

Structural and Biochemical Studies on Novel Bacterial Haem-Proteins

Sabine Schneider

Diplom-Biologist

Thesis submitted to the University of Nottingham for the
degree of Doctor of Philosophy

August 2007

“...everything that living things do can be understood in terms of the jiggings and wiggings of atoms..”

Richard Feynman, Nobel Laureate 1965

dedicated to my mother

Abstract

Haem proteins are functionally and structurally extremely diverse biomolecules and play a vital role in aerobic life. They perform a vast range of functions like transport of oxygen and electron transfer, gene regulation, redox-sensing and drug metabolism. This many-sidedness of haem-proteins is due to the extremely versatile chemical properties of the iron in the haem prosthetic group. Iron, or iron in the form of haem, plays a key role in many biological processes and it is an essential nutrient for the majority of living organisms. Despite being one of the most abundant chemical elements, iron is scarcely available under physiological conditions, because of its insolubility and toxicity. Pathogenic bacteria rely on their host as a source of haem and/or iron and a strong link between iron / haem acquisition, virulence factors and pathogenicity exists. Therefore they have evolved a set of specialised haem receptors and carriers to circumvent their iron dependency, often involving the 'stealing' of haem as a source of iron from host's haem-proteins, which in the host is the most abundant and relatively available source of iron. These proteins are both vital and unique to bacteria and so have been considered as possible drug targets. At the beginning of this thesis work, the fascinating cell and molecular biology mechanisms of these novel haem binding proteins were still largely unexplored.

In this thesis the cloning, expression, purification of four novel bacterial haem transport proteins for biochemical and biophysical characterisation and struc-

tural studies is described: HemS and HemT from *Yersinia enterocolitica* and Shp and HtsA from *Streptococcus pyogenes*. HtsA in complex with haem was crystallised and a preliminary X-ray diffraction analysis was carried out. Furthermore HemS was crystallised in its apo- and haem bound form and both structures were determined. Comparison of the apo- and haem-bound crystal structures provide penetrating insights into its mechanism of haem binding and release.

Acknowledgements

First of all I would like to thank my supervisor Dr. Max Paoli for his enthusiasm, support, encouragement, listening, confidence in me and inspiration throughout - beginning with my application for Australia to moving continent again.

Also I am grateful to the University of Nottingham for funding, without which I would not have been able to carry on with my thesis.

Further I owe many thanks to my colleagues for their good company, many pub quizzes and BBQs... Special thanks to Dr. Paul McEwan for his help getting started with the X-ray diffractor and CCP4. Many thanks also to technical assistance staff in CBS, especially to Lee Hibbet, Diane Moss, Donna Astill, Irene Reynolds, Penny Howick and Dr. Ian Withers.

Also special thanks to Prof. Eleanor Dodson for spending a day with me and my diffraction data in front of the computer, and Dr. Paul Barker for his help with the spectroscopic work. Many thanks also to Dr. Jakki Cooney for growing *Yersinia* cells for me, many helpful discussions and a great time in Ireland. Last but not least I want to thank my family and friends for their moral support, endless telephone conversations and encouragement in all my decisions despite being that far away. Especially to Isabelle for the many 'telephone beers'.

This thesis was prepared with L^AT_EX.

Contents

Abstract	i
Acknowledgments	iii
List of Figures	xiii
List of Tables	xiv
Abbreviations	xv
1 Introduction	1
1.1 Properties of haem and haem-proteins	1
1.2 Haem-acquisition in bacteria	5
1.3 General principles of X-ray crystallography	14
1.3.1 Crystallisation of proteins	14
1.3.2 Theory of X-ray diffraction by a crystal	16
1.3.2.1 Mosaicity, crystal space group and data collec- tion strategy	20
1.3.2.2 The structure factor and the phase problem .	21
1.3.2.3 Phasing techniques	22
1.3.2.4 Density modification	26
1.3.2.5 Model refinement	27
1.3.2.6 Validation of the model	32

1.4	Aims of the thesis	36
2	General materials and methods	37
2.1	Materials and chemicals	37
2.2	Methods	37
2.2.1	Bacterial strains and growth media	37
2.2.2	Polymerase chain reaction	38
2.2.2.1	PCR from genomic DNA	38
2.2.2.2	Single colony polymerase chain reaction	39
2.2.3	General cloning methods	39
2.2.4	Protein expression and purification	40
2.2.5	UV/vis absorption spectroscopy and reconstitution of proteins with haemin	41
2.2.6	Fast protein liquid chromatography	42
2.2.7	Crystallisation of proteins	42
2.2.8	Secondary and tertiary structure prediction	43
2.2.9	X-ray diffraction, data processing and structure refinement	44
3	Diversity and conservation of interactions for binding haem	45
3.1	Introduction	45
3.2	Methods	47
3.3	Results and Discussion	48
3.3.1	Diversity of haem-protein packing contacts	48
3.3.2	Re-occurring contacts: protein-haem interaction 'hot spots'	49
3.3.3	Anchoring of haem propionates by arginine residues	50
3.3.4	Conserved interactions at the 'haem face' involving leucine	51

3.3.5	Conserved interactions at the 'haem face' involving phenylalanine/tyrosine side chains	53
3.3.6	Contacts at the haem edge	55
3.3.7	Interactions with the proximal ligand	57
3.3.8	Haem orientation relative to the proximal ligand	58
3.3.9	Perspectives for the design of novel haem proteins	59
3.4	Conclusions	61
4	Biochemical and biophysical characterisation of HemS from <i>Yersinia enterocolitica</i>	62
4.1	Introduction	62
4.2	Materials and methods	63
4.2.1	Cloning of <i>hemS</i> from genomic DNA	63
4.2.2	Site-directed mutagenesis of <i>hemS</i>	64
4.2.3	Expression and purification of HemS	65
4.2.4	UV/vis absorption spectroscopy and reconstitution of HemS with haemin	65
4.2.5	His-tag pull down assays	66
4.2.6	Cloning of the <i>cyclic AMP receptor protein</i> -gene from genomic DNA	67
4.2.7	Expression and purification of the cyclic AMP receptor protein	68
4.3	Results and Discussion	68
4.3.1	Sequence alignment and structure prediction of HemS	68
4.3.2	Biochemical characterisation of HemS	71
4.3.2.1	Purification of HemS and HemS-mutants	71
4.3.2.2	Spectroscopic analysis and haem-binding prop- erties of HemS	74
4.3.3	Identification of interaction partners of HemS	79

4.4	Conclusions	86
5	Crystallisation and structure-determination of HemS from	
	<i>Yersinia enterocolitica</i>	88
5.1	Introduction	88
5.2	Materials and methods	89
5.2.1	Expression and purification of seleno-methionine derivatised HemS	89
5.2.2	Crystallisation of HemS	90
5.2.3	Diffraction analysis and structure of HemS	91
5.3	Results and Discussion	93
5.3.1	X-Ray crystal structure of HemS	93
5.3.1.1	Crystallisation of HemS	93
5.3.1.2	Structure determination of HemS	95
5.3.2	Folding topology of HemS	102
5.3.2.1	Molecular recognition of haem by HemS	104
5.3.2.2	Comparison with other haem-proteins	105
5.3.3	Iron coordination and haem-geometry	107
5.3.4	Conformational changes upon haem binding	110
5.3.5	A haem-induced fit accompanies the switch from an open, apo-state to a closed, bound-state	113
5.3.6	Structure-function relationships and evolution of the HemS fold	115
5.4	Conclusions	120
6	Cloning, expression and purification of HemT from	
	<i>Yersinia enterocolitica</i>	122
6.1	Introduction	122
6.2	Materials and methods	125

6.2.1	Cloning of <i>hemT</i> from genomic DNA	125
6.2.2	Expression and purification of HemT	126
6.2.3	UV/vis spectroscopy and reconstitution of HemT with haemin	127
6.3	Results and Discussion	127
6.3.1	Expression and purification of HemT	127
6.3.2	UV/vis spectroscopy and reconstitution of HemT with haemin	129
6.4	Conclusions	131
7	Structure-function relationship of HtsA and Shp from <i>Streptococcus pyogenes</i>	133
7.1	Introduction	133
7.2	Materials and methods	136
7.2.1	Cloning of <i>htsA</i> and <i>shp</i> from genomic DNA	136
7.2.2	Expression and purification of HtsA and Shp	136
7.2.3	Reconstitution of HtsA and Shp with haemin and UV/vis spectroscopy	137
7.2.4	Dynamic light scattering	137
7.2.5	Crystallisation of haem-HtsA and haem-Shp	138
7.2.6	Structure determination of HtsA	138
7.3	Results and Discussion	139
7.3.1	Sequence alignment and structure predictions	139
7.3.2	Purification and biochemical characterisation of HtsA .	142
7.3.2.1	Purification of HtsA and Shp	142
7.3.2.2	UV/vis spectroscopy of HtsA and Shp	145
7.3.3	Crystallisation of Shp and HtsA	146
7.3.3.1	Crystallisation trials of the haem-Shp complex	146
7.3.3.2	Crystallisation of the haem-HtsA complex . .	147

Contents

7.3.4	Diffraction analysis of HtsA	150
7.3.5	Complex-formation of Shp and HtsA	156
7.4	Conclusions	157
8	Conclusions	160
	Bibliography	164
	Appendices	xviii
A	Appendices	xviii
A.1	Amino acid abbreviations (IUPAC)	xviii
A.2	List of primer	xix
A.3	Plasmid maps	xx
A.4	Haem-propionate contacts in analysed folding topologies.	xxiii
A.5	Residues in contact with the haem-face and haem-edge.	xxiv
B	List of publications	xxv

List of Figures

1.1	Schematic representation of haem	2
1.2	Folding topologies of haem proteins	3
1.3	Schematic representation of the haem-uptake systems in bacteria	7
1.4	Organisation of the haem-transport locus in Gram-negative species	9
1.5	Organisation of the haem-transport locus in <i>S. aureus</i>	10
1.6	NEAT domains in <i>S. aureus</i> Isd-proteins.	12
1.7	Solubility curve for a protein as a function of the salt concentration	15
1.8	Schematic representation of the vapour diffusion technique . .	15
1.9	Bragg's law and diffraction condition	17
1.10	The Ewald construction	18
1.11	Locus of spots 'lunes' produced on a detector by oscillating a crystal	19
1.12	Mosaicity of a crystal	20
1.13	Friedel's law	26
3.1	Conserved protein-haem interactions	53
3.2	Packing interactions with the proximal ligand	57
3.3	'Flipped' haem orientations	59
4.1	Overlapping primer extension PCR	64
4.2	Sequence similarities and identities of HemS homologues . . .	69
4.3	Sequence alignment of HemS and HemS homologues	70

List of Figures

4.4	Expression and purification of HemS	72
4.5	Removal of the His-tag from HemS with thrombin	73
4.6	Analytical gel-filtration analysis of the haem-HemS complex	73
4.7	Reconstitution of wild type HemS with haemin	74
4.8	UV/vis spectra of wild type HemS	75
4.9	pH-dependent haem-binding by HemS	77
4.10	Spectroscopic analysis of HemS mutants and His-tagged GST	78
4.11	Pull-down assay with haem-HemS	81
4.12	Results and sequence coverage of peptide-mass fingerprinting	81
4.13	Expression and purification of CRP	83
4.14	Pull-down assay of HemS with His-CRP	84
4.15	Analytical gel-filtration of haem-HemS and CRP	85
5.1	Crystallisation of HemS	94
5.2	Diffraction pattern of the haem-HemS crystals	96
5.3	Anomalous Patterson map of haem-HemS	98
5.4	Difference Patterson map of haem-HemS	99
5.5	Seleno-methionine labeled HemS	100
5.6	Electron density of the haem pocket	103
5.7	Structure of the haem-HemS complex	103
5.8	Haem-protein interactions in the haem-HemS complex	105
5.9	Comparison of the haem-binding pocket with other haem proteins	106
5.10	Difference electron density map of the distal haem site	109
5.11	Comparison of the haem-binding pocket of the haem-HemS complex with apo-HemS	110
5.12	Difference distance matrix plot of HemS	111
5.13	Conformational switch between apo- and haem-HemS	112
5.14	Difference distance matrix plot of ChuS	114
5.15	Superposition of the haem-ChuS complex with apo-ChuS	114

5.16	Average B-factors of main chain atoms in haem-HemS crystal- structures	115
5.17	Structure comparison of haem-HemS and AGR_C_4470p	117
5.18	Secondary structure alignment of the HemS and AGR_C_4470p	118
5.19	Sequence alignment of AGR_C_4470p homologues	119
5.20	HemRSTUV operon organisation	120
6.1	Sequence alignment of HemT and HemT homologues.	124
6.2	Expression of <i>Y. enterocolitica</i> HemT and <i>S. meliloti</i> HmuT .	128
6.3	Expression and purification of <i>Y. enterocolitica</i> HemT	129
6.4	UV/vis spectrum of <i>Y. enterocolitica</i> HemT	130
6.5	Homology model of HemT.	131
7.1	Iron / haem-acquisition locus in <i>S. pyogenes</i>	134
7.2	Model of haem transport in <i>S. pyogenes</i>	135
7.3	Sequence similarities and identities of HtsA homologues	140
7.4	Sequence alignment of HtsA with BtuF	141
7.5	Purification of HtsA	143
7.6	Analytical gelfiltration analysis of HtsA	144
7.7	Dynamic light scattering analysis of HtsA	144
7.8	UV/vis spectroscopy of HtsA and Shp	145
7.9	High-throughput crystallisation screen of Shp	147
7.10	Crystallisation of HtsA	148
7.11	Partial cleavage of HtsA over time	149
7.12	Diffraction analysis of haem-HtsA	151
7.13	SSM of the molecular replacement models	153
7.14	Electron density of the haem-HtsA complex	155
7.15	Analytical gelfiltration of haem-HtsA and haem-Shp	157

8.1 Structural studies on the bacterial haem-uptake systems published in the last few years	161
---	-----

List of Tables

1.1	Distinct haem protein folding topologies	4
1.2	HemRSTUV homologues across Gram-negative bacteria	8
2.1	Bacterial growth media reference	38
2.2	Secondary and tertiary structure prediction server	43
3.1	Haem protein folds with their average solvent accessible area	49
3.2	Residues in contact with the haem propionates	51
3.3	Haem-face contacts	52
3.4	Haem-edge contacts	56
4.1	Comparison of absorption maxima of haem-proteins	76
5.1	MAD data collection details of haem-HemS crystals	92
5.2	Data processing statistics of haem-HemS Fe-MAD data	97
5.3	Data processing statistics of haem-HemS platinum derivative data	97
5.4	Data collection and refinement statistics	102
5.5	Haem geometry of HemS and selected haem-proteins	108
5.6	Centre of gravity of secondary structure elements in apo- and haem-HemS	112
7.1	UV/vis absorption spectra of HtsA and Shp-His	146
7.2	Data collection and processing statistics	152

Abbreviations

A	absorbance
Å	Ångstrom
°C	degree celsius
cAMP	3'-5'-cyclic adenosine monophosphate
DNA	deoxy nucleic acid
dNTP	deoxy nucleotide
Da	Dalton
EC	<i>E. coli; Escherichia coli</i>
EDTA	ethylene-diamine-tetra-acetate
ESRF	European Synchrotron Radiation Facility
Fe	iron
FW	molecular weight
g	gram
<i>g</i>	gravitational acceleration
h	hour
Hepes	N-(2-hydroxyethyl)-piperazine-N'-2-ethanesulfonic acid
I	intensity
Isd	iron-regulated surface determinants
IPTG	isopropyl-beta-D-thiogalactopyranoside
kV	kilovolts

Abbreviations

l	litre
m	milli
μ	micro
M	molar
mA	milliampere
MAD	multi-wavelength anomalous dispersion
MES	2-morpholinoethanesulfonic acid
min	minute
MIR	multiple isomorphous replacement
MOPS	3-(N-Morpholino)-propanesulfonic acid
MR	molecular replacement
Ni	nickel
nm	nanometer
OD	optic density
PAGE	poly acrylamid gel electrophoresis
PBP	periplasmic binding protein
PBS	phosphor buffered saline
PCR	polymerase chain reaction
PDB	protein data base
Pt	platinum
RMSD	root mean square deviation
rpm	revolution per minute
RT	room temperature
R_{factor}	R-factor
SA	<i>S. aureus</i> ; <i>Stapylococcus aureus</i>
SAD	single-wavelength anomalous dispersion
SDS	sodium dodecyl sulphate
SM	<i>S. meliloti</i> , <i>Sinorhizobium meliloti</i>

Abbreviations

SONO	sensor of nitric oxide
SP	<i>S. pyogenes, Streptococcus pyogenes</i>
sec	second
t	time
TEMED	N',N',N',N'-tetramethylethylenediamine
Tris	trishydroxymethylaminomethane
U	unit
UV	ultraviolet
V	voltage
vis	visible
v/v	volume per volume
w/v	weight per volume
YE	<i>Y. enterocolitica, Yersinia enterocolitica</i>

Chapter 1

Introduction

1.1 Properties of haem and haem-proteins

Haem (iron-protoporphyrin IX) is an extremely versatile prosthetic group widespread in biological systems and vital to aerobic life. Its biological function depends entirely on its incorporation into proteins where the protein environment determines the haem properties like its geometry, iron spin state and redox potential. Haem is an essential cofactor for the electron transfer and redox reactions of respiration and photosynthesis and the oxygen binding and transport functions carried out by the globins. Haem is also used by proteins involved in catalysis, such as the catalases and mono-oxygenases, as well as in proteins fulfilling a great variety of processes, including signal transduction and the control of gene expression (Chapman *et al.*, 1997, Paoli *et al.*, 2002, Anderson & Chapman, 2005). Figure 1.1 depicts a schematic representation of the haem.

Variations of the haem structure is also possible by modifications of the pyrroles' methyl and vinyl side chains which extend from the edge of the porphyrin. In b-type haem, iron-protoporphyrin IX is non-covalently bound to the protein; this is the prosthetic group of the familiar structures of haemoglobin, myo-

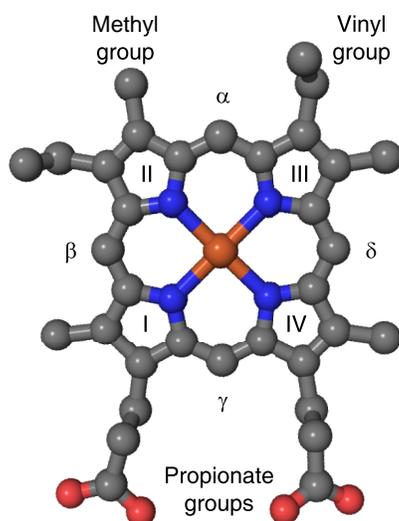


Figure 1.1: Ball and stick representation of b-type haem. Haem (protoporphyrin IX) consists of four pyrrole-rings (I-VI) that are linked by methylene bridges (α , β , γ , δ). The two propionate groups originate from pyrrol I and VI with their carboxy termini, enabling electrostatic interactions with the protein environment or the solvent. In the centre of the porphyrin plane a ferric (Fe^{3+}) or ferrous (Fe^{2+}) iron (orange) is coordinated by the four pyrrole nitrogens (blue). The haem has an asymmetric character, because of the four methyl and two vinyl groups.

globin and cytochrome b5. However, most other cytochromes contain c-type haem which differs from b-haem in the covalent attachment of the vinyl side chains to two cysteine residues by thioether bonds. Moreover there are other, less common haem derivatives including the a-type haem, which is present in cytochrome c oxidase, and the d1-type haem identified in nitrite reductase (Allen *et al.*, 2005). A summary of haem derivatives and associated haem proteins is reported in the Promise database (Degtyarenko *et al.*, 1997).

Haem-proteins show a spectacular range of distinct molecular architectures all associated with the same, chemically identical haem ligand; b-type haem is found in over 20 different folding topologies as highlighted in Table 1.1 and depicted in Figure 1.2. Moreover protein-bound haem has strong electronic absorbance bands that depend on its the geometry, spin state, redox potential and ligation. Therefore Ultra Violet/visible (UV/vis) absorption spectroscopy

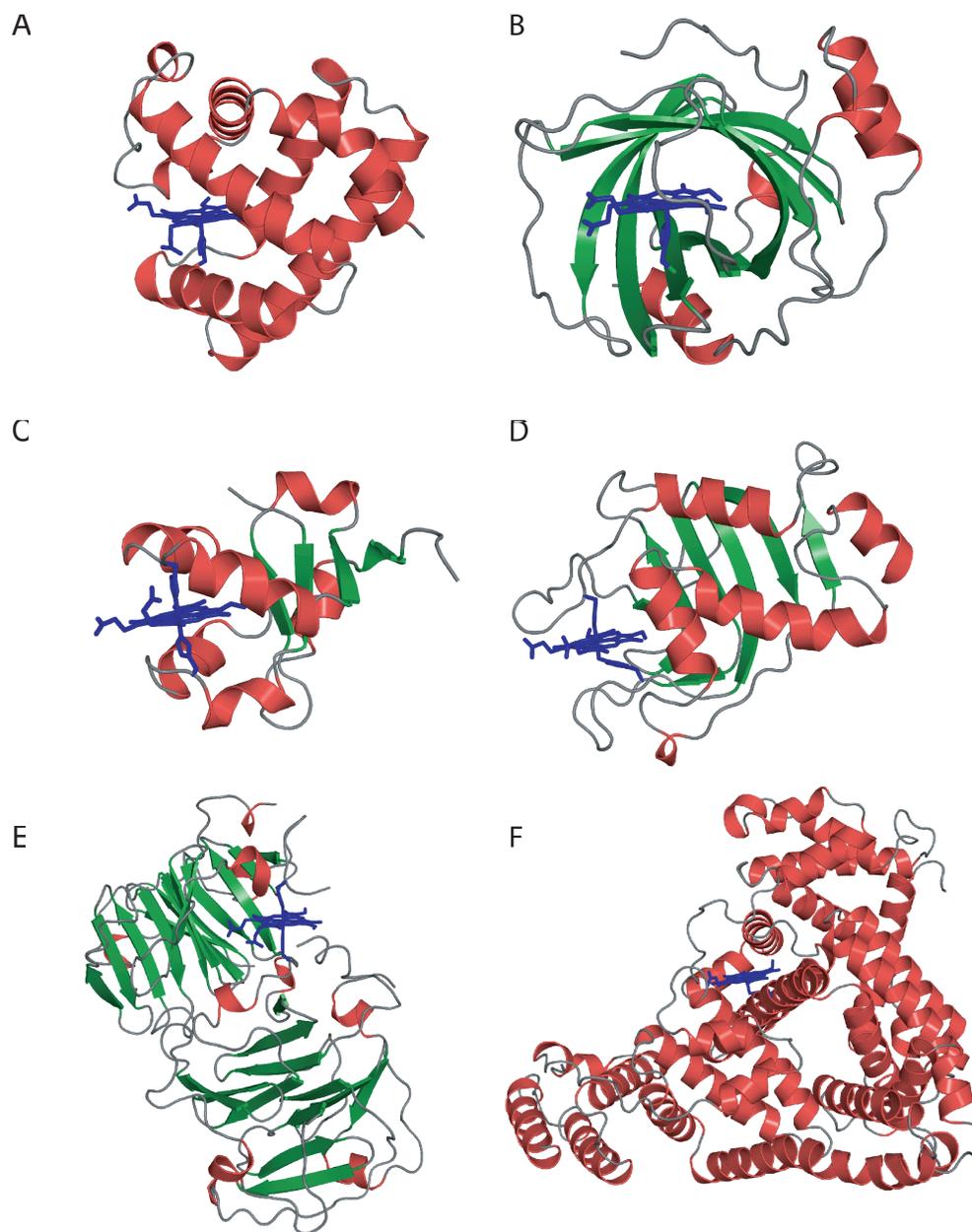


Figure 1.2: Schematic diagrams showing the structures of six representative b-type haem proteins highlighting the striking variety of their folding topologies. The haem environment can be formed by any element of secondary structure or any combination of these, as well as turns and loop regions. (A) Tuna myoglobin (1MYT), (B) Insect nitrophorin (1NP4), (C) Cow cytochrome b5 (1CYO), (D) Bacterial HasA (1DK0), (E) Rabbit haemopexin (1QHU) and (F) Human serum-albumin (1N5U).

is a powerful tool to investigate haem-protein interactions. Haem proteins have a characteristic Soret peak between 390-440 nm and Q-bands, also called

α/β -bands, between 500-700 nm. These absorption maxima and their changes under different conditions (i.e. reducing, oxidising) and the presence of ligands (i.e. carbon monoxide, cyanide, oxygen) can give information about the coordination structure and ligation state of the haem prosthetic group.

Table 1.1: Selected representatives of b-type haem proteins with distinct folding topologies. In addition to their different folds, the table highlights the diversity of their functions and phylogenetic origins as well as the ligation and coordination state of the haem-iron

Fold	Representative Proteins	Function	Origin	Ligand	PDB CODE
4-helix bundle	Cytochrome b562	electron transport	bacteria (<i>Escherichia coli</i>)	His/Met 6c	1QPU
β -propeller	haemopexin	haem binding and transport	rabbit (<i>Oryctolagus cuniculus</i>)	bis-His 6c	1QHU
CAP	CooA	CO sensing	bacteria (<i>Rhodospirillum rubrum</i>)	His/Pro 6c	1FT9
catalase	catalase	hydrogen peroxide decomposition	bacteria (<i>Helicobacter pylori</i>)	Tyr 5c	1QWL
cupredoxin-like	ubiquinol oxidase	O ₂ reduction to water	bacteria (<i>Escherichia coli</i>)	His 5c	1FFT
cytochrome b5	cytochrome b5	electron transfer	rat (<i>Rattus norvegicus</i>)	bis-His 6c	1EUE
dioxygenase	indoleamine 2,3-dioxygenase	L-Tryptophan catabolism	man (<i>Homo sapiens</i>)	His 5c	2D0T
globin-like	haemoglobin	O ₂ binding and transport	man (<i>Homo sapiens</i>)	His 5c	1A3N
haem oxygenase	haem oxygenase	haem degradation	man (<i>Homo sapiens</i>)	His/H ₂ O 6c	1N45
haem peroxidase	cyt c peroxidase	biosynthetic and catabolism	yeast (<i>Saccharomyces cerevisia</i>)	His 5c	2CYP
HemS	HemS	haem transport	bacteria (<i>Yersinia enterocolitica</i>)	His 5c	2J0P
H-NOX / SONO	SONO	NO sensing	bacteria (<i>Thermoanaerobacter tengcongensis</i>)	His 5c	1XBN
immunoglobulin -like	cellobiose dehydrogenase	lignin and cellulose degradation	fungus (<i>Phanerochaete chrysosporium</i>)	His/Met 6c	1D7B
lipocalin	nitrophorin	NO transfer	insect (<i>Rhodnius prolixus</i>)	His 5c	1NP4
meander	Has A	haem binding and transport	bacteria (<i>Serratia marcescens</i>)	His/Tyr 6c	1B2V
NO	NO synthase	catalytic	mouse (<i>Mus musculus</i>)	Cys 5c	1NOS
P450	P450 mono- oxygenase	oxidation of organic substrates	fungus (<i>Streptomyces coelicolor</i>)	Cys 5c	1ODO
PAS	FixL	O ₂ -sensor / signalling	bacteria (<i>Bradyrhizobium Japonicum</i>)	His 5c	1DRM
serum albumin -like	albumin	regulation of the colloidal osmotic pressure of blood	man (<i>Homo sapiens</i>)	Tyr 5c	1N5U
vitamin B6 family	cystathione β -synthase	redox-controlled PLP -dependent synthesis of cystathione	man (<i>Homo sapiens</i>)	Cys/His 6c	1JBQ

Optical absorption spectroscopy in combination with site-directed mutagenesis has been used to identify the haem-iron coordinating residues in a number of haem proteins, such as the N-terminal domain of the haem regulated eIF2 α -kinase (NTD-HRI) (Inuzuka *et al.*, 2004), chloroperoxidase (Yi *et al.*, 1999), SOUL (Sato *et al.*, 2004), haem oxygenase (Wilks *et al.*, 1996) cytochrome b₅₆₂ (Barker *et al.*, 1996) and cellulose dehydrogenase (Rotsaert *et al.*, 2001, 2003b).

1.2 Haem-acquisition in bacteria

Despite being one of the most abundant chemical elements, iron is scarcely available under physiological conditions due to its insolubility and toxicity, and this makes iron acquisition by any organism a remarkable challenge. Pathogenic bacteria rely on their host for iron and a first link between infection, pathogenicity of bacteria and the abundance of iron was discovered in the 1940s. It was shown that specific iron-binding proteins in human serum inhibit the growth of bacteria by preventing iron uptake through sequestration. The iron binding protein in the blood serum was identified as transferrin (Schade & Caroline, 1944, 1946). To circumvent their iron dependency microorganisms have evolved a range of specialised proteins that enables them to steal iron and iron in the form of haem from host proteins (Braun, 2001, Braun *et al.*, 1998, Faraldo-Gomez & Sansom, 2003, Genco & Dixon, 2001, Ratledge & Dover, 2000, Wilks & Burkhard, 2007). During invasive bacterial infections, erythrocytes become lysed due to lytic toxins (e.g. haemolysin) or peroxides produced by the pathogens and/or the host-defence system. Therefore the local concentration of available haem, haemoglobin and other host-haem binding proteins, which can be used by the bacteria as iron source, increases.

In Gram-negative bacteria the haem-uptake systems known to date can be

classified in two groups:

i) Systems that secrete an extracellular haem-binding protein, a so called haemophore, that scavenges haem from a variety of sources. The best genetically and structural studied example is haemophore HasA and the components (HasR, HasB, HasD, HasE) of its system from *Serratia marcescens* (Arnoux *et al.*, 1999, Wandersman & Delepelaire, 2004). When the work reported in this thesis started, HasA was the only bacterial haem transport protein the structure of which had been solved.

ii) Systems that recognise haem or specific host-haem proteins. One of the first to be studied is the HemRSTUV system from the Gram-negative Proteobacterium *Yersinia enterocolitica*. *Y. enterocolitica* is a human pathogen, can cause gastroenteritis and is closely related to *Yersinia pestis*, the causative agent of plague. The HemRSTUV haem acquisition machinery is widespread in Gram-negative bacteria and a schematic representation is depicted in Figure 1.3A. The system consists of 5 proteins. Located on the outer membrane, the receptor HemR sequesters haem from host haem proteins or directly binds free haem. When internalised, the ligand is taken up by the periplasmatic carrier HemT and passed onto the hetero-dimer HemUV, an integral inner membrane permease. In the cytosol, haem is held by the soluble protein HemS. This process is driven by the proton-motive force used by the TonB-ExbB-ExbD system as well as ATP hydrolysis by the inner membrane permease (Faraldo-Gomez & Sansom, 2003, Genco & Dixon, 2001, Stojiljkovic & Hantke, 1994, Wilks & Burkhard, 2007).

This haem-transport system is predominantly found in pathogenic bacteria. For instance the operon coding for the ChuASTUV system in *E. coli* is present in the pathogenic *E. coli* O157:H7 strain and absent in the non-pathogenic *E. coli* K12 strain (Hayashi *et al.*, 2001). The structure of the haem transport locus in some Gram-negative species, their phylogenetic distribution and their

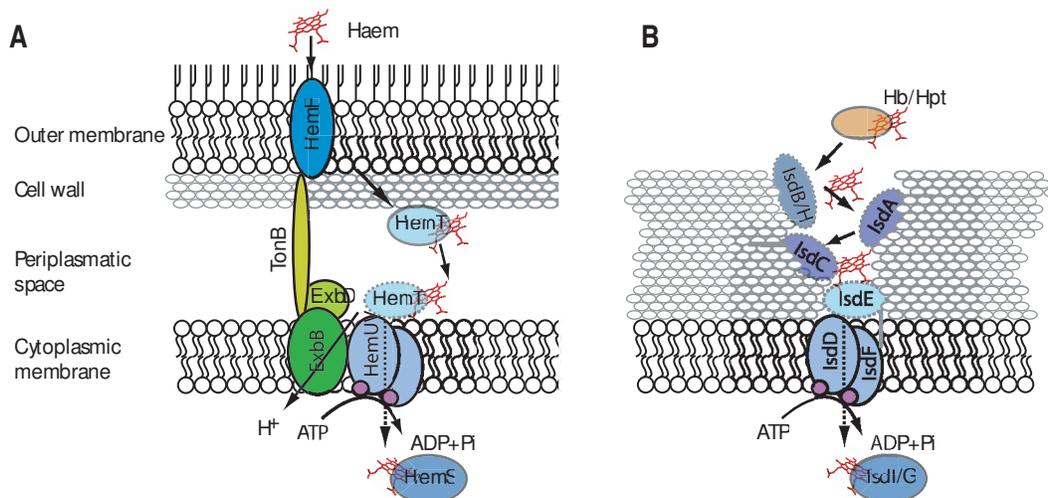


Figure 1.3: Schematic representation of the major haem-uptake systems in bacteria. (A) The HemRSTUV system of *Y. enterocolitica* (Stojilkovic & Hantke, 1994) is common to many Gram-negative species. (B) Components of the haem-uptake machinery in Gram-positive bacteria are less conserved. Currently the best studied system is the Isd system of *S. aureus* (Maresso & Schneewind, 2006).

nomenclature is shown in Figure 1.4 and listed in Table 1.2.

Homologues of the outer membrane receptor can be found across a variety of species and are assumed to share the same basic architecture of a β -barrel (Braun & Braun, 2002). Two conserved histidine residues in HemR (His 128 and His 461) are crucial for the interaction and internalisation of haem by *Yersinia* (Bracken *et al.*, 1999). ShuA from *Shigella dysenteriae* is the only outer membrane haem-receptor that has been purified and its haem-binding properties analysed (Burkhard & Wilks, 2007). It was shown that ShuA extracts haem from haemoglobin and as for HemR, two histidine residues (His 86 and His 420) are essential for haem-recognition. In addition to that it was shown that ShuA interacts with TonB (Burkhard & Wilks, 2007). *Vibrio cholerae* has two receptors (HutA and HutR) encoded at different sites on the chromosome (Wyckoff *et al.*, 2004, 2007).

No homologues in *Yersinia*, *Bordetella* or *Vibrio* can be found for the related

Table 1.2: Homologs of the HemRSTUV system in Gram-negative bacteria. CJ = *Campylobacter jejuni* (Chan *et al.*, 2006); EC = *E. coli* (Hayashi *et al.*, 2001, Perna *et al.*, 2001, Torres & Payne, 1997); PA = *P. aeruginosa* (Ochsner *et al.*, 2000); PS = *Pleisiomonas shigelloides* (Henderson *et al.*, 2001); SD = *Shigella dysenteriae* (Mills & Payne, 1995, Wyckoff *et al.*, 1998); VC = *Vibrio cholerae* (Wyckoff *et al.*, 2006); YE = *Y. enterocolitica* (Stojiljkovic & Hantke, 1994); YP = *Yersinia pestis* (Thompson *et al.*, 1999); CBP = cytoplasmic binding protein. * *V. cholerae* has a conserved hypothetical protein homologues to PhuW, ChaN etc.

	CJ	EC	PA	PS	SD	VC	YE	YP
Receptor	ChaR	ChuA	PhuA	HugA	ShuA	HutA+R	HemR	HmuR
PBP	-	ChuT	PhuT	HugB	ShuT	HutB	HemT	HmuT
Permease	-	ChuUV	PhuUV	HugCD	ShuUV	HutCD	HemUV	HmuUV
CBP	-	ChuS	PhuS	-	ShuS	-	HemS	HmuS
Others	ChaN	ChuW	PhuW	HugW	ShuW	*	-	-
	-	ChuX	-	HugX	ShuX	HutX	OrfA (OrfX)	
	-	ChuY	-	-	ShuY	-	OrfB (OrfY)	
	Cj1613c	-	-	HugZ	-	HutZ	-	YPO0285

proteins ChuW, PhuW, HugW, HutW and ChaN (Cha = *Campylobacter haem* acquisition). These proteins have homology to HemN the oxygen-independent coporphyrinogen oxidase from *E. coli* K12 (Wyckoff *et al.*, 2004, Chan *et al.*, 2006, Ochsner *et al.*, 2000), which is a member of the radical SAM superfamily of S-adenosyl-methionine-dependent enzymes (Layer *et al.*, 2002, Sofia *et al.*, 2001). HemN catalyses the oxidative decarboxylation reaction of coporphyrinogen III to protoporphyrin IX, a common step in the biosynthesis of haems and chlorophylls (Dailey, 2002). But it has been shown, that HutW from *V. cholerae* failed to complement *Salmonella thyphimurium* HemN. Therefore it is unlikely that HutW and its homologues are oxygen-independent coporphyrinogen oxidase, and the general annotation as *hemN* may need to be reviewed (Wilks & Burkhard, 2007). Moreover mutation of *phuW* had limited effect on growth with haem as sole source of iron, but was necessary for optimal haem and haemoglobin utilization (Ochsner *et al.*, 2000). Recently, the crystal structure of ChaN, identified as a haem transport protein from *Campylobacter jejuni*, has been solved, showing a novel fold, with an unusual

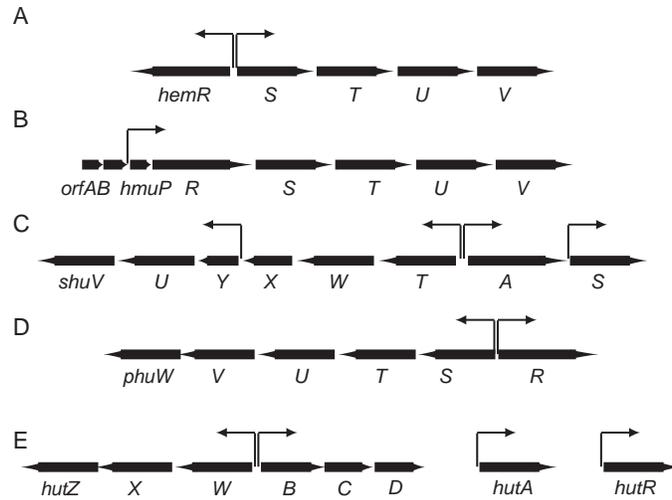


Figure 1.4: Organisation of the haem-transport locus in Gram-negative species. (A) *Y. enterocolitica* (Stojiljkovic & Hantke, 1994), (B) *Y. pestis* (Thompson *et al.*, 1999); (C) *S. dysenteriae* (Mills & Payne, 1995, Wyckoff *et al.*, 1998); (D) *P. aeruginosa* (Ochsner *et al.*, 2000). (E) *V. cholerae*; the ORF for the haem-receptors HutA and HutA are encoded on a different locus on the same chromosome (Wyckoff *et al.*, 2004, 2007). Arrows are indicating promoter and transcription direction.

cofacial haem binding arrangement. ChaN dimerises upon binding two haem molecules and the haem is 5-coordinated with a tyrosine residue as the iron's proximal ligand (Chan *et al.*, 2006).

The genes *shuX* and *shuY* in *S. dysenteriae* have homology to two ORF encoded upstream of *hmuP* in *Y. pestis* (Wyckoff *et al.*, 1998). Moreover ShuX shows homology to *V. cholerae* HutX (Wyckoff *et al.*, 2004) and *P. shigelloides* HugX (Henderson *et al.*, 2001), but no homology to open reading frames (ORFs) with known function for any could be detected. But *shuX* and *shuY* are probably not expressed due to a stop-codon after *shuW* (Wyckoff *et al.*, 2004). *V. cholerae* HutZ has been shown to be required for efficient haem utilisation (Wyckoff *et al.*, 2004) and shares homology with HugZ from *P. shigelloides* (Henderson *et al.*, 2001).

In contrast, the haem-uptake systems in Gram-positive bacteria are less conserved. The best studied is the Isd (iron-regulated surface determinant) system of the pathogen *Staphylococcus aureus*, a leading cause of a wide spectrum of

hospital acquired infections such as toxic shock syndrome (TSS) and septic endocarditis (Weems, 2001). Haem is the preferred iron-source of *S. aureus* and it appears to have therefore evolved more than one haem-uptake system, as evidenced by the recent identification of the novel *haem-regulated transport* proteins, HrtAB (Friedman *et al.*, 2006), and *haem-transport system* proteins, HtsABC (Skaar *et al.*, 2004b) shows. The acquisition of haem in *S. aureus* through the IsdDEF, HrtAB and HtsABC systems is highly efficient, since staphylococcal cells are saturated after only 15 min when incubated with exogenous haem (Mazmanian *et al.*, 2003). Gram-positive bacteria are lacking an outer membrane and therefore do not require outer membrane receptors or the TonB-ExbB-ExbD systems. A schematic representation of the interlinked *isd*-system is depicted in Figure 1.3B (Skaar & Schneewind, 2004, Skaar *et al.*, 2004b). The IsdA, IsdB, and IsdH components are attached to the cell wall and transfer their load to IsdC, which is the central conduit in this funnel-like uptake machinery. Passage through the membrane is then mediated by the IsdD/E/F complex, for final delivery to the cytosolic haem oxygenases IsdG and IsdI (Skaar & Schneewind, 2004, Skaar *et al.*, 2004a, Mack *et al.*, 2004). The *isd*-genes are encoded in three distinct regions and five transcriptional units of the *S. aureus* genome (Fig. 1.5)(Skaar & Schneewind, 2004).

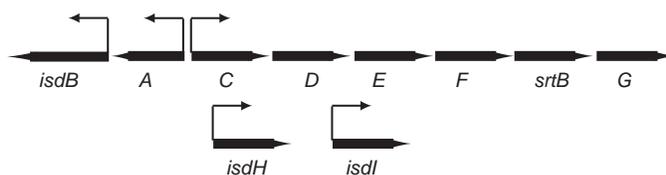


Figure 1.5: Organisation of the haem-transport locus in *S. aureus*. In *S. aureus* the Isd system is encoded on three distinct chromosomal regions. The transpeptidase SrtB, encoded by *srtB*, is involved in anchoring IsdC to the cell wall and is the only non-haem binding protein of the operon (Skaar & Schneewind, 2004).

Most components of the *isd*-system have been functionally and biochemically characterised; for instance, haem-uptake assays carried out on IsdA and IsdF staphylococcal knock-out mutants had no influence on cell growth with haem as source of iron and is most likely due to alternative haem-transport systems in staphylococci (Mazmanian *et al.*, 2003). It was shown that IsdA has a broad specificity spectrum and has binding activity to extracellular matrix proteins, such as fibronectin and fibrinogen, and also to holo-transferrin and haemoglobin (Clarke *et al.*, 2004). IsdH (also called HarA) is a receptor for haptoglobin-haemoglobin complexes, which can also bind to a smaller extent haemoglobin and haptoglobin alone, and is the sole haptoglobin-binding surface component of *S. aureus* (Dryla *et al.*, 2003); but IsdH does not bind haem on its own (Pilpa *et al.*, 2006). Furthermore, it has been shown that IsdB binds haemoglobin (Torres *et al.*, 2006). IsdA, IsdB, and IsdH are also being used to develop vaccines against *S. aureus* to reduce the occurrence of disease (Clarke *et al.*, 2004, Kuklin *et al.*, 2006).

Based on sequence comparisons, IsdA, IsdB, IsdH, and IsdC are all believed to contain one or more copies of the same haem-binding structural module, known as the NEAT domain (Andrade *et al.*, 2002)(Fig. 1.6A). Recently the structures of the apo N-terminal NEAT domain of IsdH (Pilpa *et al.*, 2006), the apo- and haem-bound NEAT domain of IsdA (Grigg *et al.*, 2007a) and of the haem-IsdC complex (Sharp *et al.*, 2007) were determined. The NEAT domain folding topology is that of a β -sandwich (Fig. 1.6B), consisting of two five-stranded, twisted, antiparallel β -sheets and is structurally related to immunoglobulin proteins. There are also other haem-proteins known with this molecular architecture, such as cellobiose dehydrogenase (Hallberg *et al.*, 2000), ethylbenzene dehydrogenase (Kloer *et al.*, 2006), and cytochrome f (Chi *et al.*, 2000). In these enzymatic proteins the haem is bound to a topologically equivalent site, which is different to the binding site in IsdA and IsdC. In IsdA

and IsdC, the haem-iron is 5-coordinate with a tyrosine ligand.

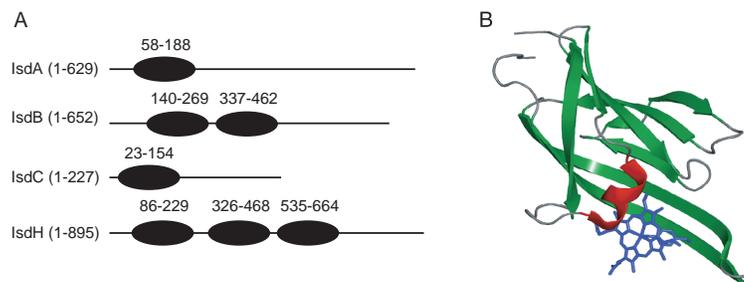


Figure 1.6: Structure and schematic representation of the NEAT domains in *S. aureus* Isd-proteins. (A) Schematic diagram of the NEAT domains in Isd-proteins (Andrade *et al.*, 2002, Pilpa *et al.*, 2006). (B) The common folding topology of NEAT domains is a β -sandwich as shown here in the cartoon representation of the structure of the haem-IsdC complex (Sharp *et al.*, 2007).

In the last few years, exciting progress has been made in the identification and functional characterisation of haem-uptake proteins, essential for bacteria. This is owing to the availability of sequence information, provided by the recent completion of multiple genome sequencing projects. Moreover these sequence data make it possible to target the proteome of a whole organism for functional and structural studies. A better mechanistic understanding of the nutrient uptake system in bacteria, in particular for an essential element such as iron, could lead to new drug targets. The knowledge gained from the study of the function and structure of these systems can help in the study of haem-transporters from other pathogens. This is of particular interest when we consider the increasing number of multi-drug-resistant bacterial strains, such as the methicillin resistant *S. aureus* (MRSA) or other pathogenic organisms like the parasitic protozoa *Plasmodium falciparum*, the causative agent of malaria. The life cycle of *Plasmodium* has steps directly involving haem interactions and metabolism and its genome sequence has been recently completed and it is therefore likely that novel haem proteins will be identified.

The impact of structural biology in biomolecular and biomedical research

is growing, for instance through structure-based mutagenesis and structure-guided drug design. Following the sequencing projects of many genomes, the area of structural genomics has grown significantly, aiming at the high-throughput structure determination of all possible folding topologies. Knowledge of all architectures would then open the door to homology modelling thus providing a basis of structural understanding for all protein molecules. This is tackled and performed in dedicated centres and consortia. A powerful technique for structure determination is X-ray crystallography, the principles of which are summarised in Section 1.3.

1.3 General principles of X-ray crystallography

The knowledge of the three dimensional structure of a protein gives precious insight into its function and can be applied to structure-based design of drugs and therapeutic agents. X-ray crystallography is a powerful technique for structure determination, where the scattering signal of a molecule is amplified by the packing into the crystal lattice. In theory there is no given size limit for the molecule to be studied. However in practice the limiting step remains the ability to produce crystals of sufficient size and a high degree of order.

1.3.1 Crystallisation of proteins

Many proteins are relatively dynamic molecules, consisting of long polypeptide chains folded into specific three dimensional structures with complex and irregular shaped surfaces. It therefore can be difficult to achieve the high level of structural order that is necessary for the molecules to pack into a crystal lattice. The molecules are held together by generally weak interactions: hydrogen bonds, salt bridges and hydrophobic interactions. Protein crystals have a high solvent content typically ranging between 30 and 90%, and usually present in large channels formed amongst the molecules packed in the lattice. Thus most protein crystals are fragile (Rupp, retrieved July 2006 from the world wide web). The formation of lattice interactions is influenced by a critical series of parameters such as temperature, protein purity, protein concentration, nature of precipitant and solvent, pH, and presence or absence of additives like cofactors, ligands and ions. The initial crucial step for crystallisation is the production of large quantities of highly purified protein. Determination of crystallisation conditions is a trial and error, iterative screening process. The balance between solubility, precipitation and the small

zone of nucleation, when regular inter-molecular interactions start to form, is delicate (Fig. 1.7).

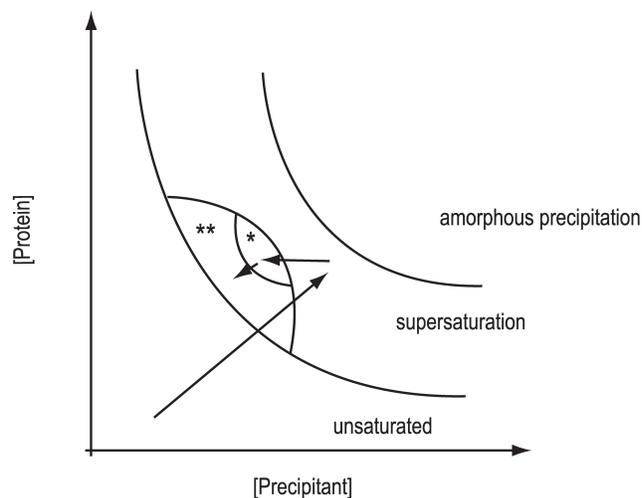


Figure 1.7: Solubility curve for a protein as a function of the salt concentration. *Zone of crystal nucleation; ** Zone of crystal growth (Drenth, 1999)

Techniques where the sample concentration increases in a controlled way have been developed to aid macromolecular crystal formation. The most popular method is that of vapour diffusion, depicted in Figure 1.8.

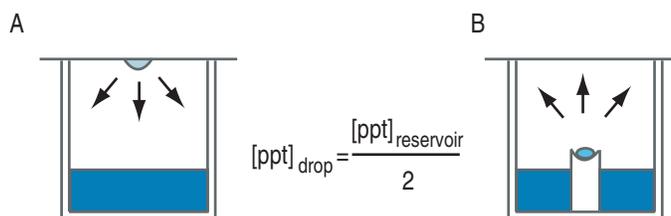


Figure 1.8: Schematic representation of the vapour diffusion technique. (A) hanging drop, (B) sitting drop vapour diffusion. The small drop containing protein, stabilizing buffer, precipitant (ppt) and other additives is placed in vapour equilibrium in a closed system with a much larger reservoir volume, with a much higher reagent concentration than in the drop. Over time water is pulled from the drop in a vapour phase until equilibrium between the both is reached. During this process, the sample in the drop is concentrated increasing its relative supersaturation (Drenth, 1999).

1.3.2 Theory of X-ray diffraction by a crystal

X-rays are scattered by the electron cloud of an atom. If the X-rays scattered by different atoms are in phase, the weak scattering vectors from each individual atom, periodically repeated in the three dimensional order of a crystal, add up to yield an overall wave of higher intensity. This phenomenon is called constructive interference and is observed as diffraction spots on a detector and is mathematically described by Bragg's law (Eq. 1.1).

$$2 \cdot d \cdot \sin\theta = n \cdot \lambda \quad (1.1)$$

where

- n is an integer
- θ is the angle of the incident beam
- d is the spacing between the reflecting planes of atoms
- λ is the monochromatic wavelength of the incident beam.

The Bragg equation (Eq. 1.1) has to be satisfied to obtain an observable diffraction pattern and in consequence the following principles can be derived:

- i) Wavelength and path-length distance are inversely related so that to maintain their relationship, a large angle of the incident beam must be coupled to a small distance between the reflecting planes (Fig. 1.9).
- ii) Because of the inverse relationship of \mathbf{d} and θ , the difference in path length is an integer of the wavelength of the X-ray beam.
- iii) For $\mathbf{n}=1$, \mathbf{d} can be substituted for d_{hkl} where h , k and l are the indices (Millers indices) that define the reflecting planes intersecting a unit cells (with the cell edges a , b and c) in $\frac{a}{h}$, $\frac{b}{k}$ and $\frac{c}{l}$.

The inverse relationship becomes more obvious by rearranging the Bragg' equation (Eq. 1.2) and the indirect relationship can be replaced by a direct one, the reciprocal lattice.

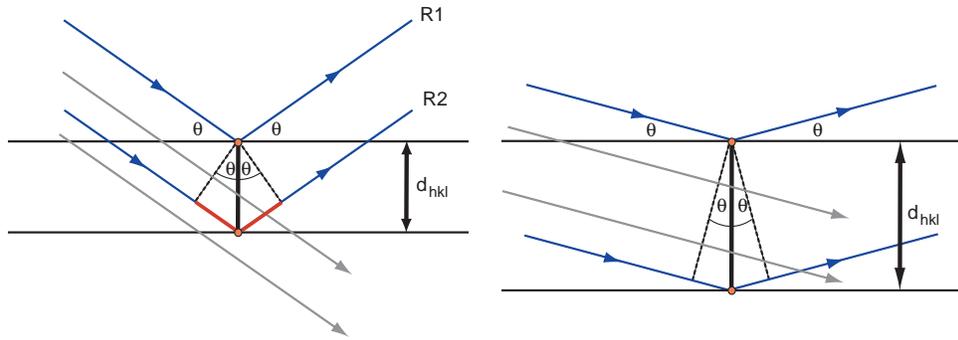


Figure 1.9: Bragg's law. When the distance additionally traveled by R2 (red) is an integer of the wavelength, then Bragg's law is fulfilled and R1 and R2 interfere constructively. The angle of the incident beam θ is inversely related to the spacing of the diffracting planes.

$$2\sin\theta = \frac{n\lambda}{d} \Rightarrow \theta \sim \frac{1}{d} \quad (1.2)$$

The reciprocal lattice is constructed based on $\frac{1}{d_{hkl}} = d_{hkl}^*$ which varies directly as $\sin\theta$ and the reciprocal lattice planes perpendicular to the real lattice planes. Therefore the larger the unit cell parameters, the smaller is the spacing of the reciprocal lattice planes. The relationship between the reciprocal unit cell axes and the real unit cell axes are described in Equations 1.3, 1.4 and 1.5 below.

$$a \cdot a^* = b \cdot b^* = c \cdot c^* \quad (1.3)$$

$$\text{or} \quad a^* = \frac{1}{V}bc \cdot \sin\alpha \quad b^* = \frac{1}{V}ca \cdot \sin\beta \quad c^* = \frac{1}{V}ab \cdot \sin\gamma \quad (1.4)$$

$$\text{or} \quad a^* = \frac{1}{d_{100}} \quad b^* = \frac{1}{d_{010}} \quad c^* = \frac{1}{d_{001}} \quad (1.5)$$

where V is the unit cell volume

a , b and c are the unit cell edges,

α , β and γ are the angles between the cell edges
and d_{100} , d_{010} and d_{001} are the distances of reflecting
planes parallel to the zero-level planes bc , ac and ab .

The Ewald construction illustrates Bragg's law geometrically in reciprocal space in three dimensions (Fig. 1.10). The radiation of the X-rays is represented by a sphere centered on the X-ray beam with the radius $\frac{1}{\lambda}$. The origin of the reciprocal lattice of the crystal is the point where the direct beam crosses the Ewald sphere. If a reciprocal-lattice point lies on an intersection of the Ewald sphere and the vector perpendicular to the originating real lattice planes (= reciprocal lattice vector) with the length $\frac{1}{d}$, then the interference condition (Bragg's law) for that particular reflection is fulfilled and gives rise to a diffracted beam along the line joining the sphere center with the reciprocal lattice point (Dauter, 1999, Rhodes, 2006).

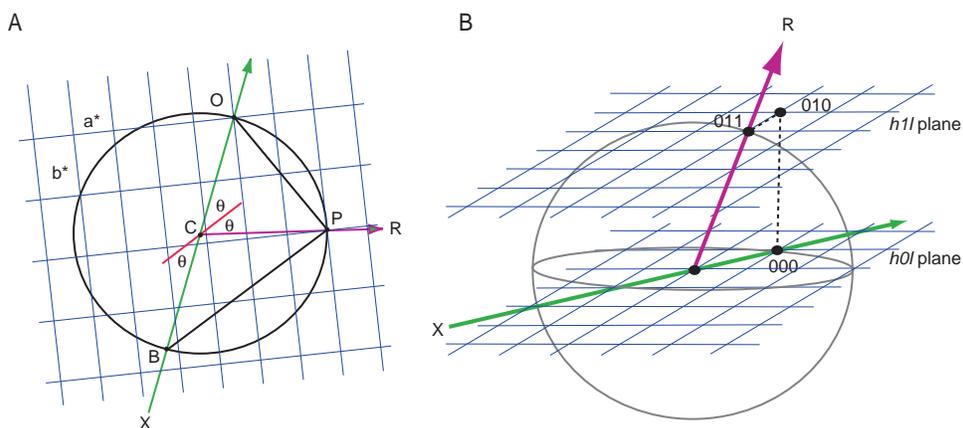


Figure 1.10: The Ewald construction. An incoming X-ray beam is diffracted at a set of lattice planes (red) and the diffracted X-ray beam is crossing the surface of the Ewald sphere (radius = $1/\lambda$) at a reciprocal lattice point. The trigonometric condition of Bragg's law is fulfilled and a reflection is observed (Rhodes, 2006).

A reflecting plane intersecting the Ewald sphere projects onto the two dimensional detector as an ellipse, since all rays diffracted by the same plane form a cone (Dauter, 1999, Rhodes, 2006). When a crystal is rotated in a X-ray

beam, different reflections from the same plane as well as different reflecting planes are brought onto the surface of the Ewald sphere, and therefore into reflection condition, forming a diffraction pattern consisting of spots arranged in concentric ellipses, the so called 'locus of spots' ('lune') on the detector (Fig. 1.11) (Dauter, 1999, Rhodes, 2006).

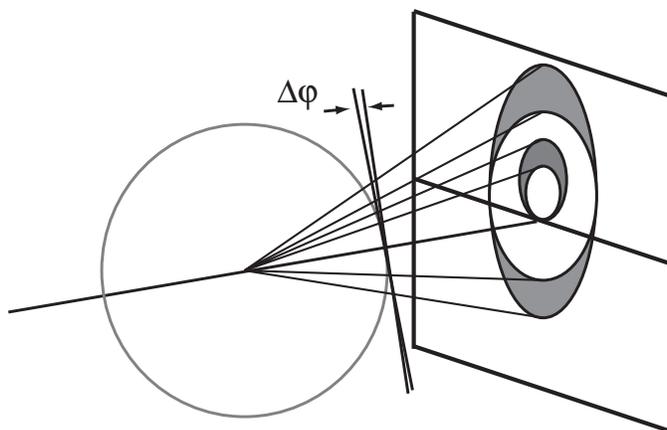


Figure 1.11: Locus of spots 'lunes' produced on a detector by oscillating a crystal. Many reciprocal lattice planes are brought into reflection condition by intersecting the Ewald sphere and many reflections are sent to the detector (Dauter, 1999).

Again due to the inverse relationship of the reciprocal lattice dimensions to the real lattice, a large structure with a large unit cell and a densely populated reciprocal lattice will give rise to a more dense reflection pattern with smaller distances between the reflections, compared with a small-molecule crystal. This is because more reflecting planes are in reflecting position at any given time. Also the finer the sampling of the reflecting planes of a unit cell, the higher are the indices of the reciprocal lattice points the larger is the reflecting angle of the scattered beam and the higher the resolution. Therefore the finer is the detail we can observe in the reconstruction of the crystal structure. The correlation between reciprocal lattice and real crystal lattice is one of the main concepts of X-ray crystallography.

1.3.2.1 Mosaicity, crystal space group and data collection strategy

Crystals are not perfectly ordered, but are a mosaic of roughly aligned submicroscopic arrays (Fig. 1.12A); the imperfections in the arrangement of molecules are more pronounced in protein crystals than in inorganic or small molecule crystals. As result of the mosaicity, the diffracted X-ray beam is not a perfect linear beam, but has a conical shape (Fig. 1.12B) so that the Bragg condition is satisfied over a larger space and the reflection is spread over more images. Therefore in protein crystallography the reflection intensities are measured over small angular ranges and the mosaic spread is measured in degree oscillation (Rhodes, 2006).

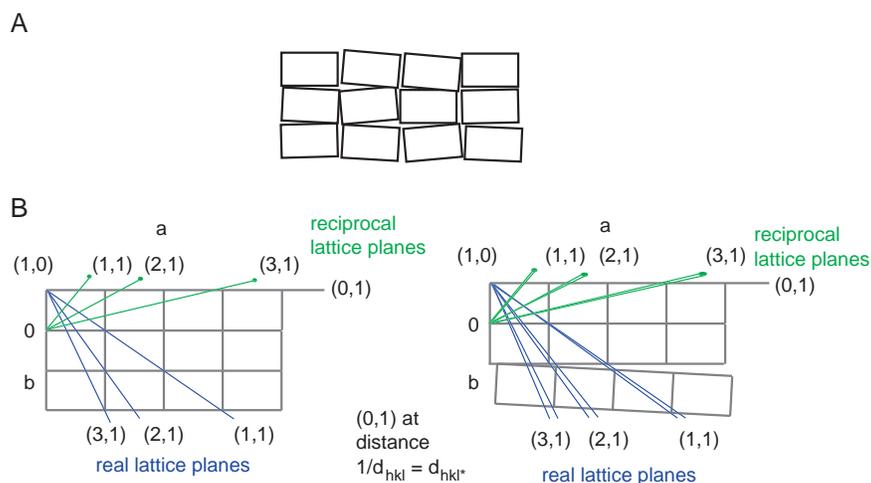


Figure 1.12: Mosaicity of a crystal. (A) Crystals are not perfectly ordered, but are a mosaic of submicroscopic arrays. (B) In case of high mosaicity the real lattice planes (blue) are 'smeared' and the reciprocal lattice points (green) become ellipsoid.

The degree of data that need to be collected to obtain the unique part of reciprocal space and therefore a complete data set, depends upon the symmetry of the repeated motif. For instance, if the crystal space group includes a three-fold axis then a unique data set consists of one sextant of the reciprocal space.

1.3.2.2 The structure factor and the phase problem

After a diffraction experiment, the position (described as Millers indices h , k , l) and the intensity (I) of the reflections are integrated in a list of figures including the (h, k, l) values, their associated intensities and standard deviations of intensity (σI). Due to the inherent non-homogeneity of the data collection, like crystal thickness and orientation, the reflections in any data collection are on different scales. In the stage of scaling, a set of mathematically determined corrections (scales) is applied and the data from each collected image are placed on a common scale. The vector representing the overall scattering from a particular set of Bragg planes is termed the structure factor (F_{hkl}). Because of the correlation between the real lattice and the reciprocal lattice space, the structure factor equation can be expressed in terms of Miller indices and fractional coordinates, which is a central concept to determine a molecular structure from a diffraction pattern. The intensity of a given reflection is proportional to the square of the structure factor amplitude, F_{hkl} , where the structure factor (Eq. 1.6) is defined as the sum of the scattering factors by a group of atoms and is dependent on the three-dimensional arrangement of the atoms in the group.

$$F_{hkl} = \sum_{j=1}^n f_j e^{[2\pi \cdot i(h_j x + k_j y + l_j z)]} \quad (1.6)$$

where f_j is the atomic scattering factor of the electron cloud of atom j ,
 x, y, z are the fractional coordinates of the electron cloud of atom j ,
 $e^{[2\pi \cdot i(h_j x + k_j y + l_j z)]}$ is the periodic function describing the reflected wave.

Taking into account the scattering of all atoms in a crystal, Equation 1.7 can be derived:

$$F_{hkl} = V \int_{\text{cell}} \rho(xyz) e^{[2\pi i(hx+ky+lz)]} dV \quad (1.7)$$

where $\rho(xyz)$ is the electron density at position x, y, z and
 V is the volume of the unit cell.

The electron density is then calculated by performing a Fourier Transform (FT) of all structure factors to give a Fourier series (Eq. 1.8) and the electron density is the graph of this function.

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{[-2\pi i(hx+ky+lz)+i\alpha(hkl)]} \quad (1.8)$$

where F_{hkl} is expressed as a complex quantity,
with the structure factor amplitude $|F_{hkl}|$
and the phase $\alpha(hkl)$.

Therefore to calculate the electron density at a given point within the unit cell, both amplitude and phase have to be known. The amplitude can be measured in the diffraction experiment, but information about the phase is lost. This is known as the phase problem.

1.3.2.3 Phasing techniques

To circumvent the phase problem various methods have been developed, to obtain phases directly and indirectly.

The Patterson function

The Patterson function is one of the most important mathematical relationships in crystallography and is calculated by inverse Fourier Transform (Eq. 1.8)

of the structure factor intensities (amplitudes squared) with the phases set to zero (Eq. 1.9).

$$P(uvw) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 \cdot e^{-2\pi(hu+kv+lw)} \quad (1.9)$$

where u, v, w locate points in the Patterson map and

V is the volume of the unit cell.

This results in a map of the vectors between all atoms of the structure. Peaks in the Patterson map are correlated to peaks in the electron density. If the structure is a small molecule with only a few atoms, it is possible to 'deconvolute' the Patterson and solve the structure. For bigger molecules this is not possible because of the large number of peaks ($N^2 - N$, where N is the number of atoms in the molecule). However, this technique can still be used to solve the substructure of heavy atoms and anomalous scatterer in the molecule, because the difference in the structure factor amplitudes ($|F_{hkl}|^2$) between the native and the derivative reflects the contribution of the heavy atom/anomalous scatterer. Additionally, using the Patterson function allows the orientation of the search model in the molecular replacement technique to be determined.

Direct methods

For small molecules with less than 100 atoms and atomic resolution data, the phases can be derived directly from the intensities in form of intensity relationships between reflections (Karle & Hauptman, 1956). Jerome Karle and Herb Hauptmann received the Nobel Prize for chemistry in 1985. For molecules larger than 100 atoms (non hydrogen atoms), the success rate drops dramatically due to the fact that intensity relationship becomes progressively weaker.

However direct methods are extremely important for solving the heavy atom 'substructure' in proteins for experimental phasing.

Molecular replacement

Molecular replacement (MR) is the technique that can be applied when the structure of a similar molecule is known, which is used as a search model. The level of sequence identity between two molecules correlates well with their structural similarity and MR works well at a sequence ID $\geq 40\%$ with a relatively complete search model. MR is carried out by rotating ('rotation function') and translating ('translation function') the search model, so that it aligns with the unknown target structure in the crystal. The known phases from the aligned search model are then used with the amplitudes from the diffraction of the target structure to calculate an initial electron density map. Traditionally rotation and translation functions were based on Patterson methods as implemented in the programs AMORE and MOLREP (CCP4, 1994). More recently methods including 'maximum likelihood' have been developed, which are taking into account the error in the data and model and improve the success rate of MR with poorer models and/or data. The program PHASER (McCoy *et al.*, 2005) uses this approach.

Single and multiple isomorphous replacement

Another method to solve the phase problem is the *single* and/or *multiple isomorphous replacement* (SIR, MIR) in which compounds containing strong scattering elements such as heavy metals (e.g. platinum or iridium) are soaked into the crystal. Taking into account the strong scattering of heavy atoms, their positions can be determined by a difference Patterson function and initial, approximate phases can be determined. It is necessary that they are occupying the same position in each unit cell. For unambiguous phase determination at

least two derivatives are often necessary, although SIR can work in some cases. The disadvantage of this method is that sufficient heavy atom-derivatised crystals have to be prepared and this often changes or even destroys the lattice. Binding of heavy atom compounds can cause rearrangement of molecules or domains/loops and so leading to non-isomorphism of the native and derivative crystal. As a result the obtained data are no longer isomorphous and the phase will be in error (Drenth, 1999, Rhodes, 2006, Rupp, retrieved July 2006 from the world wide web).

Single- and multi-wavelength anomalous dispersion

Single- and multi-wavelength anomalous dispersion (SAD, MAD) uses the property of some elements to scatter X-rays anomalously at wavelengths close to their absorption edges (or resonant energy). In the vicinity of the X-ray absorption edge the wavelength-dependent contributions f' (real) and f'' (imaginary) to the atomic X-ray scattering factor change drastically with wavelength. This is caused by inner shell electrons absorbing the incoming X-rays for a short period before re-emitting them, resulting in a small shift in phase and frequency. This phenomenon is called anomalous dispersion or scattering. Friedel's Law states that Bragg reflections related by inversion through the origin (= Friedel pair) have equal amplitude and opposite phase (Eq. 1.10).

$$|F_{hkl}| = |F_{\overline{hkl}}| \quad \varphi_{hkl} = -\varphi_{\overline{hkl}} \quad (1.10)$$

In case of anomalous scattering, the amplitudes remain equal but the phase relationship no longer holds, because the f'' term is always positive (Fig. 1.13). The measurable anomalous difference of the structure factor amplitude and phase of the X-rays scattered is maximised when comparing the observed intensities at the absorption edge and the inflection point and can be treated as a special case of MIR and so can be used to obtain phase information.

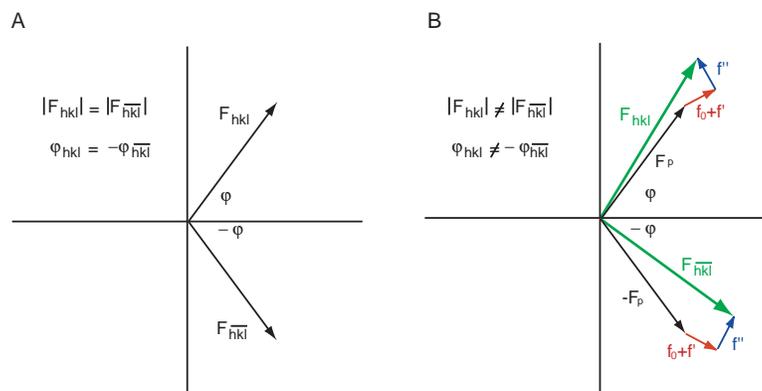


Figure 1.13: Anomalous scattering and Friedel's law. (A) Friedel's law is fulfilled when all atoms have the same scattering behaviour. (B) Break-down of Friedel's law due to anomalous scattering of some atoms. The magnitude of F_{hkl} (and thus, intensity) and phase of F_{hkl} are different for F_{hkl} and $-F_{hkl}$. F_p is the structure factor of the protein alone, $f + f_0$ is the real part and f'' is the imaginary part of the anomalous scatterer.

All atoms can scatter X-rays anomalously, but only atoms exhibiting a strong anomalous signal can be used for phasing. Examples atoms used for SAD and MAD phasing include iron, sulphur or selenium, which are naturally part of the protein e.g. in form of haem or iron-sulphur cluster or incorporated in the amino acid chain as selenium-methionine derivative. The advantage of this technique is that the collected data are derived from a single crystal and so the problem of non-isomorphism or destroying the crystal by soaking are avoided. Initial phases can also be obtained using a method which is called SIRAS (single isomorphous replacement with anomalous scattering) and is a combination of MIR with the anomalous scattering derived from the heavy atom (Drenth, 1999, Rhodes, 2006, Rupp, retrieved July 2006 from the world wide web).

1.3.2.4 Density modification

Electron density maps from phases obtained from MIR, SIRAS, MAD and other methods are often difficult to interpret due to errors in the initial phases.

Density modifications like *solvent flattening/flipping*, *histogram matching* and *non-crystallographic symmetry averaging* are techniques to improve phases for macromolecular structures. A powerful technique for phase improvement is solvent flattening and/or solvent flipping particularly for protein crystals with high solvent content. The disordered solvent density in the large solvent regions of macromolecular crystals is assumed to be 'flat', which results in a sharpened electron density map in the protein regions (Wang, 1985). Further improvement can be obtained by inverting the electron density in the solvent region using a flipping factor (Abrahams *et al.*, 1996). Histogram matching is a technique from image processing, that forces the density to contain the right distribution of values and is only applied to the density in the protein region. This method is less powerful than solvent flattening in improving phases, but useful to extent phases to higher resolutions (Zhang & Main, 1990). For crystals with more than one molecule in the asymmetric unit, non-crystallographic symmetry averaging is a very powerful technique for phase improvement. Portions of the map can, according to non-crystallographic symmetry elements in the asymmetric unit, be superimposed and averaged, resulting in better electron density maps (Vellieux, 1998).

1.3.2.5 Model refinement

Usually the accuracy of first derived model by for instance MR or MIR, is small and represents only an approximation of the real structure. So the agreement between structure factors calculated from this first model with the observed structure factors is rather poor. The *R-factor*, as defined in Equation 1.11, is used as an indicator for the refinement process and therefore the agreement between the calculated and observed structure factors.

$$R = \frac{\sum |F_{obs} - F_{calc}|}{\sum |F_{obs}|} \quad (1.11)$$

where F_{obs} is the observed structure factor amplitude and F_{calc} is the expected structure factor amplitude of the current model.

So refinement is the process of adjusting the model so that the difference between the $|F_{hkl}|_{calc}$ and $|F_{hkl}|_{obs}$ is minimized. These modifications of the model include the positional parameters (x, y, z) and the temperature factor (B-factor) of the non-hydrogen atoms of the protein. Thus the number of parameters is four times (or more) the number of atoms, but the number of observations (measured reflections) relative to the number of atoms in a protein only gives a poor observation-to-parameter ratio. This ratio can be improved by adding extra 'observations' in the refinement process, that is using stereochemical data, like chirality, bond lengths and torsion angles from small molecule structures. This stereochemical information can be used in the refinement in two ways:

- i) They can be taken rigid and only dihedral angles can be varied. So the number of parameters is reduced and the geometry and the refinement are called *constrained*.
- ii) A variation of the stereochemical parameters *restrained* around a standard value is allowed. In this case the atom positions are variables and the restraints are on bond length, bond angle, torsion angles, bond planarities and van der Waals contacts. Because a penalty is given for a disagreement with the restraint these are true observations. So small parts of the structure can be moved easily, but not a large part of the structure, like a domain (Drenth, 1999).

There are a number of commonly used mathematical refinement techniques as for example the methods of least squares, maximum-likelihood and simulated annealing. The method of least squares and maximum-likelihood are based on a weighted function to be minimized to improve the fit between $|F_{hkl}|_{calc}$ and $|F_{hkl}|_{obs}$ (Eq. 1.12).

$$\Phi = \sum_{hkl} \omega_{hkl} (|F_{obs}| - |F_{calc}|)_{hkl}^2 \quad (1.12)$$

where ω_{hkl} is an applied weight dependent on the reliability of the measured intensity

The stereochemical observations are added in the form of a standard deviation and therefore a weight, to restrain the model close to ideal geometry.

Isotropic, anisotropic temperature factor and TLS refinement

A crystal structure does not only provide a static snapshot of molecules packed in the crystal lattice, as generally thought, but a blurred view in which the blurring indicates motion and direction of the movement. Therefore it is possible to extract information about the dynamic properties of the molecules in the crystal. The two main reasons underlying the blurring effect are the dynamic and static disorder present in all crystals (Bürgi & Capelli, 2000, Bürgi, 2000). The dynamic disorder is caused by thermal vibrations and mainly gives rise to a blurred view of individual atoms or small groups of atoms. Static disorder is caused by large-scale motions due to the trapping of individual molecules in different micro conformation, which correspond to the states sampled along a trajectory of motion. In contrast to small molecule crystals in which the molecules are tightly packed, protein crystals have a high solvent content and can therefore be relatively loosely packed. Thus domains or subunits may

move relative to each other, secondary structure elements can be locally displaced and individual side chains may exhibit torsional flexibility (Painter & Merritt, 2006a).

Conventionally additional parameters have been introduced in crystallographic models to take into account dynamic and static disorder by defining the probabilistic displacement of each atom about its mean position in the crystal lattice (Trueblood *et al.*, 1996, Willis & Pryor, 1975). The so called *atomic displacement parameters* (ADPs) or *temperature factors* (B-factor) as well as the *occupancy* of each atom are included in the refinement (Eq. 1.13) to minimize the difference between F_{obs} and F_{calc} and therefore improving the fit between the observed and model structure factors.

$$F_{calc} = G \sum_j n_j f_j e^{2\pi i(hx+ky+lz)} e^{-B_j \left[\frac{\sin\theta}{\lambda}\right]^2} \quad (1.13)$$

where \mathbf{G} is an overall scale factor,

n_j is the occupancy,

f_j is the scattering factor,

\mathbf{x} , \mathbf{y} , \mathbf{z} are the coordinates,

and B_j is the temperature factor of atom \mathbf{j} .

The B-factor or ADP can be described either as *isotropic* or *anisotropic*. The isotropic ADP is described by a spherical Gaussian and requires a total of four parameters (x, y, z, B_{iso}) and is a correction of the atomic scattering factor by the displacement. The anisotropic ADP describes a trivariate Gaussian probability density function for the location of the atomic centre and requires a total of nine model parameters ($x, y, z, U^{11}, U^{22}, U^{33}, U^{12}, U^{13}, U^{23}$). The real space representation of U^{ij} , which is a probability tensor, is commonly known as the 'thermal ellipsoid'. Only very high resolution diffraction experiments

can provide enough observation to support refinement of nine parameters per atom.

An alternative way to describe the local positional displacement of the atoms in a crystal structure is TLS (Translation /Liberation /Screw), which is based on the assumption that all atoms in the TLS groups have amplitudes (in 3 dimensions) appropriate to a rigid body and that all atoms move in concert. TLS parameters describe the anisotropic motion of the TLS groups and an anisotropic U factor can be derived for each atom in a TLS group. Since the atoms belong to rigid bodies the U factors can be correlated and only 20 refinement parameters are required for each TLS group (Howlin *et al.*, 1989, Schomaker & Trueblood, 1968). Thus anisotropic displacements can be included in the refinement without the need for the large number of parameters for full anisotropic refinement. Therefore TLS refinement can improve the refinement particularly at medium resolution ($\sim 2\text{\AA}$).

Simulated annealing

In crystallography the introduction of model bias during building and refinement can be a problem. Initial phases derived from a known structure by molecular replacement or with error in experimental phases can lead to model bias overfitting, which is caused by the low parameter-to-observation ratio combined with initial errors in the model. A method to overcome overfitting and the model being trapped in a local (false) minimum is *simulated annealing*, which uses a molecular dynamics approach (Brünger, 1988). In the simulation a 'temperature' parameter is introduced by heating up the system from room temperature (300 K) to 2000- 5000 K and subsequently cooling it down to the initial value. During this simulation, the atoms fluctuate around their equilibrium positions, resulting in a non-zero gradient. The family of conformations resulting from this molecular dynamics calculation are constrained

by the X-ray data and restricted to those with lower R -factors. This method has a larger radius of convergence than restrained positional refinement, but is only useful early in the refinement procedure.

1.3.2.6 Validation of the model

A model of a protein structure may still contain errors after the refinement process, due to mistakes made in the interpretation of the electron density map, particular in regions with weak density. Therefore the assessment of the model quality is important. Here a brief overview of the main points that are important for the assessment of the quality of both the diffraction data and a model will be given.

Quality of the diffraction data

A number of criteria to access the quality of diffraction data has been developed, which are in general quoted for the entire data set as well as for the highest resolution shell. A statistical measurement for the quality of diffraction data is the R_{merge} (Eq. 1.14)

$$R_{merge} = \frac{\sum_h \sum_i |I_{h,i} - I_h|}{\sum_h \sum_i I_{h,i}} \quad (1.14)$$

where the outer sum (h) is over the unique reflection and the inner sum (i) is over the set of independent observations of each unique reflection (Drenth, 1999). The R_{merge} should reflect the spread of multiple observations of the intensity of the unique reflections. Unfortunately the R_{merge} alone is not a universally good statistic since its value increases with the increase of the redundancy of the data, despite the higher signal-to-noise ratio due to the inclusion of more reflections (Diederichs & Karplus, 1997, Weiss & Hilgenfeld, 1997). Other criteria to assess the quality of the data include *redundancy* and *completeness*. Redundancy is defined as the number of independent observa-

tions per unique reflection in the final merged and symmetry-reduced data set. Completeness is defined as the fraction of unique reflections that actually have been measured within a range of Bragg spacing, as a proportion of those that could theoretically be measured. Moreover the strength of the signal of the observed intensities as average value of $I/\sigma(I)$ is an indicator for significance of the measurements. The nominal resolution limit of a data set is chosen at the stage of data processing, taking into account not only the range of Bragg spacing but also R_{merge} , redundancy, completeness and $I/\sigma(I)$ (Kleywegt, 2000).

Model bias and R_{free}

A problem of model-building and refinement especially at low and medium resolution is the introduction of model bias by overfitting, because of the poor observation-to-parameter ratio, typical for macromolecular crystallography. That means that while apparently improving the fit of the model to the experimental data and lowering the R -factor, systematic errors are introduced by adding further parameters-fit noise. Therefore an R -factor has been introduced, which is unbiased ('free') from the refinement process (Brünger, 1992). For this so called 'cross-validation', 5-10% of unique (=free set) reflections are excluded from the refinement process, which is carried out using the remaining observations (=working set). The free R -factor is calculated only with the free reflections (Eq. 1.15).

$$R_{free} = \frac{\sum_{hkl \in free} ||F_{obs}| - K|F_{calc}||}{\sum_{hkl \in free} |F_{obs}|} \quad (1.15)$$

So in principle if the model is improved in a refinement step, R_{work} and R_{free} will decrease. Introduction of model bias, by for example fitting to many water molecules in noise peaks of the electron density map will decrease R_{work} , but increase R_{free} (Drenth, 1999, Kleywegt, 2000, Kleywegt & Jones, 1995).

Quality the model

Based on the analysis of the Cartesian coordinates of a model, its quality can be evaluated using computer programs like WHATIF (Vriend, 1990) and PROCHECK (Laskowski *et al.*, 1993). Additionally model validation tools have also been implemented in the model building program COOT (Emsley & Cowtan, 2004). Validation of the model includes the assessment of the geometry and stereochemistry by comparison of the bond lengths and angles to a library of 'ideal' values and quoting their r.m.s. deviations. Moreover the configuration and chiral centres at the C $^{\alpha}$ atom of the main chain and C $^{\beta}$ atoms of threonine and isoleucine, which is important in the case of biomacromolecules, can be assessed in terms of improper torsion angles and chiral volumes. Another important point is the planarity of the peptide bond and of moieties like carboxylate groups and phenyl rings, which can be evaluated either by the r.m.s. deviation of their torsion angles from ideal values or by fitting a least-square plane through each atom set and calculating the r.m.s. distance of each atom to the plane.

Traditionally a first indicator for the quality of the model has been the *Ramachandran plot* (Ramachandran *et al.*, 1963). The conformation of the backbone is determined by three torsion angles $\varphi(C_{i-1} - N_i - C_i^{\alpha} - C_i)$, $\psi(N_i - C_i^{\alpha} - C_i - N_{i+1})$ and $\omega(C_i^{\alpha} - C_i - N_{i+1} - C_{i+1}^{\alpha})$, which are highly constrained by steric hindrance. Due to the partial double bond character of the peptide bond, ω is restrained to values near 0° (*cis*-peptide, relatively rare) and 180° (*trans*-peptide). φ and ψ torsion angles are less restricted, but have preferred combinations of φ , ψ values because of steric hindrance (Ramakrishnan & Ramachandran, 1965). High-resolution structures have shown that in some rare cases, energetically unfavorable, unusual φ , ψ combinations occur in proteins for functional or structural reasons (Kleywegt, 2000, Lovell *et al.*, 2003).

The main stabilising forces in a protein structure are hydrophobic, electrostatic and hydrogen-bonding interactions. These can also be indicators for errors in the model, like when simple rules of physical interactions are violated and unusual short non-bonded contacts are contained in the model (Kleywegt, 2000).

1.4 Aims of the thesis

Iron haem acquisition is essential for the survival of bacteria and, especially for pathogenic bacteria, to establish and maintain infection. Knowledge of the structures of haem-sequestering proteins and their complexes, unique and essential to bacteria, will provide penetrating insights into their mechanism of binding and release. This knowledge could help to develop new antibacterial agents which would be very useful given the increasing number of multi-drug-resistant pathogens.

This project aims at investigating structure-function relationships in novel bacterial haem-binding proteins using mutagenesis, modelling, UV/vis absorption spectroscopy and X-ray crystallography. The specific objectives of the research are listed below:

- Cloning, over-expression and purification of a set of bacterial haem transport proteins.
- Biochemical characterisation of protein-haem interactions with UV/vis absorption spectroscopic methods.
- Secondary and tertiary structure prediction and homology modelling studies.
- Crystallisation and structure determination using X-ray crystallography.
- Site-directed mutagenesis based on multiple sequence alignments and homology modelling studies and/or structural data.

Chapter 2

General materials and methods

2.1 Materials and chemicals

All chemicals were purchased from Sigma and all enzymes were purchased from New England Biolabs unless otherwise stated.

2.2 Methods

2.2.1 Bacterial strains and growth media

BL21 (DE3) (Novagen) is an *Escherichia coli* B strain derived from B834 and is deficient in the ATP-dependent protease Lon, and the OmpT outer membrane protease, found to be the cause of the degradation of T7 RNA polymerase (Derman *et al.*, 1993, Dubendorff & Studier, 1991).

Genotype: F⁻ *ompT lon hsdS_B (r_B-m_B) gal dcm* (DE3)

For production of Selenomethionine labeled protein, expression plasmids were transformed into B834 (DE3) cells (Novagen) a methionine auxotroph strain (Studier & Moffatt, 1986).

Genotype: F⁻ *ompT hsdS_B (r_B-m_B) met dcm lon* (DE3)

Used growth media are listed in Table 2.1.

Table 2.1: Bacterial growth media reference

Luria Bertani (LB)^a	2xYT^b	SOB Medium^c	SOC Medium^d
10 g/l tryptone	16 g/l tryptone	20 g/l tryptone	20 g/l tryptone
5 g/l yeast extract	10 g/l yeast extract	5 g/l yeast extract	5 g/l yeast extract
5 g/l NaCl	5 g/l NaCl	10 mM NaCl	10 mM NaCl
		2.5 mM KCl	2.5 mM KCl
			20 nM glucose

^aplasmid preparation

^bprotein expression

^ccompetent cells

^delectroporation

Agar plates were prepared by adding 14 g bacto-agar per liter LB medium. Electro-competent BL21 cells were prepared as previously described (Sambrook & Fritsch, 1989).

2.2.2 Polymerase chain reaction

2.2.2.1 PCR from genomic DNA

The *polymerase chain reaction* (PCR) was optimised for annealing temperature, MgSO₄ concentration and addition of DMSO for particular template/primer pairs. A typical example for PCR reaction mix and protocol is shown below:

A 100 μ l PCR reaction mix included:	Used PCR program:
2 μ l genomic DNA	92°C, 30 sec
600 nM primer (for)	95°C, 2 min
600 nM primer (rev)	55°C, 1 min
0.25 mM dNTPs	72°C, 1.5 min
10 μ l 10 x ThermoPol reaction buffer	72°C, 1 min
2 U Vent DNA polymerase	

The annealing temperature varied depending on the used primers (see Appendix, p. xix) and template. The reactions were carried out in a Peltier Thermal Cycler (PTC2000).

2.2.2.2 Single colony polymerase chain reaction

Colonies obtained from ligation and transformation were screened by *single colony PCR*. Half a colony was picked, introduced in 10 μl H₂O and incubated for 8 min 95°C. One μl was then added to 9 μl of PCR-mix and the program below was used. For a positive result from this procedure, the plasmid of an over-night culture was purified and sent for sequencing to check the correctness of the sequence.

A 14x PCR reaction mix included:	Used PCR program (30 cycles)
1.4 μl Polymerase	2 min 95°C
1.4 μl dNTP's (25 mM each)	30 sec 92°C
1.7 μl primer forward (50 μM)	1 min annealing
1.7 μl primer reverse (50 μM)	1-2 min 72°C
14 μl 10 x Buffer	2 min 72°C
105.8 μl H ₂ O	4°C
Σ 126 μl	

2.2.3 General cloning methods

Restriction digestion

For cloning purposes approximately 10 μg DNA was incubated with restriction endonucleases for 3 h at the recommended temperature. Dephosphorylation of the backbone was carried out with SHRIMP alkaline phosphatase (Fermentas) by adding 2 units phosphatase to the restriction digestion mix following the digestion reaction and incubating the mix for additional 30 min at 37°C before heat-inactivating the enzyme at 65°C for 20 min. The DNA fragments were gel purified and their relative quantities estimated by agarose gel electrophoresis.

Ligation and transformation

For ligation DNA fragments were mixed in a one to three molar ratio (backbone to insert) in a 20 μl ligation reaction with 1 unit T4 DNA ligase (Roche)

and 2 μl 10x ligase buffer (Roche) and incubated over night at 4°C. As a negative control a ligation reaction containing water instead of the insert was set up.

An aliquot of 3 μl of the ligation reaction was used in a transformation into electro-competent *E. coli* BL21 (DE3) cells using a 0.1 mm gap cuvette (BioRad) and 1.5 kV in a BioRad MicroPulser Electroporator. After the transformation, 1 ml of SOC media (Table 2.1) was added and the cells were incubated for 30 min at 37°C. The cells were pelleted by centrifugation (10 min, 3500 rpm), resuspended in 200 μl medium, plated on LB-ampicillin agar plates and incubated at 37°C overnight. Colonies were analysed by single colony PCR.

2.2.4 Protein expression and purification

For protein expression, transformed *E. coli* BL21 (DE3) cells were grown in 2xYT media, supplemented with 50 $\mu\text{g ml}^{-1}$ carbenicillin, at 37°C and shaken at 200 rpm in baffled flasks. Once an OD_{600} of 0.6-1.0 was reached, protein expression was induced by adding IPTG to a final concentration of 0.4 mM. The induction time and temperature varied between different proteins from OD_{600} of 0.6-1.0 and 20-30°C.

Cultures were further grown overnight and shaken at 160 rpm. Cells were harvested by centrifugation at 2360 g and then resuspended in lysis buffer with a cocktail of protease inhibitors (Roche) and finally lysed using a French Press (3 times 700 Psi) or sonication (3 times, 1 min 5 sec on, 5 sec off; 10 micron). The lysate was cleared by two centrifugation steps (15,000 and 30,000 g for 30 min at 4°C), was passed through a 0.45 μm filter and incubated at 4°C for one hour with 5 ml nickel-nitrilotriacetic acid (Ni-NTA) Superflow (Qiagen) slurry per liter of culture. The nickel resin had been pre-equilibrated with lysis buffer, and imidazole was added to a final concentration of 5-10 mM to

minimise non-specific binding. The resin was applied to a column and washed by gravity flow with a series of four wash-buffer solutions (50 mM sodium phosphate, 300 mM NaCl, pH 8.0), containing 5, 10, 15 and 25 mM imidazole, respectively. An optional high-salt washing step was performed with washbuffer containing 1 - 2 M NaCl and 25 mM imidazole. The protein was finally eluted with elution-buffer containing 150 mM imidazole. The buffer was exchanged using centrifugal membrane-devices (Vivaspin), dialysation or a PD-10 column (GE Healthcare). The protein concentration was determined by spectroscopic measurement of the absorbance at 280 nm and an estimated millimolar extinction coefficient (<http://www.scripps.edu/~cdputnam/protcalc.html>).

The N-terminal tag (His and/or GST) was removed by proteolytic cleavage with thrombin (HTI). Units of thrombin per mg protein as well as incubation time and temperature were optimised for individual proteins. Usually approximately 5 U thrombin/mg protein were added to the protein solution and was incubated over night at RT. Thrombin was removed by incubation with 2 μ l p-aminobenzamidine-agarose beads per U of thrombin. The thrombin-loaded beads were separated with a column while the flow-through containing the protein was recovered. The agarose beads were washed with 2 M glycine pH 3, stored in 500 mM NaCl at 4°C. Finally, to remove uncleaved protein and the cleaved tags, the solution was incubated with the Ni-NTA resin for 1 h at 4°C, then the resin was packed on a column and the flow-through, containing un-tagged protein, was collected. Quality and integrity of the proteins were tested by SDS-PAGE.

2.2.5 UV/vis absorption spectroscopy and reconstitution of proteins with haemin

All UV/vis absorption spectra were recorded on a Cary 100-Bio Varian spectrophotometer from 260 - 800 nm in 50 mM Tris-HCl pH 7.5, 150 mM NaCl at

25°C. The haemin stock solution was prepared by dissolving haemin chloride in 0.1 M NaOH and the concentration was determined from the absorbance at 385 nm (Choi *et al.*, 1999). The stock solution was kept in the dark on ice and used within 5 hours. Cyanide binding was carried out by adding potassium cyanide to a final concentration of 5 mM to a cuvette containing a solution of reconstituted ferric protein (5 μ M).

For reconstitution of protein, the sample was saturated with haem by incubation with 3 molar excess haemin for one hour at RT or 4°C. A final gel-filtration purification step (with the resin GLC 300, Isco Inc.) was carried out to remove excess haemin. The protein was then concentrated with centrifugal membrane-devices (Vivaspin) in an appropriate buffer to the desired concentration.

2.2.6 Fast protein liquid chromatography

Fast protein liquid chromatography (FPLC) was carried out on a GE Healthcare chromatography system (Äkta Purifier). For analytical gel-filtration the system was equipped with a HiLoad Superdex 75 HR16/60 column (GE Healthcare). For ion-exchange chromatography a MonoQ HR5/5 column (GE Healthcare) was used. Representative traces were recorded both at 280 nm and the Soret maximum depending on the haem-protein.

2.2.7 Crystallisation of proteins

Solubility of proteins were first tested by a pre-crystallisation test (PCTTM, Hampton Research) to determine the most appropriate protein concentration to use. High throughput crystallisation screening was carried out using the sitting-drop vapor-diffusion method in 96-well Intelliplates (Hampton Research) or Crystalquick plates (Greiner). Manual optimisation of the hit-condition was carried out either by the sitting-drop (Crychem plates; Hamp-

ton Research) or hanging-drop (Linbro plates; Molecular Dimension) vapor-diffusion method. Crystallisation screens were purchased from Molecular Dimensions, Nextal Biotechnology/Qiagen and Hampton Research. The experiments were set up with the Hydra II micro dispensing system (Robertson Scientific).

2.2.8 Secondary and tertiary structure prediction

For the secondary and tertiary structure prediction the servers listed in the table below were used.

Table 2.2: Secondary and tertiary structure prediction server.

secondary structure prediction	link	comment
Jpred	www.compbio.dundee.ac.uk/~www-jpred/	consensus Method
PSIPRED	bioinf.cs.ucl.ac.uk/psipred/	analysis output from PSI-blast (position specific iterated -blast)
tertiary structure prediction	link	comment
SAM-T02	www.soe.ucsc.edu/research/compbio/HMM-apps/T02-query.html	secondary structure prediction with multiple sequence alignments, builds hidden markov model for searching the pdb
GenTHREADER	bioinf.cs.ucl.ac.uk/psipred/	sequence profile based on fold recognition, uses a position score matrix
FUGUE	www-cryst.bioc.cam.ac.uk/~fugue/prfsearch.html	searches sequence alignments against a library of profiles, employs environment specific substitution tables
3D-PSSM	www.sbg.bio.ic.ac.uk/~3dpssm/	sequence profile based on fold recognition, uses a position score matrix
bioingbu / ShotGun on 5	ww.cs.bgu.ac.il/~bioingbu/form.html ; //bioinfo.pl/Meta/	consensus prediction utilizing models produced by other server
ESyPred3D	www.fundp.ac.be/urbm/bioinfo/esypred/	weighting and screening of results from several multiple alignment programs

2.2.9 X-ray diffraction, data processing and structure refinement

Optimisation of cryo-conditions and initial X-ray diffraction analysis were carried out on the in-house X-ray facilities consisting of a Rigaku MicroMax 007 rotating anode generator with a rotating copper anode equipped with VariMax HF optics (Osmic, Rigaku), an R-Axis IV⁺⁺ image plate detector and the X-stream 2000 cryo-cooling vapour jet. Data were also collected at beamlines of the European Synchrotron Radiation Facility (ESRF).

Data were processed with programs of the CCP4 suite (CCP4, 1994), HKL2000 (Otwinowski & Minor, 1997) and/or XDS (Kabsch, 1988). Structural refinement was carried out with REFMAC 5.0 (Murshudov *et al.*, 1997) and TLS (Painter & Merritt, 2006a,b) Structural figures were prepared using PyMOL (Delano, 2002) or ASTEXVIEWER (Hartshorn, 2002).

Chapter 3

Diversity and conservation of interactions for binding haem

3.1 Introduction

Haem-proteins are extremely diverse in their molecular architecture and function despite containing the same prosthetic group as highlighted in Chapter 1, Table 1.1, p. 4 and Figure 1.2, p. 3. Due to the explosive growth of the structural biology field in the past decade, the structures of a large number of novel b-type haem proteins have been determined, such as the insect protein nitrophorin (Arnoux *et al.*, 1999), the enzyme NO synthase (Fischmann *et al.*, 1999), the bacterial siderophore HasA (Weichsel *et al.*, 1998), the mammalian transport protein haemopexin (Paoli *et al.*, 1999), the binding protein albumin (Wardell *et al.*, 2002) and the molecular sensors CooA (Lanzilotta *et al.*, 2000) and EcDos (Kurokawa *et al.*, 2004). Evidently, the evolutionary pressure of generating various biochemical functions, combined with the reactivity and versatile chemical properties of haem, have led to the rise of very different structural associations between haem and proteins. Even in the case of the evolution of a particular function, multiple paths can be followed that

result in proteins with distinct folds, structures and mechanisms. This is true for the mammalian blood serum glycoprotein haemopexin and the bacterial siderophore HasA which have broadly equivalent roles of sequestration, transport and receptor-mediated release of haem but are structurally unrelated.

Despite the diversity in folding topologies, it is tempting to speculate whether any interactions between the protein environment and the haem ligand are shared by the different architectures. Historically, it was assumed that the bond(s) between the haem iron and the amino acid(s) coordinating the iron is/are the major force holding the haem into the protein. However, when in globins and cytochromes the proximal histidine was mutated into a glycine and its side chain replaced by an imidazole (Barrick, 1994, McRee *et al.*, 1994), the proteins could still incorporate haem without directly ligating the iron. Hence the protein framework has to provide sufficient interactions for the haem-binding within a relatively specific binding pocket. Work on many 'designer' haem proteins such as molecular maquettes (Rabanal *et al.*, 1996) and others (Gibney & Dutton, 1999, Rojas *et al.*, 1997) has further shown the importance of these non-covalent interactions. In many engineering experiments the main focus was on the positioning of histidine residues, resulting in haem binding at the expense of structural stability. These 'designed' molecules often have low haem affinity, as well as a molten globule-like characteristics and are lacking tightly packed interiors. It was previously recognised that haem-contacting residues other than the histidine ligands are important for haem binding (Robertson *et al.*, 1994) thus one can ask if any common, key structural features or re-occurring haem-protein packing interactions might exist in distinct folding environments.

Comparison of a wide range of haem proteins known to date may give insights on the structural requirements of a haem binding site. Here a comparative analysis of structurally unrelated haem-proteins was carried out by means of

multiple structural overlays. The b-type haem proteins listed in Table 1.1 (and Table A.4, p. xxiii) are functionally dissimilar. Any shared interactions should therefore play a primarily structural role in the association with the haem. This study shows that despite the obvious differences in the haem environments which are dictated by the different functions, a specific molecular pattern emerges for the contact points between haem and the protein binding pocket.

3.2 Methods

Structures and their haem environments of 68 b-type haem proteins with less than 60% sequence identity and representatives of over 20 different folds (Table A.4, p. xxiii) have been compared by least-squares superposition of the haem atoms. In some cases where significant sequence divergence exists between two haem proteins despite their equivalent folding topologies, multiple representatives of the same fold were analysed (Table A.4, p. xxiii). For instance, molecules such as sulfite oxidase (PDB code 1SOX) were included because its cytochrome b5 domain is structurally equivalent to the rat cytochrome b5, despite the lack of significant sequence identity (Kisker *et al.*, 1997). Another example is given by the structure of soluble guanylate cyclase (PDB code 1U55) which is not included in Table 1.1, p. 4 because it has the same fold as SONO, but was considered in our analysis given the lack of significant sequence similarity (Nioche *et al.*, 2004, Pellicena *et al.*, 2004).

Atomic coordinates from the Protein Data Bank (Bernstein *et al.*, 1978) were superimposed using a subset of atoms of the porphyrin ring as a reference frame. Firstly a rotation/translation matrix was calculated from the best least-squares fit of the porphyrin atoms with the addition of, in some cases, the proximal histidine imidazole. Secondly the matrix was applied to the whole

protein. The computation of the overlays was carried out using the fitting procedures implemented in DeepView (Guex *et al.*, 1999) and visual inspections were carried out using both DeepView and the program COOT (Emsley & Cowtan, 2004). Further a list of all amino acids within 4.5 Å of any of the haem atoms was produced with the program DISTANG (CCP4, 1994).

3.3 Results and Discussion

3.3.1 Diversity of haem-protein packing contacts

Least-squares superpositions of all selected structures (using the atoms of the porphyrin ring as a frame of reference) highlight both differences and similarities between various haem-proteins. Evolutionary constraints imposed by the specific function of each protein presumably dictate many differences in the haem environments, the structure of which has been 'sculpted' to modulate the reactivity of the haem. The structural overlays highlight that the strikingly different folding topologies are associated with an entire spectrum of haem arrangements. For example, haem is buried inside the molecule like in catalases with an average solvent accessibility of 1.4 %, or is bound in a pocket near the surface, often relatively exposed to solvent, like in haemopexin with an average solvent accessibility of 25 % (Table 3.1 and Table A.4, p. xxiii). Its orientation is not fixed, in that the propionate groups can point either towards the outside (e.g. HasA PDB code 1DK0) or the interior (e.g. HemS PDB code 2J0P) of the protein. In the range of proteins surveyed, the haem environment is formed from either α -helical or β -extended structures, or both, as well as loop regions and excursions from secondary elements. Although the packing of amino acid residues around the haem is generally tight, there are examples of unusual and relatively open haem pockets, such as those found in haemopexin, cellobiose dehydrogenase, cystathione β -synthase and CoxA.

Table 3.1: Haem protein folds with their average solvent accessible area as calculated in AREAIMOL (CCP4, 1994). Percent haem exposure is calculated against free haem. *Large difference were observed between solvent accessibility in PAS domain proteins EcDos (20 %) and FixL (5 %). For a detailed list see Table A.4, p. xxiii.

Protein	No. of Structures	Average haem solvent accessibility	
		Å ²	%
free haem	-	829	100
4-helix bundle	3	166	20
<i>beta</i> -propeller	1	210	25
CAP	1	233	28
catalases	8	11	1
cupredoxin-like	1	32	4
cytochrome b5	4	184	22
dioxygenase	1	92	11
globin like	12	135	16
haem-oxygenase	6	160	19
haem-peroxidase	6	72	8
HemS-fold	1	155	19
H-NOX-fold / SONO-fold	2	19	2
immunglobulin-like	1	179	22
lipocalin	1	130	16
meander	1	191	23
NO fold	4	128	15
P450	14	22	3
PAS*	2	111	12
serum albumin-like	1	66	8
vitamin b6 family	1	148	18

3.3.2 Re-occurring contacts: protein-haem interaction 'hot spots'

Despite the structural and functional diversity of the here analysed proteins, preferences for using specific amino acids in the interactions with haem could be identified. This observed re-occurrence of some shared key contacts in distinct proteins, is discussed in more detail below and illustrated in Figure 3.1, Table 3.2, 3.3 and 3.4. These contacts were categorised in three groups, depending on whether they interact with i) the propionate groups, ii) the plane of the pyrrole rings or haem face, and iii) the haem edge defined by the perimeter of the porphyrin including the atoms of the methyl and vinyl groups.

3.3.3 Anchoring of haem propionates by arginine residues

Commonly one or both propionate groups of the haem in the 20 folds analysed, are engaged in electrostatic interactions, which are in general salt-bridges. The majority of these contacts, 38 %, are made with arginine side chains. This can be observed in proteins sharing some sequence similarity, like P450cam, P450terp and P450BM-3 (Hasemann *et al.*, 1995), as well as in functionally unrelated molecules. Principally arginine residues appear to be universal partners of the propionate groups, providing them with anchoring points to the protein (Table 3.2 and A.4, p. xxiii), independent of the haem orientation. For instance in proteins with a relatively solvent accessible haem-binding site, the propionates are directed more often towards the outside, or are lying on the protein surface, partly excluded from the solvent by interactions with arginines. In addition arginine residues also engage in interactions with the propionates in cases where the haem is buried in the protein, like in cytochrome, P450 and catalase, or when the acidic groups are pointing towards the protein interior, like in haemopexin or HemS. Residues also found in contact with the negatively charged propionates, are positively charged lysine and histidine. This can be observed in for example in the globin-fold and P450 peroxidases, but is less common with only 10 % and 17 %, respectively, of all observed contacts. Tyrosine residues can also be involved in electrostatic interactions with the propionate groups, surprisingly almost as often as lysines (Table 3.2). In some cases, as in the peroxidase domain of prostaglandin H2 synthase-1 (PDB code 1PRH), nitrous oxide synthase and haem oxygenase, the propionates are bound to the backbone amides of loop regions that embrace part of the prosthetic group.

Table 3.2: Frequencies of residues in contact with the haem propionate groups in b-type haem proteins with distinct folding topologies. Distances were calculated using DISTANG (CCP4, 1994) and residues in hydrogen bond distance of 2.5 - 3.2 Å to the propionate groups were taken into account. Percentages are relative to the number of contacts in a particular fold. A detailed list can be found in Table A.4, p. xxiii.

Fold	Propionate contacts (2.5 - 3.2Å) in %				
	Arg	His	Lys	Tyr	other
other 4-helix bundle	100.0	-	-	-	-
β -propeller	33.3	16.7	-	50.0	-
CAP	-	-	-	-	100.0
catalases	82.9	8.6	-	-	8.6
cupredoxin-like	75.0	-	-	-	25.0
cytochrome b5	-	10.0	20.0	10.0	60.0
dioxygenase	75.0	-	-	-	25.0
globin like	2.9	17.1	34.3	11.4	34.3
haem-oxygenase	37.5	-	31.3	25.0	6.3
haem-peroxidase	10.3	24.1	10.3	-	55.2
HemS-fold	57.1	-	14.3	14.3	14.3
H-NOX-fold / SONO-fold	70.0	-	-	10.0	20.0
immunoglobulin-like	-	-	-	40.0	60.0
lipocalin	-	-	66.7	-	33.3
meander	-	16.7	-	16.7	66.7
NO fold	20.0	-	-	80.0	-
P450	47.4	28.2	2.6	3.8	17.9
PAS	28.6	57.1	-	-	14.3
serum albumin-like	-	50.0	50.0	-	-
vitamine b6 family	100.0	-	-	-	-
Overall %	38.4	17.2	10.4	9.0	25.0

3.3.4 Conserved interactions at the 'haem face' involving leucine

The majority of residues making contact with the haem-face are leucines (12%), isoleucine (6%) and valine (13%). They may be found on both proximal and distal sides, making non-polar contacts to yield a close fit between haem and protein atoms (Table 3.3). The most common haem-face contact, seen in at least twelve distinct folding topologies, involves leucine and isoleucine residues located approximately over the bridge between pyrrole rings II and III, as depicted by the cluster of side chains shown in Figure 3.1A.

It is remarkable that in twelve distinct, unrelated structures side chains can be

Table 3.3: Frequencies of residues in van der Waals contact (3.6 - 4.1 Å Vriend (1990)) with the haem-face in different folding topologies of b-type haem proteins. Distances were calculated using DISTANG (CCP4, 1994) and percentages are relative to the number of contacts in a particular fold A more detailed list can be found in Table A.5, p. xxiv

Fold	Haem-face Contacts (3.6 - 4.1Å) in %							
	Ala	Ile	Leu	Phe	Trp	Tyr	Val	other
4-helix bundle	-	30.0	20.0	10.0	10.0	-	-	30.0
β -propeller	-	-	-	-	33.3	33.3	-	33.3
CAP	-	14	28.6	-	-	-	-	57.1
catalases	-	-	-	20.9	-	-	32.6	46.5
cupredoxin-like	-	22.2	-	33.3	-	-	11.1	33.3
cytochrome b5	7.9	-	10.5	15.8	5.3	5.3	7.9	47.4
dioxygenase	16.7	-	33.3	33.3	-	-	16.7	-
globin like	2.4	9.8	20.7	22.0	-	6.1	28.0	11.0
haem-oxygenase	-	-	4.3	17.4	-	-	4.3	73.9
haem-peroxidase	2.5	-	12.5	-	15.0	-	7.5	62.5
HemS-fold	-	-	-	66.7	-	-	33.3	-
H-NOX-fold / SONO-fold	-	14.3	35.7	14.3	-	14.3	-	21.4
immunoglobulin-like	25.0	-	-	-	25.0	25.0	-	25.0
lipocalin	16.7	-	50.0	16.7	-	16.7	-	-
meander	-	-	20.0	-	-	20.0	20.0	40.0
NO fold	8.3	4.2	-	16.7	16.7	4.2	8.3	41.7
P450	21.2	3.0	8.1	15.2	-	-	6.1	46.5
PAS	-	38.9	22.2	5.6	-	11.1	11.1	11.1
serum albumin-like	-	20.0	-	20.0	-	20.0	-	40.0
vitamin b6 family	14	-	14	-	14	-	-	57
Overall %	7.3	6.2	12.2	15.4	3.8	4.0	12.9	38.1

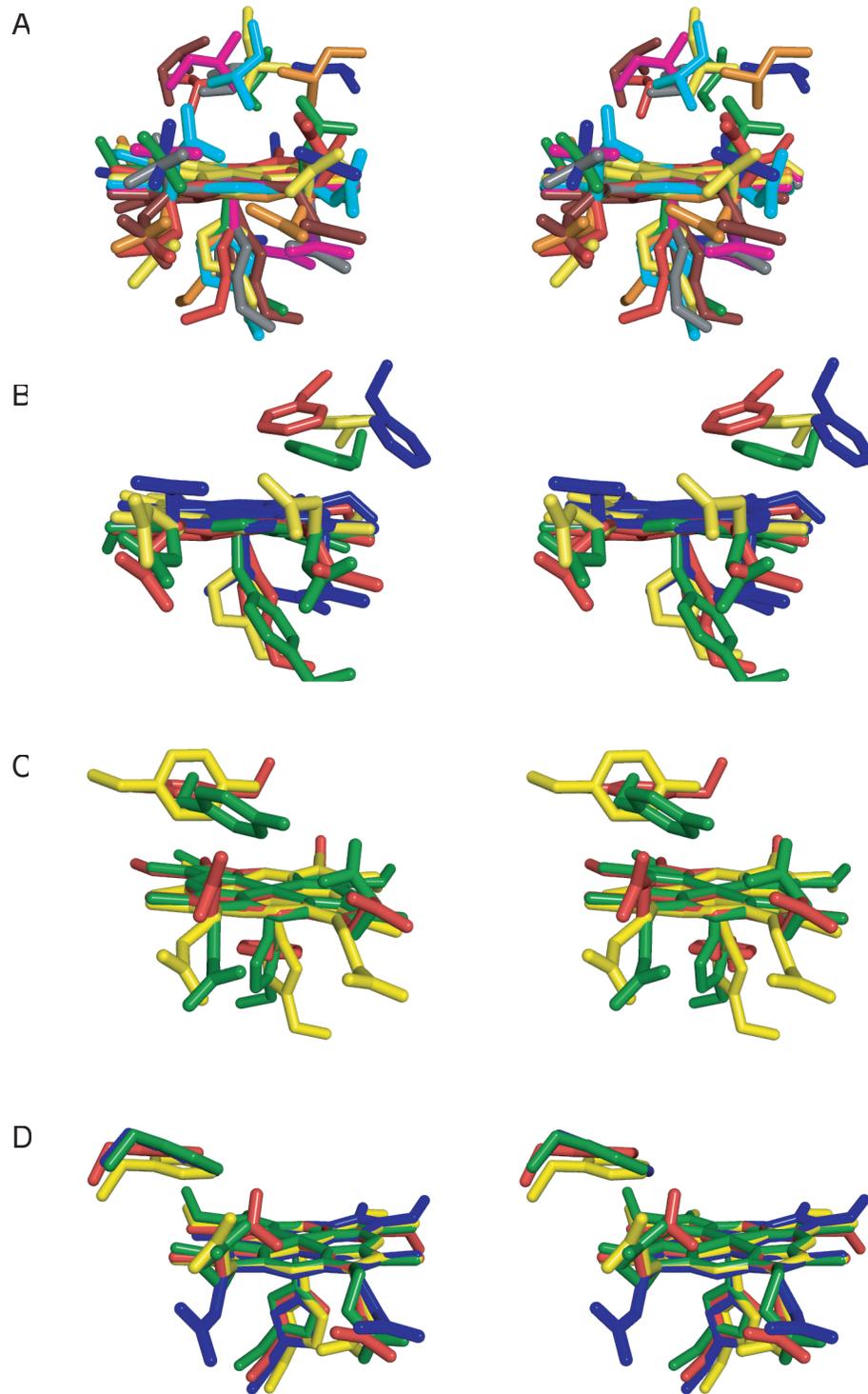
found in the same three-dimensional space with respect to the haem. Leucine and isoleucine side chains can be seen in other proteins, like the globins, at different positions over the porphyrin plane, contributing similar hydrophobic contacts. When in the globin fold the leucine in contact with the haem was mutated to alanine, an increased rate of haemin dissociation was observed (Liong *et al.*, 2001); the fact that this has mainly been attributed to hydration of the environment around the proximal histidine perhaps undermines the simple removal of favourable van der Waals interactions with the concomitant loss of the molecular fit between protein and haem. It has been shown for the NO-transport protein nitrophorin, that the two leucine residues (122 and 132) packing over the haem-face, are necessary to maintain a certain ruffling of the

porphyrin ring and thus stabilising the normally unstable ferrihaem-nitrosyl (Fe(II)-NO) complex (McGaughey *et al.*, 1998). Site-directed mutagenesis of these two leucine residues to a smaller valine causes a shift in the reduction potential of the haem-iron due to allowing a more planar haem-geometry and thus favoring the stabilisation of the Fe(III)-NO complex Samanta *et al.* (1999).

3.3.5 Conserved interactions at the 'haem face' involving phenylalanine/tyrosine side chains

The aromatic residues phenylalanine, tyrosine and tryptophane are responsible for making up a large proportion of the other observed contacts to the haem face (15 %, 4 % and 4 %, respectively; see Table 3.3). They predominantly engage in stacking interactions with the porphyrin and most of these phenylalanine-porphyrin contacts have distances in the range of 3.6-5.0 Å, with an offset face-to-face parallel geometry relative to one of the four pyrrole rings. This is similar to the pairing of aromatic side chains in proteins (Mc-

Figure 3.1 (following page): Stereoviews highlighting structurally conserved packing interactions to the 'haem-face' involving leucine/isoleucine and phenylalanine/tyrosine side chains. (A) Clustering of leucine/isoleucine residues over the haem occurs in eleven distinct proteins; for clarity nine of these structures are shown in the figure: CoxA (1FT9) Leu 112 in red, bacterioferritin (1BFR) Ile 22 in purple, cytochrome b5 (1AWP) Ile 25 in green, nitrophorin (1NP4) Leu 132 in yellow, ubiquinol oxidase (1FFT) Ile 425 in grey, FixL (1DRM) Leu 236 in cyan, serum-albumin (1N5U) Leu 139 in blue, haem oxygenase (1N45) Leu 147 in orange, SONO (1XBN) Ile 75 in dark red (not shown: Ile 57 HasA, Ile 75 guanylate cyclase, Leu 92 and Leu 94 HemS). (B) Conserved phenylalanine residues over the bridge between pyrrole rings III and IV in distinct haem proteins: SONO (1XBN) Phe 78 in red, cytochrome b562 (1QPN) Phe 65 in yellow, catalase (8CAT) Phe 160 in green and bacterioferritin (1BFR) Phe 26 in blue. (C) Conserved tyrosine residues packing over pyrrole I: human serum-albumin (1N5U) Tyr 138 in red, cellobiose dehydrogenase (1D7B) Tyr 90 in green, bacterial SONO (1XBN) Tyr 140 in yellow. (D) Conserved phenylalanine side chains packing in proximity of pyrrole I in the globin fold: rice haemoglobin (1D8U) Phe 54 in green, tuna myoglobin (1MYT) Phe 43 in yellow, mouse neuroglobin (1Q1F) Phe 42 in red, *E. coli* flavohaemoglobin (1GVH) Phe 43 in blue. Figure was prepared with PYMOL (Delano, 2002)



Gaughey *et al.*, 1998, Samanta *et al.*, 1999).

Some of the contacts made between phenylalanine side chains and the porphyrin clearly involve aromatic-aromatic π -stacking interactions and are likely to provide significant stabilisation for haem binding. Moreover examination of van der Waals radii and shapes of side chains within the pocket show that these contacts often result in excellent steric complementarity between the haem and the protein. Strikingly phenylalanine side chains can be found at specific sites on the porphyrin surface in proteins with a completely different molecular architecture and unrelated function. These structurally conserved haem-phenylalanine contacts are highlighted in Figure 3.1B-D. Phenylalanine residues are located over the bridge between pyrrole rings III and IV in the case of cytochrome b₅₆₂, catalase, bacterioferritin and SONO (Fig. 3.1B), which are all proteins with distinct topologies and unrelated functions. Another set of structurally conserved contacts is shown in Figure 3.1C, where three tyrosines pack onto pyrrole I. Similar interactions can be observed in the globin and two-over-two helical fold (Fig. 3.1D); in these cases, interestingly, equivalent phenylalanine contacts are made to different pyrrole rings. After superposition of the structures on both the proximal histidine and the haem, the haems are essentially flipped 180°, with pyrrole I of truncated haemoglobin overlaying onto pyrrole IV of myoglobin. So equivalent aromatic interactions can be observed regardless of the orientation of the haem.

3.3.6 Contacts at the haem edge

Amino acid side chains that are packing around the haem edge in different folding topologies also provide steric complementarity, albeit only limited conservation could be observed in the position of interacting groups. Nevertheless there is a striking preference for aromatic residues such as phenylalanines, which are in van der Waals contacts with the edge of the porphyrin and are

found at different locations around the non-polar sides of the haem (Table 3.4). In contrast to leucine, another side chain frequently in contact with the edge of the haem, isoleucine is less common, possibly because its numerous rotamer conformations, which make it too flexible to keep the haem in a fixed position (Table 3.4). Leucines are often found packing between pyrroles II and III at the haem-edge. Other residues packing at these sites include alanine and valine (Table 3.4). All these amino acids contribute to building relatively rigid and complementary surfaces for binding.

Table 3.4: Frequencies of residues in van der Waals contact (3.6 - 4.1 Å Vriend (1990)) with the haem-edge in different folding topologies of b-type haem proteins. Distances were calculated using DISTANG (CCP4, 1994) and percentages are relative to the number of contacts in a particular fold. A more detailed list can be found in Table A.5, p. xxiv.

Fold	Haem-edge Contacts (3.6 - 4.1Å) in %							
	Ala	Ile	Leu	Phe	Trp	Tyr	Val	other
4-helix bundle	7.1	14.3	28.6	-	7.1	7.1	-	35.7
β -propeller	-	-	-	33.3	-	-	-	66.7
CAP	-	11.1	22.2	22.2	11.1	-	11.1	22.2
catalases	14.9	1.4	10.8	9.5	-	1.4	5.4	56.8
cupredoxin-like	-	50.0	-	25.0	-	-	-	25.0
cytochrome b5	14.0	6.0	32.0	10.0	6.0	4.0	8.0	20.0
dioxygenase	-	28.6	-	28.6	-	14.3	14.3	14.3
globin like	12.5	7.5	23.8	7.5	-	12.5	10.0	26.3
haem-oxygenase	6.5	2.2	8.7	19.6	-	10.9	13.0	39.1
haem-peroxidase	1.9	13.0	20.4	20.4	7.4	3.7	9.3	24.1
HemS-fold	33.3	-	33.3	33.3	-	-	-	-
H-NOX / SONO-fold	-	-	15.4	38.5	-	30.8	-	15.4
immunoglobulin-like	14.3	14.3	14.3	14.3	-	-	14.3	28.6
lipocalin	-	16.7	16.7	16.7	-	33.3	16.7	-
meander	12.5	12.5	12.5	25.0	-	12.5	-	25.0
NO fold	3.1	3.1	12.5	9.4	12.5	6.3	6.3	46.9
P450	8.4	9.3	27.1	15.9	-	0.9	4.7	33.6
PAS	10.0	20.0	20.0	10.0	10.0	10.0	10.0	10.0
serum albumin-like	9.1	-	27.3	27.3	-	-	9.1	27.3
vitamin b6 family	-	-	-	-	-	20.0	-	80.0
Overall %	8.8	7.6	19.9	14.4	2.6	6.3	7.4	33.1

3.3.7 Interactions with the proximal ligand

In some proteins the residues next to the proximal histidine appears to play a role in the binding of haem. For instance in haemopexin as well as in cytochrome c peroxidase, a tryptophan packs face-to-face onto the histidine imidazole, also making contact with the porphyrin (Fig. 3.2A). In some cases another side chain on the other side of the proximal ligand in respect to the tryptophan, is making close contacts.

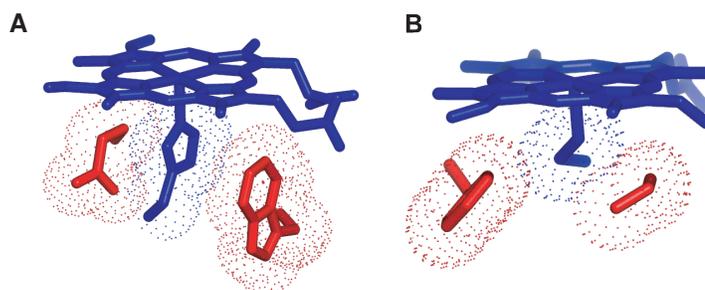


Figure 3.2: Packing interactions with the proximal ligand. (A) In haemopexin (1QHU) His 265 is sandwiched between Glu 225 (left) and Trp 267 (right). (B) In cytochrome P450 (1PQ2) Cys 435 is sandwiched between Phe 428 (left) and Gly 437 (right). Figure was prepared with PYMOL (Delano, 2002).

A previously carried out study on aromatic interactions in proteins showed that tryptophan and histidine side chains preferentially interact with a horizontally displaced stacked geometry (Samanta *et al.*, 1999). The packing of the tryptophan and the proximal imidazole provides a favourable interaction, possibly by positioning the histidine side chain and therefore stabilising the histidine-iron bond. This bond is relatively weak, with an estimated energy of about 10 kcal/mol, as revealed by spectroscopic work on globins (Stein *et al.*, 1980).

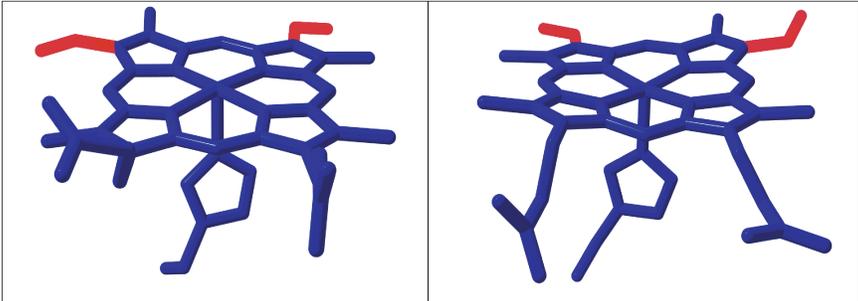
Similar π -stacking interactions can also be observed in peroxidase and nitrophorin, where a phenylalanine packs against the imidazole of the proximal histidine. In fungal peroxidase and the peroxidase domain of prostaglandin H2 synthase-1 a leucine and valine residues can be found in this position, respec-

tively. In both catalase and ubiquinol oxidase, the proximal histidine is positioned by the by extensive stacking interactions provided by two phenylalanine side chain. Finally, in the case of cytochrome P450 the cysteine haem ligand is sandwiched between a phenylalanine and a glycine, as shown in Figure 3.2B. The presence of these side chains presumably still provides a stabilising packing interaction with the proximal ligand and the haem.

3.3.8 Haem orientation relative to the proximal ligand

When structures are overlaid using central atoms of the porphyrin as well as the proximal ligand as frame of reference, the haem can be effectively rotated 180° in some proteins relative to others. Apparently the haem can therefore be bound in either of two flipped orientations defined by the asymmetry in the porphyrin, due to the positions of the vinyl substituents (Fig. 3.3). This 'flipped arrangements' were first observed in NMR studies on rat cytochrome b5 (Dangi *et al.*, 1998, Lee *et al.*, 1993, Mortuza & Whitford, 1997) and occur as well in other protein folds, such as for instance in human haemoglobin and in the haemoglobin from the ciliate *Paramecium caudatum* (reported in bold typescript in Figure 3.3). Moreover, the 1.5 Å resolution structure of neuroglobin revealed a mixed population of haem groups bound in both the conformations (Fig. 3.3 and Vallone *et al.* (2004)).

Therefore the haem-orientation, regarding the positions of the vinyl groups relative to the proximal ligand, does not affect the function as long as the protein scaffold has evolved to associate with haem in a given orientation (Dangi *et al.*, 1998, Lee *et al.*, 1993, Mortuza & Whitford, 1997, Vallone *et al.*, 2004). Indeed, examination of the protein contacts at the haem edge revealed limited conservation of interactions across different folds which reflects the fact that the different packing requirements of binding haem in either of its distinct orientations.



VINYL LEFT + BACK			VINYL BACK + RIGHT		
FOLD	PDB CODE	ORIGIN / LIGAND	FOLD	PDB CODE	ORIGIN / LIGAND
cyt b5 fold cyt b558, <i>E. vacuolata</i>	1CXY	P / bis-His	cyt b5 fold cyt b5, <i>R. norvegicus</i>	1EUE	E / bis-His
HO fold heme oxygenase, <i>H. sapiens</i>	1N45	E / His	HO fold heme oxygenase, <i>C. diptheria</i>	1IW0	P / His
globin-like hemoglobin, <i>H. sapiens</i>	1A3N	E / His	globin-like hemoglobin, <i>P. caudatum</i>	1DLW	P / His
neuroglobin, <i>M. musculus</i>			neuroglobin, <i>M. musculus</i>		
conformation I	1Q1F	E / bis-His	conformation II	1Q1F	E / bis-His
P450 fold P450, <i>H. sapiens</i>	1PQ2	E / Cys	P450 fold P450, <i>S. coelicolor</i>	1ODO	E / Cys
H-NOX fold sGC, <i>T. tengcongensis</i>	1U55	P / His	H-NOX fold SONO, <i>T. tengcongensis</i>	1XBN	P / His
β -propeller	1QHU	E / bis-His	4-helix bundle	1BFR	P / bis-Met
catalase	1A4E	E / Tyr	α/β fold	1JBQ	E / His-Cys
PAS	1V9Y	P / His	β -sandwich	1D7B	E / His-Met
serum albumin-like	1N5U	E / Tyr	CAP	1FT9	P / His-Pro
ChuS-fold	2J0P	P / His	helical membrane	1FFT	P / bis-His
			heme peroxidase	2CYP	E / His
			lipocalin	1NP4	E / His
			meander	1DK0	P / Tyr-His
			NO fold	1NOS	E / Cys

Figure 3.3: 'Flipped' haem orientations relative to the proximal ligand. Structural comparisons of haem pockets show that either of two possible 'flipped' haem orientations with respect to the proximal ligand is observed in haem-protein associations. Interestingly, as highlighted in the top panel of the table, different orientations can exist in homologous or functionally related proteins from different species. The nomenclature 'P' and 'E' refers to proteins with prokaryotic and eukaryotic origins, respectively.

3.3.9 Perspectives for the design of novel haem proteins

The here described observations on re-occurring contacts made with the haem-ligand, could provide guidelines for the design of novel haem proteins. Aromatic interactions are clearly playing a key role in stabilising the haem-protein association in many of the proteins examined here. Nevertheless aromatic contacts are not essential, as indicated by their absence, for example in CooA. Moreover steric complementarity could be more important than aromatic stacking, which is strongly suggested by the analysis of van der Waals contacts. However, a general estimation of the relative contributions of the different

interactions may not be possible, especially in those cases in which the evolutionary/functional pressures dominate the structure of the haem pocket. For instance in globins and cytochromes haem is a prosthetic group that becomes an integral part of the structure; they are less stable in their apo-forms and the folding process itself may rely on haem incorporation (Huntley & Strittmatter, 1972, Pfeil, 1993). In contrast, in HasA and haemopexin haem is a true ligand and the major function is the reversible binding of haem with high affinity. Multiple factors are combined to yield pico- and femtomolar range binding affinities, as epitomised by the structures of HasA and haemopexin (Arnoux *et al.*, 1999, Paoli *et al.*, 1999, Smith & Hunt, 1990).

Substantial efforts have been made towards the design of haem-proteins (for example Gibney & Dutton (1999), Xu & Farid (2001)). A potential role of arginine residues in anchoring the propionate groups and leucine side chains in hydrophobic contacts with the porphyrin has been discussed, but in the final design experiments a very minimalist approach was maintained (Robertson *et al.*, 1994). The majority of other design strategies emphasised on the positioning of histidine ligands to the iron. It is likely that the limited stability of the resulting designed proteins (Choma *et al.*, 1994, Gibney *et al.*, 1998, Gibney & Dutton, 1999, Rojas *et al.*, 1997) is partly due to steric interactions and non-specific packing between porphyrin atoms and protein side chains. Therefore the identification of interactions between the protein framework and the haem ligand by arginine, leucine and phenylalanine side chains could aid the design of a new generations of haem-proteins, particularly when catalysis is one of the design aims (Obataya *et al.*, 2000). In these cases, the specific haem-packing motifs presented here may be used to maximise the association of haem while allowing for a trade off in stability, often required by enzymes for scaffold flexibility and modulation of geometry at the active site.

3.4 Conclusions

In this Chapter, the haem binding sites of 68 proteins with 20 different folds and less than 60% sequence identity have been compared by means of structural superpositions. This analysis showed that multiple structural solutions are possible for binding the same, chemically identical ligand, haem. In general the interactions between protein and haem are very diverse. Non-polar interactions are made with both the edge and the flat face of the haem. The edge contacts are particularly varied, but some remarkable conservation was noticed in contacts made with the haem-face. Over two thirds of the proteins surveyed use arginine side chains to anchor the haem's propionate groups through electrostatic bridges. Strikingly, leucine and isoleucine side chains were found at a re-occurring position making van der Waals contacts with the haem face. Aromatic rings are frequently aligned at three sites over the porphyrin, engaging in stacking interactions. Other interesting contacts include hydrophobic side chains that pack against the haem iron proximal ligand, presumably helping to maintain the imidazole in a suitable orientation for coordinating the iron. Structural superpositions thus showed that residues from unrelated structures, despite the great diversity of folding topologies, cluster at particular interaction 'hot-spots' defining some common structural haem-binding motifs. The interactions and structural motifs contributing to the association and incorporation of haem by proteins identified in this study, may help to understand structure-function relationships in haem-proteins and might be useful for the design of proteins able to incorporate this versatile and ubiquitous ligand.

Chapter 4

Biochemical and biophysical characterisation of HemS from *Yersinia enterocolitica*

4.1 Introduction

HemS is the cytoplasmic recipient and presently the best studied component of HemRSTUV system which is detailed described in Chapter 1.2, p. 5. Essential for *Yersinia*, HemS was originally thought to be a haem oxygenase and was also presumed to protect the cellular environment from the toxic effects of free haem (Stojiljkovic & Hantke, 1994, 1992). Work on the homologous protein ShuS from *S. dysenteriae*, which shares 64% sequence identity with HemS, showed that ShuS binds one haem per molecule with an affinity in the micromolar range (Wilks, 2001). Moreover ShuS is essential for *Shigella* to grow using haem as sole iron source and potentially transfers it to haem-containing or degrading proteins (Wyckoff *et al.*, 2005). Recently a link between unspecific binding of ShuS to DNA and its inhibition by haem has been investigated and a potential role in preventing DNA damage and toxicity associated with

high haem concentration has been argued (Kaur & Wilks, 2007). Also recently it was discussed that the homologue from *E. coli*, ChuS, is a haem oxygenase, degrading its cargo (Suits *et al.*, 2005). In contrast it has been shown that PhuS, the homologue from *Pseudomonas aeruginosa*, is not a haem oxygenase and the break down of haem is an unspecific process caused by the presence of hydrogen peroxide. Instead PhuS transfers the haem to the haem oxygenase paOH, driven by direct protein-protein interactions (Lansky *et al.*, 2006).

In this chapter the cloning, expression, purification and site-directed mutagenesis of *Y. enterocolitica* HemS based on computational analysis is described. Moreover UV/vis absorption spectroscopy was used to characterise the haem coordinating residues of the haem-HemS complex from *Y. enterocolitica*, as well as its haem binding properties. Furthermore to identify potential interaction partners of HemS, pull-down assays with cell extracts from *Y. enterocolitica* were performed.

4.2 Materials and methods

4.2.1 Cloning of *hemS* from genomic DNA

The full-length sequence from *Y. enterocolitica hemS* (gi:2507043) was PCR-amplified from genomic DNA (genomic DNA previously prepared by Dr. M. Paoli).

For cloning purposes a maximum number of 24 amplification cycles was used.

The PCR product was gel-purified using a Qiagen or Sigma gel-extraction kit.

A standard PCR protocol was employed (see Chapter 2.2.2.1, p. 38 and Appendix A.2, p. xix).

For cloning and expression the vector pGAT2, a pGAT (Peränen *et al.*, 1996) derivative, was used (Appendix A.3, p. xx). For insertion into pGAT2, the PCR product and the vector were cleaved with HindIII and BamHI. Ligation, transformation and analysis were carried out as described in Chapter 2.2.3,

p. 39. To obtain HemS only in frame with the His₆-tag and the thrombin cleavage site linker region, YE-HemS-pGAT2 was cleaved with the restriction endonuclease SpeI, that cuts out the GST encoding region, then gel-purified and re-ligated (pSGAT2).

4.2.2 Site-directed mutagenesis of *hemS*

Single amino acid residue exchanges to alanine of *Y. enterocolitica* HemS His 75, His 196, Met 244 and His 280 were carried out by overlapping primer extension PCR (Ho *et al.*, 1989) as described in Figure 4.1. Correctness of the sequence was confirmed by sequencing.

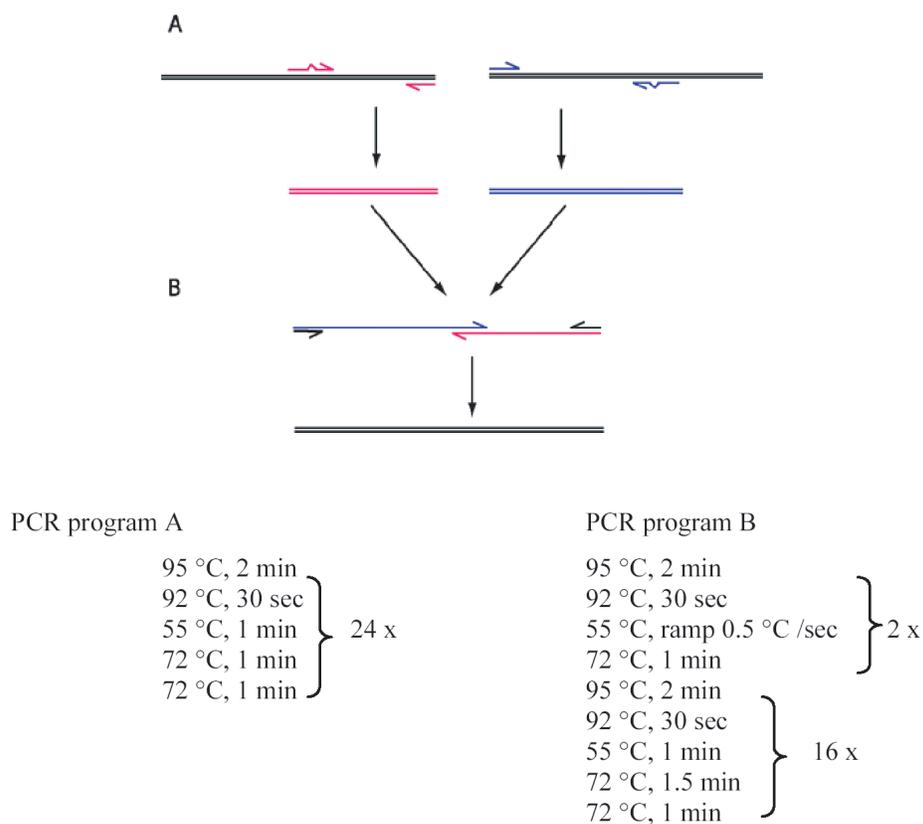


Figure 4.1: Overlapping primer extension PCR. Two separate PCR reactions with YE-HemS-pSGAT2 as template, using the reverse mutagenesis primer (Appendix A.2, p. xix) containing the nucleotide changes and the forward primer, and *vice versa* were carried out (A). The PCR products were gel-purified and equimolar amounts used in a second PCR reaction (B), to obtain the full-length mutated gene sequence that was then inserted into pSGAT2.

4.2.3 Expression and purification of HemS

HemS was expressed and purified as described in Chapter 2.2.4, p. 40 and published in Schneider & Paoli (2005). Briefly, once an OD_{600} of 0.6-1.0 was reached, the temperature was dropped to 30°C and protein expression was induced by adding IPTG to a final concentration of 0.4 mM. The cells were further grown overnight, harvested and lysed. Native HemS and HemS mutants were purified by Ni-affinity chromatography with an imidazole gradient. To determine the HemS concentration the theoretical millimolar extinction coefficient of $\epsilon_{280} = 1.13 \text{ ml cm}^{-1} \text{ mg}^{-1}$ was used (<http://www.scripps.edu/cd-putnam/protcalc.html>). The N-terminal His-tag was removed by proteolytic cleavage with thrombin. The amount of thrombin necessary to cleave 1 mg His-HemS overnight at room temperature was determined as 5 U thrombin. Thrombin, His-tag and uncleaved protein were removed as explained in Chapter 2.2.4, p. 40.

4.2.4 UV/vis absorption spectroscopy and reconstitution of HemS with haemin

All UV/vis absorption spectra were recorded in 50 mM Tris-HCl pH 7.5, 150 mM NaCl at 25°C unless otherwise stated. Experiments in which haem was titrated with the apo-protein were performed by mixing 5 μM of haemin chloride with 0.5 to 10 molar equivalents of apo-protein; these mixed samples were then allowed to equilibrate for 2 hours at room temperature. The haemin stock solution was prepared as described in Chapter 2.2.5, p. 41. Cyanide binding was carried out by adding potassium cyanide to a final concentration of 5 mM to a cuvette containing a solution of reconstituted ferric protein (5 μM). Reduction of fully reconstituted HemS was carried out by removing the oxygen from the protein solution in an anaerobic cuvette by flushing with argon, before adding

5 mM sodium dithionite. To carry out measurements on CO-bound protein, an excess of CO gas was purged through a sample previously reduced with sodium dithionite.

For reconstitution of HemS, the sample was saturated with haem by incubation with 3 molar excess haemin for one hour at room temperature and excess haemin was removed by gel-filtration (see Chapter 2.2.5, p. 41). The protein was then concentrated with centrifugal membrane-devices (Vivaspin) in 50 mM Hepes pH 8.0, 150 mM NaCl, to a concentration of 50 mg ml⁻¹.

4.2.5 His-tag pull down assays

His- and His-GST tagged HemS was over-expressed, purified over Ni-NTA resin, eluted and dialysed against 50 mM Hepes, pH 8.0, 150 mM NaCl. After reconstitution with haem, the His-tagged haem-HemS complex and His-GST-tagged haem-HemS complex were reattached to new Ni-sepharose resin and washed with phosphate-buffered saline (PBS) containing 10 mM and 25 mM imidazole.

Cell-lysate from *Y. enterocolitica* cells was prepared as follows: cells from a 125 ml ON culture grown in LB-medium (cells were grown by Dr. Jakki Cooney), were harvested by centrifugation (6000 g, 10 min, 4°C) and resuspended in 1.5 ml PBS with protease inhibitors. For cell lysis, 10 mg ml⁻¹ lysozyme was added and the cells were incubated on ice for 30 min, followed by 3 freeze/thaw cycles. After sonication (2 x 1 min of 9 cycles 10 micron) the soluble fraction was recovered by centrifugation (15,000 g, 10 min, 4°C). 600 µl of cell lysate were added to 40 µl of Ni-sepharose His-haem-HemS, Ni-sepharose His-GST haem-HemS or Ni-sepharose (blank control) resin and incubated for 1 h at RT. The resin was washed 3 times with with PBS 10 mM imidazole and 2 times with PBS supplemented with 25 mM imidazole. The resin was then resuspended in 40 µl 3xSDS sample buffer and boiled for 5 min at 95°C.

10 μl of the supernatant were loaded on a SDS-PAGE gel. Isolated bands were cut out and sent for peptide-mass-fingerprinting for identification (John Keyte, Medical School, University of Nottingham).

For verification of identified potential interacting partners for HemS, a reverse pull-down assay was carried out. Purified His-CRP protein, bound to 10 μl Ni-NTA resin was washed to remove excess CRP protein. The resin was then incubated with an excess of tag-free, haem-HemS in either 50 mM Tris-HCl, 150 mM NaCl pH 8 or PBS buffer, slowly rotating (1 h at 4°C or RT). As a control for specificity of the interaction, HtsA a haem protein from *Streptococcus pyogenes*, was included in the experiment and also incubated with His-CRP attached to the Ni-resin. The resin was then washed with buffer containing 10 and 25 mM imidazole, resuspended in 40 μl 3xSDS sample buffer and boiled for 5 min at 95°C. An aliquot 10 μl of the supernatant was loaded on a SDS-PAGE gel.

4.2.6 Cloning of the *cyclic AMP receptor protein*-gene from genomic DNA

The 632 bp gene coding for the cyclic AMP receptor protein (CRP; gi:18032025), which was identified as a potential interacting partner by pull-down assay and peptide-mass-fingerprinting, was PCR amplified from *Y. enterocolitica* genomic DNA (for primer see Appendix A.2, p. xix). The PCR product was cleaved with BamHI and HindIII and cloned into the expression vector pGAT2 (Appendix A.3, p. xx) in frame with the His₆-GST double tag. The expression construct YE-CRP-pGAT2 was digested with the restriction endonuclease SpeI in order to remove the sequence coding for GST, thus leaving the CRP sequence in frame with the His-tag (YE-CRP-pSGAT2) .

4.2.7 Expression and purification of the cyclic AMP receptor protein

Expression of the His-GST double tagged CRP was insoluble at all tested temperatures (20 - 30°C) as well as over different expression times (2 h - overnight). The expression construct with CRP only in frame with the His-tag, could be expressed soluble at 25°C overnight. The cells were lysed by french press in presence of 5 mM 3'-5'-cyclic adenosine monophosphate (cAMP). A standard His-tag purification was carried out as described in Chapter 2.2.4, p. 40.

4.3 Results and Discussion

4.3.1 Sequence alignment and structure prediction of HemS

At the beginning of the project, no structural data of HemS or its homologues were available. Therefore the conservation of the HemS protein sequence was analysed, and secondary and tertiary structure predictions were also carried out, using computational methods. In order to identify conserved amino acid residues that might play a role in haem-iron coordination, the database of the National Centre of Biotechnology Information (NCBI) was searched using BLAST (Altschul *et al.*, 1990) for homologous protein sequences to HemS. Homologous sequences were only identified in bacteria belonging to the phylum of the Gram-negative α -, β - and γ -Proteobacteria, showing a sequence identity of above 30 % (Fig. 4.2). No similar sequences from any other organism or bacterial phylum were detected, nor had the structure of any of the homologous proteins been solved at the time of this work. Alignments of these sequences show similar trends of hydrophobic regions as well as four conserved amino acids, possibly involved in haem-iron coordination could be identified:

		gamma							beta					alpha						
		YE-HemS	YP-HmuS	ECI-EhuS	EC-ChuS	ErC	PA-PhuS	PL	SD-ShuS	BA-BhuS	BC	CV	MF	RM - HmuS	BH	ML-HemS	RL	SM-HmuS	SS	
gamma	YE-HemS	100%	89%	63%	66%	65%	41%	61%	64%	39%	36%	41%	36%	38%	35%	39%	37%	37%	35%	
	YP-HmuS	95%	100%	65%	67%	67%	41%	61%	66%	40%	36%	41%	37%	39%	34%	38%	37%	37%	34%	
	ECI-EhuS	76%	77%	100%	79%	61%	40%	52%	79%	39%	33%	38%	35%	38%	33%	38%	37%	36%	33%	
	EC-ChuS	77%	78%	87%	100%	63%	39%	54%	98%	37%	35%	39%	40%	37%	34%	36%	36%	35%	34%	
	ErC	79%	79%	74%	75%	100%	38%	71%	62%	38%	37%	39%	38%	38%	34%	38%	35%	37%	33%	
	PA-PhuS	55%	55%	54%	54%	55%	100%	37%	39%	44%	45%	46%	44%	43%	35%	42%	40%	44%	38%	
	PL	75%	76%	65%	69%	58%	50%	100%	54%	40%	35%	36%	35%	33%	34%	34%	34%	34%	31%	
beta	SD-ShuS	77%	78%	87%	98%	75%	54%	68%	100%	36%	34%	39%	35%	36%	34%	36%	36%	35%	33%	
	BA-BhuS	56%	56%	53%	54%	53%	57%	52%	54%	100%	42%	45%	44%	60%	36%	42%	40%	46%	37%	
	BC	52%	52%	56%	48%	50%	59%	50%	48%	53%	100%	47%	50%	41%	37%	40%	37%	41%	38%	
	CV	59%	59%	52%	54%	56%	58%	53%	54%	58%	58%	100%	48%	44%	35%	41%	37%	44%	37%	
	MF	56%	56%	55%	56%	56%	61%	51%	54%	60%	65%	64%	100%	36%	40%	42%	40%	41%	37%	
	RM - HmuS	57%	58%	54%	54%	55%	56%	49%	54%	72%	56%	56%	54%	100%	40%	41%	40%	41%	37%	
	BH	51%	52%	50%	51%	53%	53%	52%	50%	52%	54%	52%	56%	61%	100%	45%	42%	45%	37%	
alpha	ML-HemS	57%	56%	54%	52%	52%	57%	52%	52%	56%	55%	51%	59%	59%	61%	100%	60%	60%	41%	
	RL	55%	56%	53%	54%	53%	57%	54%	53%	59%	54%	54%	61%	61%	57%	72%	100%	64%	43%	
	SM-HmuS	56%	55%	53%	51%	56%	59%	53%	51%	57%	55%	56%	59%	59%	59%	70%	78%	100%	42%	
	SS	51%	52%	47%	49%	48%	52%	49%	48%	56%	52%	50%	50%	50%	52%	58%	58%	56%	100%	
	SS	51%	52%	47%	49%	48%	52%	49%	48%	56%	52%	50%	50%	50%	52%	58%	58%	56%	56%	100%

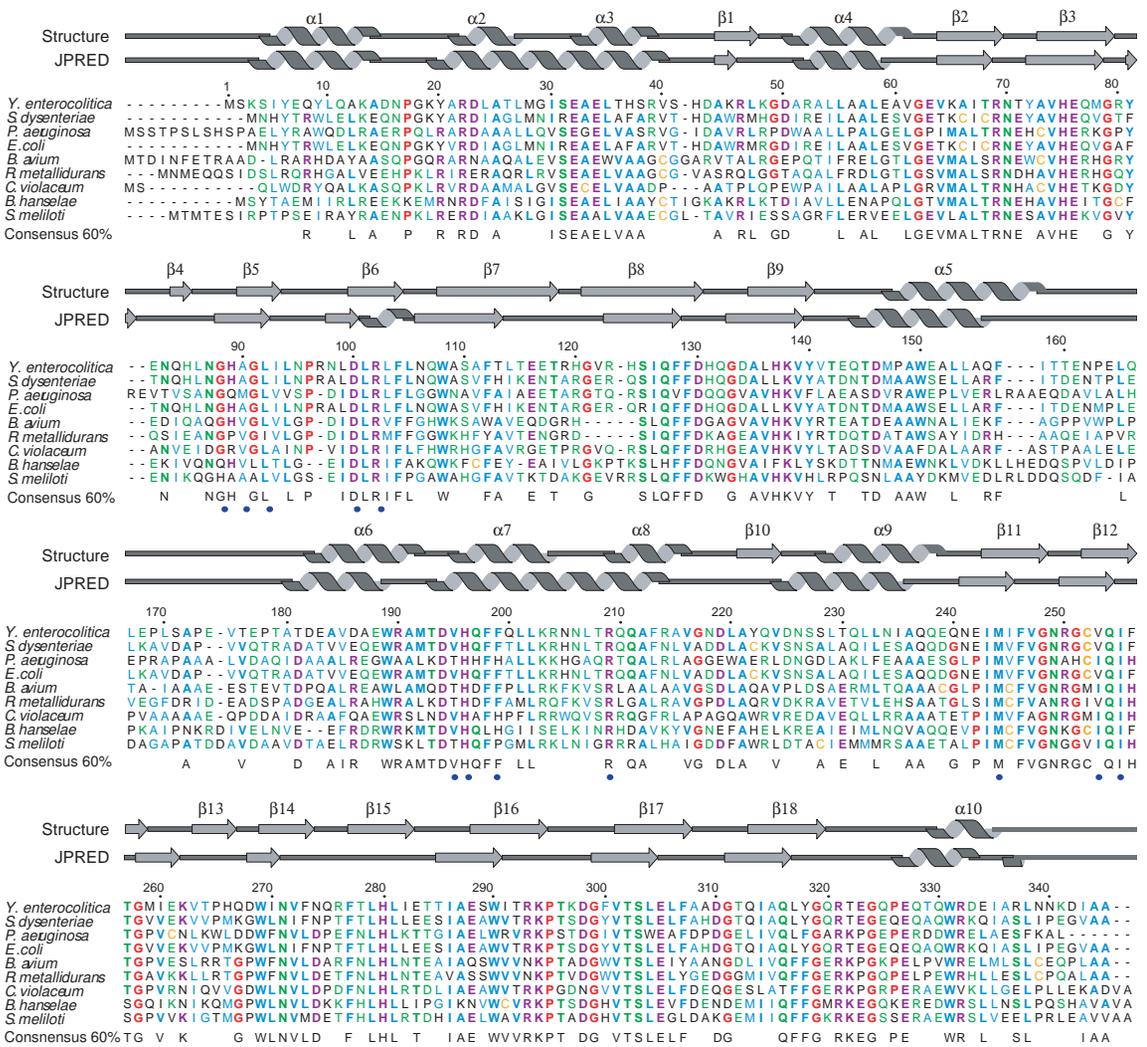
SIMILARITIES

IDENTITIES

Figure 4.2: Sequence similarities and identities of HemS homologues from different species of α -, β - and γ -Proteobacteria, calculated using EMBOSS (<http://www.ebi.ac.uk/emboss/align>). The sequences compared in the figure are from the following species: *Bartonella hanselae* (gi:49475291); *Bordetella avium* (gi:20977617); *Burkholderia cepacia* (gi:46370047); *Chromobacterium violaceum* (gi:34499352); *Enterobacter cloacae* (gi:28170649); *Erwinia carotovora* (gi:49611302); *Escherichia coli* (gi:15804043); *Mesorhizobium loti* (gi:13471236); *Methylobacillus flagellatus* (gi:45520416); *Photobacterium luminescens* (gi:37526524); *Pseudomonas aeruginosa* (gi:15599903); *Ralstonia metallidurans* (gi:22979076); *Rhizobium leguminosarum* (gi:13277332); *Shigella dysenteriae* (gi:2967533); *Silicibacter sp.* (gi:52011032); *Sinorhizobium meliloti* (gi:15966181); *Yersinia enterocolitica* (gi:2507043); *Yersinia pestis* (gi:16120621)

His 75, His 196, Met 244 and His 280 (Fig. 4.3).

Various internet-based secondary and tertiary structure prediction servers, that utilize different approaches were used to predict the structure of HemS (Table 2.2, p. 43). Secondary structure prediction methods, especially for α -helices, are relatively reliable and two different servers, JPRED (Cuff *et al.*, 1998) and PSIPRED (Jones, 1999), were used. The output from both servers was similar and the predicted secondary structural elements included seven well-defined α -helical and four β -sheet regions (Fig. 4.3). However, for fold recognition and tertiary structure prediction, despite having used different fold algorithms (Table 2.2, p. 43), no significant solution was found.



4.3.2 Biochemical characterisation of HemS

4.3.2.1 Purification of HemS and HemS-mutants

HemS from *Y. enterocolitica* was cloned from genomic DNA and the heterologous expression in *E. coli* optimised. HemS shows a highly soluble expression (Fig. 4.4A). Purification was carried out by Ni-affinity chromatography using the His-tag (Fig. 4.4B). Since the His-tag binds haem itself (Paul Barker, personal communication and Fig. 4.10D) and can interfere with crystallisation, special care was taken to remove the tag by cleavage at the thrombin cleavage site in the linker region between the N-terminal His-tag and the protein sequence. Cleavage with thrombin was optimised (Fig. 4.5A) and the resulting protein has an additional serine and glycine residue at its N-terminus. Thrombin, traces of uncleaved protein and the His-tag were removed by p-aminobenzamidine agarose and Ni-NTA resin, to obtain pure protein (Fig. 4.5B).

In order to investigate a potential role in haem-iron coordination of the four conserved residues (His 75, His 196, Met 244, His 280) identified in multiple sequence alignments, were mutated to alanine and their haem-binding properties compared with the wild type. Already upon over-expression in *E. coli* a difference could be observed: cells expressing wild type HemS, HemS-H75A and HemS-M244A were dark colored and the cell lysate was showing a purple color, indicating haem binding upon protein expression. Cells expressing HemS-H196A and HemS-H280A were not showing a dark color, indicating defective in incorporation of haem and implying an involvement in haem-iron coordination of His 196 and/or His 280. Optimised purification of the HemS mutants is shown in Figure 4.4C. The nickel resin bound with HemS was coloured dark blue; elution of the protein could be observed by eye due to the dark red color of the eluate. The yield of pure HemS protein was about 80 mg

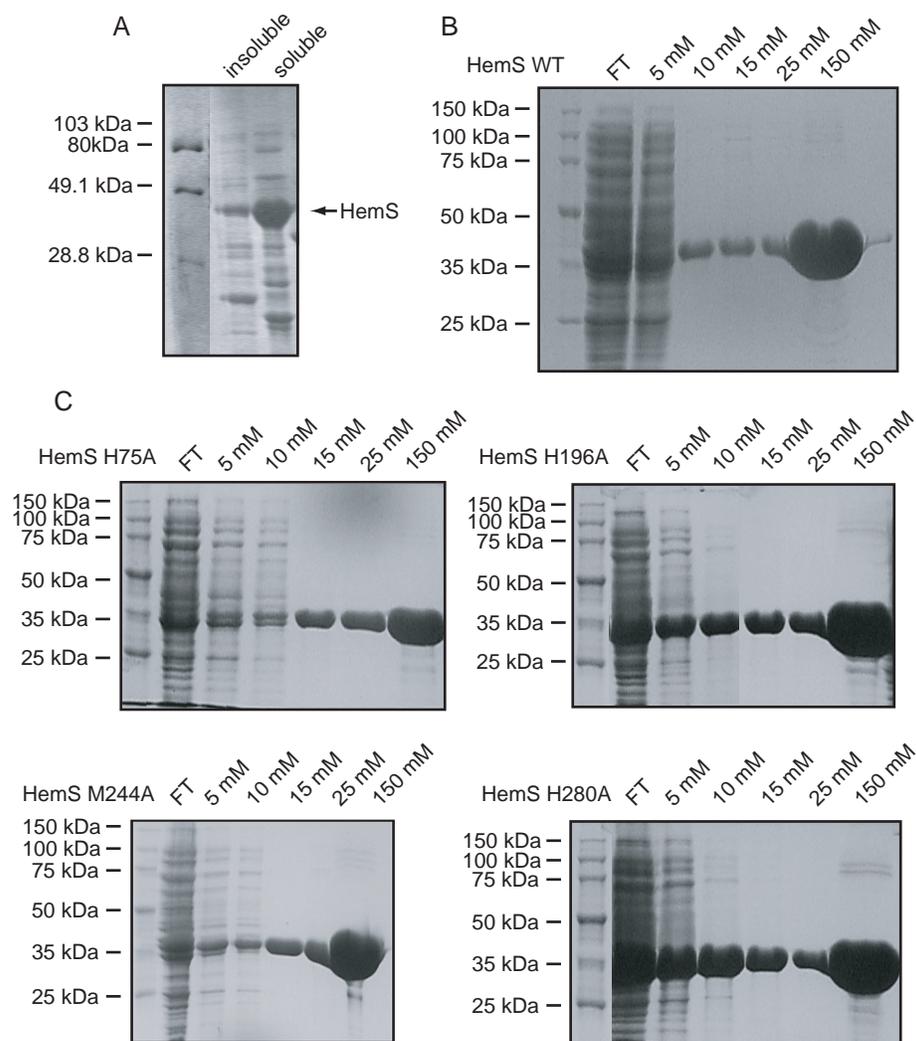


Figure 4.4: HemS expression and purification. (A) Soluble expression of *Y. enterocolitica* HemS. (B) and (C) His-tag purification of HemS wild type and HemS mutants: lane 1: flow through; lanes 2 - 5: washing with increasing imidazole concentrations; lanes 6 and 7: elution of bound HemS with 150 mM imidazole.

per liter cell culture and the protein was stable at 4°C for approximately two weeks.

Analytical gel-filtration analysis of apo- and haem-HemS with different protein and buffer/salt concentrations shows that HemS is a monomer (Fig. 4.6) and does not form dimers, as in the case of PhuS from *P. aeruginosa* (Lansky *et al.*, 2006) or oligomers, as for ShuS from *S. dysenteriae* (Wilks, 2001).

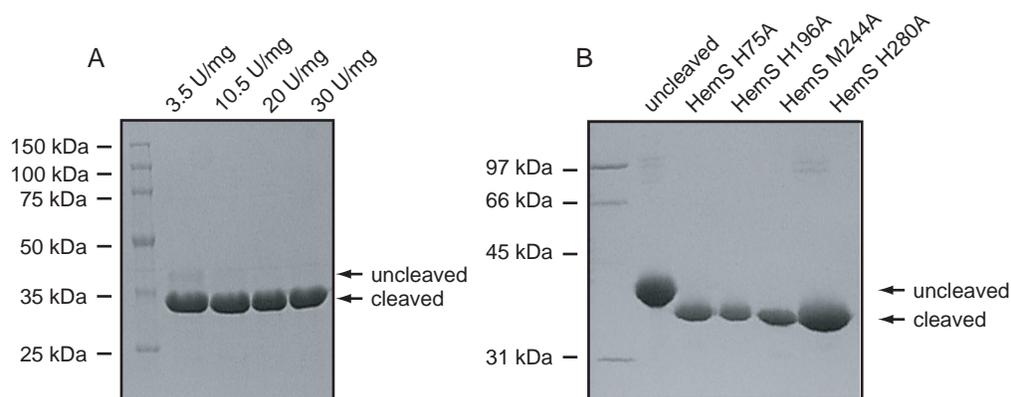


Figure 4.5: Removal of the His-tag from *Y. enterocolitica* HemS with thrombin. (A) Optimisation of thrombin cleavage of HemS over night at RT with varying thrombin concentrations: 5 U thrombin were necessary to cleave 1 mg of recombinant HemS protein. (B) Cleaved HemS mutants.

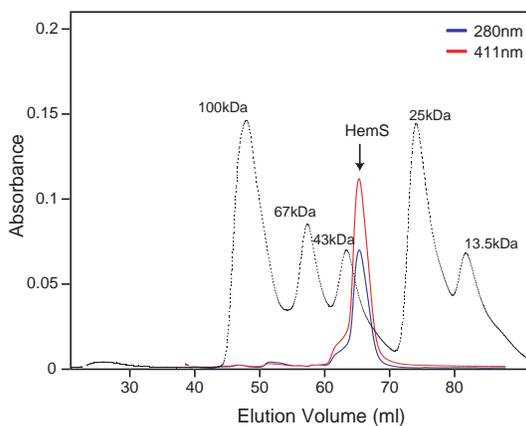


Figure 4.6: Analytical gel-filtration analysis of the haem-HemS complex. An estimated molecular mass of the protein was extrapolated from a standard calibration curve. The column was equilibrated either with 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, or 50 mM Tris-HCl, pH 7.5, 75 mM NaCl, and each was tested with two sample concentrations of 5 and 10 mg ml⁻¹. Representative traces were recorded at 280 nm for molecular weight markers (black) and HemS (blue) and trace HemS measured at Soret maximum of 411 nm (red). The data indicate that HemS is a monomer and no dimerisation is observed even at 10 mg ml⁻¹.

4.3.2.2 Spectroscopic analysis and haem-binding properties of HemS

Haem proteins have characteristic visible Soret absorption maxima between 390-440 nm and Q-bands at 500-700 nm that reveal information about the structure and coordination state of their active site.

UV/vis spectroscopic studies of wild type HemS and the HemS mutants were carried out. The fully reconstituted, ferric wild type haem-HemS complex has a Soret maximum at 411 nm and Q-bands at 548 nm and 580 nm (Fig. 4.7A). Binding of haem by HemS is relatively slow and takes approximately 40 min at room temperature before saturation is reached (Fig. 4.7B). HemS purifies red from *E. coli* but is not saturated with haem showing only a small Soret peak at 411 nm after purification (Fig. 4.8A). Upon reduction of the haem iron by addition of sodium dithionite, the absorption maximum red shifts to 420 nm, with a single Q-band at 555 nm (Fig. 4.8A). Binding of carbon monoxide to the ferrous (Fe^{2+}) haem iron causes a blue shift of the Soret peak to 420 nm and of the Q-band to 543 nm (Fig. 4.8A). Cyanide binding to the ferric (Fe^{3+}) haem iron results in a red shift of the Soret peak to 418 nm and the appearance of a single Q-band at 550 nm (Fig. 4.8B).

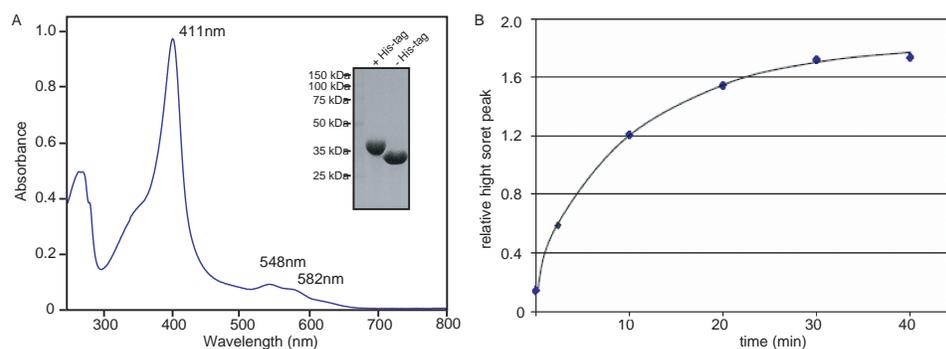


Figure 4.7: Reconstitution of wild type HemS with haemin and time course of haem binding. (A) UV/vis spectrum of haem-HemS complex and samples before and after cleavage with thrombin (inset). (B) Plot of height of the Soret peak relative to the protein peak *versus* time. Apo-HemS was incubated with 3 times molar excess of haemin chloride and a spectrum was recorded every 5-10 min.

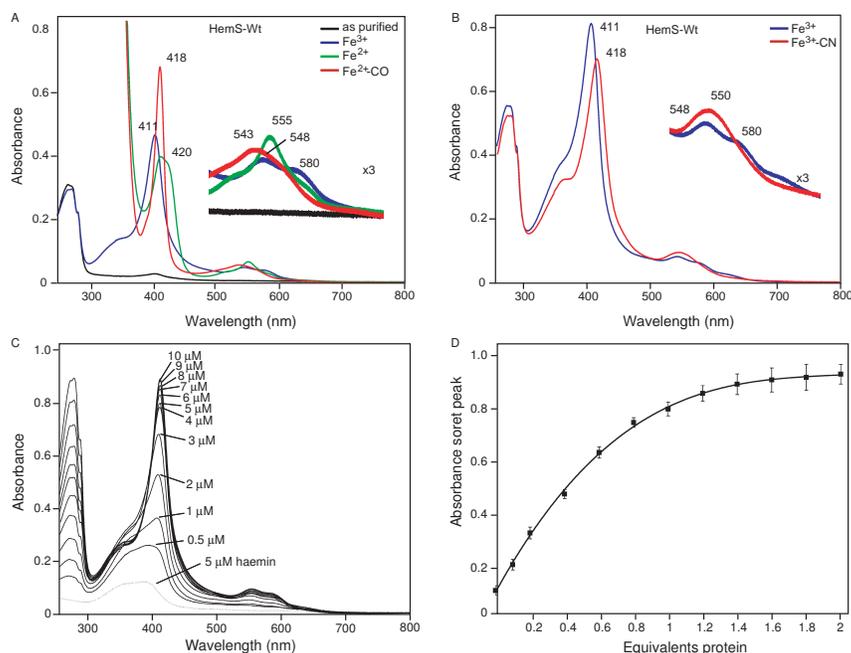


Figure 4.8: UV/vis spectra of wild type HemS. (A) Optical absorption spectra of oxidised, reduced and CO-bound HemS: as purified (black), saturated with haem as ferric (blue); ferrous (green) and CO-complex (red). (B) Optical absorption spectra of HemS in its ferric unliganded (blue) and CN-bound (red) forms. (C) Titration of haemin solution (5 μM) with apo-HemS. (D) Plot of binding curve obtained from titration of haem solution with HemS protein. The curve has been least square fitted with data points from three independent experiments.

Since cyanide as well as carbon monoxide are able to bind to the haem-iron, the distal haem side has to be accessible for a ligand and/or the distal ligand can be displaced by this strong binder. In the NEAT domain proteins IsdA and IsdC from *S. aureus*, the distal haem side is blocked by a residue and cyanide cannot bind (Grigg *et al.*, 2007a, Sharp *et al.*, 2007).

Titration of a haemin chloride solution with apo-HemS (Fig. 4.8C+D) indicates binding of one haem per molecule, as it has been shown for the homologues ShuS and PhuS (Lansky *et al.*, 2006, Wilks, 2001).

The UV/Vis spectroscopic characteristics of the haem-HemS complex were compared with other b-type haem proteins with bis-histidyl, histidine/methionine or histidine haem-iron coordination (Table 4.1). Similarities to proteins with bis-histidyl or a histidine and a water as ligands to the iron could

Table 4.1: Comparison of optical absorption maxima of haem-proteins. Absorption maxima of HemS and haem-proteins with his, bis-histidyl or his/met haem-iron ligation. (noted in parentheses are the presumed coordination structures).

Protein	Fe(III)	Fe(II)	Fe(II)-CO	ref.
HemS	411, 548, 580	420, 555	418, 543	this work
cellobiose dehydrogenase mutant M65H	411, 530, 560 (His/His 6cHS)	428, 526, 556 (His/ His 6cLS)		<i>a</i>
cyt b5	412, 533, 562 (His/His 6cLS)	423, 525, 556 (His/His 6cLS)		<i>b</i>
haemopexin	413, 525, 558 (His/His 6cLS)	426, 527, 559 (His/His6c)		<i>c</i>
neuroglobin	413 (His/His 6c)	426, 560 (His/His 6c)	420 (His/CO)	<i>d</i>
SOUL	413, 535, 565 (His, ?)	422, 527, 558	418, 536, 563	<i>e</i>
HRI-NTD	415, 534, 565 (His/His 6cLS)	429, 532, 562 (His/His 6c)	424, 539, 565 (His/CO 6c)	<i>f</i>
EcDos	416, 530, 564 (His/H ₂ O, 6cLS)	427, 532, 563 (His/Met 6cLS)	423, 540, 570 (His/CO 6cLS)	<i>g</i>
SwMb	410, 505, 635 (His/H ₂ O 6cHS)	434, 556 (His 5cHS)	423, 542, 579 (His/CO 6cLS)	<i>h</i>
haem oxygenase	404, 500, 631 (His/H ₂ O 6cHS)		421, 538, 568 (His/CO 6c)	<i>i</i>
horseradish peroxidase	403, 500, 641 (His/H ₂ O)	437, 556		<i>j</i>
sGC	393, 555 (His 5c)	431, 555 (His 5c)	423, 541, 567 (His/CO 6c)	<i>k</i>
cyt b ₅₆₂	418, 530, 564 (His/Met 6cLS)	427, 531, 562 (His/Met6cLS)	(His/CO 6cLS)	<i>l</i>
cellobiose dehydrogenase wildtype	421, 530, 570 (His/Met 6cLS)	429, 532, 562 (His/Met 6cLS)		<i>m</i>

^a(Rotsaert *et al.*, 2003a)

^b(Rivera *et al.*, 1992)

^c(Cox *et al.*, 1995, Shipulina *et al.*, 1998)

^d(Dewilde *et al.*, 2001)

^e(Sato *et al.*, 2004)

^f(Inuzuka *et al.*, 2004)

^g(Sasakura *et al.*, 2002)

^h(Antonini & Brunori, 1971)

ⁱ(Takahashi *et al.*, 1994, Wilks *et al.*, 1996)

^j(Tamura *et al.*, 1972)

^k(Stone & Marletta, 1994)

^l(Barker *et al.*, 1996)

^m(Rotsaert *et al.*, 2001)

be observed. Strikingly a recent study on cellobiose dehydrogenase, an enzyme with the his/met haem-iron coordination, where Met 65 was mutated to histidine, thus changing the coordination to two bis-histidyl, resulted in a shift of the Soret peak of the Fe^{3+} complex from 421 nm (wild type; his/met) to 411 nm (M65H; his/his)(Rotsaert *et al.*, 2003a).

Titration experiments of haemin with HemS at pH's ranging from 5.8 to 9 were carried out, to investigate whether the protonation-state of the likely involved histidine residues as proximal ligands to the iron has an influence of the HemS-haem-binding properties, since the pKa of histidine is 6.5. The graphs depicted in Figure 4.9 shows impaired haem binding at pH 5.8 compared to more alkaline pH's that seem to enhance haem-binding. Although this pH-effect on haem-binding could be caused by decreased stability of HemS at lower pHs as well as effects of pH on the propensity of haem to form ill-defined aggregates in aqueous solutions.

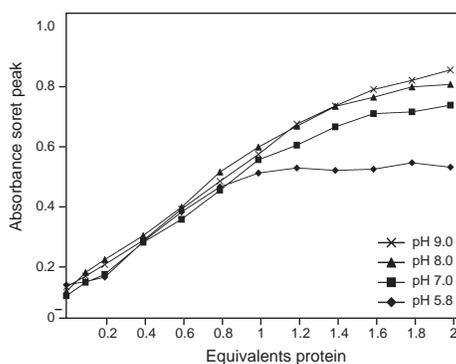


Figure 4.9: pH-dependent haem-binding by HemS. Plot showing influence of pH on haem-binding properties of HemS in titration experiments of haemin with protein.

The possible involvement in haem-iron coordination of the four conserved residues His 75, His 196, Met 244 and His 280, was investigated by mutating them to alanine and comparing the haem-binding, with the wild-type. Mutation of His 75 to Ala had no effect on the haem-binding abilities of HemS

(Fig. 4.10A+B). HemS-M244A showed a minor effects, causing a slight blue shifted as well as a reduced Soret maximum (Fig. 4.10A+B). Therefore these amino acid residues could be excluded as possible ligands to the iron. In contrast, mutation of His 196 to Ala causes a large shift in the spectra of the haem-bound form, from a low-spin (wild type) to a high-spin iron (high-spin marker band at 622 nm). Additionally, the Soret peak was reduced and blue shifted from 411 nm to 402 nm, indicating a change of the coordination state or loss of coordination of the haem-iron (Fig. 4.10A+B). Moreover the spectra of the ferric haem-HemS-H196A complex resembles the spectra of free haem in complex with cyanide (Fig. 4.10C). These changes suggest that His196 is a ligand to the haem iron.

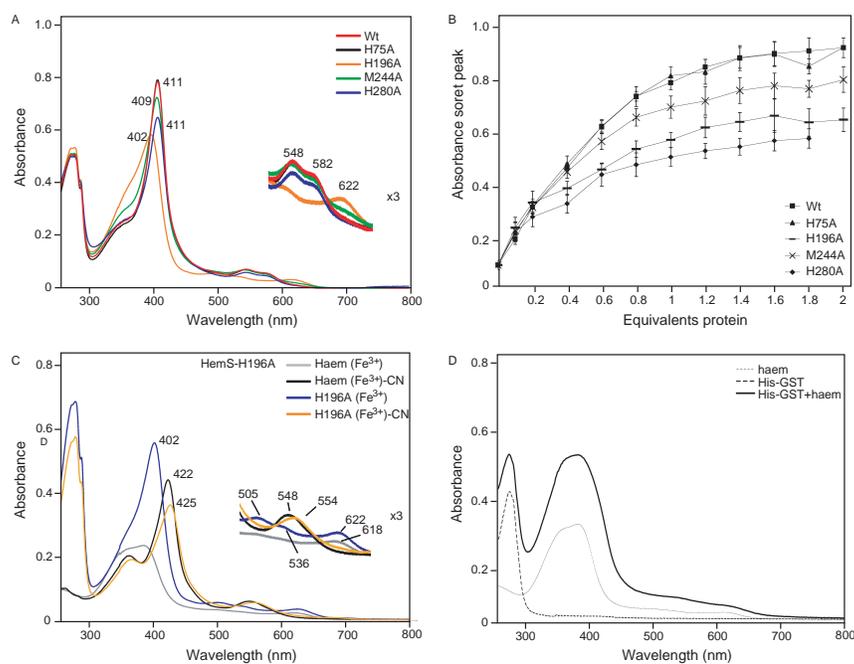


Figure 4.10: Spectroscopic analysis of HemS mutants and unspecific haem-binding by the His-tag. (A) Overlay of UV/Vis spectra of haem-bound wild type HemS (red), H75A (black), H196A (orange), M244A (green) and H280A (blue). (B) Comparison of haem-binding properties of HemS and HemS mutants in titration experiment of haemin chloride with protein. (C) Ferric HemS-H196A (blue) and its cyanide complex (orange). Haemin in solution (grey) and after addition of cyanide (black). (D) Unspecific haem-binding by His-tagged glutathione-S-transferase (GST).

The His 280 to Ala mutation seems to interfere with the haem-binding ability of HemS as well, since the mutation appears to cause a loss of the haem binding ability as investigated by the lack of colour upon purification and UV/vis absorption spectroscopy (Fig. 4.10A+B). However, when tested in haem-binding experiments, the HemS-H280A mutant can exhibit an absorption spectra comparable with the wild type, except of a reduced Soret peak. Its role might be more indirect by influencing the overall stability of the protein, because it was observed that HemS-H280A was less stable at 4°C than the wild type, which could explain the effect on the haem-binding ability. But this hypothesis needs further investigation.

Considering the comparison of UV/Vis absorption maxima with other b-type haem proteins together with the results of the site-directed mutagenesis study, it may be possible to conclude that His 196 is the proximal ligand to the haem in HemS, possibly with a water at the 6th coordination position.

A further experiment was carried out to determine whether the His-tag can bind haem by itself or contribute to haem-binding. A His-tagged version of glutathione-S-transferase (GST), which is not a haem protein and does not bind haem, was expressed and purified. After the addition of haem, His-GST displays a broad Soret peak, which is an indication of non-specific haem-binding by the tag (Fig. 4.10D). Therefore the His-tag can have an influence on haem-binding studies and should be removed before conducting experiments such as titration assays.

4.3.3 Identification of interaction partners of HemS

There are no known interaction partners for HemS from *Y. enterocolitica*. For the homologous protein PhuS from *P. aeruginosa*, it has been shown that it transfers its haem to a haem oxygenase paHO, which degrades the haem. BLAST (Altschul *et al.*, 1990) searches of the *Y. enterocolitica* genome se-

quence available at the Sanger institute (<http://www.sanger.ac.uk/Projects/>) with the sequence of paOH gave no significant hits. Recently two haem oxygenases from *Bradyrhizobium japonicum* (HmuD and HmuQ), belonging to the group of α -Proteobacteria, were identified by BLAST homology searches (Puri & O'Brian, 2006), using the sequences of the haem oxygenases from *S. aureus*, IsdG and IsdI (Skaar *et al.*, 2004a). IsdG and IsdI are structurally distinct from other haem oxygenases (Skaar *et al.*, 2004a) and share weak sequence identity with HmuD and HmuQ. In addition to that the gene coding for HmuQ is clustered in the genome with genes known to be involved in haem utilisation, such as *hmuRTUV* (Puri & O'Brian, 2006). The haem oxygenase function of HmuD and HmuQ was confirmed by *in vitro* and *in vivo* experiments (Puri & O'Brian, 2006). Homologues to HmuD and HmuQ, IsdG and IsdI can only be found in α -Proteobacteria and not in β - or γ -Proteobacteria. In order to identify proteins that interact with HemS, pull-down assays with *Y. enterocolitica* cell lysate were performed, using His-tagged haem-HemS and His-GST-tagged haem-HemS attached to Ni-sepharose resin as a bait. Three protein bands (~ 30 kDa, ~ 20 kDa and ~ 6 kDa) could be isolated from the cell lysate (Fig. 4.11) and were sent for identification by peptide-mass-fingerprinting. A band of ~ 26 kDa size is only present using His-GST-tagged haem-HemS (lane 9), but no His-tagged haem-HemS (lane 8). Therefore it is likely to be a protein that has bound to GST and not to HemS.

Two out of the three isolated bands could be clearly determined by peptide-mass fingerprinting with a sequence coverage of 58 % (Fig. 4.12), the ~ 6 kDa protein band was too small for identification. The ~ 30 kDa protein band could be determined as the *urease accessory protein* (UreE) homologue from *Y. enterocolitica* (gi:18032026)(Fig. 4.12A). The approximately 21 kDa protein band was identified as the homologue of *cAMP receptor protein* (CRP) also known as *catabolite gene activator protein* (CAP) from *E. coli* (Fig. 4.12B).

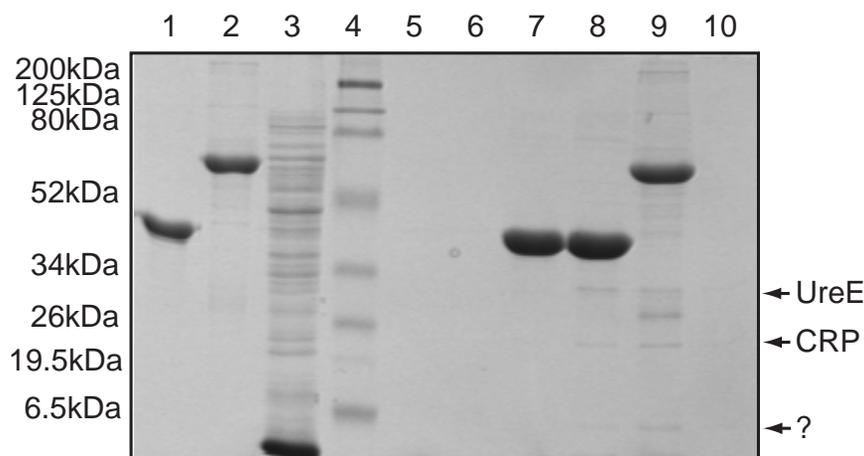


Figure 4.11: Pull-down assay with haem-HemS. Lane 1: His-haem-HemS; lane 2: His-GST-haem-HemS; lane 3: *Y. enterocolitica* cell lysate; lane 4: protein marker; lane 5: wash fraction 25 mM imidazole, blank control; lane 6: wash fraction 25 mM imidazole, His-haem-HemS; lane 7: control His-haem-HemS without cell lysate; lane 8: His-haem HemS incubated with cell lysate; lane 9: His-GST-haem HemS incubated with cell lysate; lane 10: blank control (Ni-sepharose resin without HemS incubated with cell lysate). Arrows are indicating bands identified by peptide-mass fingerprinting: cyclic AMP receptor protein (CRP), and urease accessory protein Ure3. The ~6 kDa band could not be identified by peptide-mass fingerprinting. The band between UreE and CRP is only present in lane 9, but not in lane 8 and therefore had bound to the GST in His-GST-haem HemS.

A

```

1  MILIEHILGN VKKDPVWQAK LKDATFDLLV LDQREAQKSR CRKLSTQGLD
51  LGISLDRHVV LADGDVLAWD EKTNAVAVVQ INLRDVMVID LSELKSRSPD
101 ELIKTCFELG HALGNQHWKA VTKNNEVYVP LTVATTMMS VMRTHGFOHL
151 PFRFVKGAEI LPLLSNSEAR LLFGGAEDTD THVHVASPLD EPHGSGLHIH
201 GIHSHGDGHS HSHDSSHSHS DSDHGSHSHS GDHDHDKH

```

B

```

1  MVLGKPQTD P TLEWFLSHCH IHKYPSKSTL IHQGEKAETL YYIVKGSVAV
51  LIKDEEGKEM ILSYLNQGDF IGELGLFEEG QERSAWVRK TACEVAEISY
101 KKFRQLIQVN PDILMRLSSQ MANRLQITSE KVGNLAFLDV TGRIAQTLLN
151 LAKQPDAMTH PDGMQIKITR QEIGQIVGCS RETVGRILKM LEDQNLISAH
201 GKTIVVYGR

```

Figure 4.12: Results and sequence coverage of peptide-mass fingerprinting. (A) Urease accessory protein (UreE) and (B) *cyclic AMP receptor protein* (CRP) from *Y. enterocolitica*.

UreE proteins are metallo-chaperones involved in the urease metallo-centre assembly by binding Ni^{2+} ions and transferring them into the active site of urease (Musiani *et al.*, 2004). The sequence identity between the protein from *Y. enterocolitica* and the best studied UreE proteins from *Bacillus pasteurii* and *Klebsiella aerogenes*, is 26 % and 24 % respectively. It is likely that the UreE homologue from *Y. enterocolitica* has a similar function as the proteins from *Bacillus* and *Klebsiella* and therefore binds to the Ni-sepharose resin rather than to HemS. Hence, UreE from *Y. enterocolitica* was excluded as a potential interaction partner to HemS and was not further investigated.

CRP is a highly conserved positive regulator of many genes, including the iron/haem uptake machinery in Proteobacteria and sequence identities for CRP proteins range from 60 to 99 %. CRP from *E. coli* is a 45 kDa dimer (Aiba *et al.*, 1982), is a transcriptional activator of over 150 genes (Adhya & Garges, 1990) and the structures of the cAMP-CRP complex as well as the cAMP-CRP-DNA complex have been determined (McKay & Steitz, 1981, McKay *et al.*, 1982). In *Erwinia* and *Vibrio* CRP acts as an antagonist to the *ferric uptake regulator* (FUR) and positive regulator under iron-limited conditions, for the iron acquisition machinery and pathogenicity (Franza *et al.*, 2002, Lee & Choi, 2006). The sequence identity between *Y. enterocolitica* CRP and *E. coli* CRP is 98%. Due to the involvement of CRP in iron-metabolism and if this potential interaction of CRP with HemS is true, HemS could be involved in the regulation of the HemRSTUV operon. In order to investigate this possible interaction of HemS with CRP, *crp* was cloned from *Y. enterocolitica* genomic DNA. CRP expressed insolubly in frame with a His-GST double tag (Fig. 4.13A), most likely because of the interference of the dimerisation by the GST-tag with the dimerisation of CRP. It expressed soluble with a sole His-tag (Fig. 4.13B). CRP was purified in presence of cAMP using metal-affinity chromatography (Fig. 4.13C).

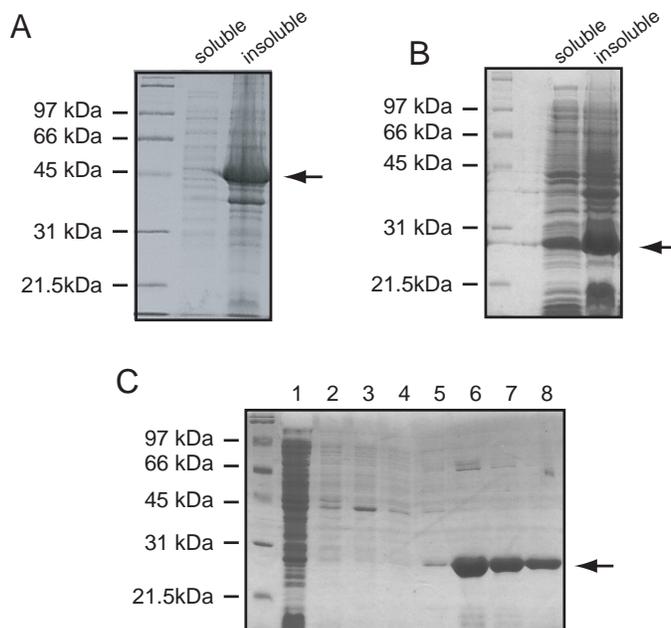


Figure 4.13: Expression and purification of the *cAMP* receptor protein (CRP). (A) Insoluble expression of His-GST-CRP. Expression trials were carried out at temperatures ranging from 20-30°C and different induction points. (B) Partly soluble expression of His-CRP at 25°C. (C) His-tag purification of CRP. Lane 1 = flow trough; lane 2-5 = washes with increasing imidazole concentration (10-25 mM); lane 6-8 = elution with 150 mM imidazole.

In order to verify the potential interaction of HemS and CRP, His-CRP bound to Ni-NTA resin was incubated with excess of fully haem-bound, tag-free HemS. HtsA, a haem protein from *Streptococcus pyogenes* was used as a negative control for the specificity of the binding. After several wash-steps, the resin was boiled with SDS-loading buffer and analysed by SDS-PAGE (Fig. 4.14). The amount of HemS bound to CRP was very limited and comparable to the amount of bound HtsA (Fig. 4.14B, lanes 5-7). The assay was carried out in two different buffers (PBS and Hepes), but no specific interaction between HemS and CRP could be identified in this reversed pull-down assay.

In addition, the potential complex formation between CRP and haem-HemS were investigated by analytical gel-filtration. Haem-HemS and CRP were mixed in a 1:1 molar ratio and incubated for one hour at 4°C before loading the sample onto a gel-filtration column. A potential complex between a

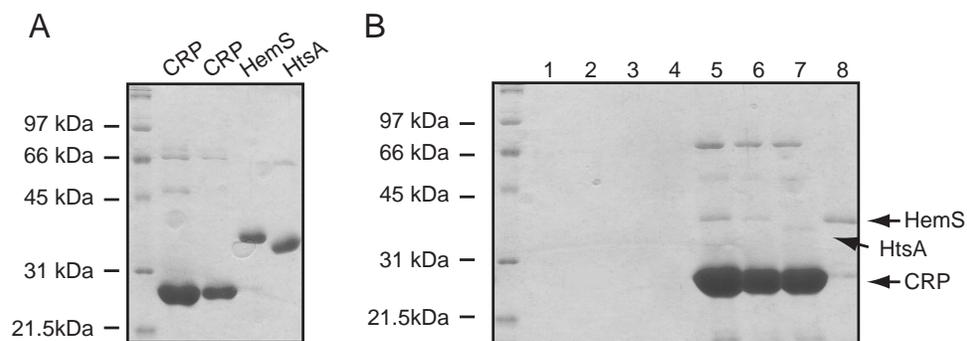


Figure 4.14: Pull-down assay of HemS with His-CRP. (A) SDS-PAGE gel showing the individual proteins His-CRP, haem-HemS, haem-HtsA. (B) Pull-down assay with His-CRP attached to Ni-NTA resin. Lanes 1 - 4 = last wash step of lanes 5 - 8; lane 5 = Ni-NTA-His-CRP incubated with haem-HemS; lane 6 = Ni-NTA-His-CRP incubated with HemS; lane 7 = Ni-NTA-His-CRP incubated with HtsA as control; lane 8 = Ni-NTA resin incubated with haem-HemS (blank control).

CRP dimer and one or two molecules of HemS would have an expected size of 84 kDa and 123 kDa, respectively. The elution of HemS was followed by the absorbance measured at its Soret maximum of 411 nm and it eluted from the column after 69 ml (Fig. 4.15), matching the elution volume of haem-HemS on its own (Fig. 4.6). The trace at the protein absorbance at 280 nm shows a shoulder, possibly corresponding to the CRP-dimer (Fig. 4.15). Thus no complex formation between HemS and CRP could be observed by analytical gel-filtration.

Given the results of the reversed pull-down assay and the analytical gel-filtration, it is unlikely that CRP interacts specifically with HemS in the pull-down assay with the *Y. enterocolitica* cell lysate. In the absence of DNA, CRP has a relative large charged surface area, which could interact with the Ni-NTA resin. Non-specific binding to affinity chromatography resins by cellular proteins is a common problem in this biochemical approach to identify target proteins. In general the term 'non-specific protein binding' is used for proteins that bind to affinity resins through physical absorption (hydrophobic

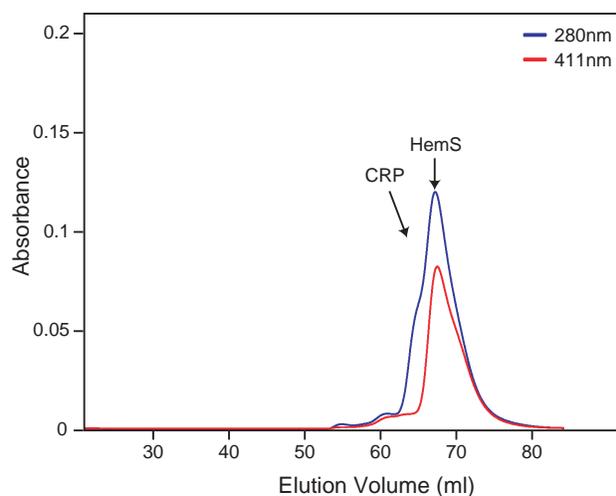


Figure 4.15: Analytical gel-filtration of haem-HemS and CRP. Equimolar amounts of Haem-HemS and CRP dimer were mixed with at a concentration of approximately 10 mg ml^{-1} . The estimated molecular mass of the proteins was extrapolated from a standard calibration curve using a HiLoad Superdex 75 HR16/60 column and a Äkta Purifier chromatography system (GE Healthcare). The column was equilibrated with 50 mM sodium-phosphate, pH 7.5, 20 mM NaCl. Traces were recorded at 280 nm (blue) and at the Soret maximum of HemS of 411 nm (red).

interactions) instead of specific ligand-receptor interaction (Hofstee, 1974, 1973, Tamura *et al.*, 2003). Another problem could be the potential low expression level of potential target proteins for HemS. This could possibly be improved by growing *Y. enterocolitica* cells under iron limited conditions, with haem as sole source of iron. Additional difficulty in determination of target proteins of HemS could be caused by a transient character of the interactions during haem transfer. Recently, short-lived protein-protein interactions and fast transfer of haem has been observed between PhuS and paOH (Lansky *et al.*, 2006) and the haem transport proteins from *S. pyogenes* (Liu & Lei, 2005, Nygaard *et al.*, 2006a). Such transient interactions could be identified for example in a bacterial yeast two hybrid assay (Joung *et al.*, 2000), which would involve generation of a gene library of *Y. enterocolitica* cells where the expression levels of potential target proteins is maximised.

4.4 Conclusions

HemS from *Yersinia enterocolitica* is a 39.2 kDa protein and the cytosolic recipient of the haem-uptake system HemRSTUV (see Fig. 1.3, p. 7). Homologues of HemS are common to the phylum of α -, β - and γ -Proteobacteria with a sequence identity of over 35%. At the beginning of this thesis work the HemRSTUV system from *Y. enterocolitica* was, biochemically, the best studied system. Multiple sequence alignments showed conservation of four potential haem-iron coordinating residues: His 75, His 196, Met 244 and His 280. Using site-directed mutagenesis and UV/vis absorption spectroscopy His 196 could be identified as ligand to the haem-iron. Moreover, in contrast to the homologues from *S. dysenteriae*, ShuS, and *P. aeruginosa*, PhuS, *Y. enterocolitica* HemS is a monomer in solution.

Originally HemS was thought to be a haem oxygenase (Stojiljkovic & Hantke, 1994), which also has been argued for the *E. coli* homologues ChuS (Suits *et al.*, 2005). Recent experiments carried out on the *P. aeruginosa* homologue PhuS showed that PhuS is not a haem oxygenase, but transfers the haem to the haem oxygenase paOH (Lansky *et al.*, 2006). It is therefore most likely that HemS and its homologues are haem-transporter facilitating transfer of haem from the uptake machinery to enzymes that either utilise haem *per se* or break it down. The *Y. enterocolitica* genome sequence was searched for a gene coding for a homologous protein to paOH, or the haem oxygenases HmuD and HmuQ from *B. japonicum* (Puri & O'Brian, 2006) or IsdG and IsdI from *S. aureus* (Skaar *et al.*, 2004a), but none could be found. To identify target proteins for HemS, pull-down assays with *Y. enterocolitica* cell lysate were performed and two potential interacting partners identified. After further investigation, these two proteins proved to be false-positives. Failure in finding target proteins for HemS could be due to their low expression levels in standard medium and/or the transient character of the protein-protein in-

teraction. The expression of potential target proteins could be increased by growing the cells under iron limiting conditions with haem as sole source of iron. In addition as an alternative method to the pull-down assay the bacterial two-hybrid system (Joung *et al.*, 2000) could be used.

Chapter 5

Crystallisation and structure-determination of HemS from *Yersinia enterocolitica*

5.1 Introduction

In Chapter 4 the role of HemS as the cytoplasmic carrier of the HemRSTUV system as well as the biochemical and biophysical properties of HemS and its homologues have been discussed. When this work was started, no structural data on HemS or its homologues were available, nor could any structural similarity to a known structure be identified (Chapter 4.3.1, p. 68).

During this work on the crystallisation and structure determination of HemS, the apo-structure of the homologous protein ChuS from *E. coli* was published, sharing 62% sequence identity with HemS and revealing a novel fold (Suits *et al.*, 2005). Despite the structural data, no haem binding site was identified. After the structural studies here carried out on apo- and haem-HemS

were published (Schneider *et al.*, 2006), the structure of the haem-ChuS complex was also published (Suits *et al.*, 2006). Recently the North East Structural Genomics Consortium solved two structures from hypothetical proteins, AGR_C_4470p from *Agrobacterium tumefaciens* and Q6d2t7_erwct from *Erwinia cartovora*, sharing strong structural similarity to HemS despite an insignificant sequence identity of less than 20% (Vorobiev *et al.*, 2007).

The purpose of the work described in this Chapter is the crystallisation and structure determination of apo- and haem-HemS from *Y. enterocolitica*, the comparisons of which will give insights into the molecular mechanism of haem-binding. Based on the structures of AGR_C_4470p and Q6d2t7_erwct, the evolution of the HemS-folding topology is discussed.

5.2 Materials and methods

5.2.1 Expression and purification of seleno-methionine derivatised HemS

HemS from *Y. enterocolitica* was cloned, expressed, purified and reconstituted with its ligand as described in Chapter 4.3.2.1, p. 71.

Selenomethionine (SeMet)-HemS was produced by transforming the expression plasmid in methionine auxotroph *E. coli* B834 (DE3) cells. 2 ml of an over-night culture of B834(DE3) cells grown in Luria-Betani medium, supplemented with 50 mg ml⁻¹ carbenicillin, were transferred to 1 ltr of LeMaster with 50 mg ltr⁻¹ SeMet (Calbiochem)(Hendrickson *et al.*, 1990) and 50 mg ml⁻¹ carbenicillin (Melford) in a baffled flask. Cells were grown at 37°C and 200 rev min⁻¹ until an absorbance (OD₆₀₀) of 0.8 was reached. Protein expression was then induced by adding IPTG to a final concentration of 0.4 mM and cultures were further grown over night. Purification of SeMet HemS was carried under

non-reducing conditions as for the native protein.

5.2.2 Crystallisation of HemS

After carrying out a pre-crystallisation test (PCTTM, Hampton Research) with the haem-HemS complex to determine the most appropriate protein concentration to use for crystallisation trials, initial high-throughput crystallisation screening was performed using the sitting-drop vapour-diffusion method in 96-well Intelliplates. The experiments were set up with the Hydra II micro dispensing system, testing conditions from five crystallisation screens from Nextal Biotechnology (ClassicsTM, PEGsTM, AmSO4TM, CationsTM, MPDTM) and a number of screens from Hampton Research (Crystal ScreenTM, Crystal Screen 2TM, IndexTM, SaltRxTM, and GRID screensTM: Ammonium SulfateTM, Sodium MalonateTM, PEG/LiClTM, PEG6000TM). The effect of volume was checked using two different drop sizes ($2\ \mu\text{l}+2\ \mu\text{l}$ and $0.7\ \mu\text{l}+0.7\ \mu\text{l}$). Plates were evaluated after 3 weeks. Crystallisation conditions were then optimised by the hanging- and sitting-drop vapour diffusion methods. Different drop sizes (4, 8 and $10\ \mu\text{l}$) and protein concentrations (10, 20, 30 and $50\ \text{mg ml}^{-1}$) were tested against a grid of precipitate concentrations. Additionally the effect of pH and buffer variation on crystallisation as well as ammonium sulphate concentration and additives like imidazole or glycerol were investigated in a 96-well grid-screen, manually designed for the Hydra II micro dispensing system. Plates were incubated at 20°C or at 4°C .

To crystallise the apo-form, Ni-affinity chromatography was followed by ion-exchange chromatography (MonoQ, GE Healthcare), equilibrated with 50 mM Bis-Tris-Propane pH 6.5), to separate molecules that had incorporated haem during expression. Initial crystals were obtained in 100 mM bicine, pH 9, 2.4 M ammonium sulphate, 5% PEG 400 in the same, manually designed 96-well grid-crystallisation screen for the Hydra II micro dispensing system, as

haem-HemS. Apo-HemS crystals could be reproduced and optimised by manual screening around the hit condition and the best crystals were obtained in 100 mM bicine, pH 9, 2.45 M ammonium sulphate and 5 % PEG 400, with a protein concentration of 30 mg ml⁻¹ in sitting-drop vapor diffusion (Crychem plates, Hampton Research)(manual optimisation of apo-HemS crystallisation was carried out by Dr. Katherine Sharp).

Initial crystallisation trials with the haem-SeMet-HemS were carried out in a grid screen with 50 mM Tris-HCl, pH 8.5, ammonium sulphate (1.8-2 M) and PEG 400 (2-5 % (w/v)). The protein concentrations tested were lower than for the native protein (10 and 20 mg ml⁻¹) due to the increased hydrophobicity of the SeMet derivative protein.

5.2.3 Diffraction analysis and structure of HemS

Prior to flash-freezing in liquid nitrogen haem-HemS crystals were transferred into an artificial, cryo-protectant mother liquor consisting of 0.1 M Tris-HCl, pH 8.5, 1.8 M ammonium sulphate, 150 mM sodium chloride, 2 % PEG 400, 1.2 M sodium malonate. Optimisation of the cryo-protectant solution was a key issue, since mixtures with higher concentrations of either PEG 400, ammonium sulphate and sodium malonate resulted in phase separation, serious crystal damage and/or unsatisfactory freezing. For phase determination with heavy atom derivatives, haem-HemS crystals were harvested and soaked over night at 4°C in artificial mother liquor containing potassium tetrachloroplatinate (II) (K₂PtCl₄) or di-*m*-iodobis(ethylenediamine) diplatinum (II) nitrate (PIP). Initial diffraction analysis on native haem-HemS crystals was carried out in-house.

Two MAD data sets on haem-HemS (Fe-MAD) and K₂PtCl₄-derivatised crystals (Pt-MAD) were collected at the European Synchrotron Radiation Facility (ESRF) beamline BM14. The exact peak and inflection wavelengths of the

iron (K-edge) and platinum (LI, LIII-edge) resonance edge were determined through an X-ray fluorescence scan. Details of the data collection are listed in the Table 5.1. In addition a data set on a PIP derivatised crystal was collected in-house with 0.5° oscillation, an exposure time of 10 min and consisting of 360 frames.

Apo-HemS crystals were soaked in artificial, cryoprotective mother liquor containing 100 mM bicine, pH 9, 2.6 M ammonium sulfate, 5 % PEG 400 and 0.8 M sodium malonate prior to flash freezing and storage in liquid nitrogen. The crystals were then diffracted at the ESRF beamline ID-14.2. 180° of data, with 1° oscillation at 0.933 \AA and 15 sec exposure, were collected.

Table 5.1: Data collection details carried out at the ESRF synchrotron beamline BM14 on haem-HemS crystals. MAD data sets were collected on a single crystal. (pk = peak; ip = inflection point)

Data set	Iron MAD			Platinum MAD (K_2PtCl_4)		
	iron pk	iron ip	remote	pt pk LIII	pt ip	pt pk LI
wavelength (\AA)	1.7393	1.7406	0.984	1.07144	1.07189	0.88
collected frames	500	360	320	100	200	200
oscillation	1°	1°	1°	1°	1°	1°
exposure time	30 sec	15 sec	20 sec	15 sec	15 sec	5 sec

Diffraction data were processed with MOSFLM (Leslie, 1992) and SCALA (Evans, 1997) or HKL2000 (Otwinowski & Minor, 1997). Initial experimental phasing trials on the haem-HemS iron- and platinum- MAD data sets were attempted with MLPHARE (CCP4, 1994), SOLVE (Terwilliger & Berendzen, 1999), SHELX (Schneider & Sheldrick, 2002) and SHARP/autoSHARP (Vonrhein *et al.*, 2006, de La Fortelle & Bricogne, 1997).

The structure of the haem-HemS complex was then solved by molecular replacement using PHASER (McCoy *et al.*, 2005) with the coordinates of the apo-ChuS structure, which were by then released (PDB code 1U9T). To solve the apo-HemS structure the coordinates of haem-HemS split in its N- and C-terminal domains were used as search models (MR and model building of

the apo-HemS structure were carried out by Dr. Katherine Sharp). Model building was carried out in COOT (Emsley & Cowtan, 2004) and restrained refinement with REFMAC5 (Murshudov *et al.*, 1997) and using TLS (Winn *et al.*, 2003). An optimal number of 10 TLS groups was determined using the TLSMD web server for the generation of multi-group TLS models (<http://skuld.bmsc.washington.edu/~tlsmd/index.html>) (Painter & Merritt, 2006b). Surface calculations were carried out using the program AREAIMOL (CCP4, 1994).

5.3 Results and Discussion

5.3.1 X-Ray crystal structure of HemS

5.3.1.1 Crystallisation of HemS

Initial high-throughput crystallisation screening of 864 conditions was carried out. Some drops in the high throughput screens showed heavy precipitation, but in many drops light precipitation could be observed (Fig. 5.1A+B). Only in one out of the 864 screened conditions, needle clusters had formed (Fig. 5.1C). A simple manual screen around the hit condition yielded larger, better clusters of crystals, though the crystal size and shape were not ideal for X-ray work (Fig. 5.1D+E).

A protocol was devised to carry out a rational grid screen with robotics to test simultaneously the effect on the crystal formation of both a range of buffers and ammonium sulphate concentrations (0.8-2.8 M). The buffers used in this screen included: MES pH 6.5, sodium cacodylate pH 6.5, Bis-Tris pH 6.5, Hepes pH 7.5, Tris-HCl pH 8.5, MOPS pH 8.0, bicine pH 9.0, CAPS pH 10.0, all at a concentration of 0.1 M. Experiments with Tris-HCl and Hepes showed a marked improvement in crystal formation relative to the MES buffer used in

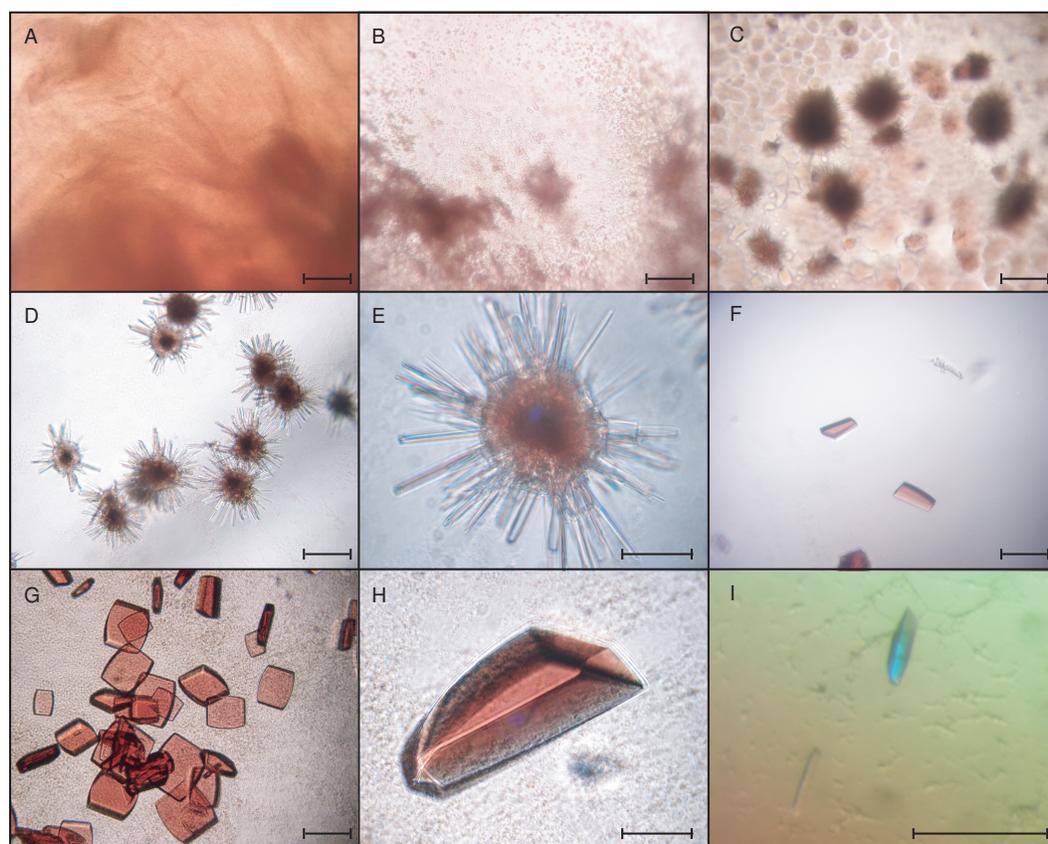


Figure 5.1: Crystallisation of *Y. enterocolitica* HemS. (A-C) high throughput screen of 864 conditions (Hampton Research, Nextal biotech) Hydra micro dispenser, sitting-drop vapour-diffusion. (HemS concentration 15 mg ml^{-1}). (A) heavy precipitation, (B) crystalline ('good') precipitation and (C) needle cluster observed in one condition out of the screened 864. 0.1 M MES, 2.0 M ammonium sulphate, 5% PEG 400, pH 6.5 (Nextal AmSO₄ suite). (D and E) reproduced needle cluster in hanging-drop vapour-diffusion in screen with 1.6-2.8 M ammonium sulphate versus 2-8% PEG 400 in 0.1 M MES, pH 6.5 and various protein concentrations. Here 1.8 M ammonium sulphate and 4% PEG 400. (F) single HemS crystals. Screen of various buffers versus various ammonium sulphate and sodium-malonate concentrations, with 4% PEG 400 using the Hydra micro dispenser. Single crystals in 0.1 M Tris-HCl, pH 8.5, 1.8 M ammonium sulphate as well as with 0.1 M HEPES, pH 7.5. (G+H) optimisation of crystallisation in hanging drop, vapour diffusion. Here: 0.1 M Tris-HCl, pH 8.5, 1.8 M ammonium sulphate, 2% PEG 400. (HemS conc. 30 mg ml^{-1}). (I) Apo-HemS crystal in 100 mM bicine pH 9, 2.45 M ammonium sulphate 5% PEG 400. (bar = $100 \mu\text{m}$).

the hit condition (Fig. 5.1F). Other screening included testing against sodium malonate (1.2-2.4 M), which has been shown to be a successful precipitant in protein crystallisation trials (McPherson, 2001). Also additives such as imida-

zole and glycerol in different concentrations were tested, as well as the effect of varying the concentration of sodium chloride, PEG 400 and protein. Finally, crystal growth was also carried out at the decreased temperature of 4°C, but this did not give improvements. After optimisation of these conditions, the best results were obtained by the hanging-drop method with drops consisting of 2 μ l HemS at 30 mg ml⁻¹, mixed with 2 μ l of 0.1 M Tris-HCl pH 8.5, 1.8 M ammonium sulphate and 2 % PEG 400, incubated at 20°C. Bright red crystals typically grew to 0.3x0.2x0.05 mm (Fig. 5.1G+H) over 48-72 h. The crystals did not grow any larger after this time. Variation of drop-size or higher protein concentration did not improve crystal size.

Small apo-HemS crystals were obtained in one condition (100 mM bicine, pH 9, 2.4 M ammonium sulphate, 5 % PEG 400) of the same rational grid screen used for the haem-HemS complex, testing various buffers against different ammonium sulphate concentrations using the crystallisation robot. Through optimisation of this condition (carried out by Dr. Katherine Sharp) more apo-HemS crystals could be obtained in 100 mM bicine, pH 9, 2.45 M ammonium sulphate, 5 % PEG 400 with a dropsize of 2 μ l + 2 μ l, but the crystal size could not be improved (Fig. 5.1I).

5.3.1.2 Structure determination of HemS

Haem-HemS crystals diffracted to 2.4 Å in-house and to 2.0 - 1.7 Å (Fig. 5.2) at the European Synchrotron Radiation Facility (ESRF). Initially structure determination of the haem-HemS complex was planned to be carried out using the multi-wavelength anomalous dispersion (MAD) method, taking advantage of the iron atom in the haem-prosthetic group. The iron-peak and inflection point were determined by an X-ray fluorescence scan around the iron-resonance edge and data sets were collected at 1.7396 Å (peak), 1.7406 Å (inflection) and 0.984 Å (remote). A highly redundant iron-MAD data set was collected which

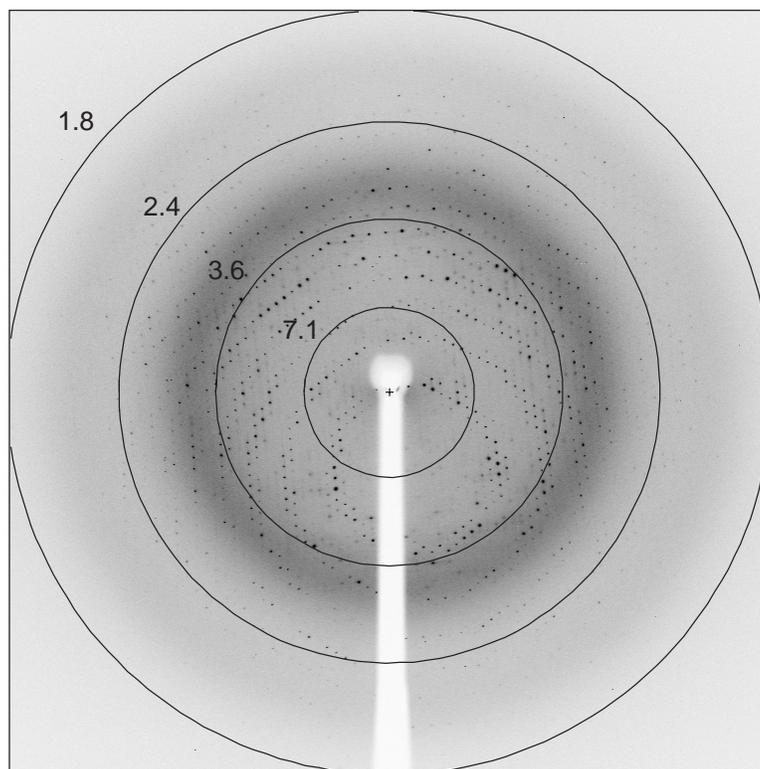


Figure 5.2: Diffraction pattern of the haem-HemS crystals at 0.984 Å. Complex crystals diffracted to 1.9 Å at the ESRF beamline BM14.

showed excellent processing statistics. Moreover a platinum-MAD data set was collected on a crystal, which was soaked over night in potassium tetrachloroplatinate (K_2PtCl_4) and diffracted to 2.6 Å. Radiation damage was observed after collection of the first two data sets (inflection, peak) on the platinum LIII-edge and data collected above the LI-edge were not used. Data processing statistics for the two MAD data sets are listed in Table 5.2 and Table 5.3. Despite the good data processing statistics of the iron-MAD data set and the clear solution of the iron-anomalous Patterson map (Fig. 5.3), phase calculations carried out using MLPHARE (CCP4, 1994), SOLVE (Terwilliger & Berendzen, 1999), SHARP/autoSHARP (de La Fortelle & Bricogne, 1997, Vonrhein *et al.*, 2006) and SHELX (Schneider & Sheldrick, 2002) resulted in poor electron density maps. Combination with phases provided by the plat-

inum derivative and solvent flattening did not significantly improve the electron density maps.

The haem-iron could be clearly located in the Harker sections of the anomalous Patterson map through a strong peak above 5σ (Fig. 5.3). Anomalous or difference Patterson map calculations, using any of the data sets collected on the platinum derivative, did not lead to any clear solutions for platinum sites.

Table 5.2: Data processing statistics of the haem-HemS Fe-MAD data. Fe-MAD data sets were collected on a single crystal on the ESRF beamline BM14. (anom. = anomalous; ip = inflection point; pk = peak).

	Iron MAD		
	peak	inflection	remote
Wavelength (Å)	1.739	1.7406	0.984
Resolution (Å)	28.5-2.15 (2.27-2.15)	28.5-2.2 (2.32-2.2)	30-1.95 (2.06-1.95)
R_{merge} (%)	6.3 (34.9)	6.7 (42.9)	6.2 (38.1)
Total No. of observations	360163 (50859)	241589 (34181)	318702 (46303)
Total No. of unique	17968 (2530)	16764 (2362)	24467 (3503)
Mean $I/\sigma I$	35.3 (8.7)	31.7 (6.1)	26.6 (6.7)
Completeness (%)	97.9 (96.2)	97.9 (96.2)	99.5 (99.1)
Multiplicity	20 (20.1)	14.4 (14.5)	13.0 (13.2)
Anom. Completeness (%)	98.2 (96.4)	98.2 (96.4)	99.6 (99.3)
Anom. Multiplicity	10.5 (10.4)	7.6 (7.5)	6.8 (6.8)

Table 5.3: Data processing statistics of the haem-HemS platinum derivative data. Platinum-MAD data sets were collected on a single crystal on the ESRF beamline BM14. (anom. = anomalous; ip = inflection point; pk = peak). Data of the PIP derivative were collected in-house.

	Platinum MAD		PIP derivative
	peak	inflection	
Wavelength (Å)	1.07144	1.07189	1.5418
Resolution (Å)	28.6-2.6 (2.74-2.6)	28.7-2.6 (2.74-2.6)	23.0-2.8 (2.95-2.8)
R_{merge} (%)	4.3 (19.2)	4.5 (28.0)	10.1 (31.5)
Total No. of observations	42613 (5810)	42825 (5839)	52750 (7570)
Total No. of unique	10724(1524)	10788 (1535)	7985 (1115)
Mean $I/\sigma I$	20.5 (6.7)	18.9 (5.2)	17.5 (6.0)
Completeness (%)	98.2 (91.7)	98.5 (91.2)	93.8 (92.6)
Multiplicity	4.0 (3.8)	4.0 (3.8)	6.6 (6.8)
Anom. Completeness (%)	98.2 (91.7)	98.5 (91.2)	
Anom. Multiplicity	2.1 (2.0)	2.1 (2.0)	

This could be caused by crystal damage during the soaking process, not fully occupied binding sites and/or deterioration of the signal during data collection, due to the fact that derivatives are more susceptible to radiation damage

because of stronger absorption of the energy of the X-ray beam. Nevertheless four to five potential platinum sites could be identified using the phases obtained by the haem-iron in MLPHARE (CCP4, 1994), albeit the occupancy was low, ranging from 0.35-0.5.

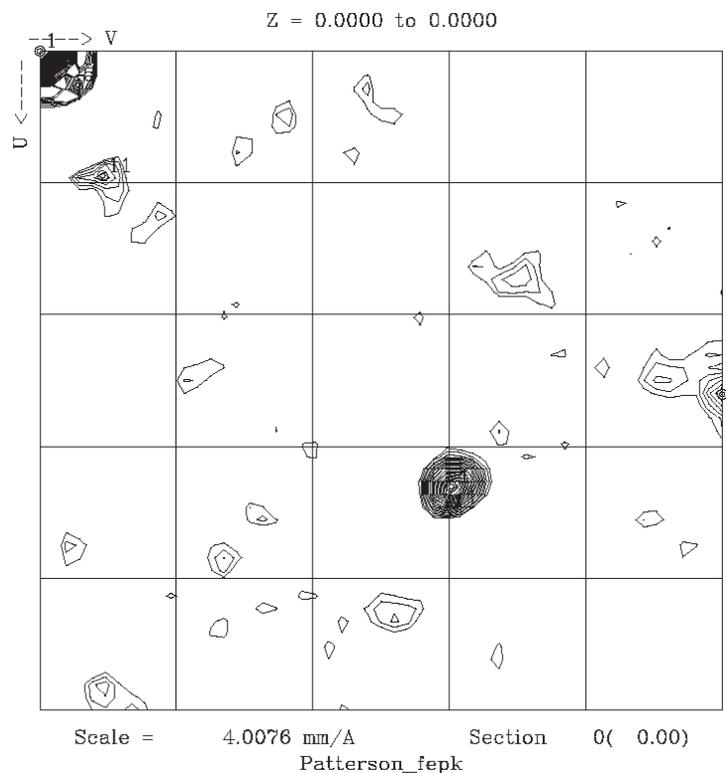


Figure 5.3: Anomalous Patterson map. Harker section of anomalous Patterson map calculated on the iron-peak data set, showing a clear peak above 5σ for the iron. The map was calculated using FFT (CCP4, 1994) and data to 2.5 \AA .

Moreover, an attempt was made to combine phases from the iron with phases obtained from the weak anomalous signal of the seven sulphur atoms in the iron-peak data set, which is relatively close to the sulphur resonance edge (sulphur-SAD data are usually collected at $1.8\text{-}1.9 \text{ \AA}$), using SHARP. This slightly improved the electron density map, but still the peptide chain could not be traced. Combining iron, platinum and sulphur phases as well as solvent flattening did not help either. Since there was only one HemS molecule in the

asymmetric unit cell, non-crystallographic averaging could not be carried out. Alternatively data were also processed with HKL2000 (Otwinowski & Minor, 1997) which did not improve the results from the experimental phasing trials. Moreover data on a PIP derivatised haem-HemS crystal were collected in-house (Table 5.3) to add information from single-isomorphous replacement to phase calculations. Two platinum sites could be identified in a difference Patterson map, calculated from the derivative and the native (remote) data set (Fig. 5.4).

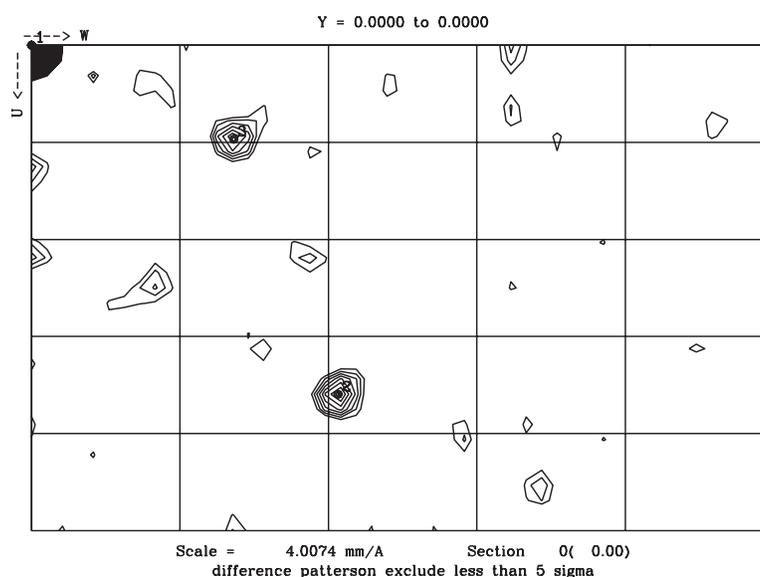


Figure 5.4: Difference Patterson map. Harker section of difference Patterson map calculated using the remote from the Fe-MAD and the PIP-derivative data sets, showing two clear peaks above 5σ . The map was calculated using FFT (CCP4, 1994) and data to 3.5 \AA .

Additionally, the 7 methionines in the HemS protein were substituted with seleno-methionine (SeMet) for structure determination by SeMet-MAD phasing. SeMet-HemS was expressed and purified (Fig. 5.5) with an overall yield of approximately 10 mg per liter culture, which is 8% of the yield for the native protein.

SeMet-HemS was purified under non-reducing conditions and reconstituted with haemin as the native protein. The redox state of the anomalous scat-

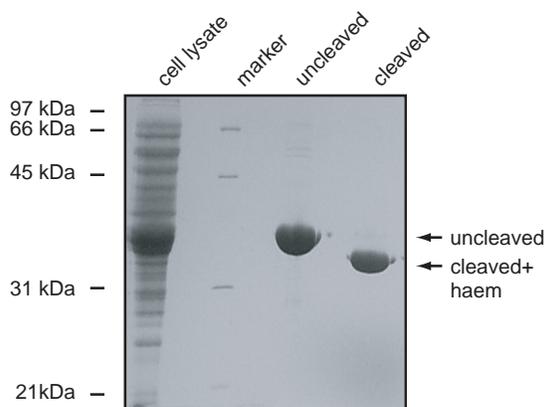


Figure 5.5: Expression and purification of seleno-methionine labeled HemS. SeMet-HemS in cell lysate (lane 1); after His-tag purification (lane 2); after removal of the His-tag and reconstitution with haemin (lane 3).

terer SeMet has a big impact on the magnitude and energy of the anomalous signal. Heterogeneity of the redox states in the crystal lattice would not only reduce the size, but also 'smear out' the resonance edge of the MAD signal and therefore could impair the precise determination of the peak and inflection point (Sharff *et al.*, 2000). Moreover it has been shown that in some cases the oxidation of the incorporated SeMet can aid the optimisation of the MAD signal (Sharff *et al.*, 2000). The haem-iron in proteins is easily oxidised and difficult to keep permanently reduced. Therefore, to ensure homogeneity of the redox state of the SeMet and the haemin, it was decided to work with a fully oxidised SeMet-HemS protein. Initial manual crystallisation screens were set up which resulted in oily drops.

Before crystallisation of SeMet-HemS and/or phase calculations using the PIP-derivative in combination with the Fe- and Pt-MAD data and the signal of the sulphurs could be tested any further, the structure of the HemS homologue ChuS from *E. coli* without its ligand haem, sharing 62% sequence identity, was published in December 2005 (Suits *et al.*, 2005). Therefore the structure of the haem-HemS complex was solved by molecular replacement using the atomic coordinates of apo-ChuS (1U9T). For the search model all residues not

shared by HemS were truncated to alanine and the residues numbered according to the HemS sequence. Furthermore the potential mobile region around His 196 (residue 193-206) was deleted and the B-factors were reset (mainchain atoms 10; sidechain 20). A clear solution was obtained with PHASER (McCoy *et al.*, 2005) with a *rotation function* Z-score of 17.7 and a *translation function* Z-score of 26.4.

The model was refined against data to 1.7 Å spacing, obtained by merging the remote of the iron-MAD data set with a data set collected at the same wavelength from a crystal with multiple ice rings diffracting to 1.7 Å. Data on an apo-HemS crystal diffracting to 1.9 Å were collected at the ESRF beamline ID14-2. The structure of apo-HemS was solved using the coordinates of the haem-HemS structure excluding the haem and residues 193-206. In addition the the N- and C-terminal domains were separated into two molecular replacement models. For apo-HemS molecular replacement and model building was carried out by Dr. Katherine Sharp. Data processing and model refinement statistics are summarised in Table 5.4. The refined models have excellent geometry and no Ramachandran outliers.

The orientation of the porphyrin was unambiguous as indicated by well-defined density peaks. The map for His196 as a ligand to the iron and for neighboring side chains such as Arg 102, Met 244 and Ile 255 were also very well defined (Fig. 5.6). His196 had previously been identified as potential ligand to the haem-iron by analysing the haem-binding properties of HemS mutants by UV/vis spectroscopy (Chapter 4.3.2.2). The vicinity of Met 244 to the porphyrin plane explains the slight difference in the Soret peak in the UV/Vis absorbance spectrum of the M244A mutant (wild type 411 nm; M244A 409 nm; Fig. 4.10, p. 78), because the mutation influences the electronic properties of the haem ligand and has therefore an impact on the absorbance spectrum.

Table 5.4: Data collection and model refinement statistics. Values in parentheses are for the highest-resolution shell. The haem-HemS data set was obtained by merging reflection data from two distinct crystals.

	apo-HemS	haem-HemS
Data Collection		
Space group	P2 ₁ 2 ₁ 2 ₁	I222
Wavelength (Å)	0.933	0.984
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	62.0, 68.3, 73.6	74.9, 77.6, 114.3
α , β , γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	30 - 1.9 (2.0 - 1.9)	30 - 1.7 (1.79 - 1.7)
<i>R</i> _{merge}	7.8 (30.7)	8.3 (22.8)
Mean I / σ I	16.9 (5.0)	13.0 (2.4)
Completeness (%)	100.0 (100.0)	97.5 (98.0)
Redundancy	6.6 (6.8)	11.7 (3.2)
Refinement		
Refinement Resolution (Å)	30 - 1.9	30 - 1.7
No. reflections	23,971	34,237
<i>R</i> _{work} / <i>R</i> _{free}	18.9 / 23.3	19.2 / 22.1
No. of Atoms		
Protein	2670	2667
Ligand	-	43
PEG/Ions	86	35
Water	134	169
B-factors		
Protein	21.9	29.9
Ligand	-	29.9
PEG/Ion	45.3	42.1
Water	23.2	34.5
R.m.s deviations		
Bond lengths (Å)	0.011	0.013
Bond angles (°)	1.367	1.474

5.3.2 Folding topology of HemS

Two topologically homologous domains join, forming a pair of large, stacked central β -sheets, which characterise the molecular architecture of HemS (Fig. 5.7A). Two distinctive pockets are created by the twist of these sheets and the helices that pack at each end of the double sheet (Fig. 5.7A+B). The haem-ligand is bound at the C-terminal domain, where the deeper, more pronounced pocket is located; the tetrapyrrole ring is buckled and distorted. The proximal side of the haem, which is the side with the coordinating

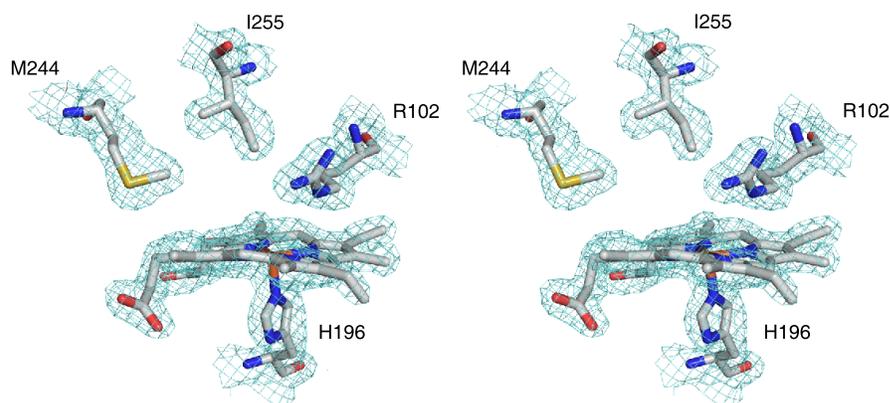


Figure 5.6: Electron density of the haem pocket. Stereo view of the haem and surrounding protein residues. The $2Fo-Fc$ density map contoured at 1σ level is shown in cyan.

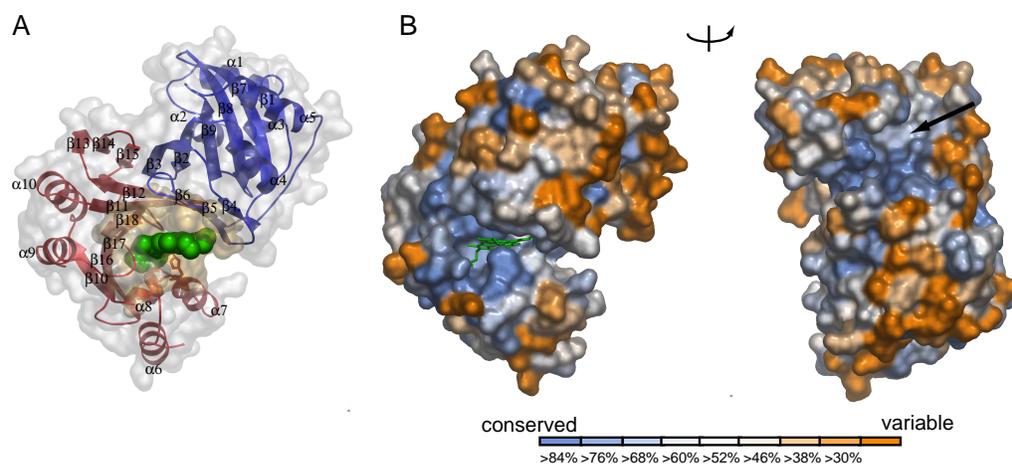


Figure 5.7: Structure of the haem-HemS complex. (A) Cartoon representation of the HemS tertiary structure with its homologous N- and C-terminal domains shown in blue and red, respectively. The deep pocket in which the haem (green) is bound is highlighted by the overlaid, semi-transparent surface representation. His 196, the ligand to the haem iron, is drawn as sticks and is clearly visible just below the haem. The striking, large β -sheet dome that embraces most of the haem is apparent. (B) The sequence conservation across Proteobacteria species is mapped onto a surface representation of the HemS molecule using ConSurf (Glaser *et al.*, 2003). The figure was prepared with ASTEXVIEWERTM (Hartshorn, 2002).

histidine, is embraced by one α -helix ($\alpha 7$) from a three-helix sub-domain ($\alpha 6$, $\alpha 7$ and $\alpha 8$); His 196 stems from the beginning of this helix. Both the haem propionate groups and its distal site are flanked by the extensive, convex face

of the twisted β -sheet. Residues that form the binding pocket clearly map onto conserved regions of the sequence (Fig. 5.7B).

In the structure of the complex, one haem ligand is bound to HemS, suggesting a one to one stoichiometry consistent with the biochemical and spectroscopic data from HemS (Chapter 4.3.2.1 and 4.3.2.2) and the homologous proteins ShuS (Wilks, 2001) and PhuS (Lansky *et al.*, 2006). The later published structure of haem-ChuS complex also showed one haem bound to the C-terminal domain (Suits *et al.*, 2006), and the authors do not mention their previously published data of one ChuS molecule binding two haems (Suits *et al.*, 2005).

5.3.2.1 Molecular recognition of haem by HemS

The haem ligand is embraced by the large, twisted β -sheet that dominates the binding pocket. All the distal residues and the groups interacting with the propionates are originating from this sheet dome (Fig. 5.8A). Over the distal side of the porphyrin plane, Arg 102 is sandwiched between Leu 92 and Ile 255, and its guanidinium group is positioned about 4 Å from the iron. Asp 100 makes a salt-bridging interaction with Arg 102 and this electrostatic contact is sealed away from the solvent by distal residues Leu 94 and Leu 92 which, together with Ala 90, align to form a barrier burying Arg 102 almost completely in the binding site. Additionally, Phe 246 packs over the propionate-bearing pyrroles and completes a line of four distal residues that starts with Leu 92, Arg 102 and Ile 255.

Placed next to Phe 246, the residues Met 244 and Val 253 are providing further hydrophobic packing (Fig. 5.8A); they form a non-polar cap shielding away a polar and electrostatic region, which includes the side chains of Arg 209, Gln 316, Tyr 318, Lys 294 and Arg 321. These residues, together with two structural water molecules, firmly anchor the propionate groups to the interior of the protein by engaging in a series of interactions (Fig. 5.8B). On the

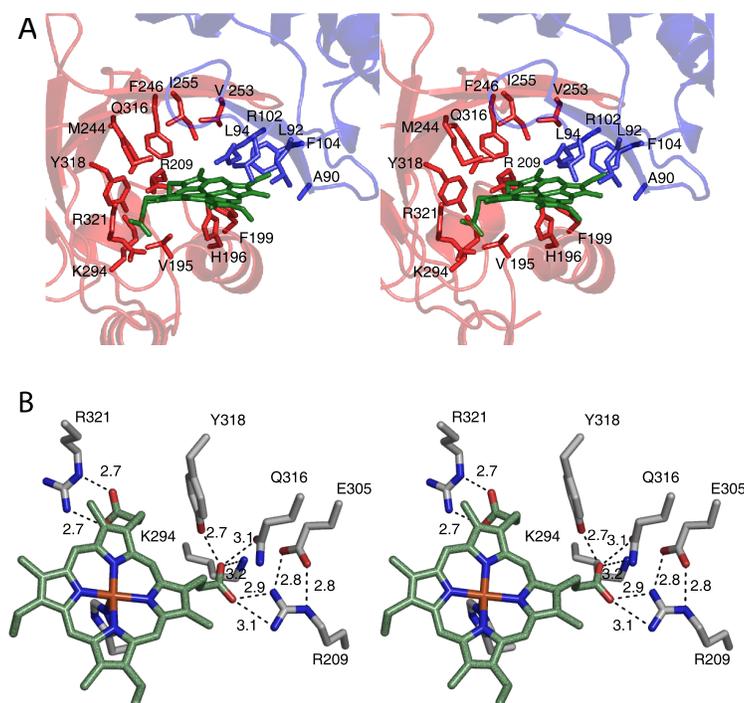


Figure 5.8: Haem-protein interactions. (A) Stereoview of the haem-binding pocket in the in the haem-HemS complex. Residues provided by the N-terminal domain are coloured in blue and residues provided by the C-terminal domain are in red. His 196 stems from the beginning of helix $\alpha 7$ and coordinates the iron. Next to Leu 94, Arg 102 extends over the porphyrin plane. (B) Residues interacting with the haem propionates. The 7 direct and 2 water-mediated (not shown) interactions build the extensive electrostatic/polar network of contacts firmly anchoring the propionate groups.

proximal side, His 196 coordinates the iron and interacts with Asp 194. Phe 199 and Val 195, packing in the vicinity of the imidazole of His 196 and making van der Waals contacts with the haem.

5.3.2.2 Comparison with other haem-proteins

A striking feature of the HemS binding site is the presence of an arginine side chain (Arg 102) which extends over the distal site of the porphyrin plane (Fig. 5.9A). Interestingly, an arginine on a topologically equivalent position is present in horseradish peroxidase (Carlsson *et al.*, 2005) (Fig. 5.9B) and other plant peroxidases (Poulos *et al.*, 1993, Smulevich, 1998). The distance from the

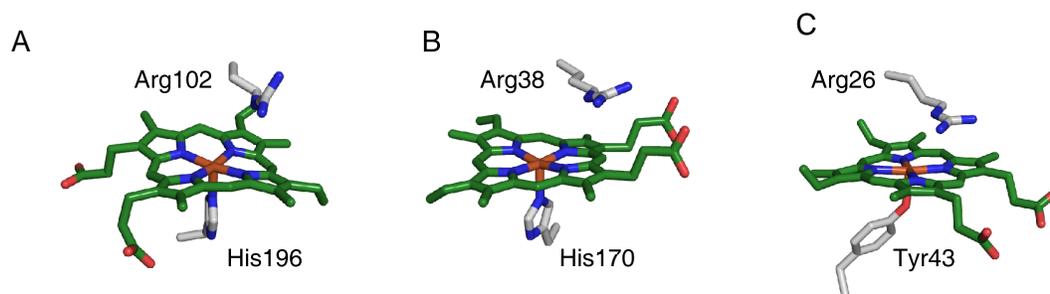


Figure 5.9: Comparison of the haem-binding pocket with other haem proteins. Arginine residues placed over the porphyrine plane in on the distal haem-side in the (A) haem-HemS complex (3.8 Å), (B) horseradish peroxidase (PDB code 1H5A) (4.5 Å) and (C) HTHP (3.7 Å)(PDB code 2OYY). Distances given are measured between the porphyrin plane and the N_{ϵ} of the arginine.

haem metal in horseradish peroxidase is similar (4.5 Å) to the distance of Arg 102 in the haem-HemS complex (3.8 Å). In contrast the N_{ϵ} of the arginine in lignin peroxidase (LiP) is more than 5 Å away from the iron (Poulos *et al.*, 1993). A recently characterised novel haem-protein, the *hexameric tyrosine-coordinated haem protein* (HTHP) from the marine bacterium *Silicibacter pomeroyi*, also has an arginine residue packing coplanar over the distal side of the porphyrin plane; its guanidinium group is placed at 3.7 Å from the iron (Jeoung *et al.*, 2007) (Fig. 5.9C). HTHP, which shows an hexameric quaternary structure, has a tyrosine to ligate the 5-coordinated haem-iron with a solvent accessible binding pocket. In all three proteins the arginine makes van der Waals contacts with the porphyrin plane.

In horseradish peroxidase the arginine exerts a strong electronic pull and, together with the distal histidine, is essential for the function (Poulos, 2006). A similar function for the arginine in HTHP has been argued, which is crucial for its catalase-peroxidase activity observed *in vitro*; however, the exact biological role remains to be determined (Jeoung *et al.*, 2007). One can only speculate about the role of the arginine in the haem-transport protein HemS. Most likely it stabilises the haem-protein complex by packing over the por-

pyrrole plane. Investigation of the function of the arginine by mutagenesis studies is currently conducted in our collaborators Dr. P. D. Barker's laboratory, Department of Chemistry, University of Cambridge.

In HemS the extensive network of 9 electrostatic contacts (7 direct and 2 water-mediated) that anchor the propionate groups is likely to provide a significant contribution to the binding (Fig. 5.8). A similarly vast set of polar/electrostatic interactions has been seen in the structure of haemopexin (Hpx), another haem-transport protein, (Paoli *et al.*, 1999), in which the propionate groups are locked into the protein by arginine, histidine and tyrosine side chains. The structure of other bacterial haem-transport proteins, like HasA, the haemophore secreted by *Serratia marcescens* (Arnoux *et al.*, 1999) and the staphylococcal IsdA (Grigg *et al.*, 2007a), IsdC (Sharp *et al.*, 2007) and IsdH (Pilpa *et al.*, 2006) proteins have a totally different folding topology and their haem pockets have no features in common with HemS.

5.3.3 Iron coordination and haem-geometry

The displacement of the haem-iron from the porphyrin plane is dependent on its ligation state and the strength of the ligands. For instance, both Hpx (bis-histidyl) and HasA (histidine and tyrosine) have a six-coordinate iron-geometry, whereas in deoxy-T-haemoglobin the five-coordination state is apparent (Table. 5.5). The iron position in HemS is intermediate between 5- and 6-coordination and spectroscopic data of the haem-HemS complex suggest a possible water at the 6th coordination position. A suitable space for a water molecule on the distal haem-site in ideal hydrogen-bonding distance with Arg 102 is available in the pocket; only weak density at the distal site was observed (Fig. 5.10), despite the excellent definition of the map for the neighboring residues and no water or hydroxide could be refined at full occupancy on the distal haem-site.

Table 5.5: Haem geometry of HemS and selected haem-proteins.

Distances	5-coordinated		6-coordinated			
	deoxy-Mb (1A6N)	FixL (1DRM)	Lignin Peroxidase (1LGN)	HemS (1J0P)	Mb-Aquo- Met(1A6K)	Hpx (1QHU)
N ϵ 2 (His) -plane	2.5	2.29	2.05	2.14	2.24	2.23
Fe-distal Ligand	3.5	-	2.7	2.4	-	2.0
Fe-plane	0.4	0.2	0.2	0.2	0.1	0.1

Interestingly, the intermediate or mixed geometry of the iron in HemS is similar to that in LiP, which is closer to a 6-coordinate system than a 5-coordinate one (Table 5.5), but electron density maps provided no evidence of ligation at the distal side (Poulos *et al.*, 1993). In many other 5-coordinate plant peroxidases the distance of the iron from the pyrrole nitrogens plane ranges between 0.01 and 0.23 Å.

Three additional data sets were collected, in order to further check on the absence of a strongly bound axial haem water ligand. In these experiments data were collected as rapidly as possible, to minimize the potential of radiation damage in the crystals which could contribute to the loss of a weakly bound water ligand. The data collection experiments were carried out at the ESRF on the ID14-2 beamline using three distinct crystals; in the best case the experiment lasted for less than 20 minutes. After molecular replacement, electron density maps were checked for peaks over the haem, at the 6th coordination position. Excellent-quality reflection data were measured to 2.2 Å spacing. Only little differences in the weak peaks of difference density over the haem were observed between the three different data sets and the previously collected data set, and again no water ligand could be satisfactorily refined. Presumably the iron in HemS has a 6th-coordinate water (or hydroxide), which agrees with the 6-coordinated UV/vis spectrum of haem-HemS (Table 4.1, p. 76);

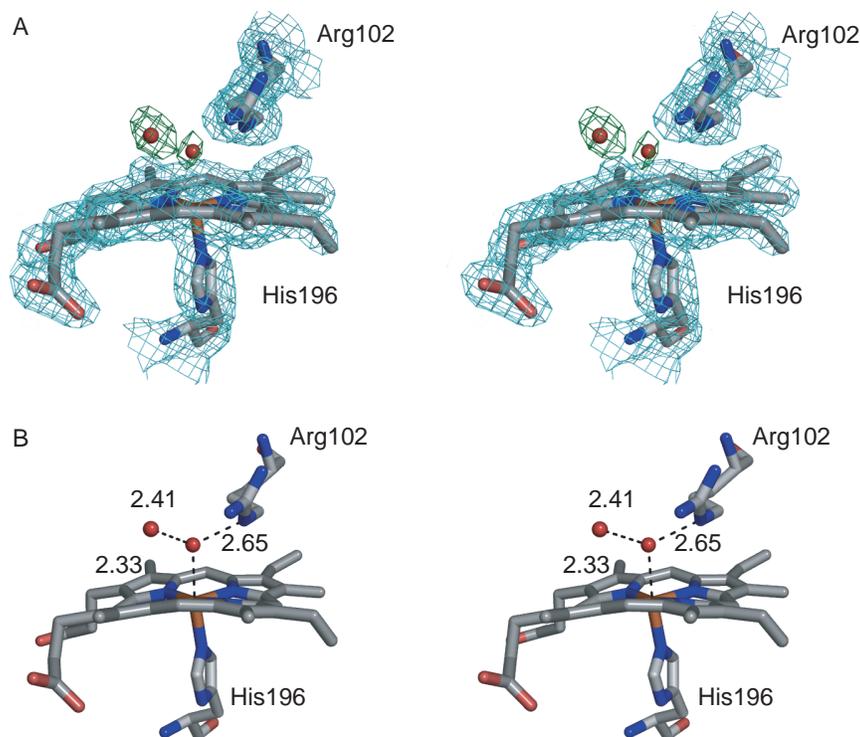


Figure 5.10: Stereo view of 6th coordination position of the haem iron. (A) The difference density map $F_o - F_c$ (green) is contoured at 3σ level and the $2F_o - F_c$ map (cyan) at 1σ level. Water molecules (red) were placed in the density peaks and (B) their distances to the haem-iron and Arg102 are shown as dotted lines.

however the porphyrin is bent by the tight clamping of the haem between His196 and Phe199 on the proximal side and Arg102, Phe246 and Leu92 on the distal side, resulting in a marked buckling of the four pyrroles away from planarity, which would make an octahedral, hexa-coordinate geometry highly distorted. Therefore the affinity for the bound water might be low and an association/dissociation equilibrium in the crystal may cause the disorder indicated by the peaks of weak difference density. In the later published haem-ChuS complex solved at 1.4 \AA resolution, the haem is slightly less buckled and two more water molecules are refined on the distal haem-site (H_2O 130 and 294)(Suits *et al.*, 2006); but as in HemS, no water is directly placed at 6th coordination position of the haem-iron.

5.3.4 Conformational changes upon haem binding

Comparison of the structure of apo-HemS, refined against data to 1.9 Å spacing, with the haem-HemS complex, revealed that local as well as global conformational changes are taking place in response to ligand binding (Fig. 5.8). Examination of lattice inter-molecular contacts confirms that these structural changes are not an artifact of crystal packing. The local changes include the switching to a different rotamer of the side chains of Leu 94 (distal), Phe 104 (equatorial) and Gln 316 (near propionates), thus becoming engaged in interactions with the ligand. In addition, many other residues shift closer to the porphyrin (Fig. 5.11). Overall, it appears that the N- and C-terminal domains (Nd; Cd) are closing into the binding pocket. Comparison of the apo- and bound-forms using a difference distance matrix (Richards & Kundrot, 1988) indeed shows that relatively few changes take place within the individual domains whilst marked shifts occur in the domains' position relative to each other (Fig. 5.12). These domain movements were also analysed

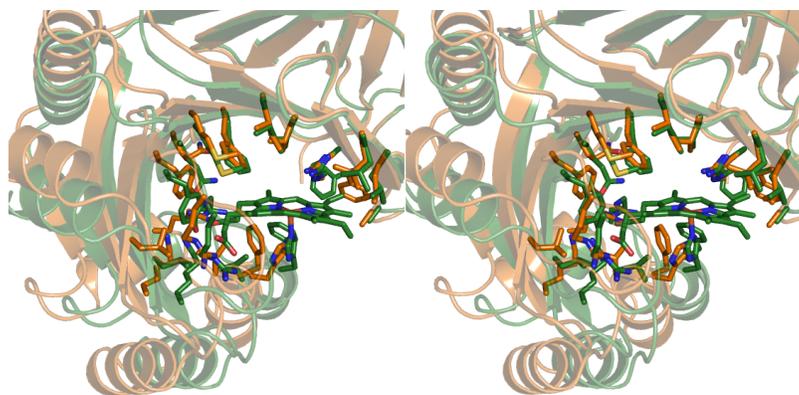


Figure 5.11: Comparison of the haem-binding pocket of the haem-HemS complex (in green) with apo-HemS (in gold). Stereo view of a superposition obtained by least-squares fitting the main chain atoms of the secondary structure elements of the N-terminal domain, as frame of reference. The flip of the Phe 104 side chain is apparent, resulting in the wedging of a haem vinyl group between Phe 104 and Ala 90. The figure also highlights the atomic shifts in the secondary structure elements surrounding the binding pocket.

by measuring the distances between the centers of gravity of secondary structure elements (Table 5.6). If HemS was separated into two distinct structural entities, it becomes apparent that the domains are affected by movements up to 4 Å, almost as rigid body shifts. Least-squares fitting of all atoms of apo- and haem-HemS gives an RMSD of 1.4 Å and a maximal displacement of main chain atoms of more than 4 Å (Fig. 5.12, Table. 5.6). Atomic superpositions using only the secondary structure elements of the Nd as a frame of reference (RMSD of 0.5 Å), show that the whole Cd and particularly the $\alpha 6$, $\alpha 7$ and $\alpha 8$ helices move towards the N-terminal distal residues and the haem (Fig. 5.13).

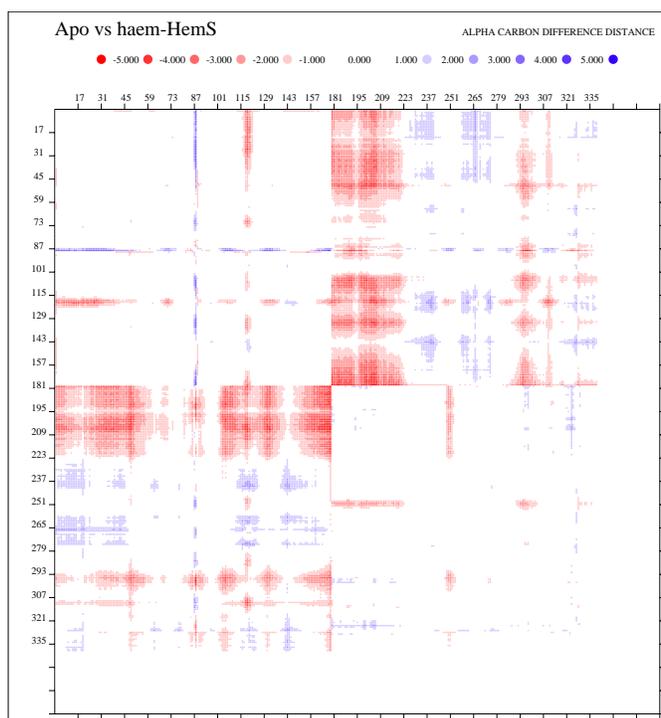


Figure 5.12: α -carbon distances comparing apo- and haem-HemS. The difference distance matrix was calculated using DDMP (Richards & Kundrot, 1988) by subtracting the α -carbon atom distances of the apo-form from the haem-HemS complex. Red indicates distances that are shorter in apo-HemS than in the ligand bound form, whereas blue indicates longer distances.

Table 5.6: Inter-domain movements upon haem binding to HemS. The centers of gravity of selected elements of secondary structure were calculated. Distances between centers of gravity (in Ångstroms) indicate that the Nd and Cd closing into the binding pocket upon association with haem. The elements of secondary structure are named as annotated in Fig. 5.7.

Nd -Cd	apo-HemS	haem-HemS
$\alpha 1 - \alpha 6$	53	51
$\alpha 1 - \alpha 7$	45	42
$\alpha 1 - \alpha 8$	43	40
$\alpha 4 - \alpha 6$	40	39
$\alpha 4 - \alpha 7$	30	28
$\alpha 4 - \alpha 8$	36	34
$\alpha 4 - \alpha 9$	36	35
$\alpha 4 - \alpha 10$	38	37
$\alpha 4 - \beta 16$	36	35
$\alpha 5 - \alpha 6$	50	48
$\alpha 5 - \alpha 7$	40	38
$\alpha 5 - \beta 16$	44	43



Figure 5.13: The HemS conformational switch between apo, open state and liganded, closed state. The superposition was prepared as in Fig. 5.11, with the apo-structure shown in orange and the complex in green. The figure shows the global inter-domain movements in HemS that effectively clamp the ligand in the binding site.

5.3.5 A haem-induced fit accompanies the switch from an open, apo-state to a closed, bound-state

Upon haem binding the Nd and Cd close in onto the ligand, effectively acting as a clamp. Therefore the HemS structure switches between an apo, open form and a liganded, closed state (Fig. 5.13). Strikingly, these global shifts also increase the twist of the large β -sheet that forms the dome of the binding site. In the open, apo-structure, haem can bind to the pocket without the need of conformational changes to remove steric barriers. The haem-pocket is partly pre-formed, but some side chains re-shape the pocket to increase the fit with the haem. Most notably, Phe104 changes rotamer to close onto the haem-edge, thus interlocking the ligand in the pocket by wedging a haem vinyl group against Ala90 (Fig. 5.11).

Equivalent conformational changes can also be observed in ChuS upon haem-binding, with movements ranging from 1 to 3 Å, as shown in the distance difference matrix plot (Fig. 5.14) and the structural superposition of apo- and haem-ChuS (Fig. 5.15).

Nevertheless the authors claimed that apart from helix $\alpha 7$ with the iron-coordinating histidine (His193), no structural changes upon haem-binding can be observed (Suits *et al.*, 2006). Since the four proteins (apo- and haem-HemS and apo- and haem-ChuS) were crystallised in 4 different crystal space groups, the observed conformational changes in HemS and ChuS are unlikely to be caused by crystal contacts. Therefore this haem-induced binding mechanism of 'clamping' onto the ligand might be universal for HemS and its homologues. The effect of ligand binding on HemS may be best described as an induced fit, as demonstrated for the glucose/hexokinase association almost described three decades ago (Bennett & Steitz, 1978). The haem-induced fit of HemS brings about a cleft closure with maximal atomic main chain displacements of

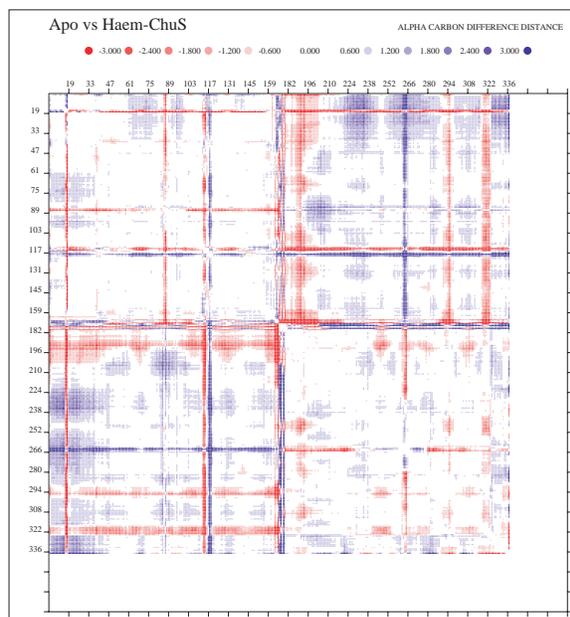


Figure 5.14: α -carbon distances comparing apo- and haem-ChuS. The difference distance matrix was calculated using DDMP (Richards & Kundrot, 1988) by subtracting the α -carbon atom distances of the apo-form from the haem-ChuS complex. Red indicates distances that are shorter in apo-ChuS than in the ligand bound form, whereas blue indicates longer distances.

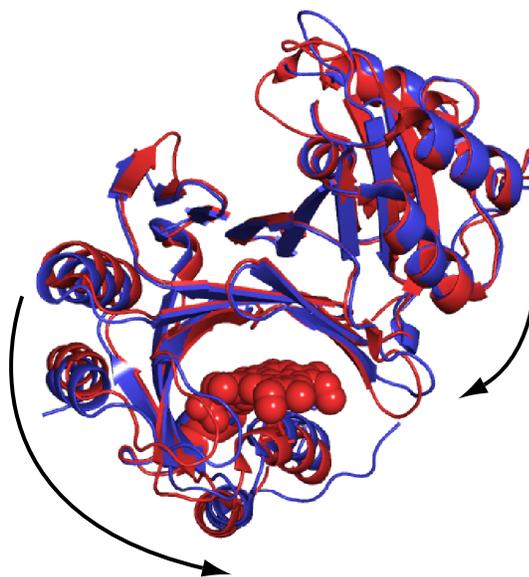


Figure 5.15: Domain movements in ChuS upon haem-binding. ChuS shows similar conformational changes as HemS between the haem-ChuS complex (in red) and apo-ChuS (in blue), when using the Nd as frame of reference. The superposition was prepared as for HemS.

over 4 Å, accompanied by the burial of 350 Å² of solvent accessible surface area additional to the area buried by the haem. The 3-helices, $\alpha 6$, $\alpha 7$ and $\alpha 8$, have higher mobility than other parts of the structure, which is indicated by analysing the temperature factors (Fig. 5.16), thus defining the 3-helix subdomain as a flexible region key for the induced-fit binding mechanism.

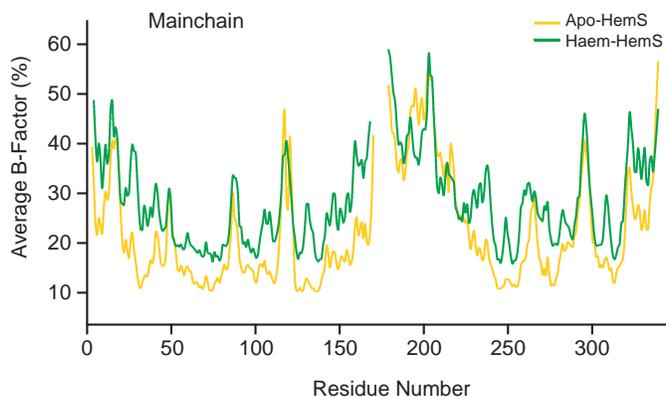


Figure 5.16: Average B-factors of mainchain atoms in the apo- and haem-HemS crystal-structures. The region with highest B-factors corresponds to the 3-helix subdomain involved in binding the proximal side of the haem.

5.3.6 Structure-function relationships and evolution of the HemS fold

In the haem-HemS complex almost 60% of the ligand's solvent accessible surface area is buried (Fig. 5.7). The haem therefore remains significantly exposed to the solvent relative to other haem proteins (e.g. catalase, haemoglobin) which incorporate haem in such a way that 85-99.5% accessible area of the prosthetic group is excluded from the solvent. However, in HemS the haem is a ligand that needs to be released as well as sequestered. The buried area of the haem in HemS is comparable to that in haemopexin (about 70%) and this reflects the haem-transfer function of these molecules. In order to release its cargo, HemS must reverse the conformational changes observed upon haem

binding. Most likely, the switching from the closed, loaded state to the apo, open form may require the binding of an effector molecule since conditions of extreme pH and high salt are not sufficient to induce haem loss (data not shown). This binding and release/transfer mechanism has been established for the homologue PhuS from *P. aeruginosa*, which interacts with the haem oxygenase paHO (Lansky *et al.*, 2006).

The similar molecular architecture of the two domains in HemS and ChuS suggests that these proteins have evolved through a gene-duplication event. This hypothesis is supported by the recently released crystal structures of the hypothetical proteins AGR_C_4470p from *Agrobacterium tumefaciens* (PDB code 2HQV) and Q6d2t7_erwct, from *Erwinia carotovora* (PDB code 2PH0). Both structures were solved by the North-East Structural Genomics Consortium and only for AGR_C_4470p a limited structural analysis was carried out (Vorobiev *et al.*, 2007). AGR_C_4470p is mostly a homodimer in solution and shares 35 % sequence identity with Q6d2t7_erwct. They superimpose with a RMSD of 1.4 Å (for 161 C α -atoms from 162 used as a reference frame) and they bear a strong structural homology to HemS, with the two monomers in the crystallographic dimer arranged in the same way as the Nd and Cd domains of HemS; but they have only a sequence identity of less than 18 % with HemS. Superposition of the Q6d2t7_erwct monomers onto each domain of apo- and haem-HemS result in RMSD ranging from 1.9 to 2.5 Å (for 169 C α -atoms out of 186 used as a reference frame) .

The two molecules form, through dimerisation, the two stacked large central β -sheets which in HemS are made up by the two domains (Fig. 5.17A). When the haem binding site of HemS is compared with AGR_C_4470p, the only common feature is the hydrophobic pocket-lining (Fig. 5.17C + 5.18). His196 is not present and therefore the authors concluded that AGR_C_4470p might not be involved in haem-utilisation, but are likely to bind a different ligand

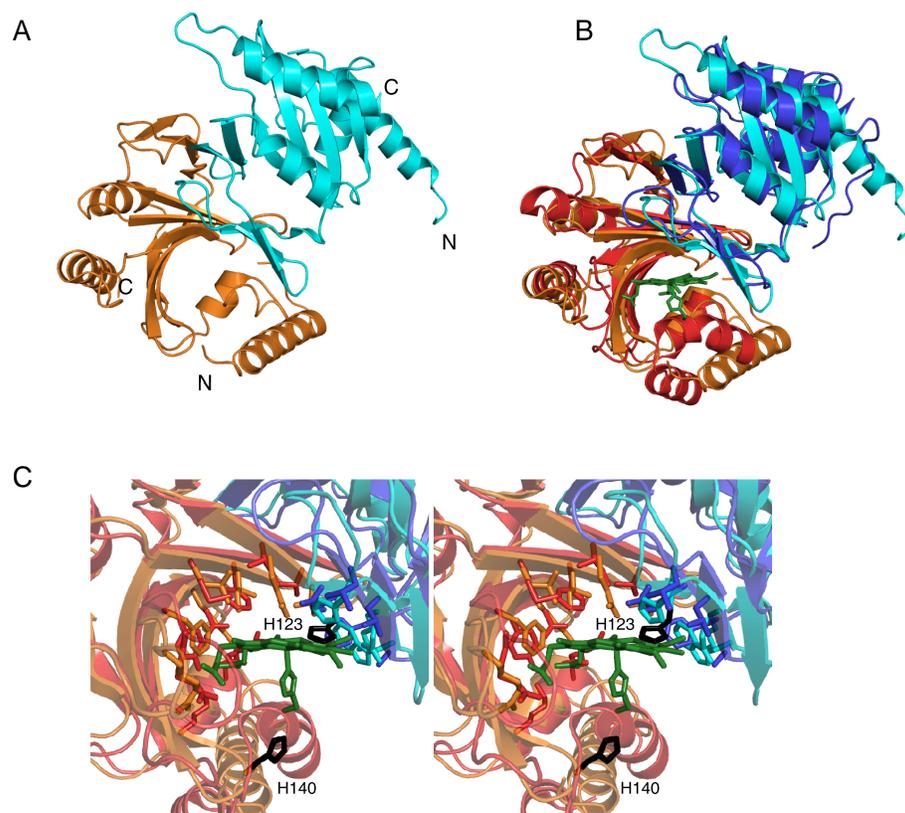


Figure 5.17: Structure comparison of HemS and AGR_C_4470p. (A) Folding topology of the AGR_C_4470p crystallographic dimer (orange and cyan) with the extended N-terminal helices. (B) Superpositions of AGR_C_4470p onto the N- (red) and C-terminal (blue) domains of haem-HemS, using SSM (Krissinel & Henrick, 2004) in COOT (Emsley & Cowtan, 2004). The arrow indicates the long, partly disordered loop connecting the two domains in HemS. The haem bound to the C-terminal domain in HemS is displayed as a stick model (green). (C) Stereo-view of the haem-binding pocket in HemS overlaid with AGR_C_4470p. Residues provided by the N- and C-terminal domains of HemS are coloured in blue and red, respectively. His 123 and His 140 of AGR_C_4470p reaching into the binding pocket are highlighted in black.

(Vorobiev *et al.*, 2007). Despite the lack in AGR_C_4470p of an equivalent to His 196, it is noticed that His 140, which originates from an adjacent loop region between $\beta 7$ and $\beta 8$, reaches into the pronounced pocket (Fig. 5.17C). The residues lining the pocket in AGR_C_4470p are conserved across its homologues identified by BLAST search (Fig. 5.19), and His 140 is replaced by a methionine in all other homologues. In addition, the conserved His 123, which

is provided by one of the two strands from the second molecule in the dimer, protrudes into the binding site from the sheet-dome (Fig. 5.17C). Interestingly, in the structural alignment His 123 is matched with His 280 in HemS and reaches into both pockets of AGR_C_4470p from the sheet-dome (Fig. 5.18). Therefore AGR_C_4470p and its homologues could potentially coordinate a haem-iron with this two residues.

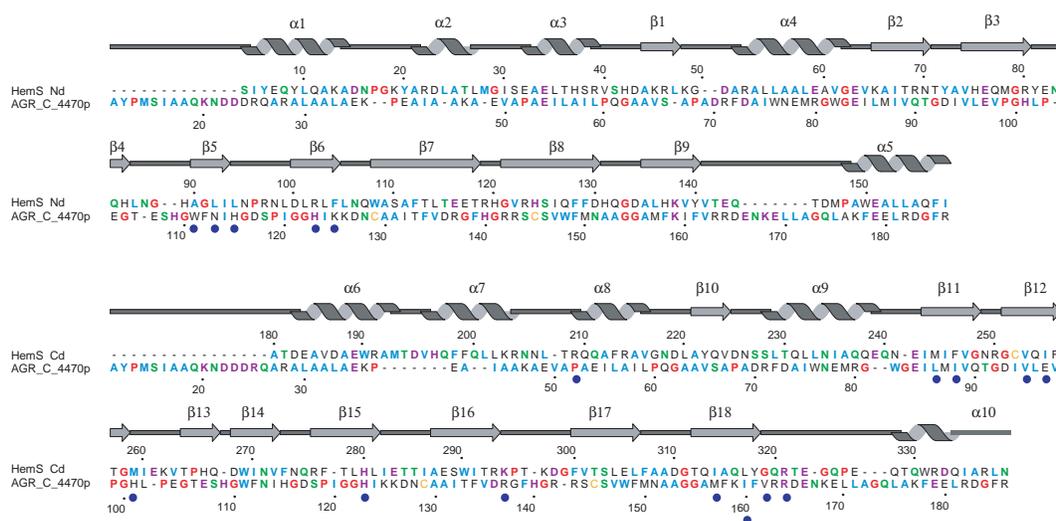


Figure 5.18: Structure based sequence alignment of HemS and AGR_C_4470p. The secondary structure annotation refers to HemS and residues lining the pockets are marked with a blue dot. The alignment of the proteins was carried out with secondary structure matching in SUPERPOSE (CCP4, 1994) and was annotated using a program kindly provided by Dr. T. Stevens (Stevens & Paoli, 2007).

AGR_C_4470p belongs to the Pfam (Bateman *et al.*, 2004) protein family DUF1008, which contains 45 proteins of unknown function. More homologues of AGR_C_4470p were identified through BLAST searches (Altschul *et al.*, 1990) of the NCBI database, sharing a sequence identity ranging from 30-40% and all belonging to the group of Proteobacteria species. Amongst these identified proteins are for instance ShuX from *S. dysenteriae* and ChuX from *E. coli*, as well as the hypothetical proteins YE0334 and OrfX (OrfA in Thompson *et al.* (1999)) from *Y. enterocolitica* and *Y. pestis*, respectively (Fig. 5.19).

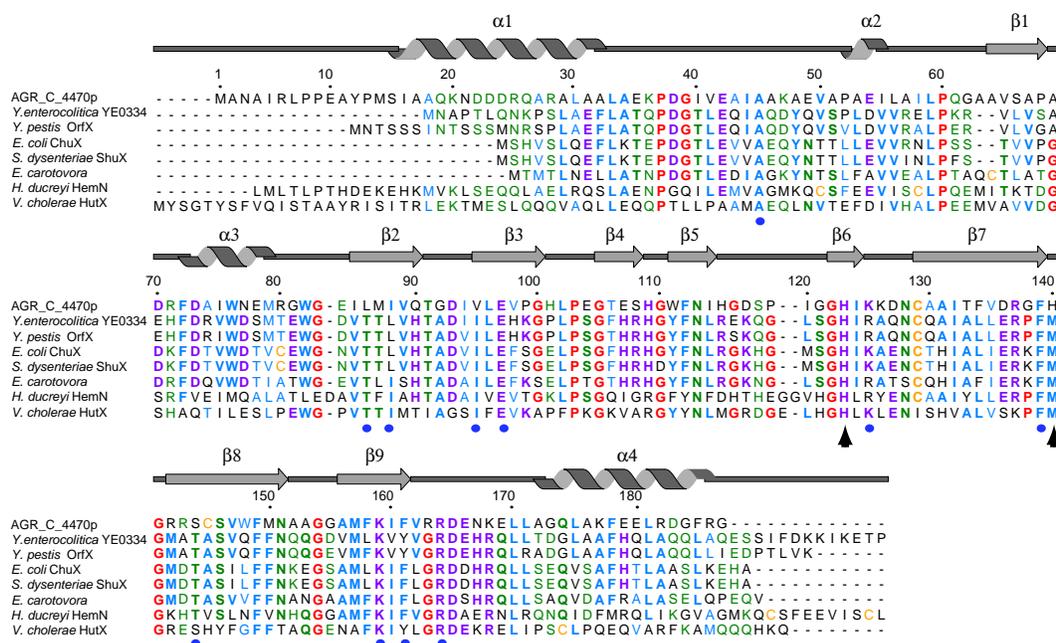


Figure 5.19: Sequence alignment of representative sequences of homologues of AGR_C_4470p. Homologues identified by BLAST search using the AGR_C_4470p protein sequence, are unique to the phylum of Proteobacteria and were aligned using ClustalX (Thompson *et al.*, 1994). The secondary structure annotation and numbering refer to AGR_C_4470p. Residues lining the binding pocket equivalent to the haem binding site in HemS are marked with a blue dot. The residues His123 and His140 are highlighted with an arrow. The sequences are: AGR_C_4470p (gi:116667786), *Erwinia carotovora* Q6d2t7_erwct (gi:146387157); *E. coli* ChuX (gi:15833638); *Y. enterocolitica* hypothetical protein YE0334 (gi:123440720); *Y. pestis* OrfX (gi:22124460); *S. dysenteriae* ShuX (gi:82778667); *Haemophilus ducreyi* HemN (resi 402-620) coproporphyrinogen III oxidase (gi:33151310); *V. cholerae* HutX (gi:121726494). The alignment was annotated using a program kindly provided by Dr. T. Stevens (Stevens & Paoli, 2007).

Search of the *Y. enterocolitica* genome sequence from the Sanger genome sequencing project (http://www.sanger.ac.uk/Projects/Y_enterocolitica/) revealed that YE0334 is encoded upstream of HemR as well as a homologue to ShuY from *S. dysenteriae*, Y0333 (Fig. 5.20). Therefore it is likely that YE0333 and Y0334 are regulated and expressed in concert with the HemRSTUV system.

No haem-binding trials with either AGR_C_4470p or Q6d2t7_erwct, or any of their homologues have been carried out so far. However, given the conservation of residues that are known to be able to coordinate a haem-iron in their

potential binding pocket, as well as the co-localisation with the HemRSTUV system in many Gram-negative species (Table. 1.2, p. 8 and Fig. 5.20), one could speculate that these proteins and their homologues are indeed involved in haem-binding and possibly haem-utilisation.



Figure 5.20: Organisation of the HemRSTUV operon. Operon organisation as annotated at the Sanger Institute (<http://www.sanger.ac.uk/>). The hypothetical proteins encoded by ORFs YE0333 and YE0334 have ~30% sequence identity to ShuY and ShuX and their homologues, respectively.

5.4 Conclusions

When the work described here was started, no structural and only limited biochemical/biophysical data were available on any bacterial haem acquisition protein, with the exception of HasA. Here HemS with and without its ligand haem were crystallised and the X-ray crystal structures determined. The structure of the haem-HemS complex shows the molecular recognition of the ligand by HemS. The haem is bound to its C-terminal domain and His 196 is the ligand to the haem-iron, which was indicated by the mutagenesis and spectroscopy work in Chapter 4.

Comparison of the apo- and haem-bound HemS structures gave precious insights into the molecular mechanism of haem-binding, which can be described as an haem-induced fit mechanism. The two domains of HemS effectively act as a clamp, closing in onto the ligand and thus HemS appears to switch from an open, apo-state to a closed, bound-state.

The binding of haem to HemS has to be reversible in order to fulfill its transport function. The conformational changes observed between the apo- and

haem-bound structures indicate that release of the haem might require the interaction with an effector molecule. Such an effector molecule has been identified for the HemS homologue PhuS from *P. aeruginosa*, the haem-oxygenase paOH (Lansky *et al.*, 2006), but homologous protein no paOH or any known haem oxygenase could be found in the genome sequences of *Y. enterocolitica*, *E. coli* or other Proteobacteria. Therefore a target molecule for HemS still remains elusive.

The recently determined structures of the hypothetical proteins AGR_C_4470p and Q6d2t7_erwct, from *Agrobacterium tumefaciens* and *Erwinia carotovora* are supporting the hypothesis that the HemS folding topology has evolved through a gene duplication and fusion event. Despite the insignificant sequence identity to HemS and its homologues, the two molecules in the homo-dimer are exactly arranged as the N- and C-terminal domains in HemS. No counterpart for the iron-coordinating residue in HemS, His 196, is present in these proteins, but two strictly conserved residues that could ligate a haem-iron were identified and seen to reach into the pocket. These observations, taken together with the finding that homologues of AGR_C_4470p and Q6d2t7_erwct are encoded next to the HemRSTUV system in many Proteobacteria, make it likely that these proteins play a role in haem / iron acquisition and utilisation. Unfortunately no haem-binding studies have been carried out with these proteins and their exact function as well as their possible role in haem-utilisation remain to be elucidated.

Chapter 6

Cloning, expression and purification of HemT from *Yersinia enterocolitica*

6.1 Introduction

In Gram-negative bacteria substrate transport across the periplasmic space is commonly carried out by a soluble periplasmic binding protein (PBP), which shuttles its cargo between the outer cell membrane receptor and inner cell membrane permease (Wilks & Burkhard, 2007). The components of the transport system (outer membrane receptor, PBP, inner membrane permease) are ligand specific and not interchangeable (Wilks & Burkhard, 2007).

The common structural feature of the PBPs characterised to date are two globular domains connected by a flexible hinge region with the ligand binding pocket lying between the two domains. PBPs are classified in three groups according to the number of domain-connecting elements (three, two or one) (Wilks & Burkhard, 2007). At present no structural information on haem PBPs is available, but they show weak sequence identity (18-22 %) to the

Vitamin B12 binding protein (BtuF) (PDB code 1N2Z) and *iron-hydroxamate binding protein* (FhuD) (PDB code 1ESZ) from *E. coli*, which have one α -helix bridging between the two domains (Fig. 6.5A, p. 131).

Comparisons of the apo-BtuF and BtuF-cobalamin (vitamin B12) crystal structures indicated a hypothetical mechanism of cargo binding and release: Local unwinding of the α -helix connecting the two domains increases the flexibility in the apo-structure and the release of the cobalamin is driven by a conformational change in the BtuF-permease interaction induced by a 'tweezer' motion of the permease upon ATP-hydrolysis. Two glutamine residues each sitting on one domain of BtuF are necessary to interact with arginine residues on the permease (Borths *et al.*, 2002).

Homologues from 52 different species to the haem PBP PhuT, from *P. aeruginosa*, were identified, showing low sequence identities of 30-40% (Tong & Guo, 2007) with a conserved tyrosine residue potentially involved in haem-iron coordination. Figure 6.1 shows a sequence alignment of selected proteobacterial haem PBPs. Recently two haem PBPs, ShuT from *S. dysenteriae* and PhuT from *P. aeruginosa*, have been biochemically and biophysically characterised. Using site-directed mutagenesis, UV/vis absorption, resonance raman and circular dichroism spectroscopy, it has been shown that they bind one haem per molecule, with the conserved tyrosine residue as the proximal ligand to a high-spin haem-iron (Eakanunkul *et al.*, 2005, Tong & Guo, 2007). ShuT binds haem with high affinity and the complex is extremely stable over a wide pH range and relatively redox inactive (Eakanunkul *et al.*, 2005). This iron-coordinating tyrosine residue in ShuT and PhuT is conserved throughout the putative periplasmic haem binding proteins (Fig. 6.1 and Eakanunkul *et al.* (2005), Tong & Guo (2007)). Additionally, far-UV CD spectroscopy and limited proteolysis assays revealed conformational changes of PhuT upon haem-binding (Tong & Guo, 2007). Moreover the glutamine residues, shown

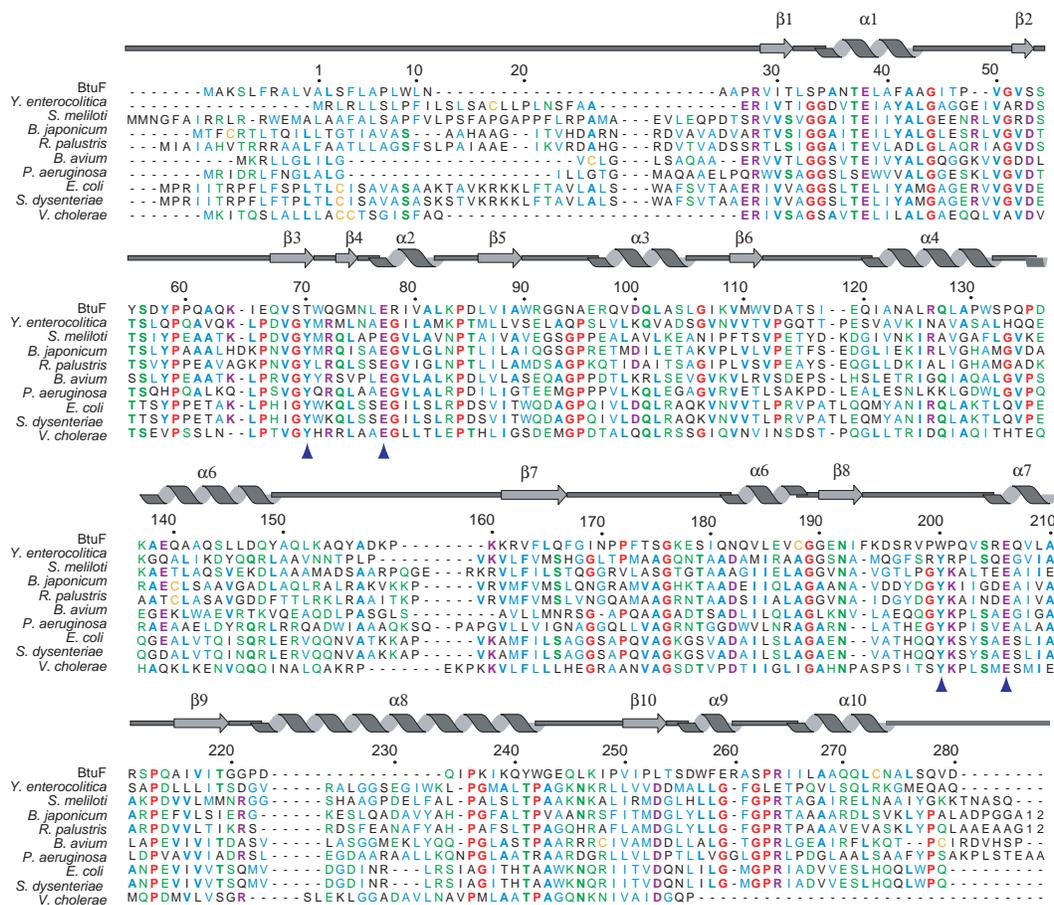


Figure 6.1: Sequence alignment of HemT and HemT homologues. Multiple sequence alignment of 9 HemT homologous sequences from representative Proteobacteria species. Homologous sequences to *Y. enterocolitica* HemT were identified by BLAST searches (Altschul *et al.*, 1990) of the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/>) and alignments were carried out with ClustalW (Thompson *et al.*, 1994). The alignment was annotated with a program kindly provided by Dr. T. Stevens (Stevens & Paoli, 2007). Numbering correspond to *Y. enterocolitica* HemT and is indicated on top of the alignment. The secondary structure assignment corresponds to the *E. coli* vitamin B12 binding protein (BtuF) crystal structure, which shares 21 % sequence identity with HemT. HemT, HmuT, ShuT and PhuT share around 30 % sequence identity and have a sequence similarity ranging from 44 to 52 %. The arrows indicate the conserved tyrosine residue, ligand to the haem-iron in the homologous proteins ShuT and PhuT from *S. dysenteriae* and *P. aeruginosa* as well as the conserved Glu residues, essential for the interaction with the permease in BtuF. Also highlighted are the conserved aromatic residues packing against the cobalamine ligand in BtuF (Borths *et al.*, 2002).

to be essential for the interaction with the permease are conserved in the haem-PBP proteins (Fig. 6.1). Therefore it is assumed that the structure and

molecular mechanism of HemT is similar to BtuF (Wilks & Burkhard, 2007). HemT also shares weak sequence identity (~22%) with the haem-binding proteins from Gram-positive bacteria, IsdE from *S. aureus* and HtsA *S. pyogenes*. IsdE and HtsA are extracellular lipoproteins that are attached to the cell wall by their C-terminus (Skaar & Schneewind, 2004, Lei *et al.*, 2003). The tyrosine residue conserved in the haem-PBPs is not conserved in IsdE and HtsA, suggesting that these proteins must employ different haem-coordinating residues, despite sharing the overall folding topology. HtsA and haem-transport in *S. pyogenes* will be discussed in Chapter 7.

In this chapter the cloning, the expression and purification for structural studies using X-ray crystallography of the haem PBP HemT from *Y. enterocolitica* and its homologue HmuT from *Sinorhizobium meliloti* are described.

6.2 Materials and methods

6.2.1 Cloning of *hemT* from genomic DNA

Initially the full-length sequence of *Y. enterocolitica hemT* (gi:1619623) and the gene coding for the homologous protein HmuT *Sinorhizobium meliloti* (gi:15966182) were PCR amplified (see Chapter 2.2.2.1, p. 38 and Appendix, p. xix) from genomic DNA (genomic DNA previously prepared by Dr. M. Paoli). *HemT* and *hmuT* were cloned into the expression vector pGAT2, a pGAT (Peränen *et al.*, 1996) derivative (Appendix A.3, p. xx). For insertion into pGAT2, the PCR products and the vector were cleaved with HindIII and BamHI. Ligation, transformation and analysis was carried out as described in Chapter 2.2.3, p. 39.

Because of insoluble expression of the full-length *Y. enterocolitica* HemT, a truncated version with its own leader sequence omitted was re-cloned into pGAT2 (YE-HemT25). Additionally *hemT* was also cloned replacing its own

leader sequence with the *E. coli* PelB sequence into pET22b (Novagen) (Appendix A.3, p. xxi) with and without a C-terminal His₆-tag.

6.2.2 Expression and purification of HemT

Expression trials for His-GST tagged YE-HemT and SM-HmuT in BL21 (DE3) cells were carried out as described in Chapter 2.2.4, p. 40. The influence of temperatures (20-30°C), induction point ($OD_{600} = 0.4 - 1.2$) as well as length of expression (4 h - ON) were tested and some soluble expression for SM-HmuT could be achieved. YE-HemT and YE-HemT25 expressed only insoluble or did not express at all.

Purification of SM-HmuT using GST-sepharose and Ni-NTA resin were carried as described in Chapter 2.2.4, p. 40, but the protein was too unstable and prone to degradation.

Expression of the YE-HemT25 construct where the PelB sequence was cloned N-terminal in frame with YE-HemT25 was carried out as described in Chapter 2.2.4, p. 40. Cells were induced when an $OD_{600} = 1.0$ was reached, the temperature dropped to 25°C and the cells were further grown overnight. Due to the PelB leader sequence, HemT is secreted to the periplasmic space and the signal peptide is cleaved off. The protein was then purified from the periplasma by osmotic shock (Shouldice *et al.*, 2003). The cells were harvested by centrifugation (2360 g, 4°C, 20 min) and resuspended in 1/10 volume of ice-cold 30 mM Tris-HCl, pH 8.0, 20% (w/v) sucrose, 5 mM EDTA. The cell suspension was rocked for 10 min at RT. After centrifugation (2250 g, 4°C, 20 min) the supernatant was discarded and all traces removed by inverting the bottles on a paper towel. The cells were then rapidly resuspended in 10 ml of ice-cold 5 mM MgSO₄ and incubated on ice for 20 min. After that the cells were pelleted by centrifugation (8000 g, 4°C, 30 min) and the supernatant containing the periplasmic osmotic shock fluid was carefully removed and the buffer

exchanged to 50 mM Tris-HCl, 200 mM NaCl, pH 8 using spin columns. The His-tagged construct was purified using Ni-NTA resin (Chapter 2.2.4, p. 40) and the buffer exchanged to 50 mM Tris-HCl, 200 mM NaCl, pH 7.5 using centrifugal filter devices (Vivascience).

6.2.3 UV/vis spectroscopy and reconstitution of HemT with haemin

Reconstitution of *Y. enterocolitica* HemT with its ligand haem and UV/visible absorption spectroscopic analysis was carried out as described in Chapter 2.2.5, p. 41, but with the protein in 50 mM Tris-HCl, 200 mM NaCl, pH 7.5.

6.3 Results and Discussion

6.3.1 Expression and purification of HemT

The sequence coding for the full-length protein HemT from *Yersinia enterocolitica* and its homologue HmuT from *Sinorhizobium meliloti* was cloned from genomic DNA into the expression vector pGAT2, in frame with a His-GST double tag. Expression trials at temperatures ranging from 20-30° were carried out (Fig. 6.2A), but no soluble expression of *Y. enterocolitica* HemT and only limited soluble expression for *S. meliloti* HmuT could be achieved. HemT was also cloned omitting its predicted leader sequence (residues 1-25), but this truncated construct (YE-HemT25) did not express because of a frame shift mutation in the primer binding site, probably caused by an error during primer production. Soluble expression of HmuT could be observed at 25°C (Fig. 6.2A). HmuT was expressed on a larger scale (Fig. 6.2B) followed by purification using the His- and GST-tag, but the protein degraded during this process despite the addition of protease inhibitors (Fig. 6.2C+D). Therefore

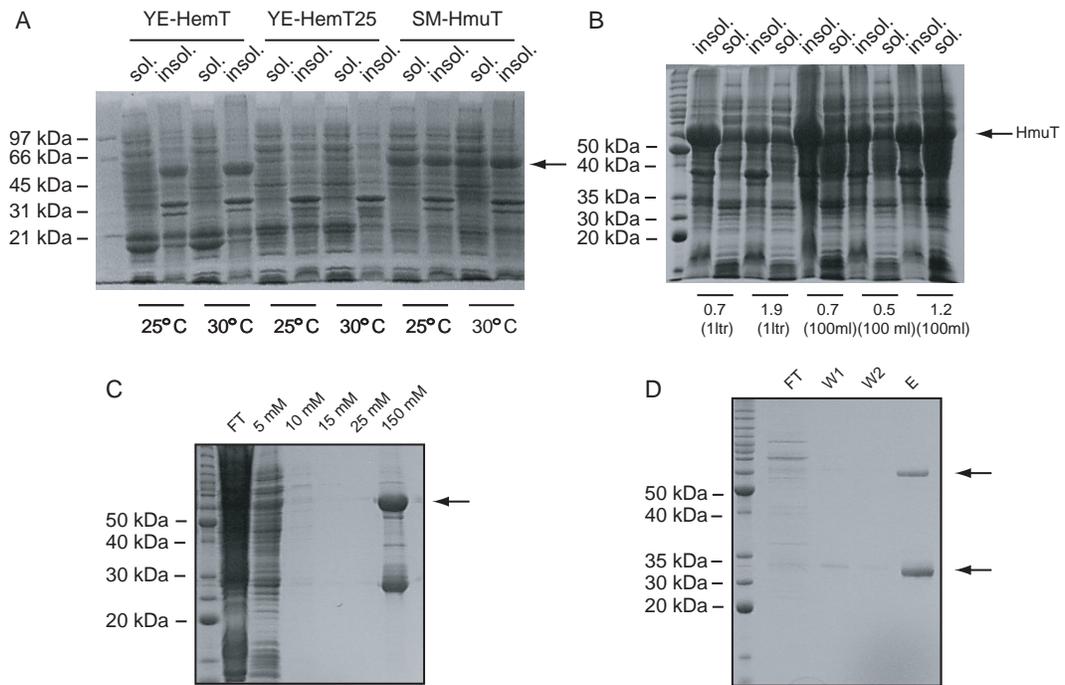


Figure 6.2: Expression optimisation and purification of *Y. enterocolitica* HemT and *S. meliloti* HmuT. (A) Small-scale expression of *Y. enterocolitica* HemT and *S. meliloti* HmuT at 25°C and 30°C. (B) Large-scale expression of *S. meliloti* HmuT. (C) Degradation of HmuT during His-tag and (D) GST purification. Arrows marking the His-GST-HmuT and the degradation product.

Y. enterocolitica HemT was recloned replacing its own leader sequence (residues 1-24) with the *E. coli* PelB leader sequence, which directs the fusion protein to the *E. coli* periplasmic space upon expression. The PelB signal sequence is cleaved off the fusion protein during the transport from the cytosol to the periplasma. The PelB-HemT construct expressed solubly and was purified from the periplasm by disruption of the outer cell membrane by osmotic shock (Fig. 6.3A).

To facilitate additional purification upon periplasmic expression, HemT in frame with the PelB sequence was also cloned with a C-terminal, non-cleavable C-terminal His-tag (Fig. 6.3B). The soluble expression of PelB-HemT was very high and not all the protein was transported in the periplasmic space. Thus in large-scale purification contamination with the cell fraction containing HemT

still in fusion with the PelB sequence was an issue and the procedure had to be optimised. The yield of pure HemT protein was approximately 60 mg per liter cell culture, was concentrated to about 40 mg ml⁻¹ and was stored at 4°C.

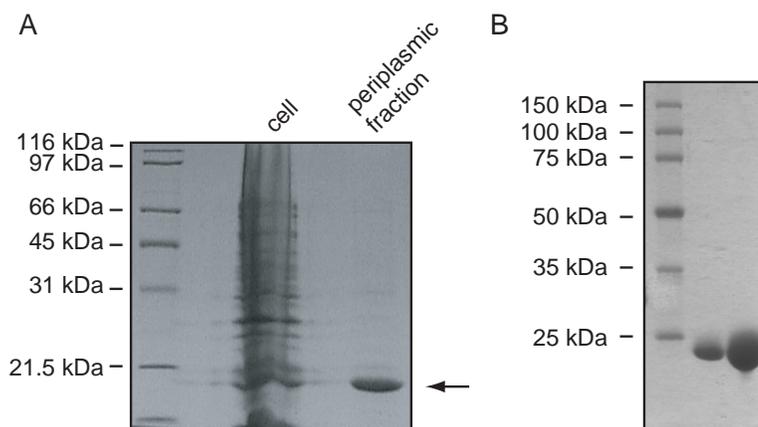


Figure 6.3: Expression and purification of *Y. enterocolitica* HemT. (A) HemT after concentration of osmotic shock fluid obtained in small scale expression and purification trials. (B) Pure HemT-His protein after optimisation of large-scale expression and purification with additional His-tag purification.

6.3.2 UV/vis spectroscopy and reconstitution of HemT with haemin

The tag-free HemT concentrated from the periplasmic osmotic shock fluid shows a Soret maximum at 400 nm and Q-bands at 498 nm and 535 nm as well as a high-spin marker band at 620 nm which increase upon addition of haemin solution (Fig. 6.4A). Figure 6.4B depicts the spectra of the reconstituted ferric haem-HemT-His complex, which has a slightly shifted absorption maxima of 398 nm, 503 nm, 537 nm and 627 nm (Fig. 6.4B), which could be caused by the additional C-terminal His-tag. The spectral maxima are similar to the homologues from *P. aeruginosa*, PhuT (400 nm, 500 nm, 534 nm, 624 nm) and *S. dysenteriae* ShuT (400 nm, 500 nm, 521 nm, 617 nm), which were both prepared as His-tag fusion proteins (Tong & Guo, 2007, Eakanunkul *et al.*, 2005).

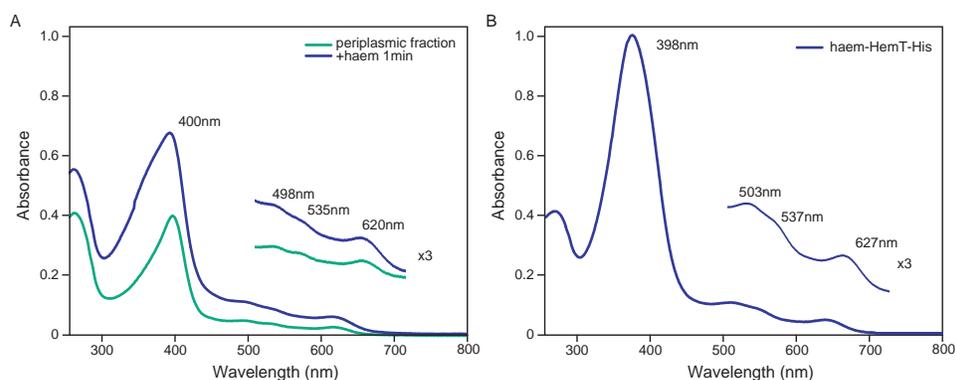


Figure 6.4: UV/vis absorption spectrum of *Y. enterocolitica* HemT. (A) Spectrum of HemT as obtained from the periplasmic fraction and after addition of haemin. (B) Reconstituted HemT-His₆.

Recently published site-directed mutagenesis carried out on ShuT and PhuT have shown that they have a 5-coordinate, high-spin haem-iron with a tyrosine (ShuT Y94; PhuT Y71) as axial ligand (Eakanunkul *et al.*, 2005, Tong & Guo, 2007). The similarities of the absorption spectra of HemT with its homologues as well as the conservation of the tyrosine in multiple sequence alignments (Fig. 6.1 and Tong & Guo (2007)) are suggesting a 5-coordinate, high-spin haem-iron with tyrosine 70 as axial ligand in *Y. enterocolitica* HemT. Due to the low sequence identity between BtuF and HemT, only a very crude homology model of HemT was generated, using the coordinates of BtuF (Fig. 6.5A) and Swiss-PDBviewer (Guex & Peitsch, 1997). No attempt was made to model the haem-ligand in the binding pocket, also due to the low sequence identity. Nevertheless, the model shows Tyr 70 in vicinity of the cobalamin ligand (Fig. 6.5B). Also highlighted are the two conserved glutamine residues (Glu 77 and Glu 206), which fall at the same topological places as in BtuF, and are essential for BtuF to interact with the inner membrane permease BtuCD (Borths *et al.*, 2002).

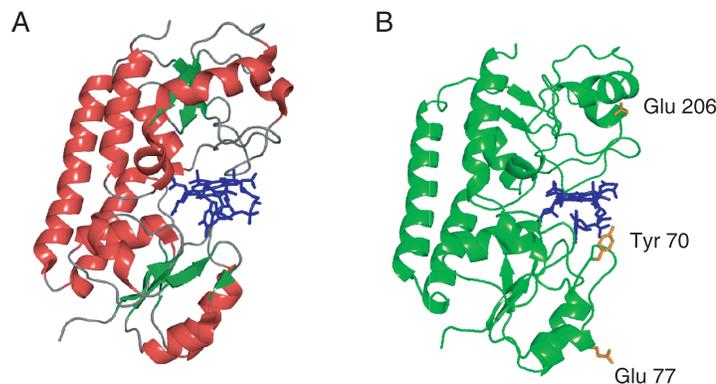


Figure 6.5: Structure of *E. coli* BtuF and homology model of HemT. (A) Cartoon representation of the crystal structure of the *E. coli* vitamin B12 binding protein (BtuF, pdb code 1NZU) in complex with cobalamin. (B) Homology modelling of HemT was carried out using Swiss-PDBviewer (Guex & Peitsch, 1997) and the coordinates of *E. coli* BtuF. The cobalamin ligand of BtuF is left in the model of HemT due to the low sequence identity between HemT and BtuF. Tyrosine 70 (orange) is in the vicinity of the cobalamin from BtuF (blue). Also in orange, residues Glu 77 and Glu 206 potentially interacting with the inner membrane permease.

6.4 Conclusions

Soluble periplasmic haem binding proteins are required in Gram-negative bacteria to transport their cargo from the outer membrane receptor to the inner membrane permease. In recently carried out biochemical and biophysical studies on the haem PBP homologues from *S. dysenteriae* and *P. aeruginosa*, a tyrosine was identified as the proximal ligand to the haem-iron in a 5-coordinate high-spin system (Eakanunkul *et al.*, 2005, Tong & Guo, 2007). This tyrosine residue is conserved across species and it can therefore be assumed that this coordination state is a common feature of these haem PBPs. The haem-transport proteins IsdE and HtsA from the Gram-positive species *S. aureus* and *S. pyogenes*, share a weak sequence identity and therefore are likely to share the overall folding topology, but not have the conserved tyrosine residue. Hence they must employ different residues to coordinate the haem-iron. This will be further discussed in Chapter 7. To date no structural data

on haem-PBPs or their homologues from Gram-positive species are available. In this thesis, the cloning and expression trials with the haem PBP homologues HemT from *Y. enterocolitica* and HmuT from *S. meliloti* are reported. Limited soluble expression of His-GST tagged fusion protein from *S. meliloti* was observed, but the protein was subject to degradation during purification. A total of four constructs were prepared to obtain soluble expression for HemT from *Y. enterocolitica* of sufficient quantity and quality for future crystallisation trials. Solution of the structure could possibly be accomplished by molecular replacement using the coordinates from BtuF (1N2Z) as a model, but might prove difficult given the low sequence identity between BtuF and HemT. Experimental phases obtained using the anomalous signal of the iron in the haem prosthetic group in a MAD experiment could be sufficient to solve the HemT structure, but might be challenging given the size of the protein to be phased with one haem-iron lone. Combination of a possible weak molecular replacement solution with experimental iron-MAD phases, are more likely to be successful for structure determination (Baker *et al.*, 1995). Alternatively, since HemT contains 12 methionine residues in 256 residues, structure determination could be carried out by SeMet-MAD.

Chapter 7

Structure-function relationship of HtsA and Shp from *Streptococcus pyogenes*

7.1 Introduction

Streptococcus pyogenes, or group A *Streptococcus* (GAS), is a Gram-positive human pathogen which causes a variety of diseases, such as pharyngitis, crochizing fasciitis, streptococcal toxic shock syndrome, scarlet fever, and postinfection sequelae including glomerulonephritis, rheumatic heart disease and rheumatic fever (Cunningham, 2000).

In the search of the GAS genome sequence for genes coding for putative extracellular and cell surface protein that could be used for vaccination trials (Lei *et al.*, 2002) or could be involved in iron / haem-acquisition (Bates *et al.*, 2003), three streptococcal proteins were identified: the cell surface proteins Shp (Lei *et al.*, 2002) and Shr (Bates *et al.*, 2003), and the lipoprotein HtsA (Lei *et al.*, 2003). They are part of a haem-specific ABC transport system, named HtsABC (haem transporter of group A streptococcus) (Lei *et al.*, 2002),

also known as SiaABC (streptococcal iron acquisition) (Bates *et al.*, 2003). The *shr*, *shp* and *htsA* genes are co-transcribed with 7 other genes (Fig. 7.1) and their expression is regulated through the abundance of iron (Lei *et al.*, 2003, Bates *et al.*, 2003). The function of the 7 other genes is yet unknown, but it is assumed that the genes named *htsB* and *htsC* are encoding for the ATP dependent permease (Lei *et al.*, 2002, 2003, Bates *et al.*, 2003). It has been argued that haem is the most important source of iron in vivo for GAS (Eichenbaum *et al.*, 1996, Francis *et al.*, 1985), and it is likely that the HtsABC system plays a considerable role in iron acquisition of GAS (Lei *et al.*, 2003).

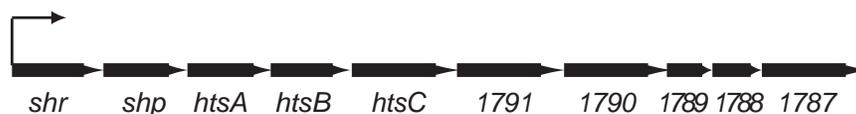


Figure 7.1: Iron / haem-acquisition locus in *S. pyogenes*. The operon is a cluster of 10 genes, and conserved across streptococci. The expression of the operon is regulated by the abundance of iron. (Bates *et al.*, 2003, Lei *et al.*, 2003).

The cell surface protein Shr is predicted to have two NEAT domains (Andrade *et al.*, 2002), like the IsdA, IsdB, IsdC and IsdH proteins from *S. aureus*, and was argued to be a receptor for haem-proteins, such as haemoglobin, myoglobin, haem-albumin and haemoglobin-haptoglobin complexes (Bates *et al.*, 2003). Shp and HtsA are the only components of the streptococcal haem acquisition machinery with a biochemical characterisation to date. It has been shown that Shp actively passes on haem from haemoglobin to HtsA (Liu & Lei, 2005). HtsA has a 5-fold higher affinity for haem than Shp, which results in a direct, affinity-driven and rapid transfer of haem from Shp to HtsA. (Nygaard *et al.*, 2006a). The proposed model for haem-uptake in *S. pyogenes* is shown in Figure 7.2. HtsA is part of the ABC transporter which then transfers the haem across the membrane (Liu & Lei, 2005, Nygaard *et al.*, 2006a).

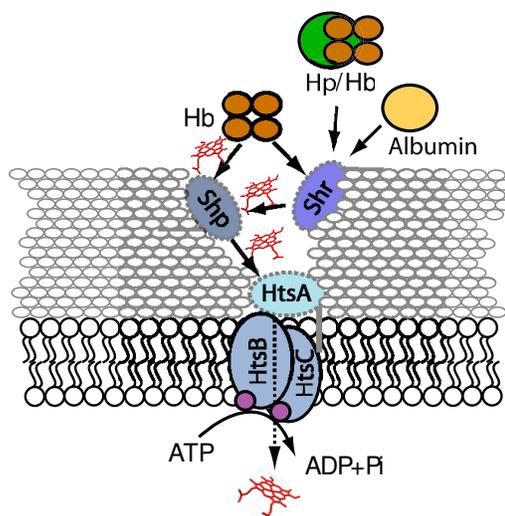


Figure 7.2: Model of haem transport in *S. pyogenes*. Shr was shown to be able to bind host haem-proteins, such as haemoglobin, haemoglobin-haptoglobin, haem-albumin and myoglobin (Bates *et al.*, 2003). Shp extracts haem from haemoglobin and passes it on to HtsA. The putative permeases HtsB and HtsC transfer the haem into the cytosol (Liu & Lei, 2005, Nygaard *et al.*, 2006a).

Structural data on these two proteins, as well as their interaction, would give precious insights into the molecular mechanism of haem transfer. Given the transient nature of the interaction it is likely to be difficult to capture a HtsA-Shp complex in a crystal lattice. Surprisingly the haem-complexes of both proteins interact with each other, forming a longer-lived complex relative to the transfer reaction (Liu & Lei, 2005).

Here HtsA and Shp from *S. pyogenes* were cloned, expressed, purified and high-throughput crystallisation screens carried out. Crystals for HtsA could be obtained and diffracted to 2.8 Å. Moreover the interaction between Shp and HtsA for possible co-crystallisation experiments was investigated.

7.2 Materials and methods

7.2.1 Cloning of *htsA* and *shp* from genomic DNA

Streptococcus pyogenes htsA and *shp* were PCR amplified from genomic DNA (strain MGAS 1718, prepared by Dr. J. Cooney). The sequence coding for residues 21 - 294 of HtsA was amplified without the N-terminal secretion signal sequence using the primer sequences as previously described (Lei *et al.*, 2003), which were modified to insert the product into pSGAT2 (Peränen *et al.*, 1996)(for plasmid map and list of primer see Appendix A.2, p. xix and A.3, p. xx) using HindIII and BamHI in frame with the cleavable His₆-tag (SP-HtsA21-pSGAT2). *HtsA* was also cloned using another forward primer starting at the sequence coding for residue 31 (SP-HtsA31-pSGAT2).

The sequence coding for Shp residues 30 - 258, therefore omitting the presumed N-terminal secretion signal sequence, the C-terminal transmembrane domain and charged tail (Lei *et al.*, 2002), was PCR amplified from genomic DNA (strain MGAS 1718 see above). The PCR product was cloned using the restriction endonucleases NdeI and XhoI into pET21a (Novagen) in frame with the C-terminal His₆-tag (SP-Shp-pET21a). The vector pET21d had been previously used by Lei *et al.* (2002) for expression of tag-free Shp. PCR, ligation, transformation and sequence analysis were carried out as described in Chapter 2, p. 2 and for the primer sequences see Appendix A.2, p. xix.

7.2.2 Expression and purification of HtsA and Shp

HtsA and Shp were expressed and purified as described in Chapter 2.2.4, p. 40. Briefly, BL21 cells containing the expression constructs were grown until an OD₆₀₀ of 0.8 - 1.0 was reached, the temperature dropped to 30°C and protein expression induced by adding IPTG to a final concentration of 0.4 mM. The cells were further grown over night, and lysed using a French press. HtsA

and Shp were purified using the His-tag and Ni-NTA resin with an imidazole gradient (10-30 mM imidazole). An additional wash-step with 1 M NaCl and 30 mM imidazole in 50 mM sodium phosphate buffer, pH 8.0 was carried out to reduce unspecific binding to HtsA and Shp. After purification the buffer was exchanged to 50 mM Tris-HCl, pH 8 200 mM NaCl using a PD10 column (GE Healthcare).

7.2.3 Reconstitution of HtsA and Shp with haemin and UV/vis spectroscopy

To reconstitute HtsA and Shp with their ligand haem, the samples were saturated with two fold excess haem and unbound haem was removed by gel-filtration (see Chapter 2.2.5, p. 41). The protein was then concentrated with centrifugal membrane-devices (Vivaspin) in 50 mM Tris-HCl pH 8, 200 mM NaCl, to the desired concentration and stored at 4°C. Spectroscopic measurements were carried out in 50 mM Tris-HCl, pH 7.5, 200 mM NaCl.

7.2.4 Dynamic light scattering

Dynamic light scattering (DLS) experiments were conducted at room temperature in a 1 ml quartz cuvette (path length 1 cm) containing protein at concentrations of 0.5 and 1 mg ml⁻¹ in PBS buffer. Prior to the DLS studies, protein samples were centrifuged at 12,000 rpm for 30 min at 4°C. Data were collected with a Viscotek model 802 DLS instrument. Measurements were programmed using the software OmniSIZE 2.0 (Viscotek Europe Ltd.) such that each experiment was averaged over 30 runs, each for 3 sec. The results were processed with the OmniSIZE software.

7.2.5 Crystallisation of haem-HtsA and haem-Shp

The most appropriate protein concentration for the haem-HtsA and haem-Shp complexes to use for crystallisation trials was determined using the pre-crystallisation test (Hampton Research). Initial high-throughput crystallisation screening was carried out using the sitting-drop vapour-diffusion method in 96-well CrystalQuick plates (Greiner). The experiments were set up with the Hydra II micro dispensing system, testing conditions from 10 crystallisation screens from Nextal Biotechnology/Qiagen (ClassicsTM, PEGsTM, PEGs IITM, AmSO4TM, CationsTM, MPDTM, pH clearTM, AnionsTM, PACTTM, MbclassTM) and a number of screens from Molecular Dimensions (Clear Strategy ITM, Clear Strategy IITM, PACTTM). For the haem-HtsA a drop size of $0.8 \mu\text{l} + 0.8 \mu\text{l}$ was used, testing the effect of protein concentration (60 mg ml^{-1} ; 100 mg ml^{-1} ; 120 mg ml^{-1}), buffer (50 mM Tris-HCl, pH 7.5; 50 mM Hepes, pH 8), temperature (20°C , 4°C) and NaCl concentration (0, 0.1 M, 0.3 M, 0.5 M, 1 M). Screens for haem-Shp were set up using either 20 or 28 mg ml^{-1} protein in either 50 mM Tris-HCl, pH 7.5, 200 mM NaCl, or 20 mM Bis-Tris-Propane, pH 6.5 or milliQ water. Plates were incubated at 20°C and examined after 2 - 3 weeks.

7.2.6 Structure determination of HtsA

Data collection was carried out at the ESRF synchrotron radiation facility beamline ID14-2. Two sweeps of data were collected on a single crystal, both consisting of 250 images collected with 1° oscillation, one with an exposure of 2 sec (low resolution), and the second with an exposure time of 60 sec (high resolution). The data were processed (low resolution data set: 30-4 Å; high resolution data set: 8.0-2.8 Å) using the programs XDS (Kabsch, 1988, 1993), MOSFLM / SCALA (Leslie, 1992, Evans, 1997) and XPREP (version 6.14 from

BRUKER) and were tested for twinning using the programs SFALL (CCP4, 1994) and XTRIMAGE (Zwart *et al.*, 2005). Molecular replacement trials were carried out using the structures of the *enterochelin uptake PBP* (CeuE) from *Champylobacter jeuni* (PDB code 2CHU), the *E. coli iron-hydroxamate binding protein* (FhuD; PDB code 1ESZ) and the *Vitamin B12 binding protein* (BtuF) (PDB code 1N2Z) as an ensembled search model in PHASER (McCoy *et al.*, 2005). All residues not shared between these structures and HtsA were truncated at the C γ -atoms and the B-factors were reset to 10 and 20 for the main- and side chains, respectively. Furthermore, alternative space groups were tested. While writing this Chapter, the structure of the *S. aureus* IsdE protein in complex with haem was determined (PDB code 2Q8Q) (Grigg *et al.*, 2007b), sharing 40% sequence identity with HtsA. IsdE was prepared as a MR model as described above.

Iron-MAD data were collected at the beamline ID23-1 at the ESRF synchrotron radiation facility. The peak and inflection points were determined by an X-ray fluorescence scan around the iron-edge. A MAD data set consisting of peak (1.7399 Å, 250°, 1° oscillation, 2 sec exposure), inflection (1.7120 Å, 240°, 1° oscillation, 2 sec exposure) and remote (0.954 Å, 180°, 1° oscillation, 1 sec exposure) were collected using different positions of a translated crystal.

7.3 Results and Discussion

7.3.1 Sequence alignment and structure predictions

The NCBI data base was searched using BLAST (Altschul *et al.*, 1990) for proteins homologous to HtsA and Shp. Homologues for Shp could only be identified from other GAS strains, sharing more than 90 % sequence identity, as well as from *Streptococcus equi*. Shp from *S. pyogenes* has a sequence identity of 75 % and a similarity of 91 % with the protein from *S. equi* (Nygaard

et al., 2006b), and appears to be unique to streptococci. In contrast, HtsA shows homology to a number of proteins from other species; for instance, it shares 40% sequence identity with the lipoprotein IsdE from *S. aureus*, which is part of the Isd haem-uptake machinery (see Chapter 1.2, p. 1.2). Three residues that could be possibly ligating the haem-iron, Met 79, His 229 and Met 238, are conserved in HtsA homologues (Fig. 7.3). Electron paramagnetic resonance (EPR) measurements carried out on haem-HtsA indicates that the iron is hexacoordinate and two histidine residues were argued to be possible ligands, because of the similarity to the EPR spectra of b-type cytochromes (Nygaard *et al.*, 2006a). HtsA also has low sequence identity to the periplasmic

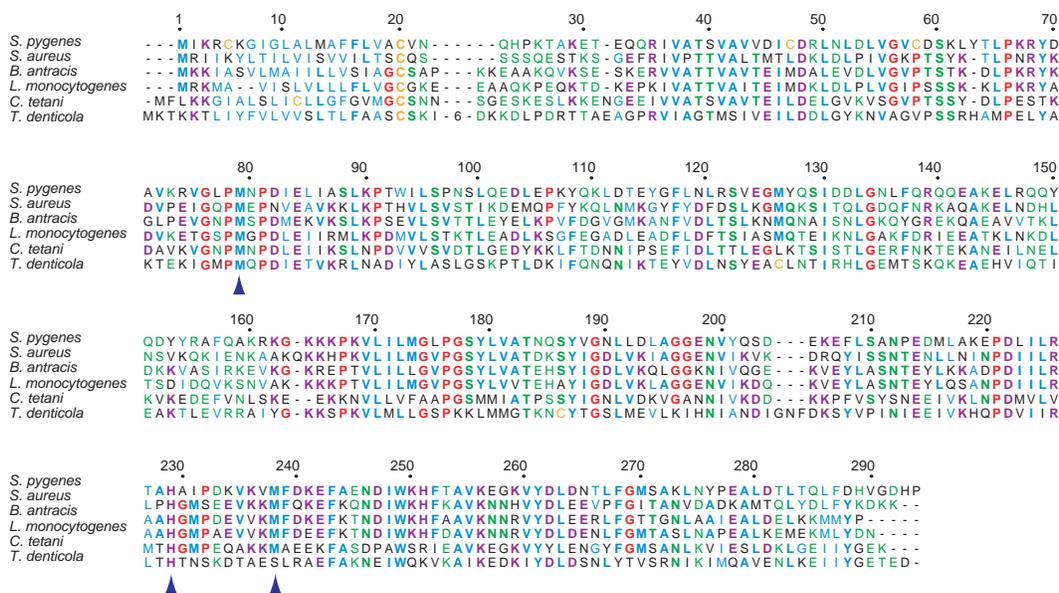


Figure 7.3: Sequence similarities and identities of HtsA homologues.

Multiple sequence alignment of 6 HtsA homologous sequences from representative Gram-positive species. Sequences homologous to HtsA from *S. pyogenes* were identified by BLAST searches (Altschul *et al.*, 1990) of the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/>) and alignments were carried out with ClustalW (Thompson *et al.*, 1994). Sequences aligned are from the following species: *Bacillus anthracis* (gi:47552030); *Clostridium tetani* (gi:28210884); *Listeria monocytogenes* (gi:16411654); *S. aureus* (gi:13700933); *Treponema denticola* (gi:42527857). The alignment was annotated with a program kindly provided by Dr. T. Stevens (Stevens & Paoli, 2007). Numbering on top of the alignment is according to the HtsA sequence.

binding proteins (PBP) HemT from *Y. enterocolitica*, ShuT from *S. dysenteriae* and PhuT from *P. aeruginosa*. In addition it shows similarity to other PBP, such as the *enterochelin uptake PBP* (CeuE) from *Champylobacter jeuni* (PDB code 2CHU), the *E. coli iron-hydroxamate binding protein* (FhuD; PDB code 1ESZ) and the *Vitamin B12 binding protein* (BtuF) (PDB code 1N2Z), of which the structures have been determined. The sequence alignment in Figure 7.4 shows that the conserved tyrosine residue coordinating the haem iron in PhuT (Tong & Guo, 2007), ShuT (Eakanunkul *et al.*, 2005) and most likely in HemT (discussed in Chapter 6) has no equivalent in IsdE and HtsA; the methionine residue conserved in the HtsA homologues is located two residues downstream from the tyrosine (Fig. 7.4). This is supporting the hypothesis

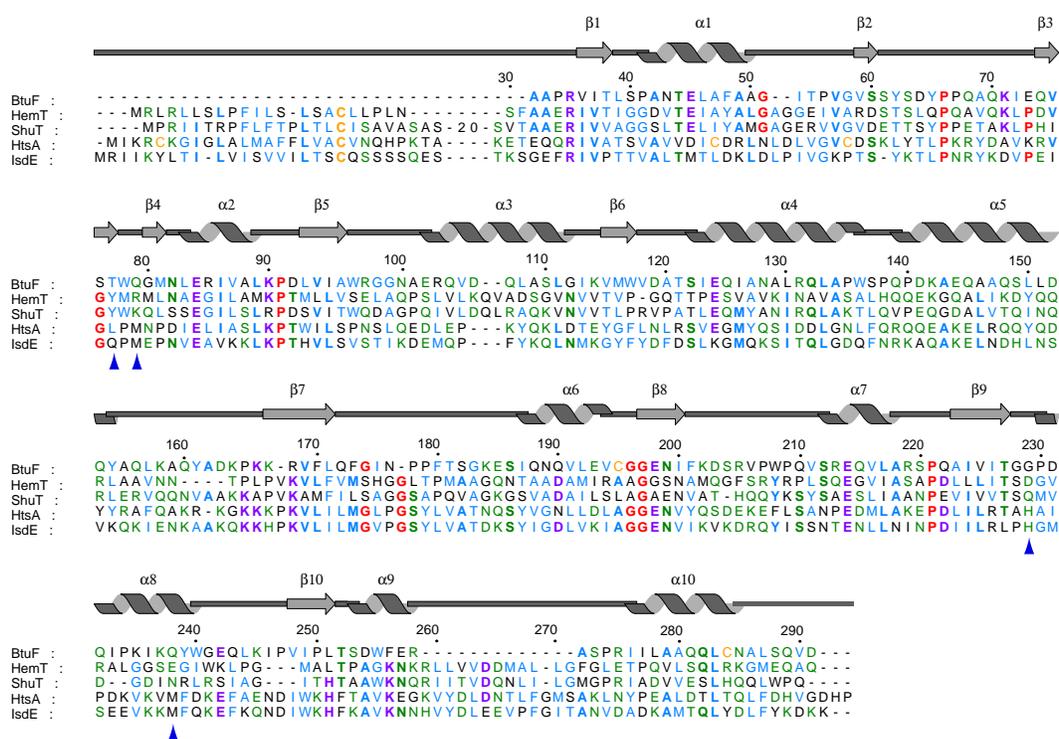


Figure 7.4: Sequence alignment of HtsA with BtuF. Annotation of the alignment was carried out by a program kindly provided by Dr. T. Stevens (Stevens & Paoli, 2007) and the numbering on top of the alignment corresponds to HtsA. Secondary structure was assigned according to the structure of BtuF. Highlighted by an arrow are residues involved or possibly involved in haem-iron coordination; the tyrosine in ShuT and HemT, and the two methionines and the histidine.

that this methionine is involved in haem-iron coordination in HtsA and IsdE. While writing this Chapter the structure of the haem-IsdE complex from *S. aureus* was published, showing a 6-coordinated haem-iron with Met 79 and His 229 as ligands, which are conserved across species (Grigg *et al.* (2007b) and Fig. 7.3). Homology modelling studies of HtsA using IsdE as template also points to this histidine and methionine as the axial ligands (Grigg *et al.*, 2007b).

7.3.2 Purification and biochemical characterisation of HtsA

7.3.2.1 Purification of HtsA and Shp

S. pyogenes *htsA* and *shp* were cloned from genomic DNA. For *htsA* two alternative constructs were generated; one coding for HtsA starting at residue 21 as previously described (Lei *et al.*, 2003) and the second starting with residue 31, where the structure of *E. coli* BtuF begins (Fig. 7.4). Shp was cloned in frame with a non-cleavable C-terminal His₆-tag and all constructs showed very high expression levels of $\sim 150 \text{ mg ml}^{-1}$ per litre cell culture. The proteins were purified using Ni-affinity chromatography, the His-tag was removed with thrombin from HtsA21 and HtsA31 (Fig. 7.5) and Shp and HtsA were reconstituted with the haem-ligand.

Analytical gel-filtration analysis was carried out on the haem-Shp and haem-HtsA complex as well as apo-HtsA and they appeared to be monomeric in solution (Fig. 7.6).

The haem-HtsA complex was also analysed by dynamic light scattering, confirming the monomeric state of the haem-HtsA complex in solution (Fig. 7.7). The observed radius in the two experiment of 3.0 and 3.1 nm (standard deviation +/- 1 nm), which is equivalent to 30 and 31 Å, respectively. This is

similar to the dimensions of the *E. coli* BtuF protein, which has a longitudinal shape with approximate diameters of 31 and 47 Å.

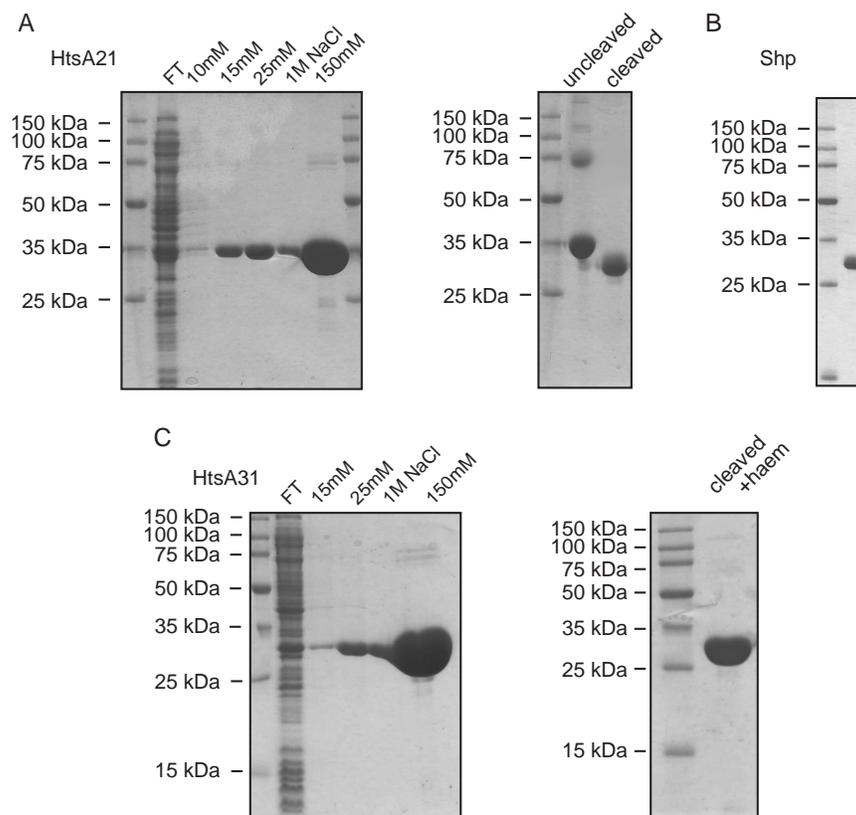


Figure 7.5: Purification of HtsA. (A) His-tag purification of HtsA21 with an additional high-salt wash step (30 mM imidazole, 1 M NaCl). His-HtsA21 after cleavage with thrombin, reconstitution with haemin and gel-filtration. (B) Pure Shp protein after His-tag purification, reconstitution with haemin and gel-filtration. (C) His-tag purification of HtsA31 and pure HtsA31 protein after cleavage with thrombin, reconstitution with haemin and gel filtration.

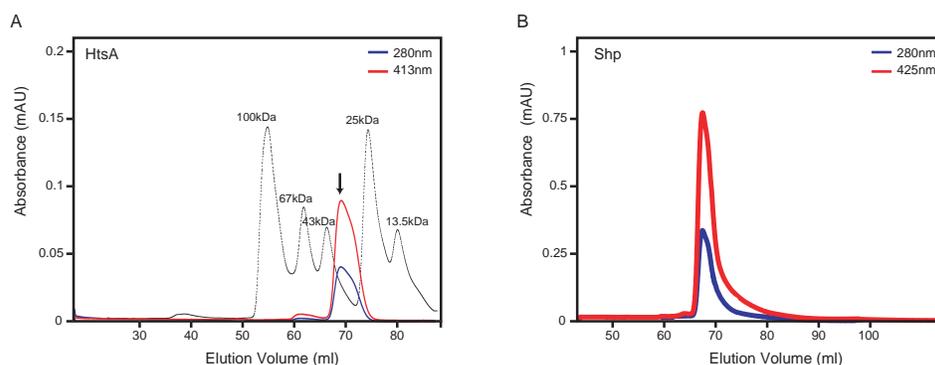


Figure 7.6: Analytical gel filtration analysis of HtsA. An estimated molecular mass of the protein was extrapolated from a standard calibration curve using an GE Healthcare chromatography system (Äkta Purifier) equipped with a HiLoad Superdex 75 HR16/60 column. The column was equilibrated with 50 mM Tris-HCl, pH 7.5, 75 mM NaCl, and was tested with a sample concentration of 20 mg ml^{-1} . (A) Representative traces were recorded at 280 nm of molecular weight markers (black) and HtsA (blue) and trace HtsA measured at Soret maximum of 413 nm (red). The data indicates that HtsA is a monomer and no dimerization is observed. (B) Haem-Shp is a monomer and its elution was followed at 280 nm (blue) and its Soret maximum 425 nm (red).

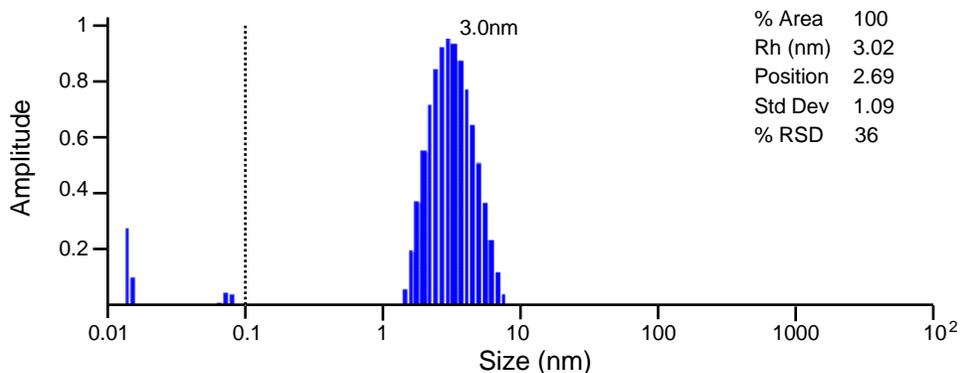


Figure 7.7: Dynamic light scattering analysis of haem-HtsA. Two different protein concentrations (1 mg ml^{-1} and 0.5 mg ml^{-1}) were analysed indicating that the haem-HtsA complex is mono dispersed in solution with an estimated diameter of 30 Å. Data shown here were obtained with a haem-HtsA concentration of 1 mg ml^{-1} .

7.3.2.2 UV/vis spectroscopy of HtsA and Shp

After reconstitution of HtsA and Shp with haem, the UV/vis absorption spectra were recorded. Fully reconstituted HtsA and Shp have similar UV/vis absorption spectra. Ferric-HtsA has a Soret maxima of 413 nm and Q bands at 534 nm and 567 nm. Upon reduction the Soret peak is red-shifted to 424 nm, with Q-bands at 529 nm and 559 nm (Fig. 7.8A). The haem-Shp complex in its oxidised form exhibits absorbance peaks at 415 nm, 532 nm and 560 nm and in its reduced form at 425 nm, 529 nm, 559 nm (Fig. 7.8B). The absorption spectra of the haem-HtsA and haem-Shp complexes both show characteristics of a 6-coordinated haem-iron. This was previously suggested by EPR measurements, which indicated two ligands to the iron (Nygaard *et al.*, 2006b). Cyanide can bind to the ferric haem-HtsA and haem-Shp complexes (Fig. 7.8). Therefore the haem has to be accessible for cyanide, which then replaces the weaker ligand to the haem iron.

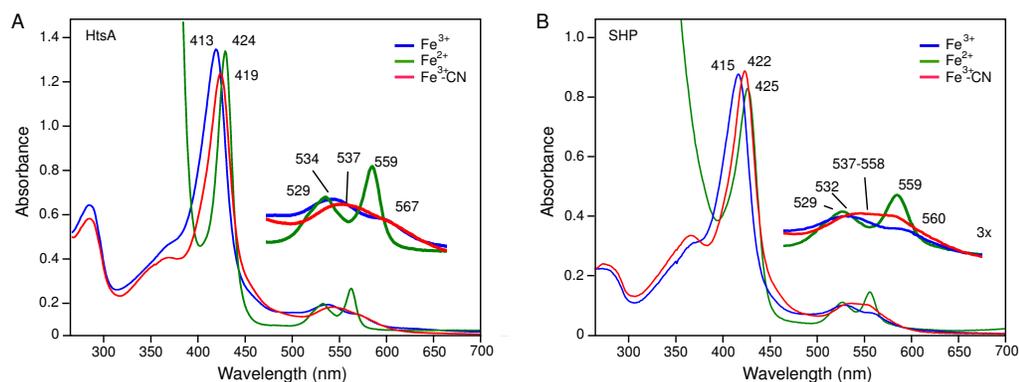


Figure 7.8: UV/vis spectroscopy of HtsA and Shp. Spectra of ferric (blue), ferrous (green) and ferric cyanide complex (red) of (A) HtsA and (B) SHP-His.

The absorption maxima of haem-HtsA and haem-Shp observed here, partly differ from the ones previously described in the literature (Lei *et al.*, 2003, Liu & Lei, 2005, Nygaard *et al.*, 2006a)(see Table 7.1). It was argued that the haem-Shp complex does not undergo auto oxidation after reduction, but

is stable in its reduced, ferrous form for weeks in Tris-buffer at pH 8 at 4°C (Nygaard *et al.*, 2006a). Here a prolonged stability of the reduced complex was noticed, but autoxidation occurred after a few days in Tris-HCl, pH 7.5. The differences in the spectroscopic characteristics compared with the ones reported in the literature could be due to slightly different buffer conditions and the absence (HtsA) or presence (Shp) of the His-tag. Deviations in the Soret peak of haem-HtsA-His can be found in the literature (Lei *et al.* (2003), Liu & Lei (2005), Nygaard *et al.* (2006a); Table 7.1), but have not been discussed. No differences in the absorption spectra of haem-HtsA beginning with residue 21 or 31 were observed.

Table 7.1: UV/vis absorption spectra of HtsA and Shp-His

this work	HtsA	SHP-His₆
ferric (Fe³⁺)	413 nm, 534 nm, 567 nm	415 nm, 532 nm, 560 nm
ferrous (Fe²⁺)	424 nm, 529 nm, 559 nm	425 nm, 529 nm, 559 nm
ferric-cyanid	419 nm, 537 nm	422 nm, 532-558 nm
Nygaard <i>et al.</i> (2006a)	HtsA-His₆	Shp
Liu & Lei (2005)		
ferric (Fe³⁺)	412 nm, 530 nm, 570 nm	420 nm, 530 nm (shoulder)
ferrous (Fe²⁺)	424 nm, 528 nm, 558 nm	428 nm, 528 nm, 560 nm
Lei <i>et al.</i> (2003)	HtsA-His₆	
ferric (Fe³⁺)	408 nm	-
as purified	417 nm 530 nm 460 nm	-

7.3.3 Crystallisation of Shp and HtsA

7.3.3.1 Crystallisation trials of the haem-Shp complex

High-throughput crystal screening of 960 conditions was carried out with haem-Shp at a concentration of 28 mg ml⁻¹ in three different solutions: i) 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, ii) 20 mM Bis-Tris-Propane, pH 6.5, and iii) milliQ water.

The trays were incubated at 20°C and inspected every one-two weeks. Crystalline precipitation, aggregation and strong phase separation could be observed (Fig. 7.9), but no crystals formed. The formation of strong phase separation and crystalline precipitate is promising. Further screening using higher protein concentrations, different temperature (4°C) and buffer conditions of the protein solution should be explored, but was not carried out due to the time limitations of this thesis work.



Figure 7.9: High-throughput crystallisation screen of haem-Shp. Commonly observed in various conditions were (A) crystalline precipitation, (B) strong phase separation and (C) aggregation.

7.3.3.2 Crystallisation of the haem-HtsA complex

Initial high-throughput crystallisation screening of 1344 conditions with various protein concentrations, buffers, sodium chloride concentrations and at different temperatures was carried out with HtsA starting at residue 21 and 31, (for details see Chapter 7.2.5, p. 138). Quite commonly crystalline aggregation and variations of strong phase separations could be observed (Fig. 7.10A+B). Figure 7.10C shows the haem-HtsA21 crystal obtained in the PEGsTM (Qia-gen / Nextal) in 0.1 M HEPES, pH 7.5, 25 % (w/v) PEG 4000, at a protein concentration of 100 mg ml⁻¹ in 50 mM Tris-HCl, pH 8 and 4°C. No crystals were obtained in the same condition at 4°C with protein solution supplemented with sodium chloride or at 20°C. HtsA crystals could be reproduced in hanging as well as sitting drop grid-screens around the initial hit condition

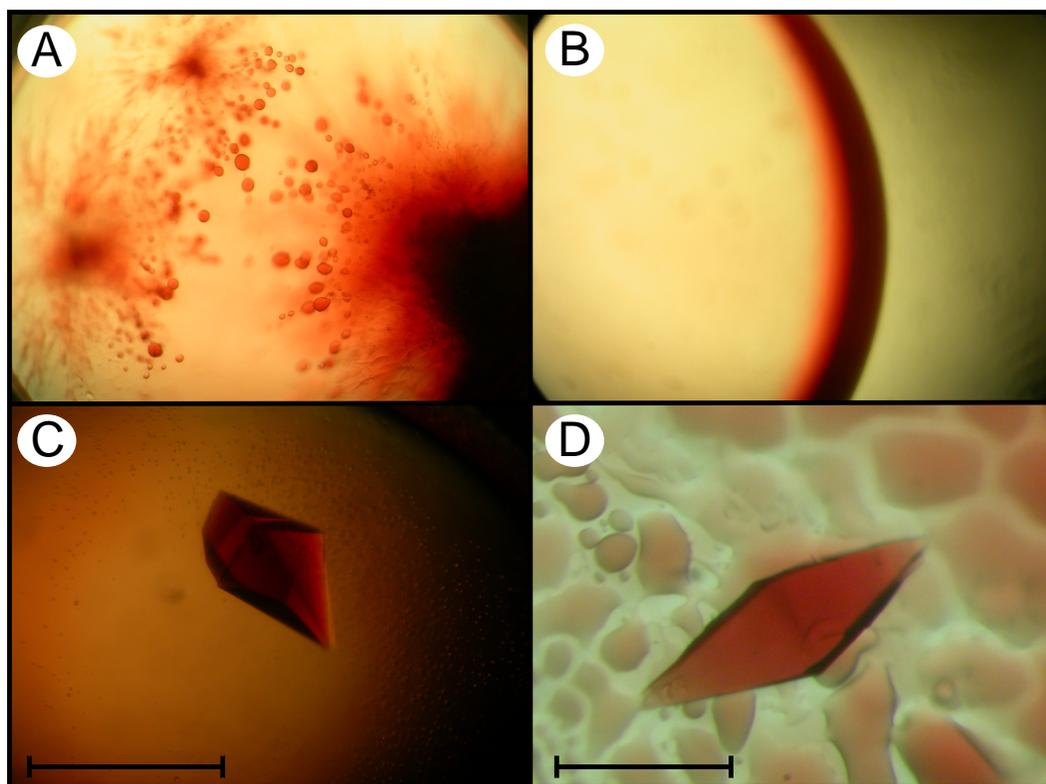


Figure 7.10: Crystallisation of HtsA. (A+B) Commonly observed strong phase separation in initial high throughput crystallisation trials. (C) HtsA crystal obtained in 0.1 M Hepes, pH 7.5, 25 % (w/v) PEG 4000, high-throughput screen PEGsTM (Qiagen), at a protein concentration of 100 mg ml⁻¹ in 50 mM Tris-HCl, pH 8.0 and 4°C. (D) Reproduced HtsA crystal in hanging-drop grid screen around the initial hit condition in 150 mM Hepes, pH 7.5, 26% PEG 4000, at a protein concentration of 125 mg ml⁻¹ and a drop size of 2 + 2 μ l. The bar indicates 100 μ m.

at 4°C (Fig. 7.10D). Crystals usually grew after 2-3 weeks and the sites of nucleation appeared to be in the zone of strong phase separation.

The crystals were very fragile and the majority of the crystals obtained in the sitting drop screens were attached to the plastic surface of the tray. Therefore handling of the crystals and transferring them to the cryo protectant was an issue. Moreover most of the crystals showed imperfections and deformations, indicating the presence of multiple lattices.

Partial cleavage of the haem-HtsA21 was observed after a few days of storage at 4°C. (Fig. 7.11, lane 1). This double band with the same ratio of un-

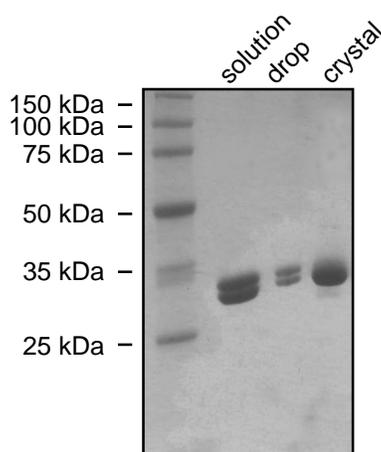


Figure 7.11: Partial cleavage of HtsA over time. SDS-PAGE gel showing partial cleavage of HtsA21 after two days storage at 4°C, a lower band appears in the protein sample (lane 1). These two bands are also present in the strong phase-separation from a drop where haem-HtsA21 crystals have grown (lane 2). Only the lower band, representing the cleaved protein can be observed in haem-HtsA crystals (lane 3).

cleaved to cleaved protein could also be seen in the strong phase-separation zone in drops where crystals had formed (Fig. 7.11, lane 2); but for crystal formation, only the cleaved protein was selected (Fig. 7.11, lane 3). It might therefore be possible to improve the packing of the crystal lattice by setting up crystallisation trials with only the cleaved, lower molecular weight protein. In order to determine the residue where the lower molecular weight protein in the crystal started, the lower band was cut out from an SDS-PAGE gel and sent for N-terminal sequencing (Altabiosciences, Birmingham). Results indicated that the protein in the crystals was truncated after Lys 25. Incorporation of uncleaved HtsA21 protein in the crystal lattice could account for the observed deformation and anisotropic diffraction of the crystals. Hence, it might be possible to improve the quality of the crystals by carrying out crystallisation trials only with the truncated HtsA protein. The previously prepared expression construct, coding for HtsA starting at residue 31 could be a suitable substitution for the truncated HtsA selected in the crystal lattice, since

cleavage of the His-tag with thrombin after expression in the pGAT2-plasmid, two additional amino acids (Gly, Ser) are left at the N-terminus, and haem-HtsA31 had not been tested in high-throughput crystallisation screens at 4°C. HtsA31 was purified and tested for improvement of crystal formation in a grid screen around the hit condition (Hepes, pH 7.5: 50-200 mM; PEG 4000: 22-30% (w/v)) with concentrations ranging from 100-160 mg ml⁻¹ at 4°C. Only strong phase separation but no crystal formation could be observed. Further screening as well as seeding using crystal fragments from HtsA21 might aid crystal formation from HtsA31, but was not tried due to time limitations of this thesis work.

7.3.4 Diffraction analysis of HtsA

The haem-HtsA crystal obtained in the initial high-throughput screen was diffracted at the European Synchrotron Radiation Facility (ESRF) beamline ID14-2. Diffraction was limited in resolution, and highly anisotropic. By reducing the beam size and screening different parts of the crystal, a complete data set to 2.8 Å (Fig. 7.12) could be obtained.

The manually reproduced haem-HtsA crystals did not show any improvement of resolution or diffraction quality at the home-source or the ESRF-beamline ID23-1. The data were processed with different programs and in different, possible space groups and the systematic absences were analysed.

Initially it appeared that the crystals belong to the tetragonal enantiomorphic space group P4₁2₁2/P4₃2₁2, but the data could be processed with a slightly better R_{merge} in P2₁2₁2₁ (Table 7.2). According to solvent content and self-rotation function analysis, there are most likely two molecules in the asymmetric unit cell.

HtsA shares low sequence identity (≤ 23%) with the periplasmic binding proteins (PBP) from Gram negative species, the structures of which have been de-

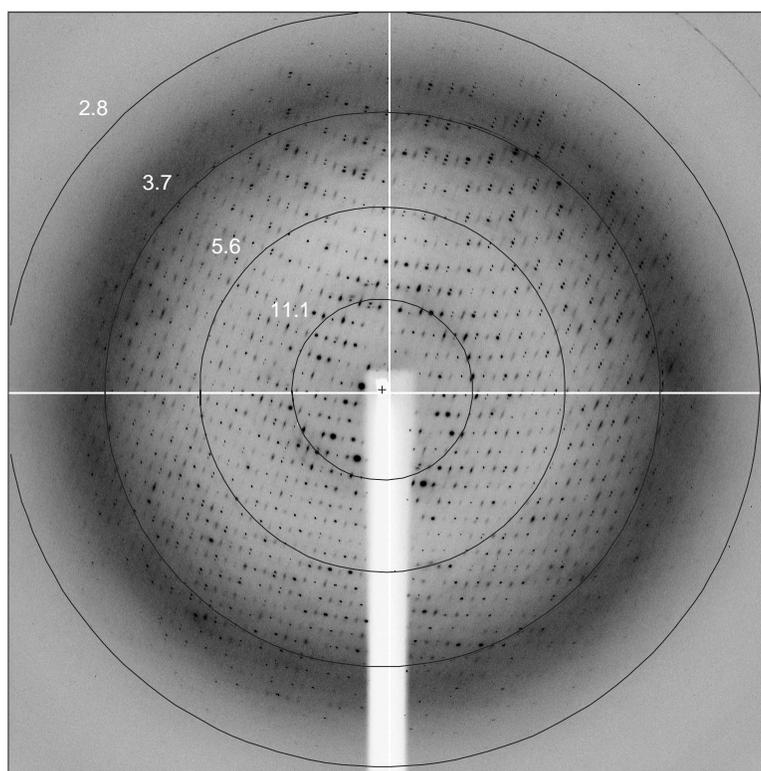


Figure 7.12: Diffraction of haem-HtsA crystal. Diffraction of haem-HtsA crystal from initial hit condition. Crystal was transferred to an artificial mother liquor of 100 mM Hepes, pH 7.5, 50 mM Tris-HCl, pH 8.0, 35 % (w/v) PEG 4000, prior to flash freezing in liquid nitrogen.

terminated; the *enterochelin uptake PBP* (CeuE; PDB code 2CHU) from *Champylobacter jeuni*, the *iron-hydroxamate binding protein* (FhuD; PDB code 1ESZ) and the *Vitamin B12 binding protein* (BtuF; PDB code 1N2Z) from *E. coli*. They superimpose onto each other with RMSDs ranging from 2.8-3.1 Å and Figure 7.13 shows a secondary structure based alignment of these molecules. Molecular replacement using all three proteins as an ensemble search model was carried out. Two weak solutions in the tetragonal space groups $P4_32_12$ (rotation function *Z*-score (RFZ) = 2.6 / 2.7; translation function *Z*-score (TFZ) = 6.0 / 5.9) and $P4_322$ (RFZ = 2.6 / 3.0; TFZ = 5.5 / 5.2) were found, with the solutions differing in the alternate arrangements of the two molecules in the asymmetric unit along the screw axis. But in both solutions the *log likeli-*

Table 7.2: Data collection and processing statistics. Values in parentheses are for the highest-resolution shell.

	haem-HtsA					
Data Collection						
Space group	P4 ₃ 2 ₁ 2			P2 ₁ 2 ₁ 2 ₁		
Wavelength (Å)	0.933					
Cell dimensions						
<i>a</i> , <i>b</i> , <i>c</i> (Å)	71.34	71.34	227.96	71.27	71.45	228.21
α , β , γ (°)	90, 90, 90			90, 90, 90		
Resolution (Å)	30 - 2.85 (3.0 - 2.85)			30 - 2.8 (2.95 - 2.80)		
R _{merge}	0.074 (0.466)			0.058 (0.469)		
Mean I / σ I	28.8 (4.5)			19.4 (3.0)		
Completeness (%)	99.9 (100.0)			99.9 (100.0)		
Redundancy	14.4 (11.6)			7.3 (6.1)		
Solvent content (%)	50.8			47.5		
Molecules in ASU	2			2		

hood gain (LLG) for the second molecule was negative. In addition there were clashes between the molecules in the asymmetric unit and symmetry related molecules.

When molecular replacement was carried out with the data processed in the orthorhombic space group P222 and tested against all possible space groups, weak solutions were found in P222₁ and P2₁2₁2₁, but the LLG was negative. No solutions were found when only one of the molecules was used as a search model instead of the ensemble. Most likely this is caused by the large error in the phases due to the low sequence identity between HtsA and the search models as well as the limited resolution.

An attempt was made to aid structure determination by taking advantage of the iron in the prosthetic group. Combination of the molecular replacement phases with phases obtained from the iron could largely improve the resulting electron density maps as well as reduce model bias. An iron-MAD data set was collected at the ESRF beam line ID23-1. The iron peak and inflection point wavelengths were determined by an fluorescence scan around the iron edge (peak = 1.7399 Å, inflection = 1.7120 Å). Unfortunately only limited resolution

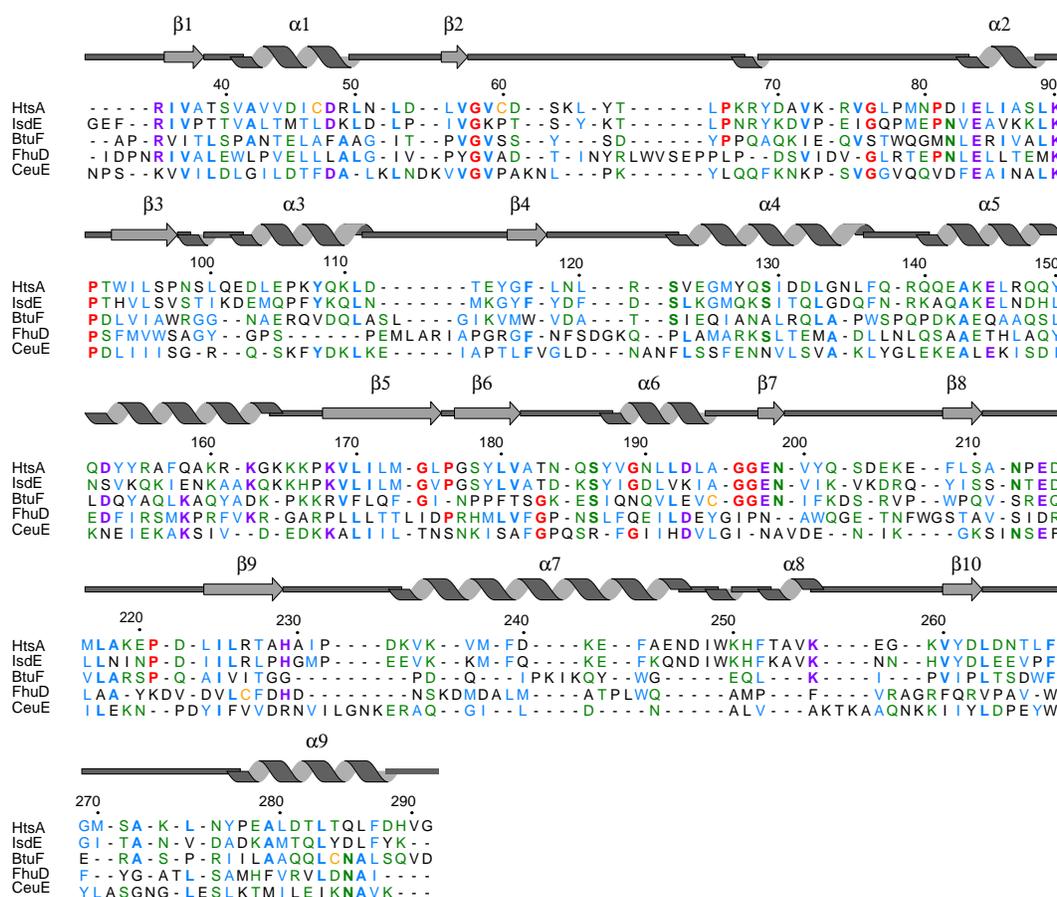


Figure 7.13: Secondary structure alignment of the molecular replacement models. An ensemble of the molecules here aligned was used for molecular replacement in PHASER (McCoy *et al.*, 2005). Secondary structure assignment corresponds to IsdE from *S. aureus* (PDB code 2Q8Q) and the numbering to HtsA. *enterochelin uptake PBP* (CeuE; PDB code 2CHU) from *Champylobacter jeuni*, the *iron-hydroxamate binding protein* (FhuD; PDB code 1ESZ) and the *Vitamin B12 binding protein* (BtuF; PDB code 1N2Z) from *E. coli*. HtsA was aligned to IsdE using ClustalW (Thompson *et al.*, 1994). The coordinates were matched according to their secondary structure using the SSM server at the European Bioinformatics Institute (EBI; <http://www.ebi.ac.uk/msd-srv/ssm/ssmstart.html>). Annotation was carried out with a program kindly provided by Dr. T. Stevens (Stevens & Paoli, 2007).

(~ 3.5 Å) and highly anisotropic diffraction, which was also depending on location of the crystal and its orientation, was obtained. The data were processed with XDS (Kabsch, 1993, 1988) and MOSFLM/SCALA (Leslie, 1992, Evans, 1997), but due to the low quality of the data the iron position could not be unambiguously determined.

The purification of SeMet derivatised HtsA was planned and attempts were made to obtain better quality crystals (see above), but while this Chapter was written, the structure of the haem-IsdE from *S. aureus* complex was published (Grigg *et al.*, 2007b), sharing 40% sequence identity with HtsA. Molecular replacement with IsdE as search model testing tetragonal space groups gave improved resulting maps ($P4_32_12$; RFZ = 7.1 / 5.3; TFZ = 21.2 / 4.6; LLG = 176 / 80), but the LLG for the second molecule was lower than for the first and there were still large errors in the calculated phases as well as clashes with symmetry related molecules. A solution resulting in better electron density maps could be obtained using the data processed in the orthorhombic space group $P2_12_12_1$ (RFZ = 4.8 / 5.2; TFZ = 10.1 / 20.9; LLG = 121 / 373), but six clashes between the two molecules occurred. Taking advantage of the two molecules in the asymmetric unit, *non-crystallographic symmetry* (NCS) averaging was carried out in DM (Cowtan, 1999), which resulted in improved phases. Further improvement of the maps was obtained by density modifications and NCS averaging using SOLOMON (Abrahams & Leslie, 1996), implemented in SHARP (de La Fortelle & Bricogne, 1997, Vonnrhein *et al.*, 2006), using the phases calculated in DM. Parts of the electron density maps were still of poor quality and only some weak density peaks for the haem could be observed (Fig. 7.14). An attempt was made to slowly improve the model by making adjustments of the positions of residues and side chains according to the NCS-averaged electron density map. Repeating the molecular replacement in PHASER with one chain of the improved model, resulted in markedly improved statistics (RFZ = 9.1 / 10.3; TFZ = 18.4 / 33.9; LLG = 348 / 1223) and the clash between the two chains was reduced to two residues. Nevertheless the *R-factors* could not be reduced below 49% in all initial refinement trials in REFMAC.

An explanation for the inability to refine the HtsA structure could be pseudo-merohedral twinning of the crystal. A merohedral twinned crystal contains two

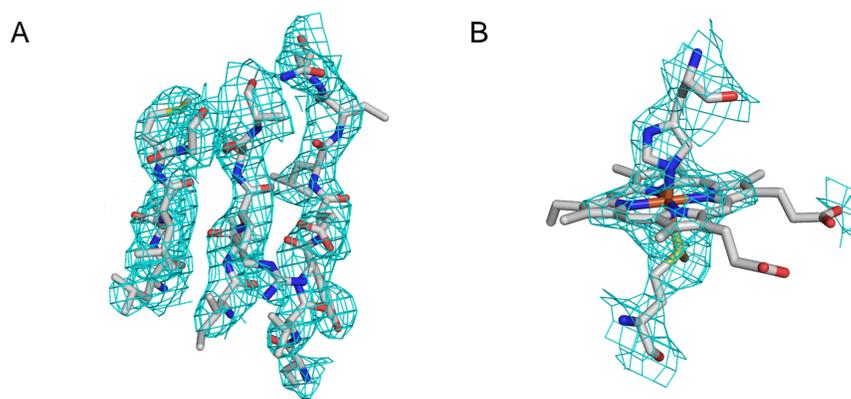


Figure 7.14: Electron density of the haem-HtsA complex after NCS averaging. Map calculated to 2.8 Å and contoured at 1 σ level. Figure showing (A) a part of the beta sheet in the core of the C-terminal domain and (B) the density peak around the haem and its ligands.

lattices of two separate twin domains, which perfectly superimpose in three dimensions. In this case the two different unit cells related by the twinning operator comply with the unit cell lattice, but the orientation of the molecules in the asymmetric unit cell is not equivalent. Usually this can sometimes be observed in higher symmetry space groups (cubic, tetragonal and hexagonal) when the space group symmetry is lower than the symmetry of the crystal lattice (Yeates, 1997). Nevertheless rare cases of merohedral twinning have been reported for lower symmetry space groups when the geometry of the unit cell is similar to a higher symmetry space group (for example Rudolph *et al.* (2004), Wittmann & Rudolph (2007), MacRae & Doudna (2007)). The native HtsA data were tested for twinning using the program XTRIAGE (Zwart *et al.*, 2005). For the data processed in $P2_12_12_1$ either a pseudo-merohedral twinning fraction of 42.3%, with the twinning operator $-k, -h, -l$, or the higher symmetry space group $P4_x2_12$ was suggested. No twinning for the data processed in tetragonal space groups could be detected.

A case of perfect pseudo-merohedral twinning of orthorhombic crystals, where the length of the a axis equal to that of b axis has been reported recently for

the protein Dicer (MacRae & Doudna, 2007). The unit cells can be orientated in two ways, which are related by a 90° rotation about the c-axis, and still fit into the crystal lattice. The structure was solved using data from crystals with reduced twinning, obtained through replacement of magnesium chloride, which was essential for crystal growth, by manganese chloride (MacRae & Doudna, 2007).

Refinement of the structure against twinned data is possible, as for example implemented in the PHENIX software package (Adams *et al.*, 2002), but special care must be taken not to introduce model bias. In the case of the Dicer structure refinement using the twinned data, which were of higher resolution than the data obtained from un-twinned crystals, resulted in apparently better electron density maps and reasonable statistics, but the maps were less reliable due to introduced model bias (MacRae & Doudna, 2007).

Given the large errors in the phases in combination with the limited resolution and the presumable pseudo-merohedral twinning of the haem-HtsA crystals, it is also likely that a better model for haem-HtsA cannot be obtained. Better quality, un-twinned crystals as well as collection of higher quality iron-MAD data would help tremendously to overcome these problems.

7.3.5 Complex-formation of Shp and HtsA

It has been shown recently that Shp transfers the haem to apo-HtsA by forming a transient protein-protein complex (Liu & Lei, 2005, Nygaard *et al.*, 2006a). A more stable interaction had been observed between the haem complexes of HtsA and Shp (Liu & Lei, 2005). In order to investigate the possibility to obtain a stable interaction for co-crystallisation trials, haem-HtsA and haem-Shp were mixed in a one to one molar ratio and analytical gel-filtration was carried out. Partly a stable haem-HtsA-haem-Shp complex was observed (Fig. 7.15), but the conditions would need to be optimised to shift the equilibrium more

towards the formation of the complex before crystallisation trials could be carried out.

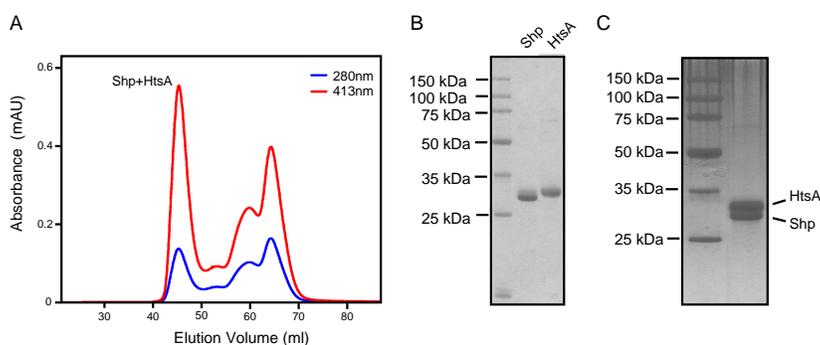


Figure 7.15: Analytical gel filtration of haem-HtsA and haem-Shp. Haem-Shp and haem-HtsA were mixed in a one to one molar ratio in 20 mM sodium phosphate buffer pH 7.5, 20 mM NaCl with a final concentration of $\sim 10 \text{ mg ml}^{-1}$ and loaded onto a Superdex 75 column (GE Healthcare). (A) Traces were recorded at 280 (blue) and the Soret maxima of 413 nm (red). Partly a haem-Shp-haem-HtsA complex was formed and eluted from the column at $\sim 45 \text{ ml}$. The protein samples were analysed using SDS-PAGE. (B) before mixing together as well as the (C) fraction containing the potential complex after gel filtration.

7.4 Conclusions

The major iron source for *Streptococcus pyogenes* during infection are host-haem proteins (Eichenbaum *et al.*, 1996, Francis *et al.*, 1985) and only recently some components of its haem-uptake machinery have been identified and biochemically characterised; the cell surface proteins Shr and Shp, as well as the lipoprotein HtsA (Lei *et al.*, 2002, 2003, Bates *et al.*, 2003, Liu & Lei, 2005, Nygaard *et al.*, 2006a), but no structural data on these proteins or their homologues were available when this work was started.

Here Shp and HtsA were cloned and purified and crystallisation trials carried out. Search of the the NCBI database for homologues of HtsA and multiple sequence alignments revealed three conserved residues that could be involved in haem-iron coordination; Met 79, His 229 and the almost conserved Met 238.

HtsA has low sequence identity to ShuT, PhuT and HemT from Gram-negative bacteria. The tyrosine residue which is coordinating the iron in these proteins (Eakanunkul *et al.*, 2005, Tong & Guo, 2007) is located two residues away from Met 79 and therefore supporting the hypothesis that this residue is ligand to the haem-iron. While writing of this Chapter the structure of the haem-IsdE complex from *S. aureus* was published, which shares 40% sequence identity to HtsA and is coordinating the iron by the conserved Met 79 and His 229 (Grigg *et al.*, 2007b). In contrast Shp appears to be unique to streptococci and homologues share a sequence identity of greater than 90%.

Analysis of the UV/vis absorption spectra of the haem-Shp and haem-HtsA showed that the haem is accessible to the binding of cyanide and to displace one of the two ligands to the haem-iron. Previous EPR measurements indicated 6-coordinated haem-iron in both proteins (Nygaard *et al.*, 2006a) which for HtsA was confirmed by the structure determination of IsdE.

No crystals for Shp could be obtained, but promising crystalline precipitation as well as strong phase separation were identified. Higher protein concentration and incubation at different temperature (i.e. 4°C) might give better results, but were not tested due to the time limitation of this work. Diffraction quality crystals for HtsA were obtained and could be reproduced. HtsA crystals diffracted X-rays to 2.8 Å using a synchrotron radiation source and most likely belong to the orthorhombic space group P2₁2₁2₁ with a pseudo-merohedral twinning fraction of 42.3%. HtsA has low sequence identity (~22%) to the PBP BtuF, FhuD and CeuE, the structures of which had been determined. Using these structures as an ensemble for molecular replacement, a solution could be found, but resulting maps were of poor quality given the low sequence identity and the resulting large error in phase.

Using the coordinates of the recently solved structure of the homologues protein IsdE from *S. aureus* (Grigg *et al.*, 2007b), as molecular replacement model,

a better solution was obtained but the errors in the phases were still large. The electron density maps could be improved by taking advantage of the two molecules in the unit cell and carrying out non-crystallographic symmetry averaging. Due to errors in the phases, the limited resolution as well as the likely pseudo-merohedral twinning of the data, improving and refining the model avoiding model bias will be challenging. Solution of the structure could be facilitated by improving the quality of the crystals to acquire better resolution data and/or combination of the phases obtained from the molecular replacement with good iron-MAD data. Moreover preparation of SeMet-derivatised HtsA could provide better means to structure determination, given the presence of 6 methionines, of which 5 appear to be buried in the molecule (by comparison with IsdE) and hence are likely to be ordered.

Additionally, complex formation of haem-Shp and haem-HtsA which had been previously reported (Liu & Lei, 2005, Nygaard *et al.*, 2006a) was investigated for co-crystallisation trials. Analytical gel-filtration showed that a proportion of the mixed two proteins form a stable interaction, but for co-crystallisation trials the equilibrium would need to be shifted more towards the complex.

Chapter 8

Conclusions

The strong link between iron / haem acquisition, virulence factors and pathogenicity in many pathogenic species emphasises the need to understand the molecular mechanisms of haem-uptake. Pathogenic bacteria have evolved specialised and unique proteins to circumvent their iron dependency and the knowledge of how these proteins bind and release haem could aid future development of new antimicrobial agents; this would be of special interest given the increasing number of bacterial drug-resistant strains.

When the work described here was started 2003 / 2004, there were only a handful of publications on genetical, biochemical and biophysical characterisation on these proteins, so important for bacteria; only the structure of the haemophore HasA from *Serratia marcescens* (Arnoux *et al.*, 1999) had been determined.

In this thesis work the haem-binding properties of the cytosolic protein HemS from *Y. enterocolitica* were biochemically characterised by means of bioinformatics-guided mutagenesis experiments and UV/vis absorption spectroscopy. Moreover the structures of the apo- and haem-HemS were determined and so precious insights were gained into the molecular mechanism of haem-binding. In addition the periplasmic binding protein HemT was cloned and its expres-

sion and purification were optimised. Finally, the streptococcal proteins HtsA and Shp were cloned, expressed, purified and crystallisation trials carried out. Diffraction quality crystals for HtsA were obtained.

During the time of this work, an extensive number of molecular biological studies as well as structure determinations on the bacterial haem-uptake systems was published and is illustrated in Figure 8.1. This is clear evidence of the interest and importance of this research area in the wide scientific community. The structures reported in Figure 8.1 are HemS from *Y. enterocolitica* (Schneider *et al.*, 2006) described in this thesis, as well as structural studies on the HemS homologue ChuS from *E. coli* (Suits *et al.*, 2005, 2006), ChaN from *Campylobacter jejuni* (Chan *et al.*, 2006), IsdH (Pilpa *et al.*, 2006), the IsdA (Grigg *et al.*, 2007a), IsdC (Sharp *et al.*, 2007) and IsdE (Grigg *et al.*, 2007b) proteins of the staphylococcal isd haem-uptake machinery and the streptococcal outer membrane protein Shp (Aranda *et al.*, 2007).

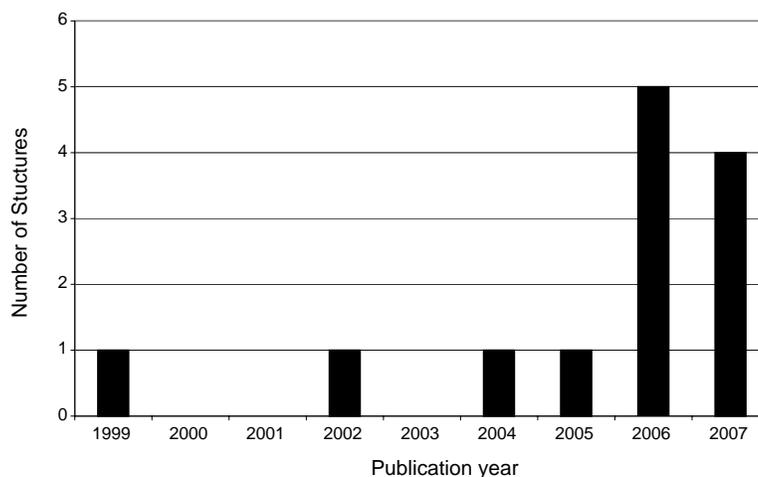


Figure 8.1: Structural studies on the bacterial haem-uptake systems in the last few years. Diagram showing the number of publications on structures of bacterial haem transport proteins. For 2007 number of publications before 20th August were taken into account.

Despite the enormous progress made on the genetical, biochemical, biophysical and structural characterisations of molecular components of the bacterial

haem-uptake machinery in the last few years, the complex relationships between its components and the exact mechanism of haem transfer remains to be determined. For instance the hypothetical model of the flow of the haem in the Isd-system from the Gram-positive pathogen *S. aureus* (Fig. 1.3B) still needs to be validated. A start has been made with the investigation of the haem transfer process from the streptococcal proteins Shp to HtsA (Liu & Lei, 2005, Nygaard *et al.*, 2006a) and the proteins PhuS and paOH from the Gram-negative pathogene *P. aeruginosa* (Bhakta & Wilks, 2006, Lansky *et al.*, 2006). Given the physical separation of the components in Gram-negative species by the two membranes (Fig. 1.3A), haem transfer experiments are challenging and biochemical and structural data on the key membrane components of these systems, like the permease, are still missing. Biochemical data are available on the outer membrane haem-receptor ShuA from *S. dysenteriae*, which has been purified and its haem-binding properties analysed (Burkhard & Wilks, 2007). Some structural data on homologues to the haem-receptors from Gram-negative species are available; the iron transporters FhuA (Ferguson *et al.*, 1998, Locher *et al.*, 1998), FepA (Buchanan *et al.*, 1999) and FecA (Ferguson *et al.*, 2002), which all share the molecular architecture of a β -barrel.

The protein-protein interactions during the haem-transfer process have only been shown for PhuS and paOH (Bhakta & Wilks, 2006, Lansky *et al.*, 2006) as well as HtsA and Shp (Liu & Lei, 2005, Nygaard *et al.*, 2006a). But conformational changes in the molecules triggered by the interactions during haem-transfer have not been investigated. No protein-protein interaction between any of the components of the Isd-system from *S. aureus* has been analysed to date.

The fate of the haem once in the cell is only partly understood; some of the haem-oxygenases involved in haem-degradation have been identified (Zhu

et al., 2000, Unno *et al.*, 2004, Skaar *et al.*, 2004a, Lansky *et al.*, 2006, Puri & O'Brian, 2006), but the effector molecules in many species, like for instance *Yersinia* and *Shigella* are still unknown.

Finally because of the importance of iron as a nutrient, pathogens have evolved multiple, phase variable acquisition systems that enables them to escape the host immune system; for instance two different haem-uptake systems have been identified in *S. aureus*; the Isd-system (Skaar *et al.*, 2004b), and the HtsABC-system (Friedman *et al.*, 2006) and the interplay between the two systems in the same cell remains to be elucidated.

Bibliography

- Abrahams, J.P. & Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr*, **D52**, 30–42.
- Abrahams, J.P., Buchanan, S.K., Raaij, M.J.V., Fearnley, I.M., Leslie, A.G. & Walker, J.E. (1996). The structure of bovine F1-ATPase complexed with the peptide antibiotic efrapeptin. *Proc Natl Acad Sci U S A*, **93**, 9420–24.
- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. & Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr*, **D58**, 1948–54.
- Adhya, S. & Garges, S. (1990). Positive control. *J Biol Chem*, **265**, 10797–800.
- Aiba, H., Fujimoto, S. & Ozaki, N. (1982). Molecular cloning and nucleotide sequencing of the gene for *E. coli* cAMP receptor protein. *Nucleic Acids Res*, **10**, 1345–61.
- Allen, J.W., Barker, P.D., Daltrop, O., Stevens, J.M., Tomlinson, E.J., Sinha, N., Sambongi, Y. & Ferguson, S.J. (2005). Why isn't 'standard' heme good enough for c-type and d1-type cytochromes? *Dalton Trans*, 3410–8.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol*, **215**, 403–10.
- Anderson, J.L. & Chapman, S.K. (2005). Ligand probes for heme proteins. *Dalton Trans*, 13–24.
- Andrade, M.A., Ciccarelli, F.D., Perez-Iratxeta, C. & Bork, P. (2002). NEAT: a domain duplicated in genes near the components of a putative Fe³⁺ siderophore transporter from Gram-positive pathogenic bacteria. *Genome Biol*, **3**, RESEARCH0047.
- Antonini, E. & Brunori, M. (1971). Hemoglobin and myoglobin in their reactions with ligands. *Elsevier/North-Holland Biomedical Press, Amsterdam*.
- Aranda, R., E., W.C., Liu, M., Bitto, E., Cates, S., Olson, J. & Lei, G., B. Phillips (2007). Bis-methionyl coordination and heme stacking in the crystal structure of the heme-binding domain of the streptococcal cell surface protein Shp. *J Mol Biol*, in press.

- Arnoux, P., Haser, R., Izadi, N., Lecroisey, A., Delepierre, M., Wandersman, C. & Czjzek, M. (1999). The crystal structure of HasA, a hemophore secreted by *Serratia marcescens*. *Nat Struct Biol*, **6**, 516–20.
- Baker, E.N., Anderson, B.F., Dobbs, A.J. & Dodson, E.J. (1995). Use of iron anomalous scattering with multiple models and data sets to identify and refine a weak molecular replacement solution: structure analysis of cytochrome *c'* from two bacterial species. *Acta Crystallogr*, **D51**, 282–9.
- Barker, P.D., Nerou, E.P., Cheesman, M.R., Thomson, A.J., de Oliveira, P. & Hill, H.A. (1996). Bis-methionine ligation to heme iron in mutants of cytochrome b562. 1. Spectroscopic and electrochemical characterization of the electronic properties. *Biochemistry*, **35**, 13618–26.
- Barrick, D. (1994). Replacement of the proximal ligand of sperm whale myoglobin with free imidazole in the mutant His-93 ->Gly. *Biochemistry*, **33**, 6546–54.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L.L., Studholme, D.J., Yeats, C. & Eddy, S.R. (2004). The Pfam protein families database. *Nucleic Acids Res*, **32**, D138–41.
- Bates, C.S., Montaez, G.E., Woods, C.R., Vincent, R.M. & Eichenbaum, Z. (2003). Identification and characterization of a *Streptococcus pyogenes* operon involved in binding of hemoproteins and acquisition of iron. *Infect Immun*, **71**, 1042–55.
- Bennett, J., W. S. & Steitz, T.A. (1978). Glucose-induced conformational change in yeast hexokinase. *Proc Natl Acad Sci U S A*, **75**, 4848–52.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, J., E. F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1978). The Protein Data Bank: a computer-based archival file for macromolecular structures. *Arch Biochem Biophys*, **185**, 584–91.
- Bhakta, M.N. & Wilks, A. (2006). The mechanism of heme transfer from the cytoplasmic heme binding protein PhuS to the delta-regioselective heme oxygenase of *Pseudomonas aeruginosa*. *Biochemistry*, **45**, 11642–49.
- Borths, E.L., Locher, K.P., Lee, A.T. & Rees, D.C. (2002). The structure of *Escherichia coli* BtuF and binding to its cognate ATP binding cassette transporter. *Proc Natl Acad Sci U S A*, **99**, 16642–7.
- Bracken, C.S., Baer, M.T., Abdur-Rashid, A., Helms, W. & Stojiljkovic, I. (1999). Use of heme-protein complexes by the *Yersinia enterocolitica* HemR receptor: histidine residues are essential for receptor function. *J Bacteriol*, **181**, 6063–72.

- Braun, V. (2001). Iron uptake mechanisms and their regulation in pathogenic bacteria. *Int J Med Microbiol*, **291**, 67–79.
- Braun, V. & Braun, M. (2002). Iron transport and signaling in *Escherichia coli*. *FEBS Lett*, **529**, 78–85.
- Braun, V., Hantke, K. & Koster, W. (1998). Bacterial iron transport: mechanisms, genetics, and regulation. *Met Ions Biol Syst*, **35**, 67–145.
- Brünger, A.T. (1988). Crystallographic refinement by simulated annealing. Application to a 2.8 Å resolution structure of aspartate aminotransferase. *J Mol Biol*, **203**, 803–16.
- Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–5.
- Buchanan, S.K., Smith, B.S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., van der Helm, D. & Deisenhofer, J. (1999). Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*. *Nat Struct Biol*, **6**, 56–63.
- Bürgi, H.B. (2000). Motion and disorder in crystal structure analysis: measuring and distinguishing them. *Annu Rev Phys Chem*, **51**, 275–96.
- Bürgi, H.P. & Capelli, S.C. (2000). Dynamics of molecules in crystals from multi-temperature anisotropic displacement parameters. I. theory. *Acta Crystallogr*, **A56**, 403–12.
- Burkhard, K.A. & Wilks, A. (2007). Characterization of the outer membrane receptor ShuA from the heme uptake system of *Shigella dysenteriae*. Substrate specificity and identification of the heme protein ligands. *J Biol Chem*, **282**, 15126–36.
- Carlsson, G.H., Nicholls, P., Svistunenko, D., Berglund, G.I. & Hajdu, J. (2005). Complexes of horseradish peroxidase with formate, acetate, and carbon monoxide. *Biochemistry*, **44**, 635–42.
- CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr*, **D50**, 760–3.
- Chan, A.C.K., Lelj-Garolla, B., Rosell, F.I., Pedersen, K.A., Mauk, A.G. & Murphy, M.E.P. (2006). Cofacial heme binding is linked to dimerization by a bacterial heme transport protein. *J Mol Biol*, **362**, 1108–19.
- Chapman, S., Daff, S. & Munro, A. (1997). Heme: The most versatile redox centre in biology? *Structure and Bonding*, **88**, 39–70.
- Chi, Y.I., Huang, L.S., Zhang, Z., Fernandez-Velasco, J.G. & Berry, E.A. (2000). X-ray structure of a truncated form of cytochrome f from *Chlamydomonas reinhardtii*. *Biochemistry*, **39**, 7689–701.

- Choi, C.Y., Cerda, J.F., Chu, H.A., Babcock, G.T. & Marletta, M.A. (1999). Spectroscopic characterization of the heme-binding sites in *Plasmodium falciparum* histidine-rich protein 2. *Biochemistry*, **38**, 16916–24.
- Choma, C.T., Lear, J.D., Nelson, M.J., Dutton, P.L., Robertson, D.E. & DeGrado, W.F. (1994). Design of a heme-binding four-helix bundle. *J Am Chem Soc*, **116**, 856–65.
- Clarke, S.R., Wiltshire, M.D. & Foster, S.J. (2004). IsdA of *Staphylococcus aureus* is a broad spectrum, iron-regulated adhesin. *Mol Microbiol*, **51**, 1509–19.
- Cowtan, K. (1999). Error estimation and bias correction in phase-improvement calculations. *Acta Crystallogr*, **D55**, 1555–67.
- Cox, M.C., Le Brun, N., Thomson, A.J., Smith, A., Morgan, W.T. & Moore, G.R. (1995). MCD, EPR and NMR spectroscopic studies of rabbit hemopexin and its heme binding domain. *Biochim Biophys Acta*, **1253**, 215–23.
- Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. & Barton, G.J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics*, **14**, 892–3.
- Cunningham, M.W. (2000). Pathogenesis of group A streptococcal infections. *Clin Microbiol Rev*, **13**, 470–511.
- Dailey, H.A. (2002). Terminal steps of haem biosynthesis. *Biochem Soc Trans*, **30**, 590–5.
- Dangi, B., Sarma, S., Yan, C., Banville, D.L. & Guiles, R.D. (1998). The origin of differences in the physical properties of the equilibrium forms of cytochrome b5 revealed through high-resolution NMR structures and backbone dynamic analyses. *Biochemistry*, **37**, 8289–302.
- Dauter, Z. (1999). Data-collection strategies. *Acta Crystallogr*, **D 55**, 1703–17.
- de La Fortelle, E. & Bricogne, G. (1997). Maximum-Likelihood Heavy-Atom Parameter Refinement for the Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. *Methods in Enzymology*, **276**, 472–94.
- Degtyarenko, K.N., North, A.C. & Findlay, J.B. (1997). Promise: a new database of information on prosthetic centres and metal ions in protein active sites. *Protein Eng*, **10**, 183–6.
- Delano, W. (2002). The PyMOL Molecular Graphics System. *on the World Wide Web* <http://www.pymol.org>.

- Derman, A.I., Prinz, W.A., Belin, D. & Beckwith, J. (1993). Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science*, **262**, 1744–7.
- Dewilde, S., Kiger, L., Burmester, T., Hankeln, T., Baudin-Creuzat, V., Aerts, T., Marden, M.C., Caubergs, R. & Moens, L. (2001). Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. *J Biol Chem*, **276**, 38949–55.
- Diederichs, K. & Karplus, P.A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol*, **4**, 269–75.
- Drenth, J. (1999). Principles of Protein X-ray crystallography. *Springer Verlag*.
- Dryla, A., Gelbmann, D., von Gabain, A. & Nagy, E. (2003). Identification of a novel iron regulated staphylococcal surface protein with haptoglobin-haemoglobin binding activity. *Mol Microbiol*, **49**, 37–53.
- Dubendorff, J.W. & Studier, F.W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol*, **219**, 45–59.
- Eakanunkul, S., Lukat-Rodgers, G.S., Sumithran, S., Ghosh, A., Rodgers, K.R., Dawson, J.H. & Wilks, A. (2005). Characterization of the periplasmic heme-binding protein shut from the heme uptake system of *Shigella dysenteriae*. *Biochemistry*, **44**, 13179–91.
- Eichenbaum, Z., Muller, E., Morse, S.A. & Scott, J.R. (1996). Acquisition of iron from host proteins by the group A *Streptococcus*. *Infect Immun*, **64**, 5428–9.
- Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr*, **D60**, 2126–32.
- Evans, P. (1997). SCALA. *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, **33**, 22–4.
- Faraldo-Gomez, J.D. & Sansom, M.S. (2003). Acquisition of siderophores in gram-negative bacteria. *Nat Rev Mol Cell Biol*, **4**, 105–16.
- Ferguson, A.D., Hofmann, E., Coulton, J.W., Diederichs, K. & Welte, W. (1998). Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. *Science*, **282**, 2215–20.
- Ferguson, A.D., Chakraborty, R., Smith, B.S., Esser, L., van der Helm, D. & Deisenhofer, J. (2002). Structural basis of gating by the outer membrane transporter FecA. *Science*, **295**, 1715–9.

- Fischmann, T.O., Hruza, A., Niu, X.D., Fossetta, J.D., Lunn, C.A., Dolphin, E., Prongay, A.J., Reichert, P., Lundell, D.J., Narula, S.K. & Weber, P.C. (1999). Structural characterization of nitric oxide synthase isoforms reveals striking active-site conservation. *Nat Struct Biol*, **6**, 233–42.
- Francis, R.T., Booth, J.W. & Becker, R.R. (1985). Uptake of iron from hemoglobin and the haptoglobin-hemoglobin complex by hemolytic bacteria. *Int J Biochem*, **17**, 767–73.
- Franza, T., Michaud-Soret, I., Piquerel, P. & Expert, D. (2002). Coupling of iron assimilation and pectinolysis in *Erwinia chrysanthemi* 3937. *Mol Plant Microbe Interact*, **15**, 1181–91.
- Friedman, D.B., Stauff, D.L., Pishchany, G., Whitwell, C.W., Torres, V.J. & Skaar, E.P. (2006). *Staphylococcus aureus* redirects central metabolism to increase iron availability. *PLoS Pathog*, **2**, e87.
- Genco, C.A. & Dixon, D.W. (2001). Emerging strategies in microbial haem capture. *Mol Microbiol*, **39**, 1–11.
- Gibney, B.R. & Dutton, P.L. (1999). Histidine placement in de novo-designed heme proteins. *Protein Sci*, **8**, 1888–98.
- Gibney, B.R., Rabanal, F., Reddy, K.S. & Dutton, P.L. (1998). Effect of four helix bundle topology on heme binding and redox properties. *Biochemistry*, **37**, 4635–43.
- Glaser, F., Pupko, T., Paz, I., Bell, R., Bechor, D., Martz, E. & Ben-Tal, N. (2003). ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information <http://consurf.tau.ac.il>. *Bioinformatics*, **19**, 163–164.
- Grigg, J.C., Vermeiren, C.L., Heinrichs, D.E. & Murphy, M.E.P. (2007a). Haem recognition by a *Staphylococcus aureus* NEAT domain. *Mol Microbiol*, **63**, 139–49.
- Grigg, J.C., Vermeiren, C.L., Heinrichs, D.E. & Murphy, M.E.P. (2007b). Heme coordination by *Staphylococcus aureus* IsdE. *J Biol Chem*, in press.
- Guex, N. & Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–23.
- Guex, N., Diemand, A. & Peitsch, M.C. (1999). Protein modelling for all. *Trends Biochem Sci*, **24**, 364–7.
- Hallberg, B.M., Bergfors, T., Backbro, K., Pettersson, G., Henriksson, G. & Divne, C. (2000). A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase. *Structure*, **8**, 79–88.

- Hartshorn, M.J. (2002). AstexViewer: a visualisation aid for structure-based drug design. *J Comput Aided Mol Des*, **16**, 871–81.
- Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A. & Deisenhofer, J. (1995). Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure*, **3**, 41–62.
- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. & Shinagawa, H. (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res*, **8**, 11–22.
- Henderson, D.P., Wyckoff, E.E., Rashidi, C.E., Verlei, H. & Oldham, A.L. (2001). Characterization of the *Plesiomonas shigelloides* genes encoding the heme iron utilization system. *J Bacteriol*, **183**, 2715–23.
- Hendrickson, W.A., Horton, J.R. & LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J*, **9**, 1665–72.
- Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–9.
- Hofstee, B.H. (1973). Immobilization of enzymes through non-covalent binding to substituted agaroses. *Biochem Biophys Res Commun*, **53**, 1137–44.
- Hofstee, B.H. (1974). Non-specific binding of proteins by substituted agaroses. *Adv Exp Med Biol*, **42**, 43–59.
- Howlin, B., Moss, D.A. & Harris, G.W. (1989). Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. *Acta Crystallogr*, **A45**, 851–61.
- Huntley, T.E. & Strittmatter, P. (1972). The effect of heme binding on the tryptophan residue and the protein conformation of cytochrome b 5. *J Biol Chem*, **247**, 4641–7.
- Inuzuka, T., Yun, B.G., Ishikawa, H., Takahashi, S., Hori, H., Matts, R.L., Ishimori, K. & Morishima, I. (2004). Identification of crucial histidines for heme binding in the N-terminal domain of the heme-regulated eIF2alpha kinase. *J Biol Chem*, **279**, 6778–82.
- Jeoung, J.H., Pippig, D.A., Martins, B.M., Wagener, N. & Dobbek, H. (2007). HTHP: a novel class of hexameric, tyrosine-coordinated heme proteins. *J Mol Biol*, **368**, 1122–31.

- Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, **292**, 195–202.
- Joung, J.K., Ramm, E.I. & Pabo, C.O. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A*, **97**, 7382–7.
- Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *J Appl Cryst*, **21**, 67–72.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst*, **26**, 795–800.
- Karle, J. & Hauptman, H. (1956). A theory of phase determination for the four types of non-centrosymmetric space groups $1P222$, $2P22$, $3P_12$, $3P_22$. *Acta Crystallogr*, **9**, 635–51.
- Kaur, A. & Wilks, A. (2007). Heme Inhibits the DNA Binding Properties of the Cytoplasmic Heme Binding Protein of *Shigella dysenteriae* (ShuS). *Biochemistry*, **46**, 2994–3000.
- Kisker, C., Schindelin, H., Pacheco, A., Wehbi, W.A., Garrett, R.M., Rajagopalan, K.V., Enemark, J.H. & Rees, D.C. (1997). Molecular basis of sulfite oxidase deficiency from the structure of sulfite oxidase. *Cell*, **91**, 973–83.
- Kleywegt, G.J. (2000). Validation of protein crystal structures. *Acta Crystallogr*, **D56**, 249–65.
- Kleywegt, G.J. & Jones, T.A. (1995). Where freedom is given, liberties are taken. *Structure*, **3**, 535–40.
- Kloer, D.P., Hagel, C., Heider, J. & Schulz, G.E. (2006). Crystal structure of ethylbenzene dehydrogenase from *Aromatoleum aromaticum*. *Structure*, **14**, 1377–88.
- Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr*, **D60**, 2256–68.
- Kuklin, N.A., Clark, D.J., Secore, S., Cook, J., Cope, L.D., McNeely, T., Noble, L., Brown, M.J., Zorman, J.K., Wang, X.M., Pancari, G., Fan, H., Isett, K., Burgess, B., Bryan, J., Brownlow, M., George, H., Mainz, M., Liddell, M.E., Kelly, R., Schultz, L., Montgomery, D., Onishi, J., Losada, M., Martin, M., Ebert, T., Tan, C.Y., Schofield, T.L., Nagy, E., Meineke, A., Joyce, J.G., Kurtz, M.B., Caulfield, M.J., Jansen, K.U., McClements, W. & Anderson, A.S. (2006). A novel *Staphylococcus aureus* vaccine: iron surface determinant B induces rapid antibody responses in rhesus macaques and specific increased survival in a murine *S. aureus* sepsis model. *Infect Immun*, **74**, 2215–23.

- Kurokawa, H., Lee, D.S., Watanabe, M., Sagami, I., Mikami, B., Raman, C.S. & Shimizu, T. (2004). A redox-controlled molecular switch revealed by the crystal structure of a bacterial heme PAS sensor. *J Biol Chem*, **279**, 20186–93.
- Lansky, I.B., Lukat-Rodgers, G.S., Block, D., Rodgers, K.R., Ratliff, M. & Wilks, A. (2006). The cytoplasmic heme-binding protein (PhuS) from the heme uptake system of *Pseudomonas aeruginosa* is an intracellular heme-trafficking protein to the delta -regioselective heme oxygenase. *J Biol Chem*, **281**, 13652–62.
- Lanzilotta, W.N., Schuller, D.J., Thorsteinsson, M.V., Kerby, R.L., Roberts, G.P. & Poulos, T.L. (2000). Structure of the CO sensing transcription activator CooA. *Nat Struct Biol*, **7**, 876–80.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*, **26**, 283–91.
- Layer, G., Verfrth, K., Mahlitz, E. & Jahn, D. (2002). Oxygen-independent coproporphyrinogen-III oxidase HemN from *Escherichia coli*. *J Biol Chem*, **277**, 34136–42.
- Lee, J.H. & Choi, S.H. (2006). Coactivation of *Vibrio vulnificus* putAP operon by cAMP receptor protein and PutR through cooperative binding to overlapping sites. *Mol Microbiol*, **60**, 513–24.
- Lee, K.B., La Mar, G.N., Mansfield, K.E., Smith, K.M., Pochapsky, T.C. & Sligar, S.G. (1993). Interpretation of hyperfine shift patterns in ferricytochromes b₅ in terms of angular position of the heme: a sensitive probe for peripheral heme protein interactions. *Biochim Biophys Acta*, **1202**, 189–99.
- Lei, B., Smoot, L.M., Menning, H.M., Voyich, J.M., Kala, S.V., Deleo, F.R., Reid, S.D. & Musser, J.M. (2002). Identification and characterization of a novel heme-associated cell surface protein made by *Streptococcus pyogenes*. *Infect Immun*, **70**, 4494–500.
- Lei, B., Liu, M., Voyich, J.M., Prater, C.I., Kala, S.V., DeLeo, F.R. & Musser, J.M. (2003). Identification and characterization of HtsA, a second heme-binding protein made by *Streptococcus pyogenes*. *Infect Immun*, **71**, 5962–9.
- Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, 26.
- Liong, E.C., Dou, Y., Scott, E.E., Olson, J.S. & Phillips, J., G. N. (2001). Waterproofing the heme pocket. Role of proximal amino acid side chains in preventing heme loss from myoglobin. *J Biol Chem*, **276**, 9093–100.

- Liu, M. & Lei, B. (2005). Heme transfer from streptococcal cell surface protein Shp to HtsA of transporter HtsABC. *Infect Immun*, **73**, 5086–92.
- Locher, K.P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J.P. & Moras, D. (1998). Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes. *Cell*, **95**, 771–8.
- Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S. & Richardson, D.C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*, **50**, 437–50.
- Mack, J., Vermeiren, C., Heinrichs, D.E. & Stillman, M.J. (2004). In vivo heme scavenging by *Staphylococcus aureus* IsdC and IsdE proteins. *Biochem Biophys Res Commun*, **320**, 781–8.
- MacRae, I.J. & Doudna, J.A. (2007). An unusual case of pseudo-merohedral twinning in orthorhombic crystals of Dicer. *Acta Crystallogr*, **D63**, 993–9.
- Marengo, A.W. & Schneewind, O. (2006). Iron acquisition and transport in *Staphylococcus aureus*. *Biometals*, **19**, 193–203.
- Mazmanian, S.K., Skaar, E.P., Gaspar, A.H., Humayun, M., Gornicki, P., Jelenska, J., Joachmiak, A., Missiakas, D.M. & Schneewind, O. (2003). Passage of heme-iron across the envelope of *Staphylococcus aureus*. *Science*, **299**, 906–9.
- McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr*, **D61**, 458–64.
- McGaughey, G.B., Gagne, M. & Rappe, A.K. (1998). pi-Stacking interactions. Alive and well in proteins. *J Biol Chem*, **273**, 15458–63.
- McKay, D.B. & Steitz, T.A. (1981). Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature*, **290**, 744–49.
- McKay, D.B., Weber, I.T. & Steitz, T.A. (1982). Structure of catabolite gene activator protein at 2.9 Å resolution. Incorporation of amino acid sequence and interactions with cyclic AMP. *J Biol Chem*, **257**, 9518–24.
- McPherson, A. (2001). A comparison of salts for the crystallization of macromolecules. *Protein Sci*, **10**, 418–22.
- McRee, D.E., Jensen, G.M., Fitzgerald, M.M., Siegel, H.A. & Goodin, D.B. (1994). Construction of a bisquo heme enzyme and binding by exogenous ligands. *Proc Natl Acad Sci U S A*, **91**, 12847–51.

- Mills, M. & Payne, S.M. (1995). Genetics and regulation of heme iron transport in *Shigella dysenteriae* and detection of an analogous system in *Escherichia coli* O157:H7. *J Bacteriol*, **177**, 3004–9.
- Mortuza, G.B. & Whitford, D. (1997). Mutagenesis of residues 27 and 78 modulates heme orientation in cytochrome b5. *FEBS Lett*, **412**, 610–4.
- Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr*, **D53**, 240–55.
- Musiani, F., Zambelli, B., Stola, M. & Ciurli, S. (2004). Nickel trafficking: insights into the fold and function of UreE, a urease metallochaperone. *J Inorg Biochem*, **98**, 803–13.
- Nioche, P., Berka, V., Vipond, J., Minton, N., Tsai, A.L. & Raman, C.S. (2004). Femtomolar Sensitivity of a NO Sensor from *Clostridium botulinum*. *Science*, **306**, 1550–3.
- Nygaard, T.K., Blouin, G.C., Liu, M., Fukumura, M., Olson, J.S., Fabian, M., Dooley, D.M. & Lei, B. (2006a). The mechanism of direct heme transfer from the streptococcal cell surface protein Shp to HtsA of the HtsABC transporter. *J Biol Chem*, **281**, 20761–71.
- Nygaard, T.K., Liu, M., McClure, M.J. & Lei, B. (2006b). Identification and characterization of the heme-binding proteins SeShp and SeHtsA of *Streptococcus equi* subspecies equi. *BMC Microbiol*, **6**, 1–10.
- Obataya, I., Kotaki, T., Sakamoto, S., Ueno, A. & Mihara, H. (2000). Design, synthesis and peroxidase-like activity of 3alpha-helix proteins covalently bound to heme. *Bioorg Med Chem Lett*, **10**, 2719–22.
- Ochsner, U.A., Johnson, Z. & Vasil, M.L. (2000). Genetics and regulation of two distinct haem-uptake systems, phu and has, in *Pseudomonas aeruginosa*. *Microbiology*, **146**, 185–98.
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, **276**, 307–26.
- Painter, J. & Merritt, E.A. (2006a). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr*, **D62**, 439–50.
- Painter, J. & Merritt, E.A. (2006b). TLSMD web server for the generation of multi-group TLS models. *J Appl Cryst*, **39**, 109–11.
- Paoli, M., Anderson, B.F., Baker, H.M., Morgan, W.T., Smith, A. & Baker, E.N. (1999). Crystal structure of hemopexin reveals a novel high-affinity heme site formed between two beta-propeller domains. *Nat Struct Biol*, **6**, 926–31.

- Paoli, M., Marles-Wright, J. & Smith, A. (2002). Structure-function relationships in heme-proteins. *DNA Cell Biol*, **21**, 271–80.
- Pellicena, P., Karow, D.S., Boon, E.M., Marletta, M.A. & Kuriyan, J. (2004). Crystal structure of an oxygen-binding heme domain related to soluble guanylate cyclases. *Proc Natl Acad Sci U S A*, **101**, 12854–9.
- Peränen, J., Rikkonen, M., Hyvonen, M. & Kaariainen, L. (1996). T7 vectors with modified T7lac promoter for expression of proteins in *Escherichia coli*. *Anal Biochem*, **236**, 371–3.
- Perna, N.T., Plunkett, r., G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamouisis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. & Blattner, F.R. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature*, **409**, 529–33.
- Pfeil, W. (1993). Thermodynamics of apocytochrome b5 unfolding. *Protein Sci*, **2**, 1497–501.
- Pilpa, R.M., Fadeev, E.A., Villareal, V.A., Wong, M.L., Phillips, M. & Clubb, R.T. (2006). Solution structure of the NEAT (NEAr Transporter) domain from IsdH/HarA: the human hemoglobin receptor in *Staphylococcus aureus*. *J Mol Biol*, **360**, 435–47.
- Poulos, T. (2006). *Peroxidases and catalases*. Biological Inorganic Chemistry.
- Poulos, T.L., Edwards, S.L., Wariishi, H. & Gold, M.H. (1993). Crystallographic refinement of lignin peroxidase at 2 Å. *J Biol Chem*, **268**, 4429–40.
- Puri, S. & O'Brian, M.R. (2006). The hmuQ and hmuD genes from *Bradyrhizobium japonicum* encode heme-degrading enzymes. *J Bacteriol*, **188**, 6476–82.
- Rabanal, F., DeGrado, W.F. & Dutton, P.L. (1996). Toward the synthesis of a photosynthetic reaction centre maquette: A cofacial porphyrin pair assembled between two subunits of a synthetic four-helix bundle multiheme protein. *J Am Chem Soc*, **118**, 473–74.
- Ramachandran, G.N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J Mol Biol*, **7**, 95–9.
- Ramakrishnan, C. & Ramachandran, G.N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys J*, **5**, 909–33.
- Ratledge, C. & Dover, L.G. (2000). Iron metabolism in pathogenic bacteria. *Annu Rev Microbiol*, **54**, 881–941.

- Rhodes, G. (2006). *Crystallography made crystal clear*. Complementary Science Series, 3rd edn.
- Richards, F. & Kundrot, C. (1988). Identification of structural motifs from protein coordinate data: Secondary structure and first-level supersecondary structure. *Proteins*, **3**, 71–84.
- Rivera, M., Barillas-Mury, C., Christensen, K.A., Little, J.W., Wells, M.A. & Walker, F.A. (1992). Gene synthesis, bacterial expression, and 1H NMR spectroscopic studies of the rat outer mitochondrial membrane cytochrome b5. *Biochemistry*, **31**, 12233–40.
- Robertson, D.E., Farid, R.S., Moser, C.C., Urbauer, J.L., Mulholland, S.E., Pidikiti, R., Lear, J.D., Wand, A.J., DeGrado, W.F. & Dutton, P.L. (1994). Design and synthesis of multi-haem proteins. *Nature*, **368**, 425–32.
- Rojas, N.R., Kamtekar, S., Simons, C.T., McLean, J.E., Vogel, K.M., Spiro, T.G., Farid, R.S. & Hecht, M.H. (1997). De novo heme proteins from designed combinatorial libraries. *Protein Sci*, **6**, 2512–24.
- Rotsaert, F.A., Li, B., Renganathan, V. & Gold, M.H. (2001). Site-directed mutagenesis of the heme axial ligands in the hemoflavoenzyme cellobiose dehydrogenase. *Arch Biochem Biophys*, **390**, 206–14.
- Rotsaert, F.A., Hallberg, B.M., de Vries, S., Moenne-Loccoz, P., Divne, C., Renganathan, V. & Gold, M.H. (2003a). Biophysical and structural analysis of a novel heme B iron ligation in the flavocytochrome cellobiose dehydrogenase. *J Biol Chem*, **278**, 33224–31.
- Rotsaert, F.A., Renganathan, V. & Gold, M.H. (2003b). Role of the flavin domain residues, His689 and Asn732, in the catalytic mechanism of cellobiose dehydrogenase from *Phanerochaete chrysosporium*. *Biochemistry*, **42**, 4049–56.
- Rudolph, M.G., Wingren, C., Crowley, M.P., Chien, Y.H. & Wilson, I.A. (2004). Combined pseudo-merohedral twinning, non-crystallographic symmetry and pseudo-translation in a monoclinic crystal form of the gamma-delta T-cell ligand T10. *Acta Crystallogr*, **D60**, 656–64.
- Rupp, B. (retrieved July 2006 from the world wide web). Crystallography 101. <http://ruppweb.dyndns.org/Xray/101index.html>.
- Samanta, U., Pal, D. & Chakrabarti, P. (1999). Packing of aromatic rings against tryptophan residues in proteins. *Acta Crystallogr*, **D55**, 1421–7.
- Sambrook, J. & Fritsch, E.F. (1989). *Molecular Cloning : A Laboratory Manual (3 Vols.)*. Cold Spring Harbor Laboratory Press.

- Sasakura, Y., Hirata, S., Sugiyama, S., Suzuki, S., Taguchi, S., Watanabe, M., Matsui, T., Sagami, I. & Shimizu, T. (2002). Characterization of a direct oxygen sensor heme protein from *Escherichia coli*. Effects of the heme redox states and mutations at the heme-binding site on catalysis and structure. *J Biol Chem*, **277**, 23821–7.
- Sato, E., Sagami, I., Uchida, T., Sato, A., Kitagawa, T., Igarashi, J. & Shimizu, T. (2004). SOUL in mouse eyes is a new hexameric heme-binding protein with characteristic optical absorption, resonance raman spectral, and heme-binding properties. *Biochemistry*, **43**, 14189–98.
- Schade, A. & Caroline, L. (1946). An iron-binding component in human blood plasma. *Science*, **11**, 340–41.
- Schade, A.L. & Caroline, L. (1944). Raw hen egg white and the role of iron in growth inhibition of *Shigella dysenteriae*, *Staphylococcus aureus*, *Escherichia coli* and *Saccharomyces cerevisiae*. *Science*, **7**, 14–5.
- Schneider, S. & Paoli, M. (2005). Crystallization and preliminary X-ray diffraction analysis of the haem-binding protein HemS from *Yersinia enterocolitica*. *Acta Crystallogr*, **F61**, 802–5.
- Schneider, S., Sharp, K.H., Barker, P.D. & Paoli, M. (2006). An induced fit conformational change underlies the binding mechanism of the heme transport proteobacteria-protein HemS. *J Biol Chem*, **281**, 32606–10.
- Schneider, T.R. & Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallogr*, **D58**, 1772–9.
- Schomaker, V. & Trueblood, K.N. (1968). On the rigid-body motion of molecules in crystals. *Acta Crystallogr*, **B24**, 63–76.
- Sharff, A.J., Koronakis, E., Luisi, B. & Koronakis, V. (2000). Oxidation of selenomethionine: some MADness in the method! *Acta Crystallogr*, **D56**, 785–788.
- Sharp, K.H., Schneider, S., Cockayne, A. & Paoli, M. (2007). Crystal structure of the heme-IsdC complex, the central conduit of the Isd iron/heme uptake system in *Staphylococcus aureus*. *J Biol Chem*, **282**, 10625–31.
- Shipulina, N., Hunt, R.C., Shaklai, N. & Smith, A. (1998). Coordination of nitric oxide by heme-hemopexin. *J Protein Chem*, **17**, 255–60.
- Shouldice, S.R., Dougan, D.R., Skene, R.J., Tari, L.W., McRee, D.E., Hua Yu, R. & Schryvers, A.B. (2003). High resolution structure of an alternate form of the ferric ion binding protein from *Haemophilus influenzae*. *J Biol Chem*, **278**, 11513–9.
- Skaar, E.P. & Schneewind, O. (2004). Iron-regulated surface determinants (Isd) of *Staphylococcus aureus*: stealing iron from heme. *Microbes Infect*, **6**, 390–7.

- Skaar, E.P., Gaspar, A.H. & Schneewind, O. (2004a). IsdG and IsdI, heme-degrading enzymes in the cytoplasm of *Staphylococcus aureus*. *J Biol Chem*, **279**, 436–43.
- Skaar, E.P., Humayun, M., Bae, T., DeBord, K.L. & Schneewind, O. (2004b). Iron-source preference of *Staphylococcus aureus* infections. *Science*, **305**, 1626–8.
- Smith, A. & Hunt, R.C. (1990). Hemopexin joins transferrin as representative members of a distinct class of receptor-mediated endocytic transport systems. *Eur J Cell Biol*, **53**, 234–45.
- Smulevich, G. (1998). Understanding heme cavity structure of peroxidases: comparison of electronic absorption and resonance Raman spectra with crystallographic results. *Biospectroscopy*, **4**, S3–17.
- Sofia, H.J., Chen, G., Hetzler, B.G., Reyes-Spindola, J.F. & Miller, N.E. (2001). Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res*, **29**, 1097–106.
- Stein, P., Mitchell, M. & Spiro, T.G. (1980). Hydrogen-bond and deprotonation effects on the resonance Raman iron-imidazole mode in deoxyhemoglobin models: implications for hemoglobin cooperativity. *Journal of the American Chemical Society*, **102**, 7795–7.
- Stevens, T. & Paoli, M. (2007). RCC1-like repeat proteins: A pangenomic, structurally diverse new superfamily of β -propeller proteins. *Proteins: Structure, Function, and Bioinformatics*, **in press**.
- Stojiljkovic, I. & Hantke, K. (1992). Hemin uptake system of *Yersinia enterocolitica*: similarities with other TonB-dependent systems in gram-negative bacteria. *EMBO J*, **11**, 4359–67.
- Stojiljkovic, I. & Hantke, K. (1994). Transport of haemin across the cytoplasmic membrane through a haemin-specific periplasmic binding-protein-dependent transport system in *Yersinia enterocolitica*. *Mol Microbiol*, **13**, 719–32.
- Stone, J.R. & Marletta, M.A. (1994). Soluble guanylate cyclase from bovine lung: activation with nitric oxide and carbon monoxide and spectral characterization of the ferrous and ferric states. *Biochemistry*, **33**, 5636–40.
- Studier, F.W. & Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, **189**, 113–30.

- Suits, M., Jaffer, N. & Jia, Z. (2006). 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. *J Biol Chem*, **281**, 36776–82.
- Suits, M.D., Pal, G.P., Nakatsu, K., Matte, A., Cygler, M. & Jia, Z. (2005). Identification of an *Escherichia coli* O157:H7 heme oxygenase with tandem functional repeats. *Proc Natl Acad Sci U S A*, **102**, 16955–60.
- Takahashi, S., Wang, J., Rousseau, D.L., Ishikawa, K., Yoshida, T., Host, J.R. & Ikeda-Saito, M. (1994). Heme-heme oxygenase complex. Structure of the catalytic site and its implication for oxygen activation. *J Biol Chem*, **269**, 1010–4.
- Tamura, M., Asakura, T. & Yonetani, T. (1972). Heme-modification studies on horseradish peroxidase. *Biochim Biophys Acta*, **268**, 292–304.
- Tamura, T., Terada, T. & Tanaka, A. (2003). A quantitative analysis and chemical approach for the reduction of nonspecific binding proteins on affinity resins. *Bioconjug Chem*, **14**, 1222–30.
- Terwilliger, T.C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr*, **D55**, 849–61.
- Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673–80.
- Thompson, J.M., Jones, H.A. & Perry, R.D. (1999). Molecular characterization of the heme uptake locus (hmu) from *Yersinia pestis* and analysis of hmu mutants for heme and hemoprotein utilization. *Infect Immun*, **67**, 3879–92.
- Tong, Y. & Guo, M. (2007). Cloning and characterization of a novel periplasmic heme-transport protein from the human pathogen *Pseudomonas aeruginosa*. *J Biol Inorg Chem*, **12**, 735–50.
- Torres, A.G. & Payne, S.M. (1997). Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7. *Mol Microbiol*, **23**, 825–33.
- Torres, V.J., Pishchany, G., Humayun, M., Schneewind, O. & Skaar, E.P. (2006). *Staphylococcus aureus* IsdB is a hemoglobin receptor required for heme iron utilization. *J Bacteriol*, **188**, 8421–29.
- Trueblood, K.N., Bürgi, H.B., Burzlaff, H., Dunitz, J.D., Gramaccioli, C.M., Schulz, H.H., Shmueli, U. & Abrahams, S.C. (1996). Atomic displacement parameter nomenclature. Report of a subcommittee on atomic displacement parameter nomenclature. *Acta Crystallogr*, **A52**, 770–81.

- Unno, M., Matsui, T., Chu, G.C., Couture, M., Yoshida, T., Rousseau, D.L., Olson, J.S. & Ikeda-Saito, M. (2004). Crystal structure of the dioxygen-bound heme oxygenase from *Corynebacterium diphtheriae*: implications for heme oxygenase function. *J Biol Chem*, **279**, 21055–61.
- Vallone, B., Nienhaus, K., Brunori, M. & Nienhaus, G.U. (2004). The structure of murine neuroglobin: Novel pathways for ligand migration and binding. *Proteins*, **56**, 85–92.
- Vellieux, F. (1998). Direct methods for solving macromolecular structures. *NATO ASI Series*, **507**, 503–12.
- Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2006). Automated structure solution with autoSHARP. *Methods Mol Biol*, **364**, 215–30.
- Vorobiev, S.M., Neely, H., Seetharaman, J., Ma, L.C., Xiao, R., Acton, T.B., Montelione, G.T. & Tong, L. (2007). Crystal structure of AGR_C_4470p from *Agrobacterium tumefaciens*. *Protein Sci*, **16**, 535–38.
- Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J Mol Graph*, **8**, 52–6.
- Wandersman, C. & Delepelaire, P. (2004). Bacterial iron sources: from siderophores to hemophores. *Annu Rev Microbiol*, **58**, 611–47.
- Wang, B. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods in Enzymology*, **115**, 90–112.
- Wardell, M., Wang, Z., Ho, J.X., Robert, J., Ruker, F., Ruble, J. & Carter, D.C. (2002). The atomic structure of human methemalbumin at 1.9 Å. *Biochem Biophys Res Commun*, **291**, 813–9.
- Weems, J.J. (2001). The many faces of *Staphylococcus aureus* infection. Recognizing and managing its life-threatening manifestations. *Postgrad Med*, **110**, 24–6, 29–31, 35–6.
- Weichsel, A., Andersen, J.F., Champagne, D.E., Walker, F.A. & Montfort, W.R. (1998). Crystal structures of a nitric oxide transport protein from a blood-sucking insect. *Nat Struct Biol*, **5**, 304–9.
- Weiss, M. & Hilgenfeld, R. (1997). On the use of the merging R-factor as a quality indicator for X-ray data. *J Appl Cryst*, **30**, 203–205.
- Wilks, A. (2001). The ShuS protein of *Shigella dysenteriae* is a heme-sequestering protein that also binds DNA. *Arch Biochem Biophys*, **387**, 137–42.
- Wilks, A. & Burkhard, K.A. (2007). Heme and virulence: how bacterial pathogens regulate, transport and utilize heme. *Nat. Prod. Rep.*, **24**, 511–22.

- Wilks, A., Ortiz de Montellano, P.R., Sun, J. & Loehr, T.M. (1996). Heme oxygenase (HO-1): His-132 stabilizes a distal water ligand and assists catalysis. *Biochemistry*, **35**, 930–6.
- Willis, B.T.M. & Pryor, A.W. (1975). *Thermal Vibrations in Crystallography*. Cambridge Univ. Press.
- Winn, M.D., Murshudov, G.N. & Papiz, M.Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol*, **374**, 300–21.
- Wittmann, J.G. & Rudolph, M.G. (2007). Pseudo-merohedral twinning in monoclinic crystals of human orotidine-5'-monophosphate decarboxylase. *Acta Crystallogr*, **D63**, 744–9.
- Wyckoff, E.E., Duncan, D., Torres, A.G., Mills, M., Maase, K. & Payne, S.M. (1998). Structure of the *Shigella dysenteriae* haem transport locus and its phylogenetic distribution in enteric bacteria. *Mol Microbiol*, **28**, 1139–52.
- Wyckoff, E.E., Schmitt, M., Wilks, A. & Payne, S.M. (2004). HutZ is required for efficient heme utilization in *Vibrio cholerae*. *J Bacteriol*, **186**, 4142–51.
- Wyckoff, E.E., Lopreato, G.F., Tipton, K.A. & Payne, S.M. (2005). *Shigella dysenteriae* ShuS promotes utilization of heme as an iron source and protects against heme toxicity. *J Bacteriol*, **187**, 5658–64.
- Wyckoff, E.E., Mey, A.R., Leimbach, A., Fisher, C.F. & Payne, S.M. (2006). Characterization of ferric and ferrous iron transport systems in *Vibrio cholerae*. *J Bacteriol*, **188**, 6515–23.
- Wyckoff, E.E., Mey, A.R. & Payne, S.M. (2007). Iron acquisition in *Vibrio cholerae*. *Biometals*, **20**, 405–16.
- Xu, Z. & Farid, R.S. (2001). Design, synthesis, and characterization of a novel hemoprotein. *Protein Sci*, **10**, 236–49.
- Yeates, T.O. (1997). Detecting and overcoming crystal twinning. *Methods Enzymol*, **276**, 344–358.
- Yi, X., Mroczko, M., Manoj, K.M., Wang, X. & Hager, L.P. (1999). Replacement of the proximal heme thiolate ligand in chloroperoxidase with a histidine residue. *Proc Natl Acad Sci U S A*, **96**, 12412–7.
- Zhang, K.Y.J. & Main, P. (1990). Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Crystallogr*, **A46**, 41–6.
- Zhu, W., Hunt, D.J., Richardson, A.R. & Stojiljkovic, I. (2000). Use of heme compounds as iron sources by pathogenic *Neisseriae* requires the product of the hemO gene. *J Bacteriol*, **182**, 439–47.

Zwart, P., Grosse-Kunstleve, R. & Adams, P. (2005). Xtrriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 newsletter*, **Contribution 7**.

A Appendices

A.1 Amino acid abbreviations (IUPAC)

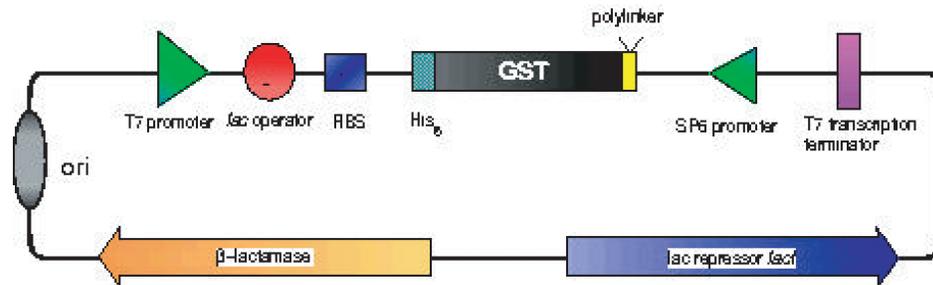
Abbreviation	Amino acid name
Ala	A Alanine
Arg	R Arginine
Asn	N Asparagine
Asp	D Aspartic acid (Aspartate)
Cys	C Cysteine
Gln	Q Glutamine
Glu	E Glutamic acid (Glutamate)
Gly	G Glycine
His	H Histidine
Ile	I Isoleucine
Leu	L Leucine
Lys	K Lysine
Met	M Methionine
Phe	F Phenylalanine
Pro	P Proline
Ser	S Serine
Thr	T Threonine
Trp	W Tryptophan
Tyr	Y Tyrosine
Val	V Valine
Asx	B Aspartic acid or Asparagine
Glx	Z Glutamine or Glutamic acid
Xaa	X Any amino acid
TERM	termination

A.2 List of primer

Name	Plasmid	Sequence (5'-3')	T
SM-HemSfor	pGAT2	ggattagcaggatccatgacgatgactgagagcattcg	55°C
SM-HemSrev	pGAT2	gctacgggtctaagcttaggccgacgacggcttcc	55°C
SM-HemTfor	pGAT2	ggattagcaggatccatgatgaacggcttcgctatcc	49°C
SM-HemTrev	pGAT2	gctacgggtctaagcttactgacgagcattgtctcttcc	49°C
SP-HtsA21for	pGAT2	ggattagcaggatccgtgaatcagaccctaaaac	55°C
SP-HtsA31for	pGAT2	ggattagcaggatccgaacagcagagaattgtagcc	55°C
SP-HtsArev	pGAT2	gctacgggtctaagcttctacggatgatctcccacgtg	55°C
SP-SHPfor	pGAT2	ggattagcaggatccgataaaaggtcaaatttatggatg	58°C
SP-SHPrev	pGAT2	gctacgggtctcgccgccttaccaaaccatacagcaaacc	58°C
SP-SHPfor	pET21a	attagcgacatatggataaaggtcaaatttatgg	58°C
SP-SHPrev	pET21a	gctacgggtctctcaggtcttttttagaccg	58°C
YE-CRPfor	pGAT2	ggattagcaggatccatggttctcgcaagcc	62°C
YE-CRPprev	pGAT2	gctacgggtctaagcttttaacgggtgccgtaaacg	62°C
YE-HemSfor	pGAT2	ggattagcaggatccatgagcaaatcaatatacagac	55°C
YE-HemSrev	pGAT2	gctacgggtctaagcttatgcccgatatccttattatg	55°C
YE-HemS-H75Afor	pSGAT2	cgcaaacctatgccgtagcagagcaaatggccggttac	55°C
YE-HemS-H75Arev	pSGAT2	gtaacggccatttgctctgctacggcataggtgtgctg	55°C
YE-HemS-H196Afor	pSGAT2	cgtgctatgactgacgtggccagttctccagttgctc	55°C
YE-HemS-H196Arev	pSGAT2	gagcaactggaagaactggccacgtcagtcatacagcag	55°C
YE-HemS-M244Afor	pSGAT2	gcaagaacagaatgaaatcgcgattttgtgggtaacctg	55°C
YE-HemS-M244Arev	pSGAT2	cacggttaccacaaaaatcgcgatttcattctgttcttgc	55°C
YE-HemS-H280Afor	pSGAT2	ccagcgcttcacgctggccctgattgaaacaacgattgc	55°C
YE-HemS-H280Arev	pSGAT2	gcaatcgttgttcaatcagggccagcgtgaagcgtgg	55°C
YE-HemTfor	pGAT2	ggattagcaggatccatgagactaaggttactgtcactcc	55°C
YE-HemTrev	pGAT2	gctacgggtctaagcttattgcgcttgttccataccttgcg	55°C
YE-HemT25for	pET22b	ggattagcaccatggcccagaaacgtatcg	61°C
YE-HemTrev-His	pET22b	gctacgggtctctcgagccttgttccatac	61°C

A.3 Plasmid maps

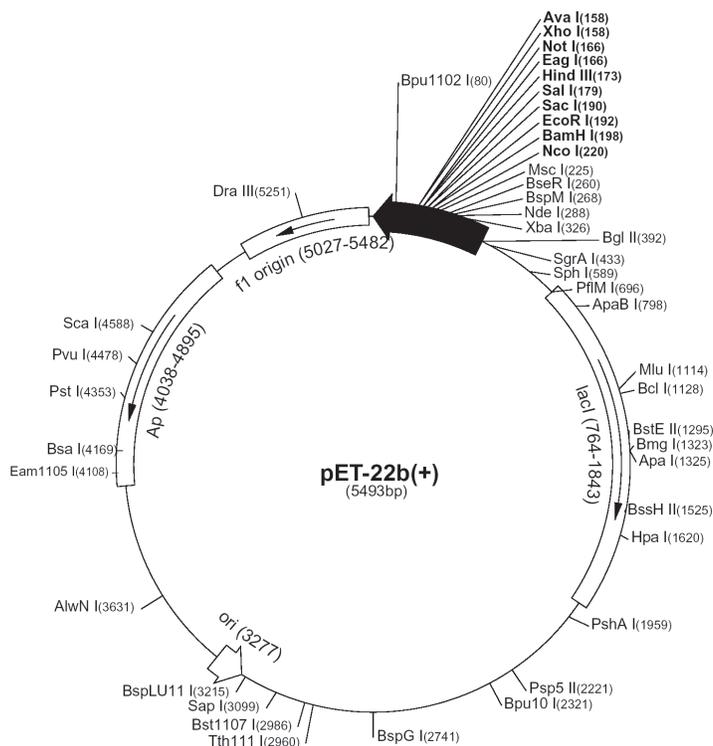
Plasmid map of pGAT2 (Peränen *et al.*, 1996).



Feature	Position (bp)
T7 promoter	1-17
Lac operator	21-42
T7 gene	10
ribosome binding site	50-80
Glutathione-S-transferase gene	83-836
Thrombin cleavage site	815-826
Polylinker	824-864
SP6 promoter	869-887
T7 transcription terminator	966-1007
LacI gene	1417-2498 (reversed)
b-Lactamase gene	3268-4118
Origin of replication	4193-4888

Plasmid map of pET22b.

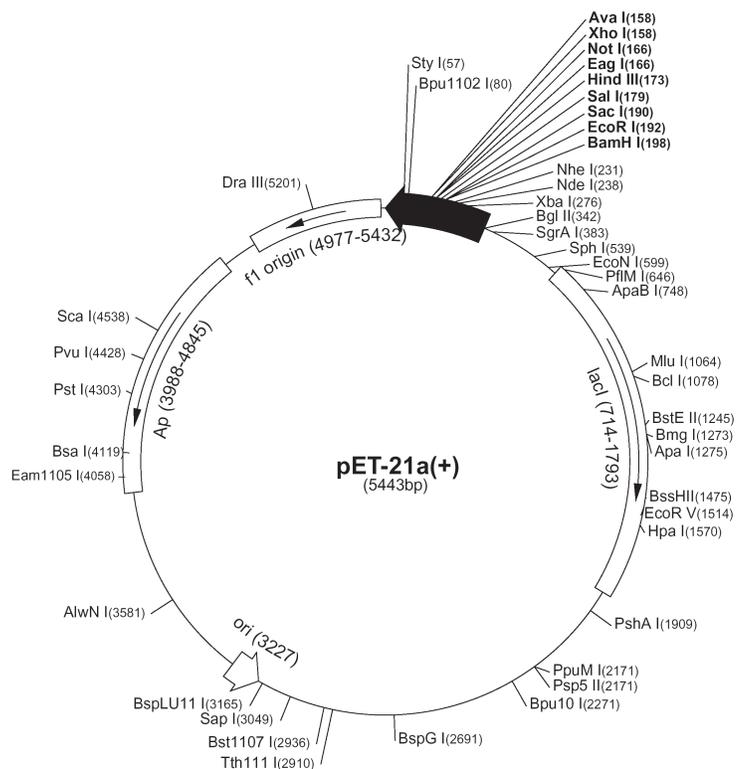
(Novagen, <http://www.merckbiosciences.co.uk/g.asp?f=NVG/home.html>)



Feature	Position (bp)
f1 origin	5027-5482
T7 promoter	361-377
T7 transcription start	360
pelB coding sequence	224-289
Multiple cloning sites	158-220
His-Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	764-1843
pBR322 origin	3277
bla coding sequence	4038-4895
f1 origin	5027-5482

Plasmid map of pET21a.

(Novagen, <http://www.merckbiosciences.co.uk/g.asp?f=NVG/home.html>)



Feature	Position (bp)
f1 origin	4977-5432
T7 promoter	311-327
T7 transcription start	310
Multiple cloning sites	158-203
His-Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	714-1793
pBR322 origin	3227
bla coding sequence	3988-4845
f1 origin	5027-5482

A.4 Haem-propionate contacts in analysed folding topologies.

FOLD	PDB	PROTEIN	ORIGIN	solvent accessibility				propionate contacts (2.5-3.2Å)					
				area	aver.	%	aver.	Arg	His	Lys	Tyr	other	
globin-like	1q1f	neuroglobin	<i>Mus musculus</i>	132		15.9		0	0	2	2	0	
	1o16	neuroglobin	<i>Homo sapiens</i>	90		10.8		0	0	2	2	2	
	1d8u	haemoglobin	<i>Oryza sativa</i>	102		12.3		1	0	0	0	1	
	1vhh	haemoglobin	<i>Vitreoscilla</i>	169		20.4		0	0	1	0	0	
	1a3n	hemoglobin	<i>Homo sapiens</i>	134		16.1		0	1	1	0	1	
	1myt	myoglobin	<i>Thunnus albacares</i>	143	134.8	17.2	16.2	0	1	2	0	1	
	1spe	myoglobin	<i>Physeter catodon</i>	147		17.7		0	2	0	0	1	
	1lhs	myoglobin	<i>Caretta caretta</i>	171		20.6		0	1	0	0	1	
	1v5h/1u0	cytoglobin	<i>Homo sapiens</i>	135		16.3		0	0	2	0	1	
	1cxq	flavo-haemoglobin	<i>Alcaligenes eutrophus</i>	219		26.4		0	0	1	0	2	
	1gvh	flavo-haemoglobin	<i>Escherichia coli</i>	39		4.7		0	1	0	0	2	
	1fsl	leghaemoglobin	<i>Glycine max</i>	136		16.4		0	0	1	0	0	
	cytochrome b5 fold	1eue	cyt b5	<i>Rattus norvegicus</i>	219		26.4		0	0	0	0	2
		1cxy	cyt b5	<i>Ectothiorhodospira v.</i>	103	184.3	12.4	22.2	0	0	2	1	2
1cyo		cyt b5	<i>Bos taurus</i>	216		26.0		0	0	0	0	2	
1sox		sulfite oxidase	<i>Gallus gallus</i>	199		24.0		0	1	0	0	0	
serum albumin-like	1n5u	albumin	<i>Homo sapiens</i>	66		7.9		0	1	1	0	0	
4-helix bundle	1jgc	bacterioferritin	<i>Rhodobacter capsulatus</i>	189		22.8		1	0	0	0	0	
	1qpu	cyt b562	<i>Escherichia coli</i>	148	166.0	17.8	20.0	0	0	0	0	0	
	1b1r	bacterioferritin	<i>Escherichia coli</i>	161		19.4		0	0	0	0	0	
catalase	8cat	catalase	<i>Bos taurus</i>	3		0.4		4	1	0	0	0	
	1qw1	catalase	<i>Helicobacter pylori</i>	15		1.8		4	0	0	0	0	
	1a4e	catalase	<i>Saccharomyces cerevisiae</i>	8		1.0		4	1	0	0	0	
	1m7s	catalase	<i>Pseudomonas</i>	5	11.6	0.6	1.4	4	0	0	0	1	
	1m85	catalase	<i>Proteus</i>	27		3.3		4	0	0	0	0	
	1dgl	catalase	<i>Homo sapiens</i>	2		0.2		3	1	0	0	0	
	1hbz	catalase	<i>Micrococcus</i>	30		3.6		4	0	0	0	1	
	1iph	catalase	<i>Escherichia coli</i>	3		0.4		2	0	0	0	1	
	haem-peroxidase	2cyp	cyt c peroxidase	<i>Saccharomyces cerevisiae</i>	24		2.9		0	3	1	0	2
		1fk	catalase-peroxidase	<i>Haloarcula m.</i>	20		2.4		0	2	1	0	2
2g1j		lactoperoxidase	<i>Bos taurus</i>	61	72.0	7.3	8.7	3	0	0	0	2	
1mwv		catalase-peroxidase	<i>Burkholderia p.</i>	30		3.6		0	2	1	0	2	
1prh		prostaglandin synthase	<i>Seminal vesicl</i>	141		17.0		0	0	0	0	4	
5cox		prostaglandin synthase	<i>Mus musculus</i>	156		18.8		0	0	0	0	4	
P450	2cyp	P450 mono-oxygenase	<i>Pseudomonas putida</i>	28		3.4		3	1	0	0	3	
	1cpt	Cytochrome P450-Terp	<i>Pseudomonas Sp.</i>	17		2.0		1	2	0	0	0	
	1pq2	P450 mono-oxygenase	<i>Homo sapiens</i>	17		2.0		5	2	0	0	2	
	1po5	P450 mono-oxygenase	<i>Oryctolagus cuniculus</i>	23		2.8		3	2	0	0	1	
	1dt6	P450 oxidoreductase	<i>Oryctolagus cuniculus</i>	2		0.2		4	3	0	0	1	
	1odo	P450 mono-oxygenase	<i>Streptomyces coelicolor</i>	16		1.9		4	2	0	1	0	
	1rom	P450 mono-oxygenase	<i>Fusarium oxysporium</i>	10		1.2		1	2	0	0	0	
	1f41	P450 mono-oxygenase	<i>Sulfolobus solfataricus</i>	39	25.5	4.7	3.1	3	1	0	0	1	
	2hpd/1fah	P450 oxidoreductase	<i>Bacillus megaterium</i>	19		2.3		2	0	0	0	1	
	1e9x	P450 α -sterol demethylase	<i>Mycobacterium tuberculosis</i>	51		6.1		2	1	1	1	3	
	1n40	P450 oxidoreductase	<i>Mycobacterium tuberculosis</i>	15		1.8		4	2	0	0	0	
	1fk	P450 oxidoreductase	<i>Amycolatopsis orientalis</i>	29		3.5		3	2	0	0	2	
	1izo	P450 oxidoreductase	<i>Bacillus Subtilis</i>	65		7.8		2	2	1	1	0	
	haem oxygenase	1n45	haem oxygenase	<i>Homo sapiens</i>	247		29.8		1	0	2	1	1
		1j77	haem oxygenase	<i>Neisseria meningitidis</i>	210		25.3		0	0	1	1	0
		1we1	haem oxygenase	<i>Synechocystis</i>	127	184.8	15.3	22.3	4	0	1	1	0
		1iw0/1	haem oxygenase	<i>Corynebacterium diphtheriae</i>	155		18.7		1	0	1	1	0
NO fold	1nos	NO synthase	<i>Mus musculus</i>	159		19.2		1	0	0	0	0	
	3nos	NO synthase	<i>Homo sapiens</i>	122		14.7		0	0	0	1	0	
	1m7z	NO synthase	<i>Bacillus subtilis</i>	102	127.8	12.3	15.4	0	0	0	1	0	
	1qw6	NO synthase	<i>Rattus norvegicus</i>	128		15.4		0	0	0	2	0	
PAS	1drm	FixL	<i>Bradorhyzobium japonicum</i>	40	105.0	4.8	12.7	2	4	0	0	1	
	1v9y/z	EcDos	<i>Escherichia coli</i>	170		20.5		0	0	0	0	0	
lipocalin	1np4	nitrophorin	<i>Rhodnius prolixus</i>	130		15.7		0	0	2	0	1	
meander	1dk0	hasA	<i>Serratia marcescens</i>	191		23.0		0	1	0	1	4	
β -propeller	1qhu	haemopexin	<i>Oryctolagus cuniculus</i>	210		25.3		2	1	0	3	0	
cupredoxin-like	1fft	ubiquinol oxidase	<i>Escherichia coli</i>	32		3.9		3	0	0	0	1	
immunoglobulin-like	1d7b	cellobiose dehydrogenase	<i>Phanerochaete chrysosporium</i>	179		21.6		0	0	0	2	3	
CAP	1f19	CooA	<i>Rhodospirillum rubrum</i>	233		28.1		0	0	0	0	1	
vitamine B6 family	1jbc	cystathione beta-synthase	<i>Homo sapiens</i>	148		17.8		4	0	0	0	0	
H-NOX / SONO fold	1u55	quanylate cyclase	<i>Thermoanaerobacter tengcongensis</i>	6	18.5	0.7	2.2	4	0	0	0	0	
	1xbn	SONO	<i>Thermoanaerobacter tengcongensis</i>	31		3.7		3	0	0	1	2	
HemS-fold	2j0p	HemS	<i>Yersinia enterocolitica</i>	155		18.7		4	0	1	1	1	
dioxygenase-fold	2d0t	indoleamine 2,3-dioxygenase	<i>Homo sapiens</i>	92		11.1		3	0	0	0	1	
Sum of contacts								103	46	28	24	67	
total number of contacts								268					
PERCENTAGE								38.4	17.2	10.4	9.0	25.0	
PERCENTAGES FOLDS								propionate contacts (2.5-3.2Å)					
								Arg	His	Lys	Tyr	other	
globin-like								2.9	17.1	34.3	11.4	34.3	
cytochrome b5 fold								0.0	10.0	20.0	10.0	60.0	
serum albumin-like								0.0	50.0	50.0	0.0	0.0	
4-helix bundle								100.0	0.0	0.0	0.0	0.0	
catalase								82.9	8.6	0.0	0.0	8.6	
haem peroxidase								10.3	24.1	10.3	0.0	55.2	
cytochrome P450								47.4	28.2	2.6	3.8	17.9	
haem oxygenase fold								37.5	0.0	31.3	25.0	6.3	
NO fold								20.0	0.0	0.0	80.0	0.0	
PAS								28.6	57.1	0.0	0.0	14.3	
lipocalin								0.0	0.0	66.7	0.0	33.3	
meander								0.0	16.7	0.0	16.7	66.7	
β -propeller								33.3	16.7	0.0	50.0	0.0	
cupredoxin-like								75.0	0.0	0.0	0.0	25.0	
immunoglobulin-like								0.0	0.0	0.0	40.0	60.0	
CAP								0.0	0.0	0.0	0.0	100.0	
vitamin B6 family								100.0	0.0	0.0	0.0	0.0	
H-NOX-fold / SONO-fold								70.0	0.0	0.0	10.0	20.0	
HemS-fold								57.1	0.0	14.3	14.3	14.3	
dioxygenase fold								75.0	0.0	0.0	0.0	25.0	

A.5 Residues in contact with the haem-face and haem-edge.

FOLD	PDB	PROTEIN	ORIGIN	haem edge (3.6-4.1A)										haem face (3.6-4.1A)																											
				Ala	Ile	Leu	Phe	Trp	Tyr	Val	other	Ala	Ile	Leu	Phe	Trp	Tyr	Val	other																						
globin-like	1q1f	neuroglobin	Mus musculus	0	1	3	0	0	1	3	1	0	0	1	3	0	0	2	1	0	1	3	0	0	1	3	0	0	2	1											
	1o9j	neuroglobin	Homo sapiens	0	0	3	0	0	0	3	1	0	0	1	2	0	0	1	2	0	0	2	0	0	1	2	0	0	2	0											
	1d8u	haemoglobin	Oryza sativa	1	1	1	1	0	2	0	4	0	0	1	2	0	1	2	0	1	3	0	1	1	1	1	0	2	0	1	3	0									
	1vnb	haemoglobin	Vitreoscilla	0	1	1	1	0	0	0	2	0	1	1	1	0	1	1	0	1	3	0	0	1	1	1	0	0	2	0	1	3	0								
	1a3n	hemoglobin	Homo sapiens	1	0	2	0	0	1	0	4	0	0	3	2	0	0	3	0	0	3	0	1	0	2	0	0	1	0	4	0	3	0								
	1myt	myoglobin	Thunnus albacares	0	1	1	0	0	0	0	2	0	1	1	2	0	0	1	2	0	0	1	2	0	1	1	0	0	0	2	0	1	2	0							
	1spe	myoglobin	Physeter catodon	1	1	1	1	0	1	0	2	0	1	1	1	0	1	1	0	0	1	2	1	1	1	1	0	1	1	0	0	1	2								
	1fhs	myoglobin	Caretta caretta	1	1	0	0	0	1	0	0	0	1	2	1	0	0	1	2	1	0	0	1	1	0	0	1	1													
	1v5h/1ut0	cytoglobin	Homo sapiens	2	0	2	0	0	1	0	0	0	0	1	2	0	0	1	2	0	0	3	2	0	0	0	0	0	1	2	0	0	3	2							
	1oqx	flavo-haemoglobin	Alcaligenes eutrophus	1	0	1	0	0	0	1	2	0	1	2	1	0	1	2	1	0	1	2	0	1	0	1	0	1	2	0	1	2	0								
	1qvh	flavo-haemoglobin	Escherichia coli	2	0	2	2	0	2	0	2	1	3	1	1	1	0	1	1	0	1	1	0	2	0	2	2	0	2	1	1	1	1	0							
	1ftr	leghaemoglobin	Glycine max	1	0	2	1	0	1	1	1	1	1	0	2	0	0	2	0	0	1	1	1	1	0	2	1	0	1	1	1	1	1	1							
	cytochrome b5	1eue	cyl b5	Rattus norvegicus	2	2	4	1	0	0	0	2	1	0	3	2	0	0	3	2	0	0	0	4	2	2	4	1	0	0	0	2	1	0	3	2	0	0	0	4	
		1cxy	cyl b5	Ectothiorhodospira v.	2	0	4	0	2	1	0	3	1	0	0	0	1	1	0	0	1	1	0	5	2	0	4	0	2	1	0	3	1	0	0	0	1	1	0	5	
		1cyp	cyl b5	Bos taurus	2	0	4	2	0	0	2	4	1	0	1	2	0	0	1	2	0	0	2	4	2	0	4	2	0	0	2	4	1	0	1	2	0	0	2	4	
1sox		sulfite oxidase	Gallus gallus	1	1	4	2	1	1	2	1	0	0	0	0	2	1	1	1	1	1	1	5	1	1	4	2	1	1	2	1	0	0	2	1	1	1	1	5		
serum albumin-like	1n5u	albumin	Homo sapiens	1	0	3	3	0	0	0	1	3	0	1	0	1	0	1	0	1	0	1	2	1	0	3	3	0	0	0	1	3	0	1	0	1	0	2			
	1lqc	bacterioferritin	Rhodobacter capsulatus	0	1	1	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1	0	1	0	0	0	0	1	0	1	0	1			
4-helix bundle	1qpu	cyl b5e2	Escherichia coli	0	0	2	0	0	0	0	4	0	0	2	0	0	2	0	0	0	0	0	2	0	0	2	0	0	0	4	0	0	2	0	0	0	0	2			
	1bfr	bacterioferritin	Escherichia coli	1	1	1	0	0	1	0	1	0	1	0	2	0	1	0	2	0	1	0	0	0	1	1	1	0	0	1	0	1	0	0	0	0	0	0	0		
catalase	8cat	catalase	Bos taurus	2	0	0	2	0	0	0	3	0	0	0	1	0	0	0	1	0	0	2	2	2	0	0	2	0	0	3	0	0	1	0	0	2	2				
	1qwl	catalase	Helicobacter pylori	1	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0	1	0	0	2	3	1	0	0	0	0	0	3	0	0	0	1	0	0	2	3		
	1e4e	catalase	Saccharomyces cerevisiae	1	0	2	1	0	1	1	4	0	0	0	0	1	0	0	1	0	0	1	4	1	0	2	1	0	1	1	0	0	1	0	0	1	4				
	1m7a	catalase	Pseudomonas	1	0	2	0	0	0	2	7	0	0	0	0	2	0	0	2	0	0	2	1	4	1	0	2	0	0	0	2	7	0	0	2	0	0	2	1		
	1m85	catalase	Proteus	1	0	1	1	0	0	0	5	0	0	0	1	0	0	1	0	0	0	1	3	1	0	1	1	0	0	0	5	0	0	1	0	0	1	3			
	1dof	catalase	Homo sapiens	2	0	1	1	0	0	1	8	0	0	0	1	0	0	1	0	0	1	0	3	2	2	0	1	1	0	0	1	8	0	0	1	0	0	3	2		
	1hzb	catalase	Micrococcus	3	0	1	1	0	0	0	6	0	0	0	0	1	0	0	1	0	0	1	4	3	0	1	1	0	0	0	6	0	0	1	0	0	1	4			
	1tjh	catalase	Escherichia coli	0	1	1	1	0	0	0	6	0	0	0	1	0	0	1	0	0	0	2	1	0	1	1	1	0	0	0	6	0	0	1	0	0	2	1			
	haem-peroxidase	2cyp	cyl c peroxidase	Saccharomyces cerevisiae	1	0	2	2	0	0	1	2	1	0	1	0	2	0	0	1	0	2	0	3	1	0	2	2	0	0	1	2	1	0	0	2	0	3			
		1tk	catalase-peroxidase	Halobaculum	0	4	0	3	1	0	1	5	0	0	1	0	1	0	2	0	1	0	1	3	0	4	0	3	1	0	1	5	0	0	1	0	2	0	1	3	
2aj1		lactoperoxidase	Bos taurus	0	0	2	1	0	0	1	2	0	0	1	0	0	1	0	0	0	0	0	6	0	0	2	1	0	0	1	2	0	0	0	0	0	6				
1mav		catalase-peroxidase	Burkholderia p.	0	3	2	2	1	0	1	4	0	0	1	0	1	0	2	0	0	0	0	4	0	3	2	2	1	0	1	4	0	0	1	0	2	0	0	4		
1pgh		prostaglandin synthase	Seminal vesicl	0	0	4	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	4	1	1	1	0	0	0	0	0	0	0	0	1	5	
5cox		prostaglandin synthase	Mus musculus	0	0	1	2	1	1	1	0	0	0	1	0	0	0	1	0	0	0	0	1	4	0	0	1	2	1	1	1	0	0	0	1	0	0	0	1	4	
P450	2cyp	P450 mono-oxygenase	Pseudomonas putida	1	0	4	2	0	0	2	4	0	0	0	0	0	1	0	0	0	0	1	4	1	0	4	2	0	0	2	4	0	0	0	1	0	0	1	4		
	1cpt	Cytochrome P450-Terp	Pseudomonas Sp.	0	1	1	1	0	0	1	1	1	0	1	2	0	0	1	2	0	0	0	1	3	0	1	1	1	0	0	1	1	1	0	0	1	3				
	1p02	P450 mono-oxygenase	Homo sapiens	0	0	3	0	0	0	0	1	1	3	0	1	1	0	0	0	0	0	0	4	0	0	3	0	0	0	0	1	1	3	0	1	0	0	0	4		
	1p05	P450 mono-oxygenase	Oryctolagus cuniculus	0	1	1	0	0	0	0	3	2	1	1	1	0	0	0	0	0	0	0	1	4	0	1	1	0	0	0	0	3	2	1	1	0	0	0	1	4	
	1d06	P450 oxidoreductase	Oryctolagus cuniculus	0	2	3	0	0	0	0	3	2	0	1	1	0	0	1	1	0	0	0	1	3	0	2	3	0	0	0	0	3	2	0	1	1	0	0	1	3	
	1odo	P450 mono-oxygenase	Streptomyces coelicolor	2	0	1	1	0	1	1	3	2	0	1	1	0	0	1	1	0	0	0	3	2	0	1	1	0	1	1	3	2	0	1	1	0	0	3			
	1rom	P450 mono-oxygenase	Fusarium oxysporum	2	1	1	2	0	0	0	2	2	1	0	1	0	1	0	0	0	0	0	4	2	1	1	2	0	0	0	2	2	1	0	1	0	0	0	4		
	1f41	P450 mono-oxygenase	Sulfolobus solfataricus	1	2	2	1	0	0	0	4	1	0	1	1	1	0	0	1	1	0	0	1	3	1	2	2	1	0	0	0	4	1	0	1	1	0	0	1	3	
	2hpd/1fah	P450 oxidoreductase	Bacillus megaterium	0	0	2	3	0	0	0	2	3	1	0	2	0	0	0	0	0	0	0	4	0	0	2	3	0	0	0	2	3	1	0	2	0	0	0	0	4	
	1efrx	P450 α -steroid demethylase	Mycobacterium tuberculosis	0	0	2	1	0	0	0	4	2	0	1	1	0	0	1	1	0	0	0	1	3	0	0	2	1	0	0	0	4	2	0	1	1	0	0	1	3	
	1n40	P450 oxidoreductase	Mycobacterium tuberculosis	1	0	3	2	0	0	0	3	1	0	0	2	0	0	0	0	0	0	0	1	4	1	0	3	2	0	0	0	3	1	0	0	2	0	0	0	1	4
	1tk	P450 oxidoreductase	Amycolopsis orientalis	1	1	4	1	0	0	1	5	2	0	1	1	0	0	1	1	0	0	0	3	1	1	4	1	0	0	1	5	2	0	1	1	0	0	0	3		
	1tzo	P450 oxidoreductase	Bacillus Subtilis	1	2	2	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	4	1	2	2	3	0	0	0	1	0	0	0	0	0	0	0	0	4	
	haem oxygenase	1n45	haem oxygenase	Homo sapiens	1	0	0	2	0	1	1	8	0	0	1	1	0	0	1	1	0	0	0	5	1	0	0	2	0	1	1	8	0	0	1	1	0	0	0	5	
		1f77	haem oxygenase	Neisseria meningitidis	0	0	1	3	0	2	2	5	0	0	0	0	1	0	0	0	0	0	1	5	0																

B List of publications

Schneider,S., Marles-Wright,J., Sharp,K. & Paoli,M. (2007). Diversity and conservation of interactions for binding heme in b-type heme proteins. *Natural Product Reports*, **24**, 621-30.

Sharp,K., Schneider,S., Cockayne,A. & Paoli,M. (2007). Crystal structure of the heme-IsdC complex, the central conduit of the Isd iron/heme uptake system in *Staphylococcus aureus*. *J Biol Chem*, **282**, 10625-31.

Schneider,S., Carneiro,M.J.V.M., Ioannou,C., Sultanas,P. & Paoli,M. (2007). Crystallization and X-ray diffraction analysis of the DNA-remodelling protein DnaD from *Bacillus subtilis*. *Acta Crystallogr*, **F63**, 1-4.

Schneider,S., Sharp,K., Barker,P.D. & Paoli,M. (2006). An induced fit conformational change underlies the binding mechanism of the heme-transport Proteobacteria-protein HemS. *J Biol Chem*, **281**, 32606-10.

Schneider,S. & Paoli,M. (2005). Crystallization and preliminary X-ray diffraction analysis of the haem-binding protein HemS from *Yersinia enterocolitica*. *Acta Crystallogr*, **F61**, 802-5.