and Equation 4.10 into Equation 4.12 gives the final fitting Equation:

$$
\lambda_{\text{obs}} = \lambda_N \frac{2P_f f_D^2}{K_1 K_2} + \lambda_I \frac{2P_f f_D^2}{K_2} + \lambda_D f_D
$$

Global fitting to the three-state dimer denaturation model

To fit the denaturant dependent data and obtain $\Delta G$ values of each step in the unfolding, $K_1$ and $K_2$ can be defined according to the linear free energy model, which states that the free energy of each unfolding process varies linearly with the concentration of denaturant [196-198]:

$$
\Delta G_x = -RT \ln K_x = \Delta G_{H_2O}^x - m [\text{denaturant}]
$$

Where $\Delta G_x$ is the free energy of the process $x$ at a specific denaturant concentration, $K_x$ is the equilibrium constant of the process $x$ (in this case $K_1$ and $K_2$), $\Delta G_{H_2O}^x$ is the free energy of process $x$ in the absence of denaturant and $m$ is a constant that is proportional to the solvent exposure difference between the initial and end state of unfolding process $x$.

Equation 4.11, with $f_D$, $K_1$ and $K_2$ defined by Equation 4.9 and Equation 4.10 (respectively), was then used to globally fit the LcrH denaturations of differing concentrations using the non-linear, least-squares fitting programme Prism. This gave global thermodynamic parameters $\Delta G_{H_2O}^1$, $\Delta G_{H_2O}^2$ (stabilities of the first and second unfolding transitions, respectively) and $m_1$, $m_2$ (the $m$-values for the first and second unfolding transitions, respectively). $\lambda_N$, $\lambda_i$ and $\lambda_D$ (the spectral signals of the native, intermediate and denatured states respectively) were allowed to vary independently with each dataset. $\lambda_i$ was assumed not to vary with denaturant concentration to minimize the number of parameters needed. Further, to reduce any baseline artifacts, the baselines ($\lambda_N$ and $\lambda_D$) were constrained to zero slope. The thermodynamic
parameters obtained are presented in Table 4.2 and confirm the identical stabilities of fl-LcrH and tr-LcrH.

Table 4.2 Thermodynamic parameters for the global fitting of fl-LcrH & tr-LcrH at 25 °C using the three state dimer model with a dimeric intermediate

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\Delta G_{H_2O}^{N_2=I_2}$ (kcal mol$^{-1}$)</th>
<th>$m_{N_2=I_2}$ (kcal mol$^{-1}$ M$^{-1}$)</th>
<th>$\Delta G_{H_2O}^{I_2=2D}$ (kcal mol$^{-1}$)</th>
<th>$m_{I_2=2D}$ (kcal mol$^{-1}$ M$^{-1}$)</th>
<th>$\Delta G_{H_2O}^{N_2=2D}$ (kcal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fl-LcrH</td>
<td>1.7 ± 0.5</td>
<td>1.0 ± 0.3</td>
<td>5.6 ± 1.8</td>
<td>1.7 ± 0.4</td>
<td>7.3 ± 1.9</td>
</tr>
<tr>
<td>tr-LcrH</td>
<td>1.7 ± 0.4</td>
<td>1.1 ± 0.2</td>
<td>5.7 ± 1.2</td>
<td>1.7 ± 0.2</td>
<td>7.4 ± 1.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>[Total Protein] (µM)</th>
<th>$b_\lambda_N$ (millideg.)</th>
<th>$c_\lambda_I$ (millideg.)</th>
<th>$d_\lambda_D$ (millideg.)</th>
<th>Cuvette pathlength (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fl-LcrH</td>
<td>50 * 1</td>
<td>-99 ± 5</td>
<td>54 ± 6</td>
<td>-5 ± 2</td>
<td>0.05</td>
</tr>
<tr>
<td>fl-LcrH</td>
<td>50 * 2</td>
<td>-95 ± 5</td>
<td>54 ± 6</td>
<td>-5 ± 2</td>
<td>0.05</td>
</tr>
<tr>
<td>tr-LcrH</td>
<td>50 * 1</td>
<td>-132 ± 5</td>
<td>-73 ± 6</td>
<td>-4 ± 2</td>
<td>0.1</td>
</tr>
<tr>
<td>tr-LcrH</td>
<td>50 * 2</td>
<td>-133 ± 5</td>
<td>-70 ± 6</td>
<td>-3 ± 2</td>
<td>0.1</td>
</tr>
<tr>
<td>tr-LcrH</td>
<td>80 * 1</td>
<td>-226 ± 8</td>
<td>-126 ± 9</td>
<td>-9 ± 3</td>
<td>0.1</td>
</tr>
<tr>
<td>tr-LcrH</td>
<td>80 * 2</td>
<td>-206 ± 7</td>
<td>-122 ± 8</td>
<td>-14 ± 4</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Errors given are from the fit of all the data and are quoted at a 95 % confidence level (2 standard errors). $\Delta G_{H_2O}^{N_2=2D} = \Delta G_{H_2O}^{N_2=I_2} + \Delta G_{H_2O}^{I_2=2D}$, with the error calculated by propagating the fitting errors. $b_\lambda_N$, $c_\lambda_I$ and $d_\lambda_D$ are the CD spectroscopic signals for native, intermediate and denatured LcrH, respectively, at 222 nm.

These show that the initial unfolding step ($N_2 \rightleftharpoons I_2$) has a $\Delta G_{H_2O}^{N_2=I_2}$ of only 1.7 ± 0.4 kcal mol$^{-1}$. Thus, the majority of LcrH’s stability resides in the dimerization structure ($I_2 \rightleftharpoons 2D$) as it requires an extra 5.7 ± 1.2 kcal mol$^{-1}$ to unfold. This means the total stability of the protein ($\Delta G_{H_2O}^{N_2=2D}$) is 7.4 ± 1.3 kcal mol$^{-1}$. Interestingly, LcrH’s stability ($\Delta G_{H_2O}^{I_2=2D}$) is less than many other dimeric proteins that have been shown to unfold via dimeric intermediate (Table 4.3). In particular, the first transition ($N_2 \rightleftharpoons I_2$) of LcrH was significantly lower than, for example, eAATase, Ure2p, SecA, OPH and AAO proteins. This confirms the non-cooperative and labile nature of LcrH’s initial unfolding.
Table 4.3 Thermodynamic parameters characteristic of chemical induced unfolding of dimeric proteins by a three-state transition with dimeric intermediate

<table>
<thead>
<tr>
<th>Protein</th>
<th>Temp. (°C)</th>
<th>( \Delta G_{H_2O}^{N_2=I_2} ) (kcal mol(^{-1}))</th>
<th>( m_{N_2=I_2} ) (kcal mol(^{-1}) M(^{-1}))</th>
<th>( \Delta G_{H_2O}^{I_2=2D} ) (kcal mol(^{-1}))</th>
<th>( m_{I_2=2D} ) (kcal mol(^{-1}) M(^{-1}))</th>
<th>( \Delta G_{H_2O}^{N_2=2D} ) (kcal mol(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>tr-LcrH</td>
<td>25</td>
<td>1.7</td>
<td>1.1</td>
<td>5.7</td>
<td>1.7</td>
<td>7.4</td>
</tr>
<tr>
<td>bEAAATase</td>
<td>25</td>
<td>12.0</td>
<td>4.8</td>
<td>24.4</td>
<td>3.4</td>
<td>36.4</td>
</tr>
<tr>
<td>cSecA</td>
<td>20</td>
<td>8.4</td>
<td>4.1</td>
<td>14.1</td>
<td>1.5</td>
<td>22.5</td>
</tr>
<tr>
<td>dOPH</td>
<td>25</td>
<td>4.3</td>
<td>1.0</td>
<td>36.1</td>
<td>4.3</td>
<td>40.4</td>
</tr>
<tr>
<td>eAAO</td>
<td>10</td>
<td>3.5</td>
<td>1.7</td>
<td>13.6</td>
<td>1.2</td>
<td>17.1</td>
</tr>
</tbody>
</table>

\( \Delta G_{H_2O}^{N_2=2D} = \Delta G_{H_2O}^{N_2=I_2} + \Delta G_{H_2O}^{I_2=2D} \). To give the best comparison with our data on LcrH the proteins chosen in this table were selected as the experimental conditions used to obtain them were similar to ours. i.e. urea as chemical denaturant, CD or fluorescence globally fitted, temperature between 10 – 25 °C and pH of between 7-8. The values quoted above are from the following studies: \(^{b}\)EAAATase - [198,199], \(^{c}\)SecA -[199,200], \(^{d}\)OPH -[200,201] and \(^{e}\)AAO -[201-203].

An apparent \( K_D^{app} \) (\( K_D^{app} \)) for the dissociation of the intermediate to denatured monomer can also be calculated from the equilibrium data. As \( \Delta G_{H_2O}^{I_2=2D} \) (from the seconds unfolding transition) is the free energy of unfolding and dissociation of the intermediate LcrH to denatured monomeric LcrH, it is related to the \( K_D^{app} \) by \( \Delta G_{H_2O}^{I_2=2D} = -RT\ln K_D^{app} \). Thus, \( K_D^{app} \) can be simply obtained by rearranging this Equation. However, two caveats need to be considered before delineating its true significance.

(i) \( K_D^{app} \) corresponds to both unfolding and dissociation of the intermediate state of LcrH. This is not exactly equivalent to the \( K_D \) value obtained from the SEC and AUC data, which correspond to the dissociation of folded dimeric LcrH to folded monomeric LcrH. (ii) the \( K_D^{app} \) obtained in different solution conditions (high urea concentration). Given these caveats, the \( K_D^{app} \) calculated was 65 ± 65 µM and is consistent with the \( K_D \) calculated from AUC data and inferred from SEC data.

4.3.6.2 Proposed equilibrium unfolding pathway:

The ability of LcrH to easily fray into partially folded structures is facilitated by its modular tetratricopeptide repeat (TPR) structure. Unlike globular proteins, TPR and other repeat containing proteins are formed from the stacking of modular secondary structure motifs [202-206]. This causes their native structures to be dominated and
stabilized by interactions from amino acid residues that are close in primary sequence, whereas globular proteins have many stabilizing interactions from amino acids that are distant in primary sequence. This arrangement can enable sections of repeat proteins to unravel without unfolding the complete structure. We have studied the folding of a number of designed TPR proteins and shown that when partial unfolding occurs the proteins tend to unravel from the less stable outer repeats first [168,204-206]. As the N-termini of LcrH form the dimeric interface (which our results show is the most thermodynamically stable part of the structure), the unraveling must occur mainly from the C-terminus. A schematic diagram of the equilibrium-unfolding pathway proposed by our results is shown in Figure 4.17.

4.3.6.3 Thermodynamic Stability of Monomeric vs Dimeric LcrH

To complete the stability studies conducted on LcrH, the stability and flexibility of the monomeric mutant L65E LcrH was investigated. In vivo studies have shown that when SycD is made monomeric through the L65E mutation, non-invasive bacteria are produced [168,207]. This suggests that the dimeric interface is important for function. When we produced L65E tr-LcrH, we also obtained a monomer (at 50 µM there is a single SEC elution peal of 10.7 ml, which is similar to an elution peak of 10.55 ml obtained for a monomeric designed TPR protein of similar dimensions (3.5 stacked TPRs, Figure 4.2). Furthermore at 25 °C, L65E has the same α-helicity by CD to dimeric tr-LcrH (Figure 4.20a). However, chemical denaturation of L65E shows that the mutant was substantially destabilised compared to dimeric tr-LcrH with a single sharp transition between native and denatured states and little native baseline (Figure 4.20b). This was confirmed, as fitting the data to a two-state model showed the midpoint of the transition ([D]50%) is only 1.6 ± 0.1 M urea, the m-value of the transition is 1.4 ± 0.1 kcal mol⁻¹ M⁻¹ and thus its stability (ΔG^N\rightarrowD_{H_2O} = m.[D]50%) is 2.1 ± 0.2 kcal mol⁻¹ (ΔG^N\rightarrowD_{H_2O} = m.[D]50%).
Figure 4.20 (a) Far-UV CD spectra comparing native fl-LcrH (black), tr-LcrH (blue), L65E tr-LcrH (green), and denatured L65E tr-LcrH in 8 M urea (red). A protein concentration of 50 µM in a 1-mm path length quartz cuvette was used in each experiment. The inset figure shows the same experiments conducted at a protein concentration of 3 µM in a 5-mm path length cuvette. Note that fl-LcrH has ~10% greater α-helical signal to tr-LcrH and L65E tr-LcrH at 222 nm. (b) Urea denaturation of L65E mutant of tr-LcrH (open blue circles, 1.5 µM, and filled blue circles, 3 µM) and tr-LcrH at 3 µM (black open and filled circles for comparison) monitored by changes in far-UV CD at 222 nm and converted to relative spectroscopic signal (see Equation 2.12). The best fit for L65E to a two state folding model is shown (black). Figure was prepared using KALEIDAGRAPH 4.0.

This stability is similar to the initial unfolding of the dimer tr-LcrH, but is substantially more cooperative and results in a complete loss of structure. Moreover, the low stability of L65E indicates that the in vivo results may well be caused by the destabilisation of LcrH to the extent that the monomers produced are unfolded and thus non-functional.
4.3.7 Conclusion

The results in this chapter show that LcrH is a weak dimer ($K_D \approx 15 \, \mu M$) that readily unfolds in a non-cooperative manner to a partially unravelled dimeric intermediate. In contrast, the monomeric mutant L65E tr-LcrH unfolds more cooperatively on chemical denaturation, resulting in a complete loss of structure. Thus, confirming the significance of the dimerization interface of LcrH for thermodynamic stability. Furthermore, the chapter shows that although deletion of the N terminus does remove some $\alpha$-helical structure (10%), it does not affect either its stability or dimerization equilibrium. Excitingly, the energy barrier for the partial unravelling of dimeric LcrH is low ($\Delta G_{H2O} \approx 1.7 \, \text{kcal mol}^{-1} \text{ at } 25 \, ^\circ C$), suggesting that partially unfolded states of LcrH (held together at the dimeric interface) are present at physiological conditions. It is therefore proposed that such a loosely folded structure has biological relevance, at the very least, when binding its cognate protein partners (the far larger membrane inserting translocator proteins).
Chapter 5 Conclusions

The Black Death was one of the most devastating pandemics in the history of human kind. It resulted in the death of nearly 50 million people in Asia, Africa and Europe in the 14th century AD [7,207]. The pathogen responsible for this deadly infection was the bacterium Yersinia pestis which causes several forms of plague [7,208]. On infection by Y. pestis, the host’s defences are overcome by a combination of stealth and active immune suppression targeting innate immune mechanisms. The type three secretion system (T3SS) of Y. pestis is crucial to the suppression arm of this strategy. This system injects a set of effector (virulent) proteins into the host cell cytoplasm when coming in contact with the bacteria compromising the host’s immune system by affecting phagocytosis and cytokine production[208,209]. Thus, T3SS provides the key virulence factor in pathogens like Y. pestis. Furthermore, the T3SS is found in over two dozen infectious Gram-negative bacteria such as various strains of E. coli and Pseudomonas aeruginosa, a potent biofilm former with a strong connection to cystic fibrosis. Without T3SS, many pathogenic bacteria are unable to cause disease [209]. This makes the T3SS an attractive target for novel antimicrobial drugs.

Recently, the World Health Organisation (WHO) published a report on antibiotic resistance and highlighted that it is an imminent threat and needs to be immediately addressed [5,210]. According to the WHO, the true cost of antimicrobial resistance will be 300 million premature deaths and up to $100 trillion (£64 trillion) lost to the global economy by 2050. This will severely impact the world’s GDP, which would be 0.5 % smaller by 2020 and 1.4 % smaller by 2030 [211]. Approximately 10 % of the microbes that would be linked to these deaths employ the T3SS making it an extremely attractive target for antimicrobial agents. In order to develop drugs more efficiently, a complete biochemical understanding of the T3SS, such as the formation of the infection machinery and the mechanism of infection are required.

To do this, I began studying the T3SS in Y. pestis and more specifically the class II chaperone LcrH to fully understand the conformational stability and biochemical behaviour of this protein. LcrH is critical for the formation of a fully functional T3SS and so too the infection process of mammalian cells by Y. pestis. Consequently, any
information regarding the biochemical characteristics of such a significant protein can provide a potential therapeutic drug target.

In order to gain insight into the protein’s behaviour, the full length sequence encoding for the protein LcrH was cloned for protein production and purification. Despite being able to produce and purify LcrH to >95% the protein proved extremely prone to aggregation. In order to overcome this problem, a construct was created with the N-terminal 20 amino acids and the last 5 amino acids removed (tr-LcrH). Here, the protein proved to be significantly less prone to aggregation. CD studies showed that the removal of the N-terminus of LcrH caused some loss in α-helical content but had no effect on the stability of the protein.

In order to determine the oligomeric state of both forms of LcrH and thus obtain a $K_D$ value for dimerisation, SEC and equilibrium sedimentations were performed. These showed that both full-length and truncated LcrH possess the same weak monomer-dimer equilibrium (low μM $K_D$), thermodynamic and conformational stability. Importantly, chemical denaturation revealed that both full length and truncated forms of LcrH unfolded in a 3-state transition, the states being (i) folded, (ii) partially folded dimeric intermediate and (iii) unfolded. This was initially surprising as most proteins in the TPR family, of which LcrH is one, unfold in a cooperative manner. Such non-cooperative behaviour must be due to the dimerisation interface. The dimer is created through interaction between the N-terminus thus the weakly folded intermediate transition must be caused by fraying at the C-terminus; as the C-terminus begins to unfold, the protein loses integrity before completely unfolding.

Next, a mutation was introduced that has been shown to disrupt the dimerisation interface and create a protein that was solely monomeric. This protein was termed LE65-tr-LcrH and was produced, purified and characterised using identical biochemical and biophysical methods as the dimer. This revealed that the mutant had the same α-helical content as dimeric tr-LcrH but, as expected, did not exist as a dimer. Furthermore, chemical denaturation studies revealed that the mutant was substantially destabilised compared to tr-LcrH with a single sharp transition between the native folded and the denatured states, a classic 2-state transition. Taken in
conjunction with the information gathered from the dimer, it is therefore evident that the dimer interface provides additional thermodynamic stability to the global structure. Consequently, some structures close to the interface (N-terminus) remain folded at equilibrium while other parts of the structure are unfolded (C-terminus).

Previous studies have illustrated the significance of the N-terminus in determining the arrangement of the subunits within the dimeric structure of class II chaperones. For example, lpgC_{10-151} produces head to head dimers whereas full length lpgC forms a back to head dimer [169]. However, the results presented here clearly show that removal of the first 20 N-terminal amino acids have no effect on stability of LcrH and therefore it is unlikely that these residues are directly involved in subunit positioning within the dimer.

Within the T3SS in Y. pestis, LcrH interacts with the large membrane translocator proteins YopB and YopD. However, the mechanism by which these small chaperones find their cognate sequences within the large membrane proteins remains unclear. Excitingly, the results presented here support one possible mechanism, termed “fly-casting”. Here, the weakly folded C-terminal repeats of the chaperone adopt a partially folded conformation allowing it to more easily find the target peptide on the translocator protein. The chaperone then refolds and remains attached to the translocator, producing a more rigid structure. It is important that the integrity of the N-terminus is retained during this process as in vivo studies have shown that a monomeric LE65 LcrH homologue in SycD renders the bacteria non-invasive [168]. Such a scenario is not without precedent. For example, the three TPR motif-containing domain of PP5 has been shown to become more structured when bound to its cognate binding partner, the C terminus of Hsp90 [212]. Moreover, studies on the six ankyrin repeat protein IκBα show it also has a weakly folded C terminus. Here, repeats five and six fold upon binding to their cognate partner, the transcription factor NF-κB [198,213]. It has been shown that the function and lifetime within the cell of IκBα is critically linked to whether this region is structured via ligand binding or not [199]. This has yet to be explored for the translocator chaperones.
In summary, this result shows that the chaperone LcrH has a very flexible and weakly folded C-terminus and forms a dimer that is held together by the dimerisation interface at the N-terminus. We propose that such a flexible structure has biological relevance that is important in binding its cognate protein partners. These results should be generally applicable to all translocator chaperones, as they all contain the same TPR motif fold. Furthermore, both the interaction between the N-terminal dimerisation interface and the C-terminus with its cognate-binding partners are potential molecular targets as an alternative to broad-spectrum antibiotic therapy.

5.1 Future work

The future work in this area must surely focus on the protein-protein interaction between the chaperone LcrH and its cognate binding partner YopB or YopD. YopB and YopD are large transmembrane proteins. So, to characterise the interaction it would be the right step forward to create a peptide of the binding domain.

To our advantage, crystallographic studies have been carried out characterising peptides binding to various domains of the cognate binding partners in *P. aeruginosa* and *Shigella flexneri* but the K$_D$ values have been weak. Therefore, the design of peptides that bind more strongly is the next step. Indeed, the ability to disable binding of the translocator to the chaperone through competitive inhibition is a key area of research.

Such studies employing the T3SS from *Y. pestis* as a model system will surely provide an excellent molecular understanding that could lead to molecularly targeted treatment. With so many potent pathogenic Gram-negative bacteria that employ the T3SS, the significance of this system cannot be ignored.
Chapter 6 Bibliography


Lu, C., Turley, S., Marionni, S. T., Park, Y.-J., Lee, K. K., Patrick, M., Shah, R.,


