Molecular insights into cobalamin-dependent enzyme systems from organohalide-respiring bacteria

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy (PhD)
in the Faculty of Life Sciences

2013

Hanno Sjuts
## Contents

1 Abstract and preface
   1.1 Abstract .................................................. 12
   1.2 Declaration ............................................... 13
   1.3 Copyright statements .................................... 13
   1.4 Preface to the alternative format ....................... 14

2 Introduction
   2.1 Halogenated organic molecules in the environment ........ 18
   2.2 The chlorine cycle ........................................ 19
   2.3 Organohalide respiration: a bacterial pathway ............. 19
      2.3.1 Energy generation during organohalide respiration .... 21
      2.3.2 Terminology and chemistry of organohalide respiration . 24
      2.3.3 Thermodynamics of reductive dehalogenation ........... 26
   2.4 Reductive dehalogenases: key enzymes in organohalide respiration ... 28
      2.4.1 Cofactors of reductive dehalogenases .................. 30
   2.5 Gene arrangements and molecular regulation of reductive dehalogenases . . 33
      2.5.1 RdhB: a membrane anchor protein for RdhA .......... 34
      2.5.2 RdhT: a specific chaperone for RdhA .............. 34
      2.5.3 Transcriptional control of organohalide respiration .... 35
   2.6 Other cobalamin-dependent enzyme systems .................. 37
   2.7 The three-component and cobalamin-dependent
      O-demethylase enzyme system .............................. 38
      2.7.1 Sequence alignments between O-demethylase proteins .... 39
      2.7.2 Metabolic links between O-demethylation and organohalide respiration . 40
   2.8 Aims and outline of this thesis ............................ 42
   2.9 References ............................................... 43

3 Methods
   3.1 General principle of X-ray crystallography .................. 52
   3.2 Crystallisation of proteins ................................ 53
## Contents

- **3.2.1** Phase diagrams ................................................................. 53
- **3.2.2** Seeding techniques ........................................................... 55
- **3.3** Physical properties of crystals ............................................... 56
- **3.4** Diffraction, Bragg’s law and crystal lattices ............................. 57
- **3.5** Reciprocal space and data collection ....................................... 61
- **3.6** Solving the phase problem ...................................................... 63
  - **3.6.1** Molecular replacement ...................................................... 64
  - **3.6.2** Isomorphous replacement .................................................. 66
  - **3.6.3** Anomalous scattering ....................................................... 69
  - **3.6.4** MAD and SAD ................................................................. 69
- **3.7** Model building, structure refinement and validation ................. 70
- **3.8** References .............................................................................. 71

## 4 PceA

- **4.1** Abstract ................................................................................. 76
- **4.2** Introduction ........................................................................... 77
- **4.3** Materials and methods ......................................................... 78
  - **4.3.1** Construct design .............................................................. 78
  - **4.3.2** Small-scale solubility screen ............................................. 78
  - **4.3.3** Large-scale expression of the PceA from the pCOLD TF system . 78
  - **4.3.4** Anaerobic protein purification and cleavage of trigger factor with HRV 3C protease .................................................. 79
  - **4.3.5** Reconstitution of 4Fe-4S clusters and cobalamin uptake ........ 79
  - **4.3.6** UV-Vis and EPR measurements .......................................... 80
- **4.4** Results .................................................................................... 80
  - **4.4.1** Validation of a cloning strategy aimed at expression of soluble holo-PceA .................................................. 80
  - **4.4.2** Purification, Fe-S cluster reconstitution and UV-Vis/EPR analysis of soluble TF-PceA fusion protein .............................. 81
  - **4.4.3** Separation of PceA from TF and reconstitution of holo-PceA ... 83
- **4.5** Discussion .............................................................................. 84
- **4.6** Conclusions ........................................................................... 86
- **4.7** Acknowledgements .................................................................. 86
- **4.8** References .............................................................................. 87

## 5 PceT

- **5.1** Abstract ................................................................................. 90
- **5.2** Introduction ........................................................................... 91
- **5.3** Materials and methods ......................................................... 92
  - **5.3.1** Cloning of expression plasmids ......................................... 92
## Cobalamin-dependent enzyme systems

### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.2</td>
<td>Expression and purification of PceT</td>
<td>95</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Expression and purification of TAT fusion proteins</td>
<td>96</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Expression and purification of the PceT-PceA fusion protein</td>
<td>96</td>
</tr>
<tr>
<td>5.3.5</td>
<td>Determination of PceT dispersity by multi-angle light scattering</td>
<td>96</td>
</tr>
<tr>
<td>5.3.6</td>
<td>Crystallisation and data collection of PceT</td>
<td>97</td>
</tr>
<tr>
<td>5.3.7</td>
<td>Isothermal titration calorimetry (ITC)</td>
<td>97</td>
</tr>
<tr>
<td>5.3.8</td>
<td><em>In vitro</em> testing of prolyl-cis-trans isomerase activity of PceT</td>
<td>98</td>
</tr>
<tr>
<td>5.4</td>
<td>Results and discussion</td>
<td>98</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Purification and crystallisation of PceT</td>
<td>98</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Purification of MBP-TAT Fusion Proteins</td>
<td>100</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Biophysical interaction between PceT and the TAT signal sequence</td>
<td>101</td>
</tr>
<tr>
<td>5.4.4</td>
<td>Expression and purification of the PceT-PceA fusion protein and co-expression of PceT::PceA</td>
<td>102</td>
</tr>
<tr>
<td>5.4.5</td>
<td>Functional characterisation of PceT</td>
<td>103</td>
</tr>
<tr>
<td>5.5</td>
<td>Conclusions</td>
<td>104</td>
</tr>
<tr>
<td>5.6</td>
<td>References</td>
<td>105</td>
</tr>
</tbody>
</table>

### 6 CobDH

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Abstract</td>
<td>108</td>
</tr>
<tr>
<td>6.2</td>
<td>Introduction</td>
<td>109</td>
</tr>
<tr>
<td>6.3</td>
<td>Materials and methods</td>
<td>110</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Cloning, expression and purification</td>
<td>110</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Crystallization and cryoprotection</td>
<td>111</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Data collection and processing</td>
<td>112</td>
</tr>
<tr>
<td>6.3.4</td>
<td>Structure determination</td>
<td>112</td>
</tr>
<tr>
<td>6.3.5</td>
<td>UV-Vis spectroscopy</td>
<td>112</td>
</tr>
<tr>
<td>6.3.6</td>
<td>EPR spectroscopy</td>
<td>113</td>
</tr>
<tr>
<td>6.4</td>
<td>Results and discussion</td>
<td>113</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Crystal structure of CobDH</td>
<td>113</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Spectroscopic analysis</td>
<td>118</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Comparison of CobDH with other methyltransferases</td>
<td>120</td>
</tr>
<tr>
<td>6.5</td>
<td>Conclusions</td>
<td>122</td>
</tr>
<tr>
<td>6.6</td>
<td>References</td>
<td>122</td>
</tr>
</tbody>
</table>

### 7 Methyltransferases

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Abstract</td>
<td>126</td>
</tr>
<tr>
<td>7.2</td>
<td>Introduction</td>
<td>127</td>
</tr>
<tr>
<td>7.3</td>
<td>Materials and methods</td>
<td>129</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Cloning and overexpression of MT1DH and MT2DH</td>
<td>129</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Crystallisation</td>
<td>130</td>
</tr>
</tbody>
</table>
7.3.3 Data collection .............................................. 131
7.3.4 Structure determination ................................... 131
7.3.5 MT2DH methyltransferase assay .......................... 132
7.3.6 Surface plasmon resonance (SPR) binding measurements .................. 133
7.4 Results and discussion ............................... 133
  7.4.1 Methyl group transfer from methylcobalamin to THF is catalysed by MT2DH ........................................... 133
  7.4.2 Crystal structure determination ........................... 134
  7.4.3 Structure comparisons .................................. 135
  7.4.4 Cofactor binding site of MT2DH .......................... 136
  7.4.5 Proposed reaction mechanism ............................. 140
  7.4.6 \textit{In vitro} stability of MT1DH .......................... 141
  7.4.7 Binding studies between MT1DH and aromatic methyl ether substrates ............................................ 141
7.5 Conclusions ..................................................... 142
7.6 References ...................................................... 143
7.7 Supplementary figures and tables ................................ 146

8 General discussion ............................................. 149
  8.1 Reductive dehalogenases involved in organohalide respiration ................. 150
    8.1.1 Large scale production of reductive dehalogenases .................... 150
    8.1.2 Catalytic activity and diversity of reductive dehalogenases ............ 151
  8.2 The \textit{O}-demethylase system .............................. 153
    8.2.1 Putative mechanism of methyl transfer reactions .................... 153
    8.2.2 Protein-protein interactions involved in \textit{O}-demethylases ............ 154
  8.3 References ...................................................... 155

9 Appendix .......................................................... 157
  9.1 Abbreviations ................................................. 158
  9.2 Amino acids ................................................... 162
  9.3 Photographic crystal images ................................... 164
  9.4 Acknowledgements ............................................. 167
  9.5 Curriculum Vitae, Hanno Sjuts ................................ 168

Word count: 49,218
List of Figures

2.1 The chlorine cycle ........................................... 20
2.2 Phylogenetic diversity of organohalide-respiring bacteria .......... 22
2.3 Energy generation in organohalide respiration ....................... 24
2.4 Mechanisms of dehalogenation ................................ 25
2.5 Organohalide respiration and O-demethylation ...................... 28
2.6 Cofactors of RdhAs ............................................ 33
2.7 The pceABCT operon .......................................... 35
2.8 O-demethylation pathway ...................................... 38
2.9 Proteins involved in O-demethylation ............................. 40
2.10 Alignment of O-demethylase components ........................... 41

3.1 Deposited structures in the PDB ................................. 52
3.2 The phase diagram ............................................. 54
3.3 Unit cell and crystal properties .................................. 56
3.4 Illustration of Bragg’s law ..................................... 59
3.5 Lattice planes and Miller indices ................................. 60
3.6 Ewald’s sphere .................................................. 62
3.7 Patterson function .............................................. 66
3.8 Argand diagram ................................................ 68
3.9 Harker construction ............................................ 68
3.10 Anomalous scattering .......................................... 70

4.1 Purification of TF-PceA ........................................... 82
4.2 EPR analysis of Fe-S cluster in PceA ............................... 83
4.3 SDS-PAGE of TF-PceA .......................................... 84
4.4 UV-Vis and EPR analysis of PceA ................................ 85

5.1 E. coli trigger factor ............................................ 91
5.2 Sequence alignment of trigger factor proteins ....................... 93
5.3 PceT purification and crystallisation ................................ 99
5.4 SDS-PAGE of MBP-TAT fusion protein .......................... 101
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>ITC between PceT and TAT signal sequence</td>
<td>102</td>
</tr>
<tr>
<td>5.6</td>
<td>SDS-PAGE of PceT-PceA fusion protein</td>
<td>103</td>
</tr>
<tr>
<td>5.7</td>
<td>Enzymatic activity of PceT</td>
<td>104</td>
</tr>
<tr>
<td>6.1</td>
<td>SDS-PAGE of purified CobDH</td>
<td>113</td>
</tr>
<tr>
<td>6.2</td>
<td>Sequence alignment of CobDH</td>
<td>115</td>
</tr>
<tr>
<td>6.3</td>
<td>Overall structure of CobDH</td>
<td>117</td>
</tr>
<tr>
<td>6.4</td>
<td>Cobalamin-binding site in CobDH</td>
<td>119</td>
</tr>
<tr>
<td>6.5</td>
<td>UV-Vis and EPR analyses of CobDH</td>
<td>120</td>
</tr>
<tr>
<td>7.1</td>
<td>Proteins involved in O-demethylation</td>
<td>128</td>
</tr>
<tr>
<td>7.2</td>
<td>Methyltransferase assay</td>
<td>134</td>
</tr>
<tr>
<td>7.3</td>
<td>Overall MT2DH structure</td>
<td>135</td>
</tr>
<tr>
<td>7.4</td>
<td>Structure alignment</td>
<td>137</td>
</tr>
<tr>
<td>7.5</td>
<td>MTHF binding site</td>
<td>139</td>
</tr>
<tr>
<td>7.6</td>
<td>Proposed mechanism</td>
<td>140</td>
</tr>
<tr>
<td>7.7</td>
<td>SPR measurement</td>
<td>142</td>
</tr>
<tr>
<td>S1</td>
<td>SDS-PAGE</td>
<td>146</td>
</tr>
<tr>
<td>S2</td>
<td>Backbone density</td>
<td>146</td>
</tr>
<tr>
<td>S3</td>
<td>Conserved Asn199</td>
<td>146</td>
</tr>
<tr>
<td>9.1</td>
<td>Amino acid list I</td>
<td>162</td>
</tr>
<tr>
<td>9.2</td>
<td>Amino acid list II</td>
<td>163</td>
</tr>
<tr>
<td>9.3</td>
<td>PceT crystals</td>
<td>164</td>
</tr>
<tr>
<td>9.4</td>
<td>CobDH crystals</td>
<td>165</td>
</tr>
<tr>
<td>9.5</td>
<td>MT2DH crystals</td>
<td>166</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Comparison of organohalide-respiring bacteria</td>
<td>23</td>
</tr>
<tr>
<td>2.2</td>
<td>Redox potentials of organohalides</td>
<td>27</td>
</tr>
<tr>
<td>2.3</td>
<td>Comparison of characterised reductive dehalogenases</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>The 65 space groups relevant in macromolecular crystallography</td>
<td>58</td>
</tr>
<tr>
<td>4.1</td>
<td>Expression plasmid list</td>
<td>81</td>
</tr>
<tr>
<td>4.2</td>
<td>TF-PceA expression and purification</td>
<td>82</td>
</tr>
<tr>
<td>5.1</td>
<td>Plasmid list</td>
<td>93</td>
</tr>
<tr>
<td>5.2</td>
<td>Crystallographic data collection statistics</td>
<td>100</td>
</tr>
<tr>
<td>6.1</td>
<td>Crystallographic data and refinement statistics</td>
<td>116</td>
</tr>
<tr>
<td>S1</td>
<td>Crystallographic data and refinement statistics</td>
<td>147</td>
</tr>
<tr>
<td>9.1</td>
<td>Abbreviations</td>
<td>158</td>
</tr>
</tbody>
</table>
Chapter 1

Thesis abstract and preface to the alternative thesis format
Chapter 1. Abstract and preface Cobalamin-dependent enzyme systems

1.1 Abstract

Molecular insights into cobalamin-dependent enzyme systems from organohalide-respiring bacteria.

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy (PhD) in the Faculty of Life Sciences by Hanno Sjuts.

The evolution of microorganisms over millions of years has led to an impressive adaptability regarding the utilisation of different environmental conditions. The identification of bacterial species with the fascinating features to use cobalamin-dependent metalloenzymes to (i) extract energy from halogenated organic compounds (organohalides) and (ii) transfer methyl groups from lignin breakdowns into central carbon pathways, are examples for this adaptability. The biochemical study of these two cobalamin-dependent enzymes is the topic of this PhD project.

For the extraction of growth energy, organohalides serve as terminal electron acceptors and are reductively dehalogenated in a respiratory manner termed organohalide respiration. Reductive dehalogenases, the key enzymes in organohalide respiration, catalyse the chemical cleavage between the halogen substituent and the carbon moiety. They use cobalamin and two Fe-S clusters as cofactors and constitute a new and distinct class of cobalamin-dependent enzymes. Their three-dimensional structure and the mechanism of catalysis are unknown, because their hydrophobicity and oxygen sensitivity have hampered their biochemical investigation. Here, a novel purification technology in Escherichia coli for the reductive dehalogenase PceA from Dehalobacter restrictus has been developed, accompanied by methods that allow the in vitro reconstitution of PceA with both cofactors, cobalamin and Fe-S clusters. It has been demonstrated that the soluble expression of PceA is dependent on the covalent fusion of the enzyme to a trigger factor chaperone. Based on these findings, the PceA specific trigger factor PceT has been studied biochemically, resulting in its successful crystallisation. The established protocols for PceA and PceT are transferable to other members of their respective families, which will therefore allow detailed studies of reductive dehalogenases and their associated chaperones in the future.

In addition to reductive dehalogenases, organohalide respiring bacteria contain another cobalamin-dependent enzyme system, termed O-demethylase, which is involved in the carbon metabolism of different anaerobic bacteria. O-demethylases are three-component enzyme systems that transfer methyl groups from aromatic methyl ethers to tetrahydrofolate via methylcobalamin intermediates. The different cofactors (substrate, cobalamin and tetrahydrofolate), bind to either of the three individual proteins involved in O-demethylation. It has been speculated that the same or similar halogenated aromatic molecules are substrates for both organohalide respiration and O-demethylation in the same bacteria. In order to test this proposal, a O-demethylase from Desulfotobacterium hafniense DCB-2 has been studied using X-ray crystallography and biochemistry. As a result, the first crystal structures of the cobalamin-binding protein in complex with cobalamin, and of the methyl acceptor protein in complex with substrate (tetrahydrofolate) and product (methyltetrahydrofolate) from a O-demethylase have been solved to resolutions of 1.5 Å, 1.8 Å and 1.6 Å, respectively. The crystal structures, in combination with spectroscopic and biophysical analyses, have led to a proposed mechanism for the catalysed methyl transfer reaction from methylcobalamin to tetrahydrofolate.
1.2 Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

1.3 Copyright statements

i) The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii) Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii) The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv) Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in The University IP Policy, in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations and in The University’s policy on presentation of Theses.
1.4 Preface to the alternative format

The thesis is being presented in the alternative format in accordance with the rules and regulations of the University of Manchester. The results chapters presented herein are in manuscript form in the style suitable for their intended journal of submission. However, elements have been reformatted (with permission from the publishers) to ensure these form a cohesive body of work.

Manuscripts included as results chapters are:

**Chapter 4: Heterologous expression, purification and cofactor reconstitution of the reductive dehalogenase PceA from *Dehalobacter restrictus***

*Hanno Sjuts*, Karl Fisher, Mark S. Dunstan, Stephen E. Rigby and David Leys


Author contributions:

*Hanno Sjuts*, Karl Fisher, Mark S. Dunstan and David Leys designed the research; *Hanno Sjuts* performed molecular biology experiments, protein overexpression trials, protein purifications, limited proteolysis, Fe-S reconstitutions, UV-Vis spectroscopy and sample preparation for EPR; Karl Fisher and Stephen E. Rigby recorded EPR spectra; all authors analysed data; *Hanno Sjuts* and David Leys wrote the paper.

**Chapter 5: Structural and functional characterisation of the trigger factor protein PceT from *Dehalobacter restrictus***

*Hanno Sjuts*, Mark S. Dunstan and David Leys

This manuscript is currently not submitted to any scientific journal.

Author contributions:

*Hanno Sjuts* and David Leys designed the research; *Hanno Sjuts* performed molecular biology experiments, protein overexpression trials, protein purification, crystallisation and biochemical experiments; all authors recorded and analysed crystallographic data; *Hanno Sjuts* wrote the manuscript.

**Chapter 6: Structure of the cobalamin-binding protein of a putative O-demethylase from *Desulfitobacterium hafniense* DCB-2***

*Hanno Sjuts*, Karl Fisher, Mark S. Dunstan and David Leys


Author contributions:

*Hanno Sjuts* and David Leys designed research; *Hanno Sjuts* performed molecular biology experiments, protein expression, protein purification, crystallisation, UV-Vis
spectroscopy and EPR sample preparation; Hanno Sjuts, Mark S. Dunstan and David Leys performed synchrotron data-collection; Hanno Sjuts solved the crystal structure with help from Mark S. Dunstan; Karl Fisher recorded EPR spectra; all author’s analysed data; Hanno Sjuts and David Leys wrote the paper.

Chapter 7: Crystal structures of the methyltransferase protein involved in $O$-demethylation that catalyses the methyl group transfer from methylcobalamin to tetrahydrofolate

Manuscript in preparation for submission to *PLoS ONE*.

**Author contributions:**

Hanno Sjuts and David Leys designed research; Hanno Sjuts performed molecular biology experiments, protein purification, protein crystallisation, UV-Vis spectroscopy and SPR measurements; Hanno Sjuts, Mark S. Dunstan and David Leys performed synchrotron data collection; Hanno Sjuts solved the crystal structures with aid from Mark S. Dunstan and David Leys; Karl Fisher grew *D. hafniense* bacteria; all author’s analysed data; Hanno Sjuts and David Leys are writing the paper.
Chapter 2

Introduction to reductive dehalogenases and \( O \)-demethylases: cobalamin-dependent enzyme systems in organohalide-respiring bacteria
2.1 Halogenated organic molecules in the environment

Both the number and diversity of halogenated organic compounds (organohalides) in the environment is astonishing. At least 3,800 different molecules, containing chloride, bromide, iodide or fluoride atoms, have been identified in nature, with those containing chloride or bromide being the most abundant (1). Due to current understanding the vast majority of these compounds are of natural and not of synthetic origin (2). In some cases, e.g. vinyl chloride (VC), the organohalide is produced both industrially and naturally (3). Naturally produced organohalides can be of either biogenic or geogenic origin. Biogenic sources are abundant, with the saline marine environment being the largest contributor (2). Organisms capable of producing these molecules include bacteria, fungi, lichens, marine sponges, worms, insects, and mammals (4). Geogenic halogenated compounds originate in volcanoes and forest fires when inorganic chloride is incorporated into organic molecules (5). However, the synthetic fraction of organohalides is increasing due to the widespread use of halogenated compounds in the industry, which started only during the 20th century. Chloroethanes, chloroethenes, chlorobenzenes and chlorophenols are produced in the largest quantities and they are used as solvents, in pharmaceutical manufacture, for chemical synthesis, etc. (6; 7). Other significant synthetic sources of halogenated organic compounds include pesticides (for both agricultural and domestic use) where their improper application results in polluted soil and groundwater, and polyvinyl chlorides (PVC’s) in the plastic industry. The use of PVC in plastic production has tremendously increased since the 1960s, giving rise to the unintentional release of other chlorinated compounds. The uncontrolled release of these compounds is a serious issue because of the negative impact they have on all life forms (8). In this respect, the Stockholm convention named twelve persistent organic pollutants (POP’s) that are all halogenated aromatic compounds and include DDT and polychlorinated biphenyls. These twelve POP’s have been termed the dirty-dozen and as a result of the convention the use of these compounds is now very restricted in many countries (9). The main chemical reason for the toxicity of these compounds is that the introduced halogen moiety in the organohalide increases its lipid solubility. Consequently its bioavailability decreases, whereas at the same time the likelihood of membrane / fat tissue incorporation increases (10). If released to the air, organohalides become destructive to the ozone layer (11). Additionally, the toxicity rises with the degree of halogenation. The low bioavailability of organohalides makes them recalcitrant for biodegradation, which is reflected in a long half-life (12). Interestingly many different bacteria evolved with the biochemical ability to dehalogenate these recalcitrant molecules. Under aerobic conditions many organohalides can be dehalogenated by the Pseudomonas genus (13), however, those with multiple halogen substituents are less sensitive to aerobic dehalogenation, as the high electronegativity of the halogen atoms prevents nucleophilic attack on the carbon backbone (14; 15). Hence, multiple-halogenated organic compounds are
more prone to anaerobic dehalogenation. In addition, most organohalides are released into anoxic environments like submerged soil, wastewater sludges, or groundwater, which might account for the evolutionary selection of a broad variety of anaerobic bacteria with the fascinating capability of coupling dehalogenation to energy conservation, a process termed organohalide respiration (16; 17). Before these processes are discussed in more detail, the next section focuses on the chlorine cycle as it plays an important role in organohalide respiration.

2.2 The chlorine cycle

The large number and diversity of naturally occurring or geologically produced organohalides indicates that the role of chlorine in the environment may be more important than previously thought. Several organisms produce enzymes which catalyse either the intra- or extracellular synthesis of chlorinated organic compounds. This synthesis can be either specific or unspecific. Thus halogens in their inorganic, ionic form are incorporated into organic compounds, which are then called organohalides, as shown in figure 2.1. Their biological functions are diverse and include e.g. chemical defence against predators, inhibition of biological collusion and as pheromones (10). As the sea is the richest source of chlorine, it is not surprising that many marine organisms are able to synthesise organohalides (10). In order to complete the chlorine cycle the covalent bond between the carbon and the halogen atom must be cleaved and the chlorine ion must be released into the environment. The main contribution to this process is achieved by microbial dehalogenation during which organohalides can either be partially or completely dehalogenated. Furthermore specialised bacteria have evolved the ability to couple this biochemical process to their energy acquisition and are therefore termed organohalide-respiring bacteria. The next chapter focuses on the molecular and biochemical description of organohalide respiration in these bacterial species.

2.3 Organohalide respiration: a bacterial pathway

Early studies by Martin Alexander have indicated that most naturally occurring organohalides are subject to bacterial dehalogenation (18). Consequently, the diversity of bacterial species contributing to this capacity is enormous. They are found in both Gram positive as well as Gram negative groups, which has been summarised by Häggblom et al. (7). Organohalide-respiring bacteria are most likely to have evolved because of the natural and not synthetic release of halogenated compounds. This is supported by the fact that during the first two billion years of life on earth, oxygen was not present. Hence, bacteria had to develop anaerobic pathways with electron acceptors other than oxygen. Organohalides are excellent electron acceptors and it is therefore not surprising that many bacterial species have been found which anaerobically dehalogenate a variety of
Figure 2.1: Overview of the geochemical chlorine cycle. Chloride ions (Cl\(^-\)) are mainly present in water and soil, where they are incorporated by microorganisms into organic molecules (R), thus forming organohalides (R–Cl). The formation of organohalides also takes place by organisms living above the surfaces of soil and water, whereas the reverse process of dehalogenation is exclusively catalysed by microorganisms in the soil and water. During the last century industrial processes led to an increased formation of organohalides and therefore induced the pollution of many soil and water environments.
organohalides (19). Figure 2.2 shows the phylogenetic diversity of organohalide-respiring bacteria. From a chemical point of view, organohalides are generally classified into aromatic and aliphatic compounds, respectively. Substrate specificity of the respective species in figure 2.2 is indicated by asterisks. Interestingly, the different substrate classes can be utilised by different members of the respective bacterial groups, indicating that the evolution of these biochemical abilities were derived upon the respective organohalide in the environment (19). The first isolated and characterised bacterial species that can anaerobically couple dehalogenation to adenosine triphosphate (ATP) generation was the proteobacterium Desulfomonile tiedjei, utilising 3-chlorobenzoate as an electron acceptor and H₂ as an electron donor. In subsequent years many pure cultures with similar capabilities have been isolated and studied. Table 2.1 gives an overview of which characterised organohalide-respiring bacteria can utilise which organohalides. The Desulfitobacterium genus was coined in the year 1994 (20). It includes only strictly anaerobic bacteria, which have been successfully isolated from soil, wastewater and freshwater sludges. Strains capable of reductive dehalogenation include Desulfitobacterium dehalogenans, hafniense, chlororespirans and frappieri. Different species within this genus use a variety of electron donors including H₂, formate, lactate and pyruvate and both aliphatic as well as aromatic organohalides are dehalogenated. The Desulfitobacterium genus with respect to its role in bioremediation has been reviewed in 2006 by Villemur et al. (21). Within the genus several species contain menaquinones, cytochromes b and c. A rather limited set of electron donors and acceptors can be utilised by the species Dehalobacter restrictus. It can only use H₂ as electron donor and tetrachlorethene (PCE) or trichlorethene (TCE) as electron acceptors and contains both menaquinones and cytochrome b (22). The species Dehalococcoides mccartyi exhibits a unique feature whereby it completely dehalogenates PCE to ethene, making it a highly efficient candidate for the detoxification of polluted environments (23). This complete dehalogenation of PCE to ethene has also been observed in mixed-culture studies, in which several bacterial species contribute to the process (24). In summary, many different bacterial species are able to anaerobically dehalogenate a variety of organohalides.

2.3.1 Energy generation during organohalide respiration

The general electron and proton transfer reactions necessary for organohalide respiration in a bacterial cell are illustrated in figure 2.3. Membrane associated hydrogenases generate the electrons by the splitting of H₂ into two protons and two electrons which are ultimately transferred to the organohalide. It has been shown that the flow of the electrons from the donor to the acceptor is used to generate ATP by a chemiosmotic gradient (37). It is not known on the molecular level how the electrons are transported through the membrane from the donor to the acceptor. Also it is unclear how the protons are pumped out of the bacterium, assuming the hydrogenase, that generates them,
Figure 2.2: Phylogenetic relationship of selected bacteria as a result of 16S rRNA gene sequence comparison. All bacteria are capable of anaerobic dehalogenation of specific organohalides. Asterisks indicate, whether aromatic (∗), aliphatic (∗∗), or both (∗∗∗) substrate groups are used. Sequence alignment was performed using ClustalX and the neighbour joining algorithm (25). The tree was subsequently calculated by the average distance method with the Jalview©2.6.1 software. The scale bar represents 1% sequence diversity between the aligned 16S rRNA sequences.
Table 2.1: Overview of selected organohalide-respiring bacteria from pure cultures. Listed are the respective organohalides which are prone to reductive dehalogenation. Both electron acceptor and donor lists are not complete. The accession numbers for the 16S rRNA sequences are from the NCBI database, which is also used for the sequence alignment. All species are strictly anaerobic, with the exception of *Sphingobium chlorophenolicum* (strictly aerobic) and *Shewanella sediminis* HAW-EB3 (facultative anaerobic). Abbreviations: CHPA = 3-chloro-4-hydroxyphenylacetate; Cl-OHPA = 3-chloro-4-hydroxyphenylacetate; DCE= dichloroethene; DCP = dichlorophenol; n. d. = not determined; PCE = tetrachloroethene; PCP = polychlorophenol; TCE = trichloroethene, TCP = trichlorophenol.

<table>
<thead>
<tr>
<th>Species</th>
<th>Halogenated electron acceptor</th>
<th>Electron donor</th>
<th>16S rDNA sequence</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Desulfomonile tiedjei</em></td>
<td>3-chlorobenzoate, PCE, TCE, PCP</td>
<td>H₂, formate, pyruvate</td>
<td>AM086646</td>
<td>(26)</td>
</tr>
<tr>
<td><em>Dehalobacter restrictus</em></td>
<td>PCE, TCE</td>
<td>H₂</td>
<td>U84497</td>
<td>(22)</td>
</tr>
<tr>
<td><em>Desulfotobacterium dehalogenans</em></td>
<td>DCP, TCP, Cl-OHPA, PCE</td>
<td>H₂, formate, pyruvate, lactate</td>
<td>L28946</td>
<td>(20)</td>
</tr>
<tr>
<td><em>Desulfotobacterium chlororespirans</em></td>
<td>DCP, TCP, CHPA</td>
<td>H₂, formate, pyruvate</td>
<td>U68528</td>
<td>(27)</td>
</tr>
<tr>
<td><em>Desulfotobacterium hafniense</em> DCB-2</td>
<td>DCP, TCP, CHPA</td>
<td>formate, pyruvate</td>
<td>X94975</td>
<td>(28)</td>
</tr>
<tr>
<td><em>Desulfotobacterium hafniense</em> PCE-1</td>
<td>DCPs, TCPs, PCP</td>
<td>pyruvate</td>
<td>U40078</td>
<td>(29)</td>
</tr>
<tr>
<td><em>Desulfotobacterium hafniense</em> PCE-S</td>
<td>PCE, TCE, TCP, PCP</td>
<td>pyruvate</td>
<td>AJ512772</td>
<td>(30)</td>
</tr>
<tr>
<td><em>Desulfotobacterium hafniense</em> Y-51</td>
<td>PCE, TCE</td>
<td>formate, lactate, pyruvate</td>
<td>AB049340</td>
<td>(31)</td>
</tr>
<tr>
<td><em>Desulfotobacterium sp.</em> PCE-1</td>
<td>PCE, CHPA</td>
<td>formate, pyruvate</td>
<td>X81032</td>
<td>(32)</td>
</tr>
<tr>
<td><em>Dehalococcoides mccartyi</em> strain 195</td>
<td>PCE, TCE, DCE, VC</td>
<td>H₂</td>
<td>DQ834686</td>
<td>(23)</td>
</tr>
<tr>
<td><em>Clostridium bifermentans</em></td>
<td>PCE, TCE, DCE</td>
<td>acetate, formate, lactate, glucose</td>
<td>EU526032</td>
<td>(33)</td>
</tr>
<tr>
<td><em>Dehalospirillum multivorans</em></td>
<td>PCE, TCE</td>
<td>H₂, formate, pyruvate</td>
<td>X82931</td>
<td>(34)</td>
</tr>
<tr>
<td><em>Shewanella sediminis</em> HAW-EB3</td>
<td>TCE</td>
<td>acetate, pyruvate, sugar</td>
<td>AY579750</td>
<td>(35)</td>
</tr>
<tr>
<td><em>Sphingobium chlorophenolicum</em></td>
<td>PCP</td>
<td>n. d.</td>
<td>NR_026249</td>
<td>(36)</td>
</tr>
</tbody>
</table>
is residing on the cytoplasmic site of the inner membrane.

![Diagram](image)

**Figure 2.3: Schematic overview of organohalide dehalogenation leading to energy generation in anaerobic bacteria.** The electron flow from the donor (in this case H₂) to the organohalide and final acceptor (here abbreviated as R–Cl) is used for the build up of a proton motive force (PMF) over the inner bacterial cell membrane, which in turn leads to the synthesis of ATP from ADP and inorganic phosphate (Pi). The involved cleavage of the carbon halogen bond is catalysed by reductive dehalogenases. Figure adopted from (38).

### 2.3.2 Terminology and chemistry of organohalide respiration

Microorganisms have evolved different strategies for the removal of the halogen substituent from the carbon moiety. Aerobic conditions allow *oxygenolytic dechlorination*, in which a hydroxy group derived from an oxygen molecule replaces the halogen atom (7). For anaerobic environments, three different dehalogenation mechanisms have been reported. *Hydrolytic dechlorination* uses a hydroxy group from a water molecule to replace the halogen substituent. Both aliphatic and aromatic organohalides can be converted by this mechanism (39; 40). During *reductive dehalogenation*, a hydrogen atom replaces the halogen atom of organohalides. It is a hydrogenolytic mechanism used for both aromatic and aliphatic substrates. Restricted to aliphatic substrates with a higher degree of halogenation is *dihaloelimination* where two adjacent halogen substituents are removed in parallel and the C–C bond is reduced to a C=C bond. Figure 2.4 outlines the different dehalogenation strategies.
This study focuses on the description of reductive dehalogenation. It refers to a chemical process during which a carbon halogen bond is cleaved and the halogen substituent is released in its ionic form into the environment. During the late 1990s, the terminology of this reaction has been subject to a controversial debate. According to Smidt et al. (19), halogenated organic compounds serve as final electron acceptors during dehalogenation. As the transferred electrons in this process are used for energy generation in the form of ATP, it was termed halorespiration. Holliger et al. (11) on the other hand proposed the term dehalorespiration, due to the fact that the final electron acceptor theory was unproven. They claim that the halogenated compound is reduced (thus losing a chloride ion) in order to serve as an electronic sink for reducing equivalents that were generated during the oxidation of the electron donor, e.g. H₂. Researchers from the field have recently agreed to use the term organohalide respiration, which will be used throughout this study (41).
2.3.3 Thermodynamics of reductive dehalogenation

An interesting question is, why reductive dehalogenation preferentially occurs in anoxic conditions? During the cleavage of the carbon halogen bond the carbon atom gets reduced. Consequently this requires the oxidation of reduction equivalents. In the presence of oxygen, the energy value of the reduction equivalents is higher than in its absence so it is more likely that reductive dehalogenation will be observed under anaerobic conditions (42). In the early 1990s, thermodynamic studies by Dolfing and Janssen showed that organohalides can serve as excellent electron acceptors. For that, they calculated their Gibbs free energy formation values (42; 43). Gibbs free energy of substrate and product formation states, whether a given chemical reaction results in free spendable energy or not. For standard conditions (1 M; 25 °C; 1 atm pressure), it can be described in the following terms:

\[ \Delta G^0 = \Delta G^0_f(\text{products}) - \Delta G^0_f(\text{substrates}) \] (2.1)

With respect to redox couples the equation of Gibbs free energy looks as follows:

\[ \Delta G^0 = -n \cdot F \cdot \Delta E^0 \] (2.2)

with \( n \) = number of involved electrons and \( F \) = Faraday constant.

Under biological conditions the equation can be rewritten as:

\[ \Delta G = \Delta G^0 + R \cdot T \cdot \ln \frac{[C]^c \cdot [D]^d}{[A]^a \cdot [B]^b} \] (2.3)

where \( R \) = gas constant; \( T \) = temperature; \([A]^a\) and \([B]^b\) are the concentrations of the substrates and \([C]^c\) and \([D]^d\) are the concentrations of the products of the reaction.

Table 2.2 summarises free energy values for reductive dehalogenation and compares them with important non-halogenated electron acceptors. The catalysed removal of chlorine gives energy yields between -130 and -180 kJ/mol, which is equal to redox potentials from 260 to 480 mV, if \( \text{H}_2 \) is considered as electron donor. These values are similar to the redox potential of \( \text{NO}_3^- / \text{NO}_2^- \), but significantly higher than for \( \text{SO}_4^{2-} \) or \( \text{HCO}_3^- \) reduction. On the other hand, because the redox potential for the reduction of oxygen is much higher, the use of oxygen as electron acceptor in nature is preferred over the use of organohalides. This explains why reductive dehalogenation has primarily been observed in anaerobic environments.
Table 2.2: Overview of Gibbs free energy values and redox potentials. Besides values for the reductive dehalogenation of selected organohalides, values for nonhalogenated redox pairs are shown for improved comparison. The values are taken from (43; 44; 45).

<table>
<thead>
<tr>
<th>Redox pair</th>
<th>$E_0$ (mV)</th>
<th>$\Delta G^0$ (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-chlorobenzoate / benzoate</td>
<td>297</td>
<td>-137.3</td>
</tr>
<tr>
<td>pentachlorophenol / tetrachlorophenol</td>
<td>455</td>
<td>-167.8</td>
</tr>
<tr>
<td>tetrachlorophenol / trichlorophenol</td>
<td>400</td>
<td>-157.1</td>
</tr>
<tr>
<td>trichlorophenol / dichlorophenol</td>
<td>319</td>
<td>-141.5</td>
</tr>
<tr>
<td>DCP / chlorophenol</td>
<td>266</td>
<td>-131.3</td>
</tr>
<tr>
<td>chlorophenol / phenol</td>
<td>438</td>
<td>-164.5</td>
</tr>
<tr>
<td>PCE / TCE</td>
<td>574</td>
<td>-55.4</td>
</tr>
<tr>
<td>TCE / cis-DCE</td>
<td>550</td>
<td>-53.1</td>
</tr>
<tr>
<td>cis-DCE / VC</td>
<td>420</td>
<td>-40.6</td>
</tr>
<tr>
<td>VC / ethene</td>
<td>450</td>
<td>-43.4</td>
</tr>
<tr>
<td>$\text{O}_2 / \text{H}_2\text{O}$</td>
<td>818</td>
<td>-180.7</td>
</tr>
<tr>
<td>$\text{NO}_3^- / \text{NO}_2^-$</td>
<td>433</td>
<td>-41.7</td>
</tr>
<tr>
<td>$\text{SO}_4^{2-} / \text{H}_2\text{S}$</td>
<td>-217</td>
<td>20.9</td>
</tr>
<tr>
<td>$\text{HCO}_3^- / \text{CH}_4$</td>
<td>-238</td>
<td>23</td>
</tr>
</tbody>
</table>

Because microorganisms in general catalyse reactions that give highest energy yields, organohalides represent an important energy source in anoxic environments (42).

Aromatic organohalides, especially when derived from lignin sources, often contain a methyl ether function. These aromatic methyl ethers are substrates for another bacterial pathway (O-demethylation) in which the methyl groups from these substrates are used for the synthesis of acetylCoA as part of the general carbon metabolism. Both pathways, organohalide respiration and O-demethylation occur in certain specific bacteria, e.g. the Desulfobacterium genus (46). Interestingly, the key proteins in both of these pathways are cobalamin-dependent metalloenzymes (reductive dehalogenases and O-demethylases, respectively). It has been proposed that aromatic molecules, which are both chlorinated and O-methylated, are consecutively used as substrates of reductive dehalogenation and O-demethylation (47) as illustrated in figure 2.5. Both of these systems are underexplored on a structural level and the main aims of this thesis are centered around the X-ray crystallographic investigation of these two enzyme families accompanied by biochemical and biophysical studies. The next sections introduce in detail the cobalamin-dependent reductive dehalogenases and O-demethylase enzymes.
Chapter 2. Introduction  
Cobalamin-dependent enzyme systems

![Chemical structures and reactions involving organohalide respiration and O-demethylation](image)

Figure 2.5: Putative metabolic link between organohalide respiration and O-demethylation. Aromatic organohalides that additionally contain a methyl ether function may serve as substrate for both reductive dehalogenases and O-demethylases.

2.4 Reductive dehalogenases: key enzymes in organohalide respiration

Organohalides are the final electron acceptors in organohalide respiration and reductive dehalogenases (RdhAs) catalyse the dehalogenation reaction of various organohalides (17). In most cases a chloride atom is replaced by a hydrogen atom. The electron transfer involved in this redox reaction is used by the respective anaerobic bacteria to build up a proton motive force (PMF), which in turn is used for the production of energy in the form of ATP (2.3) (16; 41). However, the molecular basis for the generation of the PMF is unclear. All RdhAs characterised to date appear to be monomeric and membrane associated proteins with a molecular weight range between 35 - 65 kDa. Usually, they contain one cobalamin (Vitamin B\textsubscript{12} derivative) and two Fe-S clusters as cofactors, which make them sensitive to molecular oxygen. (19). Substrates that are subject to dehalogenation by several different RdhAs are chlorinated ethenes and chlorophenols. Enzymes dehalogenating the former are termed PceA, the latter CprA. Recent availability of sequenced genomes of organohalide-respiring bacteria have shown that many strains contain a large number of rdhA genes. For example, the strain Desulfotobacterium hafniense DCB-2 harbours seven rdhA genes (48); Dehalococcoides mccartyi (previously ethenogenes) strain CBDB1 contains 32 rdhA genes (49) and the different Dehalobacter strains contain over 20 rdhA genes (50; 51). In total more than 250 rdhA genes have
been identified using different bioinformatic approaches (17), reliable biochemical information for their protein products, however, are only available for a few of them. The best studied RdhAs are PceA from *Dehalobacter restrictus* (38; 52), CprA from *Desulfotobacterium dehalogenans* (53) and PceA from *Sulfurospirillum multivorans* (54). Table 2.3 summarises and compares the main properties of the purified and biochemically studied enzymes. After translation into a polypeptide chain, the respective RdhAs need to be transported to their target destinations. RdhAs contain the TAT (twin arginine translocation) consensus sequence RRXFXK in their N-terminal located signal peptide. This sequence is recognised by chaperones that support protein folding in the cytoplasm. TAT signal sequences are often used to transport bacterial redox enzymes into the periplasmic space (55). It is therefore assumed that RdhAs are likewise biologically active in the periplasm. Usually, proteins that are subject to the TAT transport machinery are fully folded and loaded with their cofactors before they are transported across the membrane. During the transport the TAT signal sequence is recognised by the TAT translocase machinery and is cleaved off by specific proteases (56). The electrons that are needed for the reductive dehalogenation reaction must be transported horizontally through the plasma membrane into the catalytic centre of the RdhA. Hence the enzyme must be in contact with the membrane. No transmembrane regions have been identified for any RdhA yet. However, next to virtually any rdhA gene, a rdhB gene is located, that translates into a \( \sim 10 \) kDa polypeptide chain with two to three predicted membrane spanning segments. The RdhB proteins are supposed to serve as membrane anchor proteins for their corresponding RdhA (57). Beyond the determination of substrates, cofactors and activity, structural and functional studies of RdhAs have been hampered due to the inability to purify the proteins in sufficient amounts from expression in their own host. This is due to a large doubling time of these bacteria, the poor expression of specific RdhAs and their general low solubility. Furthermore, studying the biochemistry of RdhAs by homologous expression, does not allow the investigation of mutant proteins, e.g. the use of site-directed mutagenesis. At the same time, no recombinant expression and purification system exists for any RdhA, mainly because of the difficulties associated with expression, anaerobic purification and cofactor reconstitution. There are thus many open questions concerning the biochemistry of RdhAs:

- Which residues constitute the active site?
- Which residues are involved in cofactor and substrate binding?
- What is the protein fold and is it conserved between different RdhAs?
- What is the reaction mechanism, e.g. are transient radicals formed?
- How are the electrons transferred from the membrane into the catalytic centre?
Chapter 2. Introduction

Because of the low sequence similarity with other protein classes, even bioinformatic analyses provide little help in the understanding of reductive dehalogenases.

2.4.1 Cofactors of reductive dehalogenases

With the exception of one member (58), all isolated RdhAs contain one cobalamin and two Fe-S clusters as enzymatic cofactors. The best cofactor characterisation has been carried out with the enzyme PceA from *Dehalobacter restrictus*. Reversed phase high pressure liquid chromatography (HPLC) and electron paramagnetic resonance (EPR) spectroscopy have confirmed the presence of one cobalamin and two 4Fe-4S clusters (52). In general Fe-S cluster containing proteins provide a chemical basis for electron transfer reactions. Fe-S clusters are therefore important for fundamental biological concepts including photosynthesis, nitrogen fixation and respiratory complexes (71). A large variety of different clusters exist (see figure 2.6). Common properties are central iron atoms that are coordinated within the protein by the sulphur atoms of cysteine side chains and in most cases bridged by inorganic sulphur (72). A common 4Fe-4S coordinating motif of the ferredoxin type has been identified in the primary amino acid sequence of proteins (73). It has the following sequence: CXXCXXC(X)_{10-50}C (71); and is also present in characterised RdhAs (59). Furthermore, this motif is used for the identification of new *rhdA* genes (17). The cysteine residues coordinating the second 4Fe-4S cluster in RdhAs are less conserved then the first one, but they similarly reside in the C-terminal part of the protein (74). Their precise function in RdhAs remains to be elucidated. It is possible that they are exclusively used to transfer the incoming electron from the membrane into the active site. However, they may also be directly involved in the reaction mechanism leading to the substrate dehalogenation.

The second cofactor of RdhAs is the organometallic cobalamin, a vitamin B\textsubscript{12} derivative. The crystal structure of vitamin B\textsubscript{12} was solved many years ago (75), identifying it as one of the most complex enzymatic cofactors. Cobalamins contain a central cobalt atom that is coordinated by four nitrogen atoms from a planar corrin ring. From the upper axial position the cobalt atom can be coordinated to an -OH, -CH\textsubscript{3}, -CN or -adenosyl moiety, leading to the formation of hydroxo-cobalamin, methyl-cobalamin, cyano-cobalamin or adenosyl-cobalamin, respectively. The corrin ring is usually covalently connected to a dimethylbenzimidazole (DMB) base. This base can either be displaced from the cobalt atom, or, upon rotation form a sixth coordination bond to the cobalt from the lower axial site. When bound to a protein, the DMB of cobalamin is often rotated away from the cobalt and displaced by a nitrogen atom from a histidine side chain of the protein. One distinguishes the three different cobalamin conformations when bound to a protein:

- **DMB-on**: The DMB is coordinated to the cobalt atom from the lower axial site.
### Table 2.3: Comparison of reductive dehalogenases, for which successful purifications have been reported.

Unless otherwise stated the enzymes have been produced by homologous expression. Protein sequences are taken from the UniProt web server ([www.uniprot.org](http://www.uniprot.org)) and the accession numbers are listed. Used abbreviations: min = minute; n.d. = not determined; t\(_{1/2}\) = half-life.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene</th>
<th>Substrate</th>
<th>Cofactors</th>
<th>MW (kDa)</th>
<th>O(<em>2) t(</em>{1/2})</th>
<th>Protein Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desulfomonile tiedjei</td>
<td>n. d.</td>
<td>3-CB</td>
<td>heme, Fe-S</td>
<td>64 + 37 heterodimer</td>
<td>n. d.</td>
<td>n. d.</td>
<td>(58)</td>
</tr>
<tr>
<td>Dehalobacter restrictus</td>
<td>pceA</td>
<td>PCE, TCE</td>
<td>1 cobalamin, 2x 4Fe-4S</td>
<td>60 monomer</td>
<td>280 min</td>
<td>Q8GJ27</td>
<td>(59; 52)</td>
</tr>
<tr>
<td>Desulfitobacterium dehalogenans</td>
<td>cprA</td>
<td>CHPA, DCPs</td>
<td>1 cobalamin, 1x 3Fe-4S, 1x 4Fe-4S</td>
<td>48 monomer</td>
<td>90 min</td>
<td>Q0XD04-1</td>
<td>(53)</td>
</tr>
<tr>
<td>Desulfitobacterium chlorarespirans</td>
<td>cprA</td>
<td>CHPA, DCPs, TCPs</td>
<td>1 cobalamin, 2x 4Fe-4S</td>
<td>50 n. d.</td>
<td>77 min</td>
<td>Q8RQC9</td>
<td>(60; 27)</td>
</tr>
<tr>
<td>Desulfitobacterium hafniense DCB-2</td>
<td>cprA1</td>
<td>CHPA</td>
<td>1 cobalamin, 12Fe-13S</td>
<td>47 monomer</td>
<td>n. d.</td>
<td>P81594</td>
<td>(61)</td>
</tr>
<tr>
<td>Desulfitobacterium hafniense PCP-1</td>
<td>crdA</td>
<td>DCPs, TCPs, PCP</td>
<td>1 cobalamin</td>
<td>35 monomer</td>
<td>sensitive</td>
<td>Q8GEFE2</td>
<td>(62)</td>
</tr>
<tr>
<td>Desulfitobacterium hafniense PCP-1</td>
<td>cprA5</td>
<td>DCPs, TCPs</td>
<td>1 cobalamin, Fe-S</td>
<td>57 monomer</td>
<td>110 min</td>
<td>Q6V7J3</td>
<td>(63)</td>
</tr>
<tr>
<td>Desulfitobacterium hafniense PCE-S</td>
<td>pceA</td>
<td>PCE, TCE</td>
<td>1 cobalamin, 8 Fe / 8 S</td>
<td>65, homotrimer</td>
<td>50 min</td>
<td>Q8JG31</td>
<td>(64)</td>
</tr>
<tr>
<td>Desulfitobacterium hafniense Y-51</td>
<td>pceA</td>
<td>PCE, TCE, chloroethanes</td>
<td>1 cobalamin, Fe-S</td>
<td>58 monomer</td>
<td>330 min</td>
<td>Q8L172</td>
<td>(31)</td>
</tr>
<tr>
<td>Desulfitobacterium sp. PCE-1</td>
<td>n. d.</td>
<td>PCE, CHPA</td>
<td>1 cobalamin</td>
<td>48 monomer</td>
<td>n. d.</td>
<td>n. d.</td>
<td>(65)</td>
</tr>
<tr>
<td>Dehalococcoides m. cartys strain 195</td>
<td>pceA</td>
<td>TCE, PCE</td>
<td>1 cobalamin, 2x 4Fe-4S</td>
<td>51 monomer</td>
<td>n. d.</td>
<td>Q3Z9N3</td>
<td>(66; 67)</td>
</tr>
<tr>
<td>Dehalococcoides m. cartys strain 195</td>
<td>tceA</td>
<td>TCE</td>
<td>1 cobalamin, 2x 4Fe-4S</td>
<td>61 monomer</td>
<td>n. d.</td>
<td>Q3ZAB8</td>
<td>(66; 68)</td>
</tr>
<tr>
<td>Clostridium bifermentans</td>
<td>pceC</td>
<td>PCE, TCE, DCE</td>
<td>1 cobalamin</td>
<td>35 homodimer</td>
<td>20 h</td>
<td>Q93SG6</td>
<td>(69)</td>
</tr>
<tr>
<td>Sulfurospirillum multivorans</td>
<td>pceA</td>
<td>PCE, TCE</td>
<td>1 cobalamin, 8Fe / 8S</td>
<td>58 monomer</td>
<td>120 min</td>
<td>O68252</td>
<td>(54; 70)</td>
</tr>
</tbody>
</table>
• DMB-off: DMB is rotated away from the cobalt atom.

• DMB-off / His-on: The DMB is rotated away and a nitrogen atom from a histidine side chain of the protein coordinates to the cobalt from the lower axial site.

Sometimes, the DMB in cobalamin can be replaced by an adenine base that can also coordinate to the cobalt atom from the lower axial site. This form has been termed pseudo-cobalamin (76). Cobinamides are vitamin B\textsubscript{12} derivatives, in which no base is attached to the corrin ring. An example for that is dicyano-cobalamin. The central cobalt atom can change between the three different oxidation states Co\textsuperscript{+1}, Co\textsuperscript{+2} and Co\textsuperscript{+3}, leading to cob(I)alamin, cob(II)alamin and cob(III)alamin, respectively. Thus they can span a great range of different electron potentials and make them a versatile tool for enzymes mediating complex redox reactions. Cob(I)alamin has zero axial ligands bound, cob(II)alamin has one and cob(III)alamin has two (77). Cobalamin has been identified as cofactor of RdhAs mainly by spectroscopic methods. The retention time of the cobalamin during HPLC was similar to commercially purchased hydroxo-cobalamin and, from its characteristic EPR spectrum, its configuration is in the DMB-off form (52). Later on, the cobalamin of PceA from \textit{Sulfurospirillum multivorans} was isolated and identified (78). It is similar to pseudo-cobalamin and differs from it only by the lack of the methyl group on carbon 176, as confirmed by a combination of HPLC, nuclear magnetic resonance (NMR), UV-Vis spectroscopy and X-ray crystallography experiments. This cobalamin form has been named \textit{norpseudo-cobalamin}. Furthermore, the EPR spectra have revealed that the cobalamin in the oxidation state Co\textsuperscript{+2} is bound to the RdhA in the DMB-off configuration. This is in line with results from EPR studies with the RdhA from \textit{Desulfitobacterium dehalogenans} that also binds its cobalamin in the DMB-off conformation (53). The genome of \textit{Dehalococcoides} strain CBDB1 contains a few, but not all necessary genes required for the biosynthesis of cyanocobalamin in addition to cobalamin uptake genes (49), similar to the strain 195 that harbours cobalamin uptake and salvage genes (79). On the contrary, the genome of \textit{Dehalobacter restrictus} contains a complete cobalamin biosynthetic pathway (51). However, there are still many unknowns regarding cobalamins in RdhAs:

• What kind of upper axial ligand is used?

• Do all RdhAs from different organohalide-respiring bacteria use \textit{norpseudo-cobalamin}?

• Does a conserved amino acid sequence cobalamin-binding motif exist in the RdhA family?

• To which extent synthesise or modify organohalide-respiring bacteria their required cobalamins?
Recent investigations with cultures of *Dehalococoides maccartyi* strains have shown that they need to be supplemented with cyanocobalamin in order to observe high reductive dehalogenation activity. Interestingly, the same rates were observed when cyanocobalamin was replaced by dicyano-cobinamide or DMB (80; 81). More studies on cobalamins extracted from organohalide-respiring bacteria and on the putative genes involved in cobalamin biosynthesis / modification need to be performed in order to answer such questions.

![Figure 2.6: Cofactors of RdhAs.](image)

**Figure 2.6: Cofactors of RdhAs.** a) - c): Schematic drawing of common Fe-S clusters as enzymatic cofactors. a) 2Fe-2S cluster. b) 3Fe-4S cluster. This cluster is supposed to be present in CprA from *Desulfotobacterium dehalogenans*. c) 4Fe-4S cluster as found in most RdhAs. d) Chemical structure of cobalamin (vitamin B12). The central cobalt atom is laterally coordinated by a planar corrin ring. From the top the Co can be coordinated by a variety of functional groups, e.g. hydroxo, cyano or adenosyl, as indicated by the R. From the lower axial position a dimethylbenzimidazole base (DMB) can coordinate to the central cobalt atom.

### 2.5 Gene arrangements and molecular regulation of reductive dehalogenases

The development of rapid sequencing methods have enabled the identification of more than 200 *rdhA* genes in the genomes of organohalide-respiring bacteria (17). Studies of
these respective gene loci have revealed the presence of several genes associated with rdhAs.

### 2.5.1 RdhB: a membrane anchor protein for RdhA

All rdhA genes belonging to the pceA / cprA family harbour the gene pceB or cprB respectively in close proximity (less than 100 nucleotides). Usually cprB is located upstream of cprA, whereas pceB is located downstream of pceA. One exception exists for the cprA5 gene from Desulfitobacterium hafniense PCP-1, for which the cprB5 is located downstream (63). This is interesting as the sequence alignments of the RdhAs have shown a high similarity between CprA5 and PceA’s from Desulfitobacterium hafniense Y-51, PCE-S and Dehalobacter restrictus. The physiological function of the pceB / cprB genes has not been elucidated experimentally. They code for hydrophobic proteins with 100 - 110 amino acids and a MW of 11 - 12 kDa. Kyte-Doolittle plots of this protein have identified two to three transmembrane regions. Due to these characteristic features they are supposed to act as membrane anchors for the respective RdhAs. First experiments conducted in the laboratory of Gabriele Diekert seem to support the membrane anchor function of RdhBs using frozen membrane samples of organohalide-respiring bacteria (Diekert et al., unpublished data presented at the Royal Society meeting: Understanding and application of organohalide respiration: from genomes to structures in 2011). Another identified gene in close proximity to pceA / cprA has been termed pceC or cprC, respectively. This gene has also been found in most gene loci that contain the pceB / cprB genes, for example Desulfitobacterium hafniense Y51, DCB-2, Desulfitobacterium dehalogenans, and Dehalobacter restrictus (21). The pceC / cprC gene codes for a protein of unknown function. It contains five or six predicted transmembrane regions and has sequence similarities with the membrane-bound transcriptional regulator of the NosR/NirI type (82; 83).

### 2.5.2 RdhT: a specific chaperone for RdhA

For the species Desulfitobacterium hafniense TCE-1, DCB-2 and Dehalobacter restrictus the gene rdhT is situated downstream of rdhC. rdhT encodes for a trigger factor protein RdhT, which is assumed to provide functional support for the maturation process from apo-RdhA to holo-RdhA (84). Basically, RdhTs are similar to the trigger factor protein family, that acts as chaperones in the folding of nascent polypeptide chains, emerging from the ribosome exit channel; however they lack the trigger factor specific N-terminal ribosome binding domain (85). Recently, for the RdhT from Desulfitobacterium hafniense Y-51, successful expression and purification in E. coli has been reported. Functional studies suggest peptidyl-prolyl cis-trans isomerase as well as a chaperon activity. Furthermore immunoprecipitation studies lead to the assumption that RdhT binds RdhA only before its signal sequence is proteolytically cleaved off (86). Recent in vivo studies
in *E. coli* indicate an interaction between RdhT and the TAT signal sequence of RdhA in a *Desulfotobacterium* strain (87). It is unclear at which stage (if at all), RdhT is involved in the maturation of RdhA. The situation is even more confusing due to the fact that many *rdhA* genes are not associated with *rdhT* genes (57). Figure 2.7 summarises the localisation and putative function of the *rdhABCT* gene operon and their corresponding protein products.

**DNA level**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>rdhA</em></td>
<td>Reductive dehalogenase</td>
</tr>
<tr>
<td><em>rdhB</em></td>
<td>RdhA anchor?</td>
</tr>
<tr>
<td><em>rdhC</em></td>
<td>unknown</td>
</tr>
<tr>
<td><em>rdhT</em></td>
<td>RdhA chaperone?</td>
</tr>
</tbody>
</table>

**Cytoplasm**

- RdhB

**Periplasm**

- RdhA
- RdhC

**Figure 2.7:** Overview of the *rdhABCT* gene operon, which is present in this form in *Dehalobacter restrictus* and in *Desulfotobacterium hafniense* Y-51 and TCE1. Upon expression, folding, incorporation of cobalamin and cleavage of its N-terminal signal sequence, RdhA is transported through the plasma membrane by the TAT protein export system. RdhB is supposed to be a membrane anchor for RdhA and the function of RdhC is unknown; the number of their predicted transmembrane regions are two or three and five or six, respectively. RdhT might function as a trigger factor for proper folding of PceA.

Furthermore, the genes *rdhD, rdhE, rdhF, rdhG, rdhH, rdhI* and *rdhZ* have occasionally been identified in the genomes of organohalide-respiring bacteria, but nothing is known about their biological functions (57).

### 2.5.3 Transcriptional control of organohalide respiration

For the species *Desulfotobacterium dehalogenans* and *Desulfotobacterium hafniense* DCB-2 the gene *cprK* has been identified (48), encoding the transcriptional regulator CprK (a member of the CRP / FNR regulator class), for which a detailed structural and functional characterisation has been reported (88; 89). After sensing and binding of ortho-chlorophenol ligands, CprK induces the expression of CprA via DNA promoter (so called dehalobox) binding (90). Two other modes of transcriptional control do exist for
rdhA genes, however, with respect to organohalide respiration, they are biochemically less well understood than the CprK executed regulation. Firstly, two component regulatory systems, comprising a histidine kinase and a response regulator, and secondly a MarR type regulator have been found in different *Dehalococccoides* strains (49; 79) In summary, with the exception of CprK and to a lesser extent PceT, no biochemical data exist for any protein associated with reductive dehalogenase.

It will require the combined effort of microbiology, molecular biology, enzymology and structural biology in order to broaden the current understanding of this intriguing respiratory pathway.
2.6 Other cobalamin-dependent enzyme systems

Cobalamin is a versatile enzymatic cofactor, mainly due to the fact that it can make use of different upper axial ligands and thus is able to span a long range of electron potentials (91). Cobalamins are only synthesised by a number of specific bacteria (92), but are involved in catalysing a large number of evolutionarily important and biological diverse chemical reactions in all organisms from bacteria to humans (93). In addition to reductive dehalogenases, two other main classes of cobalamin-dependent enzymes are known, that catalyse methyl transfer reactions or transformation reactions, respectively (94). For transformation reactions adenosyl-cobalamin cofactors are used, which are usually bound in the DMB-off form to the protein. For catalysis, the coordination bond between the central cobalt atom and the carbon atom from the upper axial ligand adenosyl is prone to homolytic cleavage, resulting in the formation of a 5'-deoxyadenosyl radical species (95). This radical is in turn able to catalyse a large variety of chemical reactions ranging from carbon skeleton isomerisation (96) through ribonucleotide reduction (97) to aminomutations (98). Common structural motifs found in these enzyme systems are Rossmann fold domains and triosephosphate isomerase (TIM) barrels (99). Several methylcobalamin-dependent enzymes have been characterised that catalyse methyl transfer reactions. Such reactions are important in different metabolic pathways in anaerobic bacteria (100). The methyl groups are transferred from a substrate to fully reduced cob(I)alamin, leading to the formation of methylcob(III)alamin, as reaction intermediate. In a second catalytic step the methyl group is further transferred to the final methyl group acceptor of the catalysed reaction, which results in product formation and restored cob(I)alamin. Usually, the substrates and cofactors involved in the two methyl transfer steps, reside on different protein modules. As for the adenosyl-cobalamin dependent enzymes, the cofactor in methyltransferases is often bound to a Rossmann fold domain (101), whereas the substrates and methyl group acceptors are often bound to TIM barrel (102). The next section introduces in more detail a specific methyltransferase reaction, termed O-demethylation, which occurs in certain organohalide-respiring bacteria and might in fact be metabolically linked to organohalide respiration, in more detail.
2.7 The three-component and cobalamin-dependent O-demethylase enzyme system

O-demethylation is a metabolic pathway utilised by certain bacterial species that leads to the synthesis of methyltetrahydrofolate (MTHF) from tetrahydrofolate (THF) as outlined in figure 2.8. THF is then generated from MTHF for a next round of O-demethylation (103). Substrates that are used as methyl group donors in O-demethylation are usually aromatic methyl ethers, with vanillate being the best studied example (104). In O-demethylation, the covalent bond between the oxygen atom and the methyl carbon atom becomes cleaved, leading to the release of the methyl group and the formation of an hydroxo group on the phenyl moiety (100).

Figure 2.8: General outline of O-demethylation as bacterial pathway. Methyltetrahydrofolate is either obtained from formyltetrahydrofolate or via O-demethylation using phenyl methyl ethers as substrate. In a subsequent enzymatic step, the methyl group is combined with carbon monoxide (CO) to yield acetyl-CoA, which fuels central metabolic pathways while tetrahydrofolate is regenerated in the same reaction.

Both methyl transfer (MT) reactions involved in O-demethylation rely on the central protein component in the system that binds the cobalamin cofactor. The cobalamin is initially in the Co$^{+1}$ state (cob(I)alamin), which is a chemically very reactive state with a low redox potential ($\sim$ -500 mV) that allows the nucleophilic attack of the oxygen carbon bond and subsequent abstraction of the methyl group from the aromatic methyl ether, leading to the formation of methyl-cob(III)alamin. It is known that the cobalamin in methyltransferase enzymes can be sometimes oxidised into the inactive Co$^{+2}$ (cob(II)alamin) state (105). From there, a reduction reaction has to bring the cobalamin back into the active Co$^{+1}$ state. It is proposed that this reaction in the investigated O-demethylases is catalysed by an activating enzyme in an ATP dependent manner (46; 106). In the second MT reaction the methyl group is transferred from methylcobalamin to THF, leading to the formation MTHF and restored cob(I)alamin.
The methyl group of MTHF is subsequently used to synthesise acetyl-CoA, which in turn is used to fuel central metabolic pathways (48).

The substrate-, cobalamin- and THF- binding sites reside on individual protein components, respectively, as shown in figure 2.9. It is therefore imperative to assume that substantial protein-protein interactions as well as protein module rearrangement reactions must occur between the individual subunits along the catalytic pathway underpinning O-demethylation, in order to orchestrate such a delicate molecular system. In particular, the cobalamin-binding protein must be able to move back and forth between the donor- and acceptor-domains in order to bring the cobalt atom of its cofactor into close proximity to the substrate and THF, respectively. However, it is currently not clear what kind of interactions occur between the three domains and to which extent they are transient or permanent. One of the main reasons for the low level of current understanding is the lack of structural information regarding O-demethylation in general. No crystal structures exist of any O-demethylase protein. However, it has been shown using small angle X-ray scattering (SAXS) for other methyltransferase systems that the involved protein-protein interactions are of rather weak nature (107). In the case of the similar three-component methanol transferase, it has been demonstrated that the substrate binding domain and cobalamin-binding domain interact tightly with each other (108). However, the main difference between that system and O-demethylases is that the methyl acceptor domain uses a coenzyme-M (2-mercaptoethanesulfonate) instead of THF that takes up the methyl group (109).

Individual protein components of the cobalamin-dependent O-demethylases have been purified from Moorella thermoacetica (110), Acetobacterium dehalogenans (111; 112; 113) and Desulfitobacterium hafniense DCB-2 (46). The purified M. thermoacetica system has been shown to demethylate vanillate and dicamba (110), the purified A. dehalogenans O-demethylase catalyses in vitro the demethylation of vanillate, isovanillate, syringate and guaiacol (112). The substrate spectrum of the the D. hafniense O-demethylase seems to be rather broad, including vanillate, isovanillate, syringate, guaiacol and methyl chloride (46).

2.7.1 Sequence alignments between O-demethylase proteins

An alignment of the so far heterologously expressed and biochemically characterised O-demethylase components is shown in figure 2.10. It only contains protein sequences that were biochemically unambiguously identified as components of cobalamin-dependent O-demethylase systems. Amino acid sequence alignments between different cobalamin-binding domains and methyl acceptor domains within the O-demethylase family reveal that these two modules are very well conserved. Between the cobalamin-binding domains, the highest degree of conservation lies around the conserved cobalamin-binding motif DXHXXG, which is a common binding motif in the methyltransferase class (100).
A similar degree of conservation is found in the alignment of THF binding acceptor domains. The first asparagine residue (N, highlighted in bold) of the conserved block SNISFN has been implied in transition-state stabilisation in similar methyl transfer systems (114; 115; 116). The alignment of the substrate binding domains indicates that there is a large diversity between the proteins, which might indicate that the evolutionary pressure on these proteins is higher, because of the availability to use a large diversity in different aromatic methyl ether substrates. On the other hand, all of the aligned proteins are supposed to use vanillate as substrate. It would be interesting to find out which residues are involved in substrate binding and / or catalysis, however, the lack of a reliable homology model in addition to the low degree of conserved residues do not allow any sophisticated conclusions to be drawn.

2.7.2 Metabolic links between O-demethylation and organohalide respiration

A number of reasons support the possible scenario in which a metabolic relationship exists between O-demethylation and organohalide respiration. First, there is a vast amount of naturally produced organic compounds that are both halogenated and O-methylated. These molecules are chemically diverse and range from very simple to extremely complex as described above (1). Second, the recent availability of sequenced genomes has identified the presence of several operons that seem to contain all necessary genes for O-demethylation in e.g. *D. hafniense* DCB-2 (48) and *Dehalobacter* sp. CF (50). Interestingly, the *D. restrictus* P23K genome does not seem to contain any O-demethylase relevant genes. It is well documented that this *D. restrictus* strain is exclusively using the aliphatic organohalides tetrachloroethene and trichloroethene as final electron...
Cobalamin-dependent enzyme systems

Chapter 2. Introduction

Methyl donor domains

[Alignment of protein domains]

Cobalamin-binding domains

[Alignment of protein domains]

Methyl acceptor domains

[Alignment of protein domains]

Figure 2.10: Primary sequence alignment between individual protein domains of biochemically characterised O-demethylase components. Top = Substrate binding domains alignment; middle = Cobalamin-binding domains; bottom = Methyl acceptor domains alignment. 75% conservation between the sequences = light blue highlight; 100% conservation = dark blue highlight.
acceptors and does not require O-demethylation for its metabolism. O-demethylation of tetrachloroguaiacol, tetrachloroveratrol and pentachloroanisole was observed after a first step of dehalogenation, with further dehalogenation after the demethylation in *D. hafniense* PCP-1 (47; 117).

However, if the two pathways do share the same substrates, it remains to be elucidated how the substrates are transported across the plasma membrane as organohalide respiration is supposed to take place in the periplasm, whereas O-demethylation is a cytoplasmic reaction. In summary, no conclusive data exist that would allow a decision whether the two pathways share the same substrates.

### 2.8 Aims and outline of this thesis

The main aims of the PhD project undertaken were the biochemical and X-ray crystallographic study of two cobalamin-dependent enzyme systems from organohalide-respiring bacteria, namely reductive dehalogenases and O-demethylases.

In addition to the presented results chapters, the basic theory of the main method used for this study - macromolecular X-ray crystallography - is separately described in *chapter 3*.

A necessary prerequisite for the ambitious goal of crystallising a (membrane associated) reductive dehalogenase, is the availability of a method that allows the reliable and robust purification of large quantities of pure protein. Due to the fact that such a system was not available at the beginning of the project, a systematic approach was undertaken to explore the recombinant expression of the reductive dehalogenase PceA from *Dehalobacter restrictus* in *E. coli* using a variety of different solubility tags and expression conditions. The main reason why this particular reductive dehalogenase was chosen is the fact that its substrate and cofactor composition has been explored in most detail in comparison to other identified dehalogenases, making it easier to monitor any success along the pathway from cloned gene to purified holo-enzyme that would serve as promising starting material for initial crystallisation attempts.

*Chapter 4* describes the established protocol for the recombinant production and cofactor reconstitution method for PceA, that is theoretically applicable for any other reductive dehalogenase in a similar fashion. The extensive efforts undertaken resulted in the general conclusion that a recombinant expression of soluble PceA is only possible if the enzyme is covalently fused to a trigger factor protein, but that it is currently not possible to purify PceA away from this trigger factor protein making successful crystallisation unlikely.

In order to better understand the assumed functional importance of the PceA specific trigger factor protein PceT, in the maturation and transport processes of the reductive dehalogenases, PceT was characterised structurally and biochemically. *Chapter 5* describes this characterisation with special emphasis on the proposed interaction between
Cobalamin-dependent enzyme systems

PceT and the TAT signal sequence of PceA. A methodology is presented that leads to the successful crystallisation of PceT which, interestingly, relies on the presence of arginine. However, the obtained crystals did only diffract to a resolution of 3.7 Å and it was subsequently not possible to solve the three-dimensional structure of PceT.

The second part of the PhD project focused on the three-component and cobalamin-dependent $O$-demethylation pathway, as experimental evidence accumulates, that suggest metabolic links between organohalide respiration and $O$-demethylation. Here, the main aims were to solve crystal structures of the participating protein modules, investigate potential protein-protein interactions and, if possible, provide additional information regarding functional links between organohalide respiration and $O$-demethylation.

Chapter 6 describes the successful structural determination of the first protein involved in $O$-demethylation, the cobalamin-binding module CobDH from *Desulfotobacterium hafniense* DCB-2 by X-ray crystallography, along with its spectroscopic characterisation, which led to interesting findings regarding similarities and differences when compared to other enzyme systems that catalyse methyl transfer reactions.

Chapter 7 outlines the determined crystal structure of the protein component in $O$-demethylation, the methyl acceptor domain (MT2DH) that binds tetrahydrofolate, with and without bound cofactor. In combination with spectroscopic studies it has been demonstrated that MT2DH exclusively catalyses the methyl group transfer from methylcobalamin to tetrahydrofolate and a reaction mechanism for this transfer is proposed. Finally, biophysical experiments have confirmed that the proposed methyl donor protein (MT1DH) indeed binds to aromatic methyl ethers and that MT1DH only binds such substrates if they are additionally carboxylated.

Overall, the results presented in this thesis greatly enhance our knowledge on how organohalide-respiring bacteria make use of highly specified cobalamin-dependent enzyme systems for the metabolic utilisation of complex and toxic organic molecules.

2.9 References


Cobalamin-dependent enzyme systems


Cobalamin-dependent enzyme systems


49


Chapter 3

X-ray crystallography of proteins
3.1 General principle of X-ray crystallography

In order to understand on a molecular level, how proteins accomplish their complex biological functions, structural information is crucial to allow the unambiguous assignment of the amino acid side chains positions, as well as interacting molecules (enzymatic cofactors, substrates, etc.) in three-dimensional space. X-ray crystallography is the oldest method and remains the most popular one for the determination of protein structures, which is reflected in the number of deposited structures in the protein data bank (PDB, www.pdb.org). Figure 3.1 illustrates this major impact of X-ray crystallography.

![Figure 3.1: Number of deposited protein structures in the PDB. Left: Total number of deposited protein structures in the pdb. So far 75,215 structures solved by X-ray crystallography have been deposited; NMR contributed 8681 depositions and 678 structures have been reported using other methods. Right: The number of solved protein structures has been constantly increasing and still does. Numbers were taken from the PDB server in June 2013.](image)

In order to visualise the three-dimensional structure of an object, the wavelength of the incident light must not be longer than the size of the object itself. The theoretical resolution limit is half the wavelength of the incident light used. This limitation rule also applies if additional lenses, as used for examples in microscopes, are placed between the object of investigation and the human eye. Hence, visualisation techniques that rely on the wavelength range between 300 - 700 nm (visible light spectrum), have a resolution limit of 300 nm, the approximate size of cell organelles (1). Current electron microscopy (EM) techniques can be used to visualise large macromolecular complexes like ribosomes or viruses, but severe radiation damage is still limiting the resolution of EM to \( \sim 7 - 8 \, \text{Å} \) (2), which does not allow the elucidation of atomic details in proteins. X-rays have wavelengths of \( \sim 1 \, \text{Å} \) and are suited to visualise such details; for example covalent bond lengths between two adjacent carbon atoms range from 1.2 - 1.54 Å. Unfortunately, simple X-ray lenses do not exist, although stacks of strongly curved lenses
achieve modest focussing of X-rays. A main reason for this is that the refractive index for X-rays is comparable to the refractive index of a vacuum (1). On the other hand, the electrons of all atoms in a protein (or any other molecule) diffract X-rays, which can be recorded on suitable detectors. The recorded diffraction patterns thus differ from protein to protein depending on its atomic composition. Generally, diffraction intensities from a single protein are too weak in order to be detected, however, X-ray free-electron lasers (X-FELs) aim to change this using very short and intense pulses (3). In order to increase the signal to noise ratio one makes use of the inherent properties of crystals, in which many proteins are arranged in the same orientation with respect to each other. This has the effect of significant amplification of the molecular diffraction pattern at particular points governed by the lattice parameters (4). Recorded diffraction patterns are used to computationally calculate an electron density map of the protein under investigation. Based on the shape of the three-dimensional density map and previous knowledge about the geometry of the protein it is possible to model its atoms into the electron density and to thus solve its three-dimensional crystallographic structure (5).

3.2 Crystallisation of proteins

Along the experimental and computational pathway of solving the crystal structure of a particular protein, the process of obtaining well diffracting crystals is arguably a major bottleneck. However, experimental improvements are constantly made, leading to an increased likelihood of successful crystallisation of the target protein. At the same time, structural genomic consortia and the wide availability of synchrotron resources for the collection of diffraction data, led to an enormous output of solved protein crystal structures in the last decade (6). Nevertheless, it is still impossible to predict under which physical and biochemical conditions a given protein will produce well diffracting crystals, which is especially true for difficult targets, i.e. membrane proteins or oxygen sensitive proteins.

3.2.1 Phase diagrams

The physical process of protein crystallisation can be outlined by phase diagrams (figure 3.2). For successful crystallisation the aqueous protein solution must be brought into a supersaturated state, during which either nucleation (initial formation of microcrystals) or protein precipitation (denaturation) can occur. If crystal nuclei are formed, the protein concentration decreases and might enter the metastable zone, during which crystal growth from the formed nuclei might be observed. In the undersaturation zone of the diagram, crystallisation will never occur, whereas in the precipitation zone the protein undergoes rapid precipitation (7). Different techniques exist to achieve the desired transitions.
Figure 3.2: The phase diagram. For protein crystal growth, the protein first needs to form crystal nuclei (red zone), from which the growth of three-dimensional, individual crystals in the metastable zone (blue) can be achieved. VD = vapour diffusion; B = batch; D = dialysis; FID = free interface diffusion.
In the batch method a solution containing everything needed for crystallisation (supersaturated protein solution and precipitant) is sealed in an airtight environment to avoid evaporation (8).

The free interface diffusion (FID) method is usually performed in capillaries, in which the protein and precipitant solutions are placed on top of each other. This leads to a slow diffusion between the two solutions that generates a gradient, in which precipitation, nucleation and crystal growth can occur (9).

Crystallisation by dialysis is done by placing the protein solution at constant volume within a selectively permeable membrane that is surrounded by the precipitant solution. The protein concentration remains constant over the time course of the experiment and the precipitant molecules can diffuse through the membrane and might thus cause crystal nucleation (10).

In the vapour diffusion method the protein solution is mixed (usually in a drop-like fashion and in a 1:1 volumic ratio) with the precipitant solution and is then placed over an excess volume of precipitant solution (mother liquor) in a sealed environment. Slow equilibration will occur, during which water from the protein / precipitant solution in the dispensed drop will diffuse to the mother liquor, until the precipitant concentration is similar in both solutions. The vapour diffusion method can be performed as either sitting or hanging drop technique (9). Widely used is the sitting drop technique, as there are now different robotic systems available that can dispense \(\sim100\) nl drops on a 96 well scale within minutes, thus dramatically down-shifting the quantities needed for the testing of several hundred crystallisation conditions from commercially available sparse matrix screens (11; 12). Examples of such robots include the Mosquito® (TTP Labtech), Phoenix (Art Robbins Instruments) and HoneyBee (Digilab) systems.

### 3.2.2 Seeding techniques

The goal of every crystallisation experiment is to identify a condition that allows the reproducible growth of individual, stable and well diffracting crystals of the protein under investigation. Quite often the screening of many different crystallisation conditions on small sitting drops (e.g. 100 nl protein solution + 100 nl mother liquor), results in the identification of so called initial hits; the growth of crystals that in quantity or quality are not sufficient to serve the goal of the study. Reasons that might hamper their suitability can be size, shape, twinning (multiple lattices due to multiple nucleation sites in a single crystal) or diffraction to insufficient resolution. Several different seeding techniques do exist that might lead to the growth of crystals that can overcome these problems. For that, the crystal nuclei originated from the initial hit conditions can be transferred to other conditions, not necessarily related, in which they might grow into more suitable crystals. The nuclei might simply be streak seeded into already existing drops using a cat whisker or horse hair (13). Sparse matrix seeding is another possibility of crystal
nuclei transfer. Here, the crystals are resuspended in its respective mother liquor and crushed into microcrystals by the addition of a microseed bead (Molecular Dimensions) and vigorous vortexing. The microcrystals can then be used as nucleation seeds in another round of sparse matrix screening (14). Obviously, seeding does not always yield the desired results, or, even more frequently, no crystals are obtained at all during the screening of crystallisation conditions. The experimenter is then faced with a plethora of possibilities to overcome the initial hit barrier, which include (9):

- Changing of physical incubation conditions: ionic strength, pH, temperature, protein concentration, ...

- Altering the protein sequence: truncated versions, single amino acid substitutions, removal of affinity tag...

- Trying to crystallise a homolog version of the protein.

3.3 Physical properties of crystals

The unit cell is the smallest repetitive unit of the crystal, which can be described by three vectors $a$, $b$, $c$ and three angles $\alpha$, $\beta$, $\gamma$. The crystal is comprised of a three-dimensional parallel arrangement of identical unit cells (figure 3.3). Within the crystal it is always possible to get from one unit cell to the next by applying translational operations along the crystal axes $x$, $y$ and $z$.

Furthermore, the unit cell orientations with respect to each other within the crystal form another type of symmetry, the so called crystal lattice. Conventionally, the smallest possible unit cell is chosen, from which the entire crystal can be built, based on lattice symmetry operations. The eight unit cell corners are always lattice points in the crystal. If these are the only lattice points in the crystal, the lattice system is called
primitive \((P)\). In a one-face centered \((C)\) lattice system additional lattice points are at the centre of the surface opposite of the c axis, whereas in an all-face centered \((F)\) lattice system, additional lattice points lie on all surfaces of the unit cell. The lattice system is abbreviated with \(I\) (german for Innenzentriert) if an additional lattice point is located in the unit cell centre.

These additional lattice points give rise to additional crystallographic two-fold \((180^\circ)\), three-fold \((120^\circ)\), four-fold \((90^\circ)\) and six-fold \((60^\circ)\) symmetry axes (rotation), which are always oriented perpendicular to a two-dimensional unit cell plane \((10)\). With these included symmetry operators the asymmetric unit can be defined, which is the smallest volume of the crystal that can be rotated and translated to build up one unit cell using only the symmetry operators allowed by the crystallographic symmetry. The asymmetric unit can be comprised of one protein molecule or one subunit of a multimeric protein. However, it can also be more than one. Seven different crystal classes exist with different minimal symmetry elements and restrictions on their respective lattice constants. These geometric constraints result in the formation of crystal classes with different lattice systems (Bravais lattices), point groups and space groups. Space groups are abbreviated by their respective lattice system \((P, I, C \text{ or } F)\) followed by numbers that describe symmetry operations e.g. screw axes. If these operations are applied to the respective unit cell, the cell will, from a geometric point of view, remain the same. 230 space groups are mathematically possible, but biological macromolecules only comprise 65 of these groups, because of their inherent chirality that does not allow mirror symmetry. Table 3.1 gives an overview of the basic geometric properties of these 65 biological relevant space groups \((10)\).

### 3.4 Diffraction, Bragg’s law and crystal lattices

If X-rays hit a crystal, all repetitive units (e.g. all unit cells) contribute to the scattering. However, only when the waves of the scattered photons are in phase a detectable diffraction beam arises. Being in phase means that they have either the exact same path lengths, or that their difference in path length equals the wavelength multiplied by a full integer. If they are out of phase they cancel each other out and no diffraction is observed. Parallel rays that are reflected from different planes in the crystal will have different path lengths, whereas they have always the same path lengths, if they are diffracting from the same plane. In figure 3.4 the difference in path lengths between rays A and B is twice the distance of \(l\). They can still give rise to in phase diffraction, if that distance difference is a full integer multiple of the wavelength \(\lambda\) of the incoming X-ray beam, which can be shown mathematically:

The blue coloured triangle around the lattice plane distance \(d\) and the length \(l\) has an upper angle of \(\theta\), one 90° angle and one 90° - \(\theta\). This triangle is a right angled one
## Table 3.1: The 65 space groups relevant in macromolecular crystallography

<table>
<thead>
<tr>
<th>Space Group</th>
<th>No.</th>
<th>Class</th>
<th>Point Group</th>
<th>Lattice</th>
<th>Symmetry</th>
<th>Bravais</th>
<th>Crystal</th>
</tr>
</thead>
<tbody>
<tr>
<td>P432, P432, P432, P432, P432, P432, P432</td>
<td>432</td>
<td>Tetragonal</td>
<td>P4</td>
<td>F, I</td>
<td>r, a</td>
<td>4 fold</td>
<td></td>
</tr>
<tr>
<td>P622, P622, P622, P622, P622, P622, P622</td>
<td>622</td>
<td>Hexagonal</td>
<td>P6</td>
<td>F</td>
<td>r, a</td>
<td>6 fold</td>
<td></td>
</tr>
<tr>
<td>P23, P213, P213, P213, P213</td>
<td>23</td>
<td>Cubic</td>
<td>P</td>
<td>F</td>
<td>r, a</td>
<td>No symmetry</td>
<td></td>
</tr>
<tr>
<td>C23, C23, C23, C23, C23, C23</td>
<td>23</td>
<td>Cubic</td>
<td>C</td>
<td>F</td>
<td>r, a</td>
<td>3 fold</td>
<td></td>
</tr>
<tr>
<td>P4, P4, P4, P4, P4, P4, P4</td>
<td>4</td>
<td>Tetragonal</td>
<td>P4</td>
<td>F</td>
<td>r, a</td>
<td>4 fold</td>
<td></td>
</tr>
<tr>
<td>C4, C4, C4, C4, C4, C4, C4</td>
<td>4</td>
<td>Tetragonal</td>
<td>C4</td>
<td>F</td>
<td>r, a</td>
<td>4 fold</td>
<td></td>
</tr>
<tr>
<td>P21, P21, P21</td>
<td>21</td>
<td>Trigonal</td>
<td>P2</td>
<td>F</td>
<td>r, a</td>
<td>3 fold</td>
<td></td>
</tr>
<tr>
<td>C21, C21, C21</td>
<td>21</td>
<td>Trigonal</td>
<td>C2</td>
<td>F</td>
<td>r, a</td>
<td>3 fold</td>
<td></td>
</tr>
<tr>
<td>P1, P1, P1, P1, P1, P1, P1</td>
<td>1</td>
<td>Triclinic</td>
<td>P</td>
<td>F</td>
<td>r, a</td>
<td>No symmetry</td>
<td></td>
</tr>
<tr>
<td>C1, C1, C1, C1, C1, C1, C1</td>
<td>1</td>
<td>Triclinic</td>
<td>C</td>
<td>F</td>
<td>r, a</td>
<td>No symmetry</td>
<td></td>
</tr>
</tbody>
</table>

### Notes:
- The space group symbol is a combination of the point group symbol, the lattice group symbol, and the Bravais lattice symbol.
- The term "symmetry" refers to the symmetry operations that can be applied to the unit cell.
- The term "Bravais" refers to the lattice type, either primitive (P), body-centered (B), or face-centered (F).
- The term "crystal" indicates the crystal system to which the space group belongs.

### Table 3.1 continued...

<table>
<thead>
<tr>
<th>Space Group</th>
<th>No.</th>
<th>Class</th>
<th>Point Group</th>
<th>Lattice</th>
<th>Symmetry</th>
<th>Bravais</th>
<th>Crystal</th>
</tr>
</thead>
<tbody>
<tr>
<td>P432, P432, P432, P432, P432, P432, P432</td>
<td>432</td>
<td>Tetragonal</td>
<td>P4</td>
<td>F, I</td>
<td>r, a</td>
<td>4 fold</td>
<td></td>
</tr>
<tr>
<td>P622, P622, P622, P622, P622, P622, P622</td>
<td>622</td>
<td>Hexagonal</td>
<td>P6</td>
<td>F</td>
<td>r, a</td>
<td>6 fold</td>
<td></td>
</tr>
<tr>
<td>P23, P213, P213, P213, P213</td>
<td>23</td>
<td>Cubic</td>
<td>P</td>
<td>F</td>
<td>r, a</td>
<td>No symmetry</td>
<td></td>
</tr>
<tr>
<td>C23, C23, C23, C23, C23, C23</td>
<td>23</td>
<td>Cubic</td>
<td>C</td>
<td>F</td>
<td>r, a</td>
<td>3 fold</td>
<td></td>
</tr>
<tr>
<td>P4, P4, P4, P4, P4, P4, P4</td>
<td>4</td>
<td>Tetragonal</td>
<td>P4</td>
<td>F</td>
<td>r, a</td>
<td>4 fold</td>
<td></td>
</tr>
<tr>
<td>C4, C4, C4, C4, C4, C4, C4</td>
<td>4</td>
<td>Tetragonal</td>
<td>C4</td>
<td>F</td>
<td>r, a</td>
<td>4 fold</td>
<td></td>
</tr>
<tr>
<td>P21, P21, P21</td>
<td>21</td>
<td>Trigonal</td>
<td>P2</td>
<td>F</td>
<td>r, a</td>
<td>3 fold</td>
<td></td>
</tr>
<tr>
<td>C21, C21, C21</td>
<td>21</td>
<td>Trigonal</td>
<td>C2</td>
<td>F</td>
<td>r, a</td>
<td>3 fold</td>
<td></td>
</tr>
<tr>
<td>P1, P1, P1, P1, P1, P1, P1</td>
<td>1</td>
<td>Triclinic</td>
<td>P</td>
<td>F</td>
<td>r, a</td>
<td>No symmetry</td>
<td></td>
</tr>
<tr>
<td>C1, C1, C1, C1, C1, C1, C1</td>
<td>1</td>
<td>Triclinic</td>
<td>C</td>
<td>F</td>
<td>r, a</td>
<td>No symmetry</td>
<td></td>
</tr>
</tbody>
</table>
and equation 3.1 states:

$$\sin \theta = \frac{l}{d} \iff l = d \cdot \sin \theta$$  \hspace{1cm} (3.1)

In order to see scattering in phase the condition $2 \cdot l = n \cdot \lambda$ has to be fulfilled. If $l$ is substituted with $n \cdot \lambda$ from equation 3.1, it becomes:

$$n \cdot \lambda = 2 \cdot d \cdot \sin \theta$$  \hspace{1cm} (3.2)

Equation 3.2 is known as Bragg’s law of diffraction. The mentioned planes are called lattice planes or Bragg planes and act as mirrors for the incoming X-ray beams. They always intersect with crystal lattice points. Again, each unit cell of the crystal contains the same lattice planes and all lattice planes with the same Miller indices $hkl$ from all unit cells in the crystal will contribute to the same reflection. The $hkl$ numbers are always full integers and describe how often the lattice points from the unit cells are touched by the respective $hkl$ plane (5). This is illustrated in figure 3.5 in a two-dimensional example. The red $(1, 1)$ plane crosses a lattice point once for each length of $a$ and once for each length of $b$. In the case of the green $(1, 2)$ plane the length $a$ is crossed once and length $b$ twice.
Figure 3.5: Lattice planes and Miller indices. Lattice points in this illustrated two-dimensional system are black dots, the unit cell is shown in blue with sides $a$ and $b$. The $(1, 1)$ plane is shown in red and the $(1, 2)$ plane is shown in green.
3.5 Reciprocal space and data collection

There is an inverse (reciprocal) relationship between lattice planes in the unit cell and their corresponding diffraction spots on the detector surface. The shorter the distance between two planes in the unit cell, the further apart are their diffraction spots. Unit cell dimensions are defined as real space dimensions whereas the diffraction spots occur according to the lattice in the so-called reciprocal space. Hence, it is possible to calculate unit cell dimensions from the distances between the spots in the diffraction pattern.

The Fourier transformation connects the wave functions, which are observed in the reciprocal lattice, with a complex function in the real space lattice. During diffraction data collection the crystal is commonly rotated about a single axis within the X-ray beam with angular increments of 0.1 - 1° and a diffraction image is recorded for each rotation wedge. This collection strategy is termed rotation method (15). The major reason why the crystal has to be rotated in the beam is that for each orientation only a fraction of all possible reflections can be recorded. A construction of the so called Ewald’s sphere in reciprocal space helps to understand this phenomenon and is illustrated in a two-dimensional example in figure 3.6. The grey grid is a two-dimensional reciprocal lattice with axes $a^\ast$ and $b^\ast$. An X-ray beam (X) is crossing the reciprocal lattice from B to O (origin of the reciprocal lattice, which is also a real lattice point) via going through the centre of the reciprocal space (C), with a radius of $1/\lambda$. In order to observe a reflection from the lattice point A1, two conditions need to be fulfilled:

- A1 needs to be on the surface of the reciprocal space (Ewald’s sphere) in order to yield a right triangle around the points B, A1 and O.
- Both O and A1 must be lattice points in reciprocal space so that their distance is inverse to $d_{hkl}$, that is: $OA1 = 1/d_{hkl}$.

If these conditions are fulfilled Bragg’s law holds true, which can be seen in equation 3.3:

$$\sin\theta = \frac{OA1}{BO} = \frac{OA1}{2/\lambda} \iff 2 \cdot \frac{1}{OA1} \cdot \sin\theta = \lambda \quad (3.3)$$

Substituting $1/OA1$ with $d_{hkl}$ yields:

$$2 \cdot d_{hkl} \cdot \sin\theta = \lambda \quad (3.4)$$

which is Bragg’s law, for $n = 1$. The diffracted beam diverges from the point C by the angle of $2\theta$. So each time (during which the crystal is in the X-ray beam rotated over the origin O of the reciprocal space) a reciprocal lattice point comes in contact with the Ewald’s sphere surface, a reflection occurs, according to Bragg’s law (5).

Data collection is the last experimental step in protein crystallography during which the crystals inevitably get destroyed due to X-ray radiation damage. Radiation damage
Figure 3.6: Ewald’s sphere and diffraction in reciprocal space. O = origin of reciprocal lattice; C = centre of reciprocal space, which is the centre of the Ewald’s sphere; \( a^\ast, b^\ast \) = reciprocal lattice axes. Top: The lattice point A1 (red) resides on the surface of the Ewald’s sphere and thus results in the reflection R1. Bottom: After rotating the crystal around the point O, the lattice point A2 (teal) is located on the surface of the Ewald’s sphere, giving rise to reflection R2.
arises when X-rays that interact with crystal atoms eject lower shell electrons (photoelectrons), which in turn move through the crystal and result in the formation of radicals that decrease the overall stability of the crystal (16). The most important factor that needs to be considered is the completeness of collected reflection data and greatly depends on the crystal symmetry and detector position (15). For the collection of diffraction data an adequate X-ray source is required.

Two general sources are used for the analysis of protein crystals, which are in-house X-ray sources and synchrotron X-ray radiation. In-house sources usually generate X-ray radiation from a rotating copper anode. The electrons on synchrotron sources are accelerated to nearly light speed and transferred into a near-circular vacuum channel (storage ring), in which they can travel for several hours at relative constant energy. The storage ring is surrounded by bending magnets, which keep the electrons in a circular path. When the electrons change their travel direction (occurs at key points in the synchrotron), they release white light, the wavelength spectrum of which ranges from gamma rays to visible light. For macromolecular crystallography the X-rays are selected from this spectrum by the use of specific optics (i.e. wigglers and undulators) (10). Main advantages of synchrotron radiation over the rotating anode radiation are:

- Tunable wavelength, that allows the recording of anomalous signals.
- Higher X-ray intensities.
- Highly focussed and non-divergent beam.
- Much faster data collection.

The majority of protein crystals are now routinely placed in a stream of gaseous nitrogen at 100 Kelvin, which increases the life-time of the crystal in the X-ray beam (the mobility of the above mentioned radicals is decreased) without a loss in diffraction quality. Thus, cryocrystallography allows the recording of complete data sets from single crystals (17). The crystals, however, need to be cryoprotected, before they are transferred from their growth environment into a cryogenic environment in order to avoid the inclusion of ice in the crystal which has severe negative impacts on the recorded diffraction images (16). Cryoprotection can be achieved by the addition of e.g. glycerol or small molecular weight polyethylene glycol (18). Ideally, it is possible to directly grow protein crystals in a cryoprotected precipitant solution. In that respect, complete sparse matrix crystallisation screens have been developed that exclusively contain conditions that provide an cryoprotected environment (12).

### 3.6 Solving the phase problem

During the recording of diffraction data sets from a crystal, the photon intensities in reciprocal space are measured when the diffracted photons hit the detector. Each spot
Chapter 3. Methods

Cobalamin-dependent enzyme systems

on the diffraction pattern arises from all photons that are diffracted from the same
$hkl$ lattice planes of all unit cells during a specific position of the crystal in the X-ray
beam. Because of Bragg’s law all of these photons have the same angular offset. From
the photon intensities that describe each spot one can calculate the amplitudes of their
Corresponding waves, but the phase value of the respective waves, scattered from the
different atoms in the crystal, is lost. This lack of information is generally attributed as
the crystallographic **phase problem** (19).

However, this information is crucial for calculating the electron density in the crystal
for each $hkl$ lattice plane from its respective diffraction spot.

The diffracted waves are a Fourier synthesis of the electron density in the unit cell of
the crystal. In turn, the electron density, which can be described as a three-dimensional
map, can be calculated by an inverse Fourier transform (equation 3.5) from the diffracted
waves. The electron density map is therefore a transformation of reciprocal lattice point
coordinates back to real space coordinates. Into these real space coordinates (electron
density map) the three-dimensional atom coordinates $(x, y, z)$ of the crystallised protein
have to be modelled.

$$p(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F_{hkl} e^{-2\pi i (hx + ky + lz)}$$  \hspace{1cm} (3.5)

with $V = \text{volume of unit cell}$, $F_{hkl} = \text{structure factor amplitudes}$ and $2\pi i = \text{phase angles}$.

Direct phasing of diffraction data is restricted to both atomic resolution data sets
(higher resolution than 1.2 Å) and molecules of limited size, and is therefore usually
not applicable for solving protein crystal structures. Several different methods exist to
calculate the phases, and are shortly introduced within the next paragraphs.

### 3.6.1 Molecular replacement

In molecular replacement the phases are initially obtained from a protein (model), which
amino acid sequence is similar to the target protein and which structure is already solved
(20). The use of molecular replacement becomes more and more frequent due to the
constant increase of deposited macromolecular structures. The underlying principle of
molecular replacement is a trial and error approach based on translations and rotations.
All possible translations and rotations of the model protein are tested in the lattice of
the unknown target protein with the goal to find the position for which the predicted
diffraction is in the best agreement with the measured diffraction of the target structure.
Three rotational angles $(\alpha, \beta, \gamma)$ and three translational vectors $(t_x, t_y, t_z)$ define
the position of a molecule and these six parameters need to be screened during molecular
replacement. Computer programs usually divide this six-dimensional search into two
three-dimensional searches, in order to speed up the calculation. The characteristics of
the Patterson function (equation 3.6) allow the search to be split up into a rotational and translational part.

\[ P(u) = \sum_{h} |F_h|^2 e^{-2\pi ihu} \]  

(3.6)

with \( u \) being an intramolecular vector; \( h \) = individual reflections; \( i \) = imaginary number; \( F_h \) = structure factor amplitudes.

The Patterson function is a Fourier transformation of the structure factor amplitudes multiplied with itself (\( |F_h|^2 \)), neglecting their corresponding phases (they are all set to zero). It does so by treating the interatomic distances within the molecule as three-dimensional vectors. The function contains all intramolecular vectors including the vectors from every atom towards itself. In words, the Patterson function is a convolution of the scattering function with itself, that creates \( N^2 \) vectors from \( N \) original atoms. The origin of all vectors is the same for all atoms of the molecule. Visualisation of such a vector map in the simple case of a benzene molecule, aids in the understanding of the Patterson function, as shown in figure 3.7. Benzene consists of six carbon atoms \((N = 6)\), and thus its corresponding Patterson map will consist of 36 vectors. The prominent central black spot is the origin of the Patterson map and has a relative weight of six, arising from the six vectors that project each atom of the benzene ring onto itself with a vector length of zero. The twelve additional blue spots have a relative weight of two, which arise from vectors between adjacent atoms (e.g. 1 to 2) and second-nearest neighbour atoms (e.g. 1 to 3), as indicated by the blue dashed lines in figure 3.7. Finally, the six red spots with a relative weight of one, arise from the vectors between diagonal atoms in the benzene ring (e.g. 1 to 4), red dotted line. The rotation function calculates the Patterson map for both the model and the new molecule at each rotation and compares them with each other within the radius of the molecule. After the best rotation angles of the model have been found these angles are subject to a translation function. The translation function follows the same principle as the rotation function, that is, a Patterson function is calculated for each translational position of the model protein which is then compared to the Patterson function of the corresponding target protein. Eventually, the best solutions for the rotational and translational function are known, which then need to be applied to all atoms of the model protein (21). From there it is possible to calculate an electron density map with the calculated phase angles \((\alpha_{hkl})_{calc}\) from the rotated and translated model structure, combined with the observed structure factor amplitudes \((|F_{hkl}|)_{obs}\) from the target structure. With that map as a starting point, the structure solving process becomes a crystallographic refinement problem. In a later section the process of structure refinement is discussed in more detail. Another approach to perform molecular replacement calculations is by using maximum likelihood functions rather than correlation functions to score the various rotations and translations. Here, all theoretical rotations and translations are evaluated statistically in reciprocal space.
and the most likely ones are in the end selected (19). Many different programs exist that can carry out molecular replacement calculations, e.g. AMoRe, MOLREP and Phaser (22; 23; 24). Even more automated molecular replacement pipelines do exist nowadays, which provide a more or less one-button approach towards solving a crystal structure by molecular replacement. These include MrBUMP and BALBES (25; 26). In addition to the actual process of molecular replacement, these programs also perform the steps of selecting and preparing the most suitable model structure from the PDB. All mentioned programs can be executed via the Collaborative Computational Project No. 4 (CCP4) crystallographic software suite (27).

![Figure 3.7: Graphical representation of the Patterson function of benzene: Left: benzene molecule; Right: Patterson function of benzene. The figure has been adopted from (4).](image)

### 3.6.2 Isomorphous replacement

In the isomorphous replacement method, crystals are soaked with heavy-atom solutions and the differences in the diffraction patterns, that originate from the protein bound heavy-atoms, are used to calculate the location of these heavy-atoms in the unit cell. This is only possible if the unit cells and protein structures of the native and the derivative crystals are identical, that is they are isomorphous. Popular heavy-atom compounds contain mercury, platinum or gold, that can bind to specific side chain residues in the protein, e.g. cysteines, histidines or methionines (10). Single isomorphous replacement (SIR) makes use of a single heavy metal soak, whereas with multiple isomorphous replacement (MIR), more than one heavy atom are soaked into the crystal. Heavy-atoms contain many electrons (relative to standard protein atoms) and are thus strong scatterers, that lead to observable differences in a number of reflection intensities in the derivative crystal compared to the same reflections arising from the native crystal. The native reflection amplitudes are termed $|F_P|$ and the heavy-atom derivatives $|F_{PH}|$. Hence the reflection amplitude that is exclusively arising from the heavy-atom is $|F_H|$. 
Diffraction contributions can be considered as additive vectors, equation 3.7 states:

\[ F_{PH} = F_P + F_H \iff F_H = F_{PH} - F_P \] (3.7)

So called Argand or vector diagrams (figure 3.8) visualise the relationships between the structure factors of the native, heavy-atom and heavy-atom derivative (19). If only the \( F_H \) reflections in the diffraction pattern are considered, i.e. after all contributions from the protein are subtracted, the diffraction pattern contains only the information about the location of the heavy-atoms in the unit cell (if the structure is isomorphous). From the measured reflection intensity of the \( I_H \) the length of the vector \( F_H \) is known, but not its phase angle. However, as there are only a few heavy-atoms that build the heavy-atom structure, it is usually easy to locate the heavy-atom positions in the unit cell by using the already described Patterson function. With the derived phase angle \( \alpha_H \) of the heavy-atom reflections (\( F_H \)), it is possible to calculate the phase angles \( \alpha_P \) using the cosine rule (equation 3.8):

\[ F_{PH}^2 = F_P^2 + F_H^2 + 2 \cdot F_P F_H \cdot \cos(\alpha_P - \alpha_H) \iff \alpha_P = \alpha_H \pm \cos^{-1} \frac{F_{PH}^2 - F_P^2 - F_H^2}{2 \cdot F_P F_H} \] (3.8)

This calculation is repetitively performed for all \( hkl \) planes to derive their respective phase angles. Once that is done the desired electron density map (see above, equation 3.5) can be computed and the protein atoms can be modelled into this map. In a single isomorphous replacement (SIR) experiment a phase ambiguity problem arises from the cosine rule, which is best shown in a Harker construction (figure 3.9 Left). The problem in SIR is that two possible protein phase angles (\( \alpha_P \)) exist, at the points where the circle from the native structure factor (\( F_P \), teal circle) and the heavy-atom derivative (\( F_{PH} \), blue circle) intersect. This phase ambiguity problem is circumvented in multiple isomorphous replacement (MIR, figure 3.9 Right). Here, two heavy-atom derivatives (\( F_{PH1} \) and \( F_{PH2} \)) lead to only one possible protein phase angle (\( \alpha_P \)), because the circles \( F_P \) (teal), \( F_{PH1} \) (blue) and \( F_{PH2} \) (brown) intersect only at one position. In reality, experimental errors do occur that can hamper the process of solving the phase problem using isomorphous replacement, including:

- Non-isomorphism: Unit cell parameters or protein structures differ between native and derivative crystal.

- Errors in intensities measured for both \( F_P \) and \( F_{PH} \), leading to larger errors for \( F_H \).

- Inaccurate heavy atom model structure.
Figure 3.8: Structure factors described as complex vectors. Left: General structure factor ($F_{hkl}$) representation as complex number on a two dimensional plane. The length of the vector is proportional to the amplitude of its corresponding reflection and the angle $\alpha$ describes its phase. Right: Argand diagram for the native, heavy-atom and heavy-atom derivative structure factors that correspond to the reflection of a specific $hkl$ plane from the crystal.

Figure 3.9: Harker construction for single and multiple isomorphous replacement. Left: Single isomorphous replacement. The green circle and arrow correspond to the native protein; the blue one to the derivative protein. Two possible solutions exist for $\alpha_P$. Right: Multiple isomorphous replacement, resulting in the unambiguous assignment of the $\alpha_P$, as there is only position at which the circles around the origins from $F_P$ (teal), $F_{PH1}$ (blue) and $F_{PH2}$ (brown) intersect.
### 3.6.3 Anomalous scattering

Anomalous scattering is a crystallographic phenomenon that can be employed to solve the phase problem. It takes advantage of the intrinsic properties of some elements (e.g. iron, cobalt or selenium) to absorb X-ray energy of a specific wavelength (5). All elements absorb X-rays at specific wavelengths, for crystallographic purposes however, the signal intensity becomes only useful from sulphur onwards. Such absorption results in the break of Friedel’s law, which states: $|F_{hkl}| = |F_{-h,-k,-l}|$. This relationship is in fact not valid anymore, if the before mentioned absorption occurs. Different elements contain specific and wavelength dependent absorption edges, at which their X-ray absorbance significantly decreases, resulting in anomalous scattering. These anomalous scattering factors ($f''$) are 90° advanced in phase and thus lead to the described break of Friedel’s law, which is illustrated in figure 3.10 (19).

It shows that although $|F_P^+| = |F_P^-|$, and $|F_H^+| = |F_H^-|$, $|F_{PH}^+| = |F_{PH}^-|$, because of the differences between $f''^+$ and $f''^-$. $|F_P|$, $|F_H|$ and $|F_{PH}|$ are structure factor amplitudes of a native data reflection, of a derivative data reflection and of their combined reflection, respectively. + and - refer to reflection pairs that are related by a centre of symmetry through the origin of the reciprocal lattice and that contain the same structure factor amplitudes (vector lengths), but have different phase angles.

Similar to isomorphous replacement phasing, either the Patterson function or direct methods are used to determine the position of the anomalous scatterer in the unit cell, from which the subsequent calculation of the protein phases ($\alpha_{P'\delta}$) is possible. Another similarity to the described isomorphous replacement technique is the possibility to use either single or multiple anomalous scattering factors. SIRAS = single isomorphous replacement with anomalous scattering; MIRAS = multiple isomorphous replacement with anomalous scattering. An advantage is that multiple datasets can be recorded from the same crystal at different wavelengths, so that non-isomorphism is not a problem. If the target protein does not contain intrinsic anomalous scattering elements, it is often possible to grow seleno-methionine containing versions of the protein, in which seleno atoms replace the sulphur atoms in the methionine residues, without affecting the structure or function of the protein.

### 3.6.4 MAD and SAD

Multiwavelength anomalous diffraction (MAD) and single wavelength anomalous diffraction (SAD) can overcome problems that are associated with isomorphous replacement, e.g. non-isomorphism and conformational changes of the protein. In MAD experiments multiple data sets (often three) are collected from the same crystal using different wavelengths (4). Chosen wavelengths are at the absorption peak of the anomalous scatterer, at the point of inflection on the absorption curve and at a third remote wavelength. A Harker construction as described above is then used to calculate the position of the
anomalous scatterer.

In SAD experiments only one wavelength is used. The phase ambiguity (similar to single isomorphous replacement problems) has to be solved through density-modification. In both MAD and SAD it is beneficial if collected data sets are of high redundancy (19).

3.7 Model building, structure refinement and validation

The placement of the polypeptide chain(s) into the calculated electron density map can be manually performed in graphics programs like O or Coot (28), that allow the three-dimensional representation of electron density maps and atomic models. The higher resolution the diffraction data are, the better defined will the map be, and hence, the easier becomes the task to model the protein correctly into the map. Recently, several programs have been developed that aim towards an automated model building and refinement approach. These are e.g. Buccaneer (29) or ARP/wARP (30) and work very well for high resolution data sets.

One of the last computation processes involves refinement steps with the aim to optimise the agreement between modelled structure and collected diffraction data. This includes also the addition of water and other solvent molecules that are visible in the electron density map. Structure refinement can be carried out using REFMAC5 or the PHENIX suite (31; 32). The progress of the iterative model building and refinement cycles is often assessed using R-factors, that measure the agreement between observed and calculated structure-factor amplitudes (equation 3.9). Calculated numbers are usually presented as a set of $R_{\text{work}}$ and $R_{\text{free}}$, where $R_{\text{free}}$ is a cross-validation R-factor for a
test set of 5% of the reflections that have been excluded from model refinement in order to avoid overfitting of the data.

\[ R_{factor} = \frac{\sum_{hkl} |F_{obs} - |F_{calc}|}{\sum_{hkl} |F_{obs}|} \]  
(3.9)

The closer the \( R_{factor} \) approaches 0%, the more identical is the calculated structure with the model, whereas at 60% the calculated model would describe a random solution. Because of the exclusion from refinement, the \( R_{free} \)-factor will always be a few percentages higher than the \( R_{work} \)-factor.

A rather new statistical value is the CC\(_{1/2}\) value (33), which correlates the observed data set with the true signal and which is derived from the correlation coefficient between the calculated and observed structure-factor amplitudes (equation 3.10):

\[ CC = \frac{\langle F_{obs}F_{calc} \rangle - \langle F_{obs} \rangle \langle F_{calc} \rangle}{\sqrt{\langle F_{obs}^2 \rangle - \langle F_{obs} \rangle^2} \langle F_{calc}^2 \rangle - \langle F_{calc} \rangle^2}} \]  
(3.10)

Average B-factors (temperature factors) are expressed in Å\(^2\) and describe the averaged thermal motion of each atom of the model in three-dimensional space. Usually, average B-factors decrease with higher resolution data sets. Another important data validation aspect is the use of geometric assessments like Ramachandran statistics and root-mean-square deviations (rmsd), that evaluate the protein backbone dihedral angles \( \phi \) and \( \psi \), and atom bond lengths and angles, respectively. Again, programs like PROCHECK and SFCHECK exist within the CCP4, that combine different structure validation tools into a single program (34). MolProbity is another popular structure validation server that analyses both local and global atom contacts in the refined model (35). Furthermore the Twilight software can be used to analyse annotated protein-ligand complexes in the PDB (36). In the end it is always the combination of several different parameters that need to be taken into consideration when the quality of a refined protein crystal structure is judged.

### 3.8 References


Cobalamin-dependent enzyme systems


Chapter 4

Heterologous expression, purification and cofactor reconstitution of the reductive dehalogenase PceA from *Dehalobacter restrictus*

Author’s: **Hanno Sjuts**, Karl Fisher, Mark S. Dunstan, Stephen E. Rigby and David Leys

*Affiliation: Manchester Institute of Biotechnology, Faculty of Life Sciences, The University of Manchester, M1 7DN, UK*

4.1 Abstract

Organohalide respiration is used by a limited set of anaerobic bacteria to derive energy from the reduction of halogenated organic compounds. The enzymes that catalyze the reductive dehalogenation reaction, the reductive dehalogenases, represent a novel and distinct class of cobalamin and Fe-S cluster dependent enzymes. Until now, biochemical studies on reductive dehalogenases have been hampered by the lack of a reliable protein source. Here we present an efficient and robust heterologous production system for the reductive dehalogenase PceA from *Dehalobacter restrictus*. Large quantities of Strep-tagged PceA fused to a cold-shock induced trigger factor could be obtained from *Escherichia coli*. The recombinant enzyme was conveniently purified in milligram quantities under anaerobic conditions by StrepTactin affinity chromatography, and the trigger factor could be removed through limited proteolysis. Characterization of the purified PceA by UV-Vis and electron paramagnetic resonance (EPR) spectroscopy reveal that the recombinant protein binds methylcobalamin in the base-on form after proteolytic cleavage of the trigger factor, and that 4Fe-4S clusters can be chemically reconstituted under anoxic conditions. This study demonstrates a novel PceA production platform that allows further study of this new enzyme class.

**Keywords:** Organohalide respiration ◆ PceA ◆ Trigger factor ◆ EPR ◆ Cobalamin ◆ Fe-S cluster
4.2 Introduction

In the 20th century, large quantities of halogenated organic compounds (organohalides) were released to the environment by industrial applications. Two such organohalides are tetrachloroethene (also known as perchloroethylene, PCE), widely used in the dry cleaning of fabrics and an excellent solvent for organic material and trichloroethene (TCE), a volatile anesthetic. The highly toxic and potential carcinogenic properties of these compounds and their stability were only later widely documented. In fact, their relative stability has made them amongst the most common groundwater and soil contaminants to be found (1; 2). Chemical clean up of these compounds has proved very difficult but interestingly a variety of anaerobic bacteria are able to detoxify organohalides during a process recently termed organohalide respiration (3). In organohalide respiration, the organohalides serve as the terminal electron acceptor and are thus reductively dehalogenated (3; 4; 5; 6). Bacterial species that are able to detoxify organohalides belong to both Gram-positive and Gram-negative groups and are usually strictly anaerobic (7). The key enzyme that catalyzes the terminal electron transfer step leading to dechlorination is the reductive dehalogenase (RdhA) (8; 9). Recent genome sequences from organohalide respiring species have revealed some to contain up to 32 RdhA genes (10). RdhAs constitute a new class of cobalamin and 4Fe-4S dependent enzymes (11; 12). Several RdhA enzymes have been characterized in terms of substrate specificity, cofactor content and steady-state kinetics. These studies represent a considerable undertaking since the host cells grow slowly, do not reach high cell density and require the presence of large amounts of organohalides for protein expression. With one exception (13), all RdhAs are monomeric (ranging from 35 to 65 kDa) and contain one cobalamin and two Fe-S clusters (9). Arguably, the most extensive studied RdhAs are PceA from Dehalobacter restrictus (14; 15), CprA from Desulfitobacterium dehalogenans (16) and PceA from Sulfurospirillum multivorans (17). EPR studies have indicated that the cobalamin is bound in the base-off form, and established the presence of two 4Fe-4S clusters or one 4Fe-4S and one 3Fe-4S cluster. The enzymes have been reported to be oxygen-sensitive. In the case of PceA from S. multivorans the structure of the extracted cobalamin has been identified as norpseudocobalamin (18). Due to the presence of a TAT signal sequence, it is assumed that RdhAs are transported into the periplasm (19). Upon translocation, they are thought to associate with the membrane where they are anchored via a small transmembrane protein (RdhB) (20). The structure-function relationships for RdhA remain uncharacterized, as no structural model is available (21). It is therefore of great interest to develop a recombinant RdhA expression and purification system, which can serve to facilitate further study. Hitherto, efforts aimed at achieving this have met with little success (22; 23). Here we present the first successful recombinant Escherichia coli production system, using the ~61 kDa reductive dehalogenase PceA from D. restrictus. Extensive screening of solubility tags and expression trials has revealed that a soluble
form of PceA is obtained only when fused to the *E. coli* trigger factor (TF). Utilizing a C-terminal Strep-Tag II fused to TF-PceA, the protein could be recovered to > 95% purity in a single affinity purification step. We were able to demonstrate that recombinant PceA can bind methylcobalamin and that the 4Fe-4S clusters can be fully chemically reconstituted under strict anaerobic conditions.

### 4.3 Materials and methods

#### 4.3.1 Construct design

The codon optimized *pceA* gene sequence, including its TAT recognition sequence from *D. restrictus* (GenBank accession number Q8GJ27), was synthesized by Eurofins MWG Operon. PceA was subsequently cloned into several different expression plasmids by the ligation independent In-Fusion™ technology [TaKaRa Bio Inc.]. Expression vectors were derived from the pOPIN series (Oxford protein production facility, OPPF) containing either a N-terminal His$_6$-tag (pOPINF), SUMO tag (pOPINS) or MBP tag (pOPINM) [OPPF Oxford, UK]. The *pceA* gene was inserted into these vectors using the restriction sites HindIII and KpnI (New England Biolabs). Similarly *pceA* was cloned into the pET50b(+) plasmid (Novagen) and the pCOLD TF plasmid (TaKaRa Bio Inc.). All constructs are under the control of an IPTG inducible *lacZ* operon. The pCOLD vector additionally contains a cold shock promoter (*cspA*), allowing induction of gene expression at low temperatures (15 °C) (24). The Strep-Tag II was inserted downstream of *pceA* in the pCOLD TF vector using conventional cloning techniques.

#### 4.3.2 Small-scale solubility screen

In all cases BL21 (DE3) cells (Merck) were transformed and used for expression trials in 100 mL cultures. Several different growth media (LB, TB, 2xYT and autoinduction medium; all purchased from Formedium) were tested. Typically proteins were overexpressed (except for the autoinduction media) with 0.1 mM IPTG (Sigma) at temperatures of 25 °C, 18 °C or 15 °C (in case of the pCOLD TF) for 12-16 h. The 100 mL cultures were inoculated with 1% (v/v) of an overnight culture and grown at 37 °C, until mid-log phase, at which point the temperature was reduced and protein expression induced with IPTG. Soluble and insoluble fractions were prepared with BugBuster (Novagen) according to the manufacturer instructions and analyzed by SDS-PAGE.

#### 4.3.3 Large-scale expression of the PceA from the pCOLD TF system

Cultures (usually 12 or 24 L of LB medium) were inoculated with 1% (v/v) of an overnight culture. Cells were grown at 37 °C until mid log phase (0.6 OD$_{600}$ nm). Overexpression of the trigger factor PceA fusion protein was induced at 15 °C, with the
addition of 0.1 mM IPTG and 0.1 mM FeCl₃. After 12-16 h cells were harvested by centrifugation and stored at -20 °C until needed.

### 4.3.4 Anaerobic protein purification and cleavage of trigger factor with HRV 3C protease

Purification of TF-PceA was carried out under strict anaerobic conditions, with all buffers made anaerobic by sparging with nitrogen gas before use. Cell pellets were resuspended in buffer A (50 mM Hepes pH 8.0, 150 mM NaCl and 10% glycerol), supplemented with lysozyme, DNase, RNase (Sigma) and Complete EDTA free protease inhibitor cocktail tablets (Roche Diagnostics) and stirred for 30 min on ice under nitrogen atmosphere. Cells were broken by passage through a French™ pressure cell press (Thermo IEC) under a constant nitrogen gas stream. Cell membranes were removed by ultracentrifugation at 98 k x g for 60 min at 4 °C. All subsequent purification steps were carried out in an anaerobic glove box (Belle Technologies, O₂ < 1 ppm). Soluble crude extract was mixed with 10 mL anaerobic StrepTactin Sepharose (IBA) resin resuspended in buffer A for 4 h at 4 °C. The resin was added to a small Econo-Pac disposable chromatography column (BioRad) and then washed with 30 mL buffer A by gravity flow to remove unbound proteins. TF-PceA was eluted from the column with 40 mL buffer B (buffer A plus 2.5 mM D-desthiobiotin). Fractions were analyzed by SDS-PAGE and protein concentrations calculated by the Bradford method with bovine serum albumin (BSA) as a standard (25). Elution fractions containing pure TF-PceA were combined and anaerobically concentrated to ~10 μM using a 100 kDa molecular weight cut-off (MWCO) spin concentrator. Anaerobic conditions were maintained by sealing the spin concentrator in the glove box. Proteolytic cleavage of TF-PceA was performed with HRV 3C protease (Novagen) in buffer B plus 0.1% (w/v) n-octyl-β-D-glucopyranoside (BOG) for 16-24 h at 4 °C in an anaerobic sealed tube with gentle shaking.

### 4.3.5 Reconstitution of 4Fe-4S clusters and cobalamin uptake

All experiments with TF-PceA were carried out in anaerobic 50 mM Hepes pH 8.0, 150 mM NaCl and 10% glycerol, unless otherwise stated. Chemical reconstitution of the PceA 4Fe-4S clusters was performed according to (26). Briefly, 10 μM of TF-PceA was fully reduced in the glove box at 21 °C with 2% β-mercaptoethanol for 1 h under strictly anaerobic conditions. Na₂S and FeCl₃ (Sigma) were slowly added to a final concentration of 150 μM each. The reconstitution was carried out for 16 h at 4 °C with gentle shaking, in either the presence or absence of HRV 3C protease. Methylcobalamin (Sigma) was incorporated into PceA during the limited proteolysis with HRV 3C. Excess cobalamin was removed from the sample by dialysis of the sample into 500 fold excess of anaerobic buffer. The dialysis step was performed in parallel to the cobalamin uptake step.
4.3.6 UV-Vis and EPR measurements

UV-Vis absorbance spectra were recorded with a Cary UV-Vis spectrophotometer situated inside an anaerobic glove box. All spectra recorded were baseline corrected with buffer or the collected flow through. Reducing conditions were achieved by the addition of 5 mM Na-dithionite. EPR samples for cobalamin incorporation by PceA were prepared by further concentrating the protein sample from the dialysis tube with a 30 kDa cut-off spin concentrator. The 4Fe-4S clusters reconstitution in PceA was analyzed by preparing EPR samples before and after reconstitution under both reducing and non-reducing conditions. Again, reducing conditions were achieved by the addition of 5 mM Na-dithionite. Samples were transferred into 4 mm Suprasil quartz EPR tubes (Wilmad) within the glove box, anaerobically sealed, directly frozen and stored in liquid nitrogen until measured. EPR spectra were recorded at 12 and 30 K using a Bruker ELEXSYS E500/E580 EPR spectrometer (Bruker GmbH) fitted with an Oxford Instruments ESR900 helium flow cryostat coupled to an ITC 503 controller from the same manufacturer. The microwave power was 0.5 mW, the modulation frequency 100 kHz, and the modulation amplitude 5 G. The g values given were calculated using the software package supplied with the instrument.

4.4 Results

4.4.1 Validation of a cloning strategy aimed at expression of soluble holo-PceA

A variety of solubility and affinity tags were tested with the goal of obtaining soluble holo-PceA from recombinant expression in *E. coli*. Table 4.1 gives an overview of the various plasmids constructed, and for each case, the level of expression observed and whether a significant fraction of soluble protein was detected. Extensive expression trials, at various temperatures, different culture media, and different IPTG concentrations have shown that PceA can be overexpressed from all constructs. However, only when fused to either MBP or TF was soluble PceA obtained. The lack of soluble PceA expression from the other plasmids was confirmed by Western blot analysis on soluble extracts. In case of the pOPINF based plasmids, the presence of the PceA specific TAT signal sequence did not lead to the expression of soluble PceA either. Hence, the TAT signal sequence was excluded from all other constructs, because it would only increase the hydrophobic nature of PceA (18).

Although soluble protein could be obtained with the MBP-fusion system, the MBP-PceA fusion protein proved unstable and difficult to separate from MBP after proteolytic cleavage without precipitation. Soluble TF-PceA was expressed from the pCOLD TF vector at an OD_{600 nm} of 0.6 by induction with 0.1 mM IPTG and cold-shocking at 15 °C. Changes in the IPTG concentration did not significantly affect the level of TF-PceA.
Table 4.1: Overview of the various expression plasmids encoding pceA with a range of solubility and affinity tags as used in this study.

<table>
<thead>
<tr>
<th>Designed plasmid</th>
<th>Parent plasmid</th>
<th>Overexpression?</th>
<th>Soluble?</th>
</tr>
</thead>
<tbody>
<tr>
<td>His(^6)HRV(^b) PceA</td>
<td>pOPINF</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>His HRV TAT(^c) PceA</td>
<td>pOPINF</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>His SUMO(^d)HRV PceA</td>
<td>pOPINS</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>His NusA HRV PceA</td>
<td>pET50b((+))</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>His MBP(^e) HRV PceA</td>
<td>pOPINM</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>His TF(^f) HRV PceA StrepII</td>
<td>pCOLD TF</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

\(^a\) His = His\(^6\)-tag.
\(^b\) HRV = recognition sequence for the human rhinovirus (HRV) 3C protease.
\(^c\) TAT = twin arginine translocation sequence.
\(^d\) SUMO = small ubiquitin-like modifier.
\(^e\) MBP = maltose binding protein.
\(^f\) TF = trigger factor chaperone.

eexpression, but elevated temperatures did result in higher levels of insoluble expression. Interestingly, in the cases of SUMO or NusA solubility tags, large quantities of PceA were obtained at a variety of different temperatures and IPTG concentrations. However, these fusion proteins were found entirely in the insoluble fraction. Taken together, these findings suggest, that both a low induction temperature, as well as the correct solubility tag are essential for the soluble expression of PceA.

4.4.2 Purification, Fe-S cluster reconstitution and UV-Vis/EPR analysis of soluble TF-PceA fusion protein

The recombinant D. restrictus PceA, containing a C-terminal Strep-Tag II and fused to TF could be conveniently purified under anaerobic conditions to apparent homogeneity by one step affinity purification using StrepTactin Sepharose resin under mild elution conditions (2.5 mM D-desthiobiotin). The yield from the first purification step was \(\sim\) 1 mg of protein per 1 mL of resin. The binding to the resin was relatively weak, with large amounts of the overexpressed target protein eluting in the flow through and wash fractions (figure 6.1), giving a final yield of purified protein of around 25% compared to the overall expression yields as shown in table 4.2. Routinely approximately 1-2 mg of purified fusion protein could be obtained from a 1 L culture. The apparent molecular weight on SDS-PAGE (figure 6.1) was estimated to be \(\sim\)116 kDa which corresponds to the predicted size of the fusion protein. After elution from the StrepTactin resin,
TF-PceA could be concentrated under anaerobic conditions to several mg/mL. During the concentration step, the protein solution developed a characteristic golden brown colour, a common feature of Fe-S cluster containing proteins. Further investigation by EPR analysis indeed reveals the presence of a 4Fe-4S signal (figure 6.2). Under reducing conditions (5 mM Na-dithionite), a typical $[4\text{Fe}-4\text{S}]^{+1}$ cluster signal appears in the EPR spectra at 1.93 g, recorded at 12 K. This signal was oxygen sensitive and when the temperature was increased to 30 K, the signal at 1.93 g was saturated (data not shown).

### Table 4.2: Summary of the TF-PceA fusion protein expression and purification.

<table>
<thead>
<tr>
<th>Purification step $^{a}$</th>
<th>Total protein (mg)</th>
<th>TF-PceA (mg)</th>
<th>Purity (%)</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude extract $^{b}$</td>
<td>3120</td>
<td>40</td>
<td>1.28</td>
<td>100</td>
</tr>
<tr>
<td>Soluble crude extract</td>
<td>2400</td>
<td>30</td>
<td>1.25</td>
<td>75</td>
</tr>
<tr>
<td>After StrepTactin purification $^{c}$</td>
<td>10</td>
<td>10</td>
<td>&gt;95</td>
<td>25</td>
</tr>
</tbody>
</table>

$a$ The values given are from a 6 L expression in *E. coli* BL21 (DE3) an LB media using the pCOLD TF vector.

$b$ Cell wet weight $\sim 23$ g.

$c$ Performed with 10 mL StrepTactin Sepharose.

Despite the presence of a 4Fe-4S EPR signal, quantification indicated a rather low occupancy of the 4Fe-4S clusters in the recombinant TF-PceA. Hence, we attempted to chemically reconstitute the cluster formation as described above. After the reconstitution and removal of any precipitate by centrifugation, the protein solution had a strong golden brown colour. Figure 6.2 shows the respective EPR spectrum of PceA before and after chemical Fe-S cluster reconstitution. Before reconstitution, the characteristic $[4\text{Fe}-4\text{S}]^{+1}$ signal at 1.93 g was weakly developed, whereas the reconstituted sample exhibited a
strong signal. The TF-PceA concentration of the reconstituted sample was \( \sim 60 \, \mu\text{M} \). Quantification of the 4Fe-4S signals lead to an estimate of \( \sim 60 \, \mu\text{M} \) as well. As PceA is predicted to contain two 4Fe-4S binding sites, the reconstitution achieved an \( \sim 50\% \) occupancy of the total 4Fe-4S clusters. However, the pCOLD TF-PceA fusion protein failed to bind cobalamin when incubated with commercially available cobalamins (data not shown).

### 4.4.3 Separation of PceA from TF and reconstitution of holo-PceA

Proteolytic cleavage of the covalent linkage between trigger factor and PceA was necessary to observe cobalamin incorporation by PceA. Cleavage was performed with the HRV 3C protease at 4°C for 16-20 h under anaerobic conditions and gentle rotation. Shorter incubation times led to incomplete proteolysis (data not shown). Without the addition of detergent, PceA did not remain in solution. By adding 0.1% (w/v) detergent BOG to the cleavage reaction, PceA could be stabilized and equal molar ratios of TF and PceA have been obtained after cleavage. Figure 4.3 shows an SDS-PAGE of TF-PceA before and after cleavage with HRV 3C. Nearly complete cleavage has been achieved under the conditions stated. Excess methylcobalamin (100 \( \mu\text{M} \), Sigma) was added during the cleavage reaction, in order to promote the cofactor incorporation into PceA. Removal of excess (unbound) cobalamin was achieved by dialysis of the sample against anaerobic buffer A containing 0.1% (w/v) BOG. This step was performed in parallel to the HRV
3C digestion and did not lead to any loss or precipitation of PceA.

UV-Vis and EPR analyses of PceA after the cobalamin addition confirmed the formation of a PceA-cobalamin complex as shown in figure 4.4. The UV-Vis spectrum in figure 4.4 A exhibits the two characteristic cobalamin absorbance peaks at 530 and 350 nm. Furthermore, the spectrum reveals the specific Fe-S cluster absorbance peak at 420 nm, which indicates that the 4Fe-4S clusters are maintained throughout the HRV 3C cleavage and cobalamin incorporation. The cobalamin EPR spectra have been recorded under non-reducing sample conditions at 30 K. The EPR trace in figure 4.4 B gives further insights into the binding mode of cobalamin by PceA. The single g-value of 2.26 g with hyperfine splitting between Co and the nitrogen atom of the lower axial ligand indicates a base-on binding, which means that the 5,6-dimethylbenzimidazole extension of methylcobalamin or a protein-derived nitrogenous base (likely His) is coordinated to the central Co$^{2+}$ atom by a nitrogen cobalt bond.

4.5 Discussion

Until now, study of RdhA has been exclusively done on protein purified from source. However, direct protein purification from organohalide-respiring bacteria is impractical and unlikely to routinely yield the high levels required for detailed biophysical studies. This study reports for the first time the soluble expression and purification of the reductive dehalogenase PceA from *D. restrictus*, from *E. coli*. We have shown that PceA, if fused to an appropriate solubility tag (TF) and expressed at low temperatures, can be produced in large quantities in *E. coli*. The soluble nature of the TF-PceA fusion protein is in total contrast to the behaviour observed with other common solubility
tags like NusA and SUMO, that failed in promoting soluble expression of PceA. These findings suggest that the combination of an appropriate solubility tag and a low temperature rather than a specific IPTG concentration are important for soluble expression of PceA. It is interesting to note that the PceA and some other RdhA operons contain a trigger-factor like chaperone (PceT/RdhT), and the observation made here that a TF is required for stable soluble PceA (as opposed to MBP leading to unstable protein) could signify a functional role for PceT in PceA maturation. Indeed, recent microbiological studies have revealed PceT might play a role in PceA maturation (27; 28). The purified TF-PceA fusion protein only contained sub-stoichiometric amounts of 4Fe-4S clusters, despite careful purification under strict anaerobic conditions (and avoiding Ni-affinity procedures where the eluting imidazole can affect FeS cluster stability). This could be due to low 4Fe-4S incorporation efficiency in \textit{E. coli} or the fact that the clusters are labile during the aerobic growth conditions used. Reconstitution of the 4Fe-4S clusters is possible by incubation with Na$_2$S and FeCl$_3$ under anaerobic conditions, although EPR quantification of the FeS clusters reveals the occupancy to be 50%. As the signals from the individual clusters cannot be distinguished, we cannot unambiguously judge to what extent both clusters are reconstituted. Indeed, it is possible only one of both clusters is successfully incorporated, leading to only 50% overall incorporation levels. Further studies will be required to optimize FeS reconstitution and determine the relative levels for both FeS clusters. The chemically reconstituted PceA does not have any affinity for cobalamins, which is possibly due to the fact that the protein is fused at the N-terminus with a trigger factor. The cobalamin binding domain of RdhA is believed to reside in the

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4_4}
\caption{UV-Vis and EPR measurements of PceA after cleavage from TF and incorporation of methyl cobalamin. A) UV-Vis spectrum of TF-PceA after cleavage with HRV 3C, incorporation of methyl cobalamin and dialysis against buffer A. The spectrum is baseline corrected for the dialysis buffer. B) The EPR spectrum shows cobalamin base-on binding with a g-value of 2.26 as well as clear evidence of hyperfine splitting. Sample was prepared under anaerobic conditions, but was not further reduced with Na-dithionite. Estimated PceA concentration = 30 µM; Measurement temperature = 30 K.}
\end{figure}
N-terminal region, and the TF might well affect conformation and binding properties of the adjacent PceA sequence. Indeed, this appears to be the case as cobalamin binding can readily be achieved following removal of the TF through limited proteolysis. Unlike the base-off conformation reported for the RdhAs purified from source (14; 15), the fully reconstituted PceA appears to bind methylcobalamin in the base-on form. This discrepancy could be due to the fact that the cobalamin tested is not the natural cofactor for PceA, or that enzymatic turnover is required to convert the enzyme from base-on to the base-off form. Unfortunately we have not yet been able to produce catalytically active PceA (data not shown). This may be due to one or several of the following reasons: (i) the tested cobalamins are not a natural cofactor of PceA; (ii) the reconstitution worked for only one of the two 4Fe-4S clusters; and (iii) another factor, which is required by PceA for catalysis was missing and is yet to be identified. Available information that might support the hypothesis that the wrong cobalamin was incorporated is that norpseudocobalamin has been identified as the main cobalamin in the reductive dehalogenase PceA of *Dehalospirillum multivorans* (18). Pseudocobalamins differ from cobalamin in the lower axial ligand part. Cobalamins contain a 5,6-dimethylbenzimidazole and pseudocobalamins an adenine instead. This could also explain the observed base-on signal (this study), whereas previous EPR analyses of PceA from *D. restrictus* have suggested the cobalamin to be in the base-off form (14).

### 4.6 Conclusions

This study documents for the first time the successful production of soluble PceA from a recombinant *E. coli* host. This has been a long-standing goal in the field of organohalide respiration and one of the main barriers for the elucidation of structural and functional insights of reductive dehalogenases. Surprisingly, production of soluble PceA was only possible through fusion with a trigger factor protein, possibly reflecting the role of a related protein PceT in the physiological PceA maturation. Although the protein produced remains enzymatically inactive, the novel expression platform allows us to explore the PceA-PceB interaction, the nature of the cobalamin cofactor and its binding site as well as structural analysis. In addition, future research will focus on obtaining enzymatically active PceA from *E. coli*, through optimization of FeS incorporation and cobalamin screening. Finally, we will use a similar protocol to test production of other RdhAs in *E. coli*.

### 4.7 Acknowledgements

This work was supported through the ERC starting grant DEHALORES206080.
4.8 References


Chapter 5

Structural and functional characterisation of the trigger factor protein PceT from *Dehalobacter restrictus*

Author’s: Hanno Sjuts, Mark S. Dunstan and David Leys

Affiliation: Manchester Institute of Biotechnology, Faculty of Life Sciences, The University of Manchester, M1 7DN, UK

This manuscript is currently not submitted to any scientific journal.
5.1 Abstract

Chaperones from the trigger factor family assist in the folding of newly synthesised polypeptide chains, directly after their release from the ribosome into the cytosol. Genomic studies of organohalide-respiring bacteria have identified conserved trigger factor genes \((rdhTs)\) in some of the same operons as the reductive dehalogenase genes \((rdhAs)\), which encode a new family of trigger factor proteins (RdhTs). RdhTs are thought to assist in the maturation and transportation process of the reductive dehalogenases (RdhAs, the enzymes catalysing the reductive dehalogenation reactions), by binding to their N-terminal located twin arginine translocation (TAT) signal sequence. Sequence alignments of RdhT proteins show homologies to full length trigger factor, however, lacking the N-terminal ribosome binding domain. Here, the recombinant expression, purification and crystallisation of the trigger factor PceT from \em{Dehalobacter restrictus} is described. Successful crystallisation was only observed in the presence of arginine, which leads to the assumption that physiologically active PceT may bind specifically to the conserved arginine residues of reductive dehalogenase’s (PceA) TAT signal sequence. Unfortunately the obtained crystals did not diffract beyond 3.7 Å, and ultimately made solving the crystal structure impossible. Furthermore, biophysical studies aiming towards a better understanding of this complex formation did not show any specific interaction between PceT and the TAT signal sequence or PceT specifically, or L-arginine in general. Additionally it was investigated whether PceT could substitute for \em{E. coli} trigger factor, for the expression of soluble reductive dehalogenase enzyme from recombinant expressions; a task that was recently achieved and showed that soluble reductive dehalogenase expression was only achieved when covalently fused to \em{E. coli} trigger factor.

Keywords: PceT \(\diamond\) Trigger factor \(\diamond\) Chaperone \(\diamond\) X-ray crystallography \(\diamond\) Organohalide respiration \(\diamond\) TAT signal sequence
5.2 Introduction

Nascent prokaryotic polypeptide chains, exiting the ribosome into the cytosol face two main issues. First they need to be properly folded; second they need to be transported to the place where they are biologically active. In Escherichia coli folding of newly synthesised proteins is assisted by the trigger factor protein, an ubiquitously expressed cytosolic protein of 48 kDa (1). The structure and function of trigger factor has been studied extensively; the protein consists of three major domains: an N-terminal ribosome-binding domain, a central peptidyl-prolyl cis-trans isomerase (PPI) domain and a C-terminal chaperone domain as shown in figure 5.1 (1). Trigger factor assists protein folding mainly by binding to exposed hydrophobic amino acid patches and by catalysing PPI reactions. The PPI activity of trigger factor has been known for some time (2), but an in-depth understanding of this enzymatic activity is still lacking. Binding of trigger factor to the ribosome has also been confirmed, and involves the N-terminal domain of trigger factor (amino acid residues 1 - 118) which interacts with the ribosomal protein L23 on the 50s subunit (3). Crystallographic studies enabled the structural determination of several trigger factor proteins. Initially the structure of the N-terminal domain of E. coli trigger factor was reported (4). Shortly after that the structure of complete trigger factor was solved as a monomer, and in complex with the ribosome (5). Additionally, crystal structures of C-terminal truncated Vibrio cholera trigger factor and the N-terminal domain of Deinococcus radiodurans in complex with the 50S ribosomal subunit have been published (6; 7). In case of the V. cholera trigger factor, a homodimer formation in solution has been observed (6).

![Image](image_url)

**Figure 5.1: Crystal structure of E. coli trigger factor.** The N-terminal ribosome binding domain is shown in blue, the central PPI domain is shown in dark green and the C-terminal chaperone domain is shown in red. The protein chain is shown as ribbons and the protein surface is additionally presented in light blue with the transparency set to 0.3. This figure has been prepared from the PDB file 1W26 with the QtMG CCP4 program.

The trigger factor gene *tig* is conserved throughout all prokaryotes, hence it is also
present in organohalide-respiring bacteria. For example the completion of the genome sequencing of *Desulfitobacterium hafniense* has revealed the presence of two *tig* genes (8). Interestingly, many organohalide-respiring bacteria harbouring reductive dehalogenase (*rdhA*) genes contain an additional putative trigger factor gene in the same operon. This specialised trigger factor family has been named RdhT (9) and specifically CprT in e.g. *D. hafniense* and PceT in e.g. *D. restrictus* (10; 11). CprT is usually located upstream of the reductive dehalogenase gene, whereas *pceT* is located downstream of it (12). Sequence alignments of PceT from *D. restrictus* with other trigger factor proteins, for which crystal structures have been solved, are shown in figure 5.2. The alignment suggests that the N-terminal ribosome binding domain is lacking in PceT and supports the idea that PceT aids specifically in the folding / transport of selected proteins derived from its own gene operon, e.g. PceA, PceB or PceC (11). *RdhT* genes have only been identified in operons that contain the genes *rdhA*, *rdhB* and *rdhC* (13). Functional *in vivo* and *in vitro* studies suggest that PceT from *D. restrictus* is a specific chaperone for the reductive dehalogenase PceA. Co-immunoprecipitation studies have shown that PceT interacts with the twin arginine translocation (TAT) signal sequence of PceA (14). The same authors also observed PPI activity for PceT. A recent study by Maillard et. al has confirmed the *in vivo* interaction between PceT and PceA (9). This is in line with our observations, i.e. that a trigger factor protein is necessary to produce soluble PceA using a recombinant overexpression method in *E. coli* (15). However, no structural or quantitative *in vitro* biophysical techniques (e.g. isothermal calorimetry (ITC)) have been shown to support the interaction between PceT and PceA. On the other side, increasing amounts of genomic information highlight the correlated presence of *rdhT* and *rdhC* genes (13) which supports the theory that a functional relationship between RdhT and RdhC may exist. Currently, the biological function of RdhC’s is unknown, but the proteins are thought to be membrane proteins with five or six transmembrane regions. The experiments described in this chapter attempt to investigate the molecular mechanisms underpinning the putative complex formation between PceT and PceA by a structural and biophysical approach. PceT was purified on a large scale from recombinant expression in *E. coli*. A protocol for the successful crystallisation of full length PceT is described. Similarly, the TAT signal sequence from *D. restrictus* PceA, covalently fused to the maltose binding protein (MBP) was produced and tested for its ability to interact with PceT using ITC.

### 5.3 Materials and methods

#### 5.3.1 Cloning of expression plasmids

The *pceT* gene (accession number Q6EXA6) from *D. restrictus* has been cloned in the pOPINF vector (OPPF) (17) by the ligation free In-Fusion™ technology (Takara Clon-
Cobalamin-dependent enzyme systems

Chapter 5. PceT

Figure 5.2: Primary sequence alignment of trigger factor proteins for which crystal structures have been solved, compared to PceT from *D. restrictus* (no structural information available).

The alignment implies that PceT is a truncated version of full length trigger factor, and lacks the N-terminal ribosome binding domain. The sequences were obtained from the Uniprot database. The multiple sequence alignment was performed by the ClustalX program (16), the downstream colour coding and image processing with the software Jalview© 2.6.1. Coloured residues are conserved to 75% (light blue) or 100% (dark blue) between the aligned protein sequences. The N-terminal ribosome binding domain is shown within a blue box, the central PPI domain within a green box and the C-terminal chaperone domain within a red box.

Table 5.1: List of generated expression plasmids that have been used in this study. All cloning steps were performed using ligation independent cloning. The empty pOPINF/M/S plasmids have been obtained from the Oxford Protein Production Facility (OPPF); the pET28a(+) and pCOLA Duet plasmids are from Novagen.

<table>
<thead>
<tr>
<th>Construct name</th>
<th>Parent plasmid</th>
<th>Forward primer, 5' - 3'</th>
<th>Reverse primer, 5' - 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>PceT</td>
<td>pOPINF</td>
<td>AAG TTC TGT TCT AGG GCC CTA TGA</td>
<td>ATG CGT TAG AAA GCT TTA GGA AGC</td>
</tr>
<tr>
<td>MBP-TAT</td>
<td>pOPINM</td>
<td>AGG AGT TGG AAC TGG GT</td>
<td>ATG AGC TAG AAA GCT TTA GGA AGC</td>
</tr>
<tr>
<td>MBP-TAT-PceA</td>
<td>pOPINM</td>
<td>AGG AGT TGG AAC TGG GCC</td>
<td>ATG AGC TAG AAA GCT TTA GGA AGC</td>
</tr>
<tr>
<td>TAT-Sumo</td>
<td>pOPINS</td>
<td>AGG AGA TAT ACC ATG GGC ATG GGC</td>
<td>TGG CTG CTA CCC ATG GAA GGT GGC</td>
</tr>
<tr>
<td>PceT-PceA</td>
<td>pET28a (+)</td>
<td>GCC GGC GCA GGC ATA TGG AGT GGG CTC CTC GAT TAT GCC</td>
<td>TGG GCT ACT AAC GCC</td>
</tr>
<tr>
<td>PceT-PceA-Duet</td>
<td>pCOLA Duet</td>
<td>GCC AGG ATC CGA ATT CTA GGA AGC</td>
<td>CTA ACT TTA TAT TAA</td>
</tr>
</tbody>
</table>

Table 5.1: List of generated expression plasmids that have been used in this study. All cloning steps were performed using ligation independent cloning. The empty pOPINF/M/S plasmids have been obtained from the Oxford Protein Production Facility (OPPF); the pET28a(+) and pCOLA Duet plasmids are from Novagen.
The pOPINF vector allows the overexpression of proteins with N-terminal His\(_6\) tags under the control of the T7 promoter. The open reading frames of \textit{pceT} and \textit{pceA} including its N-terminal TAT signal sequence were synthesised and cloned by MWG Operon into the vector backbone pCR2.1 and pUC57, respectively. PCR amplifications of all DNA segments were performed in 50 \(\mu\)l reactions containing: 1x Phusion buffer (Finnzymes), 0.2 mM dNTP’s (Finnzymes), 0.5 \(\mu\)M forward primer, 0.5 \(\mu\)M reverse primer, 100 ng template plasmid and 0.5 \(\mu\)l Phusion polymerase (Finnzymes). Amplification was carried out in a TC-512 thermocycler (Techne) with 35 cycles consisting of: Initial denaturation: 98 \(^\circ\)C, 30 sec; denaturation: 98 \(^\circ\)C, 10 sec; annealing: 52 \(^\circ\)C, 30 sec; extension: 72 \(^\circ\)C, 30 - 120 sec (depending on the length of the sequence to be amplified); final extension: 72 \(^\circ\)C, 10 min. Afterwards, parental strands were digested by adding 20 U \textit{DpnI} (New England Biolabs) to the PCR product and incubation for 1 hour at 37 \(^\circ\)C. Restriction digest of 1 \(\mu\)g pOPINF or pOPINM plasmid was performed with 20 U \textit{KpnI} HF and 20 U \textit{HindIII} (New England Biolabs) in 1x Buffer 4 and with for 3 hours at 37 \(^\circ\)C in a total volume of 20 \(\mu\)l. Both the PCR product and the digested plasmid were cleaned by magnetic bead purification, according to the manufacturers instructions (Agencourt Bioscience Corporation) and eluted in 40 \(\mu\)l of 10 mM Tris, pH 8. DNA concentrations were determined by measuring the absorbance at 260 nm with a Nanodrop instrument (Eppendorf). The subsequent In-Fusion™ reaction was carried out in a 10 \(\mu\)l volume containing the PCR product, digested plasmid and 1 \(\mu\)l of In-Fusion enzyme in 1x In-Fusion buffer. The molar ratio of insert (PCR product) to plasmid was set to 3:1. This reaction mix was incubated for 15 min at 37 \(^\circ\)C and 15 min at 50 \(^\circ\)C. Competent \textit{E. coli} DH5\(\alpha\) cells (Stratagene) were transformed with 1 \(\mu\)l of the cloned plasmid by heat shock (45 sec, 42 \(^\circ\)C). After a 60 min recovery step in 750 \(\mu\)l SOC media, cells were plated on LB-agar plates containing 100 \(\mu\)g/ml Ampicillin (Formedium) as selective agent. Single colonies were picked and grown overnight in 5 ml LB medium containing 100 \(\mu\)g/ml Ampicillin and plasmid DNA was purified with the Qiagen Mini-Prep kit according to the manufacturers instructions, eluted in 50 \(\mu\)l 10 mM Tris, pH 8. Plasmids containing the desired clones were directly transformed into \textit{E. coli} BL21 competent cells (Stratagene) and plated on LB agar (Formedium) plates containing 100 \(\mu\)g/ml Ampicillin or 50 \(\mu\)g/ml Kanamycin (Formedium) as selective agent. For the cloning of the PceT-PceA fusion construct two separate PCR reactions were performed. The \textit{pceT} gene including the upstream His-tag was amplified from the pOPINF construct and the \textit{pceA} gene including the upstream HRV 3C protease recognition sequence and downstream StrepII tag was amplified from the pCOLD TF plasmid as described (15), leading to the expression of a His-PceT-HRV-PceA-StrepII fusion protein. The subsequent In-Fusion™ reaction was carried out in a 10 \(\mu\)l volume containing both PCR products and digested pET28a(+) plasmid (with 20 U \textit{NdeI} and 20 U \textit{BamHI} (New England Biolabs) 1x Buffer 4 and with for 3 hours at 37 \(^\circ\)C in a total volume of 20 \(\mu\)l) and 1 \(\mu\)l of In-Fusion enzyme in 1x In-Fusion buffer. The molar ratio of inserts (PCR products) to plasmid was set to 3:1.
This reaction mix was incubated for 15 min at 37 °C and 15 min at 50 °C. Transformation was performed as described above. The cloning of the different TAT plasmids was essentially performed as the PceT (pOPINF) plasmid using the same restriction enzymes. The complete pceA gene from *D. restrictus* (GenBank accession number Q8GJ27) was used as template for PCR amplification. The cloning of *pceT* and *pceA* into the pCOLA Duet plasmid (Novagen) was performed in two iterative cloning steps virtually as described above using the restriction enzymes *EcoRI* for *pceT* (multiple cloning site I), and *NdeI* and *XhoI* for the *pceA* gene (multiple cloning site II), respectively. Plasmids were verified by sequencing for the correct insertion of the desired genes by MWG Operon and are summarised in table 5.1.

### 5.3.2 Expression and purification of PceT

Single colonies of BL21 cells (Stratagene) transformed with different expression plasmids with the *pceT* gene inserted, were grown in LB media containing 100 µg/ml Ampicillin or 50 µg/ml Kanamycin as selective agent. A densely grown 5 ml starting culture was used to inoculate a 200 ml pre culture, which was grown overnight at 37 °C, with 220 rpm shaking. The main culture (usually twelve 2 litre flasks, each containing 750 ml media) was induced the next morning with the saturated pre-culture (10 ml inoculant per flask of main culture). Cells were grown at 37 °C, 220 rpm until an OD<sub>600nm</sub> of 0.5 was reached. Overexpression of the protein of interest was induced by temperature reduction to 20 °C and addition of 1 mM IPTG (Formedium). After ~16 hours of overexpression, cells were harvested by centrifugation (6,000 rpm (rotor JLA8.1, Beckman), 10 min, 4 °C) and stored at -20 °C until further use. Cell pellets were resuspended in buffer A (50 mM Hepes pH 8, 500 mM NaCl, 30 mM imidazole), supplemented with DNase, RNase (Sigma) and a protease inhibitor cocktail tablet (Roche) and twice passed through a French™ pressure cell press apparatus (Thermo IEC) at 1500 psi. Soluble crude extract was obtained by centrifugation (35,000 rpm (rotor Ti45, Beckman), 1 hour, 4 °C) in a Optima CE-80K ultracentrifuge (Beckman Coulter) and loaded on a 5 ml His Trap column (GE Healthcare) equilibrated with buffer A. The column was washed with 10 column volumes of buffer A to remove unbound proteins and PceT was eluted with a linear gradient of 20 column volumes into buffer B (50 mM Hepes pH 8, 500 mM NaCl, 300 mM imidazole). Eluted fractions containing PceT were pooled, concentrated to ~10 ml using a 30 kDa molecular weight cut-off (MWCO) spin concentrator (Sartorius). This protein sample was diluted to 100 ml with ice-cold 25 mM Hepes pH 8 in order to reduce the ionic strength and was then loaded onto a 20 ml MonoQ anion exchange column equilibrated with 25 mM Hepes pH 8 and eluted with a linear salt gradient over 20 column volumes (0 - 500 mM NaCl). Fractions containing PceT were pooled and concentrated as described. At this stage PceT was either frozen in liquid nitrogen and stored at -80 °C or further purified for crystallographic studies. For crystallographic purposes the
N-terminal His-tag was cleaved with HRV 3C protease (Novagen) in a 1:1000 molar ratio for 16 - 20 hours at 4 °C with gentle mixing. Reverse affinity purification on a 1 ml Ni-NTA column was performed for the removal of the His tag and the protease. PceT was further concentrated and then directly applied to a Superdex 200 gel filtration column (GE Healthcare), equilibrated with buffer C (10 mM Hepes pH 8 and 10 mM arginine) and purified with a flow rate of 0.3 ml/min. Fractions containing pure PceT were combined, concentrated and directly used for crystallographic experiments.

5.3.3 Expression and purification of TAT fusion proteins

Both MBP-TAT fusion proteins were expressed and purified as described for PceT with the following modifications: The purification buffer A contained 50 mM Hepes pH 8, 200 mM NaCl and 10% glycerol and the affinity purification was performed using 5 ml MBP resin for the MBP-TAT fusion protein. Fusion proteins were eluted with buffer A including 10 mM maltose. For the TAT-Sumo fusion protein expression, Kanamycin was used as selective agent. Protein purification was performed using a 1 ml His Trap column (GE Healthcare) with buffers as used for PceT containing additionally 10% glycerol.

5.3.4 Expression and purification of the PceT-PceA fusion protein

In addition to BL21, Arctic express cells (Agilent Technologies) were transformed with the PceT-PceA fusion plasmid. Protein expression and anaerobic purification using the C-terminal StrepII tag on PceA, was essentially performed as described in (15). In brief, two litre flasks, containing 750 ml of LB medium were inoculated with 10 ml saturated overnight starter culture and grown at 37 °C at 200 rpm until the mid log phase was reached. The temperature was then lowered to 12 °C and protein expression was induced by the addition of 0.1 mM IPTG supplemented with 50 µM FeCl₂. Cells were harvested 24 hours after IPTG induction.

5.3.5 Determination of PceT dispersity by multi-angle light scattering

Purified PceT was analysed by multi-angle light scattering (MALS) in order to confirm monodispersity prior to crystallization trials. 200 µg of protein in 500 µl of 25 mM Hepes pH 8, 150 mM NaCl, were applied to a gel filtration column (Superdex 200) operated by a Dionex BioLC GS 50 gradient pump, with a flow rate of 0.7 ml/min. Light scattering signals of the column eluent in addition to absorbance readings at 280 nm were recorded with a Wyatt DAWN HeleosII detector. The hydrodynamic radius and the molecular mass of the sample were calculated with the Astra software (Wyatt technology corporation).
5.3.6 Crystallisation and data collection of PceT

Monodisperse PceT was concentrated in gel filtration buffer to 15 mg/ml with a 30 kDa MWCO spin concentrator. Crystallisation screens were performed by the sitting drop vapour diffusion method and drops were dispensed with the automated mosquito liquid handling technology (Mosquito MD11-11, Molecular Dimensions) in a 96 well plate format (MRC Wheldan plates). The JCSG sparse matrix screen, PACT premier HT-96 screen, Morpheus and Clear Strategy Screens 1 and 2 (Molecular Dimensions) were tested. Crystallisation drops contained 200 nl protein solution and 200 nl precipitant solution in each well and 50 µl precipitant solution in the reservoir. Plates were covered with clear sealing sheets (Duct Tape) and incubated at 4 °C. Crystals grew in the three Morpheus conditions A2, A6 and A10 within 24 - 48 hours. All three crystallisation conditions contain 10% (w/v) PEG 8000, 20% (v/v) ethylene glycol, 0.03 M MgCl$_2$ and 0.03 M CaCl$_2$. They only differ in their buffer system (A2 = 0.1 M MES/imidazole pH 6.5; A6 = 0.1 M MOPS/HEPES-Na pH 7.5; A10 = 0.1 M bicine/Trizma base pH 8.5). When the crystallisation trays containing the PceT crystals were taken from 4 °C to room temperature, the crystals dissolved within five to ten minutes. Hence, all crystal handling steps including cryo-cooling were subsequently performed in the cold room. Due to the fact that the crystals grew in a cryoprotected solution (20% (v/v) ethylene glycol), the crystals were directly mounted from the drop and flash frozen in liquid nitrogen. Complete data sets were collected from a single crystal on beamline I04 at the Diamond Light Source synchrotron-radiation facility. Automated data processing was performed by _xia2_ (18). The crystals belonged to space group C$2\beta$, with unit-cell parameters $a = 159.59$ Å, $b = 57.07$ Å, $c = 63.2$ Å; $\alpha = 90^\circ$, $\beta = 111.84^\circ$, $\gamma = 90^\circ$, and contained one PceT protein in the asymmetric unit. Data collection statistics are given in table 5.2.

5.3.7 Isothermal titration calorimetry (ITC)

ITC experiments were performed using the VP-ITC 2000 system (MicroCal, Northampton, MA). In each experiment 20 aliquots of 15 µl containing 150 µM of ligand were injected into 1.7 ml of 15 µM PceT in 50 mM Hepes pH 8, 150 mM NaCl in the measuring cell. Injection times were set to 18 seconds with 300 seconds spacings. Measurements were performed at 25 °C with a stirring speed of 310 rpm. PceT and MBP-TAT fusion protein were dialysed into 50 mM Hepes pH 8, 150 mM NaCl by gel filtration. The TAT15 peptide (synthesised by Genosphere, containing the first 15 residues of the PceA specific TAT signal sequence) and arginine were dissolved in the same gel filtration buffer prior to the experiment.
5.3.8 *In vitro* testing of prolyl-cis-trans isomerase activity of PceT

PceT has been tested for prolyl-cis-trans isomerase activity by an assay depending on the conformational selectivity of *α*-chymotrypsin. This protease will cleave the peptide *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide, only if the Ala-Pro peptide bond is in the *trans* conformation. Cleavage then takes place between Phe and *p*-nitroanilide, which can be observed photometrically at 410 nm (2). The peptide *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (Sigma) was dissolved to a concentration of 1.56 mM in trifluoroethanol, additionally containing 0.47 M LiCl. *α*-chymotrypsin (6 mg/ml, Sigma) was dissolved in reaction buffer (100 mM Tris pH 7.5). Experiments were carried out as follows: 845 µl of reaction buffer was mixed with 100 µl of *α*-chymotrypsin solution and incubated for 10 min at 15 °C. Afterwards, 30 µl of PceT in various concentrations were added and the reaction started with the addition of 25 µl of assay peptide and absorbance at 410 nm was observed for five minutes at 15 °C.

5.4 Results and discussion

5.4.1 Purification and crystallisation of PceT

The putative trigger factor protein PceT from *D. restrictus* with N-terminal His<sub>6</sub> tag was successfully purified to homogeneity as described above after recombinant expression in *E. coli*. PceT eluted from the gel filtration column at an apparent molecular weight of ∼ 35 - 40 kDa, which is in line with the SDS-PAGE analysis (figure 5.3 A) and its theoretical molecular weight of ∼ 38 kDa. Multi angle light scattering experiments show a monodisperse protein peak at a retention time of ∼ 14.5 ml (figure 5.3 B). Therefore it was assumed that PceT, in an uncomplexed form in solution, is present as a monomer. The findings that PceT is a monomer in solution, lead to the possibility that the N-terminal domain of full length trigger factor is responsible for its dimerisation in the cytosol. This is further supported by the observation that trigger factor is bound to the ribosome in a monomeric form (5). Crystals of PceT grew within 48 hours in the Morphus screen (Molecular Dimensions) conditions A2, A6 and A10: In all three conditions the same plate like crystal morphologies were observed as shown in figure 5.3 C. Interestingly, PceT only crystallised in the presence of L-arginine in the buffer. L-arginine is a known additive that can improve the solubility / stability of proteins and enhance their ability to form crystals (19). Furthermore, the TAT signal sequence of PceA contains two consecutive arginine residues, which might imply that PceT specifically recognises that region of the TAT sequence, thus leading to a stabilisation of the protein that facilitated the process of crystallisation. These two consecutive arginine residues are the hallmark of TAT signal sequences. Studies of similar such systems have shown that these arginine residues are involved in binding to their partner chaperone (20). The obtained crystals diffracted in-house to ∼ 7 and at the synchrotron to ∼ 3.7 Å, respectively (figure 5.3 D).
Crystal annealing for up to several minutes, a technique where the nitrogen cryo-stream is switched off and the crystal allowed to warm to room temperature before reapplying the cryogenic environment, did not lead to improved diffraction of the crystals. Different seeding techniques (streak seeding and microseeding) did not lead to the identification of additional crystallisation conditions.

A full data set was collected on the I-04 Diamond light source synchrotron beamline and was autoindexed by XDS (as part of xia2) (21) into space group C2 with unit cell
Table 5.2: Crystallographic data collection statistics. Values in parentheses are for the highest resolution shell.

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source</td>
<td>Beamline I04, Diamond</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.92</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>Unit-cell parameters (Å)</td>
<td>a=159.57, b=57.06, c=63.2</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>58.66 - 3.65 (3.75 - 3.65)</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>19270 (1549)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>3.3 (3.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.7 (98.5)</td>
</tr>
<tr>
<td>$R_{merge}^a$ (%)</td>
<td>8.5 (49.3)</td>
</tr>
<tr>
<td>$[I/\sigma(I)]$</td>
<td>8.7 (2.2)</td>
</tr>
<tr>
<td>Protein molecules per asymmetric unit</td>
<td>1</td>
</tr>
<tr>
<td>Matthews coefficient $V_M$ Å³ Da⁻¹</td>
<td>3.76</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>67.28</td>
</tr>
</tbody>
</table>

\[ aR_{merge} = \frac{\sum_{hkl} |\sum_i (I_i(hkl)) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i (I_i(hkl))} \]

dimensions of $a = 159.59$ Å, $b = 57.07$ Å, $c = 63.2$ Å; $\alpha = 90^\circ$, $\beta = 111.84^\circ$, $\gamma = 90^\circ$. The calculated Matthews coefficient (22) is 3.76 for 1 molecule (35.5 kDa) in the asymmetric unit, with a solvent content of 67.28%. Such a high Matthews coefficient, although unusual (23), may explain the properties that the PceT crystals exhibit, that are: fragility, rapid melting at room temperature and diffraction to only 3.7 Å.

Different molecular replacement strategies have been tested in order to calculate phase information that can be used to build a model of PceT with the calculated structure factors. However, no solution was found by the different programs used Phaser, MrBump and Balbes that would allow the modelling of the PceT polypeptide chain into the electron density. Because of the poor diffraction properties of PceT it was not attempted to grow selenomethionine containing versions of PceT for solving the phase problem.

### 5.4.2 Purification of MBP-TAT Fusion Proteins

The TAT signal sequence of PceA has the following sequence:

MGEINRRNRFKASMLGAAAAAVASASAVKGMVSPLVADA

The sequence does not contain any residues that absorb wavelength at 280 nm and is also comprised of many hydrophobic residues, resulting in a low traceability on e.g. AKTA purification systems and low solubility in aqueous solutions. In order to investigate the putative complex formation between PceT and the TAT sequence, fusion proteins
between MBP and the PceA TAT sequence (with and without the N-terminal region of the mature PceA protein) were constructed, expressed and purified as described above. The expression and purification of these proteins was facilitated by the MBP which serves as both solubility- and affinity-tag (figure 5.4), resulting in the successful production and solubilisation of both TAT fusion proteins.

![Figure 5.4: SDS-PAGE analysis of MBP-TAT fusion proteins. A) MBP-TAT fusion protein with a calculated molecular weight of 50 kDa; B) MBP-TAT-PceAstart with a calculated molecular weight of 58 kDa. 1 = molecular weight marker in kDa; 2 = purified fusion protein. Protein bands were stained with InstantBlue dye (Expedeon).]

5.4.3 Biophysical interaction between PceT and the TAT signal sequence

Biophysical \textit{in vitro} data describing the interaction between PceT (or any other reductive dehalogenase specific trigger factor) and the corresponding TAT signal sequence are not yet available. However, \textit{in vivo} data and studies from similar protein systems give rise to the assumption that PceT aids in the maturation of the reductive dehalogenase PceA by binding to its N-terminal TAT signal sequence (24; 9). By isothermal titration calorimetry (ITC) studies, this putative interaction was investigated. Furthermore it was tested if PceT binds specifically to L-arginine, which was found to be required for the crystallisation of PceT. Unfortunately, no specific heat dissipation could be observed for any of the titrated ligands as depicted in figure 5.5, which would indicate a specific binding event. The minor spikes observed at any injection were interpreted as artefacts resulting from the mixing between the PceT and ligand solution in the ITC cell chamber. Similarly, the MBP-TAT-PceAstart protein precipitated during the concentrating step and was also not suitable for ITC analysis.
Due to scarce literature information, it is difficult to judge why no binding was observed between PceT and any binding partner tested. The negative ITC data indicate, assuming that PceT binds to PceA in vivo, that the conserved arginine residues of the TAT sequence are not sufficient to trigger the binding. However, the presence of L-arginine is crucial for successful crystallisation, as described above. It might be that L-arginine stabilises PceT either non-specifically or that a specific binding does not generate enough heat changes in order to be detected by ITC. Another explanation would be that a yet unidentified factor is necessary to trigger binding between PceT and the TAT signal sequence. The constructed MBP-TAT fusion proteins are problematic regarding their biological relevance. This is mainly due to the fact that the TAT signal sequence is located at the C-terminus of MBP and not at the N-terminus. However, this approach was necessary for the soluble expression of TAT fusion proteins. Constructs containing the TAT signal sequence as the most N-terminal element (TAT-Sumo fusion protein) did not result in any soluble expression of the TAT-Sumo protein and were hence not applicable for the biophysical studies undertaken (data not shown). Care must be taken not to over evaluate the results from the MBP-TAT fusion proteins.

5.4.4 Expression and purification of the PceT-PceA fusion protein and co-expression of PceT::PceA

Soluble production of the reductive dehalogenase PceA from a recombinant source has been observed when the full length E. coli trigger factor was covalently attached to the N-terminus of PceA lacking its N-terminal TAT signal sequence (15). In order to test whether PceT is likewise able to aid in the solubilisation of PceA, a PceT-PceA fusion was constructed 5.1. The fusion protein contains an N-terminal His-tagged PceT
and C-terminal StrepII tagged PceA (excluding the TAT signal sequence). PceT and PceA are covalently fused together by a linker region comprising the HRV 3C protease recognition sequence. Figure 5.6 shows SDS-PAGE analysis of the respective fractions from the purification process by StrepTactin affinity purification. The majority of the expressed fusion protein was found in the insoluble fraction although a small proportion of the fusion protein could be detected in the soluble fractions. However, neither the presence of 4Fe-4S clusters, nor the incorporation of cobalamin could be detected in these fractions when analysed by UV-Vis and electron paramagnetic resonance (EPR) spectroscopy (data not shown). Similarly, the co-expression experiments of PceT and PceA from the pCOLA Duet plasmid (PceT::PceA), which allows the expression of two genes individually under the control of a \textit{lac} operon), did not lead to the expression of any soluble PceA (data not shown). Hence, the question arises why full-length \textit{E. coli} trigger factor is better suited for the soluble expression of PceA than PceT. A possible explanation is that PceT binds to the TAT signal sequence at the N-terminus of PceA, whereas trigger factor binds more promiscuously to exposed hydrophobic protein stretches in general. As the TAT signal sequence has been omitted from PceA in this study, the natural binding partner of PceT would have been lost and PceT would solely serve as a solubility tag for PceA during the recombinant expression. This is in good agreement with the observation, that common solubility tags like maltose binding protein (MBP) or NusA do not promote the soluble expression of PceA (15).

**Figure 5.6:** SDS-PAGE analysis of the purification process of the PceT-PceA fusion protein. 1 = molecular weight marker in kDa; 2 = insoluble fraction containing most of the overexpressed PceT-PceA fusion protein as judged from the intense band between 100 and 130 kDa; 3 = flow through; 4 and 5 = wash fractions; 6 and 7 = fractions eluted with buffer A plus 2.5 mM d-Desthiobiotin from 5 ml StrepTactin resin. The faint band between the 100 and 130 kDa marker lanes corresponds to soluble PceT-PceA fusion protein. Protein bands were stained with InstantBlue dye (Expedeon).

### 5.4.5 Functional characterisation of PceT

For full length \textit{E. coli} trigger factor a prolyl \textit{cis}-trans isomerase activity has been confirmed (2). A single \textit{in vitro} study claims that PceT from \textit{Desulfitobacterium hafniense} Y51 likewise exhibits this enzymatic activity (14). However, no such activity has been
detected for PceT from *D. restrictus* (figure 5.7), although the protein sequences between these two proteins are the same (100% sequence identity).

![Figure 5.7: Prolyl cis-trans isomerase activity can not be detected for PceT. The photometric assay monitors the α-chymotrypsin induced release of p-nitroanilide from the peptide Ala-Ala-Pro-Phe. The absorbance of p-nitroanilide at 410 nm increases after cleavage from the peptide. Cleavage is preferred, if the Pro-Phe peptide bond is in the trans conformation. In the absence of a prolyl isomerase 60 - 90% of all peptide bonds are in the trans conformation (25). Here, no significant absorbance increases can be detected for any PceT concentration.](image)

### 5.5 Conclusions

Although the soluble expression, purification and crystallisation of PceT was successfully achieved, downstream biochemical and crystallographic experiments aiming towards the elucidation of structural and functional relationships of this novel trigger factor-like protein family, were overall unsuccessful. Multiple open questions remain regarding the physiological function of PceT, for example:

- If PceT and PceA do interact, what signal triggers the interaction?
- If PceT is a chaperone for PceA, does it support the folding, cofactor incorporation or transport of PceA from the ribosome to the TAT export machinery at the plasma membrane?
- Is the three-dimensional structure of PceT similar to that of other known trigger factors, e.g. *E. coli* trigger factor?
- Does PceT interact with RdhC?

The established crystallisation protocol for PceT, however, could pave the way for the successful crystallisation of either modified PceT proteins, or alternative members of the
RdhT trigger factor protein family. To achieve this, limited proteolysis experiments with full-length PceT could help to identify domains of PceT that are either more stable or lead to a better crystal packing interaction than the full-length and might be better suited to X-ray diffraction experiments. Alternatively, site-directed mutagenesis of charged surface residues in PceT (e.g. lysine to alanine mutations) could lead to a tighter crystal packing and therefore to improved diffraction (26). Exhibition of enzymatic prolyl cis-trans isomerase activity could not be detected for PceT from \textit{D. restrictus} and the reported result from Morita \textit{et al.} (14) are hence challenged. Further studies on both the \textit{in vivo} as well as \textit{in vitro} levels will be needed to fully elucidate the physiological functions of PceT from \textit{D. restrictus} specifically, and other RdhTs in general.

5.6 References


Chapter 6

Structure of the cobalamin-binding protein of a putative $O$-demethylase from *Desulfitobacterium hafniense* DCB-2

Author’s: **Hanno Sjuts**, Mark S. Dunstan, Karl Fisher and David Leys

Affiliation: Manchester Institute of Biotechnology, Faculty of Life Sciences, The University of Manchester, M1 7DN, UK

Reproduced with permission from the International Union of Crystallography.
6.1 Abstract

This study describes the identification and the structural and spectroscopic analysis of a cobalamin-binding protein (termed CobDH) implicated in O-demethylation by the organohalide-respiring bacterium Desulfitobacterium hafniense DCB-2. The 1.5 Å resolution crystal structure of CobDH is presented in the cobalamin-bound state and reveals that the protein is composed of an N-terminal helix-bundle domain and a C-terminal Rossmann-fold domain, with the cobalamin coordinated in the base-off/His-on conformation similar to other cobalamin-binding domains that catalyse methyl-transfer reactions. EPR spectroscopy of CobDH confirms cobalamin binding and reveals the presence of a cob(III)alamin superoxide, indicating binding of oxygen to the fully oxidized cofactor. These data provide the first structural insights into the methyltransferase reactions that occur during O-demethylation by D. hafniense.

Keywords: CobDH ⋄ Crystallography ⋄ O-demethylation ⋄ Cobalamin ⋄ Methyltransferase
6.2 Introduction

Lignin breakdown by fungal haloperoxidases in forest soil can give rise to chlorinated phenyl methyl ethers. Such compounds can serve as the terminal electron acceptors for anaerobic organohalide-respiring bacteria, leading to the corresponding dechlorinated phenyl methyl ethers (1; 2). Similar to acetogenic bacteria, certain strains of organohalide-respiring species have been found to use phenyl methyl ethers as an energy source, leading to further mineralization of lignolic compounds. A transcriptional study noted the upregulation of a gene cluster potentially involved in vanillate metabolism in Desulfitobacterium hafniense Y-51 which includes cobalamin-dependent enzymes (3).

The recent availability of genome sequences from several organohalide-respiring bacteria has revealed, in addition to a surprising wealth of reductive dehalogenase genes (RdhAs), the presence of multiple putative cobalamin-binding proteins (4). These are often located in gene clusters that appear to encode multicomponent O-demethylase-like enzyme systems (5). Additionally, it has recently been reported that D. hafniense strain PCP-1 is able to catalyse both dehalogenation and O-demethylation reactions stepwise from the same substrate precursors (e.g. tetrachloroguaiacol, tetrachloroveratrole or pentachloroanisole), providing a functional link between reductive dehalogenation and O-demethylation in organohalide-respiring bacteria (2). O-Demethylases belong to the class of cobalamin-dependent methyltransferases that catalyse methyl-group transfer from a methyl donor to a methyl acceptor via a methylcobalamin or methylcobamide acting as a reaction intermediate. Such methyl-transfer reactions are of fundamental importance in energy generation in many microorganisms (6). In the case of the O-demethylase system, four different proteins are needed for the reaction: (i) a substrate-binding domain that serves as a methyl donor, (ii) a cobalamin-binding domain that catalyses the methyl-transfer reaction, (iii) a tetrahydrofolate-binding domain that accepts the methyl group, thus forming 5-methyltetrahydrofolate, and (iv) an activating enzyme that reduces the cobalt of cobalamin back to the catalytically active +1 state in case of occasional oxidation to the inactive +2 state (7; 8). Cobalamin-dependent isomerases are responsible for the rearrangement of carbon skeletons (9). Adenosylcobalamin is usually the required cofactor of this class and it is used to generate substrate radicals as a prerequisite for the catalysed reaction (10). Less characterized cobalamin-dependent enzymes are the reductive dehalogenases, which have only been found in anaerobic organohalide-respiring bacteria (11; 12; 13). Reductive dehalogenases catalyse the terminal electron-transfer step in this pathway, which leads to dechlorination (14). The ability of cobalamin to effectively catalyse this diverse range of reactions can be attributed to its inherent properties. Firstly, the central Co atom can adopt three different oxidation states (+1, +2 and +3) and is thus able to span a wide range of electron potentials. Furthermore, the upper axial ligand of cobalamin can be substituted by several functional groups (methyl, hydroxo, cyano or adenosyl; (15), depending on
the reaction being performed. The first crystal structure of a cobalamin-binding protein to be solved was that of the *Escherichia coli* methionine synthase (MetH), which catalyses the formation of methionine from homocysteine by a methyl-group transfer via methylcobalamin (16). More recently, the crystal structure of MtaC, an enzyme that transfers a methyl group from methanol to coenzyme M, was solved (17). In both cases the cobalamin-binding domain adopts a Rossmann fold and the cofactor is bound in a base-off/His-on conformation. Structures of adenosyl-cobalamin isomerases have also been solved, the first of which, the methylmalonylcoenzyme A mutase, revealed that the cobalamin cofactor likewise binds to a Rossmann-fold domain that positions the cobalamin into a substrate-binding triosephosphate isomerase (TIM) barrel domain (18). No structural information is currently available for any of the reductive dehalogenases. Here, we provide the first crystal structure of a previously uncharacterized cobalamin-binding protein (termed CobDH) from *D. hafniense* DCB-2, which is implied to be a central part of the O-demethylation enzyme system based on its operon content. CobDH was crystallized in the presence of methylcobalamin and the structure was refined to 1.5 Å resolution. Despite its relatively low sequence homology (28%), the crystal structure of CobDH revealed significant structural homology to previously determined methyltransferase structures. The structural data are corroborated by EPR and UV-Vis spectroscopy and show that the cobalamin bound to CobDH can populate the three oxidation states +1, +2 and +3.

### 6.3 Materials and methods

#### 6.3.1 Cloning, expression and purification

The *Dhaf0720* gene was synthesized with codon optimization for expression in *E. coli* by MWG operon. PCR amplification was performed with the primers

5’-AAGTTCTGTGTTCAGGGCCCGGTCATCGACCTGAATGCG- 3’ and

5’-ATGGTCTAGAAAGCTTTAACCTACCCAACGTTGGC-3’ containing 5’ overhangs that are compatible with ligation-independent cloning (TaKaRa Bio Inc.). The gene was inserted into the pOPINF vector (OPPF, UK) encoding an N-terminal hexahistidine tag and an HRV 3C recognition sequence (LEVLFQ/GP) using the restriction sites *Hind*III and *Kpn*I. *E. coli* BL21 (DE3) cells (Merck) were transformed with the generated plasmid. The cells were grown in LB medium at 310 K and shaken until they reached mid-log phase. At this point the temperature was lowered to 295 K and protein expression was induced by the addition of IPTG (1 mM final concentration). Cells were harvested after 12-16 h by centrifugation; they were resuspended in buffer A (50 mM HEPES pH 8, 500 mM NaCl, 5% (v/v) glycerol, 30 mM imidazole) supplemented with DNase, RNase, lysozyme (Sigma) and Complete EDTA-free protease inhibitor (Roche) and stirred on ice for 30 min. Methylcobalamin (Sigma) was added to the cells before they were lysed.
by sonication (Bandelin Sonopuls) and insoluble cell membranes were removed by ultracentrifugation at 98 000 g for 45 min at 277 K. The soluble crude extract was loaded onto a 5 ml Ni-NTA column equilibrated with buffer A. After loading, the column was washed with ten column volumes of buffer A in order to remove nonspecifically bound protein. CobDH was eluted using a linear gradient to buffer B (buffer A plus 270 mM imidazole). Fractions containing CobDH were pooled, concentrated to 10 ml and then diluted to 100 ml with 50 mM HEPES pH 8. The sample was then directly loaded onto a 20 ml Mono Q anion-exchange column and eluted with a linear gradient of 0-500 mM NaCl in 50 mM HEPES pH 8. To cleave off the His tag, fractions containing the target protein were pooled, concentrated and incubated for 16-20 h with HRV 3C protease (Novagen) at a 1:100 molar ratio. The cleaved His tag was removed by reverse Ni-NTA affinity chromatography before a final purification step on an S200 gel-filtration column equilibrated with 10 mM HEPES pH 8, 100 mM NaCl. In order to assure complete cofactor reconstitution, some excess methylcobalamin was added to the sample prior to the gel-filtration step.

6.3.2 Crystallization and cryoprotection

Holo CobDH was concentrated to ∼15 mg/ml in 10 mM Hepes pH 8, 100 mM NaCl using a 10 kDa molecular-weight cutoff spin concentrator (Sartorius). Crystallization screens were performed by the sitting-drop vapour-diffusion method and drops were dispensed (200 nl protein solution plus 200 nl mother liquor equilibrated against 50 µl mother liquor in the reservoir) in a 96-well format using a high-throughput liquid-handling robot (Mosquito MD11-11, Molecular Dimensions). Small needle-shaped crystals grew from 0.01 M zinc chloride, 0.1 M HEPES pH 7, 20% (w/v) PEG 6000 within 7 d. The entire drop containing these small crystals was resuspended in 40 µl of the mother liquor from the reservoir. The crystals were crushed in this solution by the addition of a microseed bead (Molecular Dimensions) and rigorous vortexing for 2 min. The microcrystals thus generated were used without further dilution as nucleation seeds in a second round of screening using the sitting-drop vapour-diffusion method and the PACT premier matrix screen (Molecular Dimensions) as described by D’Arcy et al. (19). Seed drops consisted of 200 nl purified CobDH (in the same buffer and at the same concentration as used for the initial crystallization screen) plus 100 nl nucleation seed plus 300 nl mother liquor and were equilibrated against 50 µl mother liquor in the reservoir well.

This microseeding protocol led to the growth of single rectangular-shaped crystals within 48 h in a condition consisting of 0.1 M MMT (D-malic acid, MES and Tris base in a 1:2:2 molar ratio) buffer pH 6, 25% (w/v) PEG 1500. The crystals were cryoprotected by a direct and brief transfer into cryobuffer (16% (v/v) glycerol, 16% (v/v) ethylene glycol, 18% (w/v) sucrose, 4% (w/v) glucose) before being flash-cooled in liquid nitrogen.
6.3.3 Data collection and processing

A complete data set was collected from a single crystal on beamline I02 at the Diamond Light Source synchrotron-radiation facility. Automated data processing was performed by xia2 (20). The crystals belonged to space group $P3_1$, with unit-cell parameters $a = 69.99$, $b = 69.99$, $c = 91.68$ Å, and contained two molecules in the asymmetric unit.

6.3.4 Structure determination

The structure was solved by molecular replacement with Phaser (21) using the structure of a cobalamin-binding protein from *Moorella thermoacetica*, a protein homologous to the C-terminus of CobDH (PDB entry 1Y80; Southeast Collaboratory for Structural Genomics, unpublished work; (22; 23), as a search model and was refined with REFMAC (24; 25). The missing 86 residues of the N-terminal domain could be clearly identified in the electron-density maps which were built with several cycles of automated model building with Buccaneer (26). The cobalamin cofactor was manually built into the structure afterwards, followed by iterative rounds of manual building in Coot (27) and crystallographic refinement in REFMAC (24). An updated B12.cif parameter file was used for refinement of the cobalamin cofactor because the current definition file in REFMAC contains an incorrect double bond between C19 and N24 of the cobalamin corrin ring instead of a single bond. The final quality of the structure in complex with cobalamin was validated using PROCHECK (28) and MolProbity (29). Data-processing and refinement statistics are given in Table S1. Crystallographic figures were generated using PyMOL http://www.pymol.org.

6.3.5 UV-Vis spectroscopy

UV-Vis absorbance spectra were recorded using a Cary UV-Vis spectrophotometer. All spectra were baseline corrected with the respective buffer solution. Spectra of the protein reconstituted with methylcobalamin were recorded in the fully oxidized cob(III)alamin form (under aerobic conditions). Anaerobic spectra were recorded within an anaerobic glove box (Belle Technology, $O_2 < 5$ p.p.m.) and the bound cobalamin was reduced with titanium(III) citrate. Titanium (III) citrate was prepared in an anaerobic glove box by the addition of sodium citrate to a stirred anaerobic 12% solution of titanium(III) chloride (Sigma) at room temperature (30). This solution was neutralized by the addition of saturated sodium bicarbonate, yielding a final titanium(III) citrate stock solution of approximately 50 mM. CobDH (1 ml at 50 µM) was stepwise reduced by the addition of 0.5 ml aliquots of titanium(III) citrate. The low redox potential of titanium(III) citrate (∼ -500 mV) was sufficient to reduce cobalamin to the +1 cobalt oxidation state. The percentages of cobalamin in the different oxidation states were estimated using the extinction coefficients $\epsilon = 26 000$ M$^{-1}$ cm$^{-1}$ for cob(I)alamin at 388 nm, $\epsilon = 9470$ M$^{-1}$
cm\(^{-1}\) for cob(II)alamin at 477 nm and \(\epsilon = 9100\ M^{-1}\ cm^{-1}\) for cob(III)alamin at 540 nm \((30)\) (Goulding et al., 1997) and a CobDH protein concentration of 50 \(\mu\)M.

6.3.6 EPR spectroscopy

The EPR samples were prepared in the same manner as the UV-Vis samples. Protein concentrations were adjusted to \(\sim 2\) mg/ml (80 \(\mu\)M). Protein samples were reduced with 5 mM sodium dithionite in order to obtain the paramagnetic cob(II)alamin species. Samples were transferred aerobically into 4 mm Suprasil quartz EPR tubes (Wilmad), directly frozen and stored in liquid nitrogen until measured. EPR spectra were recorded at 12 and 30 K using a Bruker ELEXSYS E500/E580 EPR spectrometer (Bruker GmbH) fitted with an Oxford Instruments ESR900 helium-flow cryostat coupled to an ITC 503 controller from the same manufacturer. The microwave power was 0.5 mW, the modulation frequency was 100 kHz and the modulation amplitude was 5 G. The \(g\) values given were calculated using the Xepr software package supplied with the instrument.

6.4 Results and discussion

6.4.1 Crystal structure of CobDH

Recombinant production of soluble CobDH was successfully achieved in \(E.\ coli\) as described in the materials and methods. The protein consists of 212 residues on a single polypeptide chain, corresponding to a calculated molecular weight of 22.5 kDa, which is in line with SDS-PAGE analysis of purified CobDH (figure 6.1).

Given the similarity of CobDH to other cobalamin-binding proteins and the presence of the cobalamin-binding motif DxHxxG (residues 100-105; figure 6.2 A), it was anticipated that CobDH would bind cobalamin. CobDH was reconstituted during the purification process with exogenous methylcobalamin, as the heterologous \(E.\ coli\) host.
only synthesizes methylcobalamin when supplied with cobinamide (31). Reconstitution was verified by the bright pink colour of purified CobDH, which is a characteristic absorbance feature of cob(III)alamin, indicating successful in vitro reconstitution of the protein with methylcobalamin.

Microseeding with small CobDH crystals obtained from initial screening led to the growth of large rectangular crystals which diffracted to a resolution of 1.5 Å. A complete data set was recorded from a single crystal on beamline I02 at the Diamond Light Source. The data were indexed in space group $P\overline{3}_1$, with unit-cell parameters $a = b = 69.99$, $c = 91.68$ Å. Data-collection and refinement statistics are shown in table S1. The structure was solved by molecular replacement using the crystal structure of a cobalamin-binding protein from \textit{M. thermoacetica} (PDB entry 1Y80) as a search model. Two CobDH monomers were found in the asymmetric unit. The overall high-resolution structure of CobDH (figure 6.3 A-B) reveals two domains that are separated by a flexible linker. The N-terminal domain (residues 1-82) folds into an antiparallel four-helix bundle. The C-terminal domain (residues 90-212) resembles a typical Rossmann-fold with a central five-stranded parallel $\beta$-sheet surrounded by six $\alpha$-helices that form the binding site for the methylcobalamin.

The corrin ring and the central Co atom are presented on the surface of the protein, while the dimethylbenzimidazole (DMB) base protrudes deep into the central $\beta$-sheet region. The planar corrin ring is oriented perpendicular with respect to the parallel $\beta$-sheet core. His102 forms a coordination bond towards the Co atom from the lower axial site, which results in a base-off/His-on cobalamin-binding form. No electron density has been observed at the upper axial ligand site from the Co atom. This finding leads to the conclusion that the methyl group in the crystals has been cleaved off by photoreduction during crystallization, cryoprotection or X-ray exposure, a process that has also been described for other crystal structures of methylcobalamin-dependent enzymes (17; 33), resulting in the formation of cob(II)alamin. It is hence assumed that the oxidation state of cobalamin in the described CobDH crystals is +2. His102 is part of a loop that connects the most N-terminal $\beta$-strand of the Rossmann-fold domain to the first $\alpha$-helix of the domain. Structural alignments show that the entire loop is highly conserved, confirming its key role in cobalamin binding (figure 6.2). His102 NE2 forms a coordination bond with the central Co atom of cobalamin at a distance of 2.5 Å. ND1 from the same histidine side chain hydrogen-bonds to the side-chain O atom of the conserved Asp100. The side chain of the conserved residue Ser147 forms a hydrogen bond to an N atom of the DMB from the cofactor (2.7 Å distance), as shown in figure 6.4. No electron density was observed corresponding to the long linker region which connects the two domains (residues 83-89), suggesting a high degree of flexibility within the linker. This suggests that both domains might possibly undergo structural rearrangements with respect to each other as part of the catalytic cycle of CobDH. However, there are notable interactions between the domains, suggesting that the observed conformation in
Figure 6.2: Primary-sequence alignment and superposition of CobDH and homologous proteins for which crystal structures have been solved. A) Multiple sequence alignment of all deposited proteins that bind methylcobalamin in the base-off/His-on form. The histidine side chain from the conserved cobalamin-binding motif DxHxxG forms a coordination bond with the central Co atom in all cases. Conserved residues of CobDH (Asp100, His102 and Ser147) are marked with asterisks. The alignment was performed using ClustalX (32). B) Stereoview of the superposition of CobDH (orange) with MetH (PDB entry 1BMT; magenta; r.m.s. = 0.774 Å; (16) and MtaC (PDB entry 2I2X; blue; r.m.s. = 0.772 Å; (17). Protein backbone traces are shown in Cα trace representation and the cobalamins and bound histidines are shown as sticks. In MetH the N-terminal helix-bundle domain covers the cobalamin from the top, whereas in CobDH and MtaC it is oriented parallel to the cobalamin-binding domain.
Table 6.1: Crystallographic data and refinement statistics. Values in parentheses are for the highest resolution shell.

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source</td>
<td>Beamline I02, Diamond</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.946</td>
</tr>
<tr>
<td>Space group</td>
<td>P\textsuperscript{3} \textsubscript{1}</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>Unit-cell parameters (Å)</td>
<td>(a = b = 69.99, c = 91.68)</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>32.7-1.5 (1.54-1.50)</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>79674 (3997)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>9.3 (8.6)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.0 (97.3)</td>
</tr>
<tr>
<td>(R\text{merge})\textsuperscript{a} (%)</td>
<td>10 (85)</td>
</tr>
<tr>
<td>(R\text{meas})\textsuperscript{b} (%)</td>
<td>11.4 (106)</td>
</tr>
<tr>
<td>(CC_{1/2}) (%)</td>
<td>99.7 (61.1)</td>
</tr>
<tr>
<td>([I/\sigma(I)])</td>
<td>10.5 (3.5)</td>
</tr>
<tr>
<td>Wilson B factor (Å\textsuperscript{2})</td>
<td>18.50</td>
</tr>
<tr>
<td>No. of protein molecules in asymmetric unit</td>
<td>2</td>
</tr>
<tr>
<td>Matthews coefficient (V_M) Å\textsuperscript{3} Da \textsuperscript{-1}</td>
<td>2.51</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>50.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>32.7-1.5 (1.54-1.50)</td>
</tr>
<tr>
<td>(R\text{work}/R\text{free})\textsuperscript{c} (%)</td>
<td>14.7/19.4</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>3247</td>
</tr>
<tr>
<td>No. of water atoms</td>
<td>448</td>
</tr>
<tr>
<td>Overall B factor (Å\textsuperscript{2})</td>
<td>26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Favoured (%)</td>
<td>98.26</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>1.49</td>
</tr>
<tr>
<td>R.m.s.d., bonds (Å)</td>
<td>0.021</td>
</tr>
<tr>
<td>PDB code</td>
<td>4JGI</td>
</tr>
</tbody>
</table>

\(R\text{merge}\) = \frac{\sum_{hkli} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkli} \sum_i I_i(hkl)}  \text{summands}

\(R\text{meas}\) = \frac{\sum_{hkli} \sqrt{N(hkl)}}{\sum_{hkli}} ∑_{i=1}^{N(hkl)} |I_i(hkl) - \langle I(hkl) \rangle|  \text{summands}

where: \(I_i(hkl)\) are the observed intensities; \(\langle I(hkl) \rangle\) is the average intensity; \(N(hkl)\) is the multiplicity of the reflection \(hkl\)

\(R\text{work}\) = \frac{\sum_{hkli} |F_{obsi} - |F_{calc}||}{\sum_{hkli} |F_{obs}|}

\(R\text{free}\) is the cross-validation \(R\) factor for the test set (5%) of reflections omitted from refinement.
Figure 6.3: Overall structure of CobDH (cartoon representation) with bound methylcobalamin. (A, B) The N-terminal helix-bundle domain is shown in red and the C-terminal Rossmann-fold domain is shown in orange. The central Co atom of the cofactor (shown as sticks; C atoms coloured grey) is coordinated to His102 as the lower axial ligand. The black dotted line connects the domains to each other and represents the missing linker region for which no electron density was observed (residues 83-89). (C) Domain interactions between the helix-bundle domain (red) and the cobalamin-binding domain (orange). Asp54 and Asn198 form a hydrogen bond; Met46, Leu70 and Met112 are involved in hydrophobic interactions.
the crystal structure is of physiological relevance. This is further supported by the fact that the crystal packing of CobDH does not influence the position of the N-terminal four-helix bundle domain. The average B factors of the four-helix bundle domain and the Rossmann-fold domain are 29.8 and 24.04 Å², respectively, suggesting that the four-helix bundle domain is less restrained by the lattice. The interactions between both domains are centred around the side chains of Asp54 and Asn198, which form a hydrogen bond (2.9 Å distance), and hydrophobic interactions between the domains that are formed between the side chains of Met46, Leu70, Ile108 and Met112 (figure 6.3 C).

6.4.2 Spectroscopic analysis

The UV-Vis absorbance spectrum of reconstituted CobDH (figure 6.5 A) in the oxidized state revealed characteristic features of methyl(III)cobalamin (540 nm maximum), in addition to peaks at 520, 410 and 350 nm that have previously been observed for hydroxo(III)cobalamin (34). These observations indicate that the coordination bond between the cobalamin Co atom and the upper axial methyl ligand is cleaved to a large extent, leading to the formation of hydroxo(III)cobalamin or aquacobalamin. This cleavage can be attributed to the exposure of the sample to light and oxygen during the protein-purification steps. Under reducing and anaerobic conditions (the addition of titanium(III) citrate in an anaerobic glove box), distinct changes occur in the spectra. The features at 540, 520, 410 and 350 nm disappear and are replaced by cob(II)alamin-specific and cob(I)alamin-specific peaks at 477 and 388 nm. Quantification of the peaks at 540, 477 and 388 nm was performed in order to estimate the percentages of cobalamin-bound protein in the +3, +2 and +1 oxidation states, respectively. The inset in figure 6.5 A shows a plot of these percentages versus the amount of titanium(III) citrate added, revealing that the percentage of cob(III)alamin decreases from approximately 53 to 22%, whereas the percentages of cob(II)alamin and cob(I)alamin increase from 37 to 60% and from 8 to 18%, respectively. The calculated percentages are in good agreement with the assumption of complete reconstitution of CobDH with cobalamin. Similar absorbance changes have been observed for *E. coli* MetH during methyl transfer from methyltetrahydrofolate (methyl donor) to exogenous cob(I)alamin and from methyl (III)cobalamin to homocysteine (methyl acceptor; (30). EPR analysis (see figure 6.5 B) was used to further characterize the cofactor-binding mode in CobDH. The EPR spectrum of CobDH in the absence of any reductant (upper spectra in figure 6.5 B, prepared under aerobic conditions) exhibits the characteristic features of a cob(III)alamin-superoxide complex, which gives rise to an EPR signal at \( g = 2.0 \) owing to the presence of an unpaired electron that is predominantly located on the complexed oxygen molecule (35). Upon the addition of a suitable reductant (here sodium dithionite), the Co atom is reduced to the +2 state and exhibits the characteristic features of a base-off/His-on form at \( g = 2.26 \), which is in agreement with the position of His102 observed in the crystal structure. Our spectro-
Figure 6.4: Cobalamin-binding site in CobDH. The cobalamin and the conserved residues Asp100, His102 and Ser147 are shown as sticks (C atoms coloured grey) and are labelled using the single-letter amino-acid codes for clarity. Asp100 hydrogen-bonds to His102 (2.7 Å distance), which in turn is coordinated to the Co atom of the cofactor from the lower axial site (2.5 Å distance). The side chain of Ser147 forms a hydrogen bond to an N atom of the dimethylbenzimidazole (DMB) of the cobalamin (2.7 Å distance). The interactions described are indicated with green dotted lines. The blue mesh corresponds to a $2F_o - F_C$ map with the cobalamin and the key residues omitted from map calculations. The map was contoured at 2.0 $\sigma$. 
Chapter 6. CobDH

Cobalamin-dependent enzyme systems

Figure 6.5: UV-Vis and EPR analyses of CobDH. A) Anaerobic titration of oxidized CobDH with titanium(III) citrate as a reductant. The absorbance maximum of cob(III)alamin at 540 nm decreases with increasing amounts of added reductant; the absorbance of cob(II)alamin (474 nm) and cob(I)alamin (388 nm) increase correspondingly. The inset shows a plot of the percentage of CobDH in each of the three possible oxidation states of its cofactor during titration with titanium(III) citrate. The respective oxidation-state percentages were calculated as described in the methods section. The cob(III)alamin percentage in CobDH decreases from approximately 53 to 22%, whereas the percentages of cob(II)alamin and cob(I)alamin increase from 37 to 60% and from 8 to 18%, respectively. B) The oxidized EPR sample refers to CobDH as purified. The cob(III)alamin-superoxide complex signal has a g value of 2.0 (magnetic field of 3400 G; upper spectra). After sample reduction with 5 mM sodium dithionite the spectrum shows the presence of the base-off/His-on form with a main feature at g = 2.26 (magnetic field of 2900 G; lower spectra). Also present is hyperfine splitting of the signal, which arises from the interaction between the cob(II)alamin and its β-axial coordinated N atom from the His102 side chain.

Scoposcopic data indicate that the CobDH-bound cobalamin is able to cycle between different configurations: cob(I)alamin is tetracoordinated (lacking both the lower and the upper axial ligands), the paramagnetic cob(II)alamin species is pentacoordinated by His102 from the lower axial site to the central cobalt and cob(III)alamin is usually diamagnetic, but our EPR spectra suggest the presence of a paramagnetic cob(III)alamin-superoxide complex. To our knowledge, protein-bound cob(III)alamin-superoxide complexes have only previously been identified for E. coli MetH (36).

6.4.3 Comparison of CobDH with other methyltransferases

Figure 6.2 shows a multiple sequence alignment between CobDH (PDB entry 1Y80; the molecular-replacement model used) and other close homologues in the PDB. All structures were solved bound to methylcobalamin, except for the protein from Methanosarcina barkeri (PDB entry 3EZX; R. Jain, B. Hao, J. A. Soares, L. Zhang, K. Green-Church, X. Li, J. A. Krzycki & M. K. Chan, unpublished work), which was solved bound to hydroxocobalamin. The sequence alignment reveals that only the cobalamin-binding region shows high conservation (6). The cobalamin-binding motif DxHxxG is the only
common primary-sequence motif that is conserved between the methyltransferases and isomerases, while it is absent in the reductive dehalogenases. All aligned structures contain two domains. The structurally conserved C-terminal cobalamin-binding domain is responsible for the correct positioning of the cobalamin corrin ring on the surface of the Rossmann-fold domain. Hence, it is easily accessible to other protein modules that are required for catalysis, for example methyl-donor and/or methyl-acceptor domains and their respective ligands (e.g. 5-methyltetrahydrofolate, methanol, homocysteine, coenzyme M etc.). In contrast, the function of the second domain, the N-terminal four-helix bundle, is less well understood. The high flexibility of the linker domain between the domains could promote the structural rearrangement of both domains during the catalytic cycle with respect to each other. The orientation of the N-terminal four-helix bundle domain with respect to the C-terminal domain is indeed variable across the family.

The α-helices of this N-terminal domain can either be oriented parallel to the β-strands of the cobalamin-binding domain or the N-terminal domain can shield the cofactor upper axial site (as is the case in MetH; figure 6.2 B) from the environment, thus preventing putative interactions between solvent molecules and the attached methyl group (16).

It has been shown crystallographically for MetH that the N-terminal four-helix bundle domain moves 26 Å away from the cobalamin in order to allow interactions between cobalamin and the activation domain (AdoMet binding), which is the most C-terminal domain of MetH (37). MetH is a large multidomain protein with four distinct modules, whereas CobDH and MtaC only comprise the cobalamin-binding domain, which might further explain the different positions of the four-helix bundle domain with respect to the cobalamin cofactor. For CobDH and MtaC this domain is parallel to the cobalamin-binding domain and does not interact with the cofactor (figure 6.2 B). Furthermore, in the case of the methanol MtaABC system it has been proposed that the helix-bundle domain of MtaC is permanently anchored to the methanol-binding protein MtaB, allowing the cobalamin-binding domain of MtaC to rotate and bind to either the methanol-binding region of MtaB or the methyl donor region of MtaA, which is the third protein involved in the methanol-cleavage reaction (17; 38). It would be interesting to find out whether such major movements of the helix-bundle domain also occur at some stage along the catalytic reaction coordinate or during reactivation processes in the multiprotein O-demethylase system. Bioinformatic studies propose that CobDH is part of a multicomponent O-demethylase enzyme system (5). The analysis indicates that in the same operon as CobDH (gene Dhaf0720), the genes Dhaf0721 and Dhaf0722 code for methyl donor (substrate-binding) and methyl acceptor (tetrahydrofolate-binding) domains, respectively. This scenario promotes a multi-component arrangement of the O-demethylase system similar to the methanol-cobalamin methyltransferase complex (17) and would hence explain the parallel orientation of the four-helix bundle to the cobalamin-binding domain in CobDH.
6.5 Conclusions

We have identified and isolated a previously unknown cobalamin-binding protein (CobDH) from *D. hafniense* and solved its crystal structure. The structure reveals the characteristic base-off/His-on cobalamin-binding site with the cobalamin coordinated to His102 as part of a conserved sequence motif observed in other cobalamin-dependent methyltransferases and implies that CobDH likewise catalyses the transfer of a methyl group. EPR and UV-Vis spectroscopic analyses of CobDH confirmed the cobalamin binding and revealed that the CobDH-bound cobalamin cofactor is able to cycle between the cobalt(I), cobalt(II) and cobalt(III) oxidation states. However, as with the other methyltransferases studied, additional proteins or protein domains are needed to bind the substrates and products and present these to the upper axial position of the cobalamin moiety for catalysis. The genomic context of the CobDH gene suggests that it is part of a three-component *O*-demethylase enzyme system. *O*-Demethylase activity has been observed in *D. hafniense* strains DCB-2 and PCP-1 and recent studies have established *O*-demethylase activity for a related operon consisting of Dhaf4610, Dhaf4611 and Dhaf4612 when expressed in *E. coli* (5). This suggests that organohalide-respiring bacteria not only use cobalamin to support the dechlorination of halogenated compounds, but also make extensive use of this cofactor in catalysis of demethylation reactions, leading to further mineralization. With respect to bioremediation strategies and a recent study (2), it would be especially interesting to biochemically confirm the presence of shared substrates in *D. hafniense* (or any other organohalide-respiring bacteria) between *O*-demethylases and reductive dehalogenases. We hence intend to investigate the functional relationships between the Dhaf0720-0722 gene products on both a biochemical and a structural level to further explore the versatility of cobalamin-catalysed enzymatic reactions in organohalide-respiring bacteria.

This work was supported by the European Research Council (ERC) starting grant DEHALORES206080. The authors would like to thank the reviewers for valuable and critical comments on the manuscript.

6.6 References


Cobalamin-dependent enzyme systems


Chapter 7

Crystal structures of the methyltransferase protein from *Desulfitobacterium hafniense* DCB-2 that catalyses the methyl group transfer from methylcobalamin to tetrahydrofolate in O-demethylation

Author’s: Hanno Sjuts, Mark S. Dunstan, Karl Fisher and David Leys

Affiliation: Manchester Institute of Biotechnology, Faculty of Life Sciences, The University of Manchester, M1 7DN, UK

Manuscript in preparation for submission to PLoS ONE.
7.1 Abstract

$O$-demethylation refers to a bacterial pathway that is involved in the carbon metabolism of e.g. acetogenic and organohalide-respiring bacteria. The product of $O$-demethylation is the formation of methyltetrahydrofolate from aromatic methyl ethers. Methyltetrahydrofolate is ultimately used to synthesize acetyl CoA for central metabolic pathways. The demethylation reaction is catalysed by $O$-demethylases, which are cobalamin-dependent, three-component enzyme systems that catalyse methyl group transfers from aromatic methyl ethers to tetrahydrofolate via methylcobalamin intermediates. All three enzyme components involved in $O$-demethylation reside on individual polypeptide chains. The first protein binds the aromatic methyl ether, the second protein binds cobalamin and the third protein binds the tetrahydrofolate, respectively. In the first methyl transfer reaction (MT1) the methyl group from the aromatic methyl ether is transferred to cob(I)alamin, which leads to the formation of methylcobalamin and demethylated substrate. In the second methyl transfer reaction (MT2) the methyl group from methylcobalamin is transferred to tetrahydrofolate, thus forming methyltetrahydrofolate and restored cob(I)alamin. Recently, the crystal structure of the cobalamin-binding protein from an $O$-demethylase in Desulfotobacterium hafniense DBC-2 has been described (Sjuts et al., Acta Cryst. D (2013) 69, 1609-1616). In this study, crystal structures of the tetrahydrofolate binding protein (MT2DH) from the same $O$-demethylase are presented in the apo form as well as in complex with tetrahydrofolate and methyltetrahydrofolate. The structures reveal that the methyl transfer reaction mechanism from methylcobalamin to tetrahydrofolate may differ significantly from the reverse reaction (methyltetrahydrofolate to cob(I)alamin). Based on the crystal structures, a reaction mechanism is proposed in which a highly coordinated water molecule forms a proton relay system with Ser198 and which is able to abstract a proton from the positively charged methyltetrahydrofolate N5 nitrogen atom. Hence, the conserved Asn199 residue that is involved in transition state stabilisation in other methyltransferases is displaced and the reverse reaction is avoided.

Keywords: Methyltransferase ♦ MT2DH ♦ X-ray crystallography ♦ Tetrahydrofolate ♦ Methyl transfer
7.2 Introduction

Organohalide-respiring bacteria are primarily known for their ability to dehalogenate and thus detoxify chlorinated organic compounds (organohalides) in a respiratory manner (1; 2). In organohalide respiration the organohalides serve as terminal electron acceptors. The final electron transfer step is catalysed by Fe-S cluster and cobalamin-dependent enzymes, the reductive dehalogenases (3). Since a few years, and accelerated by rapid genome sequencing methods, it is becoming apparent that the genomes of specific organohalide-respiring bacteria also contain multiple gene clusters that encode O-demethylase enzyme systems (4; 5; 6). O-demethylases were previously identified in acetogenic bacteria (7; 8) and are members of the cobalamin-dependent methyl transfer enzyme class (9). Cobalamin-dependent methyl transfer enzymes catalyse the methyl group transfer from a methyl group donor to a methyl group acceptor via methylcobalamin intermediates. In the first methyl transfer reaction (MT1), the methyl group is transferred from the substrate binding domain to protein bound cob(I)alamin, which leads to the formation of methylcobalamin. In the second methyl transfer reaction (MT2) the methyl group is transferred from methylcobalamin to the final methyl group acceptor which is bound to the acceptor domain, thus forming the methylated product and restored cob(I)alamin (10). The different substrates and coenzymes involved in cobalamin-dependent methyl transfer enzymes bind to different protein domains, which can either be connected into a single polypeptide chain (for example methionine synthase), or reside on individual protein chains (11). In a broader context, cobalamin-dependent methyl transfer reactions are of pivotal importance in the carbon metabolism of mainly anaerobic bacteria (9). Substrates for O-demethylases are aromatic methyl ethers (e.g. vanillate or syringate) and the final methyl group acceptor is tetrahydrofolate (THF), a folic acid derivative. Methylation of THF occurs at nitrogen N5 of the pterin moiety of the THF cofactor and results in the formation of N5-methyltetrahydrofolate (MTHF) (12). Three different proteins are directly involved in O-demethylation:

1. A substrate binding protein.
3. A THF binding protein which is the methyl group acceptor.

A fourth protein is thought to reactivate the cobalamin cofactor in case of its occasional oxidation into the inactive +2 redox state, by reducing the cobalt atom into the +1 state in an ATP dependent manner (13; 14). Scheme 7.1 illustrates the reaction scheme and involved protein modules in O-demethylation. A few of such three-component O-demethylases have been isolated and characterised biochemically. The Moorella thermoacetica O-demethylase system has been shown to use dicamba and vanillate as substrates (15). Similar observations have been made for O-demethylases from
Acetobacterium dehalogenans and Desulfotobacterium hafniense DCB-2, using similar substrates (4; 5; 16). Aromatic methyl ethers can additionally be highly chlorinated and it has therefore been suggested that, in organohalide-respiring bacteria, reductive dehalogenation and O-demethylation reactions occur, that share the same or similar substrates (17).

Interestingly, THF or MTHF can either serve as methyl group acceptor or methyl group donor, respectively, in different methyl transfer reactions (12). It is not known if methyltransferases use the same reaction mechanism and structural active site compositions in both methyl transfer directions. However, it has been observed that the methyl transfer from MTHF to cob(I)alamin is reversible in methionine synthase (MetH) (18). Several methyl transfer enzymes have been studied extensively that use MTHF as methyl group donor, including MetH and the methyltransferase (MeTr) of the corrinoid-iron-sulfur protein (CFeSP), which is involved in the Wood-Ljungdahl pathway of anaerobic CO$_2$ fixation. Crystal structures of these MeTr’s, which fold into triosephosphate isomerase (TIM) barrels, have been solved and have revealed that they contain a conserved Asn residue in the proximity of the nitrogen atom N5 of MTHF, which seems to be crucial for catalysis and which has been implicated in transition state stabilisation (19; 20; 21). So far, no methyltransferase structures have been described, in which the methyl group of methylcobalamin is transferred to protein bound THF, thus forming MTHF. In order to evaluate mechanistic and structural similarities and differences of this methyl transfer between the two directions, three-dimensional structures of the methyl group acceptor protein (MT2DH) from a O-demethylase from Desulfotobacterium hafniense DCB-2 in the apo form and in complex with THF and MTHF have been solved by X-ray crystallography. MT2DH is from the same O-demethylase whose cobalamin-binding domain structure has recently been described (22). The solved MT2DH crystal structures in-

Figure 7.1: Schematic illustration of the O-demethylation reaction and the involved protein components. The methyl donor (substrate) binding domain is shown in orange, the cobalamin-binding domain catalysing the methyl transfer reaction in magenta and the THF binding methyl acceptor domain is shown in blue. MT I and MT II present the two iterative methyl transfer reactions that occur during O-demethylation.
dicate that the above mentioned Asn residue (Asn199 in MT2DH), does not seem to be important for the methyl group transfer in this O-demethylase. Instead, a water molecule hydrogen bonds to the nitrogen N5 atom in both the THF and MTHF structure. Ser198, in turn, hydrogen bonds to this water molecule and it is proposed that both the water and Ser198 form a proton relay system that abstracts a proton from N5, thus preventing the reverse methyl transfer reaction from MTFH to THF to take place.

7.3 Materials and methods

7.3.1 Cloning and overexpression of MT1DH and MT2DH

PCR amplifications of the Dhaf0721 (MT1DH) and Dhaf0722 (MT2DH) genes were performed with the primers:

5'-CGCCGGCGCAGCTATATGAGTGAGTGGGATCTCGAAACC-3' = 21 forward
5'-GCTCGAATTCGGATCCTCAGCTGTAGATACCG-3' = 21 reverse
5'-CGCCGGCGCAGCTATATGAGTGAGTGGGATCTCGAAACC-3' = 22 forward
5'-GCTCGAATTCGGATCCTCAGCTGTAGATACCG-3' = 22 reverse

Genomic DNA from Desulfotobacterium hafniense DCB-2 was used as PCR template, which was isolated from a 100 ml bacterial culture. Cells were pelleted by centrifugation (6,000 g, 10 min, 4 °C) and yielded ~ 450 mg of cell pellet (wet cell weight). Genomic DNA extraction was performed with the PROMEGA Kit following the manufacturers guidelines. The DNA was eluted in a total volume of 150 µl aliquoted in 10 µl and stored at - 80 °C until needed for subsequent PCR reactions. PCR amplified genes contain 5’ overhangs that are compatible with ligation-independent cloning (TaKaRa Bio Inc.). Both genes were inserted into the pET28a(+) vector (Novagen) encoding an N-terminal His-tag and a thrombin protease cleavage site using the restriction sites NdeI and BamHI. E. coli BL21 (DE3) cells (Merck) were transformed with the generated plasmids. The cells were grown (750 ml cell culture in 2 litre flasks) in LB medium (Formedium) including Kanamycin (50 µg/ml as selective reagent) at 37 °C and shaken at 200 rpm until they reached mid-log phase. At this point the temperature was lowered to 22 °C (in the case of MT2DH) and 20 °C (in the case of MT1DH). After the flasks were cooled down, protein overexpression was induced by the addition of IPTG (0.2 mM final concentration). Cells were harvested after 12 - 16 h by centrifugation and stored at - 20 °C.

Purification of MT2DH

Cell pellets were resuspended in buffer A (50 mM Tris pH 7.5, 300 mM NaCl, 5% (v/v) glycerol, 30 mM imidazole) supplemented with DNase, RNase, lysozyme (Sigma) and Complete EDTA-free protease inhibitor (Roche) and stirred on ice for 30 min. Cells were lysed by sonication (Bandelin Sonopuls instrument) and insoluble cell membranes
were removed by ultracentrifugation at 98 000 \( g \) for 45 min at 4 °C. The soluble crude extract was loaded onto a 1 ml Ni-NTA column equilibrated with buffer A. After sample loading, the column was washed with ten column volumes of buffer A in order to remove nonspecifically bound protein. MT2DH was eluted using a linear gradient to buffer B (buffer A plus 270 mM imidazole). Fractions containing MT2DH were pooled and concentrated using a 10 kDa molecular weight cut-off (MWCO) spin concentrator. MT2DH was then further purified by size exclusion chromatography using an S200 gel-filtration column equilibrated with 25 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT.

**Purification of MT1DH**

Cell pellets were resuspended in buffer A (50 mM Tris pH 9, 150 mM NaCl, 10% (v/v) glycerol) supplemented with DNase, RNase, lysozyme (Sigma) and Complete EDTA-free protease inhibitor (Roche) and stirred on ice for 30 min. Cells were lysed by sonication (Bandelin Sonopuls instrument) and insoluble cell membranes were removed by ultracentrifugation at 98 000 \( g \) for 45 min at 4 °C. The soluble crude extract was loaded onto a 1 ml Ni-NTA column equilibrated with buffer A. After loading, the column was washed with ten column volumes of buffer A in order to remove nonspecifically bound protein. MT1DH was eluted using a linear gradient to buffer B (buffer A plus 300 mM imidazole). Fractions containing MT1DH were pooled and directly desalted in the cold room using a PD10 column into buffer A, in order to avoid protein precipitation. Afterwards the protein was concentrated by a 30 kDa MWCO spin concentrator. MT1DH was then further purified by size exclusion chromatography using an S200 gel-filtration column equilibrated with 25 mM Tris pH 9, 150 mM NaCl and 10% (v/v) glycerol.

Overexpression and purification of CobDH was performed as previously described (22).

### 7.3.2 Crystallisation

**apo-MT2DH**

Full length MT2DH protein including N-terminal His\(_6\)-tag was concentrated to 7 - 8 mg/ml using a 10 kDa MWCO spin concentrator (Sartorius) in gel filtration buffer (10 mM Tris pH 7, 100 mM NaCl and 1 mM DTT). Crystallisation screens were performed at 4 °C by the sitting drop vapour diffusion method and drops were dispensed (200 nl protein plus 200 nl mother liquor over 50 \( \mu l \) mother liquor in the reservoir) using a high-throughput liquid handling robot (Mosquito MD11-11, Molecular Dimensions) in a 96 well format. Small rod shaped, needle like and rectangular crystals grew in multiple conditions from commercial crystallisation screens tested (PACT, JCSG, Morpheus, CSS1 and CSS2; all Molecular Dimensions). Additionally, plate shaped crystals grew within 72 hours in a precipitant solution containing 100 mM Tris pH 7.5, 100 mM succinic
acid and 15% (v/v) PEG 3350, which diffracted best. The plate shaped crystals were
cryoprotected by transferring them into mother liquor additionally containing 19% PEG
200. After ∼ one minute the crystals were looped and flash frozen and stored in liquid
nitrogen until measured.

**MTHF-MT2DH**

Lyophilised methyltetrahydrofolate (MTHF, Sigma) was added to a precipitation solu-
tion (to a final concentration of 1 mM) that yielded the best diffracting apo-MT2DH
crystals (100 mM Tris pH 7.5, 100 mM succinic acid, 15% (v/v) PEG 3350). Crystallisa-
tion trials (sitting drop, vapour diffusion technique, 2 µl total volume over 50 µ mother
liquor in the reservoir) were manually set up in a cold room at 4 °C with this precipitation
solution using different ratios of protein (7 - 8 mg/ml) to precipitant solution ranging
from 0.5 µl + 1.5 µl to 1.5µl + 0.5 µl. Again, plate like crystals grew within 72 hours
in these drops. Crystals were cryoprotected by transferring them into mother liquor
additionally containing 19% PEG 200 and a tiny amount of freshly dissolved MTHF
powder to ensure complete reconstitution of MT2DH with MTHF. After ∼ one minute
the crystals were looped and flash frozen and stored in liquid nitrogen until measured.

**THF-MT2DH**

Apo-MT2DH plate like crystals were grown as described above. A 8 µl cryoprotection
solution drop was prepared from the mother liquor additionally containing 19% PEG
200 and 4 mM THF (Schircks Laboratories, Switzerland). Single crystals were soaked
in this solution for 2 - 3 minutes before they were looped and flash frozen and stored in
liquid nitrogen until measured.

### 7.3.3 Data collection

Complete data sets were collected from single crystals on the beamlines I24 and I04
at the Diamond Light Source (DLS) synchrotron-radiation facility. Automated data
processing was performed by *xia2* (23). The crystals belonged to space group *P*2₁2₁2₁
and contained two molecules in the asymmetric unit. Detailed information and statistics
for the data collection of apo-, THF-, and MTHF-MT2DH are given in supplementary
table S1.

### 7.3.4 Structure determination

**apo-MT2DH**

The structure of apo-MT2DH was solved by molecular replacement with the *Balbes*
software (24), run through the *CCP4* software, by using the structure of a MTHF-
binding protein from *Clostridium thermoaceticum* (MeTrCh, PDB entry IF6Y), as a
search model and was refined with REFMAC (25; 26). After one round of manual model building in Coot, solvent water molecules were added using the automated ARP/wARP solvent building version (27). Iterative rounds of manual building were performed in Coot (28) following TLS (translation, libration, screwmotion anisotropy) refinement in REFMAC (25).

**THF-MT2DH and MTHF-MT2DH**

The MTHF-MT2DH dataset (autoprocessed fast.dp derived mtz file) was modified by Truncate to set negative intensities to zero. $R_{\text{free}}$ data were generated using the Freerflag (import) program in CCP4. Subsequently, MTHF-0722 phases were obtained by using the final apo-MT2DH dimer structure (excluding all solvent molecules) as a molecular replacement search model in Phaser (29). After one round of refinement in REFMAC, water molecules were added using the automated ARP/wARP solvent building version (27). The electron density of MTHF was clearly visible in the $2Fo - Fc$ density map when loaded into Coot. The MTHF coordinates were obtained by superimposing the coordinates of PDB entry 2OGY (moving structure) with the MTHF-MT2DH coordinates (reference structure) using the secondary structure matching (SSM) superpose tool in Coot. Superposed MTHF coordinates were copied into the MTHF-MT2DH pdb file and refined with REFMAC. The THF-MT2DH structure was solved in the same way as the MTHF-MT2DH. Both THF-MT2DH and MTHF-MT2DH were refined using TLS refinement in REFMAC (25). The final quality of all three MT2DH crystal structures were validated using PROCHECK (30) and MolProbity (31). Crystallographic figures were generated using PyMOL (http://www.pymol.org) or QtMG (CCP4).

### 7.3.5 MT2DH methyltransferase assay

Absorbance changes of the cobalamin cofactor associated with methyl group transfer from exogenous cobalamin to protein bound THF were monitored spectroscopically. The assay was performed within a glove box (Belle Technologies, $O_2 < 5$ p.p.m.) under anaerobic conditions and in the dark at 25 °C using a Cary UV-Vis spectrophotometer situated inside the glove box. DTT was omitted from the purification buffer for the methyltransferase assay and MT2DH was made anaerobic by buffer exchange within the glove box using a PD10 desalting column (Biorad). Measurements were performed in a 1 ml cuvette with a path length of 1 cm in 50 mM MES pH 6.8, 50 mM NaCl containing 5 $\mu$M MT2DH protein, 20 $\mu$M methylcobalamin and were started by the addition of 150 $\mu$M of THF. Spectra were recorded every 3 minutes for 30 minutes in total. THF methylation was monitored using the extinction coefficients 9470 M$^{-1}$ cm$^{-1}$ for cob(II)alamin at 477 nm and $\epsilon = 9100$ M$^{-1}$ cm$^{-1}$ for cob(III)alamin at 540 nm according to (18). No methyl transfer was observed in the absence of MT2DH protein.
7.3.6 Surface plasmon resonance (SPR) binding measurements

SPR measurements were performed to study the interaction between MT1DH and different aromatic methyl ethers. Measurements were performed using a *ProteOn XPR36 Protein Interaction Array System* (BioRad, Hercules, CA, USA), in phosphate buffered saline (PBS) buffer pH 7.9 containing 0.005% Tween 20. MT1DH was immobilised on a ligand channel of a GLH sensor chip using the amine coupling reagents 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDAC) and N-hydroxysulfosuccinimide (sulfo-NHS). MT1DH (0.5 µM) was prepared in acetate buffer pH 4.5 and injected onto the chip with a flow-rate of 30 µl/min. Remaining surface groups were deactivated with 1 M ethanolamine HCl. Each measurement consisted of six analyte (aromatic methyl ether) concentrations (in the range of 50 - 2000 nM) injected simultaneously for 60 seconds at a flow rate of 100 µl/min. The sensorgrams were processed for baseline alignment and reference channel subtraction. Kinetic analysis was performed by globally fitting assuming a 1:1 bimolecular reaction model to a set of six sensorgrams. The analysis was performed with the included BioRad software package.

7.4 Results and discussion

All three proteins from the *D. hafniense* gene operon *Dhaf0720 - Dhaf0722* were individually expressed in *E. coli* purified to homogenity as described above. The SDS-PAGE in supplementary figure S1 confirms the expected molecular weights of the respective proteins that are involved in O-demethylation. The following nomenclature is used: methyl donor protein = MT1DH; cobalamin-binding protein = CobDH; methyl acceptor protein = MT2DH.

7.4.1 Methyl group transfer from methylcobalamin to THF is catalysed by MT2DH

UV-Vis spectroscopy was used to demonstrate that the methyl group in the MT2 reaction is transferred from methylcobalamin (absorbance maximum at 540 nm) to THF bound to MT2DH. The associated absorbance changes are shown in figure 7.2 A. During the methyl transfer reaction cob(I)alamin is formed, however, due to the lack of a reductant in the assay mixture, cob(I)alamin is readily oxidised to cob(II)alamin (absorbance maximum of 477 nm). During the time course of the measurement the methylcobalamin signal at 540 nm is constantly decreasing, while the cob(II)alamin signal is developing. The oxidation from cob(I)alamin to cob(II)alamin is presumably occurring due to remaining traces of oxygen in the sample, and has also been observed in other studies (5). No methyl transfer reaction is catalysed in the absence of MT2DH protein (figure 7.2 B). Furthermore, methylcobalamin does not need to be bound to CobDH protein for the reaction to occur, an observation that has already been made for other O-
demethylases (15). It is therefore assumed that MT2DH exclusively catalyses the MT2 reaction and that CobDH is primarily functioning as methyl group carrier in the MT2 reaction, in agreement with other cobalamin-dependent methyltransferases (18).

Figure 7.2: Methyltransferase reaction from methylcobalamin to THF. A: The methyl transfer reaction was initiated by the addition of THF and observed by the methylcobalamin absorbance decrease at 540 nm, and cob(II)alamin absorbance increases at 477 nm. B: No decrease in methyl-cobalamin absorbance or increase of cob(II)alamin is observed in the absence of MT2DH protein. Spectra were recorded every 3 min for 30 min in total in both cases within an anaerobic glove box.

7.4.2 Crystal structure determination

MT2DH consists of 270 residues on a single polypeptide chain with a calculated molecular weight of ~ 30 kDa, which is in line with SDS-PAGE analysis of the purified protein S1. Based on the gel filtration elution profile on a S200 (10/300) size exclusion column (GE Healthcare), it was hypothesised that the protein is forming a homodimer (MT2DH protein 280 nm absorbance maximum at ~ 14 ml), similar to other described methyltransferases (32; 21). MT2DH crystallised in several conditions with different crystal shapes as described in the materials and methods section. However, only the plate shaped crystals diffracted beyond 2 Å and were used for all crystallographic experiments described here. The structure of apo-MT2DH was solved by molecular replacement (Balbes pipeline) using the MTHF-binding protein from Clostridium thermoaceticum (MeTrCh, PDB entry IF6Y, 31.2% sequence homology) as a search model. The molecular replacement solution has a Q-factor of 0.77 and \( R_{\text{work}} / R_{\text{free}} \) values of 0.35 / 0.381, respectively. A calculated Matthew’s coefficient from a collected dataset of 2.62 with a solvent content of 53% indicates the presence of two molecules (apo-MT2DH) per asymmetric unit. The solved apo-, THF- and MTHF-MT2DH structures have been refined to 1.9, 1.8 and 1.6 Å, respectively, and confirm the assumption of a homodimer formation. Both monomers of the homodimer fold into a TIM barrel domain, consisting of eight central \( \beta \)-sheets, surrounded by eight \( \alpha \)-helices. Amino acid residues of the N-terminal His\(_6\)-tag and linker region upstream of the first residue of MT2DH (Met1) appear to be
Figure 7.3: Overall structure of THF-MT2DH is a homodimer. Each monomer folds into a TIM barrel, with 8 central $\beta$-sheets surrounded by 8 $\alpha$-helices. The left monomer is shown in the cartoon representation with $\alpha$-helices in green and $\beta$-sheets in red. The THF substrate is shown as ball and stick (carbon atoms = grey; nitrogen = blue; oxygen = red). The right monomer is shown in a light blue surface representation with a transparency set to 0.5, in order to illustrate the binding site of the THF (transparency set to 1) on a groove at the protein surface. This figure was prepared using QtMG from the CCP4 suite.

Disordered and are therefore not visible in the electron density, as well as the three to five most C-terminal residues. Identified residues in the electron density are:
- Apo-MT2DH chain A: Met1 - Gly265, chain B: Met1 - Gly265;
- MTHF-MT2DH chain A: Met1 - Leu267, chain B: Met1 - Leu267;
- THF-MT2DH chain A: Met1 - Gly265 chain B: Met1 - Ile264.

The general electron density maps for all three proteins are of high quality and define the protein backbone and side chains very well as shown in supplementary figure S2 (N-terminus of MTHF-MT2DH chain A). The cofactor binding site is located on the protein surface at the C-terminal ends of the central $\beta$-sheets and is described below in greater detail. Figure 7.3 shows the overall homodimer formation of THF-MT2DH. The dimer formation is similar to other solved methyltransferase structures, e.g. MeTr from Clostridium thermoaceticum (32).

7.4.3 Structure comparisons

Structural C-$\alpha$ atom alignments have been performed between MTHF-MT2DH and two other MTHF-binding proteins that catalyse methyl transfer reactions in the different direction than MT2DH. These are the methyltransferase proteins from Carboxydothermus hydrogenoformans (MeTrCh, PDB entry 2YCJ) and Moorella thermoacetica (MeTrMt, PDB entry 2E7F), which are very similar to each other (RMS of 0.517 (227 compared atom pairs), using chain A from both structures). Both MeTr’s are part the Wood-
Ljungdahl pathway and catalyse methyl group transfers from MTHF to cob(I)alamin bound to a corrinoid iron-sulphur protein (CoFeSP). The alignment (figure 7.4 A) reveals that the structures are similar with RMS values of 1.059 (MT2DH to MeTrCh, 183 compared atom pairs) and 1.26 (MT2DH to MeTrMt, 198 compared atom pairs). The general topology (e.g. the presence and location of the eight $\alpha$-helices and eight $\beta$-strands) of MT2DH is the same as in both MeTr's. However, there are some distinct differences visible in the MT2DH structure. The central $\beta$-strand number seven and the downstream loop, comprising residues 190 - 205 in MT2DH, is rotating away from the corresponding region in the two MeTr proteins (residues 197 - 205; figure 7.4 B, red box). This is especially interesting to note as this region of the protein encompasses potentially catalytic residues, that are in close proximity to the nitrogen atom N5 of the TFH cofactor that receives the methyl group from methylcobalamin and resulting in different active sites compositions (details below) between the two types of methyltransferases. Furthermore it leads to the assumption that the forward and reverse methyl transfer reactions may be catalysed differently.

### 7.4.4 Cofactor binding site of MT2DH

Both the THF and MTHF cofactors are very well defined in the density as shown in figure 7.5 for MTHF. The cofactor is bound to the protein in an elongated shape at the C-terminal ends of the central $\beta$-strands. The binding site of the pterin ring of the cofactor is comprised of conserved residues and defined water molecules that form a sophisticated hydrogen bond network. Specifically, Asp76, Asn97, Asp161 are engaged by hydrogen bonding to the pterin heteroatoms N8, N1 and N2/N3, respectively, as shown in figure 7.5. The water molecules W2 and W3 are in analog positions to described water molecules in the MeTrMt and MeTrCh structures (20; 21). However, the conserved Asn199 residue of MT2DH is not in hydrogen bond distance to either N5 or O4 of the MTHF cofactor ($\sim$ 5 Å distance). It is assumed that this long distance is a result of the rotation of the MT2DH region (residues 197-205), relative to the two analogues MeTr regions (see figure 7.4). The position of Asn199 is the same in all three obtained MT2DH crystal structures, e.g. there is no swinging in of the Asn199 side chain upon binding of THF to the protein or after methyl group transfer (MTHF bound structure). According to our crystallographic findings, Asn199 is hence not in a position to stabilise any transition state during the methyl group transfer from methylcobalamin occurring in O-demethylation. Instead, there is a highly coordinated water molecule (water 1, W1) present in the density of all three structures that is in hydrogen bond distance to N5 (3 Å in the THF case and 2.85 Å in the MTHF case). Furthermore, W1 hydrogen bonds to Ser198 and is in putative hydrogen bond distance to the O4 atom of the cofactor and in close contact to another water molecule (W4) as shown in figure 7.5. This water is also present at the same position in the apo-MT2DH structure. A water molecule in a similar
Figure 7.4: Structure alignment of methyltransferase monomers. A: The stereo view alignment contains the respective protein monomer chains MT2DH (blue), MeTrCh (magenta, 2YCJ) and MeTrMt (green, 2E7F). Protein chains are shown as Cα traces. B: Close-up view of the protein stretch including the C-terminal end of β-strand seven that is different for MT2DH compared to the two methyltransferase structures that catalyse the reverse reaction. The β-strand that is different between MT2DH and the other two methyltransferases in both A and B is marked with red boxes.
position has been identified in the Asn199Ala mutant of the MeTrMt methyltransferase (W11m in that paper). However, in that case the water molecule is not hydrogen bonding to the Ser198 side chain (20) and therefore presumably less stabilised as W1 in MT2DH.
Figure 7.5: THF/MTHF binding to MT2DH. A) Stereo view of the MTHF ligand bound to the typical binding site of TIM barrel methyltransferases. MTHF and key residues implied to be important in ligand binding are shown as sticks (carbon grey, nitrogen blue, oxygen red). Well defined water molecules are shown as green spheres. The electron density (2Fo − Fc map, contoured at 1.5 σ) is shown as blue mesh. All polar residues of the pterin moiety of MTHF are engaged in H-bonds (indicated as magenta dots.) The protein surface is shown in light grey. Asn199 is in this position too far away in order to interact with either N5 or O4 of the MTFH cofactor. B) and C) Schematic illustration of the THF binding and MTHF binding to MT2DH. Hydrogen bonds are shown as black dotted lines and distances are given in angstrom. Notably, Asn199 in both cases is too far away from the N5 atom (red dotted lines, distance is greater than 4.75 Å) in order to be involved in methyl group transfer. R denotes the p-aminobenzoylglutamate side chain of the folate.
7.4.5 Proposed reaction mechanism

Asn199 does not seem to be involved in binding to THF / MTHF, respectively and it is therefore believed that Asn199 is not involved in methyl group transfer occurring in O-demethylation. However, a further role of Asn199 in MeTrMt has been proposed recently (33). It is argued that Asn199, when pointing away from the active site (as in the MeTrMt apo structure), is blocking the entry of cobalamin into the active site, but allows its entry upon MTHF binding and rotation of Asn199. Clearly, the situation must be different in O-demethylases. It cannot be ruled out that Asn199 is important for the correct positioning of methylcobalamin prior to methyl group transfer to N5 of the THF, which is corroborated by the fact that Asn199 is highly conserved in O-demethylases homologous to MT2DH, as shown in supplementary figure S3. Based on the crystallographic findings from this study the following reaction mechanism is proposed for the methyl group transfer from methylcobalamin to the nitrogen atom N5 of the THF cofactor bound to MT2DH as illustrated in figure 7.6:

![Figure 7.6: Proposed reaction mechanism of methyl group transfer from methylcobalamin to THF catalysed by MT2DH. THF is supposed to bind to the protein in a deprotonated form (A). The free electrons perform a nucleophilic attack onto the carbon atom of the methyl group in methylcobalamin. The methyl group is transferred as a carbocation, which results in positively charged N5 atom (B). By a rapid proton transfer step from N5 to W1, this positive charge is removed from the N5 atom (C). Thus the reverse reaction from MTHF to THF is prevented. R denotes the p-aminobenzoylglutamate side chain of the folate.](image)

N5 is thought to be deprotonated already when THF binds to MT2DH (A). The N5 free electron pair is then performing a nucleophilic attack on the carbon atom of the methyl group bound to the cobalt in cobalamin upon binding of methylcobalamin to MT2DH (B). The methyl group is subsequently transferred as a carbocation to the N5
nitrogen atom of THF. The structural data give rise to the assumption that the water coordinated between Ser198 and N5 (W1) is then in a position to rapidly abstract a proton from the now positively charged N5 (now in the MTHF state), in order to prevent the reverse reaction from MTHF to THF (B to C). The proton from W1 can efficiently be conducted to Ser198 and the bulk solvent water around. Ser198 is located on the protein surface and is therefore easily accessible by solvent water molecules. This assumption of the proton relay system is in agreement with pH dependence studies of other MTHF-dependent MeTr from C. thermoaceticum, which indicate that a protonation of THF should decrease the rate of methyl transfer from methylcobalamin to THF (34).

7.4.6 \textit{In vitro} stability of MT1DH

MT1DH protein was overexpressed in a soluble form in a recombinant \textit{E. coli} system. The protein has a calculated isoelectric point (pI) of 6.73 (without His tag) and 7.26 (including the N-terminal His tag and thrombin cleavage site). Upon purification at pH 7, 7.5 or 8 using 50 mM Hepes or Tris buffer systems, the protein rapidly undergoes precipitation, irrespective of the salt concentration, or on the addition of glycerol or potential substrates like vanillate in the buffer. When MT1DH was eluted from the Ni-NTA column with imidazole (∼150 mM) the precipitation occurred even faster and could only be prevented by immediate desalting of the sample. It was not possible to concentrate MT1DH at such pH values to higher than ∼3 mg/ml, which made the protein not very likely to crystallise. Furthermore MT1DH eluted very late (17 - 20 ml) from the S200 (10/300) gel filtration column in the pH range from 7 - 8. Gel filtration matrices are made from hydrophobic carbohydrate molecules (agarose, dextrose) that interact with hydrophobic patches from partially unfolded proteins leading to longer retention time. It is therefore assumed that MT1DH undergoes partial unfolding / precipitation in this pH range. This assumption is corroborated by the fact that MT1DH is more stable at pH 9, at which the protein elutes from the gel filtration column at 15 ml under otherwise same buffer conditions (50 mM Tris pH 9, 100 mM NaCl). Furthermore the protein could be concentrated to ∼15 mg/ml without precipitation. However, no crystallisation conditions have been identified yet.

7.4.7 Binding studies between MT1DH and aromatic methyl ether substrates

Different common \textit{O}-demethylate substrates have been tested towards their binding capacities to MT1DH using SPR as shown in figure 7.7. The experiments indicate that only carboxylated phenyl methyl ethers (e.g. vanillate, syringate) bind to MT1DH, whereas substrates that lack the carboxy function (e.g. guaiacol or anisole) do not bind at all. It is hence assumed that the carboxy function of the phenyl methyl ether substrate contributes mainly to the interaction between the substrate and MT1DH.
Figure 7.7: SPR between MT1DH and different aromatic methyl ethers. MT1DH is binding to carboxylated aromatic methyl ethers as shown by the sensorgrams for the binding to syringate, 3,4-dihydroxy benzoic acid and syringic acid (A), B), C)). The binding affinities are in the low mM range (D)). No binding interaction is observed with aromatic methyl ethers that lack an additional carboxy function as shown in the cases of anisole and guaiacol (E), F)).

Generally, substrate binding appears to be weak. The calculated dissociation constants ($K_D$) in the low mM range indicate a very rapid substrate binding and release. It is difficult to say, if such low binding also occurs at biological conditions. In the lack of a reliable activity assay (MT1DH does precipitate rapidly in the glove box, even at pH 9) it is even possible that the tested substrates here are not even demethylated by MT1DH.

### 7.5 Conclusions

This study describes the first structural characterisation of a methyltransferase (MT2DH) involved in $O$-demethylation, which catalyses the methyl group transfer from methylcobalamin to protein bound THF. Also, this is the first structure of a methyltransferase catalysing the methyl transfer in this direction. Although the structure of MT2DH is similar to methyltransferases that catalyse the reverse reaction from MTHF to THF, the reaction mechanism seems to be different due to the restructuring of active site residues. This assumption is primarily based on the fact that the conserved Asn199 residue that
is involved in transition state stabilisation in MTHF to THF reactions, is most likely not involved in the methyl group transfer in MT2DH. Even more, it is speculated that the displayed Asn199 may indeed be deliberately displaced in order to prevent the reverse reaction (MTHF to THF) to occur. A new function of the conserved Ser198 seems to emerge for the methyl group transfer from THF to MTHF. As stated above, the water W1 is assumed to abstract a proton from the THF N5 atom upon the methyl group transfer from methylcobalamin. The methyl group is transferred as methyl cation; by rapid proton abstraction from the N5, positive charge is taken from the N5. A positive charge on N5 would drive the reverse reaction from MTHF to THF. The proposed importance of Ser198 in MT2DH lies in providing an efficient proton relay system from the water W1. It would be interesting to find out, if there are structural changes associated in MT2DH upon cobalamin binding and to which extent Asn199 is involved in this binding in comparison to other cobalamin-dependent methyltransferase complex structures (33).

In the absence of a crystal structure of MT1DH or a homolog protein it is difficult to make judgements about its structure or mechanism. However, the biophysical data presented here indicate that MT1DH discriminates between its substrates based on the presence of a carboxy function.

7.6 References


Cobalamin-dependent enzyme systems


Cobalamin-dependent enzyme systems


7.7 Supplementary figures and tables

Figure S1: SDS-PAGE of all three proteins directly involved in \( O \)-demethylation. 1 = Marker in kDa; 2 = MT1DH with an apparent molecular weight of 50 kDa; 3 = CobDH with an apparent molecular weight of 22 kDa; 4 = MT2DH with an apparent molecular weight of 30 kDa. The gel was stained with InstantBlue (Expedeon).

Figure S2: Electron density map of MTHF-MT2DH. The N-terminus of chain A of the MTHF-MT2DH crystal structure is shown to illustrate that the electron density is well defined for both the protein backbone and the amino acid side chains. Protein is shown as sticks, with grey carbon atoms. The \( 2F_0 - F_c \) omit map is shown as a blue mesh, contoured at 1.5 \( \sigma \).

Figure S3: Sequence alignment of the Asn199 region between characterised \( O \)-demethylases. Both Ser198, Asn199 and downstream residues are highly conserved between MT2DH and other \( O \)-demethylase methyltransferases.
### Table S1: Crystallographic data and refinement statistics. Values in parentheses are for the highest resolution shell.

<table>
<thead>
<tr>
<th></th>
<th>apoMT2DH</th>
<th>MTHF-MT2DH</th>
<th>THF-MT2DH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X-ray source</td>
<td>I24, DLS</td>
<td>I04, DLS</td>
<td>I04, DLS</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Space group</td>
<td>(P2_12)</td>
<td>(P2_12)</td>
<td>(P2_12)</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Unit-cell parameters</td>
<td>(a=90.67, b=119.07, c=58.15)</td>
<td>(a=90.3, b=118.39, c=57.73)</td>
<td>(a=90.19, b=118.24, c=57.71)</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>27.7-1.92 (1.97-1.92)</td>
<td>29.7-1.6 (1.65-1.6)</td>
<td>30.62-1.8 (1.847-1.8)</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>48828 (2470)</td>
<td>81041 (4062)</td>
<td>57921 (2939)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>7.1(6.2)</td>
<td>6.6(6.2)</td>
<td>6.5(6.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.84 (99.7)</td>
<td>99.7(96.5)</td>
<td>99.81(99.83)</td>
</tr>
<tr>
<td>(R_{merge}) (%)</td>
<td>9.5(67)</td>
<td>6.5(67.7)</td>
<td>7.7(72.7)</td>
</tr>
<tr>
<td>(I/\sigma(I))</td>
<td>13.1(2.7)</td>
<td>16.2(2.5)</td>
<td>16.5(2.5)</td>
</tr>
<tr>
<td>Wilson B factor (Å²)</td>
<td>18.9</td>
<td>17.7</td>
<td>16.03</td>
</tr>
<tr>
<td>No. of protein molecules in asymmetric unit</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Matthews coefficient (V_M) Å³ Da⁻¹</td>
<td>2.62</td>
<td>2.57</td>
<td>2.56</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>53</td>
<td>52</td>
<td>52</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>27.7-1.92 (1.97-1.92)</td>
<td>29.7-1.6 (1.67-1.6)</td>
<td>30.62-1.8 (1.847-1.8)</td>
</tr>
<tr>
<td>(R_{work}/R_{free}) (%)</td>
<td>19.61/23.47</td>
<td>18.17/21.56</td>
<td>21.87/25.42</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>4034</td>
<td>4071</td>
<td>4041</td>
</tr>
<tr>
<td>No. of ligand atoms</td>
<td>0</td>
<td>66</td>
<td>64</td>
</tr>
<tr>
<td>No. of water atoms</td>
<td>286</td>
<td>411</td>
<td>333</td>
</tr>
<tr>
<td>Overall B factor (Å²)</td>
<td>26.02</td>
<td>27.81</td>
<td>23.03</td>
</tr>
<tr>
<td><strong>Ramachandran statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favoured (%)</td>
<td>96.93</td>
<td>97.13</td>
<td>97.88</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>2.87</td>
<td>2.68</td>
<td>1.93</td>
</tr>
<tr>
<td>R.m.s.d., bonds (Å)</td>
<td>0.02</td>
<td>0.024</td>
<td>0.019</td>
</tr>
</tbody>
</table>

\[R_{merge} = \frac{\sum_{hkl} N(hkl)\left|I_i(hkl)-\langle I(hkl)\rangle\right|}{\sum_{hkl}\sum_{i} I_i(hkl)}\]

where: \(I_i(hkl)\) are the observed intensities; \(\langle I(hkl)\rangle\) is the average intensity; \(N(hkl)\) is the multiplicity of the reflection \(hkl\)

\[R_{work} = \frac{\sum_{hkl}\left|F_{obs}-|F_{calc}|\right|}{\sum_{hkl}|F_{obs}|}\]

\(R_{free}\) is the cross-validation \(R\) factor for the test set (5 %) of reflections omitted from refinement.
Chapter 8

General discussion and future prospects
8.1 Reductive dehalogenases involved in organohalide respiration

Organohalide respiration is used by several species of anaerobic bacteria with the fascinating ability to derive growth energy from the detoxification of chlorinated organic compounds (organohalides). The enzymes that catalyse this detoxification reaction in organohalide respiration, the reductive dehalogenases (RdhAs), are the central players in this pathway and are therefore interesting candidates for bioremediation strategies (1). Furthermore, they have high scientific impact because they comprise a novel class of cobalamin-dependent enzymes whose biochemistry is only poorly understood (2). Although extensive efforts have been made into the understanding of organohalide respiration in general and RdhAs in particular, there are still many unanswered questions regarding this intriguing respiratory pathway (3). Many of these questions are associated with the molecular details regarding the structure and catalytic mechanism of RdhAs.

8.1.1 Large scale production of reductive dehalogenases

In order to answer these questions and to also unlock the full potential of RdhAs in bioremediation processes, their underpinning biochemistry and structural biology needs to be investigated in detail. For that it is necessary to obtain large quantities of pure and enzymatically active protein. The extraction of RdhAs from organohalide-respiring bacteria is very uneconomic due to the fact that the bacteria grow very slowly with doubling times of up to 24 hours and yield also poor densities (4). Furthermore, there are often several similar RdhAs expressed in a single bacterial species, which would make downstream purification and characterisation (e.g. crystallisation) of a single RdhA species an even more challenging task (5). The logical step to circumvent these problems is to express RdhAs recombinantly in *E. coli* cells or any other suitable host organism. To this end, a recombinant expression and purification strategy for the reductive dehalogenase PceA from *Dehalobacter restrictus* has been developed (6). It was found that the *E. coli* trigger factor protein was necessary to express soluble PceA. However, so far it has not been possible to physically separate trigger factor from PceA; a task that will be a necessary prerequisite for the successful crystallisation of PceA or any other RdhA. It would therefore be of great interest to develop a purification strategy that would allow such a separation without decreasing PceA stability in solution. There are examples in the literature that describe the separation of trigger factor from target protein using standard purification methods (7). It is speculated that PceA is partially unfolded or that the putative hydrophobic protein segments that associates with the membrane, form tight non-covalent complexes with trigger factor that are difficult to break apart. If these segments could be identified, it may be feasible to genetically remove / change them with less hydrophobic segments that would allow the subsequent separation from
trigger factor. Another, and arguably more elegant approach, would be the development of a novel expression strategy that would result in the soluble expression of PceA without the aid of trigger factor. However, given that RdhAs are oxygen sensitive and membrane associated metalloenzymes, it will be difficult to develop such a protocol using standard molecular biology techniques. On the other side new scientific methods are always emerging and existing ones are being improved. In that respect the development of cell free expression methods is certainly an important achievement (8). It might be feasible to adjust the experimental parameters of such an expression method for the production of biological active RdhAs, including the cobalamin and Fe-S cluster cofactors. Such expression systems have already been successfully used for the expression of human membrane proteins, which are generally difficult to produce (9).

8.1.2 Catalytic activity and diversity of reductive dehalogenases

Another important question that needs to be addressed is why recombinantly produced PceA lacks enzymatic activity. Likely explanations are that the cobalamin cofactor used for the reconstitution is not of the right type. It has been shown that the cobalamin from the reductive dehalogenase PceA from Dehalospirillum multivorans is norpseudo-cobalamin (10). Norpseudo-cobalamin has a different base ligand (adenine) than commercially available cobalamin (dimethylbenzimidazole). The cobalamin base ligand can either be coordinated to the central cobalt atom from the lower axial site or it can rotate away from the cobalt and serve as an anchor for the tight binding of cobalamin to the protein. It is quite likely that PceA from D. restrictus does likewise require the incorporation of norpsedo-cobalamin or another non-commercially available cobalamin for its catalytic activity. If this is the case, then the cobalamin will need to be either specifically synthesised or extracted from the respective organohalide respiring bacterium, for a successful reconstitution after recombinant enzyme expression. Another explanation for the non-activity would be the incorrect Fe-S clusters reconstitution. Although the 4Fe-4S clusters in PceA from D. restrictus can be reconstituted as shown by EPR spectroscopy (6), the spectra do not reveal whether the correct cysteine residues coordinate to these clusters, nor if one or both of the 4Fe-4S clusters are formed. Given that so little is known about the biochemistry of RdhAs, one can only speculate about the catalytic mechanism of the dehalogenation reaction. It seems imperative to assume that the cobalamin cofactor is directly involved in catalysis, because of its strong nucleophilic character in the oxidation state +1. During the reaction two electrons are transferred onto the carbon atom of the C-Cl moiety. The C-Cl bond is cleaved and the Cl atom is released in its anionic form and is replaced by a hydrogen atom, thus forming a C-H bond. The transfer of the two electrons results in a change in the oxidation state of the carbon atom from +I to -I. It is unclear how the electrons are transferred onto the carbon atom. It would be interesting to determine the precise function of the Fe-S clusters
in these electron transfer reactions. It is possible that they merely catalyse the electron
transfer onto the carbon atom. However, it cannot be excluded that they are involved
in changing the oxidation state of the cobalamin cobalt atom, or that they are directly
involved in catalysis and interact with the organohalide substrate. In order to elucidate
the mechanism and function of the respective cofactors a variety of different information
needs to be obtained:

- What is the structure of the active site, i.e. how is the substrate positioned in
  order to react with the cobalamin and/or Fe-S clusters?
- How far are the Fe-S clusters separated from each other, the membrane and the
  cobalamin?
- What are the redox potentials of the individual Fe-S clusters?
- What is the origin of the two electrons to be transferred onto the carbon atom?
- Does the enzyme require any additional factor(s) for catalysis (e.g. ATP, Mg$^{2+}$ or
  Ca$^{2+}$, etc.), which has (have) not yet been identified?

Most of these questions could be answered if a RdhA crystal structure was available.
However, biophysical techniques such as redox titrations combined with EPR analysis
of the Fe-S clusters could identify the redox potential of the clusters. Large quantities
(over 100 mg) of purified and reconstituted RdhA would be needed for such an analysis,
which is neither feasible with the presented method for PceA production nor with tra-
ditional homologous expression. Site-directed mutagenesis would be another method for
the identification of key residues in RdhAs. A prerequisite for that, however, would be to
establish a reliable and robust method for the production of catalytically active enzyme.
Furthermore, the large number of different reductive dehalogenase genes ($rhdA$s) identi-
fied, is astonishing. To date, more than 250 $rhdA$ genes have been found (3) and specific
organohalide-respiring bacteria contain up to 32 of them (11). It will be challenging to
find out, how and why such a broad gene repertoire evolved. In that respect it would be
especially interesting to investigate how the amino acid sequences of the RdhAs change
in respect to different substrates (e.g. aliphatic versus aromatic substrates). Ultimately,
we will have to await the first crystal structure of a RdhA enzyme in order to provide
definite answers to the questions above. The entire research field around organohalide
respiration would benefit from solving a RdhA crystal structure, which by itself will
require the combined effort from multiple disciplines including microbiology, molecular
biology, biochemistry and structural biology.
8.2 The $O$-demethylase system

Cobalamin-dependent $O$-demethylation refers to the methyl group transfer from aromatic methyl ethers to tetrahydrofolate (THF), resulting in the formation of methyltetrahydrofolate (MTHF), which is fuelled into central metabolic carbon pathways. Bacterial $O$-demethylase systems were originally identified in acetogenic bacteria and more recently in organohalide-respiring bacteria. They are comprised of three independent proteins that bind the substrate, cobalamin and THF cofactor, respectively. Here, structure-function relationships of the proteins from a *Desulfitobacterium hafniense* DCB-2 $O$-demethylase have been investigated using X-ray crystallographic and biochemical experiments. These investigations have led to the elucidation of the first crystal structures from $O$-demethylase proteins. A 1.5 Å crystal structure of the cobalamin-binding protein (CobDH) revealed that CobDH binds methylcobalamin in a similar way to other cobalamin-binding proteins that catalyse different methyl transfer reactions. Similarly, crystal structures were solved for the THF binding protein in different forms (without ligand (1.9 Å), bound to THF (1.8 Å) and bound to MTHF (1.6 Å), respectively). This has led to the identification of some interesting differences between the cobalamin-dependent methyl transfer from THF to MTHF compared to the reverse transfer from MTHF to THF, which is more common in nature and also currently better characterised.

8.2.1 Putative mechanism of methyl transfer reactions

Two independent and consecutive methyl transfer reactions (MT1 and MT2) are involved in $O$-demethylation. During the first reaction (MT1), the methyl group is transferred from the aromatic methyl ether substrate to fully reduced cobalamin, forming methylcobalamin and demethylated substrate. In the second methyl transfer reaction (MT2) the methyl group is transferred from methylcobalamin to THF, leading to the formation of MTHF and restored cobalamin (reduced state). The MT1 reaction mechanism is the subject of some controversy. One of the significant open questions is whether the substrate binding protein requires zinc for the activation of the ether bond between the oxygen and the methyl group to be transferred. Some studies suggest that zinc binds to the protein (12), whereas others do not find any evidence of zinc binding (13). For activation of the methyl group, positive charge needs to be transferred to its adjacent atom, i.e. the ether oxygen, because the methyl group itself is a poor leaving group. In case of the methanol:CoM methyltransferase, crystallography has shown that a protein bound zinc cation does serve as a Lewis acid for the transfer of such a positive charge (14). It is generally believed that cobalamin in its fully reduced form (cob(I)alamin, with a redox potential of $\sim -500\text{mV}$) performs a nucleophilic attack on the carbon atom of the methyl group to be transferred. However, it is unclear if an additional activation as in the case of the methanol:CoM methyltransferase is necessary for the activation.
of the aromatic methyl ether substrates. Due to the lack of a crystal structure from a
substrate binding protein of an O-demethylase, it is difficult to predict where and how
the substrate is bound and activated. Additionally, the protein sequences of the differ-ent substrate binding proteins are not very well conserved, making this prediction even
more difficult. Analogous to reductive dehalogenases, a crystal structure of the substrate
binding domain would certainly broaden the understanding of the MT1 reaction of O-
demethylation and would also answer the question of whether these proteins fold into
TIM barrels, a common protein fold observed for many other methyltransferases.

In the second methyl transfer reaction (MT2) the methyl group is transferred from
methylcobalamin to the N5 nitrogen atom of the THF cofactor. This reaction is ex-
clusively catalysed by the THF binding protein (MT2DH), which is proven by the fact
that the reaction takes place with exogenous supplied methylcobalamin and THF bound
protein (MT2DH). As identified by the crystallographic experiments described in the
previous chapters, both cofactors are presented on the surface of their respective pro-
teins, which allows the establishment of a close proximity between the methyl group of
the cobalamin and the N5 of the THF. Furthermore this positioning allows the planar
corrin ring of the cobalamin to be oriented parallel to the planar pterin moiety of the
THF cofactor. Thus, the Co-CH$_3$ and THF-N5-CH$_3$ bonds project out from either plane
in a perpendicular manner, which facilitates an effective methyl group transfer. It is as-
sumed that a similar arrangement will take place for the MT1 reaction between the sub-
strate binding protein and CobDH. Interestingly, the catalytically important Asn residue
in methyltransferases, catalysing MTHF to THF reactions, does not interact with the
THF/MTHF of MT2DH, although this Asn199 is similarly conserved between MT2DH
and homologue O-demethylases. It can not be discounted that this Asn is important
for the precise interaction between MT2DH and the cobalamin. It would be interesting
to test if substitutions of Asn199 would have any effect on either THF/MTHF binding
or methyl group transfer efficiency. Another interesting site-directed mutagenesis ex-
periment would be the substitution of Ser198 which forms a hydrogen bond to a water
molecule and then in turn hydrogen bonds to N5 of THF/MTHF. It is assumed that this
water molecule abstracts a proton from the N5 in order to drive the equilibrium of the
methyl transfer reactions towards MTHF and cob(I)alamin. This assumption would be
confirmed, if a disruption of this hydrogen bond network (by mutation of Ser198) leads
to decreased methyl group transfer rates, but has only minor effects on THF/MTFH
binding to the protein.

8.2.2 Protein-protein interactions involved in O-demethylases

Due to difficulties with the crystallisation of multicomponent systems, the underlying
protein-protein interactions and the protein dynamics in O-demethylases specifically and
other multicomponent systems generally are underexplored on a structural level when
compared to the investigation of their corresponding reaction mechanisms. A study, which analysed protein-protein interactions in the methyltransferase-corrinoid FeS protein (MeTr-CoFeSP) system participating in the Wood-Ljungdahl pathway, indicated that the involved protein-protein interactions are rather weak (15). In the case of the cobalamin-dependent methionine synthase it is known that protein domains that catalyse the MT1 and MT2 reaction are both folding into TIM barrels and that these two TIM barrels form a stable complex (16). Their active sites, however, are separated by $\sim 50$ Å throughout the time course of both methyl transfer reactions, indicating that the cobalamin-binding domain must be able to make impressive movements in order for the cobalamin cofactor to interact with both methyl donor and final methyl acceptor. Interesting experiments relating to the $O$-demethylase system would be the generation of fusion genes comprising either two or all three protein modules, in order to investigate whether complexes similar to the methanol:CoM system or methionine synthase system are formed. There, the substrate binding domain is tightly bound to the cobalamin-binding domain (14). Alternatively, the introduction of cysteine residues onto the surfaces of e.g. CobDH and MT2DH close to their cofactor binding sites could be used to test whether disulphide bridges are formed, resulting in stable complexes. Similar experiments could be performed with the protein pairs MT1DH / CobDH and MT1DH / MT2DH, respectively. However, it is not expected that tight interactions occur in the $O$-demethylase system, because glutaraldehyde cross-linking studies and co-purification experiments did not yield to the identification of any protein complexes (unpublished observations). Another interesting question concerns the stoichiometry of putative $O$-demethylase complexes. MT2DH is purified and crystallised as a homodimer, whereas MT1DH and CobDH were purified as monomers. It is possible that monomers of CobDH and/or MT1DH bind to both ends of the, thus central, MT2DH homodimer for the assembly of a functional $O$-demethylase complex. Again, more interdisciplinary experiments will be needed to broaden the current knowledge of protein-protein interactions involved in cobalamin-dependent $O$-demethylases.

8.3 References


Cobalamin-dependent enzyme systems


Chapter 9

Appendices
9.1 Abbreviations

Table 9.1: Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Ångström</td>
</tr>
<tr>
<td>°C</td>
<td>Degree Celsius</td>
</tr>
<tr>
<td>ε</td>
<td>Extinction coefficient</td>
</tr>
<tr>
<td>λ</td>
<td>Wavelength</td>
</tr>
<tr>
<td>σ</td>
<td>sigma</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>atm</td>
<td>Standard atmospheric pressure</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>B</td>
<td>Batch</td>
</tr>
<tr>
<td>BOG</td>
<td>n-octyl-β-D-glucopyranoside</td>
</tr>
<tr>
<td>C</td>
<td>One-face centered</td>
</tr>
<tr>
<td>calc</td>
<td>calculated</td>
</tr>
<tr>
<td>CC</td>
<td>Correlation coefficient</td>
</tr>
<tr>
<td>CCP4</td>
<td>Collaborative Computation Project No. 4</td>
</tr>
<tr>
<td>CHPA</td>
<td>3-chloro-4-hydroxyphenylacetate</td>
</tr>
<tr>
<td>CoFeSP</td>
<td>Corrinoid iron sulphur protein</td>
</tr>
<tr>
<td>cspA</td>
<td>Cold shock promoter A</td>
</tr>
<tr>
<td>D</td>
<td>Dialysis</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DCE</td>
<td>Dichloroethene</td>
</tr>
<tr>
<td>DLS</td>
<td>Diamond light source</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>DMB</td>
<td>5,6-dimethylbenzimidazole</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>EDAC</td>
<td>1-ethyl-3-(3-dimethylaminopropyl)carbodiimide</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>EPR</td>
<td>Electron paramagnetic resonance</td>
</tr>
<tr>
<td>ERC</td>
<td>European Research Council</td>
</tr>
<tr>
<td>etc.</td>
<td>et cetera, and other things</td>
</tr>
<tr>
<td>F</td>
<td>All-face centered</td>
</tr>
<tr>
<td>FID</td>
<td>Free interface diffusion</td>
</tr>
</tbody>
</table>

continued on next page
### Table 9.1 – continued from previous page

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g$</td>
<td>Force of gravity</td>
</tr>
<tr>
<td>$g$</td>
<td>gram</td>
</tr>
<tr>
<td>G</td>
<td>Gauss (magnetic flux density)</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-Transferase</td>
</tr>
<tr>
<td>h</td>
<td>hour(s)</td>
</tr>
<tr>
<td>HPLC</td>
<td>High pressure liquid chromatography</td>
</tr>
<tr>
<td>HRV 3C</td>
<td>Human rhinovirus 3C protease</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>$I$</td>
<td>Innenzentriert, body-centered</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calometry</td>
</tr>
<tr>
<td>J</td>
<td>Joule</td>
</tr>
<tr>
<td>k</td>
<td>kilo</td>
</tr>
<tr>
<td>K</td>
<td>Kelvin</td>
</tr>
<tr>
<td>Kana</td>
<td>Kanamycin</td>
</tr>
<tr>
<td>$K_d$</td>
<td>Dissociation constant</td>
</tr>
<tr>
<td>l</td>
<td>Litre</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Broth</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>MAD</td>
<td>Multiwavelength anomalous diffraction</td>
</tr>
<tr>
<td>MALS</td>
<td>Multi-angle light scattering</td>
</tr>
<tr>
<td>MBP</td>
<td>Maltose binding protein</td>
</tr>
<tr>
<td>MetH</td>
<td>Methionine synthase</td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple isomorphous replacement</td>
</tr>
<tr>
<td>MIRAS</td>
<td>Multiple isomorphous replacement with anomalous scattering</td>
</tr>
<tr>
<td>MT</td>
<td>Methyl transfer reaction</td>
</tr>
<tr>
<td>MTHF</td>
<td>Methyldihydrofolate</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>MWCO</td>
<td>Molecular weight cut-off</td>
</tr>
<tr>
<td>n.d.</td>
<td>not determined</td>
</tr>
<tr>
<td>NiNTA</td>
<td>Nickel nitrilotriacetic acid</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>obs</td>
<td>observed</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>OPPF</td>
<td>Oxford protein production facility</td>
</tr>
</tbody>
</table>

continued on next page
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>Primitive</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PCE</td>
<td>Tetrachloroethene</td>
</tr>
<tr>
<td>PCP</td>
<td>Polychlorophenol</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>pI</td>
<td>Isoelectric point</td>
</tr>
<tr>
<td>Pᵢ</td>
<td>Inorganic phosphate</td>
</tr>
<tr>
<td>PMF</td>
<td>Proton motif force</td>
</tr>
<tr>
<td>POP</td>
<td>Persistent organic pollutant</td>
</tr>
<tr>
<td>PPI</td>
<td>Peptidyl-prolyl cis – trans isomerase</td>
</tr>
<tr>
<td>p.p.m.</td>
<td>Parts per million</td>
</tr>
<tr>
<td>PVC</td>
<td>Polyvinyl chloride</td>
</tr>
<tr>
<td>R</td>
<td>Rhombohedral</td>
</tr>
<tr>
<td>RdhA</td>
<td>Reductive dehalogenase</td>
</tr>
<tr>
<td>rms</td>
<td>Root mean square</td>
</tr>
<tr>
<td>rmsd</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal ribonucleic acid</td>
</tr>
<tr>
<td>SAD</td>
<td>Singlewavelength anomalous diffraction</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small angle X-ray scattering</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>sec</td>
<td>Second(s)</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>SIR</td>
<td>Single isomorphous replacement</td>
</tr>
<tr>
<td>SIRAS</td>
<td>Single isomorphous replacement with anomalous scattering</td>
</tr>
<tr>
<td>SOC</td>
<td>Super optimal broth with catabolite repression</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>SSM</td>
<td>Secondary structure matching</td>
</tr>
<tr>
<td>sulfo-NHS</td>
<td>N-hydroxysulfosuccinimide</td>
</tr>
<tr>
<td>SUMO</td>
<td>Small ubiquitin-like modifier</td>
</tr>
<tr>
<td>t₁/₂</td>
<td>Half-life</td>
</tr>
<tr>
<td>TAT</td>
<td>Twin arginine translocation</td>
</tr>
<tr>
<td>TB</td>
<td>Terrific broth</td>
</tr>
<tr>
<td>TCE</td>
<td>Trichloroethene</td>
</tr>
</tbody>
</table>

continued on next page
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>Trichlorophenol</td>
</tr>
<tr>
<td>TF</td>
<td>Trigger factor</td>
</tr>
<tr>
<td>TF-PceA</td>
<td>Trigger factor PceA fusion protein</td>
</tr>
<tr>
<td>THF</td>
<td>Tetrahydrofolate</td>
</tr>
<tr>
<td>TIM</td>
<td>Triosephosphate isomerase</td>
</tr>
<tr>
<td>TLS</td>
<td>Translation, libration and screwmotion anisotropy</td>
</tr>
<tr>
<td>U</td>
<td>unit</td>
</tr>
<tr>
<td>UV-Vis</td>
<td>Ultraviolet-visible</td>
</tr>
<tr>
<td>V</td>
<td>Volt</td>
</tr>
<tr>
<td>VC</td>
<td>Vinyl chloride</td>
</tr>
<tr>
<td>VD</td>
<td>Vapour diffusion</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>W</td>
<td>Watt</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
<tr>
<td>X-FEL</td>
<td>X-ray free-electron laser</td>
</tr>
</tbody>
</table>
9.2 List of the 20 amino acids encoded by the universal genetic code

Glycine, Gly, G
Alanine, Ala, A
Valine, Val, V
Leucine, Leu, L
Isoleucine, Ile, I
Phenylalanine, Phe, F
Tyrosine, Tyr, Y
Tryptophan, Trp, W
Cysteine, Cys, S
Methionine, Met, M

Figure 9.1: Aliphatic, aromatic and sulphur containing amino acids.
Figure 9.2: Hydrophilic amino acids and proline.
9.3 Photographic crystal images

The following figures illustrate a couple of photographic snapshots of protein crystals that were obtained throughout the PhD project.

![Figure 9.3: Images of crystals that were obtained for PceT.](image)

Top: Sitting drop with crystallisation conditions containing 0.1 M MOPS/HEPES-Na pH 7.5, 10% (w/v) PEG 8000, 20% (v/v) ethylene glycol, 0.03 M MgCl₂ and 0.03 M CaCl₂ in a drop size of 400 nl, resulting in the growth of plate shaped PceT crystals. Bottom: Single PceT crystal in a nylon loop (Molecular Dimensions) ready for data collection on the Diamond Light Source synchrotron. The X-ray beam size is indicated by the red oval. ~ Crystal size: 200 x 50 x 5 µm.
Figure 9.4: Images of crystals that were obtained for CobDH. *Top:* Sitting drop after sparse matrix seeding. Crystallisation conditions contain 0.1 M MMT (D-malic acid, MES and Tris base in a 1:2:2 molar ratio) buffer pH 6, 25% (w/v) PEG 1500 in a drop size of 600 nl led to the growth of individual three-dimensional crystals. The presence of the cobalamin cofactor is clearly visible due to the distinct pink colour. *Bottom:* Single crystal in a nylon loop (Molecular Dimensions) ready for data collection on the Diamond Light Source synchrotron. The X-ray beam size is indicated by the red oval. ~ Crystal size: 100 x 30 x 30 μm.
Figure 9.5: Images of crystals that were obtained for MT2DH.
Top left: Needle shaped crystals. Top centre: Small crystal shower. Top right: Three dimensional, rectangular crystal. Middle: Plate shaped crystals grown in 100 mM Tris pH 7.5, 100 mM succinic acid, 15% (v/v) PEG 3350; Bottom: Single plate shaped crystal in a nylon loop (Molecular Dimensions) ready for data collection on the Diamond Light Source synchrotron. The X-ray beam size is indicated by the red oval. ~ Crystal size: 200 x 100 x 5 µm. The plate shaped crystals diffracted best (resolution higher than 2 Å), whereas the other crystal forms, e.g. needles and rectangular crystals, diffracted only to ~ 3 Å.
9.4 Acknowledgements

First of all I want to thank David Leys for his constant support throughout the last three years, his crystallographic expertise and for giving me a high degree of scientific freedom. Furthermore, I would like to thank Nigel Scrutton for his function as scientific advisor of this PhD project and the European Research Council (ERC) for generous funding of the DEHALORES project and the PhD studentship.

I could not be more grateful towards Mark Dunstan and Karl Fisher for their numerous advises and presence during the entire project. It was an honour and for me to have worked together with you two, and I will miss you a lot as bench buddies. Also I want to thank Carolina and Karl P. for sharing the burden of reductive dehalogenase research.

Special thanks go to Colin Levy for running the X-ray crystallography suite in the MIB and for his crystallographic skills, from which I was fortunate enough to learn a lot.

I want to thank all members of the MIB enzymology labs for the nice atmosphere and especially Mary, Anna and James for our friendship.

I thank Edward Snowden for leaking documents about the NSA and GCHQ mass surveillance programs, and The Guardian for publishing these documents to the general public.

Ein besonderer Dank gilt meiner ganzen Familie, insbesondere meinen Eltern und Geschwistern für ihre ständige Hilfsbereitschaft, Unterstützung und die gemeinsamen Aktivitäten auf die hoffentlich noch viele folgen werden.

Mein letzter und grösster Dank gilt Nora für ihre extrem gute Interpretation meiner Stimmungen, ihre Zuneigung und für das gemeinsame Klettern.
9.5 Curriculum Vitae, Hanno Sjuts

UNIVERSITY

09/2010 - 09/2013 PhD in Biochemistry and Structural Biology, University of Manchester, Manchester, UK


10/2008 - 04/2010 Molecular Biology M. Sc. / PhD Program, International Max Planck Research School, Georg-August-University, Göttingen, Germany

Highest Degree: Master of Science

08/2007 - 12/2007 Exchange Fall Term in Biosciences, University of New Mexico, Albuquerque, USA

10/2005 - 09/2008 Molecular Life Sciences B. Sc. Program, University of Lübeck, Lübeck, Germany

Highest Degree: Bachelor of Science

RESEARCH PROJECTS

09/2010 - 09/2013 Cobalamin-Dependent Enzyme Systems from Organohalide-Respiring Bacteria, Manchester Institute of Biotechnology (Prof. Leys), University of Manchester, UK

04/2010 - 08/2010 Structural Investigation of Carbanion Intermediates of the Enzymatic Reaction Catalysed by the Pyruvate Oxidase in Lactobacillus plantarum, Department of Bioanalytics (Prof. Tittmann), University of Göttingen, Germany

10/2009 - 04/2010 Directed Evolution and Selection of Barnase and Barstar Protein-Protein Interaction for the Development of Designer Cellulosomes External Master’s Thesis, Department of Biological Chemistry (Prof. Schreiber), Weizmann Institute of Science, Israel

01/2009 - 06/2009 3 Lab Rotations in the Course of the MSc/Ph.D Program Molecular Biology

04/2008 - 09/2008 Complex Formation between Cell-Penetrating Peptides and Oligonucleotides (Bachelor Thesis), Department of Molecular Medicine (Prof. Restle), University of Lübeck, Germany
PUBLICATIONS

**Sjuts, H.,** Dunstan, M.S., Fisher, K. and Leys, D. Crystal structures of the methyltransferase protein involved in O-demethylation that catalyses the methyl group transfer from methylcobalamin to tetrahydrofolate. Manuscript in preparation for submission to *PLoS ONE.*


AWARDS AND PRIZES

- 09/2012: FEBS Youth Travel Fund
- 09/2011: 2nd Best Poster Presentation Prize at the 8th International PhD Student Symposium: Horizons in Molecular Biology, *International Max Planck Research School, Göttingen*

SCHOLARSHIPS

- 09/2010 - 09/2013: European Research Council: PhD Studentship
- 10/2009 - 12/2009: German Academic Exchange Service (DAAD): External Master’s Thesis
- 08/2007 - 12/2007: University of New Mexico, USA: Foreign Exchange Scholarship