



<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study,
without prior permission or charge

This work cannot be reproduced or quoted extensively from without first
obtaining permission in writing from the author

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

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

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

ProQuest Number: 10390435

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10390435

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

**Structural Studies on
Histidine Metabolism
in Microorganisms**

by

Bernhard Lohkamp



**UNIVERSITY
of
GLASGOW**

A thesis submitted for the degree of Doctor of Philosophy
in the Faculty of Biomedical & Life Sciences,
Division of Biochemistry and Molecular Biology
& Protein Crystallography,
University of Glasgow

September 2003

GLASGOW
UNIVERSITY
LIBRARY:

13294

COPY 2

Declaration

This thesis has been written in accordance with the University regulations, has not been presented for a degree at any other university and is original except where indicated otherwise by reference in the text. The work contained within is the author's own except where work was done in collaboration as indicated.

© Bernhard Lohkamp, 30th September 2003.

Signed

Date

.....

.....

Bernhard Lohkamp

30.09.2003

Acknowledgements

My thanks go to a great number of people who have provided assistance, advice as well as 'a life' throughout this project.

In particular I thank my supervisors John Coggins and Adrian Laphorn for their enthusiasm and support although still letting me 'get on with it'. Adrian deserves special thanks for buying beers and being right 50% of the time. From the lab in D-floor, I am indebted to Samantha Campbell for providing PR-ATP and John Greene for initial help in protein purification and the occasional entertainment (even including the phone calls after the Old Firm matches). I wish to thank Lewis 'Pieman' Evans for the enjoyment in and outside of 'our' lab and for providing an exceptional target for me to kick some a**e. I would also like to acknowledge the contributions of 'my' students Torsten Schweiker, Fiona MacIsaac and most of all Neil Paterson, who apart from 'winning again', produced high quality crystals about which I do not want to complain. I am grateful to Gerry McDermott for the MAD data collection at the ALS and invitation to visit Berkeley, and James Foadi for useful discussion regarding ACORN.

I owe much to all the (ex-)members of the Protein Crystallography Group, especially Neil 'Bo' Hanlon, Neil 'the boy' Paterson, Paul 'Emsleyman' Emsley, 'Comedy' Dave Robinson, Alastair 'Ally' McEwen and Alan Riboldi-Tunnickliffe. Thanks for memorable synchrotron trips, data collection, advice, theoretical discussions, beers, football, ski clubbing, fun, various chats, and of course crafting to waste non-existing spare time (my thesis could have been finished so much earlier...).

The most recent thanks go to my former colleagues at PanTherix: First to Sohail Ali, for managing to teach 'a physicist' molecular biology and proofreading parts of the manuscript. Second to my bosses, Bill Primrose and John Barker, who supported me wherever possible during the writing of this thesis. And finally everyone else at PanTherix, especially the structural biology team, for further education and having a unforgettable time despite the difficult circumstances.

I like to acknowledge gratefully Nick Price for proofreading this thesis and converting my English into proper English. Additionally thanks for his help with the kinetics data.

The Wellcome Trust is recognised for funding, which gave me the opportunity to come back to Glasgow and certainly paid for my beers. Dave Barry and Bill Cushley, as organisers of the Wellcome 4-year programme, deserve a special mention particularly for always lending an open ear and giving advice.

I thank my friends all over the world for supporting me, giving me memorable times and putting up with and standing by me through difficult periods, especially during the last 5 years. This goes in particular to my sister Monika.

And last, but not least, I would like to thank my parents for supporting me all my life and especially during my university years.

Abbreviations

AICAR	Aminoimidazole carboxamide ribonucleotide
AMP	Adenosine monophosphate
asu	asymmetric unit
ATP	Adenosine triphosphate
rATP	ribo-adenosine triphosphate
BS	<i>Bacillus subtilis</i>
CIAP	Calf intestine alkaline phosphatase
Da	Dalton
dNTP	deoxyribo-nucleotide triphosphate
DMF	dimethylformamide
dsH ₂ O	distilled sterile water
DLS	Dynamic Light Scattering
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra acetate
EM	Electron microscopy
FFT	Fast Fourier transform
FoM	Figure of Merit
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
Hin	<i>Haemophilus influenza</i>
HisG	ATP-phosphoribosyltransferase
hut	Histidine utilisation
IPTG	Isopropyl- β -D-thiogalactoside
LB	Luria-Bertani medium
MAD	Multiple anomalous dispersion
MIR	Multiple isomorphous replacement
MIRAS	Multiple isomorphous replacement with anomalous scattering
MR	Molecular replacement
Mtb	<i>Mycobacterium tuberculosis</i>
Mth	<i>Methanobacterium thermoautotrophicum</i>
MW	Molecular weight
NTP	Nucleotide tri phosphate

ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Poly-ethylene-glycol
PR-ATP	<i>N</i> ⁵ '-phosphoribosyl-ATP
PRPP	5-phosphoribosyl 1-pyrophosphate
PRT	Phosphoribosyltransferase
RBS	Ribosome binding site
rmsd	root mean square deviation
SAD	Single anomalous dispersion
SDS	Sodium dodecyl sulphate
Se-Met	seleno-methionine
SIRAS	Single isomorphous replacement with anomalous scattering
SnB	Shake'n'Bake
spp	species (plural)
TAE	Tris-Acetate-EDTA
TBE	Tris-borate-EDTA
TEMED	<i>N, N, N', N'</i> -tetramethyl-ethylenediamine
Tris	Tris (hydroxymethyl)aminomethane
WT	Wild type
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
ϵ	Extinction coefficient

Summary

Histidine is an important amino acid involved in enzyme catalysis and metal binding in many proteins. Bacteria, fungi and plants synthesise histidine via the 10 step histidine pathway, whereas histidine is an essential amino acid required in the diet of animals. The synthesis of histidine is energetically expensive with about 41 ATP needed to produce one histidine. As a result the histidine biosynthetic pathway is tightly regulated and some microorganisms can consume excess histidine as a nitrogen and carbon source via a histidine utilisation (hut) pathway. The absence of a histidine biosynthetic pathway in mammals makes it a good target for the development of antimicrobial and antifungal drugs as well as herbicides.

In addition to an overall regulation of the histidine synthesis by charged histidyl-tRNA and histidyl-tRNA synthetase, the histidine pathway is regulated in various ways by the first enzyme of the pathway ATP-phosphoribosyltransferase (HisG). The enzyme catalyses the condensation of ATP and phosphoribosylpyrophosphate (PRPP) to form phosphoribosyl-ATP (PR-ATP) and is feedback regulated by histidine, the end product of the pathway. Furthermore, HisG is inhibited by its product PR-ATP and also by AMP and ppGpp, reflecting the overall energy status of the cell. Inhibition of HisG coincides with a change in aggregation state, in which the active dimers form inactive hexamers. Some bacterial genomes contain a *hisG* gene, which lacks about 80 residues at the C-terminus, but have an additional gene, *hisZ*. This short form of HisG itself is inactive and forms an active complex with HisZ.

This thesis describes the solution of the 3-dimensional structures of the hexameric forms of bacterial HisGs by X-ray crystallography. The structure of the *Escherichia coli* HisG was determined in the presence of the inhibitors AMP and PR-ATP. Additionally, a histidine-bound structure of HisG from the thermophilic bacterium *Methanobacterium thermoautotrophicum* has been obtained. The monomer of HisG comprises of three α/β domains. The first two domains show a periplasmic protein-like fold with the active site located in a cleft between the two domains. The C-terminal domain, similar to P_{II} signal transduction proteins, binds histidine and is

mainly involved in the formation of the hexamers. Additionally the structure of HisG represents a new type of phosphoribosyltransferase. The structure of the binary complex with AMP has allowed the identification of the PRPP binding site and explains the competitive inhibition by AMP with respect to both substrates. Furthermore the AMP binary structure and the PR-ATP-complexed structure identified the binding site of the ATP purine ring. The histidine-bound structure explains the histidine inhibition by disruption of the PRPP binding site on the dimer interface by elongation of the dimer and hence the hexamer. A mode of inhibition common to all the inhibitors investigated is the formation of hexamers upon inhibitor binding which buries the active sites in the inside of the hexamer and makes the sites less accessible to the substrates.

Mutational and kinetic analysis showed that the extra C-terminal domain in the long form of HisG is not essential for activity. Subsequently, it was demonstrated that the naturally short form of HisG can be activated by thermal energy. This suggests a tense conformation for the short form of the enzyme which can be relaxed and activated by forming an oligomeric complex with HisZ.

The structure of an unidentified protein in the histidine utilisation (hut) pathway from *Pseudomonas aeruginosa*, PA5104, was determined to better than 1 Å resolution. Solution of the structure was accomplished by a unique approach combining SIRAS phasing with direct methods (Sayre equation). PA5104 is an all- β protein with a bicupin fold. One of the β -barrels is open and accommodates a small molecule binding site. A different conformation of PA5104 was obtained from a second structure solved using a different crystal form. Comparison of the structures and interactions in the active site identifies glutamate or derivatives (intermediates in the hut pathway) as possible substrates, which can induce a conformational change and may give PA5104 a regulatory function.

Table of Contents

Chapter 1: Introduction	1
1.1 Histidine Biosynthesis – General Introduction	1
1.2 The Histidine Biosynthetic Pathway	4
1.3 Individual Enzymes of the Histidine Pathway	8
1.3.1 ATP-phosphoribosyltransferase (HisG or ATP-PRT)	8
1.3.2 Phosphoribosyl-ATP pyrophosphatase, phosphoribosyl-AMP cyclohydrolase (HisIE)	12
1.4 Histidine degradation	13
1.4.1 Introduction	13
1.4.2 Histidine utilisation pathways	14
1.4.3 Organisation of operons and control of expression	15
1.5 Aims of the project	17
Chapter 2: Materials and Methods	19
2.1 Materials and Methods	19
2.1.1 Chemicals	19
2.1.2 Proteins and enzymes	19
2.1.3 Oligonucleotides	19
2.1.4 Chromatography media	19
2.2 Media and supplements	20
2.2.1 Media for bacterial growth	20
2.2.2 Media to produce seleno-methionine substituted protein	20
2.2.3 Antibiotics	21
2.2.4 Buffers and solutions for molecular biology experiments	21
2.2.5 Isopropyl- β -D-thiogalactopyranoside (IPTG)	24
2.2.6 X-gal LB agar plates	24
2.3 General methods	24
2.3.1 General laboratory methods	24
2.3.2 French pressure cell	24
2.3.3 pH measurement	24
2.3.4 Determination of protein concentration	24
2.4 Bacterial strains and cloning vectors	25
2.4.1 Storage of Bacterial Strains	25

2.4.2 Bacterial strains.....	25
2.4.3 Growth of plasmid-containing <i>E. coli</i>	28
2.4.4 Growth of cells and isolation of plasmids.....	28
2.4.5 Production of competent cells.....	28
2.4.6 Transformation protocol.....	29
2.5 Manipulations of DNA.....	30
2.5.1 Gel electrophoresis.....	30
2.5.2 Restriction digests.....	30
2.5.3 Gel purification of DNA fragments.....	30
2.5.4 Alkaline phosphatase dephosphorylation.....	31
2.5.5 DNA clean up methods.....	31
2.5.6 DNA ligation.....	31
2.6 Cloning by PCR.....	32
2.6.1 Primers for PCR.....	32
2.6.2 Genomic DNA.....	33
2.6.3 PCR reactions.....	33
2.7 Over-expression analysis.....	38
2.7.1 Over-expression studies.....	38
2.7.2 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE).....	38
2.7.3 Staining for Protein.....	38
2.8 Protein Purification.....	39
2.8.1 Growth of cells for protein purification.....	39
2.8.2 Preparation of Cell Extract.....	39
2.8.3 Purification of <i>E. coli</i> HisG.....	39
2.8.4 Purification of HisG from <i>Mth</i>	39
2.8.5 Purification of His _x -tagged proteins.....	40
2.8.6 Storage of purified enzymes.....	41
2.9 Enzyme assay.....	41
2.9.1 Assay for ATP-Phosphoribosyltransferase (HisG).....	41
2.9.2 Assay for short HisG and HisZ.....	42
2.9.3 HisIE enzyme activity.....	42
2.10 Crystallisation.....	43
2.10.1 Dynamic light scattering (DLS).....	43
2.10.2 Crystallisation screens.....	43

2.10.3 Crystallisation methods.....	44
Chapter 3: Cloning, Over-expression, and Mutational Analysis of HisG, HisIE and HisZ from Multiple Microorganisms	47
3.1 Introduction.....	47
3.2 Generation of over-expression constructs from <i>B. subtilis</i> , <i>H. influenzae</i> and <i>M. tuberculosis</i>	48
3.2.1 Introduction.....	48
3.2.2 T7 expression vectors.....	48
3.2.3 Amplification of <i>his</i> genes by PCR.....	49
3.2.4 Cloning of PCR products	50
3.2.5 ECHO™ cloning.....	52
3.2.6 'New' cloning strategy.....	54
3.2.7 Generation of new over-expression vectors.....	54
3.2.8 Cloning of <i>his</i> genes into pTBL1 and pTBL2.....	56
3.3 Generation of over-expression constructs from <i>M. thermoautotrophicum</i> ..	57
3.3.1 Introduction.....	57
3.3.2 pTB361 plasmid.....	57
3.3.3 Amplification of <i>hisG</i> and <i>HisG-related protein</i> genes by PCR	57
3.3.4 Cloning of PCR product into pGEM.....	58
3.3.5 Cloning into pTB361	58
3.3.6 Generation of BL21 and C41 <i>E. coli</i> strains to increase over-expression levels of genes containing rare codons	58
3.4 Generation of mutants of HisG from <i>H. influenzae</i>	59
3.4.1 Introduction.....	59
3.4.2 Generation of deletion mutants	60
I. Appendix: Schematic of vector pUni/V5-His-TOPO	62
II. Appendix: Schematic of vector pCRT7-E.....	62
III. Appendix: Schematic of control vector pCRT7-E/Uni-CAT.....	62
Chapter 4: Over-expression, Purification, Crystallisation and Characterisation of Enzymes.....	63
4.1 Over-expression and purification of his-tagged enzymes from the histidine pathway.....	63
4.1.1 Over-expression	63

4.1.2 Purification.....	64
4.2 Over-expression and purification of HisG and HisG-related enzyme from <i>M thermoautotrophicum</i>	65
4.2.1 Over-expression	65
4.2.2 Purification of HisG	67
4.2.3 Resolubilisation, refolding and purification of HisG-related protein	68
4.3 Purification of PA5104	68
4.4 Dynamic light scattering (DLS).....	69
4.5 Crystallisation	71
4.5.1 Crystallisation of <i>E. coli</i> HisG	71
4.5.2 Crystallisation of <i>BS</i> histidine pathway enzymes	73
4.5.3 Crystallisation of <i>Mth</i> HisG	74
4.5.4 Crystallisation of PA5104.....	74
4.6 Kinetic properties of HisG.....	75
4.6.1 <i>Hin</i> HisG and truncation mutants.....	75
4.6.2 Thermophilic HisG from <i>Mth</i>	76
4.6.3 Short HisG from <i>BS</i>	77
4.7 Kinetics Discussion.....	78
Chapter 5: X-ray Data Collection, Processing and Scaling	81
5.1 Data collection and processing	81
5.1.1 HisG from <i>E. coli</i>	81
5.1.2 HisG from <i>Mth</i>	86
5.1.3 PA5104 from <i>Pseudomonas aeruginosa</i>	87
5.2 Calculation of structure factors from intensities	98
Chapter 6: Structure of HisG from <i>E. coli</i>	99
6.1 Introduction.....	99
6.2 Solving the phases of <i>E. coli</i> HisG	99
6.2.1 MAD phasing.....	99
6.2.2 Model building and refinement.....	100
6.3 HisG model	101
6.3.1 Quality of the model.....	101
6.3.2 Overall structure.....	102
6.3.3 The HisG dimer and hexamer	107
6.3.4 AMP binding.....	109

6.4 HisG with PR-ATP	110
6.4.1 Refinement	111
6.4.2 Quality of model	111
6.4.3 PR-ATP binding.....	113
6.5 Discussion	115
6.5.1 AMP inhibition at the PRPP binding site.....	115
6.5.2 Substrate binding sites	117
6.5.3 Reaction mechanism and inhibition of HisG	117
6.5.4 HisG is a new class of PRTs: Extending the PRT superfamily	118
6.5.5 The C-terminal ‘regulatory’ domain	121
Chapter 7: Structure Solution of HisG from <i>Mth</i>	124
7.1 Solving the phases of <i>Mth</i> HisG	124
7.1.1 Molecular replacement (MR) using <i>E. coli</i> HisG as a search model ...	124
7.1.2 MR solution with <i>Mtb</i> HisG as search model	125
7.1.3 Model building and refinement.....	126
7.2 <i>Mth</i> HisG model.....	128
7.2.1 Quality of the model.....	128
7.2.2 Overall structure.....	128
7.2.3 The two monomers.....	129
7.2.4 Histidine binding.....	131
7.3 Discussion	133
7.3.1 Histidine binding.....	133
7.3.2 Quaternary structure.....	135
Chapter 8: Comparison of HisGs	136
8.1 Significance of the disulfide bridge in <i>Mtb</i> HisG	136
8.2 AMP binding in HisG	137
8.3 A ‘regulatory’ loop in HisG	140
8.4 Changes in the quaternary structure of HisG.....	143
8.4.1 The HisG hexamer	143
8.4.2 The HisG monomer and dimer.....	147
Chapter 9: Structure of PA5104 from <i>Pseudomonas aeruginosa</i>	152
9.1 Solving phases of PA5104	152
9.1.1 Attempts to solve phases.....	152
9.1.2 SIR(AS) phasing and solving of phases.....	157

Table of Contents

9.1.3 Model building and refinement.....	158
9.2 PA5104 model	158
9.2.1 Quality of model	158
9.2.2 Overall structure.....	160
9.3 Structure of 2 nd crystal form of PA5104	163
9.3.1 Molecular replacement.....	163
9.3.2 Model building and refinement.....	164
9.3.3 Quality of the model.....	164
9.3.4 The structure of PA5104 from a 2 nd crystal form	165
9.4 Discussion	167
9.4.1 (Problems) solving the structure	167
9.4.2 Comparison of structures from the two crystal forms.....	169
9.4.3 Active site and function	171
Chapter 10: Conclusions	177
10.1 HisG	177
10.1.1 Substrate binding sites and reaction mechanism.....	177
10.1.2 Mode(s) of inhibition	178
10.1.3 The C-terminal 'regulatory' domain	179
10.2 PA5104	180
References	182

Table of Figures

Figure 1-1 Histidine biosynthetic pathway.....	6
Figure 1-2 Substrates and intermediates of the histidine pathway.	7
Figure 1-3 Reaction scheme of HisG.	8
Figure 1-4 Scheme of ordered mechanism of HisG.	10
Figure 1-5 Reaction scheme of HisIE.	12
Figure 1-6 Histidine utilisation pathway 1.	14
Figure 1-7 Histidine utilisation pathway 2.	15
Figure 1-8 <i>Hut</i> operon of <i>Pseudomonas putida</i>	16
Figure 3-1 Schematic diagram of TOPO reaction.	51
Figure 3-2 Sequence of pUni vector.	52
Figure 3-3 Scheme of recombination reaction in the ECHO™ Cloning System.	53
Figure 3-4 Vector diagram and sequence of pTBL1.	55
Figure 3-5 Vector diagram and sequence of pTBL2.	56
Figure 3-6 Scheme of protocol to produce deletion mutants.	61
Figure 4-1 Over-expression of <i>Mtb</i> HisI and HisE.	63
Figure 4-2 SDS-PAGE of purification of <i>BS</i> HisZ.	65
Figure 4-3 Over-expression of <i>Mth</i> HisG in different <i>E. coli</i> strains.	66
Figure 4-4 Over-expression of <i>Mth</i> HisG-related protein in different <i>E. coli</i> strains.	67
Figure 4-5 Purification of PA5104.	69
Figure 4-6 Crystals of <i>E. coli</i> HisG in presence of 2 mM AMP.	72
Figure 4-7 Crystals of <i>BS</i> HisG.	73
Figure 4-8 Crystals of <i>Mth</i> HisG.	74
Figure 4-9 Crystals of PA5104 (crystal form I).	75
Figure 4-10 Crystals of PA5104 (crystal form II).	75
Figure 4-11 Activity of <i>BS</i> HisG at different temperatures.	78
Figure 5-1 A typical diffraction pattern of ATP-PRT from <i>E. coli</i>	82
Figure 5-2 Fluorescence spectrum of Se-Met labelled HisG crystal.	84
Figure 5-3 χ^2 and R_{merge} vs oscillation.	88
Figure 5-4 Diffraction pattern from PA5104.	89
Figure 5-5 χ^2 and R_{merge} for 0.7 and 1.5 Å vs. Phi oscillation.	96
Figure 5-6 R_{merge} and completeness vs. wavelength for PA5104.	96
Figure 6-1 MAD solvent flattened electron density map of HisG.	100

Figure 6-2 Structure of <i>E. coli</i> HisG.	103
Figure 6-3 Superposition of <i>E. coli</i> HisG with glutamate receptor fragment.	104
Figure 6-4 Sequence alignment of representative HisGs.	106
Figure 6-5 HisG hexamer.	108
Figure 6-6 AMP binding in HisG.	110
Figure 6-7 Omitmap of PR-ATP and AMP.	111
Figure 6-8 Ramachandran plot of <i>E. coli</i> HisG complexed with PR-ATP.	112
Figure 6-9 PR-A(TP) binding compared with AMP binding in HisG.	114
Figure 6-10 Alignment of PRPP binding motif.	115
Figure 6-11 Comparison of PRPP binding site.	116
Figure 6-12 Comparison of different types of PRTs.	119
Figure 6-13 Comparison of PRPP binding motif of HisG and type I PRT.	120
Figure 6-14 Comparison of C-terminal domain of HisG with P _{II} -like protein.	122
Figure 7-1 AMoRe results for <i>Mth</i> HisG.	126
Figure 7-2 Electron density for histidine in <i>Mth</i> HisG.	127
Figure 7-3 <i>Mth</i> HisG monomer and dimer.	129
Figure 7-4 Rms deviation between chain A and B of <i>Mth</i> HisG.	130
Figure 7-5 Overlay of chain A and B of <i>Mth</i> HisG.	130
Figure 7-6 Trimer of C-terminal domain of <i>Mth</i> HisG with bound histidine.	131
Figure 7-7 Histidine binding in <i>Mth</i> HisG (chain A).	132
Figure 7-8 Comparison of histidine binding between HisG from <i>Mth</i> and <i>Mtb</i>	134
Figure 7-9 Histidine binding pocket.	135
Figure 8-1 Possible disulfide bond in HisG.	137
Figure 8-2 AMP binding in <i>E. coli</i> (a) and <i>Mtb</i> HisG (b).	138
Figure 8-3 AMP and PRPP binding in HisG.	139
Figure 8-4 AMP binding and binding site in HisG.	142
Figure 8-5 Surface representation of HisG hexamers.	144
Figure 8-6 Conformational change of HisG monomers and dimers.	150
Figure 9-1 R-factor distribution result from SnB.	153
Figure 9-2 Changes in correlation coefficient for successful ACORN calculation. .	157
Figure 9-3 Changes in CC _s for unsuccessful ACORN calculation.	157
Figure 9-4 Ramachandran plot for PA5104.	159
Figure 9-5 The structure of PA5104 solved at atomic resolution.	161
Figure 9-6 Surface representation of PA5104.	162

Table of Figures

Figure 9-7. Comparison for bicupins PA5104 and phosphomannose isomerase.	163
Figure 9-8 AMoRe rotation function results for PA5104.	163
Figure 9-9 PA5104 structure from the second crystal form.	166
Figure 9-10 Rms deviation between the four monomers of PA5104.	167
Figure 9-11 Electron density maps of PA5104 at different stages of phasing.	169
Figure 9-12 Superposition of PA5104 monomers.	170
Figure 9-13 Rotation of the two domains in PA5104.	171
Figure 9-14 Sequence alignment of PA5104.	172
Figure 9-15 Active site of PA5104.	174

Table of Tables

Table 1-1 Summary of enzymes involved in histidine biosynthesis and the corresponding genes from bacteria and fungi	3
Table 2-1 Growth media.....	20
Table 2-2 Antibiotic stock solutions and concentrations	21
Table 2-3 Buffers and solutions and its compositions.....	22
Table 2-4 Calculated molecular weights and absorbance at 280nm of investigated proteins	25
Table 2-5 Bacterial strains used for general cloning purposes.....	26
Table 2-6 Plasmids used for molecular cloning	27
Table 2-7 Genome Strains	33
Table 2-8 Primers used for PCR based molecular cloning.....	35
Table 2-9 Oligonucleotide to create linkers	37
Table 2-10 Oligonucleotides for PCR to create deletion mutants.....	37
Table 2-11 Buffers for purification of <i>Mth</i> HisG	40
Table 2-12 Buffers for Ni-NTA column	41
Table 2-13 Crystallisation screens.....	43
Table 3-1 Histidine pathway genes selected for cloning.....	47
Table 3-2 Conditions for optimal PCR of histidine genes	50
Table 4-1 Soluble and insoluble his enzymes obtained by over-expression in pTBL1 and pTBL2.....	64
Table 4-2 Rare codon usage in HisG and HisG-related protein of <i>Mth</i>	66
Table 4-3 DLS of <i>E. coli</i> HisG.....	70
Table 4-4 DLS of <i>Mth</i> HisG.....	71
Table 4-5 DLS of <i>Hin</i> and <i>BS</i> HisG	71
Table 4-6 Kinetic parameters of <i>Hin</i> HisG and truncation mutants.....	76
Table 5-1 Statistics for the X-ray data processed in <i>R32</i>	83
Table 5-2 Crystallographic data and phasing of <i>E. coli</i> HisG.....	85
Table 5-3 HisG <i>E. coli</i> ligand structure statistics	86
Table 5-4 Crystallographic data statistics for <i>Mth</i> HisG.....	87
Table 5-5 Data collection and scaling statistics of PA5104.....	90
Table 5-6 Comparison of PA5104 data collection and scaling statistics	91
Table 5-7 Rejections of reflection (1,1,18).	93

Table 5-8 Scaling statistics for PA5104 derivative data	94
Table 5-9 Scaling statistics for wavelength dependency of PA5104	95
Table 5-10 Scaling statistics of PA5104, crystal form 2	97
Table 6-1 Phasing statistics for <i>E. coli</i> HisG.....	100
Table 6-2 R- and free R-factor during refinement.....	101
Table 6-3 Refinement statistics of <i>E. coli</i> HisG with AMP	102
Table 6-4 Refinement statistics of HisG complexed with PR-ATP	113
Table 7-1 R- and free R-factor during refinement of <i>Mth</i> HisG	127
Table 7-2 Refinement statistics of <i>Mth</i> HisG	128
Table 7-3 Protein-histidine interactions in <i>Mth</i> HisG (chain A).....	132
Table 8-1 Buried solvent accessible areas for HisG hexamers.	146
Table 9-1 Phasing statistics for PA5104 Hg peak derivative	156
Table 9-2 R-factor and free R-factor during the refinement of PA5104	158
Table 9-3 Refinement statistics of PA5104.....	160
Table 9-4 CC and R-factor for MR in AMoRe.	164
Table 9-5 R-factor and free R-factor during the refinement of PA5104 (crystal form 2)	164
Table 9-6 Refinement statistics of PA5104 (crystal form 2).....	165

Chapter 1: Introduction

1.1 Histidine Biosynthesis – General Introduction

In microorganisms and plants histidine is synthesised via a 10 step pathway (Figure 1-1). The histidine pathway has been studied well for the bacteria *Escherichia coli* and *Salmonella typhimurium* [Winkler, 1987]. In both organisms there are 8 enzymes catalysing the 10 steps and these are encoded by 8 genes (*hisG{IE}AHFBCD*)¹ arranged as a single operon (*his* operon) [Carlomagno *et al.*, 1988]. Three of the enzymes are bifunctional (HisIE, HisD, HisB) and two form a multienzyme complex catalysing one reaction (HisH:HisF) [Alifano *et al.*, 1996]. Recent publications describe the histidine pathway enzymes from the gram-positive bacteria *Lactococcus lactis* [Delorme *et al.*, 1992], and from yeast [Bruni *et al.*, 1986] and plants [Fujimori and Ohta, 1998]. The nomenclature for the genes of the histidine pathway is different for bacteria and fungi. Table 1-1 shows the names of enzymes and the corresponding gene names for fungi and bacteria.

The histidine biosynthetic pathway is only present in bacteria, fungi, and plants, but absent in mammals. This makes it a good target for the development of antimicrobial and antifungal drugs as well as herbicides [Mousdale and Coggins, 1991]. A variety of selective inhibitors of the histidine pathway are known and used commercially as herbicides [Mori *et al.*, 1995]. These herbicidal compounds, e.g. amitrole, all inhibit imidazoleglycerol-phosphate dehydratase (the product of the *hisB* gene) in yeast and *Salmonella*, but there is evidence that there is an additional site of action in plants [Shaner, 1989]. A patent on the plant gene *hisG* emphasises the importance of the first enzyme and its role as an herbicidal target [Novartis AG, 1999]. More recently potential inhibitors of HisG were identified by computational methods [Gohda *et al.*, 2001]. Knowledge of the structures of enzymes of the histidine pathway will help with the understanding of its function and provide a basis for the development of antimicrobials and herbicides. At the beginning of this project no 3-dimensional structure of an enzyme of the histidine pathway had been determined.

¹ gene names will be displayed in '*italics*'; gene products will be displayed in 'Normal' with the first letter in capital

Although crystals had been obtained for HIS3 from *Saccharomyces cerevisiae* [Wilkinson *et al.*, 1995] and HisA and HisF from *Thermotoga maritima* [Thoma *et al.*, 1999]. Whilst conducting this study the structures of HisG, HisA, HisF and HisH have been solved by other groups [Lang *et al.*, 2000; Banfield *et al.*, 2001; Chaudhuri *et al.*, 2001]. This includes the structure of HisG from the pathogen *Mycobacterium tuberculosis* [Cho *et al.*, 2003], one of the targets organisms studied in this thesis.

Table 1-1 Summary of enzymes involved in histidine biosynthesis and the corresponding genes from bacteria and fungi.

Enzyme name	fungus gene name	bacterial gene name
ATP-phosphoribosyltransferase (EC 2.4.2.17)	<i>HIS1</i>	<i>hisG</i>
phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31)	<i>HIS4</i>	<i>hisE</i>
phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19)	<i>HIS4</i>	<i>hisI</i>
N-(5'-phospho-D-ribosylformimino)-5-amino-1-(5"-phosphoribosyl)-4-imidazole carboxamide isomerase (EC 5.3.1.16)	<i>HIS6</i>	<i>hisA</i>
imidazole glycerol phosphate synthase (EC 2.4.2.-)	<i>HIS7</i>	<i>hisH:hisF</i>
imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19)	<i>HIS3</i>	<i>hisB</i>
histidinol-phosphate aminotransferase (EC 2.6.1.9)	<i>HIS5</i>	<i>hisC</i>
histidinol-phosphatase (EC 3.1.3.15)	<i>HIS2</i>	<i>hisB</i>
histidinol dehydrogenase (EC 1.1.1.23)	<i>HIS4</i>	<i>hisD</i>

1.2 The Histidine Biosynthetic Pathway

The histidine pathway leads from the initial substrates 5-phosphoribosyl 1-pyrophosphate (PRPP) and ATP to histidine. Figure 1-1 shows the reaction pathway and names the enzymes and the genes encoding for them. Synthesis of histidine is an expensive process in energy terms, with 41 molecules of ATP needed to synthesis one molecule of histidine [Brenner and Ames, 1971].

Intermediates of the histidine pathway seem to be conserved amongst different species, however the structure and arrangement of genes and corresponding enzymes varies [Smith and Ames, 1964; Fani *et al.*, 1995; Alifano *et al.*, 1996]. In bacteria the genes are mostly arranged as an operon. Some bacteria such as *E. coli* have an operon containing all genes for histidine biosynthesis [Carlomagno *et al.*, 1988], some have an operon containing the majority of genes, e.g. *T. maritima* [Thoma *et al.*, 1998] and some have the genes scattered over the genome, e.g. *Synechocystis* [Kaneko *et al.*, 1996]. In fungi the genes are spread over 6 chromosomes [Fani *et al.*, 1995]. The arrangement of the domains of the proteins involved in histidine biosynthesis varies considerably between species. For example in *E. coli* there are 3 bifunctional enzymes HisIE, HisB and HisD whereas in *S. cerevisiae* two genes are linked to form *HIS4* containing the HisIE and HisD activities and HisB is split into two monofunctional enzymes encoded by *HIS2* and *HIS3* [Fani *et al.*, 1995].

The initial substrates of the histidine pathway PRPP and ATP play a key role in intermediary and energy metabolism. Furthermore they link the histidine pathway to the biosynthesis of folates, purines, pyrimidines, pyridine nucleotides, and tryptophan. A by-product of the histidine pathway aminoimidazole carboxamide ribonucleotide (AICAR), formed by HisH:HisF, is a purine precursor (see Figure 1-1 and Figure 1-2). For some bacteria and fungi it was found that the product of the HisI enzyme can be used in a shunt biosynthetic pathway which yields an intermediate of the purine biosynthesis [White, 1997]. The gene product of the *Snz* gene was identified as the possible enzyme for this reaction [Galperin and Koonin, 1997]. Additionally the pathway is linked to nitrogen metabolism via glutamine, a substrate for the HisH:HisF reaction (see Figure 1-2). Therefore the histidine pathway and its regulation are integrated with and influence overall cellular metabolism.

It is suggested that the histidine biosynthetic pathway was already established in the metabolism of the common ancestor of all cellular organisms [Lazcano *et al.*,

1992]. In addition, histidine might be a molecular remnant of a catalytic ribonucleotide present in the early stage of biochemical evolution in which RNA played a role in catalysis [White, 1976]. Gene duplication from a common ancestor gave rise to *hisA* and *hisF* as indicated by evolutionary studies [Fani *et al.*, 1994; Fani *et al.*, 1997] and this proposal is supported by kinetic and structural findings [Lang *et al.*, 2000]. Several evolutionary events, such as elongation, gene fusion and duplication, have occurred in the histidine pathway during its evolution from the ancestral pathway to the pathway in modern organisms [Fani *et al.*, 1995].

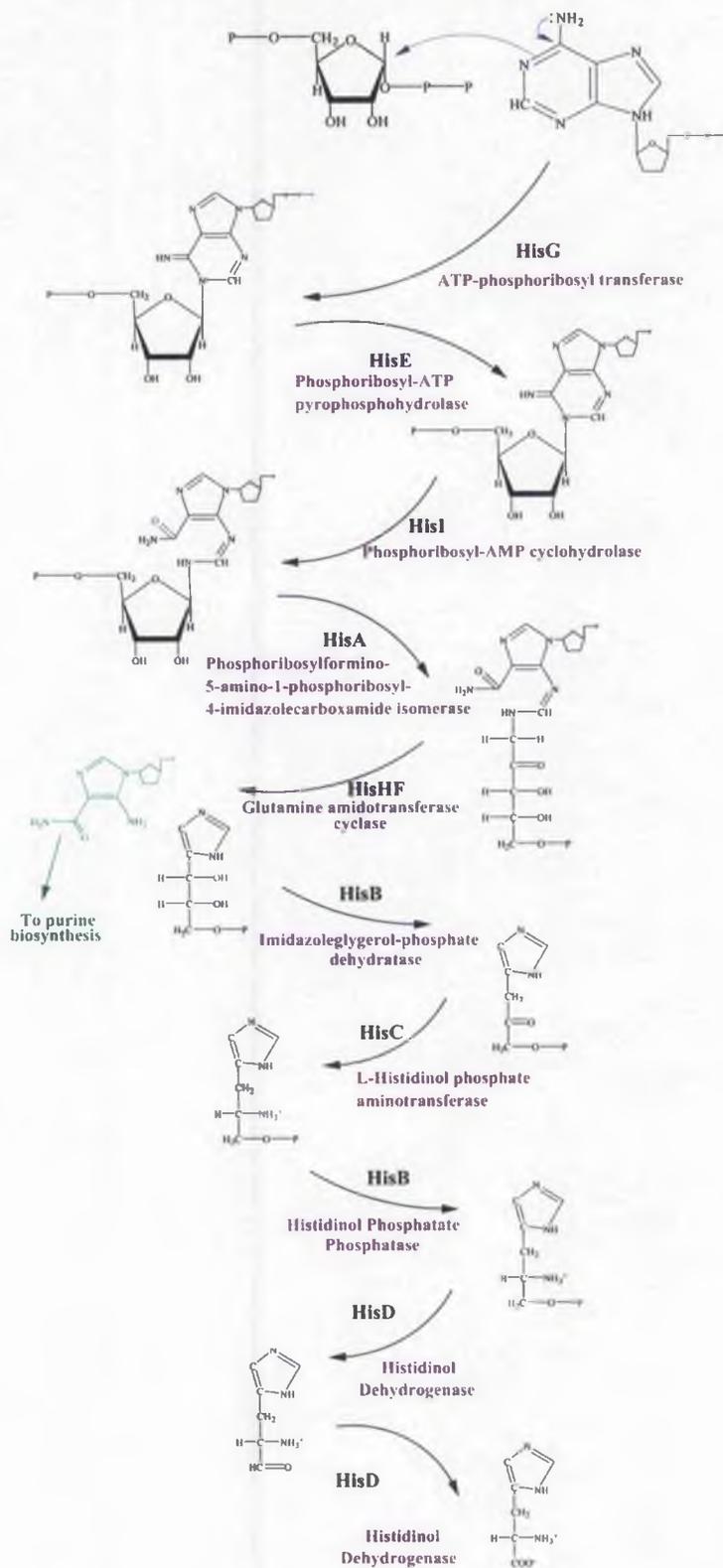


Figure 1-1 Histidine biosynthetic pathway. Shown is the reaction scheme with enzymes and genes encoding them.

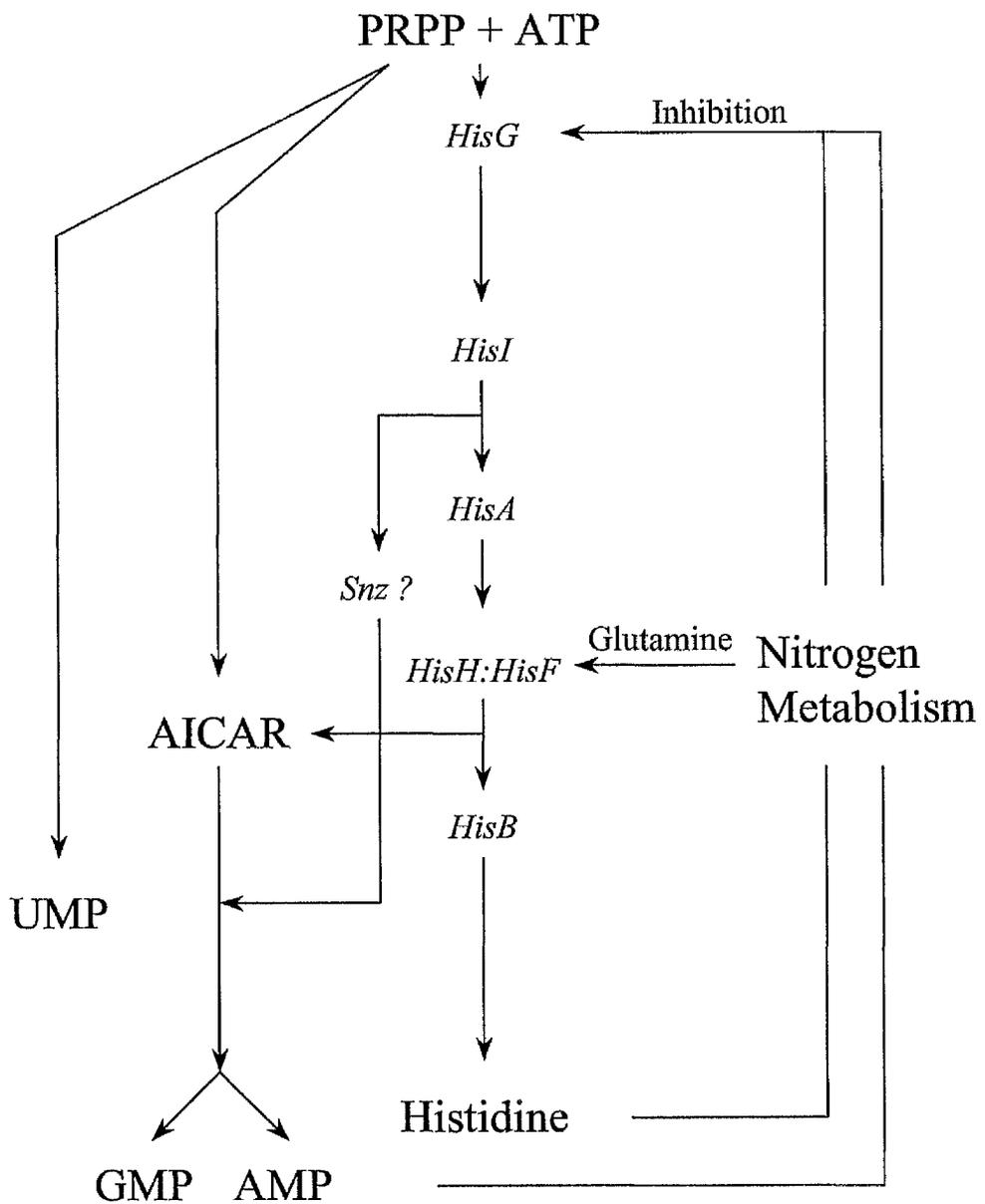


Figure 1-2 Substrates and intermediates of the histidine pathway. A simplified diagram showing the substrate and intermediates and their interaction with purine and pyrimidine biosynthesis as well as nitrogen metabolism.

1.3 Individual Enzymes of the Histidine Pathway

As research in this thesis focused on the first three enzymes of the histidine pathway catalysed by HisG and HisIE or HisI and HisE respectively, these will be described in further detail.

1.3.1 ATP-phosphoribosyltransferase (HisG or ATP-PRT)

The first enzyme of the pathway, ATP-phosphoribosyltransferase (ATP-PRT or HisG) encoded by the *hisG* gene, catalyses the condensation of ATP and PRPP to form *N*'-5'-phosphoribosyl-ATP (PR-ATP). The reaction scheme is shown in Figure 1-3. The pyrophosphate group on C-1 of PRPP is displaced by the N-1 nitrogen atom of the purine ring to form PR-ATP with the inversion of the chiral centre of C-1 [Chelsky and Parsons, 1975]. The enzyme is involved in several aspects of the regulation of histidine biosynthesis [Ames *et al.*, 1961]. ATP-PRT binds to histidyl-tRNA and it is feedback inhibited by the end product of the pathway,

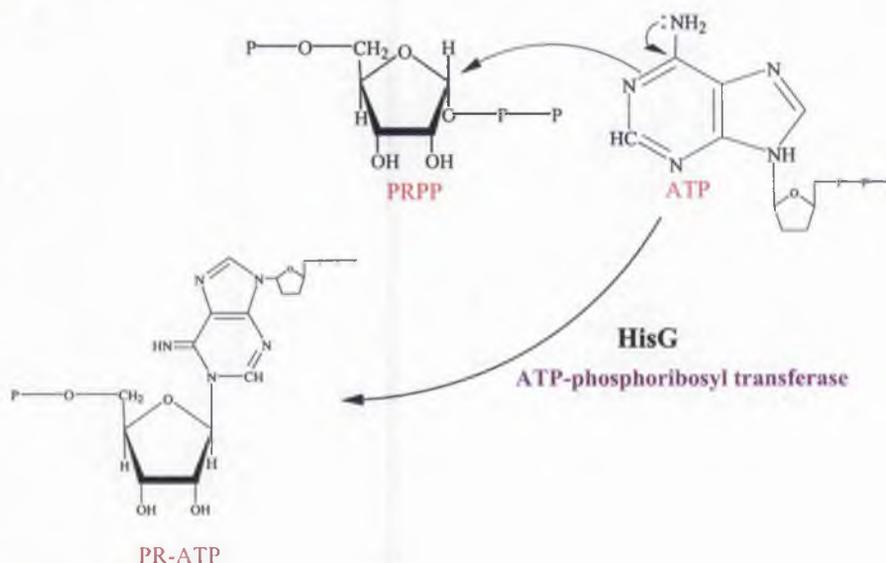


Figure 1-3 Reaction scheme of HisG.

histidine [Voll *et al.*, 1967]. Due to these properties and the connection of the histidine pathway with other biosynthetic and metabolic pathways, ATP-PRT plays an important role in the regulation of growth.

ATP-PRT was investigated in some detail in the 1970s and 1980s focusing on the enzymes from the bacteria *E. coli* [Dall-Larsen, 1988] and *S. typhimurium* [Ames *et al.*, 1961]. The enzyme from these bacteria are closely related and share a high

sequence homology of 94.7% [Carlomagno *et al.*, 1988] but have slightly different properties [Dall-Larsen, 1988].

The sources for studies of ATP-PRT from *E. coli* and *S. typhimurium* were usually depressed mutants from which the enzyme was purified [Voll *et al.*, 1967; Klungsoyr and Atkinson, 1970; Tebar *et al.*, 1973]. More recently over-expression of ATP-PRT in conjunction with the *hisIE* product was demonstrated [Davisson *et al.*, 1994]. However in this case there was no interest in purifying and studying the enzyme in more detail; the goal was to obtain histidine biosynthetic intermediates. An over-expression construct with only the *hisG* gene from *E. coli* was made and a faster purification procedure designed [unpublished results A. Elwell and J.R. Coggins; [Lohkamp *et al.*, 2000]].

From the primary structure of ATP-PRT from *E. coli* the molecular weight for the 299 amino acid polypeptide is calculated to be 33,367 Da. Experimentally determined molecular weights describe various oligomers of ATP-PRT. Ultracentrifugation and gel permeation have shown that the enzyme is in equilibrium between its active dimeric form, inactive hexameric form and higher aggregates [Klungsoyr and Kryvi, 1971; Tebar *et al.*, 1973]. The hexameric form is stabilised by histidine, AMP, PRATP, high concentrations of ATP and high enzyme concentration [Klungsoyr and Kryvi, 1971; Kryvi and Klungsoyr, 1971]. The second substrate PRPP seems to dissociate higher aggregates resulting in an increase in the concentration of dimers [Tebar *et al.*, 1973].

The secondary structure of ATP-PRT was investigated by CD spectroscopy [Kryvi, 1973]. At that time it was proposed to consist of about 33% α -helix and 20-30% β -sheet with no change observed when the inhibitors AMP and/or histidine were bound. Furthermore it was proposed that the first 120 residues of ATP-PRT contain a nucleotide-binding fold arranged in a $\beta\alpha\beta\alpha\beta\beta\alpha\beta\alpha\beta$ pattern [Argos *et al.*, 1983]. Despite the fact that a number of structures of other phosphoribosyltransferases have been determined, no comparison of ATP-PRT with these has been reported (e.g. [Focia *et al.*, 1998]).

A simple spectrophotometric assay for ATP-PRT is described in the literature which is based on the high extinction coefficient of PR-ATP at 290 nm [Smith and Ames, 1964; Martin *et al.*, 1971]. However due to the inhibition by the product PR-ATP modified assays have also been developed where the reaction is coupled with

HisIE [Kleeman and Parsons, 1975; Kronenberg *et al.*, 1975; Tebar and Ballesteros, 1976].

The reaction catalysed by ATP-PRT follows an ordered mechanism with ATP binding first and pyrophosphate being released first as shown in Figure 1-4 [Morton and Parsons, 1976]. ATP-PRT binds 3 histidine per hexamer with high affinity i.e. one histidine per dimer [Klungsoyr and Atkinson, 1970]. PR-ATP and AMP enhance the binding of histidine to ATP-PRT, the latter with strong cooperativity [Dall-Larsen and Klungsoyr, 1976]. AMP itself binds with a stoichiometry of about 3 molecules per hexamer and competes with the binding of both substrates, ATP and PRPP [Kleeman and Parsons, 1976; Morton and Parsons, 1976]. Negative cooperativity of ATP binding was found for the *Salmonella* enzyme [Bell and Koshland, 1971]. The product PR-ATP binds preferentially at 1-2 sites per hexamer depending on several factors including Mg^{2+} and PR-ATP concentration [Dall-Larsen and Klungsoyr, 1976].

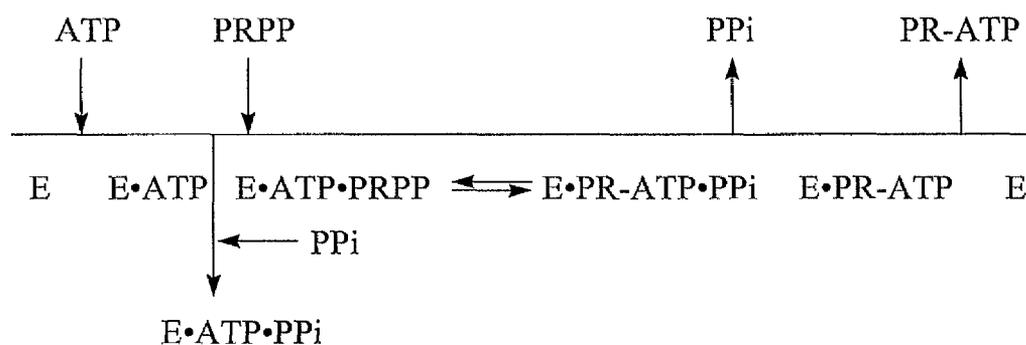


Figure 1-4 Scheme of ordered mechanism of HisG.

Mutants bearing deletions in the *hisG* gene seem to be regulated normally [Scott *et al.*, 1975], even so there is evidence for ATP-PRT being involved in regulation of the *his* operon. It has been shown that ATP-PRT from *S. typhimurium* binds to histidyl-tRNA [Kovach *et al.*, 1970] and is involved in the repression control of histidyl-tRNA synthetase [Coleman and Williams, 1974]. Furthermore it is proposed that it can bind to the leader mRNA [Ames *et al.*, 1983]. Earlier research suggested that it binds to the DNA of the *his* operon [Meyers *et al.*, 1975], but this was shown to be an error arising from incorrect DNA sequences [pers. comm. Yanofsky]. However, these experimental results suggest that ATP-PRT plays a role in regulating the histidine concentration in the cell.

The ATP-PRTs from some bacteria are by about 80 residues shorter at the C-terminus compared to the enzyme from e.g. *E. coli*². The genome of bacteria with this short form of the enzyme contains an additional gene, *hisZ*, which encodes for a histidyl-tRNA synthetase paralog [Renault *et al.*, 1995; Sissler *et al.*, 1999]. It has been shown that HisZ is necessary for the catalytic activity of the short ATP-PRT. The histidyl tRNA synthetase paralog contains the catalytic core of a class II tRNA synthetase but lacks the aminoacylation function. Both HisG and HisZ appear to play a key role in regulation of histidine biosynthesis [Sissler *et al.*, 1999]. Initial research suggested that the HisG-HisZ complex is octameric and is comprised of 6 HisG units and 2 HisZs [Champagne *et al.*, 2001]. Further investigation confirmed the octameric nature of the complex but demonstrated that it contains two of each HisG and HisZ dimers arranged as a dimer of hetero-tetramers [Bovee *et al.*, 2002].

HisG belongs to the class of phosphoribosyltransferases (PRT) some of which are linked to disease in humans [Musick, 1981]. At the beginning of the project two different types of PRT structures were known and another one predicted. Several reported structures, e.g. uracil PRT, show the type I fold of a five-stranded β -sheet surrounded by three or four α -helices (and a variable hood domain) (e.g. [Tomchick *et al.*, 1998]). Quinolinic acid PRT comprises a type II fold with two domains: a mixed α/β N-terminal domain and an $\beta_7\alpha_6$ barrel C-terminal domain, similar to known $(\beta\alpha)_8$ barrels [Eads *et al.*, 1997]. Anthranilate PRT is predicted to give a typical $(\beta\alpha)_8$ barrel fold [Wilmanns and Eisenberg, 1993]. However the crystal structure was solved recently and showed a fold distinct from the previous known ones [Mayans *et al.*, 2002]. This type III fold has a central 7-stranded sheet on both sides flanked by several α -helices.

The structure of HisG was solved as part of this project. The fold shows a new type of PRT (IV) comprising three domains. This was confirmed by the structure of HisG from *M. tuberculosis* [Cho *et al.*, 2003].

² Therefore if necessary ATP-PRT or HisG is in the following referred to as long or long form and short or short form respectively.

1.3.2 Phosphoribosyl-ATP pyrophosphatase, phosphoribosyl-AMP cyclohydrolase (HisIE)

The second and third synthesis steps are performed by the enzymes phosphoribosyl-ATP pyrophosphatase (HisE) and phosphoribosyl-AMP cyclohydrolase (HisI). Firstly HisE hydrolyses the two phosphates from the ATP part of the PR-ATP and then HisI opens the purine ring [Smith and Ames, 1965]. The product N'-[(5'-phosphoribosyl)-formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (BBMII or Pro-FAR) and intermediates of this reaction are shown in Figure 1-5.

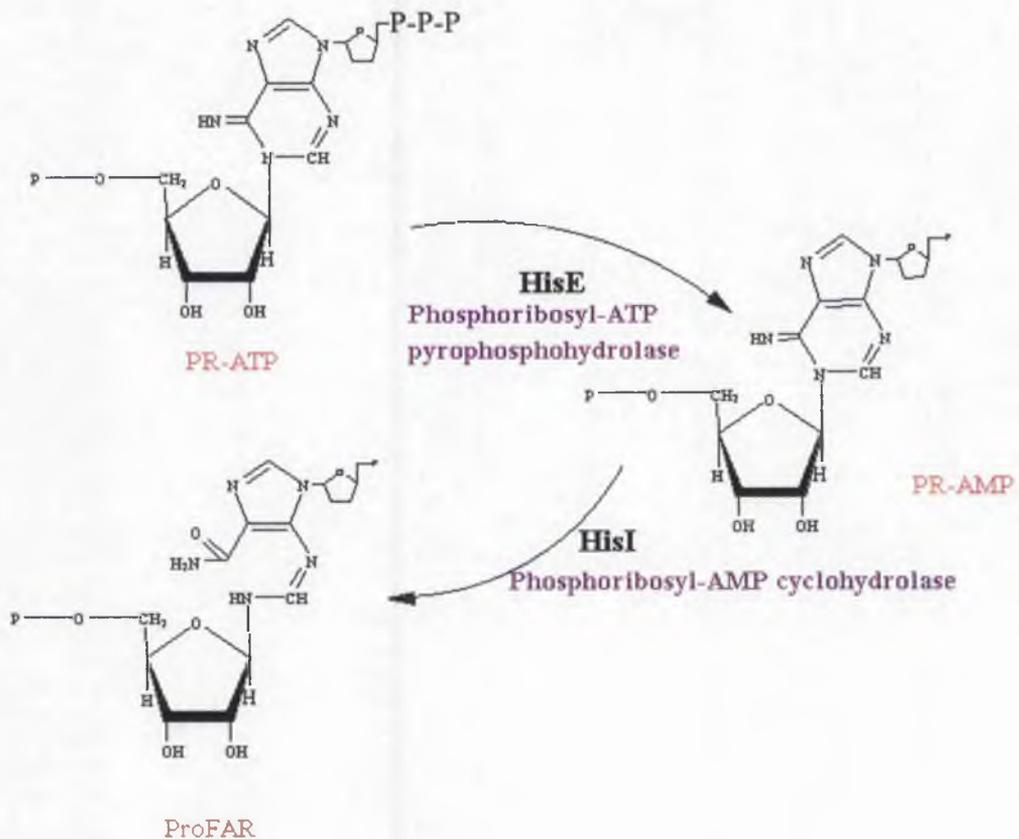


Figure 1-5 Reaction scheme of HisIE.

In bacteria the enzyme is either bifunctional (HisIE) encoded by the *hisIE* gene, e.g. *E. coli*, or there are two independent genes and corresponding monofunctional enzymes, e.g. in *Methanococcus vannielli* [Beckler and Reeve, 1986] and *Azospirillum brasilense* [Fani *et al.*, 1993]. In *Rhodobacter sphaeroides* the HisI is also monofunctional with no link to the *hisE* gene [Oriol *et al.*, 1996; Oriol *et al.*, 1997]. In the bacterial bifunctional HisIE the N-terminal domain performs the third

step (cyclohydrolase activity) and the C-terminal domain the second step (pyrophosphate hydrolase activity) [Donahue *et al.*, 1982].

In fungi the HisI and HisE activities are encoded by the multifunctional *HIS4* gene which also includes the activity for HisD [Donahue *et al.*, 1982].

Monofunctional HisI from *M. vannielli* was overexpressed and characterised [D'Ordine *et al.*, 1999]. The enzyme is present as a dimer with one equivalent of zinc ion per subunit. Both divalent cations Zn^{2+} and Mg^{2+} are required for cyclohydrolase activity of HisI [D'Ordine *et al.*, 1999].

HisIE from the plant *Arabidopsis thaliana* was cloned and identified as a bifunctional enzyme [Fujimori and Ohta, 1998].

Neither the individual enzymes HisI and HisE nor the bifunctional HisIE have been studied structurally to date.

1.4 Histidine degradation

1.4.1 Introduction

The synthesis of amino acids is energetically expensive and therefore tightly controlled in microorganisms. Excess amino acids, unlike glucose or fatty acids, cannot be stored but can be used as a carbon, nitrogen and energy source. Furthermore, amino acids can be scavenged from hosts or the environment and used *per se* or serve as nutrition. The amino acids can be degraded and used as an important metabolic intermediate and a source of energy for the microorganism.

C_5 amino acids such as histidine, arginine, glutamine and proline are converted to glutamate and subsequently to α -ketoglutarate, which then feeds into the citric acid cycle [Stryer, 1988]. The amino acids are degraded by a series of enzymatic reactions into glutamic acid. In bacteria 4 or 5 individual enzymatic steps are exploited in the histidine utilisation depending on species. Both pathways yield as products ammonia, as a nitrogen source, and glutamate. The third product is a C_1 compound, either formamide or formate. These two pathways of histidine utilisation are outlined in the following section.

1.4.2 Histidine utilisation pathways

The histidine utilisation pathway 1 is common in most bacteria that have been studied, e.g. *S. typhimurium*, *Klebsiella aerogenes*, and *B. subtilis*. The degradation is accomplished by 4 enzymes, each catalysing a single reaction (Figure 1-6). Pathway 2, which is present for example in *Pseudomonas* ssp., shares the initial three enzymatic steps with histidine utilisation pathway 1 but replaces the final enzyme with two distinct enzymes (Figure 1-7). This results in an additional intermediate and subsequently a different by-product of the final step, formate.

The first enzyme in the pathways, encoded by the *hutH* gene, is histidine ammonia-lyase (HutH). This 55.6 kDa protein catalyses the elimination of ammonia to yield the α - β -unsaturated *trans*-urocanate. Urocanase (HutU) catalyses the second step of both histidine utilisation pathways, the conversion of urocanate to 4-imidazolone-5-propionate. HutU appears to be partially stabilised by the presence of pyridoxal phosphate. Imidazolone-5-propionate hydrolase (HutI) catalyses the ring-opening reaction, converting 4-imidazolone-5-propionate into *N*-formimino-L-glutamate. Imidazolone-5-propionate hydrolase (HutI) catalyses the ring-opening reaction, converting 4-imidazolone-5-propionate into *N*-formimino-L-glutamate.

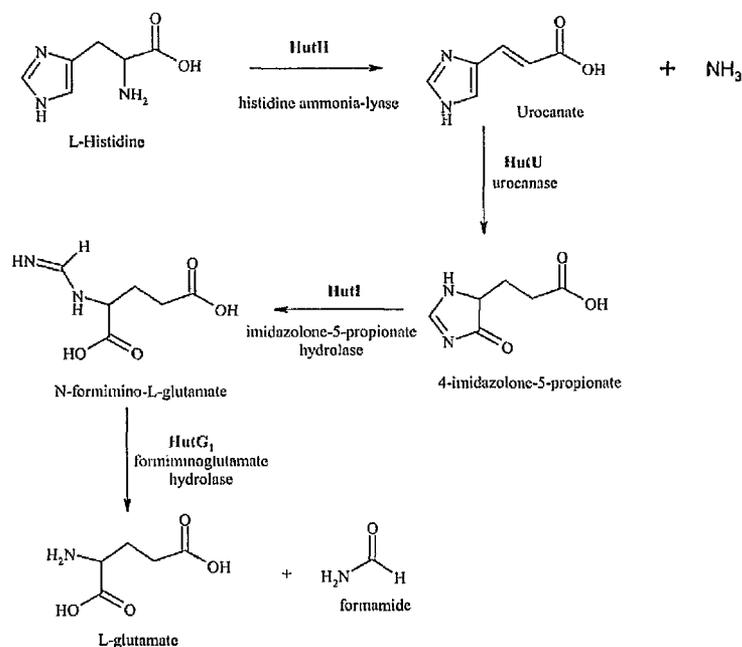


Figure 1-6 Histidine utilisation pathway 1.

The final step in histidine utilisation pathway 1 is catalysed by formiminoglutamate hydrolase (HutG) which hydrolyses *N*-formimino-L-glutamate to yield L-glutamate and formamide.

Histidine utilisation pathway 2 is different to pathway 1 in the conversion of *N*-formimino-L-glutamate to L-glutamate. The single step, performed in pathway 1 by HutG, is achieved by 2 enzymatic reactions. *N*-formiminoglutamate is initially converted to *N*-formyl-L-glutamate by formiminoglutamate deiminase (HutF) and the product of this reaction, *N*-formyl-L-glutamate, is subsequently hydrolysed to yield L-glutamate and formate by formylglutamate deformylase (HutG³).

In evolutionary terms the mutation of HutG₁ to HutF with the production of a new intermediate, *N*-formyl-L-glutamate, could have induced the recruitment of an enzyme capable of breaking it down [Hu *et al.*, 1987]. This was possibly *N*-acetylglutamate synthase, an enzyme believed to be involved in arginine biosynthesis [Früh and Leisinger, 1981].

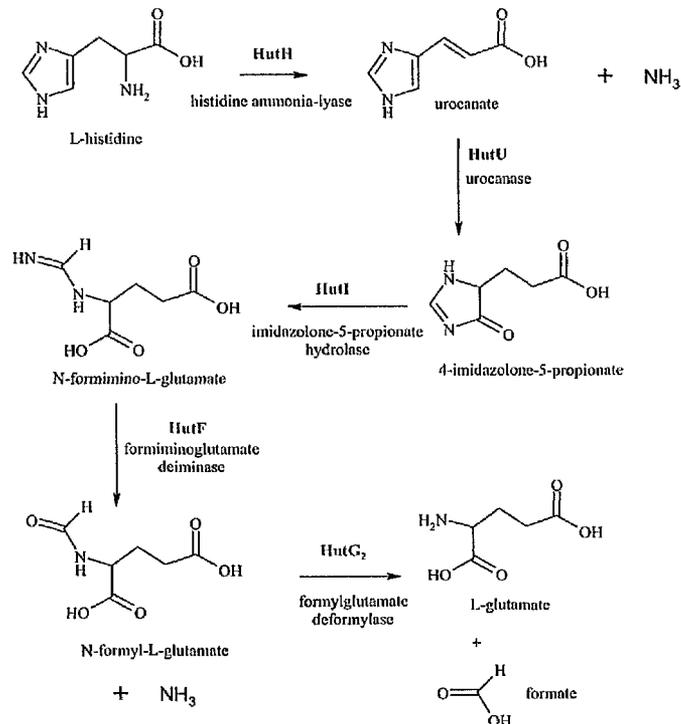


Figure 1-7 Histidine utilisation pathway 2.

1.4.3 Organisation of operons and control of expression

A number of mechanisms that respond to changes in growth conditions control the gene expression of the histidine utilization (Hut) system. One effect is positive

³ It is somewhat confusing that both pathways name one gene *hutG* even though they are clearly two different genes encoding two different enzymes. If necessary in the following discussion these genes will be differentiated by subscripts 1 and 2 which refer to the two pathways respectively.

regulation responding to cyclic AMP or the nitrogen source and repression provoked by succinate carbon metabolites. In addition to these global regulatory mechanisms, there is specific regulation of the enzymes involved in the pathway.

Genetic studies have revealed that the organisation of the *hut* genes shows considerable diversity in bacteria. In *B. subtilis* the genes are arranged in a single operon, which can be induced by histidine and is controlled by a positive regulator [Kimhi and Magasanik, 1970]. Multiple transcription units are found in *Klebsiella*, *Salmonella* and *Pseudomonas* ssp. which are under negative control by a single repressor protein, the *hutC* gene product [Smith and Magasanik, 1971]. Additionally the expression of the *hut* genes can be induced by urocanate, the product of the first step in the pathway [Smith and Magasanik, 1971; Goldberg and Magasanik, 1975]. Urocanate can originate extracellularly either *per se* or as the product of imidazolylpropionate degradation [Coote and Hassall, 1973]. In *Pseudomonas* ssp. HutF and HutG₂ are additionally induced by their respective substrates, formiminoglutamate and formylglutamate [Coote and Hassall, 1973]. In *P. putida* three operator sites for the *hut* genes were identified, the *hutUHIG*₂ operator for this operon, a *hutF* operator and a separate one for *hutG*₂ expression only [Hu *et al.*, 1989]. The inducer urocanate causes the dissociation of operator-repressor complexes for all three operators. Formylglutamate was found to inhibit the binding of the repressor to the separate *hutG*₂ operator only and has no effect on the action of urocanate on the other operators. It was proposed therefore, that formylglutamate binds to a separate site from urocanate on the repressor [Hu *et al.*, 1989].

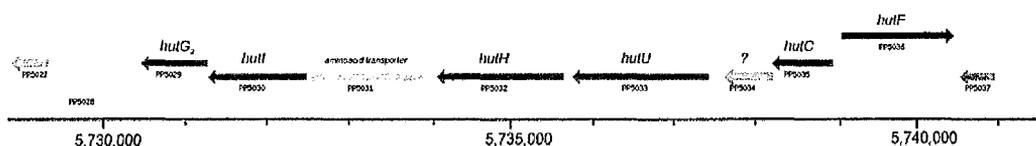


Figure 1-8 *Hut* operon of *Pseudomonas putida*. Arrows indicate ORFs and their direction of expression. The gene names are shown on top of the arrows and their genome identifier at the bottom. The scale shows the position of the operon within in the genome.

Most genes of the *hut* operon of *P. putida* have been identified, but PP5034 remains unidentified in this cluster (Figure 1-8). The gene preceding *hutC*, PP5034, is included in the same transcriptional unit as the repressor but was found to have no

effect on the binding at the *hutUHIG₂* operator site [Allison and Phillips, 1990]. However it is possible that there is a separate regulatory factor involved in enhancing or modifying the effect of the *hut* repressor. This could be related to the possible bifunctional nature of the *hut* repressor [Hu *et al.*, 1989]. The corresponding gene of PP5053 in *P. aeruginosa* is PA5104 which was studied in this thesis.

1.5 Aims of the project

The aims of this study are centred around histidine metabolism in microorganisms. The primary aims were to characterise structurally and biochemically the first two (three) enzymes of the histidine pathway, HisG and HisIE. The secondary aims were more diverse and included structural investigations on a protein involved in regulation of histidine utilisation. Most of the aims were achieved, however because of time restrictions and diversification during the project, based on the results obtained, not all could be accomplished. The aims, listed in more detail, were as follows:

Determination of the 3-dimensional structure of *E. coli* HisG, the first enzyme of histidine biosynthesis, in complex with AMP by X-ray crystallography. This would represent the first structure of an ATP-phosphoribosyltransferase. The structure would give insight into the mode of inhibition by AMP and presumably help to understand the reaction mechanism.

Cloning and over-expression of HisGs, HisIEs and HisZ from various microorganisms. The availability of enzymes from different species increases the chance of obtaining diffraction quality crystals for structural investigations. Furthermore a diversity of enzymes with different properties can be studied, e.g. long and short forms of HisG and bifunctional HisIE in comparison to monofunctional HisI and HisE.

Crystallisation and structure determination of HisGs, HisIEs and the HisG-HisZ complex. Enzymes obtained from over-expression should be crystallised and, assuming that diffraction quality crystals were obtained (ideally in the presence of inhibitors or substrates), the 3-dimensional structure is to be determined by X-ray crystallography. The structures would provide insight into the reaction mechanism and mode of inhibition of the enzymes.

Furthermore they would allow comparison of enzyme structures from different species and between different forms, aggregation states and conformations of the enzymes.

Investigation of differences between the short form and long form of HisG. The native long HisG and the short HisG, in isolation and in complex with HisZ, should be investigated structurally and kinetically. This would help to explain the lack of activity of the short form and the basis for the formation of the HisG-HisZ complex. Mutational analysis helps the investigation by separating functional and structural parts and domains of the enzyme.

Characterisation of oligomeric states of HisG. The oligomeric assemblies of HisG would be investigated by gel filtration experiments, DLS, X-ray crystallography and EM. This will help to understand the mode of inhibition by formation of hexameric or higher aggregates.

Determination of the 3-dimensional structure of PA5104 from *P. aeruginosa* by X-ray crystallography. The structure of PA5104 can give an indication of the function of the protein which is found in the operon which encodes enzymes of the histidine utilisation pathway 2.

Chapter 2: Materials and Methods

2.1 Materials and Methods

2.1.1 Chemicals

Chemicals and biochemicals were generally analytical grade or else the highest grade available. Unless otherwise stated all chemicals were purchased from the Sigma Aldrich Company Ltd, Poole, UK or Fisher Scientific, Loughborough, UK. Bacto-tryptone, bacto-yeast extract and bacto-agar were obtained from Difco Laboratories, Detroit, USA. Agarose was obtained from GIBCO BRL Life Technologies, Paisley, UK. Complete™ EDTA free Protease inhibitor cocktail tablets were obtained from Boehringer Mannheim, Lewes, UK.

2.1.2 Proteins and enzymes

Restriction enzymes and their buffers were obtained either from New England Biolabs Inc. (*KpnI*, *NdeI*, *XhoI*), Hitchin, UK; Promega Corporation (*NsiI*, *SacI*), Southampton, UK; Stratagene (*DpnI*), Amsterdam, The Netherlands; or Boehringer Mannheim (*EcoRV*).

Vent_R® DNA polymerase was obtained from New England Biolabs. Taq DNA polymerase was purchased from Promega Corporation and *PfuTurbo*® DNA polymerase from Stratagene. Bacteriophage T4 DNA ligase and Calf Intestinal Alkaline Phosphatase (CIAP), were obtained from Boehringer Mannheim.

2.1.3 Oligonucleotides

Oligonucleotides for use both in the polymerase chain reaction were purchased from MWG-BIOTECH (UK) Ltd, Milton Keynes, UK. The primers were HPLC purified and 5' phosphorylated when necessary.

2.1.4 Chromatography media

Q-Sepharose and Sephacel S-200 were obtained from Pharmacia Biotech, St Albans, UK. Reactive Green 19 agarose resin was purchased from Sigma Aldrich Company Ltd. Ni-NTA agarose resin was obtained from Qiagen Ltd, Crawley, UK.

2.2 Media and supplements

2.2.1 Media for bacterial growth

E. coli strains PIR1, PIR2, DH5 α , JM109, various C41(DE3) and BL21(DE3) were routinely cultured in Luria-Bertani medium (LB). Solid medium was obtained by adding 1.5% (w/v) bacto-agar to the liquid LB. Table 2-1 gives the composition of the growth media used.

Table 2-1 Growth media

Media	Composition per litre	Comments
LB (Luria-Bertani broth)	10 g bacto-tryptone 5 g yeast extract 5 g NaCl	Sterilised by autoclaving at 15 psi for 30 minutes.
LB agar	LB plus 15 g bacto-agar	Sterilised by autoclaving at 15 psi for 30 minutes.
SOC (for 100ml)	2 g bacto-tryptone 0.5 g yeast extract 1 ml 1 M NaCl 0.25ml 1 M KCl 1 ml Mg ²⁺ /glucose stock (1 M MgCl ₂ , 1 M MgSO ₄ , 2 M glucose) filter sterilised	Dissolve all (except Mg/glucose stock) in 97 ml dsH ₂ O, autoclave, cool to RT, add Mg ²⁺ /glucose stock, filter sterilise medium through 0.22 μ m filter.
M9 minimal media	100 ml 10x salt solution (see Table 2-3) 2 ml 1 M MgSO ₄ 0.1 ml 1 M CaCl ₂ 2 mg thiamine 5 g glucose	Additional requirements: 0.5 g all nucleotides 40 mg all amino acids

2.2.2 Media to produce seleno-methionine substituted protein

To produce seleno-methionine substituted protein, expression was carried out in the methionine auxotroph strain B834(DE3)pLysS grown at 37°C in minimal

media M9 (see Table 2-1) with glucose (5 g/l) as carbon source. Minimal media M9 was autoclaved and then supplemented with thiamine (2 mg/l), all nucleotides (0.5 g/l) and all amino acids (40 mg/l) and filter sterilised. Media were prepared devoid of methionine, with 50 mg/l methionine and 50 mg/l seleno-methionine. Antibiotics were added as required at the concentrations listed in Table 2-2.

2.2.3 Antibiotics

Four different antibiotics were used, ampicillin, tetracycline, chloramphenicol and kanamycin. Stock solutions of the antibiotics were dissolved in the given solution and, if indicated, sterilised by filtration through a 0.22 µm filter (Table 2-2). The stock solutions were stored at -20°C. The growth medium was allowed to cool to 50-55°C before the addition of antibiotic to a final concentration as given in Table 2-2. All media and plates containing antibiotics were stored at 4°C and used within 4 weeks.

Table 2-2 Antibiotic stock solutions and concentrations

Antibiotic	Concentration of stock solution	Concentration used	Solute	Filter sterilisation
Ampicillin	100 mg/ml	50-100 µg/ml	dsH ₂ O	✓
Tetracycline	12.5 mg/ml	12.5 µg/ml	80% ethanol	✗
Chloramphenicol	34.5 mg/ml	34.5 µg/ml	100% ethanol	✗
Kanamycin	10 mg/ml	50 µg/ml	dsH ₂ O	✓

2.2.4 Buffers and solutions for molecular biology experiments

A variety of different buffers and solutions were used for various molecular biology experiments. Table 2-3 gives a list of buffers with the composition.

Table 2-3 Buffers and solutions and its compositions

Buffer	Composition	Comments
RF1	100 mM RbCl 50 mM MnCl ₂ 30 mM K acetate 10 mM CaCl ₂ 15% (w/v) glycerol	Components are added to dsH ₂ O and then adjusted to pH 5.8 with dilute acetic acid and filter sterilised. For standard transformations RbCl was generally substituted with KCl.
RF2	10 mM MOPS 10 mM RbCl 75 mM CaCl ₂ 15% (w/v) glycerol	Components are added and adjusted to pH 6.8 using NaOH and filter sterilised. For standard transformations RbCl was generally substituted with KCl.
MgCl ₂ -CaCl ₂ solution	80 mM MgCl ₂ 20 mM CaCl ₂	Filter sterilised.
10x salt solution (M9 medium)	128 g Na ₂ HPO ₄ ·7H ₂ O 30 g KH ₂ PO ₄ 5 g NaCl 10 g NH ₄ Cl	Components per 1 l H ₂ O. Sterilised by autoclaving at 15 psi for 30 minutes.

50x TAE Buffer	2 M 5.71% (v/v) 50 mM	Tris base glacial acetic acid EDTA	Adjusted to pH 7.2. Used at 1x concentration.
10x TBE Buffer	900 mM 900 mM 25 mM	Tris base boric acid EDTA	Add components but slightly less than total amount of boric acid. Adjust to pH 8.3 with remaining boric acid. Bring to final desired volume. Used at 1x concentration.
DNA gel loading buffer	38% (w/v) 0.1% (w/v) 67 mM	sucrose bromophenol blue EDTA	
SDS-PAGE sample buffer	250 mM 2% (w/v) 20% (v/v) 0.0025% (w/v) DTT if required	Tris-HCl pH 6.8 SDS glycerol bromophenol blue	
SDS-PAGE 10x running buffer	3.04% (w/v) 14.4% (w/v) 1% (w/v)	Tris base glycine SDS	pH should be 8.3, can be adjusted with HCl if necessary.

2.2.5 Isopropyl- β -D-thiogalactopyranoside (IPTG)

A 1 M stock solution was made by dissolving 2 g of IPTG in a final volume of 10 ml. The solution was filter sterilised through a 0.22 μ m filter and stored at -20°C.

2.2.6 X-gal LB agar plates

LB agar medium was allowed to cool to 50-55°C before X-gal and IPTG stock solutions were added. The final concentrations were 40 μ g/ml X-gal and 0.1 mM IPTG. The stock solution of X-gal was 40 mg/ml in DMF and was stored in the dark at -20°C. The appropriate antibiotic was added at the same time.

2.3 General methods

2.3.1 General laboratory methods

General molecular biological techniques were carried out as described in Sambrook *et al.* (1989). General laboratory methods for handling proteins and enzymes were as described in Deutscher (1990).

2.3.2 French pressure cell

Cells were broken by three passages through an automatic French pressure cell at 950 psi. The pressure cell was pre-cooled on ice before use.

2.3.3 pH measurement

pH measurements were made with a Orion 420A+ Benchtop pH Meter (Thermo Electron Corporation, USA). The pH meter was calibrated prior to use with 3 different buffers at room temperature (20°C). It should be noted that given pH values are valid for room temperature and can differ when used at other temperatures due to the effect of temperature on the pK_as of certain buffer components.

2.3.4 Determination of protein concentration

Protein concentrations were estimated by measuring the absorption at 280nm. The molar extinction coefficient was calculated with the Schepartz Lab Biopolymer Calculator (<http://paris.chem.yale.edu/extinct.html>). Table 2-4 shows the calculated molecular weights and concentrations of standard solutions of the proteins investigated in this work.

Table 2-4 Calculated molecular weights and absorbance at 280nm of investigated proteins

Protein	MW / Da	MW _{tag} ¹ / Da	A _{280nm} ² / mg/ml	A _{280nm} tag ^{1,2} / mg/ml
Mth HisG	31,361.4		4.57	
Mth HisGrel	37,625.0		1.32	
Hin HisG	33,821.3	34,644.1	1.41	1.45
Hin HisG	33,821.3	34,644.1	1.41	1.45
Hin HisIE	25,096.7	25,919.5	0.68	0.71
BS HisG	23,621.3	24,444.1	1.54	1.60
BS HisZ	48,132.0	48,954.8	1.18	1.20
BS 6xG+2xZ	228,611.4	235,194.2	1.42	1.46
BS HisIE	23,901.1	24,723.9	0.81	0.84
Mtb HisG	30,481.0	31,303.8	1.38	1.42
Mtb HisI	12,465.0	14,171.9	0.86	0.97
Mtb HisE	10,274.7	11,981.6	0.71	0.82
Hin HisG short	25,347.3	26,170.1	1.51	1.56
Hin HisG 294	33,172.5	33,995.4	1.39	1.42
Hin HisG C-term	8,335.9	9,158.8	1.16	1.27

¹tag means that this calculation includes the his-tag and his-tag with cleavage site when appropriate
²this column indicates the protein concentration if the absorbance at 280nm is 1 in a 1cm cell.

2.4 Bacterial strains and cloning vectors

2.4.1 Storage of Bacterial Strains

Bacterial strains were stored as glycerol stocks. These were made by the addition of sterile 50% (v/v) glycerol/water to overnight bacterial cultures to give a final glycerol concentration of 25% (v/v). Duplicate collections of stock cultures were stored at -20°C and -80°C. Some cultures were temporarily stored by streaking onto agar plates and maintained at 4°C for up to 4 weeks.

2.4.2 Bacterial strains

The bacterial strains and plasmids used in this project are listed in Table 2-5 and Table 2-6.

Table 2-5 Bacterial strains used for general cloning purposes

Bacterial strain	Genotype	Reference
<i>E. coli</i> PIR1	F- <i>lacI69 rpoS(Am) robA1 creC510 hsdR514 endA recA1 uidA</i> (<i>Mlu</i> I):: <i>pir-116</i>	Invitrogen
<i>E. coli</i> PIR2	F- <i>lacI69 rpoS(Am) robA1 creC510 hsdR514 endA recA1 uidA</i> (<i>Mlu</i> I):: <i>pir</i>	Invitrogen
<i>E. coli</i> DH5 α	F' ϕ 80 <i>dlacZ</i> Δ M15 <i>recA1 endA1 gyrA96 thi-1 hsdR17</i> (τ_K m_K^+) <i>supE44 relA1 deoR</i> Δ (<i>lacZYA-argF</i>)U169	Hanahan, 1985
<i>E. coli</i> JM109	<i>recA1, endA1, gyrA96, thi, hsdR17</i> (τ_K - m_K^+), <i>relA1, supE44, \Delta(<i>lac-proAB</i>), [<i>F'</i>, <i>traD36, proAB, lacI q Z\DeltaM15]</i></i>	Messing, J. et al., 1981; Promega
<i>E. coli</i> B834(DE3)	F- <i>ompT hsdS</i> (τ_B - m_B -) <i>gal dcm met</i> (DE3)	Studier and Moffatt, 1986
<i>E. coli</i> BL21(DE3)	F- <i>ompT hsdS_B</i> τ_B m_B <i>dcm gal</i> λ DE3	Studier and Moffatt, 1986
<i>E. coli</i> BL21(DE3)pLysS	F- <i>ompT hsdS_B</i> τ_B m_B <i>dcm gal</i> λ DE3 pLysS Cm ^r	Studier and Moffatt, 1986; this study
<i>E. coli</i> BL21(DE3)RIL/RP	F- <i>ompT hsdS_B</i> τ_B m_B <i>dcm gal</i> λ DE3	Miroux and Walker, 1996
<i>E. coli</i> C41(DE3)	n/a	this study
<i>E. coli</i> C41(DE3)RIL/RP	n/a	this study

Table 2-6 Plasmids used for molecular cloning

Plasmid	Antibiotic resistance	Function	Reference
pTB361	tetracycline	T7 expression plasmid	Brockbank and Barth, 1993
pGEM [®] -5Zf(+/-)	ampicillin	T7 cloning vector	Promega
pUni/V5-His-TOPO	kanamycin	TOPO cloning vector	Invitrogen
pLysS	chloramphenicol	T7 lysozyme plasmid	Moffatt and Studier, 1987
RJL	chloramphenicol	Vector encoding rare codons	Stratagene
RP	chloramphenicol	Vector encoding rare codons	Stratagene
pTBL1	tetracycline	T7 expression plasmid with N-terminal His ₆ -tag	this study
pTBL1	tetracycline	T7 expression plasmid with N-terminal His ₆ -tag and 3C protease cleavage site	this study

2.4.3 Growth of plasmid-containing *E. coli*

Growth of plasmid-containing *E. coli* cells was achieved in LB or on LB agar plates containing the appropriate antibiotic. Isopropyl β -D-thiogalactoside (IPTG) was added to the medium at a final concentration of 0.8 mM to induce expression of the desired gene.

2.4.4 Growth of cells and isolation of plasmids

This was carried out using standard techniques of inoculating a 1.5 ml overnight culture of LB (containing appropriate antibiotics), with a single colony of the correct strain. In the case of low copy number plasmids the overnight culture volume was 5 ml. These were then grown overnight in a shaking incubator at 37°C. Plasmid isolation was performed using a QIAprep Spin Minirep Kit (Qiagen Ltd. See Appendix II) as described in the manufacturer's instructions, and the resulting plasmid DNA resuspended in 30 or 50 μ l of sterile water.

2.4.5 Production of competent cells

2.4.5.1 Protocol of Hanahan [Hanahan, 1985]

A single colony of an *E. coli* strain (or 1 μ l from a glycerol stock) was grown overnight at 37°C in 10 ml of LB. 6 ml of this overnight culture was then used to inoculate 74 ml LB in a 250 ml flask. The cells were grown at 37°C on a microbiological shaker at 200 rpm until they reached mid log phase ($A_{600\text{nm}}=0.35-0.5$). The cells were chilled on ice for 15 mins and kept at or below 4°C throughout the rest of the procedure. The culture was transferred to chilled centrifuge tubes and centrifuged at 3000 rpm in a Beckman J2-21 centrifuge for 15mins to pellet cells. The pellet was resuspended in 13 ml of ice cold RF1 (see Table 2-3). The cells were then incubated on ice for 60 mins and re-pelleted. The pellet was resuspended in 3 ml of RF2 (see Table 2-3). Cells were aliquoted into 200 μ l aliquots and used for transformations or flash frozen in a dry-ice/ethanol bath and stored at -80°C.

Competent cells can be stored at -80°C for an indefinite period without showing any significant reduction in transformation efficiency.

2.4.5.2 Protocol using CaCl_2

For transformation of purified plasmid RIL and RP in BL21(DE3) and C41(DE3) Mg/Ca competent cells were used. The protocol is adapted from Sambrook *et al.* (1989).

A single colony of an *E. coli* strain (or 1 μl from a glycerol stock) was grown overnight at 37°C in 1 ml of LB. 350 μl of this overnight culture was then used to inoculate 5 ml LB in a 15 ml tissue culture tube. The cells were grown at 37°C on a microbiological shaker at 200 rpm until they reached mid log phase ($A_{600\text{nm}}=0.35-0.5$). The cells were chilled on ice for 10 mins and kept at or below 4°C throughout the rest of the procedure. The culture was transferred to chilled centrifuge tubes and centrifuged at 3000 rpm for 10 mins to pellet cells. The pellet was resuspended in 1.5 ml of ice cold MgCl_2 - CaCl_2 buffer (see Table 2-3). The cells were pelleted again and resuspended in 200 μl of RF2 (see Table 2-3). Cells were immediately used for transformation.

2.4.6 Transformation protocol

An aliquot of frozen cells was allowed to thaw on ice prior to use. DNA was added to the cells and following gentle mixing the cells were incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 90 s before being returned to ice for a further 15 minutes. 800 μl of SOC medium⁴ at room temperature was added to the cells which were then incubated at 37°C for 40-60 minutes to aid recovery. The cells were pelleted for 2 mins at high speed and resuspended in 75 μl of sterile distilled water and then plated out onto selective media.

When competent cells from commercial suppliers were used instead of self-prepared ones the transformation was performed according to the manufacturers protocol.

⁴ in case pure plasmid was to be transformed LB was used instead of SOC medium

2.5 Manipulations of DNA

2.5.1 Gel electrophoresis

Mini-gels ranging from 0.5-2%(w/v) agarose content were supplemented with 0.05 µg/ml ethidium bromide for visualisation. 1X TBE or 1X TAE buffer (Table 2-3) was used for the preparation, the first for analytical purposes, the latter for extraction of DNA. Samples were loaded in a 4:1 volume ratio with DNA gel sample loading buffer (Table 2-3). A standard 1 kb ladder (Gibco-BRL) and DNA Molecular Weight Marker XIV (Boehringer) was used on all gels as a marker to identify the size of the separated bands. The gels were run at a constant voltage (typically 100 V) at room temperature until the bromophenol blue marker approached the end of the gels; DNA bands were located by viewing under UV light. Photographs were taken where appropriate on a UVP BioDoc-It™ System (Fisher Scientific).

2.5.2 Restriction digests

Analytical restriction digests contained 200 ng of DNA with 0.5-1 units of each enzyme along with the appropriate reaction buffer. Digests were carried out in a volume of 10 µl at 37°C for one hour.

Digests to obtain fragments for cloning contained 5-10 µg of DNA, 5-10 units of each restriction enzyme and the appropriate buffer. The digests were carried out in a volume of 50 µl at 37°C for at least 3 hours, preferably overnight. When double digests contained *NdeI*, this enzyme was added first and in small aliquots every 30 minutes before adding the second restriction enzyme. This was done since *NdeI* is of limited stability.

2.5.3 Gel purification of DNA fragments

The restricted plasmid or DNA fragment was separated on a low melting point TAE agarose gel and the band excised. The DNA was isolated using a QIAquick Gel Extraction Kit (Qiagen Ltd. See Appendix III) as described in the manufacturer's instructions. The resulting DNA was resuspended in 50 µl of sterile water. When visual inspection suggested a poor yield, the elution volume was lowered to 30 µl to yield DNA at an appropriate concentration.

2.5.4 Alkaline phosphatase dephosphorylation

Alkaline phosphatase (Promega) was used to dephosphorylate restricted plasmid DNA. 1 μl calf intestinal alkaline phosphatase (CIAP) (10 U/ μl) was added to the plasmid digest and digested as described above. This was only used to dephosphorylate vector DNA (not insert DNA).

2.5.5 DNA clean up methods

Sometimes it was necessary to purify further DNA from digested plasmid, insert DNA or PCR products in order to remove traces of restriction enzymes and other contaminating enzymes such as alkaline phosphatase. This was performed using the QIAquick PCR purification protocol (See Appendix III). The DNA was eluted into 30 or 50 μl of sterile water.

2.5.6 DNA ligation

The concentration of DNA was estimated by comparison with a known amount of molecular weight marker and known standard yields of DNA purification.

In the standard case approximately 0.5 μg of plasmid DNA and 1 μg of insert DNA was used per ligation reaction. DNA ligation reactions were carried out using 1 μl T4 DNA ligase (1 U/ μl) and 2 μl of the appropriate 10x reaction buffer along with plasmid and insert DNA. All ligations were carried out for 16 hours at room temperature in a final volume of 20 μl .

Ligations to produce deletion mutants were performed in a similar matter to the manufacturer's instructions for use in the Exsite Mutagenesis Kit (Stratagene). To 30 μl of purified PCR product (after *DpnI* digestion) 100 μl dsH₂O, 10 μl 10x cloned *Pfu* buffer and 5 μl rATP (10 mM) were added. 10 μl of that solution were withdrawn and 1 μl of T4 DNA ligase (4 U/ μl) added. The ligation was carried out for 1 hour at 37°C.

2.6 Cloning by PCR

2.6.1 Primers for PCR

2.6.1.1 Primers for TOPO cloning

PCR primers were designed to amplify the desired gene from genomic DNA. No special restriction sites were introduced. The forward primer contained extra bases which encode for the His₆-tag or the 3C protease cleavage site (see Table 2-8). The reverse primer contained the stop codon of the gene. The start codon of the gene was omitted from the primer after the extra bases.

2.6.1.2 Primers for conventional cloning

PCR primers were designed to incorporate two restriction sites to facilitate cloning into the expression vector. The sites used were *Kpn*I and *Nsi*I. *Kpn*I was used for the 5' end of the gene and is located directly following the ATG start codon of the cloning vector pTB361 (see Table 2-8).

2.6.1.3 Mutagenesis primers

PCR primers were designed on the basis of the vector pTBL1:*Hin*hisG. The 5' end of the reverse primers for the short forms was the stop codon. The forward primers started 5' at the last required residue of the gene *Hin*hisG running towards the 5' end of the gene. For the C-terminal domain the reverse primer 5' end was the first residue of the C-terminal domain. The forward primer started at the 3' end of the His₆-tag going in 5' direction of the tag. Both reverse primers were 5' phosphorylated for following ligations (see Table 2-10).

2.6.1.4 Linker oligonucleotides

Two complementary oligonucleotides to produce DNA linkers were designed so that the double stranded form would have sticky ends (Table 2-9). These sticky ends are compatible with the restriction sites *Nde*I and *Xho*I, making it possible to clone it into a vector restricted with *Nde*I and *Xho*I.

2.6.2 Genomic DNA

Genomic DNA from various bacterial organisms was used as DNA template for PCR reactions. Table 2-7 shows which strains were used and the source of the genomic DNA.

Table 2-7 Genome Strains

Organism	Abbreviation	Strain	Source
<i>Bacillus subtilis</i>	<i>BS</i>	168	Alex Herbert, Glasgow University
<i>Haemophilus influenzae</i>	<i>Hin</i>	KW20	American Type Culture Collection (ATCC), USA
<i>Mycobacterium tuberculosis</i>	<i>Mtb</i>	H37Rv	David Robinson/Glaxo Wellcome
<i>Methanobacterium thermoautotrophicum</i>	<i>Mth</i>	ΔH	Brian Hanzelka, Ohio State University, USA

2.6.3 PCR reactions

2.6.3.1 PCR amplification of genes from genomic DNA

PCR reactions were carried out using Vent DNA polymerase along with appropriate buffer. 2 units of enzyme were added per reaction along with 5 μl of 10X Vent reaction buffer, 1 μl of 10X dNTP mix (10 mM stock of each dNTP), 50 pmol of each primer, 100 ng of template DNA (genomic DNA, see Table 2-7) and also MgSO₄ in concentrations varying from 0-3 mM (final concentration). 50 μl reactions were set up in 0.2 ml PCR tubes without the polymerase. Polymerase was added after 3 minutes of the initial denaturation step for hotstart PCR.

The following temperature profile was used for 30 cycles in a Biorad Gene Cyclor:

- 5 min @ 94° C (initial denaturation)
- 1 min @ 94° C
- 1 min @ 55° C (annealing temperature varied, see Table 3-2)
- 1 min @ 72° C

A final elongation step of 5 minutes at 72°C was performed at the end of the amplification reaction to ensure that all amplified material was full length.

2.6.3.2 PCR for amplification of deletion mutants ('long PCR')

PCR reactions were carried out using *Pfu* turbo DNA polymerase along with the appropriate buffer. 2.5 units of enzyme were added per reaction along with 5 µl of 10X cloned *Pfu* reaction buffer, 1 µl of dNTP mix (25 mM stock), 50 pmol of each primer (Table 2-10), 20 ng of template DNA (pTBL1:*HinHisG*). To enhance the amplification of these long DNA products the method of 'long PCR' was used [Cheng *et al.*, 1994]. The final reactions contained 2%(v/v) DMSO and 8%(v/v) glycerol. 50 µl reactions were set up in 0.2 ml PCR tubes.

The following temperature profile was used for 30 cycles in a Biorad Gene Cycler:

2 min @ 95° C (initial denaturation)
30 sec @ 95°C
30 sec @ 60°C (annealing temperature, varied)
12 min @ 72°C

A final elongation step of 20 minutes at 72°C was performed at the end of the amplification reaction to ensure that all amplified material was full length.

2.6.3.3 A-tailing of blunt end PCR products

To produce 3' A-tails on blunt end PCR products the PCR reaction was treated with Taq DNA polymerase. For PCR amplifications with single bands, 1 µl Taq DNA polymerase was added to the PCR reaction. Agarose-gel purified PCR products were supplemented with the appropriate amount of 10x Taq reaction buffer, 1 µl of dNTP mix (100 mM dATP) and 1 µl Taq polymerase. The reaction was incubated at 72°C for 20 minutes.

Table 2-8 Primers used for PCR based molecular cloning

Oligonucleotide	Sequence	Comments
BShisGF	5' ATG CAT CAT CAT CAT CAT GGT AAG TTA CTC ACA ATG GCG ATG CCA	Adds a N-terminal His ₆ -tag to <i>BS</i> HisG and allows cloning into pUni
BShisGR	5' TCA TTT TGC CGT CTC TCC TTC CAC AAC	
BShisIEF	5' ATG CAT CAT CAT CAT CAT AAA CAG GCA GAT GAA CTG CGT TTT AAC	Adds a N-terminal His ₆ -tag to <i>BS</i> HisIE and allows cloning into pUni
BShisIER	5' TCT TTA CTC TTC AAT CTC AGA ATG GCG	
BShisZF	5' ATG CAT CAT CAT CAT CAT TTT ATG TTT GAA AAA CCG CAC GGC ATG	Adds a N-terminal His ₆ -tag to <i>BS</i> HisZ and allows cloning into pUni
BShisZR	5' TTA CCC ATT TTG CTC TTC CTT TCT GGC	
HinHisGF	5' ATG CAT CAT CAT CAT CAT ACA AAC ACA ACA ATG CAA CCA AAC CGC	Adds a N-terminal His ₆ -tag to <i>Hin</i> HisG and allows cloning into pUni
HinHisGR	5' TTA CTT CAT CAT CTT CTC AAT TGG TAA CAC	
HinHisIEF	5' ATG CAT CAT CAT CAT CAT AAT ATT ACT AAA ATA GAC TGG CAA AAA	Adds a N-terminal His ₆ -tag to <i>Hin</i> HisIE and allows cloning into pUni
HinHisIER	5' TTA TTT ATT TCC TCC TTC TGG ATG ATA	

MfbHisGF	5' ATG CAT CAT CAT CAT CAT CAT CTG CGG GTC GCG GTT CCC AAC AAG GGT	Adds a N-terminal His ₆ -tag to <i>Mfb</i> HisG and allows cloning into pUni
MfbHisGR	5' TCA GAA TCG GCA GAA CCT GAT GTC CGA CG	
MfbHisIF	5' CTC GAG GTA CGC GGC CAA GGT CCG ACA CTC GAC CCA AAG ATC GCG GCG CGG	Adds a N-terminal 3C protease cleavage site to <i>Mfb</i> HisI and allows cloning into pUni
MfbHisIR	5' TTA GTC GTC GGG TTC TAA CAA CAC CGC GGC	
MfbHisEF	5' CTC GAG GTA CGC GGC CAA GGT CCG CAA CAA TCG CTG GCC GTG AAG ACC TTC	Adds a N-terminal 3C protease cleavage site to <i>Mfb</i> HisI and allows cloning into pUni
MfbHisER	5' TCA CAG CTT TCG ATA GAC GTC GTC GAG G	
MthHisGF	5'GGT ACC AAA GAT AAG AAT AGC GGT TCC CTC CAA G	For cloning of <i>Mth</i> HisG in pTB361
MthHisGR	5'ATG CAT CAG GGT ATT ATC CTC TCA ATG GGC AC	
MthHisGreIF	5'GGT ACC AAA AGC TAT GAA GAT AGT TCT TG ATT A	For cloning of <i>Mth</i> HisG-related protein in pTB361
MthHisGreIR	5'ATG CAT CTA TTC TAT ATA CTG TTT GAC GTT TCC C	

Table 2-9 Oligonucleotide to create linkers

Oligonucleotide	Sequence	Comments
pTB361hisF	5' TAT GCA TAA GTC AAT AGA GC	For linker to create pTBL1
pTB361hisR	5' TCG AGC TCT ATT GAC TA TGC A	
pTB361cleaveF	5' T ATG CAT CAT CAT CAT CAC CTC GAG GTA CTC TTC CAA GGT CCG AGC	For linker to create pTBL2
pTB361cleaveR	5' TCG AGC TCG GAC CTT GGA AGA GTA CCT CGA G GTG ATG ATG ATG ATG CA	

Table 2-10 Oligonucleotides for PCR to create deletion mutants

Oligonucleotide	Sequence	Comments
HinGshortF	5' TGC CTG CTGCAC ACC TTG AAT ACG AG	For cloning of short form of HisG from <i>Hin</i>
HinGshortR	5' TAA AAG GGC AAT TCG GGA GCT CGA GTC G	5' phosphorylated; to be used with both short and short294 forward primer
HinGCtermF	5' ATG ATG ATG ATG ATG ATG CAT ATG TAT ATC	For cloning of C-terminal domain of HisG from <i>Hin</i>
HinGCtermR	5' AAA TAT ATT ATG CTC CAC GCT CCG A	5' phosphorylated
HinGshort294F	AAT TGG TAA CAC TAA AAT TGA ACT AGC TCC TGC CTC	For cloning of truncated form of HisG from <i>Hin</i> with last 5 residues deleted

2.7 Over-expression analysis

2.7.1 Over-expression studies

2.7.1.1 *Quick test of expression*

A single colony of the over-expression clone was grown in 10 ml LB for about 8 hrs at 37°C. Expression was then induced in 5 ml culture with the addition of IPTG to a final concentration of 0.8 mM. Samples were then grown for further 16 hrs at 37°C and then harvested.

2.7.1.2 *Time course of expression*

A 2 ml overnight culture, inoculated in LB, was grown overnight at 37°C. A 10 ml culture was inoculated with 675 μ l of this overnight culture. This was then grown to an $A_{600\text{nm}}$ of 0.6 before inducing expression with the addition of IPTG to a final concentration of 0.8 mM. 1 ml aliquots were removed at various time points and the $A_{600\text{nm}}$ measured and plotted. These samples were pelleted and the cells resuspended in 10 μ l of sample buffer for each 0.1 A unit. These were then subjected to SDS PAGE analysis to observe the band of over-expressed protein.

2.7.2 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-polyacrylamide gel electrophoresis was performed by the method of Laemmli [Laemmli, 1970], with a 5% stacking gel and 10 to 15% running gel (depending on the molecular weight of the protein of interest). The ratio of acrylamide to bisacrylamide in all PAGE experiments was 30:0.8 and polymerisation was induced by the addition of 0.03% (v/v) TEMED and 0.05% (w/v) ammonium persulphate. Samples were denatured by boiling for at least 5 minutes after dilution in SDS-PAGE sample buffer (Table 2-3). After electrophoresis the gels were stained for protein by the Coomassie blue method (see Section 2.7.3).

2.7.3 Staining for Protein

Gels were stained for protein with Coomassie blue for 30 minutes at 37°C. The Coomassie reagent consisted of 0.1% (w/v) Coomassie brilliant blue G250 in 50% (v/v) methanol and 10% (v/v) glacial acetic acid. Destaining was carried out in

10% (v/v) methanol and 10% (v/v) glacial acetic acid at 40°C until the background was fully destained and the bands clearly visible.

2.8 Protein Purification

2.8.1 Growth of cells for protein purification

The appropriate *E. coli* over-expression strain containing the over-expression constructs were inoculated from a glycerol stock into LB containing tetracycline and chloramphenicol and grown overnight at 37°C. The following morning, five flasks, each containing 500 ml of LB plus antibiotics⁵, were inoculated with 40 ml of this overnight culture and growth continued at 37°C to an A_{600nm} of 0.5-0.6. The cells were then induced by adding IPTG to a final concentration of 0.8 mM and grown for a further 4-5 hours before harvesting in a MSE 2L centrifuge. The cell pellet was either used immediately or stored at -20°C until required.

2.8.2 Preparation of Cell Extract

Cells of the *E. coli* strains obtained from over-expression were thawed slowly on ice and resuspended in 20 ml of the appropriate lysis or resuspension buffer (see Table 2-11 and Table 2-12). The cell paste was subjected to three passes through the French press at 950 psi. 0.1 mg of DNaseI was added and the extract stirred at 4°C for 30 minutes before centrifuging at 18,000 rpm for 1 hour at 4°C.

2.8.3 Purification of *E. coli* HisG

The purification procedure of HisG from *E. coli* is described in Lohkamp *et al.* (2000). The purification of the Se-methionine labelled protein was identical except that 0.8 mM DTT was used instead of 0.4 mM.

2.8.4 Purification of HisG from *Mth*

Cells containing over-expressed HisG from *Mth* were resuspended and lysed in buffer A. To purify HisG from *Mth* a heat step was first carried out. The soluble lysate was incubated at 65°C for 30 min. The precipitate was removed by

⁵ Usually only tetracycline; chloramphenicol was omitted in pLysS strains to boost expression, but present for RIL or RP strains.

centrifugation at 10 krpm for 1 h at 4°C. The pellet was discarded and the supernatant loaded onto a Q-Sepharose anion exchange column (30 ml bed volume, flow rate 1.5 ml/min) which was equilibrated with buffer B. The column was washed with two column volumes of buffer B. Protein was eluted with a 140 ml linear gradient of 100% buffer B to 50% buffer C. HisG eluted between 64 and 100% buffer C (340-500 mM NaCl). These fractions were pooled and dialysed against buffer B. The protein solution was loaded on a reactive green 19 dye column (50 ml bed volume, flow rate 1 ml/min) that was equilibrated with buffer B. The column was washed with two column volumes of buffer B and bound protein was eluted stepwise in two column volumes at three different concentrations of sodium chloride (250 mM, 500 mM and 1 M; 25%, 50%, 100% buffer C). HisG eluted in fractions at 500 mM and 1 M sodium chloride. The fractions containing HisG were pooled and dialysed against buffer B.

Table 2-11 Buffers for purification of *Mth* HisG

Buffer	Constituents	
Buffer A (resuspension buffer)	50 mM	Tris-HCl pH 8.5
	0.4 mM	DTT
	2 protease inhibitor tablets per 100 ml of buffer	
Buffer B (wash buffer)	50 mM	Tris-HCl pH 8.5
	0.4 mM	DTT
	1 protease inhibitor tablet per litre of buffer	
Buffer C (elution buffer)	50 mM	Tris-HCl pH 8.5
	0.4 mM	DTT
	1 M	NaCl
	1 protease inhibitor tablet per litre of buffer	

2.8.5 Purification of His_x-tagged proteins

A 10 ml Ni-NTA agarose column was pre-equilibrated with at least 3 column volumes of Buffer A (Table 2-12) prior to loading the protein. The column was loaded at a flow-rate of 1 ml per minute and then washed with 3 column volumes of buffer B before eluting with a linear gradient of 0-100% buffer C. 2.5 ml fractions were

collected at a constant flow-rate of 1 ml/min and the $A_{280\text{nm}}$ recorded on a chart recorder.

Table 2-12 Buffers for Ni-NTA column

Buffer	Constituents	Comment
Buffer A (lysis buffer)	50 mM NaH_2PO_4 pH 8.0 300 mM NaCl 10 mM imidazole 2 protease inhibitor tablets per 50 ml of buffer	pH adjusted with NaOH
Buffer B (wash buffer)	50 mM NaH_2PO_4 pH 8.0 300 mM NaCl 20 mM imidazole 1 protease inhibitor tablet per litre of buffer	pH adjusted with NaOH
Buffer C (elution buffer)	50 mM NaH_2PO_4 pH 8.0 300 mM NaCl 250 mM imidazole 1 protease inhibitor tablet per litre of buffer	pH adjusted with NaOH

2.8.6 Storage of purified enzymes

The buffer containing the purified enzymes was carefully mixed with an equal volume of glycerol so that the solution contained 50% (v/v) glycerol. This stock solution was stored at -20°C .

2.9 Enzyme assay

2.9.1 Assay for ATP-Phosphoribosyltransferase (HisG)

The enzyme activity was determined by monitoring, in a 1 cm path length cuvette, the increase in absorbance at 290 nm at 25°C . The assay mixture contained 150 mM KCl, 100 mM Tris-HCl pH 8.5, 10 mM MgCl_2 , 5 mM ATP and 0.5 mM PRPP. The enzyme assay was carried out in a final volume of 1 ml.

To obtain kinetic parameters for wild type (WT) and mutant HisG from *Hin*, assays were carried out at five concentrations of ATP (0.05, 0.15, 0.5, 1, 2 mM) and

PRPP (0.025, 0.05, 0.1, 0.15, 0.25 mM). The pH of the 25 mM ATP stock solution was adjusted to 8.5 with NaOH. Additional assays were performed with ATP concentrations up to 16 mM, but were not included in the analysis. The enzyme concentration of *Hin* WT and deletion mutant enzyme was 1.95 μM (see Section 3.4). Each measurement was carried out at least in duplicate. Individual data points were averaged and the statistical error included in the fit procedure. The data were fit to the Michaelis-Menten equation by least-square regression using Origin software to calculate k_{cat} and K_{M} . A molar extinction coefficient at 290 nm of $\epsilon=3600 \text{ M}^{-1}\text{cm}^{-1}$ for PR-ATP was used [Smith and Ames, 1964].

Inhibition experiments with histidine were carried out as described above in presence of 500 μM histidine. The assay mixture including enzyme, ATP and histidine was incubated at 25°C for 3 minutes before the reaction was started by addition of PRPP.

Assays of *BS* HisG were performed as described above with fixed substrate concentrations (5 mM ATP and 0.5 mM PRPP) and an enzyme concentration of 1.47 μM . The enzyme was incubated in assay buffer together with the ATP for 10 min at the desired temperature. During incubation the absorbance at 290 nm was measured as a reference which was subsequently subtracted from the activity measurement. The assay was started by addition of PRPP. The temperature during the assay was controlled with a heating water bath.

Assays of *Mth* HisG and HisG-related protein were routinely carried out at 65°C. The solution of Tris-HCl buffer, KCl and MgCl_2 was incubated at the higher temperature for 5 min before addition of ATP, PRPP and the enzyme.

2.9.2 Assay for short HisG and HisZ

The enzyme activity of the complex was determined as described above with the addition of 0.4 mM DTT to the assay mixture. Prior to measurements the reaction mixture (without the substrates) was incubated on ice for 1 hr.

2.9.3 HisIE enzyme activity

The activity of HisIE enzyme was monitored by measuring the increase in absorbance at 290 nm at room temperature in a 1 ml cuvette. This corresponds to the formation of the product Pro-FAR with a molar extinction coefficient at 290 nm of

$\epsilon=8000\text{M}^{-1}\text{cm}^{-1}$ [Martin *et al.*, 1971]. The assay mixture contained 50 mM Tris-HCl pH 8.5, 10 mM MgCl_2 and 50-100 μM PR-ATP (final concentrations).

2.10 Crystallisation

2.10.1 Dynamic light scattering (DLS)

DLS experiments were performed to assess the monodispersity of the protein samples prior to crystallisation and to investigate the aggregation state of the enzymes. Experiments were carried out using a DYNA-PRO, 801 dynamic light scattering/molecular sizing instrument (Protein solutions, Buckinghamshire, UK). Usually protein solutions (1 mg/ml) were in 50 mM Tris-HCl, pH 7.5 and 100 mM KCl but several other buffers and pHs were also assessed. Experiments were carried out at 20°C and protein solutions were filtered using a 0.2 μm membrane filter.

2.10.2 Crystallisation screens

Various different sparse matrix screens [Jancarik and Kim, 1991] were used for crystallisation of enzymes (see Table 2-13). Screens were either purchased from the mentioned companies or prepared according to the composition given by the companies.

Table 2-13 Crystallisation screens

Trade Name	Common name	Company
Crystal Screen	Magic 50	Hampton Research
Crystal Screen 2	B-Screen	Hampton Research
Additive Screen 1-3		Hampton Research
Wizard I	W-Screen	Emerald BioStructures
Wizard II	Z-Screen	Emerald BioStructures
Cryo I		Emerald BioStructures
Cryo II		Emerald BioStructures
Structure Screen 1	MD I	Molecular Dimensions
Structure Screen 2	MD II	Molecular Dimensions
-	N-Screen	In house screen
-	M-Screen	In house screen

2.10.3 Crystallisation methods

Crystallisation of protein was carried out by the method of vapour diffusion. Crystallisations were set up as sitting drops using Cryschem Plates (Hampton Research, Aliso Viejo, USA). The wells were sealed with Crystal Clear Sealing Tape (Hampton Research) and crystallisations were stored at 20°C.

For initial screening purposes 0.75 ml of well solution and a 2 µl drop was used. The drop contained 1 µl of well solution and 1 µl of protein solution. Refined conditions used 1 ml of well solution and a 4(2+2) µl drop volume unless otherwise stated.

To improve the diffraction quality of crystals small quantities of additional compounds were used to grow crystals. Additive Screens 1-3 from Hampton Research (a total of 72) were used according to the manufacturer's instructions.

The Seed Bead Kit™ from Hampton Research was used for seeding crystals when necessary and used according to the manufacturer's instructions.

I. Appendix: QIAprep Spin Miniprep Kit (Qiagen)

This kit allows the rapid isolation of plasmid DNA by an alkaline lysis method. The DNA is bound to a silica based resin in the presence of high salt and following washing to remove contaminants can be eluted in low salt buffer. There is no need for organic solvent extractions and the DNA is of suitable quality for use in DNA sequencing and other molecular biological applications.

Step 1 Pellet 1.5-5 ml of an overnight culture

Step 2 Resuspend cells in 250 μ l resuspension buffer P1

Step 3 Add 250 μ l cell lysis solution (P2) and mix until a clear lysate is formed

Step 4 Neutralise by adding 350 μ l potassium acetate solution (N3) and mix

Step 5 Spin to pellet insoluble material for 15 minutes at full speed in a microfuge

Step 6 Load supernatant on QIAprep column

Step 7 Centrifuge briefly and discard flow-through

Step 8 (Optional, only done when using *endA*⁺ strains) Wash with 0.5 ml PB buffer, centrifuge briefly and discard flow-through

Step 9 Wash with 0.75 ml PE buffer, centrifuge briefly and discard flow-through

Step 10 Spin the column in a microfuge for 1 min to remove any residual buffer and elute DNA with 30 or 50 μ l of sterile water (if plasmid was larger than 4 kbp the water was heated to 65°C).

II. Appendix: QIAquick PCR Purification Protocol (Qiagen)

This kit can be used as an alternative to organic solvent extractions and ethanol precipitations for the purification of PCR products from contaminating nucleotides, enzymes and primer dimers etc. or DNA from contaminating restriction enzymes or phosphatases.

The DNA is bound to a silica based resin in the presence of high salt and is eluted in low salt buffer in a similar manner to the previous protocol.

This protocol was used with QIAquick columns from QIAquick Gel Extraction kit and buffers from the QIAprep kit.

- Step 1 Take 1 volume of restriction digest or PCR product and add 5 volumes of PB buffer.
- Step 2 Load supernatant on QIAquick column
- Step 3 Centrifuge briefly and discard flow-through
- Step 4 Wash with 0.75 ml PE buffer, centrifuge briefly and discard flow-through
- Step 5 Spin the column in a microfuge for 1 min to remove any residual buffer and elute DNA with 30 or 50 μ l of sterile water (if plasmid was larger than 4 kbp the water was heated to 65°C).

III. Appendix QIAquick Gel Extraction Kit (Qiagen)

Allows the purification of DNA from a TAE agarose gel.

- Step 1 Separate DNA by electrophoresis on a TAE low melting point agarose gel and excise the relevant band
- Step 2 Add 3 volumes of QG buffer to 1 volume of gel (100 mg ~ 100 μ l).
- Step 3 Incubate the sample for 10 min at 50°C to melt the agarose.
(**Optional**, only for DNA <500 bp or >4 kbp) Add 1 gel volume of isopropanol and mix
- Step 4 Load supernatant on QIAquick column
- Step 5 Centrifuge briefly and discard flow-through
- Step 6 Wash with 0.5 ml QG buffer (preheated to 50°C), centrifuge briefly and discard flow-through
- Step 7 Wash with 0.75 ml PE buffer, centrifuge briefly and discard flow-through
- Step 8 Spin the column in a microfuge for 1 min to remove any residual buffer and elute DNA with 30 or 50 μ l of sterile water (if the plasmid was larger than 4 kbp the water was heated to 65°C).

Chapter 3: Cloning, Over-expression, and Mutational Analysis of HisG, HisIE and HisZ from Multiple Microorganisms

3.1 Introduction

Milligram quantities of protein are required for structural investigations such as protein crystallography. Cloning of the target genes and subsequent over-expression of the gene product is a convenient and relatively inexpensive method to obtain large amounts of protein that are naturally produced in very small quantities.

Here the aim was to over-express the first two (three), *hisG*, *hisI(E)*, genes from several organisms. The sequence variation between organisms significantly increases the probability of obtaining soluble, well-behaved protein suitable for growing diffraction-quality crystals which are necessary for successful structural studies. Characterization of the same enzyme from different species also provides an opportunity to investigate the role of structural differences in function. In order to study the early steps of histidine biosynthesis in detail it was decided to clone and over-express the *hisG* and *hisIE* (or where applicable *hisI* and *hisE*) genes from several microorganisms. The genes, organisms, and reasons for selecting them are summarised in Table 3-1.

Table 3-1 Histidine pathway genes selected for cloning

Organism	Gene	Reason for selection
<i>Bacillus subtilis</i> (BS)	<i>hisG</i>	Contains short form of HisG and HisZ
	<i>hisZ</i>	
	<i>hisIE</i>	
<i>Haemophilus influenzae</i> (Hin)	<i>hisG</i>	Pathogenic organism; similar to <i>E. coli</i>
	<i>hisIE</i>	
<i>Mycobacterium tuberculosis</i> (Mtb)	<i>hisG</i>	Pathogenic organism; contains two genes for <i>hisI</i> and <i>hisE</i> rather than a bifunctional <i>hisIE</i>
	<i>hisI</i>	
	<i>hisE</i>	
<i>Methanobacterium thermoautotrophicum</i> (Mth)	<i>hisG</i>	Only bacterial genome known so far to contains two distinct <i>hisG</i> genes
	<i>HisG-related</i>	

All cloned genes were tested for expression of soluble protein in *E. coli* using an IPTG-regulated T7 plasmid expression system.

3.2 Generation of over-expression constructs from *B. subtilis*, *H. influenzae* and *M. tuberculosis*

3.2.1 Introduction

To generate over-expression constructs from *B. subtilis* (*BS*), *H. influenzae* (*Hin*) and *M. tuberculosis* (*Mtb*) the ECHO™ Cloning System from Invitrogen was employed for ease of use and speed in performing multiple clonings simultaneously. All genes had a non-cleavable N-terminal His₆-tag fused for rapid purification. The exceptions were *Mtb hisI* and *hisE* which had cleavable His₆-tags included as they are small proteins and 6 extra charged residues could alter the properties of the enzyme more significantly.

Unfortunately the use of the ECHO™ Cloning System was very restricted and will be discussed in Section 3.2.5. Therefore a conventional cloning approach using a modified and well-characterized over-expression vector was employed (see Section 3.2.6).

3.2.2 T7 expression vectors

T7 expression systems allow the efficient over-expression of a protein in a controlled manner in *E. coli*. The cloning vector has a promoter for T7 RNA polymerase located upstream from a multiple cloning site which is used to clone the insert DNA. T7 RNA polymerase is able to transcribe DNA five times faster than *E. coli* RNA polymerase and as it is specific for T7 promoters it results in induction of high level expression from such promoters only [Chamberlin and Ring, 1973; Golomb and Chamberlin, 1974]. This allows the specific expression of the gene product of interest. Bacterial strains for over-expression usually contain the gene for the T7 RNA polymerase integrated into the chromosome and are labelled (DE3). The bacterial strains used for expression purposes can carry a gene for T7 lysozyme which binds to and inhibits T7 RNA polymerase [Moffatt and Studier, 1987]. This is used as a means of controlling basal levels of protein expression prior to induction with IPTG. This is particularly important if the protein is toxic to the bacterial strain. The gene for the T7

lysozyme is present on the pLysS plasmid [Studier and Moffatt, 1986], which also contains a chloramphenicol resistance gene and is transformed separately into the expression host.

In case the protein or gene is toxic to the host *E. coli* strain it can be useful to use a special over-expression strain such as C41(DE3) [Miroux and Walker, 1996]. This strain is a mutation of the well-characterised BL21(DE3) over-expression strain which was selected based on resistance to toxic genes or over-expressed proteins. It shows a delay in target mRNA production and an overall decrease in target mRNA levels.

Differences in the codon usage by different organisms can also profoundly affect the heterologous expression of proteins due to the presence of codons rarely used by *E. coli*. Genes encoding the tRNAs of codons rarely used in *E. coli* have been cloned into plasmids to help improve expression levels. Strains are available for rare Arg, Ile and Leu tRNAs (RIL) and Arg and Pro tRNAs (RP) usage from Stratagene (BL21(DE3)RIL and BL21(DE3)RP respectively).

3.2.3 Amplification of *his* genes by PCR

The genes of interest were amplified by PCR from genomic DNA using Vent DNA polymerase. The oligonucleotides in the forward direction were designed to include a His₆-tag following the start codon ATG. The exception, *hisI* and *hisE* from *Mtb*, have a 3C protease cleavage site before the gene product. This cleavage site will then be fused to a N-terminal His₆-tag encoded for by the expression plasmid. In all cases the initiation codon of the gene was omitted from the oligonucleotides to avoid an alternative translational start beginning after the tag. For the genes with the cleavage site included, care was taken to avoid frameshifts in the subsequent recombination used by the ECHO™ Cloning System. Oligonucleotides for the reverse PCR direction were matched to the gene including the stop codon. The sequences of the used oligonucleotides are shown in Table 2-8.

PCR amplification of the different genes required different reaction conditions. The conditions which gave the best results concerning a single band and amount of amplified DNA are listed in Table 3-2.

Table 3-2 Conditions for optimal PCR of histidine genes

Organism	Gene	T _m (°C)	[MgSO ₄]/mM	Comments
<i>B. subtilis</i>	<i>hisG</i>	62	0.5	
	<i>hisZ</i>	62	1.0	
	<i>hisIE</i>	62	0.5	
<i>H. influenzae</i>	<i>hisG</i>	62	1.0	Taq DNA Polymerase
	<i>hisIE</i>	56	1.0	
<i>M. tuberculosis</i>	<i>hisG</i>	68	1.0	
	<i>hisI</i>	65	2.0	
	<i>hisE</i>	65	2.0	
<i>M. thermocautotrophicum</i>	<i>HisG</i>	60	1.0	
	<i>HisG rel.</i>	55	2.0	

TOPO cloning, which is part of the ECHO™ system, requires a single 3' A-overhang and is not compatible with DNA amplification using proof-reading enzymes such as Vent DNA polymerase which generates blunt-ended PCR products. To permit cloning into TOPO vectors A-tails were generated by using Taq DNA polymerase to modify the PCR product from the initial amplification reaction.

3.2.4 Cloning of PCR products

3.2.4.1 TOPO® cloning

A schematic diagram of the TOPO reaction using the vector pUni/V5-His-TOPO with 5' T-overhangs indicated is shown in Figure 3-1. Covalently bound to the T-overhang via a phosphate group is the enzyme topoisomerase I, which uses its DNA ligase activity to join free DNA ends together. The availability of substrate with a free 3' A-overhang (the PCR product) results in the enzyme completing its catalytic cycle resulting in the joining of the two lengths of DNA (vector and PCR product) and the subsequent release of the topoisomerase enzyme.

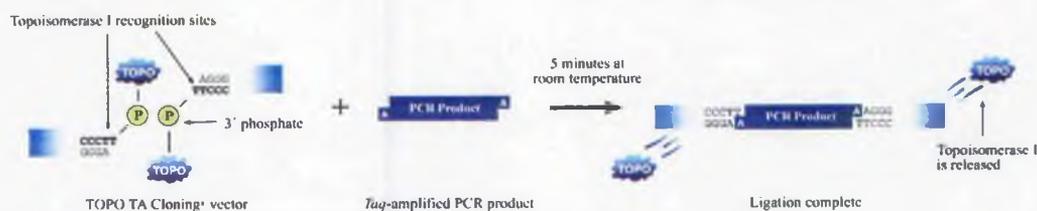


Figure 3-1 Schematic diagram of TOPO reaction.

The TOPO reaction was performed according to the manufacturer's instructions with the exception that the vector was diluted 1:10 as previous experiments had shown that diluting the vector in vector buffer still provided the user with a representative number of clones for analysis (pers. communication N. Hanlon). Comparable results were also obtained when the vector was diluted 1:10 with water. The reactions were transformed in PIR1 cells, plated on LB-agar plates and incubated overnight.

After purification of the plasmid analytical digests were performed on the pUni:*hisX* vectors to confirm the presence of the gene and also to identify the orientation of the gene in the vector, as TOPO cloning is not directional. For the subsequent ECHO cloning the 5' end of the gene has to be located just downstream of the ribosome binding site (RBS) so that the gene can be expressed.

The digests were performed with restriction enzymes which cut the vector as well as the insert. *NsiI* and *SacI* double digests were used for the His-tagged genes and single *XhoI* digests for the genes with the attached cleavage site. The recognition sequence of *NsiI* is included in the His-tag and *XhoI* is part of the 3C cleavage site. *SacI* and *XhoI* are unique restriction sites in pUni and are located closely downstream and upstream respectively of the inserted PCR product (see Figure 3-2). DNA digests were analysed by agarose gel electrophoresis and DNA fragments of the expected size produced, confirming the presence and orientation of the cloned gene (data not shown). DNA digests were analysed by agarose gel electrophoresis and DNA fragments of the expected size produced, confirming the presence and orientation of the cloned gene (data not shown).

The TOPO cloning of *Hin hisG* proved to be difficult and despite repeated attempts no positive clones could be identified. Detailed analysis revealed the absence of 3' A-overhangs on the PCR products as being responsible for the lack of positive

TOPO clones and so the gene was reamplified using Taq DNA polymerase to ensure A-tails. This approach was successful and could prove useful for further cloning.

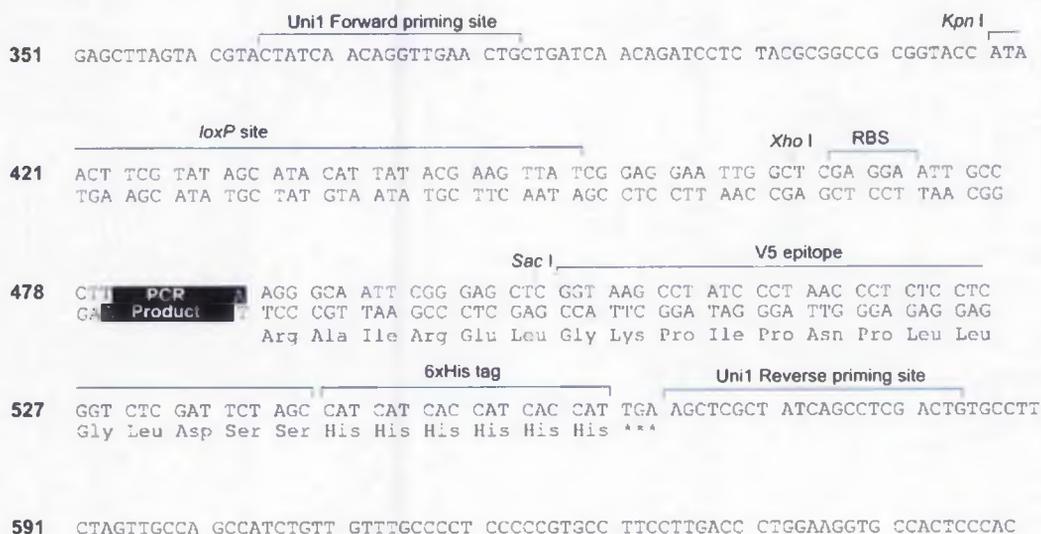


Figure 3-2 Sequence of pUni vector and characteristics in the vicinity of the inserted PCR product.

3.2.5 ECHO™ cloning

The ECHO™ Cloning System produces an over-expression construct by recombination of two vectors. One vector is pUni (donor vector) with the gene of interest cloned. The other is an acceptor vector which can have different promoters for different expression systems. The acceptor vectors used here are pCRT7-E (T7 promoter) and pBAD/Thio-E (arabinose promoter and N-terminal GST fusion) from Invitrogen as well as a pET based vector with a N-terminal His-tag (courtesy of N. Hanlon, Glasgow University). Cre recombinase recombines the two vectors with the use of the internal *lox* sites in both vectors to give the desired recombinant over-expression plasmid. A scheme of this reaction is shown in Figure 3-3.

Recombination reactions were set up according to the manufacturer's instructions and transformed into DH5α and TOP10 (only for pBAD vectors) competent cells. The pUni vectors containing the His-tagged genes were recombined with pCRT7-E and those containing the 3C cleavage site with both the pET vector and the pBAD vector.

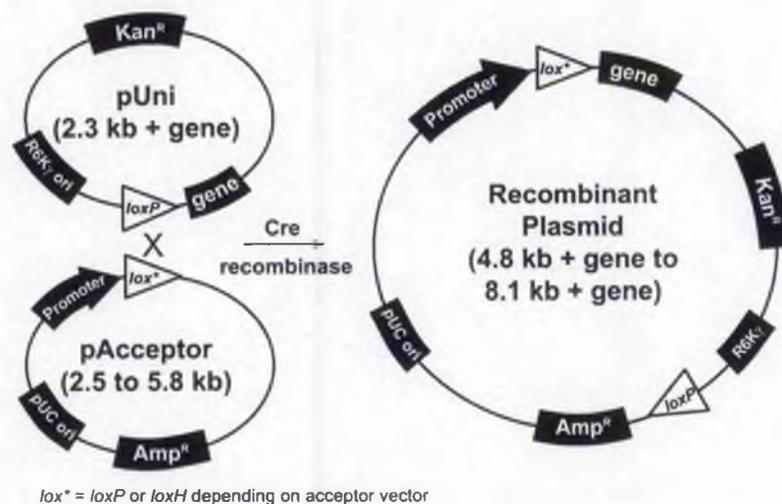


Figure 3-3 Scheme of recombination reaction in the ECHO™ Cloning System (Invitrogen).

Restriction digests of the purified plasmids confirmed that all the recombination reactions were successful and that occasionally multiple recombination events occurred (i.e. more than one acceptor and one donor recombine; data not shown).

The T7-based expression vectors were transformed into competent BL21(DE3)pLysS cells for expression studies. Expression of the pBAD vector can be induced with arabinose in TOP10 cells so no further transformation was required.

No (over-)expression was detected for any of the T7-based over-expression constructs. Various different ways of improving expression (different temperatures, different IPTG concentrations, lower levels of antibiotic etc.) as well as different *E. coli* over-expression strains such as C41 were tested with no success. Furthermore the control pCRT7-E/Uni-CAT provided by Invitrogen showed no over-expression under any of the tested conditions whereas a positive control of a gene in the T7-based vector pTB361 showed over-expression. Despite extensive help from the Technical Service from Invitrogen the problem could not be resolved.

The pBAD based expression constructs were induced with arabinose and showed very low levels of over-expression. The failure of the ECHO Cloning System to produce acceptable results led to the development of an alternative cloning strategy for the over-expression of the *hisX* genes.

3.2.6 'New' cloning strategy

Following the experience with ECHO™ Cloning System an alternative approach for over-expression of the *hisX* genes was employed. It was decided to use a 'classic' cloning strategy by sub-cloning the desired genes directly into the expression vector of choice. Since the existing pUni:*hisX* gene constructs already carried the verified *hisX* genes it was decided to use these plasmids as the source for sub-cloning the *hisX* genes into an expression vector.

To use this approach an over-expression vector with compatible restriction sites to those in pUni:*hisX* was needed. Good candidates for restriction sites would be *Nsi*I and *Sac*I for the His-tagged genes and *Xho*I and *Sac*I for the 3C cleavage site genes. These restriction sites would produce a fragment which includes the whole gene, including the His-tag or part of the 3C cleavage site respectively, and a small fragment of the pUni vector at the 3' end. Furthermore the 5' cloning site should be in close proximity of the RBS to ensure good over-expression. Unfortunately no existing T7-based over-expression vector could be found which fulfilled these criteria. It was decided to produce new over-expression vectors which could be used for subcloning with the existing pUni:*hisX* clones.

3.2.7 Generation of new over-expression vectors

The over-expression vector pTB361, which was obtained courtesy of Dr M. Horsburgh, University of Glasgow, was used as the basis for the design of new expression vectors. It was chosen for expression of the *hisX* genes because it had been used in the laboratory with great success and shows high levels of over-expression. The aim was to introduce a polylinker into the multiple cloning site of pTB361 which includes restriction sites that are compatible with the pUni:*hisX* vectors. Two different new vectors might be needed, one for the already His-tagged genes and one for the 3C cleavage site genes. For the latter sequences encoding a His₆-tag and a 3C cleavage site would need to be introduced along with the linker. It was aimed to make the new vectors as universal as possible so that their use would not be restricted to the cloning of the *his* genes of interest.

A derivative of pTB361 with a gene already cloned into the multiple cloning site was digested with *Nde*I and *Xho*I. This had the advantage that any partial digests of the plasmid could be easily identified and so only completely digested vector DNA was agarose gel purified. Overlapping oligonucleotide linkers were designed so that in

the double stranded form sticky ends complementary to *NdeI* and *XhoI* were produced. The sequence of the primers used to produce the linkers can be found in Table 2-9. Double stranded linkers were produced by mixing the two complementary oligonucleotides, heating the mixture above their annealing temperature and allowing it to anneal slowly to room temperature. This double stranded linker was then ligated into the digested pTB361 using a range of different linker concentrations in the reaction mix. The ligation was then transformed into competent DH5 α cells. The plasmids purified from colonies obtained were tested for insertion of the linker by restriction digest. Digestion of the vector with *SacI* confirmed the presence of the linker as this unique restriction site was newly introduced.

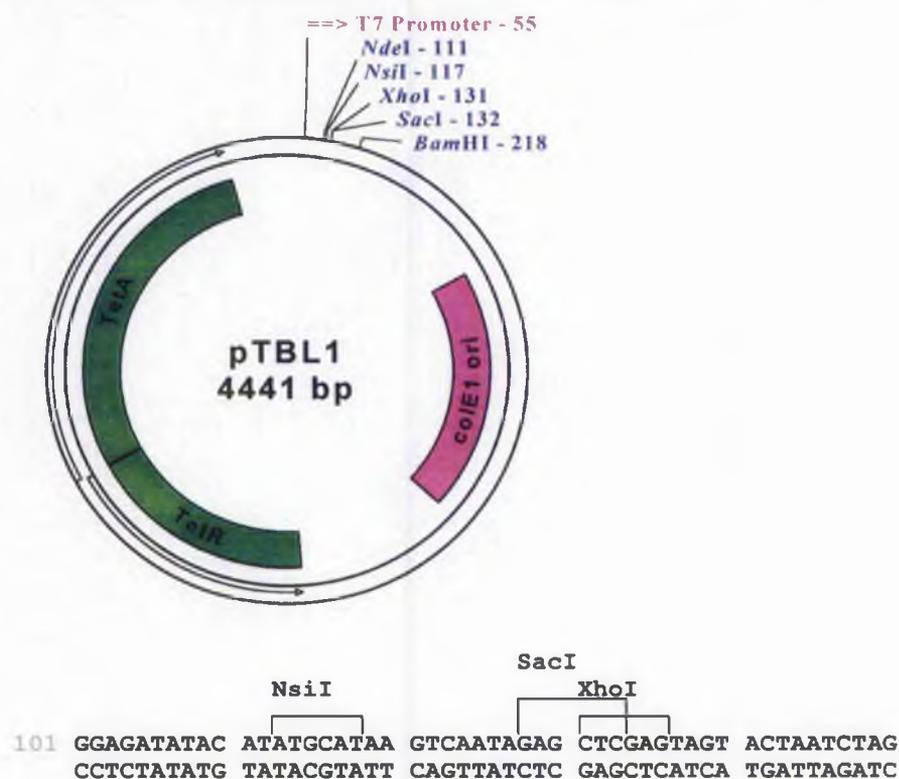


Figure 3-4 Vector diagram and sequence of pTBL1 (the diagram was produced with the program pDRAW).

Schematic diagrams of the two new vectors are shown in Figure 3-4 and Figure 3-5. The vector with the new restriction site only was named pTBL1 and the plasmid with the additional His₆-tag and 3C protease cleavage site was named pTBL2.

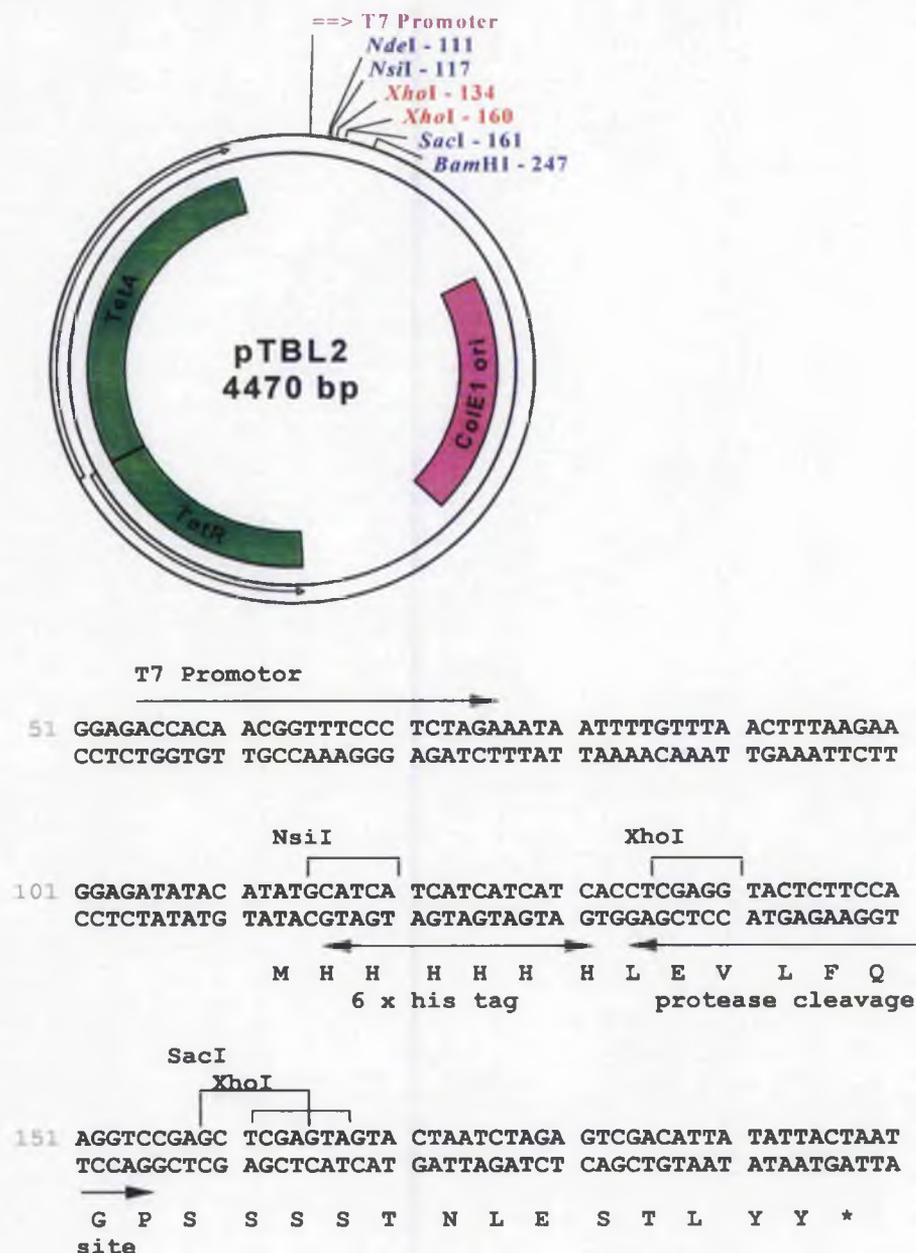


Figure 3-5 Vector diagram and sequence of pTBL2 (the diagram was produced with the program pDRAW).

3.2.8 Cloning of *his* genes into pTBL1 and pTBL2

The *hisX* genes were cloned into pTBL1 and pTBL2 as follows. The His-tagged genes were excised from pUni:*hisX* with the restriction enzymes *NsiI* and *SacI*, whereas those with the 3C cleavage site were released with *XhoI* and *SacI*. The vectors pTBL1 and pTBL2 were prepared by restriction with the appropriate restriction enzymes. The plasmid pTBL2 has two *XhoI* sites present and to eliminate

any problems which may arise from the second *Xho*I site (which overlaps with the *Sac*I site see Figure 3-5) this vector was digested with *Sac*I followed by *Xho*I. Gel-purified fragments were ligated as appropriate and transformed into DH5 α cells. The identity of recombinant clones was confirmed by restriction analysis and positive clones were then transformed in competent BL21(DE3)pLysS cells to test for over-expression.

3.3 Generation of over-expression constructs from *M. thermoautotrophicum*

3.3.1 Introduction

Interestingly the genome of *Mth* contains two genes for HisG. To investigate the biochemical and structural differences between the HisG and HisG-related protein from *Mth*, both genes were cloned into the over-expression plasmid pTB361.

3.3.2 pTB361 plasmid

The expression plasmid used for these experiments is pTB361. No affinity tag for purification purposes was used as the genes are from a thermophilic organism and a heat step should be very effective for purification. The T7 promoter is located immediately upstream of the multiple cloning site and the restriction site *Kpn*I is used to clone the 5' end. The *Kpn*I site introduces an extra Val and Pro residue between the ATG start codon and the HisG protein. This cloning strategy was favoured over using the *Nde*I site which would not add any extra residues at the N-terminus but can be difficult to digest.

3.3.3 Amplification of *hisG* and *HisG-related protein* genes by PCR

Cloning into the expression vector pTB361 required the generation of restriction sites within the HisG and HisG-related protein DNA to facilitate cloning in the correct orientation and reading frame. The *Kpn*I site was used as the 5' site and *Xho*I was used as the 3' site. These two sites were generated by PCR using primers containing the appropriate restriction sites (Table 2-8).

The genes were amplified from genomic DNA by PCR using Vent DNA polymerase. The *Mth* genomic DNA was obtained courtesy of Dr. B. Hanzelka, Ohio State University, USA.

3.3.4 Cloning of PCR product into pGEM

pGEM is a simple cloning vector with a multiple cloning site within the *lacZ α* gene which permits screening of recombinant clones using blue/white colour selection. Blue colonies indicate no insert in the multiple cloning site whereas white colonies indicate an insertion because it disrupts the *lacZ α* gene.

The multiple cloning site contains a restriction site for *EcoRV* which generates a blunt-ended linear vector. This can be used to clone blunt-ended DNA inserts, such as PCR products obtained with a proof-reading DNA polymerase.

Purified PCR product of the *Mth* genes was directly ligated into *EcoRV* digested pGEM. The ligation was transformed into competent JM109 cells which are suitable for the blue/white screening. The cells were plated on LB-agar plates supplemented with IPTG and X-Gal. White colonies were picked for analysis and purified plasmid analysed by restriction digestion.

3.3.5 Cloning into pTB361

pTB361 and the pGEM vectors with the *hisG(rel)* insert were digested with *KpnI* and *XhoI*. After purification from agarose gel the *hisG* inserts were ligated into the digested pTB361. The ligation was transformed into competent DH5 α cells. Purified plasmid was analysed by restriction digest for the correct insert and positive clones were transformed in BL21(DE3)pLysS for over-expression.

The level of over-expression of these constructs was poor. This was attributed to the presence of rare codons in the DNA sequence of the *Mth* genes. Therefore pTB361:*Mth-hisG (rel)* was transformed in *E. coli* strains designed to increase over-expression levels of genes with rare codons. The results are described in Section 4.2.

3.3.6 Generation of BL21 and C41 *E. coli* strains to increase over-expression levels of genes containing rare codons

E. coli strains BL21(DE3) which increase levels of over-expression for genes containing rare codons are commercially available (Stratagene). These strains contain a vector, RIL or RP, for expression of rare codon tRNAs. The RIL plasmid encodes

tRNA genes for Arg (AGG & AGA), Ile (ATA) and Leu (CTA), whereas the RP carries tRNA genes for Arg (AGG & AGA) and Pro (CCC). Unfortunately these strains could not be used here as pTB361 is selective for tetracycline resistance and the commercially available strains contain tetracycline resistance on their genome. Therefore new strains with no tetracycline resistance had to be created.

Firstly, the plasmids RIL and RP were isolated from the commercially available BL21(DE3) strains. This was done with extra care as BL21 strains are *endA*⁺ which means that they contain endonucleases that can degrade the plasmids. Restriction analysis confirmed the identity of the plasmid and gel analysis prove that the plasmid preparation was free from degradation products. Purified RIL and RP plasmids were then transformed in BL21(DE3) and C41(DE3) competent cells⁶. Transformations were plated on LB agar plates containing chloramphenicol to select for the RIL and RP plasmids. Strains were named BL21(DE3)RIL, BL21(DE3)RP, C41(DE3)RIL and C41(DE3)RP.

3.4 Generation of mutants of HisG from *H. influenzae*

3.4.1 Introduction

Site-specific mutagenesis of proteins is a widely used technique to investigate the role of both individual residues and protein domains in the function of a protein. Both approaches were used here to investigate structure-function relationships with the HisG protein.

The C-terminal domain of HisG is involved in the formation of the hexameric form of the enzyme but no function beyond this structural role has been elucidated⁷. Therefore a deletion mutant was created with the C-terminal domain deleted (residue 222-298 in *Hin*, corresponding to 223-299 in *E. coli*; named *Hin* HisGshort). Furthermore, a mutant with just the last 5 residues deleted (294-298 in *Hin*, corresponding to 295-299 in *E. coli*; named *Hin* HisG294) was also constructed as these residues appear to form the main interaction between the C-terminal domains (see Section 6.3 and 8.4). To investigate the binding and structural properties of the C-

⁶ Cells were made competent using the CaMg method and free of additional plasmids such as pLysS.

⁷ This study was initiated before the structure of HisG from *Mtb* confirmed the binding of histidine at the C-terminal domain.

terminal domain itself a deletion construct expressing the C-terminal domain alone was also made.

HisG from *Hin* was chosen to obtain the mutants as the vector constructs were readily available (see Section 3.2), the over-expression was very good and it is similar to the *E. coli* enzyme therefore facilitating the comparison of results. Furthermore the construct contains an N-terminal His₆-tag for rapid purification.

3.4.2 Generation of deletion mutants

Deletion mutants were created by PCR of the whole plasmid. The primers were designed so that the part of the gene to be deleted was omitted from the amplification. Figure 3-6 shows a schematic flow diagram of how the deletion mutants were created.

Firstly, the template DNA is amplified by PCR. To amplify this 5.3 kbp DNA the method of long PCR was employed [Cheng *et al.*, 1994], which requires the addition of 2% DMSO and 8% glycerol to the PCR reaction mix as well as using *PfuTurbo* DNA polymerase. A number of other PCR protocols were tried initially but none were successful.

To fragmentize the template plasmid the restriction enzyme *DpnI* was added to the PCR reaction. *DpnI* will only digest DNA which is methylated and therefore does not digest the PCR amplified DNA. The PCR product was purified from agarose gel and ligated with T4 ligase. The ligations were transformed into competent DH5 α cells and plated on LB-agar plates containing tetracycline. Plasmids were analysed by restriction digest with *NsiI* and *XhoI* and for the *Hin* HisGshort construct several positive clones were obtained, whereas only single clones were obtained for the *Hin* HisGcterm and *Hin* HisG294 constructs. Correct clones were identified by restriction analysis. The small size of the deletion (15 bp) produced by the *Hin* HisG294 construct required a high percentage agarose gel to be run to visualise the difference between the wild-type and mutant DNAs.

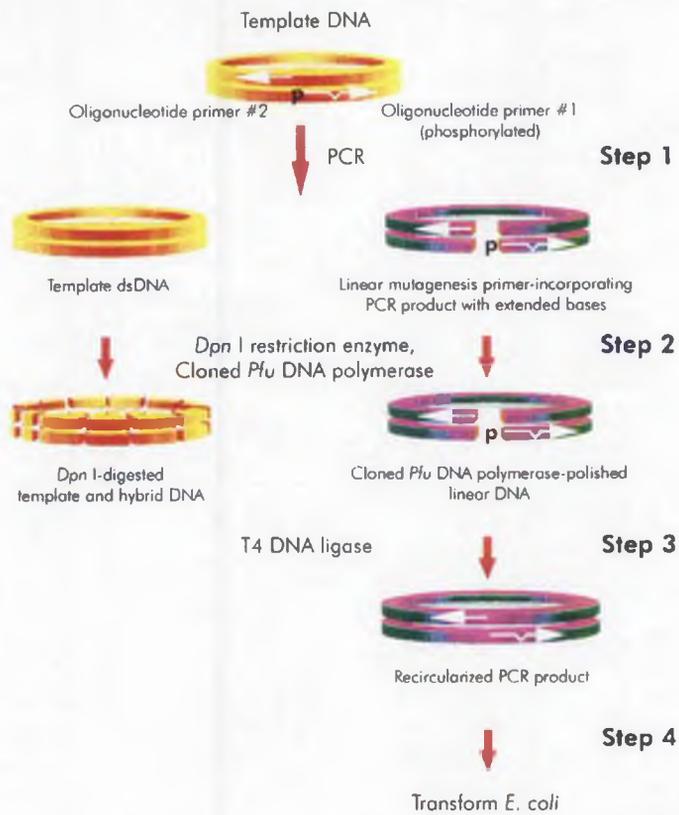
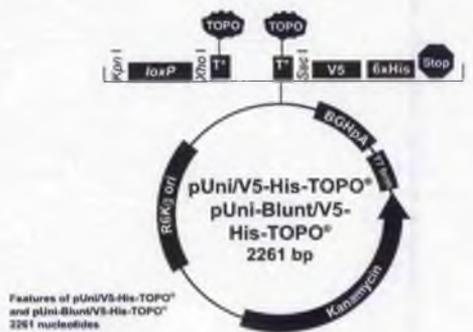


Figure 3-6 Scheme of protocol to produce deletion mutants (adapted from Stratagene)

I. Appendix: Schematic of vector pUni/V5-His-TOPO

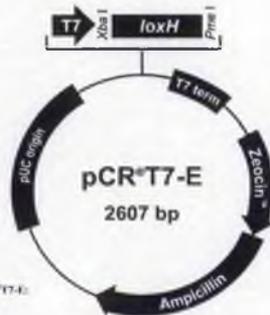


Features of pUni/V5-His-TOPO[®] and pUni-Blunt/V5-His-TOPO[®]
2261 nucleotides

R6K[®] origin: bases 6-387
 Uni1 forward priming site: bases 365-363
 loxP site: bases 418-451
 TOPO[®] Cloning site: bases 480-481
 V5 epitope: bases 500-541
 6xHis tag: bases 542-559
 Uni1 reverse priming site: bases 563-564
 BGH polyadenylation sequence: bases 582-790
 T7 transcription termination region: bases 805-833
 Kanamycin resistance gene: bases 1112-1906 (C)
 kan promoter: bases 1605-2044 (C)

*The modified T is either a 3' overhang (pUni/V5-His-TOPO[®]) or paired with an A in pUni-Blunt/V5-His-TOPO[®]

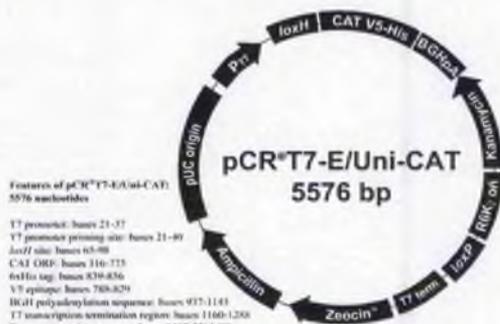
II. Appendix: Schematic of vector pCRT7-E



Features of pCRT7-E:
2607 nucleotides

T7 promoter: bases 21-37
 T7 promoter priming site: bases 21-40
 loxH site: bases 63-98
 T7 transcription termination region: bases 145-192
 Zeocin[®] resistance gene: bases 275-735
 Ampicillin resistance gene: bases 739-1616
 pUC origin: bases 1761-2434

III. Appendix: Schematic of control vector pCRT7-E/Uni-CAT



Features of pCRT7-E/Uni-CAT:
5576 nucleotides

T7 promoter: bases 21-37
 T7 promoter priming site: bases 21-40
 loxH site: bases 63-98
 CAT ORF: bases 116-775
 6xHis tag: bases 839-856
 V5 epitope: bases 789-829
 BGH polyadenylation sequence: bases 937-1143
 T7 transcription termination region: bases 1160-1208
 Kanamycin resistance gene: bases 1465-2261 (C)
 kan promoter: bases 2362-2399 (C)
 R6K[®] origin: bases 2627-3013
 Uni1 forward priming site: bases 2981-2999
 Uni1 reverse priming site: bases 3034-3052
 T7 transcription termination region: bases 3114-3161
 Zeocin[®] resistance gene: bases 3241-3704
 Ampicillin resistance gene: bases 3708-4585
 pUC origin: bases 4716-5403
 (C) indicates complementary strand

Chapter 4: Over-expression, Purification, Crystallisation and Characterisation of Enzymes

4.1 Over-expression and purification of his-tagged enzymes from the histidine pathway

4.1.1 Over-expression

The level of over-expression was tested for each of the clones by the quick protocol described in Section 2.7.1. Over-expression was observed in all cases as proved by SDS-PAGE. A typical gel demonstrating the over-expression of *Mtb* HisI and HisE is shown in Figure 4-1. It was difficult to obtain a clear band to show over-expression of *Mtb* HisE on SDS-PAGE with Coomassie blue staining; this is due to the small size of the protein.

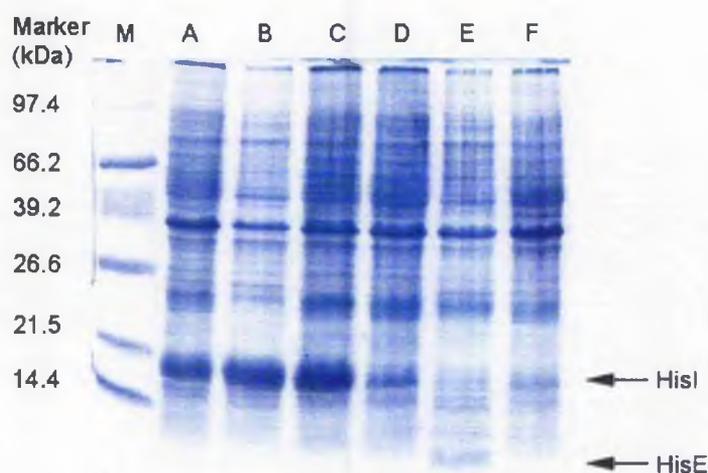


Figure 4-1 Over-expression of *Mtb* HisI and HisE.

Lane	Clone	Condition	Lanes	Clone	Condition
A	<i>Mtb</i> HisI 1	Over-night culture uninduced	D	<i>Mtb</i> HisE 1	Over-night culture uninduced
B	<i>Mtb</i> HisI 1	Over-night culture induced with IPTG	E	<i>Mtb</i> HisE 1	Over-night culture induced with IPTG
C	<i>Mtb</i> HisI 2	Over-night culture uninduced	F	<i>Mtb</i> HisE 2	Over-night culture uninduced

Comparison of soluble and insoluble fractions showed that some of the proteins are over-expressed in inclusion bodies rather than as soluble protein. Table 4-1 shows histidine pathway enzymes which expressed in a soluble form and as inclusion bodies. Enzymes with no significant amount of over-expression of soluble protein were not used any further in this work. All soluble enzymes were purified and crystallisation trials undertaken.

Table 4-1 Soluble and insoluble his enzymes obtained by over-expression in pTBL1 and pTBL2

Organism	Enzyme	Soluble	Inclusion bodies	Comment
<i>B. subtilis</i>	HisG	✓	✗	
	HisZ	✓	✗	
	HisIE	✓	✗	
<i>H. influenzae</i>	HisG	✓	✗	
	HisIE	✗	✓	
<i>M. tuberculosis</i>	HisG	✗	✓	
	HisI	✗	✓	
	HisE	✓	✓	Some soluble, some inclusion bodies

4.1.2 Purification

All soluble his-tagged histidine pathway enzymes were purified in a one-step purification procedure using a Ni-NTA column. This resulted in almost pure protein as shown by SDS-PAGE. An example of the purification of *BS* HisZ is shown in Figure 4-2. Some minor contaminants are visible with a molecular weight of about 70 kDa which result from nonspecific binding to the metal resin.

HisI and HisE from *Mtb* contained a 3C protease cleavage site for proteolytic removal of the his-tag. However the tag was not removed and enzyme including the tag was used for crystallisation and activity measurements.

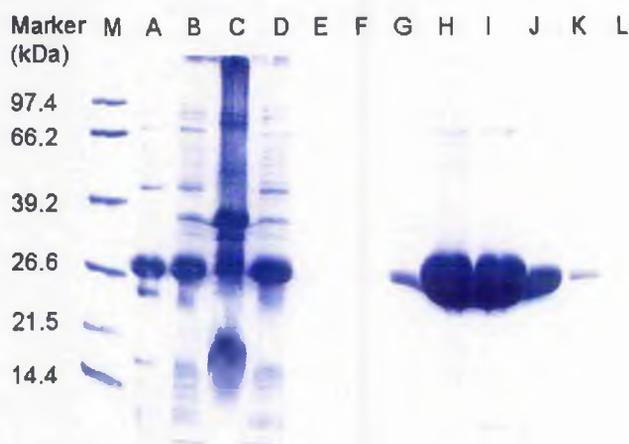


Figure 4-2 SDS-PAGE of purification of BS HisZ.

Lane	Lanes
A Whole cell extract	D Soluble fraction
B Extract after French press	E-L Fraction from Ni-NTA elution
C Insoluble fraction	

4.2 Over-expression and purification of HisG and HisG-related enzyme from *M. thermoautotrophicum*

4.2.1 Over-expression

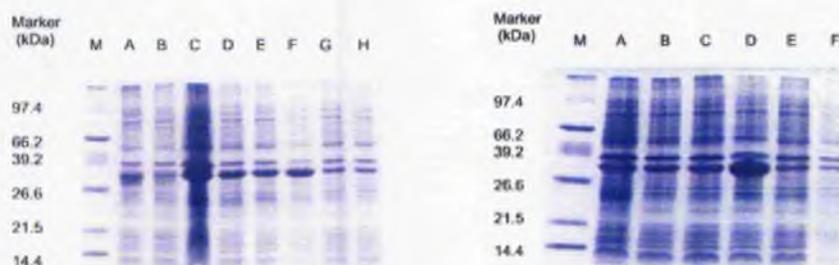
The genes encoding the HisG and the HisG-related protein of *Mth* contain several rare codons. Table 4-2 shows the percentage occurrence of rare codons in these genes and compares this with the codon usage in *E. coli*.

The table shows clearly that there will be a problem with the codons for Arg and especially Ile. Therefore the strains containing the RIL plasmid should increase the over-expression substantially.

Plasmids pTB361:*Mth-HisG/HisGrel* were transformed in the BL21 and C41 RIL and RP strains to test for expression levels. Figure 4-3 and Figure 4-4 show SDS-PAGE analysis of over-expression of HisG and HisG-related protein in the different strains.

Table 4-2 Rare codon usage in HisG and HisG-related protein of *Mth*

aa	Codon triplet	<i>Mth</i> HisG		<i>Mth</i> HisG-related protein		<i>E. coli</i>
		rare codons/ all codons for aa	fraction of rare to all codons (%)	rare codons/ all codons for aa	fraction of rare to all codons (%)	fraction of all codons (%)
Arg	AGG	14/22	4.86	14/27	3.91	0.14
	AGA	4/22	1.38	3/27	0.83	0.21
Ile	AUA	21/30	7.29	23/27	6.42	0.31
Leu	CUA	0/27	0	1/28	0.27	0.41
Pro	CCC	7/13	2.43	5/14	1.39	0.43

**Figure 4-3 Over-expression of *Mth* HisG in different *E. coli* strains. SDS-PAGE of over-expression in BL21(DE3) strains (a) and C41(DE3) strains (b).**

a) BL21(DE3)			b) C41(DE3)		
Lane	plasmid		Lanes	plasmid	
A	pLysS	3h after induction with IPTG	A	RIL	3h after start of over-expression without induction with IPTG
B	pLysS	before induction with IPTG	B	RIL	before induction with IPTG
C	RIL	3h after start of over-expression without induction with IPTG	C		as B
D	RIL	before induction with IPTG	D	RIL	3h after induction with IPTG
E		as D	E	RP	before induction with IPTG
F	RIL	3h after induction with IPTG	F	RP	3h after induction with IPTG
G	RP	before induction with IPTG			
H	RP	3h after induction with IPTG			

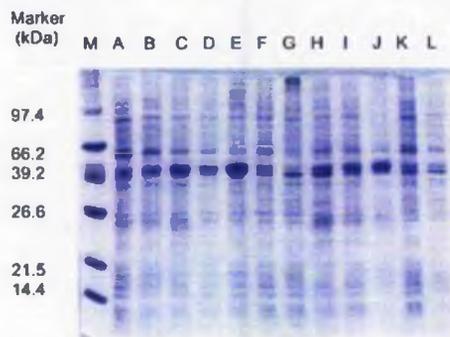


Figure 4-4 Over-expression of *Mth* HisG-related protein in different *E. coli* strains. SDS-PAGE of over-expression in BL21(DE3) strains and C41(DE3) strains.

BL21(DE3)			C41(DE3)		
Lane	plasmid		Lanes	plasmid	
A	pLysS	before induction with IPTG	H	RIL	4h after start of over-expression without induction with IPTG
B	pLysS	4h after induction with IPTG	I	RIL	before induction with IPTG
C	RIL	4h after start of over-expression without induction with IPTG	J	RIL	4h after induction with IPTG
D	RIL	before induction with IPTG	K	RP	before induction with IPTG
E	RIL	4h after induction with IPTG	L	RP	4h after induction with IPTG
F	RP	before induction with IPTG			
G	RP	4h after induction with IPTG			

For *Mth* HisG the highest levels of expression were obtained from both RIL strains with C41 giving the best results (see lane D in Figure 4-3 b). The HisG-related protein from *Mth* was expressed in the highest level in BL21(DE3)RIL (see lane E Figure 4-4). These results confirm the prediction that the RIL strains would give the best enhancement of expression. This is due to the frequent occurrence of the rare codon for Ile (ATA) in both genes. The strains with the best results were used for subsequent large scale over-expression.

4.2.2 Purification of HisG

Because *Mth* is a thermophilic bacterium with an optimal growth temperature of 65°C; purification of the over-expressed HisG was attempted using a heat step. This should denature most of the host proteins but leave the thermostable HisG in its native form. However using an incubation at 75°C for 15 minutes, which was successfully used for glutamyl-tRNA reductase from *Mth* [Hungerer *et al.*, 1996],

resulted in the denaturation of the majority of the HisG protein. Therefore the temperature in the heat step was reduced to 65°C. This removed the majority of the host protein but the purity of HisG was not adequate for crystallisation and further purification was required. A Q-Sepharose anion exchange column was an effective second step followed by a dye column (reactive green 19) which had previously been used successfully in the purification of *E. coli* HisG [Lohkamp *et al.*, 2000]. The resulting pure protein was used for crystallisation, DLS experiments and enzyme assays.

4.2.3 Resolubilisation, refolding and purification of HisG-related protein

The HisG-related protein from *Mth* was mainly over-expressed in inclusion bodies. It was shown that the inclusion bodies could be resolubilised in buffer containing 3-5 M urea. Refolding of the enzyme was accomplished by dialysis against Tris buffer at pH 7.5 containing no urea. The correct folding of the HisG-related protein was proved by activity measurements in the HisG assay.

Preliminary experiments showed that the inclusion bodies can be partially purified by washing with buffer containing 30%(w/v) sucrose and 1% Triton-X-100. However further purification steps of the resolubilised protein are required to provide pure protein. A heat step, as employed in the purification of HisG, was not useful as the majority of the HisG-related protein was denatured by heat treatment.

4.3 Purification of PA5104

PA5104 was over-expressed in *E. coli* and cells for purification were provided by N. Patterson. The first purification step was performed as described for the his-tagged histidine pathway enzymes. However the his-tag had a 3C protease cleavage site which was removed by proteolysis with his-tagged 3C protease. After cleavage, the protein was subject to a second purification on a Ni-NTA column. This bound the protease, his-tag and some nonspecific binding proteins and PA5104 was eluted in the flow through. A gel showing the purification is shown in Figure 4-5. This proved to be the first successful application of the over-expression and purification system using the vector pTBL2 created during this work.

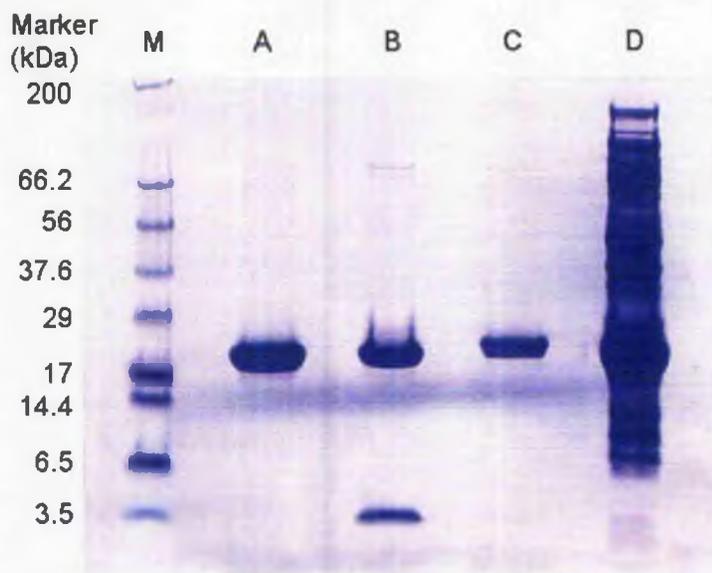


Figure 4-5 Purification of PA5104.

Lane	Lanes
A	Flow through of the second Ni-NTA column (pure PA5104)
B	Elution of second Ni-NTA column (3C protease and His ₆ tag)
C	Elution of first Ni-NTA column (= load of second column)
D	Whole cell extract

4.4 Dynamic light scattering (DLS)

DLS experiments were carried out to determine which combination of product and cofactor were appropriate to obtain a monodisperse enzyme solution. Monodispersity is not an infallible indicator that diffraction quality crystals can be obtained from a protein solution; instead it reflects the observation that monodisperse protein solutions have a much higher probability of producing crystals than those which are not [Ferré-D'Amaré and Burley, 1997]. Furthermore DLS gives an indication of the molecular weight of the protein and hence provides information about the state of aggregation.

For *E. coli* HisG several conditions were tried (Table 1) and it was shown that monodispersity could not be achieved in the presence of any of the additives tested. In the presence of AMP the lowest value for polydispersity was obtained and the molecular weight was estimated to be 186 kDa. This is about 10% lower than the total M_w of six monomers and is consistent with the expected hexameric state of the

enzyme. Other additives such as histidine and PR-ATP gave similar molecular weights and hydrodynamic radii but higher polydispersity.

Table 4-3 DLS of *E. coli* HisG

Additives	hydrodynamic radius (nm)	polydispersity (nm)	estimated M_w (kDa)	conclusion
none (no KCl)	7.3	0.3	369	polydisperse
None	6.9	1.0	321	polydisperse
AMP (0.4 mM)	5.5	0.7	186	polydisperse
Histidine (5 mM)	5.7	0.9	197	polydisperse
Histidine (5 mM)	5.3	1.0	170	polydisperse
+ AMP (1 mM)				
PRPP (1 mM) +	5.5	1.5	184	polydisperse
ATP (5 mM) +				
MgCl ₂ (10 mM)				
PRPP (1 mM) +	5.6	1.6	190	polydisperse
ATP (5 mM) +				
MgCl ₂ (10 mM) +				
Histidine (5 mM)				

None of the several conditions probed for the *Mth* HisG resulted in monodispersity (Table 4-4). The lowest polydispersity value was obtained in presence of histidine at pH 7.5. The calculated molecular weight of around 200 kDa in presence of inhibitors at low pH is in accordance with the calculated weight of 188 kDa for a hexamer. In the absence of the inhibitors histidine or AMP the polydispersity increased. In addition a higher value for the polydispersity was observed in the absence of MgCl₂ and at higher pH values.

Table 4-4 DLS of *Mth* HisG

Additives	hydrodynamic radius (nm)	polydispersity (nm)	estimated M_w (kDa)	conclusion
None	13.5	6.4	1621	polydisperse
MgCl ₂ (10 mM)	5.8	2.2	208	polydisperse
Histidine (5 mM) pH 7.5	5.7	1.1	205	polydisperse
Histidine (5 mM) pH 8.5	6.8	2.9	307	polydisperse
AMP (2 mM) pH 7.5	5.7	1.6	203	polydisperse
AMP (2 mM) pH 8.5	7.0	3.4	333	polydisperse

Table 4-5 DLS of *Hin* and *BS* HisG

Enzyme	hydrodynamic radius (nm)	polydispersity (nm)	estimated M_w (kDa)	Conclusion/ comment
<i>Hin</i> HisG WT	5.8	0.6	207	polydisperse/ filter blocked
<i>Hin</i> HisG short	3.3	0.3	54	polydisperse
<i>Hin</i> HisG 294	5.3	1.4	165	polydisperse/ filter blocked
<i>BS</i> HisG	3.2	1.6	50	polydisperse

4.5 Crystallisation

4.5.1 Crystallisation of *E. coli* HisG

Initially crystallisation experiments of HisG in the absence and presence of 2mM AMP were performed at 20°C using the sitting-drop vapour-diffusion technique. An exhaustive set of conditions, comprising commercially available and local sparse-matrix screens [Jancarik and Kim, 1991] were tried. Initial crystals were

obtained from various conditions mostly in the presence of AMP. Optimisation of these crystallisation conditions were most successful for a single condition which led to the final conditions of 1.3 M sodium tartrate, 50-200 mM MgCl₂, 100 mM citrate buffer, pH 5.6 and protein in the presence of 2 mM AMP (crystal form I). Typically, 1-2 μ l of protein (10 mg/ml) were mixed with an equal volume of the reservoir solution. Crystals appear after 3-4 days and continued to grow as trigonal prisms up to a maximum size of 0.3 x 0.2 x 0.2 mm (Figure 4-3). A second crystal form was obtained using 1.36-1.44 M ammonium sulfate, 0-0.3 M sodium chloride, 100 mM HEPES buffer, pH 7.5 and protein in the presence of 2 mM AMP. More rounded crystals appeared after 4-5 days and grew to a maximum size of about 0.2 x 0.2 x 0.2 mm (crystal form II, see Figure 4-6).

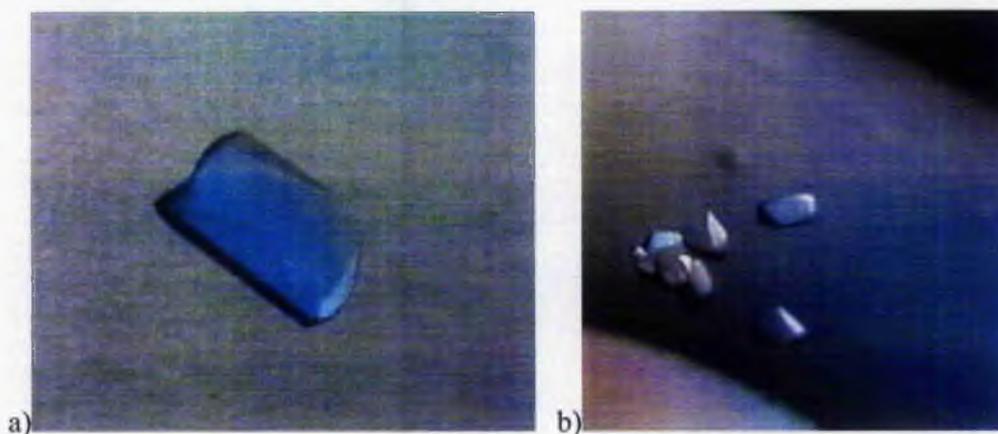


Figure 4-6 Crystals of *E. coli* HisG in presence of 2 mM AMP. a) crystal form I, b) crystal form II.

E. coli HisG was co-crystallised in various combinations of substrate and inhibitors (AMP, PR-ATP, histidine) using the conditions which yield crystal form I. Crystals were obtained in all cases where AMP or PR-ATP was present. The morphology of some of these crystals, especially in the presence of 2 mM AMP and 5 mM histidine, was clearer than the ones obtained in the presence of AMP alone. Crystals grown in the presence of 2 mM PR-ATP (no AMP) showed similar morphology to the ones grown in the presence of AMP.

At a later stage of the project, sparse-matrix screening of crystallisation conditions was tried with *E. coli* HisG (10 mg/ml) in the presence of 10 mM AMP and 10 mM histidine. Several different crystal forms of HisG complexed with AMP and histidine were found. These were refined and one condition yielded crystals of

limited diffraction quality (crystal form III) in 18% (w/v) PEG 3350, 0.2 M NaNO₃, and 100 mM HEPES buffer pH 7.0 (not refined conditions).

4.5.2 Crystallisation of *BS* histidine pathway enzymes

Sparse-matrix screening was performed on HisG, HisIE, HisZ and a complex of HisG-HisZ. The probed screens resulted in no success for HisIE, HisZ and the HisG-HisZ complex. However various crystal forms were obtained for HisG at a protein concentration of 10 mg/ml in the presence of 10 mM ATP. One crystal form was obtained from 2.5-4.5 M NaCl, 100 mM MgCl₂ and HEPES buffer at pH 8.2 (Figure 4-7). As these crystals did not diffract to better than 3 Å, crystallisations were performed using additive screens. However crystals obtained from these showed no difference in size and diffraction properties from the ones initially observed.

A second crystal form initially showed clusters of crystals which were grown in 2.0-2.4 M Na/K phosphate and Na acetate buffer at pH 4.6-5 (Figure 4-7). However using a seed-bead kit for seeding, single crystals were obtained. These were rod-like and usually very thin.

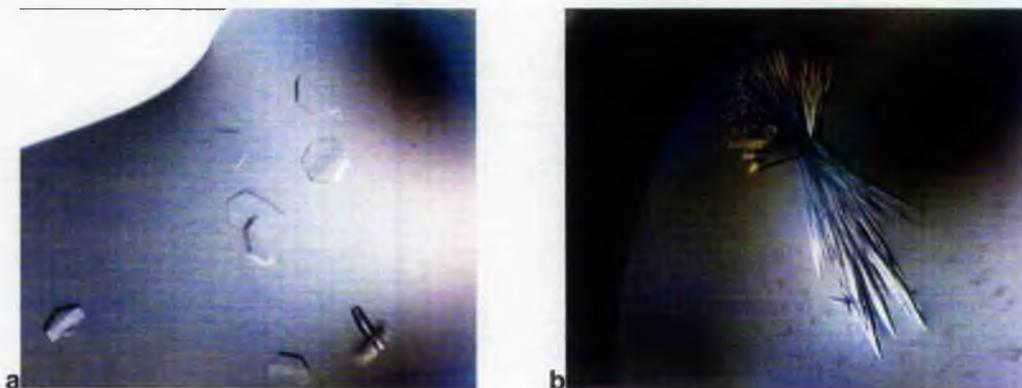


Figure 4-7 Crystals of *BS* HisG. (a) crystals of crystal form I and (b) a cluster of crystal form II.

Crystallisation of a HisG-HisZ complex was attempted with a molar ratio of 6 HisG to 2 HisZ, based on preliminary data about the quaternary structure of *L. lactis* HisG-HisZ [Champagne *et al.*, 2001]. A more recent investigation of the complex showed however that the true ratio is 4:4 [Bovee *et al.*, 2002] which could explain why no crystals could be obtained for the HisG-HisZ complex.

4.5.3 Crystallisation of *Mth* HisG

DLS showed that *Mth* HisG is monodisperse in the presence of histidine (Table 4-4). Therefore sparse-matrix screening was carried out in the presence of 5 mM histidine at a protein concentration of 10 mg/ml.

Crystals of HisG from *Mth* were grown in 32-38% (v/v) MPD, 200-350 mM MgCl₂, 100 mM imidazole, pH 8.0 and protein in the presence of 5 mM histidine (Figure 4-8). The crystals show a trigonal prism morphology and are usually fairly thin.

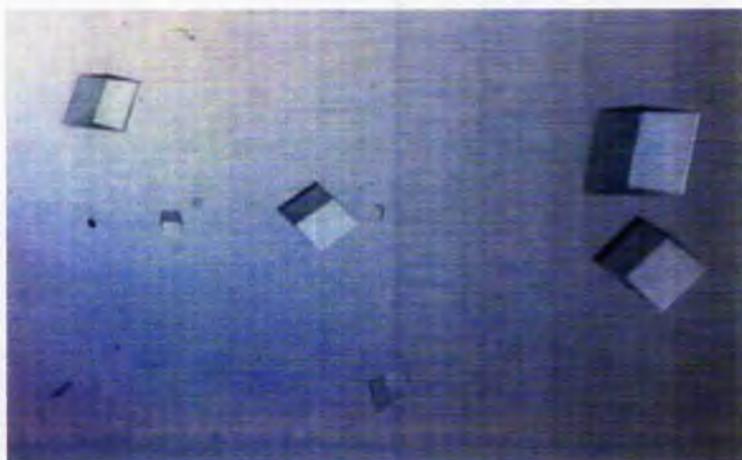


Figure 4-8 Crystals of *Mth* HisG.

4.5.4 Crystallisation of PA5104

Sparse-matrix screening for PA5104 was carried out by N. Paterson. Various conditions produced crystals of which two were refined and produced crystals suitable for X-ray diffraction. Crystals of PA5104 were grown in 24-34% (v/v) PEG 400, 200 mM MgCl₂, and 100 mM Tris-HCl buffer at pH 8.5 or 9.0 (crystal form I). The crystals appear to be rhombic, and could be nearly cubic, elongated or flat in appearance (Figure 4-9).

A second crystal form of PA5104 was obtained in 1.8-2.2 M ammonium sulfate, 200 mM NaCl, and 100 mM cacodylate buffer at pH 6.5. In many cases the crystals appeared not to be single.

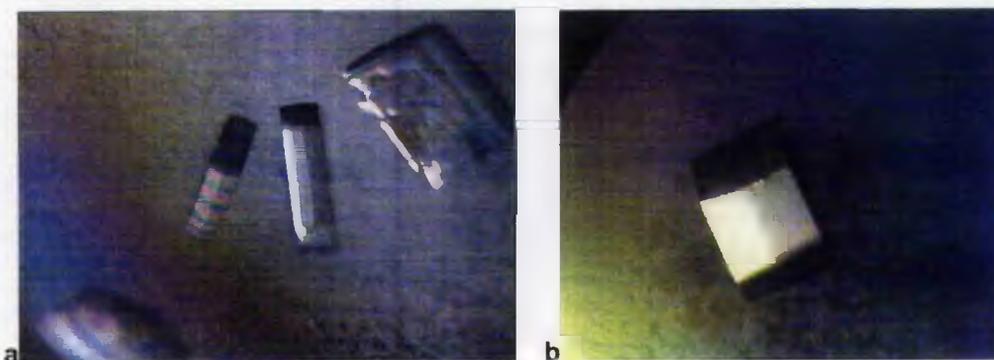


Figure 4-9 Crystals of PA5104 (crystal form I). (a) flat crystals of PA5104 and (b) more 'cubic' crystals.



Figure 4-10 Crystals of PA5104 (crystal form II).

4.6 Kinetic properties of HisG

4.6.1 *Hin* HisG and truncation mutants

Enzyme assays were carried out using the method described in the literature without the use of pyrophosphatase [Blasi *et al.*, 1971]. To obtain kinetic parameters for both substrates, ATP and PRPP, the concentration of one of the substrates in the assay was kept constant while the other was varied. The resulting plots of rate versus corresponding substrate concentration were analysed using steady-state kinetics. The data showed some degree of co-operativity which indicate a more complex kinetic behaviour. This type of behaviour had been observed previously, especially in the presence of histidine [Tebar and Ballesteros, 1976]. However the Hill coefficient was within one standard deviation of 1 and therefore a more complex analysis was omitted here. The results for the steady-state analysis are shown in Table 4-6.

Table 4-6 Kinetic parameters of *Hin* HisG and truncation mutants

Enzyme	Substrate	k_{cat} (s^{-1})	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ ($\text{M}^{-1} \text{s}^{-1}$) 10^6
<i>Hin</i> HisG WT	ATP	1188±34	185±21	6.4±0.7
	PRPP	1325±9	44±1	30.1±0.7
WT + histidine	ATP	282±9	292±28	1.0±0.1
Mutant 294	ATP	521±26	334±46	1.6±0.2
	PRPP	624±51	115±17	5.4±0.9
Short <i>Hin</i> HisG	ATP	709±60	885±195	0.8±0.2
	PRPP	769±51	147±17	5.2±0.7

For the *Hin* wild type (WT) enzyme the observed values of K_{M} for ATP (185 μM) and PRPP (44 μM) are very similar to previously observed ones. The values for the *E. coli* enzyme were determined to be 70 μM and 30 μM respectively [Tebar and Ballesteros, 1976]. For the *S. typhimurium* HisG the data vary for ATP in the range 89-340 μM and PRPP 28-52 μM [Sterboul *et al.*, 1977; Wainscott and Ferretti, 1978]. Both *Hin* mutant enzymes show activity for the HisG reaction with higher K_{M} and lower k_{cat} values for both substrates compared to the WT enzyme. Based on the K_{M} values the binding affinity of both substrates decreases from the WT enzyme to *Hin* HisG294 and the short *Hin* HisG. However the catalytic efficiency, as reflected by $k_{\text{cat}}/K_{\text{M}}$, is within the errors identical for both mutants and substrates and by a factor of about 6 lower than that for the WT HisG.

The results in Table 4-6 show that histidine is an inhibitor for *Hin* WT HisG as has been shown for other long forms of HisG [Martin, 1963]. No difference in activity could be detected for either *Hin* HisG mutant in presence of histidine (data not shown). This shows that histidine is not an inhibitor to the HisG mutants studied here.

The maximum rate for all *Hin* HisG forms studied here decreased at ATP concentrations above 5 mM (data not shown). Thus the enzymes are inhibited by higher concentrations of substrate ATP. It has been argued that this could be an effect of chelation of the required Mg ions by excess ATP [Dall-Larsen, 1988].

4.6.2 Thermophilic HisG from *Mth*

HisG from *Mth* showed almost no activity at room temperature. However, activity was observed at increased temperatures as expected for an enzyme from a

thermophilic organism. The highest activity was observed at 65°C, the optimal growth temperature of the organism. Even at higher temperatures (up to 90°C) the enzyme showed activity which only slowly decreased with increasing temperature. The activity fell with decreasing temperatures so that at room temperature the enzyme is almost inactive.

Activity measurements were carried out with the HisG-related protein obtained from resolubilisation of inclusion bodies. Enzyme activity for the condensation of ATP and PRPP to form PR-ATP was observed based on the increase in absorbance at 290 nm. Again the highest activity was observed at 65°C with a decrease in activity at higher and lower temperatures. It should be mentioned that for both enzymes control reactions without the enzyme were carried out; these showed no increase in absorbance at 290 nm.

4.6.3 Short HisG from *BS*

Assay measurements were performed on the native short form of HisG from *BS*. Under standard assay conditions at 25°C no enzyme activity could be detected. However the formation of PR-ATP at 290 nm could be observed at increased temperatures demonstrating enzyme activity. Figure 2 shows the measured increase in absorbance at 290nm following the incubation of the enzyme at various temperatures. Enzyme activity can clearly be seen at temperatures above 30°C. Further evidence that true enzyme activity is observed is provided by the proportional increase in initial rate with increasing enzyme concentration. Additionally, control reactions with no enzyme or use of HisZ, as a control enzyme (since it does not catalyse the HisG reaction), showed no increased absorbance at 290 nm (data not shown).

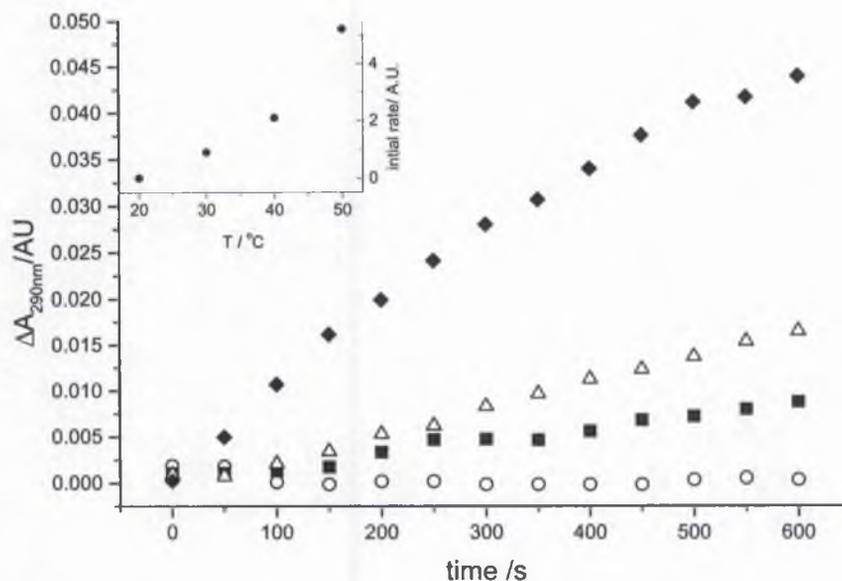


Figure 4-11 Activity of BS HisG at different temperatures. The change in absorbance measured at 290 nm is shown for 4 different temperatures (○ 20°C, ■ 30°C, △ 40°C, ◆ 50°C). The insert shows the change of the initial rate with temperature (●).

4.7 Kinetics Discussion

Two different forms of HisG are found in bacteria which are distinguished by an extra 80 residue C-terminal domain. Sequence alignment and phylogenetic analysis of the equivalent N-terminal regions, omitting this C-terminal domain, shows no significant difference between the truncated inactive form and the active canonical form. Furthermore crystal structures of long forms of HisG have revealed that the active site is located in the core, between the first and second domain. The structural function of the C-terminal extension is to form a hexamer, with mainly the last 5 residues being involved in forming an inter-subunit β -sheet (see Section 6.3). Furthermore the C-terminal domain provides a binding site for the allosteric inhibitor histidine [Cho *et al.*, 2003]. It seems that the C-terminal extension is solely involved in regulation of the enzyme. Therefore it is not clear why the short form of HisG is inactive and requires HisZ to form an active complex. To answer this question the two truncation mutants of a native long form of HisG from *Hin* were created. The first one lacks just the last 5 residues to investigate the disruption of the main interactions in

forming hexamers. In the second mutant the full length C-terminal domain was truncated to match the length of the native short forms of HisG, so that the core of the two HisG forms can be compared.

Enzyme activity was observed for both of the truncation mutants of the long HisG. However the k_{cat} for both mutants was half that of the WT enzyme and the specific activity was reduced by a factor of 6. Activity of the 294 mutant, the truncation where the last 5 residues are omitted, could have been expected as only the formation of the hexamer, which is shown to be an inactive state [Klungsoyr and Kryvi, 1971], is disrupted. DLS showed that the quaternary structure of the 294 mutant is similar to that of the WT enzyme (Table 4-5). The lack of inhibition by histidine for this mutant is again caused by the lack of stabilisation of the hexameric form. However the short *Hin* HisG mutant would not necessarily be expected to be active as it corresponds to the native short HisGs which are inactive without HisZ. As shown by DLS (Table 4-5), the quaternary structure of the short mutant is a dimer which corresponds to the oligomeric state of the native short HisG and the proposed active form of the long HisG [Klungsoyr and Kryvi, 1971]. Based on the activity shown by the short mutant *Hin* HisG we can still not understand why the native short HisG is inactive. As there is no significant difference in primary, secondary, and quaternary structure between the two forms the answer might be found in the tertiary structure. Because of the other similarities, the difference is most likely to resemble that found between the tense and relaxed conformations of allosteric enzymes. The native short enzyme therefore would be in a tense conformation and the long enzyme in a relaxed conformation. Changing the tense conformation of the native short HisG into the relaxed form and hence restoring its catalytic activity could prove this. This idea was supported by the observation that with increasing temperature the enzyme activity of the native short HisG from *BS* could be restored. This can be explained by a thermal transition from the tense inactive conformation to the relaxed conformation. Similar activations of catalytic activities at increased temperature are found in enzymes of thermophilic organisms where the highest activity is usually found around the optimum growth temperature of the organism [Rees and Adams, 1995]. Here this was observed for the HisG and HisG-related protein from *Mth*. Based on these results it appears likely that HisZ is a “macromolecular catalyst” for the short HisG by forming a HisG-HisZ complex. In this complex the short HisG is in a relaxed form and can perform its catalytic reaction whereas in isolation it is in a tense, inactive

form. Furthermore, because it is not essential for enzyme activity, the C-terminal domain of the long HisG seems only to perform regulatory functions. In organisms with short HisG and HisZ these regulatory functions must be performed either by HisZ, the HisG-HisZ complex or by another enzyme which has yet to be identified.

Chapter 5: X-ray Data Collection, Processing and Scaling

5.1 Data collection and processing

Diffraction data from various crystals have been collected using the rotation method in-house using a rotating anode X-ray generator or at a synchrotron radiation source. Although there are various programs for data processing (d*trek [Pflugrath, 1999], Mosflm [Leslie, 1992], DENZO [Otwinowski and Minor, 1997]) available DENZO has been mainly used in the course of this study and scaling of the data was done with program SCALEPACK. In some cases Mosflm and d*trek was used in an attempt to improve data statistics. Integration with Mosflm and subsequent scaling with SCALA can take absorption corrections into account. d*trek uses 3D-profile fitting to improve statistics on partially recorded reflections and the scaling procedure has various options of absorption correction. Furthermore both programs allow the use of more than one diffraction image in their auto-indexing routines which can improve the space group determination. When anomalous data were processed in order to be used in programs which allow local scaling, data were not merged with SCALEPACK.

Derivatised crystals were usually back-soaked with cryo-protectant or well solution prior to loop-mounting and cryo-cooling unless otherwise stated.

5.1.1 HisG from *E. coli*

5.1.1.1 Native (crystal form I)

Crystals of ATP-PRT (crystal form I) in presence of 2 mM AMP were used for data collection. The X-ray diffraction data were collected on beam line 9.6 at the SRS Daresbury Laboratory using a wavelength of 0.87 Å using a ADSC Quantum-4 CCD detector. The crystals were radiation-sensitive, therefore cryocooling was essential. Crystals were loop-mounted in a cryoprotectant containing 15% (v/v) glycerol and cryo-cooled to 100 K using an Oxford Cryosystems cryostream. A native dataset to 2.7 Å resolution was collected using 1° oscillation frames (Figure 5-1). A total of 109° were collected in two wedges to maximise the completeness of the observed reflections. Data were processed with DENZO and scaled with

SCALEPACK [Otwinowski and Minor, 1997]. From autoindexing using DENZO, the crystals were found to index in a primitive rhombohedral lattice, corresponding to space group $R3$, with unit-cell dimensions of $a=b=133.6 \text{ \AA}$ and $c=114.1 \text{ \AA}$. Indexing with the higher symmetry space group $R32$ resulted in comparable merging R-factor statistics (Table 5-1). An assumption of one molecule per asymmetric unit for $R32$ lead to an acceptable packing density, V_m of $2.95 \text{ \AA}^3 \text{ Da}^{-1}$ which corresponds to a solvent content of 58% [Matthews, 1968].

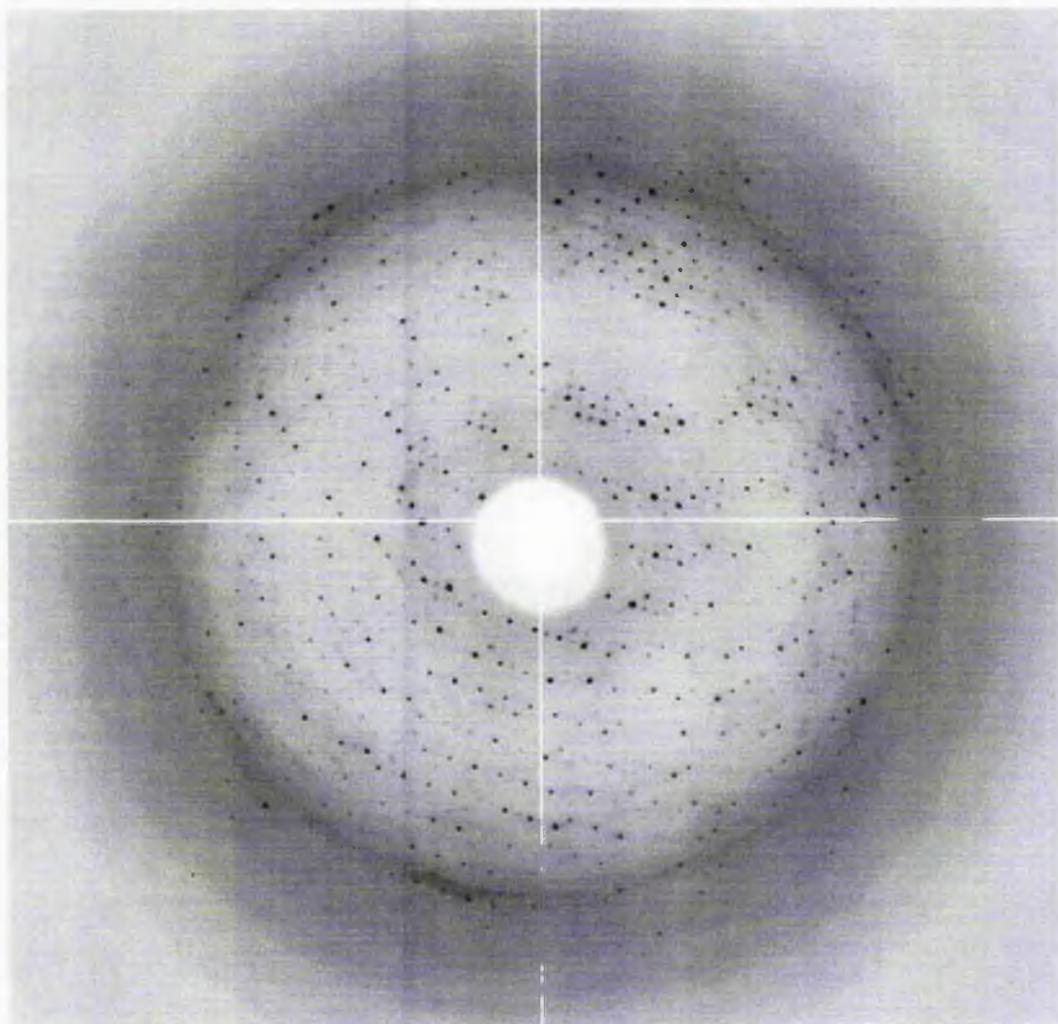


Figure 5-1 A typical diffraction pattern of ATP-PRT from *E. coli*. The data were collected on the beamline 9.6 at Daresbury synchrotron.

Table 5-1 Statistics for the X-ray data processed in R32

d_{\min} (Å)	R_{sym} (I)	(I/σ)	No. of unique reflections	Completeness (%)	Multiplicity (%)
5.82	0.041	38.2	1099	95.6	6.3
4.62	0.061	39.1	1104	99.6	6.6
4.03	0.055	36.7	1098	99.9	6.7
3.66	0.050	32.2	1094	100.0	6.7
3.40	0.071	23.6	1079	100.0	6.7
3.20	0.120	14.5	1075	100.0	6.7
3.04	0.180	10.4	1092	100.0	6.7
2.91	0.293	6.4	1076	100.0	6.7
2.80	0.469	4.3	1077	100.0	6.7
2.70	0.668	3.1	1066	100.0	6.7
Total	0.049	18.0	10860	99.5	6.7

5.1.1.2 Native (crystal form II)

Crystal form II was found to diffract to 3.7 Å on the same beam line and under the conditions as described above; processing was also performed as described previously. The crystal symmetry was determined to belong to primitive rhombohedral lattice with unit cell dimensions of $a=b=132.6$ Å and $c=113.8$ Å. The same space group and the similar unit cell dimensions indicate similar packing in the two crystal forms. As the resolution of the data from these crystals was not as good as from crystal form I no further data were collected.

5.1.1.3 MAD

SeMet labelled ATP-PRT was crystallised under conditions similar to those used for crystal form I of the native enzyme. The concentration of sodium tartrate was reduced from 1.3 M to 1.1 M, the MgCl_2 concentration was the same (50-200 mM). The crystals show a clearer morphology and grow to a larger size than their native counterparts ($0.5 \times 0.3 \times 0.3 \text{ mm}^3$); however the resolution of the diffraction was not improved.

Two independent three wavelength MAD data sets of Se-Met labelled protein were collected on a single crystal at ALS beam line 5.0.2 using a ADSC Quantum-4

CCD detector (G. McDermott). First a fluorescence spectrum of the Se-Met labelled crystal was collected. The spectrum which shows a clear peak is shown in Figure 5-2. The three wavelengths used in the experiment were 0.9796 Å (12656 eV) for the anomalous peak (f''), 0.9798 Å (12653 eV) for the inflection point (f') and 0.9611 Å (12900 eV) as a remote wavelength.

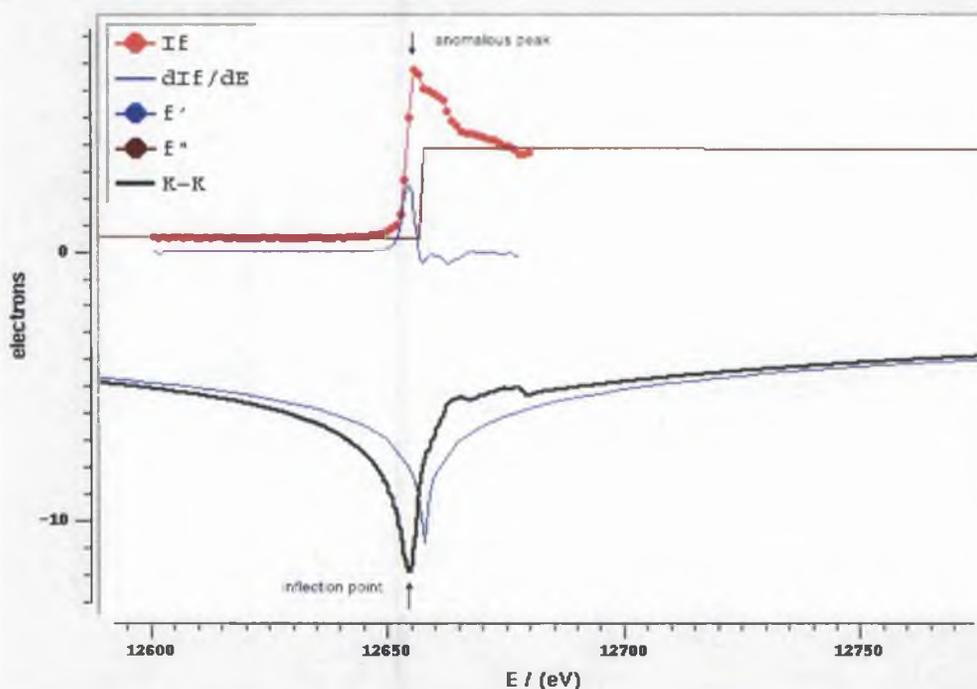


Figure 5-2 Fluorescence spectrum of Se-Met labelled HisG crystal. Shown is the experimental fluorescence $I f$ (red) which is proportional to f'' , its derivative $d I f / d E$ (blue) and the theoretical f' (blue) and f'' (brown). The Kramers-Kronig transformation is used to determine the experimental f' (K-K, green).

All diffraction data were collected with 1° oscillation. The first data set consists of two 60° wedges collected continuously for each wavelength. The wedges were separated by 180° to collect the corresponding Friedel pairs of the observed reflections. The second data set was collected in two 80° wedges again separated by 180° . Here the wedges were split in smaller units of 10° which were collected in alternation 180° apart. This method has the advantage that the Friedel pairs are collected in quick succession and possible shifts in the wavelength have a smaller effect on the anomalous differences. On the downside the crystal has to be rotated more often and possible inaccuracies in the oscillation can have a larger effect than in the first method described. All data were processed with DENZO and scaled with

SCALEPACK [Otwinowski and Minor, 1997]. Statistics for the data processing are listed in Table 5-2.

Table 5-2 Crystallographic data and phasing of *E. coli* HisG

Data set	Native AMP		SeMet	
Space group	R32		R32	
Unit cell (Å)	$a=b=133.6$ $c=114.1$		$a=b=134.1$ $c=114.5$	
Resolution (Å)	2.7		2.7	
Wavelength (Å)	0.87	(f') 0.9798	(f'') 0.9796	0.9611
Unique reflections	10860	10851	10835	10846
Completeness (%) ¹	99.5 (100.0)	99.5(97.4)	99.6(98.3)	97.7(100.0)
R _{merge} (%) ^{1,2}	5.9 (79.6)	7.7 (78.8)	8.7 (87.4)	7.8 (72.6)
Multiplicity ¹	6.7 (6.7)	7.8 (6.3)	8.6 (7.0)	8.1 (7.3)
Mean $\langle I/\sigma(I) \rangle$ ¹	18.0 (3.1)	13.0 (2.5)	13.9 (2.3)	14.2 (2.1)

¹ Number in parenthesis represents values for highest resolution bin.

$$^2 R_{\text{merge}} = \frac{\sum |I_{\text{obs}} - I_{\text{avg}}|}{\sum I_{\text{avg}}}$$

5.1.1.4 HisG complexed with histidine

Diffraction data from crystal form I, grown in the presence of 2 mM AMP and 5 mM histidine, were collected on beam line BW7A at HASYLAB, DESY, Hamburg. A dataset with 1° oscillation and a total of 69 frames was collected with the data statistic shown in Table Table 5-3.

Crystal form III, grown in the presence of 10 mM AMP and 10 mM histidine diffracted to 3 Å at best. A complete data set was collected to 3.3 Å. However diffraction was very diffuse with a high mosaic spread (up to 1.6°) which made data processing difficult. Autoindexing with DENZO suggested a primitive monoclinic lattice with cell dimensions of $a=91.19$ Å $b=113.58$ Å $c=101.58$ Å and $\beta=113.74^\circ$. Systematic absences (0,2n+1,0) indicated space group $P2_1$. Merging and scaling with SCALEPACK proved difficult with R-factors of individual frames all above 10% (up to 30%). Calculation of the Matthews coefficient suggests a hexamer in the asymmetric unit (asu). There is a possibility that the actual space group is $P1$ rather than $P2_1$. However the merging statistics are only slightly improved with an overall merging R-factor of 16.5% compared to 18.8% for $P2_1$. Additionally α and β refine to

90.01° and 90.14° respectively in *P1* which suggests that the space group has two 90° angles.

5.1.1.5 Crystal form I complexed with PR-ATP

Diffraction data were collected on beam line 9.6 at the SRS Daresbury Laboratory using a wavelength of 0.86 Å using a ADSC Quantum-4 CCD detector. Crystals were loop-mounted in a cryoprotectant containing 15% (v/v) glycerol and 2 mM PR-ATP and cryo-cooled to 100 K using an Oxford Cryosystems cryostream. A dataset was collected using 1° oscillation frames with a total of 71 frames being collected. Data were processed with DENZO and scaled with SCALEPACK [Otwinowski and Minor, 1997]. The statistics for that dataset is shown in Table 5-3.

Table 5-3 HisG *E. coli* ligand structure statistics

Data set	Native AMP	Histidine	PR-ATP
Space group	<i>R</i> 32	<i>R</i> 32	<i>R</i> 32
Unit cell (Å)	$a=b=133.6$ $c=114.1$	$a=b=133.0$ $c=113.2$	$a=b=132.1$ $c=113.4$
Resolution (Å)	2.7	2.9	2.9
Wavelength (Å)	0.87	0.8424	0.95988
Unique reflections	10860	8707	8521
Completeness (%) ¹	99.5 (100.0)	99.8 (100.0)	99.3 (99.9)
R_{merge} (%) ^{1,2}	5.9 (79.6)	5.6 (52.1)	5.8 (58.1)
Multiplicity ¹	6.7 (6.7)	4.3 (4.2)	4.2 (4.2)
Mean $\langle I/\sigma(I) \rangle$ ¹	18.0 (3.1)	20.2 (2.3)	16.9 (2.0)

¹ Number in parenthesis represents values for highest resolution bin.

² $R_{\text{merge}} = \Sigma |I_{\text{obs}} - I_{\text{avg}}| / \Sigma I_{\text{avg}}$

5.1.2 HisG from *Mth*

5.1.2.1 Native

X-ray diffraction data on a crystal of *Mth* HisG were collected on beam line 9.5 at the SRS Daresbury Laboratory using a wavelength of 1.0 Å using a 165mm MAR CCD detector. Crystals were loop-mounted in dry paraffin oil and cryo-cooled to 100 K using an Oxford Cryosystems cryostream. A native dataset to 2.6 Å resolution was collected using 0.5° oscillation frames. The smaller oscillation range

was chosen because of the high mosaic spread of the crystal (up to 1.6°) and large unit cell dimensions. Data were processed with DENZO and scaled with SCALEPACK [Otwinowski and Minor, 1997]. From autoindexing in DENZO, the crystals were found to index in a primitive rhombohedral lattice, corresponding to space group $R32$, with unit-cell dimensions of $a=b=112.10 \text{ \AA}$ and $c=247.62 \text{ \AA}$. Data statistics are shown in Table 5-4. The assumption of two molecules per asymmetric unit for $R32$ leads to a packing density, V_m of $2.5 \text{ \AA}^3 \text{ Da}^{-1}$ which corresponds to a solvent content of 49.6% [Matthews, 1968]. Assuming only one molecule V_m is calculated as $4.9 \text{ \AA}^3 \text{ Da}^{-1}$ with a solvent content of 74.9% which was considered less likely.

Table 5-4 Crystallographic data statistics for *Mth* HisG

Data set	Native histidine
Space group	$R32$
Unit cell (\AA)	$a=b=132.10 \text{ c}=247.62$
Resolution (\AA)	2.59
Wavelength (\AA)	1.0
Unique reflections	18179
Completeness (%) ¹	93.4 (84.9)
R_{merge} (%) ^{1,2}	4.0 (26.6)
Multiplicity ¹	3.2 (1.6)
Mean $\langle I/\sigma(I) \rangle$ ¹	16.8 (1.9)

¹ Number in parenthesis represents values for highest resolution bin.

$$^2 R_{\text{merge}} = \frac{\sum |I_{\text{obs}} - I_{\text{avg}}|}{\sum I_{\text{avg}}}$$

5.1.3 PA5104 from *Pseudomonas aeruginosa*

5.1.3.1 Native (crystal form 1)

One large crystal (approx. $0.5 \times 0.2 \times 0.2 \text{ mm}^3$) of crystal form I from the initial screen was loop-mounted and cryo-cooled to 100 K. The precipitant PEG 400K was sufficient as a cryoprotectant, so no further soaking of the crystals was required. A native dataset was collected on beam line 14.1 at the SRS in Daresbury (A. Riboldi-Tunnicliffe). 360 frames of 1° oscillation were recorded at a wavelength of 1.488 \AA using a ADSC Quantum-4 CCD detector. Most low resolution spots of the diffraction

pattern were overloaded and spiked into neighbouring spots (Figure 5-4). However autoindexing with DENZO suggested a C centered orthorhombic or primitive monoclinic lattice. Data were processed with DENZO in the higher symmetry space group $C222$ with unit cell dimensions of $a=40.758 \text{ \AA}$ $b=109.530 \text{ \AA}$ $c=93.730 \text{ \AA}$. Scaling of the data with SCALEPACK was problematic. Further attempts to process and scale these data with d*trek and Mosflm/SCALA to allow absorption correction were equally unsuccessful.

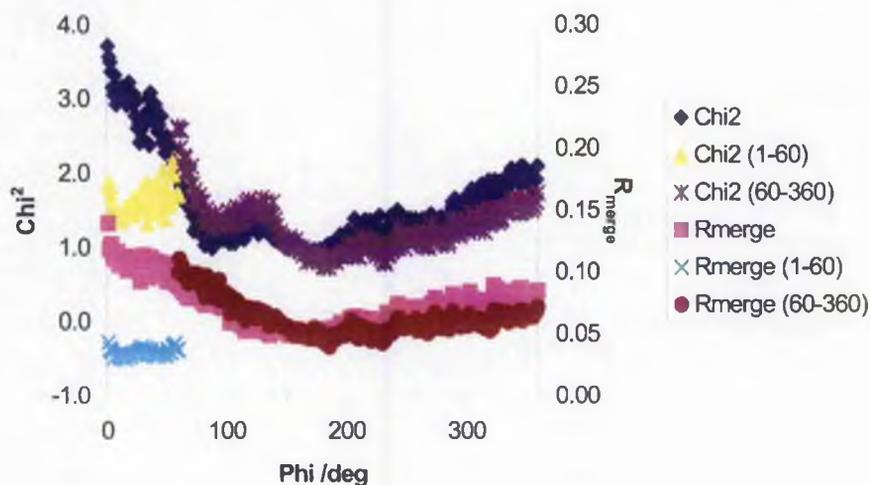


Figure 5-3 Chi^2 and R_{merge} vs oscillation.

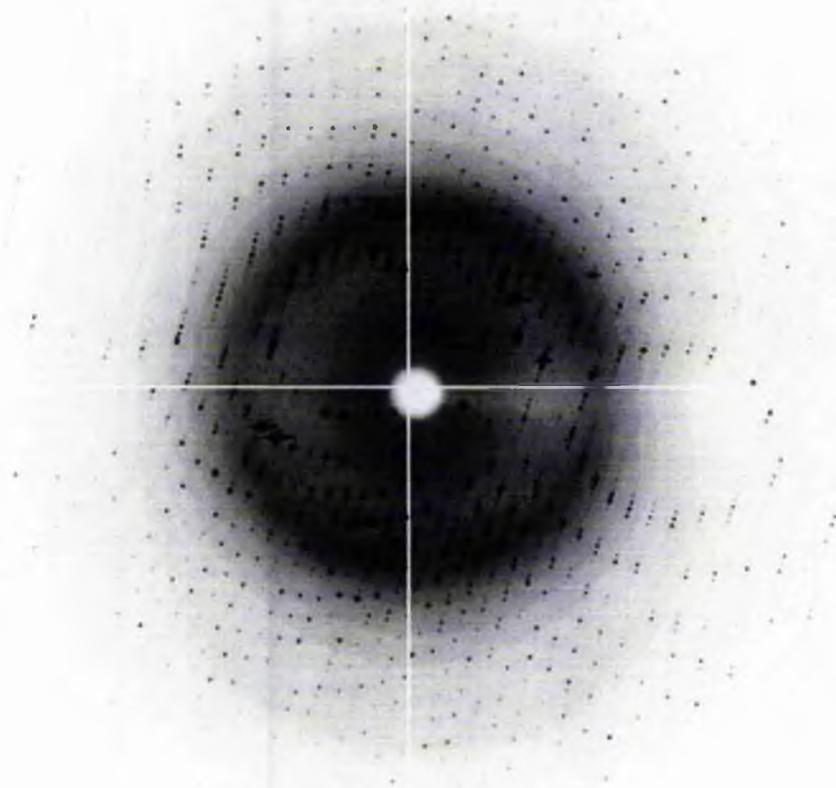


Figure 5-4 Diffraction pattern from PA5104.

To avoid the problematic overloads a smaller crystal, re-grown from the initial screen, was loop-mounted and cryo-cooled to 110 K for data collection in house on a Nonius FR591 rotating anode generator. 150 frames at 1° oscillation were collected with CuK_α radiation at wavelength 1.54 \AA using a Macscience DIP2000 image plate system. Autoindexing with DENZO confirmed the lattice to be C centred orthorhombic with similar unit cell dimensions as seen before ($a=40.527 \text{ \AA}$ $b=107.951 \text{ \AA}$ $c=93.443 \text{ \AA}$). The data were processed with DENZO in *C222* but again merging and scaling the data with SCALEPACK (and other programs) caused problems and was unsuccessful.

A third native dataset on PA5104 crystals was collected at the beam line BL 5.2 R (XRD1) at the ELLETRA synchrotron in Trieste. A medium sized, nearly cubic crystal (approx. $0.3 \times 0.3 \times 0.4 \text{ mm}^3$) was loop-mounted and cryo-cooled to 100 K.

Data were collected at a wavelength of 1.0 Å using a 165mm CCD detector from MarResearch. Two datasets were collected, first one for the lower resolution range (360 frames at 1° oscillation) and then a high resolution range (720 frames at 0.5° oscillation). Autoindexing with DENZO confirmed the C centred orthorhombic lattice, so the datasets were processed in DENZO using space group $C222_1$. Both datasets merged and scaled independently, and together, without the problems seen previously using SCALEPACK. Systematic absences at (0,0,2n+1) suggested that the actual space group is $C222_1$ rather than $C222$. Data collection and scaling statistics for the low and high resolution as well as the combined data are shown in Table 5-5.

Table 5-5 Data collection and scaling statistics of PA5104

Data set	Native all	Native low res	Native high res
Space group	$C222_1$		
Unit cell (Å)	$a=40.631 \text{ Å } b=109.180 \text{ Å } c=93.243 \text{ Å}$		
Resolution (Å)	50 - 0.96	50 - 1.55	20 - 0.96
Wavelength (Å)	1.0		
Comment	Anomalous	Anomalous	-
Unique reflections	126365	30358	126365
Completeness (%) ¹	100.0 (100.0)	99.9 (99.4)	97.8 (100.0)
R_{merge} (%) ^{1,2}	7.1 (35.7)	7.2 (16.1)	5.9 (37.2)
Multiplicity ¹	13.2 (12.0)	14.1 (13.7)	13.3 (12.1)
Mean $\langle I/\sigma(I) \rangle$ ¹	22.6 (5.5)	30.5 (16.2)	28.7 (7.4)

¹ Number in parenthesis represents values for highest resolution bin.

² $R_{\text{merge}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$

A comparison of the combined data with the previous, unsuccessfully scaled data is shown in Table 5-6.

Table 5-6 Comparison of PA5104 data collection and scaling statistics

Data set	Native	Native	Native
Source	ELLETRA	DRS	Rotating anode
Detector	MAR CCD	ADSC QUANTUM	MAR DIP2000
Wavelength (Å)	1.0	1.488	1.54
Space group	$C222_1$	$C222_1$	$C222_1$
Unit cell (Å)	$a=40.631$ Å $b=109.180$ Å $c=93.243$ Å	$a=40.758$ Å $b=109.530$ Å $c=93.730$ Å	$a=40.527$ Å $b=107.951$ Å $c=93.443$ Å
Resolution (Å)	50 - 0.96	100 - 1.38	50 - 2.6
Comment	Anomalous	Anomalous	Anomalous
Degree	360	360	150
R_{merge} (%) ^{1,2}	7.1 (35.7)	7.1 (33.1)	13.0 (30.0)
χ^2	1.0	1.6	3.2
Completeness (%) ¹	100.0 (100.0)	82.5 (25.0)	96.5 (92.2)
Mean $\langle I/\sigma(I) \rangle$ ¹	22.6 (5.5)	8.7 (2.1)	19.2 (8.8)
number of rejections / total number of reflections	$4.48 \cdot 10^{-3}$	$17.33 \cdot 10^{-3}$	$51.60 \cdot 10^{-3}$

¹ Number in parenthesis represents values for highest resolution bin.

$$^2 R_{\text{merge}} = \frac{\sum |I_{\text{obs}} - I_{\text{avg}}|}{\sum I_{\text{avg}}}$$

Assuming one molecule in the asymmetric unit, of the $C222_1$ space group with $a=40.631$ Å $b=109.180$ Å $c=93.243$ Å, V_m is calculated to acceptable 2.41 Å³ Da⁻¹ with a solvent content of 48.66%.

5.1.3.2 *S* anomalous

The same crystal, which was used to collect the native dataset, was used to collect diffraction data at higher wavelength at the XDR1 beam line at ELLETRA. The crystal had to be re-centred after changing the wavelength and 720 frames with 1° oscillation were collected using a wavelength of 1.9 Å. Data collected with this longer

wavelength and high redundancy should be sufficient to solve the phases from sulphur anomalous scattering.

The data were processed without any problems in DENZO. Scaling with SCALEPACK proved impossible as seen before with high χ^2 values (above 5 on average) and high R-factors for each frame (above 10%). Furthermore the merging R-factors were 14% and above even for the lower resolution shells. For diffraction data measured with a high redundancy (on average above 15) and a low signal to noise ratio ($I/\sigma(I)$ on average higher than 20) much lower scaling statistics would be expected. Inspection of the data showed that a large number of reflections measured multiple times were significantly different. Table 5-7 shows the rejections for reflection (1,1,18) as a typical example. It shows that the intensity of the same reflection (not just the symmetry related ones) differs above the expected deviations from a Gaussian distribution. Radiation damage as a cause could be excluded as the intensities in later frames did not always show decreased intensities. Absorption effects are possible, however the observation that the deviations and the signs thereof seem to be independent of the film number and therefore rotation angle makes it less likely.

To account for possible absorption effects data were processed, scaled and merged with Mosflm/SCALA and d*trek as these allow corrections for these effects. Scaling with d*trek without absorption correction again gave high values for the R-factors while absorption correction could not be applied (the program could not account for the effects⁸). Scaling with SCALA using absorption correction gave high values for R-factors too (above 10% in all shells).

To allow protein structures to be solved using sulphur anomalous scattering the data have to be measured with high accuracy as the anomalous signal is very low. An estimate for PA5104, which has 6 residues containing sulphur (2 Met and 4 Cys), gives an anomalous signal of about 1% for the wavelengths used here [Lemke *et al.*, 2002]. Because of the scaling problems with large deviation between related reflections this small signal is unlikely to be detected reliably above the increased noise of the data obtained here. Therefore these data were not used any further.

⁸ Presumably the differences of symmetry related and identical intensities were too large and not sufficiently systematic to calculate a correct absorption correction.

Table 5-7 Rejections of reflection (1,1,18). The total χ^2 was 872.6 for an average intensity of $\langle I \rangle = 22883$ at a resolution of 5.13 Å. Only fully recorded reflections of the first 360° of data are shown here.

h	k	l	film	Deviation/ σ	I	$\sigma(I)$	Rejection probability
1	1	18	11	-4.7	2288.3	116.8	86
			169	9.3	4434.7	131.4	100
1	-1	-18	0	1.2	3112.5	115.9	0
			206	-9.4	1634.3	110.6	100
			360	-1.4	2790.5	114.8	0
-1	-1	18	24	-5.5	2181.4	111.3	100
			176	-4.8	2281.7	114.2	90
-1	1	-18	188	-7.9	1799.5	118.2	100
-1	-1	-18	191	-14.9	885.3	108.2	100
			347	-2.2	2681.4	115.6	0
-1	1	18	26	-0.5	2892.8	116.8	0
			180	-8.8	1710.3	111.0	100
1	1	-18	204	-12.7	1187.3	108.6	100
			356	2.9	3326.7	115.4	1
1	-1	18	7	3.1	3465.2	154.8	1
			162	-1.3	2791.2	125.3	0

5.1.3.3 Derivative

5.1.3.3.1 Pt derivatives

One crystal of PA5104 was soaked with 2 mM platinum (II) 2,2':6',2''-terpyridine chloride dihydrate. A derivative dataset was collected on beam line XDR1 at ELLETRA in Trieste (A. Riboldi-Tunnicliffe). Data were processed with DENZO and merged and scaled anomalously with SCALEPACK by Alan Riboldi-Tunnicliffe. Statistics of the data can be seen in Table 5-8.

Table 5-8 Scaling statistics for PA5104 derivative data

Data set	Pt	Pt	Hg	Hg
Source	ELLETRA	SRS 9.5	SRS 9.5	SRS 9.5
Space group		C222 ₁		
Unit cell (Å)	<i>a</i> =40.551	<i>a</i> =40.594	<i>a</i> =40.588	<i>a</i> =40.609
	<i>b</i> =108.460	<i>b</i> =109.062	<i>b</i> =108.946	<i>b</i> =108.802
	<i>c</i> =93.320	<i>c</i> =92.937	<i>c</i> =92.901	<i>c</i> =93.076
Resolution (Å)	2.20	2.06	2.06	1.75
Wavelength (Å)	1.0	0.9341	1.0064	1.0050
Unique reflections	10906	13171	13151	21180
Completeness (%) ¹	83.8 (95.5)	99.0 (98.0)	99.6 (99.2)	99.6 (96.2)
R _{merge} (%) ^{1,2}	3.9 (4.2)	5.2 (8.4)	8.0 (16.5)	8.3 (58.5)
Multiplicity ¹	n/a	14.7 (14.4)	14.7 (12.3)	13.7 (7.2)
Mean <I/σ(I)> ¹	16.9 (18.7)	53.3 (32.3)	35.6 (17.7)	25.8 (2.7)

¹ Number in parenthesis represents values for highest resolution bin.

$$^2 R_{\text{merge}} = \frac{\sum |I_{\text{obs}} - I_{\text{avg}}|}{\sum I_{\text{avg}}}$$

A second Pt derivative dataset was collected on beam line 9.5 at the SRS in Daresbury. A crystal was soaked for 3 min in well solution containing 5 mM ammonium tetrachloroplatinate(II) and briefly back-soaked before being loop-mounted and cryo-cooled. Statistics of the data after processing with DENZO and scaling with SCALEPACK can be found in Table 5-8.

5.1.3.3.2 Hg derivatives

First one crystal was soaked in well solution containing 0.5 mM HgCl₂ for 10 minutes. A complete dataset was collected on beam line 9.5 at the SRS Daresbury.

A second crystal was soaked for 15 minutes and a fluorescent scan on the Hg L-III edge recorded. Only a weak peak (in fact more a shoulder) was observed and the wavelength changed to be slightly lower wavelength than the start of the peak. A complete dataset was recorded. Scaling statistics after processing with DENZO and scaling with SCALEPACK are shown in Table 5-8.

5.1.3.4 Wavelength dependency

To investigate possible absorption effects and scaling problems at higher wavelength of the PA5104 crystals 5 datasets were collected on one crystal using 5 different wavelengths (0.7, 0.9, 1.1, 1.3, 1.5 Å⁹) at beam line 9.5 at the SRS Daresbury. The crystal was approx. 0.15 x 0.4 x 0.4 mm³ in size and was loop-mounted so that the largest face was in plane with the loop. Collection of 180° with an oscillation range of 1° ensured that the path through the crystal changed maximally with the phi rotation. With the increase of the wavelength the crystal-detector distance was shortened accordingly so that the resolution at the detector edge was constant. This kept the spot size and separation constant and identical parameter files could be used for processing and scaling. Therefore differences in processing and scaling statistics at different wavelength are solely wavelength dependent (ignoring radiation damage). Data reduction statistics as obtained from the programs DENZO and SCALEPACK for the five wavelengths are shown in Table 5-9.

Table 5-9 Scaling statistics for wavelength dependency of PA5104

Data set	1	2	3	4	5
Space group	C222 ₁				
Unit cell (Å)	<i>a</i> =40.541	<i>a</i> =40.560	<i>a</i> =40.570	<i>a</i> =40.565	<i>a</i> =40.576
	<i>b</i> =108.508	<i>b</i> =108.557	<i>b</i> =108.582	<i>b</i> =108.585	<i>b</i> =108.595
	<i>c</i> =92.963	<i>c</i> =93.008	<i>c</i> =93.029	<i>c</i> =93.026	<i>c</i> =93.041
Resolution (Å)	2.05				
Wavelength (Å)	0.7	0.9	1.1	1.3	1.5
Unique reflections	13321	13291	13291	13291	13291
Completeness (%) ¹	99.7 (99.3)	99.4 (98.6)	98.9 (98.1)	98.2 (96.7)	97.3 (95.3)
R _{merge} (%) ^{1,2}	5.8 (10.5)	5.9 (10.6)	5.9 (10.7)	6.7 (15.4)	6.8 (17.8)
Multiplicity ¹	7.1 (6.3)	7.1 (6.3)	7.0 (6.6)	6.9 (6.7)	6.9 (6.8)
Mean <I/σ(I)> ¹	32.3 (16.8)	33.0 (17.6)	33.5 (17.4)	29.9 (13.4)	29.7 (12.2)
Rejections	134	145	149	225	59

¹ Number in parenthesis represents values for highest resolution bin.

$$^2 R_{\text{merge}} = \frac{\sum |I_{\text{obs}} - I_{\text{avg}}|}{\sum I_{\text{avg}}}$$

⁹ 1.5 Å was the longest wavelength used because of the set-up of the beamline it was not easily possible to gain longer wavelengths.

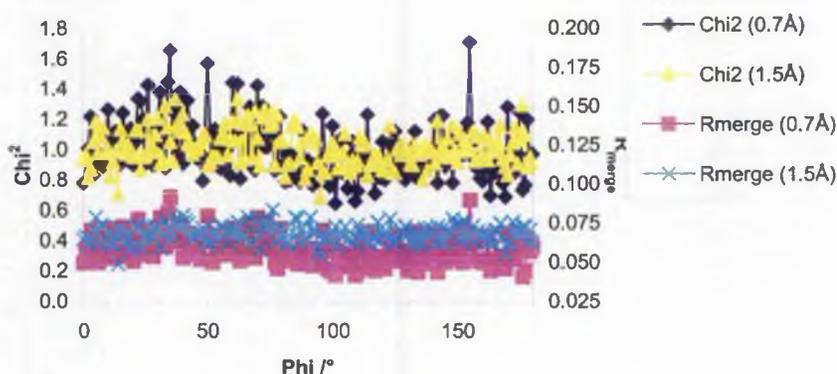


Figure 5-5 Chi^2 and R_{merge} for 0.7 and 1.5 Å vs. Phi oscillation.

The R_{merge} and Chi^2 of individual frames versus oscillation angle (frame) for the lowest (0.7 Å) and highest (1.5 Å) wavelength investigated are shown in Figure 5-5. No significant difference can be seen for the Chi^2 values and R_{merge} is marginally higher for the longer wavelength. Furthermore no difference in angular dependency can be seen, not even for the longest wavelength.

R_{merge} after scaling and completeness versus wavelength are shown in Figure 5-6. A decrease in completeness with increasing wavelength can be observed. R_{merge} increases at 1.3 Å by about one percent and to a greater extent for the outer resolution shell. It is difficult to explain these results especially as the number of rejections does not increase substantially with increasing wavelength which could indicate absorption effects. A reason why no oscillation dependency for the longer wavelength compared to the previous observed data (see Section 5.1.3.1) was observed could be the size of the crystal. Here a smaller crystal was used compared to the previous data collection at comparable wavelengths.

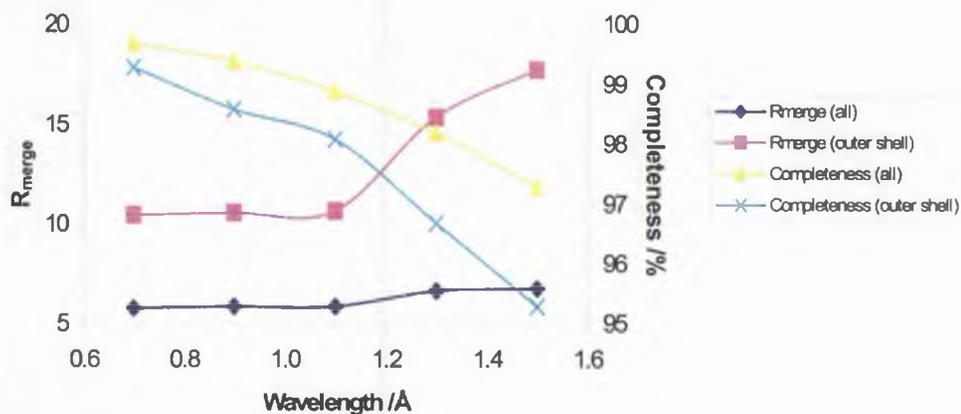


Figure 5-6 R_{merge} and completeness vs. wavelength for PA5104.

5.1.3.5 Native (crystal form II)

One crystal of crystal form II was loop-mounted and cryo-cooled to 100 K. Mother liquor containing 20% glycerol was used as a cryoprotectant. A native dataset was collected on beam line 9.6 at the SRS in Daresbury. 180 frames of 1° oscillation were recorded at a wavelength of 0.864 Å using a ADSC Quantum-4 CCD detector. Some of the diffraction spots were slightly blurred and in some frames an additional lattice was observed. This could have been due to a split or twinned crystal. However autoindexing with DENZO was successful and suggested a primitive monoclinic lattice. Data were processed with DENZO in the space group $P2_1$ with unit cell dimensions of $a=69.533$ Å $b=67.175$ Å $c=84.222$ Å and $\beta=111.932^\circ$. The data were merged and scaled with SCALEPACK. Systematic absences at $(0,2n+1,0)$ suggest the space group of $P2_1$. Data collection and reduction statistics are shown in Table 5-10.

Calculation of V_m with the above mentioned unit cell parameters gives two acceptable values. It is calculated as 2.8 Å³ Da⁻¹ with a solvent content of 55% for 3 monomers in the asu or 2.1 Å³ Da⁻¹ and 40% solvent for 4 molecules.

Table 5-10 Scaling statistics of PA5104, crystal form 2

Data set	PA 5104 crystal form 2
Source	SRS 9.6
Space group	$P2_1$
Unit cell (Å)	$a=69.533$ $b=67.175$ $c=84.222$ $\beta=111.932^\circ$
Resolution (Å)	50-1.9
Wavelength (Å)	0.864
Unique reflections	56322
Completeness (%) ¹	99.8 (99.9)
R_{merge} (%) ^{1,2}	9.0 (64.5)
Multiplicity ¹	3.5 (3.2)
Mean $\langle I/\sigma(I) \rangle$ ¹	14.0 (1.9)

¹ Number in parenthesis represents values for highest resolution bin.

² $R_{\text{merge}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$

5.2 Calculation of structure factors from intensities

Files from SCALEPACK containing miller indices (h,k,l) and corresponding intensities (sca files) were transformed in mtz format using the CCP4 program SCALEPACK2MTZ [CCP4, 1994]. Structure factors were calculated from these intensities using the program TRUNCATE [French and Wilson, 1978]. The overall B factor was calculated using Wilson statistics and twinning of the data was checked for by examining the distribution of centric and acentric reflections in the cumulative intensity distributions generated by TRUNCATE. Different datasets were combined in one reflection file (mtz format) with the CCP4 program cad [CCP4, 1994] and a free R column added. Solving of the phases for each structure is described within the appropriate sections.

Chapter 6: Structure of HisG from *E. coli*

6.1 Introduction

HisG from *E. coli* was the first enzyme to be investigated structurally in this study. An over-expression clone was already available [A. Elwell unpublished results] and together with the *S. typhimurium* enzyme is the best studied [Dall-Larsen, 1988]. Initial crystallisation conditions, crystallographic characterization and prediction of the quaternary structure has been published earlier, based on earlier work in the laboratory [Lohkamp *et al.*, 2000].

6.2 Solving the phases of *E. coli* HisG

The native crystals of HisG did not diffract to high resolution (see Section 5.1.1) presumably due to the inherent flexibility of the protein. Solving phases using MIR one has to eliminate the problems of anisomorphism associated with this method. This can be especially problematic for cases where the protein is flexible as seen for HisG. To overcome the anisomorphism, caused by using multiple crystals to solve the phases, other phasing methods such as SAD, MAD or SIRAS can be used which use data collected on a single crystal. MAD usually gives the most reliable phases as it uses at least three different wavelengths which give an anomalous and isomorphous signal. Therefore MAD phasing using Se-Met labelled protein was employed to provide phase information.

6.2.1 MAD phasing

Collection and processing of the MAD data sets is described in Section 5.1.1. The program SOLVE [Terwilliger and Berendzen, 1999] was used to find the Se sites from three wavelength MAD data. 11 out of the possible 13 Met (including N-terminal Met) of *E. coli* HisG were found with a Figure of Merit (FoM) of 0.62 and a Z score of 55.98 from resolution range 28-3 Å. Phases from these initial 11 sites were refined with MLPHARE [Otwinowski, 1991]. This resulted in an overall FoM 0.43 for the resolution range 30-2.7 Å (isomorphous differences) and 30-2.8 Å (anomalous differences). These starting phases were further improved by solvent flattening with histogram matching using the program DM [Cowtan, 1994], which gave a readily

interpretable electron density map. One more selenium site was identified during model building and was subsequently refined together with the other 11 sites in MLPHARE to improve the experimental phases. This resulted in a FoM of 0.44 and 0.86 after DM for the resolution range 30-2.7 Å (see Table 6-1).

Table 6-1 Phasing statistics for *E. coli* HisG

Wavelength (Å)	(f') 0.9798	(f'') 0.9796	0.9611
Phasing power acentric/centric	0.94/0.72	1.68/1.00	0.62/0.49
Mean figure of merit	0.4378		

6.2.2 Model building and refinement

The initial model of ATP-PRT was traced from a 2.7 Å DM map using QUANTA [Accelrys Inc.]. Figure 6-1 shows the initial electron density obtained from DM together with a built C_α trace and one Se site.



Figure 6-1 MAD solvent flattened electron density map of HisG. The C_α trace is shown in green and red with the electron density in purple. One identified selenium site is marked with a yellow sphere.

Iterative cycles of model building in QUANTA followed by refinement using CNS [Brunger *et al.*, 1998] with experimental phases included as restraints were performed. First the model was refined using simulated annealing and torsion angle dynamics followed by group B-factor refinement. Once the model was more complete energy minimization refinement in CNS was used. One AMP molecule as well as 2 tartrate molecules could be fitted to the electron density. The final refinement cycles were carried out using REFMAC5 [Murshudov, 1996] including TLS refinement [Winn *et al.*, 2001]. Three segments were used for TLS refinement corresponding to the three domains of HisG (see Section 6.3.2). The quality of the electron density map and low resolution of the data allowed only two water molecules to be build and refined, hence these were omitted and no water molecules added to the final model. Table 6-2 shows the decrease in R- and free R-factor during the steps of refinement.

Table 6-2 R- and free R-factor during refinement

Refinement step	R-factor /%	free R-factor /%
Start	42.9	46.6
Simulated annealing (CNS)	31.1	39.3
B group refinement	28.6	36.0
Energy minimization (CNS)	26.4	33.0
B group refinement	24.0	31.2
Refmac5 with TLS	22.2	27.7

6.3 HisG model

6.3.1 Quality of the model

The final model of HisG with $R_{\text{work}} = 22.1\%$ and $R_{\text{free}} = 27.7\%$ had rms deviation in bond distances of 0.017 Å and bond angles of 1.9°. The geometrical parameters were either satisfactory or better than expected for a structure of this resolution as assessed by PROCHECK [Laskowski *et al.*, 1993]. No residues were found in disallowed regions of the Ramachandran plot. Table 6-3 shows a summary of the refinement statistics.

Table 6-3 Refinement statistics of *E. coli* HisG with AMP

Refinement statistics	
Resolution range (Å)	81.65-2.70
R _{working} (%) ¹	22.2 (9975)
R _{free} (%) ^{1,2}	27.7 (883)
Number of atoms	
Protein	2187
Nonprotein (AMP, tartrate)	42
R.m.s. deviation from ideal	
Bond length (Å)	0.017
Bond angles (°)	1.9
Ramachandran statistics	
Most favoured	212
Additionally allowed	36
Generously allowed	6
Disallowed	0

¹ number in parenthesis is the number of reflections.

² R_{free} was calculated with 8.1% of reflections.

6.3.2 Overall structure

The crystals were grown with AMP which stabilises the hexameric form of the enzyme. From preliminary crystallographic analysis the quaternary structure of the hexamer was predicted [Lohkamp *et al.*, 2000]. The HisG hexamer is formed via crystallographic symmetry from a single L-shaped monomer (Figure 6-2), which comprises the asymmetric unit. The monomer is separated into three clearly defined α/β protein domains, of which the first two are linked by a two-stranded antiparallel β -sheet and form a cleft in which the active site is situated, while the third domain is connected to the first domain via an extended α -helix.

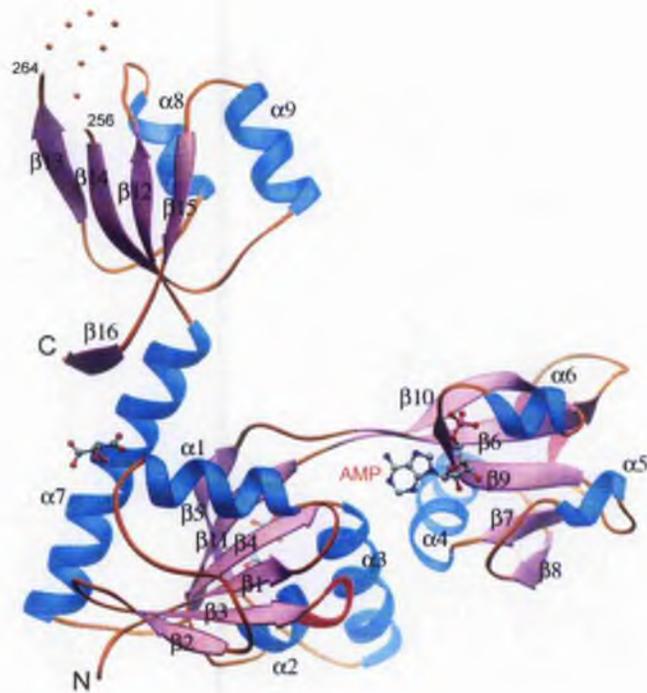


Figure 6-2 Structure of *E. coli* HisG. A ribbon representation of *E. coli* ATP-PRT monomer coloured according to secondary structure. The beginning and end of the disordered loop are labelled and the missing residues represented as spheres. One molecule of AMP and two molecules of tartrate are represented in ball and stick. The N- and C-termini are labelled with N and C, respectively. This diagram was produced using RIBBONS [Carson, 1987].

The folding of each of the three domains is distinct (Figure 6-2). The first domain consists of residues 5-103 and 192-224. It contains a core of 4 parallel and 2 anti-parallel β -strands, with a strand order of 231465 ($\downarrow\uparrow\uparrow\downarrow\uparrow$), flanked by 4 α -helices, two on each side. The second domain consists of residues 103-191. The core is built of a twisted 5-stranded β -sheet, 4 strands parallel and 1 anti-parallel with a strand order of 32415 ($\uparrow\uparrow\uparrow\downarrow\uparrow$). The β -sheet appears to wrap around helix $\alpha 4$ on one side, with the opposite side covered by two short α -helices ($\alpha 5$ and $\alpha 6$). One of the most unusual features of the folds of both domain 1 and domain 2 is the interdigitation of the β -strands. In α/β proteins containing mainly parallel β -sheets, it is most commonly found that β - α - β motifs, where the two strands are adjacent, are the preferred building blocks. In ATP-PRT this feature is rare with only $\beta 7$ - $\alpha 4$ - $\beta 8$ of the

second domain forming such a motif which is followed by a crossover connection forming a crevice where the AMP is located (Figure 6-2). Analysis of the two domains using DALI [Holm and Sander, 1995] reveals that they are not unique, together they form a "periplasmic binding protein" fold, sharing most structural homology to the glutamate receptor fragment (pdb accession code 1II5; rms deviation 4.0 Å for 219 C_α atoms) and the transcription factor CysB (pdb accession code 1AL3; rmsd 4.9 Å over 235 C_α atoms) (see Figure 6-3). The crevice between the two domains is the ligand-binding site in this fold and in ATP-PRT is undoubtedly the active site of the protein.



Figure 6-3 Superposition of *E. coli* HisG with glutamate receptor fragment. Ribbon diagram of glutamate receptor fragment (1ii5) in blue superimposed on the structure of HisG from *E. coli* (red). Ligands glutamate (Glu) and AMP are shown in stick representation. This figure was prepared using PyMOL [DeLano,]. All following figures were prepared using PyMOL unless otherwise stated.

Another feature of the first two domains and the periplasmic binding proteins is the 18 Å long pair of anti-parallel β -strands, which connects domain 1 to domain 2. This is formed by the continuation of strand β 5 into β 6 and β 10 into β 11 which is interrupted by a β -bulge at Glu 186 on strand β 10 and a larger extended bulge (residues Asp 100 - Gly 102) on strand β 5. This inter-domain anti-parallel structure is held together by five main chain hydrogen bonds and a number of salt bridges at either end. The outer face of the two β -strands is characterised by predominantly positively charged residues while the inner face comprises small residues Cys, Ser, Ala and Gly. An exception is Tyr 189, which due to the twist of the sheet, packs at the back of the interface between domains.

The interface between domains 1 and 2 buries 684 Å² of surface area with the bulk of the interactions made between helix α3 and helix α4. The relative position of the two helices seems to be oriented by the side chain of the highly conserved Glu 79 in helix α3. This interacts with the beginning of helix α4 via an electrostatic interaction with the helix dipole and by forming two hydrogen bonds to the backbone amides of His 133 and Leu 134. The residues Leu 82, Tyr 131, Leu 134 and Tyr 189 form the majority of the hydrophobic interactions at the interface, with hydrogen bonds formed by the side chains of both Arg 137 and Tyr 189 to Glu 78, and between the side chains of Tyr 131 and Asn 75. None of these residues is particularly conserved illustrating the high sequence variability of this enzyme, for example, in the first 150 amino acid residues there are only five invariant residues (4 glycines and Asp 57) (Figure 6-4). The only region of the structure where the sequence is conserved is in domain 2, residues 153-182, which contain the PROSITE ATP-PRT signature [Falquet *et al.*, 2002]. The core of this motif, residues 164-176, is similar to the 13 residue PRPP binding motif as found in type I PRTs (Figure 6-4 and Figure 6-10).

The third or C-terminal domain (residues 220-299) adopts a ferredoxin-like fold. This domain is linked to the first domain by the 22 residue helix α7 (residues 204-225) which also serves as the last helix of domain 1. The C-terminal domain contains a 4 strand β-sheet on one side flanked by two α-helices, forming a double βαβ motif, connected by an eight residue loop (residues 255-263) the majority of which is not resolved in the electron density map. By visual inspection of the 39 super families with this fold in the SCOP database [Lo Conte *et al.*, 2000], the C-terminal domain seems to be most similar to family 5, the P_{II} signal transduction proteins and family 20, the 'ACT domain' family¹⁰. This is consistent with its role in the regulation of the enzyme by histidine and tRNA binding [Hartman *et al.*, 1971; Smith *et al.*, 1974; Sterboul *et al.*, 1977].

¹⁰ 'ACT domain' stands for aspartokinase, chorismate mutase, TyrA domain

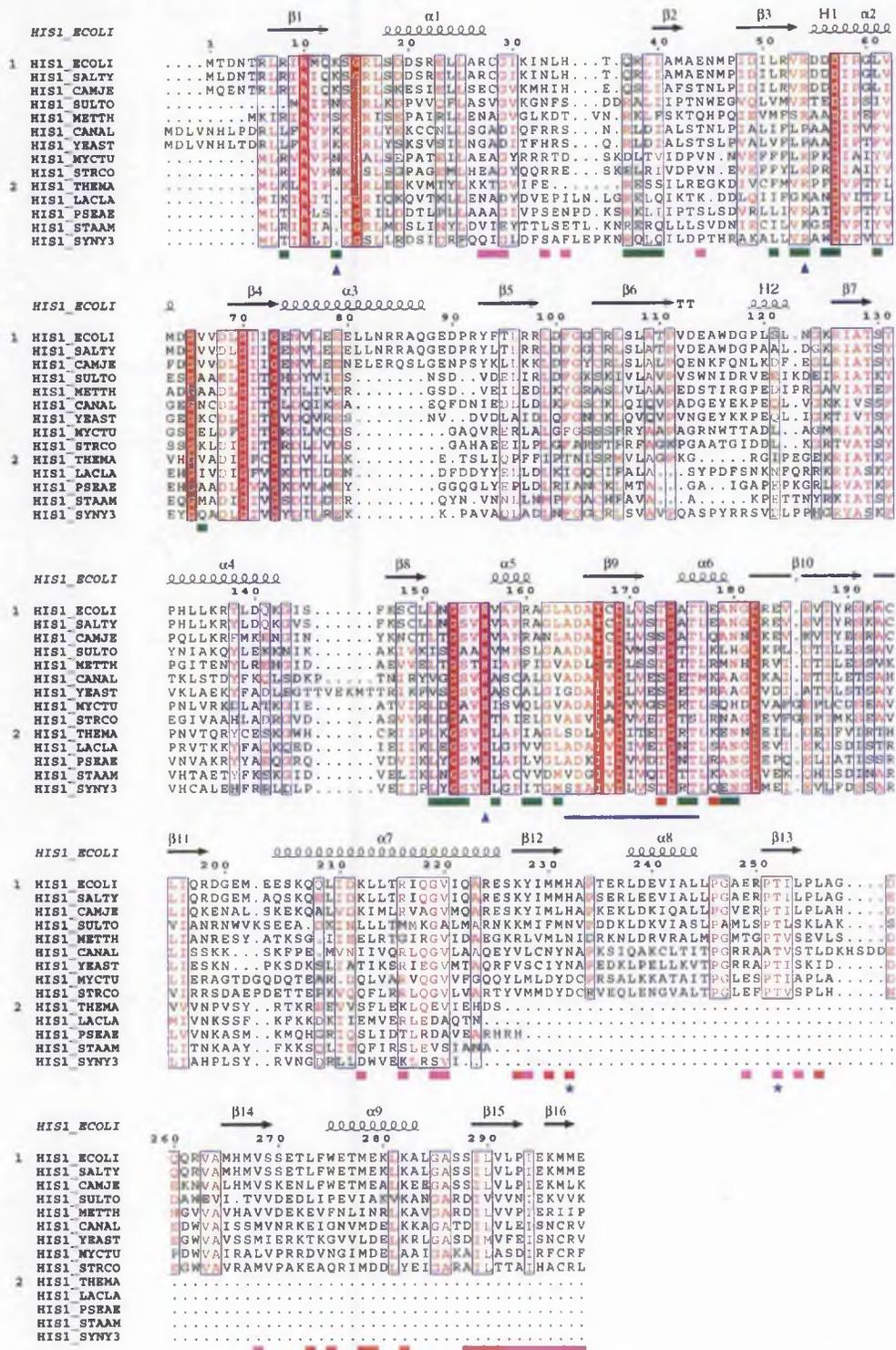


Figure 6-4 Sequence alignment of representative HisGs. A sequence alignment of representative long (1) and short (2) forms of ATP-PRT from archaea, bacteria and yeast together with the secondary structure of the *E. coli* ATP-PRT, produced using ESPrnt (51). α -helices and β -strands are represented as helices and arrows,

respectively, and β -turns are marked with TT. The residues involved in binding PRPP are indicated with ▲, the two residues shown to be involved in histidine binding with *, and the 13 residues similar to type I PRPP binding motif are marked with a blue line. The residues buried at the dimer interface are marked ■ in green, those buried at the trimer interface are coloured in red (S1 see text) and magenta (S2). This sequence alignment was created using the following sequences (Organism, SwissProt accession numbers or other source in brackets) HIS1_ECOLI (*Escherichia coli*, P10366); HIS1_SALTY (*Salmonella typhimurium*, P00499); HIS1_CAMJE (*Campylobacter jejuni*, Q9PM78); HIS1_SULTO (*Sulfolobus tokodaii*, Q970Z3); HIS1_METTH (*Methanobacterium thermoautotrophicum*, O27550); HIS1_CANAL (*Candida albicans*, P46586); HIS1_YEAST (*Saccharomyces cerevisiae*, P00498); HIS1_MYCTU (*Mycobacterium tuberculosis*, O33256); HIS1_STRCO (*Streptomyces coelicolor*, The Wellcome Trust Sanger Institute, SCO1438); HIS1_THEMA (*Thermotoga maritima*, Q9X0D2); HIS1_LACLA (*Lactococcus lactis (subsp. lactis)*, Q02129); HIS1_PSEAE (*Pseudomonas aeruginosa*, Q9HVV8); HIS1_STAAM (*Staphylococcus aureus*, Q99QW2); HIS1_SYNY3 (*Synechocystis sp. PCC 6803*, Q55503).

6.3.3 The HisG dimer and hexamer

The long ATP-PRTs of *E. coli* and *Salmonella typhimurium* have been shown to exist in an equilibrium between various oligomeric states [Klungsoyr and Kryvi, 1971; Tebar *et al.*, 1973; Parsons and Koshland, 1974]. The equilibrium is affected not only by the concentration of substrates and effector molecules but also by the protein concentration. The structure reported here is of the hexameric form of the enzyme, which is a trimer of dimers that are related to the monomer by crystallographic symmetry.

The dimer interface, which is on the crystallographic 2-fold axis, buries 1192 Å² of surface area per subunit. The interface is formed between the exposed surface of the β -sheet from β 1- β 3 together with the loop and part of helix α 2 from domain 1 interacting with strand β 8 through to helix α 5 of domain 2. The interface is made of predominantly hydrophobic side chain interactions with Leu 39, Ile 40, Leu 51, Val 53, Leu 61 and Val 66 from domain 1 interacting with Leu 151, Val 157, Ala 161, Leu 163, Ala 179 and Ser 154 from domain 2. Two residues make significant hydrogen bonds at either end of the dimer interface. The invariant Asp 57 in domain 1 hydrogen bonds with both the side chain and main chain amide nitrogen of Asn 152 in

domain 2. In addition the side chain of conserved Asn 180 in domain 2 forms hydrogen bonds to the backbone carbonyl of Glu 37 and amide nitrogen of Leu 39.

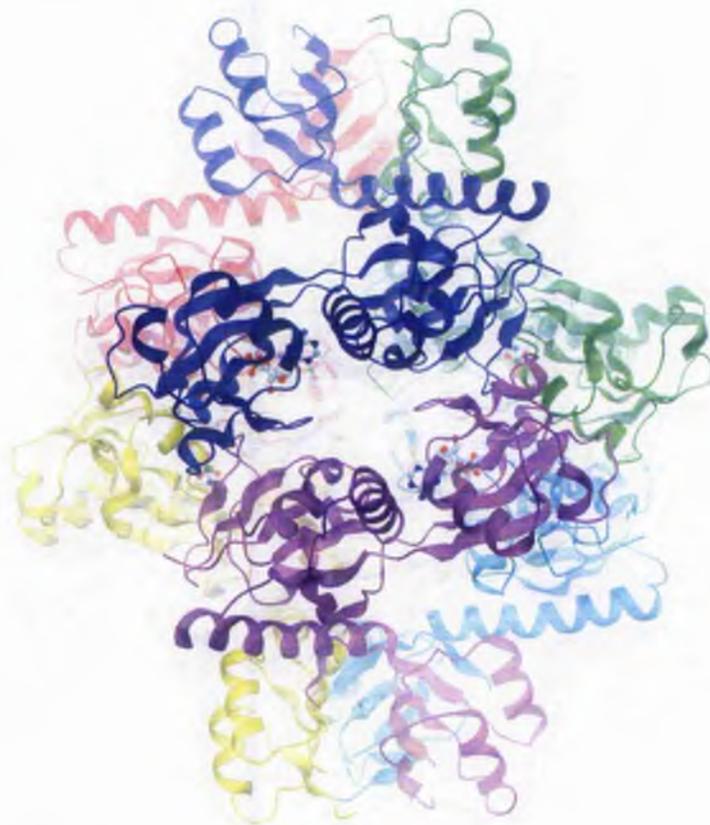


Figure 6-5 HisG hexamer. A ribbon representation of the HisG hexamer viewed down the crystallographic 2-fold axis. Individual chains are coloured uniquely and molecules of the inhibitor AMP are shown in ball and stick representation. The surface is calculated using GRASP [Nicholls *et al.*, 1993] and the figure composed using DINO [Philippsen, 2001].

The hexamer is formed from three dimers by the crystallographic 3-fold symmetry, confirming our previous quaternary structure prediction (Figure 6-5) [Lohkamp *et al.*, 2000]. The protein interactions that form the trimer are almost exclusively restricted to the C-terminal domain (Figure 6-2 and Figure 6-5). Domains 1 and 2 of the dimers contact via residues 33-38 and 173-178 with exclusively hydrophobic interactions, but in general there is a 6 Å separation between the backbone atoms of these residues. A hole of about 15-20 Å diameter is formed between the dimers and the trimers of the C-terminal domain (calculated from a

molecular surface produced using GRASP [Nicholls *et al.*, 1993]). The trimer interface in the C-terminal domain is composed of two dissimilar surfaces (S1 and S2) which bury 760 Å² and 790 Å² of accessible surface area at the interface respectively. The last eight amino acid residues (Figure 6-2) are critical for trimer formation, burying almost 70% of the S1 buried surface area and forming five out of the six interfacial hydrogen bonds. The hydrogen bonds are formed between strand β16 of S1 and β15 of S2, the last strand of the anti-parallel sheet of the C-terminal domain. In addition to this four hydrophobic residues from each monomer (Tyr 228, Met 230 from strand β12 and Leu 292 and Pro 293 from strand β15) pack together to form a core around the three-fold axis of the trimer. There is also the possibility of a salt bridge being formed between Glu 295 from S1 and Lys 227 from S2. None of the side chain interactions seems to be strictly conserved across the long ATP-PRT forms, again highlighting the sequence variability of the enzyme. Overall the buried surface area at the trimer interface is comparable to that at the dimer interface as is the number of hydrogen bonds; however from solution studies it is clear that the hexamers can be more easily dissociated into dimers than into monomers [Klungsoyr and Kryvi, 1971; Tebar *et al.*, 1973].

6.3.4 AMP binding

AMP and ADP are competitive inhibitors of ATP-PRT with respect to both ATP and PRPP [Morton and Parsons, 1977]. Low levels of histidine synergistically enhance this inhibition which is also associated with formation of the hexameric form of the enzyme [Kryvi and Klungsoyr, 1971]. Specifically, histidine lowers the dissociation constant for AMP by 20-fold and ADP by 4-fold [Morton and Parsons, 1977]. In the AMP-bound crystal structure, one molecule is bound per monomer with the phosphoribosyl moiety interacting with domain 2 and the adenine ring positioned in the cleft between domains 1 and 2. The main interactions made by the AMP to the protein are via the phosphate and the ribose groups of the nucleotide (Figure 6-6). The monophosphate binds to a loop following strand β9 and the beginning of helix α6. Hydrogen bonds are made to the AMP phosphate by the backbone amides of Ser 172, Thr 173, Gly 174 and Thr 176, and the hydroxyls of Ser 172, Thr 173 and Thr 154. The dipole of helix α6 also provides additional electrostatic binding energy. The 2' and 3' hydroxyl groups of the ribose of AMP form hydrogen bonds to the side chains

of conserved residues Glu 156 and Asp 169. The only additional interaction is the guanidinium group of Arg 16 from the N-terminal domain, which stacks against the adenine ring of AMP. This orientation of AMP is markedly different from that reported for the His-AMP *M. tuberculosis* enzyme [Cho *et al.*, 2003] (see Figure 8-2) due to the 2' and 3' hydroxyl groups of the ribose being coordinated by a non-conserved aspartate (residues Asp 30 using *M. tuberculosis* numbering) from an adjacent protein subunit.

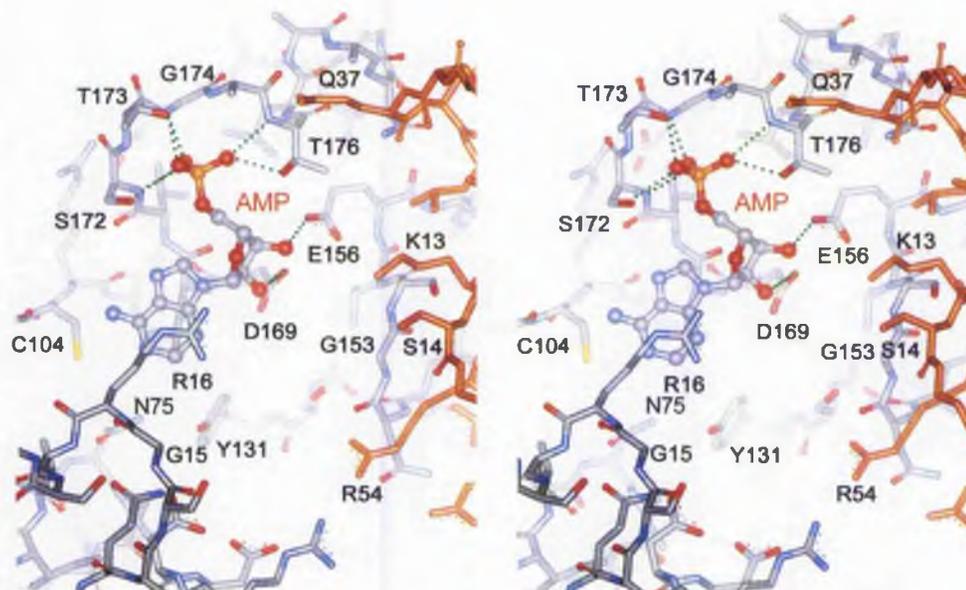


Figure 6-6 AMP binding in HisG. A stereoview of the AMP binding site with residues from the dimer related molecule coloured in orange. Hydrogen bonds formed between AMP and HisG are shown as green dotted lines. Residues D 169 to T 176 are part of a 13 residue PRPP binding motif. R 16 base stacks with the purine ring of bound AMP. This figure was prepared using WebLabviewer.

6.4 HisG with PR-ATP

Crystals of HisG co-crystallised with PR-ATP were essentially isomorphous to the ones in presence of AMP (see Section 5.1.1). Therefore phases did not need to be solved specially.

6.4.1 Refinement

The model of HisG complexed with AMP was stripped of the AMP and used as an input model for rigid body refinement in REFMAC5. This was followed by a round of refinement with REFMAC5 including TLS refinement. The monophosphate, attached ribose and the purine ring of PR-ATP were clearly identified in a Fo-Fc difference map. The ribose and tri-phosphate stemming from the ATP was not seen in the density. Figure 6-7 shows an omit difference map of PR-ATP. This part was built and the complete structure further refined with REFMAC5/TLS [Murshudov, 1996; Winn *et al.*, 2001].

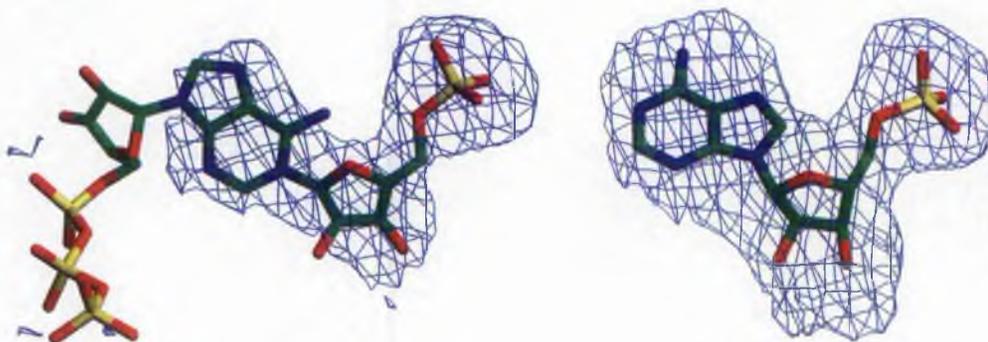


Figure 6-7 Omitmap of PR-ATP and AMP. PR-ATP and AMP shown in stick with a Fo-Fc map contoured at 2σ and 1.75σ respectively. This figure was prepared using SETOR [Evans, 1993].

6.4.2 Quality of model

The model of HisG complexed with PR-ATP had a final $R_{\text{work}} = 20.0\%$ and $R_{\text{free}} = 27.2\%$. The rms deviation in bond distances was 0.018 \AA and for bond angles of 2.031° . The assessed geometrical parameters by PROCHECK [Laskowski *et al.*, 1993] were either satisfactory or better than expected for a structure of this resolution. The Ramachandran plot shows two residues (Ala 248 and Arg 250) just outside the additional allowed region (see Figure 6-8). These residues are part of a loop in the C-terminal domain where the electron density is poorly resolved. However density is for the loop and the Arg side chain is clearly seen. Table 6-4 shows a summary of the refinement statistics.

The model with PR-ATP superimposes on that with AMP with an rms deviation of 0.346 \AA .

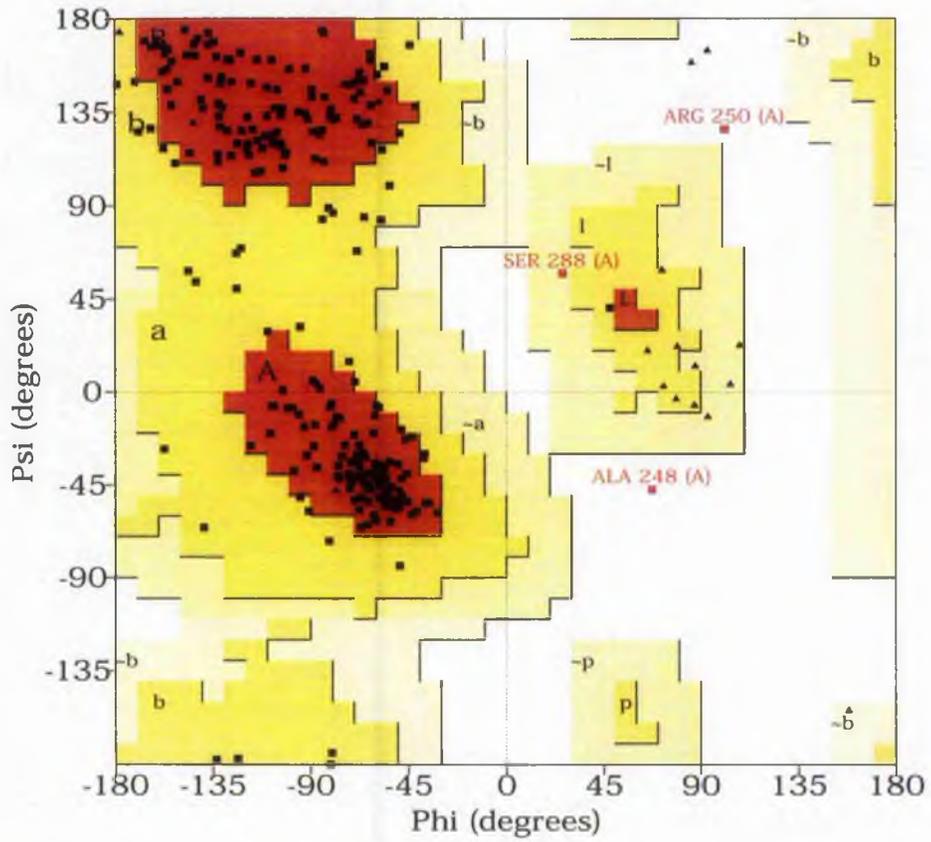


Figure 6-8 Ramachandran plot of *E. coli* HisG complexed with PR-ATP

Table 6-4 Refinement statistics of HisG complexed with PR-ATP

Refinement statistics	
Resolution range (Å)	50-2.9
R _{working} (%) ¹	20.0 (7834)
R _{free} (%) ^{1,2}	27.2 (687)
Number of atoms	
Protein	2193
Nonprotein (PR-ATP, 2 tartrate)	43
R.m.s. deviation from ideal	
Bond length (Å)	0.018
Bond angles (°)	2.031
Ramachandran statistics	
Most favoured	212
Additionally allowed	39
Generously allowed	1
Disallowed	2

¹ number in parenthesis is the number of reflections.

² R_{free} was calculated with 8.1% of reflections.

6.4.3 PR-ATP binding

HisG was co-crystallised with highly purified PR-ATP, produced by biotransformation using the *E. coli* enzyme (by S. Campbell). The structure is isomorphous with the AMP structure and only the phosphoribosyl-adenine (PR-A) part of PR-ATP is seen. PR-ATP is reasonably stable near neutral pH, however the crystallisation buffer was at pH 5.6, outside this range. Although 100 mM MgCl₂ was present in the crystallisation, tartrate was used in a more than 10-fold higher concentration which could chelate the divalent cation. Either of these effects could result in the observed lack of order or degradation of the product in the crystals. The conformation of the PR-A structure is essentially the same as that of AMP-bound protein. However, there is a small (0.5 Å) concerted shift around the monophosphate binding loop (residues 170 - 185) together with a shift in the PR-A structure relative to the AMP molecule (Figure 6-9). The ribose ring of PR-A is shifted almost 1 Å relative to AMP and the difference electron density (Figure 6-7). Due to this shift one

hydrogen bond between Asp 169 and the ribose hydroxyls is lost. This is consistent with the proposal that PR-A is bound to the crystals and not AMP, which was excluded meticulously from all stages in the protein preparation.

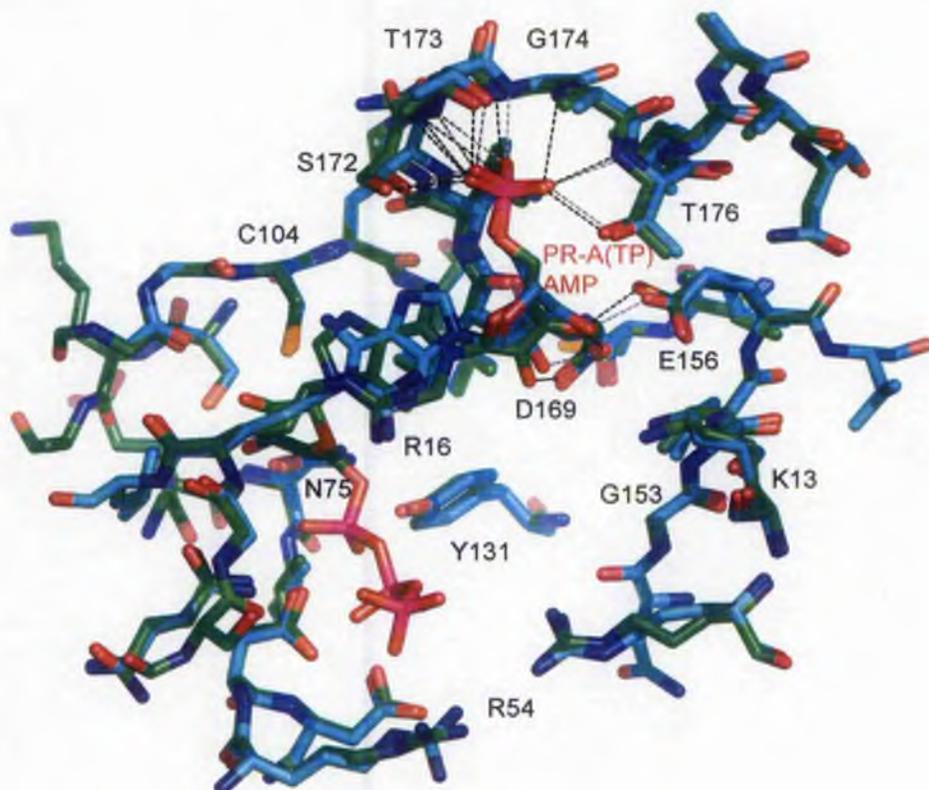


Figure 6-9 PR-A(TP) binding compared with AMP binding in HisG. The structures of HisG with PR-A(TP) and AMP are shown in stick coloured according to the elements. The residues with green carbon atoms belong to the PR-A(TP) bound structure, the ones in cyan coloured carbons to the AMP bound structure. Hydrogen bonds between HisG and PR-A(TP) are shown in black dotted lines, the ones between AMP and HisG in grey dotted lines.

6.5 Discussion

6.5.1 AMP inhibition at the PRPP binding site

The region of domain 2 involved in AMP binding (residues 164-176) is similar to the PRPP binding motif (Figure 6-10). Alignment of this motif with the type I PRPP binding motif reveals the main difference as two consecutive conserved aspartic acid residues (position 5-6 in PRPP motif) versus one in ATP-PRT (position 6, Asp 169). Residue 169 is conserved as either Asp or Glu and is part of the PRPP binding motif, whereas the invariant Glu 156 is outside the motif (Figure 6-10). The 13 residues in ATP-PRT possess a typical backbone conformation as seen for the PRPP binding motif in both type I PRTs and PRPP synthetases. In type I PRTs this loop is involved in binding of PRPP, confirmed by a crystal structure of orotate PRT with PRPP bound [Scapin *et al.*, 1995]. This identifies this loop in ATP-PRT as the potential PRPP binding site. It is therefore surprising that AMP binds to this site rather than to the ATP binding site of the enzyme which had been previously assumed [Morton and Parsons, 1977]. The only other example of a feedback inhibitor binding to the PRPP binding site, rather than an allosteric site, is that of AMP to *B. subtilis* glutamine phosphoribosylpyrophosphate amidotransferase [Smith *et al.*, 1994]. Figure 6-11 shows a comparison of the PRPP binding site in orotate PRTs and ATP-PRT.

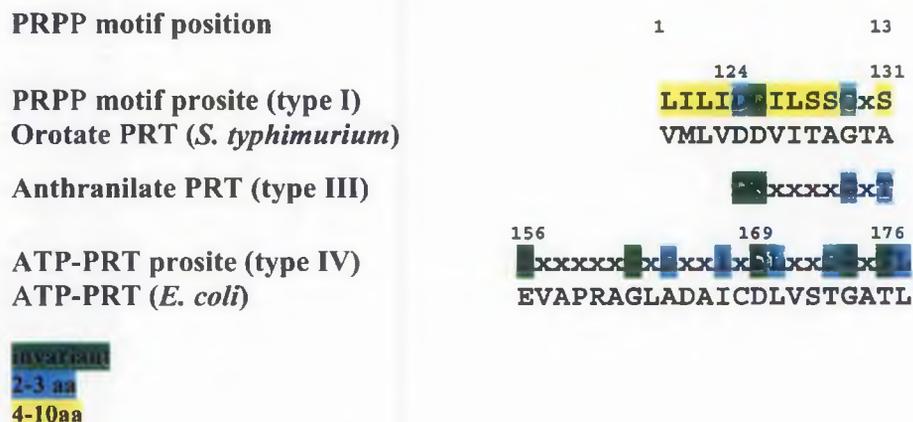


Figure 6-10 Alignment of PRPP binding motif.

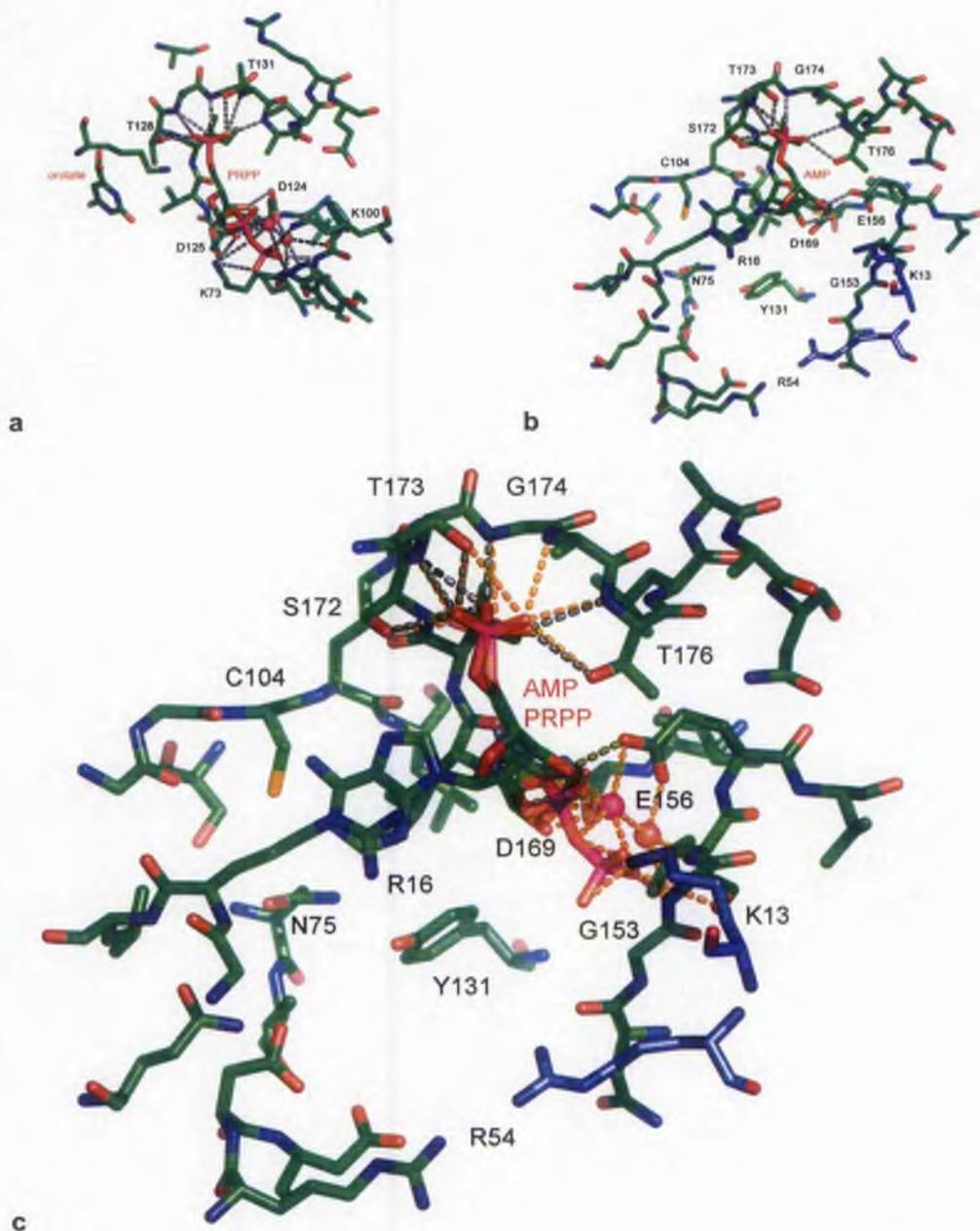


Figure 6-11 Comparison of PRPP binding site. (a) shows the PRPP binding site in orotate PRT from *S. typhimurium* (OPR1) and (b) the AMP binding site in *E. coli* HisG. The structures were superimposed using the monophosphate binding loop and adjacent helix (which includes the 13 residue PRPP binding motif). Residues from a symmetry related monomer (within the dimer) are shown with cyan carbon atoms. (c) shows the PRPP and a coordinated water (red) and Mg^{2+} ion (magenta) from the orotate PRT structure superimposed on the ATP-PRT structure. Hydrogen bonding is shown in grey dotted lines and possible ones for PRPP binding in yellow.

6.5.2 Substrate binding sites

The modified PRPP binding site of ATP-PRT can be compared with those of several representative type I PRTs, namely *Toxoplasma gondii* HG-PRT, *Trypanosoma cruzi* H-PRT and *S. typhimurium* O-PRT (1FSG, 1TC2 and 1OPR respectively, see Figure 6-11 and Figure 6-13). The binding of the PRPP to ATP-PRT requires Mg^{2+} although it has been shown that manganese can substitute functionally for magnesium [Bell and Koshland, 1971]. The structures of type I PRTs complexed with PRPP and Mg^{2+} show that four of the six coordination sites are occupied by PRPP, the 2' and 3' hydroxyl groups of the ribose and one each by the pyrophosphate moiety. The remaining two coordination sites are occupied by water molecules, one of which is bound to the protein. Comparison with these structures suggests that this arrangement will be found in ATP-PRT, the conserved water molecule binding to Glu 156 and the backbone of Val 155. The terminal phosphate of PRPP in type I PRTs is normally coordinated by the protein backbone or by arginine or lysine side chains. In ATP-PRT the amide of Ser 154 may be involved but conserved residues from domain 1 of the dimer related subunit, Lys 13 and Arg 54 are also in positions to play an important part in coordination (see Figure 6-11).

The orientation of bound AMP not only provides information about the PRPP binding but also the position of the adenine ring found in the substrate ATP and product PR-ATP. In bound PR-ATP the purine ring will be orientated in a way that the 5- and 6-membered rings swap positions compared with the bound AMP. This is possible as the only interaction of the ring is the base stacking with Arg 16 (Figure 6-6) and no hydrogen bonding or electrostatic interaction are involved. The ribose of ATP and the product are most likely to bind to conserved residues Cys 104 and Asn 75, while the triphosphate, thought to require Mg^{2+} for binding, may be positioned in the cleft between domains 1 and 2.

6.5.3 Reaction mechanism and inhibition of HisG

Detailed studies on the reaction mechanism of ATP-PRT have been undertaken for both the *E. coli* and *S. typhimurium* enzymes [Kleeman and Parsons, 1976; Morton and Parsons, 1976; Tebar and Ballesteros, 1976; Goitein and Parsons, 1978; Dall-Larsen, 1988]. The reaction proceeds via a steady state ordered Bi-Bi kinetic mechanism with the binding of ATP followed by PRPP, and PPi leaving first followed by PR-ATP (see Figure 1-4) [Morton and Parsons, 1976]. However, the

observation that both AMP and ADP are competitive inhibitors with respect to both substrates has been difficult to reconcile with this mechanism [Morton and Parsons, 1977]. From the crystal structure it is clear that the reasonable assumption that AMP and ADP inhibit by binding to the ATP binding site of the free enzyme is in error. The monophosphate and ribose of AMP are instead bound to the PRPP binding site in domain 2, while the adenine ring binds in a position occupied by ATP in domain 1. In this way AMP can compete with PRPP as well as ATP for binding to the free enzyme.

The binding of PR-ATP in the active site further confirms the binding site of PRPP. Furthermore it confirms the overlapping of the purine ring position of AMP and the product. This validates the assumption that the purine portion of ATP binds there with the ribose-triphosphate occupying the cleft in domain one. However this part could not be resolved in the structure.

AMP interacts with domain 1 through base stacking of the adenine ring with Arg 16, presumably closing the enzyme active site and forming a conformation similar to that when PR-ATP is bound which is confirmed by the structure. In this conformation the enzyme seems to favour aggregation into an inactive hexameric form as has been shown by solution studies [Klungsoyr and Kryvi, 1971]. In this aggregated form, the binding sites for substrates are much less accessible to solvent as they are located on the inner surface of the hexamer (Figure 6-5). Therefore AMP and PR-ATP inhibit the enzyme by both conventional competitive inhibition with respect to the substrates and by forming a hexamer in which the active sites are closed and inaccessible.

6.5.4 HisG is a new class of PRTs: Extending the PRT superfamily

The structures of previous reported PRTs have been classified as type I, II and III on the basis of their folded structure. ATP-PRT has no structural similarity to these enzymes and therefore should be classified as a type IV PRT. A new type of PRT would have been expected, as beyond the PRPP binding motif there is no sequence similarity between ATP-PRT and any other PRTs. However, all structurally characterised PRTs with the PRPP binding motif have been found to possess a type I fold. Although it has been proposed that a PRPP binding motif is exclusive to PRTs in the nucleotide synthesis and salvage pathways [Sinha and Smith, 2001], the histidine biosynthetic enzyme structure reported here has a similar PRPP binding motif but is not homologous to the type I PRT family of enzymes.

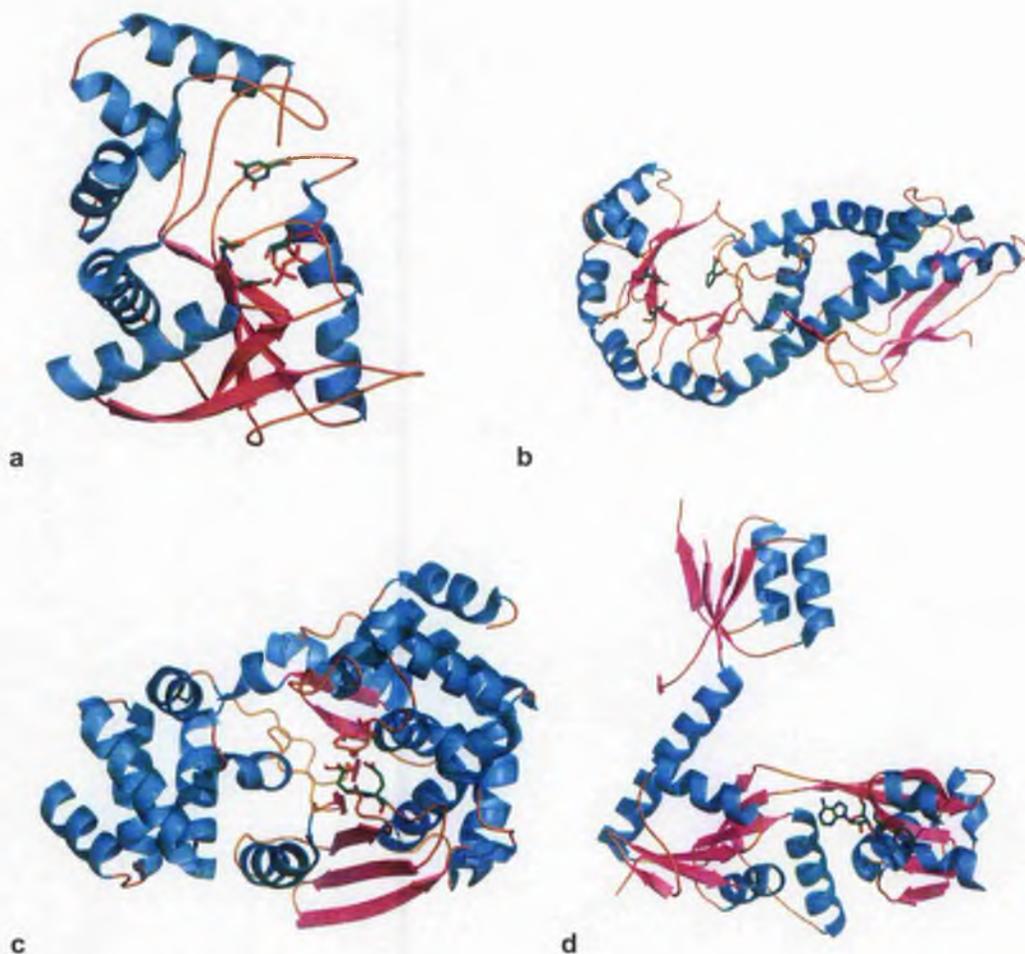


Figure 6-12 Comparison of different types of PRTs. One representative of each type of PRT coloured according to their secondary structure are shown in ribbon representation: type I - orotate-PRT (a), type II - quinolate-PRT (b), type III - anthranilate-PRT (c) and type IV ATP-PRT (d). Where available substrate, product or inhibitor are shown in stick as well as the two conserved acidic residues involved in PRPP binding.

The conserved 13 residue PRPP binding motif is common to type I PRTs (Figure 6-10). This motif is not found in the type II enzymes and is restricted in type III enzymes [Eads *et al.*, 1997]. ATP-PRTs have a modified PRPP binding motif with one of the two conserved acidic residues, which are involved in binding the PRPP ribose, being found outside the motif (Glu 156). The backbone of the PRPP binding loop and helix α_6 of ATP-PRT (residues 170-180) superimpose with a rmsd of between 0.5 Å and 0.8 Å on representative type I PRT structures highlighting the conserved nature of this motif and the common mode of binding of the

monophosphate (Figure 6-13). Although the motif is found in the central strand of the 5-stranded β -sheet of ATP-PRT domain 2, as in type I PRTs, there is no structural similarity beyond this to suggest the domains are related. For example, in the direction of the β -strands, the β -sheets of the domains are inclined at an angle 20° from each other, while perpendicular to this, the planes of the two β -sheet are at an angle of up to 60° away from each other. These large structural differences affect the position of the two acidic residues associated with the PRPP binding motif (Asp 124 and Asp 125 in *S. typhimurium* O-PRT; Glu 146 and Asp 147 in *Toxoplasma gondii* HG-PRT). In type I PRTs, these acidic side chains acids are adjacent and located at the apex of the central β -strand where they are in a position to bind the 2' and 3' hydroxyl group of the ribose respectively. ATP-PRT has only one residue, Asp 169, in an equivalent position which is situated 2.5 \AA away from the two acidic residues of the type I PRTs, however it is still in a position to bind to the 2' hydroxyl of PRPP. The residue adjacent to Asp 169 in ATP-PRT is hydrophobic and buried against helix α_4 ; the binding of the 3' hydroxyl is instead accomplished by Glu 156, a residue at the beginning of helix α_5 and outside the normal PRPP binding motif. The pairs of acidic residues in type I PRTs and ATP-PRT are separated by approximately 5 \AA and therefore are structurally equivalent.

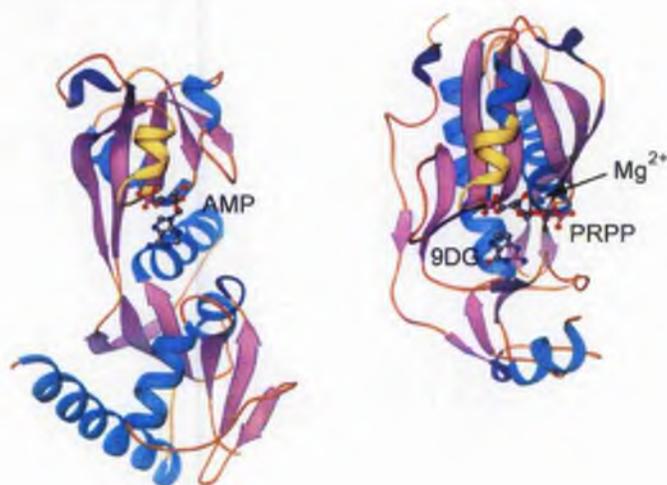


Figure 6-13 Comparison of PRPP binding motif of HisG and type I PRT. Ribbon diagram of a comparison of the 13 amino acid PRPP binding motifs of *E. coli* ATP-PRT and a typical type I PRT, HG-PRT (1FSG). The residues used to superimpose the two structures are highlighted in yellow, the rest of the structure is coloured

according to secondary structure. The ligands are represented in stick and coloured according to atom type. This figure was produced using DINO [Philippson, 2001].

Although the different types of PRTs form distinct folds, they share a number of structural similarities that probably result from catalysing a common chemical reaction. All types consist of at least two domains of which one binds the PRPP and the other is responsible for the substrate-specific recognition. The PRPP binding shows a common feature of two conserved acidic residues which bind the two hydroxyl groups of the PRPP ribose as well as coordinating the divalent cation. In type I and III PRTs these residues are adjacent in sequence whereas in type II and IV they are separated by 21 and 13 residues respectively. In type I, IV and presumably III PRTs, the PRPP monophosphate is bound in a loop with a common backbone conformation while in type II positive side chains are involved. These structural similarities appear to represent common structural solutions for the binding of PRPP. However, the differences in overall fold and the lack of amino acid sequence identity both suggest that the 4 types of PRTs are the result of convergent evolution.

6.5.5 The C-terminal ‘regulatory’ domain

The C-terminal domain of the long ATP-PRTs has been implicated in the regulation of the enzyme by histidine and tRNA binding [Hartman *et al.*, 1971; Smith *et al.*, 1974; Sterboul *et al.*, 1977]. This domain shares no obvious sequence similarity with any other proteins, but has a ferredoxin-like $\beta\alpha\beta\beta\alpha\beta$ topology, which is common to a large number of proteins. Visual inspection of the ferredoxin-like super families in the SCOP database [Lo Conte *et al.*, 2000], revealed the C-terminal domain to be most similar to P_{II} signal transduction proteins and the ‘ACT domain’ family.

The ACT (aspartokinase, chorismate mutase, TyrA domain) domains are found typically at the N- or C-termini of enzymes involved in amino acid and purine biosynthesis either as a single domain or as a tandem repeat [Chipman and Shaanan, 2001]. ACT domains are involved in oligomerisation, usually to form dimers and tetramers, with pairs of the regulatory domains forming an eight-stranded β -sheet with four helices on one side. Small effector molecules seem to bind between regulatory domains and in some cases cause feedback inhibition (e.g. serine in *E. coli* 3-phosphoglycerate dehydrogenase) [Schuller *et al.*, 1995]. This is consistent with the

evidence from mutant studies which indicates that the C-terminal ATP-PRT domain is responsible for feedback inhibition by histidine and the binding to histidyl-tRNA [Hartman *et al.*, 1971; Smith *et al.*, 1974; Sterboul *et al.*, 1977]. However, the overall fold and oligomerisation of the C-terminal domain of ATP-PRT show that it is much more similar to the P_{II} signal transduction proteins than to the ACT domains.

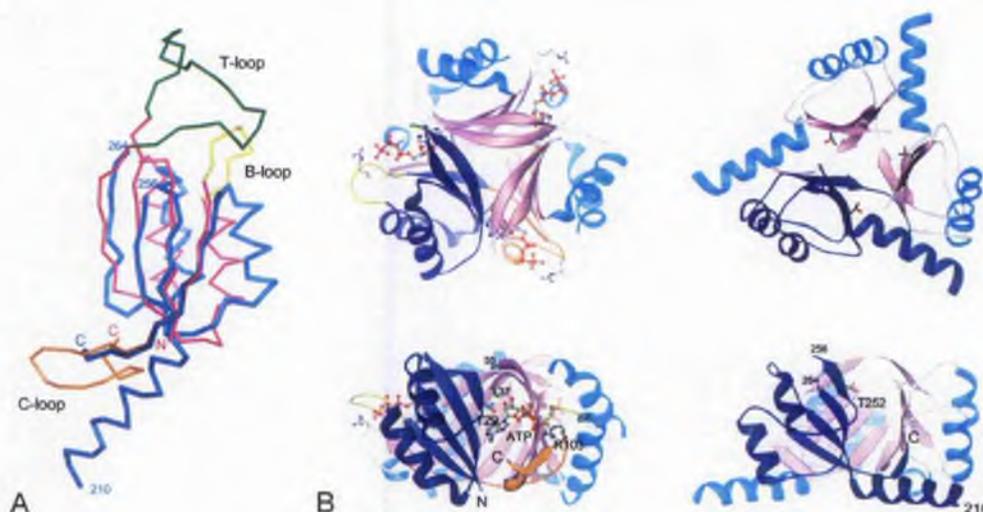


Figure 6-14 Comparison of C-terminal domain of HisG with P_{II} -like protein. (A) C_{α} superposition of the C-terminal domain of ATP-PRT compared with the P_{II} protein GlnK. ATP-PRT is shown in blue, overlaid residues of GlnK in magenta. The B-, C- and T-loop of GlnK are shown in yellow, orange and green, respectively. N- and C-termini are labelled with N and C, respectively. (B) Comparison of the quaternary structure of P_{II} protein GlnK (left) and C-terminal domain of ATP-PRT (right). The top pair of diagrams show a view down the 3-fold axis and the lower pair a side view. One subunit is coloured in blue, the other two according to their secondary structure. In GlnK the B-, C- and T-loops, which are involved in ATP binding, are shown in yellow, orange and green, respectively. Key residues involved in binding ATP in GlnK and ATP are shown in ball and stick representation. The equivalent binding cleft of ATP-PRT is shown with the conserved Thr 252 in ball and stick representation.

The P_{II} proteins form trimers with the 4 β -strands forming a barrel around the 3-fold axis, an arrangement exactly analogous to the two C-terminal domain trimers formed at either end of the ATP-PRT hexamer (Figure 6-5 and Figure 6-14). Functionally, P_{II} proteins have been shown to play a pivotal role in the coordination of carbon and nitrogen metabolism in plants, bacteria and archaea [Ninfa and Atkinson,

2000]. They are known to bind various ligands including 2-ketoglutarate and glutamate and can be uridylylated and phosphorylated. Binding of effectors or covalent modification alters the conformation of P_{II} proteins, which then interact with receptors to induce various secondary effects such as changes in gene expression, protein modification and transport of nitrogen and carbon sources. A superposition of the C-terminal domain of ATP-PRT with the P_{II} protein GlnK [Xu *et al.*, 1998] using main chain atoms from the β -sheet results in a rmsd of 1 Å over 84 atoms (Figure 6-14b). However, conserved regions of sequence found in P_{II} proteins corresponding to three loops are in each case significantly shorter in ATP-PRT. The 'B-loop' in ATP-PRT is a turn, which contains two conserved residues Gly 285 and Ala 286, while the 'C-loop' is truncated by ten residues (Figure 6-14a). The 'T-loop' as in several of the P_{II} protein structures is not completely resolved (residues 256-264 are missing); however it does not contain any conserved tyrosine or serine that could be covalently modified [Arcondeguy *et al.*, 2001], and its size is not consistent with it interacting with other protein receptors [Jiang *et al.*, 1997]. A comparison of the C-terminal domain trimer with that of GlnK complexed with ATP [Xu *et al.*, 1998] reveals that the intersubunit pocket in which ATP binds is conserved. In particular Thr 29 (GlnK numbering), which makes two important hydrogen bonds to the adenine and ribose of ATP in GlnK, is highly conserved in ATP-PRT C-terminal domains (Thr 252, Figure 6-4). In contrast residues involved in binding the tri-phosphate moiety, two glycines from the 'B-loop' and Arg 103 from the 'C-loop' of GlnK are absent in ATP-PRT, suggesting that the domains do not bind ATP.

Chapter 7: Structure Solution of HisG from *Mth*

7.1 Solving the phases of *Mth* HisG

7.1.1 Molecular replacement (MR) using *E. coli* HisG as a search model

Initially molecular replacement was attempted to solve the structure of HisG from *Mth* using the structure of the *E. coli* enzyme as a search model. A variety of search models were tried, using combinations of the first two domains, the monomer, a complete dimer, with no changes to the residues, systematically cutting residues to alanine through to a complete poly-alanine model. In addition various MR programs (AMoRe [Navaza, 1994], BEAST [Read, 2001], MOLREP [Vagin and Teplyakov, 1997] and EPMR [Kissinger *et al.*, 1999]) had been utilised and various resolution ranges of data had been assessed. Several reasons can be proposed to explain the difficulties in obtaining a correct MR solution using the *E. coli* HisG as a search model. One is the low sequence identity (38%) between the search model and the structure to be solved. Another is the coordinate error in the search model due to the low resolution and therefore high coordinate error of the *E. coli* search model. Furthermore, a conformational change between the histidine and AMP-bound structure cannot be excluded, making the search model inadequate. Finally, the search model is elongated in shape and symmetry equivalent molecules are closely associated within the hexamer. This significantly blurs the boundary between inter- and intra-molecular vectors necessary for conventional MR. Addition of (weak) phase information from derivatives was available (data not shown) and a phased MR search tried. However, despite the best efforts and use of considerable computing time, no solution was obtained.

The availability of the higher resolution HisG structure of *Mtb* initiated a new attempt to solve the *Mth* structure which was successful and is described in the following.¹¹

¹¹ At least one good thing...

7.1.2 MR solution with *Mtb* HisG as search model

A 1.8 Å resolution *Mtb* HisG structure complexed with AMP and histidine (pdb entry 1NH8) was deployed as a search model in AMoRe. Using the monomer with all residues or as a poly-alanine model gave no clear solutions (data not shown). Because there are two monomers in the asymmetric unit (asu) for the *Mth* enzyme a dimer of 1NH8 was tried as a search model. Initially the dimer was built by superposition with the *E. coli* dimer which gave no solution. Then a dimer was built using the crystallographic symmetry of the *Mtb* enzyme. Here again the full model gave no distinct solutions over various resolution ranges.

A poly-alanine model of *Mtb* HisG dimer built in this way as a search model in AMoRe gave correct solutions. Various resolution ranges were probed and solutions for ranges 10-4 Å and 17.5-4 Å are shown in Figure 7-1. Though no single distinct peak could be seen in the rotation function, the correct solutions became clear after the translational search and even clearer after rigid body fitting in AMoRe. More than one solution is found with all search models corresponding to symmetry equivalent molecules. The solutions were visualised and inspected for packing within the unit cell using the molecular graphics program PyMOL and special attention was paid to any clashes with symmetry related molecules.

A modified model was built by editing the search model so that conserved or similar residues were kept but significantly different residues were mutated to Ala. If there were one to two residues missing in the alignment then the preceding and following residues were deleted from the search model. In case more residues were missing two residues were cut from the search model. The results of MR with this modified model are shown in Figure 7-1. The results here are not as clear as for the poly-alanine models and two incorrect solutions are in the top 6 results. However this model was used for further refinement and model building as it resembles the *Mth* enzyme the most.

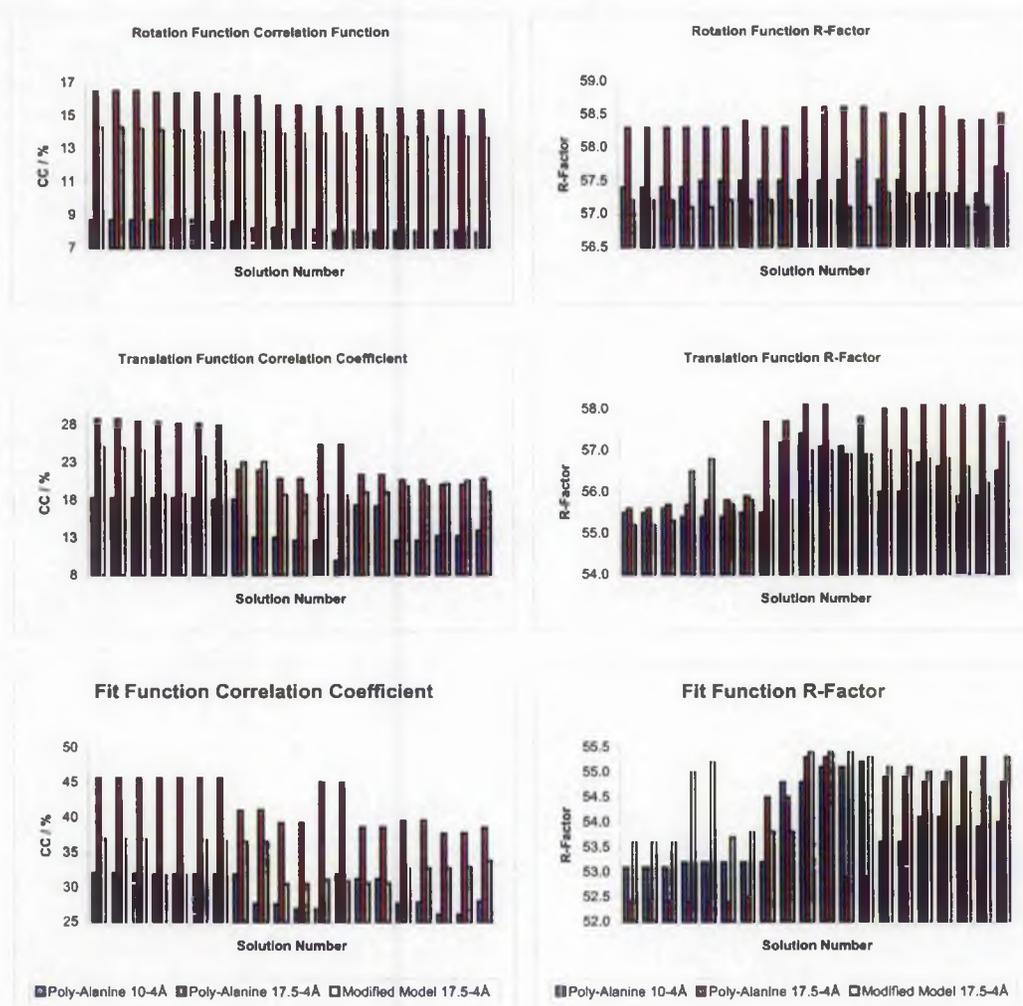


Figure 7-1 AMoRe results for *Mth* HisG

7.1.3 Model building and refinement

The modified HisG model from *Mtb*, rotated and translated according to the AMoRe solution file, was used as an input for rigid body refinement with REFMAC5 [Murshudov, 1996]. The two chains of the dimer were split in 3 individual domains as rigid body groups. After 20 cycles of refinement in the resolution range 30-4 Å the R-factors dropped by about 8% (see Table 7-1). Before model building the model was further restrained refined with REFMAC5 (40-3.5 Å). To further improve these initial phases and remove model bias, solvent flattening with histogram matching was carried out using DM [Cowtan, 1994] and an initial map was calculated.

Table 7-1 R- and free R-factor during refinement of *Mth* HisG

Refinement step	R-factor /%	free R-factor /%
Start (AMoRe model)	54.9	57.4
Rigid body (REFMAC5)	46.4	49.1
REFMAC5 (initial)	38.2	46.3
Simulated annealing (CNS)	33.3	38.1
B group refinement	30.4	35.5
Refmac5 with TLS	21.0	28.5

Iterative cycles of model building with QUANTA and refinement were carried out. Refinement for the incomplete model was accomplished with a cycle of simulated annealing, followed by grouped B-factor refinement, in CNS [Brunger *et al.*, 1998]. In early refinement stages density for histidine was found between C-terminal domains within the hexamer (see Figure 7-2). Once the model was nearly complete, histidine was modelled and refinement was carried out with REFMAC5 using TLS parameters and NCS restrains. A total of 6 TLS groups were used for the 3 domains in the 2 chains.



Figure 7-2 Electron density for histidine in *Mth* HisG. 2Fo-Fc electron density in early stages of the refinement (no histidine included) is shown at 1 σ contour level. Green and cyan carbon atom label symmetry related monomers. The position of the histidine in the final model is shown in grey.

7.2 *Mth* HisG model

7.2.1 Quality of the model

The final model of *Mth* HisG had a final R- and free R-factor of 21.0% and 28.5% respectively. The rms deviation of the bond lengths was 0.015 Å and of the bond angles 1.534°. Both chains have no residues outside the allowed region of the Ramachandran plot. The geometric parameters for the side and main chains are satisfactory or better than expected for this resolution as indicated by a PROCHECK analysis. A summary of the refinement statistics is shown in Table 7-2.

Table 7-2 Refinement statistics of *Mth* HisG

Refinement statistics	
Resolution range (Å)	47.7-2.65
R _{working} (%) ¹	21.0 (13948)
R _{free} (%) ^{1,2}	28.5 (1373)
Number of atoms	
Protein	4352
Nonprotein (Histidine, MPD, Tris, imidazole)	56
Water	99
R.m.s. deviation from ideal	
Bond length (Å)	0.015
Bond angles (°)	1.53
Ramachandran statistics	
Most favoured	447
Additionally allowed	52
Generously allowed	1
Disallowed	1

¹ number in parenthesis is the number of reflections

² R_{free} was calculated with 8.2% of reflections.

7.2.2 Overall structure

The structure of HisG from *Mth* is shown in Figure 7-3. The overall fold is the same as observed for the *E. coli* enzyme. A more detailed comparison with other

HisG structures is described in Section 8.4. The *Mth* enzyme was crystallised in the presence of histidine which stabilises the inactive, hexameric form. The two monomers in the asu form the dimer and the hexamer is formed via the 3-fold crystallographic symmetry.

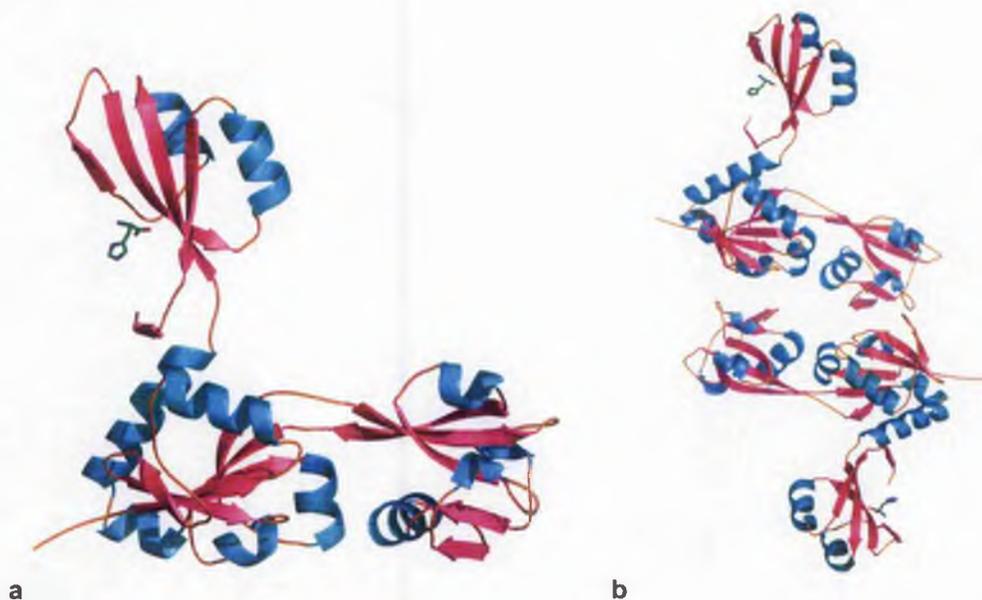


Figure 7-3 *Mth* HisG monomer and dimer. The monomer (chain A) is shown (a) and the dimer (b) in ribbons representation. The bound histidine is shown in stick representation.

7.2.3 The two monomers

The two monomers overlay with an rms deviation of 0.42 Å over all C_{α} atoms¹² and with a maximum deviation of 2.0 Å. Figure 7-4 shows the rms deviation over the whole peptide chain. A very marked deviation, including the maximum, occurs around residue 34 which is the loop before β -strand 2 (see Figure 7-5). The first 80 residues are otherwise virtually identical. For the remainder of the chain there is a deviation of around 0.5 Å throughout. From inspection of the C_{α} trace of the two superimposed chains (Figure 7-5) it is clear that this difference is caused by a tilt of domain 2 and 3 with respect to domain 1. In chain B the domains are further apart, resulting in a more open L-shape than in chain A.

¹² the coordinate error of the structure is 0.30 Å

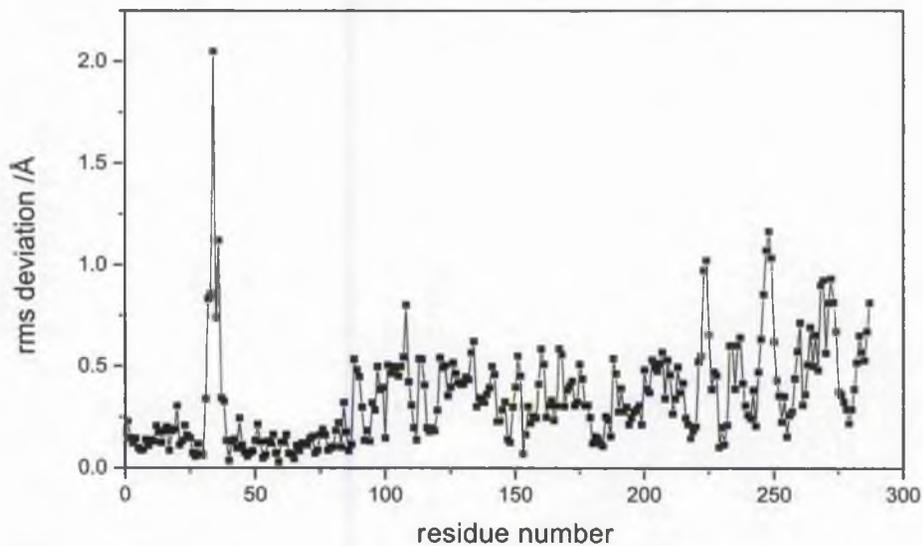


Figure 7-4 Rms deviation between chain A and B of *Mth* HisG.

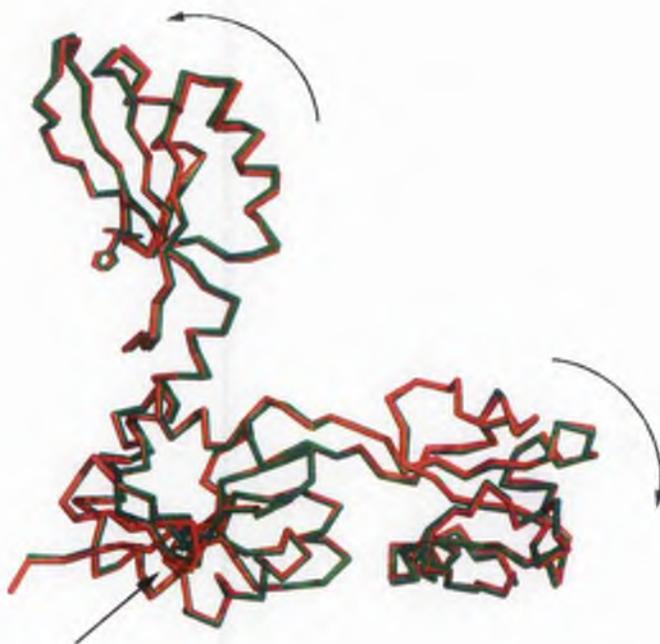


Figure 7-5 Overlay of chain A and B of *Mth* HisG. C_{α} trace of *Mth* HisG chain A is shown in red and overlaid chain B in green. The straight arrow shows the region of maximum deviation. The curved arrows show movement of the second and third domain in chain B compared to chain A and a nearly identical first domain.

7.2.4 Histidine binding

One molecule of histidine is bound per subunit of HisG. The ligand binds between two adjacent C-terminal domains within the trimer of the hexamer (see Figure 7-6). Figure 7-7 shows the interactions involved in the histidine binding to *Mth* HisG, with the distances between protein atoms and those of histidine shown in Table 7-3. Most hydrogen bonds to the histidine originate from one subunit with predominantly main chain atoms being involved (Met 236, Gly 238 & Val 256) and the side chain of conserved residue Thr 240. Additionally two residues (Asn 220' & Asp 276') of a neighbouring subunit are involved in histidine binding. Besides the main chain interactions there are two side chains (Asn 220' & Thr 240) involved in binding. One histidine side chain nitrogen (ND1) forms a hydrogen bond to a water molecule. A comparison of the histidine binding between the two chains A and B in *Mth* HisG shows that most interactions are conserved (Table 7-3). However the main difference is in the coordination of OXT which is either hydrogen bonding to the side chain nitrogen of Asn 220 (symmetry neighbour; chain A) or to the main chain nitrogen of Thr 237 (chain B). This means that the interactions of histidine with the neighbouring subunit can be reduced to two essential ones (side chain Asn 220 and main chain Asp 276).

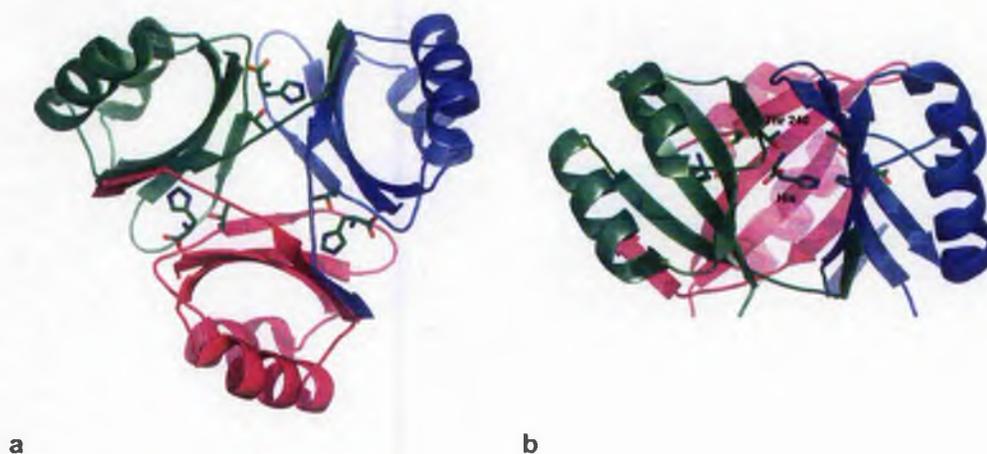


Figure 7-6 Trimer of C-terminal domain of *Mth* HisG with bound histidine. The C-terminal domains, in ribbons representation, from the three different chains are shown in green, blue and magenta. (a) shows the view from the centre of the hexamer, (b) a side view. Bound histidine is shown in stick representation and coloured by element. In addition, the conserved residue Thr 240 is shown in sticks.

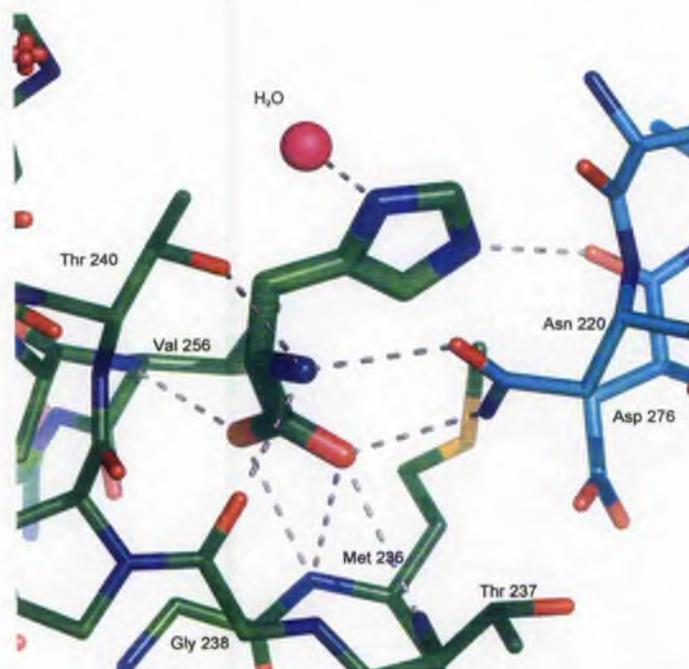


Figure 7-7 Histidine binding in *Mth* HisG (chain A). Histidine is shown in stick (thick) representation and carbon atoms in green. Two HisG subunits are shown in stick (thin) representation, the symmetry equivalent one with carbon atoms in cyan. Hydrogen bonds between histidine and the protein are shown in grey dotted lines.

Table 7-3 Protein-histidine interactions in *Mth* HisG (chain A). In parentheses are shown the numbers for chain B; italics denote symmetry related residue of the trimer within the hexamer; values in grey denote distances too long for hydrogen bonding.

histidine atom	protein residue	atom	distance / Å
ND1	H ₂ O		2.34 (2.80)
NE2	<i>Asp 276</i>	O	2.83 (2.87)
N	<i>Asn 220</i>	OD1	2.90 (3.17)
	Gly 238	O	2.76 (2.91)
	Thr 240	OG1	2.85 (2.73)
O	Met 236	N	3.05 (3.12)
	Val 256	N	3.08 (2.82)
OXT	<i>Asn 220</i>	ND2	2.95 (3.51)
	Met 236	N	3.05 (3.05)
	Thr 237	N	3.26 (3.00)

7.3 Discussion

7.3.1 Histidine binding

A comparison of the histidine binding of HisG in *Mth* and *Mtb* is shown in Figure 7-8. Overall the mode of histidine binding is identical in the enzyme from these two species. This is not surprising because most hydrogen bond interactions are by main chain atoms. Hydrogen bonds involving the two side chain residues (Asn 220 & Thr 240) are conserved as well. The side chain interactions of Thr 240 (in *Mth*), is highly conserved throughout the (long) HisG family and is found at a conserved position within the P_{II}-like C-terminal domain (see Section 6.5.5 and Figure 6-14). The residue at position 220 (here Asn) from the neighbouring subunit is conserved in size but not in charge as either His or Asp/Asn. The latter, with one hydrogen donor and one acceptor, are ideal for hydrogen bonding to the amino acid portion of histidine. In other species, like *E. coli*, where this residue is a histidine, only one hydrogen bond is possible but this does not seem to compromise the specificity¹³. Overall this shows that histidine binding in HisG relies on two side chain interactions with a conserved Thr and one other residue as a possible hydrogen bonding partner.

The histidine model in the *Mth* HisG structure (pdb code 1NH8) is incomplete with the oxygen (OXT) of the carboxylic acid being omitted. It is unclear if the atom is absent by mistake or whether it is not resolved in the electron density. No comparison of this part of the histidine is therefore attempted.

¹³ A crystal structure of *E. coli* HisG with histidine could prove this proposal. Unfortunately crystals grown in presence of histidine and AMP did not diffract beyond 3 Å and the collected data set was very difficult to scale and no solution has been obtained to date.

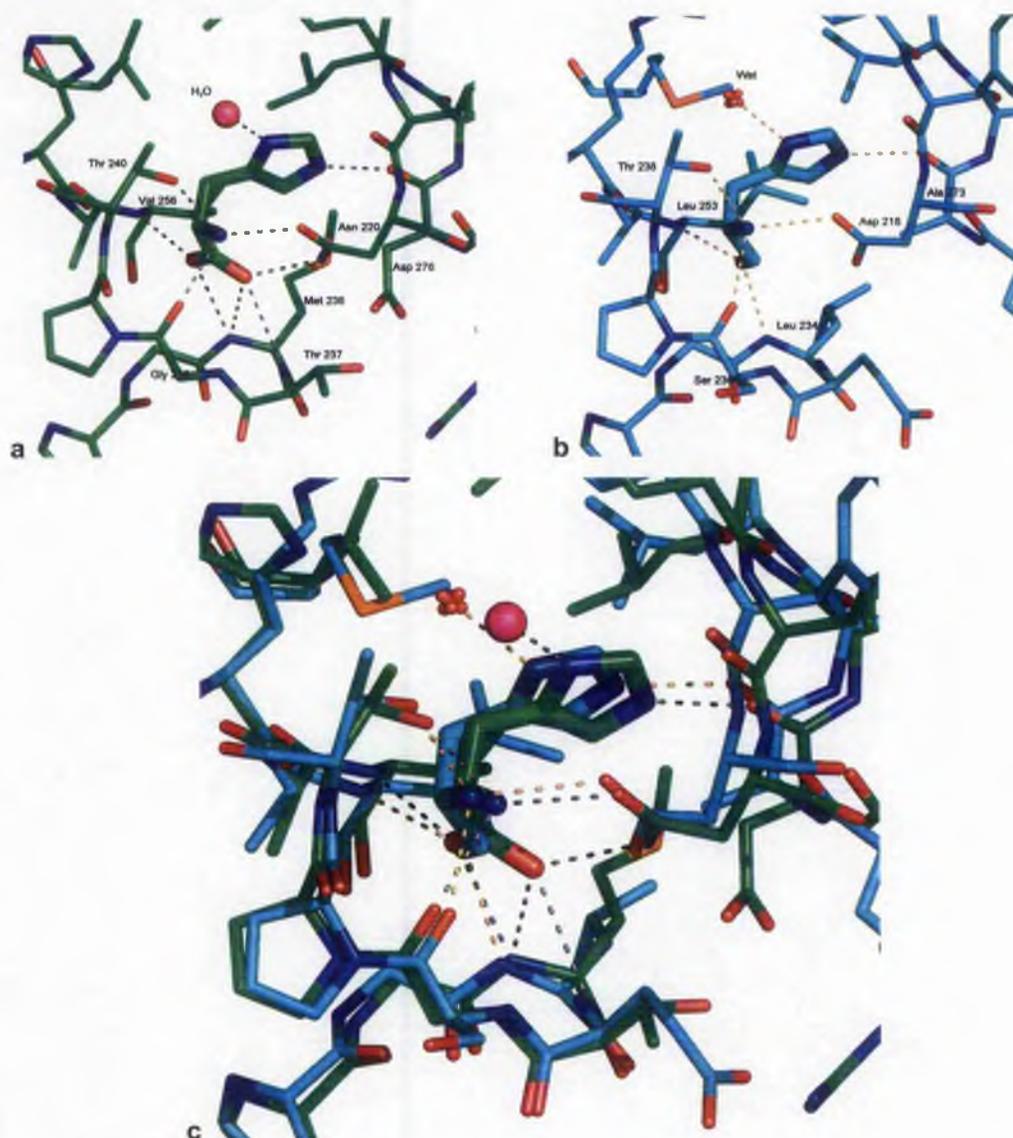


Figure 7-8 Comparison of histidine binding between HisG from *Mth* and *Mtb*. The top panels show histidine binding of *Mth* with carbon in green (a) and *Mtb* with carbon in cyan (b). Residues involved in histidine binding are labelled. The bottom panel (c) shows the two structures superimposed. Hydrogen bonds are shown in grey dotted lines (dark for *Mth*, light for *Mtb*).

Bound histidine is completely buried between the neighbouring subunits of HisG within the trimer. In comparison the histidine binding site in the *E. coli* enzyme is open and accessible (see Figure 7-9). Upon histidine binding the pocket closes with the conserved Thr residue moving towards the histidine and the ‘T-loop’ closing the binding pocket (which is not resolved in the *E. coli* structure). The C-terminal domain

in the *E. coli* structure is therefore in a favourable orientation to bind histidine which is consistent with the observation that AMP and histidine bind and inhibit HisG synergistically.

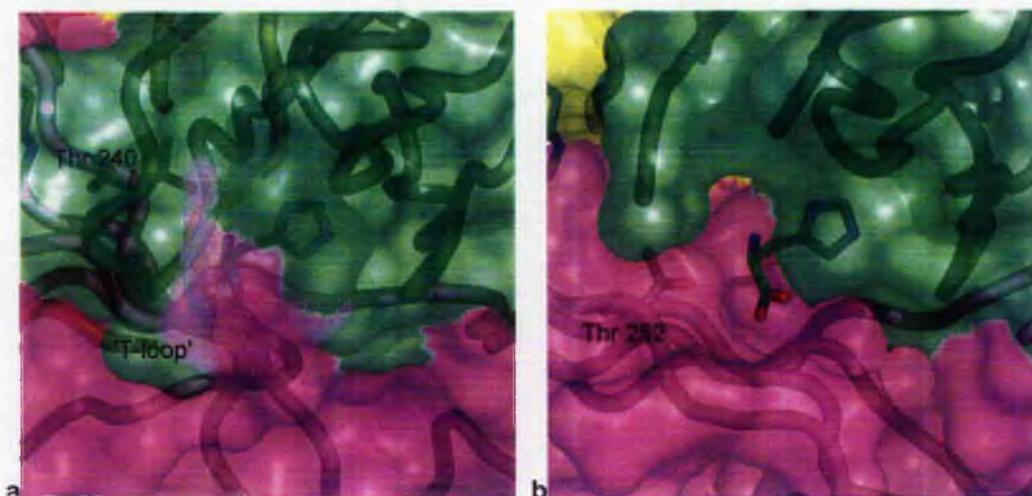


Figure 7-9 Histidine binding pocket. The buried histidine in the *Mth* structure is shown in the left panel (a), the open pocket in the *E. coli* structure in the right panel (b; histidine is shown in stick, taken from the *Mth* structure). The histidine binding portion of one subunit (green) of both structures is aligned. The conserved Thr residue from the neighbouring subunit (magenta) is shown in stick representation.

7.3.2 Quaternary structure

Conformational changes and the effect on the quaternary structure will be discussed in Section 8.4.

Chapter 8: Comparison of HisGs

8.1 Significance of the disulfide bridge in *Mtb* HisG

The structure of apo *Mtb* HisG revealed a disulfide bridge between Cys 73 from domain 1 and Cys 175 from domain 2, which was confirmed by cysteine modification experiments. This disulfide bond was not observed in the structure obtained in the presence of histidine and AMP. Cysteine modification experiments in the presence of inhibitors showed less accessible cysteines possibly due to hexamer formation. A potential role for this bond in regulation was proposed [Cho *et al.*, 2003]. This hypothesis is investigated in more detail in terms of structure and sequence alignments.

Sequence alignment shows that Cys 73 is unique to the *Mtb* enzyme (see Figure 6-4) and no cysteine residue is found in the surrounding area of position 73 (position 78 in *E. coli*). Cys 175 (position 189 in *E. coli*) is conserved among the *Mycobacterium* spp. and found in a small number of other organisms. Again, no cysteine can be found in the vicinity of position 175 to function as a substitute in the other organisms. This suggests that if there is a functional role for this disulfide bond it is an *Mtb* specific feature.

Investigation of the HisG structures of *E. coli* and *Mth* showed that no structurally equivalent cysteine residues could form a disulfide bond as observed in *Mtb*. A comparison of residues in the position of the disulfide bridge, by superposition of apo *Mtb* HisG with other HisGs, highlights the variety of interactions made between the corresponding residues (Figure 8-1). In the *E. coli* enzyme a hydrogen bond is formed between Glu 78 and Tyr 189 while solely Van der Waals interactions are observed in the *Mth* HisG between Val 75 and Leu 179. This shows that the two regions of the structure are in close contact although the nature of the interactions is not conserved. This does not support the hypothesis of a conserved regulatory function proposed for the disulfide bond.

Nevertheless the disulfide bridge observed in the apo *Mtb* HisG has a significant impact on the structure. The distance between the C_α atoms is the shortest in this structure at 7.2 Å, compared with 8.6 Å in the *Mtb* histidine+AMP ternary structure and the binary structures of AMP *E. coli* (10.7 Å) and histidine *Mth* (9.5 Å).

The formation of the disulfide bond moves strand $\beta 10$ and helix $\alpha 3$ closer together. Furthermore it stabilises the relative position of domain 1 to domain 2 beyond the more flexible interactions formed by the equivalent residues in the other species (also see Section 6.3.3).

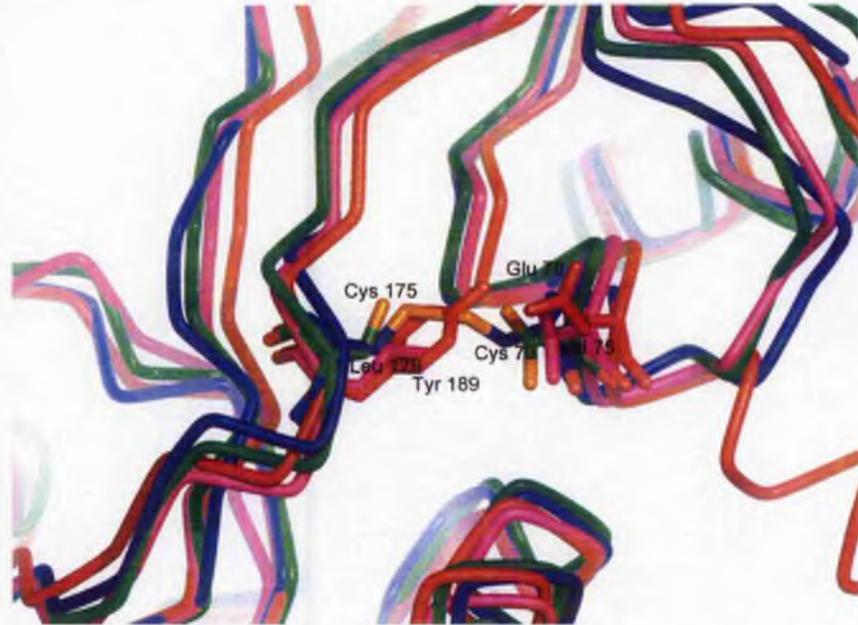


Figure 8-1 Possible disulfide bond in HisG. Backbone representation in the vicinity of the disulfide bond, shown in stick, as observed in apo *Mtb* HisG (blue). Superposed are other HisG structures with equivalent residues shown in stick: green – his+AMP *Mtb*; red: AMP *E. coli*; magenta: his *Mth*.

These observations show that a disulfide bond within HisG is not a conserved feature of the enzyme and therefore is unlikely to be involved in regulation. The disulfide bridge is more likely to be an artefact in the *Mtb* HisG formed by oxidation of cysteine residues which are structurally close and favourable positioned to form the bond. Formation of the disulfide bridge is likely to be physiologically irrelevant and most likely not to occur at all especially under reducing conditions. As a result the apo *Mtb* HisG structure has to be treated with caution, as structural differences may be the result of the formation of the disulfide bond and not due to the absence of inhibitors.

8.2 AMP binding in HisG

The binding of AMP in the *E. coli* enzyme, as observed in this study, differs significantly from the binding seen in the *Mtb* structure (see Section 6.3.4, Figure 8-2

and [Cho *et al.*, 2003]). Furthermore, based on the *Mtb* structure, it was proposed that AMP binds in the ATP binding site rather than the PRPP binding site as shown in this study.

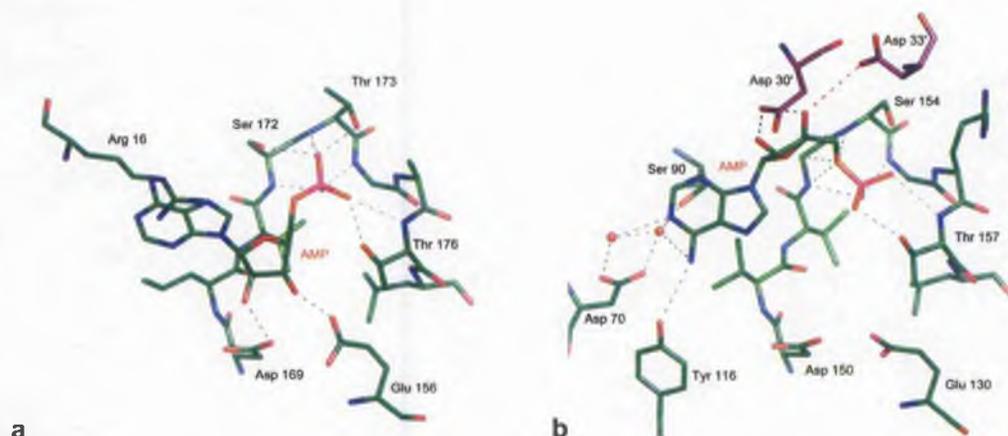


Figure 8-2 AMP binding in *E. coli* (a) and *Mtb* HisG (b). The structures are shown in the same orientation, superimposed on the PRPP binding motif. Hydrogen bonds are shown in grey dotted lines (*Mtb* - in red 3.7 Å long). Residues from a neighbouring subunit (trimer) are shown in purple.

A comparison of the AMP binding of the *E. coli* and *Mtb* (AMP+histidine) HisG enzyme shows that the binding of the monophosphate and the position of the purine ring of AMP are similar in both structures, however the orientation and binding of the ribose differ significantly (Figure 8-2).

The monophosphate binding differs slightly because in *Mtb* structure additionally O5 from the ribose is hydrogen bonding to the loop of the PRPP binding motif, rather than the phosphate oxygens only. However the residues involved in the coordination are the same in both enzymes. In the *E. coli* enzyme the ribose interacts with residues from the extended PRPP motif of the same subunit (see Section 6.3.4 and Figure 8-2 a) whereas in the *Mtb* structure the ribose hydroxyl groups hydrogen bond to side chains of a neighbouring subunit (Asp 30' and to a lesser extent Asp 33'¹⁴; see Figure 8-2 b). The two Asp residues are part of the 'regulatory' loop (see Section 8.3) which is very variable in sequence in HisG and neither residue involved in the binding is conserved beyond the *Mycobacterium* spp.. It would seem unlikely that residues from a neighbouring subunit of the trimer (not the dimer) are

¹⁴ Residue numbers from neighbouring, symmetry equivalent subunits are marked with '.

involved in specific binding of one of the substrates, ATP, especially if it is assumed that the active form is a dimer. The purine ring of AMP in the *Mtb* structure is in a comparable position but is flipped 180° and tilted by about 30° along its long axis of the plane compared to the one in the *E. coli* enzyme (see Figure 8-3 a, b). In the *E. coli* HisG the ring base stacks with Arg 16 whereas in *Mtb* the ring nitrogens form hydrogen bonds (see Figure 8-2). However the bonds to the conserved residue Asp 70 (Asn 75 in *E. coli*) are water-mediated and the direct bond of N6 is with a non-conserved residue Tyr 116 (Tyr 131 in *E. coli*). Arg 16 is conserved among almost all species, but substituted by Ala or Ser in *Mycobacterium* spp.. The lack of stabilisation of the purine ring by the base stacking Arg could be the reason why AMP is found to adopt a different binding mode in the *Mtb* structure from that in the *E. coli* structure. However it is very unlikely that HisG in the few organisms without this feature will have a different substrate and inhibitor binding site from that present in the majority of organisms. It is more likely that the different binding mode of AMP is in part the result of the conformation of the protein induced by histidine inhibition permitting at least two conformations of the AMP. This explanation accounts for the poorly resolved electron density seen for AMP in the *Mtb* structure [Cho *et al.*, 2003].

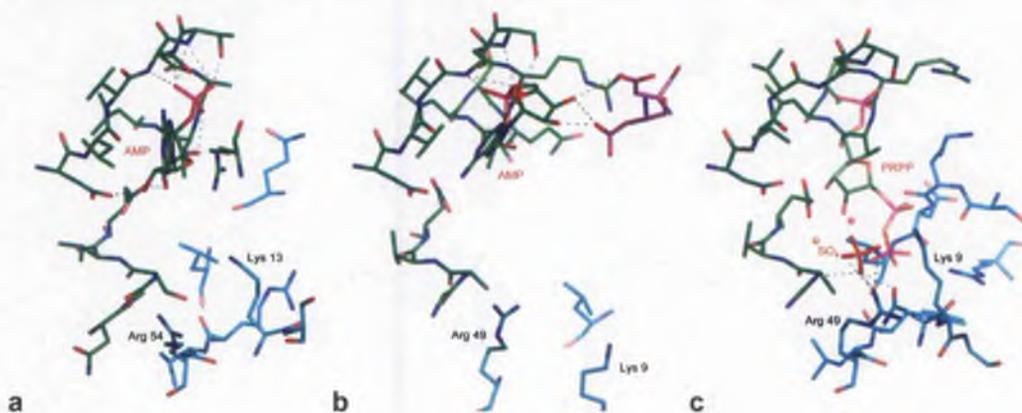


Figure 8-3 AMP and PRPP binding in HisG. In the same orientation, superimposed on the PRPP binding motif, AMP binding in *E. coli* (a) and *Mtb* HisG (b) is shown in stick representation as well as SO₄ binding in apo *Mtb* HisG structure (c). Hydrogen bonds are shown in grey dotted lines. PRPP present in orotate PRT from *S. typhimurium* (OPR1) is superimposed in (c) as shown in Section 6.5.1. Residues from dimer and trimer related subunits are shown with cyan and purple carbon atoms respectively.

A comparison of the AMP binding site in the various HisG structures with the PRPP binding site of orotate PRT (as described in Section 6.5.1) gives further insight into the PRPP binding of the HisG (Figure 8-3). The apo *Mtb* enzyme has a sulphate ion located in a position that corresponds to the 1- β -phosphate of the PRPP in the (orotate) PRT structure. The distance between the sulphate and the 1- β -phosphate is 2.3 Å in this approximation. The binding is accomplished by the side chains of residues Lys 9 and Arg 49 and main chain nitrogen of Ala 139. Both basic residues are highly conserved. In some organisms, like *E. coli*, Lys 9 (position 14 in *E. coli*) is substituted with a structurally equivalent Lys found one amino acid towards the C-terminus (Lys 13 in *E. coli*). Arg 49 (Arg 54 in *E. coli*) is conserved as Arg or Lys amongst all species with the exception of yeasts. In the AMP-inhibited *E. coli* structure these residues are found in a similar position and orientation to those in the apo *Mtb* structure; however in the two histidine-inhibited structures from *Mtb* and *Mth* these residues are much further away from the pyrophosphate binding site. Therefore the large structural change, caused by the inhibitor histidine, not only closes the hexamer (see Section 8.4), but significantly disrupts the pyrophosphate binding site. Therefore part of the action of the inhibitor can be explained by the proposal that the substrate PRPP cannot bind in this conformation.

8.3 A 'regulatory' loop in HisG

Upon binding of histidine a major structural change and repositioning of the 10 residue long loop (residue 29-39 in *E. coli* numbering) connecting helix α 1 and strand β 2 from the neighbouring subunit can be observed (see Figure 8-4). This loop will therefore hereafter referred to the 'regulatory' or 'R-loop'. In the presence of AMP alone, as seen in the *E. coli* enzyme, the 'R-loop' forms hydrophobic interactions with α 6 on the dimer-dimer interface. When histidine is bound, the 'R-loop' folds towards the PRPP binding site between domain 1 and 2. In the *Mtb* structure complexed with AMP and histidine, the 'R-loop' forms interactions with the AMP as described above. In the *Mth* histidine binary complex the 'R-loop' is even closer to the PRPP binding site and overlaps with the substrate binding sites therefore sterically hindering the binding of these. This results in a number of additional interactions of the 'R-loop' with the neighbouring subunit; these are formed between a charged residue from one of the subunits (Arg 12, Asp 30' and Ser 161, *Mth*

numbering – Arg 16, Gln 33' and Ser 172, *E. coli* numbering) with backbone atoms from the neighbouring subunit. Additionally a hydrogen bond is formed between Gln 41' (Glu 41' in *E. coli*) of the neighbouring subunit and Glu 212 (Arg 224 in *E. coli*) from the C-terminal domain. In terms of the *E. coli* numbering, only Arg 16 is well conserved; however for most charged residues which are not highly conserved and involved in the interactions a structurally equivalent residue with similar properties can be found (e.g. Glu 225 rather than Arg 224 in *E. coli*). From this we can conclude that, upon histidine binding, the dimer-dimer interactions of the 'R-loop', connecting $\alpha 1$ and $\beta 2$, change from hydrophobic to charged, forming a stronger link and tighter assembly within the hexamer.



Figure 8-4 AMP binding and binding site in HisG. (a) shows the superposition of HisG trimers from *E. coli* (red), *Mtb* (AMP+histidine; green) and *Mth* (histidine; magenta) aligned at the PRPP binding motif of one subunit. (b) shows a close up and

(c-e) the individual structures in the same colour code and with individual residues and AMP shown in stick representation. The loop between $\alpha 1$ and $\beta 2$ from a neighbouring subunit is marked by arrows in (b). Hydrogen bonds are indicated by grey dotted lines.

8.4 Changes in the quaternary structure of HisG

Four HisG structures have been determined so date, *E. coli* AMP binary¹⁵ (this study), *Mth* histidine binary (this study), *Mtb* apo and *Mtb* AMP and histidine ternary [Cho *et al.*, 2003]. These structures represent all hexameric forms of the enzyme. Therefore it is not practicable to discuss conformational changes upon inhibitor binding in terms of individual changes within the monomer or even dimer. For this reason the overall hexameric assembly and its changes will be examined first before the effects on monomer and dimer level will be discussed.

8.4.1 The HisG hexamer

The hexameric structure of the four different HisGs is shown in Figure 8-5. The conformation of the hexamer changes from “most open” to “most closed” on progression from apo, AMP-bound, AMP plus histidine-bound to histidine-bound. This is reflected by the overall dimensions of the oligomer which alters from 95 x 100 x 115 Å, 100 x 90 x 110 Å, 95 x 95 x 120 Å to 95 x 95 x 125 Å for the respective hexamers. The active site is located on the interior of the HisG hexamers (see Section 6.3.3). Access to the inside of the hexamer is restricted to a significant cavity at the dimer-dimer interface with a diameter of about 20 Å, and a smaller 5 Å diameter opening on the trimer interface along the 3-fold axis which is exclusive to the apo form. The larger cavity is seen in the AMP-bound structure with similar dimensions but a slightly modified shape. In the AMP- and histidine-bound form the cavity is almost closed with a diameter about 5 Å, while the cavity is essentially completely closed in the histidine-bound structure. There is almost no break on the surface of the histidine-bound HisG which makes it the most compact hexamer. This observation is underlined by the sum of buried surface area which is listed in Table 8-1.

¹⁵ The *E. coli* PR-A binary structure will not be taken into account here, as it is basically isomorphous with the AMP binary structure.

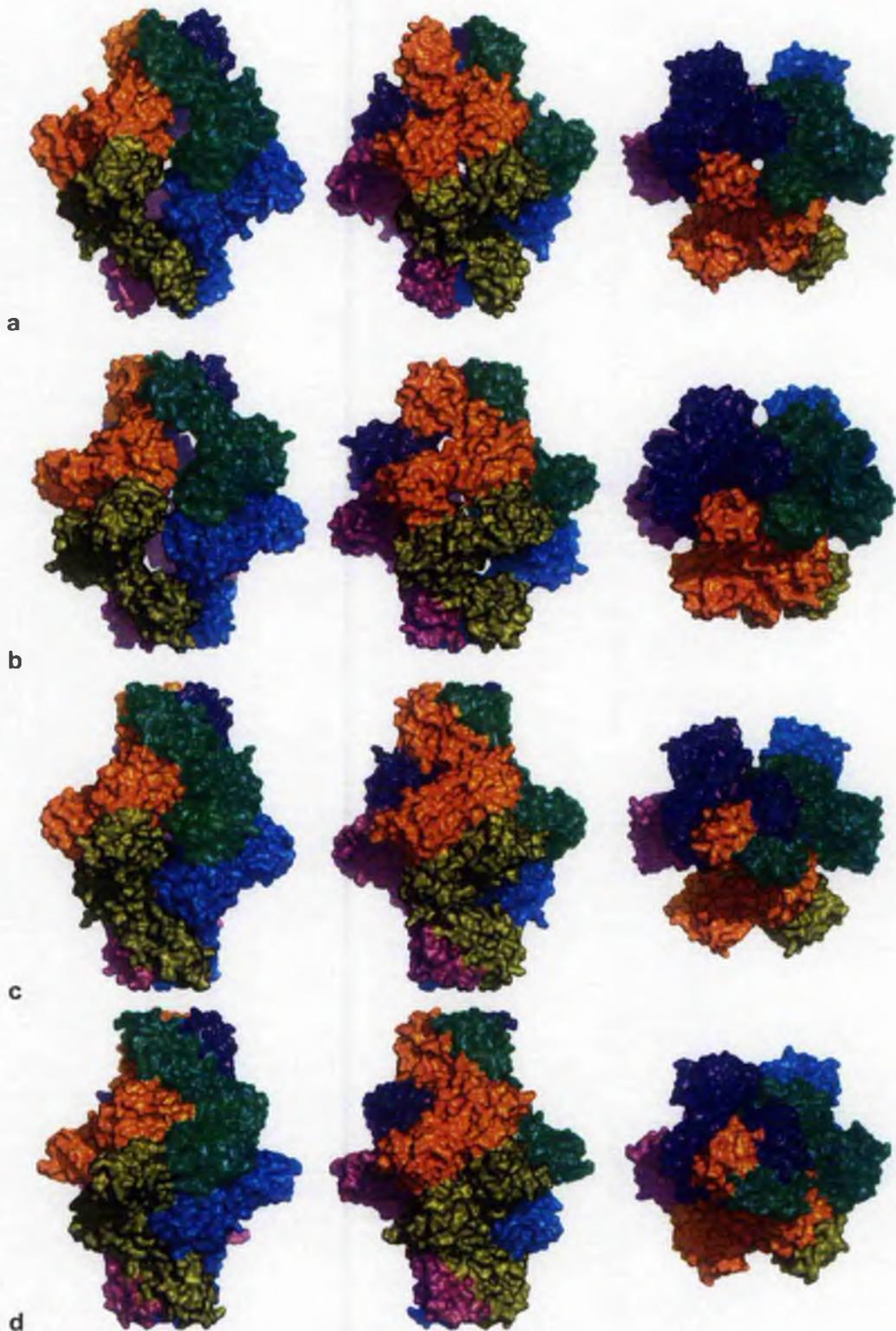


Figure 8-5 Surface representation of HisG hexamers. a) apo *Mtb*, b) *E. coli* AMP binary, c) *Mtb* AMP+histidine ternary d) *Mth* histidine binary. The surfaces of the hexamers are coloured according to the subunits and shown in three different views:

middle – along the (non-)crystallographic two-fold axis, left - rotated around 60°, right - along crystallographic 3-fold axis.

Table 8-1 gives the buried solvent accessible surface areas for the dimer, trimer and dimer-dimer interactions. In the apo HisG hexamer the largest surface area buried is at the dimer interface with little contribution made by dimer-dimer interactions. The C-terminal domains are almost fully solvent accessible with the exception of the β -sheet formed between strand $\beta 15$ and $\beta 16$ from a neighbouring subunit. This suggests that the dimer in the apo hexamer resembles most closely the active dimer of HisG. Comparison of the differences in buried surface areas for the different dimers show the conformational change is not restricted to the C-terminal domain, even though this domain is the main contributor to the hexamer interactions. For all the hexamers studied, the contributions from dimer-dimer interactions, outside the C-terminal domain and helix $\alpha 7$, are small (about 100 \AA^3). The histidine-inhibited structure represents an exception with about 400 \AA^3 buried on each side of domain 1 and 2. Additionally it represents the only structure in which interactions between a top and bottom subunit from neighbouring dimers are made. Interestingly the histidine-bound structure represents the most compact and closed conformation with the most surface area buried. This would suggest that it would be the most stable and therefore best inhibited structure. In addition, it shows the most complete surface, making the active sites in the centre of the hexamer the least accessible. This is in apparent contradiction to the finding that AMP and histidine inhibit HisG synergistically which would suggest the most closed conformation should occur in the presence of both inhibitors. However, the slightly open conformation of the AMP binary complex, especially with access to the histidine binding site between the C-terminal domains, explains the synergism well (see Section 7.3). On the other hand it cannot be excluded that the histidine-inhibited HisG structure is more compact than expected as it was solved from a thermophilic organism and enzymes from these organisms tend to be more stable than their mesophilic counterparts.

Table 8-1 Buried solvent accessible areas for HisG hexamers.

	Dimer interface (\AA^2)	trimer interface (\AA^2)		Additional dimer-dimer interface (\AA^2)	total (\AA^2)
		1	2		
<i>Mth</i> apo	1364	domain 1&2 99	domain C-terminal domain & helix $\alpha 7$ 535 634	domain 1&2 & helix $\alpha 7$ 76 587 663	2661
<i>E. coli</i> AMP	1176	104	741	83 810 893	2914
<i>Mtb</i> AMP+histidine	1144	139	1002	135 938 1073	3358
<i>Mth</i> histidine	789	423	1018	377 1059 1436	4029

8.4.2 The HisG monomer and dimer

A comparison of all HisG dimers is shown in Figure 8-6. The structures were superimposed on the first two domains of one monomer to observe the conformational change of the C-terminal domain and the relative position of the monomers within the dimer to each other. In general very little conformational change upon inhibitor binding can be seen between the first two domains. This is reflected by the low rms deviation between the structures, namely 1.0-1.6 Å for the C α atoms of the first two domains (see Figure 8-6). The conformational changes between the structures are listed here.

apo – AMP binary (Figure 8-6 a)

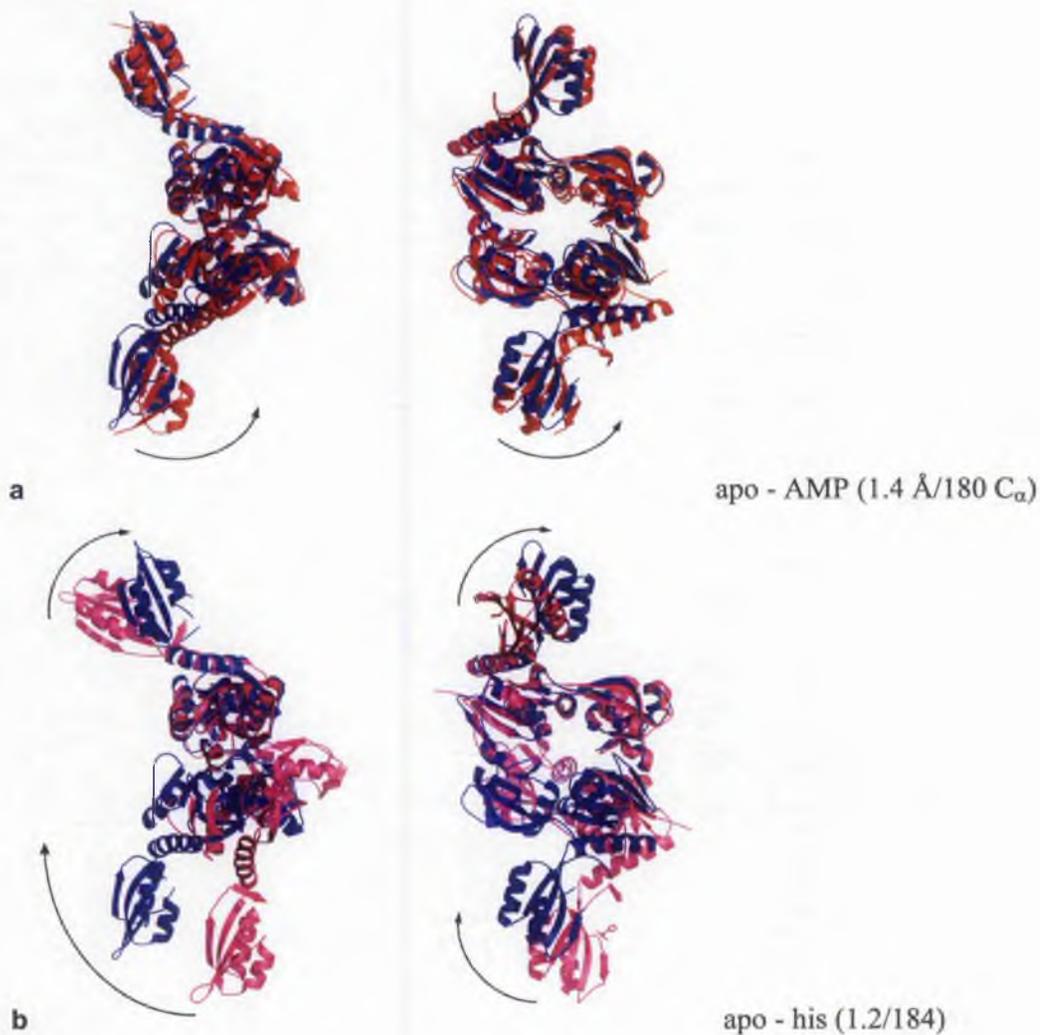
Very minor conformational changes are seen between the monomer upon AMP binding. The rms deviation increases only slightly from 1.4 Å to 1.7 Å when the whole structure is aligned rather than just the first two domains. However a small rotation of the C-terminal domain away from the 3-fold axis can be observed. The dimer interface is similar for the two structures and described for the *E. coli* enzyme in Section 6.3.3. Three hydrogen bonds are conserved on the dimer interface between the structures of which side chain oxygen Asp 57 to main chain Asn 152 is maintained in all structures discussed here. Asp 57 is conserved in all species and seems to be crucial for the formation of the dimer. Furthermore it acts as two anchor points¹⁶ around which conformational changes take place. In the AMP-bound structure the second monomer is rotated about 3° along the dimer interface. This results in the tighter packing of the C-terminal domains giving a more compact hexamer.

apo – histidine binary (Figure 8-6 b)

Upon histidine binding a significant change in the monomer can be seen with the C-terminal domain rotating 35° around the end of helix α 7 towards the 3-fold axis and the interior of the hexamer. The dimer interface is almost completely altered and the second monomer rotated about 40° around the Asp 57-Asn 152 axis (note the movement of helix α 3) to give a more open, elongated dimer which in turn results in a

¹⁶ N.B. There are two anchor points due to the (non-)crystallographic 2-fold symmetry relating the two monomers within the dimer to each other.

10 Å longer hexamer. Interactions at the dimer interface which occur on the inside of the hexamer (i.e. between the 'R-loop' connecting $\alpha 1$ and $\beta 2$ and helix $\alpha 6$ from the neighbouring subunit) are broken upon histidine binding and a new interface is formed on the outside of the hexamer between $\beta 8$ and $\alpha 2$. Because the rotation of the C-terminal relative to the first two domains is virtually in the opposite direction the monomer rotation the C-terminal domains remain close to the 3-fold axis and the hexamer stays intact.





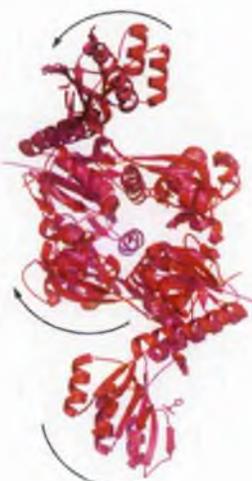
c



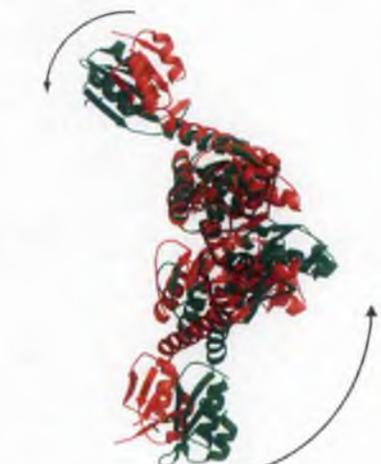
apo - AMP+his (1.0/198)



d



AMP - his (1.3/193)



e



AMP - AMP+his (1.6/192)



Figure 8-6 Conformational change of HisG monomers and dimers. HisG dimers of apo *Mtb* (blue), *E. coli* AMP binary (red), *Mth* histidine binary (magenta) and *Mtb* AMP and histidine ternary (green) are shown in ribbons representation superimposed on the first two domains of the top monomer. The right figure shows a view along the 2-fold (non-)crystallographic axis and the right one a view 90° rotated around the 3-fold crystallographic axis. Panels a-f shows all possible permutations. The numbers besides the pictures give the rms deviation in Å calculated over the stated number of C_α atoms.

apo – AMP & histidine ternary (Figure 8-6 c)

The conformational changes upon binding of AMP and histidine are comparable with those which occur when only histidine is bound. However the domain and monomer rotations are not as pronounced as in the histidine-bound structure with the C-terminal domain rotating about 20° and the second domain about 30°. Therefore the histidine-bound hexamer shortens 5 Å when AMP is bound.

AMP binary – histidine binary (Figure 8-6 d)

The binding of AMP and histidine, although both causing the formation of an inhibited hexamer, have different effects on the monomer and dimer of HisG. The C-terminal domain is rotated about 40° upon binding of histidine compared to binding of AMP. Not only is the C-terminal domain rotated around the end of helix $\alpha 7$ but $\alpha 7$ itself rotates in a concerted fashion by about 3°. Additionally the second monomer is rotated about around 40° along the Asp 57-Asn 152 axis. This results in a less

elongated hexamer than the *Mth* histidine-bound form and a narrower hexamer than in the *E. coli* AMP complexed HisG.

AMP binary – AMP & histidine ternary (Figure 8-6 e)

Binding of histidine to HisG already complexed with AMP again shows changes on monomer level with helix $\alpha 7$ tilting about 3° and the C-terminal domain about 27° both in a concerted manner towards the 3-fold axis. Again the second monomer in the dimer is rotated outwards by about 30° which overall results in a longer more compact hexamer.

Histidine binary – AMP & histidine ternary (Figure 8-6 f)

When AMP binds to the histidine binary enzyme complex the C-terminal domain shifts about 3.5 \AA perpendicular to the strands of the ferredoxin-like β -sheet away from the 3-fold axis. The second monomer rotates about 10° back towards the 3-fold axis. This shortens the hexamer by 5 \AA .

Conformational changes upon inhibitor binding can be summarised mainly as changes in the C-terminal domain and the orientation of the monomers within the dimer. Upon histidine binding the largest changes in conformation are observed with the C-terminal domain tilting towards the 3-fold axis and the second monomer rotating outwards elongating the dimer and hence the hexamer. Binding of AMP on the other hand has smaller effects with little change at the monomer level and only slight changes in the monomer orientations within the dimer. However these are large enough to change the hexamer conformation considerably, especially at the trimer interface between the C-terminal domains.

Chapter 9: Structure of PA5104 from *Pseudomonas aeruginosa*

PA5104 was cloned and initially purified and crystallised by Neil Paterson during a 4th year project which I supervised. Reported here are the crystallographic part of this project and the solution of the structure which was carried out by myself.

9.1 Solving phases of PA5104

9.1.1 Attempts to solve phases

9.1.1.1 Direct methods

Given diffraction data to beyond 1 Å, the first attempt to solve the phase problem is by direct methods. The application of two methods used in this work are described.

Direct methods are based on the linear relationship between phases as expressed in triplet (or quartet) invariants. The probability distributions of these invariants are significantly broadened for increasing number of atoms N (the variance is proportional to $N^{1/2}$ for triplets and N^1 for quartets) and therefore only reliable and practicable for structures with limited N . Phases can be calculated and refined with the tangent formula, derived from minimizing the phase probability distribution, or the minimal function as used in the program Shake'n'Bake. Furthermore computing time, especially for structure factors, increases rapidly with N . Therefore solving macromolecular structures with direct methods is challenging methodological as well as computational.

9.1.1.1.1 Shake'n'Bake (SnB)

SnB [Miller *et al.*, 1994] has been successfully used to solve protein structures *ab initio* with 1183 non-hydrogen atoms (about 100 aa). PA5104 with 196aa therefore represents a significant challenge to this method of structure solution.

First the structure factors (F) were converted in normalised structure amplitudes (E) with the program DREAR [Blessing and Smith, 1999]. Attempts to run SnB with the default parameters and arrays suggested for the amount of atoms and given resolution failed, because computational arrays limits were exceeded. Therefore

the sampling grid was broadened to reduce the size of the required arrays. SnB used about 150 days of computing time on an origin 200 server using one MIPS R10000 processor. The resulting distribution of R-factors is shown in Figure 9-1. The distribution is monomodal which indicates that no correct solution was found using this approach. The computationally expensive calculations with SnB did not allow further improved attempts to solve the structure with varied parameters¹⁷.

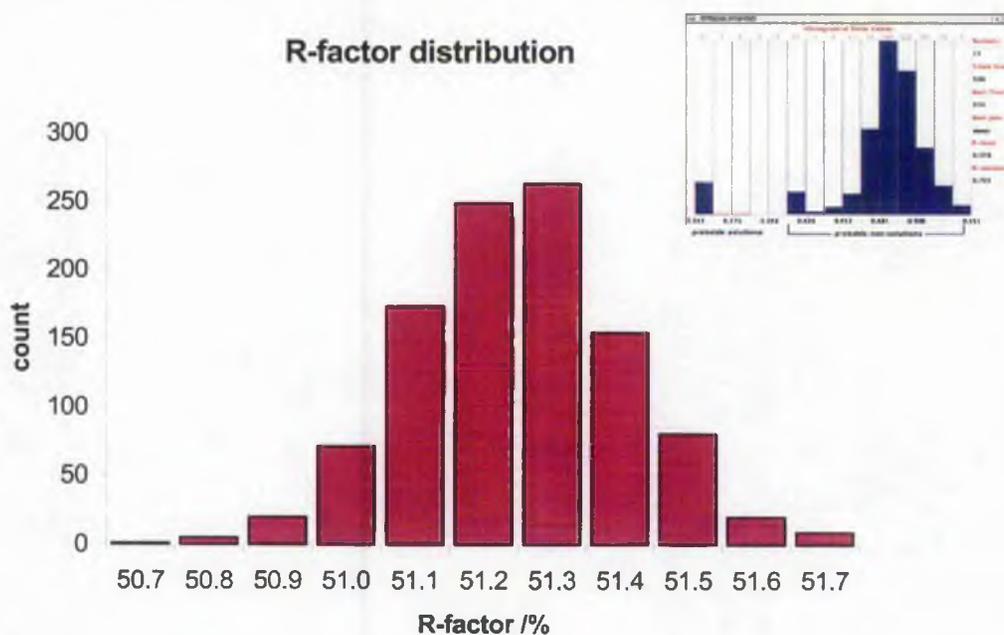


Figure 9-1 R-factor distribution result from SnB. The main panel shows the monomodal distribution of R-factors obtained for PA5104. The insert shows the bimodal distribution for a successful example. Here the R-factor represents the value for the minimal function.

9.1.1.1.2 ACORN

ACORN is a program for solving protein structures *ab initio* when atomic resolution data ($\leq 1.2 \text{ \AA}$) are available. It uses trial phases, obtained experimentally or by MR, which then are modified by iterative cycles of dynamic density modification and application of the Sayre equation until a solution is obtained [Foadi *et al.*, 2000]. A positive solution is usually indicated by a correlation coefficient for medium E-values (CC_s) of higher than 15% (and/or a significant increase from starting values). Here, initially in the absence of any phase information, trial phases were tried to be

¹⁷ The structure was solved otherwise, whilst SnB was still running.

obtained by MR. Prior to using ACORN, normalised structure amplitudes were calculated with the program ECALC [CCP4, 1994].

Positioning of a random atom using MR-ACORN yielded CC_s of below 5.5%. This can be expected for a protein structure with no heavier metal ion present. (N.B. initially it was not known that no metal ion is present, however the lack of any significant anomalous signal in the diffraction data collected at 1 Å wavelength was indicative. However a significant absorption effect was observed. To rule out the presence of any metal ions in the protein, an X-ray fluorescence scan was recorded at Station 9.5, Daresbury SRS, which showed no significant metal contribution).

A more successful approach is a MR search with a 2ndary structure element, especially an α -helix [pers. communication J. Foadi]. Various theoretical α -helices of variable lengths were tried which are available from the CCP4 library. This is justified as secondary structure prediction and CD-spectroscopy showed 11-28% α -helical contribution for PA5104. However none of these search models yielded a CC_s of greater than 6.0%. More extensive MR searches, at the expense of computing time, usually yielded a slightly higher CC_s (about 0.5%) but were still unacceptably low. Changes in the classification of Es very rarely improved the CC_s . The use of single theoretical β -strands and turns of varying lengths resulted in CC_s in ACORN of about 5.5%. Multiple β -strands extracted as sheets from high resolution structures (both parallel and anti-parallel) improved the CCs slightly (up to 1%), but still failed to give a positive solution.

Furthermore Fe-S clusters were probed as well as whole or parts of protein structures. These structures were identified with a similarity search based on the secondary structure prediction (pyruvate kinase - pdb code 1A3W, cutinase/serine esterase - 1AGY, surface glycoprotein - 1CID, ribonuclease - 1DY5, and immunoglobulin - 1VGE). None gave a positive solution.

Because tryptophan is the amino acid with the largest scattering matter and the amino acid sequence of PA5104 contains 7 tryptophan residues an attempt to use just a tryptophan ring for the MR in ACORN was made as was adding a tryptophan to a secondary structure element. However all these attempts were unsuccessful.

Further examples and combinations of secondary structure elements themselves or in combination with other amino acids could have been probed. However, this would not only be time consuming but success in solving the structure

would not be guaranteed. (N.B. currently it is not possible to place more than one fragment in MR-ACORN) Therefore phase information had to be determined experimentally.

9.1.1.2 Sulphur SAD/MAD

Collection of high redundant data at higher wavelength (1.5 Å and higher) should allow the location of sulphur atoms with anomalous differences [Dauter *et al.*, 1999]. Therefore a data set was collected using the wavelength of 1.9 Å to maximise the anomalous signal. Unfortunately neither these data nor other data, collected at longer wavelength, could be used due to scaling problems (described in Section 5.1.3). Therefore it was initially tried to obtain sufficient anomalous signal from the data set collected at a wavelength of 1 Å. Neither the program SOLVE [Terwilliger and Berendzen, 1999] nor an anomalous Patterson search with SHELX [Sheldrick, 1997] yielded a useful anomalous scatterer site. The latter sites did not refine in MLPHARE and SHARP. Results from SOLVE and subsequently RESOLVE had a low Figure of Merit (FoM) of below 0.15 and 0.60 respectively and a low 'score' of below 10. A further indication that these data did not contain enough anomalous signal was obtained from the SOLVE data analysis which indicated a signal-to-noise ratio of below 1 in all cases.

The data collected investigating the wavelength dependent diffraction of PA5104 crystals at wavelength 1.5 Å were used as a SAD data set. Again no correct solution could be found with the above mentioned approach.

Because data sets were collected at various wavelengths to investigate possible absorption effects these could be used to obtain phase information by using them as a MAD experiment. However no correct solution was found using SOLVE due to lack of anomalous signal, especially signal-to-noise ratio. This was because the data sets were not collected for this approach and therefore insufficient data of the required high redundancy were available.

9.1.1.3 Derivatives

Independently two Pt derivatives of PA5104 were obtained as well as two Hg derivatives (one collected at the L-III edge and therefore called 'peak', the other near the peak, called 'normal') and used to solve the structure of PA5104 by MIR. Heavy atoms sites were searched for in SHELX and SOLVE. No sites were found using

SHELX. SOLVE determined 2 weak Hg sites (occupancy about 0.1), which were consistent between the two Hg data sets and both refined in MLPHARE. Further sites with very low occupancy (below 0.05) could be identified but did not refine in MLPHARE. No further sites could be determined manually using difference Fourier maps¹⁸. The MIR was therefore reduced to a SIR problem.

Both Hg data sets (peak/'normal') were independently and together refined against the native data set in MLPHARE¹⁹. This resulted in a FoM of 0.28 for both data sets and 0.19 and 0.10 for the individual data sets respectively. The occupancies for the two sites were 0.18/0.12 and 0.10/0.06. The hand of the sites was confirmed based on the FoM after MLPHARE and DM. To avoid overestimating the phases, as two equivalent heavy atom sets and sites were used, only the peak data were used subsequently. The phasing statistics for this derivative are shown in Table 9-1. Following density modification and phase extension to 1.5 Å with DM the FoM increased from 0.15 to 0.60. Visual inspection of the resulting maps showed very little connected density (see Figure 9-11) and no obvious secondary structure elements, so that no initial model could be built. Therefore it was attempted to use the phases as a starting point for ARP-WARP, first in MIR mode with the aim of improving the input phases, and then to auto-build the structure. However this approach was not successful and the R- and free R-factor failed to decrease below 40%.

Table 9-1 Phasing statistics for PA5104 Hg peak derivative

R_{iso} /%		14.2
R_{cullis}	isomorphous	0.86/0.94 ¹
	anomalous	0.94
Figure of merit		0.36/0.16 ¹
	mean	0.19
Phasing power		0.51/0.62 ¹
	¹ centric/acentric	

¹⁸ SOLVE should have done that already. However maps were calculated with FFT (CCP4) and investigated further.

¹⁹ Refinement in MLPHARE always included isomorphous as well as anomalous data.

9.1.2 SIR(AS) phasing and solving of phases

Even the phases obtained from the SIRAS experiment described in the previous section were not correct enough to solve the structure. However they should be reliable, as the sites refined correctly in MLPHARE with two independent data sets, and therefore should still be better than random. Thus the phases can be used in ACORN(-PHASE) as trial phases. The density modified phases together with the atomic data were the input for ACORN and resulted in a solution with a CCs of 43% and a R-factor of 33.1%. This clearly indicates a correct solution. Figure 9-2 shows the improvement of the CCs with phasing cycles. An equivalent graph of an unsuccessful trial is shown in the same scale for comparison (Figure 9-3).

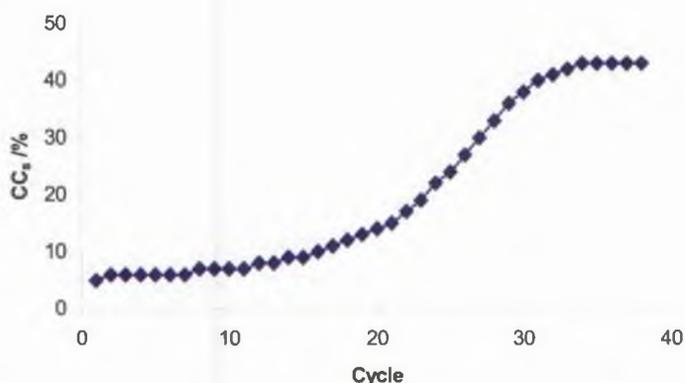


Figure 9-2 Changes in correlation coefficient CC_s for successful ACORN calculation. The solution is evident from the asymptotic behaviour of CC_s and increase over 15%. The result was obtained using atomic resolution data with SIR density modified phases.

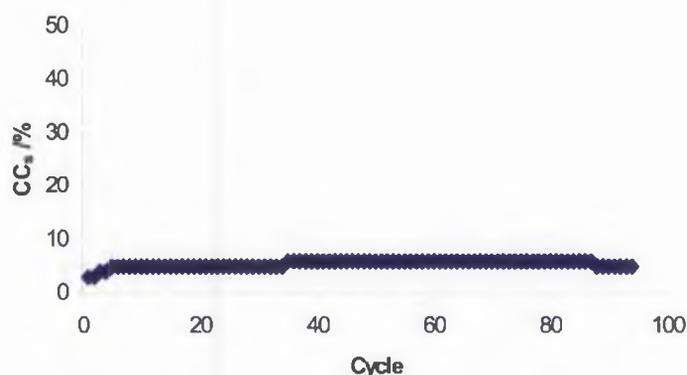


Figure 9-3 Changes in CC_s for unsuccessful ACORN calculation. No (correct) solution is found which is indicated by no increase in CC_s . The example shown is the result of an ACORN calculation with a theoretical α -helix as a MR search model.

9.1.3 Model building and refinement

The map obtained from ACORN was used in ARP-WARP for automated model building in the resolution range 20-1.5 Å. The R- and free R-factor was about 35% for the starting model and dropped to 16.2% and 20.4% respectively after model building and refinement. ARP-WARP built a continuous polypeptide from residue 3 to 195 including side chains and some additional dummy atoms. The terminal residues, to complete the model, were fitted manually using QUANTA [Accelrys Inc.]. Clear density was seen for some PEG molecules and octahedral coordinated Mg²⁺ ions. Additional water molecules were added and refined using ARP-WATER. Multiple side and main chain conformations were built manually in QUANTA. The final refinement cycles were carried out in the resolution range from 20-0.96 Å in REFMAC5. The refinements included riding hydrogens and isotropic B-factor refinement. The R- and free R-factor changes in each refinement step are given in Table 9-2.

Table 9-2 R-factor and free R-factor during the refinement of PA5104

Refinement step	R-factor /%	Free R-factor /%	Phase error /° (20-1.75Å)	Phase error /° (20-1.5Å)
SIR	n/a	n/a	79.1	n/a
After DM	n/a	n/a	73.8	75.4
ACORN	33.1	n/a	26.2	27.3
ARP-WARP	16.2	20.4	14.8	14.2
ARP-WATERS	15.5	19.9	n/d	n/d
REFMAC5	11.5	13.1	used as reference	

9.2 PA5104 model

9.2.1 Quality of model

The final model of PA5104 had an R-factor of 11.5% and a free R-factor of 13.9%. The rms deviation of the bond lengths was 0.031 Å and of the bond angles 2.93°. All residues are within the allowed regions of the Ramachandran plot (see Figure 9-4). All geometric side and main chain parameters are satisfactory or better

than expected for this resolution as indicated by a PROCHECK analysis. A summary of the refinement statistics is shown in Table 9-3.

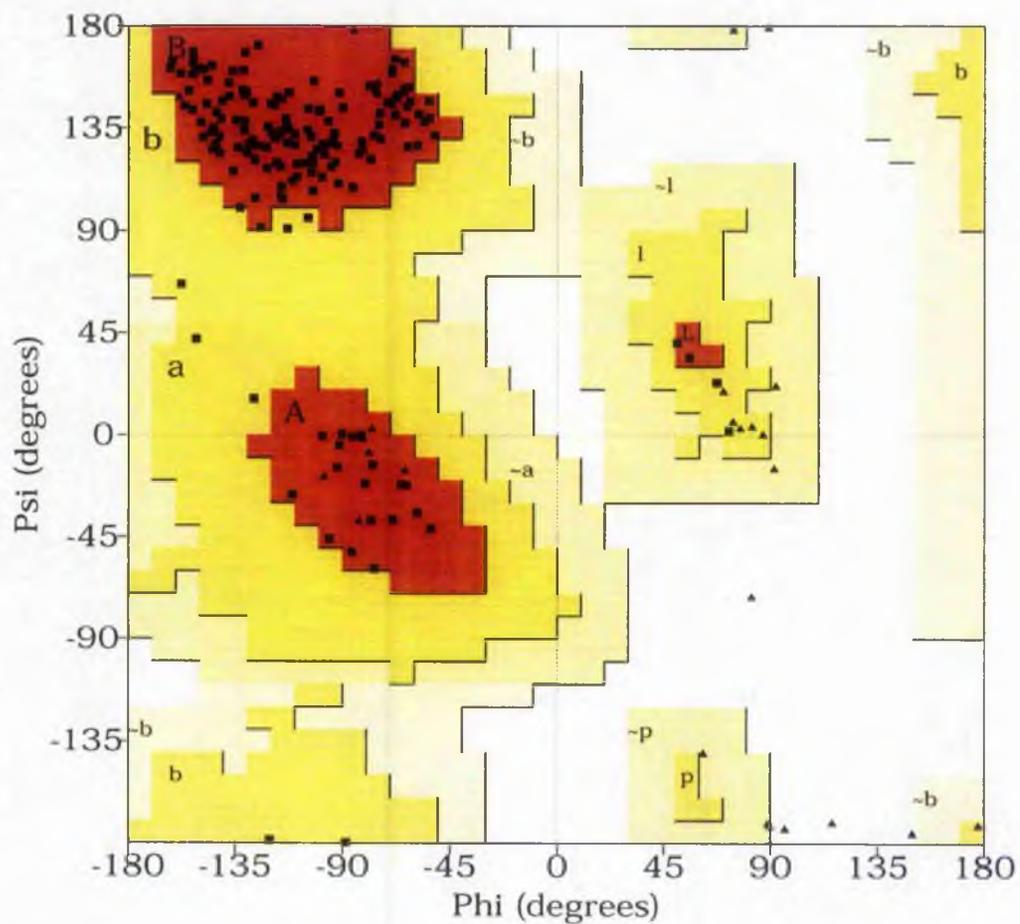


Figure 9-4 Ramachandran plot for PA5104.

Table 9-3 Refinement statistics of PA5104.

Refinement statistics	
Resolution range (Å)	20.0-0.96
R _{working} (%) ¹	11.5 (113589)
R _{free} (%) ^{1,2}	13.1 (6334)
Number of non-hydrogen atoms	
Protein	1664
Nonprotein (PEG, Tris)	79
Mg ²⁺ ions	4
Water	489
R.m.s. deviation from ideal	
Bond length (Å)	0.031
Bond angles (°)	2.93
Ramachandran statistics	
Most favoured	153
Additionally allowed	10
Generously allowed	0
Disallowed	0

¹ number in parenthesis is the number of reflections.

² R_{free} was calculated with 5% of reflections.

9.2.2 Overall structure

The structure of PA5104 is comprised of two all-β domains (Figure 9-5). These two domains are linked with a three residue ₃₁₀-helix, the only helix present in the structure. Both domains are similar in fold, showing the barrel-like double stranded β-helix structure with two antiparallel β-sheets.

The first domain consists of residues 1-11 and 118-196 and therefore contains both N- and C-termini. The 9 β-strands, β1 and β12-b19, form two antiparallel sheets with strand order 1 17 14 19 12 and 13 18 15 16 respectively. This represents a jelly roll topology with the addition of strand β1. The short ₃₁₀-helix is assigned to this domain.

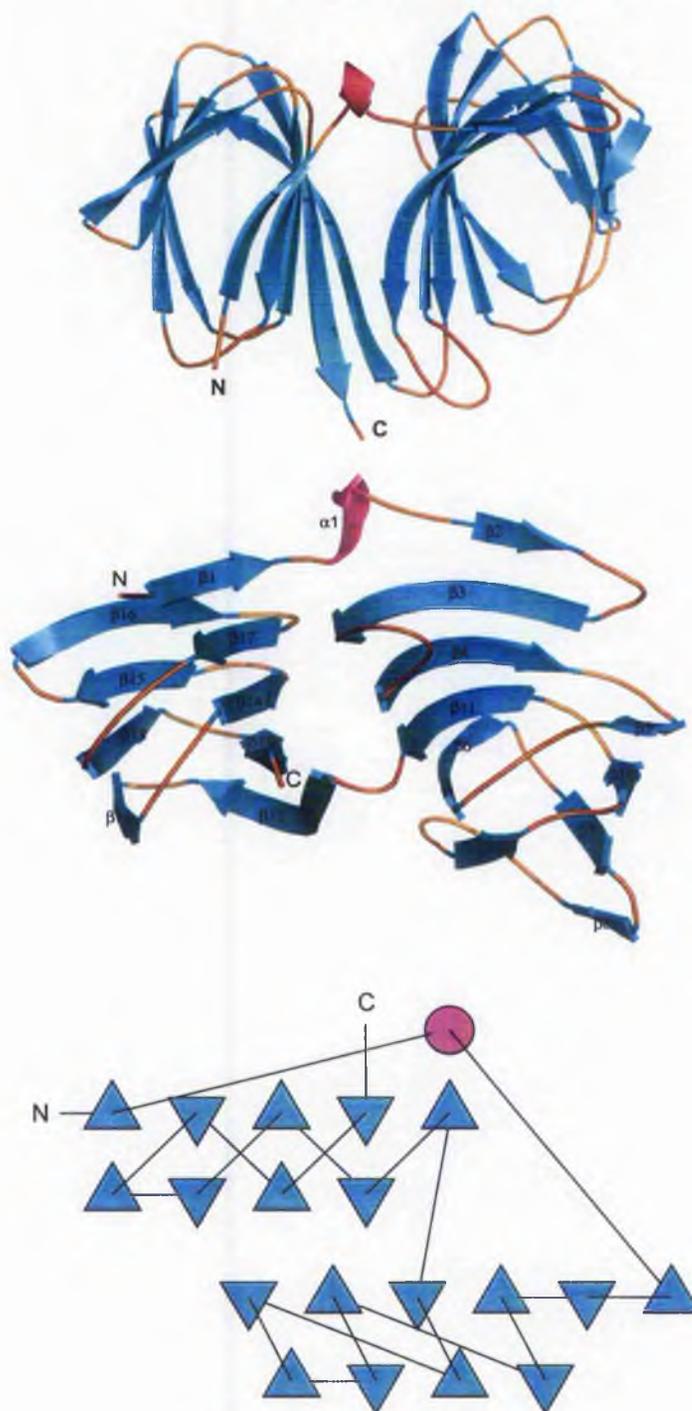


Figure 9-5 The structure of PA5104 solved at atomic resolution. Two ribbons presentation of the PA5104 monomer coloured according to the secondary structure. The top panel shows a top view and the middle panel a view from the side. The bottom panel shows a topology diagram prepared with the program TOPS [Gilbert *et al.*, 1999].

The second domain contains residues 12-117 which includes strands 10 β -sheets, β 2- β 11. The two antiparallel β -sheets are build of strands 2 3 4 11 6 9 and 5 10 7 8 respectively. Again this shows the typical jelly roll topology with addition of strands β 2-3. The barrel of this domain is open and has a cleft in centre which is filled with a PEG molecule (see Figure 9-6). The molecule is bound in a region of the protein with high sequence similarity among different species. This gives an indication for the location of an active site or small molecule binding site which will be discussed below (Section 9.4.3).

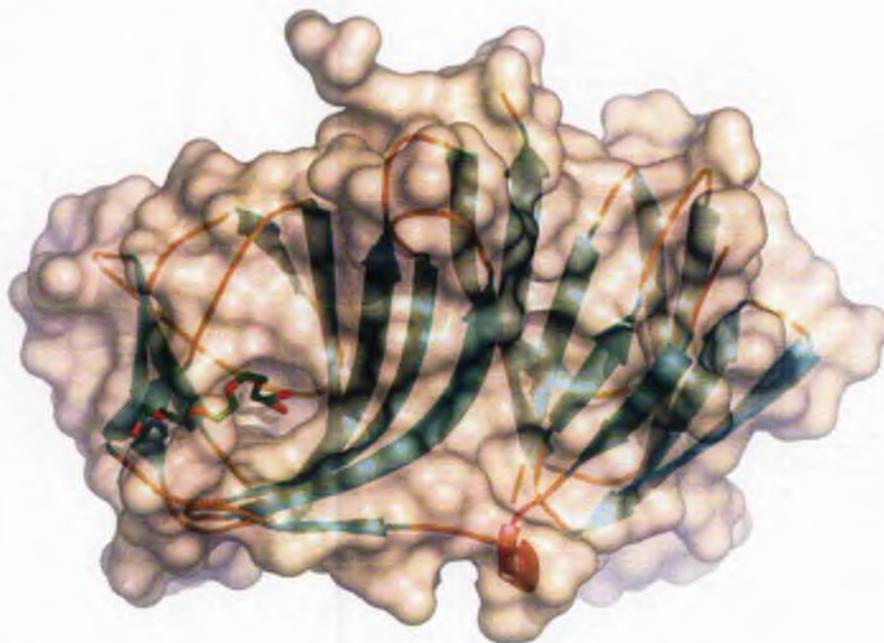


Figure 9-6 Surface representation of PA5104. Shown is a ribbon representation of PA5104 and the semitransparent surface as calculated with PyMOL. The PEG molecule binding in the cleft of the second domain is shown in stick representation.

Each domain shows a fold typical for cupins (from Latin 'cupa' for small barrel), therefore the whole protein is a bicupin. A 3D search against the SCOP database shows that the first domain is most similar to lectins/glucanases, seed storage proteins, and regulatory proteins of the AraC family (e.g. 1MAC, 2PHL, and 1ARC respectively). The second domain was determined to be most similar to the fold of glucose permease-like proteins, seed storage proteins and phosphomannose isomerase (e.g. 1F3G, 2PHL, and 1PMI respectively). Overall the fold resembles closest to the bicupin phosphomannose isomerase (see Figure 9-7).

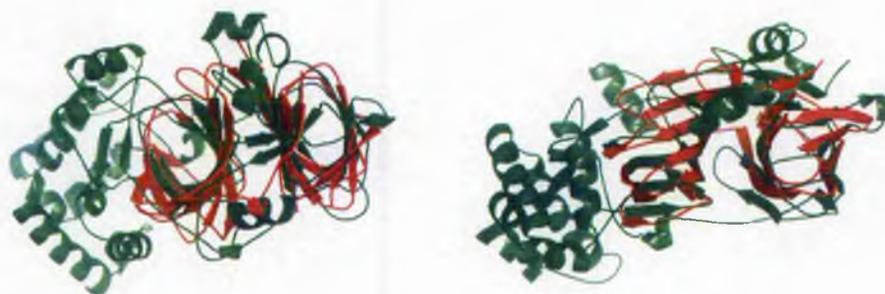


Figure 9-7. Comparison for bicupins PA5104 and phosphomannose isomerase. A ribbon representation of phosphomannose isomerase (pdb code 1PMI) in green and PA5104 in green superimposed.

9.3 Structure of 2nd crystal form of PA5104

Before the structure of PA5104 was solved a data set was collected on a second crystal form. The structure of this crystal form was solved once the atomic resolution structure was available.

9.3.1 Molecular replacement

MR was carried out with the refined structure of PA5104 as a search model in AMoRe. The rotation function showed four clear solutions (see Figure 9-8) based on the CC and R-factor. Four monomers were placed with translation search and the position and orientations were refined. The change of CC and R-factors when additional monomers are included is given in Table 9-4.

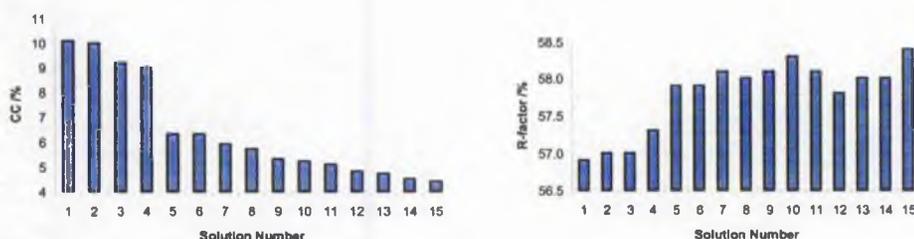


Figure 9-8 AMoRe rotation function results for PA5104. The left panel shows the CC and the right the R-factor for the top 15 peaks. The first four solutions are clear solutions as the CC is above and the R-factor below the background of false solutions.

Table 9-4 CC and R-factor for MR in AMoRe.

monomer	CC /%	R-factor /%
1	24.1	52.8
2	39.7	47.7
3	53.6	42.5
4	60.0	40.6

9.3.2 Model building and refinement

The model from the AMoRe result was first refined using rigid body refinement in REFMAC5 [Murshudov, 1996]. Iterative cycles of manual model building in QUANTA [Accelrys] and refinement in REFMAC5 with TLS [Winn *et al.*, 2001] were carried out. Each monomer was treated as one TLS group and in the initial steps of the refinement NCS restraints were imposed. Solvent molecules were added using the automated procedure in ARP-WARP. The R-factor statistics of the refinement steps are given in Table 9-5.

Table 9-5 R-factor and free R-factor during the refinement of PA5104 (crystal form 2)

Refinement step	R-factor /%	Free R-factor /%
Rigid body	36.5	36.5
REFMAC5 + TLS +NCS	24.1	28.3
With waters	19.3	24.2
REFMAC5 + TLS	18.1	23.6

9.3.3 Quality of the model

The model was refined to an R-factor of 18.1% and a free R-factor of 23.6% to 1.9 Å resolution. The geometrical parameters for main and side chains were acceptable or better than expected for this resolution as analysed by PROCHECK [Laskowski *et al.*, 1993]. The rms deviation for bond lengths was 0.016 Å and 1.436° for bond angles. All residues were in the core or additionally allowed region of the Ramachandran plot. A summary of the refinement statistics is shown in Table 9-6.

Table 9-6 Refinement statistics of PA5104 (crystal form 2)

Refinement statistics	
Resolution range (Å)	46.63-1.90
R _{working} (%) ¹	18.1 (46893)
R _{free} (%) ^{1,2}	23.6 (4707)
Number of atoms	
Protein	6159
Nonprotein (glycerol, sulfate)	127
Cl ⁻ ions	8
Water	361
R.m.s. deviation from ideal	
Bond length (Å)	0.016
Bond angles (°)	1.436
Ramachandran statistics	
Most favoured	598
Additionally allowed	50
Generously allowed	0
Disallowed	0

¹ number in parenthesis is the number of reflections.

² R_{free} was calculated with 8.4% of reflections

9.3.4 The structure of PA5104 from a 2nd crystal form

Overall the fold of the four NCS related monomers in the second crystal form of PA504 is the same as in the atomic resolution structure. The four monomers are arranged as two 'dimers' in the asymmetric unit (asu) (Figure 9-9). They interact via the bottom of the barrel of the first domain. Although the main interactions are hydrophobic, there are a significant number of hydrogen bonds. The majority of the latter occurs between main chain atoms of the 'dimer' related sheets in which the respective last strands run antiparallel to each other. Similar interactions and 'dimer' formation can be seen in the atomic resolution structure of PA5104 with symmetry related molecules.



Figure 9-9 PA5104 structure from the second crystal form. Shown in ribbon representation are the 4 monomers in the asu coloured according to the chain. The top two monomers and bottom two form a 'dimer' via the first domains.

It is questionable whether the 'dimers' observed here are physiologically relevant as DLS showed that the molecule is monodisperse as a monomer. However other cupins form real dimers which can depend on substrate binding, e.g. the regulatory protein AraC. This forms a barrel-barrel dimer, as seen here, in absence of arabinose and a dimer via its second, α -helical, domain in the presence of arabinose [Soisson *et al.*, 1997]. It remains to be proved if PA5104 can form homo-dimers upon substrate binding or if it can dimerise with another protein of a cupin fold via this mode.

The four monomers are very similar with rms deviations between the C_{α} atoms of 0.50 Å, 0.52 Å, and 0.45 Å between chain A and chain B, C, and D respectively. The main differences are in domain two in the loop between strands $\beta 3$ and $\beta 4$ (around residue 35, see Figure 9-10). Chain C shows additional differences in

the position of strand $\beta 2$ (see Figure 9-12) and the loop connecting this strand with $\beta 3$ could not be resolved in the electron density.

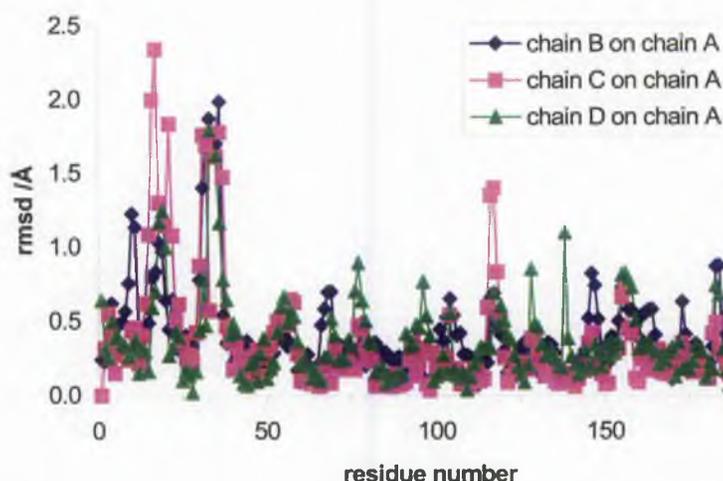


Figure 9-10 Rms deviation between the four monomers of PA5104.

9.4 Discussion

9.4.1 (Problems) solving the structure

Processing of, and extracting anomalous signal from, data collected on PA5104 crystals using wavelengths of 1.5 Å and longer proved to be extremely difficult and sometimes even impossible (see Section 5.1.3). Therefore the structure could not be solved using sulphur SAD. The precise reasons for this are still unclear, but absorption effects or radiation damage seem to be the most likely causes. However no radiation effects have been observed as judged from preliminary refinement of data collected on one crystal using different wavelengths (0.7 Å and 1.5 Å) which show no differences in refinement statistics and electron density difference maps.

All attempts to solve PA5104 using ACORN in MR mode with theoretical models failed. Various reasons can be found to explain this observation. First of all, no α -helix is found in the structure of PA5104 which is the preferred search model for use in ACORN. This is due to the compact scattering matter and orientation which helps to find a correct position in the MR routine [pers. comm. J. Foadi]. β -strands or β -sheets are more difficult to position and therefore a solution of the structure is less

likely. Furthermore β -elements show more structural diversity which makes it difficult to find a closely matched structure that can be placed with MR. Simulations have shown that a correctly placed 4 residue β -strand (20 atoms) taken from the final model has enough scattering matter to solve the structure. However the fragment and even larger ones (e.g. a 8 residue β -strand) could not be placed correctly in the MR emphasising the importance of the correct positioning of the starting model and the problem in doing so. Further simulations were performed with correctly positioned Trp residues from the final model to investigate and justify the use of single amino acids in the MR search. It was shown that at least 3 complete Trp residues (42 atoms) are needed obtain a correct solution, 2 complete Trp residues (28 atoms) and 3 Trp rings only (30 atoms) gave no positive solution. This shows that Trp residues are not a good choice as a starting model to solve the phases, especially because of the problem of obtaining correct orientations for multiple Trp residues and different possible side chain conformers. Furthermore it shows that even after correct positioning of the search model the success of ACORN does not depend on the overall size of the model (20 atoms in a sheet were successful, but 30 in 3 Trp rings were not) but on its compactness.

Attempts at solving the structure of PA5104 *a priori* with direct methods (SnB) proved too much of a challenge for a protein of this size and failed using the approach described in Section 9.1.1.1.1. However, using experimental phases combined with phase refinement using the Sayre equation in ACORN proved to be successful. This approach was still not straight forward as the phase error for the initially obtained SIRAS phases was approximately 80° (see Table 9-2) and no solution could be obtained using ACORN. Only after improving the phases using density modification by solvent flattening (which reduced the phase error by 5%) was it possible to solve the phases using ACORN. Calculations with the modified phases restricted to 1.8 Å resolution showed that extending the phases to 1.5 Å was not required to get a positive result. The high phase error of the SIRAS phases is reflected in the poor electron density maps obtained at this stage, which did not allow any model building or solution by ARP-WARP. The density lacked connectivity and no secondary structure elements could be identified, a fact which was not supported by the lack of α -helical elements in the structure as these can be easier to identify than β -sheets. However it seems noticeable that higher density is observed for the peptide

bonds at this stage (Figure 9-11 a). The phase error was reduced to about 27° after successfully using ACORN (Table 9-2). This resulted in an electron density map of high quality which only marginally improved in the following automated model building and refinement (Figure 9-11 b-d).

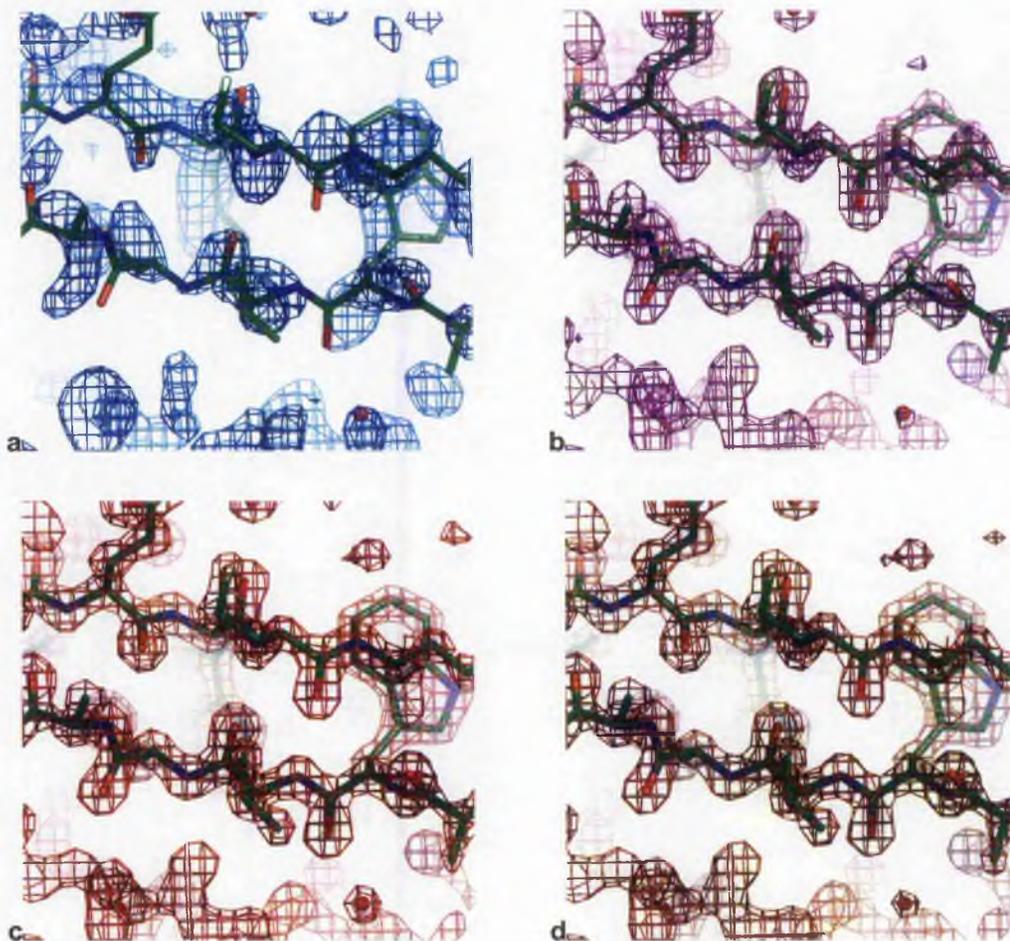


Figure 9-11 Electron density maps of PA5104 at different stages of phasing. Two representative strands are shown in stick representation with electron density maps contoured at 2σ . The maps have been phased after the SIRAS phasing (solvent flattened with DM) a), after ACORN b), after ARP-WARP c) and the final phases d).

9.4.2 Comparison of structures from the two crystal forms

A superposition of the four monomers of the second crystal form with the atomic resolution structure is shown in Figure 9-12. Overall the rms deviation of C_α atoms between the atomic resolution structure and the other four monomers is 1.52 Å, 1.48 Å, 1.63 Å and 1.78 Å for chain A, B, C, and D respectively. The difference

within the individual domain is smaller, so superimposition of both domains of chain A on the relevant domains of the atomic resolution structure gives an rmsd of 0.29 Å (excluding the difference described below).

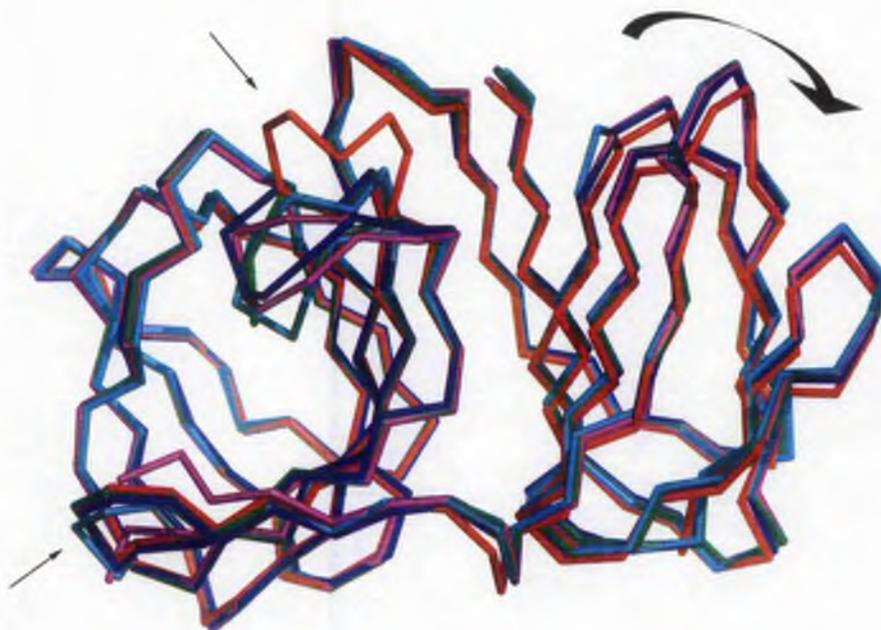


Figure 9-12 Superposition of PA5104 monomers. C_α trace of the high resolution structure (red) with the four monomers from the second crystal structure (green-chain A, blue-chain B, magenta-chain C and cyan-chain D). The monomers were superimposed on the second domain only with exception of the loops between strand β2 and β3, and β3 and β4 (indicated by thin arrows). The first loop shows a different conformation in all four monomers of the second crystal structure compared to the atomic resolution structure. The second loop is not resolved in chain C and strand β2 is tilted towards the centre of the barrel. The first domain in the atomic resolution structure is rotated with respect to the four monomers in the other structure (indicated by the boxed arrow).

The main difference between the structures is the loop between strands β3 and β4 in the second domain. In the atomic resolution structure this is open and close to the loop connecting the two domains whereas in the second crystal form the loop is moved towards the inside of the barrel in all monomers. This closes the 'active' site from the top but still leaves it open from the side (where the PEG is bound). Additionally in chain C of the second crystal form strand β2 is rotated towards the

inside of the barrel. The following loop itself is not resolved in the electron density map which indicates flexibility in this part of the structure. The two observed conformational changes could indicate possible movements in the second domain upon binding of substrate.

There are no large differences between the structures in domain 1. However a different position of the two domains towards each other can be seen between the atomic resolution structure and the second crystal form. In the first the whole domain is rotated away from the second domain by 3° . Figure 9-13 shows the movement in more detail.

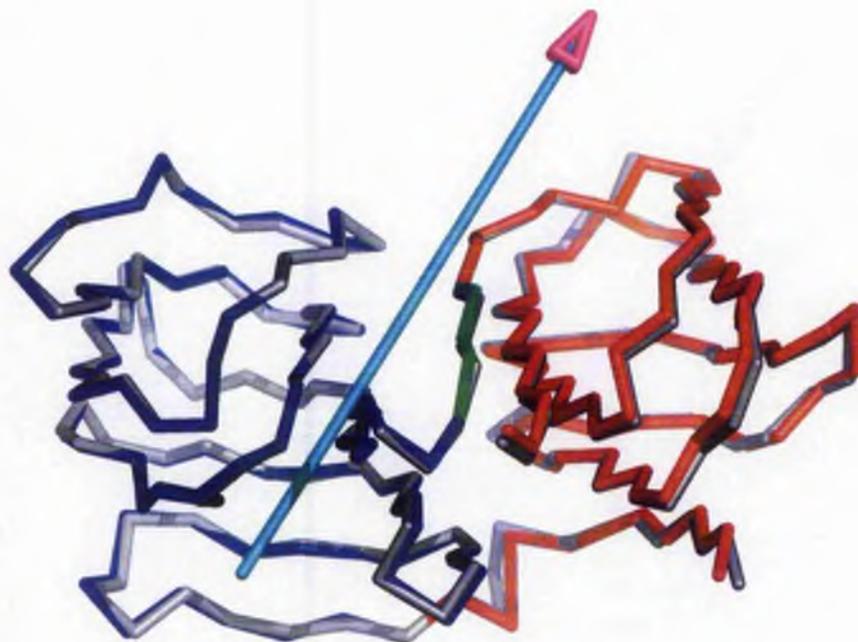


Figure 9-13 Rotation of the two domains in PA5104. A C_α trace of the atomic resolution structure (grey) superimposed with chain A of the second structure. The second domain (blue) was fixed and the first domain (red) rotated around the axis marked by the arrow by 3° . The hinge residues of strand $\beta 12$ are marked in green.

9.4.3 Active site and function

The active site or effector binding site is located in the barrel of domain 2 in PA5104. Five independent active site structures are available, one from the high resolution structure and four from the second structure. Superposition of the active sites show that very little conformational change is observed between the structures even though different small molecules are found to be bound in it (Figure 9-15). In the

high resolution structure a PEG molecule is bound, whereas in the second structure glycerol molecules and a sulphate ion are bound. The active site differs in all 4 monomers of the second structure. A number of water molecules are found to be coordinated in the active site too, of which only one is seen in all 5 structures. This water is hydrogen bonding to Arg 41, Ser 43 and Asn 110. All residues involved in contacts to bound molecules are conserved among species (Figure 9-14).

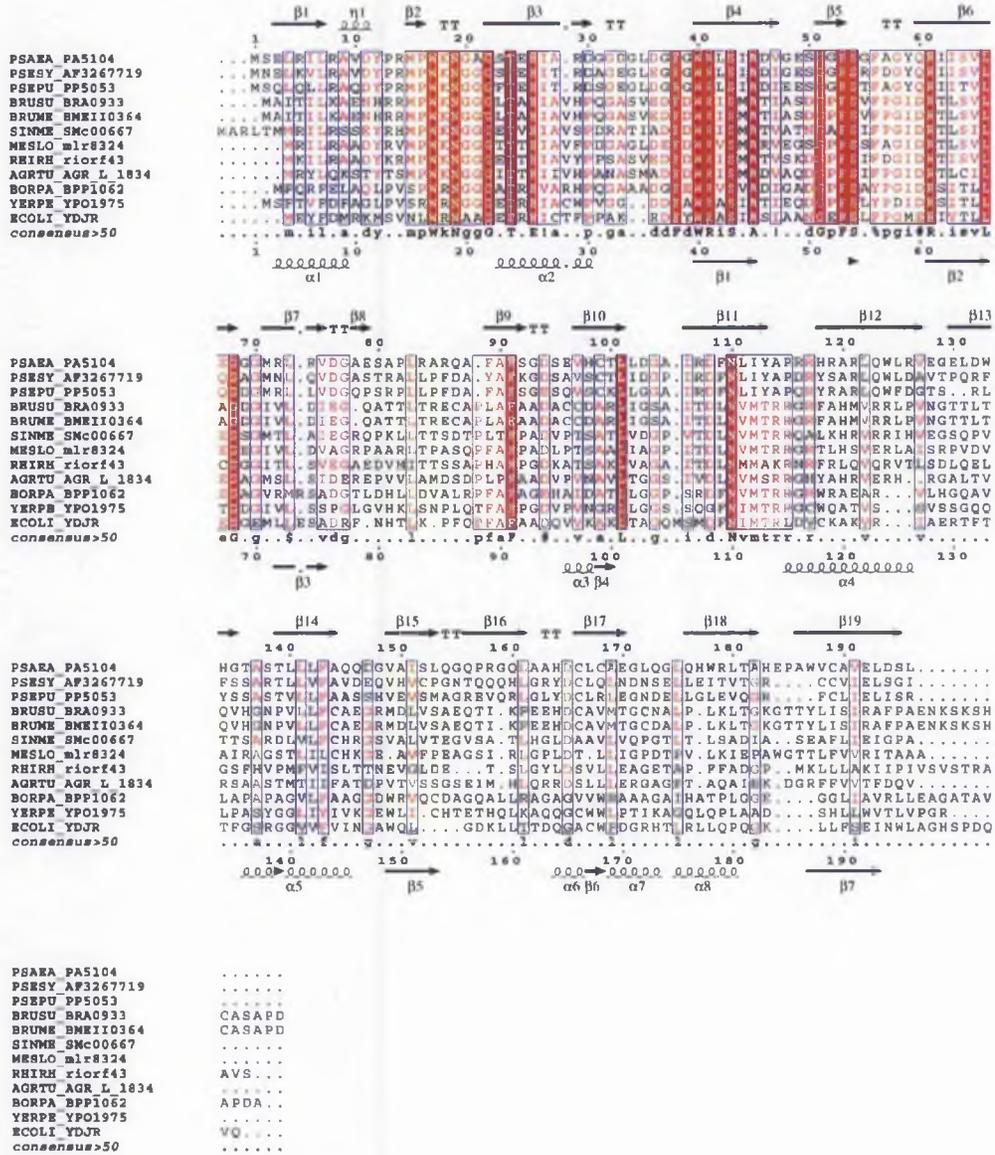


Figure 9-14 Sequence alignment of PA5104. A sequence alignment of an ORF similar to PA5104 from different bacteria together with the observed secondary structure, obtained from the structure, of the PA5104 from *P. aeruginosa* (top) and predicted secondary structure (bottom), produced using ESPrift [Gouet, 1999]. α-helices and

β -strands are represented as helices and arrows, respectively, and β -turns are marked with TT. This sequence alignment was created using the following sequences (Organism, SwissProt accession numbers or other source in brackets) PSEAE_PA5104 (*Pseudomonas aeruginosa*, NP_253791); PSESY_AF326719 (*Pseudomonas syringae*, AAK15635); PSEPU_PP5053 (*Pseudomonas putida*; P24696); BRUSU_BRA0933 (*Brucella suis*, NP_700100); BRUME_BMEI10364 (*Brucella melitensis*; NP_541342); SINME_SMc00667 (*Sinorhizobium meliloti*, NP_386811); AGRTU_AGR_L_1834 (*Agrobacterium tumefaciens*, NP_356706); MESLO_mlr8324 (*Mesorhizobium loti*, NP_108443); RHIHR_riorf43 (*Rhizobium rhizogenes*, NP_066624); BORPA_BPP1062 (*Bordetella parapertussis*, NP_883383); YERPE_YPO1975 (*Yersinia pestis*, NP_405532); ECOLI_YDJR (*Escherichia coli*, P76214).

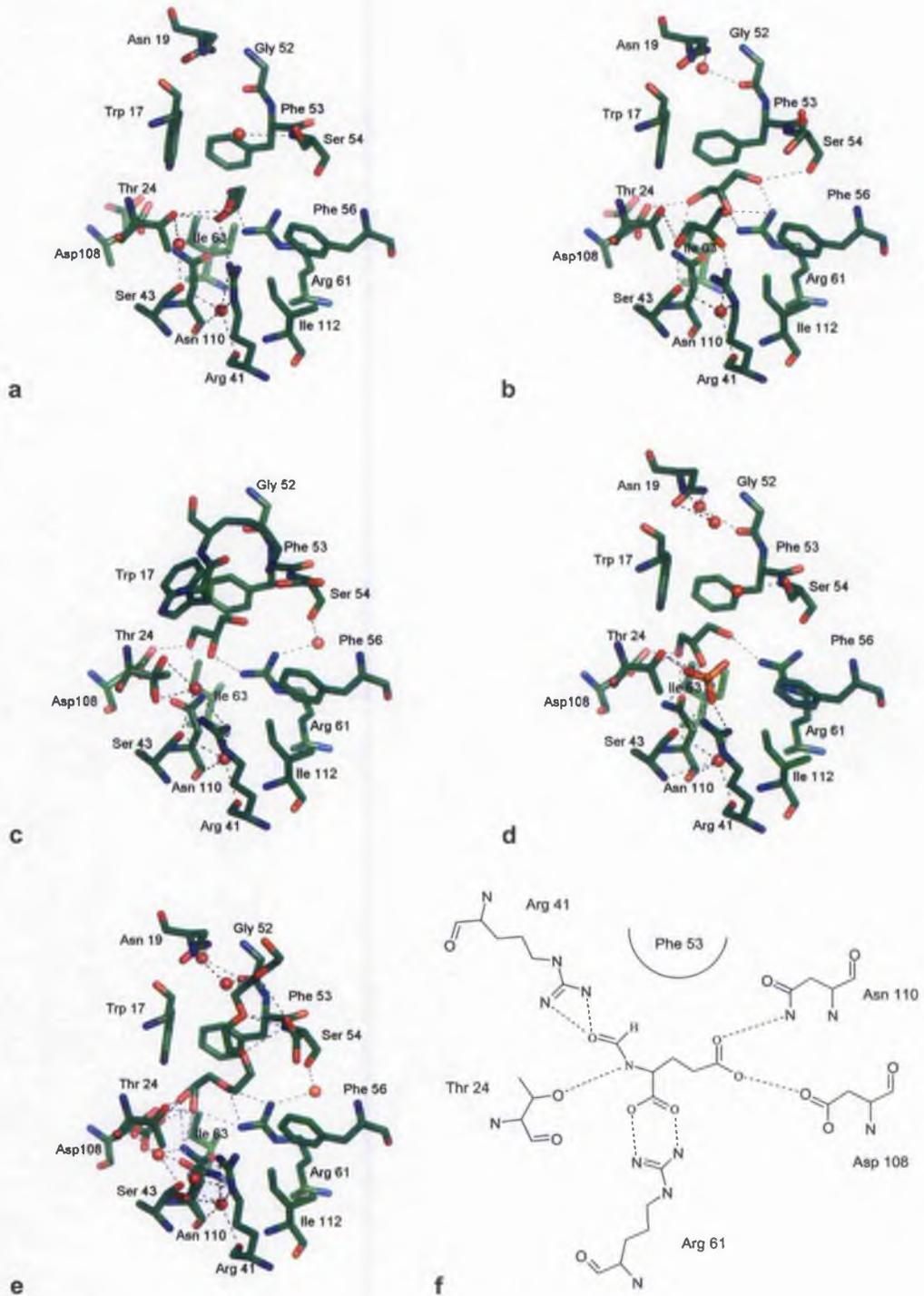


Figure 9-15 Active site of PA5104. a-d shows the active site in stick representation of the four chains of the second crystal structure and e of the high resolution structure. Hydrogen bonds between the protein and water or other molecules bound

in the active site are shown in grey dotted lines. f shows possible conserved interactions of PA5104 with N-formyl-L-glutamate.

The availability of the five different and independent active sites and their interactions with small molecules by conserved residues makes it possible to suggest potential substrates for PA5104 and therefore indicate its possible function. As PA5104 is part of the hut operon and all enzymes of the pathway are identified it seems likely that the protein is somehow involved in regulation of the pathway. Therefore the first candidates to be considered as substrates have to be intermediates in hut pathway 2. Superposition of the active sites and bound molecules therein showed that the substrates for HutF and HutG₂ (N-formimino-L-glutamate and N-formyl-L-glutamate respectively) both could be fitted in the active site. The interactions of PA5104 made with the small molecules in the 5 structures can be maintained upon binding of these substrates (Figure 9-15 f). The two intermediates are very similar with only the formimino group replaced by a formyl group in the later intermediate. However the possible interactions of this part of the molecule with the guanidinium group of Arg 41 makes it more likely to be formyl-glutamate as it can serve as a hydrogen donor for the oxygen of the formyl group. The observation that expression of HutF and especially HutG₂ both can be induced by their substrates respectively suggests that a protein binds either both or just one of the intermediates which then somehow activates expression of the respective Hut enzyme. As revealed, PA5104 could serve that function by binding formyl-glutamate which in turn causes a conformational change within the protein similar to the one observed between the high resolution structure and the structure of the second crystal form. This conformational change then might enable PA5104 to form a dimer comparable to the ones observed in other regulatory proteins, e.g. AraC (see Section 9.3.4). Because PA5104 itself has no typical DNA binding domain like a β - α - β motif, it is unlikely that the protein itself will bind to DNA to induce expression. It seems more likely that it will then bind to another, yet to be identified, protein which can bind to DNA or it disrupts DNA binding of another protein to DNA. Formation of a protein-PA5104 complex by either means can then result in induction of expression of the respective Hut enzyme.

There are however other functions and observations to be considered in the determination of the function of PA5104 in addition to those solely based on the

structures determined here and sequence alignments. First it cannot be excluded that glutamate, the end product of the hut pathways, can be bound in the active site as it shares a common scaffold with the two intermediates formimino-glutamate and formyl-glutamate. In case of binding of glutamate, PA5104 could function as a sensor reflecting the glutamate level within the cell. Additionally there is a minor possibility that a molecule could bind which is unrelated to any of the hut pathway intermediates or the end product, especially since there are a number of conserved hydrophobic residues in the active site. In the binding theory outlined above only Phe 53 and to a lesser extent Phe 56 form van der Waals interactions with the substrate but conserved Trp 17 has no function in the binding. Finally, ORFs similar in sequence to PA5104 are present in organisms which have no genes encoding any of the Hut enzymes and therefore lack the ability to utilise histidine. One example of such an organism is *E. coli* with gene ydjR being similar to PA5104 (see Figure 9-14). The amino acid sequence of ydjR shows that the residues for binding to hut intermediates, as outlined above, are conserved even though no hut pathway is present in *E. coli*. However the gene is found in a region of the genome where genes of the arginine degradation pathway are encoded. As glutamate is an end product in arginine catabolism it is possible that the protein encoded by ydjR functions as a glutamate binding protein as suggested above for PA5104. The determination of the structure of PA5104 suggests a series of experiments to elucidate the function and mode of action of PA5104. Important initial experiments would involve isothermal titration microcalorimetry to assess binding of potential ligands.

Chapter 10: Conclusions

10.1 HisG

10.1.1 Substrate binding sites and reaction mechanism

The structures of HisG determined in this work and those recently published by [Cho *et al.*, 2003], allow the substrate binding sites to be defined and provide a basis for formulating a reaction mechanism. The PRPP binding site was initially identified in the second domain of HisG by comparison of the AMP HisG structure with type I PRT structures. This comparison was conducted by superposition of the PRPP binding motif, characteristic for type I PRTs, with the modified PRPP binding motif found in HisG. The PRPP monophosphate binds to the 'P-loop' of the PRPP binding motif and the ribose hydroxyls hydrogen bond to conserved residues Glu 156 and Asp 169. In combination with the apo *Mtb* HisG structure, where a sulphate ion is bound near the active site, the pyrophosphate binding site was identified to be on the dimer interface with the conserved basic residues Lys 13 and Arg 54 from the neighbouring subunit and a Mg^{2+} ion being involved in the binding.

The binding site for the purine ring of the substrate ATP was shown to be located between domains 1 and 2 in the same position as observed in the HisG structure complexed with product PR-ATP. Because the ribose-triphosphate part of PR-ATP was not resolved in the electron density map, no exact description of this binding site can be given. However, domain 1 shows a conserved, negatively charged pocket where the triphosphate, mediated via Mg^{2+} ion(s), can bind. This would suggest that at least two Mg^{2+} ions are required for the catalysis, one for the binding of PRPP and one or more for the binding of ATP.

The catalytic reaction requires the C1 carbon of PRPP to be orientated in the proximity of the N1 of the adenine ring of ATP. In the model outlined above the distance of the modelled C1 to N1 of the product PR-ATP is 3.5 Å. However, no catalytic residue which could abstract a proton from or polarise the C1 of PRPP can be identified since there are no conserved, charged residues in the vicinity. The closest interactions possible are with the backbone carboxyl of Leu 170 and the guanidium group of Arg 16 which are 4-4.5 Å distant. It seems more likely that the reaction

mechanism proceeds via an oxocarbenium intermediate. The pyrophosphate is abstracted from PRPP first to produce the positively charged intermediate, which can be stabilised by interaction with the backbone carbonyl of Leu 170. The reaction is completed by the electrophilic attack of C1⁺ on the favourably orientated N1 of the purine ring of ATP. This mechanism can be reconciled with the observed ordered mechanism of HisG. ATP has to bind first because if PRPP were to bind first it would be hydrolysed and the oxocarbenium intermediate would not be particularly well stabilised. A second piece of evidence is that pyrophosphate has been shown to be released first. Further evidence has been provided experimentally by investigation of kinetic isotope effects in the HisG catalysis which showed the possibility of an oxocarbenium intermediate [Goitein *et al.*, 1978]. Additionally type I PRTs were shown to catalyse the transfer of the phosphoribosyl group via an oxocarbenium intermediate [Scapin *et al.*, 1995].

10.1.2 Mode(s) of inhibition

Inhibition of HisG by histidine, AMP and PR-ATP was shown to be concerted with a change in aggregation state of the enzyme from the active dimeric form to yield different, inhibited, hexameric conformations. The hexameric crystal structure of HisG explains this observation since the active sites are buried in the interior of the hexamer and made less accessible for the substrates. The hexamer is particularly compact and closed in the presence of inhibitors histidine, and AMP plus histidine, but comparatively open in the presence of AMP. This explains in part the synergistic effect between AMP and histidine and the lower K_i for histidine compared with AMP.

The competitive inhibition of AMP with respect to both substrates was impossible to reconcile with the assumption that AMP binds to the ATP binding site although this is the interpretation made from the ternary *Mtb* AMP and histidine HisG structure. However, the AMP-complexed HisG structure presented here allows the correct interpretation of the substrate binding sites, explaining the experimental observations with the ribose-monophosphate of AMP competing with the PRPP binding site and the purine ring of AMP competing with the ATP binding site. This mode of inhibition also holds for the product PR-ATP. Additionally the PR-ATP-inhibited structure is slightly more compact indicating stronger binding of the product.

Histidine binds in an intersubunit pocket between neighbouring C-terminal domains in the hexamer, resulting in a close contact between these domains and

ordering of the 'T-loop'. Surprisingly the histidine R-group forms only a single hydrogen bond to the carbonyl of Ser 288. Despite this the site has been previously shown to be highly specific for histidine [Morton and Parsons, 1977]. When histidine is bound, the hexamer is more elongated than when AMP or no inhibitor is present. This elongation is accompanied by a change in the dimer interface, with the dimers opening towards the outside so that the PRPP binding site on the dimer interface is disrupted. Therefore histidine not only inhibits by hexamer formation but also by disordering the substrate binding site. In the AMP-bound HisG structure the histidine binding pocket is exposed to solvent and in a favourable conformation to bind histidine which can explain the synergism between the binding of AMP and histidine.

10.1.3 The C-terminal 'regulatory' domain

The C-terminal domain of HisG can be described as a regulatory domain, not only because of the similarity with regulatory P_{II} signal transduction proteins, but because it provides the binding site for histidine and, in particular, is involved in the formation of inactive hexamers. Furthermore HisG mutants lacking the C-terminal domain showed activity, ruling out any involvement of the domain in catalytic activity. Although the C-terminal domain would not have been expected to play a role in substrate binding or catalysis, the native short form of HisG, without the C-terminal domain, is itself inactive and needs to form a hetero-complex with HisZ to display activity. This observation is even more surprising as there are no obvious differences in the sequence between the long and short form of HisG, apart from the C-terminus. Experiments showed that the native short HisG can be activated by thermal energy, presumably by a conformational change from a tensed, inactive form to a relaxed, active form. This same transition occurs on formation of the HisG-HisZ complex although there is no structural information on how this change occurs.

Finally, given the importance of P_{II} proteins in coordinating nitrogen and carbon assimilation in microorganisms and plants, the presence of a structural homologue as a domain of HisG is very interesting. This domain in HisG forms trimers in response to ADP and AMP which are signals of low energy levels and adopts a different trimer conformation on binding histidine. The enzyme could therefore act as a signalling molecule interacting with discrete receptor proteins to indicate the cellular levels of histidine. Alternatively the domain may have been acquired by gene fusion and now only functions as a small molecule-binding protein,

which regulates the activity of the enzyme. This might explain the presence in a number of bacteria and archaea of the short form of ATP-PRT in which the regulatory domain is absent and regulation is achieved by oligomerisation with HisZ, a histidyl-tRNA synthetase paralogue [Bovee *et al.*, 2002].

10.2 PA5104

Crystals obtained from PA5104, an unidentified protein in the hut operon of organisms with histidine utilisation pathway 2, diffracted to better than 1 Å. Therefore solution of the structure was attempted by direct methods using the programs SnB and ACORN. The structure, with 196 amino acids, proved too much of a challenge for SnB due to its size. Employment of ACORN, used in MR replacement mode, with various secondary structure elements and combinations as search models did not give a positive solution either. This failure is due to the lack of α -helices, the preferred search model in ACORN, or any heavy atoms in the structure. In addition, similarity searches and secondary structure prediction failed to identify PA5104 as a bicupin and an all- β structure. This could have focused the attempts with ACORN to various forms of β -sheets but would have been computationally very time-consuming with no guarantee of a solution. For these reasons it was necessary to obtain experimental phase information on PA5104. Diffraction data were collected at longer wavelengths to obtain anomalous signal from sulphur atoms for SAD phasing, but these data could not be scaled satisfactory for this approach, presumably due to absorption effects. Consequently, heavy atom derivatives of PA5104 were collected which proved to be difficult as the heavy atom complexes tested had severe effects on the diffraction quality of the crystals. However, it was possible for one derivative data set to locate two Hg sites with occupancies of just above 0.1 and to obtain some anomalous signal because the data were collected near the absorption edge of Hg. The acquired phase information was insufficient to allow manual or automated model building. Nevertheless, these phases were reliable and better than random, which allowed phase refinement using the Sayre equation in ACORN-PHASE to solve the structure of PA5104.

One of the cupin domains of PA5104 is closed, whereas the other has a crevice in the centre of the barrel. Conserved residues, of which most show interactions with small molecules from the crystallisation conditions, form the surface of the cleft. This

allows the identification of possible substrates for the binding site. Glutamate or its derivatives, which are the last two intermediates in the hut pathway, were recognised as possible candidates. Binding of the pathway intermediates could suggest a possible role of PA5104 in the regulation of expression of HutF and HutG₂ by their substrates respectively.

References

- Alifano, P., R. Fani, *et al.* (1996). "Histidine biosynthetic pathway and genes: Structure, regulation, and evolution." *Microbiological Reviews* **60**(1): 44-69.
- Allison, S. L. and A. T. Phillips (1990). "Nucleotide Sequence of the Gene Encoding the Repressor for the Histidine Utilization Genes in *Pseudomonas putida*." *Journal of Bacteriology* **172**(9): 5470-5476.
- Ames, B. N., R. G. Martin and B. J. Garry (1961). "The First Step of Histidine Biosynthesis." *Journal of Biological Chemistry* **236**(7): 2019-2026.
- Ames, B. N., T. H. Tsang, M. Buck and M. F. Christman (1983). "The Leader mRNA of the Histidine Attenuator Region Resembles tRNA^{his} - Possible General Regulatory Implications." *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **80**(17): 5240-5242.
- Arcondeguy, T., R. Jack and M. Merrick (2001). "P_{II} signal transduction proteins, pivotal players in microbial nitrogen control." *Microbiology and Molecular Biology Reviews* **65**(1): 80-105.
- Argos, P., M. Hanei, J. M. Wilson and W. N. Kelley (1983). "A Possible Nucleotide-Binding Domain in the Tertiary Fold of Phosphoribosyltransferases." *Journal of Biological Chemistry* **258**(10): 6450-6457.
- Banfield, M. J., J. S. Lott, *et al.* (2001). "Structure of HisF, a histidine biosynthetic protein from *Pyrobaculum aerophilum*." *Acta Crystallographica Section D-Biological Crystallography* **57**: 1518-1525.
- Beckler, G. S. and J. N. Reeve (1986). "Conservation of Primary Structure in the HisI Gene of the Archaeobacterium, *Methanococcus vannielii*, the Eubacterium *Escherichia coli*, and the Eukaryote *Saccharomyces cerevisiae*." *Molecular & General Genetics* **204**(1): 133-140.
- Bell, R. M. and D. E. J. Koshland (1971). "Allosteric Properties of the First Enzyme of the Histidine Operon." *Bioorganic Chemistry* **1**: 409-423.

- Blasi, F., S. M. Aloj and R. F. Goldberger (1971). "Effect of histidine on the enzyme which catalyzes the first step of histidine biosynthesis in *Salmonella typhimurium*." *Biochemistry* **10**(8): 1409-1417.
- Blessing, R. H. and G. D. Smith (1999). "Difference structure-factor normalization for heavy-atom or anomalous-scattering substructure determinations." *Journal of Applied Crystallography* **32**: 664-670.
- Bovee, M. L., K. S. Champagne, B. Demeler and C. Francklyn (2002). "The Quaternary Structure of the HisZ-HisG N-1-(5'-Phosphoribosyl)-ATP Transferase from *Lactococcus lactis*." *Biochemistry* **41**(39): 11838-11846.
- Brenner, M. and B. N. Ames (1971). The histidine operon and its regulation. *Metabolic pathways*. H. J. Vogel. New York, Academic Press Inc. **5**: 349-387.
- Brunger, A. T., P. D. Adams, *et al.* (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." *Acta Crystallographica Section D-Biological Crystallography* **54**: 905-921.
- Bruni, C. B., M. S. Carlomagno, S. Formisano and G. Paoletta (1986). "Primary and Secondary Structural Homologies between the *His4* Gene-Product of *Saccharomyces cerevisiae* and the *HisIE* and *HisD* Gene-Products of *Escherichia coli* and *Salmonella typhimurium*." *Molecular & General Genetics* **203**(3): 389-396.
- Carlomagno, M. S., L. Chiariotti, *et al.* (1988). "Structure and Function of the *Salmonella-Typhimurium* and *Escherichia-Coli* K-12 Histidine Operons." *Journal of Molecular Biology* **203**(3): 585-606.
- Carson, M. (1987). "Ribbon Models of Macromolecules." *Journal of Molecular Graphics* **5**(2): 103-109.
- CCP4 (1994). "The CCP4 Suite - Programs for Protein Crystallography." *Acta Crystallographica Section D-Biological Crystallography* **50**: 760-763.
- Chamberlin, M. and J. Ring (1973). "Characterisation of T7 specific ribonucleic acid polymerase I general properties of the enzymatic reaction and the template specificity of the enzyme." *Journal of Biological Chemistry* **248**: 2235-2244.
- Champagne, K. S., M. L. Bovee, *et al.* (2001). *Structural Studies of an ATP-Phosphoribosyl Transferase from Lactococcus lactis*. ACA Annual Meeting, Los Angeles.

- Chaudhuri, B. N., S. C. Lange, *et al.* (2001). "Crystal structure of imidazole glycerol phosphate synthase: A tunnel through a (β/α)₈ barrel joins two active sites." *Structure* **9**(10): 987-997.
- Chelsky, D. and S. M. Parsons (1975). "Stereochemical course of the adenosine triphosphate phosphoribosyltransferase reaction in histidine biosynthesis." *Journal of Biological Chemistry* **250**(14): 5669-5673.
- Cheng, S., C. Fockler, W. M. Barnes and R. Higuchi (1994). "Effective Amplification of Long Targets from Cloned Inserts and Human Genomic DNA." *Proceedings of the National Academy of Sciences of the United States of America* **91**(12): 5695-5699.
- Chipman, D. M. and B. Shaanan (2001). "The ACT domain family." *Current Opinion in Structural Biology* **11**(6): 694-700.
- Cho, Y., V. Sharma and J. C. Sacchettini (2003). "Crystal structure of ATP phosphoribosyltransferase from *Mycobacterium tuberculosis*." *Journal of Biological Chemistry* **278**(10): 8333-8339.
- Coleman, W. G., Jr. and L. S. Williams (1974). "First enzyme of histidine biosynthesis and repression control of histidyl-transfer ribonucleic acid synthetase of *Salmonella typhimurium*." *Journal of Bacteriology* **120**(1): 390-393.
- Coote, J. G. and H. Hassall (1973). "The Control of the Enzymes Degrading Histidine and Related Imidazolyl Derivatives in *Pseudomonas testosteroni*." *Biochemical Journal* **132**: 423-433.
- Coote, J. G. and H. Hassall (1973). "The Degradation of L-Histidine, Imidazolyl-L-lactate and Imidazolylpropionate by *Pseudomonas testosteroni*." *Biochemical Journal* **132**: 409-422.
- Cowtan, K. (1994). DM: An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*. **31**: 34-38.
- Dall-Larsen, T. (1988). "Regulation of the first step of the histidine biosynthesis in *Escherichia coli*." *International Journal of Biochemistry* **20**(3): 231-235.
- Dall-Larsen, T. (1988). "Stopped flow kinetic studies of adenosine triphosphate phosphoribosyl transferase, the first enzyme in the histidine biosynthesis of *Escherichia coli*." *International Journal of Biochemistry* **20**(8): 811-815.

- Dall-Larsen, T. and L. Klungsoyr (1976). "The binding of specific ligands to adenosine-triphosphate phosphoribosyltransferase." *European Journal of Biochemistry* **69**(1): 195-201.
- Dauter, Z., M. Dauter, *et al.* (1999). "Can anomalous signal of sulfur become a tool for solving protein crystal structures?" *Journal of Molecular Biology* **289**(1): 83-92.
- Davisson, V. J., I. L. Deras, S. E. Hamilton and L. L. Moore (1994). "A Plasmid-Based Approach for the Synthesis of a Histidine Biosynthetic Intermediate." *Journal of Organic Chemistry* **59**(1): 137-143.
- DeLano, W. L. PyMOL. www.delanoscientific.com.
- Delorme, C., S. D. Ehrlich and P. Renault (1992). "Histidine Biosynthesis Genes in *Lactococcus lactis* subsp. *lactis*." *Journal of Bacteriology* **174**(20): 6571-6579.
- Deutscher, M. P. (1990). Guide to protein purification. *Methods in Enzymology*, Academic Press. **182**.
- Donahue, T. F., P. J. Farabaugh and G. R. Fink (1982). "The Nucleotide-Sequence of the *His4* Region of Yeast." *Gene* **18**(1): 47-59.
- D'Ordine, R. L., T. J. Klem and V. J. Davisson (1999). "N-1-(5'-phosphoribosyl)adenosine-5'-monophosphate cyclohydrolase: Purification and characterization of a unique metalloenzyme." *Biochemistry* **38**(5): 1537-1546.
- Eads, J. C., D. Ozturk, *et al.* (1997). "A new function for a common fold: The crystal structure of quinolinic acid phosphoribosyltransferase." *Structure* **5**(1): 47-58.
- Evans, S. V. (1993). "Setor - Hardware-Lighted 3-Dimensional Solid Model Representations of Macromolecules." *Journal of Molecular Graphics* **11**(2): 134-138.
- Falquet, L., M. Pagni, *et al.* (2002). "The PROSITE database, its status in 2002." *Nucleic Acids Research* **30**(1): 235-238.
- Fani, R., P. Alifano, *et al.* (1993). "The Histidine Operon of *Azospirillum brasilense* - Organization, Nucleotide-Sequence and Functional-Analysis." *Research in Microbiology* **144**(3): 187-200.
- Fani, R., P. Lio, I. Chiarelli and M. Bazzicalupo (1994). "The Evolution of the Histidine Biosynthetic Genes in Prokaryotes - a Common Ancestor for the *hisA* and *hisF* Genes." *Journal of Molecular Evolution* **38**(5): 489-495.

- Fani, R., P. Lio and A. Lazcano (1995). "Molecular evolution of the histidine biosynthetic pathway." *Journal of Molecular Evolution* **41**(6): 760-774.
- Fani, R., E. Tamburini, *et al.* (1997). "Paralogous histidine biosynthetic genes: evolutionary analysis of the *Saccharomyces cerevisiae* HIS6 and HIS7 genes." *Gene* **197**(1-2): 9-17.
- Ferré-D'Amaré, A. and S. K. Burley (1997). Dynamic Light Scattering in Evaluating Crystallizability of Macromolecules. *Methods in Enzymology*. C. W. J. Carter and R. M. Sweet. New York, Academic Press. **276**: 157-166.
- Foadi, J., M. M. Woolfson, *et al.* (2000). "A flexible and efficient procedure for the solution and phase refinement of protein structures." *Acta Crystallographica Section D-Biological Crystallography* **56**: 1137-1147.
- Focia, P. J., S. P. Craig, *et al.* (1998). "A 1.4 Å crystal structure for the hypoxanthine phosphoribosyltransferase of *Trypanosoma cruzi*." *Biochemistry* **37**(43): 15066-15075.
- French, G. S. and K. S. Wilson (1978). "On the treatment of negative intensity observations." *Acta Crystallographica Section A* **34**: 517-525.
- Früh, H. and T. Leisinger (1981). "Properties and localisation of *N*-acetylglutamate deacetylase from *Pseudomonas aeruginosa*." *Journal of General Microbiology* **125**: 1-10.
- Fujimori, K. and D. Ohta (1998). "Isolation and characterization of a histidine biosynthetic gene in *Arabidopsis* encoding a polypeptide with two separate domains for phosphoribosyl-ATP pyrophosphohydrolase and phosphoribosyl-AMP cyclohydrolase." *Plant Physiology* **118**(1): 275-283.
- Galperin, M. Y. and E. V. Koonin (1997). "Sequence analysis of an exceptionally conserved operon suggests enzymes for a new link between histidine and purine biosynthesis." *Molecular Microbiology* **24**(2): 443-445.
- Gilbert, D., D. Westhead, N. Nagano and J. Thornton (1999). "Motif-based searching in TOPS protein topology databases." *Bioinformatics* **15**(4): 317-326.
- Gohda, K., D. Ohta, *et al.* (2001). "Identification of novel potent inhibitors for ATP-phosphoribosyl transferase using three-dimensional structural database search technique." *Quantitative Structure-Activity Relationships* **20**(2): 143-147.

- Goitein, R. K., D. Chelsky and S. M. Parsons (1978). "Primary ^{14}C and alpha secondary ^3H substrate kinetic isotope effects for some phosphoribosyltransferases." *Journal of Biological Chemistry* **253**(9): 2963-2971.
- Goitein, R. K. and S. M. Parsons (1978). "Rapid separation of 5-phospho- α -D-ribose-1-diphosphate from its metabolic derivatives." *Analytical Biochemistry* **87**(2): 636-40.
- Goldberg, R. B. and B. Magasanik (1975). "Gene order of the histidine utilisation (*hut*) operon in *Klebsiella aerogenes*." *Journal of Bacteriology* **122**: 1025-1031.
- Golomb, M. and M. Chamberlin (1974). "Characterisation of T7 specific ribonucleic acid polymerase IV. Resolution of the major in vitro transcripts by gel electrophoresis." *Journal of Biological Chemistry* **249**: 2858-2863.
- Gouet, P., Courcelle, E., Stuart, D.I. and Metz, F. (1999). "ESPrict: multiple sequence alignments in PostScript." *Bioinformatics* **15**: 305-308.
- Hanahan, D. (1985). Techniques for Transformation of *E. coli*. *DNA cloning*. D. M. Glover. Oxford, Washington DC, IRL Press. **1**: 109-135.
- Hartman, P. E., Z. Hartman and R. C. Stahl (1971). "Classification and mapping of spontaneous and induced mutations in the histidine operon of Salmonella." *Advances in Genetics* **16**: 1-34.
- Holm, L. and C. Sander (1995). "Dali - a Network Tool for Protein-Structure Comparison." *Trends in Biochemical Sciences* **20**(11): 478-480.
- Hu, L., S. L. Allison and A. T. Phillips (1989). "Identification of Multiple Repressor Recognition Sites in the *hut* System of *Pseudomonas putida*." *Journal of Bacteriology* **171**(8): 4189-4195.
- Hu, L., L. M. Mulfinger and A. T. Phillips (1987). "Purification and Properties of Formylglutamate Amidohydrolase from *Pseudomonas putida*." *Journal of Bacteriology* **169**(10): 4696-4702.
- Hungerer, C., D. S. Weiss, R. K. Thauer and D. Jahn (1996). "The hemA gene encoding glutamyl-tRNA reductase from the archaeon Methanobacterium thermoautotrophicum strain marburg." *Bioorganic & Medicinal Chemistry* **4**(7): 1089-1095.

- Jancarik, J. and S. H. Kim (1991). "Sparse-Matrix Sampling - a Screening Method for Crystallization of Proteins." *Journal of Applied Crystallography* **24**: 409-411.
- Jiang, P., P. Zucker, *et al.* (1997). "Structure/function analysis of the PII signal transduction protein of *Escherichia coli*: Genetic separation of interactions with protein receptors." *Journal of Bacteriology* **179**(13): 4342-4353.
- Kaneko, T., T. Matsubayashi, M. Sugita and M. Sugiura (1996). "Physical and gene maps of the unicellular cyanobacterium *Synechococcus* sp strain PCC6301 genome." *Plant Molecular Biology* **31**(1): 193-201.
- Kimhi, Y. and B. Magasanik (1970). "Genetic Basis of Histidine Degradation in *Bacillus subtilis*." *Journal of Biological Chemistry* **245**(14): 3545-3548.
- Kissinger, C. R., D. K. Gehlhaar and D. B. Fogel (1999). "Rapid automated molecular replacement by evolutionary search." *Acta Crystallographica Section D-Biological Crystallography* **55**: 484-491.
- Kleman, J. and S. M. Parsons (1975). "A sensitive assay for the reverse reaction of the first histidine biosynthetic enzyme." *Analytical Biochemistry* **68**(1): 236-241.
- Kleman, J. E. and S. M. Parsons (1976). "Reverse direction substrate kinetics and inhibition studies on the first enzyme of histidine biosynthesis, adenosine triphosphate phosphoribosyltransferase." *Archives of Biochemistry & Biophysics* **175**(2): 687-693.
- Klungsoyr, L. and D. E. Atkinson (1970). "Regulatory properties of phosphoribosyladenosine triphosphate synthetase. Synergism between adenosine monophosphate, phosphoribosyladenosine triphosphate, and histidine." *Biochemistry* **9**(9): 2021-2027.
- Klungsoyr, L. and H. Kryvi (1971). "Sedimentation behaviour of phosphoribosyladenosine triphosphate synthetase. Effects of substrates and modifiers." *Biochimica et Biophysica Acta* **227**(2): 327-336.
- Kovach, J. S., J. M. Phang, *et al.* (1970). "Interaction between histidyl transfer ribonucleic acid and the first enzyme for histidine biosynthesis of *Salmonella typhimurium*." *Journal of Bacteriology* **104**(2): 787-792.
- Kronenberg, H. M., T. Vogel and R. F. Goldberger (1975). "A new and highly sensitive assay for the ATP phosphoribosyltransferase that catalyzes the first step of histidine biosynthesis." *Analytical Biochemistry* **65**(1-2): 380-388.

- Kryvi, H. (1973). "Thermal stability of phosphoribosyladenosine triphosphate synthetase as reflected in its circular dichroism and activity properties. Effect of inhibitors." *Biochimica et Biophysica Acta* **317**(1): 123-130.
- Kryvi, H. and L. Klungsoyr (1971). "Kinetic properties of phosphoribosyladenosine triphosphate synthetase. Inhibition by aggregation at high enzyme concentrations." *Biochimica et Biophysica Acta* **235**(3): 429-434.
- Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." *Nature* **227**: 680-685.
- Lang, D., R. Thoma, *et al.* (2000). "Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion." *Science* **289**(5484): 1546-1550.
- Laskowski, R. A., M. W. Macarthur, D. S. Moss and J. M. Thornton (1993). "Procheck - a Program to Check the Stereochemical Quality of Protein Structures." *Journal of Applied Crystallography* **26**(Pt2): 283-291.
- Lazcano, A., G. E. Fox and J. F. Oro (1992). Life Before DNA: The Origin and Evolution of Early Archean Cells. *The Evolution of Metabolic Function*. R. P. Mortlock, CRC Press: 237-295.
- Lemke, C. T., G. D. Smith and P. L. Howell (2002). "S-SAD, Se-SAD, and S/Se-SIRAS using Cu $K\alpha$ radiation: why wait for synchrotron time." *Acta Crystallographica Section D-Biological Crystallography* **58**: 2096-2101.
- Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**.
- Lo Conte, L., B. Ailey, *et al.* (2000). "SCOP: a Structural Classification of Proteins database." *Nucleic Acids Research* **28**(1): 257-259.
- Lohkamp, B., J. R. Coggins and A. J. Laphorn (2000). "Purification, crystallization and preliminary X-ray crystallographic analysis of ATP-phosphoribosyltransferase from *Escherichia coli*." *Acta Crystallographica. Section D: Biological Crystallography* **56**(Pt 11): 1488-1491.
- Martin, R. G. (1963). "The First Enzyme in Histidine Biosynthesis: The Nature of Feedback Inhibition by Histidine." *Journal of Biological Chemistry* **238**(1): 257-268.

- Martin, R. G., M. A. Berberich, *et al.* (1971). "Enzymes and Intermediates of Histidine Biosynthesis in *Salmonella typhimurium*." *Methods in Enzymology* **17**(B): 3-44.
- Matthews, B. W. (1968). "Solvent content of protein crystals." *Journal of Molecular Biology* **33**: 491-497.
- Mayans, O., A. Ivens, *et al.* (2002). "Structural analysis of two enzymes catalysing reverse metabolic reactions implies common ancestry." *EMBO Journal* **21**(13): 3245-3254.
- Meyers, M., F. Blasi, *et al.* (1975). "Specific binding of the first enzyme for histidine biosynthesis to the DNA of histidine operon." *Nucleic Acids Research* **2**(11): 2021-2036.
- Miller, R., S. M. Gallo, H. G. Khalak and C. M. Weeks (1994). "Snb - Crystal-Structure Determination Via Shake-and-Bake." *Journal of Applied Crystallography* **27**: 613-621.
- Miroux, B. and J. E. Walker (1996). "Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels." *Journal of Molecular Biology* **260**(3): 289-298.
- Moffatt, B. A. and F. W. Studier (1987). "T7 lysozyme inhibits transcription by T7 RNA polymerase." *Cell* **49**: 221-227.
- Mori, I., R. Fonnepfeister, *et al.* (1995). "A Novel Class of Herbicides - Specific Inhibitors of Imidazoleglycerol Phosphate Dehydratase." *Plant Physiology* **107**(3): 719-723.
- Morton, D. P. and S. M. Parsons (1976). "Biosynthetic direction substrate kinetics and product inhibition studies on the first enzyme of histidine biosynthesis, adenosine triphosphate phosphoribosyltransferase." *Archives of Biochemistry & Biophysics* **175**(2): 677-286.
- Morton, D. P. and S. M. Parsons (1977). "Inhibition of ATP phosphoribosyltransferase by AMP and ADP in the absence and presence of histidine." *Archives of Biochemistry & Biophysics* **181**(2): 643-648.
- Mousdale, D. M. and J. R. Coggins (1991). Amino acid synthesis. *Target Sites for Herbicide Action*. R. C. Kirkwood. New York, Plenum Press: 29-56.

- Murshudov, G. N., Dodson, E. J. & Vagin, A. A. (1996). Application of maximum likelihood methods for macromolecular refinement. *Macromolecular Refinement*. E. Dodson, Moore, M. & Bailey, S. Warrington, SERC Daresbury Laboratory: 93-104.
- Musick, W. D. L. (1981). "Structural Features of the Phosphoribosyltransferases and Their Relationship to the Human Deficiency Disorders of Purine and Pyrimidine Metabolism." *CRC Critical Reviews in Biochemistry* **11**(1): 1-34.
- Navaza, J. (1994). "AMoRe - an Automated Package for Molecular Replacement." *Acta Crystallographica Section A* **50**: 157-163.
- Nicholls, A., R. Bharadwaj and B. Honig (1993). "Grasp - Graphical Representation and Analysis of Surface- Properties." *Biophysical Journal* **64**(2): A166-A166.
- Ninfa, A. J. and M. R. Atkinson (2000). "PII signal transduction proteins." *Trends in Microbiology* **8**(4): 172-179.
- Oriol, E., S. MendezAlvarez, J. Barbe and I. Gibert (1996). "Cloning of the *Rhodobacter sphaeroides hisI* gene: Unifunctionality of the encoded protein and lack of linkage to other his genes." *Microbiology-Uk* **142**: 2071-2078.
- Oriol, E., S. MendezAlvarez, J. Barbe and I. Gibert (1997). "Cloning of the *Rhodobacter sphaeroides hisI* gene: unifunctionality of the encoded protein and lack of linkage to other his genes (vol 142, pg 2071, 1996)." *Microbiology-Uk* **143**: 1779-1779.
- Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. *Isomorphous Replacement and Anomalous Scattering*. W. Wolf, P. R. Evans and I. A.G.W, SERC Daresbury Laboratory, Warrington, UK.: 80-85.
- Otwinowski, Z. and W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." *Methods in Enzymology* **276**: 307-326.
- Parsons, S. M. and D. E. Koshland, Jr. (1974). "Multiple aggregation states of phosphoribosyladenosine triphosphate synthetase." *Journal of Biological Chemistry* **249**(13): 4119-4126.
- Pflugrath, J. W. (1999). "The finer things in X-ray diffraction data collection." *Acta Crystallographica Section D-Biological Crystallography* **55**: 1718-1725.
- Philippesen, A. DINO: Visualizing Structural Biology. <http://www.dino3d.org>.

- Read, R. J. (2001). "Pushing the boundaries of molecular replacement with maximum likelihood." *Acta Crystallographica Section D-Biological Crystallography* **57**: 1373-1382.
- Rees, D. C. and M. W. W. Adams (1995). "Hyperthermophiles: taking the heat and loving it." *Structure* **3**: 251-254.
- Renault, P., J. J. Godon, *et al.* (1995). "Metabolic operons in *Lactococci*." *Developments in Biological Standardization* **85**: 431-441.
- Sambrook, J., E. F. Fritsch and T. Maniatis (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.
- Scapin, G., D. H. Ozturk, C. Grubmeyer and J. C. Sacchettini (1995). "The Crystal-Structure of the Orotate Phosphoribosyltransferase Complexed with Orotate and α -D-5-Phosphoribosyl-1- Pyrophosphate." *Biochemistry* **34**(34): 10744-10754.
- Schuller, D. J., G. A. Grant and L. J. Banaszak (1995). "The Allosteric Ligand Site in the Vmax-Type Cooperative Enzyme Phosphoglycerate Dehydrogenase." *Nature Structural Biology* **2**(1): 69-76.
- Scott, J. F., J. R. Roth and S. W. Artz (1975). "Regulation of histidine operon does not require HisG enzyme." *Proceedings of the National Academy of Sciences of the United States of America* **72**(12): 5021-5025.
- Shaner, D. L. (1989). Sites of action of herbicides in amino acid metabolism: primary and secondary physiological effects. *Plant Nitrogen Metabolism*. J. E. Poulton, J. T. Romeo and E. E. Conn. New York, Academic Press. **23**.
- Sheldrick, G. M. (1997). Patterson superposition and ab initio phasing. *Macromolecular Crystallography, Pt A*. **276**: 628-641.
- Sinha, S. C. and J. L. Smith (2001). "The PRT protein family." *Current Opinion in Structural Biology* **11**(6): 733-739.
- Sissler, M., C. Delorme, *et al.* (1999). "An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis." *Proceedings of the National Academy of Sciences of the United States of America* **96**(16): 8985-8990.
- Smith, D. W. E. and B. N. Ames (1964). "Intermediates in the Early Steps of Histidine Biosynthesis." *Journal of Biological Chemistry* **239**(6): 1848-1855.

- Smith, D. W. E. and B. N. Ames (1965). "Phosphoribosyladenosine Monophosphate, an Intermediate in Histidine Biosynthesis." *Journal of Biological Chemistry* **240**(7): 3056-3063.
- Smith, G. R. and B. Magasanik (1971). "The two operons of the histidine utilisation system in *Salmonella typhimurium*." *Journal of Biological Chemistry* **246**: 3330-3341.
- Smith, J. L., E. J. Zaluzec, *et al.* (1994). "Structure of the Allosteric Regulatory Enzyme of Purine Biosynthesis." *Science* **264**(5164): 1427-1433.
- Smith, O., M. M. Meyers, *et al.* (1974). "Defective in vitro binding of histidyl-transfer ribonucleic acid to feedback resistant phosphoribosyl transferase of *Salmonella typhimurium*." *Nucleic Acids Research* **1**(7): 881-888.
- Soisson, S. M., B. MacDougall-Shackleton, R. Schleif and C. Wolberger (1997). "Structural Basis for Ligand-Regulated Oligomerization of AraC." *Science* **276**(5311): 421-425.
- Sterboul, C. C., J. E. Kleeman and S. M. Parsons (1977). "Purification and characterization of a mutant ATP phosphoribosyltransferase hypersensitive to histidine feedback inhibition." *Archives of Biochemistry & Biophysics* **181**(2): 632-642.
- Stryer, L. (1988). Amino Acid Degradation and the Urea Cycle. *Biochemistry*. New York, W.H. Freeman and Company.
- Studier, F. W. and B. A. Moffatt (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes." *Journal of Molecular Biology* **189**: 113-130.
- Tebar, A. R. and A. O. Ballesteros (1976). "Kinetic properties of ATP phosphoribosyltransferase of *Escherichia coli*." *Molecular and Cellular Biochemistry* **11**(3): 131-136.
- Tebar, A. R., V. M. Fernandez, R. Martin Del Rio and A. O. Ballesteros (1973). "Studies on the quaternary structure of the first enzyme for histidine biosynthesis." *Experientia* **29**(12): 1477-1479.
- Terwilliger, T. C. and J. Berendzen (1999). "Automated MAD and MIR structure solution." *Acta Crystallographica Section D-Biological Crystallography* **55**: 849-861.

- Thoma, R., G. Obmolova, *et al.* (1999). "Efficient expression, purification and crystallisation of two hyperthermostable enzymes of histidine biosynthesis." *FEBS Letters* **454**(1-2): 1-6.
- Thoma, R., M. Schwander, *et al.* (1998). "A histidine gene cluster of the hyperthermophile *Thermotoga maritima*: sequence analysis and evolutionary significance." *Extremophiles* **2**(4): 379-389.
- Tomchick, D. R., R. J. Turner, R. L. Switzer and J. L. Smith (1998). "Adaptation of an enzyme to regulatory function: structure of *Bacillus subtilis* PyrR, a pyr RNA-binding attenuation protein and uracil phosphoribosyltransferase." *Structure* **6**(3): 337-350.
- Vagin, A. A. and A. Teplyakov (1997). "MOLREP: an automated program for molecular replacement." *Journal of Applied Crystallography* **30**: 1022-1025.
- Voll, M. J., E. Appella and R. G. Martin (1967). "Purification and composition studies of phosphoribosyladenosine triphosphate:pyrophosphate phosphoribosyltransferase, the first enzyme of histidine biosynthesis." *Journal of Biological Chemistry* **242**(8): 1760-1767.
- Wainscott, V. J. and J. J. Ferretti (1978). "Biochemical-genetic study of the first enzyme of histidine biosynthesis in *Salmonella typhimurium*: substrate and feedback binding regions." *Journal of Bacteriology* **133**(1): 114-121.
- White, H. B. (1976). "Coenzymes as fossils of an earlier metabolic state." *Journal of Molecular Evolution* **7**: 101-117.
- White, R. H. (1997). "Occurrence and biosynthesis of 5-aminoimidazole-4-carboxamide ribonucleotide and N-(beta-D-ribofuranosyl)formamide 5'- phosphate in *Methanobacterium thermoautotrophicum* Δ H." *Journal of Bacteriology* **179**(2): 563-566.
- Wilkinson, K. W., P. J. Baker, *et al.* (1995). "Crystallization and Analysis of the Subunit Assembly and Quaternary Structure of Imidazoleglycerol Phosphate Dehydratase from *Saccharomyces cerevisiae*." *Acta Crystallographica Section D-Biological Crystallography* **51**: 845-847.

Wilmanns, M. and D. Eisenberg (1993). "3-Dimensional Profiles from Residue-Pair Preferences - Identification of Sequences with β/α -Barrel Fold." *Proceedings of the National Academy of Sciences of the United States of America* **90**(4): 1379-1383.

Winkler, M. E. (1987). Biosynthesis of Histidine. *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. F. C. Neidhardt. Washington D.C., American Society for Microbiology. 1.

Winn, M. D., M. N. Isupov and G. N. Murshudov (2001). "Use of TLS parameters to model anisotropic displacements in macromolecular refinement." *Acta Crystallographica Section D-Biological Crystallography* **57**: 122-133.

Xu, Y. B., E. Cheah, *et al.* (1998). "GlnK, a P_{II} -homologue: Structure reveals ATP binding site and indicates how the T-loops may be involved in molecular recognition." *Journal of Molecular Biology* **282**(1): 149-165.

Bernhard Lohkamp,^a John R.
Coggins^a and Adrian J.
Lapthorn^{b*}^aDivision of Biochemistry and Molecular
Biology, Institute of Biomedical and Life
Sciences, University of Glasgow, Glasgow
G12 8QQ, Scotland, and ^bDepartment of
Chemistry, University of Glasgow, Glasgow
G12 8QQ, Scotland

Correspondence e-mail: adrian@chem.gla.ac.uk

Purification, crystallization and preliminary X-ray
crystallographic analysis of ATP-phosphoribosyl-
transferase from *Escherichia coli*Received 5 May 2000
Accepted 11 August 2000

ATP-phosphoribosyltransferase (ATP-PRT) from *Escherichia coli* has been purified and crystals were obtained by the vapour-diffusion method using sodium tartrate as a precipitant. Dynamic light scattering was used to assess conditions for the monodispersity of the enzyme. The crystals are trigonal, space group *R*32, with unit-cell parameters $a = b = 133.6$, $c = 114.1$ Å (at 100 K), and diffract to 2.7 Å on a synchrotron X-ray source. The asymmetric unit is likely to contain one molecule, corresponding to a packing density of 2.9 Å³ Da⁻¹. A model for the quaternary structure is proposed based on the crystallographic symmetry.

1. Introduction

In microorganisms and plants, histidine is synthesized in ten steps by the enzymes of the histidine pathway. Eight enzymes are encoded on eight genes (*hisG(IE)AHFBCD*) often arranged as one operon (*his* operon). Three of the enzymes are bifunctional (*hisIE*, *hisD*, *hisB*) and two form a multienzyme complex catalysing one reaction (*hisH*, *hisF*) (see, for example, Alifano *et al.*, 1996). The pathway leads from 5-phosphoribosyl 1-pyrophosphate (PRPP) and ATP to histidine.

The initial substrates of the histidine pathway, PRPP and ATP, play a key role in intermediary and energy metabolism. Furthermore, they link the histidine pathway to the biosynthesis of folates, purines, pyrimidines, pyridine nucleotides and tryptophan.

The first enzyme of the pathway, ATP-phosphoribosyltransferase (ATP-PRT; E.C. 2.4.2.17) encoded by the *hisG* gene, catalyses the condensation of ATP and PRPP to form *N*'-5'-phosphoribosyl-ATP (PRATP). In addition to its catalytic function, it has been shown to be involved in several aspects of the regulation of histidine biosynthesis (Ames *et al.*, 1961). ATP-PRT binds to the *his* operon and to histidyl-tRNA and is feedback inhibited by the end-product of the pathway, histidine (Voll *et al.*, 1967). Owing to these regulatory functions and the connection of the histidine pathway to other biosynthetic and metabolic pathways, ATP-PRT plays an important part in the regulation of growth.

From the primary structure of ATP-PRT, the molecular weight is deduced as 33 367 Da for the 299 amino acids. Experimentally determined molecular weights in the literature describe various oligomeric states of ATP-PRT. Ultracentrifugation and gel separation

have shown that the enzyme is in equilibrium between its active dimeric form, inactive hexameric form and higher aggregates (Klungsoyr & Kryvi, 1971; Tebar *et al.*, 1973). The hexameric form is stabilized by histidine, AMP, PRATP, high concentrations of ATP and high enzyme concentration (Klungsoyr & Kryvi, 1971; Kryvi & Klungsoyr, 1971). The second substrate PRPP seems to dissociate higher aggregates, resulting in an increase in the concentration of dimers (Tebar *et al.*, 1973).

The secondary structure of ATP-PRT has been investigated using CD spectroscopy (Kryvi, 1973). At that time it was proposed to consist of about 33% α -helix and 20–30% β -sheet, with no change observed when the inhibitors AMP and/or histidine were bound. Furthermore, it was proposed that the first 120 residues of ATP-PRT contain a nucleotide-binding fold arranged in a $\beta\alpha\beta\alpha\beta\beta\alpha\beta$ pattern (Argos *et al.*, 1983). To our knowledge, no comparison of ATP-PRT with other phosphoribosyltransferases (PRT) of which structures are now known has been made (see, for example, Focia *et al.*, 1998).

In common with a number of amino-acid biosynthetic pathways, the histidine pathway is present in bacteria, fungi and plants but absent in mammals. This makes it a good target for the development of antimicrobial and antifungal drugs as well as herbicides (Mousdale & Coggins, 1991). A variety of selective inhibitors for the histidine pathway are known and are used commercially as herbicides (Mori *et al.*, 1995). These herbicidal compounds all inhibit imidazoleglycerol-phosphate dehydratase, the product of the *hisB* gene. Knowledge of the structure of ATP-PRT will help with the understanding of its function and will provide a basis for the development of antimicrobial drugs and herbicides.

Table 1
Dynamic light-scattering measurements of *E. coli* ATP-PRT in the presence of different additives.

The hydrodynamic radius represents the median particle size present in the sample cell. The polydispersity value indicates the standard deviation of the spread of particle sizes about the reported average radius. A rise in polydispersity in relation to the average radius represents greater spread in the size distribution. The estimated M_w is calculated from the hydrodynamic radius (R_H) using an empirically derived relationship between the R_H and M_w values for a number of well characterized globular proteins in buffered solution. All experiments were conducted in 50 mM Tris-HCl buffer at pH 7.5 in the presence of 100 mM KCl unless otherwise stated. Samples containing PRPP and ATP were incubated at room temperature for 10 min to form PRATP.

Additives	Hydrodynamic radius (nm)	Polydispersity (nm)	Estimated M_w (kDa)	Conclusion
None (no KCl)	7.3	0.3	369	Polydisperse
None	6.9	1.0	321	Polydisperse
AMP (0.4 mM)	5.5	0.7	186	Polydisperse
Histidine (5 mM)	5.7	0.9	197	Polydisperse
Histidine (5 mM) + AMP (1 mM)	5.3	1.0	170	Polydisperse
PRPP (1 mM) + ATP (5 mM) + MgCl ₂ (10 mM)	5.5	1.5	184	Polydisperse
PRPP (1 mM) + ATP (5 mM) + MgCl ₂ (10 mM) + histidine (5 mM)	5.6	1.6	190	Polydisperse

2. Enzyme purification

ATP-PRPP from *E. coli* was overexpressed in *E. coli* B834(DE3)pLysS to 30–40% of the total cellular protein using a T7 RNA polymerase expression system (Studier & Moffatt, 1986). Cultures were grown in LB medium at 315 K to an optical density of 0.6 at 600 nm. Overexpression of protein was induced by addition of IPTG to a final concentration of 0.4 mM. Bacterial growth was continued for 4 h prior to harvesting the cells.

The cells (5.3 g wet weight) suspended in 20 ml of 50 mM Tris-HCl pH 7.5 containing 0.4 mM DTT and two protease inhibitor tablets (EDTA-free; Boehringer-Mannheim) per 100 ml (buffer A) were broken by two passes through a French pressure cell at 6.5 MPa. This material was then diluted with 40 ml buffer A and centrifuged at 18 000 rev min⁻¹ for 1 h. ATP-PRT was purified from the resulting cell-free extract. All subsequent steps were performed at 277 K.

The supernatant was applied to a DEAE-Sephacel anion-exchange column (21 × 14 cm diameter, flow rate 60 ml h⁻¹) equilibrated in 50 mM Tris-HCl pH 7.5 containing 0.4 mM DTT and one protease inhibitor tablet per litre (buffer B). The column was then washed with the same buffer until the A_{280nm} of the eluate was less than 0.3. ATP-PRT was eluted with a linear gradient of 0–500 mM KCl in 700 ml buffer B (flow rate 30 ml h⁻¹) and 10 ml fractions collected and assayed as described in Martin *et al.* (1971). Fractions containing ATP-PRT activity were pooled and dialysed overnight against buffer B.

The protein was loaded on a dye column (Reactive Green 19, Sigma; 8.5 × 2.1 cm diameter, flow rate 40 ml h⁻¹) equilibrated in buffer B. The column was then washed with 50 ml buffer B and ATP-PRT was eluted with a linear gradient (250 ml) of 200–950 mM KCl in buffer B. The flow rate was 40 ml h⁻¹ and 10 ml fractions were collected and assayed. Fractions containing ATP-PRT activity were pooled, dialysed overnight against buffer B and concentrated by ultrafiltration. At this stage, the enzyme precipitated at high concentrations, but no substantial amount of enzyme was lost owing to its irreversible precipitation.

The concentrated enzyme solution was applied to a Sephacryl S200 (superfine grade) column (170 × 1.5 cm diameter, flow rate 15 ml h⁻¹) that had been equilibrated with buffer B containing 500 mM KCl. The enzyme was eluted with buffer B containing 500 mM KCl (flow rate 10 ml h⁻¹, 10 ml fractions). Fractions containing ATP-PRT activity were dialysed against buffer B containing 50% (v/v) glycerol prior to long-term storage at 253 K.

3. Dynamic light scattering

Experiments were carried out using a DYNA-PRO 801 dynamic light-scattering/molecular-sizing instrument (Protein Solutions, Buckinghamshire, England). Protein solutions (1 mg ml⁻¹) were in 50 mM Tris-HCl pH 7.5, but several other buffers and pHs were also assessed. Dynamic light-scattering experiments were carried out to determine which combination of product and cofactor were appropriate to obtain a

monodisperse enzyme solution. Monodispersity is not an infallible indicator that diffraction-quality crystals can be obtained from a protein; instead, it reflects the observation that monodisperse protein solutions have a much higher probability of producing crystals than those which are not (Ferré-D'Amaré & Burley, 1997). Dynamic light-scattering experiments on ATP-PRT were performed using several conditions and in the presence of different inhibitors (Table 1). It was shown that monodispersity could not be achieved in the presence of any of the probed additives. The lowest value for polydispersity was obtained using 0.4 mM AMP, with a resultant molecular weight estimated as 186 000 Da. This is about 10% lower than the M_w of six monomers, but consistent with the expected hexameric state of the enzyme. Other additives, such as histidine and PRATP, gave similar molecular weights and hydrodynamic radii but higher polydispersity. Furthermore, the solubility of the enzyme was found to be greatly increased in the presence of AMP (from about 5 mg ml⁻¹ in buffer to at least 20 mg ml⁻¹). In the light of these results, crystallization trials were carried out in the presence and absence of AMP.

4. Crystallization

The enzyme was stored in 50% (v/v) glycerol, which was removed by exhaustive dialysis against 50 mM Tris-HCl pH 7.5 containing 0.4 mM DTT. The enzyme was concentrated to 11 mg ml⁻¹ in 50 mM Tris-HCl pH 7.5, 0.4 mM DTT and 2 mM AMP using Centricon-10 centrifugal concentrators (Amicon, Stonehouse, Gloucestershire, England).

Crystallization experiments were performed at 293 K using the sitting-drop vapour-diffusion technique. An exhaustive

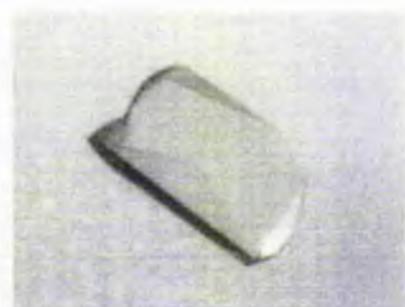


Figure 1
A crystal of ATP-PRT from *E. coli*. The crystal is approximately 0.3 × 0.2 × 0.2 mm in size (crystal form I).

Table 2
Statistics for the X-ray data processed in R32.

d_{min} (Å)	R_{sym} (I)	(I/σ)	No. of unique reflections	Completeness (%)	Multiplicity (%)
5.82	0.041	38.2	1099	95.6	6.3
4.62	0.061	39.1	1104	99.6	6.6
4.03	0.055	36.7	1098	99.9	6.7
3.66	0.050	32.2	1094	100.0	6.7
3.40	0.071	23.6	1079	100.0	6.7
3.20	0.120	14.5	1075	100.0	6.7
3.04	0.180	10.4	1092	100.0	6.7
2.91	0.293	6.4	1076	100.0	6.7
2.80	0.469	4.3	1077	100.0	6.7
2.70	0.668	3.1	1066	100.0	6.7
Total	0.049	18.0	10860	99.5	6.7

set of conditions, comprising commercially available and local sparse-matrix screens (Jancarik & Kim, 1991; Cudney *et al.*, 1994), were tried. Initial crystals were obtained from various conditions, mostly in the presence of AMP. Optimization of these crystallization conditions was most successful for a single condition which led to the final conditions of 1.3 M sodium tartrate, 50–200 mM MgCl₂, 100 mM citrate buffer pH 5.6 and protein in the presence of 2 mM AMP (crystal form I). Typically, 1–2 µl of protein (10 mg ml⁻¹) was mixed with an equal volume of the reservoir solution. Crystals appeared after 3–4 d and continued to grow as trigonal prisms to maximum dimensions of 0.3 × 0.2 × 0.2 mm (Fig. 1). A second crystal form was derived using 1.36–1.44 M ammonium sulfate, 0–0.3 M sodium chloride, 100 mM HEPES buffer pH 7.5 and protein in the presence of 2 mM AMP. Round-shaped crystals appeared after 4–5 d and grew to maximum dimensions of about 0.2 × 0.2 × 0.2 mm (crystal form II).

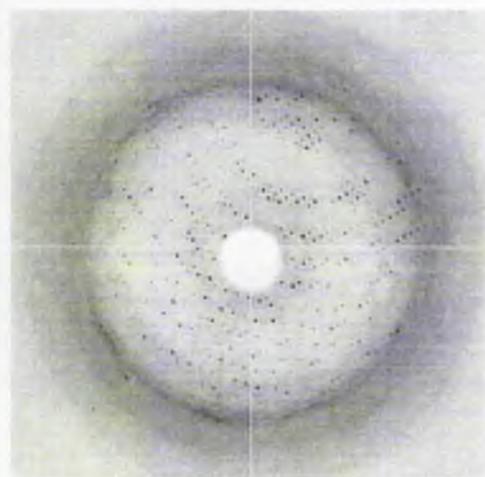


Figure 2
A typical diffraction pattern from ATP-PRT collected on beamline 9.6 at Daresbury synchrotron.

5. X-ray analysis

The X-ray diffraction data were collected on beamline 9.6 at the SRS, Daresbury Laboratory using a wavelength of 0.87 Å and a ADSC Quantum-4 CCD detector. The crystals were radiation sensitive and therefore cryo-cooling was essential. Crystals were loop-mounted in a cryoprotectant containing 15% (v/v) glycerol and cryo-cooled to 100 K using an Oxford Cryosystems Cryostream. A native data set to 2.7 Å resolution was collected using 1° oscillation frames from crystal form I (Fig. 2). Data were processed with DENZO and scaled with SCALEPACK (Otwinowski & Minor, 1997). From auto-indexing in DENZO, the crystals were found to index in a primitive rhombohedral lattice corresponding to space group R3, with unit-cell parameters $a = b = 133.6$, $c = 114.1$ Å. Indexing with the higher symmetry space group R32 resulted in comparable merging R-factor statistics (Table 2). For R32, only one monomer per asymmetric unit leads to an acceptable packing density V_M of 2.95 Å³ Da⁻¹, which corresponds to a solvent content of 58% (Matthews, 1968).

Crystal form II diffracted to only 3.7 Å; from a single frame, processing was performed as described above. The crystal symmetry was determined to be primitive rhombohedral, with unit-cell parameters $a = b = 132.6$, $c = 113.8$ Å. Because of the similar unit-cell parameters and the identical lattice, we presume that crystal form II also belongs to space group R32.

So far there are two different types of PRT structures known and another one predicted. Several reported structures, e.g. uracil PRT, show the type I fold of a five-stranded β-sheet surrounded by three or four α-helices (and a hood domain) (see, for example, Tomchick *et al.*, 1998). Quinolinate PRT comprises a type II fold with two domains: a mixed α/β N-terminal domain and an βα₆ barrel C-terminal domain, similar to known (βα)₈ barrels (Eads *et al.*, 1997). Anthranilate PRT is predicted to give a typical (βα)₈ barrel fold (Wilmanns & Eisenberg, 1993). Since there is no similarity in sequence or secondary-structure prediction of

ATP-PRT with these PRTs or other known protein structures, we intend to solve the structure by multiwavelength anomalous dispersion (MAD).

6. Quaternary structure model of ATP-PRT

A quaternary structure model of ATP-PRT based on the crystallographic data can be proposed if we assume that the enzyme in the crystal is in a hexameric arrangement (in the presence of the inhibitor AMP) as shown by light scattering (this report), gel separation (Tebar *et al.*, 1973) and ultracentrifugation studies (Klungsoyr & Kryvi, 1971). Taking into account one molecule per asymmetric unit, the symmetry in the hexamer has to be in accordance to the crystallographic symmetry of space group R32. Therefore, the twofold and threefold symmetry in this space group must be reflected by the quaternary structure model. The proposed model can be seen in Fig. 3.

Crystal structures of all type I PRTs show a dimeric arrangement (Tomchick *et al.*,

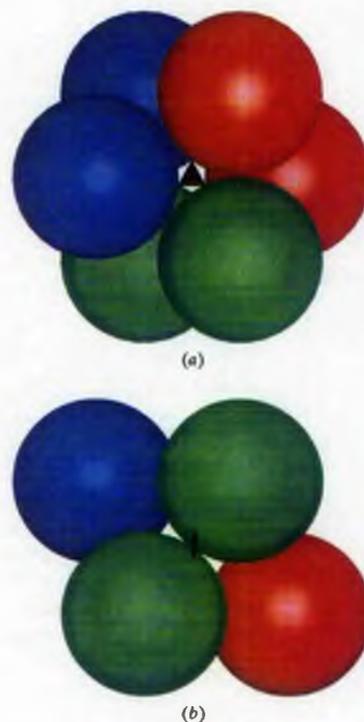


Figure 3
Proposed model of quaternary structure of ATP-PRT. (a) shows a view of the hexameric arrangement down the threefold axis. (b) is the view along the twofold axis, which is (a) rotated 90° towards the viewer. Dimers are indicated by the same colour. Threefold symmetry is marked by s and twofold symmetry by t.

1998) and uracil PRT shows a hexameric arrangement in solution (Jensen & Mygind, 1996). Therefore, the proposed model of ATP-PRT with dimers arranged within a hexamer is in accordance with known structures of other PRTs.

We would like to thank J. Greene for technical assistance, S. Ali for fruitful discussions, S. Campbell for help with purification and P. Emsley, R. Thom and A. Roszak for help with the data collection. BL would like to thank the Wellcome Trust for funding.

References

- Alifano, P., Fani, R., Liò, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M. S. & Bruni, C. B. (1996). *Microbiol. Rev.* **60**, 44–69.
- Ames, B. N., Martin, R. G. & Garry, B. J. (1961). *J. Biol. Chem.* **236**, 2019–2026.
- Argos, P., Hanei, M., Wilson, J. M. & Kelley, W. N. (1983). *J. Biol. Chem.* **258**, 6450–6457.
- Cudney, B., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst. D50*, 414–423.
- Eads, J. C., Ozturk, D., Wexler, T. B., Grubmeyer, C. & Sacchettini, J. C. (1997). *Structure*, **5**, 47–58.
- Ferré-D'Amaré, A. R. & Burley, S. K. (1997). *Methods Enzymol.* **276**, 157–166.
- Focia, P. J., Craig, S. P. III, Nieves-Alicea, R. N., Fletterick, R. J. & Eakin, A. E. (1998). *Biochemistry*, **37**, 15066–15075.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Jensen, K. F. & Mygind, B. (1996). *Eur. J. Biochem.* **240**, 637–645.
- Klungsoyr, L. & Kryvi, H. (1971). *Biochim. Biophys. Acta*, **227**, 327–336.
- Kryvi, H. (1973). *Biochim. Biophys. Acta*, **317**, 123–130.
- Kryvi, H. & Klungsoyr, L. (1971). *Biochim. Biophys. Acta*, **235**, 429–434.
- Martin, R. G., Berberich, M. A., Ames, B. N., Davis, W. W., Goldberger, R. F. & Youno, J. D. (1971). *Methods Enzymol.* **17B**, 3–44.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Mori, I., Fonné-Pfister, R., Matsunaga, S., Tada, S., Kimura, Y., Iwasaki, G., Mano, J., Hatano, M., Nakano, T., Koizumi, S., Scheidegger, A., Hayakawa, K. & Ohta, D. (1995). *Plant Physiol.* **107**, 719–723.
- Mousdale, D. M. & Coggins, J. R. (1991). *Target Sites for Herbicide Action*, edited by R. C. Kirkwood, pp. 29–56. New York: Plenum Press.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Studier, F. W. & Moffatt, B. A. (1986). *J. Mol. Biol.* **189**, 113–130.
- Tebar, A. R., Fernandez, V. M., MartinDelRio, R. & Ballester, A. O. (1973). *Experientia*, **29**, 477–479.
- Tomchick, D. R., Turner, R. J., Switzer, R. L. & Smith, J. L. (1998). *Structure*, **6**, 337–350.
- Voll, M. J., Appella, E. & Martin, R. G. (1967). *J. Biol. Chem.* **242**, 1760–1767.
- Wilmanns, M. & Eisenberg, D. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 1379–1383.

