UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

School of Biological Sciences

Structural Studies of *Burkholderia pseudomallei* invasion protein D (BipD), 2,4’-dihydroxyacetophenone dioxygenase (DAD) and L-threonine dehydrogenase (TDH)

by

Alexandra Bowyer

Thesis submitted for the degree of Doctor of Philosophy

November 2008

UNIVERSITY OF SOUTHAMPTON

Abstract

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
SCHOOL OF BIOLOGICAL SCIENCES

Doctor of Philosophy

STRUCTURAL STUDIES OF *BURKHOLDERIA PSEUDOMALLEI* INVASION PROTEIN D (BIPD), 2,4'-DIHYDROXYACETOPHENONE DIOXYGENASE (DAD) AND L-THREONINE DEHYDROGENASE (TDH)

by Alexandra Bowyer

The determination of novel protein structures is essential in advancing our understanding of how they function within their molecular environments, therefore providing some of the information necessary to exploit them in ways that will benefit society. Three very different novel proteins have been studied by X-ray crystallography, and it is hoped that the information gained can be applied to real life challenges.

Burkholderia pseudomallei invasion protein D (BipD) not only has the potential for misuse as a bioterrorist weapon, due to its air-borne infectivity, but also causes the disease melioidosis, which can kill within 48 hours. The structure of this novel protein was determined by MAD phasing methods to 2.1 Å, and has been found to be highly homologous with IpaD from *Shigella flexneri*. This knowledge has enabled proposals to be made for a homopentameric ring structure of this protein and identified possible pH induced conformational changes which may have relevance *in vivo*.

2,4'-Dihydroxyacetophenone Dioxygenase (DAD) from *Alcaligenes sp.* is of interest for quite a different reason. Aromatic hydrocarbons, are especially damaging to the environment because they are not naturally degraded. The expression and purification protocols for DAD have now been refined and the native protein has been crystallised and diffraction data collected. However, because it is a novel structure, a selenomethionine derivative must also be crystallised and data collected; these optimisation trials are continuing. Structural knowledge about the enzymes involved in aromatic catabolic pathways will result in a better understanding of the whole system and therefore present opportunities for manipulation for bioremediation purposes.

Knowledge of a protein's structure can also help identify how specific properties are conferred, such as thermostability at extreme temperatures. L-Threonine Dehydrogenase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* (TkTDH) is involved in the first of a two-step biochemical pathway conversion of threonine to glycine, by catalysing the NAD^+ dependent metabolism of L-threonine to 2-amino-3-ketobutyrate (KBL). The structure of this enzyme has been determined to 2.4 Å bound to its co-factor NAD^+. It has been proposed that a structural and catalytic zinc are present in each monomer of this homotetramer as observed in the structurally similar alcohol dehydrogenases, but both were absent in this structure. The conserved residues postulated to co-ordinate these zinc ions were however in conformation that would allow metal ion coordination. Attempts have also been made to model the TDH tetramer with 2 dimers of KBL, as it is though that these enzymes form a complex and pass the intermediate substrate between them. This modelling is continuing.

The structures of BipD and TkTDH have been solved and work is continuing on DAD This has made available a large amount of structural and functional information that would otherwise be inaccessible, contributing to our overall understanding of these macromolecules, as well as providing specific information that can be applied to particular challenges.

Table of Contents

Structural Studies of *Burkholderia pseudomallei* invasion protein D (BipD), 2,4'-dihydroxyacetophenone dioxygenase (DAD) and L-threonine dehydrogenase (TDH)

Chapter 1 Introduction.....1
- 1.1 *Burkholderia pseudomallei* Invasion Protein D.....2
  - 1.1.1 Melioidosis.....2
  - 1.1.2 Burkholderia pseudomallei.....4
  - 1.1.3 Bacterial Pathogenicity.....7
  - 1.1.4 Protein Secretion Systems.....8
  - 1.1.5 The Type III Protein Secretion System.....10
  - 1.1.6 Type III Protein Secretion System in *Burkholderia pseudomallei*.....12
  - 1.1.7 BipD Homologues.....12
  - 1.1.8 The Role of BipD within the TTSS of *Burkholderia pseudomallei*.....14
  - 1.1.9 BipD may Prevent the Leakage of Effector Proteins from the Needle.....15
  - 1.1.10 How Does BipD Fulfill its Other Functions?.....16
- 1.2 2,4'-Dihydroxyacetophenone Dioxygenase.....17
  - 1.2.1 Aromatic Compounds.....17
  - 1.2.2 Bioremediation.....18
  - 1.2.3 Microbial Methods of Degradation.....20
  - 1.2.4 Oxygenases.....20
  - 1.2.5 Dioxygenases.....21
  - 1.2.6 Cofactors.....21
    - 1.2.6.1 Mechanism of Dioxygen Activation.....21
    - 1.2.6.2 Catechol Dioxygenases.....23
    - 1.2.6.3 Cofactor Binding Sites.....24
  - 1.2.7 2,4'-Dihydroxyacetophenone Dioxygenase.....25
    - 1.2.7.1 Sequence and Structural Similarity.....26
- 1.3 L-threonine Dehydrogenase.....29
  - 1.3.1 Amino Acids.....29
  - 1.3.2 Oxidoreductases and Dehydrogenases.....30
  - 1.3.3 L-Threonine Dehydrogenase.....31
  - 1.3.4 TDH from *Thermococcus kodakaraensis*.....34
    - 1.3.4.1 Thermostability of Hyperthermophilic Proteins.....34
    - 1.3.4.2 *Thermococcus kodakaraensis*.....36

Chapter 2 Materials, Methods and X-ray Crystallography.....39
- 2.1 Molecular Biology.....40
  - 2.1.1 DNA Vectors.....40
  - 2.1.2 Restriction Endonucleases.....41
  - 2.1.3 DNA Ligase.....42
  - 2.1.4 Primer Design and Polymerase Chain Reaction.....43
  - 2.1.5 Agarose Gel Electrophoresis.....45

2.1.6 Bacterial Strains 46
2.2 Protein Purification 48
2.2.1 Sonication 48
2.2.2 Fractional Ammonium Sulphate Precipitation 48
2.2.3 Dialysis 49
2.2.4 Chromatography 49
2.2.4.1 Ion Exchange Chromatography 50
2.2.4.1.1 Q-sepharose™ Anion Exchange Chromatography Column 50
2.2.4.1.2 Hi Trap Q XL Anion Exchange Column 51
2.2.4.2 Gel Filtration Chromatography 51
2.2.4.2.1 Superdex™ 200 Gel Filtration Chromatography Column 52
2.2.4.3 Affinity Chromatography 53
2.2.4.3.1 GSTrap Affinity Chromatography Column 54
2.2.4.3.2 Benzamidine Affinity Chromatography Column 54
2.2.5 Concentration 54
2.2.6 SDS-PAGE Gel Electrophoresis 55
2.3 Crystallisation 56
2.3.1 Crystal Freezing 60
2.4 Protein X-ray Crystallography 61
2.4.1 Data Collection 61
2.4.2 Data Processing 62
2.4.3 Phase Determination 66
2.4.3.1 Molecular Replacement 67
2.4.3.2 Anomalous Diffraction Phasing 68
2.4.4 Refinement and Model Building 69
2.4.4.1 Refinement 70
2.4.4.2 Model Building 71
2.4.5 Validation and Structural Analysis 73

Chapter 3 Structural studies of Burkholderia pseudomallei Invasion Protein D from *Burkholderia pseudomallei* 75
3.1 Introduction 76
3.2 BipD Expression and Purification Protocol 77
3.2.1 DNA Manipulation 77
3.2.2 Transformation 77
3.2.2.1 Plasmid Purification 78
3.2.2.2 Quantitative Analysis of DNA 78
3.2.2.3 Primer Design 78
3.2.2.4 Polymerase Chain Reaction 79
3.2.2.5 DNA Digest 80
3.2.2.6 DNA Ligation 81
3.2.2.7 Large-scale Preparation of Plasmid 81
3.2.3 Protein Expression 82
3.2.3.1 Protein Expression of Seleno-methionine Protein 82
3.2.3.2 Over-expression of Native Protein 83
3.2.3.3 Harvest Cells 83
3.2.3.4 Sonication 84
3.2.4 Crude Purification 84

- 3.2.4.1 Ammonium Sulphate Precipitation ... 84
- 3.2.4.2 Dialysis ... 85
- 3.2.5 Column Purification (with the GST-fusion protein) ... 85
- 3.2.5.1 GSTrap Affinity Chromatography ... 85
- 3.2.5.2 Removal of Thrombin ... 86
- 3.2.6 Column Purification (with no GST-tag) ... 86
- 3.2.6.1 Q-Sepharose Anion Exchange Column ... 86
- 3.2.6.2 Superdex 200 Gel Filtration Column ... 87
- 3.3 Crystallisation ... 87
- 3.4 Structure Determination ... 88
- 3.4.1 Data Collection and Processing ... 88
- 3.4.2 Refinement and Model Building ... 89
- 3.5 Structure Analysis ... 93
- 3.5.1 General Structure Analysis ... 93
- 3.5.2 Tertiary Structure ... 95
- 3.5.3 Quaternary Structure ... 97
- 3.6 Discussion ... 98
- 3.6.1 Movement of BipD Through the TTSS Needle ... 98
- 3.6.2 The Effect of pH Changes ... 100
- 3.6.3 BipD as a Translocon Plug ... 102
- 3.7 Further Efforts at Structure Analysis ... 103

Chapter 4 Structural studies of 2,4’-Dihydroxyacetophenone dioxygenase from *Alcaligenes sp.* ... 104
- 4.1 Introduction ... 105
- 4.2 DAD Expression and Purification Protocol ... 107
- 4.2.1 DNA Manipulation ... 107
- 4.2.2 Protein Expression ... 107
- 4.2.2.1 Native Protein Over-expression ... 107
- 4.2.2.2 Seleno-methionine Protein Over-expression ... 109
- 4.2.3 Harvest Cells ... 109
- 4.2.4 Sonication ... 109
- 4.2.5 Crude Purification ... 110
- 4.2.5.1 Ammonium Sulphate Precipitation ... 110
- 4.2.5.2 Dialysis ... 111
- 4.2.6 Column Purification ... 111
- 4.2.6.1 Q-Sepharose Anion Exchange Column ... 111
- 4.2.6.2 Superdex 200 Gel Filtration Column ... 112
- 4.2.6.3 Hi Trap Q XL Anion Exchange Column ... 112
- 4.3 Crystallisation ... 113
- 4.3.1 Native DAD Protein ... 113
- 4.3.2 Selenomethionine DAD Derivative ... 115
- 4.4 Preliminary X-ray Diffraction Analysis of Native DAD ... 115

Chapter 5 Structural studies of L-threonine Dehydrogenase from *Thermococcus kodakaraensis* ... 118
- 5.1 Introduction ... 119
- 5.2 TDH Expression and Purification Protocol ... 121

5.2.1 Cloning of TkTDH 121
5.2.2 Expression and Purification 122
5.3 Crystallisation 122
5.4 Structure Determination 124
5.4.1 Data Collection and Processing 124
5.4.2 Molecular Replacement 128
5.4.3 Model Building and Refinement 133
5.5 Structure Analysis 136
5.5.1 Primary Structure Similarity 136
5.5.2 Structure of the Monomer 137
5.5.3 Structure of the Tetramer 141
5.5.4 Structural Zinc Ion 144
5.5.5 Active Site 145
5.6 Discussion 150
5.6.1 Thermostability 150
5.6.2 Complex Formation 150
5.6.3 Mechanism of Action 153

Chapter 6 Summary 155
6.1 Burkholderia pseudomallei invasion protein D 156
6.2 2,4’-dihydroxyacetophenone dioxygenase 156
6.3 L-threonine dehydrogenase 157

Appendix 158
References 162

Table of Figures

Chapter 1 - Introduction

Figure 1.1: *Burkholderia pseudomallei* 4
Figure 1.2: *B. pseudomallei* cells escaping from infected cells 6
Figure 1.3: Schematic diagram of the *B. pseudomallei* genome 7
Figure 1.4: Schematic diagram of type I, type II and type III secretion systems 9
Figure 1.5: Bacterial secretion system injectisome 11
Figure 1.6: Needle of the TTSS 11
Figure 1.7: The relative loci of type three section systems 13
Figure 1.8: A proposed pentameric structure for IpaD 16
Figure 1.9: A benzene ring consists of 6 carbons 17
Figure 1.10: Microbial degradation and bioremediation 19
Figure 1.11: Activation of dioxygen by reduced flavin 22
Figure 1.12: Proposed activation mechanism for intradiol catechol cleavage 23
Figure 1.13: Three isomeric forms of a dihydroxybenzene ring: 23
Figure 1.14: Cleavage mechanism 24
Figure 1.15: Structural analysis of IsdG and ActVA mono-oxygenase families 24
Figure 1.16: Cleavage mechanism 24
Figure 1.17: 2,4'-dihydroxyacetophenone dioxygenases reaction mechanism 26
Figure 1.18: Structural analysis of IsdG and ActVA mono-oxygenase families 28
Figure 1.19: The general structure of an amino acid 29
Figure 1.20: Amino acid enantiomers 30
Figure 1.21: Metabolism of threonine by TDH 32
Figure 1.22: Dimeric and tetrameric ADH 33
Figure 1.23: Rossmann fold 34
Figure 1.24: EM observation of *Thermococcus kodakaraensis* 37

Chapter 2 - Materials, Methods and X-ray Crystallography

Figure 2.1: Restriction enzyme recognition sites of pET11a plasmid 40
Figure 2.2: DNA agarose gel from electrophoresis 46
Figure 2.3: Chromatography types 49
Figure 2.4: Structure of sepharose cross-linked agarose media 51
Figure 2.5: Gel filtration 52
Figure 2.6: Affinity chromatography 53
Figure 2.7: An SDS gel 55
Figure 2.8: Crystallographic unit cell and lattice 56
Figure 2.9: Bragg's Law 57
Figure 2.10: Hanging-drop crystallisation method 58
Figure 2.11: A crystal growth phase diagram 59
Figure 2.12: Protein crystals 60
Figure 2.13: An X-ray crystallographic diffraction pattern 61
Figure 2.14: The 14 Bravais lattices 63
Figure 2.15: Crystallographic space group prediction by MOSFLM 64
Figure 2.16: An exert from the log file produced by Scala 66

Figure 2.17: Electron density from a $2F_o$-F_c map 71
Figure 2.18: Positive and negative F_o-F_c difference electron density 72
Figure 2.19: Modified residue in electron density 72
Figure 2.20: Ramachandran plot showing the peptide bond angles Phi and Psi 73

Chapter 3 - Burkholderia pseudomallei invasion protein D (BipD)

Figure 3.1: An agarose gel showing a plasmid ligation sample 81
Figure 3.2: SDS gel showing pre and post IPTG induced samples 84
Figure 3.3: BipD protein crystals 87
Figure 3.4: X-ray diffraction image for BipD 88
Figure 3.5: A section of the electron density map for part of helix 8 90
Figure 3.6: A Ramachandran plot for the BipD model 91
Figure 3.7: A dimer of BipD forming the crystallographic asymmetric unit 92
Figure 3.8: Schematic of the BipD protein molecule's topology 92
Figure 3.9: A ribbon model of the BipD structure 94
Figure 3.10: A CD spectrum for BipD 94
Figure 3.11: Superposition of both molecules of BipD in the asymmetric unit 95
Figure 3.12: Sequence alignment of BipD with its homologues IpaD and SipD 96
Figure 3.13: Specific contacts involved in dimer formation 97
Figure 3.14: An unfolded chaperone superimposed on IpaD 99
Figure 3.15: Alignment of BipD and IpaD at different pH 100
Figure 3.16: The electrostatic surface of the BipD dimer 101

Chapter 4 - 2,4'-dihydroxyacetophenone dioxygenase (DAD)

Figure 4.1: Reaction catalysed by 2,4'-dihydroxyacetophenone dioxygenase 105
Figure 4.2: SDS gel showing sample protein contents 108
Figure 4.3: SDS gel showing soluble protein content of samples 110
Figure 4.4: SDS gel showing the results of ASP 111
Figure 4.5: SDS gel showing protein content 112
Figure 4.6: SDS gel showing the sample purity 113
Figure 4.7: Crystals of DAD protein 115
Figure 4.8: DAD diffraction data collected to 3.1 Å resolution 116

Chapter 5 - L-threonine dehydrogenase (TDH)

Figure 5.1: Multiple TDH crystal 123
Figure 5.2: TDH crystal 123
Figure 5.3: TDH diffraction data collected to 2.4 Å 125
Figure 5.4: MOSFLM C2 space group prediction 126
Figure 5.5: Section of MOSFLM spot prediction in C2 space group 126
Figure 5.6: MOSFLM P4 space group prediction 127
Figure 5.7: Section of MOSFLM spot prediction in P4 space group 127
Figure 5.8: TDH matthews coefficient prediction 128
Figure 5.9: HKL view of TDH crystals in P2 space group 131
Figure 5.10: HKL view of TDH crystals in P4 space group 132
Figure 5.11: Electron density for the TkTDH model 135

Figure 5.12: TkTDH sequence alignment 136
Figure 5.13: Superimposition of TkTDH with PhTDH and SsADH 138
Figure 5.14: Overall topology of a TkTDH monomer 139
Figure 5.15: TkTDH nomenclature 139
Figure 5.16: Tetrameric structure of TkTDH compared to PhTDH, PaTDH and SsADH 141
Figure 5.17: TkTDH dimer of monomers A and B 142
Figure 5.18: TkTDH dimer of monomers A and C 143
Figure 5.19: Interaction of monomers A and D in TkTDH 143
Figure 5.20: NAD^+ binding pocket in TkTDH 145
Figure 5.21: Proposed active site for TkTDH 146
Figure 5.22: NAD^+ electron density in the TkTDH model 147
Figure 5.23: Conserved water molecule at the NADH binding site 149
Figure 5.24: Putative docking of KBL with TDH 151
Figure 5.25: Electrostatic surface of TkTDH and KBL 152
Figure 5.26: Proton relay mechanism for PhTDH 153

Table of Tables

Table 1: Sequence identity and similarity of known homologues of TTSS 13
Table 2: Statistics for BipD datasets 89
Table 3: Refinement statistics for the BipD structure 93
Table 4: Data collection and processing statistics for native DAD.. 117
Table 5: Data collection and processing statistics for TkTDH 133
Table 6: Refinement statistics for TkTDH 134

Acknowledgements

This thesis is the account of three years of committed work in the field of protein X-ray crystallography at the University of Southampton, which would not have been possible without the help of many.

I am most grateful to my supervisor, Professor Jon Cooper for all the time and effort he has put in to supporting me through my PhD. His guidance in the laboratory ensured that there was always plenty for me to do and his knowledge of X-ray crystallography helped me work through some difficult problems. I think perhaps he may also hold the record for the fastest reading of a PhD thesis, as all my chapters were returned to me in little over a week, with many helpful comments and suggestions.

As my third year undergraduate project supervisor, Dr Alun Coker's support and assistance nurtured my enjoyment of laboratory research, and it was perhaps at this point that I realised this was the direction in which I wished my career to progress. He also provided some much needed computational support with the various crystallographic software and initially alien Linux operating system.

It is to the people in my laboratory that I also owe a huge thanks, particularly to Halina Mikolajek who helped me in the lab and was always willing to assist when I had a problem. And with whom I have had many technical, and some not so technical, discussions in the graphic suite. To Dr Raj Gill for his assistance with the genetic aspects of my work, Jeab for putting up with me the lab, Dr Simon Kolstoe and Dr Peter Erskine for their help with the theory of X-ray crystallography, Professor Steve Wood for his help and advice, and to the undergraduate project students who have aided me in my work. A big thank you also to all my other colleagues and friends at university, who have offered a shred of sanity through the tough times.

A special thanks must also go to my husband Alex, who has put up with me during my three years as a PhD student. And particularly over the last few months when I’ve been glued to my computer from dawn till dusk writing my thesis. Without his love and support this would have been a much tougher and lonelier experience. And after three years of me talking about proteins and electron density and helices, he might just have learnt something too.

To my friends and family I also owe a huge thank you for being so supportive of my choices in life and always giving me the freedom to do what I thought was right for me. To my Mum for always being there for me and loving me for who I am. I would like to dedicate this thesis to the memory of my Dad, who lost his fight with cancer whilst I was still an undergraduate; his pleasure in learning and enthusiasm for science were an inspiration from an early age. Long may our quest for scientific knowledge continue and maybe one day it will lead us to the medical answers we seek.

List of Abbreviations:

aa	Amino Acids
ADH	Alcohol Dehydrogenase
APS	Ammonium persulphate
A. pernix	*Aeropyrum pernix*
ApADH	Alcohol Dehydrogenase from *Aeropyrum pernix*
ATP	Adenosine Triphosphate
BipD	*Burkholderia pseudomallei* invasion protein D
B. mallei	*Burkholderia mallei*
B. pseudomallei	*Burkholderia pseudomallei*
bp	base pair
B. thailandensis	*Burkholderia thailandensis*
Bsa	Burkholderia secretion apparatus
C. beijerinckii	*Clostridium beijerinckii*
CbADH	Alcohol dehydrogenase from *Clostridium beijerinckii*
CCP4	Collaborative Computational Project No. 4
CD	Circular Dichroism spectroscopy
DAD	2,4'-Dihydroxyacetophenone dioxygenase
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EcTDH	L-Threonine Dehydrogenase from *Escherichia coli*
EDTA	Ethylendiaminetetraacetic Acid
EMBL	European Molecular Biology Laboratory
ESRF	European Synchrotron Radiation Facility
EtBr	Ethidium Bromide
FAD	Flavin Adenine Dinucleotide
F. frigidimaris	*Flavobacterium frigidimaris*
FfADH	Alcohol Dehydrogenase from *Flavobacterium frigidimaris*
FMN	Flavin Mononucleotide
GSTag	Glutathione S-transferase tag
HlADH	Alcohol Dehydrogenase from Horse liver
Ipa	Invasion protein antigen (from *Shigella*)
IPTG	isopropyl-β-d-thiogalactopyranoside

KBL	2-amino 3-ketobutyrate CoA Ligase
LB	Luria-Bertani growth medium
LPS	Lipopolysaccharide
M9 media	Minimal media
Mb	Megabases
NAD^+ / NADH	Nicotinamide Adenine Dinucleotide
$NADP^+$ / NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NCS	Non-crystallographic symmetry
$^1O_2$	singlet excited state of dioxygen
$^3O_2$	ground state of dioxygen
OD	optical density
P. abyssi	*Pyrococcus abyssi*
PaTDH	L-Threonine Dehydrogenase from *Pyrococcus abyssi*
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
P. furiosus	*Pyrococcus furiosus*
P. horikoshii	*Pyrococcus horikoshii*
PhTDH	L-threonine Dehydrogenase from *Pyrococcus horikoshii*
π*	Pi electrons
RMSD	Root Mean Squares Deviation
SAD	Single wavelength Anomalous Dispersion
S. cerevisiae	*Saccharomyces cerevisiae*
ScADH	Alcohol Dehydrogenase from *Saccharomyces cerevisiae*
SDS	Sodium Dodecyl Sulphate
SeMet	Selenomethionine
Sip	Salmonella invasion proteinenediamine
S. solfataricus	*Sulfolobus solfataricus*
SsADH	Alcohol Dehydrogenase from *Sulfolobus solfataricus*
TB	Terrific Broth
T. brockii	*Thermoanaerobacter brockii*
TbADH	Alcohol Dehydrogenase from *Thermoanaerobacter brockii*
TDH	L-Threonine Dehydrogenase
TEMED	N,N,N'N'-tetramethylethyl
TkTDH	L-Threonine Dehydrogenase from *Thermococcus kodakaraensis*

TRIS	tris(hydroxymethyl)methylamine
TTSS	Type Three Secretion System
T. tengcongensis	*Thermoanaerobacter tengcongensis*
TtADH	Alcohol Dehydrogenase from *Thermoanaerobacter tengcongensis*
UV	Ultraviolet

Chapter 1

Introduction

1.1 *Burkholderia pseudomallei* Invasion Protein D

1.1.1 Melioidosis

First identified in opium addicts in Rangoon, Burma in 1911, the British pathologist Captain Alfred Whitmore (1876 – 1946), together with C.S. Krishnaswami, discovered the infectious disease melioidosis, and initially called it Whitmore's Disease. It was found to be caused by the bacterium *Burkholderia pseudomallei*, a gram-negative saprotroph. This disease is particularly prevalent during the rainy season in endemic regions, mainly Southeast Asia and Northern Australia, with the highest documented rate being in North-Eastern Thailand. Here melioidosis accounts for around 20% of all community-acquired septicaemias (Holden *et al.* 2004;Blocker *et al.* 2000;Mecsas and Strauss 1996).

The bacterium is able to enter the human body through open wounds or abrasions to the skin, or by inhalation, and mainly affects the immunocompromised and those with an underlying predisposition to infection. People who come into direct contact with wet soils, such as rice farmers, are often infected (White 2003). For example, in Thailand this bacterium can readily be recovered in abundance from more than 50% of rice paddies (Wuthiekanun *et al.* 1995;Wuthiekanun *et al.* 1995).

Melioidosis is pathologically similar to Glanders disease, a bacterial infection that predominately affects horses and other equines, caused by the closely related organism *Burkholderia mallei*, also identified over a century ago (White 2003). Glanders is now very rare in Europe and North America and is largely restricted to parts of Africa, Asia, the Middle East and Central and South America. It was previously thought the two organisms were closely related but multilocus sequence typing has shown that *B. mallei* can in fact be considered a descendant of *B. pseudomallei* (Godoy *et al.* 2003).

B. pseudomallei is known to infect almost all cell types and it is this broad cellular tropism that contributes to its versatility as a pathogen and translates into a wide spectrum of clinical manifestations, making it difficult to accurately diagnose (Gan 2005). Symptoms of melioidosis can be categorised as either acute or localised infection, acute bloodstream infection, acute pulmonary infection or chronic suppurative infection:

Acute, localised infection: generally localised as a nodule as a result of inoculation through a break in the skin. This can cause fever and general muscle aches and may progress rapidly to infect the bloodstream.

Acute bloodstream infection: this affects people with other conditions, particularly HIV, renal failure and diabetes, usually resulting in septic shock. Symptoms vary depending on the original site of infection but generally include respiratory stress, severe headache, fever, diarrhoea, development of pus-filled lesions on the skin, muscle tenderness, and disorientation. This is typically an infection of short duration, and abscesses will be found throughout the body.

Pulmonary infection: this can take the form of mild bronchitis to severe pneumonia and is typically accompanied with high fever, headaches, anorexia and general muscle soreness. Chest pain is common but a non-productive or productive cough with normal sputum is the hallmark of this type of melioidosis.

Chronic suppurative infection: this typically affects the joints, viscera, lymph nodes, skin, brain, liver, lung, bones, and spleen (US Department of Health and Human Services, 2006).

The recorded incidences of clinical melioidosis each year in endemic regions such as Thailand is estimated to be around 4.4 cases per 100,000 (Suputtamongkol *et al.* 1994). However, because the clinical manifestations of melioidosis are so varied, the symptoms are often mistaken for other diseases and so consequently it is not accurately recorded. Even with the recent improvements in diagnostic techniques and a greater awareness of the disease, it is thought current figures may still be grossly underestimated.

The fact that melioidosis is able to kill within 48 hours, coupled with the risk of the use of *B. pseudomallei* as a biochemical weapon, has meant that it is becoming more and more important to improve our understanding of the way in which this bacterium is able to infect and persist within host cells. A Canadian group has been exploring potential vaccination strategies but despite recent work demonstrating some protection in animal models following vaccination with *B. thailandensis* and other attenuated strains, it is unlikely that this strategy alone will provide sufficient protection (Druar *et al.* 2007). Indeed it was found that neither BipB, BipC nor BipD are protective antigens, because vaccination of mice with any single protein did not result in protection against experimental melioidosis (Druar *et al.* 2007). Even if a vaccine were to be developed, it is expected that economic constraints may make vaccination an unrealistic option for many regions of endemicity

(Cheng 2005). It is therefore important to further understand the method of infection in order to not only identify better vaccine targets, but also potential targets for new drugs. This will make it possible to design drugs specifically against *B. pseudomallei*, possibly by blocking the TTSS as its method of infection.

1.1.2 Burkholderia pseudomallei

B. pseudomallei is a motile, aerobic, non-spore-forming, gram-negative bacillus, shown in Figure 1.1. This resilient saprotrophic bacterium feeds by absorbing dead or decaying organic matter and is able to survive either in the soil or within human cells for many years, only re-emerging when conditions once again become favourable for its growth and replication (White 2003). Its ability to produce specialised secondary metabolites and its adaptation of complex catabolic pathways and transport systems contribute to this organism's capacity for survival in diverse, competitive environments.

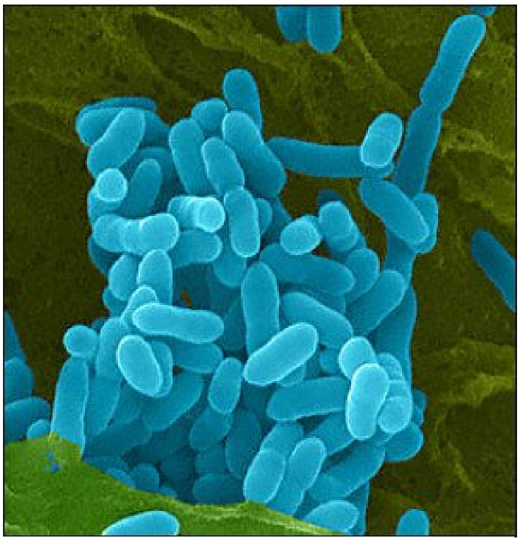


Figure 1.1: *Burkholderia pseudomallei*
This bacterium is an aerobic, gram-negative bacillus, 2-5µm in size that is able to survive in the soil or within human cells (Wellcome Trust Sanger Institute press release).

B. pseudomallei obtains its nutrition primarily from decomposing non-living organic matter and, when possible, by parasitizing soil-dwelling amoebae and absorbing the resultant soluble organic compounds. It is able to enter and survive within these free-living amoebae of the genus *Acanthamoeba* (Inglis *et al.* 2000) and it is thought that its ability to infect human cells may be an unfortunate consequence of the mechanisms that allow it to parasitize amoebae, although these are poorly understood (Aldhous 2005).

The bacterium is able to survive for many years in environments with a low nutrient concentration, illustrated by the fact that it has been successfully cultured from a sample stored in distilled water for 10 years (Aldhous 2005). It is also tolerant of adverse environmental conditions, such as low pH (Stevens and Galyov 2004). There have been case reports describing recrudescent melioidosis after an initial exposure 26 and 18 years previously (Mays and Ricketts 1975;Koponen *et al.* 1991), illustrating just how resilient this bacterium can be.

Another factor of *B. pseudomallei*'s ability to survive in adverse conditions is its reduced susceptibility to antibiotics. Having evolved to compete with other organisms in the soil that secrete antibacterial compounds, the bacterium is resistant to many drugs (Aldhous 2005). This may explain in part why relapse of melioidosis patients is common, even when treated with appropriate antibiotics (Stevens and Galyov 2004). However, in general it is susceptible to chloramphenicol, carbapenems, tetracyclines, trimethoprim-sulfamethoxazole, ureidopenicillins, third generation cephalosporins, and (unusually for a pseudomonad) amoxicillin-clavulanate. It has also been shown that the highly conserved immunodominant antigen of lipopolysaccharide (LPS), which forms part of the bacterial cell wall, is recognised by LPS2 antibodies that, in high concentrations, are associated with improved survival in cases of severe melioidosis.

In addition to this intrinsic resistance to many antibiotics, *B. pseudomallei* is able to produce a highly hydrated glycocalyx polysaccharide capsule that assists the formation of a slime capsule. This allows the development of microcolonies in which the bacteria is both protected from antibiotic penetration and phenotypically altered, as a result of which its susceptibility to antibiotics is reduced (White 2003).

Another feature of *B. pseudomallei* that makes it particularly virulent is its ability to escape endocytic vesicles and spread from cell to cell through the formation of membrane protrusions as shown in Figure 1.2 (Holden *et al.* 2004). These are caused by the initiation of actin oligomerisation at one pole of the infected cell (Stevens and Galyov 2004). In pathological specimens this can also be seen within phagocytic cells, where the bacterium is able to survive and proliferate for prolonged periods before lysing the endosome membrane and escaping (White 2003).

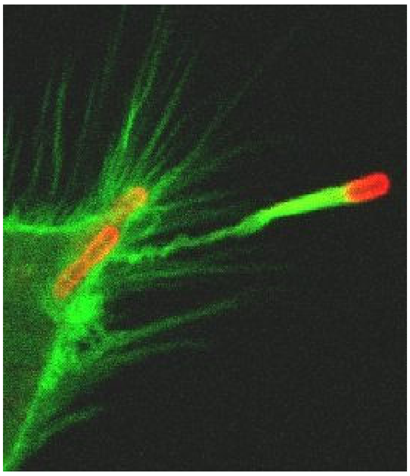


Figure 1.2: *B. pseudomallei* cells escaping from infected cells
The bacterium achieves this using actin "rockets". *B. pseudomallei* is stained red. It is thought BipD somehow plays a role in this process of rearranging the actin cytoskeleton. The actin is coloured red.

B. pseudomallei's ability to survive intracellularly is likely to be necessary for its pathogenesis and the recrudescence of melioidosis in patients. When *B. pseudomallei* moves to the periphery of the host cell, it initiates the continuous polymerisation of actin within its own cytoplasm. These actin formations create a protrusion of the host's plasma membrane with the bacterium at the tip end pushing against it (Breitbach *et al.* 2003). *B. pseudomallei* is able to project these membrane protrusions into adjacent cells, enabling cell to cell spread and inducing the fusion of adjacent cells, resulting in the formation of giant multinucleated cells in infected monolayers (Suparak *et al.* 2005).

For many years *B. pseudomallei* attracted very little attention. However, recent fears that the organism may one day be used as a bioterrorist agent, due to its worldwide availability, severe course of infection and in particular its aerosol infectivity (Stevens *et al.* 2002), have resulted in a surge of interest in the bacterium. It now appears on the category B list of critical agents, published by the US Centers for Disease Control and Prevention (White 2003).

This interest in the bacterium has resulted in the completion of the genome sequence for *B. pseudomallei*, shown schematically in Figure 1.3, published by the Wellcome Trust Sanger Institute (Holden *et al.* 2004). It was reported that, unlike most bacteria, the genome comprises two chromosomes of 4.07 megabase pairs and 3.17 megabase pairs, showing significant functional partitioning of genes between them.

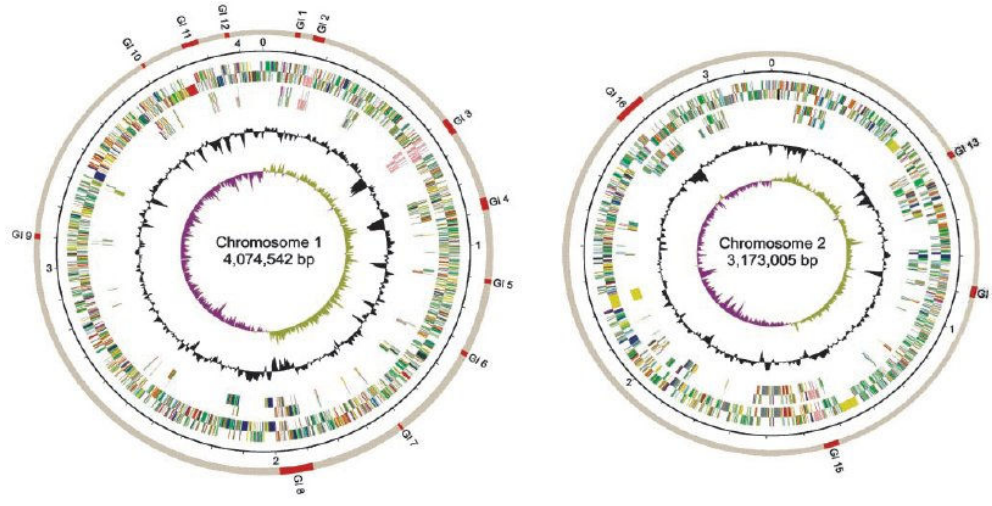


Figure 1.3: Schematic diagram of the *B. pseudomallei* genome
The complete genome of strain K96243 consists of two circular replicons, one large and one small, which encode essential and accessory functions respectively (Holden *et al.* 2004) (European Molecular Biology Laboratory accession nos. BX571965 and BX 571966).

The large chromosome encodes many of the core functions associated with central metabolism and cell growth, whereas the small chromosome carries more accessory functions associated with adaptation and survival in different environments. At 7.24 Mb the *B. pseudomallei* genome is large in comparison to a typical prokaryotic genome, perhaps accounting for the bacterium's versatility and adaptability (Holden *et al.* 2004).

1.1.3 Bacterial Pathogenicity

Bacteria are able to invade many different tissues within the human body, which consequently results in patients displaying a wide variety of symptoms. Initially it seemed reasonable to theorise that each individual disease might be generated by a different molecular mechanism (Mecsas and Strauss 1996). However, as researchers have made greater use of the molecular and genomic information now available to them, it has become apparent that the spectrum of methods is far narrower than first thought and that bacteria actually only utilise a relatively small number of common molecular tools to achieve a wide range of results (Finlay and Falkow 1989).

Obviously one of the most important components of bacterial virulence is the ability to invade host cells and then subvert the cell's machinery to its advantage. Amongst the tools used to achieve this are the protein secretion systems; type I, type II, type III, type IV and type V, with type III being the most recently identified through studies of the *Yersinia* bacteria.

When bacterial genomes were searched it was observed that virulence genes, such as those encoding these secretion systems, were often found clustered in functionally related groups. Furthermore, it appeared in many instances that these genes had been acquired from another organism, because features of this specific DNA differed from the bulk of the bacterial genome. These observations led to the concept of pathogenicity islands; discrete segments of DNA that encode virulence traits and often appear to be foreign in origin (Mecsas and Strauss 1996).

Many of these pathogenicity islands appear to have been acquired in one step. This ability to obtain complex virulence traits in one genetic event, as opposed to undergoing natural selection for many generations, provides a method for sudden radical changes in bacterial-host interactions. This flexibility is a big advantage for the bacteria's survival and rapid adaptation to new and altered environments.

1.1.4 Protein Secretion Systems

It has long been known that secreted or surface-exposed bacterial proteins (effectors) play key roles in an organism's ability to interact with host cells. Bacteria can utilise one or a combination of general secretory pathways and five specialized (Type I-V) secretion systems to transport proteins (Henderson *et al.* 2004). The function of the protein secretion systems is to actually transport these 'effector' proteins, amongst others, to the surface of the bacterium. In order to do this it must span the bacterial inner membrane surrounding the cytoplasm, and the outer envelope that encloses the periplasm and acts as a barrier to the environment (Mecsas and Strauss 1996).

The components of the secretion systems are synthesised in the cytoplasm and secreted across the membranes by a nano-machine (Mota *et al.* 2005), which can be visualised as an ATP driven molecular syringe that serves to inject bacterial proteins into target cells. The systems operate in different ways and the translocation apparatus varies in complexity and design as illustrated in Figure 1.4. The protein secretion systems themselves are well conserved amongst bacteria, whereas the specific properties of the effectors and hence the resulting symptomatic effects on the host organism vary widely

The type I system secretes proteins across both the inner membrane and cell envelope in one operation, transporting them directly from the cytoplasm to the bacterial cell surface. In contrast, the type II system utilises the cell's general secretory pathway to transport the proteins from the cytoplasm to the periplasm, and from there uses a dedicated protein channel to transport them to the outer surface of the cell envelope. Type I is a one step process whereas type II is a two step process.

The type III protein secretion system (TTSS) is present in many gram-negative bacteria and, like type I, also functions in a single step manner to transport proteins directly from the cytoplasm to the cell surface where they can interact with the target cell membrane. It is interesting to note that the genes encoding many components of the TTSS are homologous to those that encode flagella export machinery in both gram-negative and gram–positive bacteria. A schematic diagram of the type I, II and III secretion systems is shown in Figure 1.4.

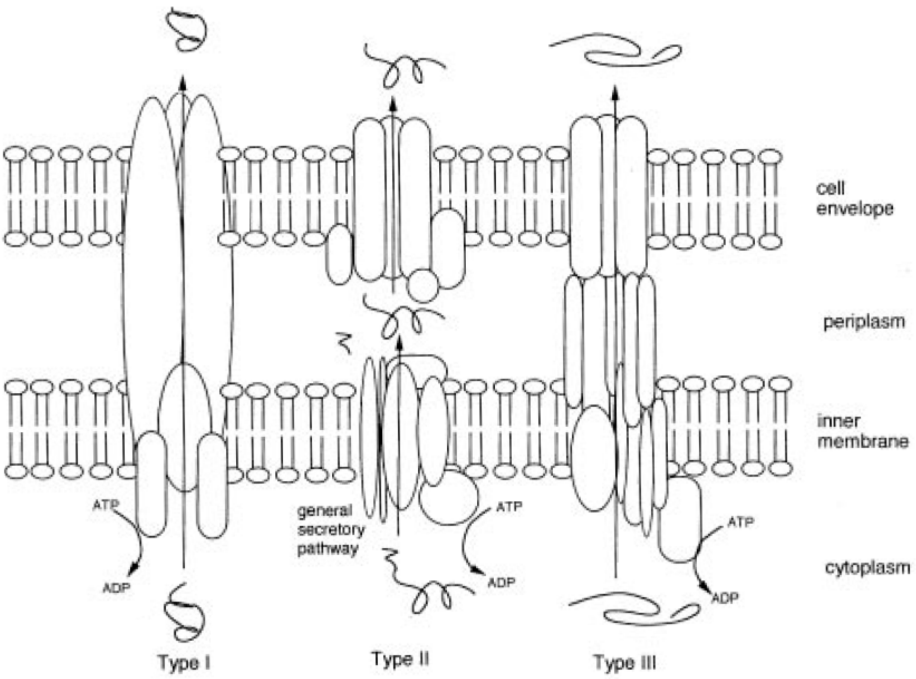


Figure 1.4: Schematic diagram of type I, type II and type III secretion systems
Type I and type III secrete proteins across both the inner membrane and cell envelope, whereas type II does this in two steps with the proteins stopping intermediately in the periplasm (Mecsas and Strauss 1996).

The type IV and type V protein secretion systems also consist of a single structure crossing both the bacterial inner membrane and the cell envelope. Of the secretion systems, type IV is considered the most versatile because it mediates conjugative DNA transfer, uptake and secretion of DNA, as well as the export of effector molecules (Schroder and Lanka 2005). Proteins secreted by the type V secretion system are referred to as autotransporters because they are translocated across the outer membrane via a transmembrane pore formed by a ß-barrel and they contain all the information required for translocation through the cell envelope. Typically, autotransporters are expressed as precursor proteins with three basic functional domains, including an N-terminal signal peptide, an internal passenger domain, and a C-terminal translocator domain (the beta-domain) (Henderson *et al.* 2004).

A strong indication that the TTSS is involved in the virulence of *B. pseudomallei*, through the translocation of effector proteins, comes from the fact that it is present in this bacterium but absent from the closely related avirulent *B. thailandensis* (Rainbow *et al.* 2002).

1.1.5 The Type III Protein Secretion System

Studies indicate that a central component of the TTSS nano-machine is the needle complex or injectisome. This organelle is comprised of three distinct parts; a cylindrical base, similar to the flagella basal body, consisting of two pairs of rings that span the inner membrane and cell envelope, joined together by a inner rod that is physically linked to a hollow, elongated, and stiff needle-like structure as shown in Figure 1.5 (Marlovits *et al.* 2004).

Upon contact of the bacterium with the target cell a signal is sent to the injectisome, which then secretes the translocator and effector proteins across the membranes to the area of contact between the two cells (Pettersson *et al.* 1996). The translocators somehow form a pore in the lipid membrane of the target cell through which the effector proteins are able to pass (Blocker *et al.* 2000). It is thought chaperones are present in the cytoplasm of the bacterial cell to ensure the delivery of the translocator and effector proteins to the injectisome when required (Mecsas and Strauss 1996).

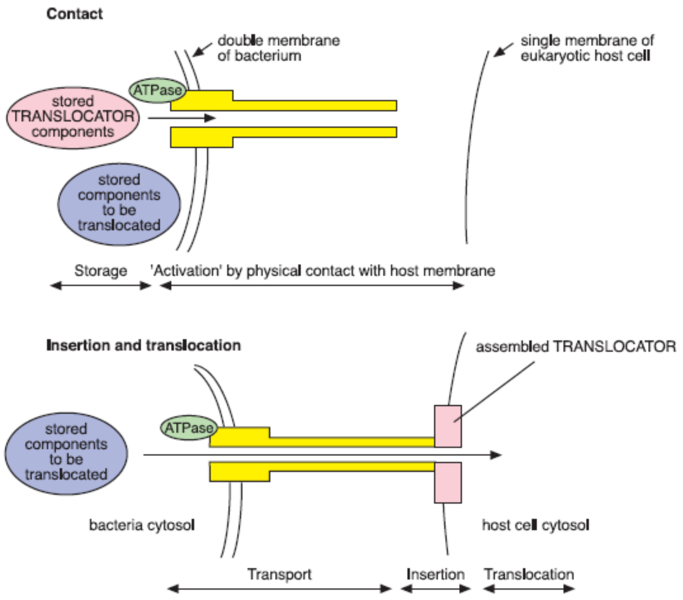


Figure 1.5: Bacterial secretion system injectisome
The needle, especially, displays great sequence and structural homology to the proteins of the flagella apparatus. It is generally thought that the injectisome serves as a tube through which the secreted proteins travel (Mota *et al.* 2005). The needle is formed through the polymerization of a major subunit as shown in Figure 1.6; BsaL in *B. pseudomallei*, MxiH in *Shigella* and Prg1 in *Salmonella* (Sekiya *et al.* 2001). The length of the needle varies between 45nm - 80nm depending on bacterial species and is approximately 70Å in diameter, traversed by a central channel 20–30-Å wide.

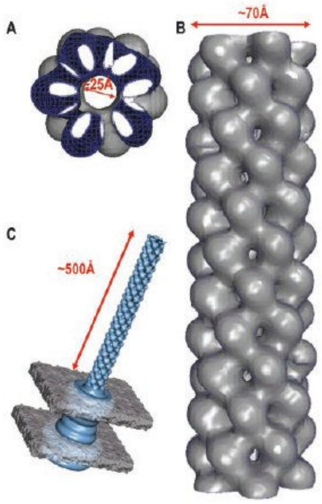


Figure 1.6: Needle of the TTSS
The needle has a central pore through which effector proteins can be transported from the bacterial cytoplasm to the external surface of the bacterial outer membrane.

These effector proteins commonly mimic eukaryotic enzymes in their function or structure, and may directly interact with cellular proteins (Stevens *et al.* 2004). It appears that often they are toxins and enzymes (kinases, phosphatases, G-proteins etc.) that are able to interfere with the signal transduction pathways of the host and ‘reprogramme’ it to facilitate bacterial infection. This is achieved, for instance, by endocytosing the bacterium, evading phagocytic cells that would otherwise recognise and destroy the bacterium, suppressing the production of proinflammatory cytokines and chemokines, inhibiting an adaptive immune response, regulating the actin cytoskeleton, anti-apoptosis and more (Viboud and Bliska 2005).

Obviously variations exist between the TTSS of different bacterial species and this is because it is thought that the pathogenicity islands encoding the TTSS gene sequence in one bacterium were acquired independently of another as no ‘common bacterial ancestor’ has been identified. Consequently the proteins involved in the TTSS of different species of bacteria differ greatly, both in number and structure, but the crucial proteins in each TTSS appear to share substantial homology (Mecsas and Strauss 1996).

1.1.6 Type III Protein Secretion System in *Burkholderia pseudomallei*

It has been found that *B. pseudomallei* contains at least 3 loci encoding putative TTSS (Rainbow *et al.* 2002), one of which shares homology with the inv/spa/prg TTSS of *Salmonella typhimurium* (Attree and Attree 2001) and the ipa/mxi/spa TTSS of *Shigella* flexneri (Stevens *et al.* 2002). This gene cluster has been designated Bsa (Burkholderia secretion apparatus) in keeping with the proposed nomenclature (Hueck 1998). The other two TTSS in *B. pseudomallei* are homologous to the hrp locus of the plant pathogen *Ralstonia solanacearum*. It is of interest to determine if these two systems are involved in pathogenic and/or symbiotic interactions with plants and the rhizosphere, as such interactions could influence the persistence of *B. pseudomallei* in the soil and aquatic environments (Stevens *et al.* 2004).

1.1.7 BipD Homologues

Within the *B. pseudomallei* TTSS that is homologous to those of *Shigella* and *Salmonella*, the effector proteins are termed Bop proteins and the translocators termed Bip; Burkholderia invasion protein (Stevens *et al.* 2004;Diaz 2006), as shown in Figure 1.7.

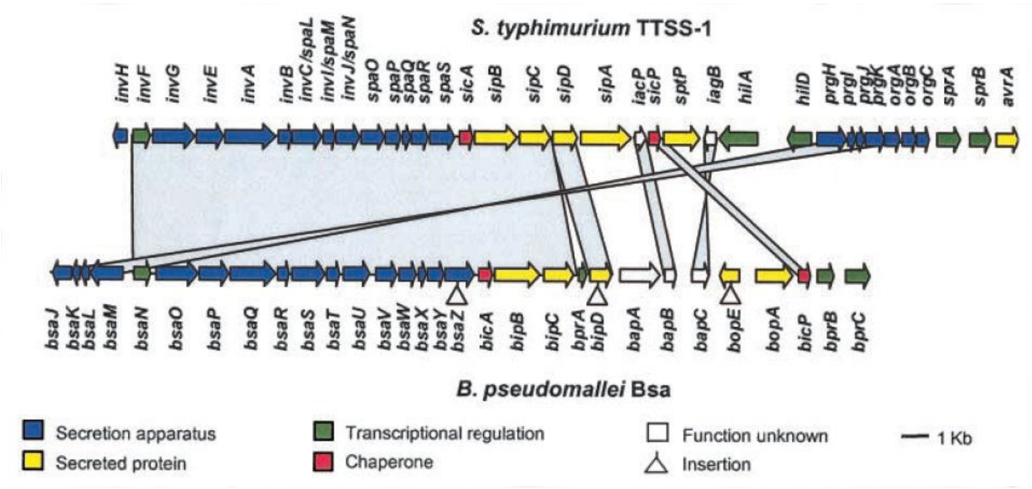


Figure 1.7: The relative loci of type three section systems
Cognate type III secretion loci in *B. pseudomallei* and *S. typhimurium*, showing the homologous translocator and effector proteins. Only part of the *B. pseudomallei bsa* putative type III secretion locus is shown (adapted from Stevens et. al. 2003).

These effector and translocator proteins appear to be homologous in structure and function, although they share relatively low sequence similarity and identity as illustrated in Table 1.

identity / **similarity**		*Burkholderia pseudomallei*					*Shigella flexneri*				*Salmonella typhimurium*			
		BipB	BipC	BipD	BopA	BopE	IpaB	IpaC	IpaD	IcsB	SipB	SipC	SipD	SopE
Burkholderia pseudomallei	BipB	100	-	-	-	-	48	-	-	-	50	-	-	-
	BipC	-	100	-	-	-	-	34	-	-	-	38	-	-
	BipD	-	-	100	-	-	-	-	36	-	-	-	39	-
	BopA	-	-	-	100	-	-	-	-	43	-	-	-	-
	BopE	-	-	-	-	100	-	-	-	-	-	-	-	37
Shigella flexneri	IpaB	30	-	-	-	-	100	-	-	-	60	-	-	-
	IpaC	-	17	-	-	-	-	100	-	-	-	48	-	-
	IpaD	-	-	22	-	-	-	-	100	-	-	-	53	-
	IcsB	-	-	-	26	-	-	-	-	100	-	-	-	-
Salmonella typhimurium	SipB	32	-	-	-	-	40	-	-	-	100	-	-	-
	SipC	-	21	-	-	-	-	30	-	-	-	100	-	-
	SipD	-	-	24	-	-	-	-	35	-	-	-	100	-
	SopE	-	-	-	-	23	-	-	-	-	-	-	-	100

Table 1: Sequence identity and similarity of known homologues of TTSS
The effectors and translocators from *B. pseudomallei*, *S. typhimurium* and *S. flexneri* are compared here. A NeedleP global alignment with an Eblosum62 matrix from the EMBL website http://srs.ebi.ac.uk was used for the analysis. For sequence data see appendix.

1.1.8 The Role of BipD within the TTSS of *Burkholderia pseudomallei*

It appears that BipD performs several roles within the TTSS of *B. pseudomallei*. It acts as a plug to prevent the leakage of effector proteins from the needle, therefore mediating their secretion on to the surface of a host cell (Menard *et al.* 1994). The location of BipD at the tip of the needle implies it may also play a role in sensing contact with potential target cells and sending a message to the bacterium (Espina *et al.* 2006). BipD may also assist the other translocator proteins form a pore in the plasma membrane of a host cell (Blocker *et al.* 2000) and has been putatively associated with affecting actin cytoskeleton rearrangement to allow *B. pseudomallei* escape from an endocytic vesicle once it has entered the cell (Stevens *et al.* 2002). In *Salmonella*, SipB, SipC and SipD have been shown to be required for the injection of effector proteins and the invasion of epithelial cells in vitro (Kaniga *et al.* 1995), and it is thought that the *B. pseudomallei* homologues BipB, BipC and BipD perform a similar function.

Although the *Salmonella* and *Shigella* TTSS have been studied in some detail, less is known about the Bsa TTSS of *B. pseudomallei*, therefore much of our current theories regarding the method by which this TTSS functions has come from studying these homologous TTSS. It appears the TTSS secretes translocator and effector proteins upon contact with the host cell (Pettersson *et al.* 1996). The translocators then facilitate the delivery of the effector proteins across the plasma membrane where they somehow bring about membrane ruffling to allow the engulfment of the bacterium (Stevens *et al.* 2002).

Experiments have shown that strains of *B. pseudomallei* containing a defined BipD deletion mutant exhibit impaired invasion of HeLa cells (Stevens *et al.* 2002;Stevens *et al.* 2004). Once the bacteria have been engulfed by the cell, the endosome is lysed and polymerisation of actin is triggered, possibly by an interaction between the effector and translocator proteins, which then enables the bacterium to move fully into the host cell where it can subvert the normal cellular process for its growth and replication (Stevens et al., 2004).

It is also believed that BipD is somehow involved in this actin tail formation that facilitates the escape of *B. pseudomallei* from endocytic vesicles. The Ipa/Mxi/Spa TTSS of *Shigella* has been shown to be necessary for lysis of the endosome membrane, particularly the secreted proteins IpaB and IpaC (High and Dobberstein 1992;Fernandez-Prada *et al.* 2000). The fact that *B. pseudomallei* contains a homologous TTSS implies that it may use a similar mechanism to escape from endosomes. Consistent with this idea, it has been reported that *B. pseudomallei* mutants lacking putative components of the Bsa secretion and translocation apparatus (BsaZ and BipD, respectively) are confined to endosomes with intact membranes following infection of J774.2 murine macrophage-like cells. It was also noted that no actin induced membrane protrusions were observed, but the growth rate of the cells was not affected (Stevens *et al.* 2002). Observations made during these experiments support the idea that a functional Bsa TTSS is required for full virulence of *B. pseudomallei*.

1.1.9 BipD may Prevent the Leakage of Effector Proteins from the Needle

It is not yet understood how the needle of a TTSS is able to detect the presence of a target cell and then send this signal to the base of the injectisome in the bacterial cytoplasm. Several laboratories have shown that cell contact by most animal pathogens expressing a functional TTSS is a prerequisite for increased production and subsequent delivery of anti-host-effector proteins across host cell plasma membranes (Holden *et al.* 2004).

IpaD (invasion plasmid antigen) a BipD homologue, has recently been shown to exist at the tip of the MxiH needle in *Shigella flexneri* (Espina *et al.* 2006). A key role for IpaD has been identified at the tip of the needle, where it is proposed to be essential for mediating the correct insertion of the IpaB / IpaC translocon into host cell membranes (Picking *et al.* 2005).

IpaD is perhaps the best studied homologue of BipD; a crystal structure has been solved and a model for an IpaD / IpaB hetero-pentamer at the tip of the MxiH needle was recently proposed, based on crystal contacts in one of the crystal forms (Johnson *et al.* 2006). This model, shown in Figure 1.8, involves the export of four copies of IpaD to the tip of the needle, where they pack into the needle via C-terminal residues. It is thought that a copy of IpaB then slots in via its central coiled-coil, locking the pentamer and preventing further secretion or leakage of other effectors. Earlier experiments have shown that the ability of IpaD to control effector protein secretion can be separated from its ability to aid cellular invasion (Picking *et al.* 2005). This implies that IpaD has more than one functional domain in order to achieve these functions and it is thought that BipD may also.

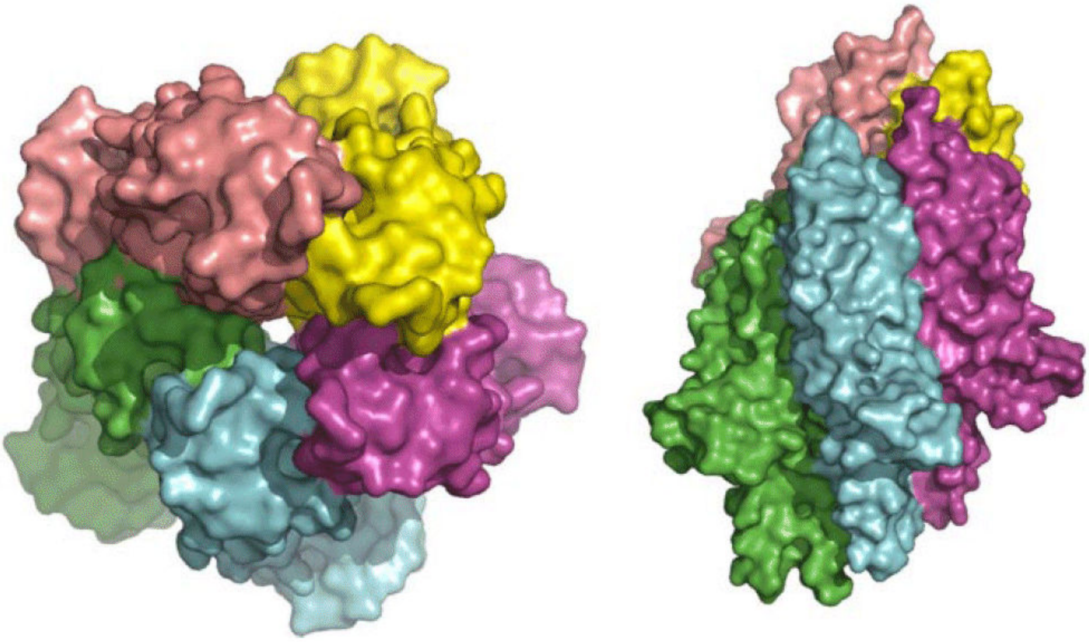


Figure 1.8: A proposed pentameric structure for IpaD
This was prepared using the non-crystallographic symmetry from crystal form 2. The model on the left shows a ribbon representation and to the right the bottom surface view (Johnson et al. 2006). Here five molecules of IpaD form a pentamer, but it has been suggested that one monomer of IpaB may replace the IpaD monomer shown in green.

Although IpaD is homologous to BipD, structural differences have been observed between them, implying that caution should be used when comparing the two proteins too closely. For example, it is known that IpaD possesses two independently folding domains (Espina *et al.* 2006;Johnson *et al.* 2006) but there is no evidence that BipD has multiple folding domains. Observations of the conformational stability and differential structural analysis have led some researchers to claim that IpaD and SipD possess properties that distinguish them from the other tip proteins of TTSS (Espina et al., 2006).

1.1.10 How Does BipD Fulfill its Other Functions?

Many questions still exist as to how BipD mediates the secretion of effector proteins from the needle, what role (if any) it plays in orchestrating the insertion of other translocator proteins into the plasma membrane to form a pore through which the effector protein can pass, and when and how it affects the actin cytoskeleton. Hopefully by elucidating the structure of BipD we will be able to glean more information about the exact role(s) of BipD within the TTSS and how it fulfils its various indicated functions.

1.2 2,4'-Dihydroxyacetophenone Dioxygenase

1.2.1 Aromatic Compounds

Aromatic compounds are molecules whose molecular structure includes one or more benzene rings; that is, a planar ring of six carbon atoms connected by delocalised electrons (Evans and Fuchs 1988). Many aromatic compounds occur naturally in the environment such as lignin, the major constituent of woody tissue that comprises approximately 25% of the earth's land-based biomass. However, the majority of aromatic compounds are synthetically produced, especially the aromatic hydrocarbons (Diaz *et al.* 2001), making them alien to the biosphere and potentially hazardous to the environment.

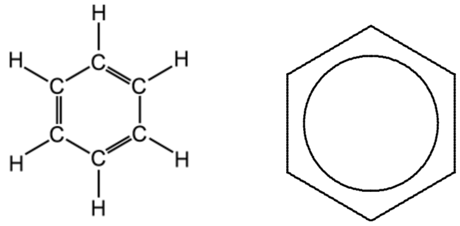


Figure 1.9: A benzene ring consists of 6 carbons These are connected by delocalised electrons as though they were alternating single and double bonds. The right hand diagram is a schematic representation of the delocalisation of electrons.

Aromatic compounds are commonly formed during the incomplete burning of coal, oil, gas and garbage as well as petroleum, solvents, pesticides and plastics. These compounds, although unsaturated, do not undergo normal reactions for linear unsaturated compounds. This is related to the way in which their bonding electrons are spread evenly over the planar ring. In addition to this, the thermodynamic stability of the benzene ring, shown in Figure 1.9, further increases its persistence within the biosphere and as such, aromatic hydrocarbon contamination of the environment has always been a big problem for industrialised countries in particular (Diaz 2006). Because of the wide range of useful substances derived from aromatic compounds, there is a huge global market for these chemicals that is increasing rapidly, therefore so too is the environmental burden they continue to pose.

Naturally occurring aromatic hydrocarbons such as lignin can be catabolised by microorganisms (Diaz *et al.* 2001). This general ability of bacteria to catabolise such compounds is a result of the presence of aromatic chemicals in the environment from the recycling of plant-derived material (Diaz 2006) and bacteria have evolved to utilise them as a carbon source. This ability of bacteria to catabolise aromatic compounds can extend to simple industrially produced chemicals if they are structurally similar to a natural substrate. In fact biodegradation of hydrocarbons by natural populations of microorganisms represents one of the primary mechanisms by which petroleum and other hydrocarbon pollutants are currently eliminated from the environment. However, although ubiquitous in both terrestrial and marine environments, the fraction of the total bacterial population represented by hydrocarbon-utilising bacteria varies dramatically between locations (Leahy and Colwell 1990).

In addition, individual bacteria are only able to degrade a limited range of hydrocarbons, therefore the presence of mixed populations representing a broad spectrum of enzymatic capabilities is necessary to degrade complex mixtures of hydrocarbons (Leahy and Colwell 1990). In situations where the aromatic compounds cannot be fully degraded by microorganisms, these toxic chemicals accumulate in the environment where they present a serious health risk to both humans and wildlife.

1.2.2 Bioremediation

Although these alien hydrocarbons can be removed from the environment by natural biological and physio-chemical processes, this is generally a slow and unpredictable method of counteracting the damage caused by the massive increase in human release of aromatic compounds into the biosphere (Diaz *et al.* 2001). In recent years bioremediation, that is the deliberate use and manipulation of detoxification abilities of living organisms, has become a useful and more commonly utilised method of removing harmful aromatic compounds from both terrestrial and aquatic environments, Figure 1.10 (Diaz 2006).

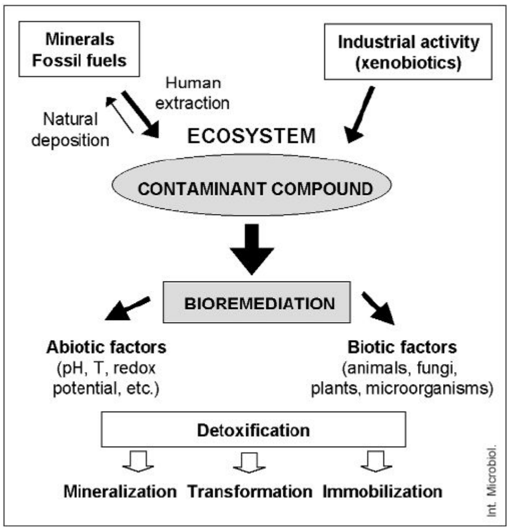


Figure 1.10: Microbial degradation and bioremediation
The main sources of environmental pollution arise from industrial activity and the combustion or accidental release of fossil fuels. Microbial degradation of these aromatic compounds forms only part of the bioremediation process (Diaz 2006).

Although most organisms possess detoxification abilities of one kind or another, such as transformation, mineralization and immobilisation of pollutants, microorganisms, and particularly bacteria, have been the most frequently used for bioremediation strategies and are the best studied (Diaz 2006). The combination of rapid growth rate, global abundance and high rate of mutations enables bacteria to evolve quickly and adapt to their ever changing environment, making them an ideal species for study and genetic manipulation (Diaz 2006). It is hoped that a better understanding of microbial degradative pathways will enable us to use the enzymes involved to greater effect, or even manipulate them to encompass degradation of chemicals that are not natural substrates, thereby greatly reducing the negative environmental impact of aromatic compounds.

In order to engineer a microbe that is capable of degrading a foreign substrate of our choosing, it is not possible simply to change a single enzyme. Molecular pathways are so intricately balanced and the enzymes involved so inter-dependent on one another that in order to equip a microorganism with the ability to degrade an aromatic compound that is not its natural substrate, rational and knowledgeable manipulation of both the host cell and the specific catabolic pathway involved is required (Diaz 2006). It is for this reason that it is not enough to simply know about individual enzymes in isolation, knowledge about the way in which they interact with their environment and the molecules around them is vital. A deeper understanding of different metabolic pathways could ultimately enable the engineering of novel hybrid pathways through the amalgamation of different sections from several well understood pathways, resulting in catabolic expansion to new substrates and the construction of new routes of degradation (Diaz 2006).

1.2.3 Microbial Methods of Degradation

The benzene ring is the second most widely distributed unit of chemical structure in nature after glucosyl residues and bacteria have developed strategies for obtaining energy from many of these aromatic compounds (Diaz 2006), under either aerobic or anaerobic conditions (using alternative final electron acceptors such as nitrate, sulphate, and ferric ions). It is a common feature of both aerobic and anaerobic degradation of aromatic compounds that structurally diverse compounds are degraded through many peripheral pathways to a relatively small number of intermediates, which can be further channelled via a few main pathways to the central metabolic machinery of the cell (Diaz 2006).

In order for microorganisms to degrade aromatic compounds they need to be able to break the carbon-carbon bonds within the molecule. The method by which this is achieved in the peripheral pathways varies depending on the level of oxygen present. Under anaerobic conditions bacterial enzymes catalyse the reductive hydrogenation of the ring, followed by the fragmentation of the cyclohexane ring skeleton but, under aerobic conditions the strategy used is oxidation of the ring followed by oxidative cleavage (Bugg and Lin 2001). The enzymes that catalyse the aerobic reactions are termed oxygenases.

1.2.4 Oxygenases

An oxygenase involved in the aerobic catabolism of aromatic rings in bacteria follows one of two paths: either it performs the first phase of degradation by preparing the aromatic compounds for ring cleavage (Karlsson *et al.* 2002) or they are involved with the second phase of degradation, where they carry out the ring-fission reactions themselves (Kita *et al.* 1999). There are 2 main classes of oxygenase: mono-oxygenases and dioxygenases (determined by whether they introduce an atom or a molecule of oxygen respectively), and it is the latter that are actually involved in the ring-fission reactions, although some pathways also use dioxygenases to catalyse the insertion of two hydroxyl groups into the aromatic ring to produce a suitable ring-fission substrate.

1.2.5 Dioxygenases

Metabolism of aromatic rings by dioxygenases involves the introduction of two oxygen atoms into the substrate. Experiments using heavy isotopes (O_2^{18}) have shown that both oxygens come from a molecule of O_2, rather than from O_2 and H_2O (Hayaishi *et al.* 1955).

1.2.6 Cofactors

Cofactors, often Fe^{2+} or Fe^{3+}, are required by many dioxygenases to act as catalysts, enabling the enzyme to reach a high enough energy state to carry out the desired reaction. This is due to the fact that the ground state of dioxygen ($^3O_2$), which contains two unpaired electrons in the highest occupied π^* orbitals, is spin-forbidden to react with spin-paired singlet species (Bugg 2003). To reach the singlet excited state of dioxygen ($^1O_2$), which contains a pair of valence electrons and is highly reactive towards alkenes and dienes, an additional 22 kcal/mol energy is required. The dioxygenase is unable to access this excited state without a cofactor.

1.2.6.1 Mechanism of Dioxygen Activation

It is believed that there are three methods by which dioxygen can be activated; orbital overlap, single electron transfer and reaction with a substrate radical.

Orbital Overlap with a metal ion occurs when a dioxygen complexes using unpaired π^* orbitals to unpaired 3d orbital electrons of a transition metal. This enables the reaction of a transition-dioxygen with a singlet organic reagent. However this can only occur if the overall number of unpaired electrons in the complex remains constant throughout.

Single Electron Transfer is believed to occur due to transition metals found in dioxygenases and indeed many metallo-enzymes having two consecutive available oxidation states, e.g. Fe^{2+}/Fe^{3+}, thus enabling the metal centre to transfer a single electron to bound dioxygen. The $^3O_2$ oxygen ground state forms a super-oxide (O_2^-) if an additional electron is added to its electron cloud, which is then able to perform several 1^- or 2^- electron chemical reactions; this process is seen in catechol dioxygenases.

Organic reagents are also able to act as stable radical intermediates in single electron transfer reactions. For example, reduced flavin co-factor found in flavoprotein hydroxylase enzymes is able to activate dioxygen via a single electron transfer forming a stable flavin semiquinone and superoxide, which then complex to form a hydroperoxide intermediate (Bugg 2003), see Figure 1.11.

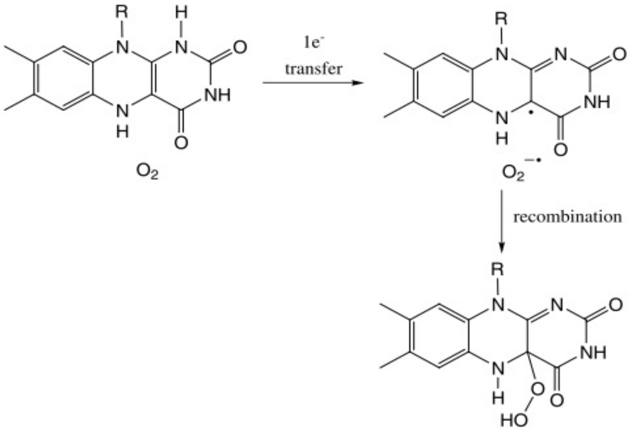


Figure 1.11: Activation of dioxygen by reduced flavin
Single electron transfer to dioxygen can also take place from organic reagents which can access a stable radical intermediate e.g.a reduced flavin cofactor is able to activate dioxygen via a single electron transfer to form a stable flavin semiquinone and superoxide, which then recombine to form a hydroperoxide intermediate. Adapted from Bugg 2003.

Reaction with a Substrate Radical: has been proposed in intradiol catechol dioxygenases (Que, Jr. and Ho 1996), where a bound catechol semiquinone intermediate attacks dioxygen to form a hydroperoxide radical, see Figure 1.12.

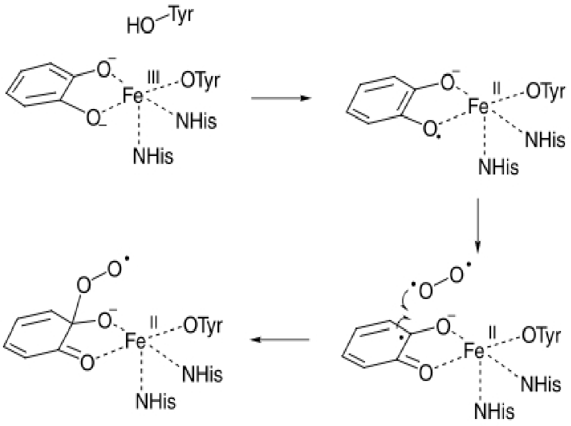


Figure 1.12: Proposed activation mechanism for intradiol catechol cleavage
Since the reaction of dioxygen via radical mechanisms is a spin-allowed process, reaction with a substrate radical is an alternative possible mechanism. It has been proposed that a substrate activation mechanism of this kind occurs in the intradiol catechol dioxygenases, where a bound catechol semiquinone intermediate attacks dioxygen to form a hydroperoxide radical. Adapted from Bugg 2003.

1.2.6.2 Catechol Dioxygenases

The hydroxyl groups of a compound can take various different conformations as shown in Figure 1.13, which determines how they are cleaved by dioxygenases.

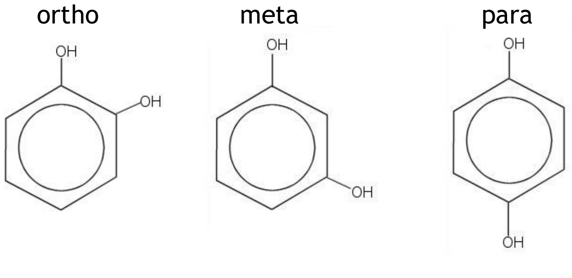


Figure 1.13: Three isomeric forms of a dihydroxybenzene ring:

ortho indicates that the substituents are adjacent e.g. at positions 1 and 2.

meta indicates that the substituents are e.g. at positions 1 and 3.

para indicates that the substituents are opposite e.g. at positions 1 and 4.

The substrate is oxygenated by the dioxygenase to form an intermediate dihydroxy aromatic compound, the ring of which is then cleaved proximal to one of the two hydroxyl groups (meta cleavage catalysed by extradiol dioxygenases) or between the two hydroxyl groups (ortho cleavage catalysed by intradiol dioxygenases) (Bugg and Lin 2001) Figure 1.14.

The intradiol dioxygenases require Fe^{3+} as a non-haem cofactor and cleave the C-C bond between phenolic hydroxyl groups in the ortho position. The extradiol dioxygenases require Fe^{2+} as a cofactor and cleave the C-C bond adjacent to the phenolic hydroxyl groups in the meta position.

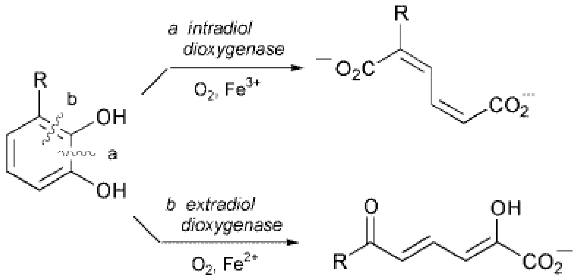


Figure 1.14: Cleavage mechanism
The mechanism of action of intradiol (cleavage occurs between the two hydroxyl groups) and extradiol (cleavage occurs adjacent to either of the two hydroxyl groups) dioxygenases. Adapted from Bugg 2003.

It appears that the vast majority of metal-based complexes designed for catechol oxidative cleavage perform intradiol cleavage, therefore it is likely that this is the kinetically favoured pathway under most conditions. Extradiol reactions have proved much harder to reproduce in the laboratory, and appear to require facial tridentate ligand binding of the catechol mono-anion, and the involvement of a proton donor, as well as being selective for Fe^{2+}. This combination of requirements makes them less favourable in nature, as well as making them increasingly difficult to study *in vivo* (Bugg 2003).

1.2.6.3 Cofactor Binding Sites

The extradiol dioxygenase BhpC was purified and crystallised by Han et al. in 1995. The structure obtained shows a funnel-shaped cavity between two domains leading to the active site where the Fe^{2+} cofactor was bound. The cofactor ligation was found to be performed by three amino acids: His 146, His 210 and Glu 260 (His_2Glu) (Han *et al.* 1995). Whilst several other extradiol dioxygenase structures have been solved in recent years there is often limited similarity between their sequences, however the active site features are always very similar, and often conserved Fe^{2+} ligands are used with the His_2Glu motif being found in most extradiol dioxygenases to date as well as in some non-haem Fe^{3+} enzymes (Miller and Lipscomb 1996;Lin *et al.* 2001).

X-ray crystallography has shown that each family of ortho pathway dioxygenases have a specific set of amino acids for each co-factor. In intradiol dioxygenases the ligands that bind the co-factors have been determined, with the non-haem Fe^{3+} ligated by four amino acid side chains: two imidazole side chains of histidine residues, and two phenolic side chains of a pair of tyrosine residues (His_2Tyr_2). An H_2O ligand was also found to create a five point trigonal bipyramidal structure (Lin *et al.* 2001).

1.2.7 2,4'-Dihydroxyacetophenone Dioxygenase

2,4'-dihydroxyacetophenone dioxygenase (DAD) was discovered when a strain of the gram-negative, aerobic soil bacterium *Alcaligenes* was grown on a medium of 4-hydroxyacetophenone (Hopper 1986). It is a homotetramer, with subunits each of 20.3 kDa, and it contains iron. This iron is of unknown redox state (although the enzyme is colourless), and it is not lost during prolonged dialysis suggesting that it is tightly bound. DAD was first purified and sequenced by Hopper et al in 1999, revealing a significant difference in sequence from other enzymes of its type (Hopper and Kaderbhai 1999). The *dad* gene was cloned and found to encode 177 amino acids including 2 cysteines, although it is not known if they form a disulphide bond in the tertiary structure. The sequence is shown below.

```
Met V A N A I S E F W R D I R P I E S P F K P D A L P E A Y I
P N A A T E D E R Y Y V P F T E T V A S R P L W I S P Q Q N R
W C D I L L A R E A G L V N R H Y H P H E V F A Y T I S G K W
G Y L E H D W T A T R G D F V Y E T P G E G H T L V A F E H E
E P Met R V F F I V Q G P L I W L D E A G N S I G H F D V H D
Y I A Met C R E H Y E K V G L G A D L V V T L F R Stop
```

This enzyme is of particular interest because its mechanism of action is unlike that of any known dioxygenase and it bears little sequence homology to other dioxygenases (Hopper and Kaderbhai 1999). What makes it different is the fact that, unlike all other dioxygenases, which introduce two atoms of molecular O_2 into the substrate and then cleave the aromatic ring, DAD introduces two atoms of molecular O_2 into 2,4'-dihydroxyacetophenone and cleavage then occurs at the side-chain (Figure 1.15).

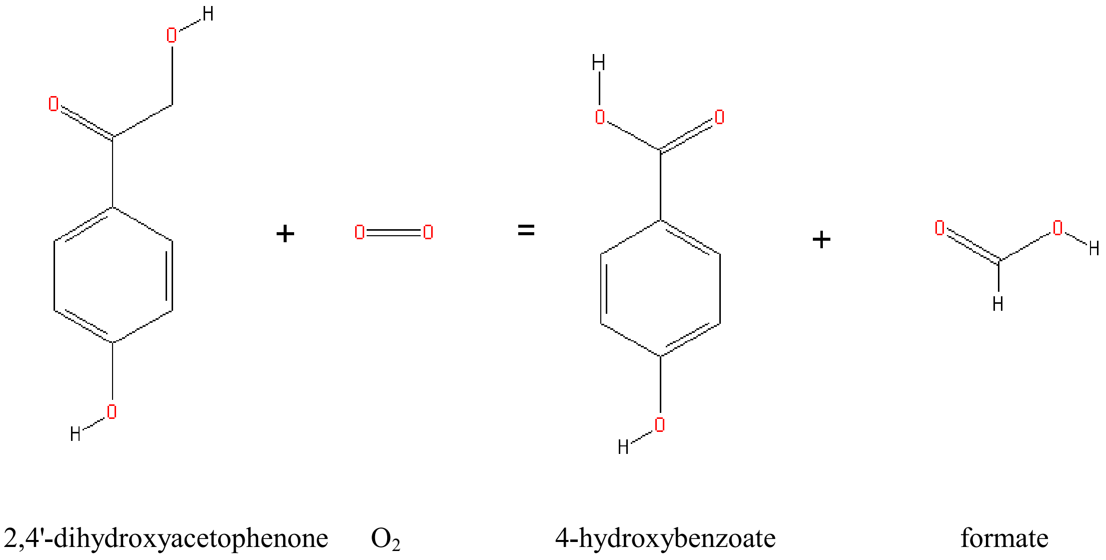


Figure 1.17: 2,4'-dihydroxyacetophenone dioxygenases reaction mechanism
2,4'-dihydroxyacetophenone dioxygenase is classed as such because it inserts 2 atoms of molecular oxygen into its substrate, 2,4'-dihydroxyacetophenone. This enzyme is unique amongst all known dioxygenases because cleavage occurs, not at the aromatic ring as normal, but at the side-chain.

This is an alternative pathway to that utilised by microorganisms such as *Pseudomonas*, *Arthobacter* and *Nocardia*. They use mono-oxygenases to metabolise aromatic ketones such as 4-hydroxyacetophenone by inserting O_2 into the substrate, forming an ester that is then hydrolysed (Darby *et al.* 1987;Cripps *et al.* 1978).

1.2.7.1 Sequence and Structural Similarity

The *dad* gene was sequenced and several reading frames were found to be highly homologous to those in various *Pseudomonas* bacteria, but the functions of the corresponding proteins are not known. However, it was also found that DAD shares sequence homology with cupin 2, a conserved barrel domain of the 'cupin' super family. This suggests that DAD, as a homo-tetramer, may form a β-barrel structure.

Threading studies (Jones 2001) were carried out by G. Beaven, which found weak sequence homology (~20 %) between DAD and the ferredoxin-like domains of several Rieske dioxygenases whose structures are known, the homology being strongest with naphthalene 1,2-dioxygenase (Karlsson *et al.* 2002). It was found through size comparison that DAD corresponded with the [Fe2-S2] cluster of this enzyme. However, it is known that this component functions solely to facilitate electron transfer and has no role in catalysis, raising the question of how DAD can catalyse a reaction if it consists almost exclusively of a structure with no catalytic function.

It has been proposed that DAD may form ferredoxin-like folds, similar to those found in ActVA-Orf6 monoxygenase from *Streptomyces coelicolor*, and two novel haem degrading enzymes from *Staphylococcus aureus*, IsdI and IsdG for which structural models have been determined (Wu *et al.* 2005) and a structural analysis is shown in Figure 1.16. A ferredoxin fold is a common αβ protein fold with a signature βαββαβ secondary structure along its backbone. Structurally, the ferredoxin fold can be regarded as a long, symmetric hairpin that is wrapped once around, so that its two terminal β-strands hydrogen-bond to the central two β-strands, forming a four-stranded, antiparallel β-sheet covered on one side by two α-helices. ActVA-Orf6 is part of a small class of mono-oxygenases that catalyse the oxygenation of their substrates without the assistance of any prosthetic groups, metal ions or cofactors normally associated with the activation of molecular oxygen. The crystal structure revealed the overall structure of ActVA-Orf6 to be a ferredoxin-like fold with a novel homodimeric assembly, indicating this structural configuration may function in an additional manner to that already known. Hydrophobic interactions from the ß-sheets stabilise the interface between the two subunits.

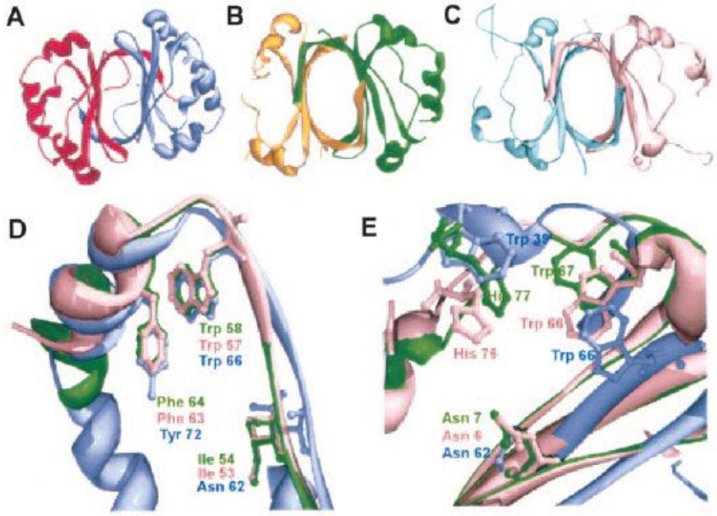


Figure 1.18: Structural analysis of IsdG and ActVA mono-oxygenase families
A, ActVA-Orf6 homodimer with subunits shaded in blue and red. B, IsdG homodimer with subunits shaded in orange and green. C, IsdI homodimer with subunits shaded in aqua and pink. D, residues predicted to be involved in binding and oxidation of the substrate in ActVA-Orf6 (blue) that are conserved in IsdG (green) and IsdI (pink). E, residues shown to be required for oxidation of haem in IsdG (green) and the equivalent residues in IsdI (pink). ActVAOrf6 residues that occupy a similar space as the catalytic residues of IsdG are shown in blue.

ActVA-Orf6 has been shown to be structurally similar to IsdI and IsdG, although they show no significant sequence similarity to any other known haem oxygenases. Their crystal structures (Wu *et al.* 2005) also show each subunit to form a ferredoxin-like fold, suggesting a way by which an enzyme formed almost entirely of this fold can be catalytically active.

Sequence alignments were conducted by G. Beaven using CLUSTAL-W, showing that DAD shares only 18.9% identity to ActVA-Orf6 and 15.1% to IsdG. However, despite this low sequence identity to one another and DAD, structurally those crystal structures appear quite similar. It is therefore possible that DAD forms similar ferredoxin folds within the tetrameric form.

It is hoped that by determining the three dimensional structure, through the use of X-ray crystallography, of both the native form of DAD and its complex with a substrate analogue, we will gain an improved understanding of its unusual catalytic mechanism and how this is determined by its structure.

1.3 L-threonine Dehydrogenase

1.3.1 Amino Acids

Amino acids are the building blocks of protein and there are 20 standard amino acids required to build the many proteins used in the growth, repair, and maintenance of our body tissues. These amino acids are biosynthesized from other molecules, but organisms differ in which ones they can synthesise and which ones must be absorbed through their diet. Those that cannot be synthesized by an organism are given the name essential amino acids. For humans, the essential amino acids are isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine and these are obtained through our diet. Another amino acid, histidine, is considered semi-essential because the body does not always require dietary sources to obtain it as it is able to synthesise small amounts. The remaining amino acids are alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, proline, serine, and tyrosine which we can synthesise.

Foods of animal origin, such as meat, poultry, fish, eggs, and dairy products are the richest dietary sources of the essential amino acids; plant proteins have an amino acid composition that is generally nutritionally less favourable than animal proteins. Metabolism of amino acids is determined by the proportion of amino acids required for protein synthesis (Friedman 1996). Oxidation rates of amino acids are low until the amount consumed exceeds the amount needed for protein synthesis; oxidation then increases rapidly.

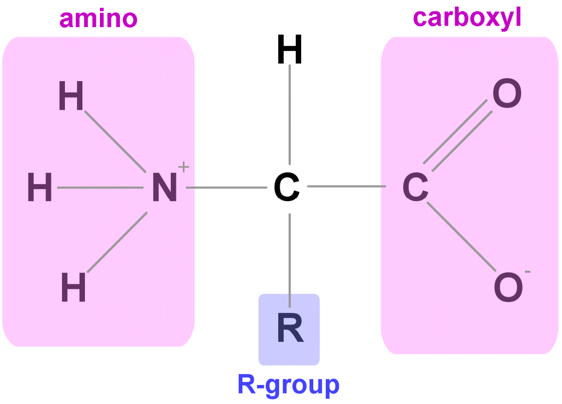
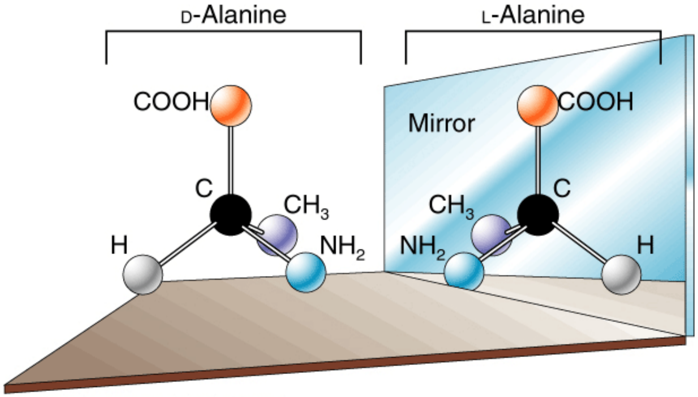


Figure 1.19: The general structure of an amino acid
This consists of an amino group and a carboxyl group co-ordinated to the α-carbon atom. The R-group is an organic side chain that differs between amino acids and confers their properties.

Alpha amino acids contain both amine and carboxyl functional groups bound to the same carbon (the α–carbon) and have the general formula $H_2NCHRCOOH$, where R is an organic side chain, as shown in Figure 1.17. It is this R-group that determines the properties of one amino acid compared to another and can vary in size from a single hydrogen atom in glycine through a methyl group in alanine to a large heterocyclic group in tryptophan. As both the amine and carboxylic acid groups of amino acids can react to form amide bonds, one amino acid molecule can react with another to form the peptide bond. It is this polymerization of amino acids that creates complex proteins.

Amino acids are chiral molecules, that is one monomer can exist as a mirror image of itself (enantiomers) and they are not super imposable, as shown in Figure 1.18 (Bada 1991). In biological systems amino acids exists almost exclusively as the L-enantiomer.



Copyright 1999 John Wiley and Sons, Inc. All rights reserved.

Figure 1.20: Amino acid enantiomers
Amino acids are chiral molecules, therefore exist as enantiomers in either the D or L forms. These enantomers are mirror images of one another and cannot be superimposed. Adapted from http://porpax.bio.miami.edu/.

1.3.2 Oxidoreductases and Dehydrogenases

Because amino acids are structurally similar, organisms utilise a range of mechanisms to metabolise one to another. For example, threonine can be metabolised to glycine in a two step enzymatic process involving L-threonine dehydrogenase, which removes a hydrogen from threonine giving 2-amino-3-ketobutyrate. In the second step this intermediate is further cleaved in a CoA-dependent reaction, catalysed by 2-amino-3-ketobutyrate CoA ligase, to produce glycine and acetyl-coA.

Generally, the enzymes that transfer electrons from one molecule to another are called oxidoreductases. These enzymes catalyze the oxidation reaction $A + B^- \rightarrow A^- + B$. In reality, free electrons do not exist as these reactions involve the transfer of hydrogen atoms. Because most metabolic oxidation reactions involve removing hydrogen from the electron donor, these enzymes are called dehydrogenases. Dehydrogenases transfer protons to an acceptor or coenzyme, i.e. small organic molecules, involved into enzymatic catalysis, such as nicotinamide adenine dinucleotide (NAD^+ or NADH), nicotinamide adenine dinucleotide phosphate ($NADP^+$ or NADPH), flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN). The term oxidase is used only for the enzymes in which the oxidation reaction has molecular oxygen (O_2) participating as the electron acceptor.

1.3.3 L-Threonine Dehydrogenase

L-threonine dehydrogenase (TDH) has been identified in all three domains of life i.e. eukarya, bacteria and archaea, from organisms such as a psychrophilic bacterium *Cytophaga sp.* KUC-1 (Kazuoka *et al.* 2003), a hyperthermophilic archaeon *Pyrococcus horikoshii* (Machielsen and van der Oost 2006), *E.coli* K-12 (Boylan and Dekker 1981), humans (Edgar 2002), chicken liver mitochondria (Aoyama and Motokawa 1981) and goat liver mitochondria (Ray and Ray 1984). The enzymes from *E.coli* and *P. horikoshii* (PhTDH) have been extensively characterised, although thus far *Pyrococcus horikoshii* (Ishikawa *et al.* 2007) and *Flavobacterium frigidimaris* are the only species from which the L-threonine dehydrogenase enzyme had been crystallised and the three-dimensional atomic structure determined (PDB codes 2DFV and 2YY7 respectively). Although TDH is not restricted to liver mitochondria, in higher organisms it is most abundant in this tissue type (Aoyama and Motokawa 1981).

L-threonine dehydrogenase belongs to the family of oxidoreductases, specifically those acting on the >CHOH group of a donor molecule with NAD^+ or $NADP^+$ as the acceptor (EC 1.1). These enzymes are divided into three classes; the short-chain dehydrogenase family (subunits with approximately 250 residues), medium-chain enzymes (with subunits of 350 - 375 residues) and long-chain dehydrogenases (of greater than 385 residues per subunit) (Esposito *et al.* 2002). It is to the medium-chain family that the 350 amino acid long TDH from *Thermococcus kodakaraensis* belongs. Typically enzymes in this family contain a catalytic zinc ion coordinated by three residues, often two Cys and one His, but variations occur and not all family members contain zinc, nor do all have proven enzyme activity; for example ζ-crystallin from a guinea pig eye lens lacks the ligands, the two metal ions and any measurable activity (Borras *et al.* 1989).

L-threonine dehydrogenase catalyses the NAD$^+$-dependent oxidation of L-threonine to L-2-amino-3-ketobutyrate, an intermediate in the metabolism of threonine to glycine (Potter *et al.* 1977), both of which are essential amino acids. The dehydrogenase catalyses the first step in the threonine degradation pathway in several microorganisms that use L-threonine as the main carbon and energy source. The product 2-amino-3-ketobutyrate is further cleaved in a CoA-dependent reaction, catalysed by 2-amino 3-ketobutyrate CoA ligase, to produce glycine and acetyl-coA (Aoyama and Motokawa 1981). TDH catalyses the first step of this reaction, shown in Figure 1.19.

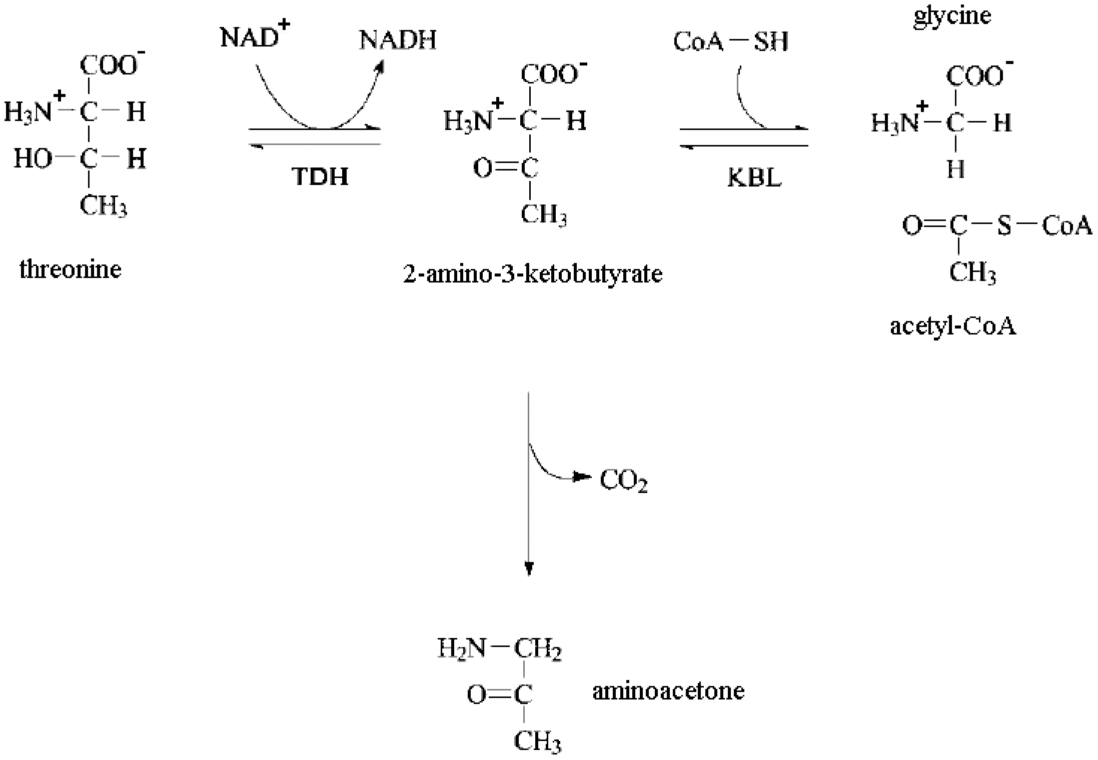
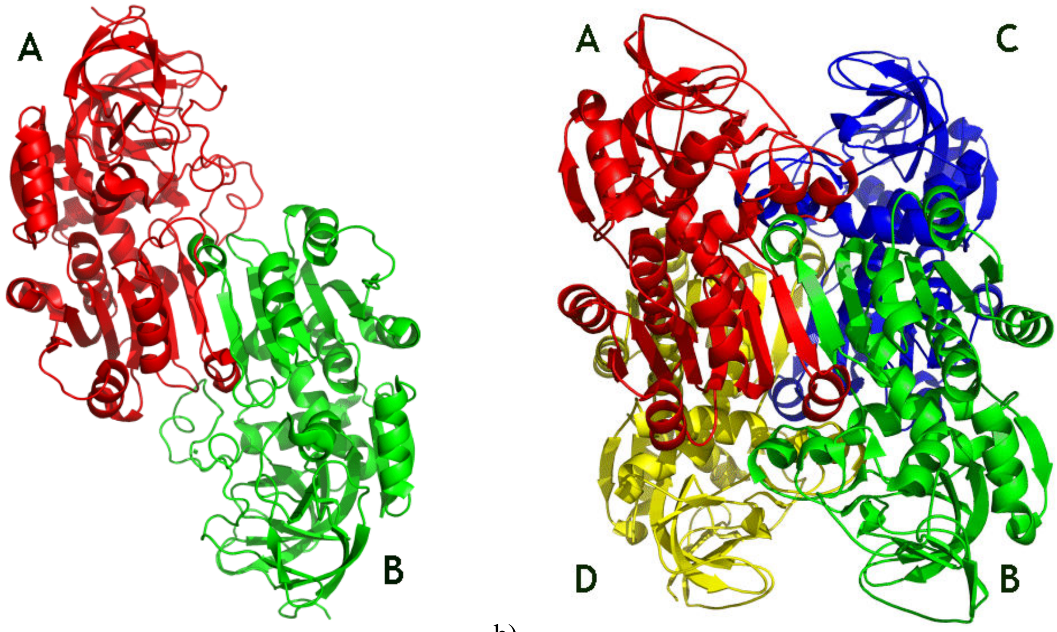


Figure 1.21: Metabolism of threonine by TDH
L-threonine dehydrogenase removes a hydrogen from threonine giving 2-amino-3-ketobutyrate. This is the first step in a two step enzymatic reaction to convert threonine to glycine as shown in Schmidt et. al 2001.

TDH is recognised as catalysing the first step in this major route of threonine utilisation in both prokaryotes (Boylan and Dekker 1981) and eukaryotes (Dale 1978). The importance of this pathway is illustrated by the fact that it accounts for 87% of the threonine catabolised in rat liver (Bird and Nunn 1983) and TDH activity is the only threonine catabolic reaction that is detected in chicken liver extracts (Aoyama and Motokawa 1981).

Within the medium-chain dehydrogenase family there exists a high degree of structural similarity between the enzymes, despite their low sequence similarity. However, there are differences between the enzymes from higher plants and animals, which are typically dimeric for example that of horse liver alcohol dehydrogenase (ADH), and those of lower organisms like yeast, bacteria and archaea, which are tetrameric such as *S. cerevisiae* ADH. This is illustrated in Figure 1.20.



a) b)

Figure 1.22: Dimeric and tetrameric ADH
Alcohol dehydrogenase exists as a dimer in higher organisms, as illustrated in **a)** by horse liver ADH, but exists as a tetramer in lower organisms, such as that shown in **b)** from *Sulfolobus solfataricus*. Because TDH shares sequence identity with ADH, it is possible that threonine dehydrogenases will exhibit the same structural difference between species.

The monomers are similar comprising two domains; a catalytic domain and a nicotinamide cofactor (NAD^+)-binding domain, which contains the α/β Rossmann fold characteristic of di-nucleotide-binding proteins (Rossman and Liljas 1974). This six stranded parallel β-sheet, flanked by α-helices, has the relative strand order of 321456 (Murzin *et al.* 1995), as shown in Figure 1.21.

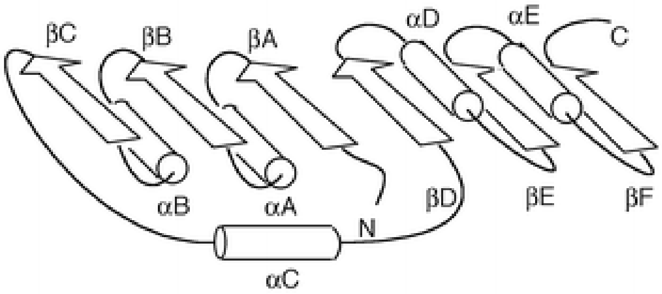


Figure 1.23: Rossmann fold
This domain is found in many nucleotide binding proteins. It consists of a parallel sheet formed by three extended parallel β-strands, connected by α-helices, forming a large sheet. Adapted from http://www.rsc.org.

In dehydrogenases this fold often forms an extended β-sheet between monomers A and B (and C and D) and in the tetramers the small helix α3, typically binding the structural zinc ion, forms contacts between the pairs of monomers: A and C (and B and D). This can be clearly seen in structures like ADH from *Sulfolobus solfataricus* (SsADH) (Esposito *et al.* 2002), ADH from *Saccharomyces cerevisiae* (ScADH) (Leskovac *et al.* 1976) and TDH from *Pyrococcus horikoshii* (PhTDH) (Ishikawa *et al.* 2007).

1.3.4 TDH from *Thermococcus kodakaraensis*

1.3.4.1 Thermostability of Hyperthermophilic Proteins

Thermophiles are now classified by their growth temperature: thermophiles above 55 ºC, moderate thermophiles above 65 ºC, extreme thermophiles above 75 ºC and hyperthermophiles above 90 ºC. Since their initial discovery in 1972 (Unsworth *et al.* 2007), many hyperthermophilic organisms have been identified and the rate of discovery is increasing as recent developments in technologies enhance our ability to access the extreme environments they inhabit. These environments, which until more recently were considered as being inhospitable to life, include volcanic areas rich in sulphur and 'toxic' metals and hydrothermal vents in the deep sea at extremely high pressures. To date the highest temperature a hyperthermophile has been observed to grow is 121 ºC; this is a microorganism, designated strain 121, that was isolated from an active "black smoker" in the Northeast Pacific Ocean (Kashefi and Lovley 2003). These developments are important because it is still yet to be understood what the underlying basis is for the extraordinary thermostability of these hyperthermophilic enzymes. Should this be discovered then a great number of opportunities to make strategic use of their thermoactivity and thermostability present themselves (Adams and Kelly 1998).

Hyperthermophilic enzymes are very similar to their mesophilic homologues, and this is supported by sequence alignments, amino acid content comparisons, crystal structure comparisons, and mutagenesis experiments (Vieille and Zeikus 2001). One hypothesis used to explain the difference in structural stability at high temperatures is that the structures of hyperthermophilic enzymes are more rigid than their mesophilic counterparts at moderate temperatures and therefore rigidity is a prerequisite for high thermostability. This hypothesis is supported by a growing body of experimental data that includes frequency domain fluorimetry and anisotropy decay (Manco *et al.* 2000), hydrogen-deuterium exchange (Bonisch *et al.* 1996;Jaenicke and Bohm 1998;Zavodszky *et al.* 1998), and tryptophan phosphorescence experiments (Gershenson *et al.* 2000). For example, at 20 ºC a much smaller fraction of the amide protons in *Sulfolobus acidocaldarius* adenylate kinase (53 %) are exchanged than in the porcine cytosolic enzyme (83 %), indicating that considerably more amide protons are involved in stable hydrogen bonds in the thermophilic enzyme. Temperatures of 80 ºC to 90 ºC must be reached before *S. acidocaldarius* adenylate kinase can show an exchange level comparable to that of the catalytically active mesophilic enzyme (Bonisch *et al.* 1996). However, not all studies support this conclusion and more work on hyperthermophilic enzyme flexibility at various temperatures is needed to gain a better understanding of the role of conformational rigidity in protein stability.

Hyperthermophilic enzymes have become model systems to study enzyme evolution, enzyme stability and catalytic mechanisms, protein structure-function relationships, and biocatalysis under extreme conditions. Not only that, they have the potential for use in a variety of industrial processes. Enzymes like DNA polymerases and DNA ligases have applications as molecular biology reagents; *Taq* DNA polymerase from *T. aquaticus* was instrumental in the development of

PCR technology. However, there is a very big difference between producing hyperthermophilic enzymes on the scale required for laboratory techniques and that required for large-scale industrial processes. Some amylases, glucosidases, pullulanases have applications (or potential applications) in starch processing, and many more are used in industrial processes such as paper pulp bleaching, chemical synthesis, cellulose degradation and ethanol production (Vieille and Zeikus 2001;Vieille and Zeikus 2001).

It is hoped that by elucidating the three-dimensional structures of hyperthermophilic enzymes it will be possible to determine the basis of their inherent properties of thermostability and optimal activity at high temperatures and the underlying atomic factors responsible. These enzymes actually share the same catalytic mechanisms with their mesophilic homologues and when cloned and expressed into such hosts, hyperthermophilic enzymes usually retain their thermal properties, implying that these are genetically encoded. Studies of primary sequences have failed to identify repeatable differences between hyperthermophilic and mesophilic enzymes; therefore it appears likely that no one mechanism is responsible for their thermostability. Instead it is likely that a small number of highly specific alterations, which often do not obey any obvious rules, may be the source of this thermostability.

Thus a concerted action of structural, dynamic and other physiochemical attributes are utilised to ensure the delicate balance between stability and functionality of proteins at high temperatures. These may include, but are certainly not limited to; charge clusters, networks of hydrogen bonds, optimization of packing, salt bridges, hydrophobic interactions, surface loop stabilization and reduction of the entropy of unfolding, each contributing to the overall thermostability (Unsworth *et al.* 2007). Identifying these minor intrinsic differences that allow hyperthermophilic enzymes to maintain structural stability and activity at elevated temperatures is a major challenge, but one that can only be aided by the elucidation of the proteins' tertiary and quaternary structures.

1.3.4.2 *Thermococcus kodakaraensis*

Thermococcales is composed of two major genera *Thermococcus* and *Pyrococcus*. They are strictly anaerobic obligate heterotrophs that grow on complex proteinaceous substrates, and their growth is strongly associated with the reduction of elemental sulphur. Alternatively, with a few exceptions, they are capable of gaining energy in the absence of elemental sulphur, by fermentation of peptides, amino acids and sugars, forming acids, CO_2 and H_2 (Fukui *et al.* 2005). The members of *Thermococcus* are ubiquitous in natural high-temperature environments and are therefore considered to play a major role in the ecology and metabolic activity of microbial consortia within hot-water ecosystems (Fukui *et al.* 2005).

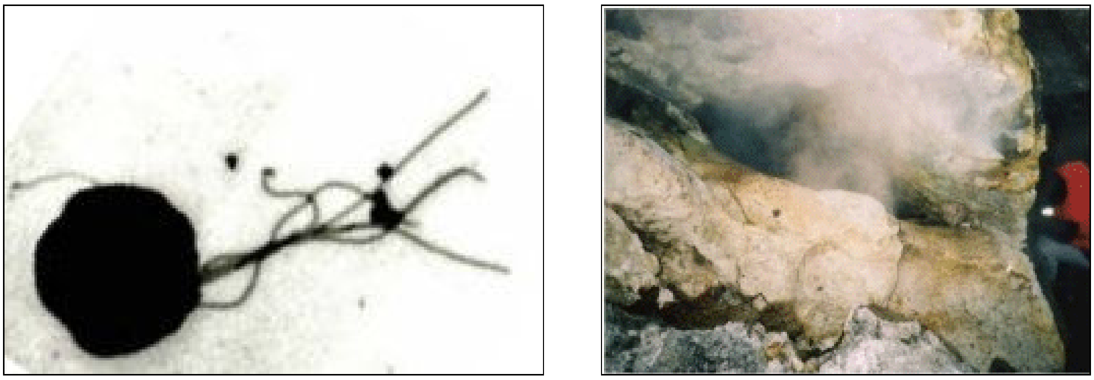


Figure 1.24: EM observation of *Thermococcus kodakaraensis*
This hyperthermophilic organism was discovered in a solfatara off the coast of Japan. It is able to survive in temperatures up to 105ºC and its entire genome has recently been sequenced. Photos taken from the *T. kodakaraensis* Genome Project.

The anaerobic hyperthermophilic archaeon *Thermococcus kodakaraensis* (KOD1) was isolated from the sediment and water of a solfatara, at over 100 ºC, on the shores of Kodakara Island off the coast of Kagoshima, Japan (Figure 1.22) (Atomi *et al.* 2004). The bacterium is of an irregular cocci shape, 1-2 µm in diameter, with several flagella. It exhibits optimal growth at 95 ºC with a growth range from 65 ºC to 100 ºC, but is unable to grow at or above 105 ºC (Morikawa *et al.* 1994). This species belongs to the order *Thermococcales*, comprising two major genera, *Pyrococcus* and *Thermococcus* (Itoh 2003), and a third genus *Palaeococcus* (Takai *et al.* 2000), which as obligate hetrotrophs are similar metabolically, growing on organic substrates usually in the presence of elemental sulphur. *Pyrococcus* however, has a higher optimum growth temperature of 95 ºC-103 ºC compared to *Thermococcus* (Fukui *et al.* 2005). The entire 2,088,737 bp genome of *T. kodakaraensis* has been sequenced and analysed (Fukui *et al.* 2005), making it one of the best characterized hyperthermophilic organisms. An orthologue search was subsequently carried out on the genome, highlighting a DNA sequence corresponding to that of L-threonine dehydrogenase (TDH; EC 1.1.1.103). A study of the enzymatic properties of the gene product revealed the enzyme to be a true L-threonine dehydrogenase of this archaeon (TkTDH) (Bowyer *et al.* 2008).

The crystal structure of TkTDH from *Thermococcus kodakaraensis* has now been solved at a resolution of 2.4 Å revealing the detailed tertiary fold and quaternary structure of the molecule. Analysis suggests that the biologically active form of the molecule is a tetramer, in agreement with ADHs from lower life forms. It was crystallised bound to the co-factor NAD^+, identifying the active site and suggesting a possible mechanism of catalysis. The TkTDH structure has been used to model a putative complex formed with 2 dimers of 2-amino 3-ketobutyrate CoA ligase, suggesting that 2 molecules of this dimeric protein bind one TkTDH tetramer. This information can be used to further our understanding of the thermostability of hyperthermophilic proteins, particularly when compared to non-thermophilic proteins.

The following chapters present the results of structural studies on three diverse proteins. The structure of BipD has been solved by X-ray diffraction, giving insights into not only its structure and function, but also how it interacts with the tip of the type three secretion system needle, and assists the invasion of host cells. The structure of TkTDH has also been determined by similar methods, allowing conclusions to be drawn as to how it interacts to form a complex with the next enzyme in the pathway and provides more evidence for the proposed catalytic mechanism of action. Studies are continuing on DAD and progress has been made in its expression and purification. Further crystallisation trials are being carried out to enable structure determination. This work has not only elucidated specific protein structures, but also contributed to our general understanding of 2 biochemical pathways.

Chapter 2

Materials, Methods and X-ray Crystallography

2.1 Molecular Biology

2.1.1 DNA Vectors

For over-expression of both the BipD and DAD proteins, it was necessary to ligate the gene into a plasmid, which was then inserted into the appropriate bacterial strain. A plasmid is an independent, circular, self-replicating bacterial DNA molecule that carries only a few genes. These occur naturally and the number of plasmids in a given cell generally remains constant from generation to generation. Plasmids are autonomous molecules and exist in cells as extra chromosomal genomes (Liu and Shi 2000). The genes they encode can endow the host organism with potentially advantageous characteristics and so can be used as DNA vectors. One of the reasons that plasmids are used in molecular biology is that as well as being readily available, they are easy to manipulate and isolate. The recombinant plasmids used today have all been engineered to contain a genetic element that provides a means by which those cells that possess it can be selected for.

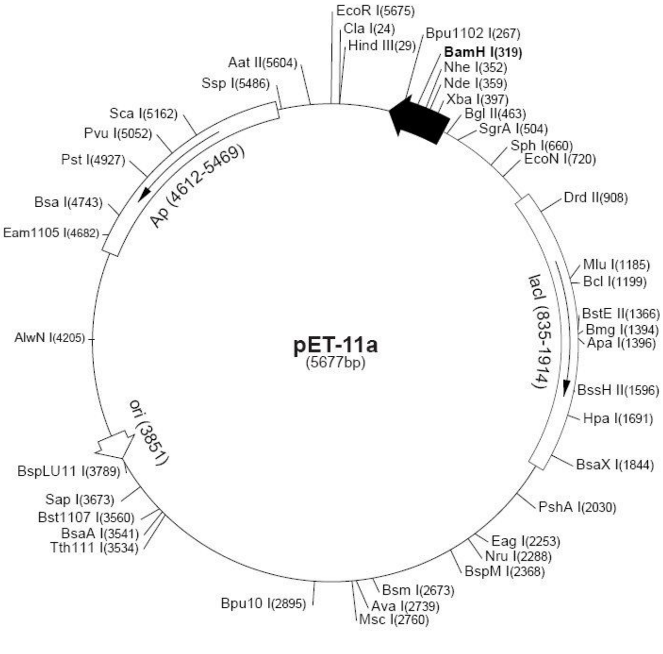


Figure 2.1: Restriction enzyme recognition sites of pET11a plasmid
A pET vector is a bacterial plasmid designed to enable the quick production of a large quantity of any desired protein when activated. This plasmid (pictured below) contains several important elements - a *lacI* gene which codes for the *lac* repressor protein, a T7 promoter which is specific to only T7 RNA polymerase (not bacterial RNA polymerase) and also does not occur anywhere in the prokaryotic genome, a *lac* operator which can block transcription, a polylinker, an f1 origin of replication (so that a single-stranded plasmid can be produced when co-infected with M13 helper phage), an ampicillin resistance gene, and a ColE1 origin of replication

The plasmid pET11a, shown in Figure 2.1, is the vector that has been used for both BipD and DAD protein expression in *E.coli*. The plasmids used throughout this research have had ampicillin resistance conferred on them as a means of selecting which bacterial cells have successfully taken up the plasmid and retained it. Ampicillin belongs to the penicillin group of beta-lactam antibiotics, which function by preventing the cross-linking of peptidoglycan chains in bacterial membrane synthesis. This results in the membrane being 'leaky' and ultimately prevents successful bacterial cell growth. The insertion of the β-lactamase gene into the bacterial plasmid is responsible for ampicillin resistance because it encodes β-lactamase, an enzyme capable of cleaving the ring in ampicillin and thereby inactivating it. Transformed cells containing the plasmid with the β-lactamase gene will be resistant to ampicillin and so will grow in the presence of the antibiotic, whereas non-engineered cells will not.

2.1.2 Restriction Endonucleases

To cut the double stranded DNA of the gene of interest and the plasmid, ready for ligation, restriction endonucleases were necessary. These enzymes function by scanning the length of a DNA molecule and once it encounters its particular recognition sequence; it will bond to the DNA molecule and make one cut in each of the two sugar-phosphate backbones of the double helix (Andrews and Patel 1998). The positions of these two cuts, both in relation to each other, and to the recognition sequence itself, are determined by the identity of the restriction endonuclease used to cleave the molecule.

Different endonucleases yield different sets of cuts, but one endonuclease will always cut a particular base sequence the same way, no matter what DNA molecule it is acting on. Once the cuts have been made, the DNA molecule will break into fragments. These enzymes are part of a cell's natural defence against foreign DNA, as well as a method of recognising and removing faulty DNA.

Today we use restriction endonucleases to cut DNA at specific sites in order to intentionally introduce modifications. There are several hundred known endonucleases, all of which function by recognising a specific base sequence in a strand of DNA and making a cut in each strand. These recognition sites are usually palindromic and contain 4-6 base pairs. There are three classes of endonucleases; Type I, II and III. Type I and III are multi-subunit enzymes that cleave with (Type I) or without (Type III) ATP (Williams 2003). After recognising a specific sequence, the actual cleavage is made away from the recognition site; Type I cuts 1-10 kb and Type III up to 100 base pairs away from the point of recognition. This is not particularly helpful in molecular biology because it is not specific. In contrast, Type II endonucleases are useful because they cleave DNA at a specified point within the recognition sequence, and these are the ones used in molecular biology (Andrews and Patel 1998).

Cleavage is achieved by two distinct methods, leaving either 'blunt' or 'sticky' ended double stranded DNA. Some enzymes cleave directly across the centre of the recognition site, giving rise to blunt ends. The endonucleases used in the research detailed here were *Nde*I and *Bam*HI, both of which are Type II enzymes that leave cohesive (sticky) ends (Williams 2003). They make a staggered cut, symmetrically placed either side of the centre of the palindromic recognition site, leaving a single-strand of DNA overhanging the end of each fragment.

↓ ↓
*Bam*HI G G A T C C *Nde*I C A T A T G
C C T A G G G T A T A C
↑ ↑
(↑ = cut site)
As with all enzymes, optimal activity is achieved at 37 ºC.

2.1.3 DNA Ligase

If restriction enzymes are the scissors of DNA manipulation, then DNA ligase is the glue. It is a particular type of ligase that can join together DNA that has a break in both complementary strands. This enzyme is able to form covalent phosphodiester bonds between 3' hydroxyl ends and 5' phosphate ends of juxtaposed nucleotides, in an ATP dependant manner. The alternative, a single-strand break, is easily fixed by DNA polymerase using the complementary strand as a template but still requires DNA ligase to create the final phosphodiester bond to fully repair the DNA (Taylor and Hagerman 1990). In this research T4 ligase (isolated from bacteriophage T4) has been used to insert DNA fragments of the bipd and dad genes, into the pET11a plasmid.

The temperature at which ligation was carried out had to be kept constant at 16 ºC. Although DNA ligase is most active at 25 ºC, the reaction temperature had to be kept low as homologous pairing of the sticky ends will not occur at high temperatures because hydrogen bonding is disrupted (Luo and

Barany 1996). The reaction was carried out over 4 hours to compensate for the low temperature and allow sufficient time for ligation to occur.

Bacteriophage T4 DNA ligase is a single polypeptide with a M.W. of 68 kDa. Maximal activity is obtained at pH 7.5 - 8.0. At pH 6.9 and pH 8.3 the enzyme exhibits 40 % and 65 % of its full activity respectively. Mg^{2+} is required and the optimal concentration is 10 mM. Sulfhydryl reagents (DTT, 2-mercaptoethanol) are also required and NaCl in concentrations exceeding 200 mM is inhibitory. For intermolecular ligation, especially when the subtrate DNAs consist of large DNA molecules PEG (concentrations of 1 % - 10 %) has been reported to stimulate the enzymatic activity (Takahashi *et al.* 1984).

2.1.4 Primer Design and Polymerase Chain Reaction

For over-expression of BipD without the GST-tag, primers for the bipd gene were designed, based on its sequence. These oligonucleotides could then be used in PCR to clone many copies of the non-tagged bipd gene sequence.

The 5' bipd sequence initially used was:

```
ggacgcgcgctgaccgtgcgcgat
G  R  A  L  T  V  R  D
```

A trial 5' primer was designed as shown here, including the recognition sequences for NdeI and XbaI:

`g tag cct |gga tct aga cat` **`atg`** `ggacgcgcgctgaccgtgcgcgat`

Tm = (1×AT) + (2×CG)
Tm = (1×18) + (2×28)
Tm = 18 + 56
Tm = 74 °C

But the melting temperature (T_m) of this sequence is too high for PCR, so some of the non-coding bases at the beginning were removed:

`gga tct aga cat` **`atg`** `ggacgcgcgctgaccgtgcgcgat`
Tm = (1×AT) + (2×CG)
Tm = (1×15) + (2×24)
Tm = 15 + 48
Tm = 63 °C

5’ end primer:

```
gga tct aga cat atg gga cgc gcg ctg acc gtg cgc gat
                 M   G   R   A   L   T   V   R   D
```

Blue = NdeI, **Green** = XbaI

The 3’ bipd sequence initially used was:

```
5′ agctatctgcagatctga 3′
```

Reverse complement:

```
tcagatctgcagatagct
```

A 3’ end trial primer: 5’ to 3’ was designed as shown here, including the recognition sequences for NdeI and XbaI:

```
g gcg caa gct tgg atc cgc cct a tcagatctgcagatagct
```

Tm = (1×AT) + (2×CG)
Tm = (1×18) + (2×23)
Tm = 18 + 46
Tm = 64 °C

3’ end primer:

```
g gcg caa gct tgg atc cgc cct a tcagatctgcagatagct
```

reverse complement:

```
agctatctgcagatctgatagggcggatccaagcttgcgcc
 S  Y  L  Q  I Stop Stop
```

Red= Hind III, **purple** = Bam HI

Sequences ordered (both going 5’ to 3’):

5’ end primer

```
ggatctagacatatgggacgcgcgctgaccgtgcgcgat
```

3’ end primer

```
ggcgcaagcttggatccgccctatcagatctgcagatagct
```

2.1.5 Agarose Gel Electrophoresis

This technique was used to separate DNA molecules by size, to visualise the products of ligation to verify they had been ligated, and following restriction enzyme digestion to check the DNA had been cut. It was also used to quantitatively analyse the products of PCR as an estimate of yield.

An agarose gel, containing ethidium bromide (EtBr), is poured (see appendix for method) and the negatively charged nucleic acid molecules move through the agarose matrix at different speeds, with an electric field (electrophoresis) (Brody *et al.* 2004). The smaller molecules are able to move more easily through the matrix, and hence faster, than longer ones, and so migrate further down the gel.

The advantage of this type of gel is that it is easily poured, does not denature the samples, and is physically firmer than polyacrylamide (used in SDS gel electrophoresis). The samples can also be recovered if required. The disadvantages are that gels can melt during electrophoresis, the buffer can become exhausted, and different forms of genetic material may run in unpredictable forms.

The most important factor affecting the rate at which samples migrate through the gel is the length of the DNA molecule, smaller molecules travel further. But conformation of the DNA molecule is also a factor. For example, a DNA plasmid that has not been cut with a restriction enzyme may be in different conformations and therefore will move with different speeds (slowest to fastest): nicked or open circular, linearised, or supercoiled (Brody *et al.* 2004).

Increasing the agarose concentration of a gel reduces the migration speed and enables separation of smaller DNA molecules. The higher the voltage, the faster the DNA migrates, but this is limited by the fact that it heats up and ultimately causes the gel to melt. High voltages can also decrease the resolution (Goedhart and Gadella, Jr. 2005).

Gel electrophoresis can be used for the separation of DNA fragments of 50 base pairs up to several megabases (millions of bases). However, it is normally used in a range of 100 bp to 20 kbp. Typical run times are about an hour. Small nucleic acids are better separated by polyacrylamide gels, large DNA molecules are only able to move end-on in a process called "reptation" and are more difficult to separate. In general lower concentrations of agarose are better for larger molecules; it will exaggerate the distances between bands (Brody *et al.* 2004). The disadvantage of higher concentrations is the long run times (sometimes days). Instead these gels should be run with a pulsed field electrophoresis (PFE), or field inversion electrophoresis.

Loading buffers are added with the DNA in order to visualize it and sediment it in the gel well. Negatively charged indicators keep track of the position of the DNA and Orange G was used in this research. The most common dye used for agarose gel electrophoresis is ethidium bromide, because it fluoresces under UV light when intercalated into DNA (or RNA). By running DNA through an EtBr-treated gel and visualizing it with UV light, distinct bands of DNA become visible (Goedhart and Gadella, Jr. 2005). Alternatives to EtBr are available but were not used in this research.

After electrophoresis the gel is illuminated with an ultraviolet lamp to view the DNA bands. The EtBr fluoresces reddish-orange in the presence of DNA. The gels were viewed and photographed using a Gene Genius Bio Imaging System and GeneSnap from Syngene to produce a black and white photo, as shown in Figure 2.2.

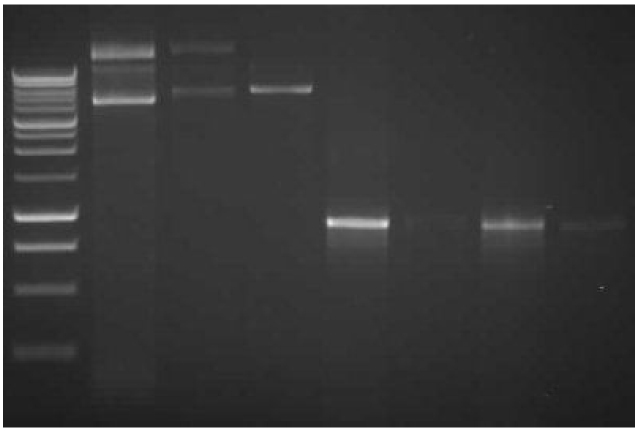


Figure 2.2: DNA agarose gel from electrophoresis
The left-hand lane shows a 1kb DNA marker and the other lanes show DNA strands of various sizes. The shorter strands have moved further down the gel than the longer ones and are visualized by staining with ethidium bromide viewed under ultraviolet light.

2.1.6 Bacterial Strains

The host strain BL21(DE3) *Escherichia coli* was used for the expression of BipD and DAD and these cells were transformed using ‘maxiprepped’ DNA. This strain is lysogenic for a λ prophage that contains an isopropyl-β-d-thiogalactopyranoside (IPTG) inducible T7 RNA polymerase. This viral polymerase is highly selective for its own promoters, which are not found to naturally occur in *E. coli*. However, this polymerase is capable of efficiently transcribing almost any DNA sequence that is linked to a T7 promoter, even when inserted in to a foreign cell (Studier and Moffatt 1986). Transcription of the T7 RNA polymerase gene is controlled by the L8-UV5 *lac* promoter and the *lac* operon.

The BL21(DE3)pLysS strain also carries the pACYC184-derived plasmid that encodes T7 lysosyme, which selectively inhibits T7 RNA polymerase by binding to it (Moffatt and Studier 1987). This serves to suppress any basal expression of target DNA under the control of the T7 promoter until over-expression is induced by IPTG. The accumulation of T7 lysosyme does not affect the growth of gram-negative *E. coli* cells, presumably because it is unable to reach its substrate, the peptidoglycan cell wall, as it cannot penetrate the outer membrane (Studier 1991). T7 RNA polymerase can elongate RNA chains approximately five times faster than the host polymerase and, reportedly, specific mRNA produced by T7 polymerase can saturate the translational machinery of the *E. coli* cell to the extent that the rate of protein synthesis from this mRNA is primarily dependent on the efficiency of its translation (Moffatt and Studier 1987).

The addition of IPTG to bacterial cultures is an established method of inducing expression of plasmid-based genes for the production of recombinant proteins under the control of the lac promoter, because the most direct way to control the expression of a gene is to regulate its rate of transcription. The lac operon is a set of genes and DNA regions that control the production of a protein and naturally it is controlled by the presence or absence of lactose and glucose. In the absence of lactose, the repressor protein binds to the lac operator and prevents transcription. Binding of lactose to the repressor causes it to dissociate from the operator, enabling RNA polymerase to transcribe the gene. IPTG mimics allolactose and switches on the lac operon, however the cell is unable to break down IPTG so the lac operon remains permanently on causing the cell to produce large quantities of the protein of interest.

The advantage of using this IPTG inducible system to over-express proteins is that the *E. coli* cells can undergo initial cellular growth and proliferation without the burden of producing the extra foreign protein. This ensures that a suitable quantity of healthy *E. coli* cells are grown up to then over-express the recombinant proteins in large amounts. The competency of transformed cells will be maintained by conferring resistance to ampicillin.

2.2 Protein Purification

2.2.1 Sonication

Sonication is the act of applying sound (usually ultrasound) energy to agitate particles in a sample. In the context of this research an ultrasonic probe was used to disrupt the *E.coli* cell membranes with inaudible ultrasound (greater than about 18 kHz) causing the release of cellular contents, particularly soluble protein of BipD or DAD, into solution.

Ultrasonication utilises the rapid sinusoidal movement of a probe within the liquid. It is characterised by high frequency (18 kHz - 1 MHz), small displacements (less than about 50 µm), moderate velocities (a few ms^{-1}), steep transverse velocity gradients (up to 4,000 s^{-1}) and very high acceleration (up to about 80,000 g) (Chaplin and Bucke 2007).

2.2.2 Fractional Ammonium Sulphate Precipitation

Fractional salt precipitation was the initial purification step carried out for both BipD and DAD to crudely remove some impurities from the soluble fraction obtained after sonication (once cell debris had been removed via centrifugation). As increasing amounts of salt are added, different proteins will precipitate out of solution. Ammonium sulphate is most commonly used for this technique because it does not have any deleterious effects on protein structure, is highly water soluble and inexpensive.

An amount of $(NH_4)_2SO_4$ sufficient to cause 15 % saturation was fully dissolved in the sample and the homogenous solution centrifuged at 18,000 rpm for 30 minutes to remove any precipitation. The pellet was resuspended, and a volume of the supernatant retained for analysis. A further 15 % $(NH_4)_2SO_4$ was added and the protocol repeated in a step-wise manner until 75 % saturation was reached. SDS-gel electrophoresis was used to ascertain at what $(NH_4)_2SO_4$ concentration the protein of interest precipitated out and this pellet sample was retained.

2.2.3 Dialysis

Following ammonium sulphate precipitation, the sample contains a high salt concentration, which could hinder chromatography purification by ion enxchange. Dialysis is used to separate molecules in solution by the difference in their rates of diffusion through a semi-permeable membrane. In this situation, salts will to move out of the dialysis tubing, in to the surrounding solution where the $(NH_4)_2SO_4$ concentration is lower.

The solution, containing many different kinds of molecules, is placed into semi-permeable dialysis tubing, such as a cellulose membrane with pores, and sealed. The sealed dialysis tube is placed in a container of 20 mM TRIS and molecules small enough to pass through the tubing move out of the dialysis tube, in to the lower concentration solution. Larger molecules such as proteins, that have dimensions significantly greater than the pore diameter, are retained inside the dialysis tube. Typically the solution was dialysed over 48 hours.

2.2.4 Chromatography

Chromatography is the principle of separating substances by their different characteristics. There are several characteristics by which proteins can be differentiated, as illustrated in Figure 2.3 (Amersham Biosciences 2007a), which can be combined to maximise purification.

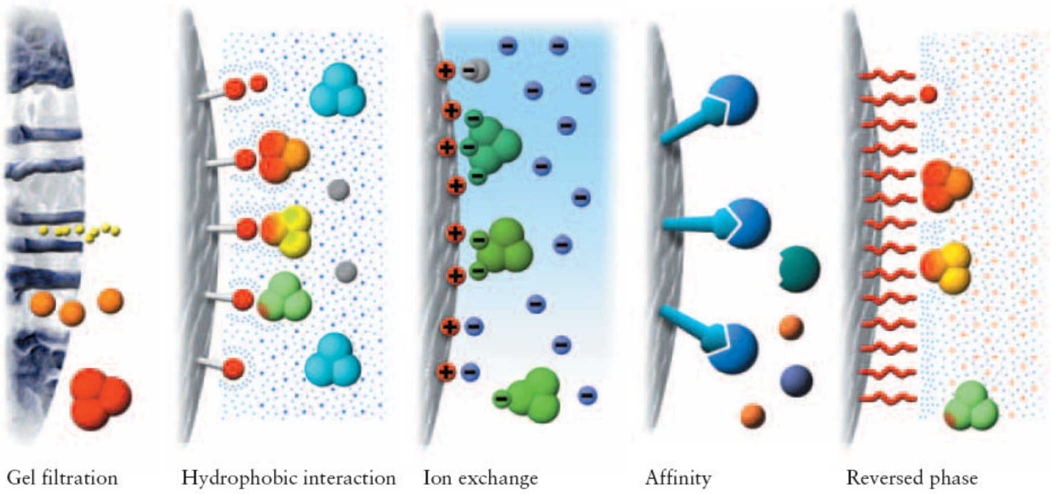


Figure 2.3: Chromatography types
Proteins can be separated by size, hydrophobicity, charge, affinity and reversed phase. More than one technique can be combined to maximise purification of a sample. Adapted from Novagen instruction manual.

2.2.4.1 Ion Exchange Chromatography

In this instance the protein of interest is separated from other proteins based on ionic charge. Because amino acids contain both the active groups of an amine and a carboxylic acid they can carry either a positive or a negative charge (or be neutral) and so are called zwitterions. At a certain pH known as the isoelectric point, the peptide carries no net electrical charge. The exact value is specific to each different amino acid. Ion exchange chromatography separates proteins according to their net charge. For example, if a protein has a net positive charge at pH 7, then it will bind to a column of negatively-charged beads, whereas a negatively charged protein would not. By changing the pH so that the net charge on the protein is negative, it too will be eluted.

A sample containing a mixture of proteins is introduced into a sample loop and a buffered aqueous solution used to carry the sample from the loop onto a column that contains some form of stationary phase material. This is typically a resin or gel matrix consisting of agarose or cellulose beads with charged functional groups covalently bonded to them. The stationary phase surface of the chromatography column displays ionic functional groups that interact with analyte ions of opposite charge. The column is washed to remove any unbound molecules, then the bound protein is eluted systematically by varying pH and/or ionic strength of the buffer solution. These changes interfere with the electrostatic interactions between the bound protein and the stationary phase material, causing the protein to be released from the column. Ion exchange chromatography retains analyte molecules based on ionic interactions. This type of chromatography is further subdivided into cation exchange chromatography and anion exchange chromatography:

Cation exchange chromatography retains positively charged cations because the stationary phase displays a negatively charged functional group such as phosphoric acid.

Anion exchange chromatography retains negatively charged anions using positively charged functional group such as a quaternary ammonium anion.

2.2.4.1.1 Q-sepharose™ Anion Exchange Chromatography Column

This column was chosen for intermediate purification because it is suitable for use at pH 8 (where both BipD and DAD are stable and are anionic). The 6 % matrix of cross-linked agarose contains a strong quaternary ammonium anion, providing a binding capacity of 120 mg/ml (Amersham Biosciences 2007c).

Figure 2.4: Structure of sepharose cross-linked agarose media
Sepharose media are based on chains of agarose, arranged in bundles and with different degrees of intra-chain cross-linking (Figure 12), to give a range of rigid, macroporous matrices with good capacity and low non-specific adsorption. Adapted from Novagen instruction manual.

2.2.4.1.2 Hi Trap Q XL Anion Exchange Column

It was found necessary to ‘polish’ the DAD sample before crystallisation to remove the few remaining trace contaminants, and this was achieved through the use of a 1 ml Hi Trap Q XL column. This contains long chains of dextran coupled to a robust, 6 % highly cross-linked agarose matrix. The dextran chains increase the exposure of the Q charged groups, resulting in very high loading capacities (Amersham Biosciences 2007c).

2.2.4.2 Gel Filtration Chromatography

This technique separates proteins by their size, or hydrodynamic volume, using what amounts to a molecular sieve. Typically a gel medium, usually porous beads of polyacrylamide, dextran or agarose, is used and filtered under low pressure; these media are also known as the stationary phase. Unlike ion exchange chromatography, these compounds contain very few charged groups to avoid unwanted electrostatic interactions that would interfere with size separation. A buffered aqueous solution is used to transport the sample through the stationary phase, in a similar manner to other chromatography columns.

The flow rate is often slow for gel filtration to allow time for better separation of similar sized proteins and hence higher resolution. Longer columns also improve resolution because they provide a greater distance over which the proteins can separate. Smaller molecules are able to pass through the porous beads, whereas larger proteins are excluded and forced to travel through the interstitial spaces. For this reason, smaller molecules take longer to travel through the column because effectively they have a greater volume available to them than larger molecules, as illustrated in Figure 2.5.

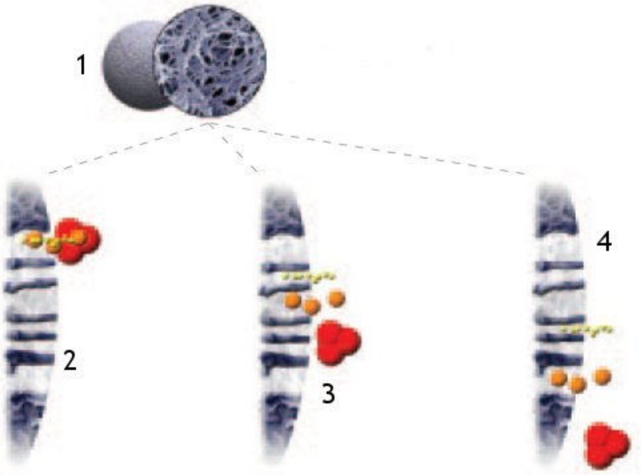


Figure 2.5: Gel filtration
This technique works on the principle of separating protein molecules according to their size.

1) many porous beads are tightly packed into the column, with only small interstitial spaces.

2) a buffer solution containing the mixture of proteins is introduced to the column, small molecules (shown in orange) are able to permeate the beads.

3) molecules diffuse in and out of the beads, becoming separated from the medium and larger sized molecules.

4) larger molecules are unable to enter the beads and so pass down the column much more quickly by moving through the interstitial spaces.

2.2.4.2.1 Superdex™ 200 Gel Filtration Chromatography Column

Superdex 200 is a spherical composite of dextran and cross-linked agarose, thereby combining the separation power of Sephadex™ with high physical and chemical stability of cross-linked Superose™ 6. The particles are approximately 13 µm in size and can separate molecules in the range of 10 kDa – 600 kDa. This column was used as a secondary step for purification of BipD and DAD after anion exchange, at a flow rate of 0.5 ml/min (Amersham Biosciences 2007b).

2.2.4.3 Affinity Chromatography

Unlike gel filtration and ion exchange chromatography, affinity chromatography does not rely on the physical difference between proteins in a mixture, but instead on a unique characteristic of the protein of interest. It is based on highly specific biological interactions; such as that between antigen and antibody, enzyme and substrate, or receptor and ligand. This is a reversible interaction between a protein and a specific ligand coupled to a chromatography matrix. Affinity chromatography combines the size fractionation capability of gel permeation chromatography with the ability to design a stationary phase that reversibly binds to a known subset of molecules. In this instance an affinity column for the GST-tag on the BipD protein was used. Affinity chromatography is a unique method by which proteins can be purified because a protein can be separated from a mixture on the basis of its biological function or individual chemical structure. Purification that would otherwise be time-consuming, difficult or even impossible using other techniques can often be achieved with affinity chromatography. The appropriate ligand is immobilised on the column and reversibly binds its substrate as a protein mixture flows through the column.

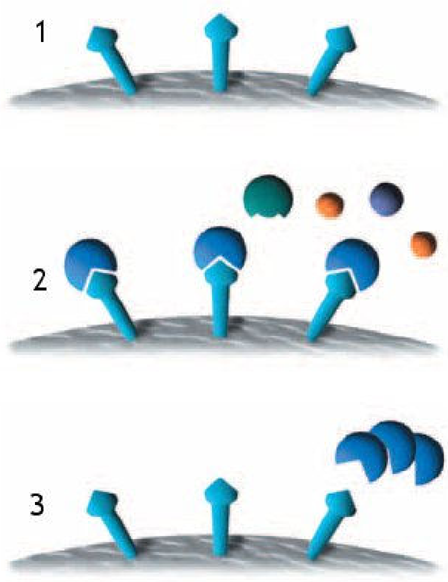


Figure 2.6: Affinity chromatography
This technique utilises a specific interaction between the protein of interest and a ligand, to purify it from the protein mixture.

1) the ligand is immobilised on the affinity medium.

2) a mixture of proteins are loaded onto the column and the one sthat is specific to the ligand will bind (reversily) to the affinity medium.

3) once all the non-bound proteins has been washed off, the protein of interest can be eluted.

2.2.4.3.1 GSTrap Affinity Chromatography Column

Immobilised glutathione is bound to the affinity medium. The solution containing the protein mixture was loaded and the GST-BipD fusion protein captured on the column. Impurities are removed by washing. Instead of eluting the protein immediately, thrombin was added to the column to cleave the GSTag from BipD. BipD and thrombin could then be washed from the column and the GST was later eluted using reduced glutathione.

2.2.4.3.2 Benzamidine Affinity Chromatography Column

To separate the thrombin and BipD, it was necessary to use another affinity column. The benzamidine column uses p-aminobenzamidine immobilised on cross-linked agarose to bind thrombin. The mixture was loaded on to the column and BipD flowed straight through whilst the thrombin bound. This was later eluted with a solution of TRIS and p-aminobenzamidine.

2.2.5 Concentration

10 kDa vivaspins from VWR biomarker were used to concentrate larger volumes of protein solution. ddH_2O was first loaded into the tube and centrifuged at 3,000 rpm until all the H_2O had passed through the membrane. The protein was then loaded in and centrifuged until it reached the appropriate volume. Frequent pipetting is required to combat concentration polarisation and maintain a reasonable flow rate.

10 kDa centricons were used to concentrate small volumes of protein. These were loaded into a standard fixed angle centrifuge and centrifuged at 5,000 rpm until the required volume was obtained. Both vivaspins and centricons work by forcing small molecules, such as H_2O and salt, through a membrane and away from the protein, leaving the protein in a smaller volume of H_2O.

2.2.6 SDS-PAGE Gel Electrophoresis

This technique is used to separate proteins, based on their size, and compare them to fragments of known molecular weight. Sodium dodecyl sulphate (SDS) is an anionic detergent that denatures proteins by "wrapping around" the polypeptide backbone, binding to proteins fairly quantitatively in a mass ratio of approximately 1.4:1. In so doing, SDS confers a negative charge to the polypeptide in proportion to its length, i.e. the denatured polypeptides become "rods" of negative charge cloud with equal charge or charge densities per unit length. It is usually necessary to reduce disulphide bridges in proteins before they adopt the random-coil configuration necessary for separation by size: this is achieved using β-mercaptoethanol. Denaturing SDS-PAGE separations therefore ensure that migration is determined by molecular weight and not by intrinsic electrical charge of the polypeptide. The proteins were run in an electric field through a polyacrylamide gel, which acts as the separating mesh.

A discontinuous system has been used in this research where a non-restrictive large pore gel (stacking gel) is layered on top of a separating gel. Each type of gel is made using a different buffer and the tank buffers are different again from the gel buffers. The resolution obtained in a discontinuous system is much greater than that obtained with a continuous system (see appendix for method). Once the gel has been stained, it is possible to see the bands of protein and compare the protein of interest against a molecular weight marker to determine its size.

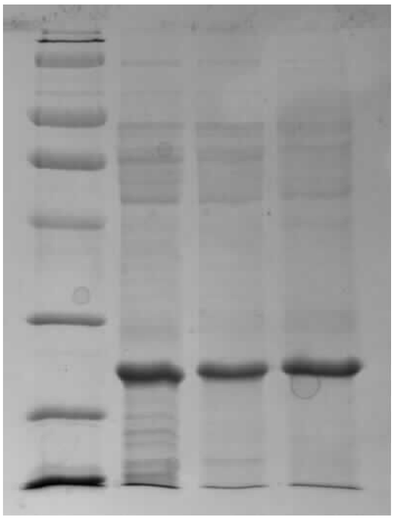


Figure 2.7: An SDS gel
SDS gel electrophoresis is an analytical technique for visualising the protein content of a solution. The first lane contains a marker with fragments of known molecular weights. The protein of interest can be run along side the marker and the bands compared to ascertain their sizes.

2.3 Crystallisation

In order to use X-ray crystallography as a method of obtaining structural data from a protein molecule, it is first necessary to grow a crystal of the protein. The use of a protein crystal results in many copies of the protein molecule being repetitively packed to form an ordered structure, sometimes up to 1mm in each dimension but often much smaller. This depth allows the X-rays to be diffracted; and recording and analysis of the resultant data to give an average structure of the protein taken from all the molecules present in the crystal. The crystal lattice can be described in terms of its unit cell, which is the smallest and simplest volume element from which the entire crystal can be constructed by repeated lattice translations. This is illustrated in Figure 2.8. The unit cell itself can be further broken down into the asymmetric unit. This is the largest aggregate of molecules that possesses no crystallographic symmetry elements, but which can be juxtaposed onto other identical entities by symmetry operations. In the Figure below, the individual molecule makes up the asymmetric unit, whilst the unit cell comprises of both molecules related by a symmetry operation. This is then repeated throughout the crystal to form the lattice.

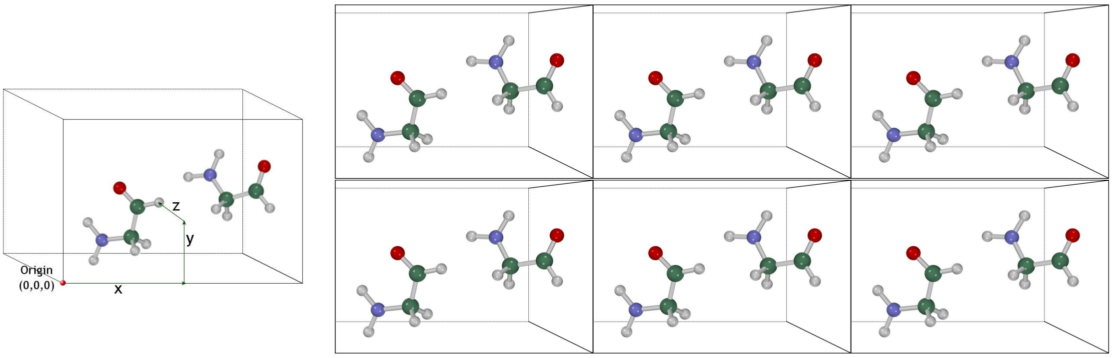


Figure 2.8: Crystallographic unit cell and lattice
The protein molecules pack together with the smallest repeating arrangement forming the unit cell, left. This unit cell is then repeated throughout the crystal, forming the crystalline lattice.

The way in which X-rays are diffracted by the electron density within crystalline structures can be explained by Braggs law $n\lambda=2d\sin\theta$. This supposes that a crystal lattice comprises multiple parallel layers or planes of molecules with constant interplanar spacing. An X-ray beam will only be diffracted when the rays reflected by successive planes emerge from the crystal in phase with one another. If they emerge out of phase, the waves destructively interfere with each other, causing graded reflections through to no reflection at all. See Figure 2.9.

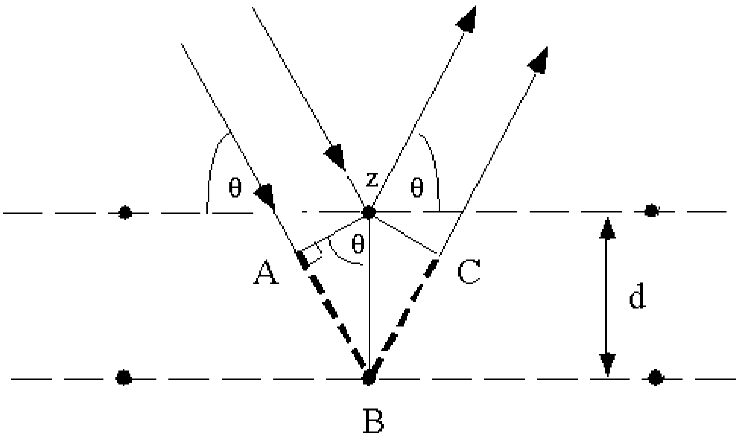


Figure 2.9: Bragg's Law
Bragg's law provides the conditions under which X-ray diffraction can occur. This diagram schematically illustrates interference between waves scattering from two adjacent rows of atoms in a protein crystal. This diagram was adapted from www.hasdeu.bz.edu.ro/softuri/fizica/mariana/Atomica/Bragg/home.html.

Once a protein has been purified, the solution is used to set up 24 well crystallisation trays. The well solution is an aqueous precipitation solution that does not denature the protein and is made up to about 1 ml total volume, and the crystallisation of biological macromolecules is achieved using a solution of the protein in precipitant as a hanging drop of about 4 μl. This is usually 2 μl of protein and 2 μl of well solution. This drop is pipetted onto a siliconised glass cover slip, which is inverted and held in place over the well by grease, as shown in Figure 2.10 (Torrii Lab 2008). This is termed the hanging-drop crystallisation method, the most common technique used for crystallising proteins.

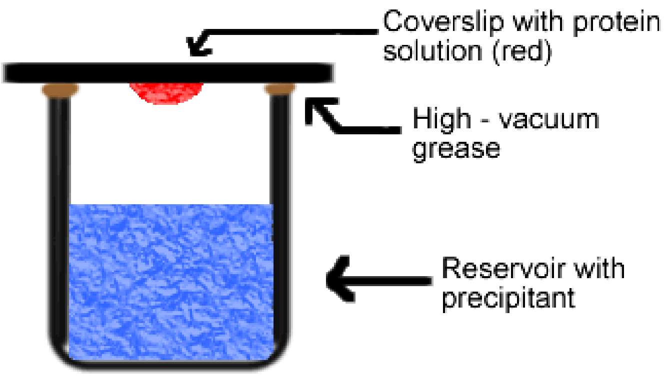


Figure 2.10: Hanging-drop crystallisation method
The hanging drop contains protein and precipitant, and reservoir only contains the precipitant solution. Vapour diffusion within the well causes the drop to become saturated with protein and at this point, nucleation of crystals occurs. Adapted from www.bio.davidson.edu/.

Once sealed by a cover slip, the well becomes a closed system. Vapour diffusion of water from the protein drop into the well solution is driven by the difference in concentration between the two solutions. This continues until equilibrium is reached. During this process, the concentration of both the protein and precipitant in the drop slowly increase until the protein solution reaches a supersaturated state. At this point, nucleation occurs and subsequently the growth of protein crystals.

Proteins differ from small molecules only in the degree of supersaturation required to induce crystallisation or to allow a useful rate of crystallisation. The supersaturation of a solution of protein may be increased in several ways such as; cooling, allowing water to evaporate, variation of the pH and the addition of an ionic solute, e.g. salt. Using evaporation to form protein crystals is much like that observed when working with common salt or sugar. As the water evaporates the protein solute begins to form crystals when the supersaturation reaches a point that the amount of water available can no longer support the high concentrations of protein. However, for small molecules this could be achieved by heating the solution until the water has boiled off. This method is obviously not appropriate for protein crystallisation, as proteins often denature above 40 ºC. Figure 2.11 (Read 2005) shows a crystallisation phase diagram. The protein and precipitant concentrations must be such that they fall into the nucleation region, where the formation of molecular clusters occurs. As the protein concentration then falls, due to some protein molecules now forming the protein crystal, the concentrations drop into the metastable region where crystals grow.

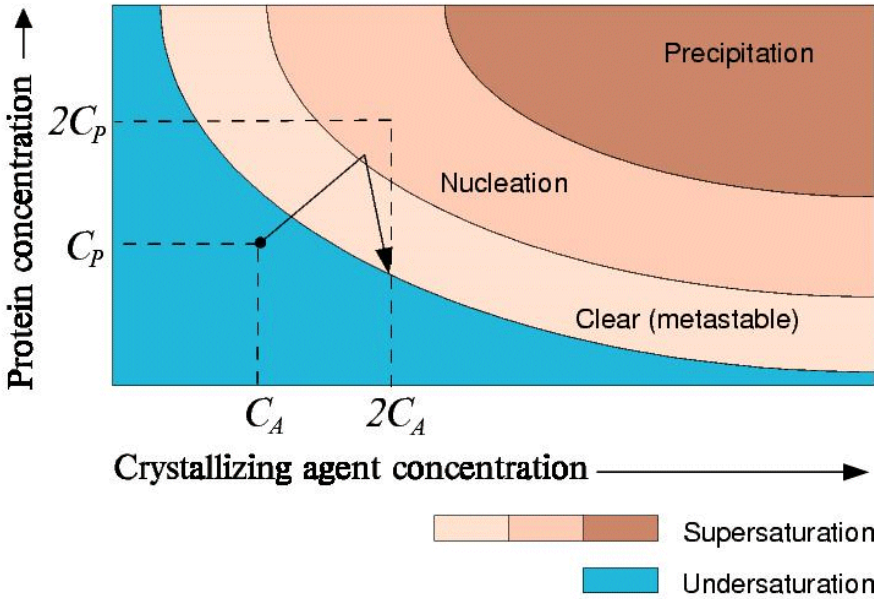


Figure 2.11: A crystal growth phase diagram
This illustrates the effect of varying precipitant and protein concentration. Proteins fail to crystallise in the unsaturated region (blue). In the nucleation region (peach) nuclei will form and in the metastable region (cream) these will grow into crystals. If the crystallisation conditions overshoot the nucleation region, that result will be precipitation of the protein (brown) to an amorphous substance. Adapted from http://kenividilaseris.wordpress.com.

A crystallisation tray is used to vary the conditions in each of the wells in order to determine the optimal conditions for growing crystals. Factors such as protein concentration, pH and solute composition and concentration can be varied. Because different pH levels (how acidic or basic a solution is) change the structural attributes of the amino acids that make up proteins, changes in pH can cause a protein to crystallise. Organic solvents also aid the precipitation of protein from solution, and a commonly used water-soluble polymer is polyethylene glycol (PEG). It is commercially available in a range of molecular masses, such as PEG 300 through to PEG 8000.

Using an ionic solute such as salt creates another problem. Ionic materials, such as sodium chloride, or common salt, dissolve readily in water. Because of the strong electric field around these ions, Na^+ and Cl^-, a large number of water molecules are loosely bound around the ions in a sphere of hydration. This is because the water molecule itself is an electric dipole, with one end slightly positively charged and the other end negatively charged. The net effect of adding salt to a protein solution is to reduce the amount of water that is free to keep the protein molecules in solution, a process called ‘salting out’. Thus, the degree of supersaturation of the protein solution is effectively increased, and crystallisation becomes more likely. Some examples of protein crystals are shown in Figure 2.12.

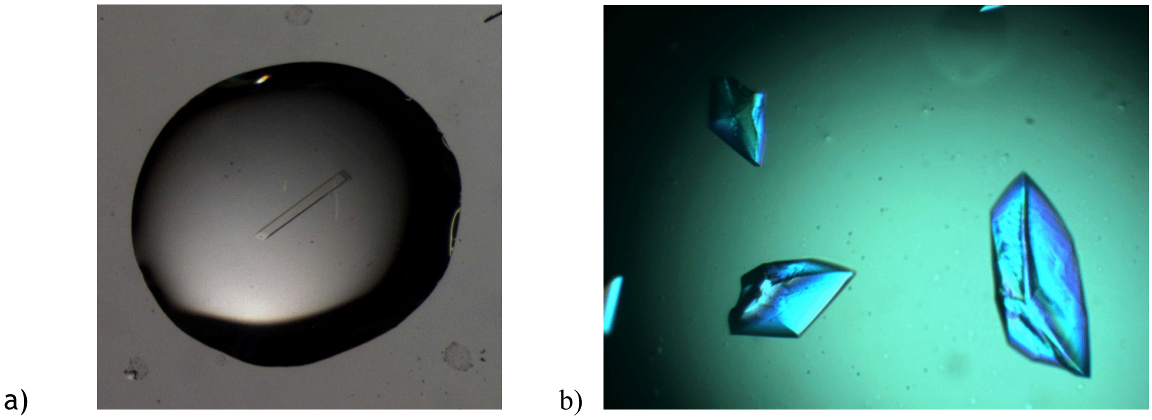


Figure 2.12: Protein crystals
Protein crystals grow in very many shapes and sizes, sometimes many grow in one drop, in others there is only a single crystals. This is determined by the protein and precipitant concentrations used, the conditions under which crystallisation occurs and how long they crystals are left to grow. **a)** shows a single crystal of TkTDH with dimensions of approximately 0.7 mm x 0.1 mm x 0.1 mm. **b)** shows several smaller SeMet BipD crystals of irregular shapes.

2.3.1 Crystal Freezing

To minimise radiation damage to the crystal whilst it is in the X-ray beam, the crystals are first frozen. There are 2 mains ways by which this can be achieved; using liquid nitrogen on the cryostream at 100K or liquid ethane. To freeze a crystal on the cryostream, a crystal is picked with a loop from the hanging drop and placed in a drop of mother liquor from that well. If the mother liquor does not contain some sort of cryo-protectant, such as glycerol or PEG 550, this can be added in a stepwise manner so as not to disrupt the crystal by rapidly altering its environment. For example, if a 30% glycerol concentration is desired, 3-4 small drops of glycerol can be pipetted next to the drop of mother liquor and each one mixed in separately, allowing time for the crystal to adjust to its new environment. The crystal is mounted and rapidly exposed to the cryostream, instantly freezing it. It is then stored in a vial of liquid nitrogen prior to exposure to the X-ray beam. To freeze the crystal in liquid ethane, the crystal is transferred to a cryo-protectant is the same manner as described above, but is frozen by first being immersed in liquid ethane, then liquid nitrogen before being stored. The higher heat capacity of liquid ethane may allow faster cooling rates.

2.4 Protein X-ray Crystallography

This is a method by which a model of a protein molecule can be made, as it is not possible to see it directly with visible light (electromagnetic radiation with wavelengths (λ) 400-700 nm). This is because in order to ‘see’ something, the object being viewed must be larger than the λ of light. Individual atoms within a protein molecule can therefore not be seen because the distance between these atoms is only about 0.15 nm (1.5 Å) and it is X-rays that have wavelengths of around this magnitude. However, X-rays cannot be focused by lenses as visible light can. To get around this problem, the direction and intensity of diffracted X-rays are measured (as diffraction patterns) at different angles around the protein crystal. Computer programs can then be used to analyse this data and construct a model of the molecule based on the observed data.

2.4.1 Data Collection

Data is collected from the protein crystal by firing X-rays at it from different angles (often 180° at 1° intervals). The X-rays diffracted by the electron clouds are recorded for each angle as diffraction patterns. The spots created by an X-ray striking the detector are called reflections, the darker it is the greater the intensity of the X-ray beam that struck the detector (Figure 2.13).

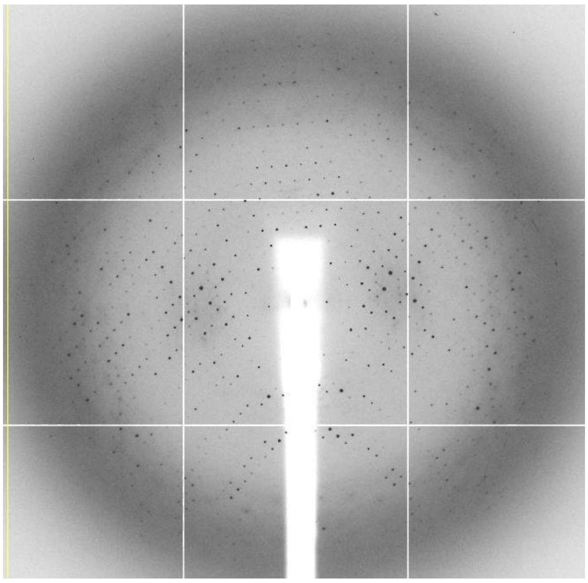


Figure 2.13: An X-ray crystallographic diffraction pattern
This 1° oscillation diffraction pattern is from 2,4-dihydroxyacetophenone dioxygenase. The crystal diffracts the source beam into many distinct rays, each of which produces a reflection (spot) on the detector. The position and intensity of these reflections contains the information necessary to reconstruct a model of the molecule.

2.4.2 Data Processing

The first step in data processing is to convert the images within a dataset to a list of reflections indexed by their position or Miller indices (h,k,l), intensity estimates (I) and the standard error of those intensities (SIGI). All of this information is recorded in a reflection file (.mtz). The program chosen to perform this function was MOSFLM (part of the CCP4 suite), although other commonly used programs include XDS and HKL2000. Although the basic principles behind these programs are similar, the methods utilised by each do differ. Firstly the unit cell for the particular crystal must be determined. This is done by measuring the intensity and position of a number of reflections, typically 100-200 from each image. This information, along with the detector distance, X-ray wavelength and detector type, is to construct the Bravais lattices and lattice type in reciprocal space.

The crystallographic unit cell dimensions are described by the length of three edges (a, b, c) and three angles (α, β, γ). The values of these dimensions will determine which of the seven crystal systems the unit cell falls into. Combinations of these seven crystal systems (cubic, tetragonal, orthorhombic, rhombohedral, hexagonal, monoclinic and triclinic) coupled with the four lattice centerings (P - primitive, C - centred, F – face-centered and I - internal) gives rise to 14 possible Bravais lattice types shown in Figure 2.14.

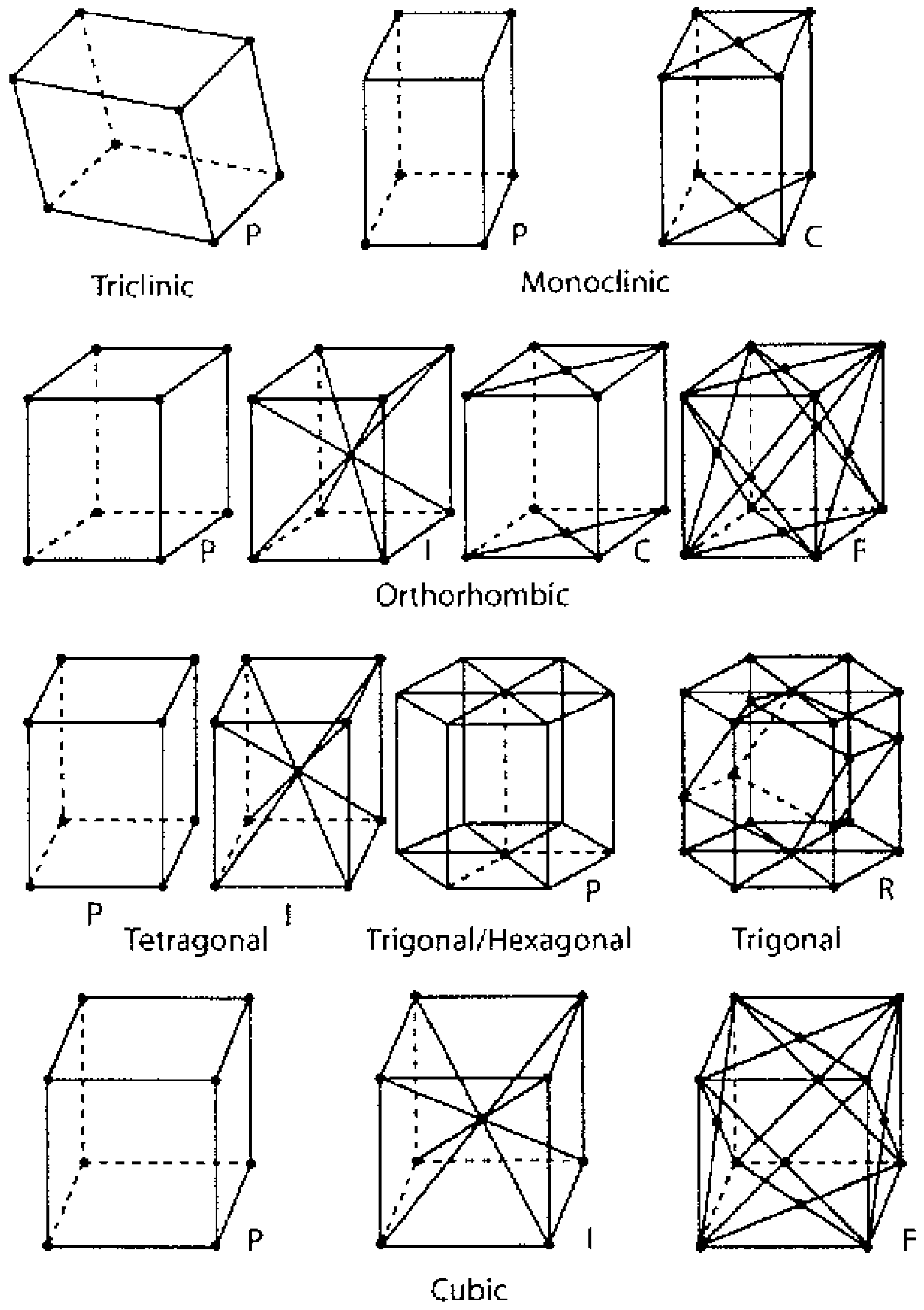
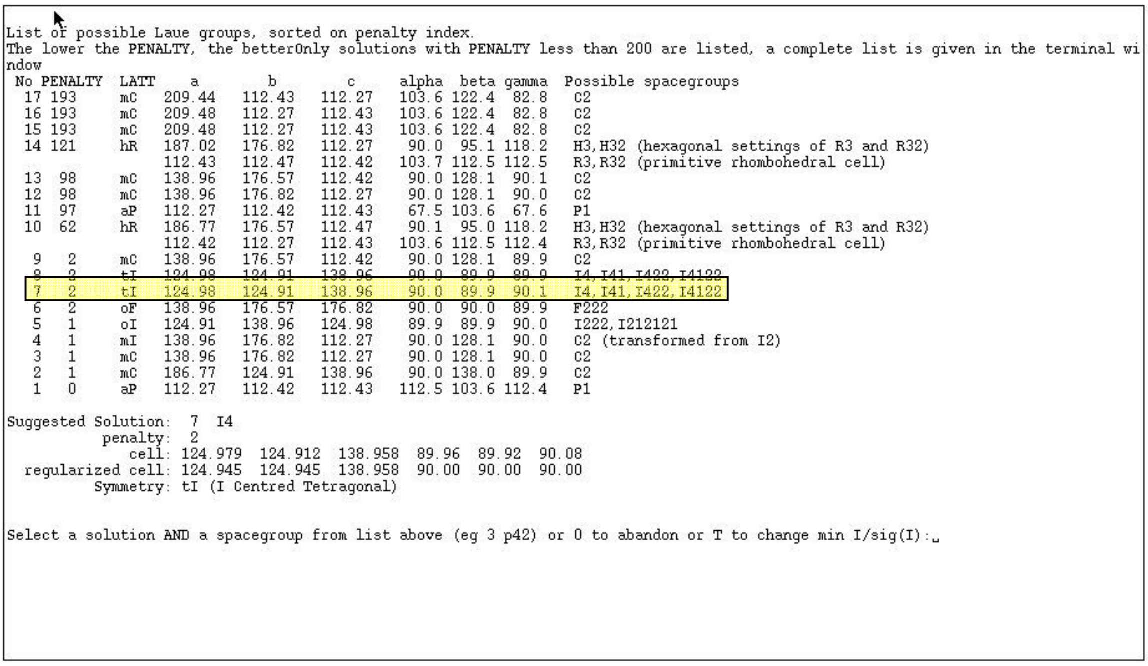


Figure 2.14: The 14 Bravais lattices

These are the distinct lattice types which when repeated can fill the whole space. There are 14 in three-dimensional space, consisting of 7 crystal systems (cubic, tetragonal, orthorhombic, rhombohedral, hexagonal, monoclinic and triclinic) coupled with the 4 lattice centerings (P - primitive, C - centred, F – face-centered and I - internal)

The Bravais lattices only describe how the unit cell can be translated in three dimensions to obtain crystal morphology; the point group provides information on how the asymmetric unit is positioned within the unit cell. The possible symmetry operations that can be imposed on protein crystals are translations, rotations and screw axes (rotations and translations combined). The combination of these possible symmetry operations of the point groups with the Bravais lattices, results in a total of 65 allowed space groups relevant to protein crystallography. More space groups are possible, but because proteins consist of L-amino acids, many of the symmetry operations are not possible in protein crystals. Each space group is represented by a symbol such as $P4_32_12$, with the letter at the beginning being the lattice type (in this example primitive) and the numbers representing the symmetry operations within the unit cell.

Once the reflection spots have been picked by MOSFLM, autoindexing can be used choose a space group with the highest symmetry. Typically the solution with the highest symmetry and lowest penalty is chosen, as can be seen from Figure 2.15 where I4 is chosen over C2.

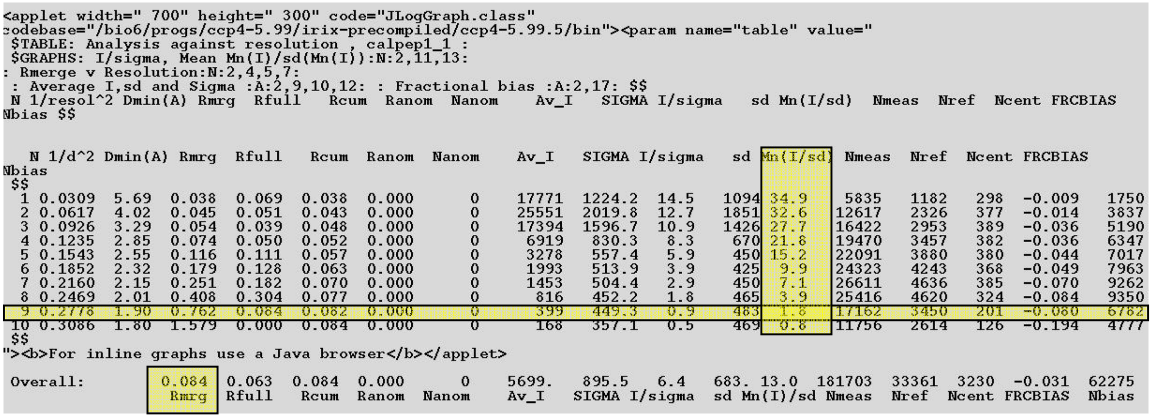


```
List of possible Laue groups, sorted on penalty index.
The lower the PENALTY, the betterOnly solutions with PENALTY less than 200 are listed, a complete list is given in the terminal wi
ndow
 No PENALTY  LATT    a         b         c      alpha  beta gamma  Possible spacegroups
  17 193      mC   209.44    112.43    112.27    103.6 122.4  82.8   C2
  16 193      mC   209.48    112.27    112.43    103.6 122.4  82.8   C2
  15 193      mC   209.48    112.27    112.43    103.6 122.4  82.8   C2
  14 121      hR   187.02    176.82    112.27     90.0  95.1 118.2   H3,H32 (hexagonal settings of R3 and R32)
                   112.43    112.47    112.42    103.7 112.5 112.5   R3,R32 (primitive rhombohedral cell)
  13  98      mC   138.96    176.57    112.42     90.0 128.1  90.1   C2
  12  98      mC   138.96    176.82    112.27     90.0 128.1  90.0   C2
  11  97      aP   112.27    112.42    112.43     67.5 103.6  67.6   P1
  10  62      hR   186.77    176.57    112.47     90.1  95.0 118.2   H3,H32 (hexagonal settings of R3 and R32)
                   112.42    112.27    112.43    103.6 112.5 112.4   R3,R32 (primitive rhombohedral cell)
   9   2      mC   138.96    176.57    112.42     90.0 128.1  89.9   C2
   8   2      tI   124.98    124.91    138.96     90.0  89.9  89.9   I4,I41,I422,I4122
   7   2      tI   124.98    124.91    138.96     90.0  89.9  90.1   I4,I41,I422,I4122
   6   2      oF   138.96    176.57    176.82     90.0  90.0  89.9   F222
   5   1      oI   124.91    138.96    124.98     89.9  89.9  90.0   I222,I212121
   4   1      mI   138.96    176.82    112.27     90.0 128.1  90.0   C2 (transformed from I2)
   3   1      mC   138.96    176.82    112.27     90.0 128.1  90.0   C2
   2   1      mC   186.77    124.91    138.96     90.0 138.0  89.9   C2
   1   0      aP   112.27    112.42    112.43    112.5 103.6 112.4   P1

Suggested Solution:  7  I4
           penalty:  2
              cell: 124.979  124.912  138.958  89.96  89.92  90.08
  regularized cell: 124.945  124.945  138.958  90.00  90.00  90.00
          Symmetry: tI (I Centred Tetragonal)

Select a solution AND a spacegroup from list above (eg 3 p42) or 0 to abandon or T to change min I/sig(I):
```

Figure 2.15: Crystallographic space group prediction by MOSFLM
MOSFLM uses the positions of the reflections, along with the beam position, to predict a space group for the data. This can be more accurately determined by refining the cell so that it picks a space group best fitted to all the data.

After the initial indexing solution has been picked, it can be further refined by comparing predictions across user defined images to try to find the "best" cell that fits the whole dataset. Often this is carried out using at least 2 segments of 2 images 90 ° apart. This solution can be tested by predicting the positions of spots based on the images used to refine the cell, which should closely match the positions of actual observed reflections. If this is the case then integration can be carried out to convert all the observed spots into a list of measured reflections. Integration extracts intensity values for all of the reflections from each image in the dataset by summing up pixel values for each spot. From the indexing parameters that were previously determined and refined, MOSFLM knows where to look on each image for all of the hkl reflections that are observed. The success of the integration can be measured by viewing the summary file to ensure the "RESID" is below 0.15, the "WRESID" is close to 1.0 and the "I/sgI_out" is around 2.

The program Sortmtz can be used to sort all the reflections before using Scala to scale all the equivalent symmetry related reflections together. By integrating the images, the intensities of typically 100,000 hkl reflections that have been extracted are combined to leave only the 20,000 or so of these that are unique. Many of the original observations will actually be separate measurements of the same hkl reflection or of symmetry related hkl reflections. All the redundant intensity measurements of the hkl reflections can be merged into single, averaged values for the intensity of each symmetrically unique reflection hkl. This is achieved by the program calculating a scale factor for each image and applying it the data to create uniformity in intensity across all images. The log file gives a good indication of the quality of the data (Figure 2.16); particularly useful are the values for mean (I/sd), multiplicity, completeness and R_{merge}. The R_{merge} value is a measure of the deviation in intensity between all the observed symmetry related reflections, which result from the range of symmetry elements in the space group. Freidel's law states that if the correct space group for the crystal has been chosen, the intensity of these reflections should be identical. The program Truncate is run with Scala to convert any small negative intensities to positive values and then convert all data into mean structure amplitudes.



```
<applet width=" 700" height=" 300" code="JLogGraph.class"
codebase="/bio6/progs/ccp4-5.99/irix-precompiled/ccp4-5.99.5/bin"><param name="table" value="
 $TABLE: Analysis against resolution , calpep1_1 :
 $GRAPHS: I/sigma, Mean Mn(I)/sd(Mn(I)):N:2,11,13:
: Rmerge v Resolution:N:2,4,5,7:
 : Average I,sd and Sigma :A:2,9,10,12: : Fractional bias :A:2,17: $$
 N 1/resol^2 Dmin(A) Rmrg  Rfull   Rcum  Ranom  Nanom    Av_I   SIGMA I/sigma   sd Mn(I/sd)  Nmeas  Nref  Ncent FRCBIAS
Nbias $$


   N 1/d^2 Dmin(A) Rmrg   Rfull   Rcum  Ranom  Nanom    Av_I   SIGMA I/sigma   sd Mn(I/sd) Nmeas  Nref  Ncent FRCBIAS
Nbias
 $$
  1 0.0309  5.69  0.038  0.069  0.038  0.000      0   17771  1224.2  14.5  1094 34.9   5835  1182   298  -0.009  1750
  2 0.0617  4.02  0.045  0.051  0.043  0.000      0   25551  2019.8  12.7  1851 32.6  12617  2326   377  -0.014  3837
  3 0.0926  3.29  0.054  0.039  0.048  0.000      0   17394  1596.7  10.9  1426 27.7  16422  2953   389  -0.036  5190
  4 0.1235  2.85  0.074  0.050  0.052  0.000      0    6919   830.3   8.3   670 21.8  19470  3457   382  -0.036  6347
  5 0.1543  2.55  0.116  0.111  0.057  0.000      0    3278   557.4   5.9   450 15.2  22091  3880   380  -0.044  7017
  6 0.1852  2.32  0.179  0.128  0.063  0.000      0    1993   513.9   3.9   425  9.9  24323  4243   368  -0.049  7963
  7 0.2160  2.15  0.251  0.182  0.070  0.000      0    1453   504.4   2.9   450  7.1  26611  4636   385  -0.070  9262
  8 0.2469  2.01  0.408  0.304  0.077  0.000      0     816   452.2   1.8   465  3.9  25416  4620   324  -0.084  9350
  9 0.2778  1.90  0.762  0.084  0.082  0.000      0     399   449.3   0.9   483  1.8  17162  3450   201  -0.080  6782
 10 0.3086  1.80  1.579  0.000  0.084  0.000      0     168   357.1   0.5   469  0.8  11756  2614   126  -0.194  4777
 $$
"><b>For inline graphs use a Java browser</b></applet>

 Overall:          0.084  0.063  0.084  0.000      0   5699.   895.5   6.4   683. 13.0  181703  33361  3230  -0.031  62275
                    Rmrg  Rfull   Rcum  Ranom  Nanom   Av_I   SIGMA I/sigma  sd Mn(I)/sd Nmeas  Nref  Ncent FRCBIAS  Nbias
```

Figure 2.16: An exert from the log file produced by Scala
This table illustrates the statistics used as a measure of data quality. The Mn (I/sd) should be about 2 and the R_{merge} low, ideally below 0.1. This indicates good data that has processed well.

2.4.3 Phase Determination

As well as knowing the position and intensity of the diffracted X-ray reflections, it is also necessary to know the phase of the wave diffracted from the electron cloud of the molecule, but this information is lost during X-ray crystallography. Knowledge of the phases is necessary to make use of the reflection data collected because both are required to compute an image of the molecule.

The recorded reflections can be described as complex three dimensional waves, which result from the superposition of the X-ray waves scattered by the protein's electron density, and can be described by a structure factor F_{hkl}:

$$F_{hkl} = V\int_0^1\int_0^1\int_0^1 \rho(x,y,z)e^{2\pi i(hx+ky+lz)}dxdydz$$

This equation is used to relate the electron density in real space ρ (x, y, z) to the reflections measured in reciprocal space, after correcting for unit cell volume (V). F_{hkl} is a periodic function describing a complicated wave and consists of a frequency, amplitude $|F_{hkl}|$ and phase α_{hkl}:

$$F_{hkl}=|F_{hkl}|\ e^{2\pi i\alpha hkl}$$

The amplitude $|F_{hkl}|$ can be obtained directly from the diffraction data because it is directly proportional to the square root of the measured intensity I_{hkl} for the reflection hkl. However, the phase angle of each reflection is unknown and is required to calculate ρ (x,y,z), and so obtain the positions of individual atoms in the protein molecule. It is therefore necessary to obtain some phase information in order to determine a protein structure.

If the structure of a similar protein has already been solved, then it can be used in molecular replacement to provide estimates of the phases for the new model. The idea behind this method is to find the rotation and translation values that position the model structure in the experimental unit cell so as to give the highest correlation between observed diffraction measurements and those calculated from the model. This method was used for the structure determination of TkTDH. However, because both BipD and DAD are novel proteins with no similar structures previously solved, this strategy is not an option and multiwavelength anomalous diffraction data must be collected from derivatives of these proteins, in order to calculate the phases.

2.4.3.1 Molecular Replacement

Molecular replacement involves borrowing the phases from a previously solved model and applying them to the unknown structure as an estimate of the phases. This is perhaps the simplest method by which phases can be determined for structure determination, so long as a similar known structure is available with at least 40 % sequence or structural similarity to the unknown structure. Molecular replacement becomes progressively more difficult as the sequence identity between the known and unknown proteins is reduced, or the completeness of the model decreases. This method involves using the known model to determine the position (translation) and orientation (rotation) of the unknown molecule in the crystallographic unit cell.

There are two main methods by which molecular replacement can be achieved. Traditional molecular replacement methods are based on the properties of the Patterson function, which is the Fourier transform of the experimental intensities. From this can be generated a Patterson map; a vector map with peaks at the positions of vectors between atoms in the unit cell and is thus unique to the protein. Programs that use this method include MOLREP (Vagin and Teplyakov 1997).

However more recent programs, such as PHASER (McCoy 2007), use maximum likelihood-based methods to perform molecular replacement, and this was the program of choice. Molecular replacement requires searching in six dimensions, so to simplify this, the rotation and translation functions are performed separately. The rotation function must first be solved to have any chance of solving translation function, because intermolecular vectors are determined by orientation as well as position. When searching for more than one monomer in the asymmetric unit, PHASER will first search for one and once it has found it will fix it in position, before searching for the next monomer.

The Log Likelihood Gain (LLG) used in PHASER, is a measure of the probability that the data would have been measured, given the model, so it allows us to compare how well different models agree with the data. In the case of molecular replacement, the model consists of the atomic coordinates plus rotation and/or translation operators applied to those coordinates. The LLG is the difference between the likelihood of the model and the likelihood calculated from a Wilson distribution, so it measures how much better the data can be predicted with your model than with a random distribution of the same atoms. The LLG allows comparison of different models against the same data set, but the LLG values for different data sets should not be compared with each other.

2.4.3.2 Anomalous Diffraction Phasing

Multiwavelength anomalous diffraction (MAD) phasing requires the presence of one or more atoms with X-ray absorption in the energy range for which X-ray intensity data is to be measured (normally around 1Å). The absorption edges of C, N, O, S and H are far away from the accessible energy range and are not generally useful for MAD phasing. However, elements with atomic numbers between roughly 20 and 40, or that are greater than 60, have accessible absorption edges. These include the transition metals, as well as atoms such as selenium, that can be artificially introduced into macromolecules. MAD phasing experiments carried out in this project used BipD in which methionine residues were replaced by selenomethionine (SeMet BipD).

The X-ray diffraction experiment consists of measuring X-ray intensity data at several different wavelengths at or near the absorption edge of the anomalous scattering atoms. Changing the wavelength changes the intensities of the individual reflections and the differences in intensities can be used to directly calculate the phase angle for each reflection. In this research the experiments were carried out on beamlines at the ESRF, Grenoble.

Once experimental data is available, the next step is to locate the anomalous scattering selenium atoms. For structures with only a few such atoms, the locations can usually be determined by inspection of the appropriate Patterson function. However if the structure has many anomalous scattering atoms then the Patterson function is too complicated for manual interpretation. Another approach is based on automated interpretation of the Patterson function in combination with difference Fourier techniques. This technique is the basis for the program SOLVE, which scales data, solves Patterson functions, calculates difference Fourier's, looks at a native Fourier to see if there are distinct solvent and protein regions, and can score partial MAD and multiwavelength isomorphous replacement (MIR) solutions to build up a complete solution. The resultant electron density map can then be improved using RESOLVE, by using a statistical approach to combine the experimental X-ray diffraction information with knowledge about the expected characteristics of an electron density map of a macromolecule.

2.4.4 Refinement and Model Building

Once an electron density map has been generated, it can be refined using programs such as REFMAC (Vagin *et al.* 2004), CNS (Brunger *et al.* 1998) and PHENIX (Adams *et al.* 2002), and model building carried out using programs such as TURBO-FRODO and COOT (Emsley and Cowtan 2007). Through a series of alternate refinement and manual model building steps, the experimental model can be modified to better fit the experimental data, and this is monitored by comparing the R_{free} value with the R_{factor}. The R_{free} is calculated using 5-10 % of the total reflections, which are not used in the refinement and thus provide an unbiased assessment of each refinement step. Both figures should be low, indicating good model agreement, and there should only be a small difference between the values, indicating good agreement between the observed data and the refined data model.

2.4.4.1 Refinement

Refinement is the iterative process of optimising the agreement between the observed structure factor amplitudes from the diffraction data and those calculated from the atomic model, also taking into consideration chemical parameters such as bond angles and length. A balance is reached between the weight carried by the experimental data versus that of known chemical parameters, depending on the quality and resolution of the experimental data; the better it is the less supplementary restraints are required to ensure a good model is produced.

Atoms are able to exist in a variety of energy states, resulting in many local energy minima as well as the global minimum, which is where the atoms should be. This problem can be solved through the use of programs like CNS and PHENIX, which are able to overcome the energy barrier through the use of a technique called simulated annealing. Annealing is a physical process in which a solid is heated until all atoms randomly arrange themselves in a liquid phase and is then slowly cooled, enabling the atoms to reach the lowest possible energy state. Simulated annealing is a computational simulation of this annealing process and increases the probability of finding a more optimal position for an atom. These local minima can also be overcome by manual model building. Refinement programs such as CNS and PHENIX are also able to refine atomic positions with the help of constraints and restraints. For example the occupancy of a particular atom can be set or constrained to a specific value during refinement. This is usually set a 1.0 (or 100 %), but for TkTDH the occupancies of the NAD^+ atoms were set at 0.5 to best account for the fact that only 50 % for the co-factor binding sites were occupied. Restraints are more flexible, for example bond angles and lengths are usually restrained to a particular range of values as opposed to being given a single set value.

Once the model had been refined a few times and therefore represented a better fit with the experimental data, another refinement program, REFMAC, was used. These programs not only differ in their user interface, but also in their method of parameter refinement. REFMAC uses the method of maximum likelihood in which atom positions are selected that maximise the probability of the model. REFMAC can be used to refine group B-factors, ligand occupancy and NCS restraints. In the final rounds of refinement the NCS restraints are relaxed and eventually removed altogether, as the model is now good enough to be refined without these.

Refinement programs produce files containing the coordinates of each atom (.pdb) as well as F_o-F_c / $2F_o$-F_c maps, which record the electron density and difference density (.mtz or .map). These are used by model building programs to manually adjust the model.

2.4.4.2 Model Building

The graphics program COOT was used for the majority of model building in these structures, although TURBO-FRODO was also utilised in some instances. The aim of such programs is to better fit the model to the observed electron density by making manual alterations not carried out during refinement. The $2F_o$-F_c electron density map depicts a molecular surface net extended over the atoms in the molecule, as shown in Figure 2.17.

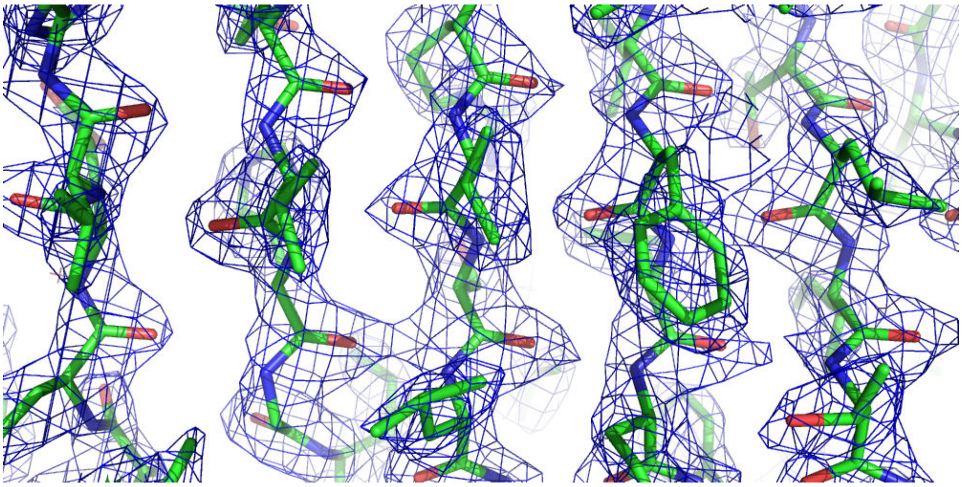


Figure 2.17: Electron density from a $2F_o$-F_c map

The F_o-F_c map is used to aid model building by highlighting areas of positive and negative density. Positive density implies that the contribution of the observed intensities (F_o) to ρ is greater than that of the calculated structure factors (F_c), therefore the unit cell contains more electron density than implied by the current model and requires the movement of atoms towards this area to increase the electron density. This often occurs if atoms are absent from a part of the model where they should actually be present. The converse is true for negative density, and suggests that atoms must be moved away from this area, i.e. they are not actually there in the protein. Figure 2.18 illustrates part of the model incorrectly built, showing negative density where a residue should not be, and positive density where a residue should be.

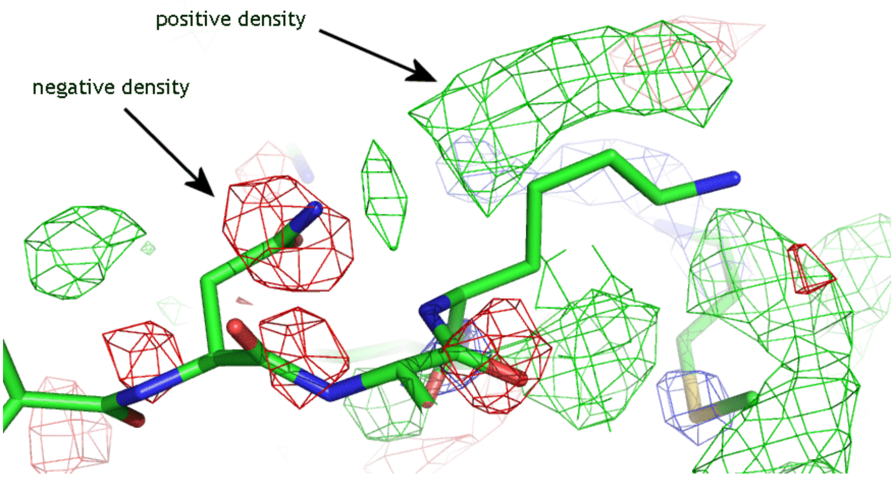


Figure 2.18: Positive and negative F_o-F_c difference electron density

COOT was first used to mutate the initial model sequence (determined by the search model) to the correct sequence for the protein structure being determined. It could then be used to overcome local energy minima in the model by manual intervention in real space to move residues into predicted locations, guided by the electron density and difference maps, as illustrated in Figure 2.19.

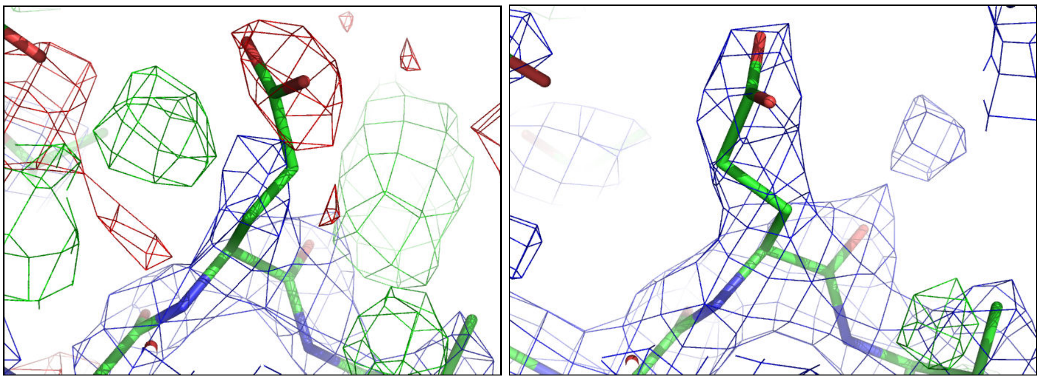
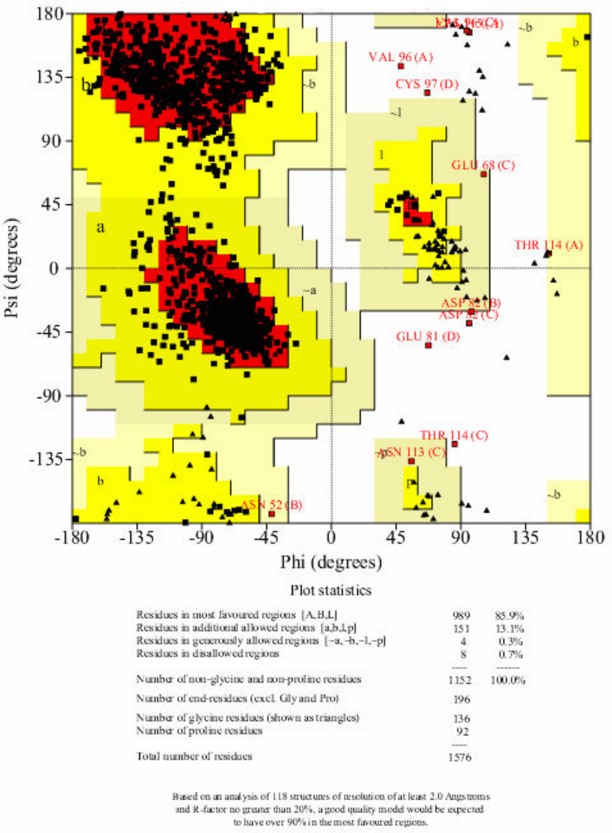


Figure 2.19: Modified residue in electron density
The picture on the left show the electron density before refinement and the picture on the right shows the improved density after refinement.

Once the model was a reasonable fit with the electron density, ligands were modelled in as necessary, i.e. for TkTDH density was observed for NAD$^+$. The RCSB PDB database was used to obtain the atomic coordinates (.pdb file) for NAD$^+$. This data was then fed into the online program PRODRG, which generated a dictionary file for the ligand (.cif file). These files could then be read into COOT and the ligand modelled into the appropriate density in each monomer. The ligand and protein model .pdb files were merged and the.cif file used to refine the ligand. Lastly, waters were added to the model.

2.4.5 Validation and Structural Analysis

Once the R_{factor} and R_{free} were as low as could be achieved, the online validation server Procheck (Laskowski *et al.* 1993) was used to assess the validity of the model. Parameters such as bond lengths were checked and the deviation of the peptide bonds φ (phi) and ψ (psi) from the standard angles shown on a Ramachandran plot (Figure 2.20).



Plot statistics

Residues in most favoured regions [A,B,L]	989	85.9%
Residues in additional allowed regions [a,b,l,p]	151	13.1%
Residues in generously allowed regions [~a,~b,~l,~p]	4	0.3%
Residues in disallowed regions	8	0.7%
	----	------
Number of non-glycine and non-proline residues	1152	100.0%
Number of end-residues (excl. Gly and Pro)	196	
Number of glycine residues (shown as triangles)	136	
Number of proline residues	92	

Total number of residues	1576	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 2.20: Ramachandran plot showing the peptide bond angles Phi and Psi

Residues in the most favoured regions lie in the red areas and include those angles that best fit the standard bond angles. Others lie in the additional allowed regions (yellow) and generously allowed regions (beige). Those residues that lie in the disallowed region are those whose φ and ψ angles deviate significantly from the standard angles. Other factors that were validated were the B-factors, cis-peptides and water molecules.

Analysis was carried out using malign (Wheeler and Gladstein 1994) and alscript (Barton 1993) to align the primary sequences. Structural figures were prepared using CueMol, PyMOL (DeLano 2008) and CCP4MG (Potterton *et al.* 2004).

Chapter 3

Structural studies of Burkholderia pseudomallei Invasion Protein D from *Burkholderia pseudomallei*

3.1 Introduction

Translocator proteins like BipD are part of the type III secretion apparatus used by bacteria to insert selected proteins into the cells of host organisms in order to infect the target cell and replicate. BipD is expressed by the gram-negative saprotroph *Burkholderia pseudomallei*, the causative agent of melioidosis. This is a serious infectious disease of both animals and humans, endemic to tropical and subtropical regions particularly Southeast Asia and Northern Australia (Gan 2005). Most commonly this disease manifests itself clinically as abscesses, pneumonia and, at worst, as a fatal septicaemia, but actually has a complex spectrum of clinical manifestations making it difficult to diagnose and treat effectively (Aldhous 2005). *B. pseudomallei* is known to infect most cell types and it appears able to evade, and perhaps actively interfere with, the host's immune system (Gan 2005). BipD may also be involved in the bacterium's ability to escape from endocytic vesicles through actin polymerisation and the formation of protrusions, facilitating cell to cell spread (Stevens *et al.* 2004).

Melioidosis does not tend to affect Western visitors to endemic countries, so consequently research into the bacterium attracts relatively little funding. However the West's awareness of *B. pseudomallei* has increased in recent years since being identified and labelled as a potential biological terrorist weapon, thanks to which research into the organism is now being promoted at the US National Institute of Allergy and Infectious Diseases through its biodefence initiative (Aldhous 2005).

The BipD protein is a component of the *B. pseudomallei* type III protein secretion system (TTSS) where it is a translocator protein that transports bacterial effector proteins across the plasma membrane into a host cell. These then subvert or inhibit the cell's processes for the benefit of the bacterium. It is thought BipD is located at the tip of the secretion needle and may act as a 'plug' to regulate the secretion of effector proteins, and as a 'sensor' to tell the bacterium when to secrete them.

B. pseudomallei strains containing mutated BipD exhibit reduced virulence but this does not eliminate the development of fatal melioidosis, indicating that it is part of a system, rather than acting alone (Stevens *et al.* 2004). It is hoped that by elucidating the structure of BipD we will better understand how this specific TTSS functions, potentially offering a target for prophylactic drugs against melioidosis. There are several known homologues of BipD in TTSS of other bacteria so knowledge of its structure will also contribute to a better understanding of how these TTSS function generally, maybe offering up potential targets against other diseases.

3.2 BipD Expression and Purification Protocol

3.2.1 DNA Manipulation

The initial BipD expressed was a GST-fusion protein, but later problems during purification necessitated the design of the bipd gene without the GST tag. Agarose gel electrophoresis was used throughout to provide qualitative analysis of DNA samples (see appendix for method).

3.2.2 Transformation

JM109 cells were used to obtain the bipd plasmid for small scale plasmid purification (although BL21(DE3) were later used for overexpression). 20 µl JM109 competent cells from the -80ºC freezer were thawed on ice. bipd DNA was mixed with the cells and left on ice for 5 minutes. These were then heat-shocked by immersing in 42 ºC water for 1 minute then put back on ice. 1 ml Luria-Bertani growth medium (LB) was added to the cells and they were grown at 37 ºC for 1 hour on a shaker.

The sample was centrifuged at 13,000 rpm for 3 minutes and 900 µl supernatant discarded. The pellet was resuspended in the remaining 100 µl and spread onto an ampicillin agar plate (see appendix for method). This was left to grow overnight at 37 ºC. 4 colonies were picked from the plate and used to inoculate 4 separate 10ml bottles of LB + 5 µl ampicillin. These were grown overnight at 37 ºC. The cultures were centrifuged at 14,000 rpm for 5 minutes and the supernatant discarded.

3.2.2.1 Plasmid Purification

A plasmid purification kit (QIAprep Spin Miniprep Kit) was used to extract the DNA. The pelleted bacterial cells were resuspended in 250 µl Buffer P1 (containing RNase A) and transferred to a microcentrifuge tube. 250 µl lysis Buffer 2 was then added and the tube gently inverted 4-6 times to mix, after which the solution turned blue. 350 µl Buffer N3 was added and the tube immediately inverted 4-6 times to mix. The microcentrifuge tube was centrifuged for 10 minutes at 13,000 rpm.

The resultant supernatant was decanted into a QIAprep spin column and centrifuged for 60 seconds. The flow-through was discarded and the column washed using 0.5 ml Buffer PB to remove trace nuclease activity. The column was centrifuged for 60 seconds at 13,000 rpm and the flow-through discarded. 0.75 ml Buffer PE was used to again wash the column and it was centrifuged for 60 seconds at 13,000 rpm. The flow-through was discarded and the column again centrifuged for 60 seconds. The column was placed in a microcentrifuge tube and the DNA eluted by adding 50 µl H_20 to the centre of the column, leaving them to stand for 60 seconds then centrifuging for 60 seconds.

3.2.2.2 Quantitative Analysis of DNA

A Pharmacia Biotech Ultrospec 3000 spectrophotometer was used to scan the DNA sample. A peak of 0.116 absorbance was observed at a wavelength of 260 nm, indicating 5.28 µg bipd DNA present. Any remaining liquid was evaporated off.

3.2.2.3 Primer Design

The bipD gene sequence is shown below:

```
CGGACCGGGTACATTTCCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGG
ACGCGCGCTGACCGTGCGCGATTGGCCGGCGCTCGAGGCGCTCGCGAAGACGATGCCGGCCGATGC
CGGCGCGCGGGCGATGACCGACGACGATCTGCGCGCAGCGGGCGTCGATCGCCGCGTGCCGGAGCA
AAAGCTCGGCGCGGCGATCGACGAATTCGCGTCGCTCCGGCTGCCCGATCGGATCGACGGGCGCTT
CGTCGATGGCCGCCGCGCGAACCTCACGGTGTTCGATGATGCACGCGTGGCGGTGCGCGGCCACGC
GCGCGCGCAGCGCAACCTGCTCGAGCGCCTGGAAACCGAGCTCCTGGGCGGCACGCTGGACACCGC
GGGCGACGAAGGCGGCATCCAGCCGGACCCGATCCTTCAGGGGCTCGTCGACGTGATCGGCCAGGG
CAAATCCGATATCGATGCGTACGCAACGATCGTCGAGGGGCTGACGAAGTACTTCCAGAGCGTCGC
```

```
CGACGTGATGAGCAAGCTGCAGGACTACATCTCGGCGAAAGACGACAAGAACATGAAGATCGACGG
CGGCAAGATCAAGGCGTTGATCCAGCAGGTCATCGACCATCTGCCGACGATGCAGTTGCCGAAAGG
GGCCGACATCGCGCGCTGGCGCAAGGAGCTCGGCGATGCCGTCTCGATCAGCGATTCGGGCGTCGT
GACGATCAATCCGGACAAGCTGATCAAGATGCGCGATTCGCTGCCCCCTGACGGCACGGTGTGGAC
ACCGCGCGCTACCAGGCCTGGAACACCGCGTTCTCCGGGCAGAAGGACACATCCAGAACGATGTGC
AGACGCTCGTCGAAAATACTCGCACCAGACTCGAACTTCGACATCTGGTCAGTGCTGAGCGGCGCG
ATCTCGACGCTCACGACACGCAGAGCTATCTGCAGATCTGATAGGCGATCGCTGCTACAAGCCGAA
GAGCTGAATGCTGCTGCACGCTGACATACTAGCATACCTGGCTCTACGTCTGAGGGATTTTGCTGA
GAGACTATCCGGATTCCGCAAGCCGCGTACGGCTACCAGCTATGCTACGATCAGGTACGTCAGATG
ACTGAGCCATGTGATTCTACCGCGCTAGTCTGCATACGCAATCGTAGCTACGGTAACCGTCACTGA
```

Translating to:

```
Met G R A L T V R D W P A L E A L A K T Met P A D A G A R A Met T
D D D L R A A G V D R R V P E Q K L G A A I D E F A S L R L P D R
I D G R F V D G R R A N L T V F D D A R V A V R G H A R A Q R N L
L E R L E T E L L G G T L D T A G D E G G I Q P D P I L Q G L V D
V I G Q G K S D I D A Y A T I V E G L T K Y F Q S V A D V Met S K
L Q D Y I S A K D D K N Met K I D G G K I K A L I Q Q V I D H L P
T Met Q L P K G A D I A R W R K E L G D A V S I S D S G V V T I N
P D K L I K Met R D S L P P D G T V W T P R A T R P G T P R S P G
R R T H P E R C A D A R R K Y S H Q T R T S T S G Q C Stop
```

Primers were designed for the bipd gene. The oligonucleotide sequences chosen were:

5' end primer (Tm 63 ºC)

`ggatctagacatatgggacgcgcgctgaccgtgcgcgat`

3' end primer (Tm 64 ºC)

`ggcgcaagcttggatccgccctatcagatctgcagatagct`

These sequences were run through the online program webcutter (http://rna.lundberg.gu.se/cutter2/) to ascertain if they contained restriction endonuclease cleavage sites for BamHI, HindIII, NdeI, XbaI. None of those endonucleases were predicted to cut the BipD sequence.

3.2.2.4 Polymerase Chain Reaction

PCR was used to amplify up the *bipd* gene. 1µl of DNA (100 ng/µl) was mixed with; 1 µl bipd 1 primer, 1 µl bipd 2 primer, 5 µl Mg-free buffer, 3 µl $MgCl_2$ 25 mM, 1 µl PCR mix nucleotides and 38 µl H_2O making a total volume of 50 µl. 0.25 µl *Taq* polymerase was added and centrifuged. The following method was programmed and the sample inserted:

Programme 1: 94 ºC for 1 min
Programme 2: segment 1: 94 ºC for 1 minute (melt)
segment 2: 55 ºC for 1 minute (anneal)] 30 cycles
segment 3: 72 ºC for 3 minutes (polymerisation)
Programme 3: 72 ºC for 10 minutes to complete all replication
Programme 4: 4 ºC for 24 hours.

The 50 μl DNA was cleaned using a QIAquick PCR purification kit. 250 μl of Buffer PB was added to the 50 μl PCR sample and mixed. This was transferred to a QIAquick column, centrifuged for 60 seconds at 13,000 rpm and the flow-through discarded. 0.75 ml Buffer PE was used to wash the column and it was centrifuged for 60 seconds. The DNA was eluted into an microcentrifuge tube by adding 30 μl elution buffer, leaving it to stand for 60 seconds then centrifuging for 60 seconds. The univap was used to evaporate off any remaining liquid.

3.2.2.5 DNA Digest

The DNA was resuspended in 10 μl H_2O and digested using the restriction endonucleases Nde1 and BamH1. The pET11a plasmid was also digested using the restriction enzymes independently as controls to check for a successful digest.

	vol DNA	BSA	H_2O	Nde1	Buffer D	BamH1
BipD	5	1	1	1	1	1
pET11a (N)	1	1	6	1	1	0
pET11a (B)	1	1	6	0	1	1

The samples were set up as shown here but initially without the BamH1. These were incubated at 37 ºC for 2 hours after which BamH1 was added to the appropriate samples. They were then incubated for a further 2 hours and an agarose gel run to check that the digest was successful.

3.2.2.6 DNA Ligation

The ligations were set up using a 1:8 ratio of plasmid to insert: 1μl digested DNA, 1 μl double cut pET11a plasmid, 1 μl ligase, 1 μl buffer, 1 μl BSA and 5 μl H_2O. This was left for 4 hours at 16 ºC.

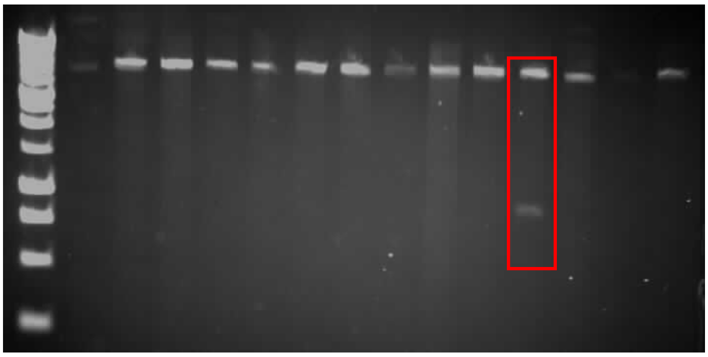


Figure 3.1: An agarose gel showing a plasmid ligation sample
Lane 12 (highlighted) shows 2 fragments, the plasmid at the top of the gel and the insert further down, indicating that the ligation was successful.

3.2.2.7 Large-scale Preparation of Plasmid

The plasmid containing the insert coding for BipD was used to transform BL21 *E.coli* cells, as described above and these were plated onto an ampicillin agar plate and grown overnight at 37 ºC. A single colony was picked and used to inoculate 5 ml LB with ampicillin and grown overnight at 37 ºC. The cells were harvested by centrifugation at 14,000 rpm for 15 minutes and the supernatant discarded.

The bacterial cell pellet was resuspended in 10 ml lysis Buffer P1 and 10 ml Buffer P2 was mixed in by gentle repeated inversion. This was incubated at room temperature for 5 minutes. 10 ml of chilled Buffer P3 was added and the solution immediately mixed before being incubated on ice for 20 minutes. The sample was centrifuged at 14,000 rpm for 30 minutes. The supernatant containing the plasmid DNA was removed and centrifuged for a further 15 minutes.

A QIAGEN-tip 500 was equilibrated by applying 10 ml Buffer QBT and allowing the column to empty by gravity flow. The supernatant was then applied and allowed to enter the resin by gravity flow. The QIAGEN-tip was washed with 2x30 ml Buffer QC to ensure the removal all contaminants and the DNA eluted using 15 ml Buffer QF. The DNA was precipitated by the addition of 10.5 ml isopropanol and centrifuged at 8,000 rpm for 30 minutes. The DNA pellet was washed with 10 ml room-temperature 70 % ethanol and centrifuged at 8,000 for 15 minutes. The supernatant was discarded and the pellet allowed to air-dry overnight.

The DNA was dissolved in 1ml autoclaved H_2O and centrifuged for 1 minute at 13,000 rpm to remove any particles. Quantitative analysis at 260 nm showed a quantity of 396 μg DNA. It was then dried in the univap, resuspended in 396 μl H_2O to a concentration of 1 μg/μl and frozen. The DNA was sequenced and the correct sequence confirmed.

3.2.3 Protein Expression

3.2.3.1 Protein Expression of Seleno-methionine Protein

Because crystals of the native protein had already been grown, it was necessary to express a selenomethionine derivative to enable structure determination by anomalous X-ray diffraction. At that time no homologues of BipD had been sequenced, hence structure determination required a heavy-atom derivative to calculate the phases of the X-rays as they diffract from the protein crystal. This was achieved through the incorporation of selenium atoms.

A starter culture of BipD (with the GST-tag for the SeMet protein and without for the native protein) was grown up overnight in 10ml with ampicillin (always at a final concentration of 100 μg/ml unless otherwise stated), at 37 ºC. 1 L LB medium was inoculated with a 10ml starter culture and ampicillin and grown at 37 ºC to mid-log phase with an optical density of approximately 0.6 OD_{600}. At this stage the culture was decanted into a sterile (ethanol washed) 1000 ml Beckman centrifuge bottle and centrifuged at 7,000 rpm for 20 minutes.

Ampicillin and 12 ml of 20 % sterile glucose was added to 1 L of sterile M9 minimal medium (see appendix for method) and 20 ml removed for later use.

Once the LB culture had been centrifuged, the supernatant was discarded and the pellet resuspended in 20 ml M9 media and kept on ice for 1 hour. This was then used to inoculate the 1 L of M9 minimal medium and incubated at 37 ºC for 45 minutes, during which time the cells are starved of many nutrients including methionine. The subsequent addition of 10 ml amino acid mix 1, 10 ml amino acid mix 2, 1 ml vitamins and 4 ml seleno – L – methionine (see appendix for methods) forces the cells to use selenomethionine and leads to the incorporation of it in place of methionine. This culture was grown for 20 minutes at 37 ºC then heat shocked by shaking in a 42 ºC water bath for 25 minutes, followed by 5 minutes on ice water. A final concentration of 0.3 M IPTG was used to induce BipD expression and the culture was grown at 16 ºC for 48 hours to encourage slow induction of protein and avoid the formation of inclusion bodies.

3.2.3.2 Over-expression of Native Protein

It was attempted to grow further native BipD crystals to improve the resolution of X-ray data collected. A starter culture of BipD was grown up overnight in 10ml and ampicillin, at 37 ºC. 1 L LB medium was inoculated with a 10 ml starter culture and ampicillin and grown at 37 ºC to mid-log phase with an optical density of approximately 0.6 OD_{600}. It was induced with filter-sterilised IPTG to a final concentration of 0.5 M and grown overnight at 37 ºC, Figure 3.2.

3.2.3.3 Harvest Cells

The 1L bacterial cell culture was centrifuged at 7,000rpm for 1 hour and the supernatant discarded. The bacterial cell pellet was resuspended in 20 ml 50 mM TRIS pH 8.

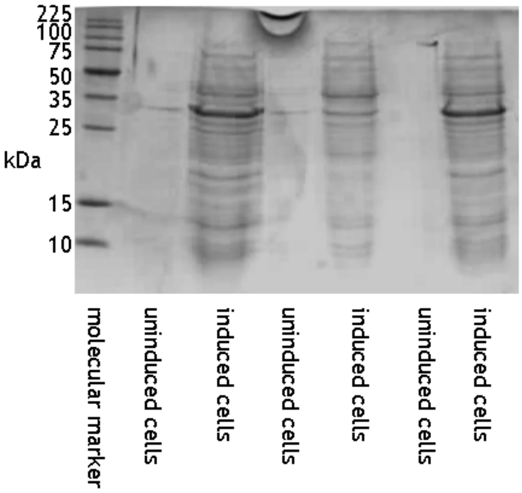


Figure 3.2: SDS gel showing pre and post IPTG induced samples
The gel shows no induction of BipD in the uninduced cells, or in lane 5, but over-expression of the protein in lanes 3 and7.

3.2.3.4 Sonication

To remove soluble fractions (including the BipD protein) from the cell, sonication was used to disrupt the cell membrane and allow the contents to spill out. The sample was put on ice and sonicated in a Soniprep 150 sonicator for 10 cycles of 30 seconds ON and 90 seconds OFF, then centrifuged at 18,000 rpm for 1 hour to remove the cell debris. The supernatant was separated off and the pellet resuspended in 5 ml 50 mM TRIS pH 8. SDS gel electrophoresis was used to confirm the induction of BipD and indicate whether it was present as soluble or insoluble protein.

3.2.4 Crude Purification

3.2.4.1 Ammonium Sulphate Precipitation

This is a crude method by which a soluble protein can be separated from the initial cell mixture because different ammonium sulphate concentrations cause different proteins to precipitate out of solution. Impurities are precipitated out by raising the ammonium sulphate concentration up to the point at which BipD just remains soluble, and then to raise the concentration again to the point beyond which very little BipD remains in solution.

Steps of 15 % $(NH_4)_2SO_4$ addition were used to determine the precipitation points. It is known that 55 % saturation requires 0.35 g/ml $(NH_4)_2SO_4$, so for a volume of 25 ml supernatant, 2.38 g was added. This was mixed well then centrifuged at 18,000 rpm for 40 minutes to remove the precipitate. A sample of both the pellet and supernatant was taken for SDS gel-electrophoresis (see appendix for method). After the supernatant had been separated from the pellet a further 2.38 g $(NH_4)_2SO_4$ was added taking the saturation to 30 %. This was again centrifuged and the above steps repeated until 75 % saturation was achieved. Analysis by SDS gel-electrophoresis showed that BipD precipitated out at 45 % ammonium sulphate saturation (data not shown).

3.2.4.2 Dialysis

Dialysis is necessary to remove the high salt content present in the sample as a result of ammonium sulphate precipitation. 1m of 12kDa pore dialysis tubing was boiled for 10 minutes with $NaHCO_3$ and EDTA to remove any residual heavy metal impurities. The 45 % pellet was resuspended in 10 ml 50 mM TRIS pH8 and pipetted into the dialysis tubing. Any air was removed and the tubing sealed at both ends. It was immersed in 2 L 20 mM TRIS pH 8 and left stirring for 48 hours at 4 ºC.

3.2.5 Column Purification (with the GST-fusion protein)

All buffers used in column purification were degassed.

3.2.5.1 GSTrap Affinity Chromatography

The column was washed with dH_2O, binding buffer and elution buffer to remove any contaminants (see appendix for methods). A 2 ml sample of BipD was diluted with 3 ml binding buffer and loaded onto the column at 0.5 ml/min using an Akta Prime machine and washed with binding buffer to remove contaminants. 1.6 mg thrombin dissolved in 5 ml binding buffer was then loaded onto the column and left for 20 hours to allow it to cleave the GST – BipD bond.

The column was washed with 30 ml binding buffer to remove the BipD and thrombin from the column. The bound GST was eluted with 30 ml 50 mM TRIS containing 10 mM reduced L-glutathione pH 8.

3.2.5.2 Removal of Thrombin

The purification of BipD from thrombin was achieved by buffer-exchange to 20 mM TRIS pH 8 containing 500 mM NaCl and passage through a 1 ml Amersham Biosciences benzamidine column. Pure BipD was eluted and the thrombin eluted from the column with 20 mM TRIS containing 500 mM NaCl and 10 mM p-aminobenzamidine.

3.2.6 Column Purification (with no GST-tag)

3.2.6.1 Q-Sepharose Anion Exchange Column

The column was washed with 1 M KCL then equilibrated with 50 mM TRIS pH8. The post-dialysis BipD was diluted to 60 ml and loaded onto the column at 3 ml/min. It was washed to remove any non-specific binding then a KCl gradient 0 M – 0.8 M was set up over 950 ml to elute the protein, which was collected in 10 ml fractions. A gel was run to determine the BipD content and the appropriate fractions pooled. This was buffer exchanged and concentrated to 5 ml using a 10 kDa Vivaspin.

3.2.6.2 Superdex 200 Gel Filtration Column

The column was washed and equilibrated with 140 mM NaCl and 50 mM TRIS pH 8. A 1 ml protein sample was loaded onto the column at 0.5 ml/min and 2 ml fractions collected over an 8 hour time period. A gel was run to determine in which fractions the protein was present.

3.3 Crystallisation

The final yield of SeMet-protein was approximately 30 mg/l of culture. For crystallization this was concentrated to 6 mg/ml using a 10 kDa vivaspin ultracentrifugation spin column.

Crystallisation screens were set up by the hanging-drop method, based on conditions used previously to obtain crystals of the native protein. Within 3 weeks promising crystals were observed and following optimisation of the crystallisation conditions good quality crystals, up to 1mm in length, Figure 3.3, were obtained under these conditions; [SeMet-BipD] = 4 mg/ml, 20-25 % PEG 4000, 5 mM nickel chloride, 0.1 M glycine, 0.1 M sodium cacodylate (pH 5.0-5.5) (Erskine *et al.* 2006). After 4 weeks growth, crystals were picked, cryo-protected with 30 % (v/v) glycerol and mounted in mohair loops or lithoicloops (Molecular Dimensions) for freezing in liquid ethane and storage under liquid nitrogen.

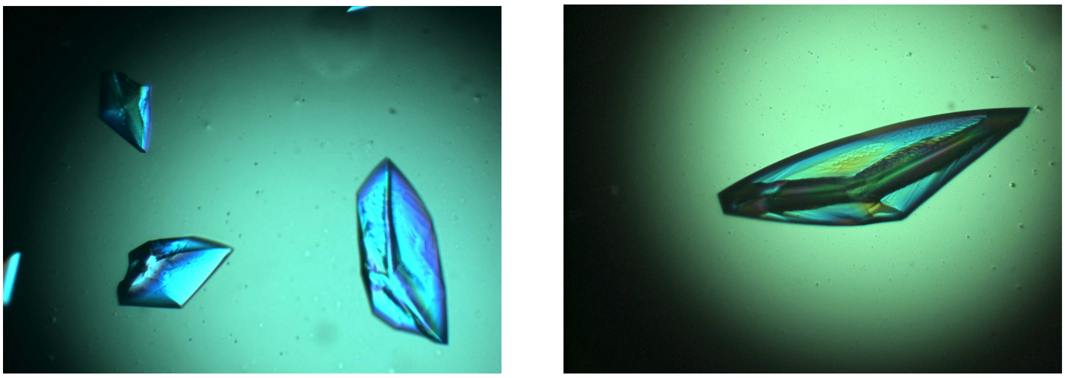


Figure 3.3: BipD protein crystals
Crystals were grown by the hanging drop method with a concentration of 4 mg/ml SeMet BipD in 20-25 % PEG 4000, 5 mM nickel chloride, 0.1 M glycine, 0.1 M sodium cacodylate (pH 5.0-5.5). The largest of these crystals measured 0.5mm in the longest dimension, and grew after 4 weeks.

3.4 Structure Determination

The structure determination of BipD was carried out by Dr Peter Erskine and Dr Jon Cooper.

3.4.1 Data Collection and Processing

The SeMet-BipD crystals were taken to the European Synchrotron Radiation Facility (ESRF) in Grenoble where they were found to diffract to a resolution of 2.1 Å using the ID23-1 beamline. A dataset, Figure 3.4, was collected at the selenium peak wavelength and processed using MOSFLM (Leslie 2006), SCALA (Evans 2006) and other programs in the CCP4 suite (CCP4 1994).

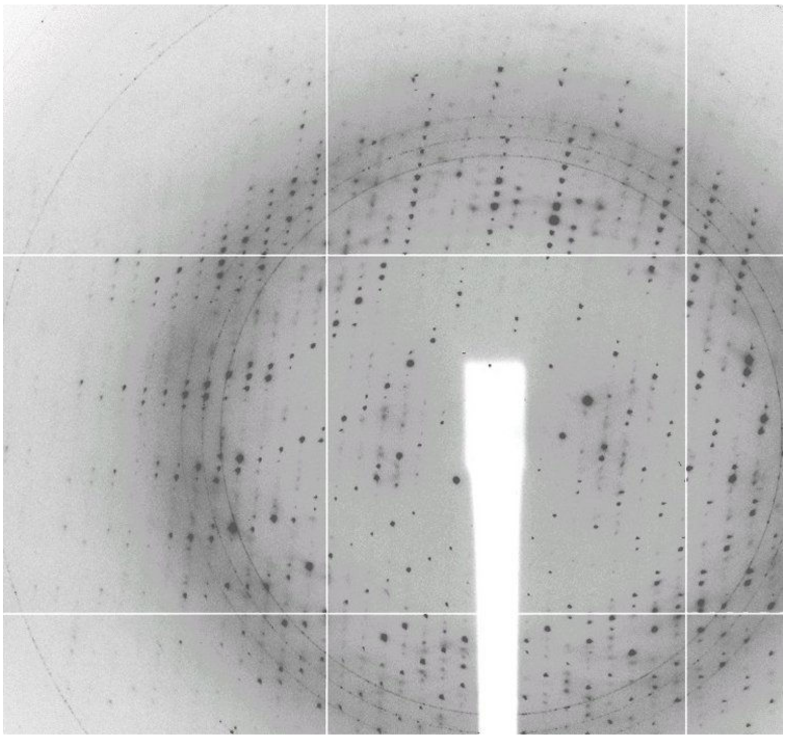


Figure 3.4: X-ray diffraction image for BipD
185 images were collected at 1° intervals at the ESRF. The crystal diffracted to 2.1 Å resolution.

This indicated the crystals belonged to the monoclinic space group $P2_1$ and had cell dimensions of a=53.5 Å, b=56.2 Å, c=84.2 Å, β= 94.5 Å. Use of the Matthews Coefficient program in CCP4 suggested 2 monomers of BipD were present per crystallographic asymmetric unit, with a solvent content of 35 %.

Dataset	Peak	Remote	Inflection	SAD 2.1 Å dataset
ESRF beam line	BM16	BM16	BM16	ID23-1
λ (Å)	0.9794	0.9077	0.9796	0.9794
Resolution (Å)	40.0-3.0 (3.2-3.0)	40.0-3.0 (3.2-3.0)	40.0-3.0 (3.2-3.0)	38.7-2.1 (2.2-2.1)
R_{merge} (%)	8.5 (16.2)	8.3 (17.8)	8.4 (21.2)	8.6 (41.1)
Completeness (%)	99.4 (100.0)	99.6 (100.0)	99.7 (100.0)	97.5 (93.4)
Average $I/\sigma(I)$	25.0 (11.3)	26.7 (12.1)	26.9 (11.1)	10.4 (2.8)
Multiplicity	9.9 (10.1)	9.8 (10.0)	9.8 (10.0)	3.7 (3.5)

Values for the outer resolution shell of each dataset are shown in parentheses
$R_{merge}=\Sigma h|(I_{hi}-I_h)|/\Sigma i\ (I_{hi})$ where I_h is the mean intensity of the scaled observations I_{hi}.

Table 2: Statistics for BipD datasets
A 2.1 Å SAD dataset was collected, but this was insufficient for structure determination, so a MAD dataset was collected from a different crystal on a different beamline.

Since attempts at SAD phasing using the above data were unsuccessful, multi-wavelength anomalous dispersion (MAD) X-ray data were collected from another crystal using beam line BM16. Data to a resolution of 3.0 Å were collected at the selenium fluorescence peak wavelength (0.9794 Å), followed by a remote dataset (0.9077 Å) and lastly an inflection point dataset (0.9796 Å). The data were processed as above and the relevant statistics are shown in Table 2.

3.4.2 Refinement and Model Building

Six out of the 12 expected selenium sites were located by use of the program SOLVE/RESOLVE (Terwilliger and Berendzen 1999), which produced an excellent electron density map as shown in Figure 3.5. Segments of polypeptide were fitted to the map using the automatic map-fitting program MAID (Levitt 2001) and were rebuilt using TURBO-FRODO (Biographics, Marseille) and WinCOOT (Emsley and Cowtan 2007). Thus one monomer with the correct sequence of BipD was built into the map and a copy of this model was then fitted by visual inspection to the electron density for the second subunit in the asymmetric unit.

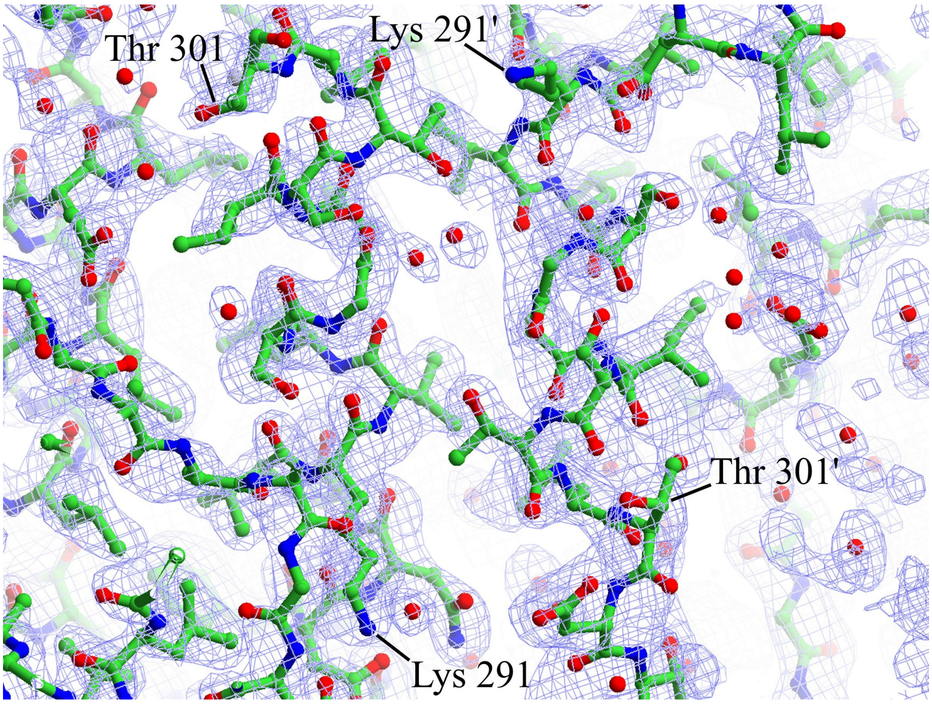


Figure 3.5: A section of the electron density map for part of helix 8
Approximately 80 % of the model was built using data from the crystal used in the MAD collection, which diffracted to 3 Å. This was then refined using the 2.1 Å dataset from the other SeMet-BipD crystal from which the SAD dataset was collected. The fit is very good in most regions, although it remains poor for some residues at the N and C termini and the loop region between residues 112 and 128, despite extensive efforts to rebuild them throughout the refinement process. This disorder in the loop region may however have functional significance.

A further 3 selenium sites were located from an anomalous difference Fourier and used in phasing. Model building was assisted by the use of the density modification program DM (Cowtan 1994) to extend the experimental MAD phases to higher resolution. The model was refined using the 2.1 Å resolution dataset from the other crystal by use of the programs SHELX-97 (Sheldrick and Schneider 1997) and CNS (Brunger *et al.* 1998), interspersed with rounds of manual rebuilding and solvent fitting. The R_{free} reflection set was chosen using SHELXPRO (Sheldrick and Schneider 1997) in thin resolution shells to avoid bias between the working and test sets due to the presence of non-crystallographic symmetry. Except for the initial round of refinement in which rigid body constraints were applied to both monomers in the asymmetric unit, non-crystallographic symmetry (NCS) restraints were not used.

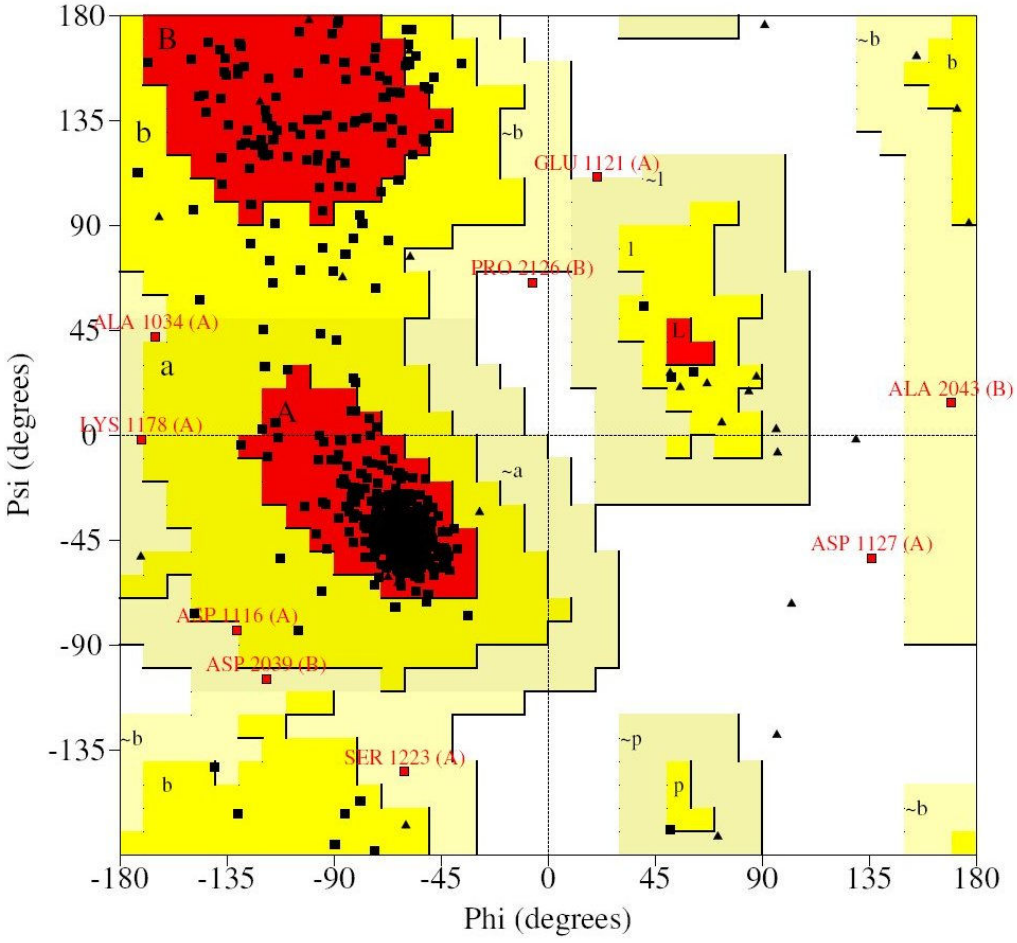


Figure 3.6: A Ramachandran plot for the BipD model
The final model has 88.8 % of its residues within the 'most favoured' regions of the Ramachandran plot, by PROCHECK criteria, and 99.6 % of residues lie within the so-called 'generously allowed' boundary. The remaining 0.4 % of residues (indicated by labelled red squares) are in parts of the molecule with poor electron density and hence their conformations are poorly defined. The black triangles indicate glycine residues.

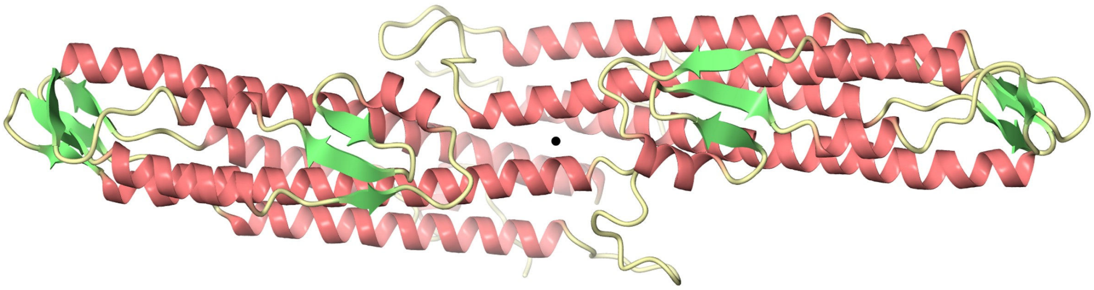


Figure 3.7: A dimer of BipD forming the crystallographic asymmetric unit
The central black dot indicates where both monomers associate with approximate 2-fold symmetry by contacts involving helices 4 (front) and 8 (back). It is yet to be determined whether this is a biologically functional structure.

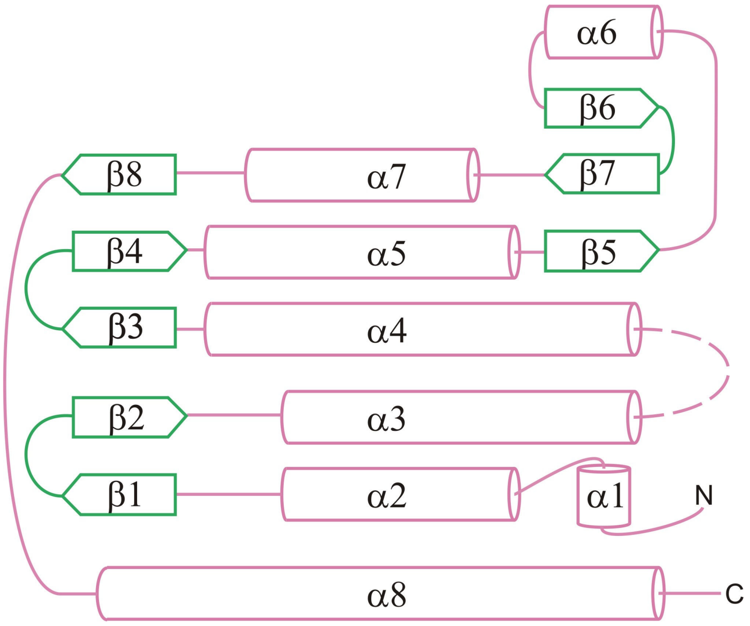


Figure 3.8: Schematic of the BipD protein molecule's topology
α-helical segments shown in pink and β-strand regions shown in green. The loop region linking helices α3 and α4 is shown dashed since this region is poorly defined by the electron density map.

The refinement statistics of the final model, Figure 3.7, are shown in Table 3. The Ramachandran plot for this model is shown in Figure 3.6.

R_{factor} (%)	22.6
R_{free} (%)	29.9
Number of reflections	28,488
RMSD bond lengths (Å)	0.004
RMSD 1-3 distances (Å)	0.017
RMSD bumps (Å)	0.014
RMSD chiral tetrahedral (Å)	0.024
RMSD planar groups (Å)	0.022
Mean B-factor for main chain atoms ($Å^2$)	43.0
Mean B-factor for side chain atoms ($Å^2$)	46.5
Mean B-factor for whole chain ($Å^2$)	44.7

Table 3: Refinement statistics for the BipD structure
All data between 10.0 Å and 2.1 Å resolution with no σ(I) cut-off were used in the refinement except for 5 % of the data which were reserved for the R-free set. Note that residues in the highly disordered loop between residues 112 and 128 were omitted from the average B-factor calculations.

3.5 Structure Analysis

3.5.1 General Structure Analysis

The final model was refined using data between 2.1 Å and 10.0 Å and has an R_{factor} of 22.6 % and an R_{free} of 29.9 % (Table 3). These slightly high figures may be attributed in part to the disorder of the loop region between residues 112 – 128 and at the N and C termini. The asymmetric unit contained 2 BipD molecules and 681 water molecules, as shown in Figure 3.7. By the PROCHECK criteria (Laskowski *et al.* 1993) 88.8 % of the residues in the final structure lie within the ‘most favoured’ regions of the Ramachandran plot and 99.6 % lie within the ‘generously allowed’ boundary. The remaining 0.4 % of residues are in parts of the molecule with poor electron density and therefore their conformations are poorly defined (Figure 3.6).

The overall structure, Figure 3.9, is appreciably extended and consists mainly of a bundle of antiparallel α-helical segments with two three-stranded β-sheets regions, one being at the end of the bundle and one being at the side.

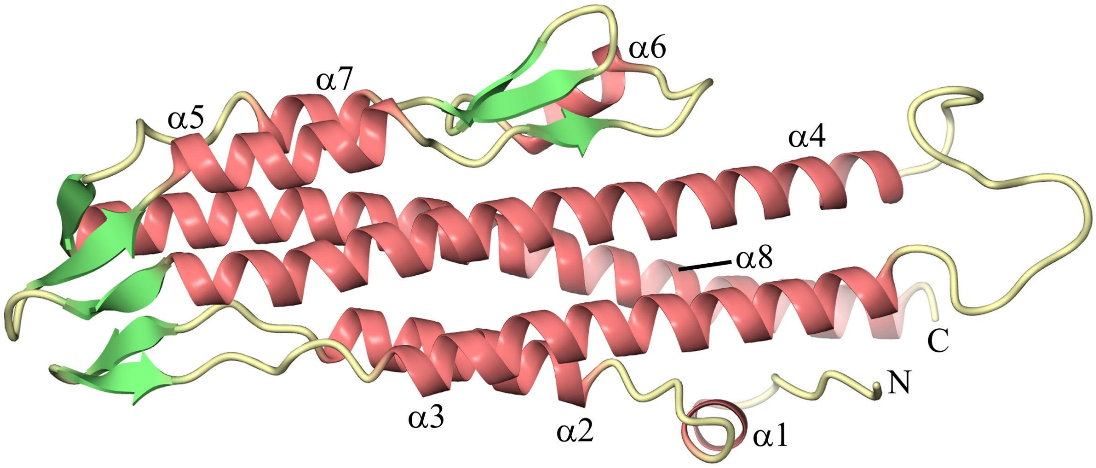


Figure 3.9: A ribbon model of the BipD structure
The longest helices (α2, α3, α4 and α8) form a 4-helix bundle and the remaining tertiary structure (3 alpha helices and 2 beta sheets) is formed by the region linking the last two helices of the 4-helix bundle.

This model is consistent with an earlier far-UV circular dichroism (CD) spectra for BipD, carried out on the protein before crystallisation and shown in Figure 3.10. Analysis of this spectra showed an estimated α-helix content of 66 % and 4 % β-sheet (X-ray analysis determined a content of 59 % α-helix and 4 % β-sheet).

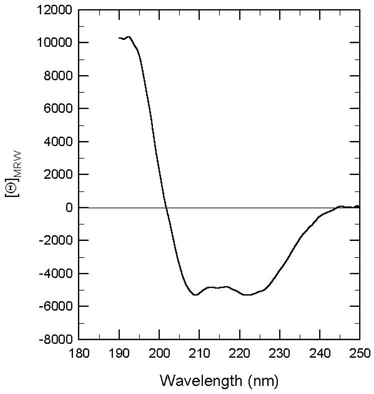


Figure 3.10: A CD spectrum for BipD
This indicates the predominance of α-helices in the structure of BipD. To record the spectrum, the protein was diluted to 0.15 mg/ml in 20 mM phosphate buffer (pH 7.6). This spectrum is in agreement with the crystallographic model.

3.5.2 Tertiary Structure

The overall topology of BipD is illustrated by the schematic diagram in Figure 3.8. The first 20-30 residues of the protein were not clearly defined in the electron density map. In one subunit of the asymmetric unit, residues 36-43 were clearly seen to form a short α-helix (helix 1) but in the other molecule this region is less well defined and adopts a slightly different conformation. This can be seen in Figure 3.11.

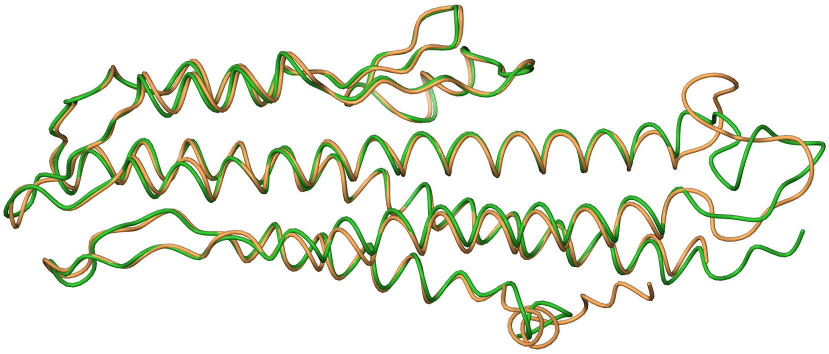


Figure 3.11: Superposition of both molecules of BipD in the asymmetric unit
The structurally conserved regions of both monomers both superimpose with an rms C^{α} deviation of 0.6 Å, emphasising the similarity of their conformations. The only regions where the structures diverge (N and C termini and the loop between helices 3 and 4) are poorly defined by the electron density map.

The polypeptide chain then goes on to a longer helical region (helix 2) from residue 47 - 63, which is followed by a slightly irregular β-hairpin involving residues 64 - 81 (β1 and β2 in Figure 3.8). The hairpin region is followed by another α-helix (helix 3) that extends from residue 82 – 111 and runs antiparallel to helix 2. The majority of the next 20 residues were found to be poorly defined in the electron density map, indicating a disordered region within the crystal, but which may be of functional biological significance.

Residue 128 is the N-terminal beginning of a very long α-helix (helix 4), which extends almost the full length of the protein molecule and ends with residue 170. It runs antiparallel to helix 3 and leads into a β-hairpin formed by residues 171 – 183 (β3 and β4 in Figure 3.9). This is followed by a short region of helix formed by residues 184 – 196 (helix 5) running antiparallel to helix 4.

Residues 197 – 203 form an extended region (β5) of which at least 3 residues take part in a three-stranded β-sheet involving a β-hairpin formed by residues 220 – 230 (β6 and β7 in Figure 3.9). There is a short region of helix formed by residues 209 – 216 (helix 6), which connects the first strand of the β5 sheet with the start of the hairpin region (β6 and β7). The topology of this region in the molecule has some parallels with a Greek key motif with one of the outer β-strands replaced by a helix (helix 6).

Helix 7 is formed by residues 233 – 241 and runs antiparallel to helix 5. It leads in to a short β-strand region formed by residues 246 – 250 (β8), which forms a three-stranded β-sheet with the β-hairpin formed by residues 171-183 (β3 and β4). The final α-helix (helix 8) extends for 50 residues from 251 – 301, running the whole length of the protein molecule. This helix is a highly conserved region of the protein family, with sequence identities of 60 % in IpaD and 52 % in SipD, suggesting that it plays an important role in the function of the type III secretion system.

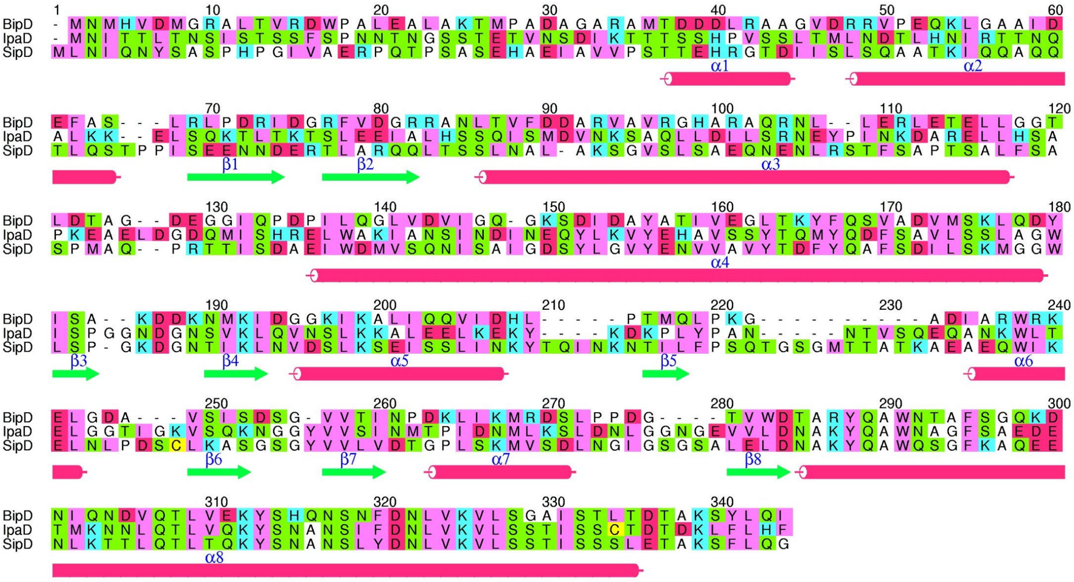


Figure 3.12: Sequence alignment of BipD with its homologues IpaD and SipD
The BipD sequence has 26 % identity and 36 % similarity with IpaD (from *S.flexneri*), and 27 % identity and 39 % similarity with SipD (from *S. typhimurium*). The amino acids are coloured according to the scheme: cyan – basic, red – acidic, green – neutral polar, pink – bulky hydrophobic, white – Gly, Ala and Pro, yellow – Cys. The secondary structure elements are indicated below the alignment and have been numbered according to the scheme in Figure 3.9.

3.5.3 Quaternary Structure

It was found that in the crystallographic asymmetric unit, two molecules of BipD associate to form a dimer via extensive contact at the C-terminal end of helix 8 and N-terminal end of helix 4 of both subunits. This is despite gel filtration of purified BipD (~3 mg/ml) on a calibrated Superdex-75 column at pH 7.5, which established that it has a V_e/V_o ratio of 1.17, which corresponds to monomeric protein.

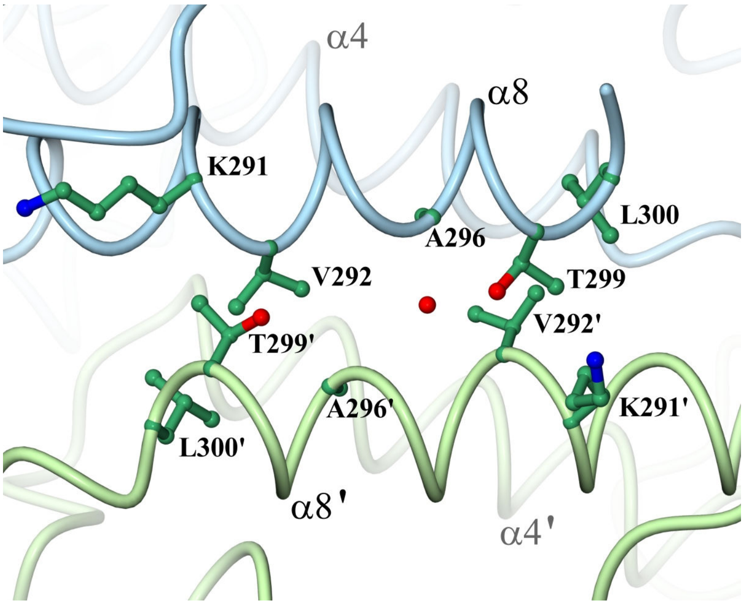


Figure 3.13: Specific contacts involved in dimer formation
Two monomers of BipD within the asymmetric unit associate to form a putative dimer. Here helix 8 (pale blue and green) can be seen paired together, with helix 4 in the background. The most extensive contacts are formed by helix 8 where residues at the C-terminal end are strongly conserved in BipD, IpaD and SipD, suggesting that the dimer may be biologically relevant.

The area of contact between the two monomers involves helix 8 pairing up with the same helix from the other monomer in an antiparallel manner, and the same for helix 4, Figure 3.13. Although gel filtration had indicated that BipD was monomeric in solution, the fact that these extensive contacts are being formed between helix 8, the most highly conserved region of the molecule, implies that the dimer may be biologically relevant, possibly forming at higher protein concentrations.

The residues of helix 8 involved in this contact region extend from Asn288 to Leu200, comprising approximately 3 turns of the helix, where 8 consecutive residues 287 - 294 (DNLVKVLS) are conserved between BipD, IpaD and SipD. This region forms an interface dominated by van der Waals contacts between aliphatic side-chains, although there are some water-mediated contacts too. These findings, combined with the known sensitivity of the protein to deletions in the region (Picking *et al.* 2005), suggest that the dimer observed in the X-ray structure may be a functionally important form of the molecule. However, the interface lacks specific direct interactions such as hydrogen bonds or salt-bridges between the monomers.

There are other stretches of conserved sequence in helix 8 that are not involved in dimer formation. It is therefore possible that these regions are involved in binding partner proteins in the translocon. In the dimer, the quaternary contacts involving helix 4 from each BipD monomer are of a similar nature (i.e. involve small hydrophobic or neutral polar side-chains) and extend over approximately four turns of the helix between Pro128 and Asp143. The sequence conservation in this region is poorer than for helix 8 and likewise the interface between helix 4 from each monomer lacks specific hydrogen bonds and salt-bridges.

3.6 Discussion

3.6.1 Movement of BipD Through the TTSS Needle

It is known that chaperones are required to assist the passage of proteins through the secretion needle, probably by partially or fully unfolding them so that they will fit through (Yip and Strynadka 2006). However, the elongated shape of the BipD molecule indicates that it may be able to pass though the pore of the needle with minimal unfolding.

The core 4-helix bundle of the molecule is formed by the longest helices 2, 3, 4 and 8. The region formed by residues 171 - 250 (i.e. helices 5, 6 and 7 and both of the 3-stranded β-sheets) is effectively an extensive elaboration of the loop linking helices 4 and 8. If the region formed by residues 171 – 250 were to unfold or swing away from the four-helix bundle, the diameter of the molecule would be reduced to approximately 25 Å, narrow enough to fit through the interior pore of the needle. Epitope mapping studies of IpaD (Turbyfill *et al.* 1998) indicate that residues in the 171 – 250 region of BipD could have considerable flexibility.

However, recent work carried out on the conformational stability and differential structural analysis of BipD and its homologues (Espina *et al.* 2007) has shown that BipD has only one thermal unfolding transition. It is also suggested that the intramolecular coil of BipD may be somewhat different from that of IpaD and SipD because it exhibits spectral characteristics between that of a double- and triple-stranded coil. This suggests the possibility that even if IpaD is able to unfold, BipD may not necessarily be able to. The N-terminal domain of BipD is much longer than seen in homologues, with more of a loop than a turn. The loop appears to form a stronger interaction with the coiled-coil, accounting for the apparent triple-stranded coiled-coil. Because the N-terminal domain is so much different to that of IpaD and SipD and has different structural characteristics, it may indicate a unique biological function for this region of the protein. Figure 3.14 (Johnson *et al.* 2006) shows a ribbon diagram of IpaD (green) superposed on *Bacillus subtilis* FliS (purple, 1VH6) demonstrating the opened four-helix bundle of this chaperone and suggesting a possible unfolded conformation for IpaD. As a homologue, it has been suggested that BipD may unfold in the same manner, although this has not been demonstrated.

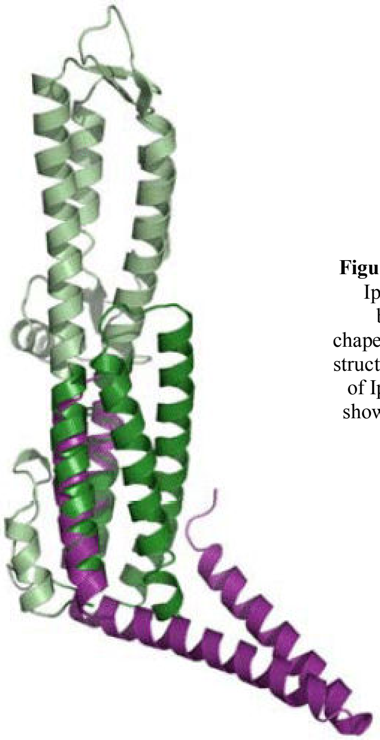


Figure 3.14: An unfolded chaperone superimposed on IpaD IpaD is shown in green, with the proposed unfolding helix bundle shown in dark green. A similar region from the chaperone of *Bacillus subtilis* FliS, has been superimposed on the structure in purple.It has been proposed that the four-helix bundle of IpaD and BipD are able to unfold in the same conformation shown for the chaperone of *Bacillus subtilis* FliS, allowing it to move through the pore of the secretion needle.

3.6.2 The Effect of pH Changes

It is known that the Ipa/Mxi/Spa TTSS of *Shigella* is necessary for lysis of the endosome membrane, once the bacterial cell has been engulfed by the host cell (High and Dobberstein 1992;Fernandez-Prada *et al.* 2000;Sansonetti 2001). The fact that *B. pseudomallei* contains a homologous TTSS implies that it may use a similar mechanism to escape from endosomes. Consistent with this idea, it has been reported that *B. pseudomallei* mutants lacking putative components of the Bsa secretion and translocation apparatus (BsaZ and BipD, respectively) are confined to endosomes with intact membranes following infection of J774.2 murine macrophage-like cells (Stevens *et al.* 2005).

It is known that some proteins, such as hemaglutinin undergo a pH-induced conformational change exposing a fusigenic peptide that facilitates entry of the influenza virus into the target cell (Bullough *et al.* 1994). The environment provided by the endocytic vesicle is around pH 5.0 and this is similar to that at which the BipD crystals grew (pH 5.8 following mixing of the protein in 50mM Tris pH 8.0 with the well solution of 0.1 M cacodylate pH 5.0 – 5.5), suggesting that the structure observed in our X-ray analysis may be very close to that which occurs in the endosome as the pH is lowered. IpaD was crystallised at pH 7.5, well above the acidic environment of the endosome. Alignment of the BipD structure (pH 5.8) and IpaD structure (pH 7.5) reveals that although the 2 proteins are structurally similar, there are marked difference, which could be accounted for by pH induced conformational changes, Figure 3.15.

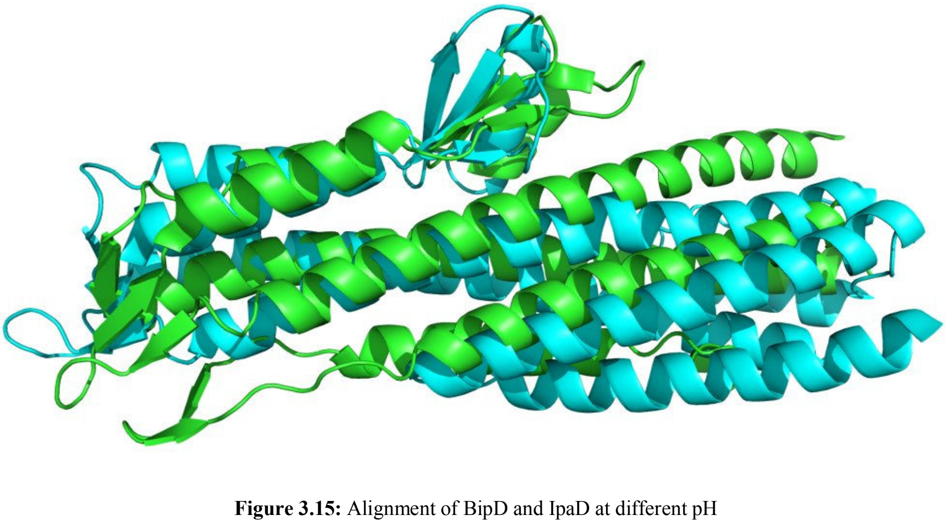


Figure 3.15: Alignment of BipD and IpaD at different pH
The structure for BipD (green), solved at pH 5.8 was aligned with IpaD (blue), solved at pH 7.5 and it can been seen that significant conformational difference exist between the two structures. Most noticeably, helix α4 of BipD is much straighter than that of IpaD, and therefore closer to the beta sheet around helix α6. This my play some role in it’s functionality.

Studies of Shigella effector proteins found that IpaD only associated with liposomes at high concentrations and that the affinity for liposomes was increased by lowering the pH below 5.0. Whilst pH-dependant denaturation of any soluble protein might be expected to expose hydrophobic regions and increase its binding to membranes, the fact that we know BipD is involved in lysing the endosome cell membrane implies that this may have some biological relevance.

All of the secondary structure elements in BipD have a high density of acidic and basic residues as shown in Figure 3.16, suggesting that as the pH is lowered in the endosome, protonation of carboxylate groups may cause repulsion of the positively charged basic groups and hence conformational change. This is especially true for the smallest exposed helices 1, 6 and 7, which have a very high abundance of acidic and basic residues.

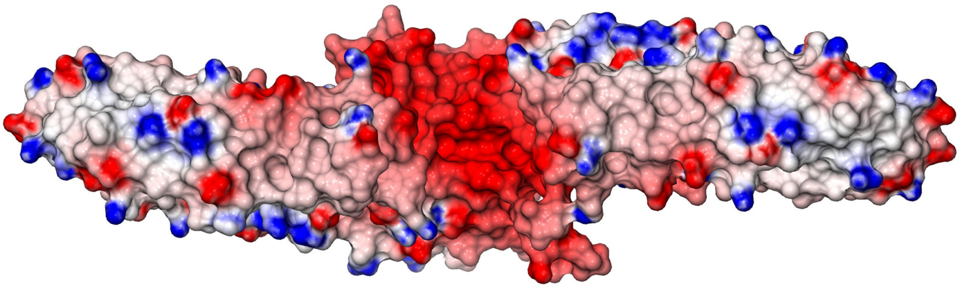


Figure 3.16: The electrostatic surface of the BipD dimer
The two monomers of BipD in the asymmetric unit associate with 2-fold symmetry contacts involving helices 4 and 8. The region of contact is highly negatively charged as shown by the large area of red.

Perhaps these regions are the most vulnerable to pH-induced unfolding in the endosome which could trigger association of the protein with the membrane. In contrast the longest and most conserved helix (helix 8) has by far the lowest density of acidic or basic residues.

3.6.3 BipD as a Translocon Plug

IpaD has been isolated from the tip of intact TTSS embedded in the bacterial membranes and on isolated needles even prior to secretion induction (Espina *et al.* 2006). Coupled with the fact that even short deletions (5 residues) in the C terminus of IpaD completely abolish the invasive phenotype and pore insertion (Picking *et al.* 2005), these observations led to suggestions that IpaD may play a role in regulating insertion of the IpaB/IpaC translocon pore from the tip of the *S.* flexneri needle (Blocker *et al.* 2003) and the secretion of effector proteins. Deletion of the first 20 residues of IpaD prevents its secretion, suggesting that these residues somehow act as a secretion signal (Picking *et al.* 2005). Unfortunately this region of BipD was not defined by the electron density map.

Removal of residues between 20 and 120 of IpaD reduced contact-mediated haemolysis of red blood cells but had no effect on invasion (Picking *et al.* 2005). Since deletions beyond residue 120 eliminate both haemolysis and invasiveness this suggested that the C-terminal region of the molecule is most important for its function as a molecular plug. All deletions (except for the first 20 residues) resulted in an increase in IpaD secretion. Deletion of just a few residues at the C-terminus also eliminated both haemolysis and invasiveness and also results in even greater levels of IpaD secretion suggesting that this region is also involved in transport regulation.

The equivalent part of BipD was not defined in the electron density map in spite of it being a strongly conserved region. It is possible that the extreme C-terminal region of the protein may only become ordered in the presence of the partner proteins BipB and BipC. There is evidence that IpaD regulates the relative rates of IpaB and IpaC transport and the reduction in haemolytic activity of IpaD mutants correlates with the reduced incorporation of IpaB and IpaC into erythrocyte membranes (Picking *et al.* 2005). Thus BipD is likely to play a similar role in regulating transport of BipB and BipC as well as their final integration into the target cell membrane.

3.7 Further Efforts at Structure Analysis

Based on the results of our structural investigation reported here, mutagenesis studies of BipD are required to define the roles of conserved and structurally important residues identified thus far. In addition, further investigation of the oligomerisation state of BipD is required, particularly as it appears to fulfil a variety of roles and may exist in different states depending on its current function. The proposed pH-dependent conformational changes in the wild-type and mutant BipD protein could be analysed by use of CD spectroscopy. In parallel with ultra-structural studies of the Burkholderia type III secretion system and other detailed studies of its protein components (e.g. the needle protein BsaL recently analysed by NMR (Zhang *et al.* 2006), the structure reported here will contribute towards a more detailed understanding of this remarkable molecular machine.

Chapter 4

Structural studies of 2,4’-Dihydroxyacetophenone dioxygenase from *Alcaligenes sp.*

4.1 Introduction

Dioxygenases catalyse the incorporation of both atoms of oxygen from O_2 into a single substrate (Hayaishi *et al.* 1955;Hayaishi *et al.* 1955), a reaction that frequently involves the cleavage of a C-C bond (Bugg and Lin 2001;Bugg and Lin 2001). 2,4'-dihydroxyacetophenone dioxygenase from *Alcaligenes* sp. is of particular interest because the way in which it catalyses the cleavage of the aromatic compound 2,4'-dihydroxyacetophenone to yield 4-hydroxybenzoate and formate is achieved in a unique manner, unlike any other known dioxygenase (Hopper and Kaderbhai 1999;Hopper and Kaderbhai 1999).

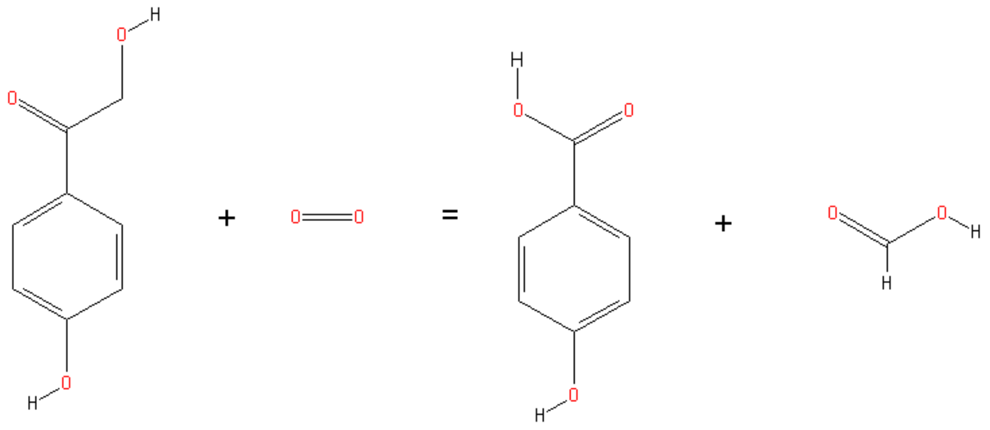


Figure 4.1: Reaction catalysed by 2,4'-dihydroxyacetophenone dioxygenase
2,4'-dihydroxyacetophenone dioxygenase acts by cleaving the side chain of 2,4'-dihydroxyacetophenone , yielding 4-hydroxybenzoate (left) and formate (right). This differs from other dioxygenases that cleave the aromatic ring itself.

Aromatic compounds are responsible for much of today's pollution and are commonly formed by industrial process during the incomplete burning of coal, oil, gas and garbage, as well as petroleum, solvents, pesticides and plastics. The majority of these chemicals are not naturally degraded and so persist in the environment for many years. The use of microorganisms to catabolise these compounds is known as bioremediation and is becoming a useful tool for removing hazardous chemicals from the land and water. By determining how the oxygenases and dioxygenases of microorganisms actually function to degrade aromatic compounds, it could then be possible to manipulate these enzymes to degrade compounds other than their natural substrates and remove aromatic pollutants from the environment.

The oxidative cleavage of an aromatic hydrocarbon ring is achieved by oxygenases; either by inserting the required oxygen atoms into the ring, thus preparing it for fission, or catalyzing the ring fission reaction themselves and cleaving the ring open. There are two classes of oxygenases, mono-oxygenases and dioxygenases, both of which can perform the above tasks but are differentiated by the number of oxygen atoms they insert into the aromatic hydrocarbon. Usually mono-oxygenases are involved in hydroxylation reactions; preparing the aromatic ring for cleavage by inserting a single atom of oxygen into the molecule. However, when hydrolyzing aromatic ketones they have been shown to insert a single atom of oxygen into the substrate to form an ester, which is then hydrolysed. The remaining atom from dioxygen is reduced to water (Tanner and Hopper 2000). Dioxygenases are involved in both the hydroxylation and fission reactions. The usual prerequisite for this oxidative cleavage is that the ring contains two hydroxyl groups, para or ortho to each other (Harwood and Parales 1996). Depending where the hydroxyl groups are positioned on the ring, the dioxygenase can utilize one or other of three different pathways (Harpel and Lipscomb 1990).

In order to engineer a microbe capable of degrading a foreign substrate of our choosing it is not possible simply to change a single enzyme. Molecular pathways are so intricately balanced and the enzymes involved so inter-dependant that in order to equip a microorganism with the ability to degrade an aromatic compound other than its natural substrate, rational and knowledgeable manipulation of both the host cell and the specific catabolic pathway involved is required (Diaz 2006). It is for this reason that it is not enough to know about individual enzymes in isolation, knowledge about the way in which they interact with their molecular environment is vital.

Through the use of techniques such as bacterial over-expression, purification and X-ray crystallography, it is hoped that it will be possible to elucidate the three dimensional structure of DAD. By achieving this it would then be possible to better understand the novel mechanism by which this dioxygenase functions and, it is hoped, a better understanding of microbial degradative pathways and detailed atomic knowledge of the enzymes involved. This will enable their use to greater effect to reduce the negative environmental impact of aromatic and for the production of chemicals by biotransformations.

4.2 DAD Expression and Purification Protocol

4.2.1 DNA Manipulation

During this project we had problems with the existing glycerol stocks of DAD in pET11a in BL21 (DE3) *E.coli* cells. This necessitated preparing the *dad* plasmid, which was then used to transform more BL21(DE3) *E.coli* cells using 0.5 µl of plasmid. This was carried out in the same manner as already described for BipD on pages 77-82.

4.2.2 Protein Expression

4.2.2.1 Native Protein Over-expression

The dad gene had been previously cloned into a pET11a expression vector in *E.coli* after efforts with the original pλ1-cyt vector proved unfavourable for the expression of a selenomethionine derivative. 15 µl of 2,4'-dihydroxyacetophenone dioxygenase BL21 *Escherichia coli* from a 15 % glycerol stock was spread on an LB-agar plate (yeast extract, NaCl, tryptone, bacterio-agar) containing ampicillin (50 µg/ml) and left to grow overnight. All bacterial growth was performed at 37 ºC unless otherwise stated. A single colony was then selected and used to inoculate 10 ml sterile LB media of the same antibiotic strength, which was grown to mid-log phase of OD_{600} 0.6, measured using a Pharmacia Biotech Ultrospec 3000 spectrophotometer. 5 ml of the culture was removed for later use as a starter culture and the remaining 5 ml induced with a final concentration of 0.5 M filter-sterilised Isopropyl β-D-1-thiogalactopyranoside (IPTG) and left to grow overnight.

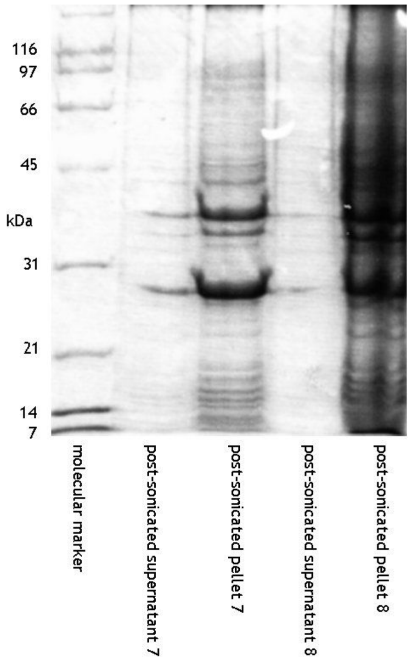


Figure 4.2: SDS gel showing sample protein contents
Cultures 7 and 8 were sonicated, as later described, and then centrifuged to separate the cell supernatant and pellet. SDS-gel analysis revealed the presence of DAD in both samples as being mainly in the cell pellet, implying that the dioxygenase is expressed as insoluble inclusion bodies.

A 300 µl sample of culture was then examined by 12.5 % SDS-PAGE gel electrophoresis at 25 mA to check the successful over-expression of DAD. Once this had been confirmed, thereby establishing the presence of the pET11a DAD construct, the 5 ml starter culture was used to inoculate 4 × 100 ml flasks of LB of the same antibiotic strength, which were induced with ITPG at OD_{600} 0.6. However, it was found that induction by this method resulted in 95 % of the DAD protein being expressed as insoluble inclusion bodies, shown in Figure 4.1. In an effort to avoid this, a heat shock method was used to grow the *E.coli*. 4 × 100 ml flasks of LB of the same antibiotic strength were grown to mid-log phase then grown for 25 minutes at 42 ºC in the hope that this would induce chaperone production, which would re-fold the denatured protein. The culture was then chilled in ice water for 5 minutes, induced with a final concentration of 0.3 mM IPTG and grown in an incubator at 16 ºC for 48 hours. However, this method too produced DAD mainly as inclusion bodies.

It was then decided to use Novagen Overnight™ Express Instant TB Medium to try to over-express DAD as soluble protein. A 1 L batch of Overnight™ Express Instant TB Medium was made by the addition of 1 L H_2O and 10 ml glycerol to the 60 g sachet. This was then autoclaved and made up to the same antibiotic strength. The media was induced with a 5 ml culture that had previously been checked for induction in the manner earlier described. Figure 4.3 shows that although this method still produces some inclusion bodies, there is a sizeable fraction of soluble protein to be worked with. The cells were harvested after 24 hours total induction time.

4.2.2.2 Seleno-methionine Protein Over-expression

To express a selemomethionine derivative of DAD, the Novagen Overnight™ Autoinduction System 2 was used, but this method produced only inclusion bodies. So instead we once again tried the heat-shock method of expression. A starter culture was grown, as previously described, and used to inoculate 1 L LB medium, which was grown at 37 ºC to mid-log phase. The cells were then centrifuged for 20 minutes and the pellet resuspended in 10 ml M9 minimal media and kept on ice for a maximum of 1 hour. A 1 L flask of M9 minimal media, containing 10 ml 40 % glucose solution and 1.5 ml ampicillin, was inoculated with the 10ml solution and incubated at 37 ºC for 45 minutes. Amino acids (with selenomethionine replacing methionine) and vitamins were added and the culture grown for a further 20 minutes (see appendix for methods). The flask was put in a shaking water bath at 42 ºC for 25 minutes, then stood in ice water for 5 minutes before being induced with a final concentration of 0.3 mM IPTG. The culture was grown at 16 ºC for 48 hours, before being harvested and purified as described for the native protein.

4.2.3 Harvest Cells

The cells were decanted into a 1 L Beckman centrifuge tube and harvested by centrifuging at 7,000 rpm for 20 minutes. The supernatant was discarded and the cell pellet resuspended in 20 ml buffer containing 50 mM TRIS at pH 8.

4.2.4 Sonication

The cells were disrupted by sonication for 2 repeats of 12 cycles, 30 seconds on, 90 seconds off, separated by a 10 minute cooling period on ice. This was centrifuged at 18,000 rpm for 60 minutes to remove cell debris. SDS gel analysis showed that although native DAD was still present as insoluble inclusion bodies in the pellet, this time it was also present in at least equal, if not greater quantities, of soluble protein in the supernatant (Figure 4.3).

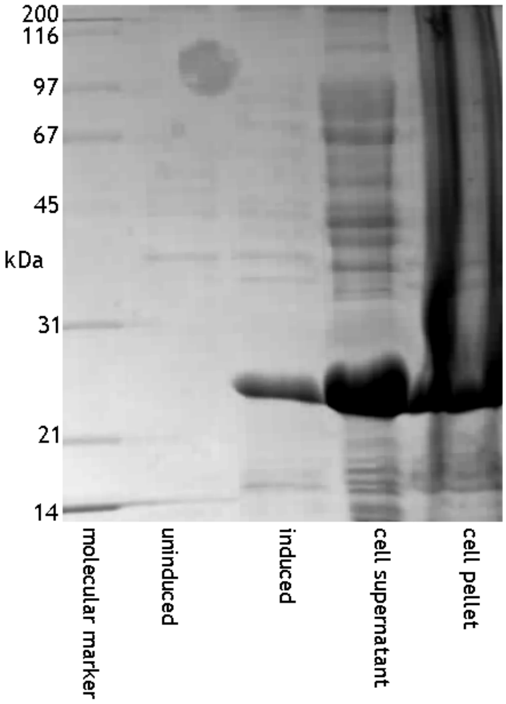


Figure 4.3: SDS gel showing soluble protein content of samples The test culture samples taken from, the Novagen Overnight™ Express Instant TB Medium culture were analysed by SDS-gel. The test culture grown in LB (uninduced) and induced with IPTG shows over-expression of DAD. The culture grown in the Instant TB medium shows massive over-expression of DAD in the pellet, probably present as insoluble inclusion bodies. However, there is an even greater quantity of DAD present as soluble protein in the cell supernatant. This soluble protein was then purified.

4.2.5 Crude Purification

4.2.5.1 Ammonium Sulphate Precipitation

Ammonium sulphate precipitation was carried out on the supernatant, initially at 30 % saturation (0.2 g/ml) to remove some of the impurities. 4 g ammonium sulphate was dissolved in the 20 ml sample and then centrifuged at 18,000 rpm for 40 minutes. The supernatant was removed and a sample taken for SDS analysis, then the pellet resuspended in 5 ml 50 mM TRIS for SDS analysis. Ammonium sulphate precipitation was then carried out at 45 % saturation by the addition of a further 2 g ammonium sulphate to the supernatant and centrifugation for 40 minutes at 18,000 rpm. Samples were taken and SDS analysis showed the presence of a very large amount of DAD in the 45 % pellet (Figure 4.4).

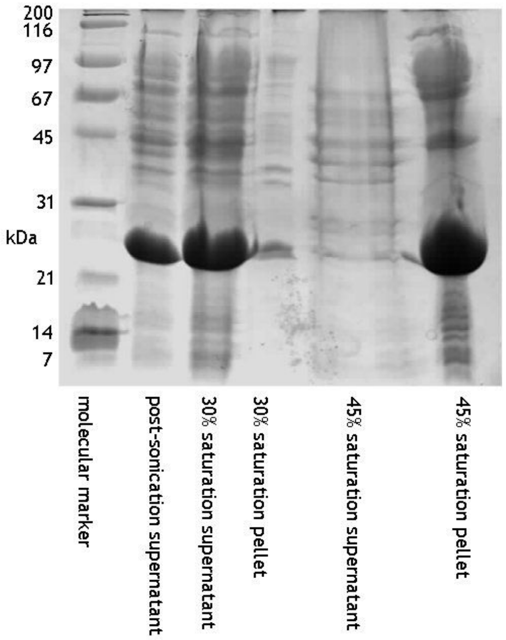


Figure 4.4: SDS gel showing the results of ASP Ammonium sulphate precipitation (ASP) at 30 % and 45 % was carried out on the cell supernatant containing soluble DAD to remove some of the impurities. It was found that DAD did not precipitate out at 30 %, but remained soluble in the supernatant. At 45 % DAD did precipitate out and was retained in the pellet.

4.2.5.2 Dialysis

The pellet was resuspended in 20 ml of 50 mM TRIS buffer. 1 m of dialysis tubing was boiled in H_2O containing sodium bicarbonate and EDTA for 10 minutes then washed in ddH_2O. The sample was dialysed in 2 L 50 mM TRIS pH 8 for 48 hours, and then centrifuged at 14,000 rpm for 20 minutes and the pellet discarded.

4.2.6 Column Purification

4.2.6.1 Q-Sepharose Anion Exchange Column

A Q-sepharose^{TM} anion exchange column was equilibrated with degassed 50 mM TRIS pH 8 until a stable base line (280 nm) was observed at 4 ºC. The sample was diluted to 70 ml in 50 mM TRIS and loaded onto the column at a flow rate of 1.5 ml/min (0.5 sensitivity) until all unbound protein had been washed off. A stable base line was then re-established. The bound proteins were then eluted on a salt gradient of 0 M – 0.8 M KCl over 950 ml at a flow rate of 1.5 ml/min and 10 ml fractions were collected. The peak fractions were analysed by SDS-PAGE and those containing DAD were pooled and concentrated using a 50 ml amicon with a cut-off membrane of 10 kDa. The concentrate was centrifuged at 4,000 rpm for 20 minutes and the pellet removed.

4.2.6.2 Superdex 200 Gel Filtration Column

A Superdex^{TM} 200 gel filtration column was washed with 1 M KCl then equilibrated with 50 mM TRIS, 200 mM KCl pH 8 until a stable base-line was observed at 4 ºC. Five separate 1 ml aliquots of protein solution were loaded onto the column at a flow rate of 1 ml/min (0.2 sensitivity) in five separate column runs. 4 ml fractions were collected and the peak fractions analysed by SDS-PAGE. Those fractions containing DAD were pooled and concentrated to 10 ml and centrifuged at 4,000 rpm for 20 minutes and the pellet discarded. Figure 4.5 shows the purity of DAD after anion exchange and gel filtration.

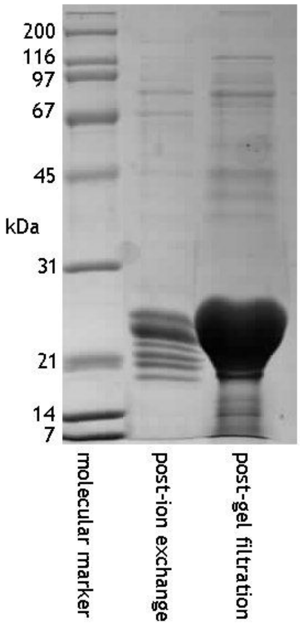


Figure 4.5: SDS gel showing protein content Even after the sample had been run on the Q-sepharose anion exchange column and Superdex 200 gel filtration column, contaminants were still present. To try to remove these, the sample was run on a Hi Trap Q XL anion exchange column.

4.2.6.3 Hi Trap Q XL Anion Exchange Column

In order to remove the remaining contaminants, a Hi trap Q XL anion exchange column was used. The column was equilibrated with 50 mM TRIS pH 8 and the 10 ml sample loaded at 0.5 ml/min (0.2 sensitivity). The protein was eluted in steps; 0.15 M KCl, 0.3 M KCl and 0.5 M KCl, and collected in 2 ml fractions.

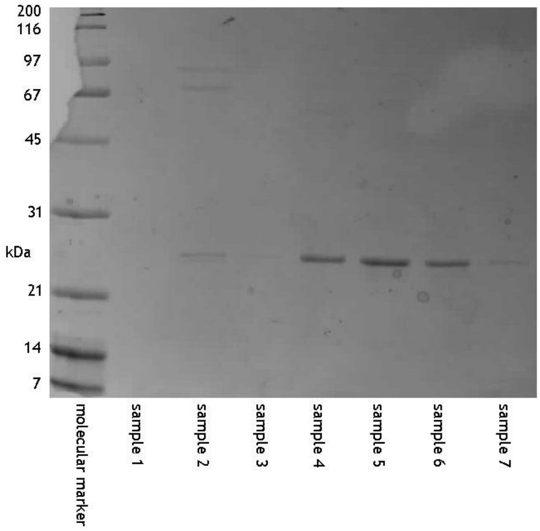


Figure 4.6: SDS gel showing the sample purity
After the DAD sample had run through the Hi-trap anion exchange column, most of the contaminants were removed. Fractions 4-6 were pooled and used in crystallisation screens.

It was found that DAD eluted at 0.3 M KCl (Figure 4.6). The fractions were pooled and concentrated using 10 kDa vivaspins and 10 kDa centricons to a volume of 600 µl. This was spectrophotometrically assayed to reveal a final concentration for the native protein of 5.8 mg/ml (absorbance at 10 times dilution = 1.342, extinction coefficient = 2.3). This was diluted by 2 to give 1.2 ml 2.9 mg/ml as it was felt that a protein concentration of around 3 mg/ml would be most suitable for crystallisation. The SeMet-DAD was concentrated to 4.3 mg/ ml for crystallisation.

4.3 Crystallisation

4.3.1 Native DAD Protein

A crystal screen was set up based around the condition used by G. Beaven with the freeze-dried protein provided by Dr D.J. Hopper of the University of Aberystwyth, although it did take 2 years for a crystal to grow. The mother liquor contained 0.7 ml 0.1 M TRIS/HCl, pH 8 and 14 % PEG 8000 and these conditions were varied:

pH

7.0 9.0

glycerol							PEG 8K
10%	50µl TRIS/HCL pH 7.0 100µl glycerol 200µl PEG 8K 650µl H_2O	100µl TRIS/HCL pH 7.0 100µl glycerol 200µl PEG 8K 600µl H_2O	50µl TRIS/HCL pH 9.0 100µl glycerol 200µl PEG 8K 650µl H2O	100µl TRIS/HCL pH 9.0 100µl glycerol 200µl PEG 8K 600µl H_2O	0.1M TRIS/HCl pH8.0 14% PEG 8K	0.1M TRIS/HCl pH8.0 14% PEG 8K	10%
	50µl TRIS/HCL pH 7.0 100µl glycerol 400µl PEG 8K 450µl H_2O	100µl TRIS/HCL pH 7.0 100µl glycerol 400µl PEG 8K 400µl H_2O	50µl TRIS/HCL pH 9.0 100µl glycerol 400µl PEG 8K 450µl H2O	100µl TRIS/HCL pH 9.0 100µl glycerol 400µl PEG 8K 400µl H2O	0.1M TRIS/HCl pH8.0 14% PEG 8K	0.1M TRIS/HCl pH8.0 14% PEG 8K	20%
0%	50µl TRIS/HCL pH 7.0 No glycerol 200µl PEG 8K 750µl H_2O	100µl TRIS/HCL pH 7.0 No glycerol 200µl PEG 8K 700µl H_2O	50µl TRIS/HCL pH 9.0 No glycerol 200µl PEG 8K 750µl H2O	100µl TRIS/HCL pH 9.0 No glycerol 200µl PEG 8K 700µl H2O	0.1M TRIS/HCl pH 7.8 14% PEG 8K	0.1M TRIS/HCl pH 8.2 14% PEG 8K	10%
	50µl TRIS/HCL pH 7.0 No glycerol 400µl PEG 8K 550µl H2O	100µl TRIS/HCL pH 7.0 No glycerol 400µl PEG 8K 500µl H_2O	50µl TRIS/HCL pH 9.0 No glycerol 400µl PEG 8K 550µl H2O	100µl TRIS/HCL pH 9.0 No glycerol 400µl PEG 8K 500µl H_2O	0.1M TRIS/HCl pH8.0 14% PEG 8K 1µ 100mM formate in drop	0.1M TRIS/HCl pH8.0 14% PEG 8K 1µ 100mM formate in drop	20%
	0.05M	0.1M	0.05M	0.1M			

TRIS/HCl

Crystallisation conditions were screened using both Molecular Dimensions screens MD 1+2 and Jena Biosciences JB screens 1-10. 24 well plates were set up using the hanging-drop vapour diffusion method at room temperature. A 1 ml reservoir was used, with a 5 µl drop containing 2.5 µl reservoir and 2.5 µl protein. Small, cubic crystals of native protein, approximately 15 µm in diameter, were observed in 5 different conditions after 10 months (Figure 4.7), suggesting the enzyme may have undergone a degree of proteolysis before crystallisation could occur. These conditions were found in screens JB1 (PEG 400 – 3000 based) and JB6 (ammonium sulphate based).

JB1	A3	15% w/v PEG 400, 200mM HEPES sodium salt pH 7.5, 200mM magnesium chloride	JB6	A1	500mM ammonium sulphate, 1.0M lithium sulphate, 100mM sodium citrate
	B3	30% w/v PEG 400, 100mM MES sodium salt pH 6.5, 100mM sodium acetate		B1	1.5M ammonium sulphate, 15% w/v glycerol, 100mM Tris-HCl pH 8.5
	B6	30% w/v PEG 400, 100mM Tris-HCl pH 8.5 200mM sodium citrate			

The crystals were mounted using Litholoops (Molecular Dimensions) and cryoprotected in the mother liquor with 10 % (v/v) glycerol. These were then frozen in a cryostream at 100 K and stored in liquid nitrogen.

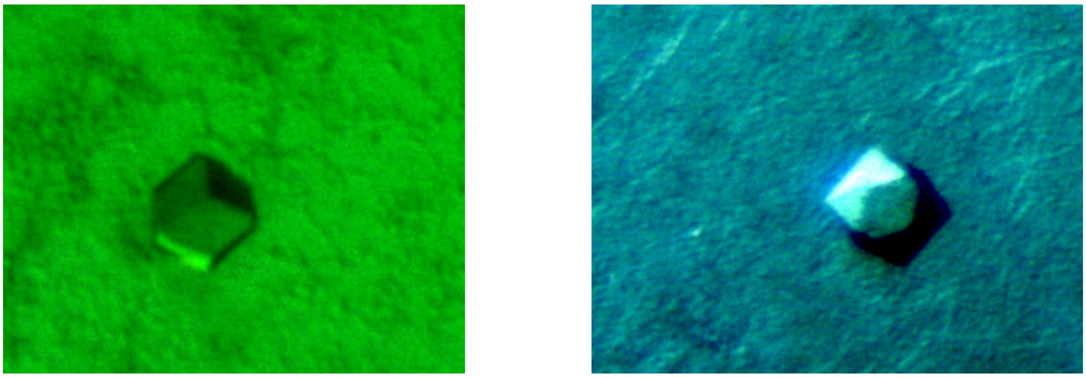


Figure 4.7: Crystals of DAD protein
These crystals of DAD protein grew under two different screen conditions (JB1 PEG 400 based, and JB6 Ammonium sulphate based) and appeared between months 10-12 after the hanging drops were set up. Both crystals were very small, not greater than 0.1mm in any dimension.

4.3.2 Selenomethionine DAD Derivative

It was found that the SeMet-DAD forms a precipitate in the native enzyme crystallisation conditions, so crystallisation trials have been set up in the same way as described above, using Molecular Dimensions screens MD 1+2 and Jena Biosciences JB screens 1-10, as well as the optimisation conditions being screened for the native protein. To date no crystals have been observed.

4.4 Preliminary X-ray Diffraction Analysis of Native DAD

The native crystals were taken to the ESRF (Grenoble, France) where the microfocus beamline ID23-2 was used to screen them. The only crystal to diffract to a reasonable resolution was that grown in condition JB1 B3, 30 % w/v PEG 400, 100 mM MES sodium salt pH 6.5, 100 mM sodium acetate. A dataset was collected to 3.1 Å resolution (Figure 4.8). 2 sets of 45 images were collected with 1° oscillations and an exposure time of 1 second per image.

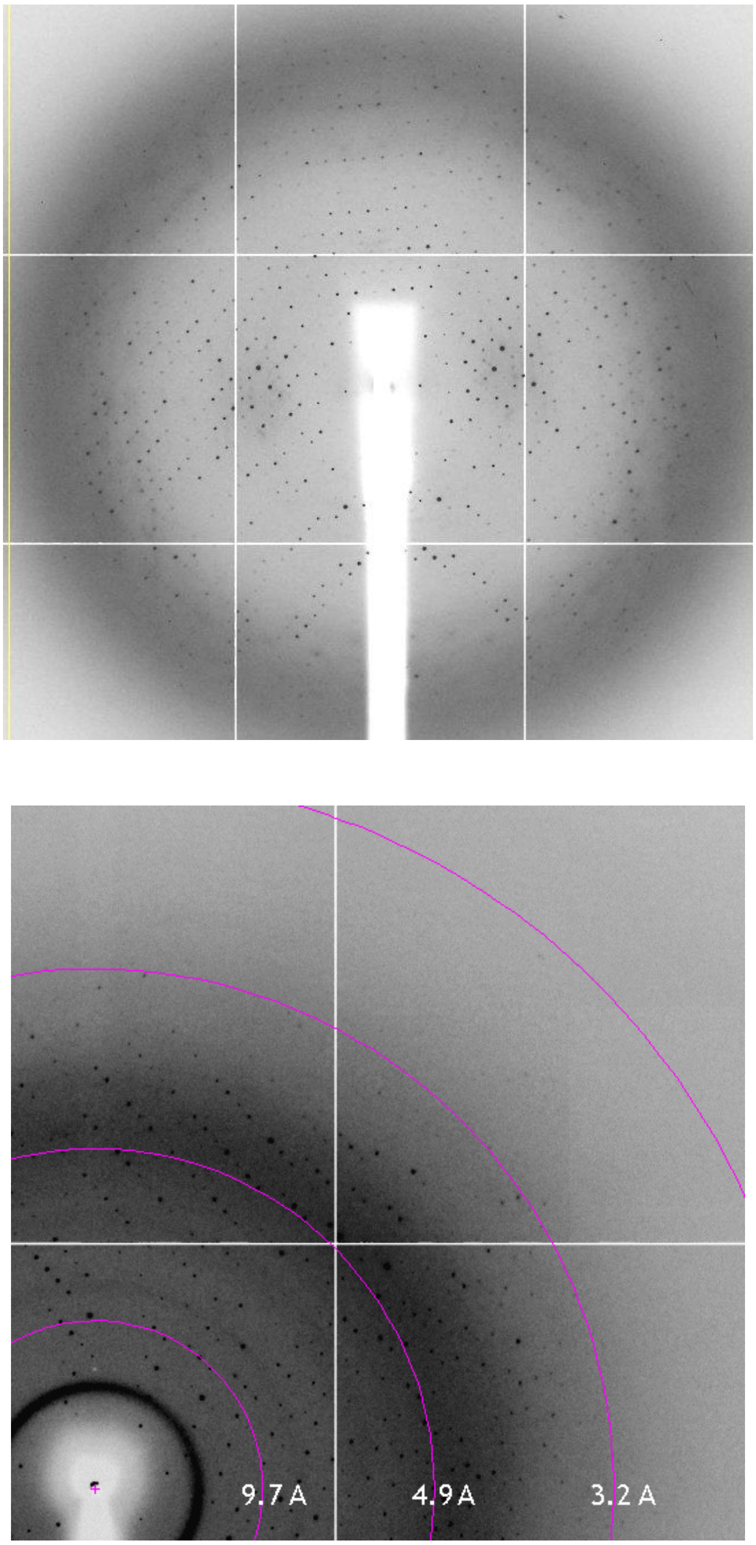


Figure 4.8: DAD diffraction data collected to 3.1 Å resolution
An image and one quadrant of a diffraction image obtained using beamline ID23-1 at ESRF (Grenoble, France) from a crystal of DAD, showing diffraction spots to 3.1 Å resolution.

Initial data processing, carried out using MOSFLM (Leslie 2006), SCALA (Evans 2006) and other programs in the CCP4 suite (CCP4 1994), indicated that the crystal belonged to the space group F23 and had unit cell parameters of a=b=c=180.1 Å, α=β=γ=90º (Table 4). The processed datasets extend to 3.1 Å resolution with an R_{merge} of 10.5 %, and completeness of 99.9 % and multiplicity of 5.3. The program Matthews_coef (CCP4 1994) was used to determine the likely number of molecules present in the asymmetric unit from which it was estimated that 2 monomers per asymmetric unit would give a solvent content of 59 %.

Beamline	ID23-2
Wavelength (Å)	0.976
Space group	*F*23
Unit Cell Parameters	
a=b=c (Å)	180.1
Resolution(Å)	104.3-3.1 (3.3-3.1)
R_{merge}^{*} (%)	10.5 (55.8)
Completeness (%)	99.9 (100.0)
Average I/σ	13.0 (3.0)
Multiplicity	5.3 (5.5)
Number of observed reflections	47183 (7078)
Number of unique reflections	8894 (1297)
Wilson's B factor (Å^2)	68.3
Matthew's coefficient	3.00
Solvent content (%)	59.00
Number of molecules in the asymmetric unit	2

Table 4: Data collection and processing statistics for native DAD. Values in parenthesis are for the outer resolution shell.

$* R_{merge} = \Sigma_h \Sigma_i \, |(I_{hi} - I_h)| \, / \, \Sigma_h \Sigma_i \, (I_{hi})$
where I_h is the mean intensity of the scaled observations I_{hi}.

Attempts were made to solve the structure using the program ACORN (CCP4 1994) and a 1.1 Å dataset collected from the freeze-dried DAD crystal, however this was unsuccessful. MrBUMP (CCP4 1994) was also used to search for a homologue and perform molecular replacement. The homologue found was a novel manganese-containing cupin from *Thermotoga maritime* (PDB code 1VJ2), but this only displayed 29 % sequence homology with DAD and is about 40 residues shorter, therefore no processable solution could be found. Because no homologous structure exists, molecular replacement cannot be used for phasing of the data, therefore structure determination will proceed when a SeMet-DAD crystal can be obtained and a MAD dataset collected.

Chapter 5

Structural studies of L-threonine Dehydrogenase from *Thermococcus kodakaraensis*

5.1 Introduction

The anaerobic hyperthermophilic archaeon *Thermococcus kodakaraensis* (KOD1) was isolated from the sediment and water of a solfatara, at over 100 ºC, on the shores of Kodakara Island off the coast of Kagoshima, Japan. The entire 2,088,737 bp genome of *T. kodakaraensis* has been sequenced and analysed (Fukui *et al.* 2005), making it one of the best characterized hyperthermophilic organisms. An orthologue search was subsequently carried out on the genome, highlighting a DNA sequence corresponding to that of L-threonine dehydrogenase (TDH; EC 1.1.1.103). TDH has been identified in all three domains of life i.e. eucarya, bacteria and archaea, from organisms such as a psychrophilic bacterium *Cytophaga sp.* KUC-1 (Kazuoka *et al.* 2003), a hyperthermophilic archaeon *Pyrococcus horikoshii* (Machielsen and van der Oost 2006), *E.coli* K-12 (Boylan and Dekker 1981), humans (Edgar 2002), chicken liver mitochondria (Aoyama and Motokawa 1981) and goat liver mitochondria (Ray and Ray 1984). The enzymes from *E.coli* and *P. horikoshii* (PhTDH) have been extensively characterised, although thus far *Pyrococcus horikoshii* and *Flavobacterium frigidimaris* are the only species from which the L-threonine dehydrogenase enzyme had been crystallised and the three-dimensional atomic structure determined (PDB codes 2DFV and 2YY7 respectively) (Ishikawa *et al.* 2007). Although TDH is not restricted to liver mitochondria, in higher organisms it is most abundant in this tissue type (Aoyama and Motokawa 1981).

L-threonine dehydrogenase belongs to the family of oxidoreductases, specifically those acting on the >CHOH group of a donor molecule with NAD^+ or $NADP^+$ as the acceptor (EC 1.1). It is to the medium-chain family that the 350 amino acid long TkTDH belongs. Typically enzymes in this family contain a catalytic zinc ion and structural zinc ion. L-threonine dehydrogenase catalyses the NAD^+-dependent oxidation of L-threonine to L-2-amino-3-ketobutyrate, an intermediate in the metabolism of threonine to glycine (Potter *et al.* 1977), both of which are essential amino acids. The dehydrogenase catalyses the first step in the threonine degradation pathway in several microorganisms that use L-threonine as the main carbon and energy source. The product 2-amino-3-ketobutyrate is further cleaved in a CoA-dependent reaction, catalysed by 2-amino 3-ketobutyrate CoA ligase, to produce glycine and acetyl-coA (Aoyama and Motokawa 1981).

Within the medium-chain dehydrogenase family there exists a high degree of structural similarity between the enzymes, despite their low sequence similarity. However, there are differences between the enzymes from higher plants and animals, which are typically dimeric for example that of horse liver alcohol dehydrogenase (ADH), and those of lower organisms like yeast, bacteria and archaea, which are tetrameric such as *S. cerevisiae* ADH.

Since their initial discovery in 1972 (Unsworth *et al.* 2007), many hyperthermophilic organisms have been identified and the rate of discovery is increasing as recent developments in technologies enhance our ability to access the extreme environments they inhabit. These environments, which until more recently were considered as being inhospitable to life, include volcanic areas rich in sulphur and 'toxic' metals and hydrothermal vents in the deep sea at extremely high pressures. To date the highest temperature a hyperthermophile has been observed to grow at is 121 ºC; this is a microorganism, designated strain 121 that was isolated from an active "black smoker" in the Northeast Pacific Ocean (Kashefi and Lovley 2003). These developments are important because it is still yet to be understood what the underlying basis is for the extraordinary thermostability of these hyperthermophilic enzymes. Should this be discovered then a great number of opportunities to make strategic use of their thermoactivity and thermostability present themselves. (Adams and Kelly 1998).

Hyperthermophilic enzymes are very similar to their mesophilic homologues, and this is supported by sequence alignments, amino acid content comparisons, crystal structure comparisons, and mutagenesis experiments (Vieille and Zeikus 2001). One hypothesis used to explain the difference in structural stability at high temperatures is that the structures of hyperthermophilic enzymes are more rigid than their mesophilic counterparts at moderate temperatures and therefore rigidity is a prerequisite for high thermostability.

It is hoped that by elucidating the three-dimensional structures of hyperthermophilic enzymes it will be possible to determine the basis of their inherent properties of thermostability and optimal activity at high temperatures and the underlying atomic factors responsible. These enzymes actually share the same catalytic mechanisms with their mesophilic homologues and when cloned and expressed into such hosts, hyperthermophilic enzymes usually retain their thermal properties, implying that these are genetically encoded. Studies of primary sequences have failed to identify repeatable differences between hyperthermophilic and mesophilic enzymes; therefore it appears likely that no one mechanism is responsible for their thermostability. Instead it is likely that a small number of highly specific alterations, which often do not obey any obvious rules, may be the source of this thermostability. Thus a concerted action of structural, dynamic and other physicochemical attributes are utilised to ensure the delicate balance between stability and functionality of proteins at high temperatures. These may include, but are certainly not limited to; charge clusters, networks of hydrogen bonds, optimization of packing, salt bridges, hydrophobic interactions, surface loop stabilization and reduction of the entropy of unfolding, each contributing to the overall thermostability (Unsworth *et al.* 2007). Identifying these minor intrinsic differences that allow hyperthermophilic enzymes to maintain structural stability and activity at elevated temperatures is a major challenge, but one that can only be aided by the elucidation of the proteins' tertiary and quaternary structures.

The crystal structure of TkTDH from *Thermococcus kodakaraensis* has been solved at a resolution of 2.4 Å revealing the detailed tertiary fold and quaternary structure of the molecule. Analysis suggests that the biologically active form of the molecule is a tetramer, in agreement with ADHs from lower life forms. It was crystallised bound to the co-factor NAD^+, identifying the active site and suggesting a possible mechanism of catalysis. The structure has been used to model the putative complex formed with 2-amino 3-ketobutyrate CoA ligase (KBL), suggesting that 2 molecules of this dimeric protein bind one TkTDH tetramer.

5.2 TDH Expression and Purification Protocol

The DNA cloning, protein expression and purification of TkTDH was carried out by Q. Bashir and colleagues at the University of Punjab, Pakistan (Bowyer *et al.* 2008).

5.2.1 Cloning of TkTDH

The Tk-TDH gene (TK0916) from the genomic DNA of *T. kodakaraensis* was amplified by PCR and ligated into the pTZ57R/T cloning vector using T4 DNA ligase (*Fermentas*). The recombinant pTZ-tdh plasmid was digested with Nco1 and EcoR1 and the Tk-TDH gene ligated with pET-8c, giving the recombinant pET-tdh plasmid that was then used to transform *E.coli* DH5α competent cells. Recombinant pET-tdh plasmid was then purified from *E. coli* DH5α and transformed into the expression strain *E.coli* BL21 (DE3).

5.2.2 Expression and Purification

Expression of the un-tagged, full-length TkTDH protein was undertaken on the 1 litre scale in LB medium and, following growth of cultures to mid-log phase, expression of the gene was induced with 0.2 mM IPTG at 37 °C for a further 4 hours. Cells were harvested by centrifugation at 6,000 x g for 10 minutes at 4 °C and washed with 50 mM Tris-HCl (pH 8). The cell pellet was resuspended in the same buffer and the protein was purified by a series of steps involving lysis of the cells by sonication, removal of the insoluble material by centrifugation (15,000 x g for 30 minutes at 4 °C) and heating of the soluble fraction to 85 ºC for 20 minutes to precipitate further impurities which were removed by centrifugation at 15,000 x g for 30 minutes at a temperature of 4 °C. This was followed by anion-exchange chromatography using a Resource Q column (Amersham Biosciences, United Kingdom) which was equilibrated with 50 mM Tris-HCl buffer at pH 8. Proteins were eluted with a linear gradient of 0 to 1.0 M sodium chloride in 50 mM Tris-HCl buffer at pH 8. Fractions with TDH activity were dialyzed against 50 mM Tris-HCl buffer pH 8 containing 2 M ammonium sulphate and applied to a hydrophobic interaction chromatography column (Resource ISO, Amersham Biosciences, United Kingdom) equilibrated with 50 mM Tris-HCl buffer at pH 8 containing 2 M ammonium sulphate. The bound proteins were eluted with a linear gradient from 2.0 to 0 M ammonium sulphate in 50 mM Tris-HCl buffer at pH 8.

Purified enzyme was stored in 50 mM Tris-HCl at pH 8 and was found to have a subunit molecular mass of 38,016 Da by mass spectrometry. It was found to be tetrameric by gel filtration, thereby making it similar to the TDHs from *E.coli*, *P.horikoshii*, *P.furiosus* and *Cytophaga sp.*

5.3 Crystallisation

Many crystallisation and optimisation trials were set up before suitable conditions were found. Crystallisation trials were set up at room temperature using the hanging-drop vapour-diffusion method on 24 well plates, and conditions were initially screened using both Molecular Dimensions screens MD 1+2 and Jena Biosciences JB screens 1-10. A 1ml reservoir was used, with a 5µl drop containing 2.5 µl reservoir and 2.5 µl protein. Crystals of TkTDH grew in 0.1 M sodium citrate pH 5.6, 2 M ammonium sulphate and 0.2 M sodium / potassium tartrate (screen MD2 well 29). Three rounds of optimisation screens were set up to refine the crystallisation conditions, as frequently it was found that fractured crystals formed as illustrated by Figure 5.1, leading to twinned data that proved too difficult to process.

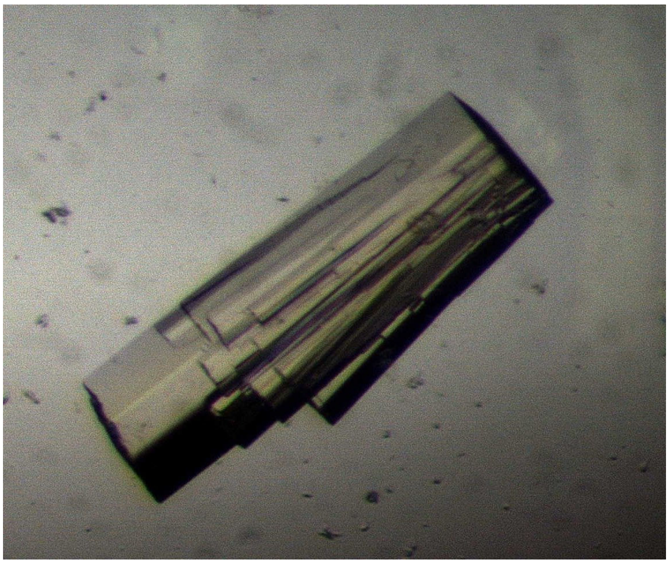


Figure 5.1: Multiple TDH crystal
Initially the TDH crystals grown by the hanging drop method were very fractured or grew as multiple crystals in the absence of glycerol, regardless of the crystallisation precipitant and additives.

Several optimisation screens were set up in the presence of 5 % (v/v) glycerol and it was found that, in general, crystals from these trays were less fractured in appearance and grew as single crystals.

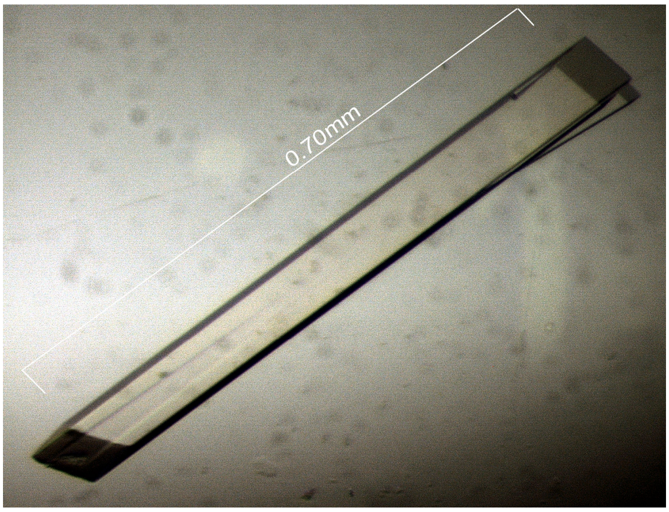


Figure 5.2: TDH crystal
The addition of 5-10 % glycerol to the TDH crystallisation conditions helped the crystals grow as less fractured, single crystals, enabling the collection of a good quality data set.

It was expected that TkTDH requires Zn^{2+} ions for structural reasons, and possibly for catalytic purposes, based on our knowledge of alcohol dehydrogenase structures (Guy et al. 2003). As it was unclear whether zinc would have remained bound to TkTDH throughout the purification, zinc chloride was added to some crystallisation trays to a final concentration of 0.1 mM. Single crystals of TkTDH eventually grew in 0.05 M sodium citrate pH 5.6, 2.4 M ammonium sulphate, 0.1 M sodium / potassium tartrate, 5 % (v/v) glycerol and 0.1 mM zinc chloride (Figure 5.2). In all cases TkTDH at a concentration of 3.4 mg/ml was used.

5.4 Structure Determination

5.4.1 Data Collection and Processing

The crystals were mounted using Litholoops (Molecular Dimensions) and cryoprotected in the mother liquor with 10 % (v/v) glycerol. These were then frozen in a cryostream at 100 K and stored in liquid nitrogen. The crystals were taken to the European Radiation Synchrotron Facility (ESRF) in Grenoble, France, where a dataset was collected to 2.4 Å resolution (Figure 5.3) using beamline ID14-3. 190 images were collected with 1° oscillations and an exposure time of 30 seconds per image with 3 passes per image.

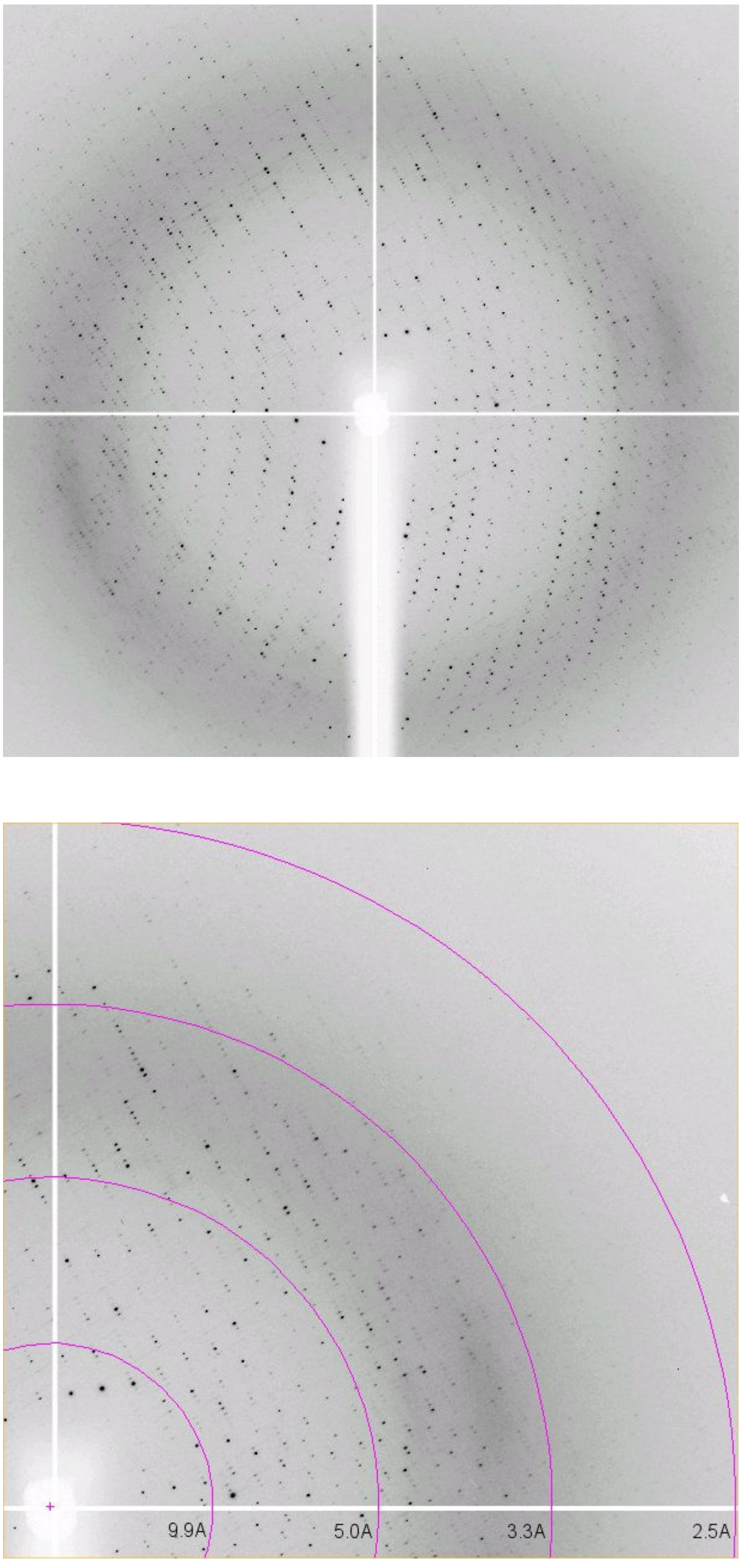
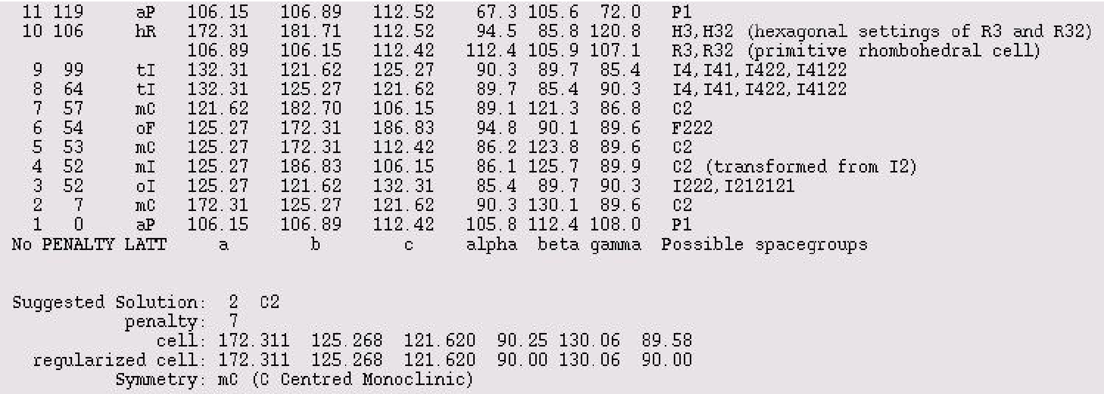


Figure 5.3: TDH diffraction data collected to 2.4 Å
An image and one quadrant of a diffraction image obtained using beamline ID14-3 at ESRF (Grenoble, France) from a crystal of TDH, showing diffraction spots to 2.4 Å resolution.

The data were processed using MOSFLM (Leslie 2006;CCP4 1994), SCALA (Evans 2006) and other programs in the CCP4 suite (CCP4 1994). However, the determination of the space group was ambiguous and varied depending on the number and angular range of the images chosen for autoindexing. When using images 1 and 90, the preferred space group prediction was C2.



```
 11 119     aP   106.15   106.89   112.52    67.3 105.6  72.0   P1
 10 106     hR   172.31   181.71   112.52    94.5  85.8 120.8   H3,H32 (hexagonal settings of R3 and R32)
                 106.89   106.15   112.42   112.4 105.9 107.1   R3,R32 (primitive rhombohedral cell)
  9  99     tI   132.31   121.62   125.27    90.3  89.7  85.4   I4,I41,I422,I4122
  8  64     tI   132.31   125.27   121.62    89.7  85.4  90.3   I4,I41,I422,I4122
  7  57     mC   121.62   182.70   106.15    89.1 121.3  86.8   C2
  6  54     oF   125.27   172.31   186.83    94.8  90.1  89.6   F222
  5  53     mC   125.27   172.31   112.42    86.2 123.8  89.6   C2
  4  52     mI   125.27   186.83   106.15    86.1 125.7  89.9   C2 (transformed from I2)
  3  52     oI   125.27   121.62   132.31    85.4  89.7  90.3   I222,I212121
  2   7     mC   172.31   125.27   121.62    90.3 130.1  89.6   C2
  1   0     aP   106.15   106.89   112.42   105.8 112.4 108.0   P1
No PENALTY LATT     a        b        c     alpha  beta gamma  Possible spacegroups


Suggested Solution:  2  C2
           penalty:  7
              cell: 172.311  125.268  121.620  90.25 130.06  89.58
  regularized cell: 172.311  125.268  121.620  90.00 130.06  90.00
          Symmetry: mC (C Centred Monoclinic)
```

Figure 5.4: MOSFLM C2 space group prediction
Initial prediction for the TDH crystal using diffraction images at 1° and 90° indicated C2 as the most probably space group.

It was attempted to process the data in C2, then later I222, but in both cases MOSFLM was unable to predict the spots correctly.

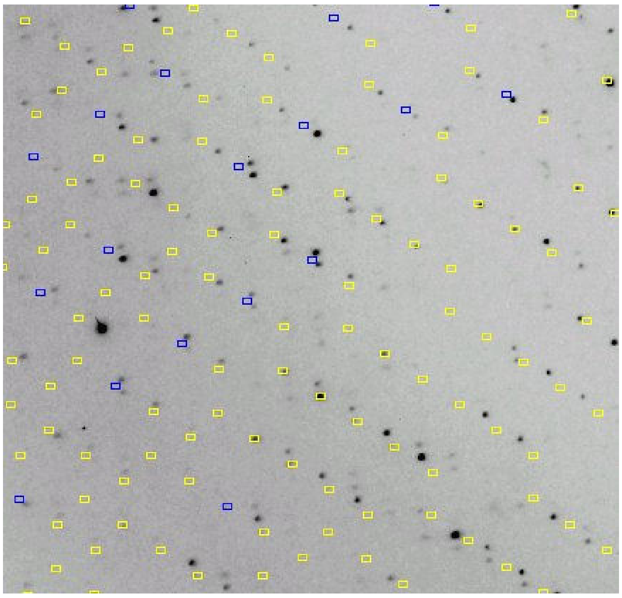
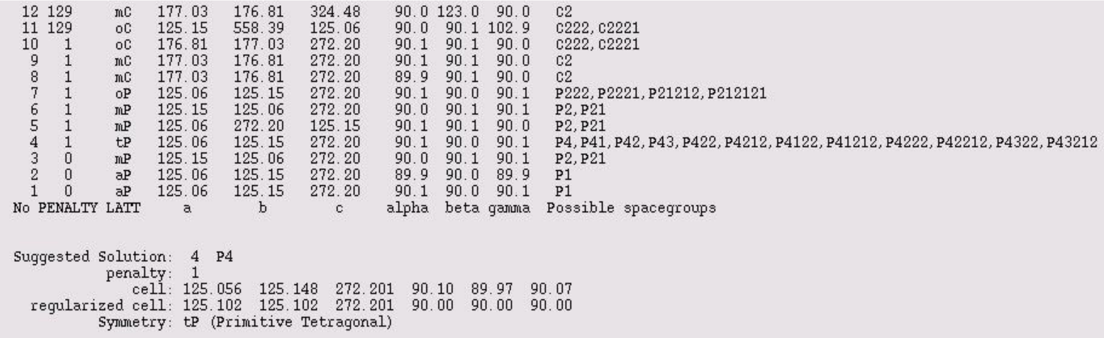


Figure 5.5: Section of MOSFLM spot prediction in C2 space group
As can be seen from the image, in the C2 space group MOSFLM is unable to correctly predict spots. The coloured squares should align with the diffraction spots, but this is clearly not the case in this example.

However, when indexing using images 1, 45 and 90, the preferred space group prediction was P4.



```
  12 129      mC    177.03    176.81    324.48     90.0 123.0  90.0    C2
  11 129      oC    125.15    558.39    125.06     90.0  90.1 102.9    C222,C2221
  10   1      oC    176.81    177.03    272.20     90.1  90.1  90.0    C222,C2221
   9   1      mC    177.03    176.81    272.20     90.1  90.1  90.0    C2
   8   1      mC    177.03    176.81    272.20     89.9  90.1  90.0    C2
   7   1      oP    125.06    125.15    272.20     90.1  90.0  90.1    P222,P2221,P21212,P212121
   6   1      mP    125.15    125.06    272.20     90.0  90.1  90.1    P2,P21
   5   1      mP    125.06    272.20    125.15     90.1  90.1  90.0    P2,P21
   4   1      tP    125.06    125.15    272.20     90.1  90.0  90.1    P4,P41,P42,P43,P422,P4212,P4122,P41212,P4222,P42212,P4322,P43212
   3   0      mP    125.15    125.06    272.20     90.0  90.1  90.1    P2,P21
   2   0      aP    125.06    125.15    272.20     89.9  90.0  89.9    P1
   1   0      aP    125.06    125.15    272.20     90.1  90.0  90.1    P1
 No PENALTY LATT      a         b         c      alpha  beta gamma  Possible spacegroups


 Suggested Solution:  4  P4
            penalty:  1
               cell: 125.056  125.148  272.201  90.10  89.97  90.07
   regularized cell: 125.102  125.102  272.201  90.00  90.00  90.00
           Symmetry: tP (Primitive Tetragonal)
```

Figure 5.6: MOSFLM P4 space group prediction
Using 3 diffraction images to index the TDH crystal, the space group predicted was P4, as opposed to C2 when using only 2 images.

This gave much better predictions and the data was processed in P4 and the data scaled, giving an R_{merge} of 0.154.

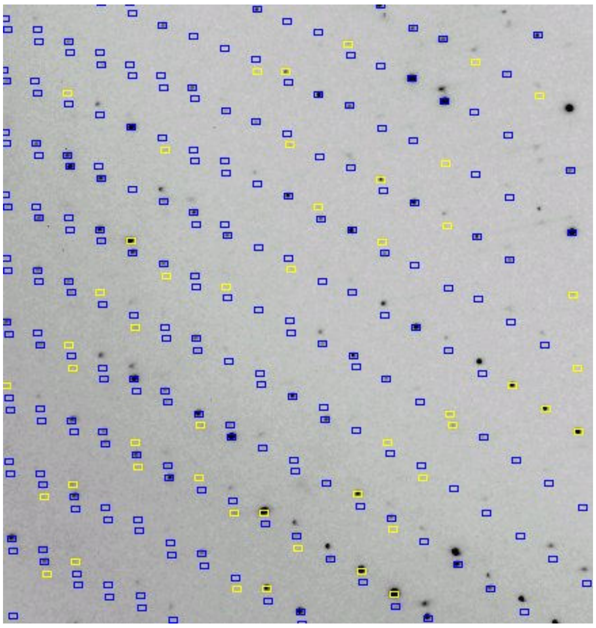
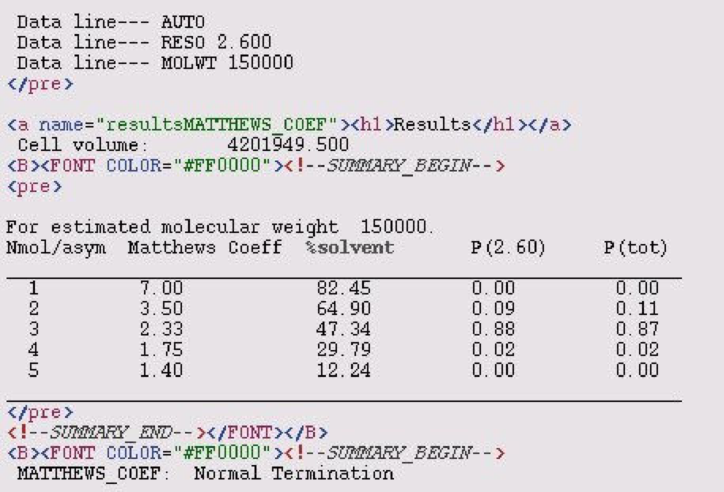


Figure 5.7: Section of MOSFLM spot prediction in P4 space group
As can be seen from the image, in this space group MOSFLM is able to correctly predict spots because the coloured squares accurately coincide with the diffraction spots.

5.4.2 Molecular Replacement

Matthews_coef (CCP4 1994) was used to predict the number of molecules present in the asymmetric unit. It was expected TkTDH would exist as a homotetramer because many other ADHs and PhTDH do, so homotetramers of approximately 150000 Da were searched for and it was predicted that there were 3 in the asymmetric unit, giving a solvent content of 47.34 % (when the data was processed in $P4_3$).



```
 Data line--- AUTO
 Data line--- RESO 2.600
 Data line--- MOLWT 150000
</pre>

<a name="resultsMATTHEWS_COEF"><h1>Results</h1></a>
 Cell volume:        4201949.500
<B><FONT COLOR="#FF0000"><!--SUMMARY_BEGIN-->
<pre>

For estimated molecular weight  150000.
Nmol/asym  Matthews Coeff  %solvent       P(2.60)     P(tot)
_____________________________________________________________
  1         7.00            82.45          0.00         0.00
  2         3.50            64.90          0.09         0.11
  3         2.33            47.34          0.88         0.87
  4         1.75            29.79          0.02         0.02
  5         1.40            12.24          0.00         0.00
_____________________________________________________________
</pre>
<!--SUMMARY_END--></FONT></B>
<B><FONT COLOR="#FF0000"><!--SUMMARY_BEGIN-->
 MATTHEWS_COEF:  Normal Termination
```

Figure 5.8: TDH matthews coefficient prediction
The program predicted the presence of three homotetramers of TDH in the crystallographic asymmetric unit, based on a molecular weight of 150 kDa.

The first TDH crystal structure was recently solved (Ishikawa *et al.* 2007), providing a molecular replacement model for our structure. That TDH comes from the hyperthermophilic archaeon *Pyrococcus horikoshii* and shares 80 % sequence homology with this TDH from *Thermococcus kodakaraensis*. A 5 % free R-flag data set was picked in thin shells using Dataman. Shells were used instead of a random selection as non-crystallographic symmetry (NCS) existed in the model.

MOLREP was used in molecular replacement with a single monomer from 2DFV as the search model looking for 3 homotetramers, but was unable to find a solution. PHASER was then used as an alternative and found a solution with a Log Likelihood Gain (LLG) of 105. In order to be more accurate with the space group, PHASER was set to run with alternative space groups, meaning that it tried $P4$, $P4_1$, $P4_2$ and $P4_3$, then picked the most promising solution and performed full molecular replacement in that space group. This resulted in $P4_3$ being the most promising space group as shown below.

	$P4$	**$P4_1$**	**$P4_2$**	**$P4_3$**
R_{merge}	0.154	0.115	0.116	0.115
LLG	105	160	No solution	5055

Still searching for 3 homotetramers, PHASER was used to find a molecular replacement solution for all choices of alternative space group as detailed below.

	$P4$	**$P4_1$**	**$P4_2$**	**$P4_3$**	**$P4_322$**	**$P422$**	**$P42_12$**	**$P4_122$**	**$P4_12_12$**	**$P4_222$**	**$P4_22_12$**	**$P4_32_12$**
R_{merge}	0.154	0.115	0.116	0.115	0.118	0.118	0.111	0.118				0.118
LLG	105	160	No solution	5055	780	-4.98	-4.21	165	125	-82	-54	521

$P4_322$ had the highest LLG of the higher symmetry space groups so a round of rigid body refinement was carried out using Refmac 5. However, this gave an R_{factor} of 0.66, so this seemed unlikely to be the correct solution.

As there was no obvious solution from the molecular replacement solutions found so far (searching for a number of homotetramers), P2 was tried. This had the lowest R_{merge} so far (0.103) and the highest LLG (12030), so it was decided to try the higher space groups within P2. It was also decided to search for monomers in the asymmetric unit. Initially 12 monomers were searched for, as this was the most favourable solution given by the Matthews coefficient, but PHASER was unable to find 12 monomers. It did however find 8 monomers.

	$P2$	**$P2_1$**	**$P222$**	**$P222_1$**	**$P22_12_1$**	**$P2_12_12_1$**
R_{merge}	0.103	0.106	0.111	0.111	0.111	0.111
LLG	12030	10310	10763	21182	low	24276

As can been seen in the table above, $P2_12_12_1$ gave the best solution. However, after a round of rigid body refinement in Refmac, $R_{factor} = 0.42$ and $R_{free} = 0.45$. This seemed very high but a round of restrained refinement in Refmac was performed, giving $R_{factor} = 0.29$ and $R_{free} = 0.34$. These statistics seemed reasonable, but when the model was viewed in COOT it was found that many of the peptide bonds were broken and it was not a good model.

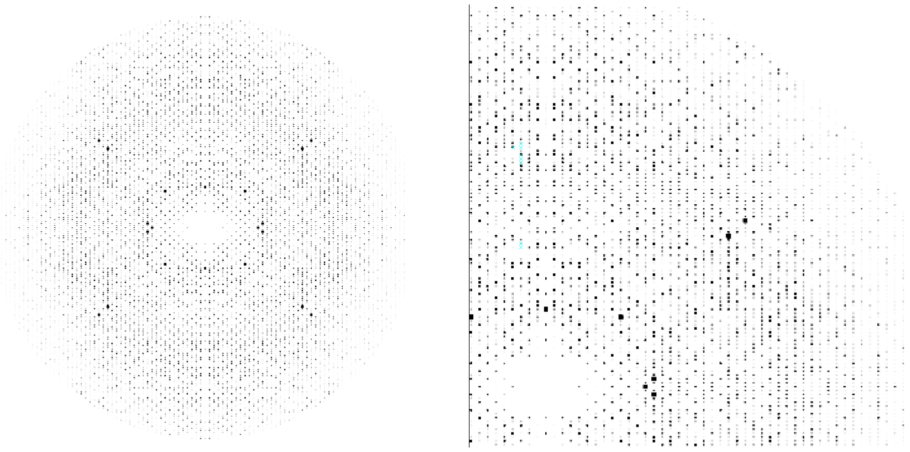
No good molecular replacement solution had been found and there was no significant difference between the statistics to distinguish a clear solution. However, with a molecular replacement model of 80% sequence similarity to TkTDH, it seemed likely that this was indeed just a case of finding the correct space group. Eventually it was looking at the HKLview for P2 and P4 in detail that helped to finally determine the space group.

It was clear from the P4 hk0 view that any given reflection was rotated 90° about the z (l) axis, suggesting P4 as opposed to P2. The 0kl and h0l views showed systematic absences every 2nd reflection along the h and k axis and every 4th along the l axis, suggesting $P4_12_12$ or $P4_32_12$. It could also be clearly seen from the hk0 view that there was a clear diagonal mirror plane suggesting P422 over P2. Pointless (CCP4 1994) was also run and this predicted P422 with the data processed in both P2 and P4.

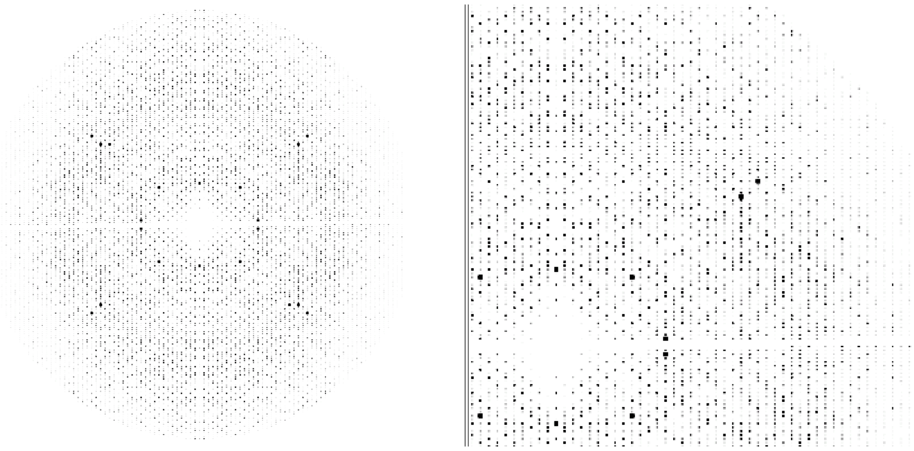
Having established that the point group was likely to be 422, and knowing that in previous runs of PHASER, $P4_3$ was preferred over $P4_2$ or $P4_1$, a molecular replacement solution was searched for in $P4_322$ and $P4_32_12$. Initially 8 monomers were searched for, based on previous runs. However, no solutions were found for either space group and both stopped after finding 4 monomers. Matthews_coef was run again, searching for 1 homotetramer of 150000 Da, which for both space groups suggested there were 4 monomers in the asymmetric unit. PHASER was again run in $P4_322$ and $P4_32_12$, this time searching for 4 monomers in the asymmetric unit.

	$P4_322$		**$P4_32_12$**	
	8 monomers	4 monomers	8 monomers	4 monomers
Matthews coefficient	1.75	3.50	1.75	3.50
Solvent Content	29.79%	64.90%	29.79%	64.90%
Log Likelihood Gain	No solution	No solution	No solution	15485

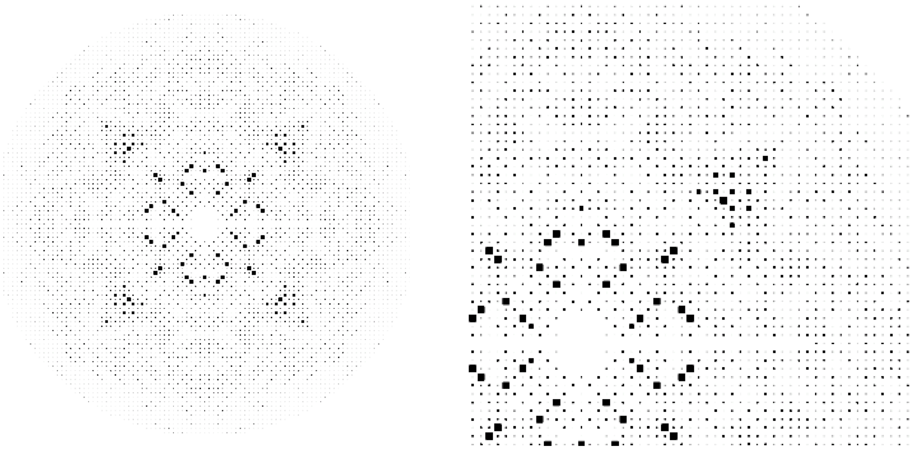
Thus $P4_32_12$ was found to be the correct space group, based on the molecular replacement, giving an R_{merge} of 0.148 and a log likelihood gain (LLG) of 15485. The corresponding unit cell dimensions are $a = b = 124.5$ Å, $c = 271.1$ Å. 5 for $P4_32_12$.



P2 0kl P2 0kl

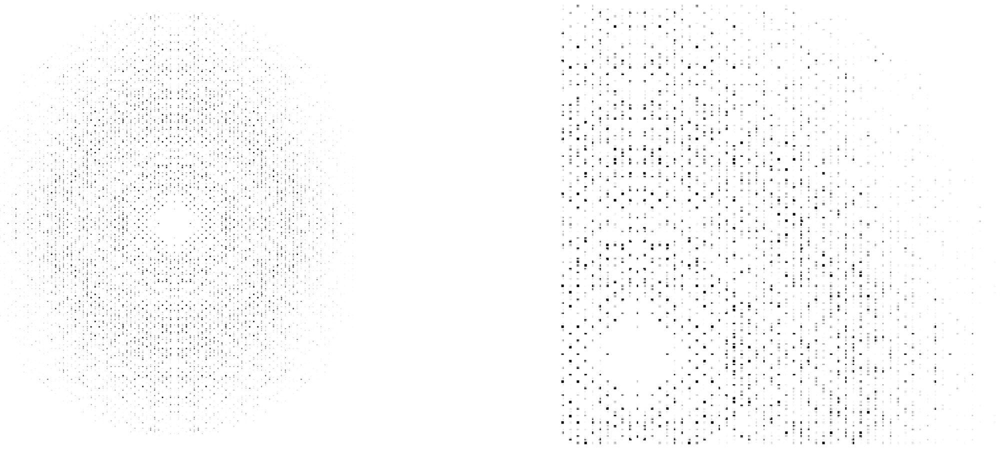


P2 h0l P2 h0l

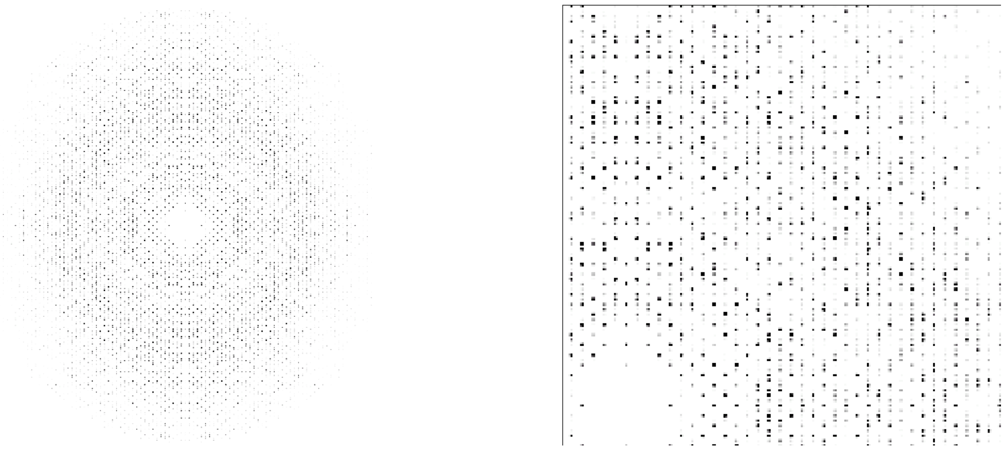


P2 hk0 P2 hk0

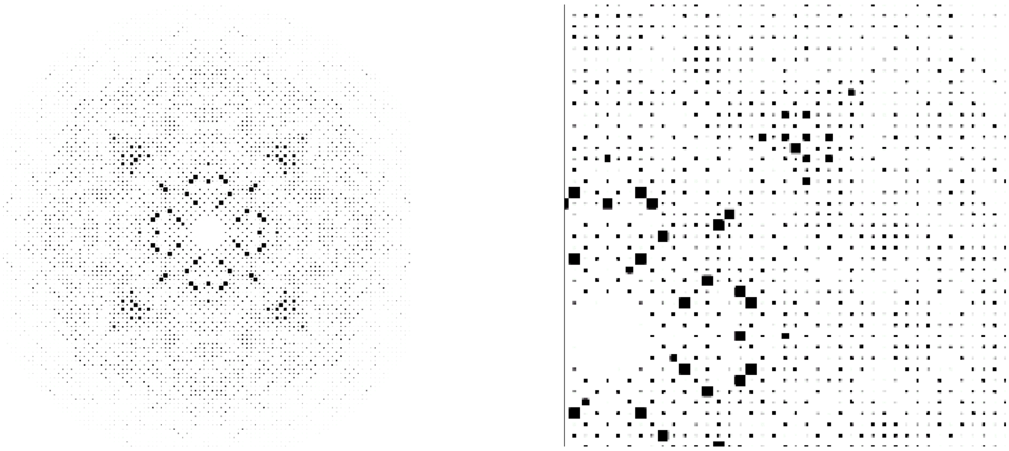
Figure 5.9: HKL view of TDH crystals in P2 space group



P4 0kl P4 0kl



P4 h0l P4 h0l



P4 hk0 P4 hk0

Figure 5.10: HKL view of TDH crystals in P4 space group
The 0kl and h0l views showed systematic absences every 2nd reflection along the h and k axis and every 4th along the l axis, suggesting $P4_12_12$ or $P4_32_12$. It could also be clearly seen from the hk0 view that there was a clear diagonal mirror plane suggesting P422 over P2.

Data collection and processing information for TkTDH is shown in Table 5

Beamline	ID14-3 (ESRF)
Wavelength (Å)	0.931
Space group	$P4_32_12$
Unit-cell parameters	
a=b (Å)	124.5
c (Å)	271.1
Resolution(Å)	73.9-2.4 (2.5-2.4)
R_{merge}^* (%)	11.4 (60.0)
Completeness (%)	100 (100)
Average I/σ	18.2 (4.8)
Multiplicity	15.1 (14.8)
Number of observed reflections	1001694 (141093)
Number of unique reflections	66425 (9511)
Wilson's B factor ($Å^2$)	51.3
Matthew's coefficient	3.50
Solvent content (%)	64.90
Number of molecules in the asymmetric unit	4

Table 5: Data collection and processing statistics for TkTDH
Values in parenthesis are for the outer resolution shell. * $R_{merge} = \Sigma_h\Sigma_i |(I_{hi}-I_h)| / \Sigma_h\Sigma_i (I_{hi})$
where I_h is the mean intensity of the scaled observations I_{hi}.

5.4.3 Model Building and Refinement

Initially rigid body refinement was carried out in Refmac to 2.6 Å, giving an R_{factor} = 0.43 and R_{free} = 0.43. Using COOT, residues in the model that differed between PhTDH and TkTDH were mutated to the correct residues and restrained refinement was done using Refmac, giving R_{factor} = 0.26 and R_{free} = 0.32. The difference between R_{factor} and R_{free} was greater than the allowed 5 %, and when viewed in COOT, it was found that the model was broken in many places.

As it appeared Refmac was not refining this structure well, it was decided to try using CNS instead. Refinement was carried out with data to a resolution of 2.4 Å, using CNS (Brunger et. al., 1998). Three rounds of simulated annealing up to 4000ºC and rigid body refinement with tight NCS restraints between the four subunits, interspersed with model building in COOT, were carried out. During the third round, group B-factors were also refined. Two rounds of rigid body refinement were then carried out using PHENIX (Adams et. al., 2002) with several cycles of tight NCS restraints, lower temperature simulated annealing and translational symmetry restraints, again interspersed with model building. The final seven rounds of restrained refinement were carried out using REFMAC 5 (CCP4 1994), with medium NCS restraints. 490 water molecules were added and the co-factor NAD^+ was built into the electron density using COOT and the structure rebuilt using both the $2F_o$-F_c and F_o-F_c maps. The final R_{factor} and R_{free} values are 22.0 % and 27.6 %, respectively, and the diffraction data and crystallographic refinement statistics are summarised in Table 6.

Resolution Range (Å)	73.9 - 2.4 (2.5-2.4)
Space group	$P4_32_12$
R_{factor} (%)	22.0
R_{free} (%)	27.6
Subunits per asymmetric unit	4
Number of non-hydrogen protein atoms	10672
Number of water molecules	490
Wilson's B-factor (Å^2)	42.6
Average B-factor (Å^2)	43.0
RMSD bond lengths (Å)	0.038
RMSD bond angles (deg.)	2.968
Matthew's coefficient	3.50
Solvent content (%)	64.90
R_{merge}^* (%)	14.6 (59.0)
Completeness (%)	100 (100)
Average I/σ	14.8 (2.7)
Multiplicity	14.5 (11.8)

Table 6: Refinement statistics for TkTDH
Values in parenthesis are for the outer resolution shell. * $R_{merge} = \Sigma_h\Sigma_i |(I_{hi}-I_h)| / \Sigma_h\Sigma_i (I_{hi})$ where I_h is the mean intensity of the scaled observations I_{hi}.

Unfortunately the lobe loop regions of helix α3 are very disordered in our model, and the electron density is poor. This high degree of disorder is likely due to the absence of a structural zinc ion, which is expected to stabilise the region. All other areas of the model shown strong electron density and the model fits the map well, as illustrated in Figure 5.11.

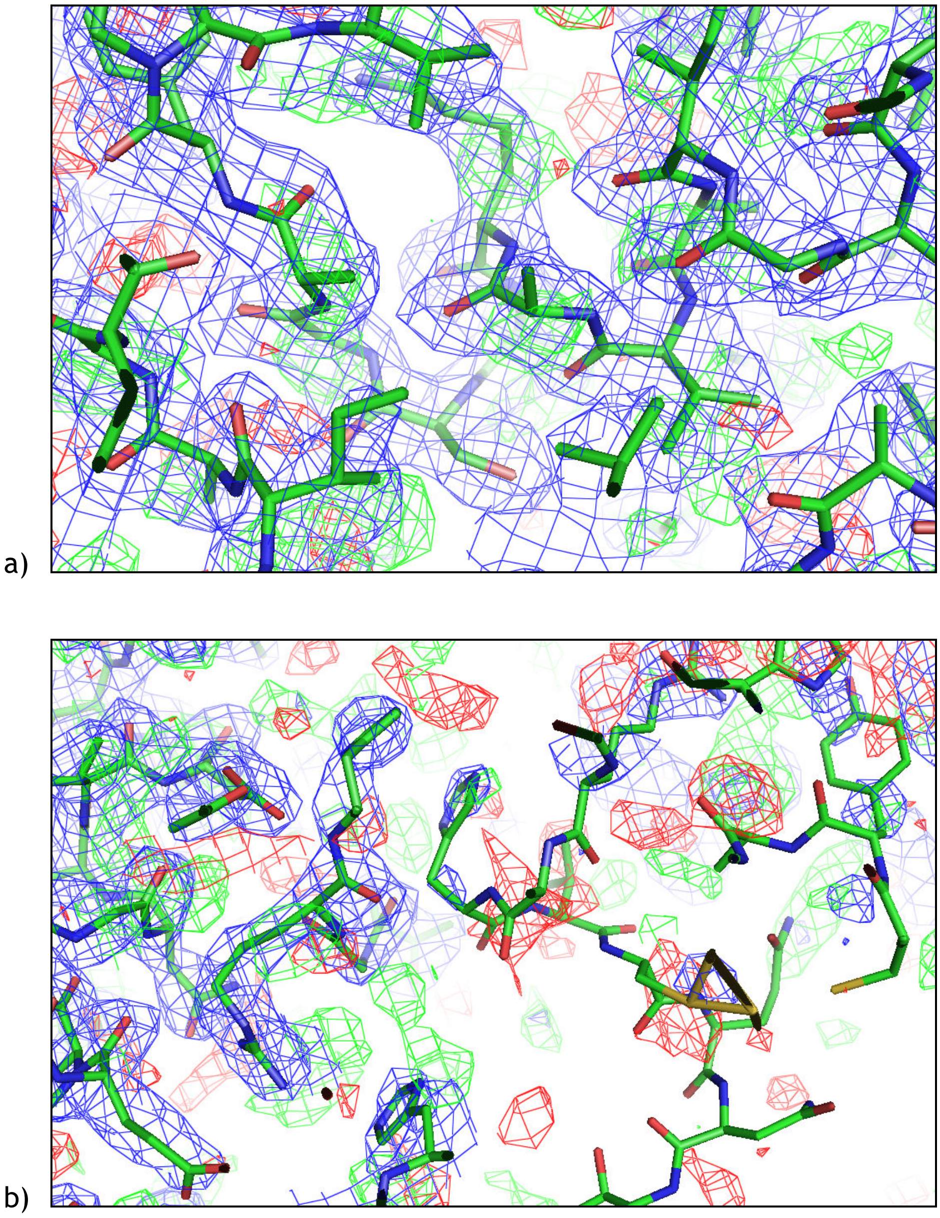


Figure 5.11: Electron density for the TkTDH model
In general there is very good electron density for the TkTDH model, as can be seen in a) for the area around residue 188. However, the electron density around the flexible loop region of α3, as can be seen in b) around residue 100. This is most likely probably due to the loops' flexibility in the absence of a structural zinc ion.
$2F_o\text{-}F_c = 1.5\sigma$ (blue), $F_o\text{-}F_c = -2\sigma$ (red) / 2σ (green)

The figures of the structure were prepared using Pymol. Areaimol (CCP4 1994) was used to calculate the surface area for an individual monomer, a dimer and the biological tetrameric protein. Buried surface areas were calculated using CNS (Brunger *et al.* 1998), and COOT (Emsley and Cowtan 2007) was used to generate RMS deviations and align subunits.

5.5 Structure Analysis

5.5.1 Primary Structure Similarity

The gene encoding L-threonine dehydrogenase, identified in the genome sequence of *T. kodakaraensis* (TkTDH), indicates that the protein comprises 350 amino acid residues. The protein has a molecular mass of 38,016 Da per monomer, as determined by electrospray mass spectrometry. As well as displaying high sequence similarity with other characterised TDHs from hyperthermophiles; *P. abyssi* (88 % identity), *P. furiosus* (88 % identity) and *P. horikoshii* (87 % identity) as determined by a BLAST search, TkTDH is homologous to other TDHs and some ADHs, as shown in Figure 5.12.

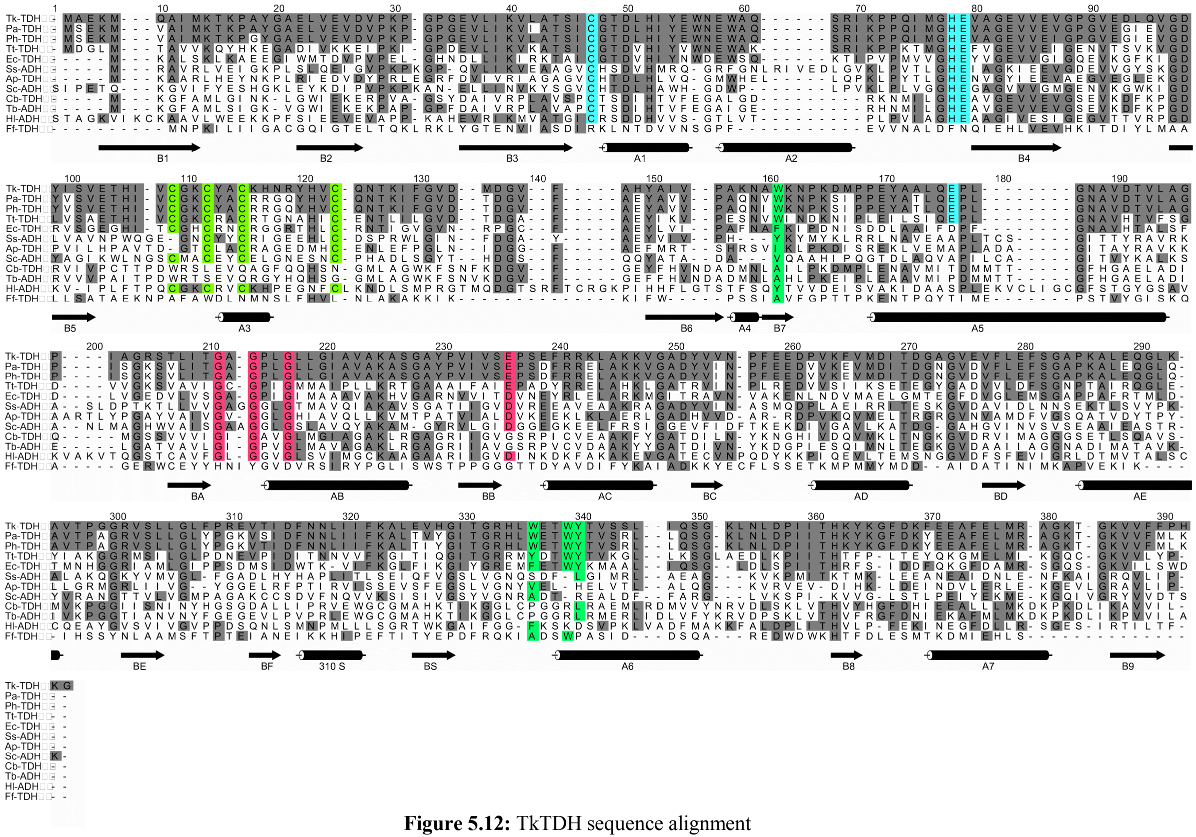


Figure 5.12: TkTDH sequence alignment

This alignment compares TkTDH with TDHs and ADHs from a variety of other organisms. TkTDH displays 87.4% sequence identity with PhTDH (*Pyrococcus horikoshii*), 87.6% identity with PaTDH (*Pyrococcus abyssi*), 51.0% identity with TtTDH (*Thermoanaerobacter tengcongensis*), 42.8% identity with EcTDH (*Escherichia coli*), 26.3% identity with Horse Liver ADH, 28.5% identity with SsADH (*Sulfolobus solfataricus*), 26.9% identity with CbADH (*Clostridium beijerinckii*), 24.9% identity with TbADH (*Thermoanaerobacter brockii*), 26.7% identity with ApADH (*Aeropyrum pernix*), 25.9% identity with ScADH (*Saccharomyces cerevisiae*) and 9.6% identity with FfADH (*Flavobacterium frigidimaris*).

The conserved cysteine residues thought to co-ordinate the structural zinc ion are shown in bright green, the conserved Cys and His and Glu residues thought to co-ordinate the catalytic zinc ion are shown in bright blue and the conserved GxGxxG NAD$^+$ binding motif is shown in pink. The four hydrophobic residues that form a cluster in some medium-chain dehydrogenases are shown in green. The α-helical and β-strand secondary structure is shown diagrammatically beneath the sequences.

Amongst the sequences of the aforementioned enzymes are some highly conserved residues, most notably the four cysteine residues that coordinate the structural zinc ion (Cys97, Cys100, Cys 103 and Cys 111), the four residues proposed to co-ordinate a catalytic zinc ion (Cys42, His67, Glu68 and Glu152) and the GxGxxG NAD$^+$ binding motif (formed by Gly175, Gly 177 and Gly180). It is worth noting that although equivalents to the first three aforementioned residues co-ordinating the catalytic zinc ion are conserved between TDHs and ADHs, Glu152 is less conserved. For example in *E. coli* ADH (EcADH) this glutamate residue is substituted for Asp152 and in ApADH this is Ala152. With regards the structural zinc, it is interesting to note that the 15 amino acid sequence of TkTDH containing the four cysteine residues ^{97}CGKCYACKHNRYHVC^{111} shows four differences (grey shaded region) with the TDH's of other hyperthermophiles and yet more differences with respect to the mesophilic homologues; shown in Figure 5.13 this loop is generally more polar in mesophilic enzymes. The difference in this lobe loop may somehow contribute to the tighter binding of the structural zinc to TkTDH compared to *P. horikoshii* and *P. furiosus* (unpublished data).

5.5.2 Structure of the Monomer

TkTDH exhibits high structural similarity with PhTDH from the hyperthermophile *P. horikoshii* but also has a similar overall fold to medium-chain ADHs from mesophilic organisms, as illustrated in Figure 5.13. Superimposition of all 348 C-α carbon atoms of the holo form of PhTDH with TkTDH reveals an RMS deviation of less than 0.57 Å, indicating how structurally similar these two enzymes are. Although a greater RMS deviation of 1.90 Å is observed between TkTDH and the holo form of SsADH, using 300 C-α atoms closer than 3.5 Å, this is unsurprising given the low sequence homology and still indicates a much higher structural homology than might be predicted from the amino acid sequence, which has 29 % identity (Figure 5.12).

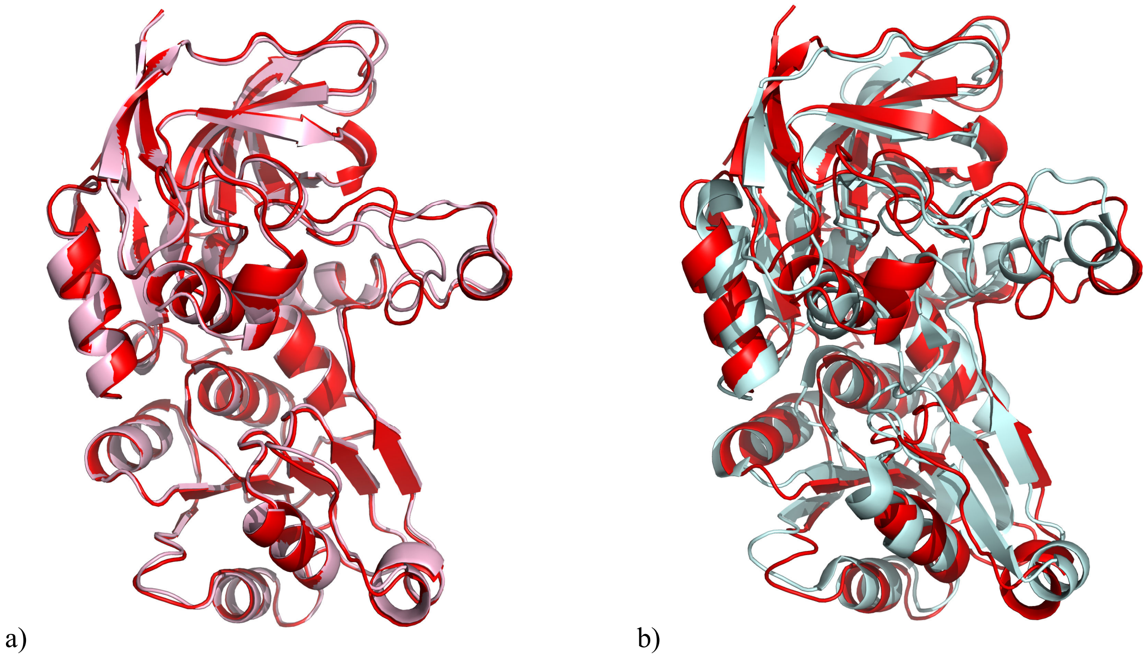


Figure 5.13: Superimposition of TkTDH with PhTDH and SsADH
a) A ribbon diagram showing the fold of a monomer of holo-TkTDH (shown in red) aligned by all 348 C-alphas with a monomer of holo-PhTDH (87 % sequence similarity) (shown in pink) with an RMS deviation of 0.57 Å. **b)** A monomer of holo-TkTDH (shown in pale blue) was superimposed on a monomer of holo-SsADH (29 % sequence similarity) (shown in blue) with an RMS deviation of 1.90 Å for 300 C-alpha carbon atoms that are closer than 3.5 Å.

The major functional difference between TDHs and ADHs is the substrate specificity. Each monomer is composed of two domains; a nicotinamide cofactor (NAD^+)-binding domain (residues 153-292), which folds into a classical α/β Rossmann fold motif, characteristic of mononucleotide-binding proteins and a catalytic domain (residues 3-152 and 293-349). In total the monomer contains 12 α-helices and 16 β-strands, making up 31 % and 23 % of the structure, respectively. The overall topology of the monomer is shown in Figure 5.14 and the assigned secondary structure labelling in Figure 5.15 is in keeping with the nomenclature for ADHs (Korkhin *et al.* 1998).

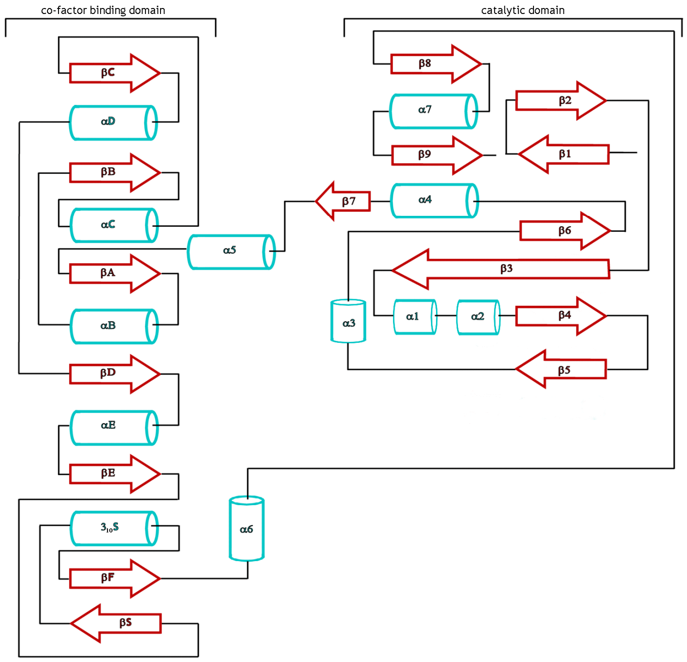


Figure 5.14: Overall topology of a TkTDH monomer
α-helical segments are shown in blue and β-strands are shown in red. The five α-helices (A, B, C, D, E) and 3_{10}S and seven β-strands (A, B, C, D, E, F and S) from residues 171 to 290 form the α/β Rossmann fold of the co-factor binding domain.

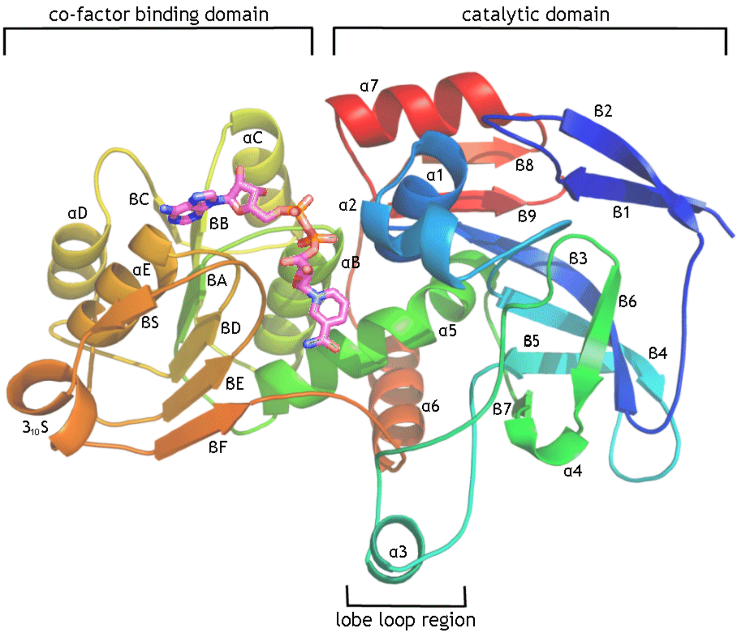


Figure 5.15: TkTDH nomenclature
This ribbon diagram shows the assignment of nomenclature to the secondary structure of TkTDH and indicates the main regions of interest. Helices 1-4, 6 and 7 and β-strands 1-7, 8 and 9 comprise the catalytic domain, helices B-E and 3^{10}S and β-strands A-F and S form the NAD^{+}-binding domain and helix 5 acts as a linker between the two.

The five helices (A, B, C, D, E) and 3_{10}S and seven β-strands (A, B, C, D, E, F and S) from residues 171 to 290 form the α/β Rossmann fold of the co-factor binding domain. Six of these β-strands make a parallel β-pleated sheet flanked by the 5 helices and the seventh β-strand. The catalytic domain consists of a core of mostly antiparallel β-strands (1-9) with seven α-helices (1-7) at the surface of the molecule.

The short α3 helix is contained within a protruding lobe loop region containing the 4 conserved cysteine residues that are expected to co-ordinate the structural zinc ion with tetrahedral geometry, on the basis of the PhTDH structure (Ishikawa *et al.* 2007). It is thought this may confer further thermostability to the enzyme, although this is also present in some mesophilic medium-chain dehydrogenases. This region appears disordered in the TkTDH structure; although it is expected that in the presence of Zn^{2+} it would be more ordered. There is a long 18 residue α5 helix (residues 145 to 162, the longest in the molecule, which forms a linker between the two domains, possibly acting as a hinge when NAD^+ and threonine bind. This helix is well defined in the electron density map, but is clearly kinked part way along at Pro153, indicating a possible bending site. Three c*is*-peptide bonds are observed in the structure preceding Pro62, Pro194 and Trp297. Pro62 is highly conserved amongst both TDHs and ADHs, as is the *cis*-peptide conformation that it adopts in the tertiary structure. However, both Pro194 and Trp297 appear to only be conserved in TDHs, but not ADHs. Trp297, a rare non-prolyl *cis*-peptide observed on the surface of the catalytic domain of both TkTDH and PhTDH appears to form part of a hydrophobic cluster with Trp137, Trp300 and Tyr301. The electron density for this region surrounding the cis peptide itself is very well defined in both structures. All four of these hydrophobic residues appear to be conserved amongst TDHs from hyperthermophiles, suggesting this may be a contributing factor to their thermostability. However, this hydrophobic cluster is not constitutively present in ADHs from hyperthermophiles and in some mesophilic ADHs alternative hydrophobic resides constitute this cluster.

Electron density was observed in the cofactor-binding domain for NAD^+. It is thought that binding of NAD^+ to TDH causes a conformational change in which residues of the active site are brought closer together, thereby enabling threonine to bind and the reaction to occur (Ishikawa *et al.* 2007). This ordered Bi Bi mechanism of binding is supported by the work done with PhTDH (Ishikawa *et al.* 2007;Higashi *et al.* 2008) and has been reported by Aoyama et. al. 1981. This binding of threonine is thought to cause a further rearrangement of α1 helix into a closed conformation, bringing it closer to the active site, as observed in horse liver ADH (HlADH) (LeBrun *et al.* 2004;Ramaswamy *et al.* 1999). The unstable intermediate 2-amino-3-ketobutyrate leaves the enzyme prior to the release of NADH.

5.5.3 Structure of the Tetramer

TkTDH exhibits a homo-tetrameric structure within the crystallographic asymmetric unit, with an estimated solvent content of 64.9 % (v/v), V_M=3.5, and this is believed to be the biologically active form of the enzyme, in keeping with many medium-chain dehydrogenases of lower organisms. The structure is very similar to PhTDH (RMS deviation 0.57 Å) and, despite low levels of sequence identity, is quite structurally similar to PaTDH (87.6 % identity) and SsADH (28.5 % identity) as can be seen in Figure 5.16.

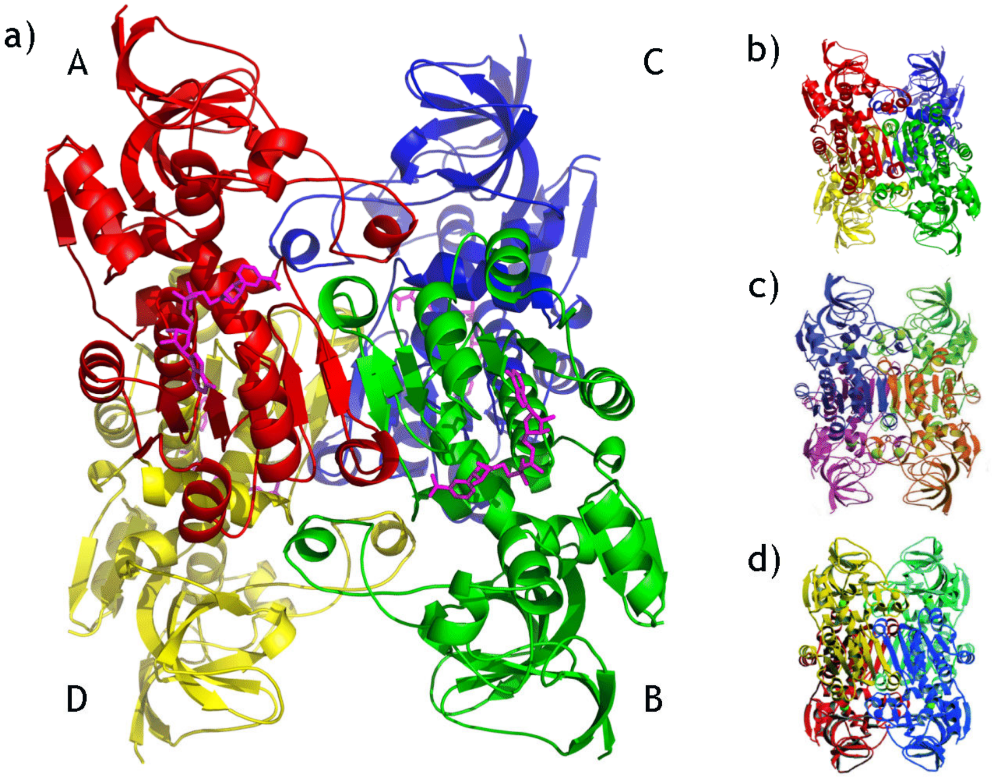


Figure 5.16: Tetrameric structure of TkTDH compared to PhTDH, PaTDH and SsADH
a) This shows the tetrameric structure of TkTDH as a dimer of dimers with the co-factor NAD^+ bound. Monomer A is shown in red, B in green, C in blue and D in yellow. The overall fold of this enzyme is very similar to that of other medium-chain dehydrogenases such as **b)** *P. horikoshii* TDH, **c)** P. *aeruginosa* ADH and **d)** *S. solfataricus* ADH, with an extended β-sheet formed by the Rossmann fold of monomers A and B (and C and D).

Essentially the quaternary structure comprises a dimer of dimers. The formation of dimers between monomers A and B (and C and D) buries 1552.6 Å^2 of the 15,425.4 Å^2 accessible surface of each monomer, resulting in a solvent-accessible surface area of 27,287.8 Å^2 for each dimer. This dimer is shown in Figure 5.17.

Figure 5.17: TkTDH dimer of monomers A and B
β-strands of the Rossmann fold form an extended 12-stranded β-sheet across the AB dimer interface to create an extensive region of contact between the monomers. This binding involves β-strands F (residues 273–275) and S (residues 287-290) of both subunits, where the βF strand of monomer A forms hydrogen bonds with the anti-parallel βF strand of monomer B, and likewise βS of A hydrogen bonds the anti-parallel βS of B. The same is true for monomers C and D.

Association of the dimers into a tetramer buries a further 5038.9 Å^2 giving a total accessible surface area of 50168.9 Å^2. Figures similar to those are found for PhTDH (Ishikawa *et al.* 2007). As has been found in many ADHs and PhTDH, the β-strands of the Rossmann fold form an extended 12-stranded β-sheet across the dimer interface to create an extensive region of contact between monomers A and B (and C and D). This binding involves β-strands F (residues 273–275) and S (residues 287-290) of both subunits, where the βF strand of monomer A forms hydrogen bonds with the anti-parallel βF strand of monomer B, and likewise βS of A hydrogen bonds the anti-parallel βS of B. In addition there is a minor interaction at the structural lobe of helix α3 (95-117) where hydrogen bonds and electrostatic interactions bridge the two dimers to form the tetrameric dimer of dimers: this interaction buries 1172.1 Å^2 accessible surface area between each monomer in the following pairings of A/C and B/D. This dimer is shown in Figure 5.18.

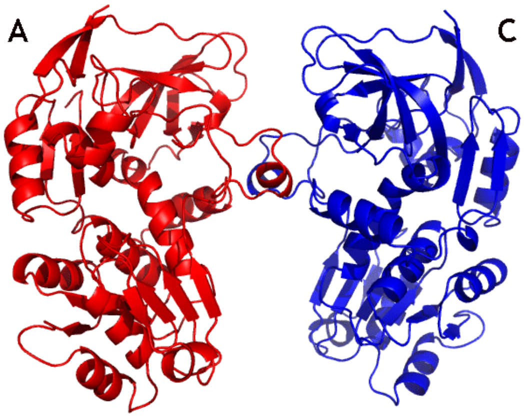


Figure 5.18: TkTDH dimer of monomers A and C
The small α3 helix, proposed to contain a structural zinc ion, from each of monomer A and C form additional bonds to stabilise the tetramer.

Conversely the opposite monomers of A/D and B/C interact primarily through helices αB and α6 as shown in Figure 5.19.

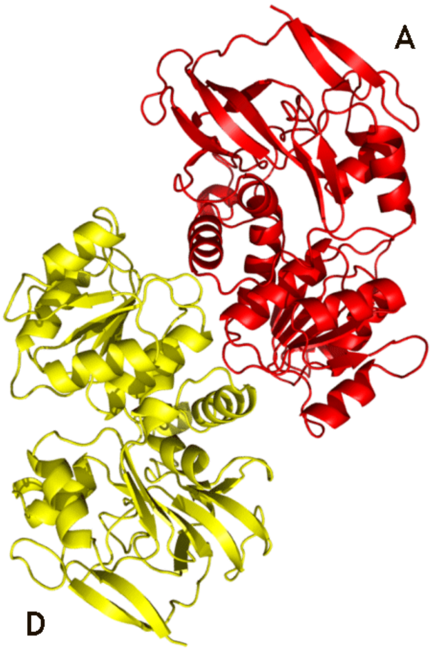


Figure 5.19: Interaction of monomers A and D in TkTDH

Specifically Gly 168 in the loop between α5 and βA of monomer A forms 2 hydrogen bonds with Ser 305 (α6) of monomer D and Gly 191 (αB) of monomer A hydrogen bonds Ala 189 (αB) of monomer D. The reverse is also true due to the 2-fold symmetry of the tetramer. This interface buries 1150.4 Å^2 solvent accessible surface on each monomer. The relatively small areas buried by these additional dimer contacts, compared with that of the greater surface area that is buried by dimer formation, does justify its description as a dimer of dimers.

5.5.4 Structural Zinc Ion

In TkTDH a structural zinc ion is proposed to bind in the lobe loop, which includes helix α3, although the zinc was not observed in this structure. Most medium-chain dehydrogenases contain both a catalytic zinc ion at the active site, and a second structural zinc ion located in the lobe loop of the catalytic domain where it most likely rigidifies the region (Jelokova *et al.* 1994), thereby conserving interfaces important in stabilising the enzymes' quaternary structure. However, the zinc content does vary between ADHs with either one or two zinc ions being present per subunit, for example *Thermoanaerobacter brockii* ADH (TbADH) (Bogin *et al.* 1997) contains only a catalytic zinc ion whereas EcADH contains both a catalytic and structural zinc ion (Sulzenbacher *et al.* 2004). The importance and functionality of this zinc ion has been demonstrated with mammalian alcohol dehydrogenase, where a structural zinc ion is naturally present, by site-directed mutagenesis of the four co-ordinating cysteine residues. These mutants were unable to ligate zinc and resulted in inactive, unstable enzymes (Jelokova *et al.* 1994). However the structural zinc ion is not present in ADHs from some organisms such as *Thermoanaerobacter brockii* (Bogin *et al.* 1997) and *Clostridium beijerinckii* (Korkhin *et al.* 1998).

Although catalytically active TkTDH was used in the crystallisation trials, no electron density is observed for zinc ions in the structure in either the proposed structural or catalytic positions. It is likely that sodium citrate in the crystallisation conditions may have chelated any bound zinc, thereby removing it from the protein. It was also observed that the 'lobe loop' region (97-115) exhibited highly disordered electron density and proved difficult to model, indicating that the structural zinc is indeed required to stabilise the loop, as observed for PhTDH (Ishikawa *et al.* 2007). It was however observed that the four conserved cysteine residues (Cys97, Cys100, Cys103 and Cys111) did co-localise to a central position large enough to accommodate a zinc ion with tetrahedral geometry without disrupting other residues. Mammalian ADHs, although most commonly dimeric, do usually contain both catalytic and structural zinc ions, whereas some bacterial ADHs adopting a tetrameric structure lack the structural zinc ion, suggesting that any correlation between the presence of a structural zinc in the lobe loop region, and the oligomeric state of the enzyme is not straightforward.

5.5.5 Active Site

Each subunit of the tetrameric TkTDH, like other medium-chain dehydrogenases, is formed of two domains, the catalytic and NAD$^+$ co-factor binding domains that are linked by helix α5 and the residues 290 to 298, forming a cleft between them, as shown in Figure 5.20a. This deep cleft forms a pocket large enough to accommodate both the NAD$^+$ co-factor and threonine substrate.

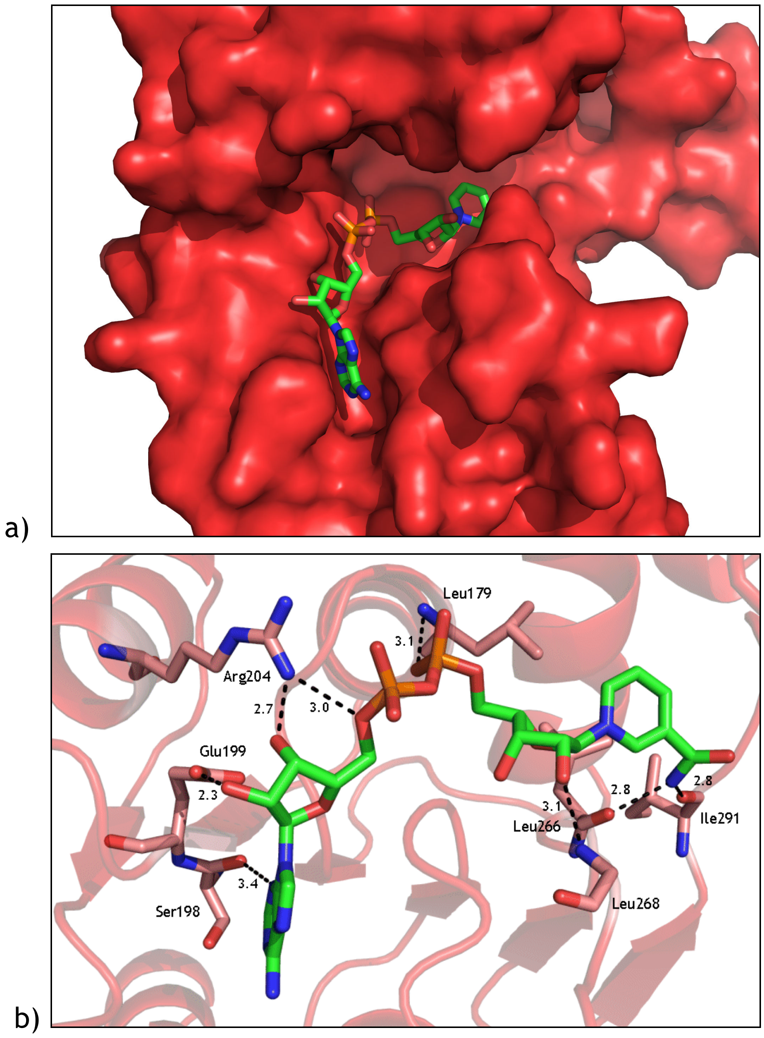


Figure 5.20: NAD$^+$ binding pocket in TkTDH
A surface representation shows NAD$^+$ bound to TkTDH in a deep cleft between the catalytic and co-factor binding domains, where threonine also binds (not shown) **b)** the co-factor forms extensive hydrogen bonds with Leu 179, Ser 198, Glu 199, Arg 204, Leu 266, Leu 268 and Ile 291, as well as many Van der Waal contacts with the surrounding residues.

In both TkTDH and PhTDH, this cleft has been observed to be less hydrophobic than that in ADHs. In ADHs the active site can be identified by the presence of a catalytic zinc ion and although this is also present in PhTDH, it was absent from the TkTDH structure. However, as already discussed, the absence of both the structural and catalytic zinc from TkTDH is likely to be due to the chelation of zinc by sodium citrate in the crystallisation conditions as opposed to it actually being absent from the native form of the enzyme. Although a catalytic zinc ion is absent from the TkTDH structure, the four residues postulated to co-ordinate it in PhTDH are conserved (Cys42, His67, Glu68 and Glu152) and these are positioned in a formation capable of liganding a zinc ion The residues proposed to form the active site in PhTDH are conserved in TkTDH, as shown in Figure 5.21

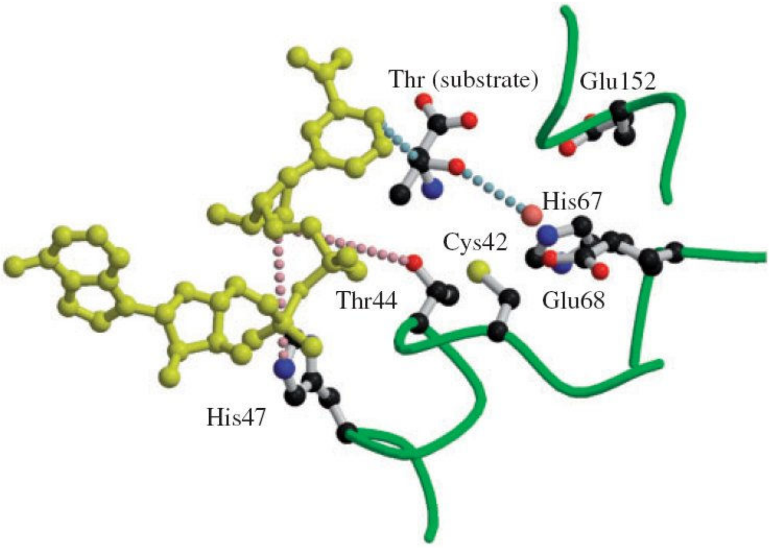


Figure 5.21: Proposed active site for TkTDH
The six residues proposed to be involved in forming the active site in PhTDH (as shown in the image above taken from (Higashi et. al. 2008) as conserved in TkTDH and are likely to act in a similar manner. A proton relay mechanism has been proposed as the catalytic mechanism for this enzyme.

The co-factor occupies the active site pocket and is bound to the protein predominantly by Van der Waal interactions and hydrogen bonds with the surrounding amino acid residues (Figure 5.19b). There are two sequence features common to NAD$^+$ binding proteins containing the Rossmann fold; the highly conserved amino acid sequence motif GxGxxG and the presence of either Asp or Glu at the C-terminal end of the second β-strand (βB) in the co-factor binding domain. This GxGxxG sequence forms a glycine-rich loop between the C-terminus of the βB strand and the N-terminus of αA, and forms extensive Van der Waal contacts with the phosphate groups of NAD$^+$. Mutations in the conserved glycine residues of this loop have been observed to cause a reduction or elimination of enzyme activity (Rescigno and Perham 1994). In TkTDH, Glu199 is present at the C-terminal end of β-strand E in the co-factor domain. This carboxyl residue is highly conserved amongst NAD^+ dependent TDHs and ADHs, but is replaced by Gly in NADPH-dependent enzymes (Korkhin *et al.* 1998). This may be because in NADPH-dependent enzymes the N-terminal of helix α3 is longer by 3 residues, causing a structural adjustment in the region of sheet βB and helix α6 and the glycine allows provides extra flexibility to accommodate the additional phosphate group. Threonine binds in close proximity to the putative zinc binding site and NAD^+ molecule. The distance between the reactive site of NAD^+ and the conserved putative catalytic zinc binding residues of Cys42, His67, Glu68 and Glu152 is approximately 7 Å.

The GxGxxG motif is important for determining substrate specificity for NAD^+, possibly by enabling the correct tertiary fold for binding at the active site, but other residues are responsible for actually binding the co-factor. As can be seen in Figure 5.20, seven residues are involved in forming hydrogen bonds and extensive Van der Waal contacts with the NAD^+. The mainchain of Leu 179 forms two hydrogen bonds, one with the nicotinamide ring and the other with an oxygen of the first phosphate group. The nicotinamide ring is also hydrogen-bonded by the mainchains of Leu 266 and Ile 291, with many contacts also being formed with Asn 156 and Gly 267. The nicotinamide ring was observed to bind in a single conformation as shown in Figure 5.22, unlike PhTDH where two alternative conformations were observed with equal occupancy (Ishikawa *et al.* 2007). In this model the NAD^+ was refined with 50 % occupancy.

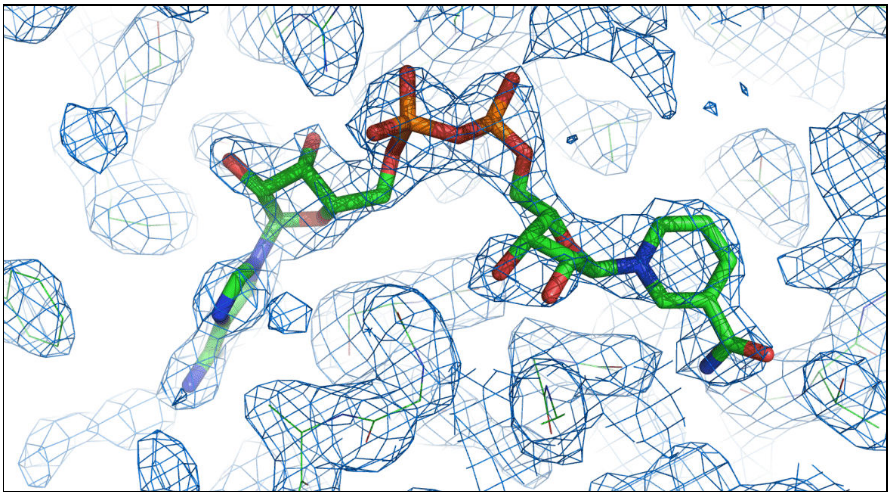


Figure 5.22: NAD^+ electron density in the TkTDH model
Strong electron density was observed at the active site for NADH and this was built into the model for each monomer. The nicotinamide ring was observed to bind in a single conformation, unlike PhTDH where two alternative conformations were observed with equal occupancy.

The mainchain of Leu 268 binds an oxygen of the first ribose ring and makes many other contacts. The second phosphate makes contacts with Pro 178 and Arg 204, and this second residue also hydrogen-bonds this phosphate and an oxygen of the second ribose ring. A second hydrogen bond is made with this ribose ring by Glu 199, as well as many contacts with Gly 175 and Phe 243. Arg 204 is highly conserved amongst TDHs but not ADHs, whereas Glu199 is present in both TDHs and ADHs. Kinetic parameters have been measured for mutant enzymes of PhTDH, in which Glu 199 and Arg204 were mutated to Ala, and this indicated these residues were indeed very important for NAD^+ binding (Ishikawa *et al.* 2007). The adenosine ring is hydrogen bound to Ser 198, as well as making significant contacts with Pro 200. A significant effect of these binding contacts is that the nicotinamide ring is bound in a slot formed by Leu 179, Leu 266 and Leu 268 and the adenine ring by Ser 198 and two water molecules.

A number of hydrogen bonds are also formed between the co-factor NAD^+ and water molecules at the active site, including one found to be conserved at this site in other di-nucleotide binding proteins containing a Rossmann fold (Bottoms *et al.* 2002). A total of eight water molecules have been found to directly hydrogen bond with NAD^+ along its length, most of which form further bonds with either residues or other water molecules. One of these waters is a conserved molecule that binds the co-factor and three residues. As can be seen from Figure 5.23, this water molecule forms four hydrogen bonds to the oxygen of Gly 175, the nitrogen of Gly 180, the oxygen of Phe 243 at the C-terminal of sheet βD and an oxygen of the first phosphate group of NAD^+ in a tetragonal co-ordination.

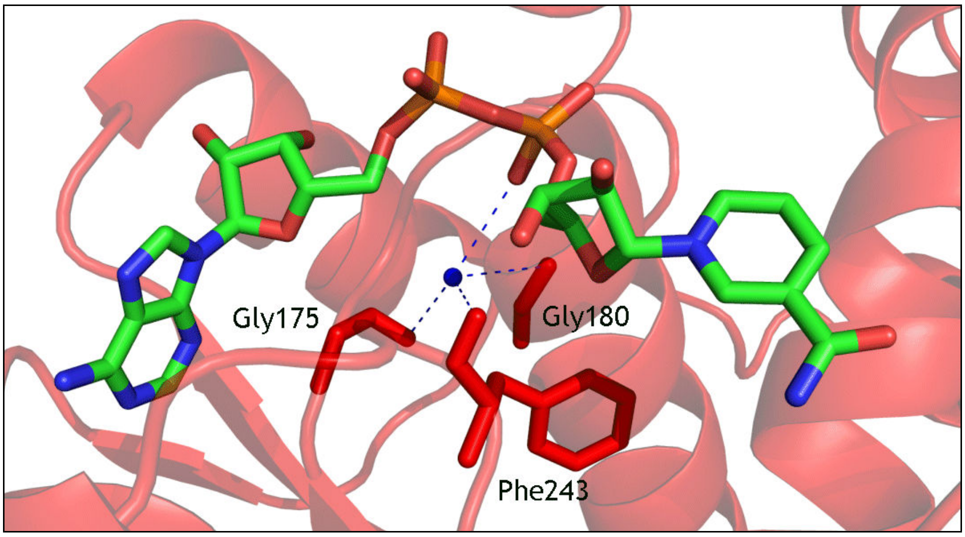


Figure 5.23: Conserved water molecule at the NADH binding site
It has been found that a water molecule conserved in di-nucleotide binding proteins containing the Rossmann fold, is present in TkTDH. This water forms four hydrogen bonds to the oxygen of Gly 175, the nitrogen of Gly 180, the oxygen of Phe 243 and an oxygen of the first phosphate group of NAD^+.

Gly 175 and Gly 180 are the first and third Gly in the sequence GxGxxG found to be conserved amongst NAD^+ binding proteins, so this conserved water may perform a role in stabilising the phosphate binding loop formed by this motif and an inherent role in the structure of the Rossmann fold. It is interesting to note that di-nucleotides have huge hydrogen bonding potential with NAD^+ containing 22 polar oxygen and nitrogen atoms. On average NAD^+ forms around 16 hydrogen bonds with the protein (Bottoms *et al.* 2002) and in the case of TkTDH seven are formed with residues and a further eight with water molecules. This illustrates how important these bridging waters may be as a component of di-nucleotide recognition and binding.

5.6 Discussion

5.6.1 Thermostability

It remains difficult to draw any firm conclusions as to the factors that confer high thermostability to hyperthermophilic medium-chain dehydrogenase enzymes. Although the extended 12 stranded β-sheet between AB and CD surely confers structural stability to the enzyme, this motif is not unique to hyperthermophilic medium-chain dehydrogenases as it is found in mesophilic enzymes. It appears likely that the rigidity of the enzymes' structure plays at least some role in their thermostability, but it is not possible to determine which residues specifically confer this property. Further mutational studies could be carried out to identify key residues.

5.6.2 Complex Formation

TDH is the first enzyme in a two step metabolic pathway that converts threonine to glycine by oxidising L-threonine to the unstable intermediate 2-amino-3-ketobutyrate. With the participation of CoA as a co-factor, the product is acted on by the second enzyme in the degradation pathway, 2-amino-3-ketobutyrate CoA ligase (KBL). The products of this final reaction are glycine and acetyl-CoA. These reactions catalysed by TDH and KBL have been shown to be coupled, both in vitro and in vivo (Marcus and Dekker 1993), which is probably due to the highly reactive nature of the intermediate. Under aqueous conditions it undergoes spontaneous decarboxylation to form aminoacetone and CO_2. Gel filtration has been used to demonstrate a physical interaction between the two enzymes and along with fluorescence experiments these findings indicate that two dimers of KBL associate with one tetramer of TDH (Tressel *et al.* 1986). In this experiment bovine TDH and KBL enzymes were used.

Since the three-dimensional structure of E.coli KBL is known (Schmidt *et al.* 2001) we used the 3D structure of TDH to model the putative interaction between the two enzymes. The active sites of KBL are approximately 24 Å apart in the dimer. This corresponds well with the active sites in TDH of monomers A and B (or C and D), which are approximately 28 Å apart. One side of the KBL dimer, where the active sites are located, has a marked depression that appears to be complimentary to the generally convex surface of the TDH tetramer on the sides where the active sites of dimers A and B (or C and D) are exposed. This indeed suggests that a symmetric complex could be formed by the two enzymes to rapidly metabolise the unstable intermediate by channelling it between them. A complex could form with one KBL dimer binding to the AB dimer of TDH and a second KBL dimer binding to the CD dimer on the opposite face of the TDH tetramer, as shown in Figure 5.24.

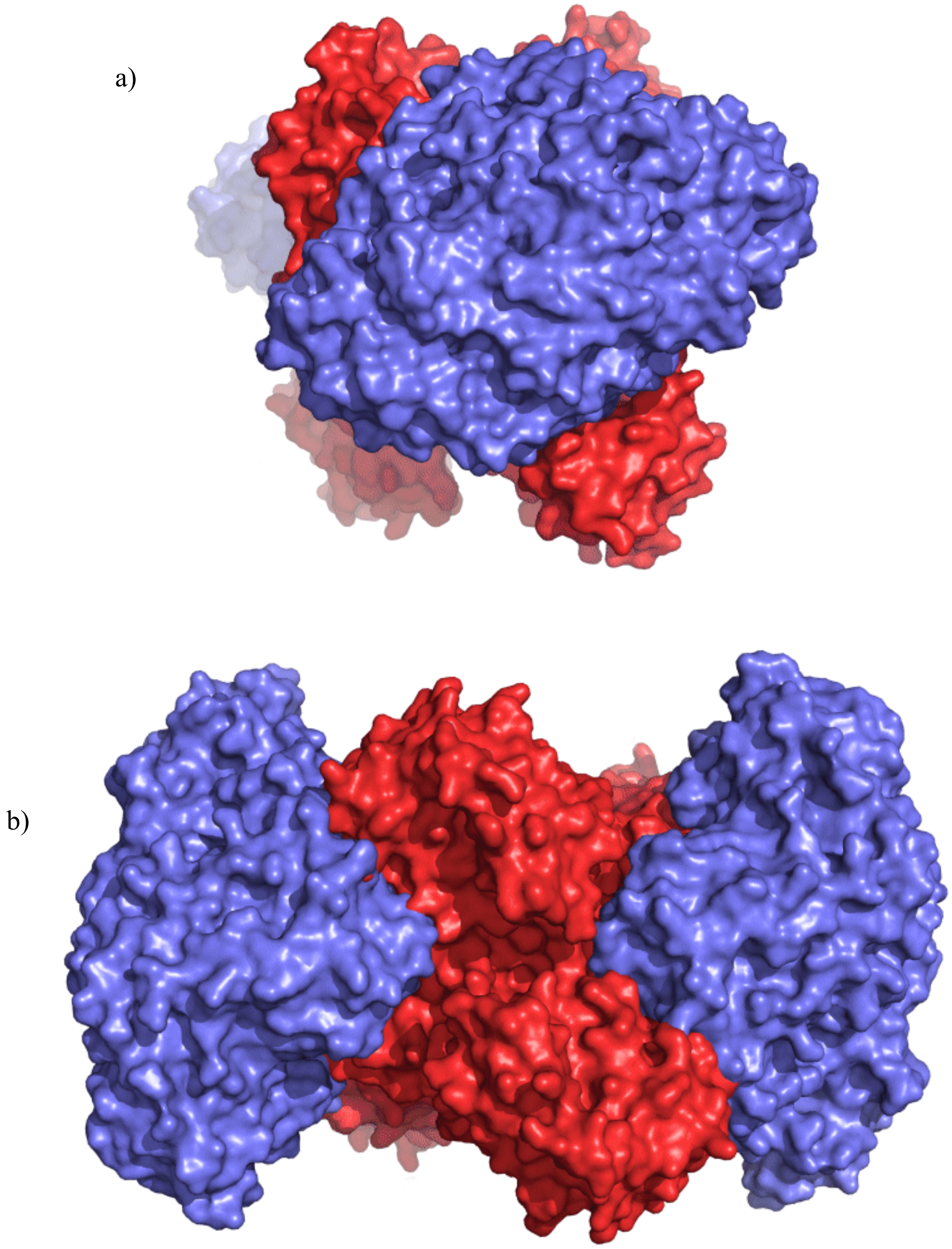


Figure 5.24: Putative docking of KBL with TDH
Two dimers of KBL have been docked with one tetramer of TkTDH, showing their proposed complex formation. Each of the two KBL dimers can dock with a dimer of TkTDH at a relative angle of approximately 45°, allowing them to fit onto opposite faces of the TDH tetramer. This brings the respective active sites into close contact (approximately 25 Å from one another). TDH is shown in red and KBL in blue. This putative complex is shown from the active site end of TDH (a) and a perpendicular view is shown in (b). The product from TDH would be able to travel 25 Å to the KBL active site and would be shielded from solvent.

An approximate model of the complex has been made by docking the two proteins manually and by the use of AutoDock (Morris *et al.* 1998) to optimise the fit. It appears that a dimer of KBL could open like a pair of jaws, to fit up against the central 12 stranded β-sheet of each TDH dimer in the tetramer. The KBL enzyme from *E.coli* has 41% sequence identity with that from *T.kodakaraensis*, similar to the level of identity between the TDHs from the two organisms, which is 43%. The SWISS-MODEL server (Arnold *et al.* 2006) was used to model the structure of KBL from *Thermococcus kodakaraensis* and the electrostatic potentials calculated, showing a large central region of positive charges on the face of the enzyme thought to bind TDH. An electrostatics model of TkTDH bound to NAD$^+$ showed a partial negatively charged surface at the face where it is thought KBL binds, suggesting that electrostatic interactions may play a role in the complex formation between TDH and KBL. Diagrams showing the electrostatic surfaces of these enzymes are shown in Figure 5.25.

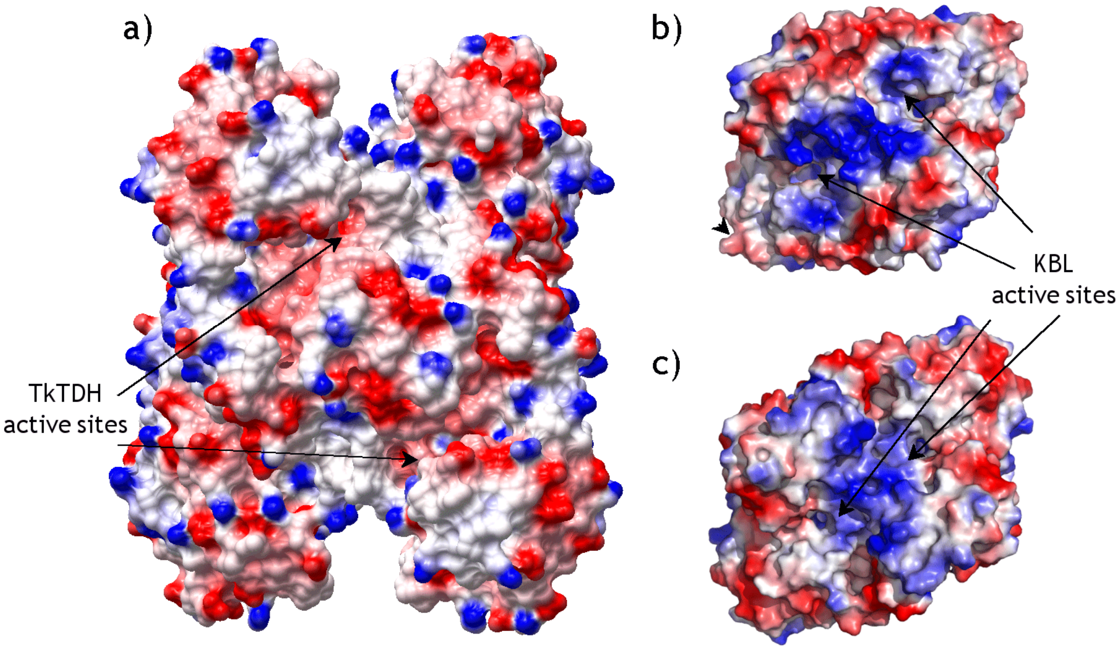


Figure 5.25: Electrostatic surface of TkTDH and KBL
The electrostatic surfaces of TDH and KBL. It can be seen from **(a)** that although in general the surface charges are relatively evenly distributed across the surface of the TDH protein, there is a region of negative charge in the vicinity of the two active sites at each end of the tetramer where it is proposed KBL binds. Conversely, the KBL proteins of both *E.coli* **(b)** and *T. kodakaraensis* **(c)** show regions of positive charge in the centre and across the active sites, suggesting electrostatic interactions may play a role in the formation of the TDH-KBL complex.

5.6.3 Mechanism of Action

Although our model of TkTDH does not contain either a structural or catalytic zinc ion, the model is still consistent with the proton relay system proposed by Higashi et al. (Higashi *et al.* 2008) as the reaction mechanism for TDHs. The triad of residues (Cys 42, His 67 and Glu 68) proposed to co-ordinate the catalytic zinc are highly conserved in both TDHs and ADHs, and in our model these residues do lie close enough to one another to co-ordinate a metal ion. However, no electron density was observed for an ion in this position in either our TkTDH model, or that published for PhTDH (Higashi *et al.* 2008). This Zn^{2+} is proposed to play a critical role in the deprotonation of the hydroxyl group of water or threonine substrate in the relay mechanism, so it is essential that efforts are made to determine whether this ion is truly present. The proposed proton relay mechanism for PhTDH is shown in Figure 5.26.

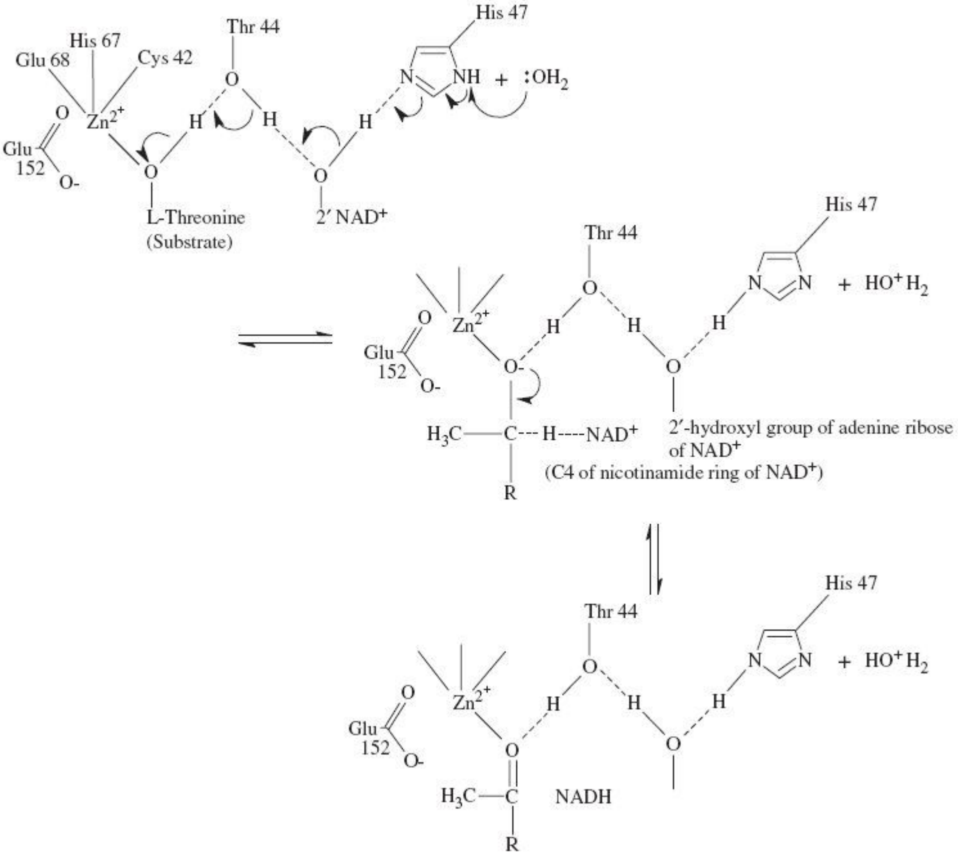


Figure 5.26: Proton relay mechanism for PhTDH
The catalytic zinc atom is supported by a catalytic triad (Cys42, His67 and Glu68) and an ionizable zinc-bound water molecule bonded to the hydroxyl group of Thr44 in the active site. Catalytic zinc has a critical role to deprotonate the hydroxyl group of water or threonine substrate, and then the deprotonation process is involved in the proton transfer through several residues.

Currently there exists conflicting evidence as to whether TDHs contain two zinc ions. Work done by Higashi et al 2008. supports the presence of two zinc ions in PhTDH, but biochemical studies carried out by Bashir et al. (unpublished work) concludes that only a single zinc ion is present in TkTDH. Further structural research must be carried out to determine whether a catalytic zinc ion is present. These could include further protein X-ray crystallographic studies on TDH crystals that have been grown in an excess of Zn^{2+} and in non-metal ion chelating conditions (such as excluding sodium citrate).

Chapter 6

Summary

6.1 Burkholderia pseudomallei invasion protein D

When work began on solving the structure of BipD, none of these Type Three Secretion System (TTSS) translocator proteins had been analysed in detail and relatively little information was known about them and how they fulfil their many proposed functions. During the time course of this PhD study, we have solved the structure of BipD from *Burkholderia pseudomallei* and structural models are now available for some of its functional homologues such as IpaD from *Shigella flexneri*, SipD from *Salmonella spp.*, LcrV from *Yersinia spp.*, and PcrV from *Pseudomonas aeruginosa*. Our 2.1 Å model of BipD is currently the highest resolution model available for this protein and has contributed to the general understanding of how TTSSs work and what role these translocators play. X-ray crystallographic structural models have enabled proposals to be made about putative quaternary structures for the protein, which include a homodimer, a heteropentamer and a homopentamer. These represent different possible forms of BipD, all of which may be correct depending on the environment in which the protein is functioning and the role that it is fulfilling. Studies are continuing to determine the quaternary structures of these translocators, and efforts are being made to establish exactly what role they play in bacterial infection of host cells. It is hoped that this structural information will reveal ways in which these translocators can be targeted to prevent invasion by the respective bacterial pathogens, thereby halting the spread of diseases.

6.2 2,4'-dihydroxyacetophenone dioxygenase

Although the structure of DAD has not yet been determined, many achievements have been made in expressing and purifying soluble protein and crystallisation conditions have been determined for the native enzyme, resulting in a low resolution dataset. It has also been possible to express and purify a selenomethionine derivative of DAD, which is necessary for structure determination using X-ray crystallography, to solve the phase problem. Unfortunately this derivative does not crystallise under the same conditions as the native protein, so crystallisation screens are being set up to determine suitable conditions. Upon the successful collection of a selenomethionine dataset it will be possible to determine the three-dimensional structure of DAD. Because this enzyme shares relatively little sequence homology with any other known dioxygenase, this structure should be particularly useful in determining how it is able to fulfil its function and cleave its substrate. Because this mechanism is unusual, it will be of benefit in improving our knowledge of how these enzymes function, as well as contributing to our overall understanding of bacterial catabolic pathways. It is hoped that eventually enough will be understood about how these enzymatic pathways function to enable the rational manipulation of enzymes to act on substrates other than the naturally occurring chemical. This would be a huge step forward in bioremediation protocols as it could allow pollutant hydrocarbons to be removed from the environment as harmless products.

6.3 L-threonine dehydrogenase

It is only recently that the three-dimensional structures of TDHs have elucidated, revealing a high degree of structural homology with alcohol dehydrogenases (ADH). Like the ADHs, it has been proposed that each monomer of TDH contains a structural and a catalytic zinc ion, however, neither our structure, nor the other published structure from *Pyrococcus horikoshii* are sufficient to answer this question. Our model contains neither of the zincs, and the PhTDH model contains only a structural zinc ion, but both enzymes have the conserved residues proposed to co-ordinate the metal ions, and these residues have been revealed to be in conformations that would enable them to fulfil this role. Further work is needed to solve the structures of these enzymes in conditions conducive to retaining any bound metal ions, as well as biochemical studies of their activity with and without zinc. The known TDH structures are all from thermophilic organisms, so it is difficult to determine which structural factors contribute to the proteins' thermostability as very high temperatures. For this to be possible, structures of non-thermophilic TDH proteins are needed such as that from *E. coli*, which we are also working on crystallising. This would allow a structural comparison between the two types and hopefully some differences could be identified. The structure of TkTDH has given us intriguing insights into a putative complex that the enzyme forms with 2-amino-3-ketobutyrate CoA ligase (KBL) and the structural nature of substrate channelling in the conversion of threonine to glycine that is carried out by these two enzymes. Modelling studies are continuing in the hope that a structurally possible complex can be found.

The structures of BipD and TkTDH have been solved and work is continuing on DAD. These achievements have made available a large amount of structural and functional information that would otherwise be inaccessible, contributing to our overall understanding of these macromolecules, as well as providing specific information that can be applied to particular challenges. It has also posed many more interesting and relevant questions that will be followed up to further elucidate how the structure of these proteins enables them to fulfil their biological function(s).

Appendix

UniProt accession numbers for sequences analysed by NeedleP global alignment with an Eblosum62 matrix:

B. pseudomallei:	*S. flexneri:*	*S. typhimurium:*
BipB – Q63K34	SipB – Q56019	IpaB – P18011
BipC – Q63K35	SipC – Q56020	IpaC – P18012
BipD – Q63K37	SipD – Q56136	IpaD – P18013
BopA – Q63K42	SopE2 – Q7CQD4	Icsb – P33546
BopE – Q63K41		

Buffers and Solutions:

Glycerol Stocks

0.3ml glycerol
0.7ml cells
dropped into liquid nitrogen
stored at -80ºC

Luria-Bertani Medium (1L)

10g bacto-tryptone
5g bacto-yeast extract
10g NaCl
dissolved in 1L H_2O

This was divided into 10ml aliquots and autoclaved. 5µl ampicillin (100mg/ml) was added before bacterial inoculation.

Ampicillin Agar Plates

LB medium was made as above and 15g bacterio-agar added. This was autoclaved, allowed to cool to 40ºC and 500µl ampicillin (100mg/ml) added. The plates were cooled, dried and stored at 4ºC.

(Novagen®) Overnight™ Express Instant TB Medium (1L)

The 60g sachet was poured into a large growth flask and dissolved in 1L H_2O and 10ml glycerol. The flask was autoclaved and 500µl ampicillin added.

Agarose DNA Gel

The gel casing was placed in the running tank to seal both ends. The gel was made up by microwaving the TAE buffer, H_2O and agarose to melt the agarose. This was allowed to cool to 50ºC and 5µl ethidium bromide was added. The molten solution was poured into the gel casing to a depth of 0.5cm, the comb inserted at one end, and allowed to set. The gel casing was rotated so the samples would run anode to cathode and covered with 1x TAE running buffer. The prepared samples were loaded onto the gel and run at 100mA until the dye was $^2/_3$ way along.

50x TAE Buffer
24.2g TRIS
5.71g glacial acetic acid
3.72g $Na_2EDTA\ 2H_2O$

Agarose Gel
2ml 50x TAE buffer
98ml H_2O
1g agarose

1x TAE Running Buffer
300ml H_2O
6ml 50x TAE buffer

DNA Sample Preparation
5µl 1kb DNA ladder + 3µl 10x Orange G + 2µl H_2O
1µl sample DNA + 3µl 10x Orange G + 6µl H_2O

SDS-PAGE Gel Electrophoresis 12%

The glass plates were cleaned firstly with H_2O then methanol, and the small and big plates clamped together with a separating strip at each outside edge. It was important to ensure the bottoms of the plates are level to avoid leaks. The separating solution was made up and cast between the plates and filled to the top with methanol to keep the gel level.

Once the gel set, the methanol was blotted off using filter paper. The stacking gel was made up cast immediately on top of the running gel. The comb was inserted and left to set. The gel was then clamped firmly in to the apparatus and the centre filled up with running buffer. 5µl Sigma® wide (M4038) molecular weight marker and 10µl samples (boiled and diluted 1:1 with LAMELI-sample buffer) were loaded into the wells. The tank was then filled to a depth of 3cm with running buffer and the gel run at 24mA for 40 minutes or until the samples reached the bottom. The gel was removed from the plates and left in coomassie blue gel stain for 15 minutes. This was then placed in destain until the bands were clearly visible.

SDS-PAGE running buffer 5X
7.5g TRIS base 0.125M
36g glycerine
2.5g SDS (0.5 w/v)
dH_20 to 500ml
titrated to pH 8.3
Store @ 4ºC

4x TRIS Separating Buffer pH 8.8
91g TRIS base (1.5M) in 300ml
H_20 adjust to pH8.8 with 1M HCl
Add dH_20 up to 500ml total volume
Filtered through 0.45ml filter paper
Add 2g SDS (0.4% w/v)

4x TRIS Stacking Buffer pH 6.8
6.05g TRIS base (0.5M) in 40ml
H_20 adjust to pH 6.8 with 1M HCl
Add dH_20 up to 100ml total volume
Filtered through 0.45ml filter paper
Add 0.4g SDS (0.4% w/v)

12% Separating Gel
3ml 30% Acrylamide/ N,N1-methylene- bis-acrylamide
1.87ml TRIS pH 8.8
2.6 ml dH_20
50ml 10% APS
10ml TEMED

12 % Stacking Gel
0.65ml 30% Acrylamide/ N,N1-methylene- bis-acrylamide
1.25ml TRIS pH 6.8
3.05ml dH_20
50ml 10% APS
10ml TEMED

LAMELI-Sample buffer
8ml of 50% glycerol
2.5% of 1M TRIS pH 6.8
0.8g of SDS
2ml β-mecaptoethanol
3mg bromophenol blue
7.5ml dH_20

Coomassie Blue Gel Stain
50% (w/v) Tricholoroacetic acid
0.25% (w/v) Brilliant Blue

Destain
30% methanol
10% acetic acid

GSTrap Affinity Column Buffer Solutions

Binding Buffer (phospho-buffered saline)
140nM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4

Elution Buffer
50mM TRIS-HCl, 10mM reduced glutathione pH 8.0

Solutions for Selenomethionine Growth

M9 Minimal Medium (1L)
2g NH_4Cl
6g KH_2PO_4
13.6g Na_2HPO_4
1g $MgSO_4$
40mg $FeSO_4 . 7H_2O$

This solution was autoclaved in 2l conical growth flasks and 12ml 20% (8g in 20ml) filter sterilised (0.20µm) glucose was added.

Amino Acid Mix 1 (20ml)
80mg of each of 16aa (Ala, Asn, Asp, Arg, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Pro, Ser, Thr, Val)
=4mg /ml
Filter sterilise and store at -20ºC

Amino Acid Mix 2 (20ml)
80mg of each of Tyr, Trp, Phe in 20ml H_2O
adjust pH to 8.0 with NaOH
=4mg /ml
Filter sterilise and store at -20ºC

Vitamins (2ml)
2ml H_2O
2mg – riboflavin, nicitinamide, thiamine and pyridoxine monohydrochloride
=1mg / ml
Filter sterilise and store at -20ºC

Selenomethionine (20ml)
80mg L-selenomethionine in 20ml H_2O
=10mg / ml
Filter sterilise and store at -20ºC

Reference List

Adams M. W. W. and Kelly R. M. (1998) Finding and using hyperthermophilic enzymes. *Trends in Biotechnology* **16,** 329-332.

Adams P. D., Grosse-Kunstleve R. W., Hung L. W., Ioerger T. R., McCoy A. J., Moriarty N. W., Read R. J., Sacchettini J. C., Sauter N. K. and Terwilliger T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallographica Section D* **58,** 1948-1954.

Aldhous P. (2005) Melioidosis? Never heard of it... *Nature* **434,** 692-693.

Amersham Biosciences (2007a) Affinity Chromatography, Principle and Methods.

Amersham Biosciences (2007b) Gel Filtration, Principles and Methods.

Amersham Biosciences (2007c) Ion Exchange Chromatography & Chromatofocusing, Principles and Methods.

Andrews K. T. and Patel K. C. (1998) FgoI, a Type II restriction endonuclease from the thermoanaerobe Fervidobacterium gondwanense. *Anaerobe* **4,** 227-232.

Aoyama Y. and Motokawa Y. (1981) L-Threonine dehydrogenase of chicken liver. Purification, characterization, and physiological significance. *J. Biol. Chem.* **256,** 12367-12373.

Arnold K., Bordoli L., Kopp J. and Schwede T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195-201.

Atomi H., Fukui T., Kanai T., Morikawa M. and Imanaka T. (2004) Description of *Thermococcus kodakaraensis sp.* nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus sp.* KOD1. *Archaea* **1,** 263-267.

Attree O. and Attree I. (2001) A second type III secretion system in Burkholderia pseudomallei: who is the real culprit? *Microbiology* **147,** 3197-3199.

Bada J. L. (1991) Amino Acid Cosmogeochemistry. *Philosophical Transactions: Biological Sciences* **333,** 349-358.

Barton G. J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6,** 37-40.

Bird M. I. and Nunn P. B. (1983) Metabolic homoeostasis of L-threonine in the normally-fed rat. Importance of liver threonine dehydrogenase activity. *Biochem. J.* **214,** 687-694.

Blocker A., Holden D. and Cornelis G. (2000) Type III secretion systems: what is the translocator and what is translocated? *Cell Microbiol* **2,** 387-390.

Blocker A., Komoriya K. and Aizawa S. (2003) Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proc. Natl. Acad. Sci. U. S. A* **100,** 3027-3030.

Bogin O., Peretz M. and Burstein Y. (1997) Thermoanaerobacter brockii alcohol dehydrogenase: Characterization of the active site metal and its ligand amino acids. *Protein Sci* **6,** 450-458.

Bonisch H., Backmann J., Kath T., Naumann D. and Schafer G. (1996) Adenylate Kinase from Sulfolobus acidocaldarius:Expression in Escherichia coliand Characterization by Fourier Transform Infrared Spectroscopy. *Archives of Biochemistry and Biophysics* **333,** 75-84.

Borras T., Persson B. and Joernvall H. (1989) Eye lens .zeta.-crystallin relationships to the family of "long-chain" alcohol/polyol dehydrogenases. Protein trimming and conservation of stable parts. *Biochemistry* **28,** 6133-6139.

Bottoms C. A., Smith P. E. and Tanner J. J. (2002) A structurally conserved water molecule in Rossmann dinucleotide-binding domains. *Protein Sci* **11,** 2125-2137.

Bowyer A., Mikolajek H., Wright J. N., Coker A., Erskine P. T., Cooper J. B., Bashir Q., Rashid N., Jamil F. and Akhtar M. (2008) Crystallization and preliminary X-ray diffraction analysis of l-threonine dehydrogenase (TDH) from the hyperthermophilic archaeon Thermococcus kodakaraensis. *Acta Crystallographica Section F* **64,** 828-830.

Boylan S. A. and Dekker E. E. (1981) L-threonine dehydrogenase. Purification and properties of the homogeneous enzyme from Escherichia coli K-12. *J. Biol. Chem.* **256,** 1809-1815.

Breitbach K., Rottner K., Klocke S., Rohde M., Jenzora A., Wehland J. and Steinmetz I. (2003) Actin-based motility of Burkholderia pseudomallei involves the Arp 2/3 complex, but not N-WASP and Ena/VASP proteins. *Cellular Microbiology* **5,** 385-393.

Brody J. R., Calhoun E. S., Gallmeier E., Creavalle T. D. and Kern S. E. (2004) Ultra-fast high-resolution agarose electrophoresis of DNA and RNA using low-molarity conductive media. *Biotechniques* **37,** 598, 600, 602.

Brunger A. T., Adams P. D., Clore G. M., DeLano W. L., Gros P. and Grosse-Kunstleve R. W. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallographica. sect D* **54,** 905-921.

Bugg T. D. H. (2003) Dioxygenase enzymes: catalytic mechanisms and chemical models. *Tetrahedron* **59,** 7075-7101.

Bugg T. D. H. and Lin G. (2001) Solving the riddle of the intradiol and extradiol catechol dioxygenases: how do enzymes control hydroperoxide rearrangements? *Chemical Communications* 941-952.

Bullough P. A., Hughson F. M., Skehel J. J. and Wiley D. C. (1994) Structure of influenze hemagglutinin at the pH of membrane-fusion. *Nature* **371,** 37-43.

CCP4 (1994) The CCP4 Suite: programs for protein crystallography. *Acta Crystallographica. sect D* **50,** 760-763.

Chaplin M. and Bucke C. (2007) website based on the book Enzyme Technology (http://www.lsbu.ac.uk/biology/enztech/index.html).

Cheng A. C. (2005) Melioidosis: Epidemiology, Pathophysiology, and Management. *Clin Microbiol Rev.* **18 (2),** 383-416.

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.

Cripps R. E., Trudgill P. W. and Whateley J. G. (1978) The metabolism of 1-phenylethanol and acetophenone by Nocardia T5 and an Arthrobacter species. *Eur. J. Biochem.* **86,** 175-186.

Dale R. A. (1978) Catabolism of threonine in mammals by coupling of -threonine 3-dehydrogenase with 2-amino-3-oxobutyrate-CoA ligase. *Biochimica et Biophysica Acta (BBA) - General Subjects* **544,** 496-503.

Darby J. M., Taylor D. G. and Hopper D. J. (1987) Hydroquinone as the Ring-Fission Substrate in the Catabolism of 4-Ethylphenol and 4-Hydroxyacetophenone by Pseudomonas-Putida Jd1. *Journal of General Microbiology* **133,** 2137-2146.

DeLano W. L. (2008) The PyMOL Molecular Graphics System. *DeLano Scientific LLC.*

Diaz E. (2006) Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. *Int. Microbiol.* **7,** 173-181.

Diaz E., Ferrandez A., Prieto M. A. and Garcia J. L. (2001) Biodegradation of Aromatic Compounds by Escherichia coli. *Microbiol. Mol. Biol. Rev.* **65,** 523-569.

Druar C., Yu F., Barnes J. L., Okinaka R. T., Chantratia N., Beg S., Stratilo C. W., Olive A. J., Soltes G., Russell M. L., Limmathurotsakul D., Norton R. E., Ni S. X., Picking W. D., Jackson P. J., Stewart D. I. H., Tsventnitsky V., Picking W. L., Cherwonogrodzky J. W., Ketheesan N., Peacock S. J. and Wiersma E. J. (2007) Evaluating *Burkholderia pseudomallei* Bip proteins as vaccines and Bip antibodies as detection agents. *FEMS Immunology & Medical Microbiology* **52,** 78-87.

Edgar A. (2002) Molecular cloning and tissue distribution of mammalian L-threonine 3-dehydrogenases. *BMC Biochemistry* **3,** 19.

Emsley P. and Cowtan K. (2007) Coot: model-building tools for molecular graphics. *Acta Crystallographica. sect D* **60,** 2126-2132.

Erskine P. T., Knight M. J., Ruaux A., Mikolajek H., Sang N. W., Withers J., Gill R., Wood S. P., Wood M., Fox G. C. and Cooper J. B. (2006) High Resolution Structure of BipD: An Invasion Protein Associated with the Type III Secretion System of Burkholderia Pseudomallei. *J. Mol. Biol.* **363,** 125-136.

Espina M., Ausar S. F., Middaugh C. R., Baxter M. A., Picking W. D. and Picking W. L. (2007) Conformational stability and differential structural analysis of LcrV, PcrV, BipD, and SipD from type III secretion systems. *Protein Sci.* **16,** 704-714.

Espina M., Olive A. J., Kenjale R., Moore D. S., Ausar S. F., Kaminski R. W., Oaks E. V., Middaugh C. R., Picking W. D. and Picking W. L. (2006) IpaD localizes to the tip of the type III secretion system needle of Shigella flexneri. *Infect Immun* **74,** 4391-4400.

Esposito L., Sica F., Raia C. A., Giordano A., Rossi M., Mazzarella L. and Zagari A. (2002) Crystal structure of the alcohol dehydrogenase from the hyperthermophilic archaeon Sulfolobus solfataricus at 1.85 A resolution. *J. Mol. Biol.* **318,** 463-477.

Evans P. R. (2006) Scaling and assessment of data quality. *Acta Crystallographica. sect D* **62,** 72-82.

Evans W. C. and Fuchs G. (1988) Anaerobic Degradation of Aromatic Compounds. *Annual Review of Microbiology* **42,** 289-317.

Fernandez-Prada C. M., Hoover D. L., Tall B. D., Hartman A. B., Kopelowitz J. and Venkatesan M. M. (2000) Shigella flexneri IpaH(7.8) facilitates escape of virulent bacteria from the endocytic vacuoles of mouse and human macrophages. *Infect. Immun.* **68,** 3608-3619.

Finlay B. B. and Falkow S. (1989) Common Themes in Microbial Pathogenicity. *Microbiological Reviews* **53,** 210-230.

Friedman M. (1996) Nutritional Value of Proteins from Different Food Sources. A Review. *J. Agric. Food Chem.* **44,** 6-29.

Fukui T., Atomi H., Kanai T., Matsumi R., Fujiwara S. and Imanaka T. (2005) Complete genome sequence of the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1 and comparison with Pyrococcus genomes. *Genome Res.* **15,** 352-363.

Gan Y.-H. (2005) Interaction between *Burkholderia pseudomallei* and the Host Immune Response: Sleeping with the Enemy? *JID* **2005:192,** 1845-1850.

Gershenson A., Schauerte J. A., Giver L. and Arnold F. H. (2000) Tryptophan Phosphorescence Study of Enzyme Flexibility and Unfolding in Laboratory-Evolved Thermostable Esterases. *Biochemistry* **39,** 4658-4665.

Godoy D., Randle G., Simpson A. J., Aanensen D. M., Pitt T. L., Kinoshita R. and Spratt B. G. (2003) Multilocus Sequence Typing and Evolutionary Relationships among the Causative Agents of Melioidosis and Glanders, Burkholderia pseudomallei and Burkholderia mallei. *J. Clin. Microbiol.* **41,** 2068-2079.

Goedhart J. and Gadella T. W., Jr. (2005) Analysis of oligonucleotide annealing by electrophoresis in agarose gels using sodium borate conductive medium. *Anal. Biochem.* **343,** 186-187.

Guy J. E., Isupov M. N. and Littlechild J. A. (2003) The Structure of an Alcohol Dehydrogenase from the Hyperthermophilic Archaeon Aeropyrum pernix. *Journal of Molecular Biology* **331,** 1041-1051.

Han S., Eltis L. D., Timmis K. N., Muchmore S. W. and Bolin J. T. (1995) Crystal structure of the biphenyl-cleaving extradiol dioxygenase from a PCB-degrading pseudomonad. *Science* **270,** 976-980.

Harpel M. R. and Lipscomb J. D. (1990) Gentisate 1,2-dioxygenase from Pseudomonas. Substrate coordination to active site Fe2+ and mechanism of turnover. *J. Biol. Chem.* **265,** 22187-22196.

Harwood C. S. and Parales R. E. (1996) The beta-ketoadipate pathway and the biology of self-identity. *Annu. Rev. Microbiol.* **50,** 553-590.

Hayaishi O., Katagiri M. and Rothberg S. (1955) Mechanism Of The Pyrocatechase Reaction. *J. Am. Chem. Soc* **77,** 5450-5451.

Henderson I. R., Navarro-Garcia F., Desvaux M., Fernandez R. C. and Ala'Aldeen D. (2004) Type V Protein Secretion Pathway: the Autotransporter Story. *Microbiol. Mol. Biol. Rev.* **68,** 692-744.

Higashi N., Tanimoto K., Nishioka M., Ishikawa K. and Taya M. (2008) Investigating a catalytic mechanism of hyperthermophilic L-threonine dehydrogenase from *Pyrococcus horikoshii. J Biochem* **144,** 77-85.

High S. and Dobberstein B. (1992) Mechanisms that determine the transmembrane disposition of proteins. *Curr. Opin. Cell Biol.* **4,** 581-586.

Holden M. T. G., Titball R. W., Peacock S. J., Cerdeno-Tarraga A. M., Atkins T., Crossman L. C., Pitt T., Churcher C., Mungall K., Bentley S. D., Sebaihia M., Thomson N. R., Bason N., Beacham I. R., Brooks K., Brown K. A., Brown N. F., Challis G. L., Cherevach I., Chillingworth T., Cronin A., Crossett B., Davis P., DeShazer D., Feltwell T., Fraser A., Hance Z., Hauser H., Holroyd S., Jagels K., Keith K. E., Maddison M., Moule S., Price C., Quail M. A., Rabbinowitsch E., Rutherford K., Sanders M., Simmonds M., Songsivilai S., Stevens K., Tumapa S., Vesaratchavest M., Whitehead S., Yeats C., Barrell B. G., Oyston P. C. F. and Parkhill J. (2004) Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. *PNAS* **101,** 14240-14245.

Hopper D. J. (1986) Oxygenase Properties of the (4-Hydroxybenzoyl)Methanol-Cleavage Enzyme from An Alcaligenes Sp. *Biochem. J.* **239,** 469-472.

Hopper D. J. and Kaderbhai M. A. (1999) 2,4'-Dihydroxyacetophenone dioxygenase (EC 1.13.11.41) from Alcaligenes sp. 4HAP: a novel enzyme with an atypical dioxygenase sequence. *Biochem. J* **344 Pt 2,** 397-402.

Hueck C. J. (1998) Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. *Microbiol. Mol. Biol. Rev.* **62,** 379-433.

Inglis T. J., Rigby P., Robertson T. A., Dutton N. S., Henderson M. and Chang B. J. (2000) Interaction between Burkholderia pseudomallei and Acanthamoeba species results in coiling phagocytosis, endamebic bacterial survival, and escape. *Infect. Immun.* **68,** 1681-1686.

Ishikawa K., Higashi N., Nakamura T., Matsuura T. and Nakagawa A. (2007) The First Crystal Structure of L-Threonine Dehydrogenase. *Journal of Molecular Biology* **366,** 857-867.

Itoh T. (2003) Taxonomy of nonmethanogenic hyperthermophilic and related thermophilic archaea. *Journal of Bioscience and Bioengineering* **96,** 203-212.

Jaenicke R. and Bohm G. (1998) The stability of proteins in extreme environments. *Current Opinion in Structural Biology* **8,** 738-748.

Jelokova J. S., Karlsson C., Estonius M., Jornvall H. and Hoog J. (1994) Features of Structural Zinc in Mammalian Alcohol Dehydrogenase. *Eur. J. Biochem.* **225,** 1015-1019.

Johnson S., Roversi P., Espina M., Olive A., Deane J. E., Birket S., Field T., Picking W. D., Blocker A. J., Galyov E. E., Picking W. L. and Lea S. M. (2006) Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD. *J. Biol. Chem.* M607945200.

Jones D. T. (2001) Evaluating the potential of using fold-recognition models for molecular replacement. *Acta Crystallographica. sect D* **57,** 1428-1434.

Kaniga K., Trollinger D. and Galan J. E. (1995) Identification of two targets of the type III protein secretion system encoded by the inv and spa loci of Salmonella typhimurium that have homology to the Shigella IpaD and IpaA proteins. *J Bacteriol.* **177,** 7078-7085.

Karlsson A., Beharry Z. M., Matthew E. D., Coulter E. D., Neidle E. L., Kurtz D. M., Jr., Eklund H. and Ramaswamy S. (2002) X-ray crystal structure of benzoate 1,2-dioxygenase reductase from Acinetobacter sp. strain ADP1. *J. Mol. Biol.* **318,** 261-272.

Kashefi K. and Lovley D. R. (2003) Extending the Upper Temperature Limit for Life. *Science* **301,** 934.

Kazuoka T., Takigawa S., Arakawa N., Hizukuri Y., Muraoka I., Oikawa T. and Soda K. (2003) Novel Psychrophilic and Thermolabile L-Threonine Dehydrogenase from Psychrophilic Cytophaga sp. Strain KUC-1. *J. Bacteriol.* **185,** 4483-4489.

Kita A., Kita S., Fujisawa I., Inaka K., Ishida T., Horiike K., Nozaki M. and Miki K. (1999) An archetypical extradiol-cleaving catecholic dioxygenase: the crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from Ppseudomonas putida mt-2. *Structure.* **7,** 25-34.

Koponen M. A., Zlock D., Palmer D. L. and Merlin T. L. (1991) Melioidosis. Forgotten, but not gone! *Arch Intern Med* **151,** 605-608.

Korkhin Y., Kalb G., Peretz M., BOGIN O., Burstein Y. and Frolow F. (1998) NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of Clostridium beijerinckii and Thermoanaerobacter brockii. *Journal of Molecular Biology* **278,** 967-981.

Laskowski R. A., MacArthur M. W., Moss D. S. and Thornton J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog* **26,** 283-291.

Leahy J. G. and Colwell R. R. (1990) Microbial degradation of hydrocarbons in the environment. *Microbiol. Mol. Biol. Rev.* **54,** 305-315.

LeBrun L. A., Park D. H., Ramaswamy S. and Plapp B. V. (2004) Participation of Histidine-51 in Catalysis by Horse Liver Alcohol Dehydrogenase. *Biochemistry* **43,** 3014-3026.

Leskovac V., Trivic S. and Latkovska M. (1976) State and accessibility of zinic in yeast alcohol dehydrogenase. *Biochem J* **155,** 155-161.

Leslie A. G. W. (2006) The integration of macromolecular diffraction data. *Acta Crystallographica. sect D* **62,** 48-57.

Levitt D. G. (2001) A new software routine that automates the fitting of protein X-ray crystallographic electron density maps. *Acta Crystallographica. sect D* **57,** 1013-1019.

Lin G., Reid G. and Bugg T. D. (2001) Extradiol oxidative cleavage of catechols by ferrous and ferric complexes of 1,4,7-triazacyclononane: insight into the mechanism of the extradiol catechol dioxygenases. *J. Am. Chem. Soc.* **123,** 5030-5039.

Liu Q. and Shi X. (2000) [Construction of prokaryotic expression vector of hALR and its expression in E.Coli]. *Zhonghua Gan Zang. Bing. Za Zhi.* **8,** 9-11.

Luo J. and Barany F. (1996) Identification of essential residues in Thermus thermophilus DNA ligase. *Nucleic Acids Res.* **24,** 3079-3085.

Machielsen R. and van der Oost J. (2006) Production and characterization of a thermostable L-threonine dehydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus. *FEBS Journal* **273,** 2722-2729.

Manco G., GiosuΦ E., D'Auria S., Herman P., Carrea G. and Rossi M. (2000) Cloning, Overexpression, and Properties of a New Thermophilic and Thermostable Esterase with Sequence Similarity to Hormone-Sensitive Lipase Subfamily from the Archaeon Archaeoglobus fulgidus. *Archives of Biochemistry and Biophysics* **373,** 182-192.

Marcus J. P. and Dekker E. E. (1993) Threonine formation via the coupled activity of 2-amino-3-ketobutyrate coenzyme A lyase and threonine dehydrogenase. *J. Bacteriol.* **175,** 6505-6511.

Marlovits T. C., Kubori T., Sukhan A., Thomas D. R., Galan J. E. and Unger V. M. (2004) Structural insights into the assembly of the type III secretion needle complex. *Science* **306,** 1040-1042.

Mays E. E. and Ricketts E. A. (1975) Melioidosis: recrudescence associated with bronchogenic carcinoma twenty-six years following initial geographic exposure. *Chest* **68,** 261-263.

McCoy A. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D* **63,** 32-41.

Mecsas J. and Strauss E. J. (1996) Molecular Mechanisms of Bacterial Virulence: Type III Secretion and Pathogenicity Islands. *Emerging Infectious Diseases* **2,** 271-288.

Menard R., Sansonetti P. J. and Parsot C. (1994) The secretion of the Shigella flexneri Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD. *The EMBO Journal* **13,** 5293-5302.

Miller M. A. and Lipscomb J. D. (1996) Homoprotocatechuate 2,3-dioxygenase from Brevibacterium fuscum. A dioxygenase with catalase activity. *J. Biol. Chem.* **271,** 5524-5535.

Moffatt B. A. and Studier F. W. (1987) T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49,** 221-227.

Morikawa M., Izawa Y., Rashid N., Hoaki T. and Imanaka T. (1994) Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic Pyrococcus sp. *Appl. Environ. Microbiol.* **60,** 4559-4566.

Morris G. M., Goodsell D. S., Halliday R. S., Huey R., Hart W. E., Belew R. K. and Olson A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* **19,** 1639-1662.

Mota L. J., Sorg I. and Cornelis G. R. (2005) Type III secretion: the bacteria-eukaryotic cell express. *FEMS Microbiol Lett.* **252,** 1-10.

Murzin A. G., Brenner S. E., Hubbard T. and Chothia C. (1995) A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *J. Mol. Biol.* **247,** 536-540.

Pettersson J., Nordfelth R., Dubinina E., Bergman T., Gustafsson M., Magnusson K. E. and Wolf-Watz H. (1996) Modulation of Virulence Factor Expression by Pathogen Target Cell Contact. *Science* **273,** 1231-1233.

Picking W. L., Nishioka H., Hearn P. D., Baxter M. A., Harrington A. T., Blocker A. and Picking W. D. (2005) IpaD of Shigella flexneri Is Independently Required for Regulation of Ipa Protein Secretion and Efficient Insertion of IpaB and IpaC into Host Membranes. *Infect. Immun.* **73,** 1432-1440.

Potter R., Kapoor V. and Newman E. B. (1977) Role of threonine dehydrogenase in Escherichia coli threonine degradation. *J. Bacteriol.* **132,** 385-391.

Potterton L., McNicholas S., Krissinel E., Gruber J., Cowtan K., Emsley P., Murshudov G. N., Cohen S., Perrakis A. and Noble M. (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallographica Section D* **60,** 2288-2294.

Que L., Jr. and Ho R. Y. (1996) Dioxygen Activation by Enzymes with Mononuclear Non-Heme Iron Active Sites. *Chem. Rev.* **96,** 2607-2624.

Rainbow L., Hart A. C. and Winstanley C. (2002) Distribution of type III secretion gene clusters in Burkholderia pseudomallei, B. thailandensis and B. mallei. *J Med Microbiol* **51,** 374-384.

Ramaswamy S., Park D. H. and Plapp B. V. (1999) Substitutions in a Flexible Loop of Horse Liver Alcohol Dehydrogenase Hinder the Conformational Change and Unmask Hydrogen Transfer. *Biochemistry* **38,** 13951-13959.

Ray M. and Ray S. (1984) L-Threonine Dehydrogenase from Goat Liver. *J. Biol. Chem.* **260,** 5913-5918.

Read R. J. (2005) Protein crystallisation theory. *http://www-structmed. cimr. cam. ac. uk/Course/Crystals/Theory/phase_methods. html.*

Rescigno M. and Perham R. N. (1994) Structure of the NADPH-binding motif of glutathione reductase: efficiency determined by evolution. *Biochemistry* **33,** 5721-5727.

Rossman M. G. and Liljas A. (1974) Recognition of Strucutral Domains in Globular Proteins. *Journal of Molecular Biology* **85,** 177-181.

Sansonetti P. J. (2001) Rupture, invasion and inflammatory destruction of the intestinal barrier by Shigella, making sense of prokaryote-eukaryote cross-talks. *FEMS Microbiol. Rev.* **25,** 3-14.

Schmidt A., Sivaraman J., Li Y., Larocque R., Barbosa J. A., Smith C., Matte A., Schrag J. D. and Cygler M. (2001) Three-dimensional structure of 2-amino-3-ketobutyrate CoA ligase from Escherichia coli complexed with a PLP-substrate intermediate: inferred reaction mechanism. *Biochemistry* **40,** 5151-5160.

Schroder G. and Lanka E. (2005) The mating pair formation system of conjugative plasmids-A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* **54,** 1-25.

Sekiya K., Ohishi M., Ogino T., Tamano K., Sasakawa C. and Abe A. (2001) Supermolecular structure of the enteropathogenic Escherichia coli type III secretion system and its direct interaction with the EspA-sheath-like structure. *Proc. Natl. Acad. Sci. U. S. A* **98,** 11638-11643.

Sheldrick G. M. and Schneider T. R. (1997) SHELXL: high-resolution refinement. *Methods Enzymol* **277,** 319-343.

Stevens M. P. and Galyov E. E. (2004) Exploitation of host cells by Burkholderia pseudomallei. *International Journal of Medical Microbiology* **293,** 549-555.

Stevens M. P., Haque A., Atkins T., Hill J., Wood M. W., Easton A., Nelson M., Underwood-Fowler C., Titball R. W., Bancroft G. J. and Galyov E. E. (2004) Attenuated virulence and protective efficacy of a Burkholderia pseudomallei bsa type III secretion mutant in murine models of melioidosis. *Microbiology* **150,** 2669-2676.

Stevens M. P., Stevens J. M., Jeng R. L., Taylor L. A., Wood M. W., Hawes P., Monaghan P., Welch M. D. and Galyov E. E. (2005) Identification of a bacterial factor required for actin-based motility of Burkholderia pseudomallei. *Molecular Microbiology* **56,** 40-53.

Stevens M. P., Wood M. W., Taylor L. A., Monaghan P., Hawes P., Jones P. W., Wallis T. S. and Galyov E. E. (2002) An Inv/Mxi-Spa-like type III protein secretion system in Burkholderia pseudomallei modulates intracellular behaviour of the pathogen. *Molecular Microbiology* **46,** 649-659.

Studier F. W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology* **219,** 37-44.

Studier F. W. and Moffatt B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **189,** 113-130.

Sulzenbacher G., Alvarez K., Van Den Heuvel R. H., Versluis C., Spinelli S., Campanacci V., Valencia C., Cambillau C., Eklund H. and Tegoni M. (2004) Crystal structure of E.coli alcohol dehydrogenase YqhD: evidence of a covalently modified NADP coenzyme. *J. Mol. Biol.* **342,** 489-502.

Suparak S., Kespichayawattana W., Haque A., Easton A., Damnin S., Lertmemongkolchai G., Bancroft G. J. and Korbsrisate S. (2005) Multinucleated Giant Cell Formation and Apoptosis in Infected Host Cells Is Mediated by Burkholderia pseudomallei Type III Secretion Protein BipB. *J. Bacteriol.* **187,** 6556-6560.

Suputtamongokol Y., Hall A. J., Dance D. A. B., Chaowagul W., Rajhanuvong A., Smith M. D. and White N. J. (1994) The Epidemiology of Melioidosis in Ubon Ratchatani, Northeast Thailand. *Int. J. Epidemiol.* **23,** 1082-1090.

Takahashi M., Yamaguchi E. and Uchida T. (1984) Thermophilic DNA ligase. Purification and properties of the enzyme from Thermus thermophilus HB8. *J. Biol. Chem.* **259,** 10041-10047.

Takai K., Sugai A., Itoh T. and Horikoshi K. (2000) Palaeococcus ferrophilus gen. nov., sp. nov., a barophilic, hyperthermophilic archaeon from a deep-sea hydrothermal vent chimney. *Int J Syst Evol Microbiol* **50,** 489-500.

Tanner A. and Hopper D. J. (2000) Conversion of 4-hydroxyacetophenone into 4-phenyl acetate by a flavin adenine dinucleotide-containing Baeyer-Villiger-type monooxygenase. *J. Bacteriol.* **182,** 6565-6569.

Taylor W. H. and Hagerman P. J. (1990) Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure : II. NaCl-dependence of DNA flexibility and helical repeat. *Journal of Molecular Biology* **212,** 363-376.

Terwilliger T. C. and Berendzen J. (1999) Automated MAD and MIR structure solution. *Acta Crystallographica. sect D* **55,** 849-861.

Torrii Lab U. o. T. (2008) Hanging-drop crystallisation method. *www. dt. k. u-tokyo. ac. jp/research/PC_2. htm.*

Tressel T., Thompson R., Zieske L. R., Menendez M. I. and Davis L. (1986) Interaction between L-threonine dehydrogenase and aminoacetone synthetase and mechanism of aminoacetone production. *J. Biol. Chem.* **261,** 16428-16437.

Turbyfill K. R., Mertz M. A., Mallet C. P. and Oaks E. V. (1998) Identification of epitope and surface-exposed domains of *Shigella flexneri* invasion plasmid antigen D. *Infect. Immun.* **66,** 1999-2006.

Unsworth L. D., Van der Oost J. and Koutsopoulos S. (2007) Hyperthermophilic enzymes;stability, activity and implementation strategies for high temperature applications. *FEBS Journal* **274,** 4044-4056.

Vagin A., Steiner R. A., Lebedev A. A., Potterton L., McNicholas S., Long F. and Murshudov G. N. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60,** 2184-2195.

Vagin A. and Teplyakov A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30,** 1022-1025.

Viboud G. I. and Bliska J. B. (2005) Yersinia outer proteins: role in modulation of host cell signaling responses and pathogenesis. *Annu. Rev. Microbiol.* **59,** 69-89.

Vieille C. and Zeikus G. J. (2001) Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiol. Mol. Biol. Rev.* **65,** 1-43.

Wheeler W. C. and Gladstein D. S. (1994) MALIGN: A Multiple Sequence Alignment Program. *J Hered* **85,** 417-418.

White N. J. (2003) Melioidosis. *The Lancet* **361,** 1715-1722.

Williams R. J. (2003) Restriction endonucleases: classification, properties, and applications. *Mol. Biotechnol.* **23,** 225-243.

Wu R., Skaar E. P., Zhang R., Joachimiak G., Gornicki P., Schneewind O. and Joachimiak A. (2005) Staphylococcus aureus IsdG and IsdI, heme-degrading enzymes with structural similarity to monooxygenases. *J. Biol. Chem.* **280,** 2840-2846.

Wuthiekanun V., Smith M. D., Dance D. A. and White N. J. (1995) Isolation of Pseudomonas pseudomallei from soil in north-eastern Thailand. *Trans R Soc Trop Med Hyg.* **89(1),** 41-43.

Yip C. K. and Strynadka N. C. J. (2006) New structural insights into the bacterial type III secretion system. *TIBS* **31,** 223-230.

Zavodszky P., Kardos J., Svingor A. and Petsko G. (1998) Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 7406-7411.

Zhang L., Wang Y., Picking W. L., Picking W. D. and De Guzman R. N. (2006) Solution Structure of Monomeric BsaL, the Type III Secretion Needle Protein of Burkholderia pseudomallei. *Journal of Molecular Biology* **359,** 322-330.