Structural determination and functional annotation of ChuS and ChuX, two members of the heme utilization operon in pathogenic Escherichia coli O157:H7

By

MICHAEL DOUGLAS LEO SUITS

A thesis submitted to the Department of Biochemistry in conformity with the requirements for the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada

April, 2007

Copyright © Michael Douglas Leo Suits, 2007
Abstract

For pathogenic microorganisms, heme uptake and degradation is a critical mechanism for iron acquisition that enables multiplication and survival within hosts they invade. While the bacterial proteins involved in heme transport had been identified at the initiation of our investigation, the fate of heme once it reached the cytoplasm was largely uncharacterized. Here we report the first crystal structures of two members of the heme utilization operon from the human pathogen *Escherichia coli* O157:H7. These are the heme oxygenase ChuS in its apo and heme-complexed forms, and the apo form of heme binding protein ChuX. Surprisingly, despite minimal sequence similarity between the N- and C-terminal halves, the structure of ChuS is a structural repeat. Furthermore, the ChuS monomer forms a topology that is similar to the homodimeric structure of ChuX. Based on spectral analysis and carbon monoxide measurement by gas chromatography, we demonstrated that ChuS is a heme oxygenase, the first to be identified in any *E. coli* strain. We also show that ChuS coordinates heme in a unique fashion relative to other heme oxygenases, potentially contributing to its enhanced activity. As ChuS and ChuX share structural homology, we extended the structural insight gained in our analysis of ChuS to purport a hypothesis of heme binding for ChuX. Furthermore, we demonstrated that ChuX may serve to modulate cytoplasmic stores of heme by binding heme and transferring it to other hemoproteins such as ChuS. Based on sequence and structural comparisons, we designed a number of site-directed mutations in ChuS and ChuX to probe heme binding sites and mechanisms in each. ChuS and ChuX mutants were analyzed through reconstitution experiments with heme and functional analyses, including enzyme catalysis by ChuS and mutants, and in culture development during heme challenge experiments by ChuX and mutants. Taken together, our results suggested that ChuX acts upstream of ChuS, and regulates heme uptake through ChuX-mediated heme binding and release. ChuS can degrade heme as a potential iron source or antioxidant, thereby contributing directly to *E. coli* O157:H7 pathogenesis. Functional implications that may be revealed from sequence and structure based information will be addressed as they pertained to our evaluation of ChuS and ChuX.
Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my supervisor, Dr. Zongchao Jia, for his guidance, support, and belief in my abilities. His clever insight and unwavering encouragement have enabled me to attain numerous goals and have instilled in me the drive to continue to seek greater scientific achievement. I would also like to acknowledge the support of past and present members of the Jia research group, including Dr. Melanie Adams for her crystallographic expertise and contagious drive, and Drs. Gour Pal, Vinay Singh and Qilu Ye, who have allowed me to draw on their vast and various scientific backgrounds. I am also grateful to Jarrett Adams, Daniel Lee, Peter Iannuzzi, Dr. Allan Matte, Dr. Chris Udell, Andrew Wong, and Jimin Zheng for their support, cooperation and technical expertise. Thank you to Drs. Kanji Nakatsu, Michael Nesheim, Steve Smith, and Glenville Jones for professional support and mentoring. I extend sincere appreciation to the Synchrotron facilities, notably the beamline scientists at APS, NSLS, and CHESS for technical support and assistance in interpreting diffraction data. I am especially grateful to Leonid Flaks and Howard Robinson, at NSLS beamlines X8C and X29, respectively.

I would like to acknowledge my siblings, parents and grandparents, to whom I am so appreciative for their unbounded love and encouragement. Their collective passion for learning, quest for knowledge, and enduring support has inspired my own academic career. Thank you, thank you, and thank you.

I would like to thank my wife Jocelyne for speaking the same language, and for encouraging me through technical insights and exceptional writing abilities to better convey ideas and concepts on paper. Thank you for sharing love and laughter in all aspects of our life.

This research has been funded by the Canadian Institutes of Health Research and various Queen’s University Fellowships including: Queen’s Graduate Award, R.S. McLaughlin, E.G. Bauman, and Franklin Bracken fellowships and Graduate Dean’s Travel Grant for Ph.D. Field Research. Further support was provided by an Ontario Graduate Scholarship.
TABLE OF CONTENTS

Abstract                                                ii
Acknowledgements                                         iii
List of Tables                                           vii
List of Figures                                          viii
List ofAbbreviations                                     x
List of Protein Mutants                                  xi

Chapter 1: General Introduction                          1
  1.1. General introduction and relevance                  2
  1.2. Bacterial iron acquisition from heme sources       4
  1.3. Role of heme and structural diversity of heme-proteins 12
  1.4. Bacterial heme oxygenases                           15
  1.5. Heme utilization protein ChuS                       17
  1.6. ChuW, ChuX, ChuY and bacterial heme proteins        18

Chapter 2: Identification of an *Escherichia coli* O157:H7 heme oxygenase with tandem functional repeats 22
  2.1. Abstract                                           23
  2.2. Introduction                                       24
  2.3. Materials and methods                              27
    2.3.1 Sequence analysis, protein expression, and purification 27
    2.3.2 Crystallization of full-length ChuS               28
    2.3.3 Data collection and structure determination       28
    2.3.4 Reconstitution with heme and absorbance measurements 29
    2.3.5 CO activity assay and inhibition                  30
    2.3.6 Dynamic light scattering of C-ChuS and N-ChuS     30
    2.3.7 EPR experiments                                   31
  2.4. Results                                            31
    2.4.1 Expression and purification                       31
    2.4.2 Structure determination and analysis              32
    2.4.3 ChuS contains a structural duplication            36
    2.4.4 Spectral properties of the heme-ChuS complex      38
    2.4.5 Heme oxygenation by ChuS                         43
    2.4.6 Detection of CO release                           45
    2.4.7 Organization of the heme uptake operon            45
    2.4.8 Dynamic light scattering                          46
    2.4.9 Inhibition by Sn(IV) mesoporphyrin                47
  2.5. Discussion                                          51
  2.6. Acknowledgements                                   54
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Summary of proteins encoded within the heme utilization operon from pathogenic <em>Escherichia coli</em> O157:H7.</td>
<td>11</td>
</tr>
<tr>
<td>Table 2.1</td>
<td>Diffraction data and refinement statistics for ChuS.</td>
<td>34</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Diffraction data and refinement statistics for ChuS-Heme.</td>
<td>66</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Diffraction data and refinement statistics for ChuX.</td>
<td>94</td>
</tr>
<tr>
<td>Table A.1</td>
<td>Diffraction data and refinement statistics for LptA.</td>
<td>157</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Structure of Heme (protoporphyrin IX) and mechanism of heme degradation catalyzed by heme oxygenases.</td>
<td>6</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Schematic of the heme assimilation system from Gram-negative bacteria at the initiation of our investigation and organization of the heme utilization operon from <em>Escherichia coli</em> O157:H7.</td>
<td>20</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Ribbon diagram of ChuS.</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>The two halves of ChuS superimposed, with an overall rmsd of 2.1 Å.</td>
<td>37</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Spectral analysis of ChuS reconstituted with heme.</td>
<td>40</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>EPR analysis of ChuS reconstituted with heme.</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>ChuS reconstituted with heme.</td>
<td>42</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Spectral change of ChuS reconstituted with heme, monitored over time following ascorbic acid addition.</td>
<td>44</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Compiled average CO detected for ChuS, N-ChuS, and C-ChuS.</td>
<td>46</td>
</tr>
<tr>
<td>Figure 2.8</td>
<td>Sequence alignment of ChuS with homologues from various Gram-negative bacteria.</td>
<td>49</td>
</tr>
<tr>
<td>Figure 2.9</td>
<td>Inhibition of ChuS heme degradation by Sn(IV) mesoporphyrin IX.</td>
<td>50</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Ribbon diagram of ChuS in complex with heme.</td>
<td>67</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Difference omit map for the heme moiety and two water molecules (blue) within the active site pocket of ChuS contoured to 2σ in the R3 space group.</td>
<td>70</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Spectral change of ChuS reconstituted with heme, monitored over time following ascorbic acid addition.</td>
<td>73</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>CO measurement via gas chromatography of ChuS, H73A, H193 and heme following ascorbic acid addition and 25 min incubation (n=3).</td>
<td>74</td>
</tr>
</tbody>
</table>
Figure 4.1  Sequence alignment of Chux with homologues from various Gram-negative bacteria. 93
Figure 4.2  Ribbon diagram of the four ChuX molecules in the crystallographic asymmetric unit. 95
Figure 4.3  Structural similarity of ChuX with ChuS and putative heme binding cleft. 98
Figure 4.4  Spectral analysis of ChuX reconstituted with heme. 102
Figure 4.5  Spectra of ChuX and mutants reconstituted with heme 103
Figure 4.6  ChuX protection against heme toxicity and heme transfer by ChuX. 106
Figure 5.1  Heme degradation by ChuS in the presence of the hHO-1 specific inhibitor (2[2-(4-chlorophenyl)ethyl]-2-[(1H-imidazol-1-yl)methyl]-1,3-dioxolane). 126
Figure 5.2  Modified skemetic of the heme assimilation system from Gram-negative bacteria and organization of the heme utilization operon from *Escherichia coli* O157:H7. 129
Figure A.1  Sequence alignment of LptA with various Gram-negative bacterial homologues, many of which are human pathogens. 154
Figure A.2  Q-TOF Mass Spectrometry result, highlighting the LptA has been processed prior to crystallization. 155
Figure A.3  Anisotropic diffraction pattern from LptA MAD (left) and two-molecule native (right) crystals. 158
Figure A.4  Ribbon diagram of LptA with two molecules in the asymmetric unit at 2.6 Å resolution. 160
Figure A.5  Ribbon diagram of LptA highlighting residues conserved in sequence alignment with other Gram-negative homologues. 161
Figure A.6  Ribbon diagram of LptA with eight molecules in the asymmetric unit at 3.25 Å resolution. 162
ABBREVIATIONS

The Abbreviations used throughout this manuscript are:

ANL  Argonne National Laboratory
APS  Advanced Photon Source
ATP  adenosine triphosphate
BGG  bovine globulin G
BV  biliverdin
BR  bilirubin
BNL  Brookhaven National Laboratory
BSA  bovine serum albumin
CCD  charge-coupled device
CHESS Cornell High Energy Synchrotron Source
CN  cyanide
CO  carbon monoxide
CPR  cytochrome P450 reductase
DLS  dynamic light scattering
DNA  deoxyribonucleic acid
EHEC  enterohemmorhagic Escherichia coli
EPR  eletroparamagnetic resonance
FUR  ferric uptake regulator
GST  glutathione S transferase
HO  heme oxygenase
HO-1  heme oxygenase-1
HO-2  heme oxygenase-2
HPLC  high-performance liquid chromatography
HUS  haemolytic uraemic syndrome
IPTG  isopropyl-1-thio-β-D-galactopyranoside
IM  inner membrane
KDO  3-deoxy-D-manno-oct-2-ulosonic acid
LB  Luria Burtani Broth
lipid A  a phosphorylated glucosamine disaccharide acylated with fatty acids
LPS  lipopolysaccharide
MacCHESS Macromolecular diffraction facilities at CHESS
MAD  multiple anomalous dispersion
MS/MS  tandem mass spectrometry
MW  molecular weight
NADH  reduced β-nicotinamide adenine dinucleotide
NADPH  reduced β-nicotinamide adenine dinucleotide phosphate
Ni-NTA  nickel nitrilotriacetate agarose
NO  nitrogen monoxide
NMR  nuclear magnetic resonance spectroscopy
NSLS  National Synchrotron Light Source
OM  outer membrane
ORF  open reading frame
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PDB</td>
<td>protein data bank</td>
</tr>
<tr>
<td>rmsd</td>
<td>root mean squared deviation</td>
</tr>
<tr>
<td>SAD</td>
<td>single anomalous dispersion</td>
</tr>
<tr>
<td>SCOP</td>
<td>structural classification of proteins</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SSM</td>
<td>secondary structure matching</td>
</tr>
<tr>
<td>STEC</td>
<td>shiga-toxin producing <em>Escherichia coli</em></td>
</tr>
<tr>
<td>TB</td>
<td>terrific broth</td>
</tr>
<tr>
<td>TCA</td>
<td>trichloroacetic acid</td>
</tr>
</tbody>
</table>
LIST OF MUTANT PROTEINS

Mutants generated as part of this investigation include:

**ChuS Mutants:**

H73A, H193N, N-ChuS (N-terminal residues M1-P171 fused to a C-terminal His$_6$-tag),
C-ChuS (C-terminal residues V172-A342 fused to an N-terminal GST-tag).

**ChuX Mutants:**

H65L, H98D, BetaX (ChuX triple mutant V69E/L71H/F73S), and DM (ChuX double
mutant H65L/H98D).
Chapter 1

General Introduction
1.1. General introduction and relevance

*Escherichia coli* is a constituent of the normal flora of the large intestine in healthy vertebrates, and is widely distributed in soil and water. However, certain pathogenic strains of *E. coli* such as O157 and CFT073 are known to cause a host of clinical diseases in humans and animals including: mild, self-limiting diarrhea; severe invasive gastro-intestinal infections such as dysentery and hemorrhagic colitis; urinary tract infections; septicemia; and meningitis (131). While many of the patients infected with these enterohaemorrhagic strains of *E. coli* (EHEC) (estimated 73,000 cases annually in the United States of America (108)) respond well to treatment and recover within a short period of time (70), the World Health Organization estimates that up to 10% of EHEC cases develop into hemolytic uremic syndrome (HUS), with an associated fatality rate of 3-5 % (3). As a result, O157 is classified as an opportunistic pathogen and a global public health concern. Other *E. coli* strains such as O26:H11 and O111:NM share a similar pathogenic potential, but it is the O157:H7 serotype that has recently been responsible for a few large EHEC outbreaks as well as a series of smaller cases worldwide (56, 99). The sporadic cases of O157 outbreaks are often due to exposure to undercooked ground beef (1), unpasteurized dairy (71), contaminated water and fruits or vegetables (4, 13, 150), person to person transfer (23), and is especially harmful to patients under five years of age (108). In Ontario, the highly publicized Walkerton tragedy (May, 2000) was due to *E. coli* O157:H7 and *Campylobacter jejuni* contaminated drinking water, which resulted in an estimated 2,300 people being infected; seven cases of which were fatal (62).
Despite that *E. coli* is the model organism for scientific investigation there still remain a large number of proteins and open reading frames (ORFs) from various strains with either unknown function or with debatable annotation based on weak sequence homology. Even in K-12, which is the most common and non-pathogenic strain of *E. coli*, an estimated 30% of identified genes have unknown function and many other genes are poorly characterized (68). A comparison of the O157 and K-12 *E. coli* genomes using Sakai oligoDNA microarray and whole genome PCR scanning revealed that the two chromosomes share 4.1 Mb of genetic information, with 1.4 Mb of O157 sequence-specific genetic information (56, 99). Much of this O157-specific DNA is thought to have arisen for horizontally transferred foreign DNA, and contains many bacteriophage elements that mediate genetic reorganization and exchange, contributing to its pathogenesis through frequent genetic change (56). Sequence analysis suggested the O157 specific DNA encodes for 1632 unique ORFs, 20 unique tRNAs, and that at least 131 of the encoded proteins are recognized virulence factors (56). These unique elements are therefore highly important, as understanding their function and interplay with other protein partners will contribute to our overall understanding of bacterial pathogenesis as well as principles and mechanism of infection. One level of protein characterization involves structural determination of O157 specific proteins via cryo-electron microscopy, NMR, and X-ray crystallography. As structural information can provide a plethora of detail into the precise 3-D folding, interaction of proteins with substrates and other molecules, mechanisms of action, and interaction with other proteins; as a whole, structural characterization has the potential to improve our ability to prevent disease and improve treatments in infected patients (74). The focus of this body of work is the
structural determination and functional annotation of ChuS and ChuX, two members of the heme utilization operon in pathogenic *E. coli* O157:H7.

### 1.2. Bacterial iron acquisition from heme sources

Iron is the fourth most abundant element in the Earth’s crust (6). However, due to the low solubility of iron(III) salts at physiological pH in the presence of oxygen, iron uptake, storage, and transport is a serious obstacle for bacteria to overcome for survival and pathogenesis. A variety of bacterial iron-acquisition and transport systems have evolved to assist survival and pathogenesis, such as the secretion of low-molecular-weight (MW) siderophores (FyuA and IreA) capable of binding tightly to free iron and mediating iron internalization via iron-specific membrane receptors (33, 100). Reinforcing the importance for bacterial iron acquisition is that the iron assimilation strategies they employ are highly redundant, for example, at least seven different mechanisms have been identified in *E. coli* (130). As part of a coordinated immune response against invading microorganisms, mammals specifically limit iron availability by producing the iron-binding proteins transferrin and lactoferrin, which depress free extracellular iron to a concentration of $\sim1\times10^{-18}$ M in serum, insufficient to sustain bacterial growth (6, 20). Furthermore, invading microorganisms are confronted with the obstacle that 99.9% of iron within the human body is sequestered as protein bound heme in hemoglobin, myoglobin, and various other heme enzymes (130).
Through the sequential participation of eight different enzymes, partly in the mitochondria and partly in the cytoplasm, nucleated human cells and many bacteria can synthesize heme from glycine and succinyl CoA (154). Heme is a propionic substituted tetrapyrrole porphyrin group complexed to a central iron atom (Figure 1.1A). As a result of the delocalized $\pi$-electron system of the porphyrin ring, the redox properties of the central iron atom, and the variety of unique interactions possible for heme coordination (104), heme is a widely used prosthetic group that enables proteins from various organisms to fulfill many vital biological roles in photosynthesis, cell differentiation and proliferation, as well as oxygen binding, transport, and utilization (89, 154). Beyond heme simply acting as a prosthetic group, in HeLa cells, succinyl acetone-induced heme deficiency was shown to increase the protein levels of the tumor suppressor gene product p53 and CDK inhibitor p21. The effect was an overall decrease in the protein levels of Cdk4, Cdc2, and cyclin D2, and diminish the activation/phosphorylation of Raf, MEK1/2, and ERK1/2-components of the MAP kinase signaling pathway (168). This result suggests a role for heme as an effector molecule beyond the regulation of proteins associated with heme and iron metabolism. Furthermore, the regulatory potential effected by hemoproteins contribute to all levels of gene expression including; DNA splicing, transcription, mRNA stability, translation, and post-translational modifications (154). The cumulative effect of heme uptake is therefore a cascade of gene up-regulation, with heme effectively acting as a signaling molecule for invading microorganisms that they have colonized a host (83). As a result, understanding the proteins and mechanisms involved heme metabolism, transfer, and utilization has the potential to contribute greatly to our understanding of many integral cellular processes as well as pathogen-host interactions.
Figure 1.1 Structure of Heme (protoporphyrin IX) and mechanism of heme degradation catalyzed by heme oxygenases.  

A) Carbon (yellow), oxygen (red), nitrogen (blue) form a conjugated planar molecule that coordinates a central iron (orange) atom. Meso-carbon positions are denoted with Greek symbols. The heme group is rotated to show that the molecule porphyrin is planar, and that the propionic side chains and vinyl groups extend out of the plane of the porphyrin ring. 

B) Heme oxygenase catalyzed degradation of heme to biliverdin requires a total of seven electrons and two equivalents of molecular oxygen. Most heme oxygenases are regioselective for the α-meso carbon position of the heme group.
Some microorganisms such as *Haemophilus* and *Bacterioides* are unable to synthesize heme and hence must acquire heme in large amounts from host organisms. Other bacteria, such as *Borrelia* and *streptococci*, are thought to neither synthesize nor require heme for growth (130). Many other pathogens have been identified which use host heme sources and have had heme utilization genes identified including: *Bordetella pertussis* (18), *Bradyrhizobium japonicum* (106), *C. jejuni* (111), *Haemophilus influenzae* (29), *Neisseria meningitides* (129), *Porphyromonas gingivalis* (124), *Serratia marcescens* (50), *Vibrio cholera* (58), *Yersinia enterocolitica* (128), *Yersinia pestis* (63), *E. coli* O157:H7 (91), and *Shigella dysenteriae* (91). Certain strains have evolved the capacity to actively seek out and utilize host heme sources though the production of host-specific hemoprotein proteases such as EspC (38), Pic (59), or Hbp (147), which separate heme from host proteins such as hemoglobin and hepatoglobin. Heme may then be taken up by bacterially produced hemophores such as HxuA (28) and HasA (50, 85), or transferred directly to outer membrane receptors that have been identified in at least 18 microorganisms, many of which are human pathogens (48). Another class of receptors interact with hemoproteins directly, mediating heme delivery without the need for hemophores (156). However, sequence comparison of human heme oxygenase-1 (HO-1) and bacterial HOs have failed to reveal homologues in some heme-utilizing bacteria such as *Escherichia, Plesiomonas, Shigella, Vibrio,* and *Yersinia* (130).

Heme has a higher solubility relative to free iron and is therefore a more readily transportable source of the metal (130). Once internalized, heme can readily be degraded by HOs, releasing iron (Fe$^{2+}$) along with the products biliverdin (BV) and carbon monoxide (CO). In mammals, the downstream products of heme catabolism have
recently been found to mediate the antioxidant, anti-apoptotic, anti-proliferative, vasodilatory, and anti-inflammatory properties of HO-1 in mammalian cells (45). For bacteria, the liberated iron atom can then be incorporated into various proteins where it acts as a biocatalyst or electron carrier, enabling organisms to fulfill many vital biological processes such as \( \text{N}_2 \) fixation, methanogenesis, \( \text{H}_2 \) production and consumption, respiration, the citric acid cycle, oxygen transport, gene regulation, and DNA biosynthesis (6, 89). Alternatively, heme cleaved at the \( \alpha \)-meso carbon position by HOs produces BV IXa, which can subsequently be converted to phytobilins, precursors of the chromophores for the light harvesting phycobiliproteins and photoreceptor phytochromes (44). The structure and function of HOs will be addressed in Section 1.4.

While heme is one of the most widespread enzyme cofactors in nature (130), the properties which contribute to its usefulness in a controlled protein associated form also make heme potentially cytotoxic in the free form. This is because heme is both a hydrophobic molecule capable of freely diffusing into membranes where it can alter the bilayer structure, thereby disrupting cell integrity (119), and can act as a potent prooxidative agent, promoting the light-dependent formation of reactive oxygen species that can cause non-enzymatic redox reactions (54). Heme synthesis and metabolism with its utilization must therefore be tightly regulated; for this reason heme is normally found tightly associated in various hemoproteins (54). Furthermore, invading microorganisms must also tightly regulate heme acquisition and metabolism by either degrading heme or storing it for later use in various processes (155). The regulation of heme uptake and utilization must also be coordinated with iron requirements in order to avoid iron induced toxicity. Using a human-based example, the disease associated with chronic iron overload
known as hemochromatosis causes rapidly progressing, multisystem disorders (32). Iron metabolism is therefore tightly regulated in host organisms through the use of iron scavenging elements such as transferrin and lactoferrin (6, 20), as well as through effects mediated by transcriptional regulatory proteins such as Fur (48), DtxR (117), or FecI (40), which bind to iron related elements on DNA known as Fur binding sites.

Enterohemorrhagic pathogens such as *E. coli* O157:H7, for which isolates from human patients have been shown to have stimulated growth in the presence of heme and hemoglobin; may induce hemolytic lesions in order to gain access to serum sources of nutrients such as heme and iron (83). It may therefore be in part as a result of attempting to acquire heme which elicits the clinical effects of pathogenic *E. coli* in HUS. Using a laboratory strain of *E. coli* (1017 (ent::Tn5)) that was unable to import heme into the cytoplasm, it was established that the outer membrane receptor ChuA (*E. coli* heme-utilization gene A) is required for heme uptake along with TonB, an energy-transducing protein associated with ExbB and ExbD that uses the proton motive force of the cytoplasmic membrane for the passage of ligands into the periplasm (146). Four other members of the heme uptake operon from *Y. enterocolitica* were shown to be required to complete heme uptake, transport, and use as an iron source by *E. coli* K-12 (127) including: the periplasmic heme-binding protein HemT, the heme permease protein HemU, the ATP-binding hydrophilic protein HemV, and the protein HemS, whose role may be to regulate the amount of non-protein associated heme in the cell (127, 128).

Restriction mapping and sequence analysis of selected regions of *E. coli* O157:H7 genome have revealed that the organization of the heme transport locus of this strain is strikingly similar to that of other enteric bacteria such as *Y. enterocolitica* and
S. dysenteriae (127, 128). This genetic organization is especially true with respect to the region encoding heme outer membrane receptors such as ChuA, ShuA, and HemR, followed by a short gap, and then the gene encoding ChuS or one of its homologues. In S. dysenteriae the intergenic region between ShuA and ShuS are separated by a 48-nucleotide region with sequence homology to the consensus Fur box, but no obvious –10 and –35 promoter elements (128). However, in response to iron limiting conditions, ShuS was upregulated, permitting heme as an iron source and preventing heme toxicity (165). A summary of the proteins encoded within the heme utilization operon from E. coli O157:H7 is summarized in Table 1.1.
Table 1.1 Summary of proteins encoded within the heme utilization operon from pathogenic Escherichai coli O157:H7.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Localization</th>
<th>Length (amino acids)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChuA</td>
<td>Outer Membrane</td>
<td>660</td>
<td>Heme Receptor</td>
</tr>
<tr>
<td>ChuT</td>
<td>Periplasm</td>
<td>304</td>
<td>Heme Transport</td>
</tr>
<tr>
<td>ChuU</td>
<td>Inner Membrane</td>
<td>330</td>
<td>Permease ABC Transport Protein</td>
</tr>
<tr>
<td>ChuV</td>
<td>Inner Membrane</td>
<td>256</td>
<td>Permease ABC Transport Protein</td>
</tr>
<tr>
<td>ChuS</td>
<td>Cytoplasm</td>
<td>342</td>
<td>Heme/hemoglobin transport protein</td>
</tr>
<tr>
<td>ChuW</td>
<td>Cytoplasm</td>
<td>445</td>
<td>Coproporphyrinogen III oxidase</td>
</tr>
<tr>
<td>ChuX</td>
<td>Cytoplasm</td>
<td>164</td>
<td>Hypothetical</td>
</tr>
<tr>
<td>ChuY</td>
<td>Cytoplasm</td>
<td>207</td>
<td>Hypothetical</td>
</tr>
</tbody>
</table>
1.3. Role of heme and structural diversity of heme-proteins

While traditional heme studies have focused on heme as a prosthetic group within hemoglobin, myoglobin, and iron processing elements, novel heme-Fcontaining proteins have recently been discovered. These include the cytochrome c chaperone CcmE (123), the iron-dependent regulator of heme biosynthesis Irr (107), and the heme-based aerotactic transducer Hem-AT (64). Other hemoproteins function through conformational changes induced as a result of modifications to the ferrous heme environment upon the coordination of small effector ligands such as NO, O₂, or CO, as in the case of guanylate cyclase (172), FixL (52), and CooA (81), respectively. Other newly identified heme-associated signaling proteins include AxPDEA1, NPAS2, and EcDos whose interaction between heme and various gases in the sensor domains mediate structural changes in the effector domains of these proteins (115). While the cellular consequences mediated by these signaling proteins are diverse, the common feature between them is that small molecules can trigger major changes to the overall heme-protein environment, which can distort the heme group itself or the interactions stabilizing the heme group, resulting in a robust structural and functional transition.

While heme helps to mediate diverse responses, it is the interaction of heme with various coordinating proteins that ultimately confer molecular function. Therefore, characterizing the interactions that contribute to heme stabilization will also help to characterize how heme mediates cell signaling. Furthermore, while heme helps to mediate diverse responses, it is the interaction of heme with various coordinating proteins that ultimately confer molecular function. Structural insights provided from vibrational
spectroscopy (EPR and Raman), nuclear magnetic resonance spectroscopy (NMR), and X-ray crystallography have indicated that heme proteins often coordinate the iron center of the heme group between a histidine side chain and a variety of other possible residues including tyrosine (8), methionine (22), cysteine (39), proline (81), histidine (103), arginine (137); while other proteins have been shown to coordinate the heme iron center without additional assistance from other side chains (52, 121). Other proximal side chains and regions of the polypeptide backbones have been shown to interact with other parts of the heme group, contributing to the stabilization of the heme moiety within the protein environment, and participating in the interactions which ultimately confers hemoprotein function. These heme-protein interactions are both hydrophobic and hydrophilic, stabilizing the porphyrin ring or propionic elements of the heme group, respectively, and are mediated through various interactions.

Given the diversity of functions attributed to heme proteins together with the various heme polypeptide interactions that have evolved, it is not surprising that heme proteins feature a broad range of structural motifs and overall architectures. The structure of sperm whale myoglobin, the first high resolution macro molecular structure solved by X-ray crystallography, consists of eight α-helices that wrap around the heme group, seating it within a hydrophobic pocket (72). The overall architecture of myoglobin is similar to that of HO-1 whose overall structure is also mainly α-helical, and sandwiches the heme group between two α-helices that are termed the proximal and distal helices (121). In the smaller hemoprotein cytochrome c, the heme group is associated within a simple, mainly α-helical fold (86) but resides closer to the surface than either myoglobin or HO-1. Heme proteins may also assume a variety of other folds. In the heme
sequestering protein hemopexin found in serum, heme is coordinated between two β-propeller domains (103), and in HasA the heme binding site is formed by loop regions which extend from an αβ-fold (8). The structure of FixL resembles an open pita or clam which coordinates heme in the center of the open mouth between a series of antiparallel β-sheets at the protein core and an α-helix (52). A unique example of heme coordination was recently reported for the iron regulated lipoprotein from C. jejuni, ChaN, whose function is poorly understood, but seems to associate with the outer membrane receptor ChuR (25). ChaN forms a homodimer which sandwiches a pair of heme molecules between a pair of conserved tyrosine residues donated from opposing monomers, that originate at the base of two α-helices. ChaN does not have a homologue in O157 or K-12, but homologues have been identified in other E. coli strains.
1.4. Bacterial heme oxygenases

Although many reports have focused on the proteins involved in heme transport across the outer membrane and periplasm, the fate of heme once it has reached the bacterial cytoplasm has only recently received attention. Investigations of eukaryotes revealed that release of iron stores requires the heme porphyrin ring to be degraded by monooxygenases known HOs. HOs are enzymatically unique, as they use heme as both a substrate and as a cofactor in the binding of oxygen for intramolecular degradation of the porphyrin macromolecule to BV, CO, and Fe$^{2+}$ (Figure 1.1B) (174). Structural insight contributed by EPR and X-ray crystallography studies suggests that heme cleavage catalyzed by HOs proceeds via similar mechanisms (73, 92). Initially, HO bound heme accepts an electron from either NADPH-cytochrome P450 reductase (CPR), yielding a ferrous (Fe$^{2+}$) heme-HO complex. Molecular oxygen binds with the complex between the iron atom and the axial histidine residue, forming a metastable oxy-form, that receives another electron from CPR and a proton from the distal water pocket, and is rapidly converted to a hydroperoxide intermediate (Fe$^{3+}$–OOH). The oxygen of the hydroperoxide intermediate attacks the $\alpha$-meso carbon (or other regioselective position), yielding a ferric $\alpha$-meso-hydroxyheme. This intermediate then receives another oxygen molecule and an electron and is converted to ferrous-verdoheme. Another oxygen and reducing equivalent are received, transforming ferrous-verdoheme to ferric-BV and is the least understood part of the mechanism (44). In the last step ferric-BV is further reduced to the ferrous state, yielding BV and Fe$^{2+}$. Overall the reaction requires an input of seven electrons and two molecules of oxygen (Figure 1.1) (73, 92).
Several bacterial and mammalian HOs have been recently identified, including HemO (173), HmuO (116), HmuQ/D (106), cyanobacterial HO-1 and HO-2 (30), IsdG/I (125), mammalian HO-1 and HO-2 (121), \textit{nm}-HO (122), PCC\_6803 (132), and PigA/BphO (46); and crystal structures have been determined for HemO (122), HmuO (61), \textit{nm}-HO (122), \textit{pa}-HO (46), PigA (46), and Syn HO-2 (132), as well as a partial structural characterization of BphO by NMR (157). Furthermore, extensive structural work has been conducted on various HOs in complex with effector molecules such as azide (135), biliverdin (134), CO and cyanide (136), ferric and ferrous forms of iron (79), imidazole, and nitrogen monoxide (79) in order to effectively capture “snapshots” along the reaction coordinate of heme degradation.

The overall topology of known bacterial HOs share strong structural similarity to mammalian HOs in that they are mainly $\alpha$-helical and sandwich heme between two helices at the base of the structures, with the propionic side chains oriented away from the protein (44). In this mode of heme coordination, HOs are regioselective for O$_2$ addition across the $\alpha$-meso position of heme, which is oriented toward a series of polar residues that do not participate directly in the reaction, but likely contribute to regioselectivity (121). The exception to this mode of coordination is \textit{pa}-HO from \textit{Pseudomonas aeruginosa} which is regioselective for the $\beta$- and $\delta$-meso positions of heme in a 30:70 ratio, which has been attributed to heme being rotated by 100° in the active site (46). The purpose of the generation of these two BV isomers is currently unknown, but is likely related to the fact that \textit{P. aeruginosa} has a second HO, BphO, with $\alpha$-meso carbon specificity (44). Despite the structural conservation across bacterial HOs identified to date, it should be noted that many of these were targeted because they shared sequence
homology with mammalian HOs. In pa-HO the propionic side chains of the heme moiety are stabilized in a unique fashion compared to other HOs identified. This highlights how a variety of protein-heme interactions can mediate heme catalysis beyond those well characterized in the hHO-1 related system. In many bacteria where sequence analysis fails to suggest the enzyme responsible for heme processing, other approaches must be utilized such as systematic gene knockouts and phenotypic characterization to identify the genes responsible for heme utilization (165). At the initiation of our investigation presented in this thesis, bioinformatics analysis had failed to identify a HO in any *E. coli* strain, although both *E. coli* O157:H7 and CFT073 were known to use heme as an iron source, suggesting that a heme degrading enzyme was necessary.

1.5. Heme utilization protein ChuS

Because of its poorly understood role in modulating heme utilization and its genetic organization downstream of the outer membrane receptor *chuA*, we directed our attention to the gene product of *chuS*, the *E. coli* O157:H7 and CFT073 homologue of *shuS*, as a candidate for intracellular heme breakdown (128). Sharing 98% identity to the *S. dysenteriae* homologue ShuS, the structure and function of these two homologues is likely highly similar. In the initial report characterizing ShuS in isolation of other members of the heme utilization operon, ShuS was shown to form a high molecular weight oligomer (650 kDa) from 37 kDa subunits, to nonspecifically bind double stranded DNA, and was not observed to exhibit any HO activity in standard NADPH or ascorbate based assays (158). In contrast, in an *S. dysenteriae* knockout, the *shuS* mutant
was defective in utilizing heme as an iron source and ShuS was required under microaerobic and anaerobic conditions for the optimal utilization of heme (165). ShuS expression also protected cells from heme toxicity at high concentrations, and was therefore suggested to bind to and potentially transfer heme from the transport proteins in the membrane to either heme containing or heme degrading proteins, or involve itself as a heme degrading enzyme (165). Furthermore, expression of the Y. enterocolitica homologue to ChuS, HemS, protected E. coli cells against heme toxicity (127).

1.6. ChuW, ChuX, ChuY and bacterial heme proteins

In addition to chuS, three other genes: chuW, chuX, and chuY are located within the heme utilization locus of O157 were poorly characterized at the initiation of our study. ChuW does not share any homology to any protein identified in heme transport, but does have weak homology to HemN, the oxygen-independent form of the heme biosynthetic enzyme co-protoporphyrinogen oxygenase III present in various bacteria including Bacillus ssp. Salmonella typhimurium and E. coli K-12 (164, 167). This homology suggests that ChuW may play a role in some aspect of heme transport or metabolism. In S. dysenteriae shuW begins 19 nucleotides downstream of shuT, and the lack of any apparent transcription termination or promoter sequences suggests that shuT and shuW are located on the same transcript (127, 128, 164, 165). In S. dysenteriae type 1, a premature stop codon exists within shuW suggesting that ShuW may be defective, although in E. coli O157:H7 this stop codon is absent in chuW suggesting that ChuW may be completely
functional (164). Furthermore, DNA microarray analysis of CFT073 strains suggested that \textit{chuW} was upregulated by more than 5-fold during urinary tract infection (126).

Twelve nucleotides downstream of \textit{shuW} is the start codon for \textit{shuX}, and the initiation condon of \textit{shuY} overlaps the stop codon of \textit{shuX}. This indicate that these two smaller genes may also be co-transcribed with \textit{shuT} (164). In a similar sequence analysis of the heme uptake locus from \textit{Y. pestis}, homologous to \textit{chuX} and \textit{chuY} (\textit{orfX} and \textit{orfY}, respectively) were determined to be located immediately downstream of the heme receptor gene \textit{hmuR} and were therefore likely co-expressed (145). In \textit{Vibrio anguillarum}, which can use heme and hemoglobin as a source of iron, using a \textit{lacZ} reporter system it was shown that a homologue of \textit{chuX}, \textit{huvX}, was co-transcribed with \textit{huvZ} under iron limiting conditions (95). An \textit{E. coli} homologue to \textit{huvZ} was not identified via NCBI Blastp sequence analysis at (www.ncbi.nlm.nih.gov/BLAST/). Furthermore, an examination of the -10 and -35 elements downstream of \textit{huvX} revealed sequence with similarity to sigma$^{70}$-promoters, as well as the presence of a Fur element (95). Characterization of either \textit{chuX} or \textit{chuY} gene products or their homologues from various heme utilizing organisms had not received any attention at the beginning of our investigation. However, based on Blastp sequence analysis ChuX was suggested to be a protein associated with heme-iron utilization, and that ChuY is a hypothetical protein or nucleoside-diphosphate-sugar epimerase (143). A schematic-diagram of the heme assimilation system of Gram-negative bacteria, as well as the structures and functions of members of the heme utilization operon at the initiation of our investigation is summarized in Figure 1.2.
Figure 1.2 Schematic-diagram of the heme assimilation system from Gram-negative bacteria at the initiation of our investigation and organization of the heme utilization operon from *Escherichia coli* O157:H7. Proteins involved in heme transport from pathogenic *E. coli* O157:H7 are presented in their locations in the outer membrane (OM), periplasm (PP), and inner membrane (IM). Where the structures of proteins are unknown, structures of homologues have been included and names followed by a question mark. Putative promoter elements within the heme utilization operon are depicted with arrows. The function of ChuS, ChuW, ChuX, and ChuY were unknown at that time. The proteins focused on as part of this body of work are ChuX and ChuS.
In light of the general importance of heme as a signaling molecule and cofactor, this thesis examines the structure and function of two members of the *E. coli* O157:H7 heme utilization operon: the utilization protein ChuS in its apo and heme bound forms, and the heme binding protein ChuX. The method of X-ray crystallography, specifically, the anomalous dispersion method using seleno-methionine labeled proteins has been used for structural determination. A combination of spectral characterization, site-directed mutagenesis of conserved residues, gas chromatography, and sensitivity of cultures challenged with heme provide biochemical evidence for the function of ChuS and ChuX. The application of structure based functional characterization as well as the implications of our findings for heme uptake and utilization in *E. coli* O157:H7 and Gram-negative bacteria in general is addressed.
Chapter 2

Identification of a novel *Escherichia coli* O157:H7 heme oxygenase with tandem functional repeats

Preface:


Michael Suits was responsible for protein expression, purification, crystallization, data collection, structure solution and refinement, and all biochemical experiments in the investigation of function. Dr. B. McLaughlin helped with gas chromatography and interpretation, Dr. B. Hill conducted EPR experimentation and interpretation, and expert help from Dr. Mike Nesheim in heme reconstitution setup and interpretation of results to calculate the $K_d$ for ChuS-heme. The *chuS* gene was cloned by Pietro Iannuzzi and the N- and C-terminal halves of ChuS by Jim Blonde. The manuscript was written by Michael Suits with editorial input from Dr. Zongchao Jia and with additional help from Dr. John S. Elce.

Data deposition footnote: Coordinates and structure factors for ChuS are deposited in the RSCB Protein Data Bank under accession code 1U9T.
2.1. Abstract

Heme oxygenases (HOs) catalyze the oxidation of heme to biliverdin, carbon monoxide (CO), and free iron. Iron acquisition is critical for invading microorganisms to enable survival and growth. Here we report the crystal structure of ChuS, which displays a novel fold and is unique compared to other HOs characterized to date. Despite only 19% sequence identity between the N- and C-terminal halves, these segments of ChuS represent a structural duplication, with a root mean square deviation of 2.1 Å between the two repeats. ChuS is capable of using either ascorbic acid or cytochrome P450 reductase-NADPH as electron sources for heme oxygenation. CO detection confirmed that ChuS is a HO, the first to be identified in any strain of *Escherichia coli*. Based on sequence analysis, this novel HO is present in many bacteria, though not in the *E. coli* K-12 strain. The N- and C-terminal halves of ChuS are each a functional HO.
2.2. Introduction

The incorporation of iron into proteins as a biocatalyst or electron carrier enables organisms to fulfill many vital biological processes including photosynthesis, \( \text{N}_2 \) fixation, methanogenesis, \( \text{H}_2 \) production and consumption, respiration, the tricarboxylic acid cycle, oxygen transport, gene regulation, and DNA biosynthesis (89). The importance of iron for bacterial survival and pathogenesis is evident by the variety of iron-acquisition and transport systems that have evolved, such as the secretion of low-molecular-weight (MW) siderophores capable of binding free iron and transporting it into the cell via specific membrane receptors (33). As part of a coordinated immune response against invading microorganisms, mammals specifically limit iron availability by producing the iron-binding proteins transferrin and lactoferrin, which depress free extracellular iron to levels insufficient to sustain bacterial growth (3). Alternatively, invading pathogens may acquire iron directly from host iron sources such as heme, hemoglobin, or hemopexin either through the production of specific outer membrane receptors that bind directly to host heme-sequestering proteins or through the secretion of heme-binding proteins (hemophores) that then deliver the heme-protein complex to bacterial cell surface receptors (48). Enterohemorrhagic pathogens such as *Escherichia coli* O157:H7, for which isolates from human patients have been shown to have stimulated growth in the presence of heme and hemoglobin, may induce hemolytic lesions in order to gain access to these serum sources of iron (83).

Using a laboratory strain of *E. coli* (1017 (ent::Tn5)) that was unable to import heme into the cytoplasm, it was established that the outer membrane receptor ChuA
(E. coli heme-utilization gene A) is required for heme uptake along with TonB, an energy-transducing protein associated with ExbB and ExbD that uses the proton motive force of the cytoplasmic membrane for the passage of ligands into the periplasm (146). Four other members of the heme uptake operon from Yersinia enterocolitica are required to complete heme uptake, transport, and use as an iron source by E. coli K-12 (127) including: the periplasmic heme-binding protein HemT, the heme permease protein HemU, the ATP-binding hydrophilic protein HemV, and the protein HemS, whose role may be to regulate the amount of non-protein-associated heme in the cell (128).

Restriction mapping and sequence analysis of selected regions of E. coli O157:H7 DNA, the serotype responsible for outbreaks of hemorrhagic colitis and hemolytic uremic syndrome, has revealed that the organization of the heme transport locus of this strain is strikingly similar to that of other enteric bacteria such as Y. enterocolitica (164). This organization is especially true with respect to the region encoding heme outer membrane receptors such as ChuA, which are followed by a short gap and then by ChuS or one of its homologues. In Shigella dysenteriae the intergenic region between ShuA and ShuS are separated by a 48-nucleotide region with sequence homology to the consensus Fur box but no obvious –10 and –35 promoter elements (127). Furthermore, a promoter was likely to be present within the intergenic region because a mini=Tn10 insertion in ShuA was not polar with respect to ShuS expression in minicells (91).

Although many reports have focused on the proteins involved in heme transport across the outer membrane and periplasm, the fate of heme once it has reached the bacterial cytoplasm has only recently received attention. Investigations in eukaryotes revealed that release of iron stores requires the heme porphyrin ring to be degraded by
monooxygenases known as heme oxygenases (HOs). HOs are enzymatically unique, as they use heme as both a substrate and a cofactor in the binding of oxygen for intramolecular degradation of the porphyrin macromolecule to biliverdin, CO, and free iron (121). Several bacterial HOs have been identified, including HemO (173), HmuO (26), IsdG/I (125), cyanobacterial HO-1 and HO-2 (132), and PigA/BphO (110). All of these were reviewed (44), and the structures of HemO (122), HmuO (61), and PigA (46) have been determined, as well as a partial structural characterization for BphO (157). The crystal structures of HemO, HmuO, and PigA all share a strong structural similarity to mammalian HOs: they are mainly α-helical and sandwich heme between two helices. At present, no sequence comparison or structural analysis has identified a HO in an *E. coli* strain.

Here we report the crystal structure of ChuS, which has a novel fold with two tandem repeats and does not show structural similarity to known mammalian or bacterial HOs. ChuS is capable of binding heme and can break down heme using ascorbic acid or cytochrome P450 reductase-NADPH (CPR) as electron sources for oxygenation. CO detection by gas chromatography confirmed that ChuS is a HO, the first to be identified in *E. coli*. In addition, HO activity occurs independently in the N- and C-terminal halves of ChuS.
2.3. Materials and methods

2.3.1 Sequence analysis, protein expression, and purification

Sequence analysis was carried out using Blast (5). Full-length ChuS and the N- and C-terminal halves of ChuS were subcloned into pET expression vectors, yielding fusion proteins to either a N- or C-terminal His tag or an N-terminal GST tag. Based on the results of small-scale expression trials, we selected the following constructs: ChuS fused to an N-terminal His$_6$ tag (ChuS), the N-terminal half fused to a C-terminal His$_6$ tag (N-ChuS), and the C-terminal half fused to an N-terminal GST tag (C-ChuS). In each case, 1-L cultures of BL21(DE3) cells carrying plasmid for the respective recombinant protein were grown at 37°C in Terrific Broth (BioShop, Burlington, Canada) supplemented with 100 µg/ml ampicillin. Protein expression was induced using 0.4 mM IPTG for 5 h. ChuS protein substituted with seleno-methionine was expressed in the metA$^-$ E. coli strain DL41 in LeMaster medium (60).

All proteins were purified using standard methods. Briefly, His$_6$-tagged proteins were purified via Ni-NTA batch purification in phosphate buffer (pH 8.0), followed by purification by Resource Q and Hi-Trap metal-chelating columns on an ATKA Explorer FPLC. Collected fractions were monitored by SDS-PAGE, for which those that contained pure protein were pooled together and dialyzed against 50 mM Tris-HCl buffer (pH 8.0), 150 mM NaCl for subsequent crystallization, or 100 mM sodium phosphate buffer (pH 7.0) for HO spectral and activity measurements. C-ChuS was purified on a GSTrap FF column.
2.3.2 Crystallization of full-length ChuS

Crystals were obtained by the hanging-drop vapor diffusion method by mixing 2.5 µl of 20 mg/ml ChuS with 2.5 µl of reservoir solution consisting of 20% (w/v) PEG 3350 and 100 mM Bis-Tris (pH 6.5). Crystals appeared after 1-3 d. ChuS crystals were further improved by microseeding using 12–20% (w/v) PEG 3350. SeMet derivative crystals were obtained in the same way. For cryoprotection, crystals were soaked in the reservoir solution containing ethylene glycol (concentration increased stepwise to 30% [v/v]), picked up using a nylon loop, and flash-cooled in the N\textsubscript{2} cold stream at 100K. Two nonisomorphous crystal forms belonging to space group \textit{P}2\textsubscript{1} were obtained. Crystals having the larger unit cell contain three molecules in the asymmetric unit, whereas those with the smaller unit cell contain only one molecule.

2.2.3 Data collection and structure determination

Native and multiple-wavelength anomalous dispersion (MAD) data sets were collected, respectively, for the two different crystal forms at beamlines 17-ID and 19-BM of the Advanced Photon Source, Argonne National Laboratory. The native data were collected for 180° in total, with 1.0° oscillation. SeMet MAD data were collected at Se peak and inflection wavelengths for 360° and remote wavelength for 180°, all with 1.0° oscillation. Data were processed with HKL2000 (102). Initial phases for the three-molecule form was determined using the MAD data with the program SOLVE (144). Density modification, including three-fold NCS averaging, was carried out using
RESOLVE (144). Automatic chain tracing was performed using RESOLVE (144), and the remainder of the model was built by manual fitting using XtalView/Xfit (90). When the model of one of the three molecules was about 80% built, it was used as a search model in molecular replacement using data obtained from the single-molecule form. This resulted in an unambiguous solution that was subsequently refined using CNS (19).

2.3.4 Reconstitution with heme and absorbance measurements

Heme complexes of ChuS, N-ChuS, and C-ChuS were prepared by dissolving heme into 0.5% (v/v) ethanolamine and adding small volumes of this mixture to protein until a final ratio of 4:1 (mol:mol) was obtained for ChuS and 2:1 for N-ChuS and C-ChuS. ChuS-heme and N-ChuS-heme were then applied to Resource Q or Superdex 200 Prep Grade columns, respectively, to remove noncoordinated heme. Reconstituted protein-heme complex concentrations were determined by adapting the reported extinction coefficient for ShuS (ε_{410}) of 79.5 and 159 mM^{-1}cm^{-1} for ChuS and N-ChuS, respectively (11). Pure fractions of C-ChuS were concentrated and used for spectrophotometric analysis and CO detection in which twice the amount of heme was added to protein prior to measurements. Absorbance data were then collected using a microplate spectrophotometer (Bio-Tek Instruments, Inc.). As a control in separate experiments, 10 μM catalase and 10 μM superoxide dismutase were added to the reaction wells, followed by ascorbic acid or cytochrome P450 reductase (CPR)-NADPH, and the difference spectra were recorded. A quadruplicate series of ChuS-heme reconstitution titrations were also conducted to examine the ChuS-heme interaction.
2.3.5 CO activity assay and inhibition

Briefly, in 2-ml vials 250 µl reaction volumes of 20 µM heme and 10 µM ChuS, N-ChuS, or C-ChuS in 100 mM phosphate buffer (pH 7.0) were incubated with constant shaking at room temperature for 6 min at which time enzymatic heme degradation was initiated by adding either CPR and NADPH or ascorbic acid at final concentrations of 17.7 nM and 135 µM, or 50 µM, respectively. Vials were then sealed with screw caps and blue silicone rubber septa and the headspace above each was immediately purged with CO-free air, and incubated for 45 min at 24°C with constant shaking. The reaction was stopped by placing the vials on powdered dry ice (-78°C), where they remained for approximately 30 min. The amount of liberated CO in the headspace of each vial was measured using a gas chromatograph equipped with an HgO reduction detector (Trace Analytical). The amount of CO resulting from ChuS-catalyzed breakdown of heme was determined by comparing peak area measurements for CO against linear CO standard curves (n = 20 determinations; average correlation coefficient 0.993). Technical details are described elsewhere (27, 153). In a parallel experiment, Sn(IV) mesoporphyrin IX dissolved in 0.5% (v/v) ethanolamine was also added during initial mixing which acted as a competitive inhibitor.

2.3.6 Dynamic light scattering of C-ChuS and N-ChuS
Dynamic light scattering (DLS) was conducted for ChuS, ChuS-heme, and N-ChuS each at 1 mg/ml concentration in 100 mM phosphate buffer (pH 7.0) using a DynaPro99 instrument (Protein Solutions). DLS was also conducted on samples of 1 mg/ml ChuS and ChuS-heme in 50 mM Tris and 150 mM NaCl (pH 8.0).

2.3.7 EPR experiments

EPR spectra were obtained on a Bruker EMX spectrometer at X-band fitted with an HSQ cavity equipped with a liquid helium flow cryostat (Oxford Instruments). The standard instrument conditions were $T = 10K$, modulation amplitude $= 20$ G, power $= 2$ mW at a gain of $1 \times 10^4$.

2.4 Results

2.4.1 Expression and purification

ChuS overexpression resulted in cells exhibiting a blue-green pigment characteristic of biliverdin accumulation (87, 157, 174). Following Ni-NTA purification and dialysis in low-ionic-strength Tris buffer, fractions containing ChuS appeared purple and turned to a straw color over time. EPR of ChuS fractions following Ni-NTA purification did not reveal the presence of any coordinated metal ion, and spectral data did not reveal the source of the pigmentation. This pigmentation was not observed in ChuS grown in LeMaster medium in DL41(DE3) cells or for the N-ChuS or C-ChuS
constructs. The purple coloration of ChuS was lost following the subsequent FPLC anion exchange step. Expression of full-length ChuS and N-ChuS yielded more than 60 mg of pure protein from 1 L culture. Recombinant C-ChuS was insoluble in pilot expression trials for N- or C-terminal His-tagged versions, and could only be expressed in soluble form as a GST-C-ChuS construct. Average yields of 2.6 mg pure GST-C-ChuS/ L culture were obtained.

2.4.2 Structure determination and analysis

Two different crystal forms were obtained from the same crystallization conditions, that shared the same overall morphology and space group (P2₁), but differed in unit cell dimensions and number of molecules in the asymmetric unit. We collected MAD data for the three-molecule crystal form and native data for the single-molecule crystal form, and the crystal structure of ChuS was solved using a combination of MAD and molecular replacement with the two crystal forms. The higher resolution (2.15 Å) structure from the single-molecule form is presented here, although the final structures for ChuS from the two crystal forms are essentially identical, with an average root mean square deviation (rsmd) of 0.92 Å. The final structure of ChuS contains residues 1–337 of a total 342 amino acids with no ligand. Due to an absence of electron density, a gap exists in the structure between residues 12 and 23 (12-QNPGKYARDI-23), and between 165 and 173 (165-KAVDAPV-173). In total, 99.7% of residues are within the allowed region of the Ramachandran plot. Only one residue, Asp 114, was in the disallowed region and
resides within a surface exposed β-turn. Final refinement statistics are summarized in Table 1.

The overall structure of ChuS comprises a central core of two large pleated β-sheets, each consisting of nine anti-parallel β-strands sandwiched together and bowing outward in a saddle motif (Figure 2.1). Each β-sheet is flanked at its N-terminus by one pair of parallel α-helices and at its C-terminus by a set of three α-helices in an α-loop-α-loop-α configuration. Two large clefts are found on opposite sides of the central core of β-pleated sheets, with the third side of the clefts delineated by the flanking sets of three α-helices. A long stretch of coil connecting the N- and C-terminal halves runs from I156 to V183, in which the gap in density occurs (Figure 2.1). Structural comparison of ChuS against structures within the SCOP database (97) was performed with the program SSM (77). This search did not reveal any significant match; hence ChuS represents a novel fold.
Table 2.1 Diffraction data and refinement statistics for ChuS

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>Peak</th>
<th>Inflection</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>P2₁</td>
<td></td>
<td>P2₁</td>
<td></td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>41.683</td>
<td></td>
<td>46.072</td>
<td></td>
</tr>
<tr>
<td>b (Å)</td>
<td>57.878</td>
<td></td>
<td>194.172</td>
<td></td>
</tr>
<tr>
<td>c (Å)</td>
<td>59.575</td>
<td></td>
<td>60.461</td>
<td></td>
</tr>
<tr>
<td>β (°)</td>
<td>96.020</td>
<td></td>
<td>100.383</td>
<td></td>
</tr>
<tr>
<td><strong>No. of molecules in ASU</strong></td>
<td>1</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.9795</td>
<td>0.9803</td>
<td>0.9804</td>
<td>0.9642</td>
</tr>
<tr>
<td><strong>Resolution range (Å)</strong></td>
<td>50–2.15</td>
<td>50–2.60</td>
<td>50–2.60</td>
<td>50–2.40</td>
</tr>
<tr>
<td><strong>Total reflections</strong></td>
<td>53,596ᵃ</td>
<td>170,485ᵇ</td>
<td>176,691ᵇ</td>
<td>149,352ᵇ</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>14,606ᵃ</td>
<td>32,694ᵇ</td>
<td>32,966ᵇ</td>
<td>40,453ᵇ</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>96.9 (82.8)ᵇ</td>
<td>99.8 (99.6)ᵇ</td>
<td>99.7 (98.4)ᵇ</td>
<td>96.3 (76.5)ᵇ</td>
</tr>
<tr>
<td><strong>R_{sym}(I) (%)</strong></td>
<td>7</td>
<td>9.7</td>
<td>9.1</td>
<td>7.8</td>
</tr>
<tr>
<td><strong>I/σI</strong></td>
<td>15.0 (2.4)ᵇ</td>
<td>17.4 (2.4)ᵇ</td>
<td>16.6 (2.0)ᵇ</td>
<td>11.4 (1.0)ᵇ</td>
</tr>
</tbody>
</table>

**Refinement (single molecule form)**

|                          |        |      |            |        |
| **Resolution range (Å)** | 50 – 2.15 |      |            |        |
| **R_{work}, (%)**       | 20.1   |      |            |        |
| **R_{free}, (%)**       | 26.3   |      |            |        |
| **No. reflections total/ R_{free}** | 12315/ 744 |      |            |        |
| **No. atoms, protein / solvent (H₂O)/ total** | 2428/ 331/ 2759 |      |            |        |
| **B factors (Å²) protein/ solvent (H₂O)/ total** | 29.7/ 48.6/ 32.0 |      |            |        |
| **RMS, bond length (Å)/ bond angle(°)** | 0.009/ 1.44 |      |            |        |

ᵃValues in parentheses are for the outermost shell (2.23–2.15 Å)
ᵇNumber of reflections following merging
ᵇNumber of reflections without merging of Bijovet pairs
Figure 2.1 Ribbon diagram of ChuS. The core of nine anti-parallel β-strands is visible in the front face. The gaps in the structure are shown with a dashed lines.
2.4.3 ChuS contains a structural duplication

On close inspection, the two halves of ChuS were found to be structurally similar. Including side chains the two halves superimpose to an rmsd of 2.1 Å (Figure 2.2). Sequence alignment of the two halves reveals only 19% identity. Nevertheless, at the structure based level, many residues are well conserved structurally, including R29 and R209, F127 and F304, and Y138 and Y315. The two halves of ChuS associate across the central portion of the β-pleated sheets. At the interface, residues from the C-terminal half are hydrophobic (V244, V250, F274, L276), point toward the interior of the C-terminal subunit, and expose regions of the backbone across the interface. In contrast, most interface residues from the N-terminal half are hydrophilic (N66, Y70, H73, N75, R95) and point across the interface between the two halves. The side chains of the N-terminal interacting residues are in contact with the backbone region of the C-terminal half. This structural similarity suggests that the two halves of ChuS may be functionally independent or that they have dual functions. Furthermore, the second gap in the structure is exactly at the midpoint of the ChuS sequence in a flexible/disordered linker connecting the two halves.
Figure 2.2 The two halves of ChuS superimposed, with an overall rSMD of 2.1 Å. The N-terminal portion of ChuS is shown in red, the C-terminal portion in blue.
2.4.4 Spectral properties of the heme-ChuS complex

Based on difference spectra comparing free heme to full-length ChuS, N-ChuS, and C-ChuS, we determined that all three formed complexes when reconstituted with heme (Figures 2.3A and 2.3B). The spectra of ChuS resemble those of other HOs in that a Soret maximum is evident at 408 nm, with a smaller set of peaks for the β-band at 545 nm and for the α-band at 580 nm, suggesting that the ChuS-heme complex formed is ferric hexacoordinate in the high-spin state at neutral pH (26). The spectrum for N-ChuS has a peak at 402 nm (which is smaller than for ChuS), a smaller peak at 530 nm, and a shoulder at 625 nm. Although the spectrum for N-ChuS with heme did not show definitive heme coordination, it remained unchanged following gel filtration, indicating a specific interaction between the two. C-ChuS with heme showed a broad peak centered at 384 nm, a smaller peak at 540 nm, and two small shoulders at 585 and 625 nm. Furthermore, EPR spectra analysis suggested the presence of two populations of heme molecules (Figure 2.4).

Despite structural similarity between the two halves, the spectra for C-ChuS and N-ChuS differed, perhaps indicating different environments for heme coordination. Based on absorbance at 408 nm during ChuS-heme reconstitution and assuming two binding sites, the estimated $K_d$ is $1.0 \pm 0.3 \ \mu M (n = 3)$ based on the following expression:

$$\Delta A = (\varepsilon_b - \varepsilon_f)0.5\{2sites \cdot P_0 + K_d + H_0 - ((2sites \cdot P_0 + H_0 + K_d)^2 - 4 \cdot 2sitesP_0H_0)^{1/2}\},$$
where $P_0$ is the amount of protein, $H_0$ is the concentration of heme, and $(\varepsilon_b - \varepsilon_f)$ is the difference in the Soret absorption spectra between the bound and free states (Figure 2.5). This estimated $K_d$ suggests a slightly weaker association than that characterized for the mammalian hHO-1 ($0.84 \pm 0.2$ µM) (159, 160) but tighter than that for other bacterial heme oxygenases: HmuO ($2.5 \pm 1.0$ µM) (162), IsdG ($5.0 \pm 1.5$ µM) and IsdI ($3.5 \pm 1.4$ µM) (125).
Figure 2.3 Spectral analysis of ChuS reconstituted with heme. Full-length ChuS and the N- and C-terminal halves of ChuS all form complexes when reconstituted with heme, with peaks at 408, 402, and 384 nm, respectively (back lines). A) Following addition of ascorbic acid as an electron donor, ChuS and its two halves were all capable of heme oxygenation, as evident from the spectral shifts (gray lines). Absorbance values corresponding to wavelengths greater than 500 nm have been multiplied by 10 to amplify absorbance signals corresponding to the β- and α-bands. B) Similar to A), but NADPH and the P450 reductase system were used as the electron source.
Figure 2.4 EPR analysis of ChuS reconstituted with heme. Based on spectra comparison with myoglobin reconstituted with heme, ChuS-heme exists in two spin states.
Figure 2.5 ChuS reconstituted with heme. Based on absorbance at 408 nm during ChuS-heme reconstitution and assuming two binding sites, the estimated $K_d$ is $1.0 \pm 0.3 \text{ µM (} n = 3 \text{).}$
2.4.5 Heme oxygenation by ChuS

Bacterial and mammalian HOs use the CPR-NADPH system, ferredoxin, flavodoxin, or ascorbic acid as reducing partners in vitro (160). To effectively monitor and quantify the HO activity of ChuS, 50 µM ascorbic acid was used. This concentration is between 100- (21) and 350-fold (26) less than that used for characterizing other HOs. The use of this lower concentration of ascorbic acid demonstrates highly efficient catalysis of ChuS with this electron donor. Furthermore, by using small amounts of ascorbic acid, heme degradation through coupled oxidation was minimized (34). To examine the heme degrading activity of ChuS, we monitored the reaction by UV-visible spectroscopy after the addition of ascorbic acid (Figure 2.6). As the reaction progressed, the Soret peak at 408 nm decreased, the small broad peak at 560 nm increased, and the absorption at 680 nm initially increased then decreased. Because the addition of catalase and superoxide dismutase did not affect the trends of the spectral change, this effect can be attributed to specific heme degradation by ChuS and its variants. These spectral shifts are similar to those observed for other HOs, which suggests the formation of biliverdin and free iron rather than the Fe$^{3+}$-biliverdin complex formed by HO-1. Furthermore, similar spectral shifts were also observed for heme reconstituted both to N-ChuS and C-ChuS, indicating that both halves of ChuS are capable of degrading heme independently (Figure 2.3A). Similar endpoint spectra were also observed for ChuS, N-ChuS, and C-ChuS when CPR was used as the electron donor (Figure 2.3B).
Figure 2.6 Spectral change of ChuS reconstituted with heme, monitored over time following ascorbic acid addition. Absorbance measurements between 500 and 700 nm have been multiplied by five to amplify the change in the β- and α-bands. Arrows follow changes in absorption at 408, 545, and 680 nm during reaction progression. The spectral change corresponds with the formation of the product, biliverdin.
2.4.6 Detection of CO release

All three ChuS constructs were capable of releasing CO (Figure 2.7). When ascorbic acid was used as the electron source, C-ChuS was 1.5 times more effective at CO production than ChuS and 10 times more effective than N-ChuS, whose release was only slightly above the background level of heme breakdown. When both N-ChuS and C-ChuS were added to the same reaction vessel, the amount of CO evolved was greater than the levels observed for ChuS alone, but not as great as those observed for C-ChuS. When CPR was used as the electron source, the amount of CO produced by ChuS and C-ChuS was five times less than when ascorbic acid was used. In contrast, N-ChuS produced more than four times as much CO as when ascorbic acid was used and was slightly more effective at producing CO than either the C-ChuS or full-length ChuS.
Figure 2.7 Compiled average CO detected for ChuS, N-ChuS, and C-ChuS. Electron donors were either ascorbic acid (red) or cytochrome P450 reductase-NADPH (blue) as an electron source for heme oxygenation. Error bars represent standard errors.
2.4.7 Organization of the heme uptake operon

Given the similarity between the heme uptake operon of *Y. enterocolitica* and *E. coli* O157:H7, it is likely that homologues of HemR, HemT, HemU, HemV, and HemS would be required for heme assimilation and use in *E. coli*. Sequence comparison between the heme uptake loci of these organisms revealed the homologous genes ChuA, ChuT, ChuU, various putative ATP-binding components of the heme transport system, and ChuS. Sequence identities ranged between 38% and 68%, and similarities were between 58% and 80%. The highest identity was conserved between HemS and ChuS, suggesting conservation of function in iron acquisition and prevention of heme toxicity. The lower identity with the other proteins may reflect differences between the periplasmic and membrane environments of the two organisms. Furthermore, sequence comparison revealed the following ChuS orthologs with strong sequence similarity: ChuS (CFT073), ShuS (*Shigella*), EhuS (*Enterobacter, Erwinia*), HemS (*Yersinia*), plu2633 (*Photorhabdus*), HmuS (*Rhizobium, Sinorhizobium*), and BhuS (*Bordetella*), suggesting that this HO is present in many other bacteria (Figure 2.8). Furthermore, conserved histidine residues at positions 73, 87, 193 and 277 may be important for heme coordination.

2.4.8 Dynamic light scattering

The derived molecular radii of N-ChuS, ChuS, and ChuS-heme in phosphate buffer and ChuS and ChuS-heme in Tris buffers were all approximately 2.90 nm,
corresponding to a MW of 41 kDa for an elongated protein. This is in agreement with that expected for an N-ChuS dimer or ChuS monomer. The longest dimension of ChuS from the crystal structure corresponds to a radius of 2.88 nm, indicating that the N-ChuS homodimer is similar in shape to full-length ChuS. The monomeric organization of ChuS was confirmed using a gel filtration column calibrated with a set of MW standards. This finding is contradictory to a previous report showing that a homologue of ChuS, ShuS (98% identity), forms an oligomer with an estimated MW of 650 kDa (11). When ChuS or ChuS reconstituted with heme were examined in the Tris buffer system following 4 d of incubation at 4°C, DLS results suggested the formation of various high-MW aggregates without any detection of a ChuS monomer. In contrast, the protein remained monomeric after 1 week in phosphate buffer.

2.4.9 Inhibition by Sn(IV) mesoporphyrin

In the presence of Sn(IV) mesoporphyrin IX, a competitive inhibitor for heme binding (152), a dose-dependent decrease in HO activity was observed from 2- to 7-fold. This level of inhibition is similar to that characterized for other HOs in the presence of competitive inhibitors (152) (Figure 2.9).
Figure 2.8 Sequence alignment of ChuS with homologues from various Gram-negative bacteria. Highly conserved residues are highlighted in red, residues with sequence similarity are presented in blue. Numbering corresponds to ChuS. Conserved histidines across the organisms examined at positions 73, 87, 193 and 277 may be important for heme coordination by ChuS.
**Figure 2.9 Inhibition of ChuS heme degradation by Sn(IV) mesoporphyrin IX.** ChuS activity was inhibited in a dose dependent manner and was similar to that observed for other heme oxygenases. Error bars represent standard error.
2.5. Discussion

ChuS is the first HO identified in an *E. coli* strain. For pathogenic organisms that cause hemolytic lesions, such as the *E. coli* O157:H7 and CFT073 strains, heme degradation by oxygenation may be a significant mechanism for iron acquisition and/or prevention of heme toxicity. Monitoring the spectral changes of ChuS reconstituted with heme when ascorbic acid or NADPH-CPR were used as electron donors provided direct evidence that ChuS catalyzes the breakdown of heme to biliverdin, CO, and free iron. *In vivo*, *E. coli* may use reductants such as ferredoxins and flavodoxin as electron sources for heme oxygenation (157). The liberation of CO and inhibition of CO release by addition of the prototypical HO inhibitor Sn(IV) mesoporphyrin IX support ChuS as a HO. While the spectra of N-ChuS and C-ChuS do not definitively suggest HO activity, unequivocal detection of CO release indicates that both halves of ChuS possess the ability to break down heme. Based on the significant sequence similarity of proteins between *Enterobacter*, *Erwinia*, *Shigella*, *Yersinia*, to *E. coli* ChuS, we can extend our identification of ChuS as an HO to these other bacterial genera. Further work will characterize heme breakdown products resulting from the action of N-ChuS and C-ChuS and the underlying mechanisms.

Although the overall structure of all bacterial and mammalian HOs characterized to date share the same mainly α-helical fold, the structure of ChuS is unique in that it is comprised of two central sets of anti-parallel β-sheets, each flanked by two pairs of α-helices. The absorption spectra of the two halves of ChuS following reconstitution with heme are notably different from each other and the full-length protein. This is especially
true for the C-terminal half, whose broad peak centered at 384 nm is unique compared to the Soret range of 402 to 412 nm observed for other HOs identified to date (44).

The structure of ChuS consists of a structural duplication, despite the fact that the two halves share only 19% sequence identity. This unique structural feature may be explained in part by the expression profiles of recombinant full-length ChuS, N-ChuS, and C-ChuS. Recombinant forms of full-length ChuS and N-ChuS were highly expressed, whereas the expression of C-ChuS was at least 30 times lower and required a large GST tag to increase protein solubility. Considering its structure, this decreased solubility of C-ChuS is not surprising. In contrast to the N-terminal half, the interface of the C-terminal half consists of hydrophobic residues (V244, V250, F274, L276), exposure of which would certainly lead to poor solubility. CO release experiments provide further insights as to why an enzyme having a structural repeat should exist when the N-terminal half alone would suffice. In the presence of ascorbic acid as an electron donor, C-ChuS produced over 18 times more CO than the weakly acting N-ChuS. When CPR was used as the electron source, N-ChuS was almost twice as reactive as C-ChuS. Together, the two halves of ChuS would enable pathogenic E. coli to use different electron donors while stabilizing the less-soluble C-terminal half of the molecule. This, in turn, would increase the chance of survival of E. coli strains in the host, thereby maximizing their virulence.

In addition to having an essential role in the acquisition of iron from host sources, ChuS may also protect E. coli against heme toxicity. There is growing evidence that heme accumulation can result in cell damage and tissue injury, as heme catalyzes the formation of reactive oxygen species, resulting in oxidative stress (9, 12). Furthermore,
because heme has a low MW and is lipophilic, it can easily intercalate into the membrane and impair lipid bilayers and organelles, such as mitochondria and nuclei, and destabilize the cytoskeleton (113, 149). Therefore, the HO activity of ChuS may be required to convert heme to biliverdin, a precursor to the less-reactive species bilirubin. This role for ChuS was also suggested through the appearance of a blue-green pigment during ChuS expression, characteristic of biliverdin accumulation. The ORF YhhX has been identified in *E. coli* K12, O157:H7, and CFT073. YhhX is downstream of two Fur binding sites (149), shares 19% sequence identity to human BVR chain A, and may be the oxido-reductase that converts biliverdin to bilirubin. The structure of a homologue to YhhX (41% identity) has recently been deposited in the PDB, and publication of its structural and functional analysis could confirm the identity of the second enzyme in the heme degradation pathway. Finally, our results from DLS demonstrated that ChuS and ChuS-heme both form high-molecular-weight aggregates after 4 d of incubation in Tris buffer at 4°C. This could be the cause of the absence of HO activity previously reported for ShuS, a homologue of ChuS (11).

In summary, we determined the structure of ChuS, which represents a novel fold and is different from that of known mammalian and bacterial HOs. The structure contains tandem structural repeats, which are both functional in terms of heme oxygenation activity, albeit with somewhat different properties. Through heme spectral analysis and CO quantification, we identified ChuS as an HO, the first to be identified in *E. coli*. 
2.6. Acknowledgments

We thank Dr. Gerald Marks for his support, Dr. B. McLaughlin for help with gas chromatography, Dr. B. Hill for EPR experimentation and interpretation, other members of the Jia lab (especially M. Adams) for help with synchrotron data collection, and Jim Blonde for help with cloning. The expert help of Dr. Mike Nesheim is greatly appreciated. We are also grateful to the Advanced Photon Source beamlines 17-ID and 19-BM, where the diffraction data were collected. This work was supported by the CIHR. MS was supported by an E.G. Bauman Fellowship, and ZJ is a Canada Research Chair in Structural Biology and an NSERC Steacie Fellow.
Chapter 3

Structure of the *Escherichia coli* O157:H7 heme oxygenase ChuS in complex with heme and enzymatic inactivation by mutation of the heme coordinating residue His-193

Preface:


Michael Suits was responsible for protein expression, purification, crystallization, data collection, structure solution and refinement, and all biochemical experiments in the investigation of function. The *chuS* gene was cloned by Pietro Iannuzzi and Neilin Jaffer was responsible for generation of the two ChuS mutants: H73A and H193N. The manuscript was written by Michael Suits with editorial input from Dr. Zongchao Jia and with additional help from Dr. John S. Elce and Jocelyne Suits.

Coordinates for ChuS-heme are deposited in the RSCB Protein Data Bank under accession code 2HQ2.
3.1 Abstract

Heme oxygenases catalyze the oxidation of heme to biliverdin, CO, and free iron. For pathogenic microorganisms, heme uptake and degradation is a critical mechanism for iron acquisition that enables multiplication and survival within hosts they invade. Here we report the first crystal structure of the pathogenic *Escherichia coli* O157:H7 heme oxygenase ChuS in complex with heme at 1.45 Å resolution. Compared to other heme oxygenases, ChuS has a unique fold, including structural repeats and a beta sheet core. Not surprisingly, the mode of heme coordination by ChuS is also distinct, whereby heme is largely stabilized by residues from the C-terminal domain, assisted by a distant arginine from the N-terminal domain. Upon heme binding, there is no large conformational change beyond fine tuning of a key histidine (H193) residue. Most intriguingly, in contrast to other heme oxygenases, the propionic side chains of heme are orientated towards the protein core, exposing the α-meso carbon position where O$_2$ is added during heme degradation. This unique orientation may facilitate presentation to an electron donor, explaining the significantly reduced concentration of ascorbic acid needed for the reaction. Based on the ChuS-heme structure, we converted the histidine residue responsible for axial coordination of the heme group to an asparagine residue (H193N), as well as a second histidine to an alanine residue (H73A) for comparison purpose. We employed spectral analysis and CO measurement by gas chromatography to analyze catalysis by ChuS, H193N and H73A, demonstrating that H193 is the key residue for the heme degrading activity of ChuS.
3.2 Introduction

Iron is an essential cofactor whose role as a biocatalyst or electron carrier enables organisms to fulfill many vital biological processes including respiration, the citric acid cycle, oxygen transport and utilization, gene regulation, and DNA biosynthesis (89). For bacteria invading mammalian hosts, iron acquisition, transport, and utilization is essential for survival and pathogenesis. However, due to the low solubility of iron(III) salts at physiological pH in the presence of oxygen, iron uptake, storage, and transport are considerable obstacles for invading bacteria to overcome. Enterohemorrhagic pathogens such as *Escherichia coli* O157:H7 may induce hemolytic lesions in order to gain access to host nutrient stores, such as heme; the most abundant source of invertebrate host iron (83). Furthermore, growth of *E. coli* O157 isolated from human patients is stimulated in the presence of heme and hemoglobin, suggesting that heme may be an important signaling molecule for the up-regulation of factors contributing to host colonization and virulence (83).

The proteins involved in heme internalization by enterohemorrhagic pathogens have been shown to involve homologues to members of the heme utilization locus (127). Using a laboratory strain of *E. coli* (1017 (ent::Tn5)) that was unable to import heme into the cytoplasm, it was established that the outer membrane receptor ChuA was required for heme uptake along with TonB, an energy-transducing protein associated with ExbB and ExbD, forming a complex that uses the proton motive force of the cytoplasmic membrane for the transport of ligands into the periplasm (146). Four other members of the heme uptake operon from *Yersinia enterocolitica* were shown to be required to
complete heme uptake, transport, and use as an iron source by *E. coli* K-12 (127) including: the periplasmic heme-binding protein HemT, the heme permease protein HemU, the ATP-binding hydrophilic protein HemV, and the protein HemS, whose role was supposed to have been involved in regulating the amount of non-protein-associated heme in the cell (128). These *Y. enterocolitica* proteins are homologous to members of the *E. coli* O157:H7 heme utilization locus *chuT, chuU, chuV,* and *chuS,* respectively.

We recently reported the crystal structure of apo-ChuS and confirmed via spectral analysis and CO measurement that ChuS was a HO, the first to be identified in any strain of *E. coli* (41). A *Shigella dysenteriae* chromosomal knockout of *shuS,* a homologue to *chuS,* was shown to be defective in utilizing heme as an iron source and that expression of ShuS was upregulated when challenged with iron-limiting conditions (22). Furthermore, *shuS* was implicated to have a cytoprotective role against heme toxicity in that growth of the *shuS* knockout mutant was impaired at intermediate concentrations of heme (15 µM) and that *shuS* was absolutely required for colony growth at high heme concentrations (40 µM) (22). Together, these studies suggest that ChuS and its homologues are up-regulated under iron-limiting conditions, and that ChuS activity enables pathogenic *E. coli* O157:H7 to use heme as an iron source, while conferring a cytoprotective role in the prevention of heme toxicity. It appears very probable that ChuS is an essential protein in heme utilization by various pathogenic, enterohemorrhagic bacteria.

The crystal structures of HOs from other Gram-negative bacteria such as HemO (174), HmuO (61), and PigA (46) all share a high structural similarity with human HO-1 (122); they are mainly α-helical and the heme is sandwiched between a histidine residue
and a glycine residue. However, the structure of ChuS revealed a unique architecture compared to other HOs in that it contains a central set of antiparallel $\beta$-sheets and that the N- and C-terminal halves are structural repeats. In light of the identification of ChuS as a novel HO and its unique overall structure, it was of interest to elucidate the mechanism of its heme binding and degradation. To this end, we have determined the X-ray crystal structure of ChuS in complex with heme at 1.45 Å resolution. We have further performed site-directed mutagenesis of the histidine residue (H193) (shown in the crystal structure to be responsible for axial coordination of the heme group within ChuS) to an asparagine (H193N), and of another conserved histidine residue (H73) to an alanine (H73A) as a control. Using spectral analysis and CO measurement by gas chromatography to analyze heme degradation catalyzed by ChuS, H193N, and H73A, we have demonstrated that H193 is a key residue for the heme degrading activity of ChuS.

3.3 Experimental Procedures

3.3.1 Protein expression and purification, site-directed mutagenesis

ChuS fused to an N-terminal His$_6$-tag was subcloned into a pET expression vector. Based on the structure of the heme complex and sequence conservation analysis, the mutations of H73A and H193N of ChuS were selected (41). Using the Quickchange® site-directed mutagenesis kit (Stratagene, CA), wild-type ChuS was mutated at these two histidine positions using the following primers. Base pair changes have been underlined.

H73A: sense 5’−CGTAATGAATATGCAGTCGCGGAGCAAGTGGTACG−3’,
antisense 5’–CGTACCAACTTGCTCCGCGACTGCATATTCCATTACG–3’, and H93N:
sense 5’–GGGCGATGACCGACGTTAATCAGTTTTTTACGTTGCTCAAG–3’,
antisense 5’–GAGCAACGTAAAAAAAAACTGATTAAACGTCGGTGTCATCGCCC–3’.
Briefly, polymerase chain reaction conditions and cycling parameters recommended by Stratagene (123) were followed using PfuTurbo polymerase for thirty cycles between 30 sec 95°C melt, 60 sec 55°C anneal, and 9 min 68°C extend. Methylated (non-PCR amplified) DNA was then digested with DpnI restriction enzyme for 1 h and the digested DNA was heat shock transformed into MC1061 cells. Transformed cells were subsequently plated onto LB-agar plates supplemented with 100 µg/ml ampicillin. Single colonies were grown overnight at 37°C in 5 ml Terrific Broth (Bioshop, Burlington, CAN) supplemented with 100 µg/ml ampicillin (TB-amp). Plasmid DNA was purified by means of a QIAprep Spin Miniprep Kit (Qiagen, Mississauga, CAN) was transformed into BL21(DE3), and the desired mutations confirmed by sequence analysis (Robarts Research Inst., London, CAN).

Cultures (1 L) of BL21(DE3) cells carrying plasmids for the respective recombinant protein were grown at 37°C in TB-amp. Protein expression was induced at a culture OD$_{600}$ between 0.6-1.0 using 0.4 mM IPTG and grown for 5 h. ChuS, H73A, and H193N were then purified via nickel nitrilotriacetate agarose (Ni-NTA) batch purification in phosphate buffer (pH 8.0) and dialyzed overnight in 30 mM Tris, 5 mM NaCl, pH 8.0. Over-expression of ChuS and H73A resulted in cells exhibiting a blue-green pigment which is characteristic of biliverdin accumulation (87, 156, 174). ChuS, H73A and H193N all expressed at high levels (>20 mg/L culture depending on batch) and were soluble in Tris (pH 8.0) or phosphate (pH 7.0) buffers following Ni-NTA purification.
Purification of each protein was monitored by SDS PAGE, and each was estimated to be >95% pure prior to spectral analysis, CO measurement, or heme reconstitution. H193N and H73A were subsequently dialyzed twice more to eliminate any residual imidazole. Protein concentrations were determined by the BioRad protein assay (Biorad, Mississauga, CAN) in 96 well format, using Bovine Globulin G as a protein standard.

3.3.2 ChuS-heme reconstitution

The ChuS-heme complex was prepared by dissolving heme (ferriprotoporphyrin IX chloride) (Sigma-Aldrich, Oakville, CAN) into 0.5% (v/v) ethanolamine for 30 min, then into 30 mM Tris, and the pH was adjusted to 8.0. Small volumes of this mixture were then added to dialyzed ChuS until a final ratio of 4:1 heme to protein was reached. The ChuS-heme mixture was then incubated for 15 min, centrifuged at 16,000 rpm for 15 min, in a JA-20 Beckman Coulter rotor (Beckman Coulter, Mississauga, CAN), and purified using a Resource Q column on an ATKA Explorer FPLC with 30 mM Tris, 1 M NaCl (pH 8.0) as the elution buffer. Collected fractions were examined by SDS-PAGE, and those that were evaluated to contain pure protein were pooled. Pure ChuS was concentrated using an Amicon Ultra 15 kDa cutoff centrifugal filtration device (Fisher, Mississauga, CAN) and used immediately for crystallization. Reconstituted ChuS-heme complex concentrations were determined by using the extinction coefficient reported for ShuS ($\varepsilon_{410}$) of 159 mM$^{-1}$cm$^{-1}$ for ChuS (11).
3.3.3 Crystallization of ChuS-heme

ChuS-heme crystals were obtained only by the sitting drop vapor diffusion method, in 96-well plates (Greiner Bio-One, NC, USA), using freshly prepared protein and reagents. Crystallization was achieved by mixing 1.75 µl of 20 mg/ml ChuS-heme with 1.75 µl of chilled reservoir solution consisting of 12-15% (w/v) PEG 3350, 0.15-0.20 M magnesium formate, and 10-20 mM NAD. Crystallization plates were incubated in the dark at 4°C, with both hexagonal and triangular prism crystals appearing after 1 to 3 d. For cryoprotection, crystals were soaked in the reservoir solution containing ethylene glycol (concentration increased stepwise to 30% [v/v]), picked up using a nylon loop, and flash-cooled in the N$_2$ cold stream at 100K. Two crystal forms belonging to space groups P6$_5$ and R$_3$ were obtained whose morphology were hexagonal and triangular prism, respectively.

3.3.4 Data collection and structure determination

Diffraction data of the P6$_5$ and R$_3$ crystals were collected at beam line X29 equipped with an ADSC Quantum-315, nine quadrant, CCD detector at the National Synchrotron Light Source, Brookhaven National Laboratory. For both the P6$_5$ and R$_3$ crystals, data were collected for 180° (0.5° and 1.0° oscillations, respectively) at the wavelength 1.100 Å. Data were processed with HKL2000 (102) and the structure of the R$_3$ crystal was solved via molecular replacement using the program Phaser (88) with apo-ChuS (PDB ID code 1U9T) as the initial probe structure. Although the P6$_5$ crystals
diffracted beyond 2.0 Å resolution, due to the long unit cell dimensions and resulting spot proximity, diffraction was only collected to 2.7 Å. The heme moiety was manually placed in the ChuS structure according to the difference density in both space groups. The ChuS-heme complex structure in R₃ space group was completed by iterative cycles of manual fitting using XtalView/Xfit (84) and refinement using Refmac5 (96, 148) at a final resolution of 1.45 Å. Coordinates and structure factors for ChuS-heme are deposited in the RSCB Protein Data Bank under accession code 2HQ2. Due to lower resolution and poor data quality in regions along the long axes, the P6₅ complex structure was not further refined beyond confirming heme the position.

3.3.5 Spectral analysis and CO measurements

Reconstitution of ChuS, H73A, and H193N with heme was achieved by dissolving heme into 0.5% (v/v) ethanolamine and diluted into phosphate buffer just prior to mixing with 10 µM protein samples to a final ratio of 1:1 protein to heme. Absorbance data were collected after 15 min incubation at room temperature in 100 mM phosphate buffer (pH 7.0) using a microplate spectrophotometer (Bio-Tek Instruments, Inc. NY, USA.). Heme degradation was initiated in parallel by adding ascorbic acid to a final concentration of 50 µM, and spectra were recorded between 300 and 700 nm at 5 nm intervals.

For CO measurement, in 2-ml amber vials, 250 µl reaction volumes of 10 µM heme and 10 µM ChuS, H73A, or H193N in 100 mM phosphate buffer (pH 7.0) were incubated with constant shaking at room temperature for 15 min at which time enzymatic
heme degradation was initiated by adding ascorbic acid to a final concentration of 50 µM. The amber vials were then sealed with screw caps and blue silicone rubber septa, the headspace above each was immediately purged with CO-free air, and the vials were incubated for 25 min at 37°C with constant shaking. The reaction was stopped by placing the vials on powdered dry ice (-78°C), where they remained for approximately 30 min. The amount of liberated CO in the headspace of each vial was measured using a gas chromatograph equipped with an HgO reduction detector (Trace Analytical). The amount of CO resulting from ChuS-catalyzed breakdown of heme was determined by comparing \( n=3 \) peak area measurements for CO against linear CO standard curves (\( n=2 \) determinations; average correlation coefficient 0.9859). Technical details of CO detection are described elsewhere (27, 153).

### 3.4 Results

#### 3.4.1 Structure determination and analysis of ChuS-heme

Crystallization of ChuS reconstituted with heme using the hanging drop vapor diffusion method turned from red to blue-green following room temperature incubation periods of over one week. In the sitting drop vapor diffusion setup, two different crystal forms of ChuS-heme were obtained from the same set of crystallization conditions over the same time period, but differed dramatically in their morphology. In the case of the P6\(_5\) crystals, which appeared more frequently at room temperature, the crystals appeared as broad (>1 mm), flat, hexagonal plates. Crystals belonging to the R\(_3\) space group appeared as clusters of long (>1 mm) triangular prisms and diffracted at 1.45 Å resolution.
Crystals from the P6$_5$ space group contained two molecules in the asymmetric unit whereas those belonging to the R$_3$ space group contained only one.

The final structure of ChuS-heme complex contains residues 0-337 of a total 354 amino acids including the His$_6$-affinity tag for the R$_3$ space group (Figure 3.1). In total, 100% of residues are within the allowed region of the Ramachandran plot. Only one residue, A96, was in the generously allowed region and resides just prior to a surface-exposed $\beta$-turn. Final refinement statistics are summarized in Table 3.1.

The ChuS-heme complex structure is virtually identical to that of apo-ChuS. The overall structure comprises a central core of two large pleated $\beta$-sheets, each consisting of nine anti-parallel $\beta$-strands sandwiched together and bowing outward in a saddle motif (Figure 3.1). Each $\beta$-sheet is flanked at its N-terminus by one pair of parallel $\alpha$-helices and at the C-terminus by three $\alpha$-helices in an $\alpha$-loop-$\alpha$-loop-$\alpha$ motif, forming a structural repeat. When ChuS-heme is superimposed with apo-ChuS, the only notable structural change is that the two $\alpha$-helices N-terminal of H193 have shifted toward the core of the protein (H193 by 3.58 Å), resulting in the appropriate orientation of this key histidine residue for heme coordination (see Section 3.4).
Table 3.1: Diffraction data and refinement statistics for ChuS-heme

<table>
<thead>
<tr>
<th>Space Group</th>
<th>P6₅</th>
<th>R₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a = b (Å)</td>
<td>58.4</td>
<td>106.5</td>
</tr>
<tr>
<td>c (Å)</td>
<td>390.0</td>
<td>90.2</td>
</tr>
<tr>
<td>α = β (°)</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>γ (°)</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>No. molecules in ASU</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>λ (Å)</td>
<td>1.100</td>
<td>1.100</td>
</tr>
<tr>
<td>Resolution Range</td>
<td>50 - 2.7</td>
<td>50 - 1.45</td>
</tr>
<tr>
<td>Total Reflectionsa</td>
<td>92612</td>
<td>361180</td>
</tr>
<tr>
<td>Unique Reflectionsa</td>
<td>20724</td>
<td>67663</td>
</tr>
<tr>
<td>Completeness (%)b</td>
<td>85.2 (97.0)</td>
<td>98.7 (98.4)</td>
</tr>
<tr>
<td>R_sym (I) (%)</td>
<td>10.0</td>
<td>6.7</td>
</tr>
<tr>
<td>I/σIb</td>
<td>23.9 (14.8)</td>
<td>38 (1.8)</td>
</tr>
</tbody>
</table>

**Refinement Statistics**

<table>
<thead>
<tr>
<th>Resolution range (Å)</th>
<th>50 – 1.45</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_work, (%)</td>
<td>17.05</td>
</tr>
<tr>
<td>R_free, (%)</td>
<td>19.30</td>
</tr>
<tr>
<td>No. reflections unique/ R_free</td>
<td>63392 / 3371</td>
</tr>
<tr>
<td>No. atoms, protein / Heme/ solvent (H₂O) / total</td>
<td>2606 / 45 / 354 / 3005</td>
</tr>
<tr>
<td>B factors (Å²) protein / heme/ solvent / total</td>
<td>21.4 / 24.3 / 32.2 / 22.7</td>
</tr>
<tr>
<td>RMS, bond length (Å) / bond angle(°)</td>
<td>0.0199 / 1.894</td>
</tr>
<tr>
<td>PDB Accession Code</td>
<td>2HQ2</td>
</tr>
</tbody>
</table>

aNumber of reflections following merging
bValues in parentheses are for the outermost shell (P6₅: 2.81-2.70 Å, and R₃: 1.50-1.45 Å)
Figure 3.1 Ribbon diagram of ChuS in complex with heme. ChuS binds to heme in a cleft region delineated by the C-terminal half (green) and N-terminal half (slate), between H193 at the base of an α-helix and R100 via two water molecules (blue) from central set of β-sheets at the core of ChuS. The residues important for heme coordination as well as the mutant control position H73 are depicted in red. Fig. 3.1 and 3.2 were generated using PyMOL (DeLano, 2006, The PyMOL Molecular Graphics System, www.pymol.org).
3.4.2 Heme binding

The heme group is coordinated in a cleft region delineated between H193 at the base of a C-terminal $\alpha$-helix and R100 from the central set of $\beta$-sheets at the core of ChuS (Figure 3.1). A network of hydrogen bonds is formed between R100 to the Fe atom of the heme group via two water molecules with distances of 2.93, 2.17, and 1.93 Å respectively. In this cleft the heme group is stabilized by the non-polar series L90, L92, F102, V192, and F243, and the polar series R100, H193, R206, M241, K291, Q313, Y315 and R318 (Figure 3.2). In contrast to other HOs, the propionic side chains of the heme moiety point toward the interior of one of the structural lobes formed between the two halves of ChuS and are stabilized by the positively charged side chains R206, K291, and R318 (Figure 3.2). These interactions are probably important in maintaining the correct orientation of heme within ChuS. The burying of the propionic side chains exposes the $\alpha$-meso carbon of the heme group which is where $O_2$ is selectively added during heme degradation in HO reactions. Although the P6$_5$ data were of lower quality, the identical heme position in this structure was unmistakable. The heme orientation in ChuS is unique compared to other HOs, such as HO-1, in which the propionic side chains point outward (117), although the heme degrading enzymes IsdG/IsdI may bind heme with the propionic side chains pointing inward (125).

Essential to the heme coordination is His193, which makes axial interaction with the central iron atom of the heme group. In ChuS, the interaction distance is 2.03 Å between the NE2 atom of H193 and the center of the iron atom, allowing H193 to clamp the heme firmly into place. R100, from the N-terminal domain, interacts with heme via
two water molecules and originates from the protein core beneath the heme group on one of the central β-sheets (Figure 3.2). These water molecules were clearly visible even at early refinement using multiple data sets, and are well defined in the final high-resolution structure.
Figure 3.2 Difference omit map for the heme moiety and two water molecules (blue) within the active site pocket of ChuS contoured to $2\sigma$ in the R3 space group. Heme and waters were omitted from refinement prior to map calculations. Residues contributing to heme stabilization include the non-polar series L90, L92, F102, V192, and F243, and the polar series R100, H193, R206, M241, K291, Q313, Y315 and R318. Heme is therefore mainly coordinated by the C-terminal half but also by a distant residue, R100 from the N-terminal half. A network of hydrogen bonds is formed between R100 to the Fe atom of the heme group via two water molecules with distances of 2.93, 2.17, 1.93 Å respectively. In this orientation the propionic side chains of the heme group point toward the protein interior, exposing the $\alpha$-meso carbon edge (black arrow). This presentation of the $\alpha$-meso edge may facilitate electron attack during heme degradation.
3.4.3 Spectroscopic properties of ChuS, H73A, and H193N in complex with heme

H193 resides at the end of an \(\alpha\)-helix, pointing into a cleft delineated by the \(\alpha\)-helix and the core set of \(\beta\)-sheets, interacting directly with iron in the heme group. Thus, the mutation of H193 was expected to inactivate the protein. On the other hand, H73 is within the central \(\beta\)-sheets, pointing between the N- and C-terminal halves and does not make contact with the heme group. Since H73 is fully conserved, but does not interact with heme, it provides a good control for H193. As previously reported, the absorbance spectrum of native ChuS exhibits a clear Soret peak at 410 nm, together with the common characteristic \(\alpha/\beta\)-bands of HOs around 550 and 580 nm (11, 41, 118). The H73A mutation was not observed to instigate any spectral changes compared to native ChuS. However, the spectrum of the H193N mutant broadened, showed a decrease in the magnitude of absorbance, and shifted the peak absorbance to 390 nm compared to the Soret absorption observed for ChuS and H73A. A similar trend of spectral flattening and peak shift to 625 nm was also observed for H193N in the \(\alpha/\beta\)-band region.

3.4.4 Heme degradation by ChuS

Bacterial and mammalian HOs use the Cytochrome P450 Reductase-NADPH system, ferredoxin, flavodoxin, and/or ascorbic acid as reducing partners \textit{in vitro} (26, 160, 174). We have previously demonstrated the ability of ChuS to use either CPR-NADPH or ascorbic acid as an electron donor for heme degradation (16, 41). It is important to note that in our analysis we employed only 50 \(\mu\)M ascorbic acid, a concentration between 100-
(77) and 350-fold (26) less than that used for characterizing other HOs, thereby minimizing the non-enzymatic degradation of heme through coupled oxidation. Furthermore, calatase and superoxide dismutase were used for additional controls. To examine the heme-degrading activity of ChuS, H73A, and H193N, we monitored the reaction by UV-visible spectroscopy for 1 h after the addition of ascorbic acid. As the reaction proceeded, the Soret peak for ChuS and H73A at 410 nm decreased, the small broad peak at 560 nm increased, and the absorption at 680 nm initially increased, then decreased (Figure 3.3). When ascorbic acid was added to H193N reconstituted with heme, the profile of the spectrum did not change, but was observed to increase slightly in absorbance. In the case of native ChuS, addition of catalase and superoxide dismutase did not affect the observed spectral change, demonstrating specific enzyme activity.

When HOs degrade heme to biliverdin, CO is released as a byproduct. Using gas chromatography, we measured CO release for wild-type ChuS, as well as the two mutants. As shown in Figure 3.4, whereas wild-type ChuS and H73A were able to generate CO as expected, the H193N mutation inactivated the enzyme. However, based on the appearance of a broad peak for H193N near the Soret region of native ChuS, this mutant seems to retain some interaction with heme.
Figure 3.3 Spectral change of ChuS reconstituted with heme, monitored over time following ascorbic acid addition. Absorbance measurements between 500 and 700 nm have been amplified five fold to emphasize the change in the $\beta$- and $\alpha$-bands. Coloured arrows corresponding to the spectra of each protein follow changes in absorption at 408, 545, and 680 nm during reaction progression.
Figure 3.4 CO measurement via gas chromatography of ChuS, H73A, H193 and heme following ascorbic acid addition and 25 min incubation ($n=3$). Mutation at conserved H193 abolished heme degradation activity by ChuS.
3.5 Discussion

The X-ray crystal structure of the *E. coli* O157:H7 HO ChuS-heme complex has revealed that the ChuS employs a unique mode of heme coordination. Whereas the structures of other HOs characterized to date coordinate heme between a set of α-helices, ChuS coordinates the central iron atom of the heme group between His193 at the end of an α-helix and R100 (via two water molecules) originating from one of the central set of β-sheets. In this fashion, the heme binding site of ChuS occurs between two distinct interfaces, delineated by the repeating structural elements contributed mainly from the C-terminal repeat and, to a lesser extent, the N-terminal half. While this heme coordination is unique compared to other HOs, it is analogous to the mode of heme coordination assumed for the heme degrading enzymes IsdG and IsdI from *Staphylococcus aureus* (125) and is also similar to the heme binding PAS domain of *BjFixL* (21) and *EcDos* (16).

Due to an absence of electron density in the data derived from the ChuS-heme complex, a gap exists in the structure between residues 168 and 176 (V168-DAPVVQT-R176), dividing the structural repeats of ChuS almost at the mid point. A similar gap was also observed for apo-ChuS in the region of K166 to V173. Interestingly, the key heme coordination residue H193 is located at the end of an α-helix towards the C-terminus of the gap. In fact, H193 is in the only region that exhibits considerable conformational change in the entire ChuS structure. It is thus conceivable that the flexible linker between the N- and C-halves provides the flexibility for the critical iron-coordinating H193 residue of the C-terminal half. In the N-terminal half, R100 is also important for
coordination of the iron atom from the β-sheet core via a hydrogen bonding network created through two water molecules. Therefore, the flexibility observed in the linker between the two halves of ChuS may be required for heme binding. Communication between the two lobes may also result in increased control over heme processing, similar to that of the hemopexin dimer (40).

In spite of the evidence that genes homologous to *chuS* have been shown to be up-regulated under iron limiting conditions, that these homologues prevented heme toxicity in *S. dysenteriae* (22), and that no other heme degrading enzyme has been identified in any *E. coli* strain, the precise role for ChuS and its homologues has been downplayed in recent reports to be that of a cytoplasmic trafficker of heme, or a non-specific oligomeric complex that binds to DNA (11, 118). By carefully examining the spectral progression and CO measurement of ChuS catalyzed degradation of heme in the presence of minimal amounts of ascorbic acid, we conclude that ChuS specifically degrades heme and that His193 is the essential amino acid responsible. The confirmation of H193 as an integral catalytic residue is affirmed through its sequence conservation, its role in coordination of the heme moiety in the crystal structure, as well as the comparison of the H73A control mutant. The spectral shifts observed for ChuS and H73A are similar to those observed for other HOs, which suggest the formation of biliverdin and free iron rather than the Fe$^{3+}$-biliverdin complex formed by HO-1. Because the addition of catalase and superoxide dismutase did not affect the trends of the spectral change, the observed activity of ChuS can be attributed only to specific heme degradation (41).

It is difficult to extrapolate the patho-physiological significance of the heme degrading ability of ChuS to its homologues in other bacteria due to the fact that other
redundant systems have been identified for iron acquisition and utilization from heme sources including proteins from *Pseudomonas aeruginosa* (PigA and BphP) and *Staphylococcus aureus* (IsdG and IsdI). Furthermore, redox enzymes are notoriously non-specific with respect to their reaction with oxygen species (52). This has led to a lack of straightforward characterization of other bacterial cytoplasmic heme binding proteins outside those with sequence homology to HO-1. Therefore, understanding the interplay between sequence and structural motifs in heme binding and utilization are of particular interest in order to clarify our understanding of heme utilization by pathogenic bacteria.

As the propionic side chains of the heme group point toward the β-sheet core of ChuS, the orientation of the heme group within ChuS may allow for specific targeting of heme for degradation. This degradation may occur through the initial activation of the two water molecules by R100, which would activate O2 in a similar fashion to the mechanism hypothesized for HO-1 and its homologues (46). The distinct orientation of heme may further explain the minimal amount of ascorbic acid needed for the degradation reaction, because the α-meso carbon atom is exposed to facilitate interaction with an electron donor.

During our investigation, a variety of practical obstacles were overcome which could help in future studies of ChuS and its homologues. As previously reported, ChuS and its homologues form high MW aggregates following incubation over a few days at 4°C (11, 16, 118). In order to overcome aggregation, fresh, unfrozen preparations of pure ChuS were absolutely necessary for crystallization and biophysical studies. Furthermore, even residual amounts of imidazole were observed to inhibit crystallization with heme
and the heme degradation reaction. Purification of ChuS must therefore ensure complete removal of imidazole from reaction buffers.

In conclusion, we have provided a first view of the heme-bound structure of ChuS; a novel HO from a pathogenic bacterium with a unique structure. The heme binding site in ChuS is very different from those of other known HO-heme complexes. H193 plays a key role in anchoring the heme group by tightly interacting with the central iron ion. The majority of the heme stabilization is contributed by the C-terminal half, along with R100 from the N-terminal half that also interacts with iron via two fully conserved water molecules. The exposed α-meso region, which is unique among known HO-heme complexes, may facilitate presentation to an electron donor and thus provide an explanation for the much lower concentration of ascorbic acid needed for the reaction. The H193N mutation, examined by spectral and CO measurements, clearly demonstrates the critical role of H193.

3.6 Footnotes

We thank Dr. B. McLaughlin and Dr. K. Nakatsu for help with gas chromatography. We would like to express our gratitude to Howard Robinson and Brookhaven National Laboratory beamline X-29, where the diffraction data were collected for both the P6₅ and R₃ crystals of ChuS-heme. Excellent equipment management was also contributed by Hans Metz. This work was supported by the CIHR. MS was supported by an E.G. Bauman Fellowship and an Ontario Graduate Scholarship. ZJ is a Canada Research Chair in Structural Biology and an NSERC Steacie Fellow.
Coordinates and structure factors for ChuS-heme are deposited in the RSCB Protein Data Bank under accession code 2HQ2.
Chapter 4

Structure of the *Escherichia coli* O157:H7 heme binding protein ChuX and its role in regulating heme uptake

Preface: Michael Suits was responsible for protein expression, purification, mutagenesis crystallization, data collection, structure solution and refinement, and all biochemical experiments in the investigation of function. Dr. Gour P. Pal conducted the initial expression and crystallization which was refined and expanded by Michael Suits. Dr. Mike Nesheim helped in heme reconstitution setup and interpretation of results to calculate the $K_d$ for ChuX-heme. The *chuX* gene was cloned by Pietro Iannuzzi. The manuscript was written by Michael Suits with editorial input from Dr. Zongchao Jia and with additional help from Brent Wathen and Jocelyne Suits.
Coordinates and structure factors for ChuX are deposited in the RCSB Protein Data Bank (www.rcsb.org) under accession code 2OVI.

4.1 Abstract

The genes encoded within the heme utilization operon enable various Gram-negative bacteria to effectively take up and utilize heme as an iron source. For many pathogenic microorganisms, iron acquisition from heme sources stimulates growth and enables successful multiplication and survival within hosts they invade. While the proteins involved in heme internalization are well characterized functionally, the fate of heme once it has reached the cytoplasm has only recently begun to be resolved. Here we report the first crystal structure of the conserved heme utilization protein ChuX from pathogenic *Escherichia coli* O157:H7 determined at 2.05 Å resolution. ChuX forms a dimer which, surprisingly, displays a very similar fold to the monomer structure of two other heme utilization operon proteins ChuS and HemS from *E. coli* and *Yersinia enterocolitica*, despite sharing less than 19% sequence identity. Absorption spectral analysis of heme reconstituted ChuX demonstrates that ChuX binds to heme in a 1:1 manner implying that each homodimer coordinates two heme molecules, in contrast to ChuS where only one heme molecule is bound. Based on sequence and structural comparisons, we designed a number of site-directed mutations in ChuX to probe the heme binding sites and dimer interface. Mutations of both H65L and H98D were required to abolish heme binding capacity of ChuX. Further, mutations within the dimer interface resulted in compromised heme binding, supporting the notion that varying ChuX juxtapositions observed in the structure may serve to modulate heme binding. Taken
together, it appears that ChuX acts upstream of ChuS and regulates heme uptake through ChuX-mediated heme binding and release.

4.2 Introduction

Iron is the fourth most abundant element in the Earth’s crust (6). However, due to the low solubility of iron(III) salts at physiological pH in the presence of oxygen, iron uptake, storage, and transport is a serious obstacle for bacteria to overcome for survival and pathogenesis. Furthermore, invading microorganisms are confronted with the obstacle that 99.9% of remaining body iron is sequestered as protein bound heme in hemoglobin, myoglobin, and various other host heme coordinating proteins (127). Through mechanisms of heme detection and liberation from host hemoproteins, bacteria have evolved the ability to effectively uptake heme through translocation machinery and then to utilize heme as a prosthetic group or as a valuable pool of iron. Additionally, the products of heme degradation, biliverdin and bilirubin, may serve a cytoprotective role in bacteria where they can act as antioxidants (73).

Enterohemorrhagic pathogens such as *Escherichia coli* O157:H7, for which isolates from human patients have been shown to have stimulated growth in the presence of heme and hemoglobin, may induce hemolytic lesions in order to gain access to this potential sources of heme and iron (83). Using a laboratory strain of *E. coli* (1017 (ent::Tn5)) that is unable to import heme into the cytoplasm, it was established that the outer membrane receptor ChuA (*E. coli* heme-utilization protein A) was required for heme uptake along with TonB, an energy-transducing protein associated with ExbB and ExbD that uses the proton motive force of the cytoplasmic membrane for the passage of
ligands into the periplasm (146). Three other members of the heme uptake operon from *Y. enterocolitica* were shown to be required to complete heme uptake and transport by *E. coli* K-12 (127) including: the periplasmic heme-binding protein HemT, the heme permease protein HemU, and the ATP-binding hydrophilic protein HemV (128).

Once internalized, heme would be rapidly degraded by bacterial proteins such as ChuS (*E. coli* O157:H7) (139), Cj1613c (*Campylobacter jejuni*) (111), HmuQ (*Bradyrhizobium japonicum*) (106), IsdG (*Bacillus anthracis*) (125), and ShuS (*Shigella dysenteriae*) (165), all of which could use the abundance of potential electron donors to liberate iron that can then be incorporated into various proteins where it acts as a biocatalyst or electron carrier. Alternatively, internalized heme could be stored or used in various vital biological processes including photosynthesis, gene regulation, oxygen utilization, N\textsubscript{2} fixation, methanogenesis, H\textsubscript{2} production and consumption, the tricarboxylic acid (TCA) cycle, or DNA biosynthesis (89, 91). A tightly regulated balance between heme uptake and degradation is of critical importance.

Restriction mapping and sequence analysis of selected regions of *E. coli* O157:H7 DNA, the serotype responsible for outbreaks of hemorrhagic colitis and hemolytic uremic syndrome, has revealed that the organization of the heme transport locus is strikingly similar to that of other enteric bacteria such as *Y. enterocolitica* and *S. dysenteriae* (127, 128). These include the three other genes clustered within the heme utilization locus, *chuW, chuX*, and *chuY* which are poorly characterized, but whose homologues appear to be co-transcribed with the periplasmic binding protein *shuT* in *S. dysenteriae* (127, 128). Twelve nucleotides downstream of *shuW* is the start codon for *shuX*, and the initiation codon of *shuY* overlaps the stop codon of *shuX* which may indicate that these two smaller
genes are co-transcribed (164). In a similar sequence analysis of the heme uptake locus from *Y. pestis*, genes homologues to *chuX* and *chuY* (*orfX* and *orfY*, respectively) were evaluated to be located immediately downstream of the heme receptor gene *hmuR* and were therefore likely co-expressed (145). In *Vibrio anguillarum*, which can use heme and hemoglobin as a source of iron, using a *lacZ* reporter system it was shown that a homologue of *chuX*, *huvX*, was co-transcribed with *huvZ* under iron limiting conditions (95). An *E. coli* homologue to *huvZ* was not identified via NCBI Blastp sequence analysis at (www.ncbi.nlm.nih.gov/BLAST/). Furthermore, an examination of the -10 and -35 elements downstream of *huvX* revealed sequence with similarity to sigma$^{70}$-promoters, as well as the presence of a Fur element (95). Characterization of either *chuX* or *chuY* gene product or their homologues from various heme utilizing organisms had not received any attention at the beginning of our investigation. However, based on Blastp sequence analysis *ChuX* was suggested to be either a hypothetical protein or protein associated with heme-iron utilization, and that *ChuY* is a hypothetical protein or nucleoside-diphosphate-sugar epimerase (143).

While traditional heme studies have focused on heme as a prosthetic group within hemoglobin and myoglobin, or general heme involvement in iron processing, many novel heme associating proteins have been discovered recently. These include the cytochrome c chaperone CcmE (123), the iron-dependent regulator of heme biosynthesis Irr (107), and the heme-based aerotactic transducer Hem-AT (64). Other heme proteins function through conformational changes induced by a modification to the ferrous heme environment upon the coordination of an effector ligand such as NO, O$_2$, or CO, as in the case of guanylate cyclase (172), FixL (52), and CooA (81), respectively. Furthermore,
heme has recently been shown to have functionality beyond simply acting as a prosthetic group or a source of iron. In HeLa cells, succinyl acetone-induced heme deficiency was shown to increase the protein levels of the tumor suppressor gene product p53 and CDK inhibitor p21, decrease the protein levels of Cdk4, Cdc2, and cyclin D2, and diminish the activation/phosphorylation of Raf, MEK1/2, and ERK1/2-components of the MAP kinase signaling pathway (168). This set of results highlights a role for heme as an effector molecule beyond the regulation of proteins associated with heme and iron metabolism.

Despite the emerging knowledge concerning hemoproteins in many Gram-negative bacteria, very little has been published concerning the three lesser characterized members of the heme utilization operon, chuW, chuX, or chuY. We have determined the crystal structure of ChuX at 2.05 Å resolution. Through site-directed mutagenesis of conserved histidine residues, spectral analysis, and structural comparison of ChuX with ChuS and HemS we have identified heme coordination sites within ChuX and the structural basis of heme binding and release. Furthermore, we have shown that ChuX plays a protective role against heme toxicity.

4.3 Experimental Procedures

4.3.1 Sequence analysis, protein expression and purification, site-directed mutagenesis

Sequence analysis was carried out using Blastp (www.ncbi.nlm.nih.gov) (5) and Multialign (http://bioinfo.genopole-toulouse.prd.fr/multalin/) (31). ChuX fused to an
N-terminal His<sub>6</sub>-tag was subcloned into a pET 15b expression vector. Based on the apo-ChuX structure and sequence conservation analysis, the two single mutants H65L and H98D, a double mutant (H65L/H98D) (termed DM) and a triple mutant (V69E/L71H/F73S) (termed BetaX) of ChuX were created. The BetaX triple mutant alters three nonpolar residues shown in the crystal structure to interact across the central β-sheets in ChuX homo-dimers, changing them to polar ones in an attempt to disrupt homo-dimerization. Using the QuickChange® site-directed mutagenesis kit (Stratagene, CA), wild-type ChuX was mutated at these three positions using the following sense primers. Base pair changes have been underlined. H65L: sense 5'-CGTCACCACGTTAGTACTTACTGCGATGTAATCC-3', H98D: sense 5'-GCACGGCATGTCCGGGGATATCAAAGCAGAAAAACTG-3', and BetaX: sense 5'-GTACATACTGCCGATGAAATCCACGAATCTAGCGCGAAACTGCTTCGGG-3', Antisense primers were reverse complements to sense primers in each case.

Briefly, polymerase chain reaction conditions and cycling parameters recommended by Stratagene (123) were followed using PfuTurbo polymerase for thirty cycles between 30 sec 95°C melt, 60 sec 55°C anneal, and 9 min 68°C extend. Methylated (non-PCR amplified) DNA was then digested with the restriction enzyme DpnI for 1 h and the digested DNA was heat shock transformed into MC1061 cells. Transformed cells were subsequently plated onto Lauria Broth (LB)-agar plates supplemented with 100 µg/ml ampicillin. Single colonies were grown overnight at 37°C in 5 ml Terrific Broth (Bioshop, Burlington, CAN) supplemented with ampicillin. Plasmid DNA purified by means of a QIAprep Spin Miniprep Kit (Qiagen, Mississauga, CAN) was transformed into BL21(DE3), and the desired mutations confirmed by sequence analysis. 1 L cultures of
BL21(DE3) cells carrying plasmids for the respective recombinant protein were grown at 37°C in Terrific Broth-ampicillin. Protein expression was induced at a culture OD$_{600}$ between 0.6-1.0 using 0.4 mM isopropyl-1-thio-β-D-galactopyranoside and grown for 5 h. ChuX protein substituted with seleno-methionine was expressed in the metA- *E. coli* strain DL41 in LeMaster medium supplemented with ampicillin (51). Over-expression of ChuX resulted in cells exhibiting a slight yellow colour compared to H65L, H98D, DM, or BetaX cultures. ChuX and mutants were all purified via nickel nitritotriacetate agarose batch purification in phosphate buffer (pH 8.0) and dialyzed overnight in 30 mM Tris, pH 8.0. Proteins were then purified further via anion exchange, using a 6 ml Resource Q column on an ATKA Explorer FPLC in 30 mM Tris, pH 8.0 with 30 mM Tris, 500 mM NaCl, pH 8.0 as the elution buffer. Purification was monitored by SDS PAGE, and each protein was estimated to be >95% pure prior to concentrated using an Amicon Ultra 15 kDa cutoff centrifugal filtration device (Fisher, Mississauga, CAN). Protein concentrations were determined by BioRad protein assay (Biorad, Mississauga, CAN) using Bovine Globulin G as a protein standard and used immediately for crystallization. ChuX, H65L, H98D, and DM all expressed at high levels (>20 mg/L culture depending on batch) and were soluble in Tris (pH 8.0) or phosphate (pH 7.0) buffers following affinity purification. BetaX expression levels (2.4 mg/L culture) were much lower than those for native ChuX or the other mutants. ChuX used in reconstitution experiments was stored in liquid nitrogen, the concentrations of which were equalized prior to spectral analysis to ensure that spectral comparisons were the result of differences in heme coordination and not due to small changes in relative protein concentrations.
4.3.2 Crystallization of ChuX

ChuX crystals were obtained using the hanging drop vapor diffusion method, using freshly prepared protein and reagents. Crystallization was achieved by mixing equal parts 40 mg/ml ChuX with reservoir solution consisting of 1.50 M ammonium sulfate, 50 mM Hepes, pH 7.5, 2% (v/v) PEG 400. Large rectangular crystals (3.0 x 1.5 x 0.75 mm) appeared after 2-4 d. For cryoprotection, crystals were looped through Paratone-N oil, and flash-cooled in the N$_2$ cold stream at 100K.

4.3.3 Data collection and structure determination

ChuX crystals were determined to belong to the space group P4$_1$ with four molecules in the asymmetric unit. Native and single-wavelength anomalous dispersion (SAD) data sets were collected, respectively, at beamlines X8C and X25, Brookhaven National Laboratory. SeMet SAD data were collected at a peak wavelength determined by fluorescence scan to be 0.97950 Å, for 360° with 1.0° oscillations. The native data were collected for 180° in total, using 1.0° oscillations. Both datasets were processed with HKL2000 (102). Heavy atom positions were initially determined using HySS (64, 168) refined, and phases determined with SOLVE (144). Density modification, solvent flattening, and automatic chain tracing were performed using RESOLVE (144). The remainder of the model was built by manual fitting using XtalView/Xfit (84). This
structure was refined using native data with the program CNS (19). A structural homology search using the SCOP database (97) was performed with the program SSM (77) and protein structures were superimposed using the CCP4 program LSQKAB (Superpose) (2, 67, 77, 78). Modeling of heme binding to ChuX was done using Autodock 3.0 (93) and crystal contacts were evaluated with CryCo (42).

4.3.4 Heme reconstitution

ChuX-heme was prepared by dissolving heme (ferriprotoporphyrin IX chloride) (Sigma, Oakville, CAN) in 0.5% (v/v) ethanolamine for 30 min, and transferring the solution into 50 mM NaH$_2$PO$_4$. The pH was adjusted to 7.0 just prior to incubation with ChuX in all reconstitution experiments. Small volumes of the heme mixture were then added until a final ratio of 2:1 heme to ChuX was reached. The ChuX-heme mixture was then incubated for 15 min, centrifuged at 16,000 rpm for 15 min at 4°C in a JA-20 Beckman Coulter rotor (Beckman Coulter, Mississauga, CAN), and purified using a Sepharose-G200 column on an ATKA Explorer FPLC in 50 mM NaH$_2$PO$_4$. The spectrum of concentrated, ChuX-heme purified by size exclusion was then recorded.

ChuX binding kinetics were determined by the addition of 15 µM ChuX to various amounts of heme (0-30 µM), followed by a 15 min incubation at room temperature, and absorbance measured at 404 nm. Spectral characterization of ChuX mutants was achieved by adding 5 µM heme to a molar equivalent of protein. In heme transfer experiments, protein heme complexes of ChuX or ChuS were setup as in the mutant spectral characterization experiments. To a set of ChuX-heme mixtures an
equivalent amount of ChuS was added, and 50 µM of ascorbic acid was added to each. Changes in absorption at 408 nm, the peak wavelength for ChuS-heme, were then monitored over time. All absorbance data were collected in duplicate using a microplate spectrophotometer (Bio-Tek Instruments, Inc. NY, USA.) for spectra between 300 and 700 nm at 1 nm intervals, 250 µl volumes, and the mean values plotted.

4.3.5 Growth of ChuX and DM in heme

Quadruplicates of BL21(DE3) cells were freshly transformed with pET15b plasmids that contained ChuX, DM, or C1528 (a 12.2 kDa hypothetical protein from *E. coli* CFTO73 unrelated to heme processing). Individual colonies were selected and grown overnight in 100 ml LB supplemented with 100 µg/ml ampicillin. Cultures were then diluted in 50 ml pre-warmed LB ampicillin to an OD 

|600 reading of 0.05 and grown for 20 min at 37°C with aeration. Protein expression was induced using 0.2 mM isopropyl-1-thio-β-D-galactopyranoside, and cultures were grown for another 20 min. Sterile filtered heme was then added to bring the solution to a concentration of 2 µM and the OD 

|650 readings were then measured over time. This wavelength was selected to minimize the amount of additional absorbance detected due to heme coordination.

4.4 Results

4.4.1 Sequence Analysis
NCBI Blastp sequence analysis suggested that ChuX belongs to a family of conserved putative heme-iron utilization proteins with the conserved domain DUF1008 (143). ChuX has high sequence similarity with proteins from human pathogens such as *S. dysenteriae* (ShuX) and various *Yersinia* strains (COG3721), as well as *Photobacterium damselae* (HutX), and *Listonella anguillarum* (HuvX). ChuX homologues also appear in photoactive bacteria such as *Halorhodospira halophila* and *Rhodopseudomonas palustris* which use either iron or the products of heme degradation in the production of light harvesting photoreceptors (47). Sequence alignment of ChuX and its homologues revealed many conserved residues (Figure 4.1), especially in the C-terminal half. Notably residues H65, H98, and the segment of residues K134 to L145 are highly conserved across ChuX homologues.

### 4.4.2 ChuX protein fold and structural analysis

The final structure of ChuX contains four molecules in the asymmetric unit, with two sets of monomers interacting to form two homodimers (Figure 4.2). Due to the absence of clear electron density for a beta turn, a gap in the structure exists in molecule B (residues 90-93). The center of each dimer is composed of two sets of nine antiparallel \( \beta \)-sheets, which bow outward to form an overall saddle motif. Each set of \( \beta \)-sheets are made up of six strands contributed by one of the ChuX monomers and three others contributed via a domain swap from the partnering ChuX molecule. Two \( \alpha \)-helices flank the small gap that is formed as a result of the bowing \( \beta \)-core, and the N-terminal 36 residues of each ChuX monomer flank each end of the central \( \beta \)-core in an \( \alpha \)-loop-\( \alpha \)-
loop-α motif. Each monomer couple interacts across equivalent polar residues from the core sets of β-sheets, forming two pairs of closely associated dimers. Molecular weight evaluation by gel filtration and dynamic light scattering suggests that ChuX forms a homodimer in solution (data not shown). Ramachandran analysis of the final ChuX structure shows 488 residues (88.1%) within the most favored region, 55 (9.9%) in additional allowed, 10 (1.8%) in generously allowed, and only 1 residue (0.2%) in the disallowed region (K11 from molecule C). Crystallographic statistics for diffraction data and final structural refinement are presented in Table 4.1.
Figure 4.1 Sequence alignment of ChuX with homologues from various Gram-negative bacteria. Highly conserved residues are highlighted in red, residues with sequence similarity are presented in blue. Numbering corresponds to ChuX. Conserved histidines across the organisms examined at positions 65 and 98 are underlined.
Table 4.1. Diffraction data and refinement statistics for ChuX

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>Native</th>
<th>Se-Met</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P4_1</td>
<td>P4_1</td>
</tr>
<tr>
<td>Cell Dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) and (b) (Å)</td>
<td>76.56</td>
<td>76.83</td>
</tr>
<tr>
<td>(c) (Å)</td>
<td>140.86</td>
<td>141.64</td>
</tr>
<tr>
<td>(\alpha), (\beta), and (\gamma) (°)</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>No. of molecules in ASU</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.1000</td>
<td>0.9795</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50 – 2.05</td>
<td>50 – 2.20</td>
</tr>
<tr>
<td>Total reflections</td>
<td>359,110(^a)</td>
<td>529,767(^a)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>50,833(^a)</td>
<td>82,255(^b)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100.0 (100.0)(^#)</td>
<td>98.6 (94.7)(^#)</td>
</tr>
<tr>
<td>(R_{sym}(I)) (%)</td>
<td>4.3</td>
<td>4.0</td>
</tr>
<tr>
<td>(I/\sigma I)</td>
<td>28.3 (1.7)(^#)</td>
<td>45.3 (6.0)(^#)</td>
</tr>
</tbody>
</table>

| Refinement Statistics                  |                 |                |
| Resolution Range (Å)                   | 50 – 2.05       |                |
| \(R_{work}\), (%)                     | 20.5            |                |
| \(R_{free}\), (%)                     | 27.4            |                |
| No. reflections total/ \(R_{free}\)   | 47915/ 3137     |                |
| No. atoms, protein / solvent (H\(_2\)O)/ total | 4833/ 360/ 5193 |        |
| B-factors (Å\(^2\)) protein/ solvent (H\(_2\)O)/ total | 49.3/ 53.8/ 49.6 |        |
| RMS, bond length (Å)/ bond angle(°)    | 0.020/1.87      |                |

\(^a\)Values in parentheses are for the outermost shells (Native: 2.12–2.05 Å, SeMet: 2.27–2.20 Å)
\(^b\)Number of reflections following merging
\(^#\)Number of reflections without merging of Bijovet pairs
Figure 4.2 Ribbon diagram of the four ChuX molecules in the crystallographic asymmetric unit. Each of the four ChuX monomers that make up two pairs of ChuX homo-dimers are depicted with different colours. ChuX homo-dimerization occurs via interactions between equivalent residues of red (A) and blue (B), yellow (C) and green (D) monomers, respectively. Interactions between ChuX monomers occur across the central sets of β-sheets, involving a domain swap between associated ChuX couples. The gap in the structure is shown with a dotted line.
Structural comparison revealed that the ChuX monomer shares structural similarity with the unpublished hypothetical protein AGR_C_4470 (2HQV, 1.3 Å) from Agrobacterium tumefaciens and overall topology with the E. coli metal chelating enzyme glyoxalase I (57) (1FA8, 3.8 Å). Furthermore, an excellent structural alignment exists between the ChuX dimer and the monomeric structures of two other members of the heme utilization operon, ChuS and HemS. ChuX aligns with these protein structures with rmsd values between 2.4 and 2.8 Å, depending on which ChuX dimer is used and whether the apo- or holoenzyme structure of ChuS or HemS is used (Figure 4.3A). This structural similarity was surprising given that ChuX shares less than 18.3% sequence identity with either the N- or C-terminal sequences of ChuS or HemS. Within each ChuX dimer, a pair of α-helices together with several segments of the β-core, delineate two sets of broad clefts. These clefts are equivalent to the heme binding positions demonstrated for ChuS (137) and HemS (120) (Figure 4.3B).

In both ChuS and HemS the histidine side chain responsible for heme coordination originates from one of the flanking α-helices. An equivalent residue is, however, absent in the structure of ChuX. Instead, the two conserved histidine residues, H65 and H98, originating from elements of the central β-core, are proximal to these structurally conserved heme binding clefts (Figure 4.3B). These histidines are contributed from opposite ChuX monomers, suggesting that domain swapping is involved in heme coordination similar to that characterized for HasA (35). The first three residues of the conserved sequence K134 to L145 reside on the medial of the three β-sheets contributed in the domain swap to the β-sheet core and may therefore be important for dimerization. This region of ChuX is also at the base of the conserved heme binding cleft, with K134
extending to only 3.76 Å from H65, potentially contributing to the stabilization of the imidazole group for iron coordination. The other residues of this conserved segment reside on a loop that delineates one edge of the heme binding cleft and may therefore be involved in guiding heme insertion, or stabilizing the heme moiety once it is coordinated by ChuX.
Figure 4.3 Structural similarity of ChuX with ChuS and putative heme binding cleft. A) Superimposition of ChuX (blue) and ChuS-heme (red) giving an rmsd of 2.8 Å. Despite low sequence similarity, the topology formed by the ChuX dimer is akin to the structural repeat of the ChuS monomer. Notably, the α-helix responsible in ChuS for heme coordination via H198 has shifted away from the heme binding cleft (*). B) Ribbon diagram of the ChuX homo dimer oriented to highlight the differences between the open cleft (top) and closed cleft (bottom) conformations. The putative heme binding clefts are emphasized with arrows. The conserved histidines shown to be responsible for heme coordination through mutagenesis and subsequent spectral analysis are highlighted (H65 in green, H98 in purple). The α-carbon positions of the central set of β-sheets used to generate the beta mutant form of ChuX in the inhibited dimer form are highlighted in green.
4.4.3 The ChuX structure may represent an open and closed conformation

Structural alignment of the two homodimers in the ChuX asymmetric unit revealed an interesting difference in that one of the four putative heme binding clefts is significantly more open compared to the other three (Figure 4.3B). In the three closed cleft conformations, the gap between α-carbons of H84 and M115 are 13.3, 15.3 and 15.5 Å across, whereas the equivalent gap distance of the open cleft is 23.3 Å. This difference is largely due to a 3.2 Å shift of the α-helix delineated by residues 16 to 26 that define the lateral edge of the cleft. In ChuX, this helix is structurally equivalent to the helix responsible for clamping heme into place in ChuS and HemS, and which shifts upon heme binding by as much as 3.8 and 4.0 Å, respectively (120, 137). Crystal contact analysis (42) revealed relatively few contacts among symmetry-related molecules. Those that were observed interacted in a similar fashion with either ChuX homodimer, suggesting that the structural difference was not due to crystal packing.

The overall effect of the gap differential is that H65 is effectively buried at the base of the putative heme binding cleft in the closed cleft conformations, while H98 remains oriented into the cleft space. When heme was modeled into each of the four clefts, the docking positions were equivalent to those characterized for ChuS and HemS, with the central iron atom proximal to an axial H98 in each of the closed conformations, but was axial to H65 in the open cleft example. In either model, the docking energy predicted for heme binding by ChuX were favorable (93), with corresponding mean values for ΔG_{pred} of -9.9 and -11.4 kcal/mol for the open and closed clefts respectively. Furthermore, this docking potential could reflect how the difference in dimer juxtapositions between the
open and closed cleft conformations of ChuX could help mediate heme uptake in that H98 is important for heme coordination in the open cleft conformations while H65 is similarly important in closed conformations. Other residues located within the heme binding cleft which could contribute to heme stabilization include T17, E19, T61, L63, I70, E72, F114, K134, and F136. All of which, except T17, are conserved across sequence homologues of ChuX.

### 4.4.4 ChuX-heme association

Heme proteins have characteristic visible spectra that convey information about the structure and coordination state of their heme binding sites. The spectra of ChuX resemble those of other heme proteins in that a Soret maximum is evident at 404 nm (Figure 4.4A) depending on buffer pH, and accompanied by a smaller peak for the $\alpha$-band at 620 nm. The spectra of ChuX-heme is consistent with the formation of a ferric hexacoordinate histidine-imidazole complex (169). Monitoring the appearance of this peak during heme reconstitution to ChuX suggests that the association occurs rapidly, reaching its maximum within five minutes. Furthermore, differential spectroscopy of heme versus ChuX-heme suggests that the association is 1:1, implying that the ChuX homodimer binds two molecules of heme (Figure 4.4B).

Spectral analysis of the ChuX mutants indicate that H65L and H98D both affect heme binding as the Soret peak for each has shifted to 414 nm, with the absorbance intensity at 360 nm decreased, while the shoulders in the $\beta$-band region intensified compared to those of ChuX (Figure 4.5). The spectrum of the BetaX mutant is similar to
ChuX with the exception of a decrease in the magnitude of the Soret peak, suggesting a reduction in heme coordination. Notably, while the independent mutations of the conserved histidines did not significantly affect heme coordination by ChuX, the spectrum of DM ChuX reflects a reduction in heme coordination due to the elimination of a clear Soret peak, as well as a reduction in absorption of the $\alpha$- and $\beta$-bands.
Figure 4.4 Spectral analysis of ChuX reconstituted with heme. A) A clear Soret peak is visible at 404 nm for reconstituted ChuX-heme purified by size exclusion chromatography, suggesting that the ChuS-heme complex formed is ferric hexacoordinate histidine-imidazole in the high-spin state at neutral pH (52). B) Differential spectroscopy of heme versus ChuX-heme at 404 nm. A transition in the slope of the line of fit for ChuX-heme inflection point is evident for ChuX to heme at a ratio of 1:1. At higher heme concentrations the slope of the line changes to one similar to that observed for free heme, suggesting that ChuX has become saturated.
Figure 4.5 Spectra of ChuX and mutants reconstituted with heme. The region between 500 and 700 nm has been amplified four-fold to highlight peak differences. The spectrum of ChuX-heme resembles that of other heme proteins in that a Soret maximum is evident at 404 and an \( \alpha \)-band at 620 nm, consistent with the formation of a ferric hexacoordinate histidine-imidazole complex (169). Mutations H65L and H98D both affect heme binding, evident in the shift in the wavelength of the Soret peak to 414 nm. The spectrum of the BetaX mutant has a decrease in the magnitude of the Soret peak, suggesting reduced heme coordination. Notably, while the H65L and H98D independent mutations did not significantly affect heme coordination by ChuX, the spectrum of DM ChuX reflects a reduction in heme coordination due to the elimination of a clear Soret peak, as well as a reduction in absorption of the \( \alpha \)- and \( \beta \)-bands.
### 4.4.5 ChuX heme binding affinity

Based on absorbance at 404 nm during ChuX-heme reconstitution, the estimated $K_d$ is $1.99 \pm 0.02 \, \mu M \, (n = 2)$ as calculated from the following expression:

$$
\Delta A = (\varepsilon_b - \varepsilon_f)0.5\left[ P_0 + K_d + H_0 - \left( \{ P_0 + H_0 + K_d \}^2 - 4 \cdot P_0 \cdot H_0 \right)^{1/2} \right]
$$

where $P_0$ is the amount of protein, $H_0$ is the concentration of heme, and $(\varepsilon_b - \varepsilon_f)$ is the difference in the Soret absorption spectra between the bound and free states. The estimated $K_d$ for ChuX-heme complex formation suggests a weaker heme association for ChuX than that characterized for other heme binding proteins such as the mammalian hHO-1 ($0.84 \pm 0.2 \, \mu M$) (159, 161), HasA of $5.0 \pm 3.1 \, nM$ (85), and ChuS ($1.0 \pm 0.3 \, \mu M$) (139), but tighter than that reported for HmuO ($2.5 \pm 1.0 \, \mu M$) (15), IsdG ($5.0 \pm 1.5 \, \mu M$) and IsdI ($3.5 \pm 1.4 \, \mu M$) (125).

### 4.4.6 ChuX protection from heme toxicity

Heme is a potentially toxic molecule due to its ability to disrupt normal redox processes and catalyze free radical formation (54, 101). As a result, uncommitted heme concentrations are normally maintained below $10^{-9}$ M (141). When challenged with low concentrations of heme ($2 \, \mu M$), LB-ampicillin cultures expressing ChuX were protected against heme toxicity and grew at an enhanced rate relative to those expressing the DM ChuX mutant or the C1528 control (Figure 4.6A).
When an equal amount of ChuS was added to ChuX reconstituted with heme, a small spectral shift in the Soret peak was evident from 404 to 408 nm (Figure 4.6B). This Soret peak shift indicates that heme was being released from ChuX and taken up by ChuS. Furthermore, when ascorbic acid, an electron donor for heme degradation was added, a decrease in the spectra for heme coordination was observed compared to either ChuX-heme or free heme, indicating that heme was degraded as previously characterized (Figure 4.6B) (139). However, because the ChuX homodimer is similar in shape, size, and molecular weight to the ChuS monomer, as well as having similar Soret peaks, it was difficult to evaluate the specific amount of heme transfer (14, 80).
Figure 4.6 ChuX protection against heme toxicity and heme transfer by ChuX. A) Progression of BL21(DE3) cells expressing ChuX, DM ChuX, and C1528. ChuX, DM, C1528 were each challenged with 2 μM heme. Cells expressing ChuX were protected relative to those expressing the ChuX double mutant or C1528. B) Transfer of heme from ChuX to the heme degrading enzyme ChuS. An initial increase in the peak for ChuX and ChuS shows that ChuS is becoming reconstituted with heme. Following ascorbic acid addition ChuS catalyzed degradation of heme was then tracked by monitoring the decrease in the Soret peak at 408 nm.
4.5 Discussion

Here we present the first crystal structure of ChuX, a member of the heme utilization operon from pathogenic *E. coli* O157:H7, and propose that ChuX may help to regulate heme utilization through structure-mediated heme binding and release. Heme is a hydrophobic molecule capable of freely diffusing into membranes where it can alter the bilayer structure and disrupt cell integrity (119). Furthermore, heme is a potent prooxidative, promoting the light-dependent formation of reactive oxygen species that can cause non-enzymatic redox reactions (54). Organisms must therefore tightly regulate heme acquisition and metabolism by either degrading heme or storing it for later use in various processes (155). Largely through gene knockouts and phenotypic characterization the transport machinery responsible for heme uptake in *E. coli* O157:H7 has been elucidated (127, 128). However, the regulation of internalized heme metabolism has only recently received attention. Furthermore, in order to avoid iron induced toxicity, heme degradation and iron storage must also be balanced. This has been shown to occur through the regulation of iron scavenging elements by transcriptional regulator proteins such as Fur (168), DtxR (116), and FecI (40). Alternatively, iron toxicity can also be avoided by reducing the amount of heme degraded within the cytoplasm through heme storage or short term sequestering until the organism is experiencing iron limiting conditions (55). A heme storage role has recently been suggested for the gene products *ght* from *Neisseria meningitides* (109) and hmsT from *Yersinia pestis* (66), as well as the protein HutZ from *Vibrio cholerae* (166). Our results suggest a cytoprotective role for ChuX based on bacterial cultures challenged with low levels of heme relative to ChuX.
mutants or non-heme processing controls. Furthermore, we demonstrated that ChuX is capable of transferring heme to other heme utilizing proteins such as ChuS, although we cannot discount at this time that heme transfer may have occurred through passive diffusion.

Structural insights provided from vibrational spectroscopy (EPR and Raman), NMR, and X-ray crystallography have indicated that heme proteins often coordinate the iron center of the heme group between a histidine side chain and a variety of other possible residues including tyrosine (8), methionine (22), cysteine (39), proline (81), histidine (103), arginine (137); some heme proteins can even coordinate their heme iron centers via a sole histidine sidechain without additional sidechain assistance (52, 121). Other proximal side chains and polypeptide backbones have been shown to interact with the propionic and vinyl parts of the heme group, contributing to the stabilization of the heme moiety within the protein environment which ultimately confers heme protein function. The structure of the dimeric oxygen sensing kinase FixL resembles an open pita or clam which coordinates heme in the center of the open mouth between a series of antiparallel β-strands arranged in a β-barrel, and an α-helix (121). This mode of heme coordination is also similar to the topology utilized by ChuS and HemS in which the central iron atom of heme is sandwiched between a histidine side chain and an arginine side chain via two water molecules (120, 137). The structural similarity shared between ChuX and ChuS/HemS suggests a similar mechanism of heme coordination. However, the axial histidine side chain responsible for heme coordination in ChuS/HemS was notably absent when the structure of ChuX was superimposed with either ChuS or HemS.
Spectral analysis of ChuX reconstituted with heme suggested that a histidine residue was involved in ChuX heme coordination and that heme coordination occurred in a 1:1 fashion. When the ChuX sequence was aligned with homologues from various Gram-negative bacteria, the conserved residues H65 and H98 were identified as potential heme coordinating residues. In the crystal structure of ChuX, both of these residues are within the putative ChuX heme binding cleft, but single residue mutants that changed these histidine residues did not abolish heme binding. However, spectral analysis of the ChuX double H65/H98 mutants demonstrated that (i) either both residues may be involved in heme coordination; (ii) that alternate histidine residues may be involved; or (iii) that heme binding sites within ChuX may be overlapping. This variability in protein-heme coordination has been detailed for the extracellular hemophore HasA from *Serratia marcescens*, which is also a domain swapping dimer (85). Lastly, the spectra of heme reconstituted with the BetaX mutant, suggested that under conditions where ChuX dimerization was impaired, heme coordination was compromised. Collectively, based on the fact that the putative heme binding sites of ChuX are contributed by two ChuX molecules, the identification of conserved residues and sequence segments through sequence homology comparison, and the evaluation of mutant ChuX variants, we conclude that ChuX dimerization is absolutely obligatory for coordinating heme molecules and that the two conserved residues (H65 and H98) located within the conserved heme binding cleft are both essential for heme coordination.

Two different ChuX homodimer conformations were observed within the ChuX asymmetric unit. One ChuX dimer assumed a structure with two similarly spaced gaps across the conserved heme binding cleft, while the second dimer had one gap similar in
size to those found in the first dimer conformation as well as a noticeably broader one. We propose that this structural plasticity may reflect an open, ligand accepting form and a closed, ligand bound form of ChuX. Similar protein flexibility has been shown to be important for hemoglobin to reversibly regulate heme coordination (53). We found ChuX-heme coordination to be conformation-dependent: when heme was docked to the closed conformation of ChuX, heme coordination was observed to cluster around H65 with favorable energy predictions, whereas when heme was docked to the open conformation of ChuX, heme coordination was found to cluster around H98. Furthermore, many conserved residues in the ChuX sequence are clustered within and proximal to the heme binding cleft. Together, the variation of ChuX juxtapositions in homodimers as well as the docking of heme to ChuX at the two conserved histidine residues H65 and H98 located within the conserved heme binding pocket, suggest a potential structural mechanism for heme binding and sequestering.

In conclusion, based on the crystal structure of ChuX we have proposed that dimer juxtaposition may serve as a structural mechanism in regulating heme binding and that ChuX may function as a cytoplasmic heme shuttling protein, thereby regulating heme processing through sequestering or storage.
4.6 Footnotes

Thank you to Dr. Michael Nesheim for his expertise in designing and interpreting heme binding studies and John Wagner for the generation of the H98D mutant. We thank Dr. Mirek Cygler and Dr. Allan Matte for their support. We are also grateful to Brookhaven National Laboratory beamline staff Leonid Flaks and Martin McMillan (X8C) as well as Anand Saxena (X12C), where the SAD and native data for ChuX were collected respectively. This work has been supported by Bracken E.G. Bauman and R. Samuel McLaughlin Queen’s Graduate Fellowships as well as CIHR and OGS. ZJ is a Canada Research Chair in Structural Biology. Coordinates and structure factors for ChuX are deposited in the RCSB Protein Data Bank (www.rcsb.org) under accession code 2OVI.
Chapter 5

General discussion and summary
5.1 Gaining functional information from protein structures

With the completion of sequencing of many microbial and eukaryotic genomes, the essential next steps toward understanding the molecular mechanisms and interactions that ultimately facilitate cellular function. To do this, one must first identify the genomic complement of ORFs, and then to determine the structures and functions of the proteins encoded. Historically, the first stage of ORF characterization has involved primary sequence homology searches to infer functionality from previously annotated proteins or segments of proteins. However, it has become obvious that current sequence homology-based functional annotation alone is not sufficient for complete genome functional coverage. During translation, protein sequences fold into conformations that serve as scaffolds for residues whose interactions establish the microenvironments for directing localization as well as protein-protein and protein-ligand interfaces that ultimately determine a protein’s function. Consequently, the global interactions present in the final fold of the protein can be better conserved than its sequence alone. Protein structure comparisons may therefore have the potential to reveal evolutionarily distant relationships that may be used for functional annotation through the identification of active sites and integral functional residues, suggesting potential ligands, and may also contribute information as to possible mechanisms of action.

Whereas traditional structural studies have often involved targeting proteins of known function in order to completely understand their mechanism of action, our efforts simultaneously combined structure determination and functional characterization of two proteins, ChuS and ChuX, which had little functional information attributed to them at the initiation of our investigation. To successfully establish this concurrent structural
determination and functional annotation methodology, a set of criteria must be established to maximize the allocation of resources. The most important of these is target selection; to highlight proteins from organisms of medical or functional interest that may be functionally significant in cell operation. Initially, the evaluation of putative genes relies on sequence comparisons and homology searches as well as investigations of the partial or complete functions attributed to genes organized into operons. This process was facilitated for ChuS and ChuX through the use of sequence analysis tools such as the Clusters of Orthologous Groups (142). Protein selection may also enrich for proteins present in pathogenic organisms which do not have homologues in more benign strains. ChuS and ChuX are genes encoded within the heme utilization operon from *E. coli* O157:H7 and CFT073, but not in the benign K-12 strain. Depending on project methodology, protein selection could also encompass well characterized proteins in order to better understand their mechanism of action or the nature of their interactions with ligands, inhibitors, and/or binding partners.

Lastly, protein selection must also apply bioinformatics-based analysis to address the inherent challenges of X-ray crystallography. As a large number of potential targets may be initially selected, a systematic ranking of targeted protein priority must also be assigned (17). This may include an evaluation of protein solubility and cellular localization. Furthermore, protein analysis may consider whether full length proteins, truncations, or co-expression with other proteins should be implemented. Other important considerations such as whether inhibitors or supplements to culture media may be necessary, as well as special considerations into growth conditions such as whether special expression systems may be better suited for certain proteins of interest. This latter
analysis may also be applied retroactively once difficulties are encountered during various phases of the investigation.

Once the structures of ChuS and ChuX had been determined, in silico functional prediction proceeded via three structure-based methods simultaneously due to the inherent strengths and weaknesses attributed to each (114). These levels of structural analysis were: (i) structure homology and overall topology searches, (ii) putative cleft identification, and (iii) cleft and surface charge distributions. Together these computational methods for probing structure have the potential to provide functional information at phenotypic, cellular and molecular levels (114). Conveniently, many of the tools used for fold and sequence alignment, structure modeling, model evaluation, and META servers are implemented as web servers, and are readily available to the public.

In structure homology searches new structures may be compared against a database of structures with or without functional annotation, or a sample of structures from the protein data bank. As the number of experimentally determined structures increases, the application of homology based or comparative modeling will likely become more frequent in structural evaluation, as it provides speed, accuracy, and detail that other modeling programs lack (94). Evidence suggests that there are a limited number of different classifications of protein folds (49). It may therefore be possible to assign specific fold and function classifications, patterns of phylogenetic occurrence, expression data, and suggest protein-protein interactions from architecture homology searches (49). Several current search engines such as CATH, CE, DALI, DEJAVU, MATRAS, PALI, SCOP, SUPFAM, and VAST provide the opportunity for this type of analysis, and have the added benefit of increasing their value as more structures have their function
annotated (49, 98, 140). Other comparisons of closely related structure families can be performed using the servers CASP, CRABP, and DDMP, which can suggest substrate specificity or protein-protein interactions. A recent evaluation of these class of databases suggested that CE, DALI, MATRAS, and VAST showed the best performance, but understandably none of the servers achieved a 100% success rate (98). For both ChuX and ChuS, a DALI search was the most informative, but only when the homologous structures were subsequently evaluated in greater detail; including sequence and structural investigation.

The second method for predicting function would be to use software such as pvSOAR, DOCK, GRASP, LigPROF, PROFUNC, PROSITE, SLIDE, SUMO and SURFNET that compare highlighted regions against catalogues of chemical structures and functional mechanisms (49). Akin to randomly fitting molecular pieces into a protein jigsaw puzzle, this method involves docking libraries of molecules to regions on the “virtual screening” structure. When the program DOCK is applied, factors such as rigidity, energy, contact, and solvation scorings are considered to determine how well a molecule fits sites on the structure. For both ChuS and ChuX, pvSOAR correctly predicted that both were heme binding proteins, and even suggested putative residues responsible for heme coordination. However, in both cases the correct pvSOAR predictions were not the top results, or even in the top 10. This experience highlights that where some functional information has already been acquired, it can facilitate the processing of the vast amount of information output during *in silico* characterization. An advantage of PROFUNC is that it combines multiple analysis of a protein structure in parallel including: a prediction of cellular localization, sequence and structure alignments,
and cleft identification. PROFUNC therefore provides the advantage of a complete summary of sequence and structural based analysis in a single output, which facilitates comparisons and may highlight more distant relationships that would otherwise be overlooked.

In the final technique, a system such as principal component analysis would be used to compare subtle surface and cleft characteristic differences across orthologous groups of structures (24), such as our analysis of apo ChuS with heme-ChuS, and ChuX with ChuS and HemS. Key regions of conformational changes were highlighted, in addition to consensus pockets and surface regions for interactions. These analyses were particularly informative in suggesting the position of heme coordination in ChuX and potentially suggested a mechanism of heme uptake. More broadly applied, this method could provide information concerning substrate specificity and inhibitors that could be tested in binding studies, applied in co-crystallization experiments, or in catalytic investigations with the potential to postulate molecular mechanisms for action (24). However, this method is often more computationally expensive than the previous two; requiring time and computing power, but has the potential to yield a more complete characterization of a class of structures with careful analysis.

With the initiation of structural genomics projects, many examples of structure based functional annotation have emerged. The crystal structure of a previously uncharacterized H. influenzae protein YbaK was reported and through the implementation of combination of sequence and structure homology searches to demonstrate that YbaK was a tRNA synthetase involved in nucleotide/oligonucleotide binding (171). Alternatively, structures of previously uncharacterized proteins may
provide insight into their function through the shape and location of the active site pockets or through the presence of unexpected cofactors in the solved structures. The structure of MJ0577 was solved with an ATP molecule bound and shared the corresponding binding motif with a family of ATP binding structures (170). These attributes strongly suggested that the protein was an ATPase or ATP-mediated molecular switch, which was later confirmed through biochemical experimentation. Furthermore, obtaining protein structures that share sequence homology to families of proteins with annotated function may complete our understanding of their molecular mechanisms, highlight key regions important for function and substrate specificity, or reveal novel mechanisms of action. An example of this last class of structures was demonstrated for the G-domain of Rnd3/RhoE, a protein member within a subfamily of Rho GTPases (43). The structure of Rnd3/RhoE shared a similar fold to other Rho GTPases but revealed striking differences with respect to charge distribution, GTPase center, ligand binding sites, and interfaces for known regulators and effectors. Together these examples, as well as the plethora of positive cases across a variety of prokaryotes, eukaryotes, and archaeabacteria, demonstrate the feasibility, pitfalls, and potential impact for functional assignment following structural solution of ORFs and poorly annotated proteins (10).

5.2 ChuS and ChuX: structural and functional implications for heme utilization

Due to their localization within the heme binding operon, and phenotypic evaluation of gene knockouts, a putative function had been ascribed for ChuS and suggested for ChuX prior to the initiation of our studies. It was therefore important for
our investigation to develop testable hypotheses as to particular mechanistic and \textit{in vitro} functions for these two proteins. Primarily, information was gathered based on literature evaluation and through the incorporation of as much bioinformatics-based information as possible into the study. In order to process the large amount of computational based data generated, a combined approach of bioinformatic analysis, experimental observations, and structural characterization improved our overall analysis. In the case of ChuS, biochemical clues were revealed prior to a successful structure determination. During ChuS over-expression, a blue-green pigment developed and persisted in fractions containing ChuS following affinity purification (139). As ChuS is organized into an operon responsible for heme uptake and utilization, the formation of this pigment was suggested to be the result of the production of biliverdin (BV) or bilirubin (Figure 1.1B). Furthermore, the colour causing pigment was eliminated following subsequent purification indicating that the pigment producing agent was not tightly coordinated or covalently bound by ChuS. These observations suggested that while bilirubin or biliverdin may be responsible for the observed pigment produced during ChuS over-expression, neither was the preferred substrate or ligand for ChuS.

Concurrent with efforts to obtain the structure of ChuS, we initiated experiments to reconstitute ChuS with heme and conducted spectral analysis of the protein. When hemoproteins coordinate heme, a characteristic spectrum results that is dependent on the nature of the protein-heme interactions. The spectra of ChuS reconstituted with heme has a Soret maximum at 408 nm, with a smaller set of peaks at 545 and 580 nm, suggesting that the ChuS-heme complex formed was ferric hexacoordinate and that a histidine was responsible for this coordination (26). A comparison of the ChuS-heme spectra with other
hemoproteins revealed similarities with HOs (26, 158). HOs use molecular oxygen together with an electron donor such as cytochrome P450 reductase-NADPH or ascorbic acid to degrade heme. When either electron donor was added, the spectra of ChuS-heme changed over time in a similar fashion to that characterized for other HOs, producing BV, CO, and free iron. Furthermore, we used sensitive gas chromatography to measure CO production during heme degradation, confirming that ChuS was a HO. In order to confirm that this change was not due to non-enzymatic coupled oxidation, catalase and superoxide dismutase were also added prior to the addition of either electron donor and similar trends in spectral analysis and CO measurement were observed, demonstrating specific HO activity.

Concurrent with ChuS-heme interaction study, the structure of the apo-ChuS had initially been solved by our group prior to experimental characterization of ChuS to be a HO. However, as ChuS represented a novel fold at that time, structural information alone failed to suggest any immediate functional information. Based on the appearance of a Soret peak in heme coordination studies, we postulated that an axial histidine side chain may be responsible for heme coordination. An examination of the structure of ChuS in light of highly conserved residues from this protein family, revealed four candidate histidine residues (H73, H87, H198, H277), all in proximity to two broad clefts. These clefts in ChuS are delineated on opposite sides of the central core of β-pleated sheets, with the third side of the clefts delineated by the flanking sets of three α-helices (Figure 2.1). The exception was H73 which was proximal to the putative cleft, but was oriented toward the interior of the structure. A careful examination of the ChuS structure revealed that ChuS is a structural repeat (Figure 2.2). This would not have been predicted from
sequence analysis, as the N- and C-terminal halves of ChuS share low sequence identity (139). Notably, the structure of ChuS does not share any structural homology to characterized heme degrading enzymes, which are mainly α-helical and coordinate heme by sandwiching it between a histidine residue and a glycine, such as observed in hHOs (structures reviewed in (105)), but has since been determined to share homology with ChuX, HemS, AGR_C_4470, and glyoxalase I.

In an attempt to identify the key heme coordinating residues, we initiated mutagenesis of these conserved histidine residues. Simultaneously, we obtained two isoforms of ChuS-heme co-crystals, which led to structure determination of ChuS in complex with heme and provided a detailed view of the heme environment. Spectral analysis and CO detection by gas chromatography of native ChuS together with H73A and H193N mutants was conducted. Together, the mutagenesis study and the structural solution of ChuS-heme indicated that H193 was important for axial ligation of heme, and that R100 further stabilized the heme group from the medial side of the protein via two water molecules. Our initial hypothesis was that ChuS bound to two equivalents of heme. Despite that the conserved histidine 277 was in proximity to the second putative heme binding cleft it originates from the central set of β-sheets, and not the α-helix shown to be responsible for heme coordination by ChuS. This difference between the two clefts may explain why ChuS and HemS have been shown through crystal structure analysis to only bind one equivalent of heme (120, 137). Furthermore, we must amend our evaluation of the $K_d$ for ChuS-heme to $0.4 \pm 0.1 \mu M$ from our previously estimated $K_d$ of $1.0 \pm 0.3 \mu M$.

In summary, ChuS represented a novel fold; it was the first HO to be identified in any strain of *E. coli*, and is therefore an important factor in heme utilization as an iron
source. Furthermore, the mode of heme coordination was unique compared to other heme degrading enzymes which potentially could be related to the increased activity observed for ChuS relative to other bacterial enzymes.

The axiom of sequence based structure and function prediction is that similar sequences have similar folds. The corollary of this statement, that dissimilar sequences will fold into slightly different structures, has been shown to be untrue in the cases of AGR_C_4470, ChuS (137, 139), ChuX (138), and HemS (120). Despite sharing low sequence similarity, all four structures have very similar topology. Furthermore, as the monomeric structures of ChuS and HemS resemble the homodimer formed by ChuX and AGR_C_4470, this could imply an evolutionary relationship where ChuS has evolved from a chuX gene duplication event. Absorption spectral analysis of ChuX reconstituted heme demonstrates that ChuX binds to heme in a 1:1 manner, implying that each homodimer coordinates two heme molecules, in contrast to ChuS where only one heme molecule is bound. Due to the structural similarity observed between the two proteins, we extended the mechanism of ChuS heme binding to ChuX, and suggested the putative location in the protein responsible for heme coordination. Based on sequence comparisons of ChuX and its homologues, as well as structural comparisons with ChuS and HemS, we designed a number of site-directed mutations to probe heme binding sites as well as the dimer interface. In ChuS a single His mutation abolished the capacity to coordinate heme, but in ChuX, two mutations (H65L and H98D) were required to achieve the same result. Further, dimer interface mutations resulted in compromised heme binding, supporting the notion that varying ChuX juxtapositions observed in the structure may serve to modulate heme binding and release. We also demonstrated that
ChuX can transfer heme to ChuS for degradation although currently we cannot exclude the possibility that this transfer is occurring through diffusion. Furthermore, we demonstrated that ChuX expression had a cytoprotective role against heme toxicity. In conclusion, ChuX appears to act upstream of ChuS and may regulate proper heme uptake and utilization through ChuX-mediated heme binding and release.

5.3 Inhibition of heme utilization as a potential mode for treatment

HOs catalyze the degradation of heme to BV, CO, and free iron. In bacteria, this liberated iron can be incorporated into various proteins where it can act as a cofactor in many reactions that contribute to pathogenesis. Beyond a role in iron metabolism, it has also been revealed that the two other products of oxidative heme cleavage, BV and CO, are important signaling molecules in mammalian systems. BV is subsequently reduced by BV reductase, forming BR, a potent cellular antioxidant (12, 113). While excessive levels of BR cause toxicity, adequate levels play an important role in cytoprotection form oxidative stress (69). Furthermore, approximately 85% of the physiologically derived CO is produced through HO-1 mediated heme degradation (112). CO has been shown to act as an important signaling molecule that plays a role in modulating vasodilation, neurotransmission in the central or peripheral nervous systems, immune responses, and anti-inflammatory, anti-apoptotic, and/or anti-proliferative potentials (7, 36, 75). In addition, HO-1 activity has been shown to be important for tumor growth, implicating a role for HO-1 inhibitors as anticancer agents (37).
Because of the broad range of effects mediated by the products of heme degradation, HO-1 specific inhibitors have the potential to modulate the function of many physiological systems. Traditional HO-1 inhibitors have involved the use of various metalloporphyrins, where the central iron atom of heme is replaced with various other metals including chromium, tin, and zinc (7). However, these heme derivatives will also interact with other important hemoproteins such as nitric oxide synthase and soluble guanylyl cyclase (76). New classes of HO-1 inhibitors have emerged that are based on azalanstat and azole (76, 151), as well as targeted delivery of HO-1 inhibitors (65). Investigation into these class of compounds include the structure and biochemical properties of a ternary complex of rat HO-1, heme, and an imidazole-dioxolane compound (2[2-(4-chlorophenyl)ethyl]-2-[(1H-imidazol-1-yl)methyl]-1,3-dioxolane) was recently described (133). The usefulness of these classes of chemicals in HO inhibition stems from their similarity to the natural imidazole group of the preferred coordinating histidine residue. Thereby contributing additional stabilization of the heme moiety by interacting with the iron atom and shielding the heme group from the oxygen and electron addition required for heme degradation.

We have shown that in the presence of Sn(IV) mesoporphyrin IX, a competitive inhibitor of heme binding and degradation (152), the activity of ChuS decreased between 2- to 7-fold in a dose-dependent manner (139) (Figure 2.7). This level of inhibition was similar to that characterized for HO-1 (152). However, when we added the HO-1 specific inhibitor (2[2-(4-chlorophenyl)ethyl]-2-[(1H-imidazol-1-yl)methyl]-1,3-dioxolane) to ChuS reconstituted with heme, we did not detect any inhibition during spectral analysis (Figure 5.1). The inability of this compound to inhibit the activity of ChuS, is likely due
to the structural differences between ChuS and HO-1, and their different modes of heme coordination and overall stabilization. From a therapeutic perspective these differences may potentially be exploited to specifically target the mechanism of iron acquisition mediated by ChuS during *E. coli* O157:H7 infections, while not disrupting normal HO-1 function and signaling. As the structure of ChuS has been solved in complex with heme, it may be used for inhibitor design and *in silico* screening. This could involve parallel screening of the existing catalogue of azalanstat andazole inhibitors of HO-1, or could involve the synthesis of unique inhibitors that potentially would interact with ChuS-heme.
Figure 5.1 Heme degradation by ChuS in the presence of the hHO-1 specific inhibitor (2-[2-(4-chlorophenyl)ethyl]-2-[(1H-imidazol-1-yl)methyl]-1,3-dioxolane). Reaction progression by monitoring the decrease in the Soret peak (408 nm) for ChuS were similar in the presence or absence of the hHO-1 inhibitor.
5.4. Conclusion

At the initiation of our experiments, the proteins involved in heme uptake through the OM, periplasm, and IM had been identified largely through genetic knockouts (91, 127, 128). However, the fate of heme once it reached the cytoplasm was lacking in detail. It was recognized that heme was a signaling molecule that would simultaneously stimulate increases in DNA splicing, transcription, mRNA stability, translation, and post translational modifications (154) and, therefore, the cumulative effect of heme uptake was effectively a signal for successful colonization of a host (83). While many bacterial heme degrading enzymes had been identified (130), and it had been hypothesized that a heme binding or storage protein also existed in *E. coli* (158, 165), the genes responsible had not been identified. We have therefore contributed both structural and functional insight for two proteins involved in cytoplasmic heme processing in pathogenic *E. coli* O157:H7, and can extend our findings for ChuS catalyzed degradation of heme for the production of antioxidants and as a source of iron. Based on conclusions drawn from this structural and function insight, we can therefore augment the structures and mechanism of heme uptake in pathogenic *E. coli* O157:H7 (Figure 5.2). Having demonstrated that homologues to ChuS and ChuX are present in many Gram-negative bacteria, including *Enterobacter, Erwinia, Shigella*, and *Yersinia*, we can extend our findings for these two proteins across various organisms with respect to heme utilization. Due to the implications of heme degradation for iron acquisition, and heme binding as it pertains to prevention against heme toxicity and proper utilization of heme, our characterization of these two proteins has contributed to our overall understanding of the pathogenesis of *E. coli* O157:H7.
Future work will involve characterization of the interactions of the cytoplasmic proteins ChuS, ChuW, ChuX, and ChuY and their roles in heme processing and utilization in \textit{E. coli} O157:H7 or its homologues. Through improved characterization of the heme utilization proteins and the interplay in mediating heme utilization, we will likely be able to extend our findings in bacterial systems to fungal, protozoan, and mammalian heme transports systems, which at this point are poorly understood. Follow-up investigations of \textit{E. coli} O157:H7 system may involve tracking heme internalization (82, 163) as well as a characterization of protein expression profiles during heme limiting and heme surplus conditions. In order to characterize the protein-protein interactions involved a synthetic heme derivative may be used where two heme molecules are covalently attached by a linker. This may be used in a type of tethered pull down experiment to explore possible protein-protein interactions. Conversely, understanding the structural interplay and modes of heme coordination in infectious bacteria may also be exploited to mediate medical treatments through screening of chemical compounds that may interact with these proteins. As the structures of ChuS and ChuX revealed unique modes of coordination compared to mammalian hHO-1, structure based drug design may be viable to specifically inhibit Gram-negative bacteria. At this time these azole based inhibitors seem to be the most suitable candidates for inhibiting heme uptake.
Figure 5.2 Modified schematic of the heme assimilation system from Gram-negative bacteria and organization of the heme utilization operon from *Escherichia coli* O157:H7. Proteins involved in heme transport from pathogenic *E. coli* O157:H7 are presented in their locations in the outer membrane (OM), periplasm (PP), and inner membrane (IM). Where the structures of proteins are unknown, structures of homologues have been included and names followed by a question mark. Putative promoter elements within the heme utilization operon are depicted with arrows. The structure and function of ChuS, ChuS-heme, and ChuX have been included, while the function of ChuW and ChuY is still unknown at this time.
Reference List


67. **Kabsch, W.** A solution for the best rotation to relate two sets of vectors.


78. Krissinel, E. and K. Henrick. 2004. Protein structure comparison in 3D based on secondary structure matching (SSM) followed by Cα alignment, scored by a new structural similarity function., p. 88. In Andreas J.Kungl & Penelope J.Kungl (Eds.) (ed.).


158. **Wilks, A.** 2001. The ShuS protein of *Shigella dysenteriae* is a heme-sequestering protein that also binds DNA. Arch Biochem Biophys **387**:137-142.

159. **Wilks, A., S. M. Black, W. L. Miller, and P. R. Ortiz de Montellano.** 1995. Expression and characterization of truncated human heme oxygenase (hHO-1) and a fusion protein of hHO-1 with human cytochrome P450 reductase. Biochemistry **34**:4421-4427.


Appendix I

Crystal structure of the conserved Gram-negative lipopolysaccharide transport protein LptA

Michael D. L. Suits and Zongchao Jia


Michael Suits was responsible for protein expression, purification, crystallization, data collection, structure solution and refinement. Dr. Christian M. Udell cloned LptA into the expression vector. Dr. Howard Robinson collected the LptA MAD diffraction data. Dr. Clemens Vonrhein helped with the application of autoSHARP that ultimately led to a structural solution. Dr. Michael Sawaya provided useful discussions on how to overcome anisotropy and developed the anisotropic correction server employed herein. This chapter was written by Michael Suits with editorial input from Dr. Zongchao Jia and with additional help from Jocelyne Suits.
A1.1 Abstract

LptA is an essential lipopolysaccharide (LPS) transport protein that is conserved across various Gram-negative bacteria, many of which are human and plant pathogens. LptA has been implicated in LPS transport from the inner membrane to the outer membrane, thereby contributing to maintaining cell envelope integrity and outer membrane assembly. Here we present the first crystal structure of the essential LPS periplasmic transport protein LptA in two crystal forms. This is the processed form of LptA which has been exported to the periplasm and has concurrently had the N-terminal 27 residues removed during the export process. The structure of LptA represents a novel fold which comprises a set of eight antiparallel \(\beta\)-sheets, bent in the middle to resemble a ribcage or \(\beta\)-jellyroll. The structure of LptA is not completely symmetrical along the length but opens slightly at both the N and C-terminal ends, creating two putative clefts. The N-terminal cleft is slightly larger than the C-terminal cleft, and is surrounded by a conserved segment of residues from P35-K83. Other conserved residues align along the length of the \(\beta\)-sheets, implicating a structural contribution. Severe anisotropy in diffraction data was corrected in order to obtain the final refined structure.
A1.2 Introduction

Gram-negative bacteria such as *Escherichia coli* are compartmentalized via an inner membrane (IM) and outer membrane (OM) into the cytoplasm, periplasm, and extra cellular environment. This compartmentalization facilitates efficient cellular operation by concentrating functional components, while avoiding potentially disruptive cross talk, similar to the system of organelles in eukaryotic cells. Focusing on the outermost components, the IM, periplasm, and OM make up the cell envelope, a barrier which separates bacteria from their environments, while affording efficient chemical transport, detoxification, and cell protection. Between the two membranes is the highly viscous, oxidizing environment of the periplasm. This compartment houses the peptidoglycan or cell wall, consisting of a polymer mesh of glycan chains cross-linked by oligopeptides that helps to maintain structural integrity (9, 25). Proteins are targeted to the periplasm through type II secretion, a process whereby N-terminal signal sequences direct translated peptides to IM transport proteins such as the Sec class of proteins (8). In the process of translocation, the targeting signal sequences are removed from periplasmic proteins, resulting in a mature form of the protein that can then participate in various functions such as preventing cellular toxicity through small molecule transport, chemical modifications, as well as participating in polymer synthesis and degradation (31). Furthermore, periplasmic proteins interact with the IM and OM, maintaining structural integrity and proper membrane activity.

While both the IM and OM are highly dynamic lipid bilayers with embedded proteins and other factors, they have significant differences in their compositions which
contribute to their specialized roles as cytoplasmic barrier and semipermeable cell surface layer, respectively. In the IM, both leaflets are composed of phospholipids such as phosphatidylethanolamine, with embedded proteins that are typically α-helical bundles. It has been estimated that as many as 20% of the open reading frames identified in *E. coli* encode for IM proteins; the roles of these proteins are diverse, but relate to cellular metabolic process such energy generation and conversion in the respiratory chain, cell division, signal transduction, and transport processes (18). The OM however is an asymmetric bilayer, with an inner leaflet composed of phospholipids and an outer leaflet composed primarily of lipopolysaccharide (LPS) (33). Proteins embedded in the OM are primarily β-barrels, and although they are encoded by only 2-3% of the genome in Gram-negative bacteria (40), they makeup about 90% of all cellular lipoproteins (38). Furthermore, lipoproteins are attached to the inner leaflet of the OM through post transcriptional modifications such as acylation (19). This asymmetry of the OM contributes to its role as a protective, semi-permeable barrier, which effectively interacts with its environment. However, the asymmetry of the OM also poses a unique challenge in the actual assembly of the lipid bilayer, in protein transport and insertion, as well as effective extracellular communication and pathogenesis.

The unique feature of the OM is the essential component LPS, a distal O-antigen polysaccharide attached to an endotoxin lipid A core (a phosphorylated glucosamine disaccharide acylated with fatty acids). The O-antigen can be divided further into an outer core oligosaccharide of various lengths, and an inner core sugar KDO (3-deoxy-D-manno-oct-2-uloseonic acid) (30). While various *E. coli* mutants have been identified which lack O-antigens, Lipid A and KDO have been found to be essential for proper
growth (31). Largely through characterization of gene knockouts, more than 50 genes have been identified to be required for synthesis and assembly of LPS at the cell surface (42), many of which are validated drug targets including the LpxC inhibitor L-161,240 (11, 22). Furthermore, two genes, kdsD and kdsC (formerly yrbH and yrbI, respectively) have recently been implicated in KDO assembly (23, 41), and that their action is genetically redundant (35).

While the genes that contribute to proper LPS assembly and transport have been highlighted, the mechanisms of how these factors interact to actually facilitate the process are poorly understood. The exceptions are the proteins MsbA which is responsible for flipping LPS across the IM (27, 29), and the Imp/RlpB complex involved in LPS targeting to the OM (42). However, the mechanism and factors involved in LPS transport across the periplasm, through the peptidoglycan remain largely uncharacterized. Two of the four other genes in the same operon as kdsD and kdsC, are the lipopolysaccharide transport genes A and B (lptA and lptB, formerly yhbN and yhbG) have been implicated in maintaining cell envelope integrity (35) and LPS transport to the OM (34). lptB is immediately downstream of the sequence for rpoN encoding for sigma factor $\sigma^5$. The other two genes in this operon are yrbG and yrbK, which have been suggested to be homologous to cation/calcium exchanger and IM anchored, respectively.

The lptA gene encodes for a 185 amino-acid protein that has an N-terminal 27 residue periplasmic targeting sequence and is co-transcribed with lptB (35). LptA has been shown to be essential for growth in E. coli (32, 34), and is upregulated during swarming in Salmonella, contributing to pathogenesis in Gram-negative bacteria (13). Conditional gene knockouts of lptA were shown to be sensitive to hydrophobic chemicals,
suggesting that these mutants had defective outer membranes (35). Furthermore, mutants depleted in LptA produce an anomalous form of LPS, resulted in GlcNAc accumulation in a novel membrane fraction, and were defective in LPS transport to the OM (34). In order to characterize the essential LPS assembly protein LptA, we have determined its crystal structures in two crystal forms: one with two molecules in the asymmetric unit at 2.15 Å resolution and another at 3.25 Å resolution with eight molecules in the asymmetric unit.

A1.3 Materials and methods

A1.3.1 Sequence analysis, protein expression and purification

Sequence analysis was carried out using Blastp (www.ncbi.nlm.nih.gov) (2) and Multialign (http://bioinfo.genopole-toulouse.prd.fr/multalin/) (7). GeneContext was used to highlight other proteins in the same operon as LptA (6). LptA fused to a C-terminal His$_6$-tag (LptA-SGRVEHHHHHHH) was subcloned into a pET 21b expression vector and transformed into BL21(DE3) cells. Cultures (1L) carrying plasmids for the recombinant protein were grown at 37°C in Terrific Broth-ampicillin (100 µg/ml), with five times the suggested amount of glycerol. Protein expression was induced at a culture OD$_{600}$ between 1.0-1.2 using 0.1 mM isopropyl-1-thio-β-D-galactopyranoside and grown for 16 h at 16°C. LptA protein substituted with seleno-methionine was expressed in a similar fashion, in the metA- E. coli strain DL41, in LeMaster medium supplemented with ampicillin (10). LptA was purified via nickel nitritotriacetate agarose batch purification in phosphate
buffer (pH 7.0) supplemented with 2% glycerol and dialyzed overnight in 20 mM sodium acetate (pH 5.5), 2% (v/v) glycerol. LptA was then purified further via anion exchange, using a 6ml Resource Q column on an ATKA Explorer FPLC in dialysis buffer, with 20 mM sodium acetate (pH 5.5), 2% (v/v) glycerol, 500 mM NaCl as the elution buffer. During each stage, purification was monitored by SDS PAGE, and LptA was estimated to be >95% prior to dialysis in 20 mM sodium acetate (pH 5.5), 50 mM NaCl, 2% glycerol pure prior to concentrated using an Amicon Ultra 5 kDa cutoff centrifugal filtration device (Fisher, Mississauga, CAN). Protein concentration was determined by BioRad protein assay (Biorad, Mississauga, CAN) using BGG as a protein standard and used immediately for crystallization. LptA expressed at low levels (~2.5 mg/L culture depending on batch). A Micromass Q-TOF mass spectrometer (Mississauga, CAN) was used to evaluate the processed form of purified LptA.

A1.3.2 Crystallization of LptA

Two different LptA crystal forms were obtained; one with two molecules in the asymmetric unit and another with eight molecules in the asymmetric unit (see Section A1.3.3). Crystallization of both crystal forms were obtained by hanging drop vapour diffusion method using freshly prepared protein and reagents. Crystallization of the two-molecule form of LptA was achieved by mixing equal parts 15 mg/ml LptA with reservoir solution consisting of 15% (v/v) PEG 3500, 17% (v/v) glycerol, 25 mM MgSO$_4$, 60 mM BisTris (pH 6.5), 2% Methanol. Rectangular prism crystals (2.2 x 1.0 x 1.0 mm) appeared after 14-21 d. The eight-molecule form of LptA crystallized in similar conditions to the two molecule form, except that reservoir solution was supplemented
with 5 mg/ml Ra-LPS or lauric acid. Crystals corresponding to the eight-molecule form had a morphology that resembled an American football (1.6 x 0.9 mm) and developed after 14-21 d. Where these crystals nucleated from, or grew into the cover-slip, the cross section of the football shaped crystals were hexagonal in their appearance. For both two- and eight-molecule forms, crystals were briefly soaked in reservoir solution supplemented with 30% glycerol as a cryoprotectant, and flash-cooled in the N₂ cold stream at 100K prior to diffraction data collection.

A1.3.3 Data collection and structure determination

LptA crystals were determined to belong to the P2₁2₁2₁ and P3₂2₁ space groups with two and eight molecules in the asymmetric unit, respectively. The multiple-wavelength anomalous dispersion (MAD) data sets were collected on the P2₁2₁2₁ form at beamline X29, Brookhaven National Laboratory. Data collection strategy involved data collection at the peak \( \lambda \) first, followed by the remote \( \lambda \), and then at the inflection \( \lambda \), each with 1.0° oscillations for 360°. Native two- and eight-molecule data was collected at beamlines F1 and A1, respectively, Cornell High Energy Synchrotron source. In both cases diffraction data was collected for 180°, using 1.0° oscillations. For both native and MAD datasets diffraction data were processed with HKL2000 and scaled with Scalepack (28). Heavy atom positions were determined and refined with autoSHARP, including density modification with SOLOMON, and solvent flattening (39). Electron density and structural solutions were attempted with many other software packages, but failed due to a lack of suitable anisotropic correction to correctly identify and refine heavy atom
positions. Automatic chain tracing was performed using RESOLVE (37). The remainder of the model was built by manual fitting using XtalView/Xfit (16). Due to the absence of any tryptophan residues in LptA which would facilitate assigning protein sequence, the heavy atom positions of selenomethionine were relied upon, and were later evaluated to be quite accurately estimated by autoSHARP (39). Once a suitable model of LptA was built using the peak wavelength dataset in the P2₁₂₁₂₁ form, it was used as a molecular replacement model to place the second molecule in the asymmetric unit prior refinement of the final structure (21). A single LptA molecule from the high resolution refined LptA structure was used for locate 8 molecules by molecular replace in the P3₂2₁ form.

For both native datasets, an anisotropic correction was applied via Michael Sawaya’s Anisotropy Correction Server (www.doe-mbi.ucla.edu/~sawaya/anisoscale/) (36), that facilitated both molecular replacement and improved the quality of the electron density obtained. In the eight-molecule form, Phaser successfully found six LptA molecules, and the other two were revealed by RESOLVE (37) through automatically chain tracing. The extra two molecules revealed by RESOLVE were then used together with the program LSQKAB (Superpose)(15) to correctly place two more LptA molecules in the asymmetric unit. The structure of LptA was refined based on both two- and eight-molecule datasets using CNS (5). Structural homology search using the SCOP database (24) was performed with the program SSM (14) and protein structures were superimposed using the CCP4 program LSQKAB (Superpose) (1, 12, 14, 15).
A1.4 Results and Discussion

A1.4.1 Sequence analysis

NCBI Blastp sequence analysis suggested that LptA contains sequence similarity with the Imp and OstA domains that have been implicated in tolerance to organic solvents (3, 42). Furthermore, LptA is highly conserved across Gram-negative bacteria, with homologues detected in various strains of *E. coli*, *Erwinia*, *Haemophilus*, *Salmonella*, *Shigella*, and *Yersinia*, many of which are human pathogens (Figure A1). While LptA mutants have been shown to have defective OMs (34), LptA overexpression was also an important obstacle to overcome. Sequence alignment suggested residues with high homology across the species examined are present throughout the sequence of LptA with a clustered stretch of residues extending from L45 to P77. Including residues which have less similarity throughout the sequences examined, we can extend sequence conservation throughout LptA, with the exception of the N-terminal 27 residues and the C-terminal 15 residues. Furthermore, LptA was evaluated to contain an N-terminal signal sequence (M1-A27) (20, 26), that is removed when LptA is exported into the periplasm (17). This result was corroborated by our own mass spectrometry analysis in which purified LptA was evaluated to have a molecular mass of 18603 Da, corresponding to a processed recombinant LptA-His$_6$ protein which has had the N-terminal 27 residues (M1-A27) removed (Predicted MW= 18645 (4)) prior to crystallization (Figure A2). LptA processing was also corroborated through size exclusion chromatography which further suggested that LptA is a monomer in solution at pH 5.5.
Figure A.1 Sequence alignment of LptA with various Gram-negative bacterial homologues, many of which are human pathogens. Residues highlighted with red represent high homology, and those highlighted with blue represent sequence similarity. While regions with high homology are sporadic throughout protein homologous to LptA, a notable cluster of conserved residues corresponding to residues P35-K83 are evident. Numbering corresponds to LptA, with gaps. Alignment was performed with Multalign.
Figure A.2 Q-TOF Mass Spectrometry result, highlighting the LptA has been processed prior to crystallization. The predominant protein present in the sample is 18603 Da, corresponding to a processed recombinant LptA that has had its N-terminal 27 residues (M1-A27) cleaved. Four other peaks are ~20 Da apart.
A1.4.2 Overcoming severe anisotropy in diffraction data

In all three LptA crystals (Table 1), diffraction data was severely anisotropic, which has the effect of varying the maximum resolution of diffraction data along different directions (a*, b*, and c*) (Figure A3). For the MAD dataset the diffraction limits were 4.2, 3.9, 3.2 Å along a*, b*, and c*, respectively, and in the native two-molecule form the diffraction limits were 2.7, 2.9, and 2.1 Å along a*, b*, and c*, respectively. In the eight-molecule form, the data was still severely anisotropy, with diffraction limits of 3.3, 3.3, 3.9 Å along a*, b*, and c*, respectively. The effect of this inherent anisotropy was discontinuous electron density especially for loop regions, poor side chain density for longer residues such as lysine and arginine, and the absence of clear density for solvent atoms. In order to overcome the problem of severe anisotropy and correct the structure factors associated with the LptA diffraction data, an anisotropy correction was applied (36). Effectively, this correction truncated any data outside of a defined ellipse instead of a sphere which is normally used, the data was then scaled into the defined ellipse, and a negative b-factor correction was applied in order to restore the magnitude of the high resolution reflections to their original values. The structural improvement following these corrections improved R_work and R_free for the two-molecule native dataset from 36% and 40% for the uncorrected data, to 30% and 37%, respectively, prior to solvent addition. Ramachandran analysis of LptA was also improved. A similar improvement in structural refinement was observed for the eight-molecule form. Data processing and refinement statistics are summarized for MAD and native crystals, as well as structural refinement in Table 1.
### Table A.1 Diffraction data and refinement statistics for LptA

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>Peak</th>
<th>Inflection</th>
<th>Remote</th>
<th>Native 1</th>
<th>Native 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of reflections following merging</td>
<td>158.7</td>
<td>159.8</td>
<td>159.8</td>
<td>159.8</td>
<td>159.8</td>
</tr>
<tr>
<td>a (Å)</td>
<td>45.759</td>
<td>45.866</td>
<td>45.884</td>
<td>46.762</td>
<td>46.21</td>
</tr>
<tr>
<td>b (Å)</td>
<td>55.457</td>
<td>55.658</td>
<td>55.624</td>
<td>57.473</td>
<td>46.21</td>
</tr>
<tr>
<td>c (Å)</td>
<td>149.276</td>
<td>149.567</td>
<td>149.572</td>
<td>148.907</td>
<td>186.93</td>
</tr>
<tr>
<td>No. of molecules in ASU</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9972</td>
<td>0.9974</td>
<td>0.9971</td>
<td>0.9124</td>
<td>0.9770</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>30 - 3.25</td>
<td>30 - 3.25</td>
<td>30 - 3.25</td>
<td>30 - 3.25</td>
<td>30 - 3.25</td>
</tr>
<tr>
<td>Total reflections</td>
<td>86,991</td>
<td>76,526</td>
<td>79,509</td>
<td>129,161</td>
<td>318,457</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>6,474</td>
<td>5,953</td>
<td>6,020</td>
<td>22,591</td>
<td>35,797</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (99.9)</td>
<td>99.9 (99.8)</td>
<td>99.9 (100)</td>
<td>99.9 (99.9)</td>
<td>97.3 (81.8)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>13.3</td>
<td>16.3</td>
<td>15.0</td>
<td>8.2</td>
<td>12.7</td>
</tr>
<tr>
<td>Rwork (%)</td>
<td>24.9</td>
<td>25</td>
<td>30.7</td>
<td>44</td>
<td>33.1</td>
</tr>
<tr>
<td>No. reflections total / Rmerge</td>
<td>10,510 / 913</td>
<td>10,510 / 913</td>
<td>10,510 / 913</td>
<td>10,510 / 913</td>
<td>10,510 / 913</td>
</tr>
<tr>
<td>No. atoms, protein / solvent (H2O) / total</td>
<td>2,054 / 293 / 2,347</td>
<td>15,158 / 0 / 15,158</td>
<td>15,158 / 0 / 15,158</td>
<td>15,158 / 0 / 15,158</td>
<td>15,158 / 0 / 15,158</td>
</tr>
<tr>
<td>B factors (Å²) protein / solvent (H2O) / total</td>
<td>45.1 / 61.1 / 47.1</td>
<td>66.0 / 66.0</td>
<td>66.0 / 66.0</td>
<td>66.0 / 66.0</td>
<td>66.0 / 66.0</td>
</tr>
<tr>
<td>RMSD, bond length (Å) / bond angle (°)</td>
<td>0.01 / 1.48</td>
<td>0.01 / 1.48</td>
<td>0.01 / 1.48</td>
<td>0.01 / 1.48</td>
<td>0.01 / 1.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement Statistics</th>
<th>2 Molecule Form</th>
<th>8 Molecule Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution Range (Å)</td>
<td>50 - 3.25</td>
<td>50 - 3.25</td>
</tr>
<tr>
<td>Rwork (%)</td>
<td>24.9</td>
<td>25</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>30.7</td>
<td>44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran Analysis</th>
<th>2 Molecule Form</th>
<th>8 Molecule Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in most allowed region # (%)</td>
<td>201 (82.2)</td>
<td>201 (82.2)</td>
</tr>
<tr>
<td>Residues in generously allowed region # (%)</td>
<td>28 (11.8)</td>
<td>402 (45.0)</td>
</tr>
<tr>
<td>Residues in additionally allowed region # (%)</td>
<td>6 (2.5)</td>
<td>142 (16.2)</td>
</tr>
<tr>
<td>Residues in disallowed region # (%)</td>
<td>1 (0.4)</td>
<td>63 (7.2)</td>
</tr>
</tbody>
</table>

*Number of reflections following merging
*Values in parentheses are for the outermost shell
Figure A.3 Anisotropic diffraction pattern from LptA MAD (left) and two-molecule native (right) crystals. Red circle highlights spherical area for scaling data and blue enclosed areas highlight diffraction which would normally not have been included. Anisotropic correction allowed for the inclusion of the higher resolution data.
A1.4.3 LptA protein fold and structural analysis

The overall refined structure of LptA resembles a series of eight antiparallel \( \beta \)-sheets that wind back along the path of the previous peptide stretch, throughout the length of the protein. These antiparallel sheets are folded approximately in the middle along their length, creating an overall topology where LptA resembles a ribcage or \( \beta \)-jellyroll (Figure A4), where each rib is approximately seven residues in length. Furthermore, the structure of LptA is not completely symmetrical along the length but opens slightly at both the N- and C-terminal ends, creating two clefts. The N-terminal cleft is slightly larger than the C-terminal cleft, and is surrounded by the conserved segment of residues from P35-K83 (Figure A5). Other conserved residues align along the length of the \( \beta \)-sheets, implicating a structural contribution. An evaluation of the surface electrostatics of LptA suggested a mainly negatively charged molecule, with a pronounced cleft delineated by the N-terminal cleft. In the eight-molecule form LptA is organized in a head-to-tail fashion, effectively forming fibers of LptA that intertwine (Figure A6). In this case, there was an absence of density for LPS or lauric acid in the partially refined structure of LptA. At this time it is therefore unclear whether the change observed for LptA has been influenced due to an interaction with LPS or lauric acid. Structural comparison of LptA against structures within the SCOP database (24) did not reveal any significant structural matches; hence, LptA represents a previously uncharacterized fold, although the structure of LptA does resemble a left-handed \( \beta \)-helix or \( \beta \)-jellyroll in general.
Figure A.4 Ribbon diagram of LptA with two molecules in the asymmetric unit at 2.15 Å resolution. Each of the LptA molecules is presented with different colours and gaps with dotted lines. LptA has also been rotated by 90° to highlight the gap along the length of LptA.
Figure A.5 Ribbon diagram of LptA highlighting residues conserved in sequence alignment with other Gram-negative homologues. Two clefts are present at the ends of LptA due to opening of the folded β-sheets that run along its length. A notable cluster of conserved residues flank the N-terminal cleft delineated at the end of the LptA structure, corresponding to residues P35-K83.
Figure A.6 Ribbon diagram of LptA with eight molecules in the asymmetric unit at 3.25 Å resolution. LptA molecules are organized head to tail fashion, forming a set of intertwined fibres. There was no clear density that resembled LPS or lauric acid in the partially refined model of LptA.
A1.4.4 Conclusions

Here we present the crystal structure of the processed form of the periplasmic localized, LPS transport protein LptA at 2.6 Å resolution with two molecules in the asymmetric unit and at 3.25 Å resolution with eight molecules in the asymmetric unit in the presence of LPS or lauric acid. The structure of LptA forms a unique set of eight antiparallel β-sheets which open at their ends. Conserved residues are not clustered to a particular structural location in LptA, with the exception of the N-terminal residues P35-K83 which are clustered proximal to one of the terminal clefts in LptA. This could potentially be a position that contributes to ligand binding or to mediate protein-protein contact. Based on sequence homology we can extend the structural information for LptA to various Gram-negative bacteria. Furthermore, due to the implication of LptA function in maintaining membrane integrity and LPS transport we can extend our findings to multiple organisms, with implications for an LptA contribution for pathogenesis.

A1.4.5 Footnotes

Thank you to Dr. Christian M. Udell for cloning LptA into the expression vector. Thank you to Dr. Robert Sweet for organizing the Brookhaven National Laboratory (BNL), Rapid Data collection course 2006 that ultimately led to a structural solution for LptA. LptA has not been deposited yet in the Protein Data Bank, pending further refinement. We are also grateful to Howard Robinson X-29 where the MAD dataset was collected, and the staff at MacCHESS where native data for both isoforms was collected.
Thank you to Dr. Clemens Vonrhein for expert assistance in implementing autoSHARP, and Dr. Michael Sawaya for useful discussions and anisotropy correction server to improve the map quality obtained from LptA data. This work has been supported CIHR. MS is supported by Bracken E.G. Bauman and R. Samuel McLaughlin Queen’s Graduate Fellowships as well as OGS. ZJ is a Canada Research Chair in Structural Biology.

A1.4.6 Appendix References


12. **Kabsch, W.** A solution for the best rotation to relate two sets of vectors.


15. **Krissinel, E. and K. Henrick.** 2004. Protein structure comparison in 3D based on secondary structure matching (SSM) followed by Cα alignment, scored by a new structural similarity function., p. 88. In Andreas J.Kungl & Penelope J.Kungl (Eds.) (ed.).


