X-ray crystallographic studies on Particulate Methane Monooxygenase, Thioredoxin A and Arginine Decarboxylase

Juni Andrél

Membrane Protein Crystallography Group
Division of Molecular Biosciences
Imperial College

Thesis submitted for the degree of doctor of Philosophy
2008
Abstract

The work presented in this thesis describes the X-ray crystallographic studies of particulate methane monoxygenase (pMMO) from *Methylococcus capsulatus* (Bath), thioredoxin A (BsTrxA) from *Bacillus subtilis* and arginine decarboxylase (AdiA) from *Escherichia coli*.

1. pMMO is a respiratory enzyme that catalyses the first step in the metabolic pathway in methanotrophic bacteria by converting methane to methanol. The crystal structure of this integral membrane protein was determined by molecular replacement to 3.5 Å resolution. The three metal sites in pMMO were confirmed to be a mononuclear copper site, a dinuclear copper site and a mononuclear zinc site.

2. Thioredoxin is a ubiquitous protein present in nearly all known organisms. Its purpose in the cell is to maintain cysteine-containing proteins in the reduced state by converting intramolecular disulfide bonds to dithiols in a redox reaction. The crystal structure of an active site mutant of BsTrxA was determined by molecular replacement to 1.5 Å resolution. The structure shows a homodimer that resembles enzyme-substrate reaction intermediates.

3. AdiA is a vitamin B6-dependent enzyme that catalyses the decarboxylation of arginine into agmatine. It forms a part of an enzymatic system in *E. coli* that contribute to making this organism acid resistant. The structure of arginine decarboxylase (AdiA) from *E. coli* was determined by multiple isomorphous replacement and anomalous scattering (MIRAS) methods to 2.4 Å resolution. The structure revealed a ~800 kDa decamer composed as a pentamer of five homodimers. AdiA becomes active as the cellular environment becomes more acidic. The structure of AdiA suggests how functional decamers associate with decreasing pH or disassociates into inactive homodimers with increasing pH. The enzyme mechanism and determinants for substrate specificity are discussed within the framework of the structure and comparisons with related structures are made.
# Table of contents

- **Abstract**  
  - Page 2

- **Table of Contents**  
  - Page 3

- **Abbreviations and Nomenclature**  
  - Page 8

- **Figure list**  
  - Page 10

- **Table list**  
  - Page 11

- **Thesis introduction**  
  - Page 12

- **Crystallography**  
  - Page 14

  1.1 **Overview**  
  - Page 15

  1.2 **Protein Crystallisation**  
  - Page 16

    1.2.1 Principle of protein crystallisation  
    - Page 16

    1.2.2 Protein crystallisation by vapour diffusion  
    - Page 17

    1.3.3 Membrane protein crystallisation  
    - Page 18

  1.3 **X-ray Crystallography**  
  - Page 19

    1.3.1 Unit cell  
    - Page 19

    1.3.2 X-ray diffraction  
    - Page 20

      1.3.2.1 Miller indices  
      - Page 20

      1.3.2.2 Bragg’s law  
      - Page 20

      1.3.2.3 The reciprocal lattice and the Ewald sphere  
      - Page 21

      1.3.2.4 Atomic scattering and structure factors  
      - Page 23

    1.3.3 Data collection and processing  
    - Page 24

      1.3.3.1 Data collection  
      - Page 24

      1.3.3.2 Collecting anomalous data  
      - Page 25

      1.3.3.3 Data processing  
      - Page 26

    1.3.4 Phase determination  
    - Page 26

      1.3.4.1 Multiple isomorphous replacement with anomalous scattering (MIRAS)  
      - Page 27

      1.3.4.2 Molecular replacement (MR)  
      - Page 28

    1.3.5 Solvent flattening  
    - Page 29

    1.3.6 Non-crystallographic symmetry averaging  
    - Page 29

- Page 3
### 1.3.7 Model building and refinement

29

### 1.3.8 Model quality

31

### PARTICULATE METHANE MONOOXYGENASE

32

### 2.1. INTRODUCTION

33

#### 2.1.1 Methane monooxygenase (MMO)

33

#### 2.1.2 The pMMO structures

34

- 2.1.2.1 X-ray structure of pMMO from *M. capsulatus* (Bath)

- 2.1.2.2 X-ray structure of pMMO from *M. trichosporium* OB3b

- 2.1.2.3 EM structures of pMMO and a pMMO supercomplex

36

#### 2.1.3 pMMO active site and metal content

37

#### 2.1.4 Aims of this study

39

### 2.2 METHODS AND MATERIALS

39

#### 2.2.1 Protein preparation

39

- 2.2.1.1 Protein purification

- 2.2.1.2 Suicide substrates

- 2.2.1.3 Crosslinked pMMO-MDH

39

- 2.2.2 Crystallisation

40

- 2.2.2.1 1YEW crystallisation condition

- 2.2.2.2 Novel crystallisation of pMMO: Conditions A-G

- 2.2.2.3 Further optimisation trials of pMMO crystals

- 2.2.2.4 Suicide substrates

- 2.2.2.5 Derivitisation with xenon and krypton

- 2.2.2.6 Crystallisation trials of pMMO-MDH

41

#### 2.2.3 Data collection and structure determination

44

- 2.2.3.1 A 3.5 Å dataset from a crystal grown in condition G

- 2.2.3.2 A 8.2 Å dataset from a crystal grown in condition D

- 2.2.3.3 A 4.0 Å dataset from a crystal potentially co-crystallised with propargylamine

- 2.2.3.4 A 4.7 Å dataset from a crystal pressurised with krypton gas

44

### 2.3 RESULTS

48

#### 2.3.1 Crystallisation of pMMO

48

- 2.3.1.1 Conditions A-G

- 2.3.1.2 Optimisation trials of G-type conditions

- 2.3.1.3 Crystallisation trials of pMMO-MDH

49

#### 2.3.2 The 3.5 Å pMMO model

52

- 2.3.2.1 Quality of the model

- 2.3.2.2 Metal site identities

52

#### 2.3.3 Other datasets

54

- 2.3.3.1 Propargylamine complex dataset at 4.0 Å

- 2.3.3.2 Krypton derivative dataset at 4.7 Å

54
## 2.4 DISCUSSION

### 2.4.1 The pMMO crystallisation process
- 2.4.1.1 The initial optimisation process: Condition A-G
- 2.4.1.2 Further optimisation trials of the G-type conditions
- 2.4.1.3 Screening for a pMMO-MDH crystallisation condition

### 2.4.2 Comparison of the 3.5 Å pMMO model with the 1YEW structure
- 2.4.2.1 Model bias
- 2.4.2.2 Metal site identities
- 2.4.2.3 The importance of zinc

### 2.4.3 Identifying the pMMO active site

### 2.4.4 Summary

## THIOREDOXIN A

### 3.1 INTRODUCTION
- 3.1.1 Thioredoxin
- 3.1.2 Thioredoxin active site and reaction mechanism
- 3.1.3 Thioredoxin structures
- 3.1.4 Thioredoxin A from *Bacillus subtilis*
- 3.1.5 Aim of this study

### 3.2 METHODS AND MATERIALS
- 3.2.1. Crystallisation
- 3.2.2 Data collection, structure determination and model building
- 3.2.3 Structure analysis

### 3.3 RESULTS
- 3.3.1. *B. subtilis* TrxA homodimers from mixed disulfide fishing experiments
  - 3.3.1.1 Mixed disulfide fishing
  - 3.3.1.2 Redox state of *B. subtilis* TrxA
- 3.3.2. Quality of the *B. subtilis* TrxA homodimer model
- 3.3.3 *B. subtilis* TrxA homodimer structure
- 3.3.4. The C32S mutation of the resolving cysteine

### 3.4 DISCUSSION
- 3.4.1 C29S and C32S *B. subtilis* TrxA homodimers are distinct from other thioredoxin homodimers
- 3.4.2 The C32S and C29S mutations of the active site cysteines
3.4.3 Comparison with thioredoxin-substrate complexes representing mixed disulfide intermediates

3.4.3.1 Comparison with human thioredoxin in complex with substrate peptide Ref-1 81
3.4.3.2 Comparison with human thioredoxin in complex with substrate peptide NFκB 82
3.4.3.3 Comparison with the BsTrxA-ArsC heterodimer 84
3.4.3.4 Comparison with the HvTrxh2-BASI complex 84
3.4.3.5 Thioredoxin substrate binding 85

3.4.4 Comparison with thioredoxin-thioredoxin reductase complexes 87

3.4.5 Conclusion 88

ARGININE DECARBOXYLASE

4.1 INTRODUCTION 90

4.1.1 Acid resistance in E. coli 90
4.1.1.1 AR1 system in E. coli 90
4.1.1.2 AR2 and AR3 systems in E. coli 90

4.1.2 Arginine dependent acid resistance (AR3) 91
4.1.2.1 pH homoestasis 91
4.1.2.2 Reversal of the electrochemical gradient Δψ 92

4.1.3 Amino acid decarboxylases are PLP enzymes 92
4.1.3.1 Pyridoxal-5′-phosphate 92
4.1.3.2 Group III decarboxylases 93
4.1.3.3 Group II decarboxylases 93
4.1.3.5 Amino acid decarboxylases in acid resistance 95

4.1.5 Aims of this study 95

4.2 METHODS AND MATERIALS 97

4.2.1 Expression and purification of AdiA 97
4.2.2 Crystallisation 98
4.2.3 Heavy atom derivatives 98
4.2.4 Data collection 99
4.2.5 Structure determination 99
4.2.6 Model building 100
4.2.7 Structure analysis 100

4.3 RESULTS 100

4.3.1 Quality of the AdiA model 100
4.3.2 The AdiA decamer and the molecule AC* dimer 102
4.3.3 The AdiA monomer 104
4.3.4 The active site

4.3.5 Stabilisation of the decameric structure
   4.3.5.1 Interactions in the AC* dimer
   4.3.5.2 Wing domain packing

4.4 DISCUSSION

4.4.1 The AdiA decamer – a pentamer of dimers

4.4.2 AdiA decamer formation and enzyme activity

4.4.3 AdiA active site and the decarboxylation reaction
   4.4.3.1 Conserved features in the AdiA active site
   4.4.3.2 AdiA decarboxylation reaction
   4.4.3.3 Identity of the proton donating residue in the decarboxylation reaction
   4.4.3.4 AdiA carboxy-terminal domain and substrate specificity

4.4.4 AdiA and OrnDC comparison
   4.4.4.1 Oligomeric states and enzyme activity
   4.4.4.2 AdiA and OrnDC active sites

4.4.5 AdiA and GadB – decarboxylases in *E. coli* acid resistance

4.4.6 Summary

THESIS SUMMARY

ACKNOWLEDGEMENTS

REFERENCES
### Abbreviations and Nomenclature

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdiA</td>
<td>Arginine decarboxylase from <em>E. coli</em></td>
</tr>
<tr>
<td>AdiC</td>
<td>Arginine:agmatine antiporter</td>
</tr>
<tr>
<td>AMS</td>
<td>4-acetamido-4′-maleimidyl-stilbene-2,2′-disulfonate</td>
</tr>
<tr>
<td>AR1</td>
<td>Acid resistance system 1</td>
</tr>
<tr>
<td>AR2</td>
<td>Acid resistance system 2</td>
</tr>
<tr>
<td>AR3</td>
<td>Acid resistance system 3</td>
</tr>
<tr>
<td>ArsC</td>
<td>Arsenate reductase</td>
</tr>
<tr>
<td>AspAT</td>
<td>Aspartate aminotransferase</td>
</tr>
<tr>
<td>BASI</td>
<td>alpha-amylase/subtilisin inhibitor</td>
</tr>
<tr>
<td>BsTrxA</td>
<td><em>B. subtilis</em> thioredoxin A</td>
</tr>
<tr>
<td>CadA</td>
<td>Lysine decarboxylase</td>
</tr>
<tr>
<td>CMC</td>
<td>Critical micelle concentration</td>
</tr>
<tr>
<td>DDC</td>
<td>Aromatic L-amino acid decarboxylase</td>
</tr>
<tr>
<td>DDM</td>
<td>Dodecyl-β-D-maltopyranoside</td>
</tr>
<tr>
<td>DG</td>
<td>Decyl-β-D-glucopyranoside</td>
</tr>
<tr>
<td>DGD</td>
<td>Diakylglycine decarboxylase</td>
</tr>
<tr>
<td>DM</td>
<td>Decyl-β-D-maltopyranoside</td>
</tr>
<tr>
<td>DOPA</td>
<td>3,4-dihydroxy-L-phenylalanine</td>
</tr>
<tr>
<td>DTSSP</td>
<td>Dithiobis-sulfosuccinimidyl propionate</td>
</tr>
<tr>
<td>DTT</td>
<td>1,4-Dithio-DL-threitol</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>ESRF</td>
<td>European Synchrotron Radiation Facility</td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast protein liquid chromatography</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>FTR</td>
<td>Ferredoxin-thioredoxin reductase</td>
</tr>
<tr>
<td>GABA</td>
<td>Gamma-aminobutyric acid</td>
</tr>
<tr>
<td>Gad67</td>
<td>Human glutamate decarboxylase</td>
</tr>
<tr>
<td>GadA</td>
<td>Glutamate decarboxylase from <em>E. coli</em> isoform A</td>
</tr>
<tr>
<td>GadB</td>
<td>Glutamate decarboxylase from <em>E. coli</em> isoform B</td>
</tr>
<tr>
<td>GadC</td>
<td>Glutamate:GABA antiporter in <em>E. coli</em></td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine triphosphate</td>
</tr>
<tr>
<td>HEPES</td>
<td>N-2-Hydroxyethylpiperazine-N′-2-ethanesulfonic acid</td>
</tr>
<tr>
<td>HgCl₂</td>
<td>Mercuric chloride</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>HvTrh2</td>
<td>Barley thioredoxin</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>MDH</td>
<td>Methanol dehydrogenase</td>
</tr>
<tr>
<td>MES</td>
<td>2-(N-morpholino)ethanesulfonic acid</td>
</tr>
<tr>
<td>MIRAS</td>
<td>Multiple isomorphous replacement anomalous scattering</td>
</tr>
<tr>
<td>MMO</td>
<td>Methane monooxygenase</td>
</tr>
<tr>
<td>NCS</td>
<td>Non-crystallographic symmetry</td>
</tr>
<tr>
<td>NFkB</td>
<td>Nuclear factor κB</td>
</tr>
<tr>
<td>NG</td>
<td>Decyl-β-D-glucopyranoside</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>OG</td>
<td>Decyl-β-D-glucopyranoside</td>
</tr>
<tr>
<td>OM</td>
<td>Octyl-β-D-maltopyranoside</td>
</tr>
<tr>
<td>OrnDC</td>
<td>Ornithine decarboxylase</td>
</tr>
<tr>
<td>PA</td>
<td>Propargylamine</td>
</tr>
<tr>
<td>PbMe₃</td>
<td>Trimethyllead acetate</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PIPES</td>
<td>1,4-piperazinediethanesulfonic acid</td>
</tr>
<tr>
<td>PLP</td>
<td>Pyridoxal-5’-phosphate</td>
</tr>
<tr>
<td>pMMO</td>
<td>Particulate methane monooxygenase</td>
</tr>
<tr>
<td>RPM</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SIRAS</td>
<td>Single isomorphous replacement anomalous scattering</td>
</tr>
<tr>
<td>sMMO</td>
<td>Soluble methane monooxygenase</td>
</tr>
<tr>
<td>TDM</td>
<td>Didecyl-β-D-maltopyranoside</td>
</tr>
<tr>
<td>TLS</td>
<td>Translation liberation screw-rotation</td>
</tr>
<tr>
<td>TR</td>
<td>Thioredoxin reductatse</td>
</tr>
<tr>
<td>Trx</td>
<td>Thioredoxin</td>
</tr>
<tr>
<td>Trx-(SH)₂</td>
<td>Reduced thioredoxin</td>
</tr>
<tr>
<td>Trx1</td>
<td>Human thioredoxin-1</td>
</tr>
<tr>
<td>TrxA</td>
<td><em>E. coli</em> thioredoxin A</td>
</tr>
<tr>
<td>Trx-f</td>
<td>Spinach chloroplast thioredoxin f</td>
</tr>
<tr>
<td>Trx-m</td>
<td>Spinach chloroplast thioredoxin m</td>
</tr>
<tr>
<td>Trx-S₂</td>
<td>Oxidised thioredoxin</td>
</tr>
<tr>
<td>UDM</td>
<td>Undecyl-β-D-maltopyranoside</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
Figure list

<table>
<thead>
<tr>
<th>Figure</th>
<th>Figure title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Overview of X-ray crystallography</td>
<td>15</td>
</tr>
<tr>
<td>1.2</td>
<td>Crystallisation phase diagram</td>
<td>17</td>
</tr>
<tr>
<td>1.3</td>
<td>Crystallisation by vapour diffusion techniques</td>
<td>18</td>
</tr>
<tr>
<td>1.4</td>
<td>Membrane protein crystal contacts</td>
<td>19</td>
</tr>
<tr>
<td>1.5</td>
<td>Unit cell</td>
<td>20</td>
</tr>
<tr>
<td>1.6</td>
<td>(234) planes in unit cell</td>
<td>20</td>
</tr>
<tr>
<td>1.7</td>
<td>Bragg’s law</td>
<td>21</td>
</tr>
<tr>
<td>1.8</td>
<td>Ewald sphere construction</td>
<td>22</td>
</tr>
<tr>
<td>2.1</td>
<td>Methane oxidation pathway</td>
<td>33</td>
</tr>
<tr>
<td>2.2</td>
<td>Structural architecture of pMMO</td>
<td>34</td>
</tr>
<tr>
<td>2.3</td>
<td>The metal sites of the pMMO αβγ protomer</td>
<td>35</td>
</tr>
<tr>
<td>2.4</td>
<td>EM structure of the pMMO-MDH complex</td>
<td>37</td>
</tr>
<tr>
<td>2.5</td>
<td>Optimisation process of Conditions A-G</td>
<td>49</td>
</tr>
<tr>
<td>2.6</td>
<td>pMMO diffraction pattern to 3.5 Å resolution</td>
<td>51</td>
</tr>
<tr>
<td>2.7</td>
<td>The 3.5 Å pMMO model</td>
<td>53</td>
</tr>
<tr>
<td>2.8</td>
<td>Metal sites in the pMMO trimer</td>
<td>55</td>
</tr>
<tr>
<td>2.9</td>
<td>Zn mediated crystal contacts in the pMMO trimer</td>
<td>56</td>
</tr>
<tr>
<td>3.1</td>
<td>Thioredoxin reaction</td>
<td>65</td>
</tr>
<tr>
<td>3.2</td>
<td>Thioredoxin system</td>
<td>66</td>
</tr>
<tr>
<td>3.3</td>
<td>Mixed disulfide fishing results</td>
<td>70</td>
</tr>
<tr>
<td>3.4</td>
<td>Redox states of BsTrxA monomers and dimers</td>
<td>72</td>
</tr>
<tr>
<td>3.5</td>
<td>C32S BsTrxA homodimer and crystal packing</td>
<td>75</td>
</tr>
<tr>
<td>3.6</td>
<td>C32S BsTrxA dimer interface</td>
<td>77</td>
</tr>
<tr>
<td>3.7</td>
<td>Ser32 interactions in the C32S BsTrxA homodimer structure</td>
<td>78</td>
</tr>
<tr>
<td>3.8</td>
<td>Ser32 in C32S BsTrxA and Cys32 in wild type BsTrxA</td>
<td>80</td>
</tr>
<tr>
<td>3.9</td>
<td>Thioredoxin substrate binding regions</td>
<td>85</td>
</tr>
<tr>
<td>3.10</td>
<td>Comparison of thioredoxin protein/peptide binding modes</td>
<td>86</td>
</tr>
<tr>
<td>3.11</td>
<td>Comparison with thioredoxin reductase binding modes</td>
<td>88</td>
</tr>
<tr>
<td>4.1</td>
<td>Arginine dependent acid resistance in <em>E. coli</em></td>
<td>91</td>
</tr>
<tr>
<td>4.2</td>
<td>Pyridoxal-5´-phosphate (PLP)</td>
<td>93</td>
</tr>
<tr>
<td>4.3</td>
<td>PLP enzymes discussed in this study</td>
<td>96</td>
</tr>
<tr>
<td>4.4</td>
<td>AdiA decamer crystallises as a decamer</td>
<td>102</td>
</tr>
<tr>
<td>4.5</td>
<td>AdiA decamer</td>
<td>103</td>
</tr>
</tbody>
</table>
4.6 AdiA homodimer 104
4.7 AdiA domains 105
4.8 PLP coordination in the AdiA active site 107
4.9 Packing of the AC* dimer 108
4.10 Wing domain packing in the AdiA decamer 109
4.11 Wing domain interactions in the AdiA decamer 110
4.12 The physiological AdiA decamer is a pentamer of dimers 111
4.13 Electrostatic surface representation of AdiA AC* homodimer 112
4.14 AdiA decarboxylation reaction 114
4.15 AdiA active site cleft 116
4.16 AdiA active site in comparison with other PLP enzyme active sites 117
4.17 Wing domain packing interactions in AdiA and OrnDC 119
4.18 Comparison of PLP coordination in the AdiA and OrnDC active sites 121
4.19 Comparison of OrnDC-Glu532 and AdiA-Thr571 regions 121

**Table list**

<table>
<thead>
<tr>
<th>Table</th>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Crystallisation conditions A-G</td>
<td>42</td>
</tr>
<tr>
<td>2.2</td>
<td>Low salt PEG vs pH screens for pMMO-MDH crystallization</td>
<td>45</td>
</tr>
<tr>
<td>2.3</td>
<td>Data collection, processing and refinement statistics for the 3.5 Å pMMO model</td>
<td>46</td>
</tr>
<tr>
<td>2.4</td>
<td>Data collection and processing statistics for a 8.2 Å dataset, PA dataset and Kr dataset</td>
<td>48</td>
</tr>
<tr>
<td>2.5</td>
<td>Comparison of crystallisation conditions</td>
<td>61</td>
</tr>
<tr>
<td>3.1</td>
<td>C32S BsTrxA data collection and refinement statistics</td>
<td>74</td>
</tr>
<tr>
<td>3.2</td>
<td>Thioredoxin complex structures</td>
<td>83</td>
</tr>
<tr>
<td>3.3</td>
<td>Substrate binding regions in thioredoxin</td>
<td>85</td>
</tr>
<tr>
<td>4.1</td>
<td>PLP enzymes discussed in this study</td>
<td>94</td>
</tr>
<tr>
<td>4.2</td>
<td>AdiA data collection and refinement statistics</td>
<td>101</td>
</tr>
</tbody>
</table>
Thesis introduction

In this work three enzymes have been structurally characterised using X-ray crystallography: particulate methane monooxygenase (pMMO) from *Methylococcus capsulatus* (Bath), thioredoxin A from *Bacillus subtilis* (BsTrxA) and arginine decarboxylase (AdiA) from *Escherichia coli*.

The first project presented in this thesis is the X-ray crystallographic work on the integral membrane protein pMMO from *M. capsulatus* (Bath). *M. capsulatus* (Bath) is a methanotrophic bacteria that survives by oxidising methane in a series of steps to provide carbon and energy for the bacterium. The first essential reaction in this pathway is the oxidation of methane to methanol, which is catalysed by pMMO. How pMMO catalyses methane oxidation is of great interest, with possible commercial applications in industry and bioremediation. The mechanism of the reaction is, however, not known. The active site of pMMO is thought to be a number of metal sites in the enzyme. However, although an earlier crystal structure of pMMO identified three metal sites in the enzyme, the active site could not be located and identified (Lieberman & Rosenzweig, 2005). In addition, the number of metal sites and the precise metal content of these sites in pMMO remain controversial. Hence the aims of the structural work on pMMO in this thesis were to identify the active site and to confirm the metal sites and metal content of pMMO, which would be the first step in elucidating the precise mechanism by which pMMO oxidises methane. In order to improve the limited functional understanding of pMMO an additional aim was to determine the structure of a pMMO in complex with methanol dehydrogenase (MDH) that catalyses the oxidation of methane to formaldehyde.

The second project of this thesis was aimed at characterising the structure of a thioredoxin homodimer in *B. subtilis* thought to mimic a reaction intermediate. Thioredoxin is a ubiquitous enzyme that catalyses the reduction of intramolecular disulfide bonds to dithiols on a range of different substrates in the cell. This process of ‘thiol-redox-control’ is thought to regulate important cellular processes, such as DNA synthesis and protein repair. Thioredoxin A in *B. subtilis* (BsTrxA) is essential to the organism’s survival, highlighting the importance of thioredoxin substrates being converted to active reduced forms. During a study aimed at identifying the substrates of BsTrxA, a homodimer of a BsTrxA active site mutant was identified. It was thought that this homodimer was formed through a disulfide bond between
active site cysteines in a manner mimicking the reaction intermediate between thioredoxin and its substrate. No thioredoxin-substrate protein structures or structures of the BsTrxA were available at the time this project was initiated. Hence the structural work on the active site mutant BsTrxA homodimer was carried out in order to (i) confirm that the homodimer was indeed formed by a disulfide bond between active site cysteines and (ii) to analyse the dimer interface in order to infer how BsTrxA would bind substrate proteins.

Structural studies on the enzyme arginine decarboxylase (AdiA) form the third project described in this thesis. AdiA is important for survival of *E. coli* in acidic environments and becomes activated as the pH in the cell decreases. The active AdiA converts arginine to agmatine, using up a proton in the process. The agmatine is shuttled out of the cell in exchange for more arginine by the arginine-agmatine antiporter (AdiC). Together the AdiA and AdiC enzymes form an acid resistance system that effectively shuttles out one proton from the cell per agmatine molecule. The aim of the work on AdiA was to investigate the molecular mechanism of AdiA in *E. coli* acid resistance. A detailed analysis of the structure is presented together with a comparative analysis of AdiA with homologous enzymes.

pMMO, BsTrxA and AdiA are un-related proteins from separate species of bacteria with different functions in each organism. Hence for clarity, the studies on each enzyme will be presented in a separate chapter with an introduction, methods and materials, results and discussion section. The structural studies of pMMO, BsTrxA and AdiA were carried out using the method of X-ray crystallography, which will be introduced in the first chapter of this thesis.
CRYSTALLOGRAPHY
1.1 OVERVIEW

An understanding of the atomic structure of a protein can help in elucidating how it functions chemically and structurally. Since the distances between atoms are very small, of the order of $10^{-10}$ meters, it is impossible to look at the atomic components of a protein using a normal light microscope, because the wavelength of light is too long. In contrast, X-rays have smaller wavelengths, in the range of 0.5-2.5 Å (where 1 Å is $10^{-10}$ meters), which corresponds to the distances involved in the spacing of atoms in a protein. Hence X-rays can give us the resolution required to separate one atom from another and thus look at the structure of a protein in atomic detail.

X-rays are electro-magnetic waves and when they meet an atom a proportion of them is scattered, or diffracted, by the electron clouds of the atoms. In a light microscope the waves scattered by an object are scattered as the Fourier Transform (FT) of the object. The lens in the microscope recombines the scattered light and performs a second FT to turn it into the image of the object. One of the obstacles in X-ray crystallography is that there are no X-ray lenses to perform the second FT, so it is not possible to construct an X-ray microscope. Instead a lens has to be mimicked by recording a diffraction pattern of the scattered X-rays, i.e. the first FT of the object, and use computational methods to calculate the second FT (Figure 1.1).

Another major obstacle to be overcome in X-ray crystallography is the requirement for protein crystals. In order to record a diffraction pattern it is necessary to reinforce the signal produced by scattered waves coming from a single protein molecule. A single atom or molecule would produce a signal that would be too weak to record. Many protein molecules in solution would

![Figure 1. Overview of X-ray crystallography.](image-url)
have different orientations, which would also produce too weak a signal. By using crystals of the protein it is possible to amplify the signal, making it strong enough to be detected. Protein crystals are a requirement in X-ray crystallography and the process of growing them is known as protein crystallisation (see section 1.2). A protein crystal contains millions of copies of the protein arranged at specific orientations in regular, symmetric intervals repeated throughout the crystal in three dimensions. Waves scattered from the crystal and traveling in the same direction, will form a resultant wave that will be strong enough to be recorded as a spot, or a reflection, in the diffraction pattern.

In order to combine the waves into an image three properties of the scattered waves must be known: the direction, the amplitude and the phase of each wave. Two of the properties, the direction of the wave and its amplitude, can be found out directly by recording the diffraction pattern of the scattered X-rays. From the position of each reflection in the diffraction pattern it is possible to calculate the diffraction angle, i.e. the direction of the wave, and from its intensity it is possible to calculate the amplitude of the wave. It is, however, not possible to calculate the phase of the wave from the diffraction pattern. This is known as the phase problem, which can be the largest obstacle to be overcome in X-ray crystallography. Additional methods are required to calculate the phases of the scattered waves, some of which will be described below. Once all three properties of the scattered waves are known, the direction, the amplitude and the phase, it is possible to perform the second FT and calculate an image of the protein. In X-ray crystallography this image consists of a three-dimensional map of the electron density of the protein. It is possible to build a model of the protein into this electron density by using the known structures of amino acids as building blocks (Figure 1.1).

1.2 PROTEIN CRYSTALLISATION

1.2.1 Principle of protein crystallisation

Well-ordered protein crystals are a requirement for structure determination using X-ray crystallography and the process of growing protein crystals is known as crystallisation. The process of crystallisation can be illustrated in a phase diagram (Figure 1.2). When crystallising a protein, it is necessary to decrease the protein solubility in a controlled fashion
that will yield crystals. The protein molecules exist in solution in the undersaturated state. As precipitant or protein concentration is increased the solution will become supersaturated, which causes nucleation to occur. Nucleation is the formation of small ordered aggregates of protein, which can grow into crystals by the ordered addition of protein molecules from the solution.

The formation of crystals is dependent on many factors in addition to protein and precipitant concentrations, such as the crystallisation method used, salt and buffer concentration, pH and temperature. It can also be dependent on the presence of additives, metal ions and ligands. In the case of membrane proteins (see section 1.2.3), the detergent choice is also an important factor. By varying these parameters, there are, in practice, an infinite number of different crystallisation conditions that can be screened to obtain well-diffracting crystals.

1.3.2 Protein crystallisation by vapour diffusion

One of the most common crystallisation methods is known as the vapour-diffusion technique. In this work, vapour diffusion techniques by the hanging-drop method and by the sitting-drop method were used. In the hanging-drop method a small aliquot of protein solution, typically 1-5 µl, is placed on a cover glass slide to which an equal volume of crystallisation solution is added. The crystallisation solution commonly contains a buffer and a precipitant, such as salt and/or polyethylene glycol (PEG), at a concentration just below that required for nucleation to occur. The cover slide is placed over a reservoir, typically containing 50 µl – 1000 µl
crystallisation solution, and the system is sealed with vacuum grease. The crystallisation solution placed in the reservoir is commonly referred as reservoir solution. The mixing of the protein solution with the crystallisation solution in the drop halves both the protein concentration and the concentration of the reagents in the crystallisation solution. In order to achieve a vapour pressure equilibrium between the drop and the reservoir solution, water will evaporate from the drop until the reagent concentrations in the drop are approximately the same as that in the reservoir. Hence the protein and the precipitant concentration will slowly increase in the drop, which may cause the protein to form crystals. In the sitting-drop method the drop is placed on a platform in the reservoir instead of hanging from a cover glass slide (Figure 1.3).

![Figure 1.3](image.png)

**Figure 1.3. Crystallisation by vapour diffusion techniques.** Schematic representation of two vapour diffusion methods.

### 1.3.3 Membrane protein crystallisation

The additional difficulty in crystallising membrane proteins compared to soluble proteins is the requirement for detergent to keep the protein in solution. The detergent interacts with the hydrophobic surface of the protein, forming a detergent-protein micelle, thus mimicking the cell membrane environment. The type of detergent used is an additional variable in the crystallisation equation compared to soluble proteins and it is often necessary to prepare protein with different detergents to optimize the crystallisation. The presence of detergent, although critical for keeping the membrane protein in solution, has an adverse effect on the crystallisation process because the detergent micelle covering the hydrophobic surface of membrane protein limits the region of the protein available to interact with other protein molecules and form crystal contacts (Figure 1.4). Hence, the more hydrophobic the membrane protein surface is, the weaker the interactions between protein molecules are. Consequently, membrane protein crystals tend to have few protein-protein contacts and to show high solvent contents, which negatively affects the quality of the crystal (Iwata, 2003).
1.3 X-RAY CRYSTALLOGRAPHY

1.3.1 Unit cell

In a protein crystal, the protein molecules are arranged in specific orientations with regular intervals repeated throughout the crystal in three dimensions. The unit cell is the smallest building block of the protein crystal and by applying translation repeatedly, it could generate the whole crystal. By definition, the distribution of atoms in each unit cell is identical. The unit cell is described by the length of the edges \((a, b, c)\) and the angles of the corners \((\alpha, \beta, \gamma)\) (Figure 1.5). The asymmetric unit is the unique part of the unit cell and this may contain one or several copies of the protein. The asymmetric unit is the smallest entity that can be copied by crystallographic symmetry operations to give the rest of the unit cell. Non-crystallographic symmetry (NCS) occurs if there is symmetry between multiple copies of the protein in the asymmetric unit. This type of symmetry is not repeated throughout the crystal, but only within the asymmetric unit. The unit cells repeated throughout the crystal define the crystal lattice.

The crystal lattice can be triclinic, monoclinic, orthorhombic, trigonal, tetragonal, hexagonal or cubic depending on the symmetry within the unit cell. Since proteins consist of L-amino acids, and are thus chiral, only two sets of crystallographic symmetry operators are allowed in protein crystals: rotations and translations. There are 65 possible combinations of symmetry elements in a protein crystal corresponding to the 65 space groups to which a crystal can belong.
1.3.2 X-ray diffraction

1.3.2.1 Miller indices

The reflections produced in a diffraction pattern are directly related to the properties of the crystal. When X-rays are scattered by the electrons in the crystal, they behave as if they were reflected from different sets of parallel planes in the unit cell. An entire set of parallel planes acts as a single diffractor and scatters a resultant wave that produces one reflection in the diffraction pattern. The Miller index \((hkl)\) of the reflection defines the set of parallel planes from which the reflection originates. \(h\) gives the the number of planes in the \(x\)-direction per unit cell, \(k\) gives the number in the \(y\)-direction and \(l\) gives the number in the \(z\)-direction. Hence reflection \((234)\) originates from a wave scattered from the set of planes that intersects the \(x\)-axis of the unit cell twice, intersects the \(y\)-axis three times and intersects the \(z\)-axis four times (Figure 1.6). Many sets of parallel planes are present in the crystal and each set of planes will produce one reflection in the diffraction pattern.

1.3.2.2 Bragg’s law

In order for X-rays scattered by a set of parallel planes to add up to a resultant wave and produce a reflection they must be in phase with each other. The waves are only in phase with each other if they satisfy Bragg’s law (Figure 1.7). Bragg’s law shows that two parallel waves only add up in phase when the extra distance the second wave has to travel in the crystal \((2d\sin \theta)\), in comparison to the first wave, is a multiple of the wavelength (Equation 1.1). \(\lambda\) is the wavelength of the X-ray, \(d_{hkl}\) is the distance between Bragg planes and \(\theta\) is the angle between the incident beam and the Bragg plane.
\[ 2d_{hkl} \sin \theta = n\lambda \]  

(1.1)

If a crystal is put in the X-ray beam at a fixed orientation only some sets of parallel planes will satisfy Bragg’s law and produce reflections. In order to record the reflections corresponding to all planes in the three-dimensional crystal, the crystal must be rotated through a certain angle which is determined by the space group of the crystals.

![Bragg's law diagram](image)

**Figure 1.7.** Bragg’s law.

### 1.3.2.3 The reciprocal lattice and the Ewald sphere

There is a reciprocal relationship between the distance \(d_{hkl}\) between a set of Bragg planes in the crystal (real space) and the distance \(d'\) between reflections in the diffraction pattern (reciprocal space) where \(d' = 1/d_{hkl}\). This reciprocal relationship is represented in the reciprocal lattice (Figure 1.8). Each set of parallel planes in the crystal is represented by a point in the three-dimensional reciprocal lattice, e.g. the (234) planes produces a (234) point in the reciprocal lattice. The lattice is defined so that if the distance between the origin (O) of the crystalline lattice and a set of parallel planes with indices (hkl) is \(d_{hkl}\), then the distance between the origin (O) of the reciprocal lattice and the reciprocal lattice point (hkl) is \(1/d_{hkl}\). The reciprocal lattice together with the geometric construction of the imaginary Ewald sphere, can be used as a tool to visualize Bragg’s law and to predict the direction of the diffracted X-rays from all sets of parallel planes in the crystal. The Ewald sphere is centered on the crystal.
Figure 1.8. Ewald’s sphere construction.

at point C and has a radius of $1/\lambda$, which represents the inverse of the wavelength (Figure 1.8). The incident X-ray beam passes through point O, which is the origin of the reciprocal lattice ($hkl=000$). The reciprocal lattice in this figure is at an orientation where the reciprocal lattice point $P(hkl)$, representing the set of parallel planes ($hkl$) in the crystal, contacts the Ewald sphere. When $P(hkl)$ contacts the Ewald sphere the real lattice planes ($hkl$) will align parallel to the IP vector at an angle of $\theta$, which is the angle between the incident beam and the ($hkl$) plane. The distance between the origin O of the reciprocal lattice and the reciprocal lattice point P (OP) is $1/d_{hkl}$. This relationship can be represented by Equation 1.2, which is Bragg’s law with $n=1$.

$$OP = \frac{1}{d_{hkl}} = IO \sin \theta = \frac{2 \sin \theta}{\lambda}$$  (1.2)
Hence Bragg’s law is satisfied when the reciprocal lattice point \( P(hkl) \) contacts the Ewald sphere and the set of parallel planes \((hkl)\) will diffract X-rays in the direction of the CP vector, producing reflection \((hkl)\) in the diffraction pattern (Figure 1.8). Rotation of the crystal in the beam around \((C)\), rotates the reciprocal lattice around \((O)\), which will bring other reciprocal lattice points in contact with the Ewald sphere. As they come in contact with the sphere, Bragg’s law will be satisfied and diffraction will occur from the set of planes represented by the reciprocal lattice points.

The position of reflection \( hkl \) in the diffraction pattern is thus only dependent on which set of parallel planes it arises from. The Bragg planes are in turn defined by the unit cell dimensions. Hence the position of the reflection in the diffraction pattern is only dependent on the unit cell. The intensity of the reflection, however, is dependent on the distribution of atoms on the Bragg planes within the unit cell.

### 1.3.2.4 Atomic scattering and structure factors

Each reflection \( hkl \) in the diffraction pattern is produced by a resultant wave, which is the sum of all X-rays diffracting in phase with each other from a set of Bragg planes \((hkl)\) in the unit cell. The structure factor \( F_{hkl} \) of a reflection is a complex number that represents the resultant wave, and as such it has both an amplitude and a phase. All atoms in the unit cell contribute to each diffracted X-ray and thus to every reflection in the diffraction pattern. Hence the structure factor \( F_{hkl} \) for each reflection can be calculated by summing the diffractive contribution by every atom in the unit cell. The diffractive contribution of one atom \( j \), in position \( x_j, y_j, z_j \), to the reflection \( hkl \) is the atomic structure factor \( f_{hkl} \) (Equation 1.3). The term \( f_j \) is the atomic scattering factor, which is dependent on the element of the atom, the wavelength \( \lambda \) and the diffraction angle \( \theta \) of each wave scattered by the atom \( j \). The value of the atomic scattering factor \( f_j \) increases the heavier the atom is, i.e. the more electrons the atom has, and decreases as the scattering angle \( \theta \) increases. The exponential term indicates the position, \( xyz \), of the atom \( j \) in terms of fractions of the unit cell axis lengths.

\[
f_{hkl} = f_j \exp[2\pi i(hx_j + ky_j + lz_j)]
\]  

(Equation 1.3)

Since the structure factor \( F_{hkl} \) of reflection \( hkl \) is the sum of diffractive contributions from all atoms in the unit cell, it can be calculated by summing the atomic structure factors for every atom in the unit cell. This is known as the structure factor equation (Equation 1.4).
\[ F_{hkl} = \sum_{j=1}^{n} f_j \exp[2\pi i(hx_j + ky_j + lz_j)] \]  \hspace{1cm} (1.4)

The structure factor equation can alternatively be written as the integral of contributions from small volume elements of electron density in the whole unit cell on the planes (hkl) (Equation 1.5).

\[ F_{hkl} = \int \int \int \rho(xyz) \exp[2\pi i(hx + ky + lz)] dx dy dz \]  \hspace{1cm} (1.5)

This equation is the Fourier transform of the electron density, \( \rho \), in the unit cell. Reversibly, the electron density is the Fourier transform of the sum of the structure factors in the unit cell (Equation 1.6)

\[ \rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| \exp\left(-2\pi i(hx + ky + lz - i\alpha_{hkl})\right) \]  \hspace{1cm} (1.6)

In order to calculate the electron density in the unit cell from Equation 1.6, we need to know the value of each structure factor \( F_{hkl} \) and add them together. Each structure factor \( F_{hkl} \) describes the resultant wave producing the reflection \( hkl \). To describe this wave, and to add it to other waves, in order to calculate the electron density, we need to know its structure factor amplitude \( |F_{hkl}| \) and its phase \( \alpha_{hkl} \).

**1.3.3 Data collection and processing**

**1.3.3.1 Data collection**

The intensity of each reflection can be measured from the diffraction pattern and from the intensity the amplitude of the wave, \( |F_{hkl}| \), can be calculated. The aim of data collection is to record the intensities of a complete set of reflections, i.e. a full dataset, usually at as high a resolution as possible. Preferably, each reflection is measured several times in order to get as accurate a measure of the intensity as possible and of the experimental error associated with the intensity measurement. The output at the end of the data collection and processing is a list of reflections, \( hkl \), their intensities (I) and the error on the intensity (\( \sigma I \)). In order to collect a dataset that is complete, goes to high resolution and measures each reflection as many times (high redundancy/multiplicity) it is important to have a good data collection strategy.
How many images to collect for a complete data set depends to a large extent on the symmetry of the crystal. If there is symmetry in the crystal, it is only necessary to collect the unique part of the diffraction pattern since applying the symmetry will generate all possible reflections. If there is no anomalous scattering (see below) then reflection \( hkl \) will be the same as reflection \(-h-k-l\), which is known as Friedel’s law. Hence 180° of data will result in a complete dataset for all possible space groups. However, whether this will be sufficient for a complete dataset also depends on the orientation of the crystal, with respect to the rotation axis used for data collection, and the starting \( \varphi \). In this work pMMO crystals grew in a tetragonal space group, hence the diffraction pattern contained four-fold symmetry and a minimum of 45° had to be collected for a full dataset. AdiA crystals grew in a hexagonal space group and hence the diffraction pattern contained six-fold symmetry and a minimum of 60° had to be collected for a full dataset. In contrast, BsTrxA crystals contained no symmetry and 300° data was collected. The data in excess of the 180° of data required for a full dataset was collected in order to improve the multiplicity of the data.

The resolution of the data is in most cases dependent on the quality of the protein crystal where disordered and imperfect crystals diffract less well. For instance the crystal is often composed of smaller mosaic units. The units cells may pack well within each mosaic unit, but the alignment of the mosaic units in relation to each other may be poor. Hence each reflection represents the average diffraction for the different orientations of the mosaic units. This is known as the mosaicity of a crystal. How susceptible the protein crystal is to radiation damage will also affect the resolution and how long and how intensely the crystal can be exposed in the X-ray beam. The completeness and the redundancy of the data collected at a certain resolution will thus be dependent on how long the crystal survives in the beam.

1.3.3.2 Collecting anomalous data

At certain wavelengths the electrons of an atom absorb part of the energy of an X-ray and reemits it with an altered phase with respect to the normally scattered wave. This phenomenon is referred to as anomalous scattering or anomalous dispersion. This occurs at different wavelengths for different types of atoms and is known as the absorption edge of an element. The phase difference in the emitted wave causes Friedel’s law to break down and the intensities of Bijvoet pairs, the \( hkl \) and \(-h-k-l\) reflections, are now different, i.e. \( I(hkl) \neq I(-h-k-l) \). This inequality between the \( F_{hkl} \) and \( F_{-h-k-l} \) reflections can be used to determine the position of the anomalous scattering atom, which can contribute to the estimation of phases (see below). Mainly heavier atoms, such as metals ions, have absorption edges within the range of the X-ray wavelengths used in a diffraction experiment. Since metal ions have more electrons,
they typically have large atomic scattering factors, which increase the anomalous signal from these atoms. Consequently the intensity differences between the Bijvoet pairs, referred to as the anomalous signal, are larger and more measurable for heavy atoms.

1.3.3.3 Data processing

After data collection the reflections on each image are indexed according to their Miller indices \(hkl\) and their intensities are integrated. Scale factors are then applied to the reflections in order to make equivalent reflections (such as multiply measured reflections and their symmetry related equivalents) comparable by correcting for variations between images, for instance in the strength of the beam and in the radiation damage to the crystal. Finally the data is merged when the intensities of equivalent reflections are averaged. For this project the program DENZO (Otwinowski, 1997) was used to index and integrate the data and the integrated data was scaled and merged by the program SCALEPACK (Otwinowski, 1997). The intensities were then converted to structure factor amplitudes using the program TRUNCATE (French, 1978). When processing anomalous data it is important not to merge Bijvoet pairs, since this would cause the anomalous signal to be lost.

\(R\)\(_{\text{sym}}\) is an indicator of data quality from comparing the intensities of symmetry equivalent reflections (Equation 1.7). Because the \(R\)\(_{\text{sym}}\) is dependent on the redundancy of the data, where more redundant data gives a higher \(R\)\(_{\text{sym}}\) value, this value can be a misleading when judging data quality. The average intensity of reflection \(hkl\) is represented by \(\bar{I}_{(hkl)}\) and the intensity of the \(i\)th measurement by \(I_{(hkl)}\).

\[
R_{\text{sym}} = \frac{\sum_{hkl} \sum_{i} | I_{(hkl)} - \bar{I}_{(hkl)} |}{\sum_{hkl} \sum_{i} I_{(hkl)}}
\]  

(1.7)

1.3.4 Phase determination

The structure factor amplitude \(|F_{hkl}|\) for each measured reflection was obtained from the recorded diffraction patterns. In order to calculate the electron density, the value of the phase of each reflection \(\alpha_{hkl}\) needs to be determined. There are different methods that can be applied to obtain this phase information. This work used multiple isomorphous replacement with anomalous scattering (MIRAS) to obtain the phase information for the structure determination of AdiA and molecular replacement (MR) for determining the structures of pMMO and BsTrxA.
1.3.4.1 Multiple isomorphous replacement with anomalous scattering (MIRAS)

In isomorphous replacement crystals are derivatised with heavy metals that have large atomic scattering factors. The prerequisite for this method is that the native crystal (without heavy atoms) and the derivative crystal (with heavy atoms) are isomorphous, i.e. the same in all respects except for the addition of the heavy atoms in the derivative crystal. An indication of how isomorphous data from native and derivative crystals are, is given by the $R_{\text{deriv}}$ value (Equation 1.8), where a low $R$-factor indicates that the datasets are isomorphous. $|F_{PH}|$ is the structure factor amplitude for each reflection in the derivative dataset and $|F_P|$ is the structure factor amplitude for each reflection in the native dataset.

$$R_{\text{deriv}} = \frac{\sum_{hkl} |F_{PH}| - |F_P|}{\sum_{hkl} |F_P|}$$ (1.8)

The $R_{\text{deriv}}$ value can also give an indication as to whether a dataset is likely to be heavy atom derivative or not. Since the atomic scattering factors of heavy atoms decrease with increasing resolution, the isomorphous differences between two datasets, i.e. those difference due solely to the presence of the heavy atoms, would also decrease with increasing resolution. In contrast, differences between datasets due to non-isomorphism would increase with increasing resolution. Consequently, the $R_{\text{deriv}}$ value for a non-derivative dataset would increase with increasing resolution, while an $R_{\text{deriv}}$ value for a derivative dataset would decrease with increasing resolution.

The intensities of the equivalent reflections between the dataset collected on the native crystal and the dataset collected from the derivative crystal are compared. Since the crystals are isomorphous, the observed differences in the intensity of equivalent reflections between the two datasets are attributed to the presence of the heavy atoms. Scattering by the derivative crystal (PH) is the sum of the scattering by the native crystal (P) and the scattering from the heavy atoms (H) (Equation 1.9). Solving Equation 1.9 for the structure of the heavy atoms only, i.e. finding the heavy atom positions and obtaining values for each structure factor $F_H$, will be a first step towards calculating the phase for each reflection.

$$F_{PH} = F_P + F_H$$ (1.9)

The heavy atom positions can be found with Patterson methods or direct methods. Program SHELXD (Uson & Sheldrick, 1999), which was used to find the lead (Pb) atom positions in
the Pb-AdiA derivative, uses a combination of both Patterson methods and direct methods. From the position of the heavy atoms the heavy atom structure factors $F_H$, including their phases, can be calculated. This information allows for the calculation of two estimates of the values of the phase $\alpha_{\text{obs}}$ for each reflection $F_r$.

If the heavy atom in the derivative crystal scatter anomalously at the wavelength the data is collected at, it is possible to use the anomalous scattering to determine which of the two estimates of the phase for each reflection is the most likely. Under anomalous scattering Friedel’s law $F_{\text{obs}} = F_{-h,k,l}$ does not hold true and the intensity differs between the Bijvoet pairs, reflections $hkl$ and $-h-k-l$. The data from the $-h-k-l$ reflection may now be treated as if coming from another isomorphous derivative. With this additional phase information the most likely phase for each reflection can be determined and an initial electron density map can be calculated using Equation 1.6. By including phase estimates from a second derivative the overall phase estimation in the electron density equation can be improved. In this work, for instance, the mercury (Hg) sites in the Hg-AdiA derivative was used to improve the initial phase estimates from the Pb-AdiA derivative.

1.3.4.2 Molecular replacement (MR)

In molecular replacement the initial estimates of the phases are taken from a previously known homologous structure, the search model. This is done by finding the orientation and position of the search model in the unknown crystal where the predicted diffraction pattern from the search model in the new crystal form best matches the observed diffraction. The orientation and position of the search model is defined by six parameters, three rotation angles and three translations $(\alpha, \beta, \gamma, t_x, t_y, t_z)$. The program PHASER (Read, 2001), which was used for molecular replacement in this work, uses maximum likelihood methods to first find the best orientation of the search model in a rotation function followed by a translation function to find the best position for that orientation. Once the best orientation and position of the search model is found the calculated phases, $\alpha_{\text{calc}}$, from the search model are combined with the observed structure factor amplitudes, $|F_{\text{obs}}|$, to calculate the electron density from Equation 1.6. The initial phase estimates from the search model are improved during rebuilding and refinement of the new model so that it more closely resembles the target structure.

A challenge in MR can be to identify a good search model since it needs to resemble the target protein in the crystal sufficiently for a correct orientation/position to be identified. In this work the MR solution of pMMO was straightforward since the search model was a
pMMO structure from the same organism that crystallised in the same space group. In the MR solution for BsTrxA a search with the native TrxA *E. coli* was unsuccessful, while a polyserine model of the same enzyme gave a solution.

### 1.3.5 Solvent flattening

The initial estimates of phases determined experimentally, for instance by MIR, usually contains many errors. This makes the electron density maps noisy and difficult to interpret. By applying density modification methods the estimates of the phases can be improved. The method of solvent flattening identifies the protein and solvent regions in the cell and assigns the solvent regions a low electron density value and the protein regions a high electron density value. This improves the phase estimate and an improved electron density map can be calculated that is less noisy. In this work, solvent flattening was carried out by the program DM (Cowtan, 1994) for the experimentally determined electron density maps of AdiA. Since protein crystals contain approximately 30-80% solvent, this can be a powerful method of obtaining better electron density maps.

### 1.3.6 Non-crystallographic symmetry averaging

If the asymmetric unit contains several copies of the protein the NCS between these copies can be exploited to improve the phase estimates. If the copies of the protein in the asymmetric unit are identical the electron density for each molecule should be very similar. Based on this assumption, a NCS mask can be created around one molecule and applied to the other molecules in the asymmetric unit related by the NCS. The correlation of the electron densities enclosed by the NCS masks is maximised in a process of *NCS averaging*. NCS averaging of the electron density can improve the phase estimation depending on how many copies there are in the asymmetric unit. For AdiA, the NCS averaging of the electron density for the ten copies of the monomer in the asymmetric unit improved the overall quality of the electron density map.

### 1.3.7 Model building and refinement

Combining NCS averaging with solvent flattening will improve the quality of the experimentally determined phase estimates. These phase estimates are combined with the observed structure factor amplitudes, $|F_{\text{obs}}|$, from the diffraction data to generate structure factors $F_{\text{obs}}$. An electron density map is calculated by Fourier Transform from these structure factors. The quality of the density modified electron density map is dependent not only on the
experimentally determined phase estimates but also on the quality of the diffraction data and the resolution limit.

In order get a better quality map the only variable that can be improved upon at this point is the phase estimate. The phase estimates are improved by building a model of the protein into the electron density. The electron density maps used in model building are usually weighted maps between the observed structure factors $F_{\text{obs}}$ and the calculated structure factors from the model that is being built $F_{\text{calc}}$. In this work the models were built into $(2F_{\text{obs}} - F_{\text{calc}}, \alpha_{\text{calc}})$ maps and difference density maps $(F_{\text{obs}} - F_{\text{calc}}, \alpha_{\text{calc}})$ were used to evaluate regions in the map that contained either more observed than calculated density or more calculated than observed density.

From the model built into the electron density structure factor amplitudes, $|F_{\text{calc}}|$, are calculated. These are refined against the experimentally observed structure factor amplitudes, $|F_{\text{obs}}|$, in reciprocal space. The aim of the refinement is to reduce the differences between calculated structure factor amplitudes, $|F_{\text{calc}}|$, and the experimentally observed structure factor amplitudes $|F_{\text{obs}}|$. The agreement between $|F_{\text{calc}}|$ and $|F_{\text{obs}}|$ indicates how well the model correlates with the observed diffraction data, which is measured by an $R$-factor (Equation 1.10).

$$R_{\text{work}} = \sum_{hkl(\text{work})} \frac{\| F_{\text{obs}} \| - \| F_{\text{calc}} \|}{\| F_{\text{obs}} \|}$$  \hspace{1cm} (1.10A)

$$R_{\text{free}} = \sum_{hkl(\text{free})} \frac{\| F_{\text{obs}} \| - \| F_{\text{calc}} \|}{\| F_{\text{obs}} \|}$$  \hspace{1cm} (1.10B)

$R_{\text{work}}$ refers to the majority of the reflections used in the model building and refinement process. As the model converges with the data, i.e. the observed and calculated structure factor amplitudes are more in agreement with each other, the $R_{\text{work}}$ value decreases. $R_{\text{free}}$ refers to the small fraction of reflections, usually ~5%, that is set aside and not used in model building in order to provide an unbiased measure of the convergence.

The agreement between the calculated and observed structure factors are improved upon by refining the atom positions, atom occupancies and atom temperature factors of the model. Additional restraints, such as bond angles, bond lengths, torsion angles, planarity, chirality and van der Waals contacts, are applied during refinement, which improves the data to
parameter ratio and improves the accuracy of the model. As model building progresses, through iterative cycles of model building and refinement, the phase estimates are improved, resulting in better quality electron density maps, which allows for the atom positions to be refined with greater accuracy. In this work the program COOT (Emsley & Cowtan, 2004) was used for model building and the program REFMAC5 (Murshudov et al, 1997) for reciprocal refinement.

1.3.8 Model quality

The quality of the finished model can be evaluated by its stereochemistry and geometry, in addition to the $R$-factors. Structure validation programs, like MOLPROBITY (Davis et al, 2007), assesses values such as bond lengths, bond angles and dihedral angles. The $\phi$ and $\psi$ dihedral angles of the polypeptide chain can be plotted as a Ramachandran plot to visualize favoured, allowed and disallowed values of these angles in the protein structure. The temperature factor, $B$, of an atom can also be used to evaluate the quality of the finished model, where a high $B$ factor can indicate a low occupancy, disorder, or dynamic movement in the protein.
PARTICULATE METHANE MONOOXYGENASE

FROM METHYLOCOCCUS CAPSULATUS (BATH)
2.1. INTRODUCTION

2.1.1 Methane monooxygenase (MMO)

Methanotrophs are gram-negative bacteria that grow anaerobically using methane as their only source of carbon and energy. The first step of this metabolic pathway is the oxidation of methane to methanol by methane monooxygenases (MMOs) (Figure 2.1). There are two types of methane monooxygenase, a cytoplasmic soluble methane monooxygenase complex (sMMO) and an integral membrane protein complex, the particulate methane monooxygenase (pMMO). pMMO is the predominant methane oxidation catalyst since it is expressed in almost all known methanotrophs (Hanson & Hanson, 1996), while sMMO is expressed only in a small number. Despite similarity in function, the two proteins are remarkably different in fold, substrate profile and subcellular location (Colby et al, 1977; DeWitt, 1991; Lieberman & Rosenzweig, 2005; Nguyen et al, 1998; Prior, 1985; Rosenzweig et al, 1993). In addition, sMMO is only expressed under conditions of low copper availability and has an active site containing a non-heme binuclear iron cluster (Murrell et al, 2000; Rosenzweig et al, 1993). In contrast, copper is essential for both pMMO stability and activity and the availability of copper in the medium upregulates the expression of pMMO in extensive intracytoplasmic membranes (Stanley, 1983).

![Figure 2.1. Methane oxidation pathway.](image)

The methane monooxygenases sMMO and pMMO have been extensively studied due to their ability to activate the inert CH bond and convert methane to methanol under ambient temperature and pressure. The conversion of methane to methanol is a difficult process industrially that requires expensive technology employing high temperatures and pressures. There is a considerable interest in generating CH activation based synthetic catalysts (Periana, 2004). A greater understanding of the molecular mechanism of pMMO may provide insights into how to improve the efficiency of such synthetic catalysts. There is also potential for pMMO to be used in bioremediation since it also oxidises chlorinated hydrocarbons, the most common groundwater pollutants (Dispirito, 1992; Lontoh & Semrau, 1998).
This work on pMMO, as well as most other studies, has been carried out on pMMO from *Methylococcus capsulatus* (Bath). Hence any discussion of pMMO, unless otherwise specified, will refer to the enzyme from *M. capsulatus* (Bath).

### 2.1.2 The pMMO structures

There are four structures of pMMO, (i) one X-ray crystallographic structure of pMMO from *Methylococcus capsulatus* (Bath) at 2.8 Å resolution, (ii) one X-ray crystallographic structure of pMMO from *Methylosinus trichosporium* OB3b at 3.9 Å resolution, (iii) one EM structure of pMMO from *M. capsulatus* (Bath) at 23 Å resolution and (iv) one EM structure of a pMMO supercomplex from *M. capsulatus* (Bath) at ~16 Å resolution.

#### 2.1.2.1 X-ray structure of pMMO from *M. capsulatus* (Bath)

pMMO consists of three integral membrane protein subunits: pmoB (α subunit, 45 kDa), pmoA (β subunit, 27 kDa) and pmoC (γ subunit, 23 kDa) in a stoichiometry of 1:1:1. The overall architecture of the enzyme was revealed by the X-ray crystallographic structure of pMMO from *M. capsulatus* (Bath) at 2.8 Å resolution (PDB entry 1YEW, (Lieberman & Rosenzweig, 2005)). This structure, referred to below as the 1YEW structure, shows pMMO as a cylindrical αβγ trimer ~105 Å in height and ~90 Å in diameter (Figure 2.2). The pmoA

![Figure 2.2. Structural architecture of pMMO. A. pMMO trimer (sideview) B pMMO trimer (topview). The three protomers are shown in yellow, green and darkblue.](image)
Figure 2.3. The metal sites of the pMMO αβγ protomer. pmoB (α-subunit) is shown in limegreen, pmoA (β-subunit) in orange and pmoC (γ-subunit) in lightblue. Copper ions are shown as gold spheres and the zinc ion as grey sphere. The residues coordinating the metal ions are shown as sticks. Oxygen atoms are coloured red, nitrogen atoms are coloured blue and carbon atoms are coloured according to the subunit.

and pmoC subunits comprise the bulk of the membrane domain, contributing seven and five transmembrane helices, respectively, while pmoB contributes only two transmembrane helices. The pmoB subunit forms the hydrophilic domain of the enzyme with two β-barrels connected to the transmembrane helices by a long loop (Figure 2.3). The 1YEW structure has three metal centers per αβγ protomer: (1) a mononuclear copper site located in pmoB ~25 Å above the membrane, (2) a dinuclear copper site located in pmoB ~10 Å above the membrane and (3) a mononuclear zinc site located in the membrane domain (Figure 2.3). This third metal centre, which has zinc bound, is likely to be an artefact from the zinc acetate in the crystallisation condition rather than being naturally present in pMMO (Lieberman &
Rosenzweig, 2005) and is hence referred to below as the “zinc site”. The physiological content of this “zinc site” is unknown. The histidine ligands coordinating the two copper ions in the dinuclear copper site (pmoB-His33, pmoB-His137, pmoB-His139) are strictly conserved as are the residues coordinating the zinc ion (pmoA-Glu195, pmoC-Asp156, pmoC-His160, pmoC-His173). Of the three residues in the mononuclear copper site (pmoB-His38, pmoB-His72, pmoB-Gln404) only pmoB-His72 is strictly conserved (Hakemian & Rosenzweig, 2007). Although the structural architecture of pMMO was shown in the 1YEW structure, little was revealed about the function of the enzyme. The active site of pMMO is still unidentified and the metal content of the enzyme remains controversial, see below.

2.1.2.2 X-ray structure of pMMO from *M. trichosporium* OB3b

More recently the Rosenzweig lab determined the structure of pMMO from *M. trichosporium* OB3b at 3.9 Å by molecular replacement (Hakemian et al, 2008). *M. trichosporium* OB3b pMMO is also an $\alpha_3\beta_3\gamma_3$ trimer with an additional transmembrane helix present in the pmoC subunit. The dinuclear copper site observed in the structure of *M. capsulatus* (Bath) pMMO was also found in the *M. trichosporium* OB3b pMMO. One significant difference between the two enzymes is that only the pmoB-His72 ligand from the mononuclear copper site in *M. capsulatus* (Bath) pMMO is conserved in the *M. trichosporium* OB3b pMMO structure, and no copper is found at this site. The other significant difference between the two enzymes is that the “zinc site” in *M. capsulatus* (Bath) pMMO is occupied by a copper in *M. trichosporium* OB3b pMMO, which suggests that copper may be the physiological metal ion at this site (Hakemian et al, 2008).

2.1.2.3 EM structures of pMMO and a pMMO supercomplex

The structure of pMMO from *M. capsulatus* (Bath) has also been determined by electron microscopy (EM) to 23 Å resolution (Kitmitto et al, 2005) and as part of a pMMO-supercomplex to ~16 Å resolution (Myronova et al, 2006). The EM structure of pMMO confirmed the trimeric arrangement of pMMO and showed good consensus with the 1YEW structure. The EM structure of the pMMO supercomplex shows the pMMO trimer in complex with a trimer of methanol dehydrogenase (MDH) molecules. MDH is a 65 kDa soluble protein, which catalyses the second step of methane metabolism by oxidizing methanol to formaldehyde. The pMMO-MDH complex thus oxidizes methane to formaldehyde. From a comparison of the pMMO crystal structure and the pMMO-MDH EM structure it is apparent that major conformational changes take place when MDH binds to the pMMO trimer (Figure 2.4). The activity of the pMMO-MDH complex is approximately twice that of pMMO on its own, which was suggested to be due to the structural stabilisation of pMMO within the
pMMO-MDH complex (Myronova et al, 2006). Hence the pMMO-MDH complex represents a more active form of pMMO and a crystal structure of the pMMO-MDH complex could thus reveal more functional aspects of the enzyme.

2.1.3 pMMO active site and metal content

One of the metal sites in pMMO is thought to be the active site, but it has not been possible to positively identify which from the studies on pMMO to date. In order to correctly identify the active site, and to elucidate the chemistry of the pMMO catalysed methane to methanol

![Figure 2.4. EM structure of the pMMO-MDH complex. A. Comparison of the 3D structures of pMMO-MDH and pMMO. pMMO (PDB entry 1YEW) fitted into the EM envelope (sideview). B. 3D volume of pMMO-MDH matched with the X-ray structure of pMMO. The arrows indicate that the EM density of pMMO-MDH is rotated with respect to the hydrophilic domain of pMMO. C. Comparison of EM structures of pMMO (left) and pMMO-MDH (right) illustrating a major conformational rearrangement of the arm densities of pMMO for interaction with the MDH trimer. Figures and legend from (Myronova et al, 2006) and from personal communication with Natalia Myronova.]
conversion, it is important to confirm the number of metal sites, identify the specific metal content and reveal the metal ion coordination. This has proven to be a challenge and is consequently the subject of controversy. The copper content of pMMO has been reported at 2 ions per αβγ protomer (Basu et al, 2003), 2-3 ions per αβγ protomer (Lieberman et al, 2003), 8-10 ions per αβγ protomer (Choi et al, 2003) and ~15 ions per αβγ protomer (Nguyen et al, 1998; Yu et al, 2003). The disagreement regarding the copper content of pMMO has been attributed to the difficulty in purifying an active and stable enzyme that retains its metal cofactors. The presence of iron in pMMO from *M. capsulatus* (Bath) has also been the subject of much debate for the same reason. It has been reported that it is iron free (Nguyen et al, 1998) has 1 per αβγ protomer (Basu et al, 2003) or 2 per αβγ protomer (Zahn & DiSpirito, 1996). EPR studies of pMMO also indicate the presence of iron and a non-heme diiron center has been proposed at the “zinc site” (Myronova, unpublished results, (Martinho et al, 2007).

The 1YEW structure identified three metals sites: a mononuclear copper site, a dinuclear copper site and a mononuclear “zinc site” (Figure 2.3). The number of metal sites present in *M. capsulatus* (Bath) pMMO and the identity of the copper species are, however, still under debate. The debate regarding the metal sites in *M. capsulatus* (Bath) pMMO is particularly relevant since the mononuclear copper site is not present in *M. trichosporium* OB3b pMMO and copper has been found in the “zinc site” (Hakemian et al, 2008). Chan and coworkers suggest the presence of ~15 copper ions in pMMO. Nine of these copper ions are proposed to be sequestered into a Cu (I) domain in the aqueous exposed part of pmoB and act as a buffer of reducing equivalents for the reaction. Six of the copper ions are proposed to form catalytic clusters that readily react with molecular oxygen (Chan et al, 2007; Yu et al, 2007). The identity of the catalytic copper clusters are suggested to be the mononuclear and dinuclear copper sites, identified in the 1YEW structure, and an additional tricopper cluster at a “D-site” in pmoA (Chan et al, 2007; Chan & Yu, 2008). A catalytic site in pmoA is supported by studies with the suicide substrate acetylene. Acetylene is thought to render the enzyme inactive by covalently binding to amino acids in the active site. The pmoA subunit was labelled by radiolabelled acetylene in three different studies (Cook & Shemke, 1996; Prior, 1985; Zahn & DiSpirito, 1996). The pmoA subunit is the least involved in metal coordination in the 1YEW structure, where the three metal sites are located primarily in pmoB and pmoC, with only one ligand (Glu-195 in the “zinc site”) contributed by pmoA. The “D-site” in pmoA, suggested to contain a tricopper cluster by Chan and coworkers, was also identified as a potential metal site by Rosenzweig and coworkers since it contains a cluster of conserved hydrophilic residues (His38, Met42, Asp47, Asp49, Glu100 from pmoA and Glu154 from pmoC) (Balasubramanian & Rosenzweig, 2007). Both Chan and Rosenzweig argue that the
absence of this site in the 1YEW structure could be due to its depletion of metals during purification and/or crystallisation. If so, it is possible that additional metal sites are present in pMMO other than the ones identified in the 1YEW structure. Hence a structure of pMMO from *M. capsulatus* (Bath) produced from a different purification protocol and in a novel crystallisation condition could provide further knowledge on the number of pMMO metal sites, their metal content and the identity of the active site.

**2.1.4 Aims of this study**

The aims of this study were: (1) Obtain the atomic structure of the pMMO-MDH complex by X-ray crystallography in order to reveal more functional aspects of pMMO. (2) Identify the pMMO active site by co-crystallisation of pMMO with suicide substrates or by derivitising crystals with xenon or krypton gas. (3) Analyse the number of metal sites in pMMO and their metal content using anomalous difference Fourier maps.

**2.2 METHODS AND MATERIALS**

**2.2.1 Protein preparation**

**2.2.1.1. Protein purification**
Native pMMO and pMMO-MDH protein samples were purified from *Methyllococcus capsulatus* (Bath) by Natalia Myronova at Warwick University (Basu et al, 2003) and supplied in a buffer containing 25 mM PIPES pH 7.2, 1 mM benzamidine and 0.03% dodecyl-ß-D-maltopyranoside (DDM) (Buffer A). The samples were concentrated using an Amicon ultracentrifugation device with a molecular weight cutoff of 100 kDa. Crystallisation trials were carried out with protein at final concentration of 15-20 mg/ml as evaluated by UV spectrophotometry at 280 nm.

**2.2.1.2 Suicide substrates**
Acetylene (HC≡CH) acts as a suicide substrate, covalently modifying a residue close to the active site and rendering the enzyme inactive (Prior, 1985). Propargylamine (HC≡CCH₂NH₂) has been suggested as another potential suicide substrate acting in a similar mode to acetylene (Takeguchi, 2000). pMMO and pMMO-MDH protein samples were incubated with the suicide substrate acetylene and the potential suicide substrate propargylamine by Dr Natalia
Myronova at Warwick University and supplied in Buffer A. Samples were concentrated as described.

2.2.1.3 Crosslinked pMMO-MDH
Cross-linking of pMMO-MDH was performed by means of dithiobis-sulfosuccinimidyl propionate (DTSSP) (Myronova et al, 2006). DTSSP is a membrane impermeable cross linker, hence it cross-links between the pMMO hydrophilic domain and MDH. Samples of crosslinked pMMO-MDH with DTSSP were prepared by Dr Natalia Myronova at Warwick University and supplied in Buffer A. Samples were concentrated as described.

2.2.2 Crystallisation

2.2.2.1 1YEW crystallisation condition
The buffer of pMMO protein samples was exchanged from Buffer A to the protein buffer of 50 mM Hepes pH 7.5 and 0.12% Cymal-5 by washing the sample several times using an Amicon ultracentrifugation device with a 100 kDa cutoff. Grid screens around the published condition for pMMO crystallisation of 200 mM zinc acetate and 8-12% PEG 8000 (Lieberman & Rosenzweig, 2005) were set up in a 24-well plate format at 20°C using the hanging drop vapour diffusion technique.

2.2.2.2 Novel crystallisation of pMMO: Conditions A-G
Crystallisation conditions were screened at 293 K, according to the sparse matrix method (Jancarick, 1991), using commercially available screens (Crystal Screen 1 and 2 from Hampton Research, Wizard I and II from Emerald BioSystems, MemStart/MemSys from Molecular Dimensions) and the sitting drop vapour diffusion technique. Screens were set up in 96-well Corning round bottom plates using a Cartesian Microsys robot. Drops consisted of pMMO-MDH protein in Buffer A (0.2 µl) with an equal volume of reservoir solution. Initially, crystals were observed in condition 37 of Crystal Screen 2 (0.1 M Hepes pH 7.5, 10% PEG 8000, 8% ethylene glycol), here referred to as Condition A, and condition 11 of the Wizard II screen (0.1 M sodium cacodylate pH 6.5, 0.2 M zinc acetate, 10% 2-propanol), here referred to as Condition B. Crystals were optimized through a sparse matrix additive screen where 25 µl of commercial screen solution was added to 75 µl of condition A in the reservoir (Birtley & Curry, 2005). The crystallisation drops were setup using 96-well plates, as described. Visually improved crystals were found in condition C (0.075 M Hepes pH 7.5, 0.025 M sodium cacodylate pH 6.5, 6.5% PEG 8000, 6.75% ethylene glycol, 0.05 M zinc acetate, 2.5% 2-propanol), which was identified as the combination of conditions A and B at a
ratio of 3:1. Crystal optimisation grid screens were set up manually in 24-well Linbro plates, using the hanging drop vapour diffusion technique, and left to equilibrate at 20°C. Drops consisted of, as previously, pMMO-MDH protein in Buffer A (1.0 µl) with an equal volume of reservoir solution. Further optimisation in the 24 well format included exchanging 2-propanol for dioxane (Condition D), removing the HEPES buffer and increasing the concentration of the sodium cacodylate buffer (Condition E), lowering the pH of the sodium cacodylate buffer (Condition F) and identifying spermidine as an additive (Condition G). Additive screening was carried out using the Additive Screen from Hampton Research where 1 µl of additive solution was mixed with 9 µl of reservoir solution and 1 µl of this mixture was added to the protein solution in a 1:1 ratio as previously described. Condition G produced crystals of typical dimensions of 0.3 x 0.1 x 0.08 mm³ in 0.1 M sodium cacodylate pH 6.0, 6% PEG 8000, 6% ethylene glycol, 50 mM zinc acetate, 3% dioxane and 0.01 M spermidine. Typically, crystals appeared after 2-3 days and grew to their maximum dimensions in 7-10 days. For a summary of Conditions A-G, see Table 2.1.

2.2.2.3 Further optimisation trials of pMMO crystals

2.2.2.3.1 G-type condition
Grid screening based on Condition G produced a G-type condition in which pMMO usually crystallised: 0.1-0.2 M sodium cacodylate pH 6.0-6.5, 6-8% PEG 8000, 6-10% ethylene glycol, 0.05 M zinc acetate, 3% dioxane, 0.01-0.02 M spermidine.

2.2.2.3.2 Detergent screens
The DDM detergent in pMMO samples was exchanged by diluting approximately 250 µl of concentrated pMMO sample at 15-20 mg/ml concentration with 15 ml of protein buffer, containing 25 mM PIPES pH 7.2, 1 mM benzamidine and 2-3x the critical micelle concentration (CMC) of the new desired detergent, followed by concentration of the sample using an Amicon Ultra centrifugation device. This procedure was repeated three times to ensure maximum replacement of DDM by the new detergent. pMMO was exchanged into the following detergents: tridecyl-β-D-maltopyranoside (TDM), undecyl-β-D-maltopyranoside (UDM), decyl-β-D-maltopyranoside (DM) and nonyl-β-D-maltopyranoside (NM), octyl-β-D-maltopyranoside (OM), decyl-β-D-glucopyranoside (DG), nonyl-β-D-glucopyranoside (NG) and octyl-β-D-glucopyranoside (OG) purchased from Anatrace. Crystallisation trials were performed in a 24-well plate setup using identified crystallisation conditions.
### Table 2.1. Crystallisation conditions A-G

<table>
<thead>
<tr>
<th>Condition</th>
<th>Solution Components</th>
</tr>
</thead>
</table>
| **Condition A** | 0.1 M Hepes pH 7.5  
10% PEG 8000  
8% Ethylene Glycol |
| **Condition B** | 0.1 M sodium cacodylate pH 6.5  
0.2 M zinc acetate  
10% 2-propanol |
| **Condition C** | 0.075 M Hepes pH 7.5  
0.025 M sodium cacodylate pH 6.5  
6.5% PEG 8000  
6.75% ethylene glycol  
0.05 M zinc acetate  
2.5% 2-propanol |
| **Condition D** | 0.075 M Hepes pH 7.5  
0.025 M sodium cacodylate pH 6.5  
7% PEG 8000  
6.75% ethylene glycol  
0.05 M zinc acetate  
3% dioxane |
| **Condition E** | 0.1 M sodium cacodylate pH 6.5-6.9  
6% PEG 8000  
6% ethylene glycol  
0.05 M zinc acetate  
3% dioxane  
0.01 M spermidine |
| **Condition F** | 0.1 M sodium cacodylate pH 6.0C  
6% PEG 8000  
6% ethylene glycol  
0.05 M zinc acetate  
3% dioxane |
| **Condition G** | 0.1 M sodium cacodylate pH 6.0  
6% PEG 8000  
6% ethylene glycol  
0.05 M zinc acetate  
3% dioxane  
0.01 M spermidine |
| **G-type condition** | 0.1-0.2 M sodium cacodylate pH 6.0-6.5  
6-8% PEG 8000  
6-10% ethylene glycol  
0.05 M zinc acetate  
3% dioxane  
0.01-0.02 M spermidine |

#### 2.2.2.3.3 Dehydration experiments

Two strategies for dehydrating pMMO crystals were employed in 24 well plate setups: (1) pMMO crystals grown in a G-type condition were transferred to drops containing mother liquor with increased amounts of the precipitant PEG 8000 (7-20%). The drops were left to equilibrate over reservoir solutions containing respectively increased amounts of PEG 8000 (7-20%) for 2-24 hours before being cryo-protected and cryo-cooled in liquid nitrogen as
described. (2) pMMO crystals grown in a G-type condition were transferred to drops containing mother liquor and left to equilibrate over reservoir solutions containing increasing amounts of PEG 8000 (7-20%) for 24-48 hours before being cryo-protected and cryo-cooled in liquid nitrogen as described.

2.2.2.3.4 Crystal soaks at lower pH
pMMO crystals grown in G-type conditions were transferred to drops containing mother liquor with 0.1 M sodium cacodylate buffer at pH 4.5, 5.0 and 5.5. The drops were left to equilibrate for 1-48 hours over reservoir solutions of corresponding pH prior to the cryo-protection and cryo-cooling of the crystals.

2.2.2.3.5 Crystal soaks with iron salts
pMMO crystals grown in G-type conditions were transferred to drops containing mother liquor with added 0.1 M of FeCl₂/FeSO₄/Fe(NO₃)₃ salt for 5min-2hours prior to cryo-protection and cryo-cooling.

2.2.2.4 Suicide substrates
Crystallisation trials with pMMO samples incubated with the suicide substrate acetylene and the proposed suicide substrate propargylamine (PA) were set up in a 24 well plate setup, as described. The protein + PA samples generally crystallised in G-type conditions.

2.2.2.5 Derivitisation with xenon and krypton
Xenon or krypton atoms bound to pMMO can potentially mimic where the gaseous substrate methane would bind to the protein, for instance in the active site and/or in a channel leading to the active site (Prange et al, 1998). Hence pMMO crystals were pressurized with 6.9 - 13.8 bar xenon (Xe) or krypton (Kr) gas for 1-3 minutes using a xenon gas pressure cell prior to immediate cryo-cooling. These pMMO crystals were mounted in a loop on a magnetic base, which is inserted into the gas pressure cell chamber. The gas pressure cell chamber is kept at a humidity mimicking the drop with 100 µl reservoir solution placed at the bottom of the cell in order to avoid dehydration of the crystal. The pressure in the cell was increased to 6.9 - 13.8 bar for 1-3 minutes before releasing the gas from the cell. Crystals were cryo-cooled within 2-4 seconds after the pressure cell was opened.
2.2.2.6 Crystallisation trials of pMMO-MDH

2.2.2.6.1 Cross-linked pMMO-MDH
Crystallisation conditions were screened at 20°C, according to the sparse matrix method (Jancarick, 1991), using commercially available screens (Crystal Screen 1 and 2 from Hampton Research, Wizard I and II from Emerald BioSystems, MemStart/MemSys from Molecular Dimensions) and the sitting drop vapour diffusion technique. Screens were set up at room temperature in 96-well Corning round bottom plates using a Cartesian Microsys robot and stored at 20°C. Drops consisted of cross-linked pMMO-MDH protein in Buffer A (0.2 µl) with an equal volume of reservoir solution.

2.2.2.6.2 Low salt screens
PEG vs pH grid screens were designed that contained no added salt in order to minimize the ionic strength of the conditions (Table 2.2). These low salt screens were set up at room temperature with a Cartesian Microsys robot in a 96 well setup as described, and incubated at 20°C and 4°C. Crystals that grew in 0.1 M Tris pH 9.0, 15% PEG 8000 were cryoprotected by a quick immersion in a solution containing 0.1 M Tris pH 9.0, 15% PEG 8000 and 30% glycerol prior to cryo-cooling.

2.2.3 Data collection and structure determination

2.2.3.1 A 3.5 Å dataset from a crystal grown in condition G
A pMMO crystal grown from pMMO-MDH protein solution in condition G was cryoprotected by transferring the crystal to a drop containing the reservoir solution with an ethylene glycol concentration of 20% (6% PEG 8000, 3% dioxane, 0.05 M zinc acetate, 0.1 M sodium cacodylate pH 6.0, 0.01 M spermidine, 20% ethylene glycol) immediately prior to cryo-cooling. Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, at the ID14-4 beamline. 90 images were collected for a full data set, corresponding to 45° of data. The crystal was translated every 15° in order to minimize the effects of radiation damage. A suitable data collection strategy was determined using the program MOSFLM (Leslie, 1992) and the data was scaled and processed using the program packages DENZO and SCALEPACK (Otwinowski, 1997). Molecular replacement was performed using the program PHASER (Read, 2001) from the CCP4 program suite and the 1YEW published structure of pMMO (Lieberman & Rosenzweig, 2005) was used as the search model. Initial rigid body and restrained refinement, applied with tight non-crystallographic symmetry and geometry restraints, was carried out in REFMAC5.
Table 2.2. Low salt PEG vs pH screens for pMMO-MDH crystallisation

<table>
<thead>
<tr>
<th>PEG</th>
<th>Range</th>
<th>Buffer</th>
<th>pH range</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>20-40%</td>
<td>Glycine, Tris, Hapes, Sodium cacodylate, MES, Sodium citrate</td>
<td>9.0-10.0, 7.0-9.0, 6.8-8.4, 5.5-7.0, 5.0-6.5, 3.0-5.0</td>
</tr>
<tr>
<td>2000</td>
<td>2-25%</td>
<td>Glycine, Tris, Hapes, Sodium cacodylate, MES, Sodium citrate</td>
<td>9.0-10.0, 7.0-9.0, 6.8-8.4, 5.5-7.0, 5.0-6.5, 3.0-5.0</td>
</tr>
<tr>
<td>3350</td>
<td>5-25%</td>
<td>Glycine, Tris, Hapes, Sodium cacodylate, MES, Sodium citrate</td>
<td>9.0-10.0, 7.0-9.0, 6.8-8.4, 5.5-7.0, 5.0-6.5, 3.0-5.0</td>
</tr>
<tr>
<td>6000</td>
<td>3-23%</td>
<td>Glycine, Tris, Hapes, Sodium cacodylate, MES, Sodium citrate</td>
<td>9.0-10.0, 7.0-9.0, 6.8-8.4, 5.5-7.0, 5.0-6.5, 3.0-5.0</td>
</tr>
<tr>
<td>8000</td>
<td>5-20%</td>
<td>Glycine, Tris, Hapes, Sodium cacodylate, MES, Sodium citrate</td>
<td>9.0-10.0, 7.0-9.0, 6.8-8.4, 5.5-7.0, 5.0-6.5, 3.0-5.0</td>
</tr>
</tbody>
</table>

(Murshudov et al, 1997) and gave residuals of $R_{work} = 30.7 \%$ and $R_{free} = 32.0 \%$. The data processing and refinement statistics are given in Table 2.3. Anomalous data was obtained from the same crystal at ESRF on beamline ID29. Data was collected near the zinc absorption edge at 9.686 keV, near the copper absorption edge at 8.985 keV, near the iron absorption edge at 7.167 keV and below the iron absorption edge at 7.085 keV, corresponding to wavelengths of 1.28 Å, 1.38 Å, 1.73 Å and 1.75 Å, respectively. For each dataset 90 images, corresponding to 45° of data, were collected and data was scaled and processed using the program packages DENZO and SCALEPACK (Otwinowski, 1997). The data collection and processing statistics are presented in Table 2.3. Anomalous difference Fourier density maps were calculated with the program FFT (Read, 1988) from the CCP4 program suite.
Table 2.3. Data collection, processing and refinement statistics for the 3.5 Å pMMO model

<table>
<thead>
<tr>
<th>Data set</th>
<th>pMMO</th>
<th>ZnANOM</th>
<th>CuANOM</th>
<th>FeANOM</th>
<th>Fe low energy ANOM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection and processing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>$P4_{2,2}$</td>
<td>$P4_{2,2}$</td>
<td>$P4_{2,2}$</td>
<td>$P4_{2,2}$</td>
<td>$P4_{2,2}$</td>
</tr>
<tr>
<td>Cell dimensions a, c (Å)</td>
<td>270.2, 150.7</td>
<td>270.2, 150.3</td>
<td>270.1, 150.5</td>
<td>270.4, 150.5</td>
<td>270.3, 150.4</td>
</tr>
<tr>
<td>ESRF beamline</td>
<td>ID14-4</td>
<td>ID29</td>
<td>ID29</td>
<td>ID29</td>
<td>ID29</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.94</td>
<td>1.28</td>
<td>1.38</td>
<td>1.73</td>
<td>1.75</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.0-3.5</td>
<td>50.0-3.9</td>
<td>50.0-3.9</td>
<td>50.0-3.9</td>
<td>50.0-3.9</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>165919</td>
<td>141974</td>
<td>142178</td>
<td>144102</td>
<td>145329</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>60671</td>
<td>46560</td>
<td>46045</td>
<td>464102</td>
<td>46730</td>
</tr>
<tr>
<td>Completeness (%)$^a$</td>
<td>85.9 (70.2)</td>
<td>90.6 (74.6)</td>
<td>90.2 (72.7)</td>
<td>91.3 (77.7)</td>
<td>91.6 (72.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.7</td>
<td>3.0</td>
<td>3.1</td>
<td>3.1</td>
<td>3.1</td>
</tr>
<tr>
<td>$I/\sigma(I)$</td>
<td>8.8 (1.6)</td>
<td>8.2 (1.3)</td>
<td>12.6 (3.0)</td>
<td>11.3 (2.3)</td>
<td>9.0 (1.5)</td>
</tr>
<tr>
<td>$R_{sym}$ (%)$^b$</td>
<td>10.2 (44.3)</td>
<td>12.6 (71.0)</td>
<td>8.0 (31.0)</td>
<td>9.0 (39.5)</td>
<td>11.1 (54.3)</td>
</tr>
<tr>
<td><strong>Model refinement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{work}$ (%)$^c$</td>
<td>30.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{free}$ (%)$^d$</td>
<td>32.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average $B$ factor (Å$^2$)</td>
<td>86.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stereocchemistry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root mean square deviation for bond length (Å)</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root mean square deviation for bond angles (°)</td>
<td>1.303</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Residues in the Ramachandran plot$^e$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored regions (%)</td>
<td>87.4 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allowed regions (%)</td>
<td>97.4 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Statistics for the highest resolution shell are given in parentheses.

$^b R_{sym} = \frac{\sum_h \sum_i |I_{hi}| - \langle I_h \rangle}{\sum_h \sum_i |I_{hi}|}$

$^c R_{work} = \frac{\sum |F_{obs} - F_{calc}|}{\sum F_{obs}}$ for all reflections.

$^d R_{free} = \frac{\sum |F_{obs} - F_{calc}|}{\sum F_{obs}}$ calculated using randomly selected reflections (5%).

$^e$ Ramachandran plot values as assessed by MOLPROBITY.

The ANOM datasets were used for the calculation of anomalous maps.
2.2.3.2 A 8.2 Å dataset from a crystal grown in condition D

A pMMO crystal grown from pMMO-MDH protein solution in condition D (0.075 M HEPES pH 7.5, 0.025 M sodium cacodylate pH 6.5, 7% PEG 8000, 6.75% ethylene glycol, 0.05 M zinc acetate, 3% dioxane) was cryoprotected by briefly immersing the crystal in a drop containing the reservoir solution with an ethylene glycol concentration of 20% immediately prior to transfer into liquid nitrogen. Diffraction data were collected at the ESRF at the ID14-4 beamline. 180 images were collected, corresponding to 90° of data. The data was scaled and processed using the program packages DENZO and SCALEPACK (Otwinowski, 1997). Molecular replacement was performed as described. Initial rigid body and restrained refinement were carried out in REFMAC5 (Murshudov et al, 1997). Data collection and processing statistics are presented in Table 2.4.

2.2.3.3 A 4.0 Å dataset from a crystal potentially co-crystallised with propargylamine

A crystal grown in a G-type condition from pMMO protein solution incubated with propargylamine was cryoprotected by transferring the crystal to a drop of mother liquor (6% PEG 8000, 6% ethylene glycol, 3% dioxane, 0.05 M zinc acetate, 0.1 M sodium cacodylate pH 6.0, 0.01 M spermidine, 20 mM PIPES pH 7.2, 1 mM benzamidine, 0.03% DDM) and increasing the ethylene glycol concentration in the drop to a final concentration of 20% in ca. 7% increments. The crystal was briefly immersed into a second drop (1 μl) containing 20% ethylene glycol, immediately prior to transfer into liquid nitrogen. 140 images were collected at ESRF on beamline BM-14, corresponding to 70° of data. The data was scaled and processed using the program packages DENZO and SCALEPACK (Otwinowski, 1997). Molecular replacement and initial refinement was performed as described. Data collection and processing statistics are presented in Table 2.4.

2.2.3.4 A 4.7 Å dataset from a crystal pressurised with krypton gas

A crystal grown in a G-type condition was taken from the drop and directly pressurized with krypton gas, as described. The crystal was cryo-cooled within 2-4 seconds after pressurization without any additional cryoprotection. Data were collected at ESRF on beamline ID23-1 near the krypton absorption edge at 14.345 keV, corresponding to a wavelength of 0.86 Å. Alternating 5° batches were collected at angle φ and angle φ-180°. 2x90 images, corresponding to 2x45° of data, were collected in total using this inverse beam strategy in order to maximize the anomalous signal between the Bijvoet pairs. The data was scaled and processed using the program packages DENZO and SCALEPACK (Otwinowski, 1997). Molecular replacement and initial refinement was performed as described. Anomalous
difference Fourier density maps were calculated with the program FFT (Read, 1988) from the CCP4 program suite. Data collection and processing statistics are presented in Table 2.4.

**Table 2.4. Data collection and processing statistics for a 8.2 Å dataset, PA dataset and Kr dataset**

<table>
<thead>
<tr>
<th>Data set</th>
<th>8.2 Å native</th>
<th>PA dataset‡</th>
<th>Kr dataset‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection and processing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P4₁,2,2</td>
<td>P4₁,2,2</td>
<td>P4₁,2,2</td>
</tr>
<tr>
<td>Cell dimensions a, c (Å)</td>
<td>269.2, 152.4</td>
<td>267.4, 151.5</td>
<td>270.6, 151.5</td>
</tr>
<tr>
<td>ESRF beamline</td>
<td>ID14-4?</td>
<td>BM14</td>
<td>ID23-1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.07</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.0-8.2</td>
<td>50.0-4.0</td>
<td>50.0-4.7</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>35567</td>
<td>249046</td>
<td>190826</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>5268</td>
<td>45987</td>
<td>55606</td>
</tr>
<tr>
<td>Completeness (%)ᵃ</td>
<td>99.2 (99.2)</td>
<td>99.0 (97.5)</td>
<td>99.3 (98.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.8</td>
<td>5.4</td>
<td>3.4</td>
</tr>
<tr>
<td>I/σ(I)ᵇ</td>
<td>27.3 (8.0)</td>
<td>12.8 (2.8)</td>
<td>8.9 (1.9)</td>
</tr>
<tr>
<td>R&lt;sub&gt;sym&lt;/sub&gt; (%)ᵃ</td>
<td>11.0 (35.2)</td>
<td>11.4 (49.5)</td>
<td>17.5 (78.5)</td>
</tr>
</tbody>
</table>

ᵃ Statistics for the highest resolution shell are given in parentheses.
ᵇ R<sub>sym</sub> = ΣₙΣₖ|Iₙₖ| - Σₙ|Iₙₖ|/ΣₙΣₖ|Iₙₖ|
ᶜ PA dataset - data from crystal grown from pMMO sample incubated with propargylamine.
ᵈ Kr dataset - data from crystal pressurized with krypton gas.

**2.3 RESULTS**

**2.3.1 Crystallisation of pMMO**

**2.3.1.1 Conditions A-G**

2.3.1.1 The crystallisation and optimisation processes

Efforts to reproduce pMMO crystals in grid screens based on the published crystallisation condition using protein sample in 1YEW protein buffer were unsuccessful. In parallel with these efforts, crystallisation trials for pMMO-MDH were set up using the Cartesian Microsys robot and commercial screen kits in a 96-well sitting drop setup. Condition A, which produced oval colourless crystalline discs, and condition B, which gave rise to dense crystal needle clusters were identified (Figure 2.5). Grid screen optimisation trials based on these
Figure 2.5. Optimisation process of Conditions A-G.
conditions and set up in a 24-well hanging drop setup were unsuccessful. A sparse matrix additive screen in a 96-well sitting drop setup produced single rod shaped crystals in Condition C, which was a combination of Condition A and B in a 3:1 ratio. These rod crystals were reproduced in a 24 well hanging drop setup and were amenable to optimisation in several steps of screening (Figure 2.5): (i) The 2-propanol was replaced by dioxane (Condition D), another organic solvent that produced significantly larger crystals with approximate dimensions of 80 x 30 x 10 µm. (ii) The HEPES buffer was removed and the sodium cacodylate concentration increased to 0.1 M (Condition E). This step made the crystals more reproducible, in contrast to previous conditions, where the precipitation of the zinc acetate when the pH exceeded 7.0 prevented reproducible crystal formation. The presence of zinc acetate in the crystallisation condition was essential and non-replaceable for crystal formation to occur. (iii) The pH was decreased to 6.0 (condition F) and the crystals grew larger, which improved the resolution of the diffraction significantly. (iv) Spermidine was identified as an additive that helped crystal formation and improved the quality of the crystals (Condition G). After this process of sparse-matrix, additive and grid crystallisation trials the final crystallisation condition obtained (Condition G) yielded a crystal that produced a 3.5 Å dataset.

2.3.1.1.2 Data collection
The initial crystals from Conditions A and B were confirmed as protein crystals as they diffracted weakly to approximately 18-25 Å. A full dataset could be collected from a crystals grown in Condition D to 8.2 Å resolution with a tetragonal space group of dimensions a = b = 269.1 Å, c = 152.5 Å. These unit cell dimensions are very similar to the unit cell of the published pMMO structure (a = b = 264.1 Å, c = 150.0 Å) (Lieberman & Rosenzweig, 2005). The similarity between these unit cell dimensions indicated that pMMO had crystallised instead of the full pMMO-MDH complex. This was confirmed to be the case by a successful molecular replacement using the published pMMO structure (PDB entry 1YEW) as a model. The diffraction limit of the crystals increased significantly from ~ 8 Å to 3.5 Å when they grew larger in condition F and G. A native dataset was collected from one of the largest crystals from Condition G at beamline ID14-4 at ESRF, France (Figure 2.6). The data was processed and scaled to a resolution of 3.5 Å, as summarised in Table 2.3.
Figure 2.6. pMMO diffraction pattern to 3.5 Å resolution. A. pMMO crystal diffraction pattern, the edge of the plate is at 3.0 Å. B. Crystal of pMMO used for structure determination. C. Close up of the pMMO diffraction pattern.

2.3.1.2 Optimisation trials of G-type conditions

2.3.1.2.1 Detergent screening
pMMO samples were exchanged into several different maltoside or glucoside detergents. Crystals grew from pMMO samples exchanged into maltoside detergents tridecyl-β-D-maltopyranoside (TDM), undecyl-β-D-maltopyranoside (UDM), decyl-β-D-maltopyranoside (DM) and nonyl-β-D-maltopyranoside (NM), but did not grow if the detergent was exchanged to the shorter chain octyl-β-D-maltopyranoside (OM) or in the glucoside detergents decyl-β-D-glucopyranoside (DG), nonyl-β-D-glucopyranoside (NG) or octyl-β-D-glucopyranoside (OG). No direct correlation between carbon chain length in the maltoside detergents and crystal quality could be conclusively determined. The best quality crystals grew in DDM, one of the longer chain detergents, in G-type conditions.

2.3.1.2.2 Dehydration experiments
The dehydration experiments reduced or abolished diffraction from these crystals.
2.3.1.2.3 Crystal soaks at lower pH
Soaking crystals in conditions containing a sodium cacodylate buffer with a lower pH (pH 4.5-5.5) had no effect on crystal quality in comparison to crystals not soaked in a condition with sodium cacodylate buffer of lower pH.

2.3.1.2.4 Crystal soaks with iron salts
pMMO crystals were soaked in mother liquor containing FeCl$_2$/FeSO$_4$/Fe(NO$_3$)$_3$ with the aim that, if the third metal site naturally contains iron, it would preferentially bind to iron instead of zinc. Crystals remained intact and visually undamaged during iron chloride soaks, but they did not diffract.

2.3.1.3 Crystallisation trials of pMMO-MDH

2.3.1.3.1 Cross-linked pMMO-MDH
No crystals were obtained from cross-linked pMMO-MDH in the crystallisation screens prepared to date.

2.3.1.3.1 Low salt screens
Several new pMMO crystallisation conditions were identified in the low salt PEG vs pH screens. Crystals of approximate dimensions 15 x 15 x 30 mm$^3$ grew in 0.1 M Tris pH 9.0, 15% PEG 8000. These crystals showed diffraction to approximately 8 Å at the beamline I04 at Diamond Synchrotron Light Source (data not shown).

2.3.2 The 3.5 Å pMMO model

2.3.2.1 Quality of the model
Molecular replacement using the 1YEW structure as the search model was straightforward since it is the same protein (pMMO) from the same organism (M. capsulatus (Bath)), crystallised in the same space group (P4$_2$2$_1$2) with one $\alpha_3\beta_3\gamma_3$ trimer in the asymmetric unit. For the same reasons initial refinement with tight geometry and non-crystallographic symmetry restraints produced reasonable residuals of $R_{\text{work}} = 30.7\%$ and $R_{\text{free}} = 32.0\%$ despite the low 3.5 Å resolution data. These values compare well with the 1YEW residuals of $R_{\text{work}} = 27.3\%$ and $R_{\text{free}} = 30.2\%$ at 2.8 Å resolution. In spite of applying tight geometry and non-crystallographic symmetry restraints the 3.5 Å model shows questionable geometry with only 87.4% in the favoured Ramachandran region and 97.4% in the allowed region as assessed by MOLPROBITY (Davis et al, 2007). This is a reflection of the geometry values of the starting
model 1YEW used in molecular replacement, which has 84.3 % and 95.8% in the favoured and allowed Ramachandran regions respectively. As expected after molecular replacement with 3.5 Å data, the pMMO structure does not differ from that of 1YEW in its overall architecture, which is highlighted by an rmsd value of 0.43 Å for a global Cα superposition of the two structures. The $\alpha_3\beta_3\gamma_3$ trimer consists of pmoB (chain A, E, I), pmoA (chain B, F, J) and pmoC (chain C, G, K). The structural architecture of 1YEW, also seen in the 3.5 Å pMMO model presented here (Figure 2.7), has been described in detail elsewhere (Balasubramanian & Rosenzweig, 2007; Lieberman & Rosenzweig, 2005).

2.3.2.2 Metal site identities

Data were collected near the absorption edges of zinc (Zn), copper (Cu) and iron (Fe), and below the iron absorption edge. By calculating anomalous difference Fourier density maps any zinc, copper or iron bound to the protein can be visualized. Since the Zn absorption edge

Figure 2.7. The 3.5 Å pMMO model. The $\alpha_3\beta_3\gamma_3$ protomer of the pMMO trimer is highlighted in colour. pmoB is shown in cyan, pmoA in pink and pmoC in yellow. The metal centres, as identified in the 1YEW structure, are shown as spheres. Copper ions are shown as gold spheres and zinc as a grey sphere.
is at a higher energy than the Cu absorption edge, the anomalous signal from Cu will also be present in the Zn anomalous difference Fourier maps. Zn sites can be distinguished from the Cu sites by the peaks in the Zn anomalous difference Fourier maps not present in the Cu anomalous difference Fourier maps. Similarly any Fe sites can be distinguished from other elements, such as calcium or potassium, by comparing the anomalous difference Fourier maps from data collected above and below the Fe absorption edge. By analysing anomalous difference Fourier density maps for copper, zinc and iron, it was possible to confirm a mononuclear copper site, a dinuclear copper site and a zinc site in \textit{M. capsulatus} (Bath) pMMO, as found in the 1YEW structure (Figure 2.8). No iron was found in the enzyme, nor any tricopper cluster or any other detectable metal site. Six zinc sites mediating trimer crystal contacts were identified (Figure 2.9).

2.3.3 Other datasets

2.3.3.1 Propargylamine complex dataset at 4.0 Å

This data was obtained from a crystal grown from a pMMO sample incubated with propargylamine. It was not possible to identify any bound propargylamine in the 4.0 Å resolution electron density or difference(147,903),(877,921)(147,921),(877,938)(147,938),(877,955)(147,956),(877,973)(147,973),(877,991)(147,991),(877,1008) density maps.

2.3.3.2 Krypton derivative dataset at 4.7 Å

This data was obtained from a crystal pressurized with krypton gas. It was not possible to positively identify any krypton binding sites in the 4.7 Å anomalous difference Fourier density map.

---

**Figure 2.8. Metal sites in the pMMO trimer. A1-A3.** ‘Mononuclear’ copper sites in chain A, E and I, respectively. \textbf{B1-B3}. ‘Dicopper’ sites in chain A, E and I, respectively. \textbf{C1-C3}. ‘Zinc sites’ in chain C, G and K, respectively. Chain A (cyan), chain B (pink), chain C (yellow), chain E (violet), chain F (salmon), chain G (teal), chain I (orange), chain J (limon), chain K (brown). Metal ions from the 1YEW structure. Copper ions are shown as gold spheres and zinc ions are shown as grey spheres. The 2FoFc electron density map is contoured at 1σ and shown in pale cyan. Positive difference density contoured at 3σ is shown in green. Negative difference density contoured at 3σ is shown in red. The anomalous difference Fourier map for copper contoured at 3σ is shown in orange and contoured at 5σ in magenta. The anomalous difference Fourier map for zinc contoured at 3σ is shown in grey and contoured at 5σ in purple. In the ‘monocopper’ sites the anomalous difference Fourier map for zinc contoured at 3σ is shown in black for clarity.
Figure 2.8. Metal sites in the pMMO trimer. For figure legend, see previous page.
Figure 2.9  Zn mediated crystal contacts in the pMMO trimer. A. The pMMO trimer coloured according to chains: Chain A (cyan), chain B (pink), chain C (yellow), chain E (violet), chain F (salmon), chain G (teal), chain I (orange), chain J (limon), chain K (brown). The residues coordinating the zinc ions in the crystal contacts are shown as spheres and coloured according to zinc site 1-6, see table. B. The pMMO trimer (blue) forms crystal contacts with the pmoB hydrophilic domains of five other trimers. In addition one crystal contact is formed between pmoC (chain C) and pmoC (chain C) of another trimer in the crystal. The residues involved in crystal contacts mediated by zinc ions are shown as spheres and coloured according to zinc site 1-6, see table.
2.4 DISCUSSION

2.4.1 The pMMO crystallisation process

2.4.1.1 The initial optimisation process: Condition A-G
The process of optimisation of the initial pMMO crystals highlighted the importance of several factors required to produce well-diffracting pMMO crystals. The first important step was to lower the pH from 7.5 to 6.0-6.5, which eliminated zinc acetate precipitation and made the crystals larger and more reproducible (Condition D → E). The second important factor was the screening of additives, exemplified by the switch of 2-propanol to dioxane (Condition C → D) and the effect of spermidine (Condition G). Spermidine improves the quality of the diffraction pattern by making the spots smaller and more well defined, indicating that it improves the long-range order in the crystal. Spermidine also facilitated crystal nucleation, as drops containing spermidine as an additive more often produced crystals than drops containing no spermidine.

The most important factor in improving the diffraction quality from pMMO crystals was the crystal size. pMMO crystals, like many membrane protein crystals, show very weak diffraction, especially at higher resolution. These weak spots can be impossible to distinguish from the background, which lowers the resolution limit of the data collected. The thickness of the crystal, i.e. the size of the crystal, amplified the signal to noise in the diffraction patterns since more protein particles are involved in X-ray scattering.

2.4.1.2 Further optimisation trials of the G-type conditions
The larger size of pMMO crystals could, however, improve the diffraction limit up to a point only. By increasing the sodium cacodylate concentration from 0.1 M to 0.2 M (Condition G → G-type conditions) larger crystals with a thickness of ~150 µm were generally grown, but the diffraction limit of approximately 4 Å did not improve compared to crystals with a thickness of ~100 µm.

Other approaches were tried to improve the diffraction quality of crystals from G-type conditions: (i) Since the quality of membrane protein crystals often depends on the type of detergent, specifically the detergent chain length, the detergent in the pMMO protein solution was exchanged from DDM to detergents with either shorter or longer chain and with
maltoside or glucoside head groups. The detergent of choice, however, remained DDM since the best diffracting crystals were obtained with this detergent and no visible change in diffraction pattern quality could be detected in crystals grown with TDM, UDM, DM or NM detergents. (ii) The 3.5 Å pMMO model was obtained from a crystal with unit cell dimensions of \( a = b = 270.2 \, \text{Å}, \, c = 150.7 \, \text{Å} \). In contrast the 2.8 Å resolution 1YEW structure was obtained from a crystal with slightly smaller unit cell dimensions of \( a = b = 264.1 \, \text{Å}, \, c = 149.9 \, \text{Å} \). It was therefore attempted to dehydrate these crystals, as this has been shown on occasion to decrease the unit cell and increase the order in the crystals. In practice, however, dehydrating the crystals caused them to diffract poorly or not at all. (iii) Since sodium cacodylate buffer of pH 6.0 seemed to produce better diffracting crystals compared to crystals grown in sodium cacodylate buffer of pH 6.5, crystals from G-type conditions were soaked in mother liquor containing sodium cacodylate buffer with a lower pH of 4.5-5.5. Crystal soakings like these produced no visible change in the diffraction pattern quality. (iv) It was found that G-type conditions (0.1-0.2 M sodium cacodylate pH 6.0-6.5, 6-8% PEG 8000, 6-10% ethylene glycol, 0.05 M zinc acetate, 3% dioxane, 0.01-0.02 M spermidine) were cryocompatible. By eliminating the cryo-protection step and cryo-cooling crystals directly from the drop the quality and resolution of the diffraction patterns was visibly improved. It did not, however, extend the resolution limit beyond 3.5 Å for the crystals screened.

Unfortunately, many results from these optimisation processes are uncertain because the variability in the observed diffraction resolution limits seemed to stem not from the various crystallisation processes tried, but from different protein preparations. This “prep-to-prep” variability was the main problem in optimizing pMMO crystals to obtain a higher resolution structure. Some pMMO preps gave rise to large well-diffracting crystals with good reproducibility, while other preps produce needles in the same condition. For every sample it was almost always necessary to re-screen around the condition identified for the previous prep, and there was no guarantee that the resulting crystals, if any, would be of the same quality as those from the previous prep. There are no clearly visible differences between these preps looking at a gel or their absorbance scan, so an explanation of their different behaviour is not readily available. It is common for different protein preparations to vary and it could be anything in the process of expression, protein purification or crystallisation plate setup that is the cause.

2.4.1.3 Screening for a pMMO-MDH crystallisation condition
A crystal structure of the pMMO-MDH complex would represent a more active form of pMMO and could thus reveal more functional aspects of the enzyme. A major obstacle to
overcome in the crystallisation of the pMMO-MDH complex is its tendency to fall apart in the crystallisation drop, as evidenced by the formation of the pMMO crystals from protein solution containing the full pMMO-MDH complex. The interaction between MDH and the soluble region of pmoB, as seen in the EM structure of pMMO-MDH (Figure 2.4) is probably ionic in nature and the high salt and buffer concentrations typically present in a commercial crystallisation screen (0.1-0.2 M salt and 0.1 M buffer) could be causing the complex to dissociate. In order to prevent this dissociation our collaborator Natalia Myronova provided pMMO-MDH samples where pMMO and MDH were covalently crosslinked with DTSSP. No crystallisation condition of crosslinked pMMO-MDH could, however, be identified in commercially available crystallisation screens. It is likely that the crosslinking introduced heterogeneity into the system, which is not beneficial for crystallisation to occur. Since the high ionic content of commercial crystallisation screens could be causing the dissociation of pMMO-MDH, low salt PEG vs pH screens were designed. Crystals grew in several conditions in these screens. One of these conditions produced crystals with weak diffraction visible to ~8 Å. It has not yet been confirmed that these crystals contain the full pMMO-MDH complex. Even if these crystals are of pMMO only, they are still of interest since the conditions do not contain zinc. Hence, their study could be valuable in identifying physiological metal that binds to the third metal site in pMMO (see section 4.2.2).

2.4.2 Comparison of the 3.5 Å pMMO model with the 1YEW structure

2.4.2.1 Model bias

The 3.5 Å pMMO model presented here is heavily model biased towards the 1YEW structure, which was used as the search model for the molecular replacement solution. In addition the 3.5 Å pMMO model and the 1YEW structure represent the same protein (pMMO) from the same organism (M. capsulatus (Bath)), crystallised in the same space group (P4_2_2_1) with comparable unit cell dimensions and one αβδγ trimer in the asymmetric unit. This is the reason of the relatively low residuals (R_w = 30.7 % and R_p = 32.0), in spite of the limited resolution. This is also the reason of the low rmsd (0.43 Å) between the two models and the similarities in problematic regions in the models.

2.4.2.2 Metal site identities

The data collected near the absorption peaks of zinc, copper and iron, together with the available biochemical data, allow the positive identification of the metal site identities in pMMO. However the resolution of the data at 3.9 Å does not allow for the identification of the number of atoms bound at each site, nor is sufficient to provide detailed information on
metal coordination. The monocopper site in the 1YEW structure of *M. capsulatus* (Bath) pMMO was not conserved in the *M. trichosporium* OB3b pMMO structure (Hakemian et al, 2008). The presence of the mononuclear copper site in *M. capsulatus* (Bath) pMMO is confirmed in our structure. We also confirm the dinuclear copper site and the third metal site, which similarly to the 1YEW structure, contained zinc (Figure 2.8). No iron, no tricopper cluster nor any other detectable metal site was present in the structure. However, the electron density maps and anomalous difference Fourier maps obtained here do raise some question marks regarding the metal sites in *M. capsulatus* (Bath) pMMO. For instance, the mononuclear copper site seems to additionally bind a zinc ion (Figure 2.8A1-3). In addition, the largest negative difference density peak in the map of 6.8σ, is centered on the loop containing the chainF-pmoA-Glu195 residue coordinating the Zn atom in the 1YEW structure. Negative difference density peaks can also be found around the chain B-pmoA-Glu195 (5.2σ) and chainJ-pmoA-Glu195 residues (4.7σ) in the trimer (Figure 2.8C1-3). It is difficult to evaluate these differences since the map resolution is too limited for meaningful analysis. These results highlight the need for an independent higher resolution structure of pMMO, preferably obtained in the absence of zinc, in order to further analyse and verify the metal site identities of pMMO.

### 2.4.2.3 The importance of zinc

The pMMO crystallisation condition obtained here (Condition G) and the 1YEW crystallisation condition share certain similarities, most notably the presence of zinc acetate (Table 2.5). Similar to the situation of the 1YEW structure, the zinc in the third site is likely to be an artefact from the crystallisation condition. Zinc is also found mediating crystal contacts at six other sites in the trimer (Figure 2.9), which explains why zinc acetate was essential in obtaining pMMO crystals. Soaking pMMO crystals in different iron salts abolished any diffraction, which could be due to a disruption of the zinc mediated crystal contacts. The crystal contacts are formed primarily between the pmoB soluble domains of the trimer (Figure 2.9), which effectively excludes any binding of MDH to the pMMO trimer and thus explains the preferential crystallisation of pMMO over pMMO-MDH.

### 2.4.3 Identifying the pMMO active site

In order to identify the pMMO active site positively, pMMO samples were incubated with acetylene (a suicide substrate) and propargylamine (a potential suicide substrate) prior to crystallisation. A dataset of pMMO incubated with propargylamine was obtained to 4.0 Å.
Table 2.5. Comparison of crystallisation conditions

<table>
<thead>
<tr>
<th></th>
<th>3.5 Å pMMO model</th>
<th>1YEW</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protein buffer</strong></td>
<td>25 mM PIPES pH 7.2</td>
<td>50 mM Hepes pH 7.5</td>
</tr>
<tr>
<td></td>
<td>1 mM benzamidine</td>
<td>0.12% Cymal-5</td>
</tr>
<tr>
<td></td>
<td>0.03% DDM</td>
<td></td>
</tr>
<tr>
<td><strong>Detergent structure</strong></td>
<td><img src="image" alt="Detergent structure" /></td>
<td><img src="image" alt="Detergent structure" /></td>
</tr>
<tr>
<td><strong>Crystallisation condition</strong></td>
<td>0.2 M sodium cacodylate pH 6.0</td>
<td>200 mM zinc acetate</td>
</tr>
<tr>
<td></td>
<td>6% PEG 8000</td>
<td>8-12% PEG 8000</td>
</tr>
<tr>
<td></td>
<td>6% ethylene glycol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05 M zinc acetate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3% dioxane</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01 M spermidine</td>
<td></td>
</tr>
</tbody>
</table>

resolution (PA dataset, Table 2.4). Efforts were also made to derivitise pMMO crystals with xenon or krypton gas in order to obtain information regarding the pMMO active site. Xenon or krypton atoms bound to pMMO can potentially mimic where the gaseous substrate methane would bind to the protein, for instance in the active site and/or in a channel leading to the active site (Prange et al, 1998). This approach has previously been successful in identifying binding sites in sMMO (Whittington et al, 2001). The pressurization of pMMO crystals with xenon and krypton gas did not noticeably affect the diffraction quality of the crystals compared to non-pressurized crystals. A 4.7 Å resolution dataset from a crystal pressurized with krypton gas was obtained (Kr dataset, Table 2.4). No significant different density peaks in the map calculated from the PA dataset, or any anomalous signal from krypton in anomalous difference Fourier maps, could be found in the vicinity of the three metal sites or at any other interior site of the protein. There was therefore no evidence of krypton or propargylamine binding and the active site of pMMO could not be identified from these datasets. It is unclear whether propargylamine has bound to the protein or not, since the resolution of the data is limited and propargylamine with a molecular weight of only 55 Da, would not be expected to produce a very strong difference density peak. Krypton binding should be obvious even at 4.7 Å resolution since krypton would be expected to give a very high peak in the maps. The results of these studies, aimed at identifying the pMMO active site, are therefore inconclusive.
2.4.4 Summary

A novel crystallisation condition was obtained for pMMO, which yielded crystals from which a 3.5 Å resolution dataset could be collected. Data were collected at the absorption peaks of zinc, copper and iron and anomalous difference Fourier maps calculated in order to verify the metal site identities in pMMO. The results obtained here, with crystals resulting from a different purification protocol and crystallisation conditions compared to the 1YEW structure, confirms the three metal sites in *M. capsulatus* (Bath) as a mononuclear copper site, a dinuclear copper site and a zinc site, with the latter likely to be a crystallisation artefact. No new metal binding sites, relevant for the enzymatic function were found in the structure. A potential crystallisation condition for the pMMO-MDH complex has been obtained in the absence of salt. This condition will be pursued with the aim of solving the pMMO-MDH structure and to identify the physiological metal bound in the third site of *M. capsulatus* (Bath) pMMO.
THIOREDOXIN A

FROM *BACILLUS SUBTILIS*
3.1 INTRODUCTION

3.1.1 Thioredoxin

Thioredoxin (Trx) is a ubiquitous protein present in most organisms from archaea to humans. It is a ~12 kDa soluble protein active in the cytosol to maintain cysteine-containing proteins in the reduced state by converting intramolecular disulphide bridges to dithiols via a redox reaction (Holmgren, 1984). Thioredoxin is thus important in maintaining the reducing environment in the cytoplasm (Stewart et al., 1998). The important thiol-disulfide oxidoreductase function of thioredoxin has been studied extensively since the enzyme was first described in 1964 (Laurent et al., 1964) By maintaining substrate proteins in the reduced state thioredoxin effectively protects cytoplasmic proteins from inactivation or aggregation via the formation of non-specific intra or intermolecular disulfide bonds. Hence thioredoxin is a key enzyme in thiol redox control, which is the mechanism by which protein function can be regulated via the redox state of structural or catalytic thiol groups (Holmgren, 1985). For instance, thioredoxin activates ribonucleotide reductase, which synthesises deoxyribonucleotides, by reducing structural disulfides and donating hydrogens to the enzyme’s reaction cycle (Holmgren, 1989; Thelander, 1974). The oxidoreductase activity of thioredoxin also provides the electrons necessary for the catalytic cycles of methionine sulfoxide reductases (Weissbach et al, 2005) and peroxiredoxins (Chae et al, 1999) as well as having several other additional roles in the cell (for an overview, see Table 1 in Arner and Holmgren, 2000). In general, the specific physiological roles of thioredoxins in different organisms are directly related to the specific functions of their substrates in these organisms.

3.1.2 Thioredoxin active site and reaction mechanism

A defining characteristic of thioredoxin proteins is the presence of two cysteine residues within the active site motif Trp-Cys-Gly-Pro-Cys (Holmgren, 1968). The thioredoxin-catalysed reaction is believed to follow the reaction scheme shown in Figure 3.1 (Jeng et al, 1994; Kallis & Holmgren, 1980). An oxidized disulfide-containing substrate protein binds to a conserved hydrophobic surface of thioredoxin. The lower pKa of the Trx N-terminal cysteine promotes its existence as a thiolate, the reactive deprotonated form of thiol, and this thiolate can act as a nucleophile to combine with a protein substrate forming a covalently linked mixed disulfide intermediate (thioredoxin-S-S-substrate). The hydrophobic environment provided by the active site and the close proximity (i.e. high local concentration)
Figure 3.1. Thioredoxin reaction. Thioredoxin (Trx) is represented as an orange circle, the substrate protein as a green rectangle, thioredoxin reductase (TR) as a yellow oval. The electron flow during the reaction is represented by red curly arrows.

of the thiol group of the second, C-terminal, cysteine, often referred to as the resolving cysteine, facilitate the next step of the reaction where a deprotonated resolving cysteine attacks the intermolecular disulfide bond. At the end of the reaction fully reduced protein and oxidised disulfide containing thioredoxin (Trx-S₂) are generated (Figure 3.1). The net result is the interconversion of a disulfide and a dithiol through the transfer of two hydrogens in the form of two protons and two electrons from thioredoxin to its substrate. This reaction is reversible depending on the redox potential of the substrate protein. However, since the redox potential of thioredoxin is ~270 mV at pH 7.0 (Krause et al, 1991) thioredoxin usually acts as the electron donor, reducing the substrate protein. Oxidised thioredoxin (Trx-S₂), containing an intramolecular disulfide bridge between its two vicinal active site cysteines, is reduced back to its active form (Trx-(SH)₂) by thioredoxin reductase (TR) using reducing equivalents from NADPH (Moore et al, 1964) (Figure 3.1, 3.2)

3.1.3 Thioredoxin structures

The first thioredoxin structure, Trx-S₂, from *Escherichia coli* in the oxidised state, was obtained to 2.8 Å resolution (Holmgren et al, 1975). *E. coli* Trx-S₂ was characterized in detail in further studies (Jeng et al, 1994; Katti et al, 1990) in addition to the reduced thioredoxin,
Figure 3.2. Thioredoxin system. The cofactor FAD and redox-active disulfide of thioredoxin reductase (TR), together with reducing equivalents from NADPH, is used in reducing oxidized thioredoxin (Trx-S₂) back to its active reduced state. In turn, reduced thioredoxin (Trx-(SH)₂) reduces a disulfide in the substrate protein.

Trx-(SH)₂ (Dyson et al, 1990; Jeng et al, 1994). Human Trx1 has also been structurally characterized in both reduced and oxidized states by X-ray crystallography and NMR (Qin et al, 1994; Weichsel et al, 1996). More recently, while the work described in this thesis was in progress, the Bacillus subtilis thioredoxin A (BsTrxA) structure was obtained in both the oxidized and reduced states by NMR (Li et al, 2007). Several thioredoxin structures from other organisms are available in the protein data bank (summarized in (Maeda et al, 2006).

In addition, three-dimensional structures of thioredoxin variants complexed with substrate protein or peptides have also been determined. At the time this work was initiated, only two thioredoxin-substrate peptide complexes had been characterised: the NMR structures of human Trx1 with bound peptides from substrate proteins NFκB and Ref-1 (Qin et al, 1995; Qin et al, 1996). While the work described in this thesis was in progress, the structures of two thioredoxin-substrate protein complexes were obtained. The first was a crystal structure for the plant thioredoxin-substrate complex between barley thioredoxin and alpha-amylase/subtilisin inhibitor (BASI) (Maeda et al, 2006). The second was an NMR structure of the B. subtilis thioredoxin A (BsTrxA) mutant C32S as a heterodimer complex with the substrate protein arsenate reductase (ArsC) (Li et al, 2007).

Other protein interactions with thioredoxin have been characterized. For instance the interaction between the E. coli thioredoxin and thioredoxin reductase (TR) (Lennon et al, 2000) and the interactions between ferrodoxin-thioredoxin reductase (FTR) and spinach chloroplast thioredoxins (Trx-f and Trx-m) (Dai et al, 2007). The complex of thioredoxin with DNA polymerase from bacteriophage T7 (Doublie et al, 1998) is involved in the processivity of DNA replication. This complex is different from other thioredoxin complexes in that thioredoxin acts as a subunit of DNA polymerase and its oxidoreductase activity is not
involved. The available structures of thioredoxin complexes open up the possibility of comparisons in this work, from which more general information can be derived.

### 3.1.4 Thioredoxin A from *Bacillus subtilis*

Most studies on thioredoxins have been carried out on the human thioredoxin-1 (Trx1), thioredoxin A (TrxA) from *E. coli* and plant chloroplast thioredoxins. In contrast, relatively little is known about thioredoxin A from *B. subtilis* (BsTrxA). The *trxA* gene is essential for growth and viability of *B. subtilis* and the expression of BsTrxA is elevated by multiple kinds of stress conditions (Kobayashi et al., 2003; Scharf et al., 1998). Depletion of BsTrxA levels represents a form of stress that impacts on a wide range of physiologically and developmentally important processes in *B. subtilis* (Smits et al., 2005). In contrast to the *trxA* gene in *B. subtilis*, the two genes coding for thioredoxins in *E. coli* (*trxA* and *trxC*) are not essential (Holmgren et al., 1978). In addition to the thioredoxin system, *E. coli* has the complementary glutathione system (Holmgren, 1976; Prinz et al., 1997), which is absent in *B. subtilis* (Newton et al., 1996). This is likely why thioredoxin is essential in *B. subtilis* but not in *E. coli*. Hence some of the substrates of BsTrxA in *B. subtilis* will be essential for growth and viability of the organism.

In order to identify these essential substrates of BsTrxA in *B. subtilis* our collaborator Prof Jan Maarten van Dijl and his group from Groningen University utilized a technique named mixed disulfide fishing (Kouwen et al., 2008). This technique has been employed previously to identify thioredoxin substrates in various biological systems (Hisabori et al., 2005; Verdoucq et al., 1999). In order to carry out mixed disulfide fishing the second C-terminal cysteine in the thioredoxin active site is replaced with serine. The resulting mutant thioredoxin is still able to bind its substrates via the N-terminal (attacking) cysteine but the subsequent release of a reduced substrate is precluded by the absence of the C-terminal (resolving) cysteine. Consequently, normally very short-lived covalent enzyme-substrate reaction intermediates accumulate, which can then be purified and analysed to identify the bound substrates.

### 3.1.5 Aim of this study

During the mixed disulfide experiments to identify essential BsTrxA substrates, homodimers of single-cysteine mutant BsTrxA were identified that displayed a so far uncharacterized auto-oxidation. It was believed that homodimer formation occurred via a disulfide bond between the active site cysteines, thus mimicking the mixed-disulfide intermediate state in the
thioredoxin reaction. The aim of this study was to verify this suggestion by determining the structures of the single-cysteine mutant BsTrxA homodimers by X-ray crystallography. At the time this study was initiated no structures of thioredoxin-substrate protein complexes were available. Hence an additional motivation to study these homodimers was that they provided an opportunity to characterise a thioredoxin-protein complex that potentially could mimic a mixed-disulfide intermediate of the thioredoxin reaction.

3.2 METHODS AND MATERIALS

Expression and purification of BsTrxA mutants, mixed disulfide fishing experiments and the determination of redox states of BsTrxA monomers and dimers is work carried out by Thijs R.H.M Kouwen, a PhD student in Prof Jan Maarten van Dijl’s group at Groningen University, our collaborators on this project (Kouwen et al, 2008).

3.2.1. Crystallisation

Purified wild-type BsTrxA, C29S single mutant, C32S single mutant and C29S-C32S double mutant protein samples were provided by Thijs R.H.M Kouwen in a buffer containing 20 mM Tris-HCL pH 7.6 with 150 mM NaCl (Buffer A). Samples were concentrated to 10 mg/ml using an Amicon Ultra centrifugation device with a molecular weight cutoff of 10 kDa. Some concentrated protein samples were buffer-exchanged into a solution containing 25 mM sodium phosphate pH 8.0 and 400 mM NaCl (Buffer B) by diluting approximately 100 µl of concentrated thioredoxin sample with 15 ml of Buffer B followed by concentration of the sample to 10 mg/ml using an Amicon ultracentrifugation device. This procedure was repeated three times to ensure maximum replacement of Buffer A into Buffer B. Crystallisation conditions were screened at room temperature, according to the sparse matrix method (Jancarick, 1991), using commercially available screens (Crystal Screen 1 and 2 from Hampton Research, Wizard I and II from Emerald BioSystems, MemStart/MemSys from Molecular Dimensions) and the sitting drop vapour diffusion technique. Screens were set up in 96-well Corning round bottom plates using a Cartesian Microsys robot. Drops consisted of purified wild-type BsTrxA/C29S single mutant/ C32S single mutant/C29S-C32S double mutant in Buffer A or Buffer B (0.2 µl) with an equal volume of reservoir solution. Crystallisation plates were incubated at 20°C and 4°C. Hanging drop vapour diffusion
crystallisation experiments were performed with 1 µl each of protein in Buffer B and reservoir solution (0.1 M Tris pH 8.5, 0.2 M magnesium chloride, 4% acetonitrile and 30% PEG 4000) in 24-well Linbro plates that were incubated at 4°C. The reservoir solution, the protein buffer and the temperature choice were based on the published condition used to crystallise a TrxA mutant from E. coli (Collet et al, 2005). Crystals of the C32S single mutant grew in 0.1 M Tris pH 7.8, 0.1 M magnesium chloride, 4% acetonitrile and 35% PEG 4000. The crystals appeared in 1-2 weeks with dimensions of approximately 100x50x20 µm. No crystals were obtained for wild type BsTrxA, the C29S single mutant or the C29S-C32S double mutant in any of the crystallisation trials performed.

3.2.2 Data collection, structure determination and model building

C32S BsTrxA crystals were cryo-cooled directly from the drops without any added cryo-protectant. 300 images corresponding to 300° of data were collected at the ESRF at the BM14 beamline. Data collection statistics are summarized in Table 3.1. The data were scaled and processed using the program HKL2000 (Otwinowski, 1997). Molecular replacement was performed using the program PHASER (Read, 2001) with a polyserine model of thioredoxin A (TrxA) from E. coli as the initial model (PDB entry 2TRX, (Katti et al, 1990)). The sequence identity between B. subtilis BsTrxA and E. coli TrxA is 47%. After rigid body and restrained refinement using REFMAC-5 (Vagin et al, 2004), the model-building program ARP-WARP was used to build a starting model from the BsTrxA C32S sequence (Perrakis et al, 1999). The model was improved by several rounds of manual model building and refinement, using the programs COOT (Emsley & Cowtan, 2004) and REFMAC-5. The coordinates and structure factors were deposited with the Protein Data Bank as PDB entry 2VOC.

3.2.3 Structure analysis

Model geometry was analyzed using COOT and the program MOLPROBITY (Davis et al, 2007). Figures were generated using the program Pymol (DeLano, 2002). Rmsd values between structures were obtained using Dalilite (Holm & Park, 2000). Buried surface area in interfaces were calculated by the protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html), authored by E. Krissinel and K. Henrick (Krissinel, 2005).
3.3 RESULTS

3.3.1. BsTrxA homodimers from mixed disulfide fishing experiments

The studies on the biochemistry of BsTrxA homodimers are the work of Thijs R.H.M Kouwen, a PhD student in Prof Jan Maarten van Dijl’s group at Groningen University, our collaborators on this project (Kouwen et al, 2008).

3.3.1.1 Mixed disulfide fishing

His-tagged wild-type BsTrxA, a mutant of the resolving cysteine (C32S BsTrxA) a mutant of the attacking cysteine (C29S BsTrxA) and a double mutant (C29S/C32S BsTrxA) were purified (Figure 3.3A). The ability of the BsTrxA variants to bind cytoplasmic proteins from B. subtilis WB800 ItrxA strain, depleted for BsTrxA, was tested by mixing them with B. subtilis cytoplasmic proteins. The mixture was incubated and separated by SDS-PAGE under non-reducing conditions followed by Western blotting using antibodies against BsTrxA (Figure 3.3B). From this experiment it was found that the C32S BsTrxA, which contains the

![Figure 3.3. Mixed disulfide fishing results. A. Silver stained gel of BsTrxA wildtype (WT), C29S single mutant (C29S), C32S single mutant (C32S) and C29S/C23S double mutant (C29S/C32S). B. BsTrxA protein complexes. Background: Cytoplasmic extract treated with water instead of BsTrxA protein as a control. C. BsTrxA protein complexes extracted by mixed disulfide fishing. Figure from (Kouwen et al, 2008) used with permission from Thijs R.H.M. Kouwen.](image-url)
attacking Cys29 residue, was able to form stable interactions with potential substrate proteins, whereas the C29S/C32S double mutant did not form these interactions. This observation is consistent with the known reaction mechanism of thioredoxins and the anticipated accumulation of enzyme-substrate complexes if the resolving Cys32 is absent. An unexpected observation was that the C29S BsTrxA variant (lacking the attacking Cys29 residue) was also able to form reaction intermediates that could be separated by SDS-PAGE, albeit with a slightly lower efficiency than the C32S BsTrxA (Figure 3.3B). This suggested that Cys32 can also act as a nucleophile in BsTrxA-substrate interactions. The stability of these BsTrxA-complexes was tested in a mixed disulfide fishing experiment. The C-terminal His6-tag on the BsTrxA variants was used to immobilize the BsTrxA-substrate complexes on magnetic beads precharged with nickel. The beads were extracted after an incubation time of 10 minutes using a magnet. The beads were washed nine times and the BsTrxA proteins were eluted from the beads with buffer containing imidazole. The eluted proteins were separated by non-reducing SDS-PAGE and the BsTrxA-substrate bound complexes were visualized by Western blotting with antibodies against BsTrxA (Figure 3.3C). Figure 3.3C shows that only the mixed disulfides formed by the C32S BsTrxA mutant protein were sufficiently stable for fishing. In contrast, all complexes with the C29S mutant or the wild-type BsTrxA protein were lost during the washing of the beads, which implies that the respective interactions were significantly less stable than those observed for the C32S mutant and its potential substrates.

The BsTrxA-substrate interaction analyses indicated that all purified BsTrxA proteins (C29S, C32S, C29S/C32S and wild-type) were capable of dimer formation when analysed under non-reducing conditions, as evidenced by the presence of bands with twice the mass of monomeric BsTrxA (~25 kDa) (Figure 3.3, B and C). Such potential dimers were present when purified BsTrxA proteins were mixed with potential substrates and they remained stable upon mixed disulfide fishing with magnetic beads. Further investigation with purified protein under non-reducing conditions, in the absence of cytoplasmic proteins, showed dimers only for the two single mutants, in particular for the C32S mutant, not for the wild type or double mutants.

3.3.1.2 Redox state of BsTrxA

Several observations supported the idea that the observed ~25 kDa species consisted of two identical BsTrxA held together by a disulfide bond via one of its active site cysteine residues. (1) The ~25 kDa species were confirmed to consist of BsTrxA by Western blotting using antibodies against BsTrxA. (2) The ~25 kDa band was not observed in the presence of the reducing agent DTT (1,4-Dithio-DL-threitol), i.e. the two subunits are held together by a bond that can be broken with DTT (Figure 3.4). (3) Cross-linking experiments with the thiol
Figure 3.4. Redox states of BsTrxA monomers and dimers. A. Purified His6-tagged BsTrxA proteins were separated by capillary electrophoresis using a 2100 Bioanalyzer (Agilent). To monitor the presence of free thiols in the purified BsTrxA proteins, samples were incubated in the presence or absence of 0.3 mM AMS (lanes marked +AMS or -AMS). To test whether the C29S and C32S dimers are formed by disulfide bonding, these proteins were incubated with 10 mM DTT (lanes marked +DTT). DTT was absent from all other samples. The image of the Bioanalyzer chromatograms was generated using the 2100 Expert Software package (Agilent). B. ~2.5 µg C32S BsTrxA protein was reduced with increasing concentrations of DTT (shown on top of the panel) and separated by capillary electrophoresis as described for panel A. The dimer to monomer ratios (Dimer [%]) are shown at the bottom of the panel. Figure and legend from (Kouwen et al, 2008), used with permission from Thijs R.H.M. Kouwen.

Specific reagent AMS (4-acetamido-4’-maleimidyl-stilbene-2,2’-disulfonate) decreased the electrophoretic mobility of the wildtype BsTrxA by ~1.0 kDa and BsTrxA C29S and C32S mutants by ~0.5 kDa respectively. This indicated that the wild type had two thiol groups present in the reduced state and that the C29S and C32S BsTrxA variants had one thiol group each present in a reduced state (Figure 3.4A). By peak integration of the band for wild type BsTrxA bound to AMS (upper band, WT +AMS, Figure 3.4A) it was estimated that only 33%
of the wild type BsTrxA was in a reduced state. (4) Cross-linking experiments with AMS showed that the ~25 kDa species had no free thiol groups to bind to AMS with (Figure 3.4A). Since Cys29 and Cys32 are the only two cysteine residues in BsTrxA, these observations imply that the dimerisation of the single-cysteine mutant proteins was due to disulfide bond formation between the active site cysteine residues of two monomers, thus forming a homodimer.

3.3.2. Quality of the BsTrxA homodimer model

In order to verify the suggestion that the C32S BsTrxA protein forms a homodimer via a disulfide bond between the attacking cysteines, we crystallised this protein and subsequently determined its three-dimensional structure to 1.5 Å using molecular replacement (PDB entry 2VOC, (Kouwen et al, 2008)). The crystals had a Matthews coefficient \(V_m\) of 2.1, corresponding to a solvent content of 42.4%, and were space group \(P1\), with unit cell dimensions \(a = 36.8\,\text{Å}, \, b = 38.4\,\text{Å}, \, c = 41.9\,\text{Å}, \, \alpha = 83.3^\circ, \, \beta = 66.6^\circ, \, \gamma = 78.1^\circ\) (Table 3.1).

There was one copy of the dimer in the asymmetric unit consisting of chains A and B, related to each other by a two-fold non-crystallographic symmetry axis. Chain A consists of residues 2-111, where residues 105-111 belong to the His6-tag used in the purification of the protein. Chain B consists of residues 2-105, where residue 105 belongs to the His6-tag. The first methionine residue is not present in the electron density map for neither chain A nor chain B. The two Pro31 side chains in chain A and B respectively, are only 2.6 Å apart in the model, which is the best fit to the electron density. This clash would not normally occur in the wild-type enzyme. It is likely that the two proline residues adopt alternative non-clashing positions in solution, but these are not apparent in the averaged electron density map and so they are not included in the model. Refinement of the model converged with residuals \(R_{\text{work}} = 18.5\%, \, R_{\text{free}} = 23.7\%\) and an average temperature factor of \(B = 18.2\,\text{Å}^2\). The model shows good geometry with 98.1% of the residues in the most favoured region of the Ramachandran plot as assessed by MOLPROBITY (Davis et al, 2007). The two chains A and B superpose with an rmsd value of 0.4 Å for 104 common \(\text{C}_\alpha\) atoms. Altogether the asymmetric unit contains 214 protein residues, 254 water molecules and one PEG molecule from the precipitant solution.

3.3.3 BsTrxA homodimer structure

Each C32S BsTrxA monomer, similar to other thioredoxin structures, has a central hydrophobic core with a five-stranded \(\beta\)-sheet (residues 4-5, 19-25, 51-56, 74-79, 82-88) surrounded by four \(\alpha\)-helices on the external surface (residues 11-15, 30-46, 63-66, 93-101)
Table 3.1. C32S BsTrxA data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data set</th>
<th>C32S BsTrxA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection and processing</strong></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
</tr>
<tr>
<td>Unit cell: a, b, c (Å)</td>
<td>36.8, 38.4, 41.9</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>83.3, 66.6, 78.1</td>
</tr>
<tr>
<td>ESRF beamline</td>
<td>BM-14</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.98</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.0-1.5</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>86633</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>31147</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>94.5 (90.1)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.8</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>15.6 (1.9)</td>
</tr>
<tr>
<td>Rsym (%)</td>
<td>7.0 (33.9)</td>
</tr>
</tbody>
</table>

| **Model refinement** | |
| Resolution (Å) | 23.0-1.5 |
| Rwork (%) | 18.5 |
| Rfree (%) | 23.7 |
| Average B factor (Å²) | 21.2 |

| **Stereocchemistry** | |
| Root mean square deviation for bond length (Å) | 0.021 |
| Root mean square deviation for bond angles (°) | 2.027 |

| **Residues in the Ramachandran plot** | |
| Favored regions (%) | 98.1% |
| Allowed regions (%) | 100.0% |

---

Statistics for the highest resolution shell are given in parentheses.

\[ R_{\text{sym}} = \frac{\sum |I_h| - \langle |I_h| \rangle}{\sum |I_h|} \]

\[ R_{\text{work}} = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|} \text{ for all reflections.} \]

\[ R_{\text{free}} = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|} \text{ calculated using randomly selected reflections (5%).} \]

Ramachandran plot values as assessed by MOLPROBITY.

The secondary structure elements are arranged in a \( \beta_1\alpha_1\beta_2\alpha_2\beta_3\alpha_3\beta_4\alpha_4 \) topology, which includes the highly conserved thioredoxin fold (\( \beta_2\alpha_2\beta_3\alpha_3\beta_4\alpha_4 \)). Each monomer in the C32S homodimer closely resembles the NMR structure of C32S BsTrxA (Li et al, 2007) with an rmsd of 1.3 Å for 103 common Cα atoms. It is also similar to known human and *E. coli* structures (PDB entries 2TRX, 1XOA, 1XOB, 1ERU, 1ERT, 1CQG, 1CGH, 1MDJ) with rmsd values of 1.2-1.5 Å for global Cα superpositions. The C32S BsTrxA homodimer packs closely in space group P1, as expected from the 42.4% solvent content estimated from the Matthews coefficient (Figure 3.5C). The His6-tag on chain A
Figure 3.5. C32S BsTrxA homodimer and crystal packing. Overall fold of the C32S BsTrxA homodimer. Chain A is shown in orange and chain B in green. The C29-C29 disulfide bond between the chains is shown in yellow. A. View shown down the two-fold non-crystallographic symmetry axis. B. 90° rotation around the x-axis from orientation in (A). The β-sheet strands are numbered from 1-5 in chain A. All strands are parallel except strand β₄, which runs antiparallel. The α-helices are numbered from 1-4 in chain B. C. The His6-tag on chain A is important in forming crystal contacts between homodimers.
(residues 105-111) is well defined in the electron density map as it is involved in the crystal packing via the formation of eight hydrogen bonds to a neighbouring molecule in the crystal. In contrast, the His6-tag on chain B is not observed in the electron density map, presumably due to an inherent flexibility of the tag as it is not participating in crystal contact interactions. The active site in each chain consists of residues Trp28-Cys29-Gly30-Pro31-Ser32, where the resolving cysteine (Cys32) has been mutated into a serine. The numbering of the *B. subtilis* BsTrxA residues differs by -3 from the commonly used *E. coli* numbering, in which Cys32 is the N-terminal attacking cysteine and Cys35 is the C-terminal resolving cysteine. The active site in each chain forms a protruding loop linking the end of the β2-strand with the beginning of the α3-helix. The dimer is, as expected, held together by a disulfide bond between the attacking Cys29 residues from the active site of each chain in the homodimer. Dimerisation is further facilitated by intermolecular hydrogen bonds and hydrophobic interactions contributing to tight packing between the monomers (Figure 3.6A). Residues from the active site loop binds into a shallow hydrophobic groove on the thioredoxin surface of the opposite chain. The groove is formed by side chains of residues Ala26, Trp28, Val57, Ala64, Val69 and Ile72 from the β3-β4 region of thioredoxin, which is comprised of the β3-α3 loop, helix α3 and the α3-β4 loop. The large hydrophobic residue Trp28, which precedes the catalytic Cys29, interacts with this shallow hydrophobic groove in addition to the backbone of residues Met70 and Ser71 that pack against the Trp28 ring. The cis conformation of the conserved Pro73 residue causes a change in chain direction. This exposes the carbonyl and amide groups of Ile72, making them available to form interactions with the amide of Gly30 and the carbonyl of Trp28 on either side of the disulfide bond (Figure 3.6B). This series of backbone-backbone hydrogen bonds and hydrophobic interactions, buries 2x438.6 Å² of the surface area of chain A and chain B.

### 3.3.4. The C32S mutation of the resolving cysteine

In wild-type BsTrxA residue Cys29 would act as the attacking cysteine, i.e. attacking a substrate disulfide and form an intermolecular disulfide bond with the substrate. The Cys32 residue would act as the resolving cysteine in the reaction, i.e. attacking the intermolecular disulfide bond formed between Cys29 and the substrate protein, thus forming an intramolecular disulfide bond with Cys29 and releasing the substrate. In the C32S BsTrxA homodimer structure the Ser32 mutant is unable to resolve the intermolecular disulfide or form an intramolecular disulfide bond with Cys29. The C32S mutation therefore ensures that the reaction does go to its completion. The crystal structure shows that the sidechain of the...
Figure 3.6. C32S BsTrxA dimer interface. A. Hydrophobic interactions. B. Polar interactions. The chain B (green) Trp28 has two alternative conformations whereas the chain A (orange) Trp28 has only one orientation. Trp28 interacts with the same surface on the opposite chains. Chain A is shown as a partially transparent electrostatic surface. Oxygen atoms are shown in red, nitrogen atoms in blue and the disulfide bond atoms are shown in yellow. The $\beta_1$-$\beta_4$ region is shown in blue in panel A. Hydrogen bonds are shown as blue dotted lines in panel B.
Figure 3.7. Ser32 interactions in the C32S BsTrxA homodimer structure. A. The hydroxyl sidechain of Ser32 is forming hydrogen bonds to the amide and carbonyl of Ala26. The active site loop is further stabilized by a hydrogen bond between the amide of Ser32 and the carbonyl of Cys29. B. The environment surrounding residue Ser32 is shown. The view is rotated ~90° in comparison to panel A. Residues 26-32 and 72-73 in chain A are shown as sticks, other parts of the structure are shown as cartoons. Chain A is shown in orange and chain B in green. The disulfide bond is shown in yellow, oxygens are shown in red and nitrogens are shown in blue. Hydrogen bonds are shown as dotted blue lines and hydrogen bond distances are shown in blue text.

mutated residue Ser32 points away from the intermolecular disulfide Cys29-Cys29 bond, forming hydrogen bonds to the backbone amide and carbonyl of Ala26 (Figure 3.7A). The Ser32 residue is packed between the side chains of Trp25, Pro31, Ile72 and Pro73 so it is buried inside the protein (Figure 3.7B).

3.4 DISCUSSION

3.4.1 C29S and C32S BsTrxA homodimers are distinct from other thioredoxin homodimers

C32S BsTrxA crystallises as a dimer stabilised by the formation of a disulfide bond between the two active site attacking Cys29 residues, as expected from the biochemical results. C32S and C29S BsTrxA homodimers are formed by thiol oxidation between two identical BsTrxA molecules, most likely through auto-oxidation. However, the C29S and C32S BsTrxA
homodimers were detected in interaction studies performed in vitro under non-reducing conditions, not in vivo. The cross-linking studies with AMS to free thiols showed that approximately 2/3 of the purified wild type BsTrxA molecules were oxidized under the conditions used, which is consistent with the fact that reduced thioredoxin is oxidized by molecular oxygen in air at neutral pH (Laurent et al, 1964). Even though it is not known whether BsTrxA molecules interact with each other in vivo, it is a possibility since oxidized BsTrxA can be detected in growing cells (Hochgrafe et al, 2005). Other thioredoxin homodimers have been reported to form by (auto)oxidation, such as for an E. coli mutant TrxA lacking the resolving cysteine (Jeng et al, 1998) or for human Trx1 where a third cysteine residue outside the active site (Cys73) oxidises to form homodimers (Ren et al, 1993; Weichsel et al, 1996). This Cys73-based dimerisation is believed to function as a means to down-regulate thioredoxin activity, since such dimers were found to be inactive (Ren et al, 1993). Since BsTrxA only has the two active site cysteines it appears to lack this regulatory mechanism. Thiol-independent dimerisation was not observed for the purified BsTrxA, but has been observed under reducing acidic conditions via non-covalent binding for E. coli and human thioredoxins (Andersen et al, 1997; Dyson et al, 1994; Holmgren & Roberts, 1976; Holmgren et al, 1975). A thioredoxin homodimer formed between active site cysteine residues has been observed for mutants of E. coli TrxA with an altered active site (CACA instead of CGPC) (Collet et al, 2005). The change in sequence altered the conformation of the active site by partly unwinding a helix and exposing the C-terminal cysteine of the CACA motif. Both cysteine residues in the active site became disulfide bonded. This makes the E. coli CACA TrxA dimer distinct from the C32S BsTrxA dimer reported here. In summary, the dimerisation of C29S and C32S BsTrxA mutants by thiol oxidation differs substantially from previously reported thioredoxin homodimers.

3.4.2 The C32S and C29S mutations of the active site cysteines

The Ser32 sidechain in the C32S BsTrxA homodimer is oriented away from the intramolecular disulfide bond as it is forming a series of hydrogen bonds (Figure 3.7). Similar interactions are not made by the Cys32 residue in the wildtype BsTrxA-(SH)2 structure solved by NMR (PDB code 2GZY, (Li et al, 2007)). The Cys32 sidechain in reduced wildtype BsTrxA enzyme adopts an alternative rotamer compared to Ser32, which makes it ideally positioned for attacking an intermolecular disulfide bond between Cys29 and the substrate protein (Figure 3.8). The Ser32 mutation does not cause any structural rearrangement of the Cα-backbone in comparison with the wild type BsTrxA. In fact, the serine mutation stabilizes
Figure 3.8. Ser32 in C32S BsTrxA and Cys32 in wild type BsTrxA. Comparison of the Ser32 residue in the C32S BsTrxA homodimer crystal structure with the Cys32 residue in the reduced wildtype BsTrxA NMR structure (PDB entry 2GZY). The C32S BsTrxA chain A is shown in orange and chain B in green. The wild type BsTrxA residues 28-32 are shown in purple. Sulfur atoms are shown in yellow, oxygens are shown in red and nitrogens are shown in blue. Hydrogen bonds are shown as dotted blue lines.

The conformation of the active site loop, residues 26-30, by forming hydrogen bonds from its hydroxyl side chain to the backbone amide and carbonyl of Ala26 (Figure 3.8). The main difference between the C32S homodimer structure and the wild type BsTrx-(SH)$_2$ structure in this region is the different rotamer of Ser32, compared to Cys32, caused by the polar hydroxyl side chain forming hydrogen bonds. Both the wild-type Cys32 residue and the mutant Ser32 residue are buried inside the protein being packed between the side chains of Trp25, Pro31, Ile72 and Pro73. The Cys32 residue is thus not exposed at the protein surface to the same extent as Cys29.

The results from Jan Maarten van Dijl’s group show that both the C32S and the C29S BsTrxA mutant proteins are able to form mixed disulfide intermediates when provided with potential substrate proteins. The finding that BsTrxA mutant proteins lacking the attacking cysteine residue (Cys29 in BsTrxA) can form such intermediates has not been reported before. In previous interaction profiling studies, mixed disulfide fishing experiments were only carried out using mutant thioredoxin lacking the resolving C-terminal cysteine residue, not with thioredoxin lacking the attacking N-terminal cysteine (Balmer et al, 2003; Motohashi et al, 2001; Verdoucq et al, 1999). Both variants (C29S and C32S) were used in the mixed disulfide
fishing experiments carried out by Jan Maarten van Dijl’s group (Kouwen et al, 2008). The data obtained through these experiments in Jan Maarten van Dijl’s group suggests that the resolving cysteine (Cys32) in the C29S mutant protein is also able to engage in substrate binding. In *E. coli* TrxA the attacking cysteine has an unusually low pK\(_a\) of 6.7 (Kallis & Holmgren, 1980) and the resolving cysteine an unusually high pK\(_a\) of 9.5 (Jeng et al, 1995). The low pK\(_a\) value of the attacking cysteine is required for the thioredoxin catalysed thiol-disulfide exchange reaction to occur, since it promotes the existence of the attacking cysteine as a reactive thiolate ion at the neutral pH of the cell (Jeng et al, 1995). The unusual pK\(_a\) values of the thioredoxin active site cysteines have been suggested to be due to the two cysteines sharing the proton normally attached to the resolving cysteine (Dyson et al, 1990; Jeng et al, 1995). If the attacking cysteine is mutated, as in the C29S BsTrxA variant, it may lower the pK\(_a\) of the resolving cysteine, thus making it more reactive towards possible substrates than in the wild-type protein. In the wild type enzyme the Cys32 residue is buried within the protein, hence a change of conformation to make this residue more surface accessible may be necessary for Cys32 to act as the attacking cysteine in the C29S mutant. Possibly the C29S mutation could cause such a change in conformation since the Ser29 side chain could form extra interactions, which could cause a modification of the active site loop. A rearrangement of the active site loop would change any interaction with substrates, which could explain why the C29S BsTrxA mutant does not react as efficiently with substrate proteins relative to the C32S mutant protein. A high-resolution structure of the C29S BsTrxA mutant, which is not available at present, would be required to determine whether this is indeed the case.

### 3.4.3 Comparison with thioredoxin-substrate complexes representing mixed disulfide intermediates

C32S BsTrxA crystallises as a dimer held together by a disulfide bond between the two active site attacking Cys29 residues. This dimer mimics the mixed disulfide intermediate state in the thioredoxin reaction, where an intermolecular disulfide bond is formed between thioredoxin and a substrate protein (Figure 3.1). In the dimer one BsTrxA chain would represent the enzyme and the other one the substrate. Representations of mixed disulfide intermediates of thioredoxin have previously been characterized at the structural level by NMR of human thioredoxin with two substrate peptides derived from Ref-1 and NF-κB (PDB entries 1CQG, 1CQH and 1MDJ; (Qin et al, 1995; Qin et al, 1996)), by NMR of the BsTrxA-ArsC complex (Li et al, 2007) and by X-ray crystallography of barley thioredoxin in complex with the substrate protein barley alpha-amylase/subtilisin inhibitor (HvTrxh2-BASI) (Maeda et al,
2006). The resolving cysteine in these complex structures, similar to the C32S BsTrxA homodimer structure, was mutated to prevent the reaction proceeding, thus “trapping” the mixed disulfide intermediate. For comparison of the protein/peptide binding modes in thioredoxin complexes the structures of the thioredoxins were superimposed on chain A of the BsTrxA homodimer structure. Thioredoxin complex structures used in this comparison are listed in Table 3.2.

3.4.3.1 Comparison with human thioredoxin in complex with substrate peptide Ref-1

The first structures representing the mixed disulfide intermediate state were human Trx1 in complex with two substrate peptides derived from Ref-1 and NFκB (PDB codes 1CQG, 1CQH and 1MDJ). The direction of the polypeptide chain through the thioredoxin active site is different in the Ref-1 peptide complex and the NFκB peptide complex. The C32S BsTrxA homodimer structure has the same peptide chain orientation as the NFκB peptide. Nevertheless, the interactions between the Ref-1 peptide and Trx1 are similar to the interactions found in the C32S BsTrxA homodimer. In both complexes a tryptophan residue (Trp28-BsTrxA/Trp67-Ref1) binds the shallow hydrophobic groove on the thioroedxin surface formed by the β3-β4 region. From this region residue Thr74-Trx1, in an equivalent position and role to the Ile72-BsTrxA residue, forms a backbone-backbone hydrogen bond to Cys65-Ref1, the residue that is involved in the intermolecular disulfide bond to Cys32-Trx1. In contrast to the C32S BsTrxA homodimer, Ref-1 forms an interaction with the thioroedxin β3-α4 loop as the carbonyl of Lys63-Ref1 is within hydrogen-bonding distance to the backbone amide of Ala92-Trx1 and their aliphatic side chains pack against each other.

3.4.3.2 Comparison with human thioredoxin in complex with substrate peptide NFκB

Similarly to the C32S BsTrxA homodimer and the Trx1-Ref1 complex, the NFκB peptide has a large hydrophobic residue, Tyr60-NFκB, binding into the shallow hydrophobic groove on the Trx1 surface formed by the β3-β4 region. The intermolecular disulfide bond is formed between Cys32-Trx1 and Cys62-NFκB. The Thr74-Trx1 residue occupies the same position as Ile72-BsTrxA in the homodimer and forms a similar interaction in making a backbone-backbone hydrogen bond to the residue next to the intermolecular disulfide bond, here Glu63-NFκB. In addition to these interactions, the NFκB peptide forms five hydrogen bonds and five hydrophobic interactions to Trx1, involving residues from the Trx1 active site loop, the β3-β4 region and the β2-α4 loop. In contrast to the Trx1-Ref-1 complex and the C32S BsTrxA homodimer, almost every residue in the NFκB peptide is involved in interactions with Trx1. This could be a reflection of a higher conformational freedom of the NFκB peptide (Qin et al., 1995).
<table>
<thead>
<tr>
<th>PDB code</th>
<th>Thioredoxin (mutation)</th>
<th>Trx catalytic residue: substrate target residue</th>
<th>Other proteins in complex (mutation)</th>
<th>Method</th>
<th>Resolution (Å)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1CQG</td>
<td>Human Trx1 (C35A, C65A, C69A, C73A)</td>
<td>C32 : C65</td>
<td>Peptide consisting of residues 59-71 of the human substrate NFκB</td>
<td>NMR</td>
<td>N/A</td>
<td>(Qin et al, 1996)</td>
</tr>
<tr>
<td>1MDJ</td>
<td>Human Trx1 (C35A, C65A, C69A, C73A)</td>
<td>C32 : C62</td>
<td>Peptide consisting of residues 56-68 of the human substrate NFκB</td>
<td>NMR</td>
<td>N/A</td>
<td>(Qin et al, 1995)</td>
</tr>
<tr>
<td>2IPA</td>
<td>BsTrxA (C32S)</td>
<td>C29 : C89</td>
<td>Arsenate reductase (ArsC) (C10S, C15A, C82S)</td>
<td>NMR</td>
<td>N/A</td>
<td>(Li et al, 2007)</td>
</tr>
<tr>
<td>2IWT</td>
<td>Barley HvTrxh2 (C49S)</td>
<td>C46 : C148</td>
<td>α-Amylase serine proteinase inhibitor (BASI) (C144S)</td>
<td>X-ray</td>
<td>2.3</td>
<td>(Maeda et al, 2006)</td>
</tr>
<tr>
<td>2VOC</td>
<td>BsTrxA (C32S)</td>
<td>C29 : C29</td>
<td>BsTrxA homodimer (C32S)</td>
<td>X-ray</td>
<td>1.5</td>
<td>(Kouwen et al, 2008)</td>
</tr>
<tr>
<td>1F6M</td>
<td>E. coli TrxA (C35S)</td>
<td>C32 : C138</td>
<td>E. coli thioredoxin reductase (TR) (C135S)</td>
<td>X-ray</td>
<td>3.0</td>
<td>(Lennon et al, 2000)</td>
</tr>
<tr>
<td>2PU9</td>
<td>Spinach chloroplast (Trx-f) (C49S)</td>
<td>C46 : C57</td>
<td>Synechocystis ferredoxin-thioredoxin reductase (FTR)</td>
<td>X-ray</td>
<td>1.65</td>
<td>(Dai et al, 2007)</td>
</tr>
<tr>
<td>2PUK</td>
<td>Spinach chloroplast (Trx-m) (C40S)</td>
<td>C37 : C57</td>
<td>Synechocystis ferredoxin-thioredoxin reductase (FTR)</td>
<td>X-ray</td>
<td>3.0</td>
<td>(Dai et al, 2007)</td>
</tr>
</tbody>
</table>
3.4.3.3 Comparison with the BsTrxA-ArsC heterodimer

Comparison of the BsTrxA homodimer structure and the BsTrxA-ArsC heterodimer structure shows that the binding mode is similar to a certain extent but involves more extensive interactions in the BsTrxA-ArsC complex. In both the C32S homodimer and in the BsTrxA-ArsC complex the residues located on either side of the intermolecular disulfide bond (Cys29-Cys29 in the homodimer, Cys29-Cys89 in the BsTrxA-ArsC heterodimer) are important in complex formation. The Trp28 residue in the BsTrxA homodimer and the Met91 residue in ArsC are both binding to the shallow hydrophobic groove on thioredoxin formed by the β3-β4 region. Similarly, the Ile72 residue on the α3-β4 loop is forming backbone-backbone hydrogen bonds to the residue next to the disulfide cysteine, i.e. the Gly30 residue in the BsTrxA homodimer and the Lys88 residue in ArsC. The Met70-BsTrxA residue binds into a groove on the ArsC surface similar to the Met91-ArsC residue binding to the thioredoxin surface. In addition to these interactions in BsTrxA-ArsC, the β3-β4 region forms one hydrogen bond with the carbonyl of the Met91-ArsC residue through the sidechain amine group of Gln61-BsTrxA and a second hydrogen bond from Gly68-BsTrxA to the sidechain hydroxyl group of residue Ser69-ArsC. A third additional hydrogen bond is formed by Val88-BsTrxA carbonyl on the β3-α4 loop to the sidechain imine group of the Lys88-ArsC. In summary, ArsC form more extensive interactions than the BsTrxA homodimer that includes the β3-α4 loop, which is not involved in the binding in the homodimer structure. Hence the two complexes show similarities, but ultimately the more extensive interactions in the BsTrxA-ArsC heterocomplex make the binding modes different to each other.

3.4.3.4 Comparison with the HvTrxh2-BASI complex

The BASI substrate does not bind to the shallow hydrophobic groove formed by the β3-β4 region on the thioredoxin surface as seen in the other thioredoxin complexes. Instead the Trp147-BASI residue binds between the α3-β4 loop and the β5-α4 loop on the other side of the intermolecular disulfide bond by making hydrophobic interactions with residue Ala87-HvTrxh2 and residue Val104-HvTrxh2 respectively. The β3-α4 loop is further involved in binding through a hydrogen bond formed between the backbone amide of residue Ala106-HvTrxh2 to the carbonyl of Asp146-BASI. The Met88-HvTrxh2 residue, next to Ala87-HvTrxh2, is in an equivalent position to residue Ile72-BsTrxA on the α3-β4 loop, and it forms two intermolecular hydrogen bonds between its backbone carbonyl and amide to the amide and carbonyl of residue Cys148-BASI, respectively. The HvTrxh2-BASI complex thus has a different binding mode in comparison to the other thioredoxin complexes.
Table 3.3. Substrate binding regions in thioredoxin

<table>
<thead>
<tr>
<th>Substrate-binding region</th>
<th>Thioredoxin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BTrxA</td>
</tr>
<tr>
<td>1. Active site loop b2-a2</td>
<td>Ala26-Gly30</td>
</tr>
<tr>
<td>2. Helix-loop region b3-b4</td>
<td>Val57-Pro73</td>
</tr>
<tr>
<td>(loop b3-a3, helix a3, loop a3-b4)</td>
<td>Gly89-Pro92</td>
</tr>
<tr>
<td>3. Loop b5-a4</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.9. Thioredoxin substrate binding regions. Superimposed BTrxA (C32S) is shown in orange, human Trx1 is shown in yellow and barley HvTrxh2 is shown in green. The substrate binding region b2-a2, corresponding to the active site loop, is shown in magenta. The b3-b4 region, comprised of loop b3-a3, helix a3, and loop a3-b4 is shown in skyblue and the b5-a4 loop is shown in dark blue. A. Side view. B. View tilted ~60° around y axis compared to the view in panel A.

3.4.3.5 Thioredoxin substrate binding

The comparisons between these five different thioredoxin complexes show that interactions between thioredoxin and its substrates generally involve three substrate binding regions: 1) the active site loop, 2) the b3-b4 region forming the shallow hydrophobic groove and 3) the b5-a4 loop (Table 3.3, Figure 3.9). The b5-a4 loop does not interact with the other chain of the dimer in the BTrxA homodimer structure. The directions of the polypeptide chain through the thioredoxin active site are different in the thioredoxin substrate complexes. The C32S BTrxA homodimer structure has the same parallel peptide chain orientation as the NFκB peptide (Figure 3.10). However, in the BTrxA-ArsC and HvTrxh2-BASI complexes the chain direction is anti-parallel, similar to the Ref-1 peptide. In general the complexes have large hydrophobic residues on the substrate (either Trp, Tyr or Met) that interact with a shallow hydrophobic groove: Trp28 in the BTrxA homodimer, Met91 in the BTrxA-ArsC heterodimer, Trp67 in the Ref-1 peptide, Tyr60 in the NF-κB peptide and Trp147 in the
Figure 3.10. Comparison of thioredoxin protein/peptide binding modes. The thioredoxins from *B. subtilis* (BsTrxA) is shown in orange, human (Trx1) is shown in yellow and barley (HvTrxh2) is shown in light green. The electrostatic surface for chain A of C32S BsTrxA is shown. The substrate binding regions on the thioredoxins are highlighted in magenta for the active site, light blue for the β3-β4 region and dark blue for the β5-α4 loop. The intermolecular disulfide containing region from chain B from the C32S homodimer is shown in green, in red for the Ref-1 peptide, in cyan for the NFXB peptide, in black for arsenate reductase (ArsC) and in purple for the α-amylase serine proteinase inhibitor (BASI). The disulfides and the hydrophobic residues contributing to binding are shown as sticks. The chain directions of bound peptides and proteins are indicated by N (N-terminus) and C (C-terminus). The structures were superimposed on the thioredoxin moiety using the align command in Pymol.

HvTrxh2-BASI complex. In four of the structures (the peptide complexes, the C32S BsTrxA homodimer and the BsTrxA-ArsC structures) the hydrophobic interactions occur in the β3-β4 region (Figure 3.10). In contrast, the hydrophobic interaction between Trp147 in the BASI substrate and HvTrxh2 occurs at the other side of the intermolecular disulfide bond and involves the β5-α4 loop (Figure 3.10). Residue Ile72-BsTrxA/Thr74-Trx1/Met88-HvTrx occupy equivalent positions in the thioredoxin structures on the α3-β4 loop (part of the β3-β region) and as a rule this residue forms intermolecular backbone-backbone hydrogen bond(s) to the substrate cysteine residue involved in the intermolecular disulfide bond, or to the residue next to it. The amide and carbonyl groups of this residue are exposed and available for hydrogen bonding due to the adjacent conserved *cis* proline residue (Pro73-BsTrxA/Pro75-Trx1/Pro89-HvTrxh2), which causes a change in direction of the main chain (Maeda et al, 2006). This *cis* proline is an invariant feature of thioredoxins and backbone-backbone
hydrogen bonds from its preceding residue to substrate proteins seems to be an invariant feature of thioredoxin complexes. These two general interactions, the hydrophobic residue binding to the thioredoxin surface and the backbone-backbone hydrogen bonding by the residue preceding the conserved cis proline, are displayed in the C32S BsTrxA homodimer. It is thus similar to the thioredoxin-substrate complexes. Apart from these two general interactions the thioredoxin binding modes in the substrate complexes vary greatly in the number of additional hydrogen bonds, the hydrophobic interactions formed and the residues involved. The diversity of binding modes observed in these five complexes illustrates how thioredoxin can bind and react with a variety of different substrates.

3.4.4 Comparison with thioredoxin-thioredoxin reductase complexes

Thioredoxin catalyses an interconversion of a disulfide and a dithiol. As the substrate protein is reduced, thioredoxin becomes oxidized forming an intramolecular disulfide bond between its two vicinal active site cysteines. Oxidised thioredoxin, Trx-S₂, is then reduced back to its active form, Trx-(SH)₂, by thioredoxin reductase (TR). Structures of complexes of E. coli thioredoxin with thioredoxin reductase (PDB entry 1F6M (Lennon et al, 2000)) and spinach chloroplast thioredoxins (Trx-f and Trx-m) with Synechocystis ferredoxin-thioredoxin reductase (FTR) (PDB entries 2PU9, 2PUK, (Dai et al, 2007)) are available. These complexes show different thioredoxin binding modes compared to the thioredoxin-substrate complexes.

E.coli thioredoxin reductase (TR) enfaolds TrxA on two sides forming extensive interactions. E.coli TR has a large hydrophobic residue, Phe142, which binds to the α₃-β₄ loop in approximately the same position as the hydrophobic residues in the substrate complexes. E.coli TR further interacts with the thioredoxin substrate-binding regions through four hydrogen bonds to the α₃-β₄ loop and two hydrogen bonds to the β₅-α₄ loop, thus ensuring tight packing between the molecules. TR also forms a hydrogen bond to the TrxA C-terminal α₄-helix (residues 95-106) and a salt bridge to the long α₃-helix (residues 32-49).

In the spinach chloroplast Synechocystis ferredoxin-thioredoxin reductase (FTR) complex with Trx-f and Trx-m, the FTR interacts with the thioredoxin active site loop and the β₅-β₄ region with two separate loops (Figure 3.11) forming a total of five hydrogen bonds. The ring of the tryptophan residue preceding the attacking cysteine residue in Trx-f and Trx-m packs against a long helix in FTR. There is no large hydrophobic residue from ferredoxin thioredoxin reductase interacting with the thioredoxin surface, nor are any interactions made with the third substrate-binding region in thioredoxin, the β₅-α₄ loop.
Figure 3.11. Comparison with thioredoxin reductase binding modes. The thioredoxins and their binding partners are shown in colours according to the colour key. The disulfides and the hydrophobic residues contributing to binding are shown as sticks. The chain directions of bound peptides and proteins are indicated by N (N-terminus) and C (C-terminus). The N1/C1 of FTR indicate chain direction of an N-terminal loop binding to the thioredoxin surface, and N2/C2 indicate the chain direction of a C-terminal loop binding to the thioredoxin surface.

Both the *E. coli* TrxA-TR complex and the spinach chloroplast *Synechocystis* Trx-*fl*-FTR complexes thus show distinctly different binding modes compared to the thioredoxin-substrate complexes. The C32S BsTrxA homodimer complex has a large hydrophobic residue (Trp28) interacting with the shallow hydrophobic groove on the thioredoxin surface and it has a chain direction through the active site similar to that of the NFκB peptide. Hence the C32S BsTrxA homodimer is more closely related to the thioredoxin-substrate complexes rather than the thioredoxin-thioredoxin reductase complexes.

### 3.4.5 Conclusion

The binding mode in the C32S BsTrxA homodimer resembles those of thioredoxin substrate complexes. It can thus be regarded as a representation of a mixed disulfide BsTrxA reaction intermediate. As such it is the first example of a thioredoxin protein-protein complex with a chain orientation similar to that of the peptide from NFκB in complex with human Trx1. The C32S BsTrxA homodimer structure, in comparison with thioredoxin substrate complexes, illustrates the common factors in thioredoxin-substrate binding modes as well as the variability of these binding modes that enables thioredoxin to bind to and react with a variety of different substrates.
ARGININE DECARBOXYLASE

FROM ESCHERICHIA COLI
4.1 INTRODUCTION

4.1.1 Acid resistance in *E. coli*

In order to infect a human host, enteric bacteria must pass through the stomach, which has a pH of around 2.0, and survive there for approximately two hours before the stomach is emptied. *Escherichia coli*, *Shigella flexneri* and *Salmonella enterica* serovar Typhimurium are enteric bacteria that are capable of surviving in the highly acidic environment of the stomach since they possess systems that make them acid resistant. Three such acid resistance systems have been identified in *E. coli*: (1) oxidative glucose-repressed acid resistance (AR1), (2) glutamate-dependent acid resistance (AR2) and (3) arginine-dependent acid resistance (AR3) (Lin et al, 1996). The three AR systems work independently and are redundant, ensuring the survival of the stationary phase organism in a variety of acidic environments.

4.1.1.1 AR1 system in *E. coli*

The AR1 system requires oxidative growth and successfully protects cells to pH 2.5 in unsupplemented minimal media (Castanie-Cornet et al, 1999; Lin et al, 1995). However, the exact components comprising the AR1 system and the mechanism of action of the system remains largely uncharacterised. A study has shown that AR1 is repressed by glucose, which allowed the discovery of the glutamate- and arginine-dependent acid resistance systems, AR2 and AR3 respectively (Lin et al, 1995; Lin et al, 1996).

4.1.1.2 AR2 and AR3 systems in *E. coli*

The AR2 and AR3 systems are also referred to as the amino acid dependent acid resistance systems, since they are dependent on the presence of glutamate and arginine, respectively. Each system consists of a decarboxylase and an integral membrane protein antiporter. AR2 consists of the glutamate decarboxylase of which there are two isoforms (GadA and GadB) and the glutamate:GABA antiporter (GadC). AR3 consists of the acid inducible biodegradative arginine decarboxylase (AdiA) and the arginine:agmatine antiporter (AdiC) (Gong 2003, Iyer 2003), which are expressed at high levels in the cell at low pH in rich media (Auger et al, 1989). AdiA catalyses the decarboxylation of one molecule of arginine into agmatine, consuming one proton in the process, and the integral membrane protein AdiC (Gong et al, 2003) transports the agmatine out of the cell in exchange for arginine in a 1:1 stoichiometry (Iyer et al, 2003). Hence AdiA and AdiC are effectively shuttling out one...
4.1.2 Arginine dependent acid resistance (AR3)

4.1.2.1 pH homestasis

The consumption of protons during the AdiA catalysed decarboxylation reaction coupled with the extrusion of the product agmatine in exchange for the substrate arginine seem to protect the cell pH homeostasis by effectively shuttling out one proton from the interior of the cell per decarboxylation reaction (Figure 4.1). Arginine-dependent acid resistance is only active at acidic pH levels harmful to the cell (Gong et al, 2003), so when the external pH = 2.5 the rate of agmatine extrusion is at a maximum, while when the external pH = 3-4.0 almost no agmatine is exported (Gong et al, 2003; Iyer et al, 2003). When maximum extrusion rate of agmatine at an external pH = 2.5 occurs, the internal pH is elevated by the arginine-dependent
AR3 system from 3.6 to 4.7 (Richard & Foster, 2004), which correlates with the pH optimum of 5.2 of the arginine decarboxylase AdiA enzyme (Blethen et al, 1968). Similarly, the glutamate-dependent AR2 system raises the internal pH from 3.6 to 4.2 during a pH 2.5 acid challenge (Richard & Foster, 2004). This suggests that the amino acid dependent acid resistance systems contribute to the bacteria’s survival by raising the internal pH of the cell.

4.1.2.2 Reversal of the electrochemical gradient \( \Delta \psi \)

However, a specific internal pH is not the only requirement for surviving at low pH. Richard & Foster found that during extreme acid stress at pH 2.5 *E. coli* reverses its electrochemical gradient (\( \Delta \psi \)) from a negative inside charge to a positive inside charge (Richard & Foster, 2004). This is a strategy also utilized by acidophiles, which is proposed to help repel protons and maintain a higher internal pH. Richard & Foster hypothesize that this positive inside change is due to the decarboxylation and antiport processes of AR2 and AR3. This is based on findings that in the absence of glutamate and arginine the \( \Delta \psi \) went from \(-86\) mV to zero, while the \( \Delta \psi \) increased dramatically to \(+30\) mV and \(+80\) mV in the presence of glutamate and arginine respectively. This reversal of the transmembrane potential was suggested to be caused by two processes that happens when the cell is exposed to an external pH of 2.5: (i) the initial influx of protons into the cell that causes the intracellular pH to drop from 7.5 to 4.5 and (ii) a build up of positively charged product (agmatine/GABA) from the decarboxylation required to drive the antiport. These two processes have an overall result of reversing the intracellular charge from negative to positive (Richard & Foster, 2004).

4.1.3 Amino acid decarboxylases are PLP enzymes

4.1.3.1 Pyridoxal-5'-phosphate

In extreme acid environments the internal pH of *E. coli* is elevated by the arginine-dependent AR3 system from 3.6 to 4.7 (Richard & Foster, 2004). This correlates with the pH optimum of 5.2 of the AdiA, which is 90% active at pH 4.5 and at pH 6.0 (Blethen et al, 1968). The activity of AdiA is dependent on the vitamin B6 derived cofactor, pyridoxal-5'-phosphate (PLP) (Figure 4.2), a common cofactor in enzymes involved in amino acid biosynthesis. PLP enzymes are capable of catalyzing a broad range of chemical reactions such as transamination, racemization, elimination and replacement reactions in addition to decarboxylation. In addition, the PLP cofactor is capable of slowly catalyzing many of these reactions in the absence of enzyme. The function of the protein is to provide substrate and reaction specificity and to enhance the catalytic properties of the PLP (Eliot & Kirsch, 2004; Hayashi, 1995; John, 1995).
4.1.3.2 Group III decarboxylases

PLP enzymes are generally divided into five groups based on their fold type or whether they catalyse a reaction at the α, β, or γ position of an amino acid. The α-family and most PLP-dependent amino acid decarboxylases belong to Fold Type I, also referred to as the aspartate aminotransferase (AspAT) fold after the first PLP enzyme structure determined (Table 4.1) (Ford et al, 1980). PLP-dependent amino acid decarboxylases can be further classified as belonging to one out of four different classes of decarboxylases (Group I-IV) based on sequence similarity (Sandmeier et al, 1994). Acid induced biodegradative arginine decarboxylase (AdiA) belongs to group III decarboxylases together with biodegradative ornithine decarboxylase (OrnDC) and biodegradative lysine decarboxylase (CadA). The group III decarboxylases use basic amino acid substrates and are large enzymes with approximately ∼ 730 amino acids per monomer (Morris & Boeker, 1983; Sabo & Fischer, 1974). The structure of OrnDC from Lactobacillus 30a (PDB entry 1ORD, (Momany et al, 1995) has until now been the only structure available from this group of enzymes (Table 4.1). The structure of OrnDC will be further discussed in the context of the AdiA structure presented here.

4.1.3.3 Group II decarboxylases

Three structures are available from the group II decarboxylases: glutamate decarboxylase (GadB) from E. coli (Capitani et al, 2003a), human glutamate decarboxylase (Gad67) (Fenalti et al, 2007) and aromatic L-amino acid decarboxylase, also called DOPA decarboxylase (DDC), from pig kidney (Burkhard et al, 2001)(Table 4.1). These enzymes have an N-terminal domain of variable length, a large PLP-binding domain and a small AspAT-like domain. The PLP-binding domain exhibits the typical α/β fold of the large domain of AspAT with a characteristic seven stranded β-sheet flanked by α-helices (Eliot & Kirsch, 2004). In common with most Fold Type I PLP enzymes, the group II and group III decarboxylases

---

**Figure 4.2. Pyridoxal-5'-phosphate (PLP).**
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Organism</th>
<th>Classification</th>
<th>PDB entry</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartate amino transferase</td>
<td>AspAT</td>
<td>Chicken</td>
<td>1AAT</td>
<td>(Malashkevich et al, 1982)</td>
</tr>
<tr>
<td>Diakylglycine decarboxylase</td>
<td>DGD</td>
<td>B. cepacia</td>
<td>2DKB</td>
<td>(Toney et al, 1993)</td>
</tr>
<tr>
<td>Aromatic L-amino acid decarboxylase (DOPA)</td>
<td>DDC</td>
<td>Pig</td>
<td>1JS3,1JS6</td>
<td>(Burkhard et al, 2001)</td>
</tr>
<tr>
<td>Group II decarboxylase</td>
<td>Gad67</td>
<td>Human</td>
<td>2OKJ</td>
<td>(Fenalti et al, 2007)</td>
</tr>
<tr>
<td>Glutamate decarboxylase</td>
<td>GadB</td>
<td>E. coli</td>
<td>1PMM,1PMO</td>
<td>(Capitani et al, 2003a)</td>
</tr>
<tr>
<td>Ornithine decarboxylase</td>
<td>OrnDC</td>
<td>Lactobacillus 30a</td>
<td>1ORD,1C4K</td>
<td>(Momany et al, 1995) (Vitali et al, 1999)</td>
</tr>
<tr>
<td>Arginine decarboxylase</td>
<td>AdiA</td>
<td>E. coli</td>
<td>2VYC</td>
<td>This study</td>
</tr>
</tbody>
</table>
decarboxylases associate into homodimers, or homodimers assembled into oligomers. There are two PLP-containing active sites per homodimer, located on the dimer interface with each monomer contributing essential residues to both active sites (Figure 4.3).

### 4.1.3.5 Amino acid decarboxylases in acid resistance

The biodegradative ornithine- and lysine decarboxylases are induced at low pH (Morris & Fillingame, 1974). Hence they have been suggested to regulate cell pH through their consumption of protons in the decarboxylation reaction (Morris & Fillingame, 1974) in a similar way to AdiA and GadB. However, only the systems involving biodegradative arginine decarboxylase (AdiA) (AR3) and glutamate decarboxylase (GadB) (AR2) confer acid resistance to *E. coli* when the external pH is 2.5 (Lin et al, 1996). This has been suggested to be due to the lower pH optima of AdiA and GadB (5.2 and 4.6 respectively) compared to the higher pH optima of the *E. coli* ornithine- and lysine decarboxylases (6.9 and 5.7 respectively) (Richard & Foster, 2004). This hypothesis was based on the observation that when the external pH is 2.5 the AR2 and AR3 systems elevate the internal pH of the bacterial cell from 3.6 to 4.2 and 4.7 respectively, which correlates with the pH optima of 4.6 and 5.2 of GadB and AdiA respectively.

### 4.1.5 Aims of this study

The acid resistance systems in enteric bacteria allow the organism to survive in hostile acidic environments, like the stomach. The arginine-dependent AR3 system forms an important part of the acid resistance of *E. coli*. This system is proposed to protect the cell at an external pH of 2.5 by raising the internal pH and reversing the $\Delta \psi$ from negative to positive inside the cell (Richard & Foster, 2004). The biodegradative arginine decarboxylase (AdiA) is an essential part of the AR3 system. In order to investigate the molecular mechanism of AdiA in *E. coli* acid resistance the structure was determined to 2.4 Å resolution by X-ray crystallography. A detailed analysis of the structure is presented together with a comparative analysis of AdiA with homologous enzymes.
Figure 4.3. PLP enzymes discussed in this study. Fold Type I enzyme dimers are shown in red and green. PLP cofactors are shown as yellow spheres.
4.2 METHODS AND MATERIALS

4.2.1 Expression and purification of AdiA

The expression and purification of AdiA described in this section is the work of Dr Matthew G. Hicks who worked with Prof Tracy Palmer John Innes Centre, Norwich.

AdiA was purified instead of a target his-tagged outer membrane protein. E. coli strain BL21(DE3) (Studier et al, 1990) was transformed with plasmid pREP4, which encodes the lacI⁹ allele of the LacI repressor protein. This was transformed with plasmid pQE70 (Qiagen) containing the gene for the target outer membrane protein. The cells were grown at 37 °C with shaking at 180 RPM in Luria Bertani medium (1% Tryptone, 0.5% yeast extract, 1% NaCl, supplied by Formedium of Norwich) supplemented with 0.2% glucose, ampicillin (100 μg/ml) and kanamycin (30μg/ml) in unbuffered conical flasks. Typically, an overnight culture was used to inoculate 1 L of medium in a 2 L flask at a 1:100 dilution. Cultures were grown for 2-4 h to an OD₆₆₀ of ~0.4 when overexpression of the target outer membrane protein was induced with 2mM IPTG. Cells were grown for a further 3h then harvested at 5,500 RPM in a Sorvall SLA6000 rotor. Cell pellets were combined and washed in 20 mM Na-HEPES pH 8.0, 150 mM NaCl (Buffer 1). Cells were harvested again and pellets routinely frozen at –80 °C. Washed cell pellets were re-suspended in Buffer 1 at 3 ml/L of culture. Cell suspensions were disrupted by passage through a French pressure cell in the presence of Complete™ EDTA-free protease inhibitor (Roche), DNase and lysozyme (Sigma). Cell debris was removed by centrifugation for 10 min at 7000 RPM in a Sorvall SS34 rotor. Cell envelopes were isolated by ultracentrifugation at 45,000 RPM in a Beckman Ti70 rotor for 90 minutes. Protein concentration in crude cell envelopes resuspended in Buffer 1 was determined using the Bio-Rad detergent-compatible assay kit, based on the method of Lowry (1951) (Lowry et al, 1951). Crude cell envelopes were agitated for 1 hour in the presence of 1% N-lauroyl sarcosine in Buffer 1, at 5-10mg/ml total cell envelope protein. AdiA was present mainly in a form resistant to N-lauroyl sarcosine extraction. Insoluble material was removed by further ultracentrifugation at 45,000 RPM in a Beckman Ti70 rotor for 60 minutes, and then agitated for 1 hour in 1% C₈E₄ detergent in Buffer 1 at a protein concentration at 5-10mg/ml protein. Insoluble material was removed by a further 60 minute ultracentrifugation step, as previously. Soluble material was then applied to a 5ml Hi-Trap chelating Ni-column (GE Lifesciences) using an ÄKTA-FPLC. For unknown reasons, AdiA bound to this nickel-chelating column. The column was washed in 12 mM Imidazole (Fluka) in Buffer 1 supplemented with 0.05%
C<sub>8</sub>E<sub>4</sub>. A gradient of 12-150 mM Imidazole in Buffer 1 supplemented with 0.05% C<sub>8</sub>E<sub>4</sub> over 6 column volumes was applied. AdiA eluted at between 60-90 mM Imidazole. Fractions containing AdiA were analyzed by SDS-PAGE using 10% polyacrylamide gels, with AdiA displaying a characteristic heat modifiability, converting to its monomeric mass of 87 kDa from a mass of >250 kDa after 5 minutes heat treatment in the loading buffer of Laemmli (1970) (Laemmli, 1970). Fractions containing AdiA were pooled and concentrated to a volume <500µl and applied to a Superdex 200 HR10/30 gel-filtration column equilibrated in 20 mM Na-HEPES, pH 8.0, 50 mM NaCl, 0.05% C<sub>8</sub>E<sub>4</sub> and eluted at 0.5 ml/min. AdiA eluted at >99% purity at between 8-11 ml. Samples containing AdiA were analyzed by SDS-PAGE.

4.2.2 Crystallisation

Purified AdiA in a buffer containing 20 mM HEPES pH 8.0, 50 mM sodium chloride and 0.05% C<sub>8</sub>E<sub>4</sub> detergent was provided by Dr Matthew G. Hicks (Buffer A). Samples were concentrated to 13 mg/ml using an Amicon ultracentrifugation device with a molecular weight cut-off of 100 kDa. Crystallization conditions were screened at 293 K, according to the sparse matrix method (Jancarick, 1991), using commercially available screens (MemStart/MemSys from Molecular Dimensions) and the sitting drop vapor diffusion technique. Screens were set up in 96-well plates (MRC 2well plates from Swisscl) using a Cartesian Microsystems robot. Drops consisted of protein (0.2 µl) with an equal volume of reservoir solution. Initially, crystals were observed in condition 43 of the MemStart screen (100 mM MES pH 6.5, 10% PEG 8000 and 200 mM sodium acetate). Further refinement of the crystallization conditions produced crystals of typical dimensions 0.1 x 0.2 x 0.4 mm³, which were grown by mixing the protein solution (0.2 µL, 13 mg/mL in 20 mM HEPES pH 8.0, 50 mM sodium chloride and 0.05% C<sub>8</sub>E<sub>4</sub> detergent) with an equal volume of reservoir solution (100 mM MES pH 6.5, 13% PEG 8000 and 400 mM sodium acetate). Typically, crystals grew to their maximum dimensions in 3-5 days. Crystals of AdiA were cryoprotected by transferring each crystal into drop solution (1 µl, 100 mM MES pH 6.5, 13% PEG 8000, 400 mM sodium acetate, 20 mM HEPES pH 8.0, 50 mM sodium chloride and 0.05% C<sub>8</sub>E<sub>4</sub>) and increasing the glycerol content of the drop to a final concentration of 20% in ca. 7% increments. The crystal was briefly immersed into a second drop (1µl) containing 20% glycerol, immediately prior to transfer into liquid nitrogen.

4.2.3 Heavy atom derivatives

Lead and mercury-derivatized crystals of AdiA were prepared by transferring individual crystals into drop solution (2-3 µl, 100 mM MES pH 6.5, 13% PEG 8000, 400 mM sodium
acetate, 20 mM HEPES pH 8.0, 50 mM sodium chloride and 0.05% C₈E₄ detergent) containing trimethyllead acetate (PbMe₃) (2 mM final concentration) or mercuric chloride (HgCl₂) (1 mM final concentration), respectively. The crystals were soaked for a total of 45 hours, and were cryoprotected as described, including PbMe₃ or HgCl₂ in the cryo-buffer.

### 4.2.4 Data collection

X-ray data were collected from frozen crystals at 100 K at the ESRF. Diffraction data were collected from a crystal of native AdiA (‘Native-AdiA’; Table 4.2) on beam line ID14-4 on a Q315r ADSC CCD image-plate detector at a fixed wavelength of 0.93 Å. A dataset from a crystal soaked in HgCl₂ (‘Hg-AdiA’; Table 4.2) was collected on beamline ID14-1 on a ADSC Q210 CCD image-plate detector at a fixed wavelength of 0.93 Å. A dataset from a lead-soaked crystal (‘Pb-AdiA’; Table 4.2) was collected on beamline ID14-3 on a Q4R ADSC CCD image-plate detector at a fixed wavelength of 0.93 Å. The data was processed using the HKL suite of programs, DENVZO and SCALEPACK (Otwinowski, 1997). Data collection statistics are presented in Table 4.2.

### 4.2.5 Structure determination

The 2.4 Å resolution structure of AdiA was solved by MIRAS methods using data from crystals soaked in solutions containing PbMe₃ or HgCl₂. The positions of 10 Pb atoms in the Pb-AdiA data were determined by direct methods, with SHELXD (Uson & Sheldrick, 1999), using isomorphous differences between the Pb-AdiA and Native-AdiA data. These sites were used to calculate initial SIRAS phases to the resolution limit of the Pb-data (3.0 Å) using the program SHARP (Bricogne et al, 2003). A 5-fold NCS averaging matrix was calculated by RESOLVE (Terwilliger, 2000) using the coordinates of the Pb heavy atom positions. Density modification carried out to 3.0 Å resolution using the program DM and solvent flattening histogram mapping and averaging options (Cowtan, 1994) was used to improve the initial Pb-SIRAS phases. This yielded an interpretable electron density map. In order to improve the quality of the experimental electron density map further, an additional 20 Hg atom positions were identified in anomalous difference Fourier maps calculated using FFT (Read, 1988) with the Hg-AdiA data and the SIRAS phases from the Pb derivative. SHARP was used to calculate MIRAS phases using a total of 10 Pb and 20 Hg sites. The refined Pb coordinates were used to calculate a 10-fold NCS averaging matrix, and density modification was carried out using the program DM, with phase extension to 2.4 Å, the resolution of the Native-AdiA data.
4.2.6 Model building

An incomplete initial model of AdiA was built using automatic modeling programs BUCCANEER (Cowtan, 2006) and ARP-WARP (Perrakis et al, 1999). Further model building, to complete the trace of AdiA chain A, was carried out in COOT (Emsley & Cowtan, 2004), using the structure of the paralogue ornithine decarboxylase (PDB entry 1ORD, (Momany et al, 1995), modified with CHAINSAW (Schwarzenbacher et al, 2004), as a template. Applying individual components of the 10-fold NCS matrix using PDBSET generated AdiA chains B-K. Initial rigid body and restrained refinement in REFMAC5 (Murshudov et al, 1997) gave residuals for the initial model of \( R_{\text{work}} = 40\% \) and \( R_{\text{free}} = 37\% \). Tight NCS restraints were applied throughout subsequent rounds of refinement in REFMAC5. TLS refinement (Painter & Merritt, 2006) was applied for each chain in the last cycles of refinement. The model was improved upon by iterative cycles of model building and map calculation. The NCS restraints were removed in the final cycles of refinement. Data to 2.4 Å resolution was included for the native dataset in spite of a high \( R_{\text{sym}} \) in the highest resolution shell (78.5%) since the refined structure shows reasonable residuals of \( R = 25.4\% \), \( R_{\text{free}} = 31.8\% \) in the highest resolution shell (Table 4.2). The coordinates and structure factors were deposited with the Protein Data Bank as PDB entry 2VYC.

4.2.7 Structure analysis

Model geometry was analyzed using COOT and the program MOLPROBITY (Davis et al, 2007). Figures were generated using the program Pymol (DeLano, 2002). Rmsd values between structures were obtained using Dallilite (Holm & Park, 2000). Buried surface area in interfaces were calculated with the program AREAIMOL (Lee & Richards, 1971). Chemical structure figures were generated with the program ChemDraw.

4.3 RESULTS

4.3.1 Quality of the AdiA model

The structure of AdiA was solved by the MIRAS method using data from a native crystal and crystals derivatized with lead and mercury. The structure was refined to 2.4 Å resolution. The asymmetric unit contains ten molecules, chains A-J. They are organized as two pentameric
Table 4.2. AdiA data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection and processing</th>
<th>Native-AdiA</th>
<th>Pb-AdiA</th>
<th>Hg-AdiA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>P6$_4$</td>
<td>P6$_4$</td>
<td>P6$_4$</td>
</tr>
<tr>
<td><strong>Cell dimensions a, c (Å)</strong></td>
<td>197.7, 450.3</td>
<td>198.3, 449.8</td>
<td>197.7, 449.9</td>
</tr>
<tr>
<td><strong>Soaking concentration and time</strong></td>
<td>-</td>
<td>2 mM, 45 h</td>
<td>1 mM, 45 h</td>
</tr>
<tr>
<td><strong>ESRF beamline</strong></td>
<td>ID14-4</td>
<td>ID14-3</td>
<td>ID14-1</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td><strong>Resolution range (Å)</strong></td>
<td>50.0-2.4</td>
<td>50.0-3.0</td>
<td>50.0-3.5</td>
</tr>
<tr>
<td><strong>Measured reflections</strong></td>
<td>1913962</td>
<td>1147010</td>
<td>904555</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>385411</td>
<td>388656</td>
<td>117365</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong>$^a$</td>
<td>99.9 (100.0)</td>
<td>98.6 (97.4)</td>
<td>93.4 (87.2)</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>5.0</td>
<td>3.0</td>
<td>7.7</td>
</tr>
<tr>
<td><strong>I/σ(I)</strong>$^a$</td>
<td>10.7 (2.3)</td>
<td>9.8 (2.2)</td>
<td>14.3 (3.1)</td>
</tr>
<tr>
<td><strong>R$_{sym}$ (%)$^b$</strong></td>
<td>15.0 (78.5)</td>
<td>10.4 (38.5)</td>
<td>15.7 (58.3)</td>
</tr>
</tbody>
</table>

| **Model refinement**           |            |         |         |
| Resolution (Å)                 | 2.4        |         |         |
| $R_{work}$ (%)$^c$             | 17.7 (25.4) |         |         |
| $R_{free}$ (%)$^d$             | 22.9 (31.8) |         |         |
| Average B factor (Å$^2$)       | 23.6       |         |         |
| Phasing power ISO [acentric/centric] | 1.21/0.95 | 0.78/0.70 |         |
| Phasing power ANO              | 0.24       |         |         |

| **Stereochemistry**            |            |         |         |
| Root mean square deviation for | 0.014      |         |         |
| bond length (Å)                |            |         |         |
| Root mean square deviation for | 1.464      |         |         |
| bond angles (°)                |            |         |         |

| **Residues in the Ramachandran plot$^f$** |            |         |         |
| Favored regions (%)            | 97.3%      |         |         |
| Allowed regions (%)            | 100.0%     |         |         |

$^a$ Statistics for the highest resolution shell are given in parentheses.

$^b$ $R_{sym} = \sum_h \sum_i |I_{hi}| - \langle |I_{hi}|\rangle / \sum_h \sum_i |I_{hi}|$

$^c$ $R_{work} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$ for all reflections.

$^d$ $R_{free} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$ calculated using randomly selected reflections (5%).

$^e$ Phasing power: $1 F_{calc} / phase-integrated lack of closure.

The overall figure of merit is [acentric/centric] = 0.25/0.28.

$^f$ Ramachandran plot values as assessed by MOLPROBITY.
ring structures, chains A-E and F-J (Figure 4.4A). Refinement of the model converged with residuals $R_{work} = 17.7\%$, $R_{free} = 22.9\%$. The model shows good geometry with 97.3 % in the most favored region of the Ramachandran plot as assessed by MOLPROBITY (Davis et al, 2007) (Table 4.2). The ten identical molecules, chains A-J, in the asymmetric unit superpose with an average rmsd value of 0.2 Å for 755 common Cα atoms, which is consistent with the resolution of the structure and the fact that non-crystallographic symmetry restraints were applied until the last cycles of refinement. Altogether, the asymmetric unit contains 7550 protein residues, 2556 water molecules and 10 pyridoxal-5’-phosphate (PLP) molecules. For the bulk of the structure, the electron density is very well defined, which allowed the modeling of all 755 protein residues per chain. The average temperature factor of the model (all atoms) is $B = 23.6 \text{ Å}^2$.

Figure 4.4. AdiA decamer crystallises as a decamer. A. Two AdiA pentameric rings are found in the asymmetric unit. Molecules A-E are shown in skyblue and molecule F-J are shown in yellow. B. Sideview of the AdiA decamer. The two pentameric rings in the AdiA decamer are related by a 2-fold crystallographic symmetry axis. Molecules A-E are shown in skyblue and molecules A*-E* are shown in pink. C. Bottomview of the AdiA decamer.

4.3.2 The AdiA decamer and the molecule AC* dimer

AdiA crystallizes as two pentameric rings in the asymmetric unit in space group $P6_1$ (Figure 4.4A). The application of the $P6_1$ symmetry operators creates six decamers per unit cell. Each decamer is composed of two pentamers related by a crystallographic two-fold axis (Figure 4.4B). For example, the decamer composed of chains A-E and chains A*-E* includes the A*-E* pentamer, which is generated by symmetry operator (-x+1, -y, z) (Figure 4.5). This decamer has a diameter of $\sim 180 \text{ Å}$, is $\sim 100 \text{ Å}$ high and contains a central hole of $\sim 30 \text{ Å}$.
Figure 4.5. AdiA decamer. A. Molecules A-E, bottomview, same view as in C. B. Molecules A*-E*, topview, same view as in C. C. AdiA decamer. Molecules A/A* (blue), molecules B/B* (purple), molecules C/C* (red), molecules D/D* (yellow), molecules E/E* (orange).

diameter. Within the decameric structure, each molecule in the A-E pentamer interacts with only one partner in the neighbouring A*-E* pentamer. For example, molecule A interacts with molecule C* only and this unit can be considered as an AC* dimer (Figure 4.6). The AdiA decamer can thus be described as a pentamer of dimers (dimers AC*, BB*, CA*, DE* and ED*). The structure of the AC* dimer in isolation, shows a compact globular “core” structure protruding from the wing domains of each monomer (Figure 4.6). The dimer has dimensions of approximately 100 x 55 x 55 Å and the two molecules are related by a crystallographic two-fold axis. 5233.2 Å² of the monomer becomes buried upon dimer formation and an additional surface area of 2731.5 Å² of the monomer becomes buried when the dimers assemble into a decamer.
4.3.3 The AdiA monomer

The structure of AdiA is homologous to the structure of biodegradative ornithine decarboxylase (OrnDC) from *Lactobacillus* 30a (PDB entry 1ORD, (Momany et al, 1995) (Figure 4.3F). For clarity the same terminology used by Momany et al will be employed when appropriate when discussing the structural organization of AdiA. Since the ten AdiA molecules in the asymmetric unit are identical, within experimental error, only molecule A will be discussed here.

Molecule A consists of one polypeptide chain, residues 1-755, which can be divided into five domains: (1) the amino-terminal wing domain (residues 1-139), (2) the linker domain (residues 140-192), (3) the PLP-binding domain (residues 193-439), (4) the aspartate aminotransferase (AspAT)–like small domain (residues 440-608) and (5) the carboxy-terminal domain (residues 609-755). The PLP-binding domain, the AspAT-like small domain and the carboxy-terminal domain form an open bowl-like structure, where each domain roughly forms a third of the “bowl”. The wing domain extends from the “bowl” like a handle, while the linker domain protrudes away at approximately a 90° angle relative to the other domains (Figure 4.7).

The wing domain, residues 1-139, consists of an α/β Rossmann fold. It has a central parallel β-sheet consisting of five β-strands. The β-sheet is flanked by six α-helices: two on one face of the sheet (α1α6), two on the other face (α2α3) and two at the C-terminal edge of the sheet close to strand β5 (α4α5). The end of helix α6 of the wing domain connects to the linker.
Figure 4.7. AdiA domains. The AdiA monomer (molecule A) is shown in the middle in the same view as in Figure 6C. The wing domain, residues 1-139, is shown in purple, the linker domain, residues 140-192, is shown in red, the PLP binding domain, residues 193-439, is shown in orange, the AspAT-like small domain, residues 400-608, is shown in blue, the carboxy-terminal domain, residues 609-755, is shown in green and the PLP cofactor is shown as a yellow sphere or as yellow sticks. The α-helices are numbered in black (except in the wing domain where they are numbered in orange) from 1-22 in the N-terminal to C-terminal direction. β-strands are numbered in red from 1-20 in the N-terminal to C-terminal direction.

domain. The linker domain, residues 140-192, is a small domain that consists of three α-helices (α7, α9, α10) and a 3<sup>10</sup>-helix (α8) connected by loops and turns. The PLP-binding domain, residues 193-439, is the largest domain and it displays a type I PLP enzyme fold based on its structural homology to the large domain of aspartate aminotransferase (AspAT). The AdiA PLP-binding domain superposes with the large domain of AspAT from <i>E. coli</i> (PDB entry 1C9C, Ishijima et al, 2000) with an rmsd of 3.3 Å for 198 aligned Cα atoms. The AdiA PLP-binding domain is an α/β complex consisting of a seven-stranded mixed β-sheet (β6-β12) and seven α-helices (α11-17) in a conserved type I fold. The PLP cofactor is bound at the C-terminal end of the β-sheet in a manner similar to other type I PLP enzymes. The AspAT-like small domain, residues 440-608, consists of four α-helices (α18-21), an antiparallel three-stranded β-sheet (β15β17β16) and a small parallel two-stranded β-sheet (β13β14). The fold of this domain is similar to that of the small domain of AspAT (PDB entry 1C9C, Ishijima 2000) in that the cores of the structures superpose with each other with an rmsd of 2.6 Å for 82 aligned Cα atoms. However, the extensive loop regions (loop α19-β15: residues 506-535, loop β15-α20: residues 540-556) and the small two-stranded beta sheet and small helix α19 present in this domain do not exist in the small domain of AspAT. The carboxy-terminal domain, residues 609-755, consists of five α-helices (α22-27) and a curved three-stranded anti-parallel β-sheet (β18-β20). A domain equivalent to the AdiA carboxy-terminal domain is found in the homologous structure of OrnDC from <i>Lactobacillus</i> 30a (PDB entry 1ORD, Momany et al, 1995), but not in the AspATs nor in any other protein in the protein data bank according to an SSM search (Krissinel & Henrick, 2004). Hence this domain appears unique to the group of basic amino acid decarboxylases to which AdiA and OrnDC belong. The five domains are linked together mainly through hydrogen bonds and electrostatic interactions.
4.3.4 The active site

The formation of the AC* dimer effectively buries the active sites of each molecule ~ 30 Å from the dimer surface. There are two active sites per dimer that are located at the dimer interface ~ 20 Å apart. The PLP-cofactor is partially solvent accessible through a deep cleft formed between the carboxy-terminal domain of molecule A and the PLP-binding domain of molecule C* in the AC* dimer. The PLP-cofactor is covalently bound to the conserved residue Lys A386 forming a Schiff’s base. Residues from both molecules A and C* form a total of nine hydrogen bonds to the PLP phosphate group (Figure 4.8). The PLP pyridine ring is sandwiched between the imidazole side chain of residue His A255 and the methyl group side chain of Ala A349. The N atom of the pyridine ring forms an electrostatic interaction with Asp A347 and the indole of Trp A350 forms a hydrogen bond to the pyridine ring 3-hydroxyl. The PLP cofactor is thus tightly bound and oriented in the AC* dimer via a covalent bond, tight coordination of the phosphate group and several interactions with the pyridine ring.

Figure 4.8. PLP coordination in the AdiA active site. The phosphate group of the PLP cofactor is coordinated by nine hydrogen bonds. Atom OP1 is coordinated by three hydrogen bonds to the side chains of residues His A385, Ser A383 and the main chain amide of residue Thr A230. Atom OP2 is coordinated by three hydrogen bonds to the side chain of Thr C*421 and the side chain and main chain amide of Ser C*422. Atom OP3 is coordinated by three hydrogen bonds the side chain of Thr C*420 and the side chain and main chain nitrogen of Ser A231. The N atom of the pyridine ring forms a salt bridge with Asp A347 and the indole of Trp A350 forms a hydrogen bond to the pyridine ring 3-oxygen. The figure was generated by the program ChemDraw.
4.3.5 Stabilisation of the decameric structure

The decameric structure as a pentamer of dimers is stabilized by the tight packing within dimers and by interactions made between the wing domains of neighbouring molecules in each pentameric ring.

4.3.5.1 Interactions in the AC* dimer

The buried surface area between the two molecules in the AC* dimer is extensive at 5233.2 Å² per molecule, which corresponds to 16.4 % of the solvent-accessible area of the molecule. Molecule A and C* form an extensive network of interactions with each other, consistent with their tight packing. Due to the two-fold crystallographic symmetry between molecule A and C* the interactions between molecule A to molecule C* are mirrored in equivalent interactions between molecule C* to molecule A. The molecule A and C* linker domains stabilize the AC* dimer by interlocking with each other. The linker domain of molecule A/C* also forms interactions with the wing domain, PLP-binding domain and the AspAT-like small domain of molecule C*/A, respectively (Figure 4.9). The PLP-binding domain of molecule A/C* interacts extensively with the PLP-binding domain of molecule C*/A, and this is where the largest interface between the molecules is located in the interior of the dimer. The PLP-binding domains of molecule A/C* also pack against the carboxy-terminal domain of molecule C*/A, respectively (Figure 4.9).

Figure 4.9. Packing of the AC* dimer. A. Sideview of AC* dimer. B. Surface representation of the AC* dimer. The view is rotated 90° clockwise around the y-axis compared to the view in A. C. Surface representation of the AC* dimer. The view is rotated 90° counterclockwise around the y-axis compared to the view in A. The dimer is coloured according to domains: wing domain (purple), linker domain (red), PLP-binding domain (orange), AspAT like small domain (blue), carboxyl-terminal domain (green).
4.3.5.2 Wing domain packing

In the decamer, the dimers pack against each other via the wing domains of the molecule where they interlock above and below in each pentameric ring. For example, in pentamer A-E, molecule A is flanked by molecule B and molecule E (Figure 4.10). The molecule A wing domain forms close interactions with both molecules B and E. The two small helices α4 and α5 in the molecule A wing domain interact closely with the long α18 helix in the molecule E AspAT-like small domain via two salt bridges and three hydrogen bonds. In addition, residues in the α5-α6 loop in the molecule A wing domain form two hydrogen bonds with the α21 helix in molecule E AspAT-like small domain. The molecule A wing domain also contacts the wing domains of both molecule B and E via a salt bridge between molecule A/B helix α1 and molecule E/A helix α2 and a hydrogen bond between molecule A/B loop β5-α6 and molecule E/A loop α2-β3, respectively. Altogether, the molecule A wing domain forms four salt-bridges and seven hydrogen bonds to molecules B and E (Figure 4.11). Equivalent contacts are made between the wing domains of each molecule to neighbouring molecules in the pentameric ring.

The extensive interactions within dimers and between dimers stabilize the AdiA decameric structure. Overall, 24.9 % (7964.7 Å²) of the surface area of molecule A is buried within the decamer through these extensive interactions with the flanking molecules B and E, and with the symmetry-related molecule C*.

![Figure 4.10. Wing domain packing in the AdiA decamer. A. One of the two pentameric rings of the AdiA dimer, bottomview. B. The AdiA wing domain packs against the AspAT-like small domain of a neighbouring molecule and against two flanking wing domains of neighbouring molecules. Molecule A (blue), molecule B (purple), molecule C (red), molecule D (yellow), molecule E (orange).](image-url)
Figure 4.11. Wing domain interactions in the AdiA decamer. A. AdiA pentameric ring coloured according to domains: wing domain (purple), linker domain (red), PLP-binding domain (orange), AspAT like small domain (blue), carboxy-terminal domain (green). An oval highlights the part of the structure shown in B. B. Schematic of the interactions between domains of molecules A and E and the wing domain of molecule B. Hydrogen bonds are represented by black lines, salt bridges are represented by red lines and hydrophobic packing is represented by a dotted line. The colour of the domain is the same as in A.
4.4 DISCUSSION

4.4.1 The AdiA decamer – a pentamer of dimers

The AdiA structure determined at 2.4 Å resolution shows that AdiA assembles as a ~ 800 kDa decamer, where each decamer can be considered as a pentamer of dimers. There are approximately three times as many interactions between the monomers in the dimer than interactions between neighbouring monomers in the pentameric rings. This suggests that the physiological AdiA decamer is assembled from five homodimers arranged with five-fold non-crystallographic symmetry, rather than from two separate pentameric rings (Figure 4.12). The assembly of the AdiA decamer as a pentamer of dimers is consistent with analytical ultracentrifugation results and kinetic studies of dimer dissociation (Boeker, 1978; Boeker et al, 1969). Overall, the structural organization of the AdiA decamer corresponds well with the model proposed by Boeker et al in 1969 and with what has previously been observed in electron micrographs (Boeker et al, 1969; Boeker & Snell, 1968).

![Figure 4.12](image1.png)

*Figure 4.12. The physiological AdiA decamer is a pentamer of dimers.*

4.4.2 AdiA decamer formation and enzyme activity

The AdiA decameric assembly is required for enzyme activity, which is optimal at pH 5.2. The enzyme associates from inactive dimers into functional decamers as the pH of the cell decreases below pH ~6.0, and they dissociate again into inactive dimers at pH 6.5 or above (Boeker & Snell, 1968; Nowak & Boeker, 1981). The dissociation of the AdiA decamer into inactive dimers at neutral pH is likely to serve as a regulatory mechanism to ensure that AdiA
is only active in the cell as the internal pH becomes more acidic (Boeker, 1978). The AdiA crystal structure was determined at pH 6.5, which should favour the dimeric form of the enzyme. However, decamer association is also favoured by an increase in the concentration of the enzyme itself, an increase in substrate concentration and/or an increase in ionic strength, such as an increased concentration of sodium ions (Boeker & Snell, 1968; Nowak & Boeker, 1981). The AdiA crystallization condition contained the protein at high concentration (~13 mg/ml) and a sodium ion concentration of 400 mM, which is likely to promote the decameric form of the enzyme in the crystal. It is more common for increasing ionic strength to promote subunit dissociation than association (Frieden, 1971). The fact that an increase in ionic strength promotes the association of AdiA from dimers to decamers suggests that a charge repulsion needs to be overcome for assembly to take place (Sabo & Fischer, 1974). Indeed, the AdiA homodimer has a net negative charge of -80 at neutral pH, resulting from the presence 184 solvent accessible acidic residues compared with 104 solvent accessible basic residues (Figure 4.13). The AdiA homodimer has a pI of 4.44 (Boeker et al, 1969), which is due to this excess of acidic residues. 25% of the solvent accessible acidic residues of the homodimer are located in the two wing domains. In each wing domain there are 23 solvent accessible acidic residues (11 Glu + 12 Asp) compared to 11 solvent accessible basic residues (7 Arg + 4 Lys). Hence, the wing domains carry a strong negative electrostatic potential at neutral pH. Decamer formation, and thus the activity of the enzyme, is structurally dependent on the interactions of the wing domains of the homodimers. The acidic surface residues in the

![Image](Figure 4.13. Electrostatic surface representation of AdiA AC* homodimer at neutral pH as rendered by Pymol. A. Sideview of the AC* homodimer. The PLP cofactor can be seen in yellow. B. 90° rotation around the y axis from the view shown in A. Negative electropotential is shown in red, positive electropotential in blue and neutral in white.)
wing domains, and in the homodimer as a whole, would start to become protonated as the pH decreases. This would decrease the net negative charge of inactive homodimers, which is likely to promote the association of the active AdiA decamers. This is supported by the observation that AdiA decamers form when the ionic strength increases and when the pH decreases to below ~6.0 (Boeker & Snell, 1968). AdiA thus combats the decreasing pH not only by consuming protons in the decarboxylation reaction, but also by acting as a biological buffer in its uptake of protons.

The AR3 system in *E. coli* is mimicking strategies commonly seen in acidophiles to cope with acidic pHs. For example, the AR3 system is known to cause the reversal of the membrane potential from negative to positive, which is a strategy observed in acidophiles proposed to deter the entry of protons into the cell (Richard & Foster, 2004). In addition, the abundance of acidic surface residues in AdiA is also a common feature for many acidophilic proteins (Bonisch et al, 2002; Cooper et al, 1990; Fushinobu et al, 1998; Kashiwagi et al, 1997). At low pH, an abundance of acidic surface residues increases protein stability by reducing positive electrostatic repulsion caused by the increase in proton concentration. At neutral pH, the abundance of acidic residues will lead to an increase in negative electrostatic repulsion with a corresponding reduction in protein stability, which is consistent with the dissociation of the AdiA decamer around pH 6.5.

### 4.4.3 AdiA active site and the decarboxylation reaction

#### 4.4.3.1 Conserved features in the AdiA active site

The active site coordination of the PLP cofactor in AdiA shows features conserved among fold type I PLP-dependent enzymes. For instance, (1) the PLP phosphate group is bound to the N-terminus of an α-helix (AdiA-α12), where the helix dipole stabilizes the negative charge of the phosphate group, (2) a lysine residue (AdiA-Lys386) forms a covalent Schiff-base linkage with C4’ atom of the pyridine ring, (3) an aromatic residue (AdiA-Trp350) and an alanine residue (AdiA-Ala349) are involved in binding the pyridine ring and (4) an aspartate residue (AdiA-Asp347) forms a salt bridge with the protonated nitrogen of the pyridine ring (Figure 4.8).

#### 4.4.3.2 AdiA decarboxylation reaction

The AdiA decarboxylation reaction most likely proceeds according to the generally accepted reaction coordinate of PLP enzymes (Eliot & Kirsch, 2004; John, 1995) as shown in Figure 4.14: (1) In the resting state of the enzyme, as observed in the AdiA structure, Lys386 forms an
internal aldimine with the PLP cofactor. The Lys386 residue will be displaced in a transimination reaction by the substrate L-arginine, forming an arginine aldimine (2) with the PLP cofactor through a covalent bond between the C4’ position of the pyridine ring and the amino group of L-arginine. The L-arginine Cα-COO⁻ bond to be broken in the decarboxylation reaction is predicted to align perpendicularly to the PLP pyridine ring (A), based on the Dunathan stereoelectric hypothesis and structural studies of substrate-bound dialkylglycine decarboxylase (Toney et al., 1995). This orientation allows for maximum overlap between the σ orbitals of the bond to be broken and the π-system of the PLP pyridine ring. This minimizes the energy of transition state for bond breaking, which makes the Cα-COO⁻ bond the most likely Cα bond to be broken. As the Cα-COO⁻ bond is broken the developing charge at the Cα position, caused by the loss of CO₂, is stabilized by delocalization within the extended conjugated π-system of the pyridine ring. The pyridine nitrogen acts as an electron-withdrawing group facilitating the breakage of the Cα-COO⁻ bond. The Asp347 residue provides a stabilizing negative charge to the protonated pyridine nitrogen, which enhances the electron withdrawing properties of the cofactor. The Cα-COO⁻ bond breakage produces the quinonoid intermediate (3), in which the conjugated π-system is extended from the pyridine ring to the Cα of the substrate aldimine (B). The quinonoid intermediate is protonated at Cα by an unidentified residue in the active site to form the agmatine aldimine
(4). The product agmatine is released from the enzyme as the internal aldimine between Lys386 and the PLP cofactor is reformed.

4.4.3.3 Identity of the proton donating residue in the decarboxylation reaction

The proton consumed in the AdiA decarboxylation reaction is donated by an as yet unidentified residue in the active site. It has been shown in kinetic experiments for fern L-methionine decarboxylase that a monoprotic acid protonates Cα in the decarboxylation reaction (Akhtar et al, 1990). It was suggested that a histidine residue, which can act as a monoprotic acid, would be the most likely candidate for protonating Cα (Gani, 1991). The histidine next in sequence to the internal aldimine forming lysine (AdiA-His385) is an unlikely candidate since it is involved in coordinating the phosphate. A more likely candidate, upon examining the AdiA active site, is the His255 residue. His255 would be ideally positioned as a proton donor as it is the active site histidine residue located closest to the C4’ position of the pyridine ring with its side chain imidazole amine within 4 Å. The His255 residue is conserved among PLP-dependent decarboxylases (Ishii et al, 1996) and is found in this position in all the fold type I PLP-dependent decarboxylase structures determined to date (Table 4.1), with the exception of the inducible glutamate decarboxylase (GadB) from E. coli that has a glutamine residue in this position. The absence of a histidine in this position in GadB together with a catalytically active DOPA decarboxylase mutant, where the histidine residue in this position was mutated into a glutamine (Bertoldi et al, 2001), suggests a different role for His255 than as the proton-donating residue. Hence, the identity of the proton-donating residue in AdiA remains to be confirmed.

4.4.3.4 AdiA carboxy-terminal domain and substrate specificity

The structure of AdiA shows the enzyme with an open active site where the PLP cofactor is clearly visible through a cleft formed between the carboxy-terminal domain of one monomer in the dimer and the PLP-binding domain of the other monomer in the dimer (Figure 4.15). In AdiA the ridge of the cleft formed by the carboxy-terminal domain is lined with negatively charged residues Glu735, Glu737 and Glu739 along the α26-β19 loop that runs from the surface of the molecule towards the interior of the active site. It is possible that these negatively charged glutamate residues could serve to attract the positively charged substrate L-arginine into the active site. In AspAT from E. coli the main determinants for enzyme substrate specificity are two arginine residues that coordinate the side chain dicarboxylic groups of the substrates glutamate and aspartate (Malashkevich et al, 1993). It is reasonable to expect that in AdiA negatively charged glutamate and/or aspartate residues would coordinate
Figure 4.15. AdiA active site cleft. A. Electrostatic surface representation of the AdiA homodimer. The square highlights the active site cleft. B. Closer view of the active site cleft between the carboxyl-terminal domain from molecule A and the PLP-binding domain of molecule C*. The cleft is lined with negatively charged residues Glu735, Glu737, Glu739 along a loop running from the surface to the active site. The PLP cofactor is shown as yellow spheres.

the positively charged guanidinium group of the L-arginine substrate. An agmatine aldimine molecule can comfortably be modelled into the active site of AdiA (molecule A) by superposition of the PLP cofactors (Figure 4.16). This approximate model puts the side chain guanidinium group of the agmatine aldimine approximately 5-6 Å from the side chain dicarboxylic groups of residues Glu A739 and Asp C*202. These are the closest glutamate and/or aspartate residues in the active site to the C4’ atom of the pyridine ring. PLP-dependent enzymes generally form a closed conformation around the bound substrate (Eliot & Kirsch, 2004). In AdiA, closing the cleft around the active site is likely to bring residues Glu A739 and Asp C*202 within coordinating distance of the guanidinium group of arginine or agmatine aldimine. This would be analogous to the situation in AspAT where the closed conformation brings the substrate-coordinating residue Arg386 closer to the coenzyme (Figure 4.16). From these observations it is possible to hypothesize that the carboxy-terminal domain would contribute to AdiA substrate specificity in two ways: (1) it would attract the doubly charged L-arginine species to the active site through glutamate residues lining the active site cleft and (2) one of these residues (Glu739) would be within a reasonable distance to coordinate the guanidinium group of the arginine or agmatine aldimine reaction intermediates in a substrate-bound closed conformation of the enzyme. If this is so,
Figure 4.16. AdiA active site in comparison with other PLP enzyme active sites. A. The AdiA active site in the open conformation modeled with an agmatine aldimine. Molecule A PLP binding domain (orange), molecule C* PLP binding domain (magenta), molecule A carboxy-terminal domain (green), PLP (yellow), agmatine aldimine model (beige). The distances between sidechain groups are indicated by black lines. The pyridine ring of the agmatine aldimine is tilted, which is known to occur in AspAT upon binding of the substrate. The His255 residue is positioned within 4 Å of the Cα atom of the agmatine aldimine. B. The open and closed conformations of the AspAT active site with bound substrate/substrate analogue (PDB entries 1MAP, 1C9C). Open state dimer with bound substrate analogue (gray), closed state monomer A (yellow), closed state monomer A* (redbrown), bound substrate glutamate ketimine (pink). C. Dialkylglycine decarboxylase active site in a closed conformation with bound aminophosphonate inhibitor (PDB entry 1M0Q). Molecule A (bluegreen), molecule A* (red), bound AMPP (R-1-amino-methylpropanephosphonate) (orange). The phosphate group of the inhibitor is coordinated by the enzyme in a manner that is likely to resemble that of the coordination of the substrate glycine carboxylate group. In all: nitrogen atoms (blue), oxygen atoms (red), phosphate atoms (grey) and hydrogen bonds are shown as blue dotted lines. D. DOPA decarboxylase active site in a closed conformation with bound inhibitor (PDB entry 1JS3). Molecule A
(purple), molecule B (green), bound inhibitor carbiDOPA (cyan). Residues from both monomers in the dimer contribute to the binding of the inhibitor. The aromatic ring of the inhibitor is accommodated by a hydrophobic pocket consisting of residues Trp A71, Phe A79 (not shown), ThrA82, Ile B101 and Phe B103.

it is possible that the maximal activity of AdiA at pH 5.2 reflects the pH at which Glu739 and Asp202 retain their negative charge for coordination of the substrate, while the optimum proportion of acidic surface residues are protonated to allow the assembly of functional decamers.

4.4.4 AdiA and OrnDC comparison

AdiA belongs to group III decarboxylases, which consists of AdiA, ornithine decarboxylase (OrnDC) and lysine decarboxylase (CadA) (Grishin et al, 1995; Sandmeier et al, 1994). The group III decarboxylases use basic amino acid substrates and are large enzymes with approximately ~ 730 amino acids per monomer (Morris & Boeker, 1983; Sabo & Fischer, 1974). As the cell pH becomes acidic, AdiA and CadA form decamers as pentamers of dimers and OrnDC forms a dodecamer as a hexamer of dimers (Momany et al, 1995; Sabo & Fischer, 1974). Two structures of the Lactobacillus 30a OrnDC have been determined as a dodecamer at 3.0 Å (PDB entry 1ORD, (Momany et al, 1995) and a mutant dimer at 2.7 Å resolution (PDB entry 1C4K, (Vitali et al, 1999). The 2.4 Å resolution structure of AdiA presented here allows for a more detailed comparison to be made within the group III decarboxylases than has previously been possible.

4.4.4.1 Oligomeric states and enzyme activity

The E. coli AdiA decamer and the Lactobacillus OrnDC dodecamer are homologous, they have a similar fold, similar domains, they superpose with an rmsd of 1.74 Å for 649 aligned Ca. positions and their sequence identity is 31%. The main difference between the AdiA and OrnDC structures is their oligomerization state. AdiA forms a decamer and OrnDC forms a dodecamer (Figure 4.17A,B). As a consequence, the wing domains of AdiA and OrnDC mediate different interactions in the AdiA decamer or OrnDC dodecamer. The OrnDC wing domain lacks the first two α-helices present in the AdiA wing domain. This prevents the OrnDC wing domain from forming the wing-wing contacts seen in AdiA between helices α1 and α2 of neighbouring AdiA molecules (Figure 4.17E, F). The wing-wing interactions in AdiA contribute to tighter packing between monomers, which is necessary for a pentameric
Figure 4.17. Wing domain packing interactions in AdiA and OrnDC. **A.** One of the two pentameric rings of the AdiA decamer, bottomview. **B.** One of the two hexameric rings of the OrnDC dodecamer, bottomview. **C.** The AdiA wing domain packs against the AspAT-like small domain of a neighbouring molecule and against two flanking wing domains of neighbouring molecules. **D.** The OrnDC wing domains pack against the AspAT-like small domain of a neighbouring molecule. **E.** Two flanking wing domains (molecule A in blue, molecule E in orange) pack against each other through helix α1 and helix α2, highlighted by a square in the figure. The helices are numbered 1-6 in the N- to C-terminal direction. The molecule A wing domain (blue) makes no interaction with the linker domain from molecule D* (green), part of the ED* dimer, as highlighted in the figure by a circle. **F.** The wing domains from neighbouring molecules in the OrnDC dodecamer (molecule A1 is shown in lime, molecule A2 is shown in red) do not form any interactions with each other, as highlighted in the figure by a circle. The molecule A1 wing domain (lime) packs against the molecule B2 (cyan) linker domain in the A1B2 dimer, highlighted by a square in the figure. The view is the same as in (E).
ring formation in contrast to a hexameric ring formation (Figure 4.17C, D). The wing domain of OrnDC, similar to AdiA, packs against the AspAT-like small domain of a neighbouring molecule (Figure 4.17C, D). In contrast to AdiA, the OrnDC wing domain also interacts with the linker domain in the neighbouring dimer (Figure 4.17F). The wing-linker interactions between neighbouring dimers in OrnDC are necessary for dodecamer association. A mutation in the linker domain disrupts the packing of adjacent dimers and prevents dodecamer formation, as shown by the Gly121Tyr mutant structure of OrnDC (Vitali et al, 1999). OrnDC binds GTP, which activates OrnDC homodimers (Oliveira et al, 1997). In contrast, AdiA does not display this type of wing-linker packing (Figure 4.17E), nor does it bind GTP or other nucleotides (Applebaum et al, 1977) and the AdiA homodimers are inactive. It is possible that the AdiA homodimers are inactive because they are unstable. In particular, the protruding wing-domains may have a flexibility that could cause long-range conformational changes affecting the enzyme activity. In the AdiA decamer the wing domains of the homodimers are stabilised by tight wing-wing interactions and a close packing of the complex as a whole (Figure 4.17). That the activity of the AdiA decamer is dependent on its decamer formation may thus be a reflection of the increased stability of the homodimers.

4.4.4.2 AdiA and OrnDC active sites

The AdiA and OrnDC active sites superpose well and the active site residues are conserved (Figure 4.18). Only one region of the AdiA active site, corresponding to the β16-strand and the preceding α20-β16-loop in the AspAT-like small domain, show significant differences to the equivalent region in OrnDC. This is due to a shift in the position of the AdiA β16-strand, which is caused by the preceding α20-β16 loop being two residues shorter than the equivalent loop in OrnDC. Interestingly, the first residue in this β-strand in OrnDC is Glu532, which has been suggested by Momany et al to be a determinant of OrnDC substrate specificity in coordinating the ε-amino group of the substrate ornithine aldimine (Momany et al, 1995). The equivalent residue to OrnDC-Glu532 in AdiA is Thr571. The distance between the PLP cofactor and OrnDC-Glu532 / AdiA-Thr571 is too short to accommodate an arginine aldimine, which could be why OrnDC shows no activity with arginine as a substrate (Guirard & Snell, 1980). In addition, a threonine residue cannot satisfy the positively charged guanidinium group of the arginine as well as a glutamate or aspartate residue. Hence Thr571 is unlikely to determine substrate specificity in AdiA. The difference in residue identity and the relatively poor structural alignment of the OrnDC-Glu532 and AdiA-Thr571 regions, compared to the rest of the AdiA and OrnDC active sites (Figure 4.19), is likely to reflect the different substrate specificities of these decarboxylases.
Figure 4.18. **Comparison of PLP coordination in the AdiA and OrnDC active sites.** The superposition of the AdiA and OrnDC PLP binding domains here illustrates the high degree of conservation of the PLP coordinating residues and the active site between the two decarboxylases. For clarity only the AdiA residues are labeled and hydrogen bonds are shown in blue dotted lines between the PLP cofactor and the AdiA residues only. AdiA molecule A PLP binding domain (orange), AdiA molecule C* PLP binding domain (magenta), PLP cofactor (yellow). OrnDC molecule A PLP binding domain (cyan), OrnDC molecule B PLP binding domain (purple), OrnDC-PLP cofactor (green).

Figure 4.19. **Comparison of OrnDC-Glu532 and AdiA-Thr571 regions.** The OrnDC PLP binding domain is shown in cyan, the AspAT like small domain in green and the PLP cofactor as green sticks. The AdiA PLP binding domain is shown in orange, the AspAT like small domain in blue and the PLP cofactor as yellow sticks.
4.4.5 AdiA and GadB – decarboxylases in *E. coli* acid resistance

The inducible biodegradative glutamate decarboxylase (GadB) is part of the glutamate dependent acid resistance system (AR2) in *E. coli* (Lin et al, 1996). The principle behind the AR2 system is the same as that of the arginine dependent acid resistance system (AR3) with a decarboxylase working in tandem with an antiporter. Both AdiA and GadB belong to the type I fold of PLP-dependent enzymes and as such they both show homology to the PLP-binding domain and the small domain in AspAT. Typically for PLP enzymes both decarboxylases form dimers with the active sites at the dimer interface. Superposing AdiA with GadB (PDB entry 1PMM, (Capitani et al, 2003a) gave an rmsd of 3.6 Å for 330 aligned Cα atoms. This reflects a gross structural homology between the two enzymes while being consistent with the fact that the sequence identity between GadB and AdiA is only 12%. The main difference between AdiA and GadB, and by extension between group III and group II decarboxylases, is the presence of the wing domain and the carboxy-terminal domain in AdiA. None of the group II decarboxylases for which structures have been determined to date (Table 4.1, Figure 4.3) form such large oligomeric assemblies as the AdiA decamer and the OrnDC dodecamer of the group III decarboxylases.

GadB becomes activated and recruited to the membrane at acidic pH (Capitani et al, 2003b). This has been linked to a conformational change in the structure of the GadB N-terminal chain (Capitani et al, 2003a; Gut et al, 2006) from a flexible loop at pH 7.6 (PDB entry 1PMO) to a helix at acidic pH 4.6 (PDB entry 1PMM). In the GadB hexamer (PDB entry 1PMM) this helix forms a triple helix bundle (Figure 4.3F) that is involved in the recruitment of GadB to the membrane when the cellular pH decreases (Capitani et al, 2003a). It was suggested that this recruitment to the membrane would allow GadB to more efficiently counter the influx of protons at acidic pH. In contrast, the N-terminus of the AdiA decamer (the active form of the enzyme at acidic pH) forms a wing domain involved in the stabilization of the decameric structure through extensive interactions. The average *B* value of the wing domain (21.1 Å² in molecule A) compares well with the average *B* value of the molecule as a whole (23.6 Å²). The wing domain is thus unlikely to show conformational flexibility, or to undergo a similar structural rearrangement as proposed for the N-terminal domain of GadB. There are no biochemical studies that would confirm whether or not AdiA is recruited to the membrane at acidic pH. However, from comparison of the AdiA and GadB structures, any AdiA recruitment to the membrane at acidic pH would likely occur in a different manner to that of GadB.
4.4.6 Summary

The structure of AdiA was solved by MIRAS methods and revealed a ca. 800 kDa decamer composed as a pentamer of five homodimers. The abundance of acidic surface residues on the AdiA homodimer mimic acidophilic proteins and is likely to regulate the assembly of functional decamers from inactive homodimers as the pH decreases. Comparison of AdiA with the homologous structure of biodegradative ornithine decarboxylase (OrnDC) from Lactobacillus 30a reveals the structural basis of their different oligomerization states and how this difference is reflected in their enzymatic activities. The present work also identifies potential substrate specificity determinants in the AdiA active site. The AdiA structure represents a first step towards a detailed mechanistic understanding of the arginine-dependent acid resistance system in E. coli.
Thesis summary

This thesis presents the structural studies on three enzymes using X-ray crystallography: the particulate methane monooxygenase (pMMO) from *Methylococcus capsulatus* (Bath), thioredoxin A (BsTrxA) from *Bacillus subtilis* and arginine decarboxylase (AdiA) from *Escherichia coli*.

**Particulate methane monooxygenase (pMMO) from *M. capsulatus* (Bath)**

The work on particulate methane monooxygenase (pMMO) from *M. capsulatus* (Bath) identified a novel crystallisation condition for pMMO. A dataset to 3.5 Å resolution was collected from the best crystal obtained using this condition. The published 2.8 Å resolution structure of pMMO from *M. capsulatus* (Bath) (Lieberman & Rosenzweig, 2005) was used for molecular replacement to obtain a pMMO structure from the 3.5 Å resolution data. Although the crystallisation conditions were different, the crystal unit cells were similar and as expected, the 3.5 Å structure the two structures were essentially identical. One of the aims of this project was to confirm the metal sites and metal content of pMMO, hence data were also collected on the same crystal at the absorption peaks of zinc, copper and iron. The anomalous difference Fourier maps calculated from this data confirms the three metal sites in pMMO from *M. capsulatus* (Bath) as a mononuclear copper site, a dinuclear copper site and a mononuclear zinc site, with the latter likely to be a crystallisation artefact. No new metal binding sites, relevant for enzymatic function, were found in the structure. Unfortunately the active site of pMMO could not be identified in this study, despite efforts involving co-crystallisation of pMMO with suicide substrates acetylene and propargylamine and the derivitisation of pMMO crystals with xenon or krypton gas.

Substantial efforts were made to improve the diffraction limit of the pMMO crystals including changing the detergent, dehydration experiments and improvement of the cryo conditions. A major obstacle in the optimisation of the pMMO crystallisation condition was the variability between different batches of purified protein. As a result of this variability the resolution limit of the pMMO crystals could be either 4 Å or 8 Å for crystals grown in the same crystallisation condition but from different batches of protein sample.

Initial crystallisation trials of the pMMO-MDH complex were unsuccessful, probably as a result of dissociation of the complex in the high salt crystallisation condition. However, a
potential crystallisation condition for the pMMO-MDH complex was obtained in the absence of salt. This condition will be pursued with the aim of solving the pMMO-MDH structure and to identify the physiological metal bound in the third site of *M. capsulatus* (Bath) pMMO.

**Thioredoxin A from *B. subtilis* (BsTrxA)**

In mixed disulfide experiments aimed at identifying substrates of thioredoxin A (BsTrxA) in *B. subtilis*, it was found that the active site mutant C32S formed homodimers. This homodimer form via a disulfide bond between active site cysteines, as revealed by the 1.5 Å resolution structure of the C32S BsTrxA homodimer presented here. The C32S mutation does not cause any structural rearrangement of the thioredoxin backbone, indeed the only difference compared to the wild type BsTrxA seem to be a change in the rotamer of the Ser32 side chain caused by the polar hydroxyl chain forming hydrogen bonds. This manner of dimerisation differs substantially from previously reported thioredoxin homodimers. It was proposed that the C32S BsTrxA homodimer mimics the mixed disulfide reaction intermediate between thioredoxin and its substrate. At the time this project was initiated no thioredoxin substrate protein complexes were available and the aim of studying the C32S BsTrxA homodimer was therefore to analyse the dimer interface in order to infer how BsTrxA binds substrate proteins. While the work described in this thesis was in progress two thioredoxin substrate protein complexes became available (Li et al, 2007; Maeda et al, 2006). From a comparison with the structures of these complexes and thioredoxin substrate peptide complexes (Qin et al, 1995; Qin et al, 1996) it can be concluded that the binding mode of the C32S BsTrxA homodimer resembles that of the thioredoxin substrate complexes. It can thus be regarded as a representation of a mixed disulfide reaction intermediate. The comparison between the C32S BsTrxA homodimer structure with thioredoxin substrate complexes, illustrates the common factors and variability in thioredoxin-substrate binding modes that enables thioredoxin to bind to and react with a variety of different substrates.

**Arginine decarboxylase (AdiA) from *E. coli***

The structure of the arginine decarboxylase (AdiA) from *E. coli* was determined by MIRAS methods and revealed a ~800 kDa decamer comprised of a pentamer of homodimers. AdiA is part of *E. coli* acid resistance and becomes active as the pH in the cell decreases. AdiA becomes active upon assembly of the decamer from inactive homodimers, which occurs as the pH decreases. The surface of an AdiA homodimer has an abundance of acidic surface residues. At neutral pH the negatively charged acidic surface residues keep the homodimers
apart. As the proton concentration increases with decreasing pH the negatively charged acidic surface residues are neutralized allowing the formation of the active decamers.

In contrast to the homologous enzyme ornithine decarboxylase (OrnDC) from *Lactobacillus 30a*, which is active both as the homodimer and in the larger oligomeric form as a dodecamer, AdiA is only active in its decameric form while the dimeric form is inactive. It is possible that the AdiA homodimers are inactive because they are unstable. In particular, the protruding wing-domains may have a flexibility causing long-range conformational changes affecting the enzyme activity. In the AdiA decamer the wing domains of the homodimers are stabilised by tight wing-wing interactions and a close packing of the complex as a whole. The activity of the decamer may thus be a reflection of the increased stability of the homodimers.

The AdiA structure shows the open conformation of the enzyme with a cleft leading into the active site. By modelling a product molecule (agmatine aldimine) into the AdiA active site two residues Glu739 and Asp202, were identified as potential substrate specificity determinants. In parallel with other PLP decarboxylases, the dicarboxylic groups of Glu739 and Asp202 would coordinate the guanidinium group of the arginine/agmatine substrate/product in the closed form of the enzyme. It would thus be possible that the maximal activity of AdiA at pH 5.2 reflects the pH at which Glu739 and Asp202 retain their negative charge for coordination of the substrate, while the optimum proportion of acidic surface residues are protonated to allow the assembly of functional decamers.
Acknowledgements

First of all I would like to thank my supervisor So Iwata for giving me this wonderful opportunity. Your unwavering support and encouragement these years have been truly amazing. Thank you! My heartfelt thanks also go to my supervisor Liz Carpenter for helping me out whenever difficulties arose. There is none better to have at your side at 3 am when you are driving the wrong way in Italy. Great thanks are also due my supervisor Bernadette Byrne, in particular for the occasional life-saving talk in her office. A very special thank you goes to Megan Maher for showing me the practical side of crystallography. Your skill and support have been invaluable for the work on this thesis. Thank you all for sharing your knowledge, teaching-skills and passion for science with me.

There have been many past and present members of the MPC group contributing to this work and I would like to extend my thanks to all of you. Tina Iverson, for introducing me to the field and getting me hooked on protein crystallography during my masters project. Mika Jormakka, for good advice and a healthy Swedish perspective on things. Jonathan Ruprecht, for joining in the crystallisation angst and synchrotron-fun. David Drew, for sharing his fantasy books as readily as his vast knowledge on membrane proteins. Simon Newstead and Momi Iwata for helping hands. Simone Weyand and Mikio Tanabe, for pep talks, friendship and hello kitty frustrations. Good advice and good company are always welcome, for which I thank Osman Mirza, Kuakarun ‘Ying’ Krusong, Rob Horsefield, Manami Kanazawa, Rie Tanaka, Rohini Rana, Nien-Yen Hu and all members of the MPC group.

I would like to express my gratitude especially to my parents Laila and Leif Andréll for their invaluable support and trust. My thanks also go to my sister Paulin Andréll and my brother Hampus Andréll for believing in me and my work.

Last but never least I want to thank my husband Alexander, who means the world to me. None of this would have been possible without you. Thank you for being there.
References


Boeker EA (1978) Arginine decarboxylase from *Escherichia coli* B: mechanism of dissociation from the decamer to the dimer. *Biochemistry* 17(2): 258-263


Lieberman RL, Shrestha DB, Doan PE, Hoffman BM, Stemmler TL, Rosenzweig AC (2003) Purified particulate methane monooxygenase from Methylococcus capsulatus (Bath) is a dimer with both mononuclear copper and a copper-containing cluster. Proc Natl Acad Sci U S A 100(7): 3820-3825


