

Structure to function studies in hybrid cluster proteins and glucose-1-phosphate uridylyltransferases

David Guerra Aragão

Doctoral Thesis



Dissertação apresentada para obtenção do grau de
doutor em Bioquímica pelo Instituto de Tecnologia
Química e Biológica da Universidade Nova de Lisboa

(Release Candidate 2)

Oeiras, September 2007

Author Edition

Copyright © 2007 by

D. G. Aragão, Lisbon, Portugal.

All rights reserved

Todos os direitos Reservados

The project described in this thesis was carried out as a collaboration between the Macromolecular Crystallography Group at the European Synchrotron Research Facility (ESRF) in France and the Protein Crystallography Laboratory in Instituto de Tecnologia Química e Biológica (ITQB) from the Universidade Nova de Lisboa (UNL) in Portugal. The author acknowledges the grant BD/6480/2001/ from Fundação para a Ciência e Tecnologia. This work has been partially funded by ESRF, FEDER and Fundação para a Ciência e a Tecnologia (FCT), Portugal (POCTI/BME/38859/2001, POCTI/BME/44441/2002 and POCTI/BIO/58041/2004)

ISBN 978-989-20-0807-3

This thesis was written using the [GPL](#) software:

[OpenSuse 10](#)

[OpenOffice 2.2](#)

[GIMP 2.0](#)

[BkChem 0.11](#)

**Structure to function studies
in hybrid cluster proteins and
glucose-1-phosphate uridylyltransferases**

David Guerra Aragão
Doctoral Thesis

Table of Contents

1. Introduction.....	1
1.1. Scope of the thesis.....	1
1.2. Hybrid cluster protein (HCP).....	1
1.2.1. Metals in proteins.....	1
1.2.2. Iron in proteins.....	2
1.2.3. Iron co-ordination in proteins.....	4
1.2.4. Iron sulphur based enzymes.....	5
1.2.5. Introduction to the hybrid cluster protein.....	6
a) Classification of HCPs	7
b) Description of the structure.....	14
c) Structures at different oxidation states.....	17
d) Biological function.....	17
1.3. UDP-glucose uridylyltransferase (UGP).....	19
1.3.1. Sugar activation.....	19
1.3.2. Sphingomonas elodea and gellan.....	23
1.3.3. Glucose-1-phosphate nucleotidylyltransferase mechanism...27	
1.3.4. Glucose-1-phosphate uridylyltransferase vs glucose-1-phosphate thymidylyltransferase.....	29
1.3.5. Glucose-1-phosphate thymidylyltransferase (RmlA).....	31
2. Reduced hybrid cluster proteins (HCP) from <i>Desulfovibrio desulfuricans</i> ATCC 27774 and <i>Desulfovibrio vulgaris</i> (Hildenborough)	
X-ray structures at high resolution using synchrotron radiation.....	33
2.1. Abstract.....	33
2.2. Introduction.....	34
2.3. Materials and methods.....	37
2.3.1. Crystallisation.....	37
2.3.2. Reduction of the crystals.....	38
2.3.3. Assay for hydroxylamine reductase activity.....	39
2.3.4. Immuno-blotting.....	39
2.3.5. X-ray diffraction data collection.....	40
a) The reduced form of the HCP from <i>D. desulfuricans</i>	40
b) The reduced form of the HCP from <i>D. vulgaris</i>	41
2.3.6. Structure Solution	42
a) The reduced form of the HCP from <i>D. desulfuricans</i> : Ab Initio phasing.....	42
b) The reduced form of the HCP from <i>D. vulgaris</i>	44
2.3.7. Structure refinement.....	44
a) The reduced form of HCP from <i>D. desulfuricans</i> at 1.25 Å resolution.....	44
b) The reduced form of the HCP from <i>D. vulgaris</i> at 1.55 Å	

resolution.....	48
2.3.8. Quality of the refined structures.....	49
2.4. Results and discussion.....	50
2.5. Acknowledgements.....	58
3.The crystal structure of the hybrid cluster protein (HCP) from Desulfovibrio desulfuricans ATCC 27774 containing molecules in the oxidized and reduced states.....	61
3.1. Abstract	61
3.2. Introduction.....	62
3.3. Materials and methods.....	63
3.3.1. Crystallisation.....	63
3.3.2. X-ray data collection.....	64
3.3.3. Structure solution and refinement.....	66
a) Refinement with REFMAC.....	67
b) Refinement with SHELXL	70
3.3.4. Quality of the refined structures.....	71
3.4. Results and discussion.....	73
3.5. Acknowledgements.....	79
4.Structure of Desulfovibrio vulgaris anaerobically purified hybrid cluster protein at 1.35 Å resolution using MAD at the iron K-absorption edge...82	82
4.1. Abstract.....	82
4.2. Introduction.....	82
4.3. Materials and methods.....	85
4.3.1. Purification of D. vulgaris HCP.....	85
4.3.2. Crystallisation of D. vulgaris HCP.....	86
4.3.3. Data collection	86
4.3.4. Phase determination.....	88
4.3.5. Model construction and refinement.....	88
4.3.6. Three-dimensional structure comparison.	90
4.4. Results and discussion.....	90
4.4.1. Quality of the structure.....	90
4.4.2. Structure.....	92
4.5. Conclusions.....	106
4.6. Acknowledgements.....	107
5.Cloning, expression, purification, crystallisation and preliminary structure determination of Glucose-1-phosphate uridylyltransferase (UgpG) bound to glucose-1-phosphate from Sphingomonas elodea ATCC31461.....	109
5.1. Abstract	109
5.2. Introduction	110
5.3. Materials and Methods.....	112
5.3.1. Cloning and expression of UgpG.....	112
5.3.2. Purification of UgpG.....	114

5.3.3. Crystallisation of UgpG protein.....	116
5.3.4. X-ray diffraction analysis and structure determination.....	119
5.4. Results.....	122
5.5. Acknowledgements.....	123
6. The complex of <i>Sphingomonas elodea</i> ATCC 31461 glucose-1-phosphate uridylyltransferase with glucose-1-phosphate reveals a novel quaternary structure, unique amongst NDP-sugar pyrophosphorylase members	125
6.1. Abstract.....	125
6.2. Introduction.....	126
6.3. Material and methods.....	128
6.3.1. Expression, purification and crystallisation of UgpG.	128
6.3.2. Data collection, structure determination and refinement. .	129
6.3.3. Three-dimensional structure comparison.....	131
6.3.4. Coordinate accession numbers.....	131
6.4. Results.....	131
6.4.1. Crystal contents and structure quality.....	131
6.4.2. Overall polypeptide fold.....	132
6.4.3. Quaternary structure and oligomerisation interfaces.....	137
6.4.4. Catalytic cavity.....	140
6.4.5. Substrate affinity and specificity.....	143
6.4.6. Structural neighbours.....	144
6.4.7. Modelling of nucleotide binding.....	145
6.5. Discussion.....	146
6.6. Acknowledgements.....	150
7. Conclusions.....	151
7.1. Hybrid cluster protein (HCP).....	151
7.1.1. Structure of anaerobically purified oxidised structure vs that of anaerobically purified dithionite reduced HCP.....	153
7.1.2. Structure of aerobically purified oxidised vs that of anaerobically purified oxidised HCP.....	154
7.1.3. Three dimensional comparison of HCPs and carbon-monoxide dehydrogenases.....	155
7.2. Glucose-1-phosphate uridylyltransferase (UgpG).....	157
8. Appendices	161
8.1. Genomic era and structure determination	161
8.2. Overview of X-ray crystallographic structure determination....	165
8.2.1. Protein crystallisation.....	167
8.2.2. Space groups.....	169
8.2.3. X-ray diffraction experiment.....	170
8.2.4. Phase determination	171
a) Molecular Replacement (MR).....	171
b) Multiwavelength Anomalous Diffraction (MAD) and Single	

wavelength Anomalous Diffraction (SAD).....	172
c) Multiple Isomorphous Replacement (MIR).....	172
d) Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS).....	173
e) Ab initio/brute force.....	173
8.2.5. Model construction and crystallographic refinement.....	174
8.2.6. Structure quality.....	176
References.....	177
A to G.....	177
H to L.....	180
M to Z.....	183

Abstract

The isolation and crystallisation of Hybrid Cluster Protein (HCP) under anaerobic conditions was carried out and resulted in the elucidation of the high resolution structures of both the anaerobically purified and dithionite reduced HCP forms from *D. vulgaris*. These structures complete the set of information on HCP in various oxidation states: aerobically purified oxidised, anaerobically purified oxidised and anaerobically purified dithionite reduced. This set allows the nature of the hybrid cluster to be followed as the oxidation state changes. The knowledge of both the oxidised and reduced structures allowed the interpretation of the asymmetric unit contents and in the crystallographic refinement of a crystal containing a mixture of oxidised and reduced HCP protein. The anaerobically purified structure was determined by multiwavelength anomalous dispersion at the iron K-edge allowing the determination of the nature and position of the anomalous scattering atoms of and around the hybrid cluster, and served as well as an independent confirmation that the hybrid cluster is not an artefact of *in vitro* oxidation procedures. HCPs share sequence and structure homology to anaerobic carbon-monoxide dehydrogenases (CoDs). A detailed comparison between the HCP proteins and the anaerobic CoDs was carried out and showed that the main differences between HCP and CoD proteins lie in the N-terminal domain where the cubane cluster (B-cluster in CoDs) binding region is apparently absent for CoD. However, the structures of dimeric and tetrameric CoDs show that this region is present, via monomer B, and has the same fold as in HCP. Both HCPs and CoDs have intriguing and unusual clusters at the interface of their three domains. These clusters, although different, share striking similarities and are linked to the polypeptide chain via homologous ligands.

The glucose-1-phosphate uridylyltransferase (UgpG) protein expression, isolation and crystallisation, and subsequent diffraction data collection and phasing was carried out and resulted in the elucidation of its tetrameric structure with bound glucose-1-phosphate. This new structure presents similar characteristics to the known glucose-1-phosphate timydylyltransferases (RmlAs) models, with a similar overall fold and shows the conserved active site cavity with similar spatial orientations for the catalytic residues side-chains. A loop closes the active site in all of the RmlA structures, whether substrates are bound or not, whilst in the UgpG structure this loop adopts an “open” conformation allowing access to the active site where G1P was found already bound in place. Based on its mobility, this loop has been suggested to be the gate keeper to the active site, which on binding both substrates may then help to stabilise the ternary complex. Although space is available in the catalytic cavity of UgpG to accommodate either UTP or TTP, the presence of this flexible loop that could close the active site, suggests that protein conformation changes may play an important role in determining the different specific activities (as observed in vitro) and thus the UgpG preference for UTP.

Sumário

Foram realizados o isolamento e a cristalização em condições anaeróbicas da proteína de cluster híbrido (HCP). Determinaram-se estruturas de alta resolução da HCP de *D. vulgaris*, purificada anaerobicamente e reduzida com ditionito. Estas estruturas completam um conjunto de informações sobre a HCP em vários estados de oxidação: oxidada purificada aerobicamente, oxidada purificada anaerobicamente e reduzida com ditionito. Este conjunto de informações permite seguir a natureza do cluster híbrido nos diferentes estados de oxidação. O conhecimento da estrutura oxidada e reduzida permitiu a interpretação do conteúdo da unidade assimétrica e o refinamento criptográfico de um cristal contendo uma mistura da proteína HCP, de *D. desulfuricans*, na forma oxidada e reduzida. A estrutura da proteína purificada anaerobicamente foi determinada por dispersão anômala usando vários comprimentos de onda (MAD) na banda K do elemento ferro. Determinou-se a natureza e posição dos átomos no cluster híbrido e foi possível confirmar que o cluster híbrido não é um artefacto do procedimento de oxidação durante o processo de isolamento da proteína. As HCPs partilham uma homologia de sequência e estrutura com as desidrogenases do monóxido de carbono (CoDs). A comparação detalhada das proteínas HCPs e das CoDs anaeróbicas mostrou que a principal diferença estrutural, entre estas proteínas, ocorre no N-terminal, onde a região de ligação do cluster de cubano (cluster-B nas CoDs) está ausente na CoD. Contudo, as estruturas diméricas e tetraméricas das CoDs mostram que essa região está presente, através do monomero B e apresenta o mesmo fold que nas HCPs. Ambas as proteínas apresentam clusters interessantes e pouco comuns na interface dos seus três domínios.

Estes clusters, apesar de diferentes, partilham semelhanças significativas e estão ligados à cadeia polipeptídica através de ligandos homólogos.

A expressão, purificação e cristalização da proteína glucose-1-fosfato uridiltransferase (UgpG) de *S. elodea*, bem como a subsequente recolha de dados de difracção e determinação das fases foi realizada e resultou na elucidação da estrutura tetramérica da UgpG ligada a glucose-1-fosfato (G1P). Esta nova estrutura apresenta características semelhantes aos modelos da glucose-1-fosfato timidiltransferase (RmlAs). Esta semelhança é notória no *fold* da proteína bem como no facto do centro activo apresentar orientação espacial semelhante, resíduos catalíticos homologous e com cadeias laterais na mesma orientação. Uma zona estruturalmente desordenada (*loop*) fecha o centro activo de todas as estruturas de RmlAs, estado o substrato presente ou não; enquanto que na estrutura da UgpG a zona estruturalmente desordenada da proteína adopta uma conformação “aberta” que permite acesso ao centro activo, onde se encontrou a G1P ligada. Baseado nesta mobilidade, sugere-se que esta zona estruturalmente desordenada funciona como um regulador da entrada para o centro activo e que a ligação a ambos os substratos poderá permitir a estabilização do complexo ternário. Apesar de haver espaço disponível na cavidade catalítica da UgpG para acomodar tanto os nucleótidos UTP como o TTP, a presença desta zona estruturalmente desordenada sugere que alterações conformacionais na proteína podem desempenhar um papel fundamental na determinação das diferentes actividades específicas (observadas *in vitro*) bem como a preferência da UgpG pelo UTP.

Acknowledgements.

The work here reported was supported by many people that provided their personal expertise, valuable assistance and many times friendship.

First, I would like to thank my two official supervisors, Dr. Carlos Frazão and Dr. Edward Mitchell. They started the project with me and were present, as they could, all the way through. I am grateful to Carlos for teaching me to be persistent. Persistence is the only way to overcome difficult tasks. From Carlos I also learned to never forget how the small details can be incredibly important. I'm also grateful for the time Ed always had for my research and my doubts. Particularly, I will never forget the late nights, early mornings and weekends collecting data in between synchrotron user time.

I am grateful to my non-official supervisors, Professor Peter Lindley and Professor Maria Arménia Carrondo. I thank Maria Arménia for highlighting my path for a science & research careerer. I also thank her for all the help in the first years of my thesis. I thank Peter for his committed work on the HCP project and for putting research as the main goal even when problems were raised.

I thank Dr. Francisco Enguita for teaching, correcting and improving my techniques while working on the lab. From him I learned good laboratory practices but also had particular help on setting the first selenium-methionine expression system in the Protein Crystallography group at ITQB.

I thank Dr. Célia Romão and Dr. Ana Rita Marques for their work on the HCP and UgpG isolation, respectively. Some of the techniques I learn from both were used all the way through the thesis work.

I thank Ulrike Kapp and Ricardo Coelho from the help both gave me as laboratory technicians in the ESRF and ITQB, respectively.

I thank Cyril Dian, Dora Alves, Helena Ferreira, Joana Rocha, Sofia Caria and Dr. Sofia Macedo for many valuable discussions, helping me in laboratory work and some corrections to the text and figures of my thesis.

I'm also grateful to all the researchers and, in particular, to both group leaders from the Macromolecular Crystallography Group in the ESRF and the Protein Crystallography Laboratory in the ITQB, Dr. Sean McSweeney and Professor Maria Arménia Carrondo.

A special thank to Professor António Câmara, Professor Fernando Caria, Professor Isabel Carvalho Guerra, Carmo Guerra and Margarida Guerra for their support in particularly important phases of the writing.

As for my funding I acknowledge a FCT grant SFRH (BD/6480/2001), the ITQB directorate, Protein Crystallography Laboratory in the ITQB and the ESRF.

Index of Tables

Table 1.1. Structures of bacterial glucose-1-phosphate thymidyltransferase and their respective organism.....	31
Table 2.1. Data collection statistics	41
Table 2.2. Refinement statistics for reduced HCPs from <i>D. desulfuricans</i> and <i>D. vulgaris</i> at resolutions of 1.25 Å and 1.55 Å respectively.....	47
Table 2.3. Quality of reduced HCP structures.....	49
Table 2.4. Geometry (bond distances in Å) of the cubane and hybrid clusters in the as-isolated oxidised and reduced HCP proteins from <i>D.</i> <i>desulfuricans</i> and <i>D. vulgaris</i>	54
Table 3.1. Data collection statistics.....	65
Table 3.2. Refinement statistics.....	70
Table 3.3. Progress of the SHELXL refinement.....	72
Table 3.4. Quality of reduced HCP structure.....	72
Table 4.1. Data collection statistics for anaerobically purified <i>D. vulgaris</i> HCP.....	87
Table 4.2. R-factor progress through refinement.....	89
Table 4.3. Final refinement statistics.....	91
Table 4.4. Structure similarity and identity between HCPs and CoDs.....	98
Table 4.5. Root mean square (RMS) differences between HCPs and CoDs homologous C-alpha.....	99
Table 4.6. Structurally homologous residues that bind the HCP hybrid cluster and the CoD C-cluster.....	105
5.1: Table 5.1. Crystallisation conditions for UgpG.....	117
Table 5.2. Data collection statistics for the several crystal types.....	121
Table 6.1. Refinement statistics for UgpG at 2.65 Å.....	130
Table 6.2. RMSD between UgpG monomers.....	137
Table 6.3. Inter-monomers interactions for tetramers 1 (a) and 2 (b).....	139
Table 6.4. Structural alignment statistics and PDB codes from the DALI structural search.....	145
Table 6.5. Active site residues - substrates interactions.....	146
Table 6.6. Three dimensional superposition of UgpG and RmlA structures.....	148
Table 8.1. Main features by levels of the SCOP protein classification...	162
Table 8.2. Main features by levels of the CATH protein classification.	163
Table 8.3. The seven crystal systems (adapted (Blundell and Johnson 1976)).....	169
Table 8.4. Non-centrosymmetric space groups (adapted from (Blundell and Johnson 1976)).....	169
Table 8.5. Typical data collection parameters at a synchrotron beamline of the ESRF.....	171

Illustration Index

Figure 1.1. The transition metal elements and some of their oxidation states.....	3
Figure 1.2. The most common co-ordination in iron-sulphur proteins.....	5
Figure 1.3. Prismane cluster.....	7
Figure 1.4. HCP sequence comparison.....	13
Figure 1.5. Topology diagram of HCP.	16
Figure 1.6. The hybrid cluster in <i>D. vulgaris</i> HCP.....	16
Figure 1.7. The HCP-HCR operon.	19
Figure 1.8. Haworth projection of UDP-glucose.....	21
Figure 1.9. Sequence dissimilarity for 39 members of the NTP transferase family.	22
Figure 1.10. UDP-glucose metabolic pathways in prokaryotes and eukaryote (Chandrasekaran and Radha 1995; Dutta et al. 1998; Nishikawa et al. 1998). Figure adapted from Marques Pina, 2003.....	23
Figure 1.11. Gellan gum structure.....	25
Figure 1.12. Pathway leading to the nucleotide sugar precursors.....	26
Figure 1.13. Proposed mechanisms for the glucose-1-phosphate nucleotidyltransferase family.....	28
Figure 1.14. Nucleotides.....	30
Figure 2.1. Ribbon representation of <i>D. vulgaris</i> HCP.....	34
Figure 2.2. As-isolated oxidised hybrid cluster in HCPs.....	36
Figure 2.3. Reduced hybrid cluster in HCPs.....	43
Figure 2.4. Superposition of the as-isolated oxidised and reduced clusters in <i>D. desulfuricans</i>	51
Figure 2.5. The environment of site Y in the reduced form of the hybrid cluster in <i>D. desulfuricans</i>	53
Figure 2.6. Immuno-blotting, using a polyclonal antibody against <i>D. vulgaris</i> HCP, of <i>D. desulfuricans</i> ATCC 27774 cell extracts from cultures grown with sulphate, nitrate and combinations of the two.....	56
Figure 3.1. The $ F_o - F_c $ electron density synthesis at the hybrid cluster in molecule A prior to the refinement stage.....	67
Figure 3.2. An overall view of the hybrid cluster protein.....	73
Figure 3.3. Models of the as-isolated oxidised (a) and reduced (b) forms of the hybrid cluster.....	74
Figure 3.4. SigmaA weighted electron density syntheses showing the presence of the as-isolated oxidised and reduced forms of the HCP in the vicinity of the hybrid cluster.....	78
Figure 4.1. The HCP monomer.....	92
Figure 4.2. Topology diagram of HCP.....	93

Figure 4.3. The hybrid clusters in HCP and CoD.....	94
Figure 4.4. The superposition of <i>D. vulgaris</i> and <i>D. desulfuricans</i> HCPs.	97
Figure 4.5. The superposition of HCP with CoDs.....	98
Figure 4.6. Topology diagram for a dimer of the beta-subunit of CoD..	101
Figure 4.7. Three-dimensional structure alignment.....	105
Figure 5.1. UgpG expression and purification.	113
Figure 5.2. UgpG selenium-methionine protein purification.....	115
Figure 5.3. UgpG native protein purification.....	116
Figure 5.4. Selenium energy scan (fluorescence counts vs keV) for a crystal grown from selenium-methionine protein.....	118
Figure 5.5. Crystal of native UgpG.....	119
Figure 5.6. Crystal of UgpG soaked in potassium osmate.....	119
Figure 5.7. Experimental electron density map at 3.5 Å resolution.....	122
Figure 6.1. UgpG monomer.....	136
Figure 6.2. UGP tetramer.....	139
Figure 6.3. Three-dimensional structure alignment.....	141
Figure 6.4. UgpG catalytic cavity.....	142
Figure 8.1. The growth of deposited proteins structures in the PDB.....	161
Figure 8.2. The diffraction experiment.....	167
Figure 8.3. The vapour diffusion technique.....	168

1. Introduction

1.1. Scope of the thesis

The thesis is divided into two parts corresponding to structural studies on two different proteins. The first part concerns the study of a metalloprotein, the hybrid cluster protein (HCP), with the goal of increasing knowledge regarding its structural properties in order to find new clues to its, as yet, unknown function. This thesis reports the protein isolation and crystallisation under oxygen free conditions, determination of the initial phases by multiwavelength anomalous dispersion (MAD) of the oxidised anaerobic structure, additional phasing with molecular replacement (MR) of the reduced structure, the high resolution refinement and structural analysis and extended comparison between the HCP and carbon-monoxide dehydrogenase (CoD) proteins. The second part of this thesis concerns structural studies on an enzyme with a known function, glucose-1-phosphate uridylyltransferase (UgpG), to improve knowledge of the structural properties of the whole glucose-1-phosphate nucleotidylyltransferase family. It reports the expression, purification and crystallisation of both the native protein and the selenium methionine derivatised protein, the structure determination using multiple isomorphous replacement with anomalous scattering (MIRAS), the structure model building, the refinement and analysis, and a comparison with related proteins.

1.2. Hybrid cluster protein (HCP)

1.2.1. Metals in proteins

In living organisms there are four groups of biological molecules: proteins,

lipids, sugars and nucleic acids, of which proteins are by far the most abundant. Their main constituents are carbon, nitrogen, oxygen, hydrogen and sulphur. However, for specific tasks proteins can incorporate prosthetic groups or co-factors in the form of small molecules and/or metals. In proteins many different metals, such as iron, nickel, cobalt, molybdenum and magnesium, are known to bind as simple ions or as metal containing co-factors. From the perspective of evolution, the transition metals (Figure 1.1) are particularly interesting since whilst remaining good Lewis acids they can be both π -electron donors and show redox activity, giving them increased functional capacity (Greenwood and Earnshaw 1994). During evolution, the use of transition metals in organisms underwent a separation between organisms heavily dependent upon nickel, iron and cobalt: the archaeobacteria; and other prokaryotes where the dominance of iron is more obvious (Fraústo da Silva and Williams 1991). The role of copper is also quite reserved in most prokaryotes relative to its multitude of roles in multicellular eukaryotes where the roles of nickel and cobalt are further diminished. Transition metals have undergone large changes in availability during evolution and biological systems have developed selected organic partners, e.g. Fe (haems), Co (corrin, B12), Ni (F-430), and Mo (pterin), whilst leaving much of the other transition elements in a freer but more vulnerable form to changes in the environment. These organic partners have lent versatility to the metals by increasing their redox potential sensitivity due to changes in ligand co-ordination and environment (Fraústo da Silva and Williams 1991).

1.2.2. Iron in proteins

The one transition metal element common to all life is iron. Iron plays a major role in protein chemistry being present alone, in haem groups or other complex arrangements (e.g. together with other metals such as nickel or

copper or as iron-sulphur clusters) (Andrews 1998; Crichton 2001). This thesis work concentrates on the most prominent non-haem

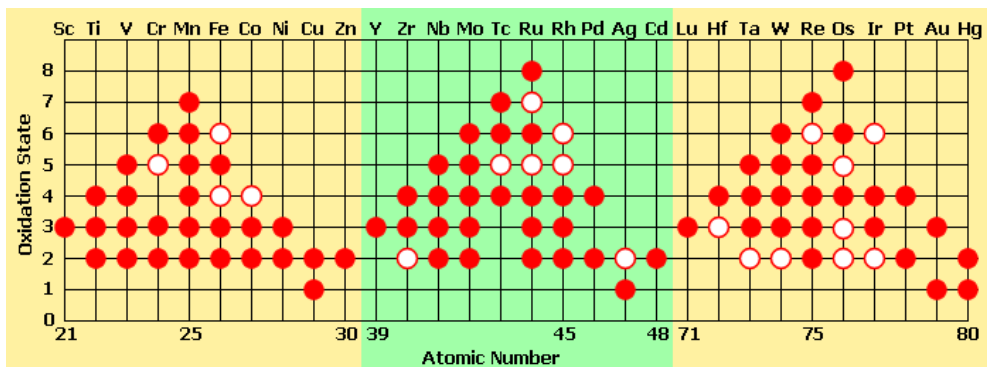


Figure 1.1. *The transition metal elements and some of their oxidation states.* In a broad definition, transition metals are the 40 chemical elements with atomic numbers 21 to 30, 39 to 48 and 71 to 80. The name transition comes from their position in the periodic table of elements. In each of the four periods in which they occur, these elements represent the successive addition of electrons to the *d atomic orbitals* of the atoms. In this way, the transition metals represent the transition between group 2 elements and group 13 elements. The solid circle represents a common oxidation state, and the ring represents a less common (less energetically favourable) oxidation state. (figure may be distributed either under licence [FDL](#) or [Attribution-ShareAlike 2.5](#))

proteins, the iron-sulphur proteins which are widespread throughout nature and are perhaps the richest class of metalloproteins due to the high affinity of iron for sulphur atoms and the availability of sulphide ions. It is thought that iron-sulphur clusters were the first redox complex systems, used by nature, with the ability to transfer electrons and store redox energy. Ferrous iron, Fe(II), is easily oxidised in the current atmosphere of the Earth, but was stable in the initial reducing conditions some billions of years ago, before the availability of oxygen. It is important to note that Fe(II) can readily participate in reactions leading to the production of the pernicious hydroxide radical, whereas ferric iron, Fe(III), is almost insoluble at physiological pH. Nature has therefore had to devise methods of maintaining Fe(II) in a “safe” environment and solubilising Fe(III) for transport and storage (Harrison and Arosio 1996; Andrews 1998). Some of the more important biological functions of iron-containing proteins are iron uptake and storage, electron

transfer chains (principal role), oxidase activity and dioxygen carrier.

1.2.3. Iron co-ordination in proteins

The main types of co-ordination of iron found in iron-sulphur proteins are mononuclear (e.g: rubredoxins), Fe_2S_2 ferredoxins, Fe_3S_4 ferredoxins, Fe_4S_4 ferredoxins and some higher clusters, in which all of the iron atoms are bound either to thiolate in the form of cysteine or sulphur atoms (Figure 1.2). The structures of a number of this family of proteins are known. To a first approximation, in the vast majority of them each iron is bound tetrahedrally by four RS^- moieties⁽¹⁾, alone as in rubredoxins or together with S^{2-} donors (ferredoxins). The fold of these proteins is based on beta-strands forming small beta-sheets; the iron co-ordination sphere is usually saturated and is held in a stable condition by this fold and is located in a protective groove in the protein. The exact stereochemistry of the iron, even in rubredoxins, is clearly dependent upon the exact protein fold. The iron-sulphur proteins are variously soluble in water or membrane phases and in this stable form and in either phase they appear overwhelmingly to have a simple single function: they are part of electron-transfer chains (Fraústo da Silva and Williams 1991). The redox switch of high-spin Fe(II) to high spin Fe(III) involves only a small change in Fe-S bond length due to the fact that it occurs in an interactive sulphide cluster resulting in little change in the relaxation energy so that the sites are ideal for functional activities. Conformational strain induced by the protein fold, which is undoubtedly observed in some Fe/S proteins, is not necessary for the basic function, but may be used to control redox potentials as may the general environments of the iron centres (Fraústo da Silva and Williams 1991). For the most part, the iron sulphur proteins have very negative redox potentials and typically in the range from 0.0 to -0.5V. Nevertheless, there are some unusual iron-sulphur proteins such as the

1 - Where R is typically a cysteine residue and S is the cysteine S_γ atom.

HiPIP (high potential iron protein) which are part of the mitochondrial and chloroplast redox chain, or a group of iron-sulphur proteins that are enzymes as opposed to electron-transfer proteins.

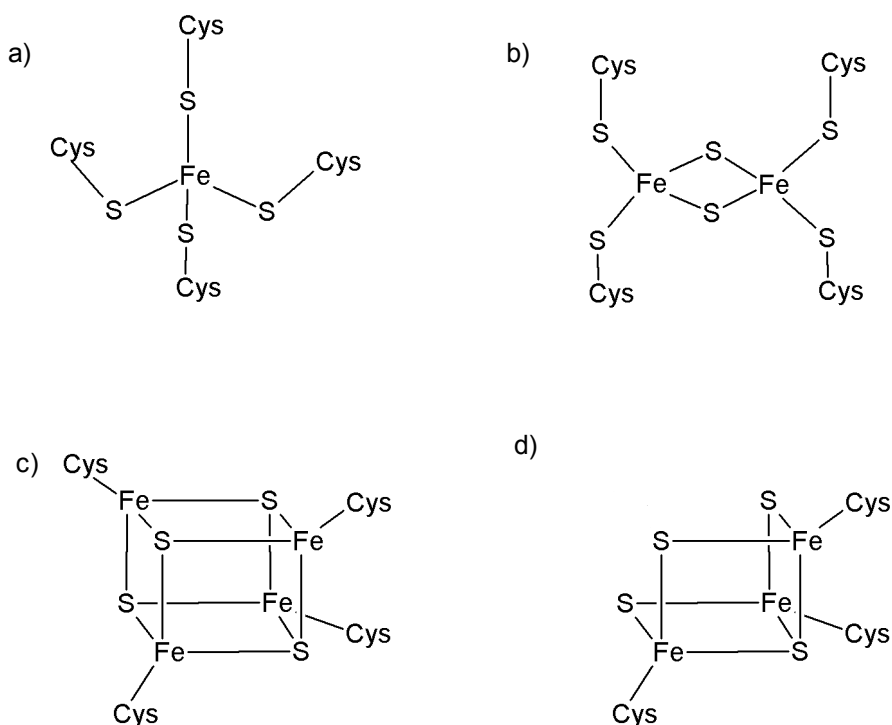


Figure 1.2. *The most common co-ordination in iron-sulphur proteins.*

(a) rubredoxin; (b) Fe_2S_2 ferredoxins. The Rieske-like proteins have two of the cysteine ligands replaced by histidines; (c) Fe_4S_4 ferredoxins. Iron hydrogenases have a cysteine ligand replaced by a histidine. d) Fe_3S_4 unit is like that of the Fe_4S_4 with one iron removed. Bacterial ferredoxins have a cysteine ligand replaced by an aspartate and there are also other variations.

1.2.4. Iron sulphur based enzymes

The iron sulphur proteins which bind substrates rather than transfer electrons are much more complicated since the acid/base properties of the iron atoms become important (Beinert and Kennedy 1989; Beinert 2000). The iron must be to some degree exposed to solvent, and may lead to dissociation of some of it linking iron containing proteins to iron sensing (Fraústo da Silva and Williams 1991). The peculiarity of these sites rests not just in the type of

ligation between the polypeptide chain and the iron atoms, but also in the site geometry, in the number of iron atoms in a cluster, and in the exposure of at least one iron atom to exchange with iron in the medium. The type of coordination varies and includes one bound histidine and one bound carboxylate from the protein. When the substrate binds, the single exposed iron of, for example, a Fe_4S_4 cluster, may become six co-ordinated and the centre probably generates a series of trapped free radical intermediates. Aconitase, which catalyses the citrate-isocitrate isomerisation in the tricarboxylic acid cycle, is one of these examples (Beinert and Kennedy 1989). One particular but important aspect of iron-sulphur clusters, as indicated above, is that they are very sensitive to dioxygen and their inorganic-chemistry resembles anaerobic biology (Holm 1977). The problem for the biochemist is to appreciate how such species can be formed today in the interior of cells which are exposed to dioxygen. Indeed, it is not always possible even to extract the ferredoxins without damage from dioxygen, and special methods for purification and manipulation of proteins in anaerobic environment have to be used.

1.2.5. Introduction to the hybrid cluster protein

Hybrid cluster protein (HCP), formerly known as the prismane protein, was first identified due to its highly unusual EPR signal, which was attributed to a, then still unknown in biology, type of iron-sulphur cluster (Hagen 1989). Initially it was thought to be a prismane cluster (Figure 1.3) due to the strong similarities between its EPR signal and that of synthetic $[\text{Fe}_6\text{S}_6]$ clusters. However, subsequently it was found to contain two different clusters, an $[\text{Fe}_4\text{S}_4]$ cubane cluster and the so-called hybrid cluster. HCP was first isolated from the strictly anaerobic sulphite reducing bacterium *Desulfovibrio vulgaris* (Hildenborough) and later from *Desulfovibrio desulfuricans*.

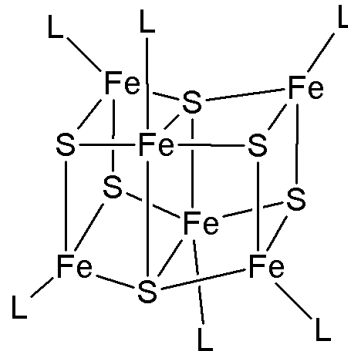


Figure 1.3. Prismane cluster.

Proposed iron-sulphur cluster present in prismane proteins on the basis of the similarity of the EPR spectra to those of the synthetic compounds with a [Fe₆S₆] core.

a) Classification of HCPs

The increasing availability of genomic sequence data and its analysis has shown that HCPs are spread throughout all kingdoms of life and not just in bacteria. They have been identified in archaea, e.g. *Methanococcus jannaschii* and eukaria, e.g. *Giardia intestinalis*, amongst others (Figure 1.4).

		Cubane binding region			
a)	Dvu	MF	FO	QETAKNT	GVKVG---MGKP
	Dde	SNAMF	YO	QETVGNK	GTQVG---VGKK
	Ctep	MH	DO	QESIKG	GVNVRG---VGKD
	Cac	MSMF	YO	QEAAGCK	GVTVRG---VGKT
	Ctet	MSMF	YO	QEAAGCK	GVTVRG---VGKT
	Cpe	MSMF	FO	QETAGCK	GVTVRG---VGKT
	Cth	MSMF	FO	QETAGCK	GVTVRG---VGKT
	Dha	MSMF	FO	QETAGCK	GVTVRG---VGKT
	Mja	MF	FO	QEAANE	GVTKG---VGKD
	Mma	MF	NO	QEALNVI	GVTKG---VGKK
	Tte	MKMF	YO	QEALKNT	GVTKG---VGKT
	Bth	MSMF	YO	QETAMGT	GVTLKG---VGKT
	Pgi	MEKMF	YO	QETAGNK	GVTLKG---VGKD
	Fnu	MDKMF	YO	QETAKGT	GVTSIG---VGKT
	Tva	MF	AO	EETNNNT	GVTKKG---VGKQ
Class II	Ehi	MEKMF	NO	QETAKGT	GVTVCG---VTKE
	Gla	MY	AO	EQTLHPG	GVVRAG---VSKS
	Eco	MIMF	VO	EQTIRTPAGNG	SYAQ---GMGKT
	Ype	MF	VO	EQTIRTPAGNG	SYAQ---GMGKT
	Sen	MF	VO	EQTIRTPAGNG	SYAQ---GMGKT
	Mmo	MY	VO	EQTMRTVPVNG	SYAQ---GMGKT
	Sty	MF	VO	EQTIRTPAGNG	SYAQ---GMGKT
	Sfl	MIMF	VO	EQTIRTPAENG	SYAQ---GMGKT
	Vpa	MF	IO	EQTIQTTPAVKG	SFAQ---GMGKT
	Vvu	MF	IO	EQTIQTTPAVKG	SFAQ---GMGKT
Class III	Pph	MF	IO	EQTLEAPKKKG	AYAT---GMGKT
	Son	MF	IO	EQTIRTPAGNG	SYAQ---GMGKL
	Tfe	MMF	YO	EQTRSPAGIG	TSEP---GTGKD
	Ban	MIGRGWDDMF	YO	EQTPXG	GKVMG---VGKN
	Bce	MIGRGWDDMF	YO	EQTPTG	GKVMG---VGKN
	Wsu	MSNIDLAMF	NO	EMSAPM	GKAGQEKGT---GKD
	Pab	MF	NO	SMSLSG	GVTVRG---VGKD
	Pfu	MEMAIRVPEAFDML	NO	SMSLAG	GVTVRG---VGKD
	Tma	MMF	YO	SQTANGT	GVTEYG---VGKS
	Mth	MGDPRSGIEPGTFFELPDDWLCPCVNVGKDQFEPLRGEIKRVRPEDIDMF	YO	SQTVRGR	AVTIRG---VGKE
		Rubredoxin domain			

	Dvu	EETANLQDILLFVLRGIAIYGEKLEKELGQPDERS----	NDDFVLQGLFATITNANWDDARFEAMISEGLARRDKLRNAFLA
	Dde	PETAALQDALLIYVTKGLGQIATRLRAEGKAVDH----	RIDRLVTGNLFATITNANFDDDDILAERVMTCAAKKELASLTD
	Ctep	ELTAKLQDILLIYAAEGIALCAEGF--EGKIDRK----	YGQFISECLFVTVTNTNFDDVAIVEQIRKALAMDEVR-----
	Cac	EDLAKKQDILLIYTLKVVSLYNVKARKLN-LVKN----	EIDNYIVDSLFAITTNANFDENIFTERIEKGLKYKEELKKQISA
	Ctet	EDVAKSQDILLIYILKGISYVGVKGREVG-VINK----	EVDNFMVGEGLFATITNANFDREVFIERIKRGLQLRQELKEQVIK
	Cpe	EDVAKIQDILLIFVTKGLATVANEGRKVG-IVDK----	KVNRMIIDNLFITITNANFDFKAIKERVKDTLVAREELKERVQA
	Cth	EEVAKIQDILLIYTVKGISYIVTKGNIDA-AKLG----	NTNHEVLSLFFMTITNANFDDGSGIEKQIRKMLAVRDEMKKSVOA
	Dha	ADVANKDILLIYVMKGIANSIQARELG-IVRQ----	DVDKFMVGEGLFATITNANFDNARFVALVREGALRDLKADIVK
	Mja	DVAVANLQDILLIYITKGLCYVCDKGNYL--DDE----	VMDYIP-KALFVTVITNANFDDKDVINWIKGVALREKIEKN---
	Mma	GEVADLQDILLIYVLKSVSYNLKARELG-LNEE----	TADKIVLDAFFATITNANFDDKQAIKESYIKKGEIRDSIKAK---
	Tte	ADVANKDILLIFTKGISFLNLKAREVG-VNKE----	KTRDFLIEGLFSTITNANFDRNFFLRKIEKAVLRDEIKEDLRN
	Bth	SEVANLQDILLIFVVRGIAVYNEHLRREGHPSE----	EADRFIYDALFITITNANFDKAAIIRKIEGQLKRELASKVITI
	Pgi	FSTANLMDLLFVFNKGLIAIIMTSMRAGVKADY----	RKADKAIMESLFAITTNANFDYSIAKRVETPAKKAELYSIAFT
	Fnu	SETSGIQDILLIYTEKGVAAYSTVFRKNGKAKELLEKVNRYLINSLFITITNANFDDNALDEIKAGLRELKALATD	
	Tva	ASTANIEDEMVOIVKEISIALGDGAVSI-----	EMGELLNKTLFMTITNANFDDIALKRLQDYCKEIKSKLTAKN---
	Ehi	PIVSNMTMDVLIYALRMNALVALKAKNGIEVDN----	EYHVLTCALFSTLTNANFSEKDISQIVQVFIQRKKLIEKLP-
	Gla	PEAAATQDAMIYGATLLGGFQLEED-----	HVLILOQSIFATLTNANFSAVDLRTRFLGPVTSRLEALSKESP
	Eco	AETSDLQDILLIAALQGLSAWAKAREYGIINH-----	VDSFAPRAFFSTLTNANFDSPRIVGYAREALREALKACQLA
	Ype	AETSDLQDILLIAALQGLSAWAKAREYGIIDSQ----	IDSFAPRAFFSTLTNANFDSRIVEYAKDAILLRHSLAVRCRL
	Sen	AETSDLQDILLIAALQGLSAWAKAREYGIINH-----	VDNFAPRAFFSTLTNANFDSRIVGYAREALREALKACQLS
	Mmo	AETSDLQDILLIAALQGLSAWAKAREYGIIDHD----	TDSFAPRAFFSTLTNANFDSRIVGYAREALREALKACQLA
	Sty	AETSDLQDILLIAALQGLSAWAKAREYGIINH-----	VDNFAPRAFFSTLTNANFDSRIVGYAREALREALKACQLS
	Sfi	AETSDLQDILLIAALQGLSAWAKAREYGIINH-----	VDSFAPRAFFSTLTNANFDSRIVGYAREALREALKACQLA
	Vpa	SEVSDLQDVLVYITLQGVSWASKALEFNIINDE----	INQWAPKAFFSTLTNANFDPERILELTSQAANYKALKKEQVMS
	Vvu	AEVSDLQDVLVYIMLQGVSVYAVKARELGVDTE----	INQWAPKAFFSTLTNANFDPERILQTSQAQFKAQLTKDQVLT
	Pph	AEISDLQDVLVYIILQGVSWADLAHKFDIIDDE----	INHAPRAFFATLTNANFDRERILALANLAENYKRLKQAVIA
	Son	AATSDLQDVLVYIMLQGVSVYAVKARELGVDTE----	VDTFVPEKAFFSTLTNANFDDERIIYAKAQAAQYRESLKNAYEA
	Tfe	EATAVLQDILLIYTHLMKGIASMR-AGRAMGVADRR----	TDDFIYGLFTTLTNANFTATRFVHLIQEASKRRERIKLLYEE
	Ban	ETIASLQDITIVFGLKGLIAAYRTHAAQLGYTDAF----	VDATTQEALYMTLTNSNFN-----
	Bce	ETIASLQDITIVFGLKGLIAAYRTHAAQLGYTDGF----	VDATTQEALYMTLTNSNFN-----
	Wsu	STLARIQDITIVFGLKGLIAAYRTHAAQLGYTDKFS----	VDDVMSETLYFTLTNANFN-----
	Pab	PDLSLQDALLIYIGIKGTAAAYYHHALEMGYDNPE----	IGHFLAEALYSTLTNANFN-----
	Pfu	PDLSLQDALLIYIGIKGTAAAYYHHALEMGYDDPE----	IGHFLSKALYSTLTNANFN-----
	Tma	PTVARLQDNLVFAIKGISAYYHHALEMGYDDPE----	IAGFLDEALYSTLTNANFN-----
	Mth	PTVARLQDNLVFAIKGISAYYHHALEMGYTDEE----	VDNFLERGFYSTLTNANFN-----

	Dvu	YVYKAKNGKDF-----	SEPLPEAATWTG-DSTAF--AEKAKSVGILATENE-DVRSRELLIIGLKG
	Dde	KSGLSDA-----	ALWEASEKSA--MLAKAGTVGMATDD-DVRSRLWLITFGLKG
	Ctep	---ALAG-----	TPAHDAANWSGSTKEEF--LAKAAACSIDLSADPDLRSKSLIYGLKG
	Cac	AGGNVN-----	LNKTEAVGVGLSTENE-DIRSLRELLITYGKIG
	Ctet	AGGNVNSHANKNWLKMLSFVGLKKEEETKLPDAATWADNVEE----	FNKAKEVGVGLATKDE-DIRSLRELLITYGKIG
	Cpe	KGGNPIG-----	SDFKGCATWTATTSEE--MMEKASQVGVGLATENE-DIRSLRELLITYGKIG
	Cth	EG-----	LHDAAVFSVDSRES--MLKKADSVGVGLSTQNE-DIRSLRELLITYGKIG
	Dha	AGGVLP-----	NLHDSATWTADSAAE--FEQKAQVGVGLATENE-DVRSRELLIYGLKG
	Mja	---NLNKEE-----	LPYCATWAYETDEDLI--NLANTKEVSVLAEDNE-DIRSLKELITYGKIG
	Mma	---LPAGT-----	LPAAEDLPDVAKVTPE--NITG-MOVAVHSTQNE-DVRSRELLITYGMKG
	Tte	KGVGVDE-----	YKNIDAI SWTYGTDAD--IEAISQEVGVGLSTEDNE-DIRSLKELITYGKIG
	Bth	A-----	NAPDECLWDGNEDE--FEEKARTVGVGLSTPDE-DIRSLKELVHVGKIG
	Pgi	Q-----IE-----	LPENEAVTMQGKPEE--YDRLALSVGLIRETDE-DIRSLKELITYGKIG
	Fnu	EKKKEAE-----	KYGADLVNWWYESNEDLIKFSQENQSVGVGLSTENE-DVRSRELLITYGKIG
	Tva	-----	ITAPLGVNSHNSNP-DVRSRELLITYGKIG
	Ehi	-----	KNELNGVEDVEDVAHYIETHLSDKRPNKWINQDQISGLMQTIIIFGLKG
	Gla	SOAMLDY-----	LALAKRILTESTDSWIKACRSTLAPDLTRKYGPTSAQRLYLTLGMKG
	Eco	VDANAR-----	VDNPMADLQLVSDDLGELQRAAEFTPNKDKAAIGENILGLRLCLYGLKG
	Ype	LDSTIT-----	VDHPLAEQLVADDIPSLQSQSQFALNNDKADVGDDIHGLRMLCLYGLKG
	Sen	VDANAR-----	CNPNMADLQLVSDDLGELQRAAEFTPNKDKAAIGENILGLRLCLYGLKG
	Mmo	KNPALV-----	VNHPLANLQLKGTDIASLVKQADEFALNNDKQAGDDIHGLRMLCLYGLKG
	Sty	VDANAR-----	CNPNMADLQLVSDDLGELQRAAEFTPNKDKAAIGENILGLRLCLYGLKG
	Sfi	VDANAR-----	VDNPMADLQLVSDDLGELQRAAEFTPNKDKAAIGENILGLRLCLYGLKG
	Vpa	AATLSNTN-----	LADIPAVANFELPNSAEAILAFAPQVAVNRKGDQVHEDVIGRLRLCLYGLKG
	Vvu	ASSLANSP-----	LSEVPAAEFELPENQAAILAFAPQVAVNRKGEVHEDVIGRLRLCLYGLKG
	Pph	ATTIAQQP-----	LPTLTPVAEFTLPNTIDAILTAPQVAVNRGYETINEDVIGRLRLCLYGLKG
	Son	TCEQSQKT-----	AEQMPVPAQLVLGTSKLEMLSQAPISLLNNDKNNIHEDVIGRLRLCLYGLKG
	Tfe	AAREQKKT-----	PEILSGFALFQPADSLEQLLRQAPSVAINADVEHLGSDVIGARALILYGMKG
	Ban	-----	-----
	Bce	-----	-----
	Wsu	-----	-----
	Pab	-----	-----
	Pfu	-----	-----
	Tma	-----	-----
	Mth	-----	-----

	1	2
	Dvu VNIGVGKNPGILISGDLKDMAELLKQTEG---TGVDVYTHGMLPANYYPAFKKYPHFVGNYGGSWWQQNPEFESFNG	
	Dde VNIGVGSNPGILISGDLRDLEMLLKQTEG---TGVDVYTHGMLPAHYYPAFKKYAHFVGNYNAGWKKQKEEFESFNG	
	Ctep VKTGVGTRPGILISGDLRDMEDLLKQTEG---TGVDVYTHGMLPAHYYPAFKKYDHFVGNYGNSWWSQDREFEFNG	
	Cac VNIGTRNPNPGILISGDLKDIEELLKQTEG---TGVDVYTHGMLPANYYPAFKKYSHFVGNYNAGWKKQKEEFESFNG	
	Ctet VNIGVRNPNPGILISGDLRDLEQLLEQTKG---TGVDVYTHGMLPAHYYPAFKKYSHFAGNYGNAGWKKQTEFEKFNNG	
	Cpe VNIGVRNPNPGILISGDLKDLEMLLEQTEG---TGVDVYTHGMLPGHYYPAFKKYAHFAGNYGNAGWKKQKEEFESFNG	
	Cth VNIGVRNPAIIVSGDLTDLEQLLEQTKG---TGVDVYTHGMLPAHYYPAFKKYDNFVGNYNAGWKKQVEEFESFNG	
Class I	Dha VNIGVRNPAIIVSGDLRDLEMLLKQTEG---TGVDVYTHGMLPAHYYPAFKKYAHFVGNYNAGWKKQKEFDSFNG	
	Mja VNLGVRDRPGILISGDLKDLEQLLEQSKD---AGVDIYTHGMLPAHYYPFFKKYEHFVGNYGGSWFFQREFEKFNNG	
	Mma VNIGVRNPNPGILISGDLLDLEQLLEQTKG---TGIDVYTHGMLPANSYYPAFKKYDNFAGNYGNAGWKKQKEEFESFNG	
	Tte VNIGVRNPNPGILISGDLKDLEQLLEQTAG---TGVDVYTHGMLPAHYYPAFKKYPHFVGNYNAGWKKQKEEFELFNG	
	Bth VNLGVRNPNPGILISGDLKDLEELLEQTEG---TGIDVYTHGMLPAHYYPQLKKYKHLAGNYGNAGWKKQKEEFESFNG	
	Pgi VNIGVGSRPGILISGDLKDMEMLEQTEG---TGIDVYTHGMLPANYYPFKKKYNHFFVGNYNAGWKKQKEEFESFNG	
	Fnu VKIGAGKRPILISGDLWDLKQLEQSKD---SGIDIYTHGMLPGHYYPFKKKYSHFVGNYNAGWKKQKEEFESFNG	
	Tva VNIGVGTRPGILISGDLRDLEMLLEQTKD---SGVDVYTHGMLPARGYYPFKKYPHLRGNYGGSWFFQDREFFPFNG	
	Ehi VLIGHKAGKCVLTGDLHDLEMLLKQTEG---KGVNVYTHGMLPAHYYPFLKKYKHLGHFVGNYNAGWKKQKEEFESFNG	
	Gla VPVRCMPGKCVLTGDMVDMLEMLTAAQE---KGVNVYTHGMLPAYTYPLRKSFCFKGHFVGNYNAGWKKQKEEFESFNG	
	Eco VNVKATAGKCVLTGDLKDLNMLEQTEG---TGVDVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
	Ype VNVVPVAGKAILISGDLKDLHMLEQTAG---TGVDVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
	Sen VNVKATAGKCVLTGDLKDLNMLEQTEG---TGVDVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
Class II	Mmo VNVVPVAGKAILISGDLKDLHMLEQTEG---TGVDVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
	Sty VNVKATAGKCVLTGDLKDLNMLEQTEG---TGVDVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
	Sfl VNVKATAGKCVLTGDLKDLNMLEQTEG---TGVDVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
	Vpa VNVKPVGKCVLTGDLHDLEMLLKQTEG---KGINVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
	Vvu VNVKPVGKCVLTGDLHDLEMLLKQTEG---KGINVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
	Pph VNVKPVGKCVLTGDLHDLEMLLKQTEG---KGINVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
	Son VNVKPVGKCVLTGDLHDLEMLLKQTEG---KGINVYTHGMLPAHYYPFLRKFHLVGNYGGSWQWQKEEFARFPNG	
	Tfe VRIISPQKAILISGDLHDLKQILEQTKD---QGINVYTHGMLPANAYPLKKYPHLAGNKGAWDQDQREFFADFPNG	
	Ban ITQNRVEGKAIIVVTGDLFALEELLKQTEG---KEINIYTHGMLPAHYYPQLKKYKHLKGNIGKAWYDQDQREFFADFPNG	
	Bce ITQNRVEGKAIIVVTGDLFALEELLKQTEG---KEINIYTHGMLPAHYYPQLKKYKHLKGNIGKAWYDQDQREFFADFPNG	
Class III	Wsu VSQNRVEGKAIIVVTGDLFALEELLKQTEG---KEINIYTHGMLPAHYYPQLKKYKHLKGNIGKAWYDQDQREFFADFPNG	
	Pab VPTGTAEHGILVTHGSHKALYELLNQIREMGLLEEEIKVYTHGMLPAHYYPQLKKYKHLKGNIGKAWYDQDQREFFADFPNG	
	Pfu VPTGTAEHGILVTHGSHKALYELLNQIREMGLLEEEIKVYTHGMLPAHYYPQLKKYKHLKGNIGKAWYDQDQREFFADFPNG	
	Tma VETGTAKKGIIVTGHNLKALEELLKQTEG---TNVYVYTHGMLPAHYYPQLKKYKHLKGNIGKAWYDQDQREFFADFPNG	
	Mth VRVGALEPAIATGSHSLALEELLKQTEG---TGIDVYTHGMLPAHYYPQLKKYKHLKGNIGKAWYDQDQREFFADFPNG	

```

      3
Dvu PILMTTNCLVPLKKENTYLDRLYTTGVVGYEGAKHIADRPAGGA--KDFSALIAQAKKCP-----P
Dde FVLMTTNCLVPPK--DSYKDRVYTTGIVGFTGCKHIPGEIGEH--KDFSALIAHAKTCP-----A
Ctep FILMTTNCLVFPVR--ESYGRMFTTGMAGYFGLKHIPARPDGGS--KDFSIIELAKTCK-----P
Cac FILMTTNCLVPPK--ASYIDRMYYTTGVTFYGVVGHIEGE--F--KDFSALIEHAKKCA-----P
Ctet FILMTTNCLVPPK--DSYKDRVYTTGVVGFPGGLKHIDADEKGE--KDFSIIELAKTCK-----A
Cpe FILMTTNCLVPPK--DSYGRGRIFTTGSAGFPGCKHIEADENGK--KDFSQIIELAKTCK-----A
Cth PILFTTNCLVPPRS--EEVRRRIFTTGSAGFPGCKHIEADENGK--KDFSIIELAKTLP-----A
Dha ALLTTNCLVPPK--DSYKDRIFTTGSVGVYEGVKHIPAREAGKV--KDFSIVIELAKTLP-----A
Mja PIVMTTNCLVPPK--DSYKDRVYTTGVVGFYGLKRIPVKEDGT--KDFSIVIEHAKKCK-----P
Mma FVLMTTNCLIPPR--ESYRNRIYTTGVVGFEGVGLPHIEKEDGT--KDFSIIIEHAKKCK-----P
Tte FILMTTNCLIPPR--DSYKDRVYTTGVVGFEGVGLPHIEKEDGT--KDFSIIIEHAKKCK-----P
Bth PILFTSNCLVPPRANASYKDRIFTTGNATGYGFKYIESDQHGR--KDFSALIAHAKQCC-----P
Pgi PILFTSNCLVPPKANASYKDRIFTTGNATGYGFKYIESDQHGR--KDFSIIELAKTCC-----P
Fnu PIIFTTNCLVPPKANASYKDRIFTTGNATGYGFKYIESDQHGR--KDFSIIELAKTCC-----P
Tva PIFMTTNCLVPPVR--KEYKDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Ehi AILFTTNCLMKPLP--SYKDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Gla PIVFTSNCLMKPLP--SYKDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Eco PIVMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Ype PILMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Sen PIVMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Mmo FILMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Sty PIVMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Sfl PIVMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Vva PIVMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Vvu PIVMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Pph PIVMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Son PIVMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Tfe PIVMTSNCLIDPTV--GAYDDRIFTTGAIVGVPELKHIPDRKDVNGQKDFSEIIVAKTCC-----P
Ban AILATTNCLVMPKIG--SYSDRFFSYDIAGLEGVQKIENDN--FAPLIQKALEP-----E
Bce AILATTNCLVMPKIG--SYSDRFFSYDIAGLEGVQKIENDN--FAPLIQKALEP-----E
Wsu AILVNTNCLVPPRSNCGYLDRLYTYKIVGVGEGAKKIENDN--FEPLIAQKALS-----S
Pab VIIGTSNCLVQOPTK--AYQDRIFTVGIAGLEGVPHIEDYD--FEPAIKRALET-----K
Pfu VIIGTSNCLVQOPTK--AYQDRIFTVGIAGLEGVPHIEDYD--FEPAIKRALET-----K
Tma AILGTSNCLVLPSE--SYRDRMFTTGIARLPGVKHIDGYD--YTEVIEKAKSLP-----D
Mth AILGTSNCLVLPSE--SYRDRMFTTGIARLPGVKHIDGYD--YTEVIEKAKSLP-----D

```

	Dvu	PVEIETGS-IVGGFAHNQVLA--LADKVVEAVKSGAIKRFVVMAG	4	DGRQKSRSYYTEVAENLPKDTVILTGA	5	AKYRYN
	Dde	PTEIESGE-IIGGFAHNQVLA--LADKVVIDAVKSGAIKRFVVMAG		DGRAKSRSYYTDFAEGLPKDTVILTGA		AKYRYN
	Ctep	PTEIENG-IVGGFAHTQVLA--LADKVVDVAVKSGAIRRFVVMAG		DGRHNSRQYYTDAETLTPKDTVILTGA		AKYRYN
	Cac	PTEIEKGE-IIGGFAHNQVLA--LADKVVEAVKSGAIKRFVVMAG		DGRAKTRNYYTDFAKALPKDAVILTGA		AKYRYN
	Ccet	PTEIEKGE-IVGGFAHNQVQ--LADKVVEAVKTGAIRRFVVMAG		DGRAKSRYNYYTEFAQKLPKDTVILTGA		AKYRYN
	Cpe	PTEIETGE-IVGGFAHNQVLA--LADQVVDVAVKSGAIRRFVVMAG		DGRAKSRYNYYTEFAQKLPKDTVILTGA		AKYRYN
	Cth	PDEIETGS-IVGGFAHNQVMA--LADKVVEAVKSGAVKRFVVMAG		DGRMKSRSYYTEFAQNLPKDTVILTGA		AKYRYN
Class I	Dha	PTEIETGE-IVGGFAHNQVFA--VADKVVEAVKSGAVKRFVVMAG		DGRMKSRYNYYTEFAKALPKDITILTGA		AKYRYN
	Mja	PTPLENGK-IVGGFAHNQVLA--LADKVIEAVKSGKIRKRFVVMAG		DGRHKTRYYTEFAKALPKDTVILTGA		AKYRFI
	Mma	PEQLES-IMGGFAHEAALS--VADKIIDAVKTGKISRFFVVMAG		DGRHKERAYNYYTEFAKALPKDTVILTGA		AKYRYN
	Tte	PVEIERGE-IIGGFAHNQVLA--LADKIVEAVKTGAIRRFVVMAG		DGRMKSRYNYYTEFAKALPKDTVILTGA		AKYRYN
	Bth	PTAIESGT-LIGGFAHAQVVA--LADKVVEAVKSGAIKRFVVMAG		DGRMKSRYNYYTEFAKALPKDTVILTGA		AKYRYN
	Pgi	PTEIESGT-IIGGFAHNQVLS--IADKVVEAVSSGAIKRFVVMAG		DGRQSGRNYNYYTEFAALPSDITILTGA		AKFRYN
	Fnu	PVEVESGE-ITVGGFAHNQVLS--LADKVVENIKSGAIKRFVVMAG		DGRMSQRNYYTEFAENLPKDTIILTGA		AKYRYN
	Tva	PKELEKQ-VTVGLAHDQVLA--VADKIVEGVKNGQIKKFCVMGG		DGRHKERDYTKVAKALPKDTVILTGA		AKYRYN
	Ehi	FTDDLKEDTVTGFSHHAILEEKTANKLVSLIKEGKIRNIFVVG		DGAKAGRNYYTELQVQIKPKDSIILTGA		CKFRFF
	Gla	AEGFRASEIRGAGFNHRVLKQPDILTNNVNIKTGQITCIQFNGG		DGASKDRSVFSKKALQADQKTLILTGA		CKFRFN
	Eco	FPYSEIPLHITVGFGRQTLG--AADTLIDLVSREKLRHIFLVGG		DGARGERNYYTDFATSVPPDCLILTGA		CKYRFN
	Ype	FPYSEIPLHITVGFGRQTLG--AADTLIDLVSREKLRHIFLVGG		DGSRTERSYYTDFARSVPDCCIIMTLGA		CKYRFN
	Sen	FPYSEIPLHITVGFGRQTLG--AADTLIDLVSREKLRHIFLVGG		DGARGERNYYTDFATSVPPDCLILTGA		CKYRFN
	Mmo	FPYSEIPLHITVGFGRQTLG--AADTLIDLVSREKLRHIFLVGG		DGSRGERSYYTDFARQVPDCLIMTLGA		CKYRFN
Class II	Sty	FPYSEIPLHITVGFGRQTLG--AADTLIDLVSREKLRHIFLVGG		DGARGERNYYTDFATSVPPDCLILTGA		CKYRFN
	Sfl	FPYSEIPLHITVGFGRQTLG--AADTLIDLVSREKLRHIFLVGG		DGARGERNYYTDFATSVPPDCLILTGA		CKYRFN
	Vpa	FQHDEIEHHITVGFGRNALMN--AAPAVIDQVKQGNIKHFFLVGG		DGDKAERSYYTDFATEAPEDTLILTGA		CKFRFN
	Vvu	FQHDEIEHHITVGFGRNALMN--AAPAVIDQVKQGNIKHFFLVGG		DGDKAERSYYTDFATEAPEDTLILTGA		CKFRFN
	Pph	FQHDEIEHHITVGFGRNALMN--AAPAVIDQVKQGNINHHFFLVGG		DGDKAERSYYTDFATEAPQDLSILTGA		CKYRFN
	Son	FHYDEIHPNITVGFGRNALMA--AAPTVENVKNGSIKHFFLVGG		DGDKSERSYFTDLAKSAPKDSIILTGA		CKYKFN
	Tfe	FTDAKEQRITVGFHHHTLLG--VADKIVDAVKHGIKRFVVIAG		DGVSPARNYYTEVADNAPDSVVMTLGA		CKYRFN
	Ban	VH-MESDEQLVTGFHHHTVLS--LAPEIINAVKEGKIRRFVVIAG		DAPGKGGEYYRELATSLPETVILTGA		CKFRFN
	Bce	VH-MESDEQLVTGFHHHTVLS--LAPEIINAVKEGKIRRFVVIAG		DAPGKGGEYYRELATSLPETVILTGA		CKFRFN
	Wsu	VDRFNAETTLVTGFHHHTVLS--LAPEIINAVKEGKIRRFVVIAG		DAPGKGGEYYRELATSLPETVILTGA		CKFRFN
Class III	Pab	MGRK-EGEKIVTGFHHHTVLA--MKDKLIELINEGKIRHIFVVG		DVPNPKMSYYEKLTELVPDAIILSAA		CKFRFN
	Pfu	MKAY-DGKLLTGFHHHTVLA--MKDKLIELINEGKIRHIFVVG		DTPHKGMYEKLTELIPKDAIILSAA		CKFRFN
	Tma	LEEKPGSYKLRGTGFTSVVVS--LADKIKELVEAGKIRHIFLVGG		DVFFRKNYYEKLTELIPKDAIILSAA		CKFRFN
	Mth	LSEE-DATTITGFGLSITLS--LADKIRELVEDGKIRRFVVG		DSPLPKRYNYYTEFVRKLPEDTVILTGA		CKYRFN

Class I { Dvu KL-NLGDIGG--IPRVLDAGG--NDNSYSLAVIALKLKEVFGLEDINDLPVSYDIANWYEQKAVAVLLALLFLGVKGIRLGP
 Dde KL-NLGDIGG--IPRVLDAGG--NDNSYSLAVIALKLKEVFGLEDVNDLPVYNIAWYEQKAVIVLLALLSLGVKNHILGPT
 Ctep KL-ELGDIGG--IPRVLDAGG--NDNSYSLAVIAMKLQEVFGLENINDLPISYDIANWYEQKAVIVLLALLFLGVKGIRLGP
 Cac KL-DLGDIGG--IPRVLDAGG--NDNSYSLVVIKALKLKEVFGLEDINELPIAYNIAWYEQKAVIVLLALLHLGVKNHILGPT
 Ctet KL-NLGDIGG--IPRVLDAGG--NDNSYSLALIALKLKEVFELEDINELPISYNIAWYEQKAVIVLLALLHLGVKNHILGPT
 Cpe KL-PLGDING--IPRVLDAGG--NDNSYSLVVIKALKLAEVFGTDSVNELPISYNIAWYEQKAVIVLLSLHLGVKNHILGPT
 Cth KL-GLGDIGG--IPRVLDAGG--NDNSYSLAVIALKLKEVFGLEDINKLPIAFNIAWYEQKAVIVLLALLYLGVKNHILGPT
 Dha KL-ALGDIGG--IPRVLDAGG--NDNSYSLAVIALKLKEVFGLEDVNLQPLISYNIAWYEQKAVIVLLALLYLGVKNHILGPT
 Mja KL-DLGDIGG--IPRVLDAGG--NDNSYSLVVIKALKLAEVFGLEDINELPISYNIAWYEQKAVIVLLALLSLGVKNIVLGP
 Mma KL-DLGDIGG--IPRVLDAGG--NDNSYSLVVIKALKLAEVFGLEDINELPISYNIAWYEQKAVIVLLALLSLGVKNIVLGP
 Tte KL-NLGDIGG--IPRVLDAGG--NDNSYSLAVIAMKLKEIFGLNDINELPISYNIAWYEQKAVIVLLALLYLGVKNHILGPT
 Bth KL-ALGDING--IPRVLDAGG--NDNSYSLAVIALKLKEVFGLEDVNLKPLIVYNIAWYEQKAVIVLLALLHLGVKNHILGPT
 Pgi KL-QLGDIGG--IPRVLDAGG--NDNSYSLAVIALKLKEVFGLEDINKLPIVYNIAWYEQKAVIVLLALLSLGVKNHILGPT
 Fnu KL-NLGDING--IPRVLDAGG--NDNSYSLAVIALKLKEVFGLEDINELPIVFNIAWYEQKAVIVLLALLYLGIKNHILGPT
 Tva KL-DLGDING--IPRVLDAGG--NDNSYSLAVIALKLKEVFGLEDVNLKPLAFDIAWYEQKAVIVLLALLFLGVKNHILGPT
 Ehi RE-KLGDIEG--IPRLDLGG--NDNSYSAIVVIALKLAEIFKCG--VNELPLHLILSWYEQKAVAVLLALLYLGIKNHILGPT
 Gla RLQDYGEIN--IPRLDLGG--NDNSYSAIVVIALKLAEIFKCG--VNELPLHLILSWYEQKAVAVLLALLYLGIKNHILGPT
 Eco KL-EFGDIEG--LPRLDVGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Ype KL-DLGDIEG--LPRLDVGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Sen KL-EFGDIEG--LPRLDVGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Mmo KL-EFGDIEG--LPRLDVGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Sty KL-EFGDIEG--LPRLDVGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Sfl KL-EFGDIEG--LPRLDVGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Vpa KN-FFGDING--IPRLDLGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Vvu KN-FFGDING--IPRLDLGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Pph KN-FFGDING--IPRLDLGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Son KL-EFGDIEG--LPRLDVGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Tfe KH-EFGDIEG--IPRLDLGG--NDAYSAILAVTLAEKLGCG--VNDLPLSLVLSWFEQKAVIVLLTLLSLGVKNIVTGP
 Ban DV-DYGVVGTETIPRYDLGG--NNISITVKIAAALADAFQCE--VNELPVSIVLSWFEQKAVAILLGLFSLGIQDIRIGPK
 Bce DV-DYGVVGTETIPRYDLGG--NNISITVKIAAALADAFQCE--VNELPVSIVLSWFEQKAVAILLGLFSLGIQDIRIGPK
 Wsu DI-DFGTIEGTIPRYDLGG--NNISITVKIAAALADAFQCE--VNELPVSIVLSWFEQKAVAILLGLFSLGIQDIRIGPK
 Pab RR-DYGSIEG--IPRFMDGG--NNVYSIIEIALALAKELMD--INSLPVSIVLSWFEQKAVAILLGLFSLGIQDIRIGPK
 Pfu AR-NYGTIEG--IPRFMDGG--NNVYSIIEIALALAKELMD--INSLPVSIVLSWFEQKAVAILLGLFSLGIQDIRIGPK
 Tma DL-DLGDIGG--IPRLDLGG--NDTIVAIIEIAQALAKVFGVE--VTEPLTLVLWMEQKAVAILLGLFSLGIQDIRIGPK
 Mth SM-DLGDIEG--VPRLDLGG--NDTIVAIIEIAQALAKVFGVE--VTEPLTLVLWMEQKAVAILLGLFSLGIQDIRIGPK

Class I { Dvu LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Dde LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Ctep LPEFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Cac LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Ctet LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Cpe LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Cth LPEFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Dha LPEFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Mja LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Mma LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Tte LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Bth LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Pgi LPAFLSPNVAKVLVFNENFIKK--IGTVQDDIAAMMAGK-----
 Fnu LPEFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Tva LPEFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Ehi LPAFLTPGVFKYLDNLDLMM--TTTPDNDLANL-----
 Gla LPEFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Eco APGFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Ype APGFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Sen APGFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Mmo LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Sty APGFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Sfl APGFLTPTNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Vpa LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Vvu LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Pph LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Son LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Tfe LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Ban LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Bce LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Wsu LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Pab LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Pfu LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Tma LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----
 Mth LPAFLTPNIAITLVKLEFGLK--IGTVQDDIAAMMAGK-----



Figure 1.4. HCP sequence comparison.

(a) Aligment of known and putative HCP protein sequences - *Desulfovibrio vulgaris* (Dvu), *Desulfovibrio desulfuricans* (Dde), *Chlorobium tepidum* (Ctet), *Clostridium acetobutylicum* (cac), *Clostridium tetani* (Ctet), *Clostridium perfringens* (Cpe), *Clostridium thermocellum* (Cth), *Desulfotobacterium hafniense* (Cha), *Methanococcus jannaschii* (Mja), *Methanosarcina mazei* (Mma), *Thermoanaerobacter tengcongensis* (Tte), *Bacteroides thetaiotaomicron* (Bth), *Porphyromonas gingivalis* (Pgi), *Fusobacterium nucleatum* (Fnu), *Trichomonas vaginalis* (Tva), *Entamoeba histolytica* (Ehi), *Giardia lamblia* (Gla), *Escherichia coli* K12 (Eco), *Yersinia pestis* (Ype), *Salmonella enterica* (Sen), *Morganella morganii* (Mmo), *Salmonella typhimurium* (Sty), *Shigella flexneri* (Sfl), *Vibrio parahaemolyticus* (Vpa), *Vibrio vulnificus* (Vvu), *Photobacterium phosphoreum* (Pph), *Shewanella oneidensis* (Son), *Thiobacillus ferrooxidans* (Tfe), *Bacillus anthracis* (Ban), *Bacillus cereus* (Bce), *Wolinella succinogenes* (Wsu), *Pyrococcus abyssi* (Pab), *Pyrococcus furiosus* (Pfu), *Thermotoga maritima* (Tma) and *Methanobacter thermotrophicus* (Mth) prepared using the program ClustalX/ClustalW (Higgins *et al.* 1992; Thompson *et al.* 1994; Thompson *et al.* 1997). The cubane binding region is highlighted in red with residues binding the cubane cluster boxed in grey. The rubredoxin domain for Mth HCP is highlighted in yellow. Residues binding the hybrid cluster are numbered 1 to 6 and boxed in black. **(b) Sequence dissimilarity representation.** Tree representation build from clustalX alignment, distance matrix by the tree-puzzle method (Schmidt *et al.* 2002) and neighbour-joining method (Saitou and Nei, 1987). The colour code highlight represents in green organisms from the Eukarya kingdom, in orange organisms from the Archaea kingdom and in yellow bacteria with HCP reductase (partner for HCP). In blue are highlighted the X-ray structures determined to date.

Sequence discrepancies between organisms have been noted, in particular the sequential arrangement of four cysteines ligating the cubane cluster. These led to a proposal for the classification of HCPs based on the cysteine spacings (van den Berg *et al.* 2000). However, not only the sequential arrangement of the four cysteine ligands of the cubane cluster is unusual but HCPs also show different sequence arrangements from organism to organism (Bailey *et al.* 2001). Class I HCPs, found in strictly anaerobic bacteria and in the methanogenic archeon *Methanococcus jannaschii*, have the characteristic C-X₂-C-X_{7,8}-C-X₅-C spacing; class II HCPs, found in facultative anaerobic bacteria, are characterised by a C-X₂-C-X₁₁-C-X₆-C spacing; class III HCPs, found in thermophilic bacteria and archaea, have spacings similar to those of class I HCP proteins, but lack more than one hundred residues downstream of the N-terminal cysteine region (Figure 1.4a,b). There is one exception, HCP from *Wolinella succinogenes*, that has the C-X₂-C-X₇-C-X₉-C spacing and has the same deletion of the other class III HCP. Interestingly, the HCP protein from *Methanobacter thermautotrophicus* has an additional rubredoxin domain at the N-terminal (Figure 1.4a).

b) Description of the structure

HCP contains about 550 amino acids (553 in *D. vulgaris*; the residue numbering scheme of *D. vulgaris* will be used throughout this section), arranged in three domains, and include two iron-sulphur clusters, a cubane cluster and a hybrid cluster. Domain 1 (residues 1 to 221) comprises a metal binding region and two three-helix-bundles. The two bundles are approximately orthogonal to each another and are linked together through a 36 residue polypeptide segment that contains a short helix. Domains 2 and 3 (222 to 375 and 376 to 553, respectively) have central 6-stranded β -sheets

with helices on either side. These two domains are homologous, superposing by alignment of their central β -sheets with a 1.9 Å RMSD for the 48 equivalent C α atoms of the central beta-sheet (Figure 1.5). This first domain contains in its interior the cubane-type iron-sulphur cluster whilst the hybrid cluster is located at the interface point of the three domains. The cubane cluster is bound to the protein through four covalent bonds between its iron atoms and the cysteine S γ atoms of residues 3, 6, 15 and 21, all from the N-terminus. This contrasts with a typical cubane to protein linkage in ferredoxins where the fourth ligand usually belongs to the C-terminal part of the sequence. The hybrid cluster is a non-symmetric iron-sulphur cluster found so far only in HCPs and with similarities to the C-cluster found in some carbon-monoxide dehydrogenases (CoDs). The HCP hybrid cluster can be described as a 2Fe-2S (Fe5 and Fe6) square base capped by one iron (Fe8) at the expected vertex to form a cube and a second iron (Fe7) displaced from the last vertex (Figure 1.6). Atoms Fe5 and Fe6 are bound to the protein through cysteine S γ atoms, from residues 434 and 312 respectively. However, Fe8 is bound to a less common per-cysteine sulphur atom S δ of Cys406 and to a carboxylic oxygen from Glu494. Furthermore, Fe7 is bound to the S γ atom from Cys459, to a carboxylic oxygen of Glu268 and to the imidazole N ϵ of His244. The ligation of Fe7 to the cluster is obtained through two bridging cluster oxygen atoms. In *D. vulgaris*, the nature and position of O10 is not completely clear since there are two electron density peaks 0.8 Å apart and each can be modelled by an half occupied oxygen atom. However, both resonance Raman (de Vocht *et al.* 1996), crystallographic refinement and EXAFS spectra (Arends *et al.* 1998; Cooper *et al.* 2000; Macedo *et al.* 2002) all point to O10 being an oxygen species.

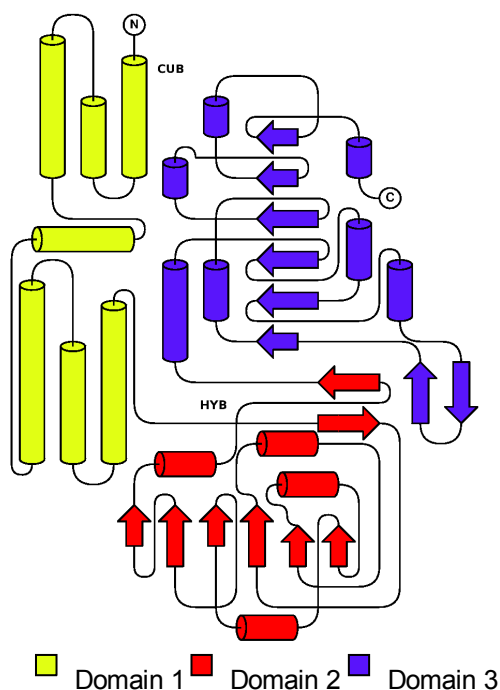


Figure 1.5. Topology diagram of HCP.

The molecule is divided into three domains, numbered 1 to 3, and coloured in yellow, red and blue. The locations of the cubane and hybrid clusters are labelled CUB and HYB, respectively.

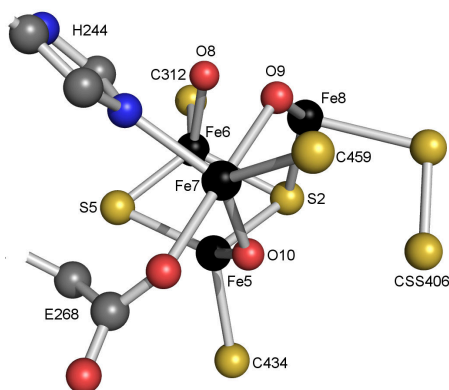


Figure 1.6. The hybrid cluster in *D. vulgaris* HCP.

Ball and stick representation with carbon atoms coloured in grey, nitrogen in blue, oxygen in red, sulphur in yellow and iron in black. The position and nature of O10 was under debate because of an elongated electron density (not shown), with peaks at 0.8 Å apart, at this location.

c) *Structures at different oxidation states*

Several X-ray crystal structures of HCPs became available in 2001 (Arendsen *et al.* 1998; Cooper *et al.* 2000) but all of these structures were obtained from proteins purified and crystallised under an aerobic environment. There was a debate as to whether the observed hybrid cluster truly represented *in vivo* reality, or was a result from an oxidation event which occurred during the purification and crystallisation processes. This resulted in a recognised need for an anaerobic purification, crystallisation and structural study, as well as the determination of a fully reduced structure, in order to address the puzzling questions that the oxidised structures had raised. Additionally, such data potentially could add insights into the still unknown physiological functions of HCP. Purification in an anaerobic environment greatly increases the complexity of the process. The availability of several experts on this subject, and in particular the late Jean Legall, was critical for the success of the project. Anaerobic as-isolated and reduced structural studies were accomplished independently for *D. vulgaris* (this thesis) and for *D. desulfuricans* (Macedo *et al.* 2002; Aragao *et al.* 2003).

d) *Biological function*

Despite all the wealth of structural and spectroscopic data available, the physiological role of HCP is still under considerable discussion. HCP is known to be able to perform redox reactions and has been shown to be reduced by an NAD(P)H oxidoreductase (van den Berg *et al.* 2000), which is a product of the *hcr* gene immediately downstream of the *hcp* gene in *Escherichia coli*. The two genes may form an operon in *E. coli* and related species (van den Berg *et al.* 2000) (Figure 1.7). However, there is no evidence for a similar redox partner in the genomes of obligate anaerobes

and in (hyper)thermophilic archaea, where the *hcp* gene does not appear to be part of an operon (van den Berg *et al.* 2000; Bailey *et al.* 2001). Two main hypotheses for a functional role have been raised. The first concerns the involvement of HCP in protection against reactive nitrogen species whereas the second suggests an involvement in oxygen related oxidative protection mechanisms. The first hypothesis gained substance with the report of the conversion of the HCP related, NiFeS carbon monoxide dehydrogenase to a hydroxylamine reductase (Heo *et al.* 2002), the report of hydroxylamine reductase activity for *E. coli* HCP (Wolfe *et al.* 2002) as well as increased anaerobic tolerance to hydroxylamine of *E. coli* cells overproducing *Rhodobacter capsulatus* HCP (Cabello *et al.* 2004). However the very high optimum pH (Cabello *et al.* 2004) and high K_m at neutral pH (Aragao *et al.* 2003) determined for this activity do not allow complete confidence in assigning the physiological function for HCP as a hydroxylamine reductase. In *E. coli* it has been reported that *hcp* transcription is regulated in response to nitric oxide and to the regulatory protein, NsrR, involved in reactive nitrogen stress response (Filenko *et al.* 2005; Flatley *et al.* 2005; Filenko *et al.* 2007). This reinforces previous reports pointing to increased *hcp* transcription in response to nitrogen oxides in *Salmonella enterica* (Kim *et al.* 2003), *Shewanella oneidensis* (Beliaev *et al.* 2005) and *D. vulgaris* (He *et al.* 2006). In the light of these reports, a role of HCP in protection against a reactive nitrogen compound, other than hydroxylamine, cannot be ruled out. Reports of HCP involvement in other types of stress (Briolat and Reysset 2002) and the *hcp* gene being present in a large number of organisms not involved in the biological nitrogen cycle led to the suggestion of involvement in oxidative stress response (Aragao *et al.* 2003). Almeida and co-workers (2006) have reported both HCP induction by hydrogen peroxide, an induction to be regulated by the redox-sensitive transcriptional activator OxyR, and peroxidase activity for *E. coli* and *D. desulfuricans*

HCPs using ascorbate as the electron donor (Almeida *et al.* 2006). Despite a K_m within the values measured for typical peroxidases, the low V_{max} obtained for both HCPs, from *E. coli* and *D. desulfuricans*, still poses doubts about the true physiological function for this protein.

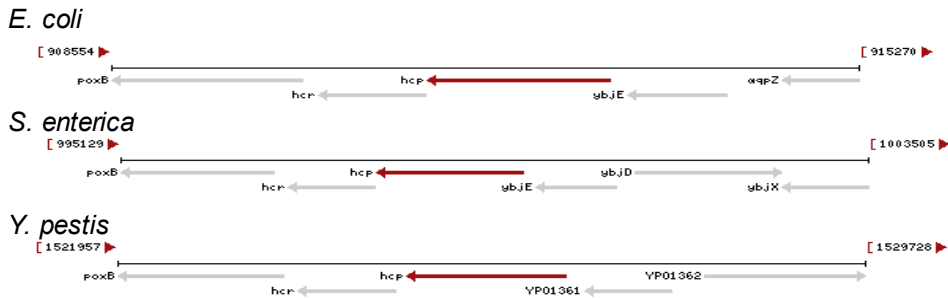


Figure 1.7. The HCP-HCR operon.

Salmonella enterica, *Escherichia coli* and *Yersinia pestis* have the two genes, *hcp* and its reductase *hcr*, close in their genome, forming a common transcription unit.

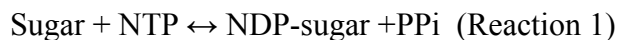
1.3. UDP-glucose uridylyltransferase (UGP)

Sugars (carbohydrates) are aldehyde or ketone compounds with multiple hydroxyl groups. They make up a large percentage of organic matter on Earth because of their multiple roles in all forms of life. Their functions vary from structural elements in the cell walls of bacteria and plants, ribose and deoxyribose components of DNA and RNA, highly polar surface coats for proteins, energy stores, fuels and metabolic intermediates (Stryer 1995). All organisms have evolved to be able to capture, process, store and sometimes synthesise sugars.

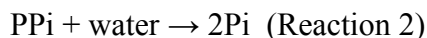
1.3.1. Sugar activation

One important step for the use of sugars is their activation. This is a process

whereby the sugar is bound to an activation agent and gains propensity for reaction. One example is ATP, the universal currency of free energy, which is a phosphorylated sugar derivative. As mentioned above, sugar activation is an important reaction pathway in living organisms. One particular sugar activation process is the reversible transfer of nucleotide triphosphate (NTP), which consists of the transfer of a nucleotide diphosphate (NDP) group to a sugar molecule with the loss of pyrophosphate (PPi) (Reaction 1).



However, pyrophosphate can be irreversibly hydrolysed, *in vivo*, to orthophosphate by enzymes such as inorganic pyrophosphatase (Reaction 2).



This can lead to an overall irreversibly driven reaction (Reaction 3) leading to a recurring theme in cellular chemistry: the hydrolysis of pyrophosphate driving a certain desired biochemical reaction.



One such sugar activation is the production of UDP-glucose, an activated form of glucose (Figure 1.8). This last example has broad significance since nucleoside diphosphate sugars serve as glycosyl donors in the biosynthesis of many disaccharides and polysaccharides (Stryer 1995; Sá-Correia *et al.* 2002).

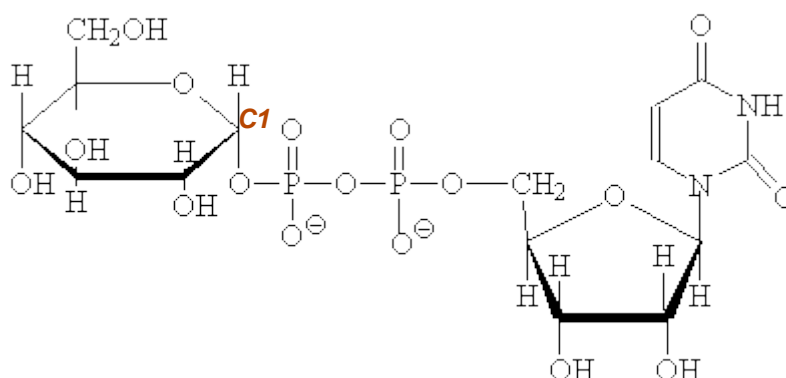


Figure 1.8. *Haworth projection of UDP-glucose.*

The C-1 carbon atom (in red) of the glycosyl unit of UDP-glucose is activated because its hydroxyl group forms an ester link to the diphosphate moiety of UDP.

In living cells the protein responsible for the reversible catalysis of glucose-1-phosphate into UDP-glucose is glucose-1-phosphate uridylyltransferase (UGP; EC 2.7.7.9). Sequence comparisons have shown that this protein is distinct between the bacteria and eukaryotic domains of life. Some authors proposed that this happened because prokaryotic UGPs have diverged during the evolution process (Hossain *et al.* 1994) but others argue that eukaryotic UGPs did not evolve from the same template (Eimert *et al.* 1996). This difference, between proteins from bacteria and eukaryotic domains is seen in all the nucleotidylyl-family with the exception of glucose-1-phosphate adenylyltransferase (Figure 1.9).

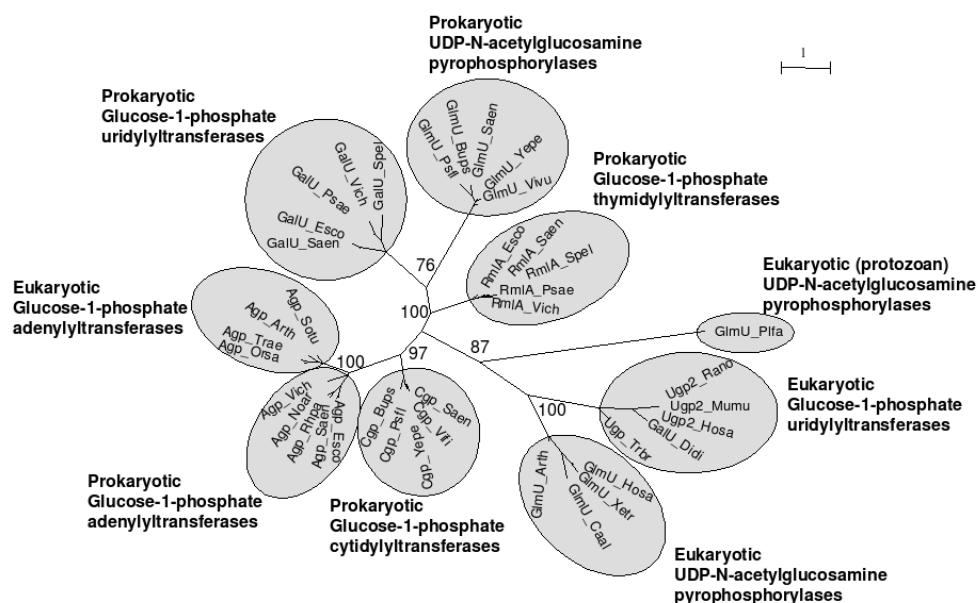


Figure 1.9. Sequence dissimilarity for 39 members of the NTP transferase family.

Tree representation built from a clustalX/ClustalW alignment (Higgins *et al.* 1992; Thompson *et al.* 1994; Thompson *et al.* 1997), distance matrix by the tree-puzzle method (Schmidt *et al.* 2002) and neighbour-joining method (Saitou and Nei 1987). The figure shows large differences between the NTP transferases from eukaryotic and prokaryotic counterparts, which form two distinct branches. The exception is glucose-1-phosphate adenyltransferase where both eukaryotic and prokaryotic proteins share the same branch. There are neither eukaryotic glucose-1-phosphate thymidylyltransferases or glucose-1-phosphate cytidylyltransferases in Nature.

The UGP enzyme in prokaryotic organisms is typically involved in the synthesis of polysaccharides, many of them important for bacterial pathogenesis. This has led to UGP being a potential target for inhibitors (Mollerach *et al.* 1998). UDP-glucose, the product of the reaction catalysed by UGP, is also a precursor in the production of threulose, which is a disaccharide important for protection against osmotic stress (Giaever *et al.* 1988; Hengge-Aronis *et al.* 1991; Weissborn *et al.* 1994), of gluten, which is a sugar storage polysaccharide, and of catabolism of galactose into glucose via the Leloir pathway (Frey 1996) (Figure 1.10).

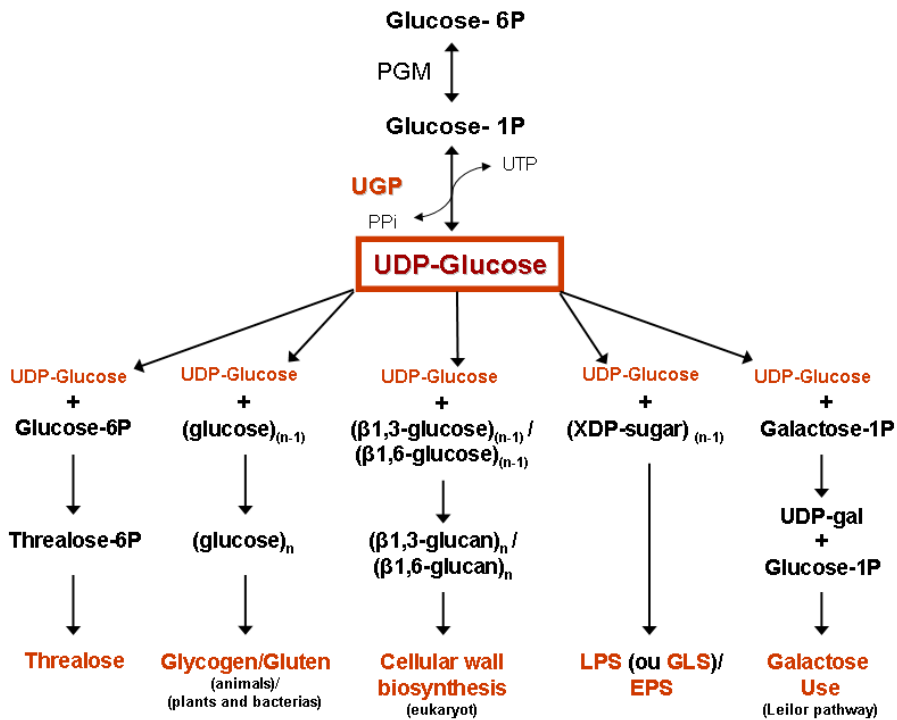


Figure 1.10. UDP-glucose metabolic pathways in prokaryotes and eukaryote (Chandrasekaran and Radha 1995; Dutta *et al.* 1998; Nishikawa *et al.* 1998). Figure adapted from Marques Pina, 2003

1.3.2. *Sphingomonas elodea* and gellan

The Gram-negative bacteria *S. elodea* has great potential for bioremediation since it is able to colonise different environments, including water, soil and sediments, thanks to its broad range of catabolic capabilities (Fredrickson *et al.* 1995; Dutta *et al.* 1998; Nishikawa *et al.* 1998). However, this bacterium has also gained importance as *the* industrial method to produce gellan. This is a very efficient and low cost process since the large scale industrial fermentation of *S. elodea* from glucose gives a typical gellan yield of 40 to 50%. Gellan, also known as E418, is an exopolysaccharide (EPS) used as a gelling, stabilising and suspending agent that is approved in both the US and the EU for application in the food, pharmaceutical and other industries

(Chandrasekaran *et al.* 1995; Sutherland 1998). In the US, the production of gellan gum by the leading global producer, CP Kelco, is 100 tonnes/year (data from 2003) with a value of tens of billions of US dollars and the Environmental Protection Agency estimated, in 2004, an annual exposure of gellan gum to the US population of about approximately 0.94 mg/person/day (EPA 2004).

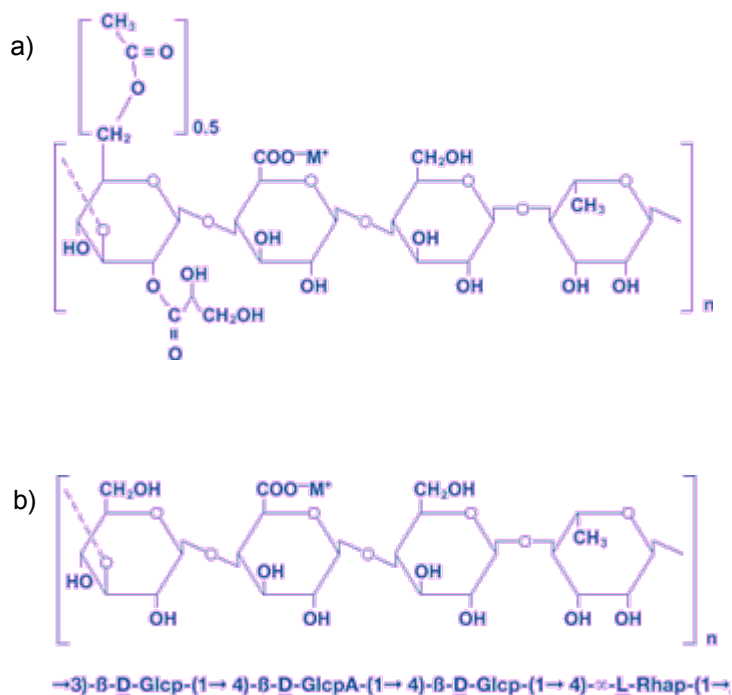


Figure 1.11. *Gellan gum structure.*

The molecular structure of gellan gum is a straight chain based on repeating glucose, rhamnose and glucuronic acid units. In its native, or high acyl form, two acyl substituents acetate and glycerate are present. Both substituents are located on the same glucose residue and, on average, there is one glycerate per repeat and one acetate per every two repeats. The high acyl form (a) produces soft, elastic, non-brittle gels, whereas the low acyl form (b) produces firm, non-elastic, brittle gels. Figure adapted from CP Kelco, Huber company marketing brochure.

In its native form, gellan is a linear anionic heteropolysaccharide based on a tetrasaccharide repeat unit composed of two molecules of D-glucose (Glc),

one of D-glucuronic acid (GlcA) and one of L-rhamnose (Rha) (Figure 1.11). The gellan biosynthetic pathway has been partially elucidated (Martins and Sa-Correia 1991; Sá-Correia *et al.* 2002) at the level of the enzymes necessary for the formation of the sugar nucleotides UDP-D-Glc, UDP-D-GlcA and TDP-L-Rha, the donors of the monomers for the assembly of the tetrasaccharidic unit (Figure 1.12).

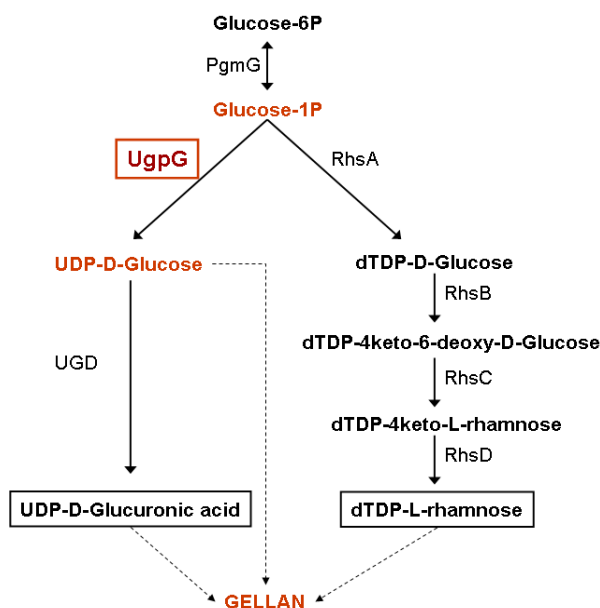


Figure 1.12. Pathway leading to the nucleotide sugar precursors.

UDP-D-glucose, UDP-D-glucuronic acid and dTDP-L-rhamnose are involved in gellan gum biosynthesis. The genes that are predicted or proved to encode some of the enzymes are in parenthesis. Abbreviations: PgmG, phosphoglucomutase (*pgmG*); UgpG, UDP-D-glucose pyrophosphorylase (*ugpG*); TGP, dTDP-D-glucose pyrophosphorylase; UGD, UDP-Dglucose dehydrogenase; TRS, dTDP-L-rhamnose enzyme system [dTDP-D-glucose-4,6-dehydratase (*rhsB*)+dTDP-4-dehydrorhamnose-3,5-epimerase (*rhsC*)+dTDP-4-dehydrorhamnose reductase (*rhsD*)]. Adapted from Sá-Correia *et al.* (2002).

1.3.3. Glucose-1-phosphate nucleotidyltransferase mechanism

Recently, X-ray crystal structures of the eukaryotic UGPs from *Arabidopsis thaliana* (McCoy *et al.* 2007), *Leishmania major* (Roeben *et al.* 2006) and *Saccharomyces cerevisiae* (Steiner *et al.* 2007) have been determined but they are unrelated to the bacterial ones. Structural and kinetic similarities between cytidylyltransferases and thymidylyltransferases have pointed to a possible similar mechanism in spite of their low amino-acid sequence similarities, circa 15% (Koropatkin and Holden 2004). No structure of a bacterial uridylyltransferase is yet available to help clarify the catalytic mechanism but a similar mechanism may be expected.

Two conflicting hypothesis for this type of nucleotidyltransferase catalysis have been suggested. The first consists of a ping-pong bi-bi mechanism (Figure 1.13a) ⁽²⁾, whereby the first substrate binds to the protein forming a covalent intermediate, followed by nucleophilic attack of the second substrate on to the intermediate giving the final product. This mechanism was proposed for both cytidylyltransferase (Lindqvist *et al.* 1994) and thymidylyltransferase (Lindquist *et al.* 1993).

2 *Ping-pong bi-bi* mechanism is a multisubstrate reaction in which, for a two-substrate, two-product (i.e., bi-bi) system, an enzyme reacts with one substrate to form a product and a modified enzyme, the latter then reacting with a second substrate to form a second, final product, and regenerating the original enzyme. This type of mechanism is also commonly called *double displacement mechanism*. (CancerWEB, On-Line Medical Dictionary.)

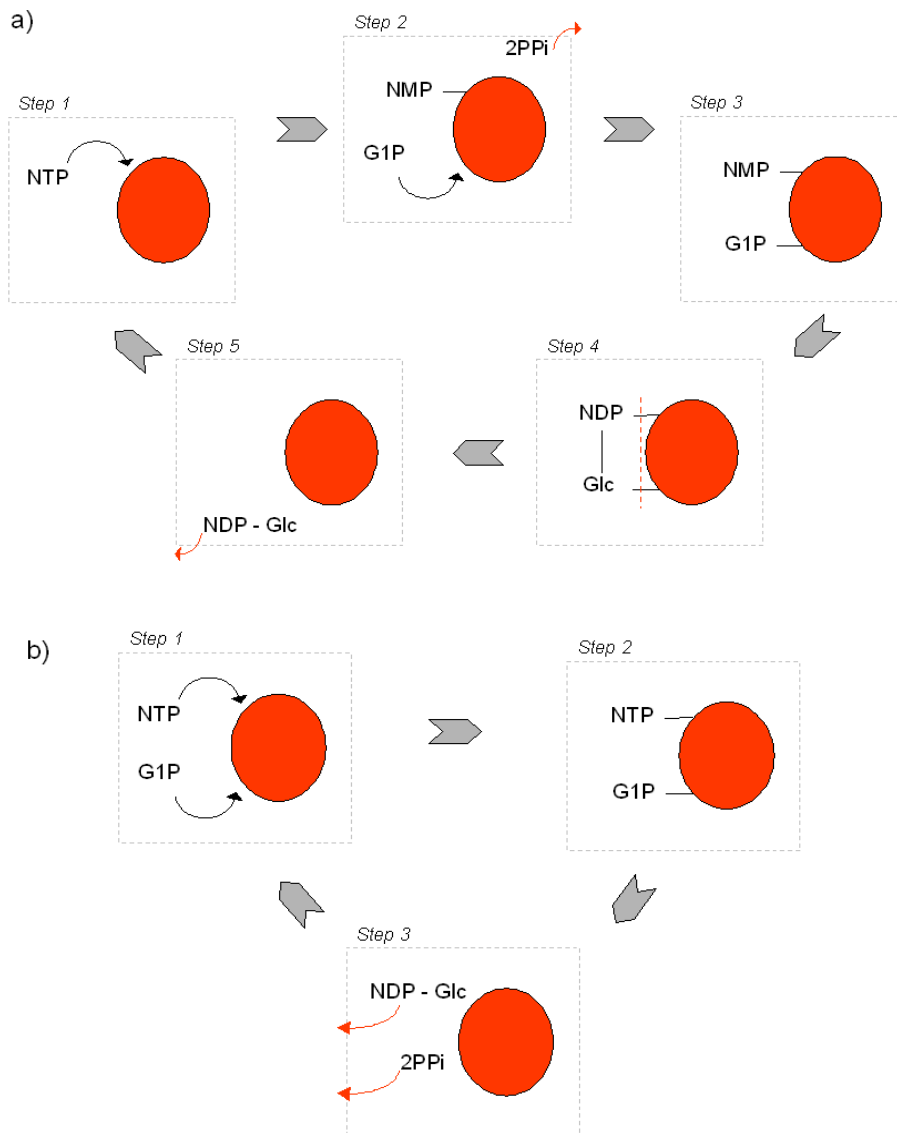


Figure 1.13. Proposed mechanisms for the glucose-1-phosphate nucleotidyltransferase family.

(a) The ping-pong bi-bi mechanism. This mechanism consists of nucleotide triphosphate (NTP) binding, followed by the transfer of a nucleotide monophosphate (NMP) moiety to the enzyme (red ellipsoid) with release of pyrophosphate (PPi). Only then would the NMP and glucose-1-phosphate (G1P) condense to form NDP-glucose. **(b) The sequential ordered bi-bi mechanism.** This mechanism consists of NTP and G1P binding to the protein followed by a direct SN2 attack of the phosphoryl oxygen of G1P to the alpha phosphate of NTP, which would displace beta and gamma phosphates. These two phosphate groups would leave as pyrophosphates, together with NDP-glucose, and diffuse away from the protein

The second consists of a sequential ordered bi-bi mechanism (Figure 1.13b) in which both substrates bind, react, and then diffuse away (Melo and Glaser 1965; Lindqvist *et al.* 1994). Detailed kinetic data (Paule and Preiss 1971; Barton *et al.* 2001; Koropatkin *et al.* 2005), evidence for a typical SN2 inverted geometry ⁽³⁾ about the alpha-phosphate upon attack by G1P (Sheu *et al.* 1979; Blankenfeldt, Asuncion *et al.* 2000; Barton *et al.* 2001) and isotope exchange data (Blankenfeldt, Asuncion *et al.* 2000) are in favour of a single displacement mechanism, such as the sequential ordered bi-bi mechanism, for the entire nucleotidyltransferase family. However, there still remain questions on the reasoning behind the different activities for the various nucleotides.

1.3.4. Glucose-1-phosphate uridylyltransferase vs glucose-1-phosphate thymidylyltransferase

The closest sequences in terms of overall similarity to the prokaryotic UDP-glucose uridylyltransferases, UGP, are the prokaryotic TDP-glucose thymidylyltransferases (EC 2.7.7.24, RmlAs), with sequence percentage identities, nonetheless, below 20%, followed by the prokaryotic UDP-N-acetylglucosamine pyrophosphorylases, with identities of circa 15%. Both UGP and RmlAs use G1P as the sugar substrate but different nucleotides, UTP and dTTP respectively, as the acceptor. However, in *S. elodea*, both proteins recognise both UTP and dTTP as substrates *in vitro* (Silva *et al.* 2005) although other UPGs have been reported for not being able to use dTTP as substrate *in vitro* (Bonofiglio *et al.* 2005).

UGP and RmlA also share the highly conserved sequence X-E-K-P (Silva *et al.* 2005) which is known to be responsible for G1P binding in glucose-1-phosphate NTP transferases (Thorson *et al.* 1994). The carboxyl group of

3 Further reading about SN2 inverted geometry see (Streitwiser *et al.* 1992).

glutamate binds to a hydroxyl group of G1P and the nitrogen from the side chain of lysine is responsible for stabilising the negative charge of the alpha-phosphate. This is not unusual since UTP and TTP are very similar, differing only by one CH₃ group (Figure 1.14). On the other hand, these proteins cannot use the nucleosides ATP, GTP or CTP as substrates. This is not unexpected since ATP and GTP have a purine type ring system and the active sites will need more space to be able to bind them. Furthermore, CTP has an amine group where UTP and TTP have a carbonyl oxygen (Figure 1.14c,d and e) at their aromatic rings. NTP to protein hydrogen bonds are made through this amine group (Koropatkin and Holden 2004; Koropatkin *et al.* 2005), in glucose-1-phosphate cytidylyltransferases, and through the corresponding carbonyl group, in glucose-1-phosphate thymidylyltransferases.

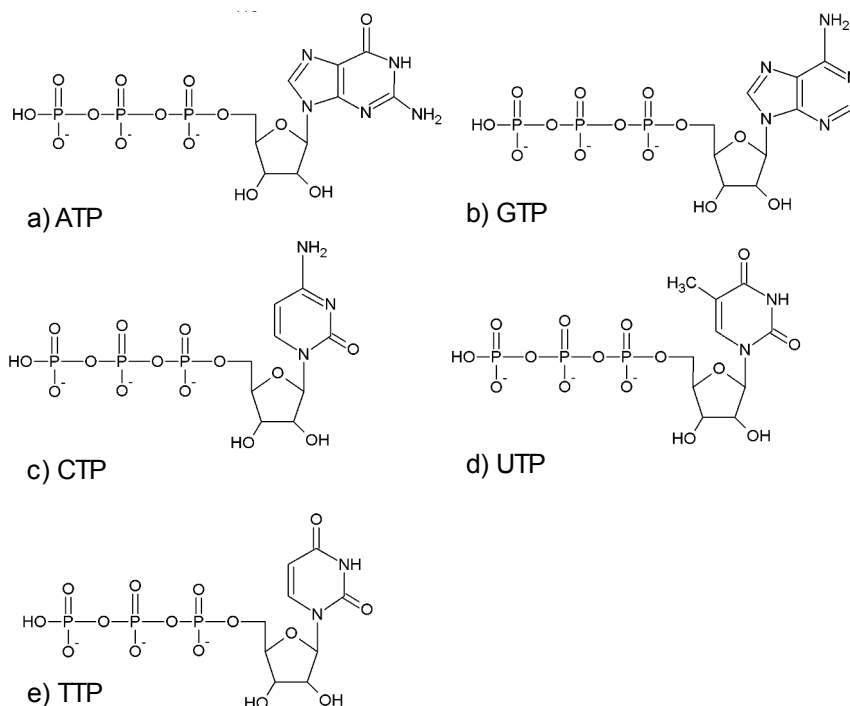


Figure 1.14. Nucleotides.

(a) adenosine triphosphate (ATP), (b) guanine triphosphate (GTP), (c) cytosine triphosphate (CTP), (d) uridine triphosphate (UTP) and (e) thymine triphosphate (TTP).

1.3.5. Glucose-1-phosphate thymidyltransferase (RmlA)

When the present studies started, no structures of eukaryotic or prokaryotic UGPs were available, although several RmlA structures bound to different substrates, products and inhibitors had already been published for *Salmonella enterica*, *Pseudomonas aeruginosa*, *Escherichia coli* and *Methanobacterium thermoautotrophicum* (Table 1.1).

Table 1.1. Structures of bacterial glucose-1-phosphate thymidyltransferase and their respective organism.

Bacteria	PDB code
<i>S. enterica</i>	1IIN, 1IIM, 1MP3, 1MP4 and 1MP5
<i>P. aeruginosa</i>	1G0R, 1G23, 1G3L, 1G1L and 1FXO
<i>E. coli</i>	1H5R and 1MC3
<i>M. thermoautotrophicum</i>	1LVW

These structures presented high structural similarity, revealed their substrate binding sites, defined the arrangement of the substrates and the products in the active site, and showed that dTTP binds to the enzyme in a strained conformation. Briefly, the RmlA molecule is classified by CATH (see Appendices) as an alpha/beta complex fold, built from several β -strand/ α -helix units that assemble into a central seven-stranded β -sheet, with six parallel and one antiparallel strands and helices lining at both sides of the sheet. The C-terminus is capped with a bundle of three antiparallel α -helices, each being two turns long. Two additional two stranded β -sheets are formed, one along the edge of the central sheet, and the other on top, near the C-terminal ends of the central sheet. This enzyme crystallised, in all structures, as a tetramer and other biochemical techniques have confirmed that indeed

this enzyme is a tetramer in solution.

The studies reported here were aimed to over express and purify both *S. elodea* UGP and RmlA in quantities suitable for X-ray protein crystallography studies, and to determine both the protein structures.

2. Reduced hybrid cluster proteins (HCP) from *Desulfovibrio desulfuricans* ATCC 27774 and *Desulfovibrio vulgaris* (Hildenborough) X-ray structures at high resolution using synchrotron radiation

(Based on a published refereed paper. Reference: Aragão D, Macedo S, Mitchell EP, Romao CV, Liu MY, Frazao C, Saraiva LM, Xavier AV, LeGall J, van Dongen WM, Hagen WR, Teixeira M, Carrondo MA, Lindley P. J Biol Inorg Chem. 2003 May;8(5):540-8)

2.1. Abstract

The hybrid cluster proteins from the sulphate reducing bacteria, *D. desulfuricans* ATCC 27774 and *D. vulgaris* strain Hildenborough have been isolated and crystallised anaerobically. In each case, the protein has been reduced with dithionite and the crystal structure of the reduced form elucidated using X-ray synchrotron radiation techniques at 1.25 Å and 1.55 Å resolution for *D. desulfuricans* and *D. vulgaris* respectively. Although the overall structures of the proteins are unchanged upon reduction, there are significant changes at the hybrid cluster centres. These include significant movements in the position of the iron atom linked to the persulphide moiety in the oxidised as-isolated (⁴) proteins and the sulphur atom of the persulphide itself. The nature of these changes is described and the implications with respect to the function of hybrid cluster proteins are discussed.

4 The term as-isolated was used extensively in the published article where this chapter was based on. As-isolated should be understood as anaerobically purified in contrast with the structures purified in the presence of oxygen.

2.2. Introduction

Genes encoding hybrid cluster proteins (HCP) have been identified in a large number of prokaryotes, including archaea and bacteria, both strict anaerobic and facultative anaerobic, as well as in eukaryotic protozoa. No specific function has yet been assigned to this family of proteins, although its involvement in the biological nitrogen cycle (van den Berg *et al.* 2000; Beliaev *et al.* 2002; Wolfe *et al.* 2002), and in protection against oxidative stress (Briolat and Reysset 2002) has been proposed. The similarity of the hybrid cluster in HCPs to the nickel containing cluster present in the catalytic domain of the carbon monoxide dehydrogenase from *Carboxydotherrmus hydrogenoformans* (Dobbek *et al.* 2001) has also been noted.

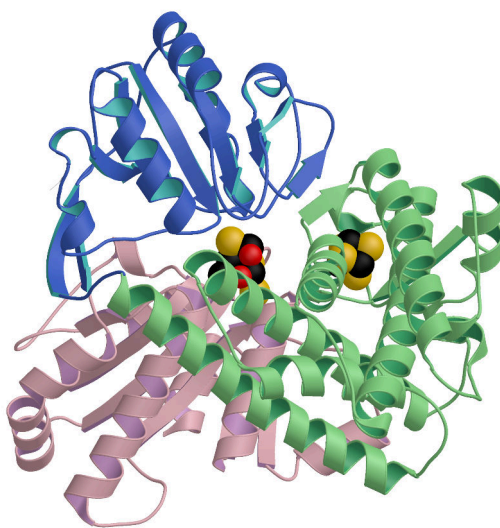


Figure 2.1. Ribbon representation of *D. vulgaris* HCP.

The cubane cluster and hybrid cluster are shown with irons in black, sulphurs in yellow and oxygens in red. The molecule is divided into three domains which are coloured in green (domain 1), blue (domain 2) and light pink (domain 3).

In previous structural studies (Arendsen *et al.* 1998; Cooper *et al.* 2000; Macedo *et al.* 2002) it has been shown that the HCPs from the sulphate reducing bacteria, *D. desulfuricans* and *D. vulgaris*, in the as-isolated oxidised form, contain an unusual iron-sulphur cluster, whose structure appears to be independent of whether the proteins are isolated aerobically or anaerobically. This hybrid cluster is located at the interface of the three domains comprising the HCP and some 11-12 Å distant from a [4Fe-4S]^{2+/1+} cubane cluster which is located near the outside of the molecule at the N-terminus in domain 1 (Figure 2.1). In the as-isolated protein, the hybrid cluster can be considered as being derived from a cube, but with an iron atom, Fe7, at one corner peeled back – see Figure 2.2(a). It is bound to the protein by seven amino acid residues lying at the interfaces of the three domains, so that:

- Fe5 and Fe6, linked to the protein by Cys427 (Cys 434 in *D. vulgaris*) and Cys308 (Cys 312 in *D. vulgaris*) respectively, together with S5 and S6, form a [2Fe-2S] moiety that constitutes a face of the cube.
- Fe8 lies at another corner of the cube, linked to the protein by Glu487 (Glu494), but is also bound to S6 of the [2Fe-2S] moiety and a persulphide moiety at Cys399 (Cys406).
- Fe7, linked to the protein by His240 (His 244), Glu264 (Glu268) and Cys452 (Cys 459) can be considered as the fourth corner of the cube peeled away so that it lies close to the diagonal of the face formed by Fe5, S6 and Fe8 that bisects the two iron atoms.

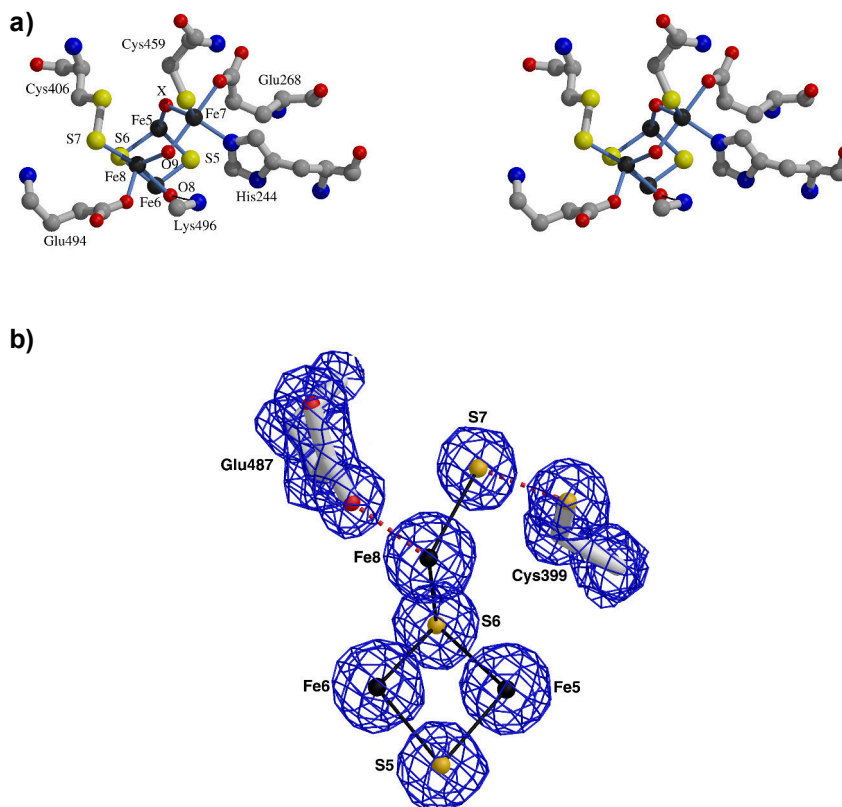


Figure 2.2. *As-isolated oxidised hybrid cluster in HCPs.*

a) A model representation, in stereo view, of the hybrid cluster in the as-isolated oxidised form of *D. vulgaris* (see chapter 4 for protein isolation and structure determination: PDB ID 1W9M) showing the bridging oxygen atoms O8, bridging Fe6 and Fe8, and O9, bridging Fe7 and Fe8. The moiety X is also represented as an oxygen atom bridging Fe5 and Fe7. Residues Cys434 and Cys312 bound to Fe5 and Fe6 respectively have been omitted for clarity.

b) Electron density at the hybrid cluster in the as-isolated oxidised form of *D. desulfuricans* [(Macedo *et al.* 2002): PDB ID 1GNL] showing the persulphide moiety at Fe8. The contour level is 1.0 rms in SigmaA weighted $2|F_o| - |F_c|$ Fourier syntheses. This is a different orientation to (a) and Fe7 and its environment have been omitted for clarity.

The Fe-Fe separations range between 2.8 Å and 5.1 Å and the hybrid cluster contains both μ -S and μ -O bridges between pairs of iron atoms. Thus, O8 and O9 bridge Fe6 and Fe8 and Fe7 and Fe8 respectively. A further moiety, X, appears to bridge Fe5 and Fe7 and in this case the environments of the iron atoms can be described as distorted tetrahedral for Fe5 and Fe6, and distorted trigonal bipyramidal for Fe7. The coordination of Fe8 is distorted trigonal bipyramidal, but if X is also considered, it becomes distorted octahedral. In the oxidised forms of the HCP from *D. desulfuricans* and *D. vulgaris*, X has been assigned as an oxygen atom on the basis of Resonance Raman studies with two possible positions in the *D. vulgaris* structure.

In the crystalline state it is possible to reduce the HCP proteins with dithionite without loss of diffraction quality. This paper reports the X-ray crystal structures of the reduced HCPs from *D. desulfuricans* and *D. vulgaris*. Unprecedented large structural changes in the hybrid cluster are observed upon reduction and rationalised with the enzyme's putative role as a reductase.

2.3. Materials and methods

2.3.1. Crystallisation

The HCP from *D. desulfuricans* was anaerobically purified as previously described (Macedo *et al.* 2002); the protein from *D. vulgaris* was obtained from a *D. vulgaris* clone overproducing HCP from the recombinant plasmid pJSP104 (Stokkermans *et al.* 1992) and purified as for the *D. desulfuricans* protein (for further details on *D. vulgaris* HCP isolation see chapter 4). Crystals of both HCPs were grown in a nitrogen environment inside a glove box system ($pO_2 < 1$ ppm) at different temperatures by the sitting drop vapour-diffusion method. For *D. desulfuricans* crystallisation trials were set up at 279K, 283K and 293K, yielding needle-shaped pale yellow crystals in

PEG 4000 (concentration range 25-35%) and MES (pH range 6.0-6.7). The best crystals grew at 279K in 35% PEG 4000 and 0.1M MES buffer at pH 6.5. Protein solution, 1.5 μl , (12.6 mg.ml^{-1} in 20mM Tris-HCl pH 7.6 and 0.2M NaCl) and an equal amount of precipitant solution were allowed to equilibrate against 500 μl of the latter in the reservoir. Crystals take around 1 month to grow to a maximum size of approximately 0.2 mm x 0.1 mm x 0.08 mm.

For the *D. vulgaris* enzyme crystallisation trials at 277K yielded rectangular-shaped laminar pale yellow crystals in PEG 4000 (concentration range 15% to 30%) and MES (pH range from 6.0-6.5). The best crystals were obtained in 25% PEG 4000 and 0.1M MES buffer at pH 6.5. A protein solution, 2.0 μl (25 mg.ml^{-1} in 20 mM Tris-HCl at pH 7.6 and 0.05 M NaCl) and an equal amount of precipitant solution were allowed to equilibrate against 500 μl of the latter in the reservoir. Crystals take around 6 weeks to grow to a maximum size of 0.2 mm x 0.2 mm x 0.05 mm.

2.3.2. Reduction of the crystals

Crystals obtained as previously described were reduced by soaking with sodium dithionite. For *D. desulfuricans* a solution of 0.22 M $\text{Na}_2\text{S}_2\text{O}_4$ in freshly prepared mother liquor was added to drops containing crystals, resulting in an approximate final concentration of sodium dithionite in the drop of 0.18 M. For *D. vulgaris* the same procedure was applied, but with a higher concentration of reducing agent, resulting in an approximate final concentration of sodium dithionite in the drop of 0.30 M. In both cases the crystals (previously cryo-protected by transfer to a solution obtained by adding 15% of glycerol to the mother liquor) were flash-cooled in liquid nitrogen after 2 hours soaking and stored. Addition of sodium dithionite to the drop and flash freezing were carried out inside the glove box system and

all solutions used were de-aired under vacuum beforehand.

For *D. desulfuricans* the oxidation state of the crystals was tested by comparison of the UV-visible spectra measured from crystals in a microspectrophotometer (Bourgeois *et al.* 2002) before and after reduction with sodium dithionite.

2.3.3. Assay for hydroxylamine reductase activity

Assays were performed under anaerobic conditions, in a Coy model a-2463 anaerobic chamber, using a Shimadzu UV-1203 spectrophotometer equipped with a stirring cell. A solution of hydroxylamine (1 M) adjusted to pH 7.0 was used in all assays. The assay mixture (1.2 ml) contained HEPES buffer (100 mM, pH 7.0), EDTA (10 μ M), and methyl viologen (0.2 mM), was reduced with sodium dithionite until an absorbance at 600 nm ($\epsilon_{600} = 13 \text{ mM}^{-1}\text{cm}^{-1}$) close to 1 was obtained. Hydroxylamine was added (final concentration 100 mM) to the assay solution, and the reaction was started by the addition of HCP; no change in the activity was detected by changing the order of addition of the hydroxylamine or protein solutions. The pH-dependence assays for the NH_2OH reduction were also determined at pH 7.5 and 8.8, using Tris-HCl buffer (100 mM) adjusted to the desired pH. Activities are expressed as micromoles of NH_2OH reduced per minute per milligram of protein, taking into account that for the reduction of hydroxylamine, two equivalents of reduced methyl viologen are consumed.

2.3.4. Immuno-blotting

Immuno-blotting with antibodies against *D. vulgaris* HCP (van den Berg *et al.* 2000) was used to monitor the expression of the *D. desulfuricans* HCP grown using either nitrate or sulphate as terminal electron acceptors.

2.3.5. X-ray diffraction data collection

a) *The reduced form of the HCP from D. desulfuricans*

X-ray data were collected from a single cryo-cooled crystal at 100 K on beamline ID14-4 at the ESRF synchrotron radiation source using a wavelength of 0.939 Å. The ADSC Q4R CCD detector was positioned at a distance of 100 mm from the sample and the beam size used was 30 µm x 30 µm. The effects of radiation damage were assessed using the individual frame B-factors referred to the first image and comparing images taken from identical ϕ rotation ranges at the start and end of the data collection; there appeared to be no significant radiation damage.

Data were processed using DENZO (Otwinowski and Minor 1997) and the indexed intensities were merged and converted to structure factors using the programs ROTAPREP, SCALA and TRUNCATE in the CCP4 program suite (CCP4 1994). Details of data collection are given in Table 2.1.

The asymmetric unit accommodates two molecules of molecular weight 60 kDa, corresponding to a V_m of 2.03 Å³ Da⁻¹ (Matthews 1968), and a solvent content of some 39 %.

Table 2.1. Data collection statistics

Data collection statistics	<i>D. desulfuricans</i>	<i>D. vulgaris</i>
Beam line	ID14-4 : ESRF	ID14-2 : ESRF
Space group	P1	P2 ₁ 2 ₁ 2 ₁
Number of molecules in <i>a.u.</i>	2	1
Resolution (Å)	1.25	1.55
Highest res. shell (Å)	1.28 - 1.25	1.59-1.55
Wavelength (Å)	0.939	0.933
Cell dimensions		
a (Å)	57.75	64.05
b (Å)	61.99	67.46
c (Å)	72.80	134.95
α (°)	82.72	90.00
β (°)	73.74	90.00
γ (°)	87.44	90.00
Total no. of hkl	448 850	262 651
No. of batches & rotation angle (°)	360 / 0.5	425 / 0.2
Number of unique hkl	246 621	81856
Average multiplicity §	1.8 (1.8)	3.2 (3.8)
R _{merge} ^{5 §}	0.040 (0.098)	0.071 (0.254)
Average I/σ(I) §	7.8 (3.6)	13.2 (3.6)
% completeness §	92.7 (78.7)	95.7 (93.5)
Average B (Wilson) Å ²	7.1	19.3

(§ - Values in parentheses refer to highest resolution shell)

b) The reduced form of the HCP from *D. vulgaris*

X-ray data were collected from a single cryo-cooled crystal at 100K on beamline ID14-2 at the ESRF, Grenoble, using a wavelength of 0.933 Å and an ADSC Q4R CCD detector placed at a distance of 120 mm from the sample; the beam size used was 25 μm x 25 μm. Radiation damage effects were assessed in a similar manner to that used for *D. desulfuricans* and again there appeared to be no significant radiation damage. The diffraction data were also processed as for *D. desulfuricans* and details are given in Table 2.1.

The asymmetric unit accommodates only one molecule of *D. vulgaris*, corresponding to a V_m of 2.04 Å³ Da⁻¹ (Matthews 1968), and a solvent

5 R_{merge} is $\Sigma |I - \langle I \rangle| / \Sigma \langle I \rangle$;

content of some 40%.

2.3.6. Structure Solution

a) *The reduced form of the HCP from D. desulfuricans : Ab Initio phasing*

The *ab initio* phasing program ACORN (Foadi *et al.* 2000) was used to generate preliminary experimental phases. The *E*-values (normalised structure factors) required by ACORN were generated from the SCALA output using the CCP4 (CCP4 1994) program ECALC. No reflections were excluded in the calculation of *E*-values. In ACORN the initial phases are generated from a fragment, such as a set of heavy atoms, with a previously determined position. It was assumed that the [4Fe-4S] cubane clusters, one in each monomer in the asymmetric unit, would be geometrically identical and in the same position in the unit cell in the reduced and oxidised forms of the HCP protein. Increasing numbers of iron atoms from the two cubane clusters in the asymmetric unit of the as-isolated *D. desulfuricans* anaerobic structure at 1.25 Å resolution (PDB code 1GNL) were used as a starting fragment until successful phasing of the structure was accomplished. In all cases default settings for ACORN were used. For this particular protein and at this resolution, a clearly interpretable electron density map was obtained using all of the eight iron atoms from the cubane clusters of both monomers. However, a second run with only six Fe atoms in the cluster also gave a positive result, albeit with a significant increase in computational time. In both cases, only the cubane cluster iron atoms were used in order to minimise any possible bias in the resulting electron density map; no assumptions were made concerning the nature of the hybrid clusters.

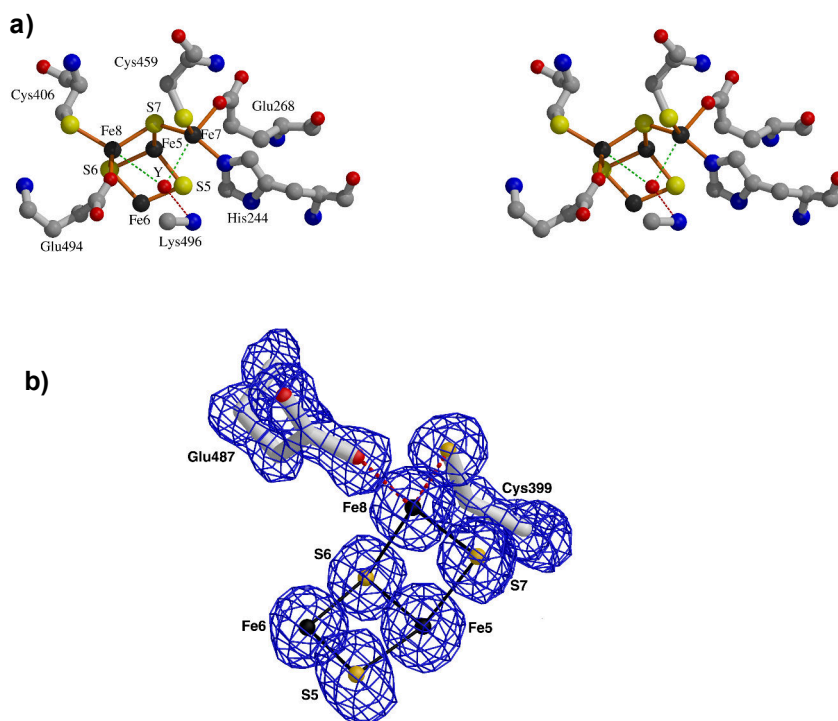


Figure 2.3. *Reduced hybrid cluster in HCPs.*

a) A model representation, in stereo view, of the hybrid cluster in the reduced form of *D. vulgaris* (PDB ID 1OA1) showing the Y moiety between the positions previously occupied by O8 and O9. Residues Cys434 and Cys312, bound to Fe5 and Fe6 respectively, have been omitted for clarity.

b) Electron density at the reduced hybrid cluster in *D. desulfuricans* (PDB ID 1OA0). The contour level is 1.0 rms in SigmaA weighted $2|F_o| - |F_c|$ Fourier syntheses. The SG atom of Cys399 is directly bound to Fe8 and S7 has moved towards the X position in the as-isolated enzyme, and is bound to Fe5, Fe7 and Fe8. This is a different orientation to (a) and Fe7 and its environment have been omitted for clarity.

The initial Fourier map was of high quality and clearly showed electron density consistent with the two molecules of the *D. desulfuricans* protein in the asymmetric unit. In each molecule nearly all the main chain and side chain residues could be positioned directly. However, in the hybrid cluster large positional changes were evident for both Fe8 and the persulphide S

atom to which it is attached in the oxidised form of the protein. Figures 2.2(b) and 2.3(b) compare the electron density at the hybrid cluster in the as-isolated and reduced forms of the HCP from *D. desulfuricans*. The [2Fe-2S] moiety was also slightly moved with respect to the oxidised as-isolated protein and there were no indications of any bridging oxygen atoms in the hybrid cluster.

b) The reduced form of the HCP from D. vulgaris

Since the resolution, 1.55 Å, was not sufficient for *ab initio* phasing, an alternative molecular replacement procedure was used as described by Perrakis *et al.* (2001). The structure of the anaerobically as-isolated protein at 1.35 Å resolution (PDBID 1W9M: see chapter 4) was used to generate phase information and the resulting data interpreted by ARP/wARP (Lamzin *et al.* 1999).

2.3.7. Structure refinement

a) The reduced form of HCP from D. desulfuricans at 1.25 Å resolution.

Refinement was performed using the maximum likelihood functions implemented in REFMAC (Murshudov *et al.* 1997), starting with the model determined for *D. desulfuricans* at 1.25 Å resolution (Macedo *et al.* 2002), but with new positions for Fe8 and S7 and no bridging oxygen atoms in the hybrid cluster. At this stage no solvent molecules were included. Rounds of conjugate gradient sparse matrix refinement with bulk solvent modelling according to the Babinet principal (Tronrud 1996) were alternated with model building using the O program (Jones *et al.* 1991) in combination with SigmaA weighted $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ electron density maps (Read 1986).

For the hybrid clusters, attempts were made to refine the Fe8 atoms as sulphur and the S7 as oxygen, but these resulted in anomalously small temperature coefficients, indicating that the assignment of Fe and S respectively was correct. The assignment was also verified using anomalous dispersion techniques. The geometries of the cluster atoms were unrestrained throughout the refinement, and in the final stages all the Fe and S atoms (inorganic S and cysteine S) associated with the two clusters in each molecule were allowed to refine anisotropically. After the first round of refinement, solvent molecules were added to the model based on standard geometrical and chemical restraints; they were subsequently deleted if their isotropic thermal coefficients exceeded $50 \text{ \AA}^2$ or they were not visible at the 0.8 rms level in SigmaA weighted $|F_o| - |F_c|$ electron density maps. In several regions, solvent molecules were close together indicating either partially occupied sites or possible stretches of polyethylene glycol density. Where solvent peaks occurred within $2.0 \text{ \AA}$ of one another, they were treated as partially occupied sites with occupancies (not refined) roughly proportional to the peak volumes. In addition to the solvent peaks assigned to water, there were two regions of density that were clearly different to the normal solvent structure. The first of these lay in the vicinity of the cubane cluster in molecule A. It was unclear what it represented but, for example, the density could accommodate a dithionite anion (DTN). This was considered reasonable since dithionite was used in the protein reduction process. If this were the case, one oxygen atom of the anion would have a closest approach of about $7.0 \text{ \AA}$ to Fe3 in the cubane cluster; the remaining oxygens of the anion would form hydrogen bonds with neighbouring solvent molecules. However, attempts to refine this moiety indicated that if dithionite was present then the occupancy was only around 0.5. The density did not appear to be present near to the cubane cluster in molecule B. The second region of density lay between molecules A and B in the asymmetric

unit. Although the density was not completely contiguous, it could clearly be ascribed to a MES molecule; 0.1M MES buffer was used in the crystallisation process. One of the sulphite oxygen atoms interacts with the side chain of Lys313 in molecule B. A second interacts with the main chain amino atom of Asp314, also in molecule B, and with a neighbouring solvent molecule, whilst the third interacts with a solvent molecule only. The morpholino moiety fits into a pocket in molecule A in the vicinity of Lys434.

For the protein, hydrogen atoms were included as 'riding' atoms. Details of the final refinement statistics are given in Table 2.2. Several residues were modelled in more than one conformation. A number of the surface residue side chains, particularly lysines, had no visible electron density after the CG atom, but as far as possible these were modelled in geometrically feasible positions. Details are shown in Table 2.2.

Table 2.2. Refinement statistics for reduced HCPs from *D. desulfuricans* and *D. vulgaris* at resolutions of 1.25 Å and 1.55 Å respectively.

	<i>D. desulfuricans</i>	<i>D. vulgaris</i>
PDB ID	1OA0	1OA1
Amino acids	2 x 544	1 x 553
Protein atoms	4176 A : 4181 B	4252
Solvent atoms	2044	866
No. in dual conformations	173	-
Other atoms	2 x 16 + 1xDTN + 1xMES	1x 16 + 2x GOL
Resolution Limits (Å)	29.49 – 1.25	19.50 – 1.55
Working set R (Observations)	0.131 (240495)	0.131 (73710)
Test set R (Observations)	0.148 (6123)	0.153 (4074)
Highest resolution shell R (Obs) Free R (Obs)	0.150 (14346) 0.166 (365)	0.168 (5497) 0.199 (295)
Cruickshank's DPI (Å) (CRUICKSHANK 1999)	0.038	0.072
Average B _{iso} (Å ²)		
All atoms	11.9	15.1
Protein atoms	8.1 A : 8.6 B	12.4
Main chain	7.3 A : 7.8 B	11.3
Side chain	8.8 A : 9.5 B	13.5
Solvent	25.3	29.2
	<i>D. desulfuricans</i>	<i>D. vulgaris</i>
PDB ID	1OA0	1OA1
Amino acids	2 x 544	1 x 553
Protein atoms	4176 A : 4181 B	4252
Solvent atoms	2044	866
No. in dual conformations	173	-
Other atoms	2 x 16 + 1xDTN + 1xMES	1x 16 + 2x GOL
Resolution Limits (Å)	29.49 – 1.25	19.50 – 1.55
Working set R (Observations)	0.131 (240495)	0.131 (73710)
Test set R (Observations)	0.148 (6123)	0.153 (4074)
Highest resolution shell R (Obs) Free R (Obs)	0.150 (14346) 0.166 (365)	0.168 (5497) 0.199 (295)

Table 2.2 (cont.) Refinement statistics for reduced HCPs from *D. desulfuricans* and *D. vulgaris* at resolutions of 1.25 Å and 1.55 Å respectively.

	<i>D. desulfuricans</i>	<i>D. vulgaris</i>
Residues not found	G544 A & B	-
Residues in alternative conformations		D109, I151, R171, T173, E335, K338, E390, K411, N443 GOL1
Residues not complete		- -
Distance deviations: [†]		
Bond distances (Å)	0.008	0.011
Bond angles (°)	1.245	1.292
Planar groups (Å)	0.006	0.006
Chiral volume deviation (Å ³)	0.138	0.090

[†] rms deviations from standard values.

b) The reduced form of the HCP from *D. vulgaris* at 1.55 Å resolution.

Refinement was also performed using the maximum likelihood function implemented in REFMAC, starting with the model produced by ARP/wARP, but inserting atoms one at a time for the hybrid cluster; no solvent structure was included in the initial stages. Rounds of conjugate gradient sparse matrix refinement with bulk solvent modelling according to the Babinet principal were alternated with model building using XtalView (McRee 1992) in combination with SigmaA weighted electron density syntheses.

In a similar manner to the *D. desulfuricans* protein attempts were made to refine the Fe8 atom in the hybrid cluster as sulphur and the S7 atom as oxygen, but again these resulted in anomalously small thermal coefficients, suggesting that the original assignments were correct. Density in the solvent region, although not completely contiguous, could be clearly ascribed to two glycerol molecules (GOL); 25% glycerol buffer was used as a cryo-protector. Further details of the refinement are given in Tables 2.2

2.3.8. Quality of the refined structures

The quality of the refined structure of *D. desulfuricans* is high with final values for R and R_{free} of 13.1 % and 14.8 %, respectively and an overall G factor of 0.21 (Laskowski *et al.* 1993), Table 2.3. As in the oxidised structure two residues in each monomer, Ser263 and Asn299 lie in the disallowed region of the Ramachandran plot (Ramakrishnan and Ramachandran 1965). In all cases these residues lie in well defined density at bends between helices and strands.

Table 2.3. *Quality of reduced HCP structures.*

	<i>D. desulfuricans</i>	<i>D. vulgaris</i>
Overall G factor	0.21	0.20
Ramachandran analysis % (No.)		
Favourable	93.7 (888)	94.1 (450)
Additional	5.6 (53)	5.6 (27)
Generous	0.3 (3)	0 (0)
Disallowed	0.4 (4)	0.2 (1)
No. of glycine residues	2 x 49	50
No. of proline residues	2 x 18	23

The quality of the refined structure of *D. vulgaris* is also high with final R and R_{free} values of 13.1% and 15.6% respectively and an overall G factor of 0.19. Residue Asn303 (equivalent to Asn 299 in *D. desulfuricans*) lies in the disallowed region of the Ramachandran plot, but the density is very well defined.

Information for the two data sets and the associated structures has been deposited in the Protein Data Bank (Berman *et al.* 2000) with deposition codes 1OA0 and 1OA1 for *D. desulfuricans* and *D. vulgaris* respectively.

2.4. Results and discussion

The overall architecture of the reduced forms of the HCP proteins prepared anaerobically from *D. desulfuricans* and *D. vulgaris* is almost identical to that of the as-isolated oxidised forms, prepared both anaerobically [(Macedo *et al.* 2002) (*D. desulfuricans*), chapter 4 (*D. vulgaris*)], and aerobically [(Cooper *et al.* 2000) (*D. vulgaris*)]. Each molecule contains three domains and two iron-sulphur clusters; one cluster is a conventional cubane cluster situated at the N-terminus of domain 1, and the hybrid cluster is located at the interface of the three domains. It is in this hybrid cluster that the major differences between the reduced and oxidised forms of the protein occur. Throughout this discussion the atom and residue numbering scheme appropriate to *D. desulfuricans* will be utilized. Geometrical parameters cited for *D. desulfuricans* are the average values for the A and B molecules in the asymmetric unit.

Thus, in the hybrid cluster of the reduced protein the Fe7 atom linked to the protein through residues His240, Glu264 and Cys452 is virtually unchanged in position, the [2Fe-2S] moiety involving Fe5 and Fe6, and linked to the protein through cysteine residues 427 and 308 respectively is slightly tilted, but Fe8 and S7 of the persulphide moiety at Cys399, undergo large movements in position and coordination. Concomitantly, residues Cys399 and Glu487, attached to Fe8, also undergo significant positional changes. Figures 2.2(a) and 2.3(a) show the arrangement of the atoms in the hybrid cluster for the as-isolated and reduced forms of the HCP from *D. vulgaris*. Figure 2.4 highlights the almost identical changes for the HCP from *D. desulfuricans* by the superposition of the as-isolated and reduced forms of the hybrid cluster.

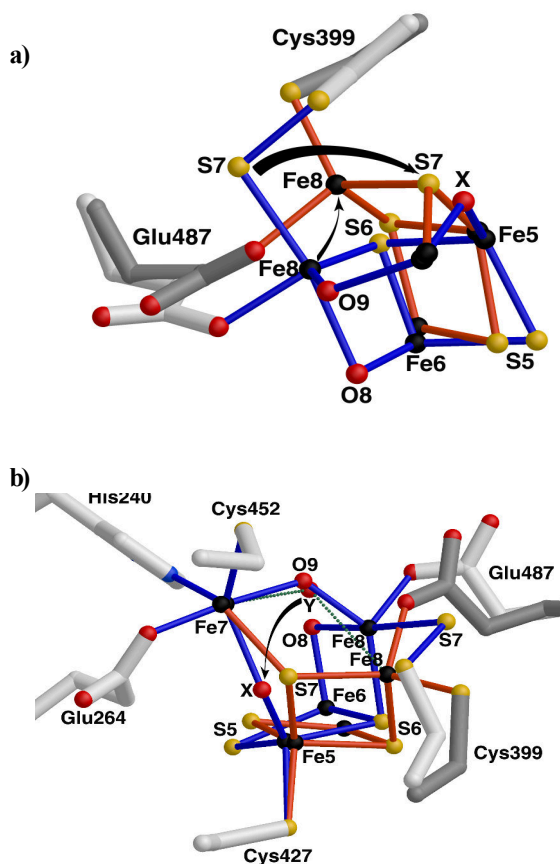


Figure 2.4. Superposition of the as-isolated oxidised and reduced clusters in *D. desulfuricans*.

The as-isolated oxidised form is in blue and the reduced form in red. The superposition was achieved using the LSQ option in the O program suite and rotating and translating the whole of the molecule from the asymmetric unit of *D. vulgaris* onto molecule A in the asymmetric unit of *D. desulfuricans*.

(a) A view from the top of the [2Fe-2S] moiety comprising Fe5 and Fe6. For clarity residues Cys427, (bound to Fe5), Cys308 (Fe6), and His240, Glu264 and Cys452 (Fe7) have been omitted, as well as the Y site in the reduced protein.

(b) A view almost parallel to the plane defined by the [2Fe-2S] moiety. Residue Cys308 (bound to Fe6) has been omitted for clarity.

The persulphide S atom, S7, moves very close to the position of atom X, defined as an oxygen atom or OH⁻ anion in the as-isolated form of the

protein, and appears to replace it. Thus, S7 forms the second side of a cube in conjunction with atoms Fe5, Fe8 and S6 and is only 2.4 Å from Fe7. The movement of S7 in *D. desulfuricans* is 3.6 Å and the SG of Cys399 now forms a direct bond to Fe8. In turn Fe8 has moved by some 2.1 Å. The geometries of the iron atoms in the reduced form of the hybrid cluster can now be described as tetrahedral for Fe5, Fe7 and Fe8, and trigonal planar for Fe6 in sharp contrast to those observed in the as-isolated oxidised cluster. Table 2.4 compares Fe-Fe, Fe-S and other important distances in the as-isolated, oxidised and reduced forms of the HCP proteins.

The hybrid cluster in the reduced form of the protein also lacks the μ -bridging oxygen atoms O8 and O9, between Fe6 and Fe8 and between Fe7 and Fe8 respectively, present in the as-isolated form. However, there is a peak in the electron density of the reduced protein, Y, not present in the as-isolated form and roughly in between the positions formerly occupied by O8 and O9, as shown in Figure 2.5. Y has been refined as a single oxygen atom (OH⁻), but the electron density at the 1 rms level is not spherical and could possibly accommodate a di-atomic (in a similar manner to X in the as-isolated protein) or even tri-atomic species. Y is some 3.1 Å and 3.4 Å from Fe7 and Fe8 respectively, too long for a direct bond. It is, however, well within hydrogen bonding range of the NZ of Lys489 (2.9 Å) and the ND2 of Asn307 (3.1 Å); both of these residues are conserved in most of the known HCP sequences.

It is likely that the significant structural changes between the aerobically, anaerobically and reduced forms of the HCP proteins reflect their function, which nevertheless remains to be established. In the cases of *E. coli* and *Morganella morganii*, two facultative anaerobes, expression of HCP was only observed when these bacteria were grown in nitrate or nitrite as terminal electron acceptors. This suggested a role for HCPs in the

metabolism of nitrogen oxides in these organisms (van den Berg *et al.* 2000). For another facultative anaerobe, *S. oneidensis*, an increase in the expression of HCP was also observed when the cells were grown in nitrate as electron acceptor, as compared with the expression in aerobically grown cells (Beliaev *et al.* 2002). However, disruption of the HCP gene in either *E. coli* or *M. morganii* had no effect on anaerobic cell growth, in the presence of nitrate or nitrite (minimal or rich media) (van den Berg *et al.* 2000). In the case of *D. desulfuricans* ATCC 27774, using antibodies against *D. vulgaris*, HCP, immuno-blotting of cell showed only a small increase in the expression of *D. desulfuricans* HCP when the cells were grown in the presence of sulphate/nitrate mixes (Figure 2.6, lanes 2-4). In addition, for both *D. vulgaris* and *D. desulfuricans*, HCP is still expressed in significant amounts when the cells are grown using sulphate as the only terminal electron acceptor (Figure 2.6, lane 1)

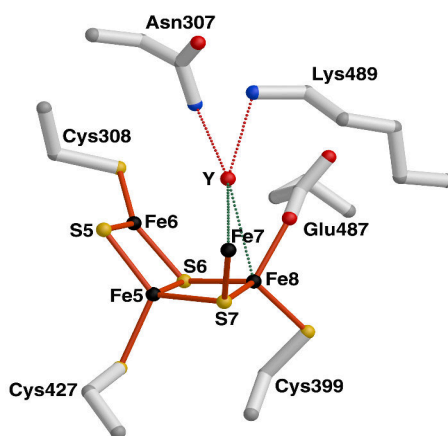


Figure 2.5. The environment of site Y in the reduced form of the hybrid cluster in *D. desulfuricans*.

Y has been modelled as an oxygen/hydroxide and, in this case, can form hydrogen bonds with the side chains of Lys489 and Asn307. The previous position for X, in the as-isolated oxidised form, is now occupied by a well refined sulphur atom, S7. Residues His240, Glu264 and Cys452 bound to Fe7 have been omitted for clarity.

Table 2.4. Geometry (bond distances in Å) of the cubane and hybrid clusters in the as-isolated oxidised and reduced HCP proteins from *D. desulfuricans* and *D. vulgaris*.

Residue numbering refers to the Dd protein. The values for the as-isolated Dd protein are from reference (Macedo *et al.* 2002); values for the as-isolated Dv protein are unpublished results at a resolution of 1.35 Å (see chapter 4). We estimate using the Cruickshank (Cruickshank 1999) method that the estimated uncertainties in the bond lengths involving Fe and S are less than 0.03 Å and 0.02 Å for Dd and Dv respectively. The unpublished structure determination for *D. vulgaris* as-isolated (PDID 1W9M) is described in chapter 4.

a) Cubane cluster

	<i>D. desulfuricans</i> (A) as-isol.	<i>D. desulfuricans</i> (B) as-isol.	<i>D. desulfuricans</i> (A) reduced	<i>D. desulfuricans</i> (B) reduced	<i>D. vulgaris</i> as-isol.	<i>D. vulgaris</i> reduced
PDB ID	1GNL		1OA0		1W9M	1OA1
Fe1 - Fe2	2.73	2.73	2.74	2.74	2.74	2.71
- Fe3	2.69	2.70	2.71	2.71	2.70	2.70
- Fe4	2.77	2.76	2.79	2.79	2.78	2.79
- S1	2.30	2.31	2.30	2.32	2.32	2.27
- S2	2.29	2.28	2.31	2.30	2.26	2.30
- S3	2.32	2.33	2.34	2.33	2.33	2.32
- Cys 6 SG	2.28	2.29	2.32	2.32	2.29	2.27
Fe2 - Fe3	2.71	2.72	2.72	2.71	2.69	2.69
- Fe4	2.69	2.70	2.69	2.69	2.66	2.64
- S1	2.28	2.26	2.30	2.29	2.29	2.26
- S2	2.36	2.35	2.37	2.36	2.34	2.31
- S4	2.32	2.33	2.34	2.32	2.32	2.33
- Cys 9 SG	2.26	2.26	2.28	2.29	2.27	2.31
Fe3 - Fe4	2.73	2.74	2.75	2.75	2.71	2.69
- S1	2.33	2.34	2.34	2.36	2.33	2.33
- S3	2.32	2.32	2.35	2.33	2.33	2.31
- S4	2.27	2.27	2.30	2.29	2.27	2.27
- Cys 18 SG	2.31	2.30	2.33	2.33	2.36	2.32
Fe4 - S2	2.36	2.34	2.37	2.37	2.33	2.34
- S3	2.26	2.26	2.30	2.29	2.28	2.28
- S4	2.34	2.35	2.36	2.37	2.35	2.32
- Cys24 SG	2.30	2.29	2.34	2.30	2.31	2.31

b)Hybrid cluster

	<i>D. desulfuricans</i> (A) as-isol.	<i>D. desulfuricans</i> (B) as-isol.	<i>D. desulfuricans</i> (A) reduced	<i>D. desulfuricans</i> (B) reduced	<i>D. vulgaris</i> as-isol.	<i>D. vulgaris</i> reduced
PDB ID	1GNL		1OA0		1W9M	1OA1
Fe5 - Fe6	2.76	2.77	2.76	2.76	2.75	2.73
- Fe7	3.63	3.67	3.98	3.97	3.64	3.89
- Fe8	3.35	3.36	2.84	2.84	3.12	2.79
- S5	2.31	2.31	2.31	2.30	2.31	2.28
- S6	2.32	2.31	2.35	2.33	2.30	2.35
- S7	-	-	2.34	2.35	-	2.29
- Cys 427 SG	2.32	2.33	2.36	2.36	2.32	2.34
- X	2.00	1.99	-	-	1.94	-
Fe6 - Fe7	4.94	4.86	5.60	5.58	5.12	5.57
- Fe8	3.02	3.02	3.79	3.78	3.15	3.74
- S5	2.25	2.25	2.22	2.22	2.19	2.22
- S6	2.27	2.26	2.22	2.21	2.25	2.23
- Cys 308 SG	2.33	2.33	2.30	2.31	2.31	2.28
- O8	1.98	1.93	-	-	2.11	-
- Y	-	-	3.81	3.80	-	4.00
Fe7 - Fe8	3.55	3.56	4.13	4.11	3.37	4.11
- S7	> 4.5	> 4.5	2.35	2.35	> 4.5	2.34
- His 240 NE2	2.13	2.16	2.18	2.17	2.18	2.12
- Glu 264 OE2	2.15	2.16	2.08	2.09	2.16	2.06
- Cys 452 SG	2.45	2.42	2.35	2.37	2.45	2.34
- O9	2.13	2.14	-	-	2.17	-
- X	2.06	2.08	-	-	2.10	-
- Y	-	-	3.14	3.15	-	2.78
Fe8 - S6	2.42	2.44	2.35	2.34	2.42	2.29
- S7	2.54	2.54	2.36	2.36	2.56	2.35
- Cys 399 SG	-	-	2.36	2.35	-	2.32
- Glu 487 OE2	2.04	2.02	2.07	2.07	2.09	2.01
- O8	2.19	2.17	-	-	2.24	-
- O9	1.93	1.95	-	-	1.89	-
- X	3.31	3.33	-	-	3.26	-
- Y	-	-	3.44	3.41	-	3.33
S7 - Cys 399 SG	2.09	2.09	4.06	4.04	2.10	4.00
O8 - Asn 307 ND2	3.02	3.04	-	-	2.88	-
O9 - Lys 489 NZ	2.99	2.95	-	-	2.95	-
Y - Asn 307 ND2	-	-	3.14	3.13	-	3.29
Y - Lys 489 NZ	-	-	2.89	2.88	-	2.93

Furthermore, genes encoding HCP proteins are present in the genomes of a wide variety of organisms that are not involved in the biological nitrogen cycle (for example, *Thermotoga maritima* and *Pyrococcus furiosus*), and present quite variable genomic organisations. Thus, at the present stage, it is not clear whether HCPs are directly involved in the metabolism of nitrogen species.

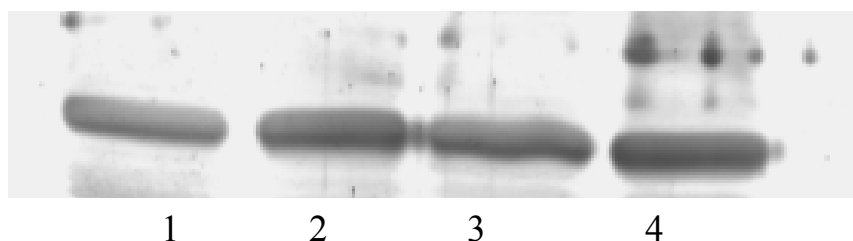


Figure 2.6. Immuno-blotting, using a polyclonal antibody against *D. vulgaris* HCP, of *D. desulfuricans* ATCC 27774 cell extracts from cultures grown with sulphate, nitrate and combinations of the two.

The lanes contain lysates of equal amounts of *D. desulfuricans* ATCC27774 cells grown in Saunders medium with lactate and (1) 50 mM sulphate; (2) 50 mM sulphate + 50 mM nitrate; (3) 50 mM sulphate + 10 mM nitrate and (4) 50 mM nitrate. Only a small increase in the expression of *D. desulfuricans* HCP is observed when the cells were grown in the presence of nitrate (lane 4).

Quite recently, Wolfe *et al.* (2002) reported that the recombinant HCP from *E. coli* exhibits *in vitro* hydroxylamine reductase activity, again suggesting a link of this protein to nitrogen metabolism. Such reductase activity can be rationalised by the structural changes undergone by the HCP enzymes as they transform from the reduced state to the as-isolated oxidised state. Thus, it can be envisaged that the substrate will initially bind to the reduced form of the enzyme close to the position of Y. Reduction of the substrate, with concomitant oxidation of the enzyme, will then take place as the substrate moves from the Y position to the X position in the as-isolated oxidised form. The product will then diffuse away from the active site and eventually away from the enzyme through the extensive hydrophobic cavity that gives ready access to the hybrid cluster (Cooper *et al.* 2000). The large movement of S7

will not only make accessible the X position in the oxidised form of the enzyme, but it may also assist in the electron transfer process. The possibility also exists that this sulphur atom may even provide an “S-Y” intermediate to assist in the translocation of the substrate at Y to the product at X.

However, in the present work, hydroxylamine reductase activity measurements have been performed for the anaerobically purified *D. desulfuricans* HCP (see Materials and methods). The specific activity determined, $3.6 \mu\text{mol NH}_2\text{OH reduced min}^{-1}\text{mg}^{-1}$, is approximately 10 times lower than that reported for the *E. coli* recombinant enzyme, and independent of pH between pH 7.0 and 8.8. It should be noted that hydroxylamine is quite unstable, at basic pH (Greenwood and Earnshaw 1994). The turnover value for the *E. coli* HCP (ca 40 s^{-1}) should also be compared with those from other *E. coli* enzymes known to reduce hydroxylamine, albeit having as their main physiological substrate, nitrite. Thus, for the periplasmic pentahaem nitrite reductase and the cytoplasmic sirohaem-containing nitrite reductase, expressed under nitrate and nitrite respiring conditions, this is reflected in high K_{MS} for hydroxylamine, in the millimolar range, compared with the micromolar K_{MS} for nitrite. The pentahaem nitrite reductase has, at neutral pH, a turnover for hydroxylamine reduction almost 50 fold higher than HCP [2380 s^{-1} (Bamford *et al.* 2002)]; values for the *E. coli* cytoplasmatic enzyme could not be found in the literature, but very similar sirohaem-containing enzymes, the spinach nitrite reductase and the *E.coli sulphite* reductase, have turnovers of about $500\text{--}750 \text{ s}^{-1}$ (Krueger and Siegel 1982; Siegel *et al.* 1982). These facts, together with the very high K_{M} for the *E.coli* HCP (30 mM at neutral pH), therefore argue against a real physiological role of HCPs as hydroxylamine reductases.

The hybrid cluster in HCP's exhibits a complex redox behaviour and four

redox can be detected by, for example, by EPR titration. On the basis of the differences in structures of the as-isolated oxidised and fully reduced proteins, which reveal the reduction of the cysteine 399 persulphide moiety, it cannot be excluded that at least one of the EPR detected transitions is due to this process, and not to a reduction based on the iron atoms in the cluster.

In conclusion, although the crystallographic studies alone cannot precisely define the natures of X and Y, the changes observed in the structure of the hybrid cluster as it transforms between the reduced and oxidised states give strong credence to the role of the HCPs as reductases. What is clearly indicated, by the structural changes in the HCPs in moving from the reduced to the oxidised states, is that the enzyme acquires two bridging oxygen atoms. Another possible function for the enzyme could be to sequester oxygen and/or its by-products (eventually together with reactive nitrogen species) in an anaerobic environment. That is, the HCP could protect the organisms against oxidative stress (Briolat and Reysset 2002) or against reactive nitrogen species. Further biochemical, spectroscopic and crystallographic studies are in progress to examine these hypotheses.

2.5. Acknowledgements

We thank the European Synchrotron Radiation Facility, Grenoble, France, and the staff of Macromolecular Crystallography Group for the provision of synchrotron radiation facilities without which the high-resolution studies would not have been possible. The ESRF also provided a Ph.D. studentship for SM and kindly provided partial support for DGA. DGA acknowledges a grant from Fundação para a Ciência e Tecnologia (FCT), SFRH (BD/6480/2001). CVR acknowledges a grant from Praxis XXI (BD/19879/99). PFL would like to acknowledge the assistance and

cooperation of the Department of Crystallography, Birkbeck College, University of London, UK. This work was partially funded by FCT, SAPIENS (POCTI/BME/38859/2001). The authors would also like to thank to Dora Alves for her valuable help on some aspects of *D. vulgaris* protein purification.

Figures 2.1 to 2.4 were produced with the help of the programs Molscript (Kraulis 1991), Bobscript (Esnouf 1999) and Raster3D (Merritt and Bacon 1997).

David Aragão was only involved in the work on *D. vulgaris* HCP. This protein purification was performed by Célia V. Romão, Dora Alves and David Aragão. The crystallisation, data collection, phase determination, structure refinement, structural analysis and comparison was done by David Aragão. The work on *D. desulfuricans* HCP was done mainly by Sofia Macedo and Célia V. Romão. Final refinement stages and part of the structure analysis of the reduced *D. desulfuricans* HCP were done by Peter F. Lindley.

3. The crystal structure of the hybrid cluster protein (HCP) from *Desulfovibrio desulfuricans* ATCC 27774 containing molecules in the oxidized and reduced states.

(Based on a published refereed paper but now includes a more detailed description of the refinement using SHELXL and other minor points: S. Macedo, D. Aragão, E. P. Mitchell and P. Lindley. Acta Crystallogr D Biol Crystallogr. (2003). D59, 2065-2071).

3.1. Abstract

The hybrid cluster protein from the sulphate reducing bacteria, *D. desulfuricans* ATCC 27774 (*D. desulfuricans*) has been isolated and crystallised anaerobically. The protein sample used in the crystallisation studies was several months old, having been stored at 193 K, and initial crystal structure studies were unable to fully resolve details of the hybrid cluster despite the use of high resolution data to 1.25 Å collected at the ESRF, Grenoble, France. The full elucidation of the structure has only become possible with the complete knowledge of the as-isolated and fully reduced crystal structures. The analysis clarifies the significant movements in the position of the iron atom linked to the persulphide moiety in the oxidised as-isolated protein and the sulphur atom of the persulphide itself, as the protein is reduced. The structures of the as-isolated and reduced states are discussed in terms of the putative function of the HCP proteins.

3.2. Introduction

Hybrid cluster proteins (HCP) are encoded by genes in a large number of prokaryotes, including archaea and bacteria, both strictly anaerobic and facultative anaerobic, as well as in eukaryotic protozoa. However, despite a wealth of biochemical, spectroscopic and high resolution X-ray structural information no specific function has yet been assigned to this family of proteins. Involvement in the biological nitrogen cycle has been proposed, (van den Berg *et al.* 2000; Beliaev *et al.* 2002; Wolfe *et al.* 2002), as well as an adaptive response to oxidative stress (Briolat and Reysset 2002). Similarity of the hybrid Fe-S cluster in HCPs to the nickel containing cluster present in the catalytic domain of the carbon monoxide dehydrogenase from *C. hydrogenoformans* has also been noted, (Dobbek *et al.* 2001; Macedo *et al.* 2002) a recent study Kim *et al.* (2003), have studied the modulation of virulence by two acidified nitrite-responsive loci of *Salmonella enterica* Typhimurium. Authors note that the nitrite-inducible promoters are located upstream of loci designated *nipAB nipC* which correspond to *hcp-hcr* (hybrid cluster protein) of *E. coli* and *norA* of *Alcaligenes eutrophus* respectively. This study refocused attention on a putative role for the HCP in the nitrogen cycle.

After the initial discovery of the hybrid cluster (Arendsen *et al.* 1998) some debate arose as to whether the structure was really an artifact resulting from the aerobic preparation and crystallisation. This debate occurred despite considerable spectroscopic and biochemical evidence to suggest that the as-isolated form of the protein was independent of the aerobic or anaerobic conditions used in its preparation. In order to unequivocally resolve the debate it was decided to isolate, purify and crystallise the protein from *D. desulfuricans* under strict anaerobic conditions. This was eventually

achieved using fresh samples of the protein isolated at the ITQB in Portugal (Macedo *et al.* 2002). However, in some early attempts using protein which had been prepared by the late Jean LeGall and his colleagues (University of Georgia, Athens, USA), transferred to the ITQB, and kept in a freezer for several months at around 193 K, crystals were produced containing a structure that could not be fully interpreted. It was suspected that the crystals contained the structure of the as-isolated oxidised form of the protein, but heavily contaminated with another form. Only in the full light of the as-isolated oxidised (Macedo *et al.* 2002) and dithionite reduced (Aragao *et al.* 2003) forms of the protein has it been possible to interpret the electron density resulting from data collected to high resolution on the old anaerobically treated protein sample. The present paper reports the X-ray crystal structure of the HCP from *D. desulfuricans* containing both the oxidised and reduced forms of the protein.

3.3. Materials and methods

3.3.1. Crystallisation

The sample of HCP from *D. desulfuricans* was anaerobically purified by the late Jean LeGall and colleagues at the University of Georgia, USA, in a similar manner to that previously described (Macedo *et al.* 2002), transferred to the ITQB, Portugal, and kept at around 193 K for several months before use. The crystals were grown in a nitrogen environment inside a glove box system ($pO_2 < 1$ ppm) using the sitting drop vapour-diffusion method. The pale yellow plate-like crystals grew at 279 K in 25% PEG 8000 and 0.1 M MES buffer at pH 6.0. Protein solution, 1.5 μ l, (12.6 mg ml⁻¹ in 20 mM Tris-HCl pH 7.6 and 0.2 M NaCl) and an equal amount of precipitant solution were allowed to equilibrate against 500 μ l of the latter in the reservoir. Crystals took some 4 – 5 weeks to grow and reached a maximum size of

approximately 0.2 mm x 0.1 mm x 0.08 mm. Crystals for diffraction experiments were then mounted in loops and cryo-frozen at 100 K within the anaerobic environment of the glove box and transferred to a synchrotron in a nitrogen Dewar. Crystals were also obtained with 25%PEG 4000, 0.1 M HEPES at pH 7.0 with 0.2 M calcium acetate, but these gave very poor unusable diffraction patterns.

3.3.2. X-ray data collection

X-ray data were collected from two crystals on beam line ID14-2 at the ESRF synchrotron radiation source, Grenoble, France, using a wavelength of 0.933 Å. For the first crystal the ADSC CCD detector was positioned at a distance of 135 mm from the sample to collect data between 30.0 Å and 1.4 Å resolution. The beam size used was 50 µm x 50 µm with an oscillation range of 0.75°, 4 passes per frame with a frame exposure time of 2 sec. For the second crystal, the detector was positioned 60 mm from the sample with a potential data collection range down to 0.97 Å resolution. The beam size was chosen as 65 µm x 65 µm with an oscillation range of 0.5° using 1 pass and 10 seconds exposure per frame.

Data for both crystals were processed using DENZO (Otwinowski and Minor 1997). The program SCALEIT was used to check that the crystals were isomorphous and then 280 batches from the first crystal were merged with 244 batches from the second using SCALA, and the indexed intensities were converted to structure factors using the program TRUNCATE; SCALEIT, SCALA and TRUNCATE are part of the CCP4 program suite, (CCP4 1994). Details of the data collection are given in Table 3.1. In hindsight, it is clear that the data collection strategy adopted was not the optimum, but lack of crystals and accessibility to the synchrotron prevented remedial actions. As a consequence, it was decided that because of the high

values of the R_{symm} and the overall weakness and low multiplicity of the high resolution data, further work would proceed with data only to 1.25 Å resolution; the R_{symm} in the range 1.29 to 1.25 Å is 24.1%. The data used comprised 227130 unique reflections corresponding to an overall completeness of 88.9%.

Table 3.1. Data collection statistics.

Beam line	ID14-2 : ESRF
Space group	P1
Z	2
Resolution (Å) : crystal 1	30.0 – 1.40
: crystal 2	6.0 – 1.15
Wavelength (Å)	0.933
Cell dimensions : a (Å)	57.30
: b (Å)	61.20
: c (Å)	72.00
: α (°)	82.80
: β (°)	73.70
: γ (°)	87.30
No. of batches & rotation angle (°)	
: crystal 1	280 / 0.75
: crystal 2	244 / 0.5
No. of unique (hkl)	
: crystal 1	157 207
: crystal 2	232 683
Merged data from crystals 1 and 2	
No. of unique hkl to 1.15 Å resolution	280898
Average multiplicity [§]	2.3 (1.3)
R_{merge} ^{6 §}	0.082 (0.271)
Average $I/\sigma(I)$ [§]	3.6 (1.3)
% completeness [§]	84.8 (70.0)
Average B (Wilson) Å ²	6.0

§ - Values in parentheses refer to the highest resolution shell collected, 1.19 – 1.15 Å.

The asymmetric unit (equivalent to the unit cell for space group P1) accommodates two molecules of molecular weight 60 kDa, corresponding to a V_m of 2.03 Å³ Da⁻¹, (Matthews 1968), and a solvent content of some 39 %.

⁶ R_{merge} is $\Sigma |I - \langle I \rangle| / \Sigma \langle I \rangle$;

3.3.3. Structure solution and refinement

The structure was elucidated by the method of molecular replacement using the program AMORE (Navaza 1993) in a similar manner to that used for the anaerobically isolated, oxidised structure, (Macedo *et al.* 2002). The starting model was the oxidised structure of the HCP from *D. vulgaris* (Arendsen *et al.* 1998) from which the two Fe-S clusters and the solvent molecules had been removed; the *D. desulfuricans* and *D. vulgaris* proteins share a 64% sequence identity and differences were not considered at this stage. AMORE enabled the location of both non crystallographic symmetry (NCS) related molecules in the asymmetric unit and solvent modification was then performed using DM (Cowtan and Main 1998); the R-factor and correlation coefficients were 43.1% and 44.2 respectively. The initial Fourier map was of high quality and clearly showed electron density consistent with the two molecules of the *D. desulfuricans* protein in the asymmetric unit. In each molecule nearly all the main chain and many side chain residues for *D. desulfuricans* could be positioned directly. However, in the hybrid clusters the electron density could not be explained by the presence of the as-isolated oxidised form of the protein alone. Extra density was particularly evident in the vicinity of Fe8 and near to the moiety bridging Fe5 and Fe7 and assigned as an oxygen atom (OH⁻ or water) on the basis of resonance Raman spectroscopy (de Vocht *et al.* 1996). Figure 3.1 shows part of the electron density at the hybrid cluster in molecule A and, superimposed, the as-isolated, oxidised form of the HCP from *D. desulfuricans*. However, when the structure of the reduced form of the protein became available (Aragao *et al.* 2003) it was clear that the key features of the extra density could be matched well with the reduced form. A consideration of the respective peak volumes suggested a 2:1 ratio of the reduced and oxidised forms respectively.

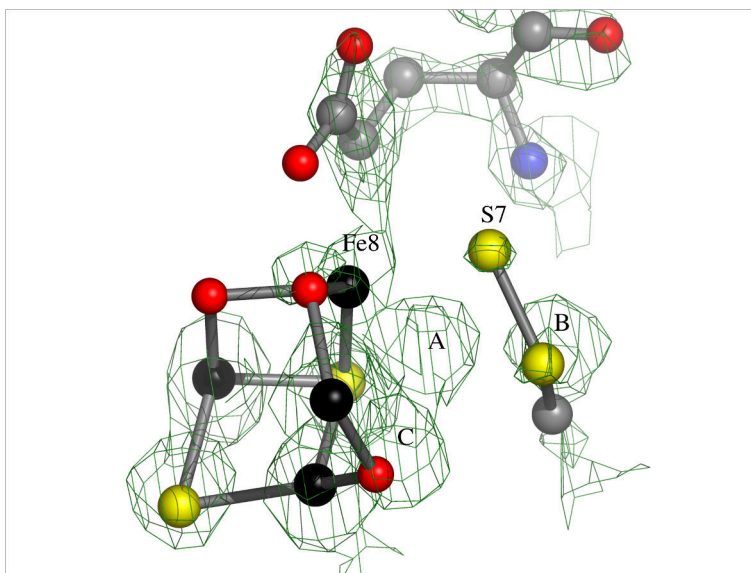


Figure 3.1. The $|F_o| - |F_c|$ electron density synthesis at the hybrid cluster in molecule A prior to the refinement stage.

At this time only the as-isolated oxidised form of the HCPs was known (Macedo *et al.* 2002) and the superposition of this structure on the electron density clearly highlights many unexplained features including the maxima at A, B and C. After the elucidation of the reduced structure (Aragao *et al.* 2003) it was possible to account for most of the additional density. Thus, the peaks labelled A, B and C are occupied by Fe8, the SG of Cys399 and S7 (part of the persulphide moiety in the oxidised form) in the reduced structure respectively. The contour level is +3.0 rms.

a) Refinement with REFMAC

Refinement was performed using the maximum likelihood functions implemented in REFMAC (Murshudov *et al.* 1997). Rounds of conjugate gradient sparse matrix refinement with bulk solvent modelling according to the Babinet principal (Tronrud 1996) were alternated with model building using the O program (Jones *et al.* 1991) in combination with SigmaA weighted $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ electron density maps (Read 1986). For the hybrid clusters, only Fe5, Fe6, Fe7, S5 and S6 of the reduced form were initially included with full occupancy to give values of R and R_{free} (0.9% of the data) of 29.3 and 29.7% respectively. As the refinement process

proceeded, the corresponding atoms of the oxidised form were added and partial occupancies for both forms fixed in the ratio 0.7 reduced to 0.3 oxidised. The two positions for Fe8 and S7 were then added together with the respective conformations for residues 398-402 (Cys399 is directly bound to Fe8 in the reduced form and through a persulphide moiety in the as-isolated oxidised form) and E487 (directly bound to Fe8 in both forms) with the same relative occupancies. Finally the various oxygen atoms in the two forms of the protein were also included. Since the program REFMAC did not allow the refinement of occupancies, an independent refinement using SHELXL, (Sheldrick and Schneider 1997) was performed in which group occupancies for the reduced and oxidised forms of the cluster were refined (see section SHELXL refinement). As a consequence, the occupancies were adjusted to 0.65 and 0.35 for the reduced and oxidised form of the cluster respectively in the final cycles of the REFMAC refinement. This resulted in a better agreement between the temperature factors of equivalent atoms in the two forms of the hybrid cluster. Nevertheless, some individual adjustments were made for the oxygen atoms in the oxidised and reduced forms of the protein as discussed in the next section.

The geometries of the cluster atoms were unrestrained throughout the refinement, and in the final stages all the Fe and S atoms (inorganic S and cysteine S) associated with the cubane cluster in each molecule were allowed to refine with anisotropic *B*-factors. After the first round of refinement, solvent molecules were added to the model based on standard geometrical and chemical restraints; they were subsequently deleted if their isotropic thermal coefficients exceeded 45 Å² or they were not visible at the 0.8 rms level in SigmaA weighted $|F_o| - |F_c|$ electron density maps (Read 1986). In several regions, solvent molecules were close together indicating either partially occupied sites or possible stretches of polyethylene glycol density.

As far as possible, where solvent peaks occurred within 2.0 Å of one another, they were treated as partially occupied sites with occupancies (not refined) roughly proportional to the peak volumes. In addition to the solvent peaks assigned to water, there are two regions of density that are clearly different to the normal solvent structure. The first of these lies in the vicinity of the cubane cluster in molecule A and was tentatively assigned as a dithionite anion (DTN) in the reduced structure (Aragao *et al.* 2003), since dithionite was used in the protein reduction process. In the current structure it has been modelled by two water molecules (W1182 and W1334), both occupying two positions. The density does not appear to be present near the cubane cluster in molecule B. The second region of density lies between molecules A and B in the asymmetric unit and can clearly be ascribed to a MES molecule; 0.1M MES buffer was used in the crystallisation process. One of the sulphite oxygen atoms interacts with the side chain of Lys313 in molecule B. A second interacts with the main chain amino atom of Asp314, also in molecule B, and with a neighbouring solvent molecule, whilst the third interacts with a solvent molecule only. The morpholino moiety fits into a pocket in molecule A in the vicinity of Lys434.

For the protein, hydrogen atoms were included as 'riding' atoms. Details of the final refinement statistics are given in Table 3.2. Several residues were modelled in more than one conformation. A number of the surface residue side chains, particularly lysines, had no visible electron density after the CG atom, but as far as possible these were modelled in geometrically feasible positions. Details are shown in Table 3.2.

Table 3.2. Refinement statistics.

Amino acids	2 x 544
Protein atoms	4195A : 4205 B
Solvent atoms	1645
No. in dual conformations	29
Other atoms	2 x cubane and hybrid clusters + 1 x MES
Resolution Limits (Å)	27.52 – 1.25
R (All data)	0.140 (227130)
Test set R (Observations)	0.153 (2141)
Highest resolution shell, 1.29-1.25 R (working set)	0.206 (13493)
Free R (test set)	0.221 (125)
ESU based on max. likelihood (Å)	0.026
Average B _{iso} (Å ²)	
All atoms (10138)	11.6
Protein atoms	8.8 A : 9.0 B
Main chain	7.9 A : 8.1 B
Side chain	9.7 A : 10.0 B
Solvent	25.1
Residues not found	G544 A & B
Residues in alternative conformations (Residues in bold type are linked to Fe8)	A: V21, R61, M90, S101, K165, N168, M174, S345, I347, E364, G398, C399, D400, G401, R402 , E414, N436, E487 . B: V21, Q46, D59, K165, E167, M174, K250, K290, V302, I338, E340, S345, N436, G398, C399, D400, G401, R402 , N436, S455, I483, E487, I529 .
Distance deviations: [†]	
Bond distances (Å)	0.008
Bond angles (°)	1.351
Planar groups (Å)	0.005
Chiral volume deviation (Å ³)	0.147

[†] rms deviations from standard values.

b) Refinement with SHELXL

Refinement was performed using the least squares functions implemented in SHELXL (Sheldrick and Schneider 1997) using the same intensities for cross validation as in the REFMAC refinement. The Engh and Huber stereochemical dictionary (Engh and Huber 1991) was used for the

polypeptide moiety. NCS restraints were imposed on dihedral angles and isotropic *B-factors* between the two monomers of the asymmetric unit but were released later on. Initially only the cubane cluster was included in the protein model and no restraints were imposed to any of the iron-sulphur clusters through the whole refinement. The initial electron-density fitted model was refined in increasing resolution steps of 0.05 Å, beginning from 1.9 Å resolution data and refining isotropic *B-factors*. At 1.35 Å resolution data refinement with anisotropic *B-factors* was introduced for the iron and sulphur atoms of the cubane cluster and of sulphur containing aminoacids. The refinement was applied using default “rigid-bond restraints” (DELU) and “similar ADP restraints” (SIMU) on neighbouring atoms (Sheldrick and Schneider 1997). Then the hybrid cluster was included in the model as two conformations, with 0.6 occupancy for the reduced form and 0.4 occupancy for the oxidised form with a restraint on the sum of the occupancies as 1. At the end of the refinement, hydrogen atoms were introduced using the riding model. The structure was checked against SigmaA weighted maps (Read 1986) using XtalView (McRee 1992) and corrected and completed with discretely alternating side chains. Further solvent waters were added and the protonation assignment of basic residues was eventually corrected according to their immediate environment and the bump errors due to atoms close distance. See Table 3.3 for details of the progress of the refinement. In the last cycle of refinement, at 1.25 Å resolution, the occupancy was set at 0.65 and 0.35 for the reduced and oxidised form of the cluster respectively.

3.3.4. Quality of the refined structures

The quality of the refined structure of *D. desulfuricans* is high with final values for R and R_{free} of 14.0 % and 15.3 %, respectively and an overall G factor of 0.21 (Laskowski *et al.* 1993), Table 3.4.

Table 3.3. *Progress of the SHELXL refinement.*

Refinement	Resolution (Å)	Solvent + alternative side chains	R (%)	Rfree (%)
All isotropic	1.90	0	28.99	24.13
Solvent added	1.90	1381+3	23.00	18.26
	1.50	1381+15	20.58	17.80
	1.35	1394+15	20.11	17.68
Cubane atoms anisotropic	1.35	1394+15	18.70	16.80
Hybrid cluster modelled	1.35	1394+15	15.92	14.12
H atoms (riding position)	1.25	1394+15	15.72	14.10

Table 3.4. *Quality of reduced HCP structure.*

	<i>D. desulfuricans</i>
Overall G factor	0.21
Ramachandran analysis % (No.)	
Favourable	93.6 (887)
Additional	5.7 (54)
Generous	0.3 (3)
Disallowed [‡]	0.4 (4)
No. of glycine residues	2 x 49
No. of proline residues	2 x 18

[‡] The residues in the disallowed region of the Ramachandran plot, Ser263 and Asn299 in both molecules A and B lie in well defined regions of electron density at sharp bends between helices and strands.

Information for the data set and the associated structure, REFMAC refinement, has been deposited in the Protein Data Bank (Berman *et al.* 2000) with deposition code 1UPX.

3.4. Results and discussion

The HCPs from the sulphate reducing bacteria *D. desulfuricans* ATCC27774 and *D. vulgaris* (strain Hildenborough) have three domains as shown in Figure 3.2. The N-terminal domain, residues 1-220, comprises two three-helix-bundles arranged perpendicular to one another. The first 21 residues encompass a standard [4Fe-4S] cluster which is readily accessible to the external environment and may play a role in electron transfer. The two remaining domains comprise a central β -sheet with helices above and below.

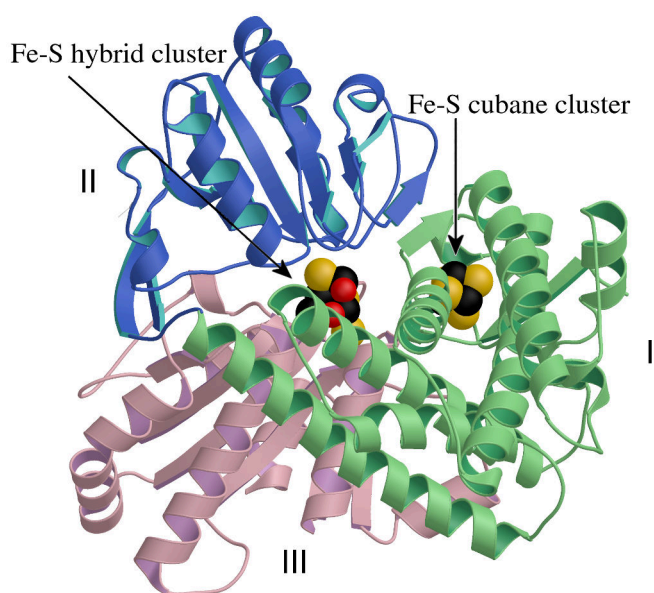


Figure 3.2. An overall view of the hybrid cluster protein.

The three domains and the two Fe-S clusters are shown; a cubane cluster situated near the out side of the protein in the vicinity of the N-terminus in domain 1 and the hybrid cluster at the interface of the three domains.

The hybrid cluster lies at the interface of the three domains and is some 11-12 Å distant from the cubane cluster. In the as-isolated oxidised form, whether the protein is isolated aerobically or anaerobically (Macedo *et al.*

2002), the hybrid cluster can be considered as being derived from a cube, but with an iron atom, Fe7, at one corner peeled back as shown in Figure 3.3(a), (Macedo *et al.* 2002). The hybrid cluster contains both μ -S and μ -O bridges between pairs of iron atoms. Thus, O8 and O9 bridge Fe6 and Fe8 and Fe7 and Fe8 respectively. A further moiety, X, appears to bridge Fe5 and Fe7 and in this case the environments of the iron atoms can be described as distorted tetrahedral for Fe5 and Fe6, and distorted trigonal bipyramidal for Fe7. The coordination of Fe8 is distorted trigonal bipyramidal, but if X is also considered, it becomes distorted octahedral. In the as-isolated oxidised form of the protein, X has been assigned as an oxygen atom, O11.

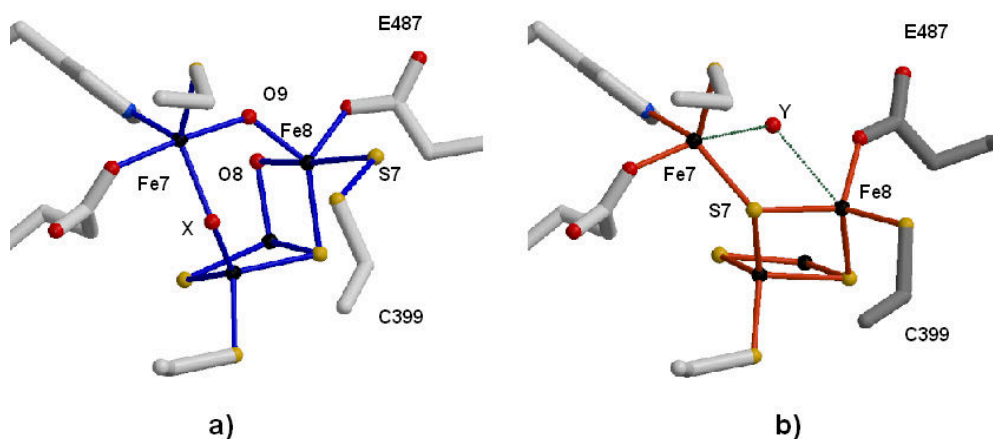


Figure 3.3. Models of the as-isolated oxidised (a) and reduced (b) forms of the hybrid cluster.

The large movements of Fe8 and S7 and the loss of the bridging oxygen atoms can be clearly envisaged. Cys308, bound to Fe6, has been omitted for clarity in both (a) and (b), as have Asn307 and Lys489 that may form hydrogen bonds with Y in the reduced structure.

In the reduced form (Aragao *et al.* 2003) the overall molecular structure is virtually the same, but there are major changes at the hybrid cluster. As shown in Figure 3.3(b) the Fe7 atom linked to the protein through residues

His240, Glu264 and Cys452 is virtually unchanged in position. The [2Fe-2S] moiety involving Fe5 and Fe6, and linked to the protein through cysteine residues 427 and 308 respectively undergoes a small movement, but Fe8 and S7 of the persulphide moiety at Cys399, undergo large movements in position and coordination. Concomitantly, residues 398 to 402 and including Cys399, and Glu487 attached to Fe8, also undergo significant positional changes. The persulphide S atom, S7, moves very close to the position of the moiety X, so that S7 forms the second side of a cube in conjunction with atoms Fe5, Fe8 and S6 and is only 2.4 Å from Fe7. The movement of S7 in *D. desulfuricans* is 3.6 Å and the SG of Cys399 now forms a direct bond to Fe8. In turn Fe8 has moved by some 2.1 Å. The geometries of the iron atoms in the reduced form of the hybrid cluster can now be described as tetrahedral for Fe5, Fe7 and Fe8, and trigonal planar for Fe6 in sharp contrast to those observed in the as-isolated oxidised cluster. The hybrid cluster in the reduced form of the protein also lacks the μ -bridging oxygen atoms O8 and O9, between Fe6 and Fe8 and between Fe7 and Fe8 respectively, present in the as-isolated form. However, there appears to be a moiety binding at Y, not present in the as-isolated form and roughly in between the positions formerly occupied by O8 and O9. Y has currently been assigned as a single oxygen atom (O12), but it could be a larger moiety. Y is some 3.1 Å and 3.4 Å from Fe7 and Fe8 respectively, too long for a direct bond. It is, however, well within hydrogen bonding range of the NZ of Lys489 (2.9 Å) and the ND2 of Asn307 (3.1 Å); both these residues are conserved in most of the known HCP sequences.

In the present structure both the as-isolated and reduced forms of the HCP appear to be present in the crystal. This is clearly illustrated in Figure 3.4 which shows electron density syntheses corresponding to the two forms individually and combined. It is not clear how this has occurred since the

protein was prepared anaerobically in the as-isolated oxidised state and no deliberate attempts to reduce it were made. Radiation damage is known to cause reduction, but a close examination of the data gave no clear evidence that radiation damage was significant. The reduction has presumably occurred as a result of transportation between continents and/or prolonged storage! A model with 65% and 35% of the reduced and oxidised structures explains most of the electron density in SigmaA weighted $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ electron density maps at the hybrid clusters with the exception of that in the vicinity of the O8 and O9 atoms in the oxidised and the O12 atom in the reduced structure respectively. Increasing the occupancies of O8 and O9 and decreasing the occupancy of O12 alleviates any discrepancies, but whether such adjustments have any experimental significance is debatable. If they are real, then they may indicate that the crystal also includes some structures intermediate between the as-isolated oxidised and reduced states.

In a previous paper, (Aragao *et al.* 2003), a reductase enzyme mechanism was proposed whereby the substrate initially binds to the reduced form of the enzyme close to the position of Y (O12), Figure 3.3(b). Reduction of the substrate, with concomitant oxidation of the enzyme, will then take place as the substrate moves from the Y position to the X (O11) position in the as-isolated oxidised form, Figure 3.3(a). The product will then diffuse away from the active site through the extensive hydrophobic cavity that gives ready access to the hybrid cluster, (Cooper *et al.* 2000), and eventually away from the enzyme. The large movement of S7 makes the X position accessible in the oxidised form of the enzyme and may have some role in the substrate movement and its reduction by electron transfer. As a consequence of the translocation of the substrate and its reduction, two oxygen atoms are introduced into the cluster at O8 and O9. Whether these arise from one or

more substrate molecules and/or are derived from solvent molecules is not yet known. However, an oxidase activity for the HCP proteins cannot be discounted at this stage. In this case the substrate would initially bind at X in the as-isolated oxidised form of the enzyme and then trans-locate to Y, gaining one of more oxygen atoms from O8 and O9 as the enzyme is reduced. Until the nature of X and Y are clarified the function of the HCPs will remain undefined.

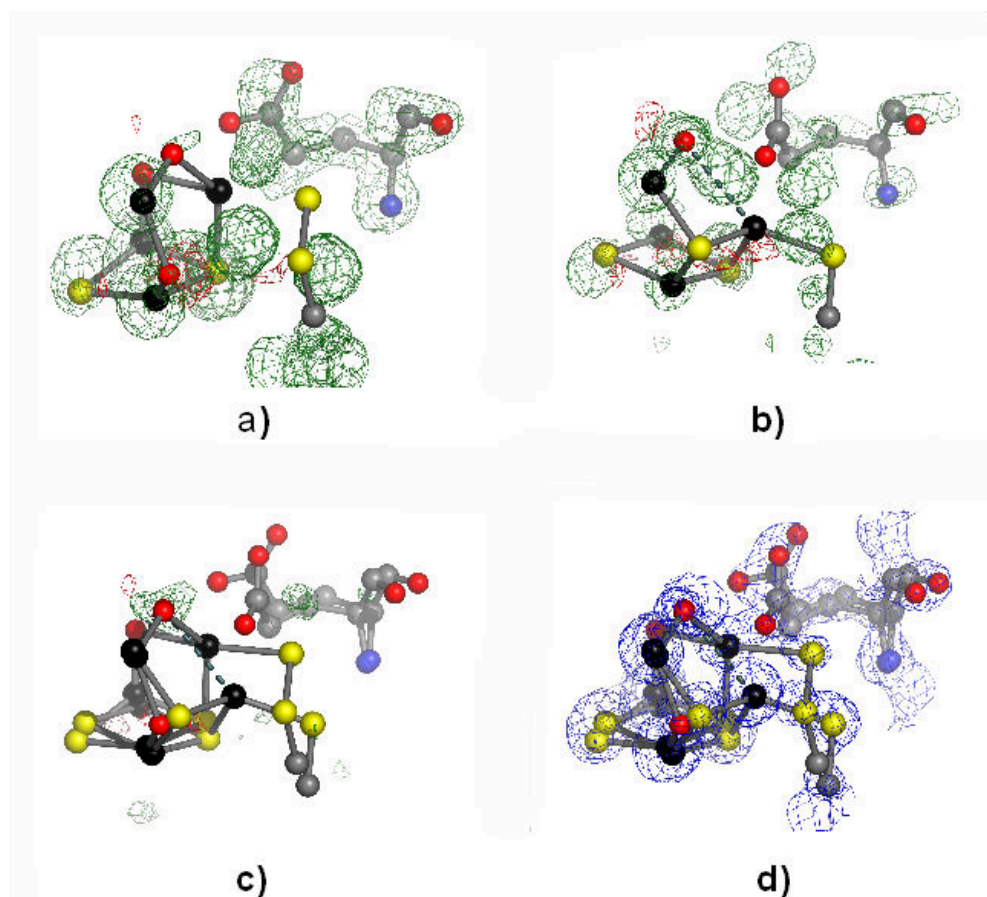


Figure 3.4. *SigmaA weighted electron density syntheses showing the presence of the as-isolated oxidised and reduced forms of the HCP in the vicinity of the hybrid cluster.*

In (a), (b) and (c) the Fourier coefficients used were $|F_o| - |F_c|$ and phases were based on (a) the as-isolated oxidised form, (b) the reduced form, (c) both forms in the ratio 65% reduced to 35% oxidised; the contour levels are +3.5 rms (green) and -3.5 rms (red). The only feature of chemical interest in (c) is a peak at 4-5 rms close to O9 in the oxidised form, and Y (O12) in the reduced form. This peak may correspond to a partial substrate or product. Figure 3.4(d) is the same as (c) with coefficients $2|F_o| - |F_c|$; the contour level is 1.0 rms (blue).

Studies on the HCP proteins enable a number of observations to be made. Firstly, the adage that it is highly desirable to use freshly prepared protein for crystallisation studies appears to be strongly reinforced. Secondly, crystallographic studies alone, even if they are derived from high quality, high resolution data, cannot always precisely define what function an

enzyme performs, or indeed, how it achieves this function. Proteomics specialists beware! In this case, even with extensive biochemical and spectroscopic studies, the nature of X and Y and the meaning of the changes observed in the structure of the hybrid cluster as it transforms between the reduced and oxidised states remain to be determined, although they may give strong credence to the role of the HCPs as reductases. What is clearly indicated by the structural changes in the HCPs in moving between the reduced and the oxidised state and *vice versa*, concomitantly with the large movements of Fe8 and S7, is that they involve a gain or loss respectively of the two bridging oxygen atoms, O8 and O9. In the former case this would indeed be consistent with the sequestration of oxygen and/or its by-products (eventually together with reactive nitrogen species) in an anaerobic environment. That is, the HCP could protect the organisms against oxidative stress (Briolat and Reyssset 2002) or against reactive nitrogen species. As the enzyme returns to its reduced state the oxygen atoms would be released as two water molecules.

Clearly additional studies are required and further biochemical, spectroscopic and crystallographic studies are in progress to attempt to define the function of the HCPs and to examine the hypotheses mentioned above.

3.5. Acknowledgements

We thank the European Synchrotron Radiation Facility, Grenoble, France, and the staff of Macromolecular Crystallography Group for the provision of synchrotron radiation facilities without which the high-resolution studies would not have been possible. The ESRF also provided a Ph.D. studentship for SM and kindly provided partial support for David Aragão. David Aragão acknowledges a grant from Fundação para a Ciência e Tecnologia (FCT),

SFRH (BD/6480/2001). We would also like to acknowledge our colleagues at ITQB, M^a Armenia Carrondo, Miguel Teixeira, Ligia Saraiva and Carlos Frazão for helpful advice and the late Jean LeGall and his colleagues in the USA for providing the sample of the HCP protein. PFL would like to acknowledge the assistance and cooperation of the Department of Crystallography, Birkbeck College, University of London, UK. This work was partially funded by FCT, SAPIENS (POCTI/BME/38859/2001).

Figures were produced with the help of the programs Raster3D (Merritt and Bacon 1997), and PYMOL (DeLano 2002).

David Aragão was involved in protein expression and purification together with Jean Legall and co-workers while in University of Georgia, Athens, USA. The crystallisation, data collection and phase determination was done by Sofia Macedo. The refinement using REFMAC was performed by Peter F. Lindley. The refinement using SHELX was performed by David Aragão. Structure analysis of was done by Peter Lindley together with David Aragão.

4. Structure of *Desulfovibrio vulgaris* anaerobically purified hybrid cluster protein at 1.35 Å resolution using MAD at the iron K-absorption edge

(David Aragão, Célia V. Romão, Dora Alves, Miguel Teixeira, Maria Arménia Carrondo, Edward P. Mitchell, Carlos F. Frazão and Peter F. Lindley. *to be published*)

4.1. Abstract

The hybrid cluster protein (HCP) from the sulphate reducing bacterium *Desulfovibrio vulgaris* Strain Hildenborough (*D. vulgaris*) has been isolated and crystallised anaerobically. The phase problem was solved for a $P2_12_12_1$ crystal form using multiple wavelength anomalous diffraction data collected in the vicinity of the iron K absorption edge. Although the structure is essentially the same as previously obtained when aerobically isolated, it shows that the nature of the hybrid cluster has particular differences when isolated and crystallised in the absence of oxygen. This gives insights about structural features associated with changes in the oxidation state of the HCP oxygen sensitive iron-sulphur clusters. A comparison between HCPs and carbon-monoxide dehydrogenases (CoDs) shows a similar fold and a related cluster at the HCP hybrid cluster position. This helps to understand the nature of the hybrid cluster.

4.2. Introduction

Hybrid cluster protein (HCP) is a 60 kDa protein (approximately 550 residues) present in a wide variety of organisms from the Bacteria, Archaea and Eukarya kingdoms (see chapter 1). HCP contains two types of Fe-S clusters; one variable and present as either [2Fe-2S] or [4Fe-4S] and another

so called hybrid cluster, [4Fe-2S-3O]. The latter is the main feature in the HCP structure and was initially thought to be a [6Fe-6S] moiety, in a “prismane” configuration (Hagen 1989). Later studies, when the first crystallographic structures were obtained to 1.72 Å resolution, suggested that it may have resulted from aerobic oxidation (Arendsen *et al.* 1998). Furthermore, not all of the ligands to this cluster could be reliably assigned. Further crystallographic studies led to higher resolution, 1.25 Å, structures obtained using protein prepared and crystallised anaerobically, for *Desulfovibrio desulfuricans* and aerobically for *D. vulgaris* (Macedo *et al.* 2002). Subsequently dithionite reduced structures for both bacteria were also determined (Aragao *et al.* 2003). Despite all these crystallographic structures, detailed spectroscopic analyses and many other studies, the precise function(s) of HCP is still subject to discussion. Two main hypotheses for its function have been proposed. Firstly it is thought to have a role in protection against reactive nitrogen species and secondly in oxygen related oxidative protection mechanisms. The first hypothesis gained ground with the publication of three reports; (a) the conversion of the HCP related NiFeS carbon monoxide dehydrogenase (CoD) to a hydroxylamine reductase by amino acid mutation or nickel to iron replacement at the CoD active site (Heo *et al.* 2002), (b) the report of hydroxylamine reductase activity for *Escherichia coli* (*E. coli*) HCP (Wolfe *et al.* 2002), and (c) increased anaerobic tolerance to hydroxylamine of *E. coli* cells overproducing *Rhodobacter capsulatus* HCP (Cabello *et al.* 2004). However the very high optimum pH (Cabello *et al.* 2004) and high K_m at neutral pH (Aragao *et al.* 2003) determined for hydroxylamine reductase activity do not give complete confidence in the suggestion that the true physiological function of this protein is as a hydroxylamine reductase. More recently it has been reported that HCP transcription (in *E. coli*) is regulated in response to nitric oxide or

to the regulatory protein, NsrR, which is involved in reactive nitrogen stress response (Filenko *et al.* 2005; Flatley *et al.* 2005; Filenko *et al.* 2007). These findings are reinforced by other reports pointing to increased HCP transcription in response to nitrogen oxides in *Salmonella enterica* (Kim *et al.* 2003), *Shewanella oneidensis* (Beliaev *et al.* 2005) and *D. vulgaris* (He *et al.* 2006). In light of these facts, protection against a reactive nitrogen compound, other than hydroxylamine, cannot be ruled out. The second hypothesis of HCP function has been linked to an oxidative stress response (Briolat and Reysset 2002). Almeida and co-workers (Almeida *et al.* 2006) have reported HCP induction by hydrogen peroxide, with regulation by the redox-sensitive transcriptional activator OxyR, and peroxidase activity for *E. coli* and *D. desulfuricans* HCPs using ascorbate as the electron donor. Despite a K_m within the values measured for typical peroxidases, the low V_{max} obtained for both HCPs, from *E. coli* and *D. desulfuricans* still poses doubts about the true physiological function for this protein. A search for the physiological electron donor and for other substrate oxidative species with or without nitrogen could lead to the true function.

HCPs share sequence and structure homology to anaerobic CoDs but not to aerobic CoDs that use a molybdopterin cofactor. CoD is an enzyme responsible for the reduction of CO_2 to CO. Anaerobic CoD can be a single monomer as in *Carboxydotherrmus hydrogenoformans* (Dobbek *et al.* 2001) or a homo-dimer as in *Rhodospirillum rubrum* (Drennan *et al.* 2001). In both cases three clusters are present. However, the monomeric enzymes have a hybrid $[Ni-Fe_4S_4]$ (also called the C-cluster), a cubane $[Fe_4S_4]$ (B-cluster) and a simple $[Fe_2S_2]$ (D-cluster) cluster whilst the dimeric forms contain another $[Fe_4S_4]$ cubane, which replaces the $[Fe_2S_2]$ cluster, with ligands from both monomers composing the dimer. In addition, the CoD dimer can also form $\alpha_2\beta_2$ hetero-tetramers as in *Moorella thermoacetica* (Doukov *et al.*

2002; Darnault *et al.* 2003). The latter is a bi-functional enzyme with catalysis occurring at the C- and A-clusters. In the C-cluster (corresponding to the hybrid cluster in HCP), located in the beta subunit, CO is generated by reducing CO₂ whilst in the A-cluster, located in the alpha subunit, CO is combined with CoA and a methyl group to form acetyl-CoA (Doukov *et al.* 2002).

4.3. Materials and methods

4.3.1. Purification of *D. vulgaris* HCP

HCP was overexpressed using *D. vulgaris* cells harbouring the recombinant plasmid pJSP104. Cells were resuspended in a degassed solution of 10 mM Tris-HCl pH 7.6 and disrupted in an oxygen free French Press previously flushed with argon. Crude cell extract was obtained by ultracentrifugation at 40,000 rpm for 5 h and followed by centrifugation at 18,000 rpm for 1 h. All purification steps were performed at 277 K in an anaerobic chamber, model A-2463 (Coy Labs, MI - USA) under a nitrogen atmosphere. The supernatant was applied to a Q-sepharose anion exchange column (Pharmacia XK26 x 40 Fast-Flow) pre-equilibrated with buffer A (20 mM Tris-HCl pH 7.6). The column was washed with 400 ml of the same buffer and a gradient to 100% of buffer B (20 mM Tris-HCl pH 7.6, 400 mM NaCl) was applied (1.6 l at 4 ml.min⁻¹). HCP eluted together with other contaminants at approximately 120 mM NaCl. The eluted fractions were immediately concentrated and desalted by ultrafiltration on an Amicon YM-30 membrane before application to the next column, a ceramic hydroxyapatite HTP column (Pharmacia XK 3 x 28) previously equilibrated with buffer C (5 mM KH₂PO₄ pH 7.6). This column was washed with 500 ml of buffer C and then a linear gradient of buffer D (200 mM KH₂PO₄ pH 7.6) was applied (500 ml at 0.8 ml.min⁻¹). HCP eluted together with contaminants

at approximately 40 mM. These fractions were desalted, transferred into buffer A and applied to a mono-Q column (Pharmacia Mono-Q). A linear gradient to 50% buffer E (20 mM Tris-HCl pH 7.6, 200 mM NaCl) was applied (120 ml at 4 ml.min⁻¹). HCP eluted at approximately 50 mM NaCl. HCP containing fractions were concentrated and dialysed against buffer E and then applied to a Superdex 75 column (Pharmacia XK 16). The column was eluted with the same buffer at a flow rate of 0.3 ml.min⁻¹. The purity was verified by a 15% SDS-PAGE gel and UV-visible spectra. A total of 27.5 mg of pure HCP protein was concentrated to 25 mg.ml⁻¹ and stored in buffer F (50 mM Tris-HCl pH 7.5, 50 mM NaCl) at 193 K.

4.3.2. Crystallisation of *D. vulgaris* HCP

Initial crystallisation trials based on the original conditions obtained for aerobic HCP (Arendsen *et al.* 1998) were set-up immediately using fresh (less than 5 days old) protein. All crystallisation trials were performed with protein at a concentration of 25 mg.ml⁻¹, at 278 K in a nitrogen anaerobic chamber in the presence of less than 3 ppm of oxygen. All buffers were previously degassed. Screens were made with 0.1 M MES buffer (pH range 6.0 - 6.5), PEG 4000 (concentration range 15 - 30%) and drops of protein plus precipitant of: 1.5 µl + 2.5 µl, 2 µl + 2 µl, 2 µl + 3 µl, 3 µl + 2 µl. The best crystals were obtained using 1.5 µl protein plus 2.5 µl of 22% PEG 4000 and 0.1 M MES buffer at pH 6.5. The crystals took about 8 weeks to grow to a maximum size of 0.18 x 0.18 x 0.05 mm³. A solution with 15% glycerol and the conditions of the mother liquor solution was used as cryo-protectant.

4.3.3. Data collection

X-ray diffraction data were collected on ESRF beam line ID29 using X-rays with wavelengths of 1.742 Å, 1.736 Å and 1.284 Å corresponding to the

inflection point and peak of the iron K absorption edge and a high energy remote point, respectively. For the three data sets images of 0.5° oscillation and 5 s exposure for a total of 112° were taken which resulted in diffraction patterns with reflections to better than $2.2 \text{ \AA}$ resolution. Data images were integrated with MOSFLM (Leslie 1992) and scaled, merged and converted to structure factors with SCALA and TRUNCATE (CCP4 1994). For structure refinement, data were collected using X-rays of $0.933 \text{ \AA}$ wavelength on ESRF beamline ID14-2 with 0.2° oscillation and 5 s exposure for a total of 107° giving a $1.35 \text{ \AA}$ resolution data set. Data images were integrated with DENZO and reduced and scaled with SCALEPACK (Otwinowski and Minor 1997). Data collection statistics are given in Table 4.1.

Table 4.1. Data collection statistics for anaerobically purified *D.vulgaris* HCP.

	Pi (inflection point)	Pk (peak)	Rm (remote)	High (native)
ESRF Beamline	ID29	ID29	ID29	ID14-2
Resolution limits ($\text{\AA}$) ⁷	38.20 – 2.20 (2.32 – 2.20)	38.20–2.20 (2.32–2.20)	47.40 – 2.20 (2.32–2.20)	67.40 – 1.35 (1.36–1.35)
Wavelength ($\text{\AA}$)	1.742	1.736	1.284	0.933
Space Group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit Cell Dimensions				
a ($\text{\AA}$)	64.3	64.3	64.1	64.27
b ($\text{\AA}$)	66.8	66.8	67.5	66.84
c ($\text{\AA}$)	134.8	134.9	135.0	134.76
Images x oscillation range ($^\circ$)	224 x 0.5	224 x 0.5	224 x 0.5	535 x 0.2
Total reflections	111349 (8684)	112006 (9145)	110922 (9242)	710927
Unique reflections	26707 (2546)	26842 (2657)	26296 (2501)	127 993
Multiplicity	4.2 (3.4)	4.2 (3.4)	4.2 (3.7)	5.6 (1.7)
R _{merge} ⁸	0.039 (0.062)	0.040 (0.062)	0.033 (0.039)	0.059 (0.35)
Wilson B factor ($\text{\AA}^2$)	-	-	-	17.7
Average I/ σ (I)	13.9 (9.3)	13.4 (9.0)	17.2 (14.9)	12.2 (3.5)
% Completeness	98.8 (98.8)	98.8 (98.8)	96.9 (92.4)	100.0 (100.0)

⁷ Numbers in parenthesis refer to the highest resolution shell.

⁸ $R_{\text{merge}} = \Sigma |I - \langle I \rangle| / \Sigma \langle I \rangle$;

The crystals belong to space group $P2_12_12_1$ with cell dimensions $a=64.27 \text{ \AA}$, $b=66.86 \text{ \AA}$, $c=134.76 \text{ \AA}$ and accommodate one molecule of molecular weight approximately 60 kDa in the asymmetric unit, corresponding to a V_m of $2.07 \text{ \AA}^3 \text{ Da}^{-1}$ (Matthews 1968) and a solvent content of 41 %.

4.3.4. Phase determination

The structure was determined by the multiwavelength anomalous diffraction (MAD) method using data sets collected to $2.2 \text{ \AA}$ resolution at the peak, inflection point and remote wavelengths for the iron K-absorption edge. The anomalous signal of the iron atoms within the structure was sufficient for automated interpretation of anomalous Patterson maps by SOLVE (Terwilliger 1994a; Terwilliger 1994b; Terwilliger and Berendzen 1999) which located six of the eight irons expected to be present in HCP. The locations of the remaining two irons were determined using the corresponding anomalous difference map calculated with the phases obtained from the six iron substructure. The eight iron positions were then recycled into SOLVE for further phase improvement.

4.3.5. Model construction and refinement

The phases obtained were extended, in steps, up to $1.35 \text{ \AA}$ resolution with DM (Cowtan 1994). Initial model building was performed automatically with ARP/wARP (Lamzin and Wilson 1993; Perrakis *et al.* 1999). After 50 cycles of refinement and 10 cycles of model building 525 residues out of a possible total of 553 were constructed with a global connectivity index of 0.91. At the end of the automatic model building the refinement converged to 28.5% and 30.2%, for R-factor and R-free respectively. The amino acid side chains were placed with the side dock script from the ARP/wARP suite. After this procedure, SigmaA-weighted (Read 1986) electron density $2F_o-F_c$

and Fo-Fc maps were inspected with XTALVIEW (McRee 1992) and the missing sections built. The iron-sulphur clusters were added with the help of the iron anomalous map to position the iron ions. The model obtained was then used to perform refinement with SHELXL (Sheldrick and Schneider 1997). Positional and isotropic *B*-factor parameters were initially refined for each atom. Alternative conformations for side chains were added where suggested by $|Fo| - |Fc|$ maps and finally accepted if the occupancy did not refine to below 0.25. Twelve parameters for an overall anisotropic *B*-factor were added (parameter HOPE in SHELXL) and when the refinement stabilised these were fixed. After this round of refinement, solvent molecules were added to the model based on standard geometrical and chemical restraints; they were subsequently deleted if were not visible at the 0.8 rms level in sigmaA-weighted $2|Fo| - |Fc|$ electron density maps. Refinement, using anisotropic *B*-factors on the iron atoms on a first step and later extended to all atoms, led to a decrease of 1.2 % in R and 2.0% in R-free (see Table 4.2). After several cycles of refinement and manual rebuilding where necessary the refinement converged to a final R-factor of 12.8% (115017 reflections) and R-free of 17.7% (6422) reflections).

Table 4.2. *R-factor progress through refinement.*

	R-free (%)	R-factor (%)
Arp-wARP (no irons) + side chains	25.2	23.5
Add 8 irons + 6 sulphurs	24.8	23.0
SHELXL (HOPE) + 200 waters	23.2	19.6
Add 789 waters + 11 alternative side chains	19.4	15.1
SHELXL (all <i>B</i> -factors anisotropic)	18.2	13.1
Final (more manual rebuilding)	17.7	12.8

4.3.6. Three-dimensional structure comparison.

All of the HCP and the carbon-monoxide dehydrogenase coordinate files deposited in the Protein Data Bank were aligned with the anaerobically purified *D. vulgaris* structure presented here, using the secondary structure matching algorithm (SSM) (Krissinel and Henrick 2004) implemented in COOT (Emsley and Cowtan 2004). This initial fit was further refined with Modeller (Sali and Blundell 1993) where residues are aligned by their C α coordinates and considered structurally homologous if within a threshold of 3.5 Å. Two residues of different structures are considered structurally identical if they are of the same residue type and structurally homologous.

4.4. Results and discussion

4.4.1. Quality of the structure

The stereochemistry of the final model was analysed with PROCHECK (Laskowski *et al.* 1993) giving an overall G factor of 0.09 (see Table 4.3 for detailed refinement statistics). The Ramachandran plot shows 93.7% (448) and 6.1% (19) of the amino acids in the most favourable regions and additional allowed regions respectively. As reported before, Asn303 (Macedo *et al.* 2002; Aragao *et al.* 2003) lies in the generously allowed regions but is supported by electron density. Information for the data set and associated structure has been deposited in the Protein Data Bank (Berman *et al.* 2000) with deposition code 1W9M.

Table 4.3. Final refinement statistics.

	<i>D. vulgaris</i>
PDB ID	1W9M
No. amino acids	1 x 553
No. protein atoms	4259
No. solvent atoms	789
No. other atoms	1 x S ₄ Fe ₄ + 1 x S ₂ Fe ₂ O ₃
Resolution limits (Å)	67.4 – 1.35
R-factor (No. observations)	0.128 (115017)
R-free (No. observations)	0.177 (6422)
Highest resolution shell (Å)	1.36 – 1.35
Cruickshank's DPI (Å)	0.048
Average B _{iso} (Å ²)	
All atoms	16.3
Protein atoms	13.6
Main chain	11.8
Side chain	15.4
Solvent	30.8
Residues with dual conformations	I35, D60, T173, E174, D189, V314, K338, P363, E366, N533 and I536.
RMS distance deviations:	
Bond lengths (Å)	0.012
Angles distances (°)	0.027
Distances from restraint planes (Å)	0.026

4.4.2. Structure

Overall Structure. Briefly, the protein comprises three domains: an N-terminal mainly alpha-helical domain (domain 1) and two similar domains comprising a central beta-sheet flanked by alpha-helices showing an alpha-beta Rossmann fold-like geometry (domains 2 and 3) (Figures 4.1 and 4.2). The protein contains two iron-sulphur clusters that are approximately 11 Å apart.

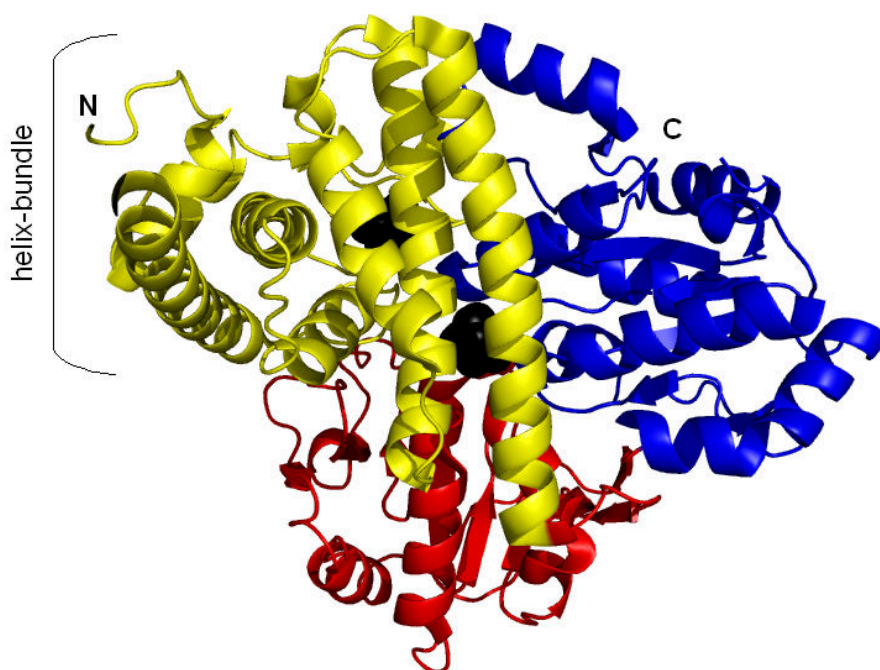


Figure 4.1. *The HCP monomer.*

Ribbon representation of anaerobically purified *D. vulgaris* HCP showing the cubane cluster and hybrid cluster as black spheres. The molecule is divided into three domains which are coloured in yellow (domain 1), red (domain 2) and blue (domain 3). The labelled helix bundle absent in carbon-monoxide dehydrogenase monomers but a similar helix-bundle is formed upon dimerisation.

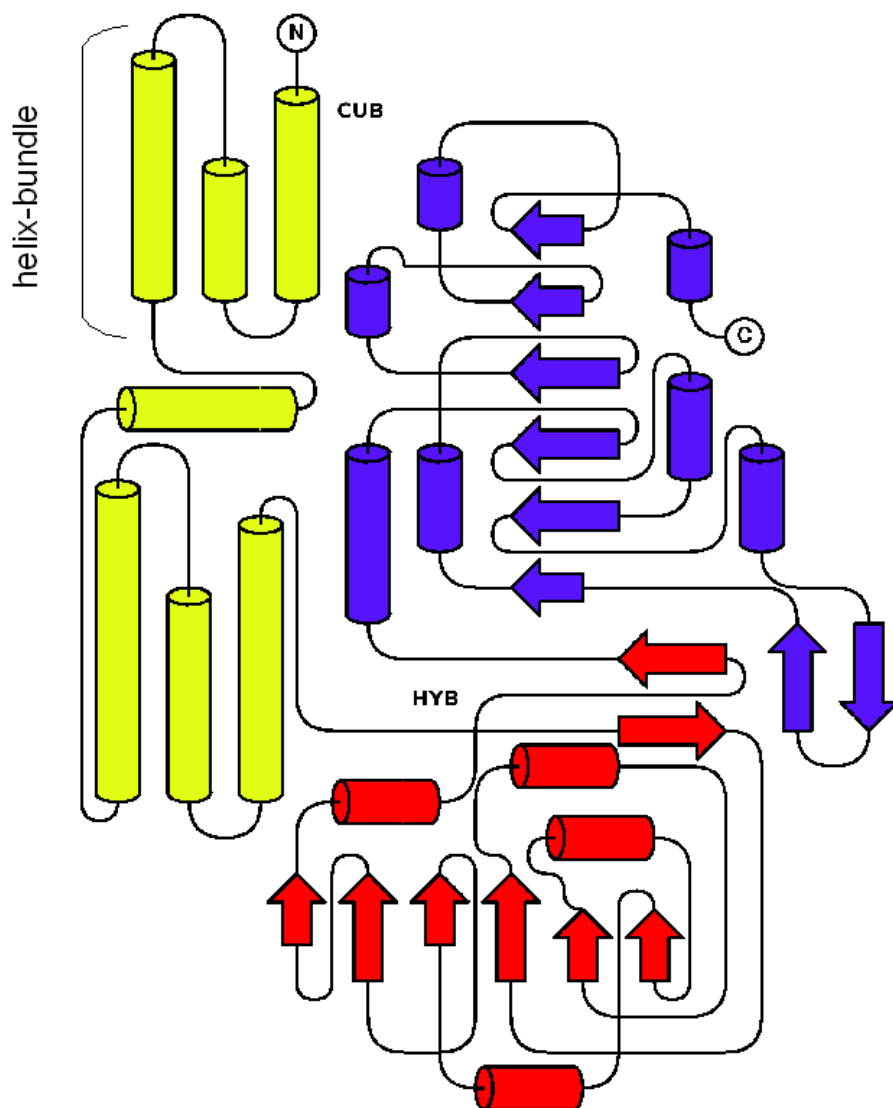


Figure 4.2. *Topology diagram of HCP.*

The same colour scheme as Figure 4.1 was used for domain identification. The labels CUB and HYB indicate the location of the cubane and hybrid cluster, respectively.

A conventional [4Fe-4S] cubane cluster binds through its irons to four cysteine residues at the N-terminus of the protein and a second unusual and

so-called hybrid cluster binds other four cysteine residues located near the geometrical centre of the protein and at the interface of the three domains. The hybrid cluster $[4\text{Fe-2S-3O}]$ is composed of a square-like base with two irons, FE5 and FE6, and two sulphur atoms, S5 and S6, with the irons bound to cysteine residues, Cys434 and Cys312, and a cap with two irons, FE7 and FE8, and three oxygen atoms, O8, O9 and O10, with FE8 iron bound to Cys406 through a per-sulphide bond and FE7 bound to a Glu494 (Figure 4.3).

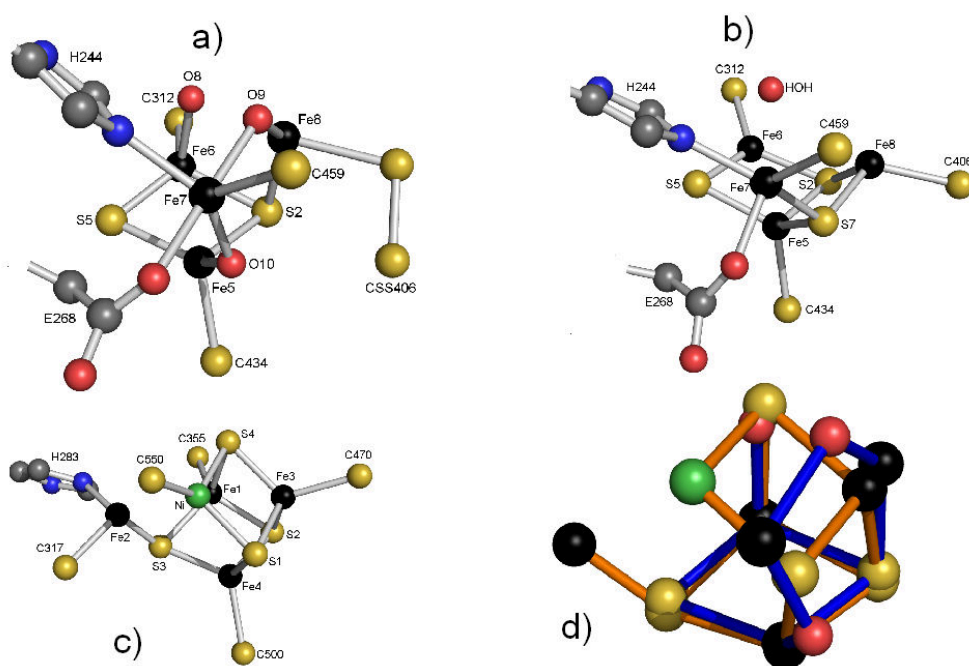


Figure 4.3. The hybrid clusters in HCP and CoD.

a) Model representation of the hybrid cluster in the *D. vulgaris* HCP anaerobically purified oxidised form (this work, PDB ID 1W9M). **b)** Model representation of the hybrid cluster in the *D. vulgaris* HCP dithionite reduced form (PDB ID 1OA1). The main chain and part of the side chain for residues binding the hybrid cluster, H244, E268, C312 C406 and C434 are omitted for clarity. **c)** Model representation of the C-cluster in the *Moorella thermoacetica* CoD (Doukov *et al.* 2002, PDB ID 1MJG). The main chain and part of the side chain for residues binding the hybrid cluster, H283, C355, C470, C500 and C550 are omitted for clarity. **d)** Superposition of the anaerobically purified HCP hybrid cluster (bonds in blue) and CoD C-cluster (bonds in orange). The ligands binding Fe5 (C434), Fe6 (C312) and Fe8 (C406) in HCP are structurally identical to ligands binding Fe4 (C500), Fe1 (C355) and Fe3 (C470) in the *Moorella thermoacetica* CoD C-cluster (see also Table 4.6 and Figure 4.7). Atom colouring: nickel (green), carbon (grey), nitrogen (blue), oxygen (red), sulphur (yellow) and iron (black).

To confirm the nature of atoms in the cluster, an anomalous difference map was calculated using data taken at the peak of the iron K-absorption edge. As

expected, all of the irons and the two sulphurs, S2 and S5, were visible in the anomalous difference map whilst oxygen atoms, O8, O9 and O10 were not (the oxygen anomalous signal is negligible at this wavelength). Moreover, these oxygen atoms refined to similar *B*-factor values as the atoms to which they are bound. In the initial *D. vulgaris* structure (Arendsen *et al.* 1998), aerobically prepared, O10 was modelled as a single oxygen atom. Later Cooper *et al.* (2000) proposed an oxygen atom disordered between two distinct positions around the co-ordinates of O10. Finally, structures at 1.25 Å resolution (Macedo *et al.* 2002) showed an elongated electron density peak between Fe5 and Fe7 for aerobically prepared *D. vulgaris*, whilst the anaerobically prepared *D. desulfuricans* structure showed a single position for this oxygen atom. It was previously reported that the major difference between *D. desulfuricans* HCP and *D. vulgaris* HCP lays in the moiety of O10, bound to Fe5 and Fe7 (Macedo *et al.* 2002; Aragao *et al.* 2003). The careful anaerobic protein preparation and crystallisation presented here shows that the hybrid cluster for *D. vulgaris* is not an oxidation artefact due to aerobic purification and crystallisation. The hybrid centre has four oxidation states, ranging from the most oxidised state (+6) containing four Fe(III) atoms to the fully reduced state (+3) with three Fe(II) and one Fe(III) atoms (Macedo *et al.* 2002). Upon successive one-electron reduction and starting from the fully oxidised protein, four redox states are detected for the hybrid cluster. It was shown that the anaerobically purified *D. desulfuricans* HCP is EPR silent and therefore suggested to have its hybrid cluster in the +4 oxidation state (Macedo *et al.* 2002) and had a single oxygen position for O10. Here it is shown that when anaerobic purified *D. vulgaris* HCP also shows also a single position for oxygen. The hybrid cluster for the aerobically purified *D. vulgaris* HCP had an oxygen disordered in two positions and was in +5 oxidation state (Arendsen *et al.* 1998; Macedo *et al.* 2002). If so, these results suggest that the presence or absence of an

elongated electron density in the O10 position is a result of the different oxidation state in the hybrid cluster and this might be important for the elucidation of the catalytic role of the HCPs.

Comparison with other structures. It has been reported that both *D. vulgaris* and *D. desulfuricans* strains have carbon monoxide dehydrogenase activity (Meyer and Rohde 1984) and CoD share both fold and hybrid cluster similarity (Macedo *et al.* 2002) which makes pertinent to compare both structures in detail. Briefly, CoD are present in three different arrangements: monomeric (*Carboxydotherrmus hydrogenoformans*), dimeric (*Rhodospirillum rubrum*) and tetrameric bifunctional (*Moorella thermoacetica*). The tetrameric form is an $\alpha_2\beta_2$ heterotetramer that catalyses two different functions, the β -subunit generates CO from CO₂ while in the α -subunit CO is combined with CoA and a methyl group to form acetyl-CoA (Ragsdale and Wood 1985; Lu *et al.* 1990). The β -subunit has the same arrangement as the dimeric CoD structure. From hereafter only the CoD structures of the monomeric, dimeric and the β -subunit of the tetrameric form will be referred to since the α -subunit is unrelated to HCPs. The CoD monomer comprises three domains: one domain, containing a three helix bundle, and two other similar domains comprising a central beta-sheet flanked by alpha-helices (Figure 4.6). CoDs have one cubane-like iron-sulphur cluster (B-cluster) and one variable iron-sulphur cluster (D-cluster) bound to the first domain, and one nickel containing hybrid iron-sulphur cluster at the interface of the three domains (C-cluster).

The *D. vulgaris* and *D. desulfuricans* HCP structures share over 60% amino-acid identity as based on their 3D alignment (Figure 4.4), whilst HCPs and

carbon-monoxide dehydrogenases (CoD) share 10% or less structural amino-acid identity (Table 4.4, Figure 4.7).

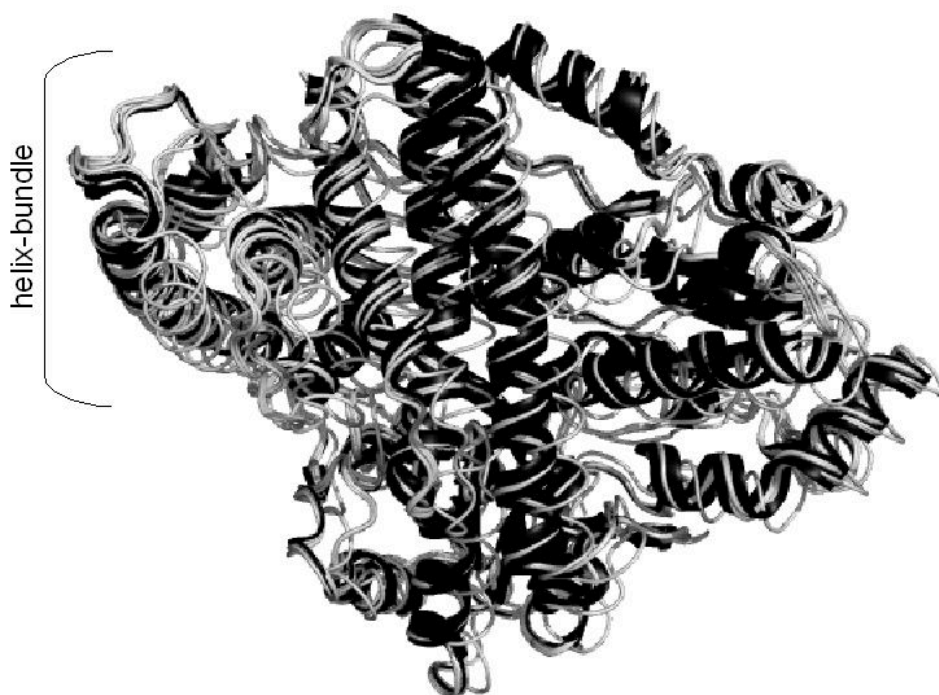


Figure 4.4. The superposition of *D. vulgaris* and *D. desulfuricans* HCPs. Superposition of anaerobically purified *D. vulgaris* in black ribbon (this work, PDB ID 1W9M) and *D. desulfuricans* HCP in grey coil (PDB IDs 1UPX, 1GN9, 1GNL and 1OA0). These observations are consistent with the fact that the anaerobically purified *D. vulgaris* HCP structure, presented here, has a maximum root mean square difference of 0.4, 0.8 and 1.9 Å for *D. vulgaris* HCPs, *D. desulfuricans* HCPs and CoDs respectively (Table 4.5). However, remarkably, the overall folds of HCPs and CoDs are similar (Figures 4.2, 4.6, 4.4 and 4.5) and their superposition shows about 59% of structurally homologous residues between HCPs and CoD (Table 4.4).



Figure 4.5. *The superposition of HCP with CoDs.*

Superposition of anaerobically purified *D. vulgaris* HCP (black ribbon) and carbon-monoxide dehydrogenase structures in grey (PDB IDs 1SU7, 1SU8, 1SUF for *C. hydrogenoformans*, 1JQK for *R. rubrum*, 1MJG and 1OAO for *M. thermoacetica*) showing their overall similar fold.

	1E1D	1E2U	1E9V	1GNT	1OA1	1W9M	1UPX	1GN9	1GNL	1OA0	1SU6	1SU7	1SU8	1SUF	1JQK	1MJG	1OAO
1E1D	-	100	100	100	100	100	95	95	95	95	58	58	59	59	58	57	58
1E2U	100	-	100	100	100	100	95	95	95	95	58	58	59	59	58	57	58
1E9V	99	99	-	100	100	100	95	95	95	95	59	59	59	59	59	58	58
1GNT	99	99	100	-	100	100	95	95	95	95	59	59	59	59	59	58	58
1OA1	99	99	100	100	-	100	95	95	95	95	59	59	59	59	59	58	58
1W9M	99	99	100	100	100	-	95	95	95	95	59	59	59	59	59	57	58
1UPX	65	65	65	66	66	65	-	99	99	100	59	59	60	60	59	59	59
1GN9	66	66	66	66	66	66	100	-	100	100	60	60	60	60	59	59	59
1GNL	66	66	66	66	66	66	100	100	-	100	60	60	60	60	59	59	59
1OA0	65	65	65	66	66	65	100	100	100	-	60	59	60	59	59	58	58
1SU6	10	10	10	10	10	10	10	10	11	10	-	100	100	100	95	97	97
1SU7	10	10	10	10	10	10	10	10	11	10	100	-	100	100	95	97	97
1SU8	10	10	10	10	10	10	10	10	11	10	100	100	-	100	95	97	97
1SUF	10	10	10	10	10	10	10	10	11	10	100	100	100	-	95	97	97
1JQK	10	10	10	10	10	10	11	11	11	11	57	57	57	57	-	98	98
1MJG	9	9	9	9	9	9	8	8	8	8	49	49	49	49	47	-	100
1OAO	9	9	9	9	9	9	8	8	8	8	49	49	49	49	47	100	-

Table 4.4. Structure similarity and identity between HCPs and CoDs.

Values above the diagonal represent the percentage of structure similarity between proteins (an amino-acid is considered structurally homologous to another amino-acid when, after Ca protein to protein fit, their Ca are at 3.5 Å or less from each other). Values below the diagonal represent the percentage of structural homology between proteins (an amino-acid is considered structurally identical when it is structurally homologous and of the same type of amino-acid). The *D. vulgaris* HCP structures are shown in dark grey (PDB IDs: 1E1D, 1E2U, 1E9V, 1GNT, 1OA1 and 1W9M), the *D. desulfuricans* HCP structures in light grey (PDB IDs: 1UPX, 1GN9, 1GNL and 1OA0), *C. hydrogenoformans* CoD structures in darker orange (PDB IDs: 1SU6, 1SU7, 1SU8 and 1SUF), the *R. rubrum* CoD structure in orange (PDB ID 1JQK) and the *M. thermoacetica* CoD structures in light orange (PDB IDs: 1MJG and 1OAO).

	1E1D	1E2U	1E9V	1GNT	1OA1	1W9M	1UPX	1GN9	1GNL	1OA0	1SU6	1SU7	1SU8	1SUF	1JQK	1MJG	1OAO
1E1D	-	0	0,3	0,3	0,4	0,4	0,9	0,8	0,8	0,8	1,9	1,9	1,9	1,9	1,9	1,9	1,9
1E2U	-	-	0,3	0,3	0,4	0,4	0,9	0,8	0,8	0,8	1,9	1,9	1,9	1,9	1,9	1,9	1,9
1E9V	-	-	-	0,1	0,3	0,2	0,8	0,8	0,8	0,8	1,9	1,9	1,9	1,9	1,9	1,9	1,9
1GNT	-	-	-	-	0,3	0,3	0,8	0,8	0,8	0,8	1,9	1,9	1,9	1,9	1,9	1,9	1,9
1OA1	-	-	-	-	-	0,2	0,7	0,7	0,7	0,7	1,9	1,9	1,9	1,9	1,9	1,9	1,9
1W9M	-	-	-	-	-	-	0,8	0,8	0,7	0,8	1,9	1,9	1,9	1,9	1,9	1,9	1,9
1UPX	-	-	-	-	-	-	-	0,2	0,1	0,2	1,9	1,9	1,9	1,9	1,9	1,9	1,9
1GN9	-	-	-	-	-	-	-	-	0,2	0,2	1,9	1,9	1,9	1,9	1,9	1,9	0,0
1GNL	-	-	-	-	-	-	-	-	-	0,2	1,9	1,9	1,9	1,9	1,9	1,9	1,9
1OA0	-	-	-	-	-	-	-	-	-	-	1,9	1,9	1,9	1,9	1,9	1,9	1,9
1SU6	-	-	-	-	-	-	-	-	-	-	-	0,3	0,1	0,1	0,8	0,8	0,8
1SU7	-	-	-	-	-	-	-	-	-	-	-	-	0,2	0,3	0,8	0,8	0,8
1SU8	-	-	-	-	-	-	-	-	-	-	-	-	-	0,2	0,8	0,8	0,8
1SUF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,8	0,8	0,8
1JQK	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,9	1,0
1MJG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,3
1OAO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 4.5. Root mean square (RMS) difference between between HCPs and CoDs homologous C-alpha.

Same colour code as in Table 4.4.

Domain 1 is responsible for the binding of one and two [Fe₄-S₄] cubane clusters in HCP and CoD, respectively. Although the CoD domain 1 lacks the first 3-helix bundle, responsible in HCP for the binding of its cubane cluster, this bundle is replaced in multimeric CoDs (*Rhodospirillum rubrum* dimeric CoD, β-subunit of the bifunctional CoD/Acetyl-CoA synthase in

Moorella thermoacetica) by an equivalent helix bundle from a different monomer in order to create a similar arrangement (Figure 4.6).

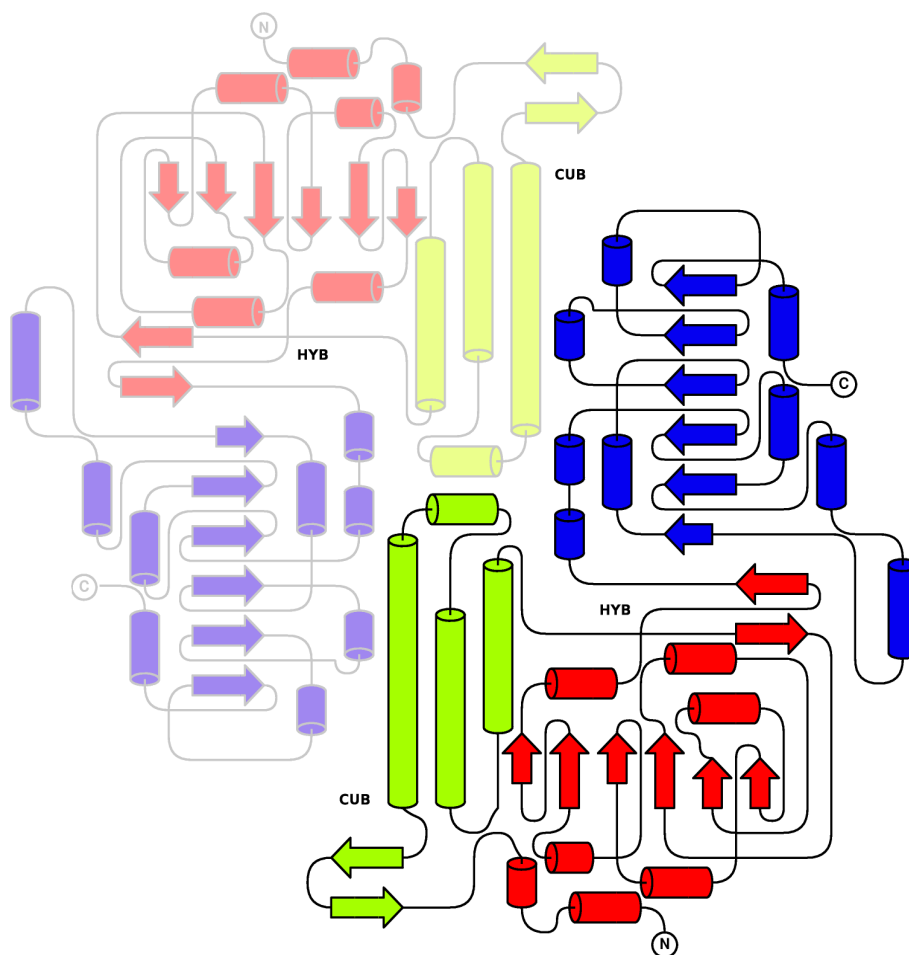


Figure 4.6. Topology diagram for a dimer of the beta-subunit of CoD. The same colour scheme as in Figure 4.1 and Figure 4.2 was used for domain identification and comparison with HCPs. The labels CUB and HYB indicate the location of the CoD B- and C-clusters respectively. The CoD D-cluster is located at the two monomers interface and was omitted for clarity.

Domains 2 and 3 contribute with ligands to the [4Fe-2S-3O] hybrid cluster in

HCP and to the [Ni-4Fe-4S] C-cluster in the CoD. The CoD C-cluster can be interpreted as a [3Fe-4S] subsite and a [Ni-Fe] subsite. Two of the three facial sulphides of the [Fe₃-S₄] subsite coordinate the Ni whilst the third coordinates the unique Fe of the [NiFe] subsite (Doukov *et al.* 2002; Darnault *et al.* 2003; Kim, Feng *et al.* 2004; Volbeda and Fontecilla-Camps 2005). The fit of the two hybrid clusters after superposition of both molecules, HCP and COD, shows that 6 out of 9 atoms overlap (Figure 4.3): HCP atoms Fe5, Fe6, Fe8, O8, S5 and S6 (Fe4, Fe1, Fe3, S4, S3 and S2 in *Moorella thermoacetica*, CoD numbering (Doukov *et al.* 2002)). Iron atoms Fe5, Fe6 and Fe8 (HCP numbering) are bound by structurally identical cysteine residues (Figure 4.7, Table 4.6). The equivalent position to O10 of HCP is occupied by a sulphur atom in CoD. This is the configuration found in reduced HCP structures (Aragao *et al.* 2003). CoD structural studies have clearly shown catalysis to occur on the top and mobile part of the cluster in the vicinity of Ni and Fe2 (Darnault *et al.* 2003; Lindahl 2004; Volbeda and Fontecilla-Camps 2005). The positions of the Ni and Fe2 atoms in CoD are vacant in HCP structures.

[illegible]

[illegible]

```

551      -MA-----
551      -MA-----
550      -MA-----
551      -MA-----
553      -MAGK-----
552      -MAGK-----
544      -FG-----
540      -FG-----
541      -FG-----
544      -FG-----
633      A--AINERRAGLG--LPW
633      A--AINERRAGLG--LPW
633      A--AINERRAGLG--LPW
633      A--AINERRAGLG--LPW
610      A--AIQERRAGL-----
672      D---ALEYRTWKLGHVHKEVAERYETKLCQGY---
673      D---ALEYRTWKLGHVHKEVAERYETKLCQGY---
1E1D      -MA-----
1E2U      -MA-----
1E9V      -MA-----
1GNT      -MA-----
1OA1      -MAGK-----
1W9M      -MAGK-----
1UPX      -FG-----
1GN9      -FG-----
1GNL      -FG-----
1OA0      -FG-----
1SU6      A--AINERRAGLG--LPW
1SU7      A--AINERRAGLG--LPW
1SU8      A--AINERRAGLG--LPW
1SUF      A--AINERRAGLG--LPW
1JQK      A--AIQERRAGL-----
1MJG      D---ALEYRTWKLGHVHKEVAERYETKLCQGY---
1OAO      D---ALEYRTWKLGHVHKEVAERYETKLCQGY---

```

Figure 4.7. Three-dimensional structure alignment.

Alignment of all available HCP and CoD 3D structures, upon superposition using Modeller (Sali and Blundell 1993). Residues boxed in black and labelled 1-6 bind to the hybrid cluster and the C-cluster in HCP and CoD, respectively (see Table 4.6 for residue numbering). The particular helix-bundle discussed in the text is highlighted in blue. Label A groups HCP PDB codes 1E1D, 1E2U, 1E9V, 1GNT, 1OA1, 1W9M from *D. vulgaris*. Label B groups 1UPX, 1GN9 and 1GNL from *D. desulfuricans*. Label C groups CoD PDB codes 1SU6, 1SU7, 1SU8 AND 1SUF from *C. hydrogenoformans*. Label D shows CoD PDB code 1JQK from *R. rubrum* and label E groups CoD PDB codes 1MJG and 1OAO from *M. thermoacetica*.

Table 4.6. Structurally homologous residues that bind the HCP hybrid cluster and the CoD C-cluster.

	HCP		COD		
	<i>D. vulgaris</i> (A)	<i>D. desulfuricans</i> (B)	<i>C. hydrogenoformans</i> (C)	<i>R. rubrum</i> (D)	<i>M. thermoacetica</i> (E)
1	His 244	His 240	His 261	His 265	His 283
2	Glu 268	Glu 264	Cys 295	Cys 300	Cys 317
3	Cys 312	Cys 308	Cys 333	Cys 338	Cys 355
4	Cys 406	Cys 399	Cys 446	Cys 451	Cys 470
5	Cys 434	Cys 427	Cys 476	Cys 481	Cys 500
6	Cys 459	Cys 452	Cys 526	Cys 531	Cys 550

Label A groups HCP PDB codes 1E1D, 1E2U, 1E9V, 1GNT, 1OA1, 1W9M from *D. vulgaris*. Label B groups 1UPX, 1GN9 and 1GNL from *D. desulfuricans*. Label C groups CoD PDB codes 1SU6, 1SU7, 1SU8 AND 1SUF from *C. hydrogenoformans*. Label D shows CoD PDB code 1JQK from *R. rubrum* and label E groups CoD PDB codes 1MJG and 1OAO from *M. thermoacetica*.

and 1OAO from *M. thermoacetica*. For residues labelled 1-6 refer to Figure 4.7.

This structural similarity between HCP and CoD may hint for some degree of similarity in their catalytic mechanism. In HCP this would therefore involve the substrate binding to the hybrid cluster reduced structure in the vicinity of the water molecule shown in Figure 4.3b and forming hydrogen bonds with the conserved residues Asn311 (Gln in CoDs) and Lys496 (Aragao et al 2003). In the case of a dioxygen containing substrate reduction would yield two hydroxyl species, giving rise to the oxygen positions O8 and O9, as in the anaerobically purified oxidised form (Figure 4.3a). Sequential release of these would lead to a water product in the O10 position. The product could then diffuse away from the active site and eventually away from the enzyme through the extensive hydrophobic cavity that gives ready access to the hybrid cluster (Cooper *et al.* 2000).

4.5. Conclusions

The iron MAD phasing allowed the complete structure elucidation of the anaerobically purified HCP protein and its iron-sulphur cluster. Structures of anaerobically purified HCP from both *D. desulfuricans* (Macedo *et al.* 2002) and *D. vulgaris* (this study) show a single oxygen position for O10 whilst structures determined using samples prepared aerobically have an elongated density at the same position. In some cases this density has been modelled as two half occupied oxygen positions.

The main sequence differences between these HCP and CoD proteins lie in the N-terminal domain where the cubane cluster (B-cluster in CoDs) binding region is apparently absent for CoD (Figure 4.7). However, the structures of dimeric and tetrameric CoDs show that this region is present, via monomer B, and has the same fold as in HCP (Figure 4.6). Both HCPs and CoDs have intriguing and unusual clusters at the interface of their three domains. These clusters, although different, share striking similarities and are linked to the

polypeptide chain via homologous ligands. The HCP reaction mechanism may also be similar to CoD reaction mechanism with substrate binding to the top mobile part of the cluster followed by its reduction and release of an oxygen ion or hydroxyl group. This mechanism would fit both hydroxylamine reduction and hydrogen peroxide reduction as well as other small or linear substrates with potential to be reduced. In light of current knowledge, the HCP substrate could be either a reactive nitrogen or oxygen species and further biochemical studies are needed to access and understand the true nature of the substrate for this enzyme in order to clarify its role in either protection against oxidative stress or reactive nitrogen stress.

4.6. Acknowledgements

The purification of *D. vulgaris* HCP was performed by Célia V. Romão, Dora Alves and David Aragão. The crystallisation, data collection, phase determination, structure refinement, structural analysis and comparison with CoDs were made by David Aragão. The staff of ESRF stations ID-14 and ID-29 are acknowledge for the use of synchrotron radiation in diffraction data collection.

5. Cloning, expression, purification, crystallisation and preliminary structure determination of Glucose-1-phosphate uridylyltransferase (UgpG) bound to glucose-1-phosphate from *Sphingomonas elodea* ATCC31461

(Based on a published refereed paper but now including a more detailed description of relevant points on the materials and methods with addition of new figures: D. Aragão, A. R. Marques, C. Frazão, F. J. Enguita, M. A. Carrondo, A. M. Fialho, Sá-Correia and E. P. Mitchell. *Acta Crystallograph Sect F Struct Biol Cryst Commun.* (2006). 1;F62:930-4)

5.1. Abstract

We report here the cloning, expression, purification, crystallisation and preliminary crystallographic analysis of glucose-1-phosphate uridylyltransferase (UgpG) from *Sphingomonas elodea* ATCC 31461 bound to glucose-1-phosphate. Diffraction data sets were obtained from seven crystal forms, within five different space groups, with highest resolutions ranging from 4.20 Å to 2.65 Å. The phase problem was solved for a P2₁ crystal form using multiple-isomorphous replacement with anomalous scattering from an osmium and a Se-Met derivatives. The best native crystal in space group P2₁ has cell dimensions $a = 105.5\text{Å}$, $b = 85.7\text{Å}$, $c = 151.8\text{Å}$ and $\beta = 105.2^\circ$. Model building and refinement are currently underway.

5.2. Introduction

Glucose-1-phosphate uridylyltransferases ([EC 2.7.7.9](#)) are enzymes that mediate the reversible conversion of glucose-1-phosphate (G1P) and UTP into UDP-glucose and pyrophosphate. These proteins are also known as pyrophosphorylases, based on their ability to catalyse the reverse reaction. This enzymatic activity is present in all living organisms, although prokaryotic enzymes diverge significantly from those of eukaryotic origin (Daran *et al.* 1995; Eimert *et al.* 1996). The glucose-1-phosphate uridylyltransferases of prokaryotic organisms are essential to the synthesis of several polysaccharides, many of them involved in the pathogenesis of those organisms. In fact, glucose-1-phosphate uridylyltransferases are required for the synthesis of the virulence factor capsular polysaccharide (CPS) in *Streptococcus pneumoniae* (Bonofiglio *et al.* 2005), and have been suggested to be required for corneal infection by *Pseudomonas aeruginosa* and to be necessary for efficient systemic spread following lung infection (Priebe *et al.* 2004). These findings have prompted an increasing interest in prokaryotic glucose-1-phosphate uridylyltransferases enzymes as potential drug targets (Bonofiglio *et al.* 2005).

The UDP-glucose produced by glucose-1-phosphate uridylyltransferase can be interconverted into other sugar nucleotides or be used as the glucosyl donor in the biosynthesis of various cellular carbohydrates, such as exopolysaccharides (EPS), lipopolysaccharides (LPS), capsular polysaccharides (CPS) and storage compounds like trehalose. Glucose-1-phosphate uridylyltransferases are also involved in the catabolism of galactose into glucose in the LeLoir pathway (Frey 1996).

In this work we report the crystallisation of a glucose-1-phosphate

uridylyltransferase (commonly referred to as UgpG) from *Sphingomonas elodea* ATCC 31461 (Marques *et al.* 2003). This bacterium produces EPS gellan gum in high yield. Gellan gum produced by this method has been approved for use as a gelling and suspending agent in the USA and EU. In its native form gellan is a linear anionic heteropolysaccharide based on a tetrasaccharide repeat unit composed of two molecules of D-glucose, one of D-glucuronic acid and one of L-rhamnose. The native gellan is partially esterified with acyl substituents: 1 mole of glycerate and 0.5 mol of acetate per repeat unit (Jay *et al.* 1988). Gellan biosynthesis starts with the intracellular synthesis of the nucleotide-sugar precursors: UDP-glucose, UDP-glucuronic acid and dTDP-L-rhamnose. The repeat unit is then constructed by sequential transfer of the sugar donors to an activated lipid carrier by committed glycosyltransferases, followed by gellan polymerisation and export (Sá-Correia *et al.* 2002).

In fact, UgpG recognises both uridine (UTP) and deoxythymidine (dTTP) nucleotides substrates *in vitro* (Silva *et al.*, 2005). However, *in vivo* a different enzyme, glucose-1-phosphate thymidylyltransferase (EC 2.7.7.24), is responsible for the formation of dTDP-glucose.

Mutation studies, sequence comparisons and the four available glucose-1-phosphate thymidylyltransferase structures from *E. coli* (Blankenfeldt, Asuncion *et al.* 2000; Blankenfeldt, Giraud *et al.* 2000), *P. aeruginosa* (Sivaraman *et al.* 2002), *S. enterica* (Barton *et al.* 2001; Barton *et al.* 2002) and *M. thermoautotrophicum* (PDB ID: 1LVW) have identified the possible key residues for substrate (G1P, dTTP/UTP) and product (TDP-glucose/UDP-glucose) binding, but there is still no report of any glucose-1-phosphate uridylyltransferase structure.

5.3. Materials and Methods

5.3.1. Cloning and expression of UgpG

The complete sequence of the *ugpG* gene from *Sphingomonas elodea* ATCC 31461 was amplified by PCR using the oligonucleotides PUGP19 (5'-AAAGGATCCATGACGATCAAGC-3') and PUGP29 (5'-AAAAAGCTTTCAGCCGAGCGCCTT-3') and cosmid pC22 (16) as the template DNA. The PCR product (888 bp) was digested with *Bam*HI and *Hind*III (underlined in the primers used) and cloned into compatible sites of pWH844 (Schirmer *et al.* 1997), generating pUgpG2. This recombinant plasmid carries the *ugpG* gene preceded by a sequence coding for six histidine residues followed by two residues, a glycine and a serine, before the initiating methionine. The insert cloned in pUgpG2 was then sequenced to confirm the fidelity of DNA amplification.

To over express the *ugpG* gene, transformants of *E. coli* BL21 Gold carrying pUgpG2 were cultivated at 298 K in 4x100 ml LB medium until an OD_{640nm} of 0.6 ± 0.1 . The cells were then induced with 0.1 mM isopropyl β -D-thiogalactoside (IPTG) for 3 h, harvested by centrifugation (5000 x g, 10 minutes, 277 K) and stored at 253 K overnight. SDS-Page gels for the insoluble fraction (position 8, Figure 5.1a upper panel) and different induction times (positions 1, 2 and 3, Figure 5.1a lower panel) showed that the protein is greatly soluble and has a good yield for 3 h induction. For the production of the SeMet-UgpG derivative protein (Molecular Dimensions protocol), methionine auxotrophic strain *E.coli* B843 (DE3) harbouring pUgpG2 was cultivated at 298 K in 4 x 100 ml LB until an OD_{640nm} of 0.6 ± 0.1 . Cells were then harvested by centrifugation (5000 x g, 10 min, 278 K) and gently resuspended in SelenoMet Medium Base™ (Molecular Dimensions, UK). This process was repeated three times to remove traces of LB containing methionine. Finally, selenomethionine solution (Molecular Dimensions, UK) 250x concentrated was added and the cells were induced for 5 h with 0.1 mM IPTG.

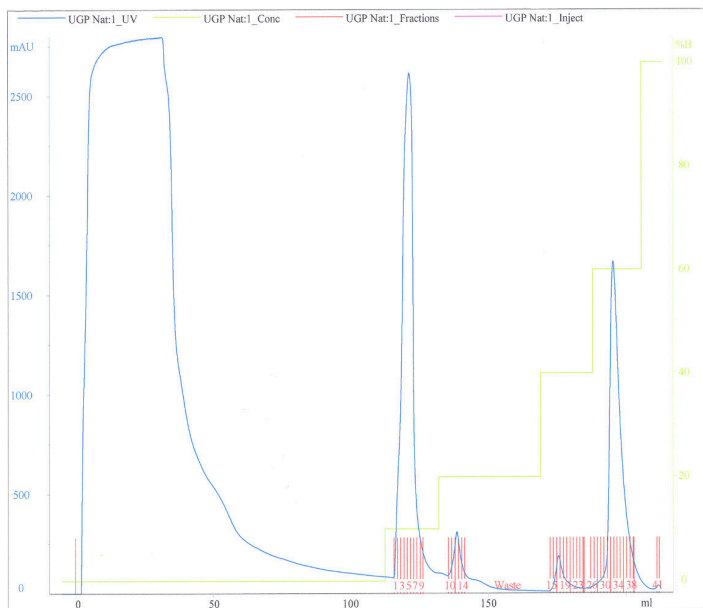
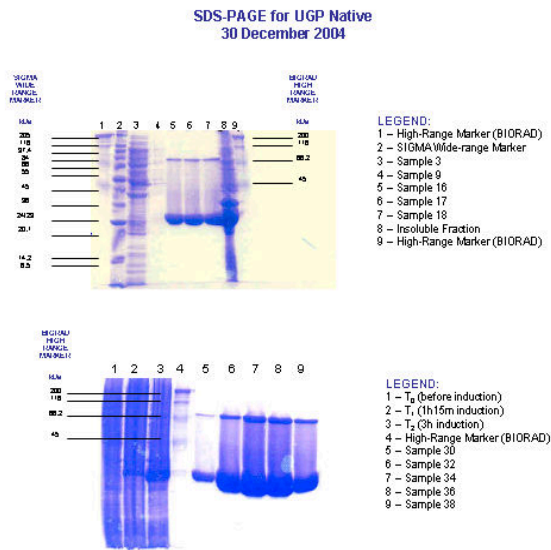


Figure 5.1. *UgpG* expression and purification.

a) Native UgpG expression and purification gels. b) Chromatogram of the UgpG native protein purification starting from crude cell extract applied to a 5 ml HisTrap column. The protein eluted in fractions 30-38 was almost pure and in a symmetric chromatographic peak.

5.3.2. Purification of UgpG

Cells were resuspended in 50 mM Tris-HCl pH 8.0, 10 mM imidazole, 50 mM NaCl and disrupted in a French Press. Crude cell extract was obtained by centrifugation at 17000 x g for 25 min and the supernatant applied to a 5 ml HisTrap column (Amersham Pharmacia, UK) pre-equilibrated with 50 mM Tris-HCl pH 8.0, 10 mM imidazole, 50 mM NaCl. The column was washed in steps with 10 column volumes of buffer A (50 mM Tris-HCl pH 8.0, 10 mM imidazole, 500 mM sodium chloride), 5 column volumes 10% buffer B (50 mM Tris-HCl pH 8.0, 500 mM imidazole, 500 mM sodium chloride), 10 column volumes 20% buffer B and 5 column volumes with 40% buffer B to remove any unbound proteins. Then, two final steps with 60% buffer B and 100% buffer B were applied. UgpG was eluted at approximately 300 mM imidazole in a symmetrical chromatography peak (SDS-PAGE gel in Figure 5.1a, chromatogram in Figure 5.1b,) with an absorbance for the protein peak of approximately 1700 mAU (280 nm wavelength). The eluted fractions were immediately transferred into 75 mM Tris-HCl buffer pH 8.0 and 40 mM sodium chloride using a PD10 desalting column (Amersham Pharmacia, UK) and the protein concentrated to 10 mg/ml before storage at 253 K.

The SeMet-UgpG protein was purified, concentrated and stored in a similar way as described above. However, this protein form, eluted at lower imidazole (approximately 200 mM) and gives a lower yield (260 mAU absorbance at 280 nm wavelength for the protein peak) for same volume amount of starting bacterial cells (Figure 5.2). MALDI-TOF mass spectrometry confirmed 100% selenomethionine incorporation and a molecular weight of 32 kDa in the UgpG produced by this method.

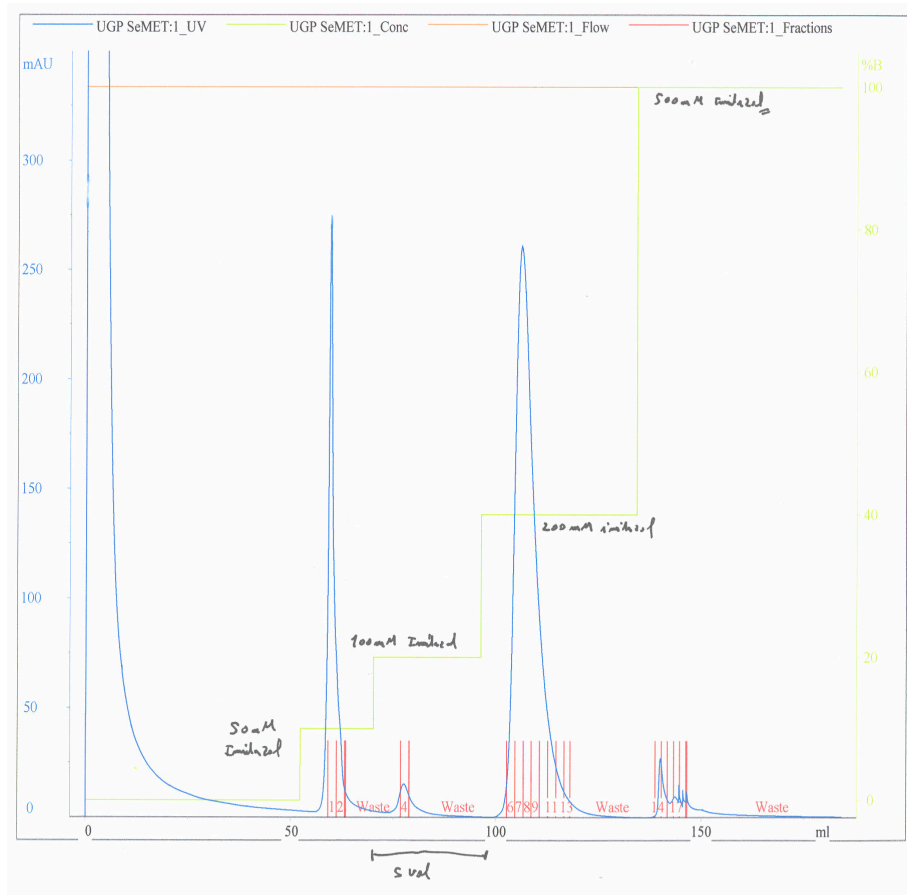


Figure 5.2. *UgpG selenium-methionine protein purification.*

Chromatogram of UgpG selenium-methionine protein purification starting from crude cell extract applied to a 5 ml HisTrap column

The purity of both native and Se-labelled proteins was confirmed by SDS-PAGE (not shown). Protein concentrations were determined by the method of Bradford (Bradford, 1976) using bovine serum albumin fraction V (Sigma) as the standard.

Gel filtration chromatography (Figure 5.3) and dynamic light scattering (not shown) confirmed the protein was stable and mono-disperse at the working conditions. Both techniques show UgpG to have a molecular mass between 130 and 140 kDa in solution.

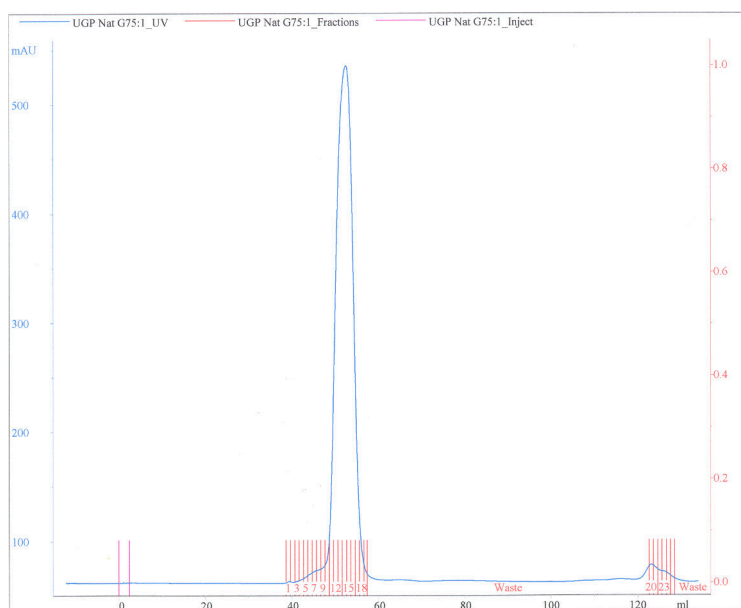


Figure 5.3. *UgpG native protein purification.*

Gel filtration chromatogram, the the second step UgpG native protein purification, at pH = 8.0. The protein elutes in a symmetric and unique peak at about 55 ml (corresponding after column calibration to weight of approximately 133 kDa).

5.3.3. Crystallisation of UgpG protein

Crystallisation screens with native protein were performed on a TECAN Genesis robot using the vapour diffusion method. Drops with 1 μ l native protein solution (75 mM Tris-HCl pH 8.0, 40 mM sodium chloride) at 5 or 10 mg/ml and 1 μ l precipitant solution were equilibrated against 576 conditions based on crystallisation screens: Crystal Screen, Crystal Screen 2, Crystal Screen Lite, Peg/ion Screen, Quick Screen, Index, Crystal Screen Ammonium Sulphate, Screen Malonate, Screen Formate, PEG 6K, PEG LiCl, MPD and MME 5000 from Hampton Research (Aliso Viejo, US). Crystalline material was found in several of these screens. In Crystal Screen Lite, 5 drops had crystals that were visible through a microscope. From those, solution 9 was the most promising in terms of diffraction and was subsequently refined (crystallisation Condition A1, see Table 5.1).

5.1: Table 5.1. Crystallisation conditions for UgpG..

Crystal	Crystallisation conditions				Co-crystallisation	Drop size/ratio	Crystal size	Time to grow	Temp.	Cryoprotectant	Observations
	Salt/Alcohol	Buffer	Precipitant	Protein + precipitant (µl)							
A1	0.2M ammonium acetate	0.1M sodium citrate pH=4.6	33% PEG 4K	-	1+1	0.02 x 0.02 x 0.08	15	281	25% v/v glycerol/mother liquor		-
A2	20% 2-propanol	0.1M sodium citrate pH=6.0	20% PEG6k	-	2+2	0.03 x 0.03 x 0.10	15	291	25% v/v glycerol/mother liquor		Macroseeding
B*	0.1M ammonium acetate	0.1M sodium citrate pH=5.0	31% PEG 4k	-	3+3	0.05 x 0.05 x 0.10	21	277	25% v/v glycerol/mother liquor		SeMet protein
C	1.6M ammonium sulphate	0.1M bicine pH=9.0	-	-	3+2	0.06 x 0.03 x 0.03	4	293	35% w/v sucrose/mother liquor		-
D	1.6M ammonium sulphate	0.1M bicine pH=9.0	-	-	3+2	0.08 x 0.04 x 0.04	7	293	35% w/v sucrose/mother liquor		-
E*	-	0.1M sodium citrate pH=5.6	10% PEG MME	+0.5 µl dioxane/drop	5+4.5	0.12 x 0.08 x 0.03	2	293	35% w/v sucrose/mother liquor		0.1% LM agarose gel, SeMet protein
F	-	0.1M sodium citrate pH=6.0	15% PEG 4K	10mM UDP-Glc	2+2	0.02 x 0.03 x 0.05	15	293	35% w/v sucrose/mother liquor		-
G1**	0.1M ammonium acetate	0.1M sodium citrate pH=4.6	15% PEG MME	5mM GIP	5+2	0.20 x 0.05 x 0.05	15	293	Not needed		30 min soak in 10mM K ₂ OsO ₄
G2*	0.1M ammonium acetate	0.1M sodium citrate pH=4.6	13% PEG MME	5mM GIP	5+2	0.10 x 0.04 x 0.04	15	293	Not needed		SeMet protein
G3	0.1M ammonium acetate	0.1M sodium citrate pH=4.6	15% PEG MME	5mM GIP	5+2	0.20 x 0.05 x 0.05	15	293	Not needed		-

In Crystal Screen I, solution 40 was optimised by the use of macroseeding and temperature screens between 277 K and 308 K which led to larger needles (Condition A2, 5.1). Since the initial screens showed amorphous precipitate in numerous conditions, lower buffer and salt in the protein solution (50 mM Tris-HCl pH 7.6, 30 mM sodium chloride) were used hereafter. A further set of crystallisation screens, this time for both native and selenated proteins, were performed using the vapour diffusion method as implemented on the EMBL Cartesian HTX platform, which led to over 30 conditions producing crystals. Manual reproduction of the condition giving crystals with promising diffraction (again solution 9 from Crystal Screen Lite) and protein/precipitant composition and ratio optimisation led to the first good sized needle-like crystals of the SeMet protein (B in Table 5.1). Optimisation of crystals from solution B6 of Crystal Screen Ammonium Sulphate led to rod-like crystals of the native protein (C and D in Table 5.1). SeMet-UgpG crystals growing in condition C2 of Screen MME 500 grew very fast, in less than two hours reaching $0.1 \times 0.08 \times 0.03 \text{ mm}^3$, but unfortunately did not deliver diffraction higher than 4.5 Å resolution. Temperature, precipitant composition, co-crystallisation with glycerol and

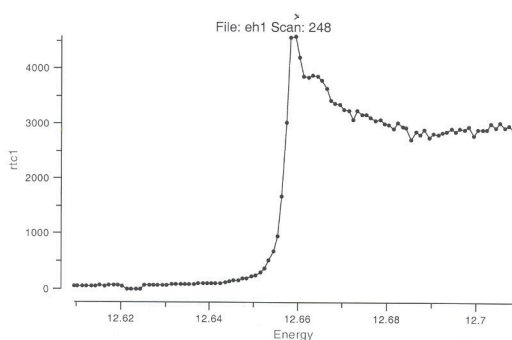


Figure 5.4. *Selenium energy scan (fluorescence counts vs keV) for a crystal grown from selenomethionine protein.*

other additives were not successful to slow down nucleation. The use of 0.1% LM agarose (Hampton Research) in the drops retarded nucleation by

24 h before the appearance of the first crystals. For one of these crystals selenium incorporation was confirmed by energy scanning at the K Se absorption edge (Figure Figure 5.4). Optimisation of this condition led to condition E in 5.1. A similar condition with the native protein was also found to be successful, condition F in Table 5.1.

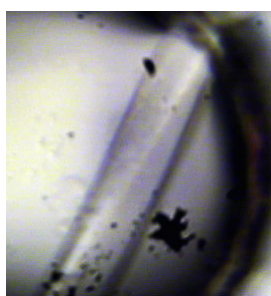


Figure 5.5. *Crystal of native UgpG*

Crystallised with 5 μ l protein at a concentration of 9 mg/ml plus 2 μ l buffer solution (G3, 5.1).

Since crystal quality had not been sufficient to enable structure determination, another set of 576 crystallisation trials for the native and selenated UgpG were performed using the EMBL Cartesian HTX platform, this time also attempting co-crystallisation with 1 and 5 mM G1P, UDP-Glc or UTP.

Again, the best conditions were found to be similar to

those already obtained before and optimisation led

to condition G3 (Figure 5.5). Crystal soaks and co-crystallisation trials with salts of Hg, Pt, Sm, W and

Os were screened in order to search for heavy atom derivatives. Soaking times ranged from 10 min to 3 h and soaking concentrations varied from 2 to 50

mM. The crystals soaked with potassium osmate (VI) darkened (Figure 5.6) with increasing incubation time, suggesting metal incorporation (crystals G1, G2 and G3). Only W and Os produced diffracting crystals.

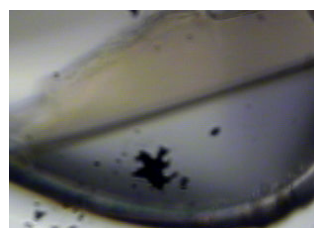


Figure 5.6. *Crystal of UgpG soaked in potassium osmate*

Soaked with 10 mM potassium osmate (IV) for 30 minutes. The crystal started to darken after 15 minutes soaking time.

5.3.4. X-ray diffraction analysis and structure determination

All diffraction data were collected at the European Synchrotron Radiation Facility (ESRF). Conditions A1, A2, B, C, D, E and F gave crystals that diffracted to a maximum of 2.9 Å resolution when cryo-cooled (Table 5.2).

Cryo-protection was always achieved by soaking the crystals in cryoprotectant for the shortest possible time, longer soaks did not result in improved diffraction. Flash-cooling was made by rapidly placing crystal loops under a 100 K nitrogen stream. Heavy atoms screens were made with salts of Hg, Pt, Sm, W and Os but neither suitable derivatives were obtained nor molecular replacement (using homologous Rm1A structures and a number of software packages) was successful in structure determination. Other attempts to use the homologous Rm1A structures as search models for glucose-1-phosphate uridylyltransferase molecular replacement had also been reported to fail (Kim, Wu *et al.* 2004). The best crystals grew from type G conditions, from which were obtained an osmium and a selenomethionine derivative, and the highest resolution native datasets, G1, G2 and G3 respectively.

Table 5.2. Data collection statistics for the several crystal types.

Crystal Type	A1	A2	B	C	D	E	F	G1	G2	G3
Source	ESRF, ID14-1	ESRF, ID14-2	ESRF, ID14-2	ESRF, ID14-4	ESRF, ID14-4	ESRF, ID14-4	ESRF, ID14-2	ESRF, ID29	ESRF, ID14-2	ESRF, ID29
λ (Å)	0.934	0.934	0.933	0.933	0.933	0.979	0.933	1.141	0.933	0.976
Special Observations	-	-	SeMet	-	-	SeMet	UDP-Glc	Os, G1P	SeMet, G1P	G1P
Resolution (Å)	24.50-3.50 (3.65-3.50)	81.60-3.10 (3.22-3.10)	77.40-4.20 (4.34-4.22)	76.40-3.00 (3.10-3.00)	76.20-2.90 (3.04-2.90)	76.40-3.00 (3.17-3.02)	83.60-4.00 (4.28-4.00)	74.70-3.50 (3.67-3.50)	75.20-3.50 (3.59-3.50)	102.0-2.65 (2.67-2.65)
Space group	P3 ₁ 21	P3 ₁ 21	P312	P1	P1	P2 ₁ 2 ₁ 2 ₁	P3 ₁ 21	P2 ₁	P2 ₁	P2 ₁
Cell a, b, c, α , β , γ (Å, °)	92.6	93.0	122.6	92.3	89.7	76.6	92.1	105.3	106.6	105.5
	92.6	93.0	122.6	92.7	124.3	102.2	92.1	85.7	85.9	85.7
	257	256.6	159.7	124.8	167.7	152.9	245.6	152.3	152.3	151.8
	90.0	90.0	90.0	96.6	86.7	90	90.0	90.0	90.0	90.0
	90.0	90.0	90.0	90.1	85.9	90	90.0	104.8	105.4	105.2
	120.0	120.0	120.0	115.6	83.6	90	120.0	90.0	90.0	90.0
No. Ref.	177577	296577	72806	144310	536358	162303	98321	192656	440196	490748
Unique Ref	16645	27833	10841	82535	140323	24180	18944	33222	33342	85566
Redundancy	10.7	10.6	6.7	1.7	3.8	6.7	5.2	5.8	13.2	5.7
MnI/Sig	15.4	20.3	18.4	9.0	10.5	13.8	13.8	12.0	18.1	14.1
(outer shell)	(3.0)	(4.0)	(5.5)	(2.1)	(2.3)	(3.1)	(3.8)	(2.4)	(4.5)	(2.0)
Rmerge	0.11	0.08	0.11	0.10	0.11	0.13	0.11	0.13	0.15	0.10
(outer shell)	(0.60)	(0.51)	(0.44)	(0.25)	(0.50)	(0.48)	(0.50)	(0.64)	(0.65)	(0.65)
Completeness (outer shell) %	99.5 (98.9)	99.9 (99.8)	100.0 (100.0)	90.6 (90.1)	90.9 (90.2)	100 (100)	98.4 (98.1)	99.6 (98.5)	99.2 (99.0)	100 (100)
Mosaicity (°)	0.4	0.4	1.0	0.6	1.5	0.7	1	0.5	0.5	0.5
% solvent (V_m)	50.5	50.7	54.6	50.5	57.5, 49.0	47.4	47.7	52.6	53.0	52.4
A ³ /Dalton	(2.48)	(2.50)	(2.71)	(2.49)	(2.89, 2.41)	(2.34)	(2.35)	(2.60)	(2.62)	(2.59)
Estimation no. molecules in a.u.	4	4	4	12	20, 24	4	4	8	8	8
isomorphous phasing power	-	-	-	-	-	-	-	0.73	0.53	-
Figure of merit (centric/acentric reflections)	-	-	-	-	-	-	-	0.25/0.27	0.19/0.10	-

All of the data reduction was made with MOSFLM (Leslie 1992) and SCALA and TRUNCATE (CCP4 1994). Estimation of number of molecules in asymmetric unit (a.u.) were based on normalised probabilities for the occurrence of multimerisation states of the Matthews coefficient (V_m) and solvent content (V_s) (Kantardjieff and Rupp 2003).

Crystals cryo-cooled as grown (no cryoprotectant required) to 100 K from condition G1 diffracted to 3.5 Å resolution, G2 to 3.5 Å resolution and G3 to 2.65 Å resolution. Gel filtration and dynamic light scattering studies show that UgpG oligomerises as a tetramer, 4 x 32 kDa, in solution as reported by others (Chang *et al.* 1999). Taking this into consideration, the Matthews

Coefficient ($V_m = 2.6 \text{ \AA}^3/\text{Dalton}$) suggests two tetramers in the asymmetric unit with 53 % solvent content. Diffraction data were collected at the Os L-III absorption peak on a type G1 crystal soaked with 10 mM potassium osmate (VI) for 60 min. AutoSHARP (Vonrhein 2005), using data from the G3 type native and osmium derivative, was directed to search for osmium sites and successfully located 12 sites. To improve the electron density maps 80 selenium peaks were located from the phased anomalous difference map calculated using diffraction data collected at the peak of the selenium K absorption edge from a G2 type crystal and input into SHARP (Bricogne 1997) together with the phase information from the osmium data. The resulting electron density map is shown in Figure 5.7. Monomers were built by manual model building using Coot (Emsley and Cowtan 2004) on 8-fold

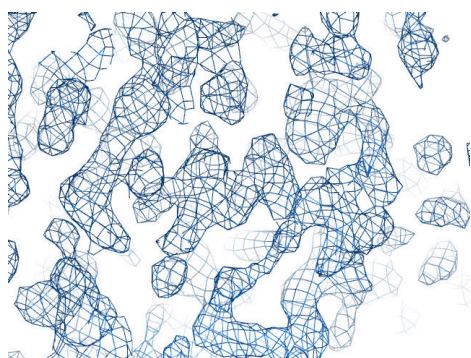


Figure 5.7. Experimental electron density map at 3.5 Å resolution

Produced by SHARP after density modification by Solomon (12 osmium sites and 83 selenium sites for phasing).

averaged maps calculated with DM (Cowtan 1994).

5.4. Results

Glucose-1-phosphate

uridylyltransferase

from

Sphingomonas elodea ATCC 31461

was cloned and expressed. Extensive

crystallisation screens were performed

in order to address the phase problem.

Diffracting crystals were obtained in

five different space groups and seven crystal forms, from which diffraction data were collected and characterised. SeMet and heavy metals derivatives were screened. The phase problem was finally solved by MIRAS using Os and SeMet derivatives and the best diffracting crystal form, a complex with one G1P bound molecule in a monoclinic system that diffracted to 2.65 Å

resolution. In the crystal, with 8 UgpG molecules per asymmetric unit and a solvent content of 52% (v/v), the protein oligomerises as tetramers, matching that observed in solution by other biochemical techniques.

5.5. Acknowledgements

The authors thank the staff at IBS and EMBL-Grenoble for the use of high-throughput crystallisation robots and ESRF for the provision of synchrotron radiation. The authors also thank Ana Varela Coelho and the ITQB/IBET Mass Spectrometry Laboratory for mass spectrometry analysis. D.A. acknowledges a grant from FCT, SFRH (BD/6480/2001). This work was partially funded by FEDER and Fundação para a Ciência e a Tecnologia (FCT), Portugal (grants: POCTI/BME/44441/2002, POCTI/BIO/58401/2004).

The cloning was done by Ana Rita Marques (ARM). The protein expression and purification, including seleno-methionine incorporated, was done by David Aragão following ARM experience and protocols and making the necessary adjustments for large scale preparation. The crystallisation, data collection and phase determination was done by David Aragão.

6. The complex of *Sphingomonas elodea* ATCC 31461 glucose-1-phosphate uridylyltransferase with glucose-1-phosphate reveals a novel quaternary structure, unique amongst NDP-sugar pyrophosphorylase members

(Based on a published refereed paper now including more detailed data on the results and discussion sections and replacing some of the black and white for colour pictures: David Aragão, Arsénio M. Fialho, Ana R. Marques, Edward P. Mitchell, Isabel Sá-Correia and Carlos Frazão. J. Bacteriol. 2007 Jun;189(12):4520-8.)

6.1. Abstract

Gellan gum is a widely used commercial material, available in many different forms. Its economical importance has led to studies into the biosynthesis of exopolysaccharide gellan gum which is industrially prepared in high yields using *Sphingomonas elodea* ATCC 31461. Glucose 1-phosphate uridylyltransferase mediates the reversible conversion of glucose-1-phosphate and UTP into UDP-glucose and pyrophosphate which is a key step in the biosynthetic pathway of gellan gums. Here we present the X-ray crystal structure of the glucose-1-phosphate uridylyltransferase from *S. elodea*. The *S. elodea* enzyme shares strong monomeric similarity with glucose-1-phosphate thymidylyltransferase, several structures of which are known, although the quaternary structures of the active enzymes are rather different. A detailed comparison between *S. elodea* glucose-1-phosphate uridylyltransferase and available thymidylyltransferases is described, and shows remarkable structural similarities despite the low sequence identity between the two divergent groups of proteins.

6.2. Introduction

The bacteria *Sphingomonas (S.) elodea* ATCC 31461 (Marques *et al.* 2003; Silva *et al.* 2005) is an industrially used micro-organism for the high yield synthesis of exopolysaccharide gellan gum, approved in the USA and EU as food gelling, stabilising and suspending agent, either on its own or in combination with other hydrocolloids (Chandrasekaran and Radha 1995). In its native form, gellan is a linear anionic heteropolysaccharide based on a tetrasaccharide repeat unit composed of two molecules of D-glucose (Glc), one of D-glucuronic acid (GlcA) and one of L-rhamnose (Rha). Native gellan is partially esterified with acetate and glycerate substituents on the D-glucosyl residue adjacent to the D-glucuronyl residue (1 mole of glycerate and 0.5 mole of acetate per repeat unit) (Jay *et al.* 1988) with an average molecular mass of about 0.5 MDa. The rheological properties of gellan are significantly modified upon the removal of acyl substituents. They might be tuned from soft and elastic gells to more brittle and harder ones, depending on the level of glycerate substituents present in the repeating unit (Jay *et al.* 1988).

Gellan gum is available commercially in three chemical forms: no, low and high acyl content with the respective denominations of Gelrite®, Kelcogel® F and Kelcogel® LT100. Gelrite® is used as a substitute of agar in microbiological and tissue culture media. Kelcogels are food grade gellan gums mainly used as gelling agents in foods and personal care applications. Other biomedical applications of gellan include its use as a pharmaceutical excipient for nasal and ocular drug delivery applications (Sultana *et al.* 2006) and as material for the construction of three-dimensional scaffolds for tissue engineering (Ciardelli *et al.* 2005).

The commercial applications of gellan have been a stimulus to study its biosynthesis. The production pathway is a multi-step process starting with

the intracellular formation of the nucleotide-sugar precursors, uridine 5'-diphosphate-glucose (UDP-Glc), UDP-glucuronic acid (UDP-GlcA) and dTDP-L-rhamnose (dTDP-L-Rha). This is followed by the formation of the repeat unit, with sequential transfer of the sugar donors to an activated lipid carrier by committed glycosyltransferases and ultimately by gellan polymerisation and export (Sá-Correia *et al.* 2002). The genes coding for the proteins involved in the synthesis of dTDP-L-Rha, glycosyltransferases and those required for gellan polymerisation and export are located together in the *gel* cluster (Videira *et al.* 2000; Sá-Correia *et al.* 2002). However, the genes (*pgmG*, *ugpG* and *ugdG*) coding the enzymes required for the synthesis of the sugar-nucleotides (UDP-glucose and UDP-GlcA) from glucose 1-phosphate are located at different positions on the *S. elodea* genome (Videira *et al.* 2000; Sá-Correia *et al.* 2002; Marques *et al.* 2003), perhaps reflecting the more broad function of these enzymes beyond the specific gellan pathway.

Glucose 1-phosphate uridylyltransferase (UGP, EC 2.7.7.9) mediates the reversible conversion of glucose-1-phosphate (G1P) and UTP into UDP-glucose and pyrophosphate. This enzyme is widely distributed throughout all domains of life although the eukaryotic UGPs are unrelated in terms of amino acid sequence to their prokaryotic counterparts (Daran *et al.* 1995; Eimert *et al.* 1996; Flores-Diaz *et al.* 1997). UGPs belong to the superfamily of nucleoside diphosphate sugar pyrophosphorylases. The protein sequence most closely homologous to bacterial UGP is glucose-1-phosphate thymidylyltransferase (RmlA, [EC 2.7.7.24](#)), a bacterial enzyme that mediates the reversible production of dTDP-glucose from G1P and dTTP, the first of a four step pathway necessary for the formation of the nucleotide sugar dTDP-L-rhamnose (Silva *et al.* 2005). The protein under study - UGP from *S. elodea* (UgpG) - recognises both UTP and dTTP nucleotides substrates *in*

vitro (Silva *et al.* 2005) whilst other UGPs have been reported as being unable to use dTTP as substrates *in vitro* (Bonofiglio *et al.* 2005). Studies have been performed to understand why some of these proteins are promiscuous toward both their nucleotide triphosphate (NTP) and the sugar substrates (Blankenfeldt, Asuncion *et al.* 2000; Barton *et al.* 2001) but the reasons are yet to be established. We have previously described the kinetic properties (Marques *et al.* 2003; Silva *et al.* 2005) and preliminary structure determination of UGP from the industrial micro-organism *S. elodea* ATCC 31461 (Aragao *et al.* 2006). Recently, the coordinates of a crystallographic structure of *Escherichia (E.) coli* glucose-1-phosphate uridylyltransferase were deposited in the PDB, entry 2E3D, but no associated publication was reported ⁽⁹⁾. The study presented here reports the structure analysis of UgpG and highlights the main differences between UGPs and RmlAs.

6.3. Material and methods

6.3.1. Expression, purification and crystallisation of UgpG.

UgpG was expressed, purified and crystallised as previously described (Aragao *et al.* 2006). Briefly, for the production of both native and selenomethionine derivatised protein, the expression vector was transformed into *E. coli* BL21 Gold (DE3) and B834 (DE3) respectively, grown to an OD_{640nm} of 0.6 and induced with 0.1 mM IPTG for 3 h at 37 °C. Bacteria were harvested, lysed (French press) and cell debris removed. The supernatant was applied to a nickel affinity column and eluted with a gradient of increasing imidazole concentration. The chromatographic peak corresponding to UgpG was separated and concentrated to 10 mg/ml in a

9 When the paper from where this chapter is based was at proof stage a publication with the 1.9 Å resolution description of the glucose-1-phosphate uridylyltransferase from *E. coli* has been published by Thoden and Holden (2007a).

buffer containing 30 mM NaCl and 50 mM Tris-HCl, pH 7.6 and crystallised by vapour diffusion in hanging drops at 20 °C. The UgpG-G1P crystals were obtained by the vapour diffusion method equilibrating drops of 5 µl of the protein solution complemented with 5 mM G1P and 2 µl of reservoir solution, against the reservoir containing 500 µl of 0.1 M sodium citrate pH 4.6, 0.1 M ammonium acetate and 15% PEG MME. A heavy atom soak was carried out on the native crystals with 10 mM K₂OsO₄ for 30 minutes.

6.3.2. Data collection, structure determination and refinement.

X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, beamlines ID14-2 and ID29. The diffraction images were integrated with MOSFLM (Leslie 1992) and the intensities scaled, merged together and reduced to magnitudes using the CCP4 suite of programs (CCP4 1994), which was also used for further data handling and scaling purposes. Reflections for the calculation of R_{free} were grouped in thin resolution shells, due to the presence of 8-fold non-crystallographic symmetry (NCS). Diffraction data statistics are presented in Table 6.1.

The UgpG-G1P structure was determined (Aragao *et al.* 2006) by the multiple-isomorphous replacement and anomalous scattering phasing method (MIRAS). Data were collected from a native crystal, from an osmium salt soaked crystal (collected at the peak of the osmium L-III absorption edge) and from a seleno-methionine derivatised crystal (collected at the peak of the selenium K absorption edge). The program autoSHARP (Vonrhein 2005) identified the locations of 12 osmium atoms and produced a set of initial phases with an overall isomorphous phasing power of 0.73. With these initial phases, 80 selenium sites were found by SHARP (Bricogne 1997) using a phased anomalous difference map calculated with the

selenium data.

Table 6.1. *Refinement statistics for UgpG at 2.65 Å.*

Resolution Limits (Å)	102.0-2.65							
Highest resolution shell	2.67-2.65							
R _{work} (No. observations)	0.245 (76179)							
R _{Free} (No. observations)	0.295 (1992)							
Highest resolution shell	0.384 (1883)							
R _{work} (Obs)								
R _{free} * (Obs)	0.403 (492)							
Solvent water molecules	160							
GIP atoms	8 x 16							
Residues not located								
A ₁	1, 162-172, 187-200, 223-230							
A ₁ '	223-229							
B ₁	1, 223-229							
B ₁ '	1, 223-230							
A ₂	1, 162-182, 205-219, 221-229							
A ₂ '	1, 222-229							
B ₂	1							
B ₂ '	1, 224-229							
Average B _{iso} (Å ²)	54 (39)							
All atoms (solvent)								
Monomer	A ₁	A ₁ '	B ₁	B ₁ '	A ₂	A ₂ '	B ₂	B ₂ '
B _{iso} Protein atoms (Å ²)	61	52	56	53	52	48	54	59
B _{iso} Main chain (Å ²)	61	50	55	52	50	46	52	57
B _{iso} Side chain (Å ²)	62	54	56	56	54	49	56	61
Nr. Amino acids	255	282	281	280	244	280	288	282
Nr. Protein atoms	1940	2145	2135	2126	1866	2131	2174	2144
Deviations from ideality:								
Bond distances (Å)	0.012							
Bond angles (°)	1.84							

*The R-factor highest resolution shell differs from that of the R-free set, 2.75Å -2.77 Å, because the last was selected in thin shells in order to minimise the bias due to the restrained non-crystallographic symmetry refinement.

The combined phases were improved using the NCS information in DM (Cowtan 1994), to average the electron density maps. A preliminary model was produced with COOT (Emsley and Cowtan 2004) and was iteratively fed back to SHARP (Bricogne 1997) allowing an improvement of the overall isomorphous phasing power from 1.3 to 1.8 in the resolution range 100 – 3.5 Å. The molecular model was completed and refined with CNS (Adams *et al.* 1997) using the NCS restraints, simulating annealing and a maximum

likelihood target.

6.3.3. Three-dimensional structure comparison

All of the RmlA and the *E. coli* Ugp coordinate files deposited in the Protein Data Bank were aligned with the UgpG structure presented here using the secondary structure matching algorithm (SSM) (Krissinel and Henrick 2004) implemented in COOT (Emsley and Cowtan 2004). This initial alignment was further refined with Modeller (Sali and Blundell 1993) with a threshold of 3.5 Å for inter-protein residues matching.

6.3.4. Coordinate accession numbers

Coordinates and structure factors of *S. elodea* UgpG have been deposited in the Protein Data Bank (PDB accession numbers, 2UX8 and 2UX8SF).

6.4. Results

6.4.1. Crystal contents and structure quality

The *S. elodea* UgpG crystals belong to space group P2₁, cell dimensions $a = 105.5 \text{ Å}$, $b = 85.7 \text{ Å}$, $c = 151.8 \text{ Å}$ and $\beta = 105.2^\circ$. They diffract to 2.65 Å resolution and contain 8 molecules in the asymmetric unit, which corresponds to a calculated solvent content of 51.7% (Matthews 1968). The asymmetric unit contains 2219 residues, out of a possible total of 2312, with monomers A₁ and A₂ being those with fewer residues visible in the electron density, with only 86% and 85% of their possible atoms modelled respectively (Table 6.1). The histidine-tag present in the protein sequence is not visible in the density for any of the monomers. Other undetected residues belong to loops and were not built into the model when not visible in 1 σ electron density maps, probably due to disorder. The asymmetric unit consists of two tetramer units, with one G1P molecule per UgpG monomer and a total of 160 solvent molecules modelled as waters.

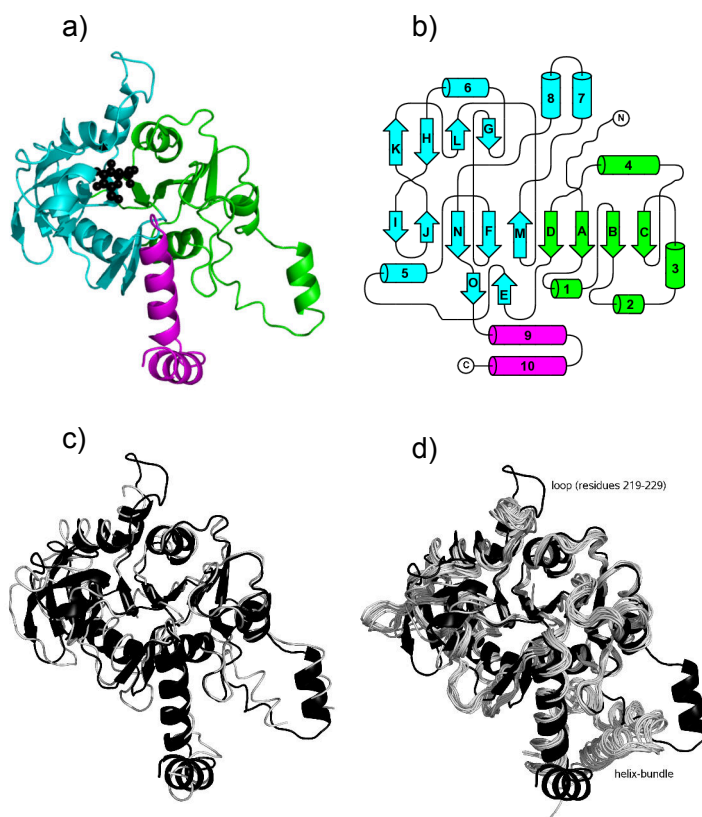
Enzyme gel filtration profile and dynamic light scattering experiments show a single species in solution of about ~120 kDa mass (data not shown). This corresponds to approximately four times the expected molecular weight of the 289 residues protein monomer, and suggests that the crystallographic tetramers are the functional units of UgpG.

The refined structure converged to final R_{work} and R_{free} of 24.5% and 29.5%, respectively. A stereochemical analysis of the refined model using PROCHECK (Laskowski *et al.* 1993) revealed 86%, 13%, 1% and 1% residues in the most favourable, additional allowed, generously allowed and disallowed regions of the Ramachandran plot, respectively. The residues in the disallowed areas are partially supported by electron density and lie in regions with high B-factors. The overall geometry was better than for the set of PROCHECK comparative structures at 2.65 Å resolution, with a G-factor of 0.13.

6.4.2. Overall polypeptide fold

The monomer (Figure 6.1a) has an overall size of around 60 Å, 45 Å, 35 Å and is classified by CATH (Pearl *et al.* 2003; see Appendices) as an alpha/beta complex fold (CATH classification of 3.90). Each monomer is made up of ten helices and fifteen β -strands (Kabsch and Sander 1983) labelled consecutively from 1 to 10 and A to O, respectively (Figure 6.1b).

Figure 6.1. UgpG monomer.



a) Ribbons representation of UgpG showing G1P as ball-and-stick (black). The molecule may be divided into three functional regions, the nucleotide binding, the sugar binding and the dimerisation sub-domains, which are coloured in green, cyan and pink, respectively. **b)** Topology diagram of UgpG using the same colour scheme for sub-domain identification. **c)** Superposition of UgpG (black ribbon) and the recently available *E. coli* UGP (PDB ID 2D3D, grey). **d)** Superposition of UgpG (black ribbon) and RmlA structures (grey) showing their overall similarity. The loop 219-229 lies above the catalytic cavity. The labelled helix bundle, only present in the RmlA structures, contains a second allosteric NTP binding site.

The UgpG monomer is built by a large, mixed β -sheet, reminiscent of the Rossman-fold but including nine β -strands, spanning the whole molecule and surrounded by ten helices and two additional anti-parallel β -sheets. The

molecule may be divided into three functional regions or sub-domains (Creighton 1992). The nucleotide binding sub-domain (NBSD, residues 1-132), is composed of the first four β -strands of the main β -sheet sandwiched between their interconnecting helices. It resembles other nucleotide binding domains (Brandeen and Tooze 1991) and is strongly involved in inter-monomer interactions. Subdomain 2, the sugar binding sub-domain (SBSD, residues 133-256), binds G1P and is the most compact of the three sub-domains. It is composed of the second half of the main β -sheet including five mixed strands, the two extra β -sheets with two and four strands, and four α -helices. Finally, sub-domain 3, the dimerisation sub-domain (DSD, residues 257-289) is composed of two α -helices arranged like a “V” and, together with sub-domain 1, is responsible for the formation of the active tetramer (see below).

6.4.3. Quaternary structure and oligomerisation interfaces

The asymmetric unit contains two homotetramers that can be described as dimers of dimers (A_1A_1' and B_1B_1' , and A_2A_2' and B_2B_2' with subscripts denoting the tetramer, letters identifying primary dimers, and primes used to identify monomers) with the monomers showing positional root-mean-square differences (RMSDs) between homologous $C\alpha$'s within 0.2-0.4 Å for 237 positions to a cut-off of 3.5 Å (Table 6.2).

Table 6.2. *RMSD between UgpG monomers.*

Top left to bottom right diagonal, number of visible residues per monomer. Above the diagonal, $C\alpha$ atoms positional RMSD, below the diagonal, number of 3D homologous $C\alpha$

positions within 3.5 Å.

<i>Monomer</i>	A₁	A₁'	B₁	B₁'	A₂	A₂'	B₂	B₂'
A₁	255	0.4	0.3	0.4	0.2	0.3	0.4	0.3
A₁'	253	281	0.4	0.4	0.3	0.3	0.3	0.4
B₁	255	279	281	0.3	0.2	0.3	0.3	0.2
B₁'	254	278	280	280	0.3	0.2	0.3	0.4
A₂	240	242	244	243	244	0.2	0.2	0.2
A₂'	254	278	280	279	244	280	0.2	0.3
B₂	254	278	280	279	244	279	288	0.4
B₂'	255	279	281	280	244	280	280	282

This assembly is responsible for an overall buried surface area upon tetramer formation of 12 kÅ², slightly larger than the surface of one monomer alone, 10 kÅ² (AreaMol (CCP4 1994)). The inter-dimer assembly that produces the tetramer (Figure 6.2a, b) is dominated by the stacking of helices 2 and β -strand C between monomers A and B (in a total of 36 to 46 van der Waals interactions, 13 to 20 hydrogen bonds, Table 6.3). The inter-dimer interfaces are strongly bound due to two particular interactions made by the NBSD and DSD. From the NBSD, residues 15 to 39 form a set of hydrogen bonds (considered until 3.2 Å) and favourable apolar interactions (van der Waals interactions, vdW, considered until 3.6 Å) with their symmetry equivalents, and in particular methionines 27 and 31 and 27' and 31' create a four methionine apolar region (Figure 6.2c). These interactions are responsible for 28 to 52 vdW interactions and 6 to 10 hydrogen bonds between monomers, depending on the particular dimer. The DSD is responsible for the helix-helix interactions between helices 9 and 10 and 9' and 10', with one set with the “V” shape fitted into the inverted “V” of the neighbour monomer (Figure 6.2b, middle top and middle bottom and Figure 6.2d). This interaction is responsible for 8 to 17 vdW interactions and 1 or 2 hydrogen bonds within each dimer. The two interactions account, on average, for 90% of the total vdW interactions and 70% of the total hydrogen bonds that create

the dimer. Upon tetramer formation the parallel alignment of the C β -strands allows the extension of the monomers' wide β -sheet into two, inter-dimers, supra- β -sheets (Figure 6.2e, stacks of near parallel strands represented by lines) contributing to the tetramer stability. These interactions account for 82% of the vdW interactions and 98% of the hydrogen bonds of the total dimer-dimer contacts interface (Table 6.3).

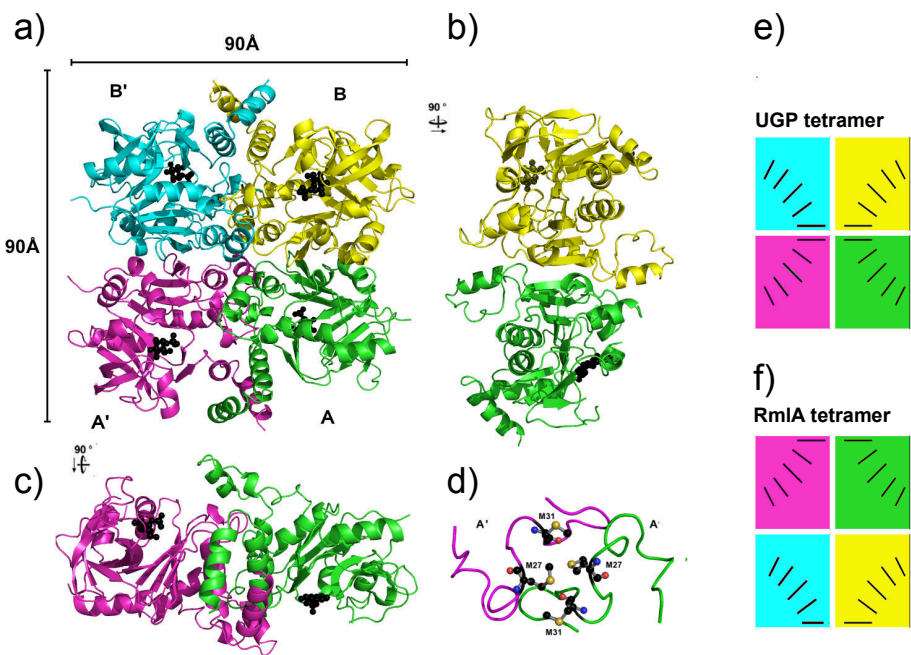


Figure 6.2. UGP tetramer.

a, b and d) Perpendicular views of the UgpG oligomer coloured by monomer, in ribbon representation with G1P in ball-and-stick. Dimers AA' and BB' are represented by the green-magenta and yellow-blue monomers. While (b) shows the whole tetramer (a) and (d) represent only the front monomers for clarity purposes. **c)** A - A' interactions (residues 15:37) with residues 27 and 27' and 31 and 31' creating a four methionine apolar region. **e)** and **f)** schemes of UGP and RmlA tetramer arrangements, respectively. Rectangles are representing monomers, coloured coded as in (b). Lines inside monomers represent β -strands, which in UGP form an extended β -sheet across monomers AB. The schemes highlight the different quaternary structures adopted by UGP and RmlA.

Table 6.3. Inter-monomers interactions for tetramers 1 (a) and 2 (b).

Top-right to the diagonal, number of vdW interactions up to 3.6 Å maximum length and in brackets number of hydrogen bonds up to 3.2 Å maximum length. Bottom-right to the diagonal, number of vdW interactions and in brackets the number of hydrogen bonds due to

a) A-A' and B-B' dimerisation sub-domains interactions (residues 257-289), b) A-A' and B-B' interaction in four methionine apolar region (residues 15-39) and c) A-B and A'-B' helix 2 (residues 60-72) and β -sheet C (residues 97-107) interactions. Calculated with CONTACT

(a)	A_I	A_I'	B_I	B_I'
A_I	-	51 (14)	41 (17)	1 (0)
A_I'	A17 (1)/b28 (10)	-	14 (0)	40 (15)
B_I	C40 (17)	-	-	65 (15)
B_I'	-	C36 (14)	A12 (2)/b45 (6)	-

(b)	A_2	A_2'	B_2	B_2'
A_2	-	70 (13)	49 (20)	3 (0)
A_2'	A12 (1)/B52 (8)	-	7 (0)	37 (13)
B_2	C46 (20)	-	-	45(16)
B_2'	-	C36 (13)	A8(2)/B32(10)	-

6.4.4. Catalytic cavity

A loop between the K and L β -strands over the catalytic cavity contains the UGPs and RmlAs conserved sequence V190-E191-K192-P193 (UgpG numbering) (Figure 6.3, (Thorson *et al.* 1994; Silva *et al.* 2005)), where K192 is thought to stabilise the negative charge of the phosphate group when the reaction takes place. The catalytic cavity spreads through both the NBSD and SBSD, and is solvent accessible (Figure 6.4a).

ugp1	-AI----	KKP-LRKAVFVAGL-GTRFLPATKAMPKEMLPVDRPLIQYAVDEAVEAGIEQMFVTGRGK-SALEDHFDIAEAEATMAARGKS--LDVL-D	90
2E3D	N-----	TKVKKAVIPVAGL-GTRMLPATKAMPKEMLPVDRPLIQYVNECIAAGITEIVLVTHSSK-NSIENHFDSFELEAMLEK--RQLLD-EVQS	89
1I1N	-----	MK-TRKGIILAGGS-GTRLYPVTMAVSQQLLPIDYDKPMIYPLSTMLAGIRDILIIISTPDQTPRFQQLLG	69
1I1M	-----	MK-TRKGIILAGGS-GTRLYPVTMAVSQQLLPIDYDKPMIYPLSTMLAGIRDILIIISTPDQTPRFQQLLG	69
1MP3	-----	K-TRKGIILAGGS-GTRLYPVTMAVSQQLLPIDYDKPMIYPLSTMLAGIRDILIIISTPDQTPRFQQLLG	68
1MP4	-----	MK-TRKGIILAGGS-GTRLYPVTMAVSQQLLPIDYDKPMIYPLSTMLAGIRDILIIISTPDQTPRFQQLLG	69
1MP5	-----	K-TRKGIILAGGS-GTRLYPVTMAVSQQLLPIDYDKPMIYPLSTMLAGIRDILIIISTPDQTPRFQQLLG	68
1H5R	-----	RM-----	68
1MC3	-----	KGIIILAGGS-GTRLHPITRQVSKQLLPIDYDKPMIYPLSVLMLAGIREILIIISTPDQTPRFQQLLG	65
1G0R	-----	K-TRKGIILAGGS-GTRLHPATLAISKQLLPVVDKPMIYPLSTMLAGIREILIIISTPDQTPRFQQLLG	67
1G23	-----	K-TRKGIILAGGS-GTRLHPATLAISKQLLPVVDKPMIYPLSTMLAGIREILIIISTPDQTPRFQQLLG	67
1G3L	-----	K-TRKGIILAGGS-GTRLHPATLAISKQLLPVVDKPMIYPLSTMLAGIREILIIISTPDQTPRFQQLLG	67
1FX0	-----	K-TRKGIILAGGS-GTRLHPATLAISKQLLPVVDKPMIYPLSTMLAGIREILIIISTPDQTPRFQQLLG	67
1G1L	-----	K-TRKGIILAGGS-GTRLHPATLAISKQLLPVVDKPMIYPLSTMLAGIREILIIISTPDQTPRFQQLLG	67
1LVW	-----	MKGIVLAGGSG-TRLYPITRAVSKQLLPIDYDKPMIYPLSVLMLAGIRDILIIISTPRDLPLRYDLDL	66
ugp1	GTR-LKPG-----	NIAVYRQCEPMQGHAV-NCARDIVGDEPFVLPDDFMF--GQPGCL--K-Q--MVDAY-N-KVG-----GNLICAEV-PDDQ	166
2E3D	-----	PHVTIMQVYRQC--KQGHAV-LCAHPVVGDEPVAVT--PDVILDEYES--DL--SQDNLA--MIRR--FDET-GHSGIMVEPV-A-D	165
1I1N	-----	DGS--QW--GNLNLQYKVPSPDGAQ-AFTIGEEFIGHDDCALV--GDNIFY--G--HDLPKL--M-EAARN--K--ESGATVFAYHV-N-D	142
1I1M	-----	DGS--QW--GNLNLQYKVPSPDGAQ-AFTIGEEFIGHDDCALV--GDNIFY--G--HDLPKL--M-EAARN--K--ESGATVFAYHV-N-D	142
1MP3	-----	DGS--QW--GNLNLQYKVPSPDGAQ-AFTIGEEFIGHDDCALV--GDNIFY--G--HDLPKL--M-EAARN--K--ESGATVFAYHV-N-D	141
1MP4	-----	DGS--QW--GNLNLQYKVPSPDGAQ-AFTIGEEFIGHDDCALV--GDNIFY--G--HDLPKL--M-EAARN--K--ESGATVFAYHV-N-D	142
1MP5	-----	DGS--QW--GNLNLQYKVPSPDGAQ-AFTIGEEFIGHDDCALV--GDNIFY--G--HDLPKL--M-EAARN--K--ESGATVFAYHV-N-D	141
1H5R	-----	DGS--QW--GNLNLQYKVPSPDGAQ-AFTIGEEFIGHDDCALV--GDNIFY--G--HDLPKL--M-EAARN--K--ESGATVFAYHV-N-D	141
1MC3	-----	DGS--EF--GTLQLEYAEPSPDGAQ-AFTIGEEFINGDEPSCIV--GDNIFY--G--QGSFVK--L-RHVA--R--TEGATVFAYHV-M-D	138
1G0R	-----	DGS--NW--GLDLQYAVOPSPDGAQ-AFLIGESFIGNDLSALV--GDNLYY--G--HDFHEL--L-GSASQ--R--QTGASVFAYHV-L-D	140
1G23	-----	DGS--NW--GLDLQYAVOPSPDGAQ-AFLIGESFIGNDLSALV--GDNLYY--G--HDFHEL--L-GSASQ--R--QTGASVFAYHV-L-D	140
1G3L	-----	DGS--NW--GLDLQYAVOPSPDGAQ-AFLIGESFIGNDLSALV--GDNLYY--G--HDFHEL--L-GSASQ--R--QTGASVFAYHV-L-D	140
1FX0	-----	DGS--NW--GLDLQYAVOPSPDGAQ-AFLIGESFIGNDLSALV--GDNLYY--G--HDFHEL--L-GSASQ--R--QTGASVFAYHV-L-D	140
1G1L	-----	DGS--NW--GLDLQYAVOPSPDGAQ-AFLIGESFIGNDLSALV--GDNLYY--G--HDFHEL--L-GSASQ--R--QTGASVFAYHV-L-D	140
1LVW	-----	DGS--QF--GVRFYRVDPEPRGLAD--FTVKGDFIGDSKVALV--GDNIFY--G--HRFSEI--L-RRAAS--L--EDGAVIFGYVRD--	139
ugp1	-THRYGIIITP-----	GTQDGVLT-TEVKGLEKVP-AP-GT--APSN-LSVIGRMILQPEVMRILENAGKGAGGEI-----QTDAMORMI--G	239
2E3D	-VTAYGVVDCKG--VELAPG-----	ES-VPMGVVVEK-P-KA--DV-APSNLAIVGRVLSADIWPLLAK-----T-----QTDADIMLEK--	234
1I1N	P-ERYGVVEF-DQKG-----	T-----AVSLEEK-P-LQ-----PK-SNYAVTGLFVYDNSVVEAMKN-----LKPARGELDTINRIYMEQG	211
1I1M	P-ERYGVVEF-DQKG-----	T-----AVSLEEK-P-LQ-----PK-SNYAVTGLFVYDNSVVEAMKN-----LKPARGELDTINRIYMEQG	211
1MP3	P-ERYGVVEF-DQAG-----	T-----AVSLEEK-P-LQ-----PK-SNYAVTGLFVYDNSVVEAMKN-----LKPARGELDTINRIYMEQG	210
1MP4	P-ERYGVVEF-DQAG-----	T-----AVSLEEK-P-LQ-----PK-SNYAVTGLFVYDNSVVEAMKN-----LKPARGELDTINRIYMEQG	211
1MP5	P-ERYGVVEF-DQAG-----	T-----AVSLEEK-P-LQ-----PK-SNYAVTGLFVYDNSVVEAMKN-----LKPARGELDTINRIYMEQG	210
1H5R	P-ERYGVVEF-DKNG-----	T-----AISLEEK-LE-----PK-SNYAVTGLFVYDNSVVEAMKN-----LKPARGELDTINRIYMEQG	210
1MC3	P-ERFVGVEF-----	DD-----NFRAISLEEK-P-KQ--P-K-SNYAVTGLFVYDQVVDIARD-----VKPSRGELDTINQMYLEAG	207
1G0R	P-ERYGVVEF-DQGG-----	K-----AISLEEK-LE-----PK-SNYAVTGLFVYDQVVDIARD-----LKPSPRGELDTVNRAYLERG	209
1G23	P-ERYGVVEF-DQGG-----	K-----AISLEEK-LE-----PK-SNYAVTGLFVYDQVVDIARD-----LKPSPRGELDTVNRAYLERG	209
1G3L	P-ERYGVVEF-DQGG-----	K-----AISLEEK-LE-----PK-SNYAVTGLFVYDQVVDIARD-----LKPSPRGELDTVNRAYLERG	209
1FX0	P-ERYGVVEF-DQGG-----	K-----AISLEEK-LE-----PK-SNYAVTGLFVYDQVVDIARD-----LKPSPRGELDTVNRAYLERG	209
1G1L	P-ERYGVVEF-DQGG-----	K-----AISLEEK-LE-----PK-SNYAVTGLFVYDQVVDIARD-----LKPSPRGELDTVNRAYLERG	209
1LVW	PRP-FGVVEF-----	DS-----EGRVISIE-EKP-S--R-PKSNYVVPGLFVYDQVVEIARR-----TEPSDRGELDTVNEEYLRMG	208
ugp1	DQPFHGV-TFQ-----	GTRYDCGCKAGFIQANLAVALS--RPDL-EPAVR-AFAV-KALG-----	288
2E3D	ET-VEAYHM-K-----	GKSHDCGNKLGMYMAFVEYGI-RHN--TLGTE-FKAW-LEEE-M-----	281
1I1N	-R-LSVAMM--GRGYAWLDTGTHQSLEASNFATIE--E-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	283
1I1M	-R-LSVAMM--GRGYAWLDTGTHQSLEASNFATIE--E-----	RQGLKVSCEPIAFRKNFINAQQVIELAGPL-SKND-YGK	283
1MP3	-R-LSVAMM--GRGYAWLDTGTHQSLEASNFATIE--E-----	RQGLKVSCEPIAFRKNFINAQQVIELAGPL-SKND-YGK	282
1MP4	-R-LSVAMM--GRGYAWLDTGTHQSLEASNFATIE--E-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	283
1MP5	-R-LSVAMM--GRGYAWLDTGTHQSLEASNFATIE--E-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	282
1H5R	-R-LSVAMM--GRGYAWLDTGTHQSLEASNFATIE--E-----	RQGLKVSCEPIAFRKGFDVQVRLKAV-FLKN-NYQG	282
1MC3	-N-LTVELL--GRGYAWLDTGTHQSLEASNFATIE--E-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	282
1G0R	-Q-LSVEIM--GRGYAWLDTGTHQSLEAGQFIATLE--N-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	281
1G23	-Q-LSVEIM--GRGYAWLDTGTHQSLEAGQFIATLE--N-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	281
1G3L	-Q-LSVEIM--GRGYAWLDTGTHQSLEAGQFIATLE--N-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	281
1FX0	-Q-LSVEIM--GRGYAWLDTGTHQSLEAGQFIATLE--N-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	281
1G1L	-Q-LSVEIM--GRGYAWLDTGTHQSLEAGQFIATLE--N-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	281
1LVW	-K-LRVELM--GRGYAWLDTGTHQSLEAGQFIATLE--N-----	RQGLKVSCEPIAFRKNFINAQQVIELAG-PLS-KNDYGK	280
ugp1	-----	288	
2E3D	-----	281	
1I1N	YLLKMV-----	289	
1I1M	YLLKMV-----	289	
1MP3	YLLKMV-----	288	
1MP4	YLLKMV-----	289	
1MP5	YLLKMV-----	288	
1H5R	YLYKMTKD-----	290	
1MC3	YLELL-R-----	289	ARP
1G0R	YKRLLTE--TVY--	292	
1G23	YKRLLTE--TVY--	292	
1G3L	YKRLLTE--TVY--	292	
1FX0	YKRLLTE--TVY--	292	
1G1L	YKRLLTE--TVY--	292	
1LVW	YLRDLAEGNFHG-----	292	

Figure 6.3. Three-dimensional structure alignment.

Alignment of all available RmlA and UGP structures, upon superposition with Modeller (Sali and Blundell 1993). Residues boxed in white form H-bonds with G1P, residues highlighted in grey form H-bonds with nucleotides and residues highlighted in black constitute a hydrophobic cap to the base of the sugar ring at the catalytic cavity. RmlA PDB codes 1I1N, 1I1M, 1MP3, 1MP4 and 1MP5 are from *Salmonella enterica*, 1G0R, 1G23, 1G3L, 1G1L and 1FX0 are from *Pseudomonas aeruginosa*, 1H5R and 1MC3 are from *E. coli* and 1LVW is from *Methanobacterium thermoautotrophicum*. UGP PDB codes 2UX8 (*S. elodea*) and 2E3D (*E. coli*).

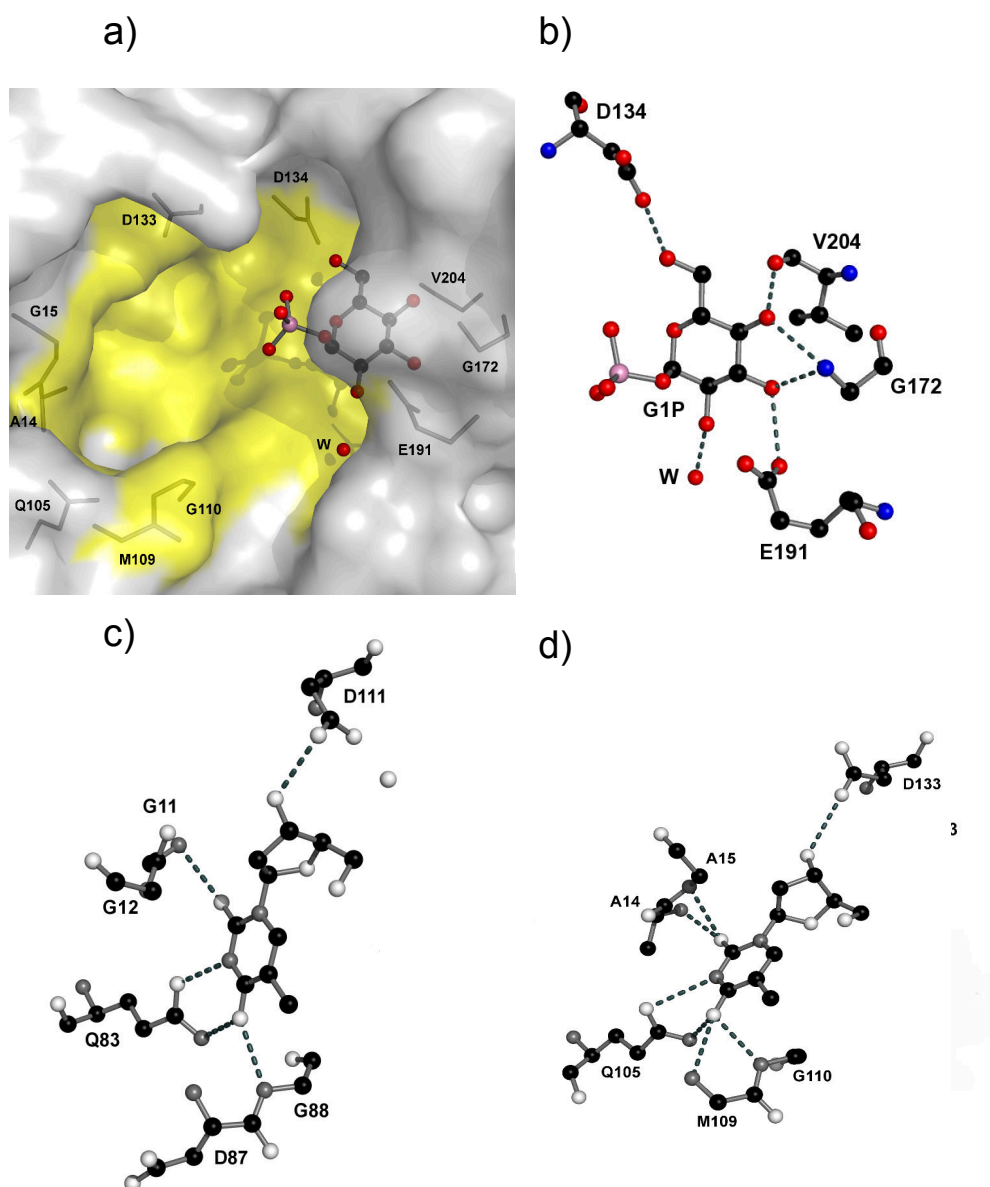


Figure 6.4. *UgpG catalytic cavity.*

a) UgpG transparent solvent accessible surface around the catalytic cavity, with bound G1P (ball-and-stick) and catalytic H-bonding residues (sticks). On the left of G1P there is enough space for a NTP molecule. Side chains atoms of residues M109 and V204, and main chains atoms of residues Q105, D133, D134 and E191, were removed for clarity. **b)** G1P H-bonding network with UgpG catalytic residues, where dashed lines represent proton-donor to proton-acceptor distances up to 3.2 Å maximum distance. **c)** The Rmla-deoxytimidine complex structure 1H5R (Zuccotti *et al.* 2001) with its H-bonding network. **d)** A deoxytimidine molecule was manually fitted, using a molecular graphics workstation, into the empty UgpG catalytic cavity using the (c) complex structure as template. M109 is represented by its main-chain atoms only, for clarity purposes. Carbon is represented in black, oxygen in red, nitrogen in blue and phosphorous pink.

In this structure the substrate, G1P, is located to one side of the cavity, and although it does not form any direct interactions with K192, hydrogen bonds are formed with residues D134, G172, E191 and V204 (Figure 6.4b). Glucose oxygen O3 forms a hydrogen bond to the carboxyl moiety of E191 and to the nitrogen of G172. Oxygen O4 is also bound the same G172 main chain atom and an additional bond to the carbonyl of V204. Oxygen O6 makes a hydrogen bond with OD1 of D134 and, finally, in four of the eight monomers O2 forms a hydrogen bond to a visible water molecule bridging to T234. The side chains of residues L111, L131, V204, Y208 and L231 form the boundaries of the catalytic cavity by providing a hydrophobic cap to the base of the sugar ring and are generally conserved in RmlA sequences (Figure 6.3). The NBSD part of the catalytic cavity is empty, with sufficient space to allow a nucleotide molecule to bind. (Figure 6.4ac, see *Nucleotide modelling*).

6.4.5. Substrate affinity and specificity

Trials to co-crystallise UgpG with thymidine triphosphate (TTP), uridine 5'-triphosphate (UTP) and uridine 5'-diphosphate-glucose (UDP-glc) failed. To check that our protein could indeed bind nucleotides, preliminary isothermal calorimetry studies using the available the protein were made at 298 K with UTP in the presence of 20 mM Mg^{2+} and, separately, with G1P (D. Aragão, unpublished results). A qualitative analysis of the obtained results hinted at favourable binding with similar free energy gains for both substrates. UTP binding has a small enthalpy contribution and is mainly controlled by an increase in entropy upon binding, probably from protein conformation changes and water displacement. Whilst on the other hand, G1P binding is controlled by a considerable enthalpy gain. A possible reason for our failure to co-crystallise UgpG with any of the nucleotides tested is

that their binding is favoured by the presence of Mg^{2+} as a co-factor and, unfortunately, successful crystallisation conditions were only found at low pH where protonated side chains thwart Mg^{2+} coordination.

6.4.6. Structural neighbours

A structural similarity search using DALI (Holm and Sander 1993) against the UgpG monomer shows a number of structurally related proteins, many of which belong to the superfamily of bacterial nucleotide diphosphate sugar pyrophosphorylases. The twelve structures with z-scores between 22 (best) to 14 (worse) and RMSDs from 2.2 to 3.3 Å, respectively, are shown in Table 6.4. The model with the highest z-score and, as it happens, highest sequence identity (21% for 214 3D homologous C α positions) to UgpG is PDB entry 1G0R, a *Pseudomonas aeruginosa* RmlA structure complexed with thymidine and G1P. Many of the other high scoring structures share the NTP binding subdomain with UgpG.

Table 6.4. *Structural alignment statistics and PDB codes from the DALI structural search.*

All hits with a z-core above 14, RMSD below 3.5 and more then 150 aligned residues are shown. (No. Align – number of aligned residues, No. Resi – Total number of residues). PDB ID: 1G0R (Blankenfeldt, Asuncion *et al.* 2000), 1FWY (Brown *et al.* 1999), 1TZF (Koropatkin and Holden 2004), 1JYK (Kwak *et al.* 2002), 1YP2 (Jin *et al.* 2005), 1W55 (Gabrielsen *et al.* 2004), 1EYR (Mosimann *et al.* 2001), 1I52 (Richard *et al.* 2001), 1QWJ (Krapp *et al.* 2003), 1JV1 (Peneff *et al.* 2001), 1H7E (Jelakovic and Schulz 2001), 1FR9 (Lake *et al.* 2000).

PDB ID	z-score	R MSD (Å)	No. Align	No. Resi.	% Identity	Function	Quaternary Structure
1G0R	21.7	2.6	214	292	21	G1P thymidyltransferase	tetramer
1FWY	20.4	3.1	214	326	21	N-acetylglucosamine 1-phosphate uridyltransferase	monomer
1TZF	18.6	2.8	201	251	16	alpha-D-glucose-1-phosphate cytidyltransferase	monomer
1JYK	18.0	2.6	195	229	16	CTP:phosphocholine cytidyltransferase	monomer
1YP2	17.5	3.1	215	428	16	ADP-glucose pyrophosphorylase	tetramer
1W55	16.5	2.6	180	369	12	Methylethritol 4-phosphate cytidyltransferase	monomer*
1EYR	16.2	2.4	168	225	12	Acylneuraminate cytidyltransferase	dimer
1I52	16.1	2.7	184	225	13	4-diphosphocytidyl-2-c-methylethritol synthase	monomer
1QWJ	15.5	2.6	174	228	12	Murine CMP-5-N-acetylneuraminic Acid Synthetase	tetramer
1JV1	14.5	3.3	203	490	14	Udp-n-acetylglucosamine pyrophosphorylase	dimer
1H7E	14.5	2.2	192	245	13	3-deoxy-manno-actulosonate cytidyltransferase	dimer
1FR9	14.2	2.6	163	185	11	Molybdopterin guanine dinucleotide biosynthesis	monomer

* two domains with one active site each

6.4.7. Modelling of nucleotide binding

The above structure superpositions show that their nucleotide binding sub-domains are very similar. This fact, together with the existence of several complexes with nucleotides, was used to help to model thymidine in the UgpG catalytic cavity in the proximity of the bound G1P. Using the *E. Coli*

RmlA structure complexed with both deoxythymidine and G1P (PDB code 1H5R (Zuccotti *et al.* 2001)) as a reference (Figure 6.4c) an homology model was built (Figure 6.4d). The model shows that also in UgpG thymidine is likely to form hydrogen bonds with residues A14, G15, N105, M109, G110 and D133 in accordance with the three-dimensional structure alignment (Figure 6.3 and Table 6.5). UgpG can employ either thymidine or uridine as substrates. They differ in their structures only by a methyl group and our structure has space to accommodate it, leading to little specificity between the two bases. One can expect the binding of either nucleotide to be very similar.

Table 6.5. Active site residues - substrates interactions.

Substrate-Prot	Substrate	<i>S. elodea</i> (UgpG)	<i>S. enterica</i> (RmlA)	<i>E. coli</i> (RmlA)	<i>P. aeruginosa</i> (RmlA)	<i>M. thermophilus</i> (RmlA)
O2-N O2*-O	UTP	Ala 14 †	Gly 11	Gly 11	Gly 10	Gly 8
O3*-NE2	UTP	Glu 30 (OE2) †	Gln 27	Gln 27	Gln 26	Gln 24
O4-NE2 N3-OE1	UTP	Gln 105 †	Gln 83	Gln 83	Gln 82	Gln 80
O4-N	UTP	Met 109 †	Asp 87	Asp 87	Asp 86	Arg 84
O4-N	UTP	Gly 110 †	Gly 88	Gly 88	Gly 87	Gly 85
O3*-OD2	UTP	Asp 133 †	Asp 111	Asp 111	Asp 110	Asp 108
O4-N	G1P	Gly 172	Gly 147	Gly 147	Gly 146	Gly 144
O2-OE2 O3-OE1	G1P	Glu 191	Glu 162	Glu 162	Glu 161	Glu 159
O1B-NZ	G1P	Lys 192	Lys 163	Lys 163	Lys 162	Lys 160
O4-O	G1P	Val 204	Val 173	Val 173	Val 172	Val 170

† - Proposed *S. elodea* UgpG ligands to UTP substrate based on Three-dimensional structure alignment. (Figure 6.3).

6.5. Discussion

Both UgpG and RmlA from *S. elodea* ATCC 31461 are able to catalyse, *in vitro*, the same enzymatic reactions (Silva *et al.* 2005). Several structures of RmlAs have been determined to date and display the same Rossmann-like fold with a nucleotide binding sub-domain plus a sugar binding sub-domain. The three-dimensional structural comparison between the monomers of 13 published X-ray crystal structures for RmlAs with UgpG shows a similar

monomer fold, with RMSDs within 1.5-1.7 Å for over more than 200 fitted C α positions with 3D structural identities below 20% (see Table 6.6). Indeed, the sequence analysis made by Silva *et al.* (2005) on 17 RmlA proteins and three putative or confirmed UGPs showed the two enzymes to form two divergent groups in a phylogenetic tree. The UgpG structure strongly suggests that dTDP and UDP glucose pyrophosphorylase enzymes share a common three-dimensional structure, as Silva *et al* suggested, albeit with low sequence conservation. Local inspection of the structure superpositions shows that the active site cavity and side chain spatial orientations of the catalytic residues are conserved (Figures 6.3 and 6.4). A loop, consisting of residues 219 to 229 and located between helices 7 and 8, closes the active site in all of the RmlA structures, whether substrates are bound or not, whilst in the UgpG structure this loop adopts an “open” conformation allowing access to the active site where G1P is already bound in place (Figure 6.1c). It should be noted, though, that this loop in the UgpG structure is visible in the electron density and unambiguously modelled only for one monomer. The lack of clear electron density for the remaining monomers is a sign of the mobility of this loop, acting as gatekeeper to the active site, which on holding both substrates may then help to stabilise the ternary complex.

Compared to UgpG, all RmlAs have a C-terminal extension of 30 residues forming a three helix bundle that is absent in UgpG (Figure 6.1c, bottom right). The volume corresponding to the missing helix bundle is partially occupied by helix 3 in UgpG. This region is particularly significant as it accommodates a second, presumed to be allosteric NTP binding site on some RmlAs (Sivaraman *et al.* 2002). There is, however, no biochemical, sequence or structural evidence to date that UgpG has a similar second site.

Table 6.6. *Three dimensional superposition of UgpG and RmlA structures.*

Top left to bottom right diagonal, number of residues. Above the diagonal: the number of CA positions less than 3.5Å apart (root mean square difference in brackets); Below the diagonal: the percentage of identical residues with CA position less than 3.5Å apart. PDB IDs: 1IIN, 1IIM (Barton *et al.* 2001), 1MP3, 1MP4, 1MP5 (Barton *et al.* 2002), 1H5R (Zuccotti *et al.* 2001), 1MC3 (Sivaraman *et al.* 2002), 1G0R, 1G23, 1G3L, 1FXO, 1G1L (Blankenfeldt, Asuncion *et al.* 2000).

	1IIN	1IIM	1MP3	1MP4	1MP5	1H5R	1MC3	1G0R	1G23	1G3L	1FXO	1G1L	1LVW	ugp1
1IIN	289	286 (0.78)	285 (0.79)	289 (0.46)	288 (0.42)	284 (0.77)	280 (0.99)	283 (0.79)	283 (0.75)	283 (0.84)	283 (0.80)	283 (0.81)	273 (1.11)	210 (1.48)
1IIM	98	289	288 (0.26)	286 (0.74)	285 (0.86)	284 (0.60)	277 (1.04)	285 (0.50)	286 (0.58)	285 (0.57)	285 (0.57)	285 (0.53)	274 (1.08)	210 (1.54)
1MP3	97	99	288	285 (0.75)	285 (0.85)	284 (0.61)	277 (1.04)	285 (0.49)	286 (0.59)	285 (0.61)	285 (0.56)	285 (0.52)	274 (1.08)	209 (1.57)
1MP4	99	97	97	289	288 (0.65)	284 (0.76)	280 (1.05)	283 (0.75)	283 (0.75)	283 (0.80)	283 (0.79)	283 (0.77)	273 (1.16)	209 (1.47)
1MP5	99	97	97	99	288	284 (0.88)	280 (1.03)	283 (0.88)	283 (0.85)	283 (0.94)	283 (0.89)	283 (0.92)	273 (1.15)	209 (1.55)
1H5R	90	91	91	90	90	290	279 (1.21)	286 (0.63)	286 (0.68)	286 (0.59)	286 (0.57)	286 (0.62)	273 (1.21)	207 (1.54)
1MC3	62	61	61	62	62	63	289	277 (0.98)	277 (1.02)	278 (1.12)	277 (1.07)	278 (1.02)	277 (1.11)	207 (1.64)
1G0R	72	74	74	72	72	75	63	292	291 (0.35)	292 (0.46)	292 (0.34)	292 (0.17)	278 (1.11)	207 (1.60)
1G23	72	74	74	72	72	75	63	100	292	291 (0.46)	291 (0.43)	291 (0.35)	279 (1.09)	107 (1.54)
1G3L	72	74	74	72	72	75	63	100	100	292	292 (0.40)	292 (0.45)	276 (1.10)	207 (1.58)
1FXO	72	74	74	72	72	75	63	100	100	100	292	292 (0.33)	278 (1.13)	207 (1.62)
1G1L	72	74	74	72	72	75	63	100	100	100	100	292	277 (1.10)	207 (1.60)
1LVW	56	57	57	56	56	57	59	57	57	57	57	57	292	201 (1.61)
ugp1	19	19	19	19	19	19	17	19	19	19	19	19	17	288

The results of the DALI search to locate structural neighbours to the UgpG monomer show similar proteins to cover diverse quaternary structures with monomers and multimeric associations to form dimers and tetramers. However, none of the tetrameric structures shows a quaternary arrangement similar to that found in UgpG, which reveals a new type of oligomerisation to any of its DALI neighbours and indeed to any of the RmlA structures

published to date. Oligomers 1G0R (glucose-1-phosphate thymidyltransferase (Blankenfeldt, Asuncion *et al.* 2000)) and 1QWJ (murine CMP-5-N-acetylneuraminic acid synthetase (Krapp *et al.* 2003)) have RmlA-like tetrameric arrangements (see below *Comparison with existing RmlAs*), the tetrameric arrangement of 1YP2 (ADP-glucose pyrophosphorylase (Jin *et al.* 2005)) is different to that of UgpG, due to an extra subdomain in 1YP2 that is the principal component for forming its monomer-monomer interactions.

Whilst the UgpG tetramer has the four symmetrically related helices (helix 2), one from each monomer, arranged in the centre of the tetramer, all RmlA structures show their oligomerisation equivalent to be helix 9. The interface that creates the dimer in UgpG remains the same as in RmlA, but the interface that builds the dimer of dimers is from the opposite side of the monomer (Figure 6.2 e, f). Consequently, the RmlA tetramers do not contain the long continuous stack of β -chains between adjacent monomers (see *Quaternary structure* section).

A comparison of UgpG with the recently deposited *E. coli* UGP (PDB entry 2D3D) shows 1.4 Å RMSD between homologous C α 's to a cut-off of 3.5 Å, including 238 positions with a percent identity of 35% in a 3D-based sequence alignment. It is noteworthy that none of the four crystallographically independent molecules in the structure of *E. coli* UGP could be completely modelled. Most significantly, their ternary and quaternary structures coincide with those of UgpG.

The analysis of the UgpG catalytic cavity structure and its comparison with that of known RmlA structures reveals considerable similarities in sugar and nucleotide binding, as might be expected for a similar catalytic function. Moreover, in mechanistic terms, conserved residues can be identified in 3D-equivalent positions to those proposed to act as the catalytic residues in

RamlA's, which points to a similar catalytic mechanism. Although space is available at the catalytic cavity of UgpG to accommodate either UTP or TTP, the presence of a flexible loop (residues 219 to 229) that could close the active site and the calorimetry results suggest that protein conformation changes may play an important role for the different specific activities observed *in vitro* (Marques *et al.* 2003; Silva *et al.* 2005), which may be determining for the UgpG preference for UTP.

Note added in proof: Since this paper was submitted the 1.9 Å resolution description of the glucose-1-phosphate uridylyltransferase from *E. coli* has been published by Thoden and Holden (2007a).

6.6. Acknowledgements

The authors thank IBS and the EMBL Grenoble Outstation for the use of high-throughput crystallisation robots, the ESRF for provision of synchrotron radiation and CERMAV for ITC experiments. The authors gratefully acknowledge Anne Imberty (CERMAV) for valuable discussions. D.A. acknowledges a grant from FCT, SFRH (BD/6480/2001). This work was partial supported by FEDER and Fundação para a Ciência e a Tecnologia (FCT), Portugal (POCTI/BME/44441/2002 and POCTI/BIO/58041/2004).

The cloning was done by Ana Rita Marques (ARM). The protein expression, purification, crystallisation, data collection, phase determination, modelling of nucleotide binding, structural comparisons and analysis was done by David Aragão.

7. Conclusions

The work described in this thesis brings new information to the research community in both the understanding of the nature of the hybrid cluster and the possible reductase function for hybrid cluster proteins and the first structure of a bacterial glucose-1-phosphate uridylyltransferase. The key results from four years of scientific research are summarised below.

7.1. *Hybrid cluster protein (HCP)*

Pathogenic organisms are a continuous threat to humans and therefore new methods to efficiently fight them are a permanent issue. Superoxide, nitric oxide and other free radicals are used to kill pathogens by the immune system inside the phagolysosome. They are killed by the activity of digestive enzymes or following a respiratory burst that releases free radicals into the phagolysosome (Ryter 1985; Langermans *et al.* 1994). Some pathogens have developed strategies to avoid the free radicals created by the immune system. The HCP protein is spread through all domains of life and many of the organisms where it is found are responsible human diseases [e.g.: *Clostridium tetani* (tetanus), *Clostridium perfringens* (gas gangrene), *Porphyromonas Gingivalis* (periodontitis), *Trichomonas vaginalis* (trichomoniasis), *Entamoeba Histolytica* (dysentery) , *Bacillus anthracis* (anthrax), *Yersinia pestis* (black death/bubonic plague) and others]. The true function of HCP is not yet established but implications in organism protection against harmful agents have been reported [e.g. reactive nitrogen species (Cabello *et al.* 2004) or oxygen oxidative stress (Almeida *et al.*

2006)]. Hence HCP is a pertinent target for structural and biochemical studies.

The isolation and crystallisation of HCP under anaerobic conditions was carried out and resulted in the elucidation of the high resolution structures of both the anaerobically purified⁽¹⁰⁾ and dithionite reduced HCP forms from *D. vulgaris*. These structures complete the set of information on HCP in various oxidation states: aerobically purified oxidised, anaerobically purified oxidised and anaerobically purified dithionite reduced. This set allows the nature of the hybrid cluster to be followed as the oxidation state changes. The knowledge of both the oxidised and reduced structures allowed the interpretation of the asymmetric unit contents and in the crystallographic refinement of a crystal containing a mixture of oxidised and reduced HCP protein. The anaerobically purified structure was determined by multiwavelength anomalous dispersion at the iron K-edge allowing the determination of the nature and position of the anomalous scattering atoms of and around the hybrid cluster, and served as well as an independent confirmation that the hybrid cluster is not an artefact of *in vitro* oxidation procedures. HCPs share sequence and structure homology to anaerobic CoDs. A detailed comparison between the HCP proteins and the anaerobic CoDs was carried out and showed that the main differences between HCP and CoD proteins lie in the N-terminal domain where the cubane cluster (B-cluster in CoDs) binding region is apparently absent for CoD (Figure 4.3). However, the structures of dimeric and tetrameric CoDs show that this region is present, via monomer B, and has the same fold as in HCP (Figure 4.1c). Both HCPs and CoDs have intriguing and unusual clusters at the interface of their three domains. These clusters, although different, share striking similarities and are linked to the polypeptide chain via homologous

¹⁰ The use of “anaerobically purified” has the same meaning as the expression “as-isolated” in chapter 2.

ligands.

The overall structure of HCP remains the same at different oxidation states. Briefly, the protein comprises three domains: domain 1, a N-terminal mainly alpha-helical domain of two three-helix-bundles and, domains 2 and 3, two similar domains with a central beta-sheet flanked by alpha-helices in a Rossmann fold-like geometry. The protein core contains two iron-sulphur clusters that are approximately 11 Å apart. A conventional [4Fe-4S] cubane cluster binds through its irons to four cysteine residues at the N-terminus of the protein and a second unusual and so-called hybrid cluster binds other four cysteine residues located near the geometrical centre of the protein and at the interface of the three domains.

7.1.1. Structure of anaerobically purified oxidised structure vs that of anaerobically purified dithionite reduced HCP

The anaerobically purified hybrid cluster [4Fe-2S-3O] is composed of a square-like base of two irons, Fe5 and Fe6, and two sulphur atoms, S5 and S6, with the irons bound to cysteine residues, Cys434 and Cys312, and a cap with two irons, Fe7 and Fe8, and three oxygen atoms, O8, O9 and O10, with Fe8 iron bound to Cys406 through a per-sulphide bond and Fe7 bound to Glu494. The anaerobically purified dithionite reduced hybrid cluster [4Fe-3S-O] is composed of the same square-like base with Fe7 virtually unchanged in position. However, the cap Fe8 moves by 2.1 Å and is now directly bound to the SG of Cys406, the position of O10 is now occupied by a sulphur atom (S7), and the two oxygen atoms, O8 and O9, are replaced by a single oxygen species roughly half way between the positions formerly occupied by O8 and O9 in the anaerobically purified oxidised structure.

Comparison of the anaerobically purified oxidised and reduced states, shows that although the protein fold remains the same, there are clear structural changes at the hybrid cluster of the HCP protein upon reduction. A large displacement of Fe8 is associated with the loss of S δ by the per-cysteine 406, and a new covalent bond is formed between its S γ and Fe8. At the position of O10, absent in the reduced structure, now resides a sulphur atom, S7, possibly the S δ of the anaerobically purified oxidised structure, bridging simultaneously Fe5, Fe7 and Fe8.

7.1.2. Structure of aerobically purified oxidised vs that of anaerobically purified oxidised HCP

The structural differences between aerobically isolated HCP protein and anaerobically isolated HCP protein lie mostly in the hybrid cluster. The main change is centred on oxygen O10 (also referred to as X in some publications). For the aerobically isolated structures of HCP from *D. vulgaris*, it was refined in two positions, each with half occupancy and separated by $\sim 1.1\text{\AA}$ (Arendsen *et al.* 1998; Cooper *et al.* 2000; Macedo *et al.* 2002), whilst O10 was refined as a single oxygen position for the anaerobically isolated structures of HCP from two bacteria *D. desulfuricans* (Macedo *et al.* 2002) and *D. vulgaris* (D. Aragão unpublished results - see chapter 4). The structures show that the changes observed around O10 are not due to the fact that the structures come from proteins of *different bacteria*, as initially hypothesised (Macedo *et al.* 2002), but possibly arise from the fact that the structures of the two homologous proteins were obtained in *different oxidation states*. Furthermore, these results raise the possibility of an important role for Fe7 in the binding of substrates, since it is directly involved in binding the labile O10.

7.1.3. Three dimensional comparison of HCPs and carbon-monoxide dehydrogenases

HCP and carbon-monoxide dehydrogenases (CoDs) show a similar polypeptide fold and share a similar non-canonical iron-sulphur cluster, at the interface of their three domains. The most striking differences lie in domain 1 where CoDs lack one of the two orthogonal three-helix bundles present in HCPs, which are responsible for the binding of the HCP cubane cluster. Nevertheless, the formation of the CoD biologically active dimers leads to the localisation of the three-helix bundles at a position equivalent to that found in HCPs where equivalent cubane Fe_4S_4 clusters are located. Additionally, one observes a similar arrangement of clusters, with equivalent distances between the cubane cluster and the main central cluster (C-cluster in CoDs, hybrid cluster in HCPs). However, CoDs have an extra iron-sulphur cluster (Fe_2S_2 in monomeric CoDs and cubane Fe_4S_4 in dimeric or tetrameric CoDs) bound to domain 1, which does not exist in HCPs. The non-canonical iron-sulphur cluster of both HCP and CoD proteins is located at the three-domain interface. They are built of a similar Fe_2S_2 base bonded to the protein through structurally equivalent cysteine residues. In both proteins a third Fe atom (Fe3 in CoDs or Fe8 in HCPs) binds to one of the bridging, inorganic sulphurs, approximately perpendicularly to the common Fe_2S_2 base. However, in CoDs the other bridging sulphur of the base is further bonded to another Fe atom (Fe2), which is also coordinated by a histidine and a cysteine. A structurally based sequence alignment of the two protein types shows these residues to correspond to HCP His244 and Asp268, respectively. However, in HCP they bind Fe7, differently located in the cluster when compared to CoDs, as Fe7 is not directly bound to the Fe_2S_2 base. The upper part of the CoD C-cluster shows clear differences to both the anaerobically purified oxidised, and dithionite reduced hybrid clusters in HCP, but is more closely related to the reduced hybrid cluster. CoDs contain

a Ni atom located close to the Fe7 position in HCP. The Ni atom is covalently bound to the protein through a cysteine residue structurally equivalent to *D. vulgaris* HCP Cys459 (that binds Fe7), and is bound to the cluster through two sulphur atoms. It has been shown that the CoD substrate, CO, binds to the Ni atom. If a parallel situation were to be expected for HCPs, Fe7 would therefore be the Ni equivalent, where the substrate would be expected to bind.

The HCP reaction mechanism may also be similar to CoD reaction mechanism (Lindahl 2004) with the substrate binding to the upper, mobile part of the cluster, followed by its reduction and release of an oxygen atom. Such a mechanism would fit both hydroxylamine reduction (Wolfe *et al.* 2002) or hydrogen peroxide reduction (Almeida *et al.* 2006), as well as other small substrates with potential to be reduced. In light of current knowledge, the HCP substrate could be either a reactive nitrogen or oxygen species and further biochemical studies are needed to access and understand the true nature of the substrate for this enzyme in order to clarify its role in protection against oxidative stress, protection against reactive nitrogen species stress or as a bi-functional enzyme with roles in both.

As shown in the introduction (see Section 1.2.5.a – HCP classification), the HCP protein sequences for *Bacillus anthracis*, *Bacillus cereus*, *Methanobacter thermautotrophicus*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *Thermotoga maritima* and *Wolinella succinogenes* have a deletion of about 110 residues when compared to the HCPs of other organisms, including that of *D. vulgaris*. This gap commences after the cubane binding region at residue 31 (*D. vulgaris* numbering) and comprises a major part of HCP domain 1. Surprisingly, *Methanobacter thermautotrophicus* HCP has an

additional 51 residue insertion at the N-terminus, which is predicted to be a rubredoxin domain by sequence comparisons (Figure 1.4a). It has been shown that CoDs have a similar overall structure to HCPs with the striking differences in the first of the two orthogonal three-helix bundles present in HCP domain 1 (Macedo *et al.* 2002). It has been shown also that upon dimerisation some CoDs are able to produce a homologous three-helix bundle (this work, chapter 4). It is conceivable that for some of class III HCPs, with a gap at the N-terminus, a similar mechanism of quaternary structure formation would enable them to build up such a three-helix bundle motif.

The clarification of the HCP function is now at an important stage. Two different functions have been reported, but a clear answer to the true physiological function is not yet available. To be able to improve our knowledge about this intriguing protein, the search for new substrates from both reactive nitrogen species and oxidative stress generators has to continue. Obtaining the structure and the biochemical characterisation of new HCP proteins could also prove to be important giving new clues and insights about this problem. Probably the most interesting HCP structures to study are those that show peculiarities when compared to the available structures, such as the previously mentioned *Methanobacter thermautotrophicus* HCP.

7.2. Glucose-1-phosphate uridylyltransferase (UgpG)

Glycosylation is pharmacologically important due to the fact that sugar moieties are often determinant in the biological activity of pharmaceutically active metabolites (*e.g.* calicheamicin, amphotericin, megalomicin/erythromycin, mithramycin, doxorubicin and vancomycin).

The inexistence of large and efficient methods to systematically alter target carbohydrate ligands by glycosylation has led to research in this area. Published structural engineering studies on glucose-1-phosphate thymidyltransferases (RmlAs) (Barton *et al.* 2001; Barton *et al.* 2002) are based on nucleotidyltransferase, the first step of a procedure ⁽¹¹⁾ to glycosylate a sugar substrate prior to the glycosyltransferase-catalysed attachment to the aglycon, the nonglycosylated core structure (Barton *et al.* 2001). This type of study is also an important possible route to enhance the industrial use of UgpG since this protein is easily purified in large quantities. Both UgpG and RmlA from *Sphingomonas elodea* ATCC 31461 are able to catalyse *in vitro* the same enzymatic reactions (Silva *et al.* 2005). Several structures of RmlAs from other organisms have been determined to date and all display the same Rossmann-like fold with a nucleotide binding sub-domain plus a sugar binding sub-domain (Blankenfeldt, Asuncion *et al.* 2000; Barton *et al.* 2001; Zuccotti *et al.* 2001; Barton *et al.* 2002; Sivaraman *et al.* 2002).

In this thesis work, protein expression, isolation and crystallisation (both native and seleno-methionine proteins), and subsequent diffraction data collection and phasing (using both Se and Os anomalous data) resulted in the elucidation of the tetrameric structure of glucose-1-phosphate uridylyltransferase with bound glucose-1-phosphate (G1P). This new structure presents similar characteristics to the known RmlA models, with a similar overall fold and shows the conserved active site cavity with similar spatial orientations for the side chains of the catalytic residues. A loop,

11 Substrate glycosylation procedure is a multi-step procedure that starts with sugar activation by nucleotidyltransferase, followed by several enzymatic processing reactions (e.g. deoxygenation, transamination, oxidation/reduction, epimerisation, alkylation and decarboxylation).

formed by residues 219 to 229 and located between helices 7 and 8, closes the active site in all of the RmlA structures, whether substrates are bound or not, whilst in the UgpG structure this loop adopts an “open” conformation allowing access to the active site where G1P was found already bound in place. Based on its mobility, this loop has been suggested to be the gatekeeper to the active site, which on binding both substrates may then help to stabilise the ternary complex (Aragao *et al.* 2007; Thoden and Holden 2007a). Although space is available in the catalytic cavity of UgpG to accommodate either UTP or TTP, the presence of this flexible loop that could close the active site, suggests that protein conformation changes may play an important role in determining the different specific activities (as observed *in vitro*) and thus the UgpG preference for UTP.

Compared to UgpG, all RmlAs have a C-terminal extension of 30 residues forming a three-helix bundle that is absent in UgpG. The volume corresponding to the missing helix bundle is partially occupied by helix 3 in UgpG. This region is particularly significant as it accommodates a second, presumed to be allosteric NTP binding site on some RmlAs. There is, however, no biochemical or sequence comparison evidence to date that UgpG had a similar second site, which is supported by the structure obtained in this work.

Whilst the UgpG tetramer has the four symmetrically related helices (helix 2), one from each monomer, arranged in the centre of the tetramer, all RmlA structures show their oligomerisation equivalent to be helix 9. The interface that creates the dimer in UgpG remains the same as in RmlA, but the interface that builds the dimer of dimers is from the opposite side of the monomer (see section 6.5 and Figure 6.2ef). This results in tetrameric oligomerisation that is different from all the known RmlA. Molecular

replacement methods have not been successful in determining the first structure of a bacterial UGP using the model of any available RmlA structures (Kim, Wu *et al.* 2004; Aragao *et al.* 2006). This is probably due to the different quaternary structure together with small but important differences in the protein fold such as the lack of a second binding site for UTP.

With the determination of the first UGP X-ray structures (Aragao *et al.* 2007; Thoden and Holden 2007a), future structural studies of UGP can focus on the determination of complexed structures with NTP and/or UDP-Glc. For this, crystallisation conditions at high pH are required, since Mg^{2+} is important for nucleotide stabilisation in the active site (Barton *et al.* 2001; Barton *et al.* 2002). Another important objective towards the understanding, at molecular basis, of the UGP properties is to rationalise the UgpG substrate preference for UTP rather than TTP. Mutation studies within the flexible loop (residues 219 to 229) may help to understand its influence on substrate preference, together with complementary structural and calorimetric analysis of the binding of G1P, UTP and TTP.

During this thesis work, the X-ray structure of the UGP from *Corynebacterium glutamicum* crystallised in the presence of both magnesium and UDP-glucose has been reported and supports for the role of the conserved magnesium ion in the catalytic mechanisms of the sugar-1-phosphate nucleotidyltransferases (Thoden and Holden 2007b).

8. Appendices

8.1. Genomic era and structure determination

Over the last ten years there has been a great revolution in terms of the knowledge and availability of protein structures. The number of protein structures deposited in the Protein Databank (PDB) (Berman *et al.* 2000) is growing exponentially (Figure 8.1) giving rise to the need of structure classification methods.

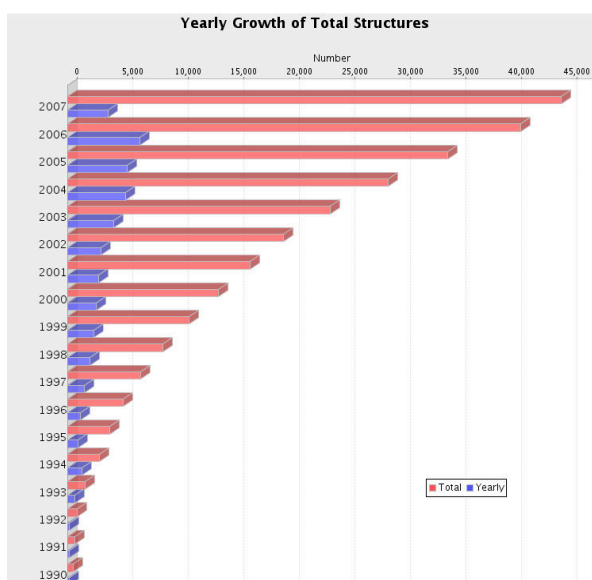


Figure 8.1. *The growth of deposited proteins structures in the PDB.*

Over the last two decades the number of structures deposited has grown exponentially, giving rise to an new wealth of protein structure knowledge (PDB 2007).

The two most used are the Structural Classification of Proteins (SCOP - (Murzin *et al.* 1995)) and the Class, Architecture, Topology and Homologous (CATH - (Orengo *et al.* 1997)) classification. In SCOP, proteins are classified to reflect both structural and evolutionary similarity. The principal levels of classification are family, superfamily and fold (Table 8.1). The exact boundary between these levels is, to some degree, subjective. SCOP

evolutionary classification is generally conservative: where any doubt about similarity exists, new divisions are made at the family and superfamily levels. Thus, some researchers prefer to focus on the higher levels of the classification tree, where proteins with structural similarity are clustered.

Table 8.1. *Main features by levels of the SCOP protein classification*

	SCOP
Fold	Proteins are defined as having a common fold if they have the same major secondary structures in the same arrangement and with the same topological connections. Different proteins with the same fold often have peripheral elements of secondary structure and turn regions that differ in size and conformation. In some cases, these differing peripheral regions may comprise half the structure. Proteins placed together in the same fold category may not have a common evolutionary origin: the structural similarities could arise just from the physics and chemistry of proteins favouring certain packing arrangements and chain topologies.
Superfamily	Proteins that have low sequence identities, but whose structural and functional features suggest that a common evolutionary origin is probable are placed together in superfamilies. For example, actin, the ATPase domain of the heat shock protein, and hexakinase together form a superfamily.
Family	Proteins clustered together into families are clearly evolutionarily related. Generally, this means that pair-wise residue identities between the proteins are 30% and greater. However, in some cases similar functions and structures provide definitive evidence of common descent in the absence of high sequence identity; for example, many globins form a family though some members have sequence identities of only 15%

In CATH, domains are automatically sorted into classes and clustered on the basis of sequence similarities. These groups form the homologous superfamily level of the classification. The topology level is formed by structural comparisons of the homologous groups. Finally, the Architecture level is assigned manually (Table 8.2).

Table 8.2. *Main features by levels of the CATH protein classification*

	CATH
(C) Class	The overall secondary-structure content of the domain. Four classes are defined: mostly-alpha, mostly-beta, alpha and beta, few secondary structures.
(A) Architecture	A large-scale grouping of topologies which share particular structural features.
(T) Topology	High structural similarity but no evidence of homology.
(H) Homologous superfamily	Indicative of a demonstrable evolutionary relationship.

Some of the main hierarchy levels superpose between the two methods (SCOP Fold *vs* CATH Topology and SCOP Superfamily *vs* CATH Homologous Superfamily). The main differences between the methods rely on the subjective definition of what is a domain. A detailed discussion about the resemblances and differences between these two methods of classification can be found in the literature (Hadley and Jones 1999; Day *et al.* 2003).

By the end of June 2007 there were over 40,000 three-dimensional protein structures deposited in the PDB, of which about 37,000 were determined using single crystal X-ray protein crystallography (PDB on-line resource). There are many different factors contributing to this tremendous growth, but the three most striking are:

- easily usable molecular biology techniques for gene cloning and protein expression;
- the increased availability and brilliance of synchrotron X-ray sources;
- jumps in readily accessible computing power.

Since the complete elucidation of the DNA sequence of a number of genomes, new big projects making use of this vast wealth of information have emerged: the structural genomic projects. The aim of these projects is to determine structures for examples of all existing protein folds and get insights about their function. It is thought that this will allow more frequent sequence to structure prediction for unknown structures. In practice this is being carried out by tackling all the protein structures for a given organism or all the protein structures for a given biosynthesis pathway, often across several organisms. This has led in turn to the development of strong automation in areas such as genetic engineering to clone, express and purify as well as crystallise proteins and thence to X-ray diffraction data collection and structure determination. Furthermore, new synchrotrons have brought increased amounts of routine beam time to protein crystallography. Synchrotron radiation is far superior in brightness and collimation than in-house radiation sources. This new exciting era has opened the technique of crystallography to a wide base of scientists who now use crystallography as a base line tool in their research. We can say that protein structure is now reachable to a broad community of researchers.

As mentioned above, sample production and isolation evolved concurrently and keeping pace with the developments in crystallogenesis, data collection and structure solution. New and improved techniques, automation and many other standard protocols are now available to express and purify proteins. One of the techniques that is most used is genetic manipulation of the gene to attach a DNA sequence coding for a tag. A tag is usually a small peptide sequence with high affinity to a particular chromatographic column matrix. A cleavage site can be added between the last protein residue and the peptide sequence tag allowing its removal after purification. The use of this

technique is a major improvement for the purification of large amounts of proteins since it permits purities above 90% to be obtained after the first chromatography step. Reduction of the number of purification steps decreases the cost, isolation time and sample loss. However, tag addition is a protein modification and can pose new problems to crystallisation due to the potential flexibility of the tag and hence related increased heterogeneity in crystal packing, though it can also, in some cases, aid crystallisation. In addition, tags can also affect protein function or protein folding if not removed. This leads to good reasons for the continuation of the use of untagged purification in particular cases. The drawbacks are the use of large volume cell cultures and preparative protein purifications. Another major breakthrough is the availability of reagents and routine methods for the preparation of proteins incorporating selenium-methionine residues. This method can speed up the crystallographic phase determination with the use of the selenium anomalous signal and therefore alleviate the burden of finding suitable conditions for heavy metal incorporation through co-crystallisation or soaking (see below).

8.2. Overview of X-ray crystallographic structure determination

The understanding of protein structure and function has greatly benefited from the developments in X-ray protein crystallography, a technique that can reveal the precise three-dimensional positions of most of the atoms in a protein molecule. Firstly, pure protein is needed for the initial step: the crystallisation. After the first crystal is produced there are three components to an X-ray diffraction experiment: a source of X-rays, the protein crystal and a detector (Figure 8.2). The diffracted intensities measured at the detector are due to three phenomena; namely [for further reading see (Blundell and Johnson 1976; Drenth 1999)]:

- the direct proportionality between the amplitude of the scattered wave and the number of electrons in each atom;
- the atomic arrangement of the structure is responsible for the recombination of the scattered waves;
- all atoms in the crystal contribute to each diffracted beam.

A typical single diffraction pattern of a macromolecular single crystal recorded on one rotation image consists of several thousand spots (Figure 8.2) and as stated above. . Each spot corresponds to a set of diffracted photons resulting from the coherent superposition of the scattered waves from each elemental volume within the whole crystal, and may be mathematically represented by a vector structure factor. The Fourier transform of the amplitudes and phases of all structure factors yields the electron density distribution within the crystal asymmetric unit. The amplitude of the structure factor is proportional to the observed intensity at the spot and can be directly measured. However, the phase cannot be experimentally detected and has to be calculated by a phase determination process (see section 8.6). Once phases are available electron density maps may be calculated, representing the density of electrons for all elemental volumes of the structure being determined. The interpretation of the three-dimensional electron density map leads to the construction of an atomic model. The process by which the model is adapted iteratively corrected by minimising the differences between the observed electron density and the calculated electron density, that produces is known as crystallographic refinement (McRee 1993).

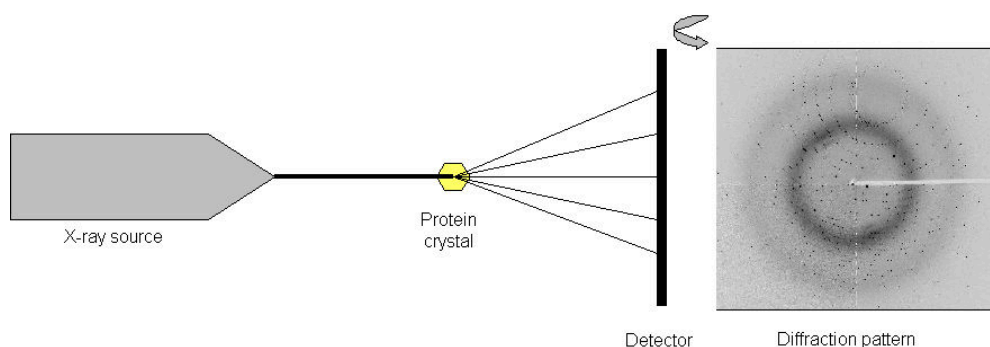


Figure 8.2. *The diffraction experiment.*

Data collection consists of illuminating a protein crystal with a narrow beam of X-rays. Part of this beam goes straight through the crystal but some rays are diffracted and their intensity recorded by an X-ray detector. The image collected at the detector – diffraction pattern – is then saved. The crystal is rotated (typically in steps of one degree or less) and a new image is recorded.

8.2.1. Protein crystallisation

One of the most time consuming steps in the determination of a protein structure is obtaining crystals of sufficient quality (and size) for the diffraction experiment. One can define crystallisation as the process in which a protein is transferred from a solubilised state to a highly organised solid state. The solid, crystalline, state is composed of the three dimensional packing of the protein molecules in a repetitive geometrical arrangement. To succeed in crystallisation one needs to control many different experimental variables from both thermodynamic or to kinetic nature, but their refinement unfortunately still requires essentially only empirical procedures. Crystallisation through vapour diffusion is the most commonly used method requiring quite small amounts of protein, thereby allowing many permutations of variables to be screened with the protein available (Figure 8.3).

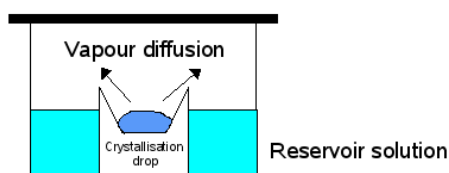


Figure 8.3. *The vapour diffusion technique.*

In a sealed compartment, protein in a lower concentration buffer is allowed to equilibrate against the higher concentration reservoir buffer. In the instance shown above, the configuration is a sitting drop. The protein drop may equally be set hanging from the compartment cover.

This protocol is repeated hundreds or thousand of times with different combinations of buffers and precipitating agents. Variations on parameters such as protein concentration, temperature, pH, ionic strength and additives can also be essayed once a starting point crystallisation has been found. Nowadays, automation is playing an important role in crystallisation since robots are able to produce routinely and reliably protein drops of 100 nl or less and are able to set thousands of conditions in a few hours. The crystallisation experiments have to be checked one by one for crystalline material at a microscope. The process of identifying protein crystals is being automated but non-human identification of crystals has created challenging problems that are expected to be overcome in the near future. Once a suitable sized crystalline result has been found it should be checked for X-ray diffraction and reproducibility. Again, automation has evolved tremendously over the last few years and sample changers allowing auto-mounting of crystals in standardised holders are now routine at many synchrotron beamlines allowing rapid sample screening (Leslie *et al.* 2002).

8.2.2. Space groups

There are 230 space groups (Tables for Crystallography, vol I) that can be organised in seven different crystal systems on the basis of the highest rotation symmetry present (Figure 8.3).

Table 8.3. *The seven crystal systems (adapted (Blundell and Johnson 1976)).*

System	Possible Bravais Lattices	Axes of symmetry	Lattice
Triclinic	P	No axes of symmetry	$a \neq b \neq c$ $\alpha \neq \beta \neq \sigma$
Monoclinic	P, C	1 dyad axis (parallel to b)	$a \neq b \neq c$ $\alpha = \sigma = 90^\circ \neq \beta$
Orthorhombic	P, C, I, F	3 dyad axes mutually orthogonal	$a \neq b \neq c$ $\alpha = \beta = \sigma = 90^\circ$
Tetragonal	P, I	1 tetrad axis (parallel to c)	$a = b \neq c$ $\alpha = \beta = \sigma = 90^\circ$
Trigonal	P (or R)	1 triad axis (parallel to c)	$a = b \neq c$ $\alpha = \beta = 90^\circ$, $\sigma = 120^\circ$ $a = b = c$ $\alpha = \beta = \sigma < 120^\circ$, $\neq 90^\circ$
Hexagonal	P	1 hexad axis (parallel to c)	$a = b \neq c$ $\alpha = \beta = 90^\circ$, $\sigma = 120^\circ$
Cubic	P, I, F	4 triad axes (along diagonals of the cube)	$a = b = c$ $\alpha = \beta = \sigma = 90^\circ$

Fortunately for the protein crystallographer, proteins are chiral and therefore crystals of proteins cannot contain a centre of symmetry which reduces the number of possible space groups down to 65 (Table 8.4).

Table 8.4. *Non-centrosymmetric space groups (adapted from (Blundell and Johnson 1976)).*

System	Class	Space group
Triclinic	1	P1
Monoclinic	2	P2, P2 ₁ , C2
Orthorhombic	222	C222, P222, P2 ₁ 2 ₁ 2 ₁ , P2 ₁ 2 ₁ 2, P222 ₁ , C222 ₁ , F222, I222, I2 ₁ 2 ₁ 2 ₁
Tetragonal	4 422	P4, P4 ₁ , P4 ₂ , P4 ₃ , I4, I4 ₁ P4222, P4 ₂ 1 ₂ , P4 ₁ 22, P4 ₁ 2 ₁ 2, P4 ₂ 22, P4 ₂ 2 ₁ 2, P4 ₃ 2 ₁ 2, P4 ₃ 22, I422, I4 ₁ 22
Trigonal	3 32	P3, P3 ₁ , P3 ₂ , R3 P3 ₁ 2, P3 ₂ 1, P3 ₁ 2 ₁ , P3 ₁ 12, P3 ₂ 12, P3 ₂ 2 ₁ , R32
Hexagonal	6 622	P6, P6 ₅ , P6 ₄ , P6 ₃ , P6 ₂ , P6 ₁ P622, P6 ₁ 22, P6 ₂ 22, P6 ₃ 22, P6 ₄ 22, P6 ₅ 22

System	Class	Space group
Cubic	23 432	P23, F23, I23, P2 ₁ 3, I2 ₁ 3 P432, P4 ₁ 32, P4 ₂ 32, P4 ₃ 32, F4332, F432, F4 ₁ 32, I432, I4 ₁ 32

8.2.3. X-ray diffraction experiment

A diffraction data collection can be broken into several basic steps:

0. Cryocooling. To minimise destructive radiation damage whilst collecting X-ray diffraction data, it is common to cryocool the crystalline sample to 110K or below. This is usually done by soaking the protein crystal in a cryoprotectant prior to crystal flash cooling to avoid the formation of water ice.

1. Indexing. Before a full diffraction data collection is started crystals should be characterised to find their space group and unit cell parameters. This is often done by collecting a few frames at different crystal orientations. For this thesis different software packages were used namely, Mosflm (Leslie 1992), Denzo (Otwinowski and Minor 1997) and XDS (Kabsch 1993). The development of automation is also bringing software to automatically screen, index, and propose collection strategies to X-ray beamlines (for example DNA (2004)).

3. Fluorescence Edge scans. At this point, if the protein contains elements suitable for phase determination (see below), it is important to do a fluorescence wavelength scan in order to decide the most appropriate wavelength(s) for anomalous scattering data collection.

4. Radiation damage. The exposure time is set depending on the main goal of the data collection. For phase determination lower exposure times are best to minimise error in the determination of the Bijvoet reflection pairs whilst, for refinement, higher exposures may result in higher resolution, given the radiation sensitivity of the crystal (which may need to be assessed with a

trial data collection).

5. *Data reduction.* Whilst collecting data it is important to perform data reduction with the partial data being collected to ensure the best data and to allow to corrections to the experiment. For this thesis different software packages were used namely, Scala (CCP4 1994), Scalepack (Otwinowski and Minor 1997) and XDS (Kabsch 1993).

An example of parameters used at a synchrotron beamline is shown in Table 8.3.

Table 8.5. *Typical data collection parameters at a synchrotron beamline of the ESRF.*

Data collection parameters	Value
Wavelength (keV)	9-14
Exposure time (seconds)	0.1-1.0
Oscillation range (°)	0.5-1.0
Number of degrees of data	60-360

8.2.4. Phase determination

Diffraction from a crystalline object results in information that is recovered in the form of intensities. Unfortunately the phases for the diffracted X-ray radiation are not directly measurable although they are crucial to the reconstruction of the structural information. The phase determination step is the process in which that information is recovered. There are several methods that can be used and often they require diffraction experiments with precise experimental conditions such as different wavelengths or upon chemical changes to the proteins under analysis.

a) *Molecular Replacement (MR)*

This method requires prior knowledge of a similar protein model. This is the

case when structures of homologous proteins are accessible. The growing number of structures currently available increased the possibility to use this method. For this thesis different software packages were used namely, Amore (Navaza 1994), Molrep (Vagin and Teplyakov 1997) and Phaser (Storoni *et al.* 2004).

b) Multiwavelength Anomalous Diffraction (MAD) and Single wavelength Anomalous Diffraction (SAD)

These methods require the presence of an ordered element, in the protein crystal, scattering with an anomalous signal sufficiently large to be measured and to tune the X-ray wavelength to a value that depends on the anomalous scatterer. The two methods method became more popular with the availability of molecular biology methods to incorporate selenium, via selenium-methionine, in proteins and the increased availability of tuneable beamlines at synchrotrons. The method relies in the breakage of Friedel's law in the presence of anomalous scatterers. A very accurate determination of the differences between the Bijvoet pairs, usually obtained by increasing the multiplicity of the data, allows the calculation of experimental phases for the protein structure. For this thesis different software packages were used namely, SHELXD/SHELXE (Sheldrick and Schneider 1997), Solve/Resolve (Terwilliger and Berendzen 1999; Terwilliger 2000) and SHARP (Vonnrhein 2005).

c) Multiple Isomorphous Replacement (MIR)

This method requires at least two data sets from a derivative and a native protein crystal. The derivative protein crystal is obtained by soaking with a heavy atom compound (for example mercury chloride). The search for a suitable heavy atom derivative can be difficult since the binding of the heavy atom to the protein has to be such that both crystals remain isomorphous

(they should maintain, to a certain degree, the same crystal unit cell dimensions). This method requires the determination of the heavy metals substructure, which is usually feasible due to the relatively low number of parameters required. With the model of the substructure it is possible to calculate the corresponding structure factors, which in association of both the native and derivative data allows the calculation of starting, experimental phases of the native protein. These phases are used as a starting point to build the protein model.

d) *Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS)*

This method is a combination of the MAD and MIR methods and is used when the heavy atom gives a measurable anomalous signal at the working wavelength. Both information from anomalous scattering and differences in the diffraction pattern intensities due to the heavy metal are used in combination to determine the position of the subset of heavy atoms and calculate experimental phases.

e) *Ab initio/brute force*

This method starts from random phases and uses several assumptions, such as that the electron density is always positive and the crystal contains discretely placed atoms, to act as boundary conditions. The success of such a method is heavily dependent on protein size and the availability of atomic resolution data since the complexity of the problem increases with the number of atoms but the number of experimental data increases with resolution. Commonly used software packaged for this step are SHELXD/SHELXE (Sheldrick and Schneider 1997), SnB (Weeks and Miller 1999) and Acorn (Foadi *et al.* 2000; Yao 2002).

For further reading about the phase problem determination see Drenth

(1999).

8.2.5. Model construction and crystallographic refinement

Once phases are available, the protein model can be built either manually using interactive graphics software, such as COOT (Emsley and Cowtan 2004) or, if the data are sufficiently good, in terms of resolution and phase error, with automatic software, such as ArpWARP (Lamzin and Wilson 1993; Perrakis *et al.* 2001). Model construction is usually not completed in a single step, but is an iterative process combined with refinement of the model against the diffraction data.

Crystallographic refinement is the process by which a protein model is iteratively altered with the objective of minimising the differences between the collected intensities F_{obs} and the intensities calculated from the model under refinement F_{calc} . In this process standard interatomic bond distances, angles (in the form of crystallographic dictionaries) are targets for the final structure, where relative weights between these geometrical parameters and the X-ray data are used to refine the model.

In practice, refinement is the minimisation of a function which may be thought of as composed of several energy terms incorporating both the X-ray information and *a priori* geometric information:

$$E_{\text{total}} = w_c E_{\text{crystallographic}} + w_{bd} E_{\text{bond distances}} + w_{ba} E_{\text{bond angles}} + w_{ta} E_{\text{torsion angles}} \\ + w_{nbc} E_{\text{non-bonded contacts}} + w_{pg} E_{\text{planar groups}} + w_{cv} E_{\text{chiral volumes}}$$

where the w are the weighting terms and E is the total energy for that particular term.

$$E_{\text{crystallographic}} = \sum_{hkl} (F_{\text{obs}} - F_{\text{calc}})^2$$

The crystallographic energy term is calculated from all the differences between F_{obs} and F_{calc} whilst the stereochemical energy terms are calculated using the differences between the actual geometrical value and an ideal

stereochemistry values. These ideal values have been compiled by analysis of high resolution structures of proteins, amino acids and small peptides. The refinement program is set to minimise E_{total} by applying small shifts to all atom coordinates and to the atomic displacement parameters (B -factors) in a method called least squares refinement. The equations for the derivatives necessary for the minimisation of this function are non-linear and which means that they can only be solved iteratively. The main implication of non-linear equations is that the point of convergence may be a local minimum. This problem increases if the starting model is far from the final model. To try to avoid this, refinement should ideally start at medium resolution and gradually add data by increasing resolution. Other methods exist that increase the radius of convergence. In simulated annealing, for example, crystallographic refinement is combined with “Monte Carlo” molecular dynamics overcoming local minimums (McRee 1993).

Today it is rare to use the particular case of least squares refinement and refinement using maximum likelihood ⁽¹²⁾ approaches have been generalised. These permit robust and more reliable refinements using all available data. Several programs have implemented methods of maximising the likelihood function to obtain the most reliable set of parameters, for examples and a more detailed discussion see Bricogne (1997) and Drenth (1999).

The progress of refinement is monitored. The traditional measure of the quality of this fit of the structure to the experimental data is the R-factors. Despite of the stereochemical restraints, it is possible to overfit or 'misfit' the diffraction data: an incorrect model can be refined to fairly good R-factors. In the early nineties, Brünger proposed the use of a statistical quantity (Rfree) that measures the agreement between observed and computed structure factor amplitudes for a test set of reflections that is omitted in the

12 Further reading see the review Aldrich (1997) or the original article Fisher (1912).

modelling and refinement process (Brunger 1992; Brunger 1993). His studies showed high correlation between Rfree and the accuracy of the atomic model phases. This is useful because experimental phase information can be poor, incomplete or unavailable. The R-factor and Rfree are now used broadly in refinement of protein structures.

8.2.6. Structure quality

The overall quality of the final structure can be assessed using the R value, Rfree and the difference between the two. For deeper assessment it is important to check the quality of the fit between the model and the obtained electron density maps, and to check the model geometry against databases of good quality protein structures. These provide averages of bond lengths, bond angles, chiral volumes, statistical Ramachandran analyses among other computed quantities. There are easy to use software packages that perform these checks, and many of them are used in routine validation and in the deposition process of structures to the PDB (Berman *et al.* 2000). Commonly used software packages are Procheck (Laskowski *et al.* 1993) and What If (Vriend 1990; Hooft *et al.* 1996) and checks on the electron density fit may be performed using the Electron Density Server (Kleywegt *et al.* 2004).

References

A to G

- (2004). "DNA: Automated data collection and processing at synchrotron MX beamlines." from <http://www.dna.ac.uk/>.
- Adams, P. D., Pannu, N. S., Read, R. J. and Brunger, A. T. (1997). *Proc Natl Acad Sci U S A* 94(10): 5018-23.
- Aldrich, J. (1997). *Statist. Sci.* 12(3): 162-176.
- Almeida, C. C., Romão, C. V., Lindley, P. F., Teixeira, M. and Saraiva, L. M. (2006). *J Biol Chem* 281(43): 32445-50.
- Andrews, S. C. (1998). *Adv Microb Physiol* 40: 281-351.
- Aragão, D., Fialho, A. M., Marques, A. R., Mitchell, E. P., Sá-Correia, I. and Frazão, C. (2007). *J Bacteriol* 189(12): 4520-8.
- Aragão, D., Macedo, S., Mitchell, E. P., Romão, C. V., Liu, M. Y., Frazão, C., Saraiva, L. M., Xavier, A. V., LeGall, J., van Dongen, W. M., Hagen, W. R., Teixeira, M., Carrondo, M. A. and Lindley, P. (2003). *J Biol Inorg Chem* 8(5): 540-8.
- Aragão, D., Marques, A. R., Frazão, C., Enguita, F. J., Carrondo, M. A., Fialho, A. M., Sá-Correia, I. and Mitchell, E. P. (2006). *Acta Crystallograph Sect F Struct Biol Cryst Commun* 62(Pt 9): 930-4.
- Arendsen, A. F., Hadden, J., Card, G., McAlpine, A. S., Bailey, S., Zaitsev, V., Duke, E. H. M., Lindley, P. F., Kröckel, M., Trautwein, A. X., Feiters, M. C., Charnock, J. M., Garner, C. D., Marritt, S. J., Thomson, A. J., Kooter, I. M., Johnson, M. J., van den Berg, W. A. M., van Dongen, W. M. A. M. and Hagen, W. R. (1998). *J Biol Inorg Chem* (3): 81-95.
- Bailey, S., Cooper, S. J., Hagen, W. R., Arendsen, A. F. and Lindley, P. F. (2001). *Hybrid Cluster Protein. Handbook of metalloproteins*. A. Messerschmidt, R. Huber, T. Poulos and K. Wieghardt, Wiley. 1: 593-601.
- Bamford, V. A., Angove, H. C., Seward, H. E., Thomson, A. J., Cole, J. A., Butt, J. N., Hemmings, A. M. and Richardson, D. J. (2002). *Biochemistry* 41(9): 2921-31.
- Barton, W. A., Biggins, J. B., Jiang, J., Thorson, J. S. and Nikolov, D. B. (2002). *Proc Natl Acad Sci U S A* 99(21): 13397-402.
- Barton, W. A., Lesniak, J., Biggins, J. B., Jeffrey, P. D., Jiang, J., Rajashankar, K. R., Thorson, J. S. and Nikolov, D. B. (2001). *Nat*

- Struct Biol 8(6): 545-51.
- Beinert, H. (2000). *J Biol Inorg Chem* 5(1): 2-15.
- Beinert, H. and Kennedy, M. C. (1989). *Eur J Biochem* 186(1-2): 5-15.
- Beliaev, A. S., Klingeman, D. M., Klappenbach, J. A., Wu, L., Romine, M. F., Tiedje, J. M., Nealson, K. H., Fredrickson, J. K. and Zhou, J. (2005). *J Bacteriol* 187(20): 7138-45.
- Beliaev, A. S., Thompson, D. K., Khare, T., Lim, H., Brandt, C. C., Li, G., Murray, A. E., Heidelberg, J. F., Giometti, C. S., Yates, J., 3rd, Nealson, K. H., Tiedje, J. M. and Zhou, J. (2002). *Omics* 6(1): 39-60.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). *Nucleic Acids Res* 28(1): 235-42.
- Blankenfeldt, W., Asuncion, M., Lam, J. S. and Naismith, J. H. (2000). *Embo J* 19(24): 6652-63.
- Blankenfeldt, W., Giraud, M. F., Leonard, G., Rahim, R., Creuzenet, C., Lam, J. S. and Naismith, J. H. (2000). *Acta Crystallogr D Biol Crystallogr* 56(Pt 11): 1501-4.
- Blundell, T. L. and Johnson, L. N. (1976). *Protein Crystallography*. London, Academic Press.
- Bonofiglio, L., Garcia, E. and Mollerach, M. (2005). *Curr Microbiol* 51(4): 217-21.
- Bourgeois, D., Vernede, X., Adam, V., Fioravanti, E. and Ursby, T. (2002). *J. Appl. Crystallogr* 35: 319-326.
- Brandeen, C. and Tooze, J. (1991). *Introduction to Protein Structure*. New York and London, Garland Publishing Inc.
- Bricogne, E. d. L. F. G. (1997). *Methods in Enzymology, Macromolecular Crystallography*, vol 278, pp472-494. New York, Academic Press.
- Briolat, V. and Reysset, G. (2002). *J Bacteriol* 184(9): 2333-43.
- Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreux, D., Cambillau, C. and Bourne, Y. (1999). *Embo J* 18(15): 4096-107.
- Brunger, A. T. (1992). *Nature* 355: 472-475.
- Brunger, A. T. (1993). *Acta Crystallogr D Biol Crystallogr* 49(Pt 1): 24-36.
- Cabello, P., Pino, C., Olmo-Mira, M. F., Castillo, F., Roldan, M. D. and Moreno-Vivian, C. (2004). *J Biol Chem* 279(44): 45485-94.
- CCP4 (1994). *Acta Crystallogr D* 50: 760-763.
- CancerWEB. "On-Line Medical Dictionary." Retrieved 13 June 2007, from <http://cancerweb.ncl.ac.uk/omd/index.html>.
- Chandrasekaran, R. and Radha, A. (1995). *Trends in Food Sci Technol* 6: 143-148.

- Chang, H. Y., Huang, H. C., Lee, J. H. and Peng, H. L. (1999). *Proc Natl Sci Counc Repub China B* 23(2): 74-84.
- Ciardelli, G., Chiono, V., Vozzi, G., Pracella, M., Ahluwalia, A., Barbani, N., Cristallini, C. and Giusti, P. (2005). *Biomacromolecules* 6(4): 1961-76.
- Cooper, S. J., Garner, C. D., Hagen, W. R., Lindley, P. F. and Bailey, S. (2000). *Biochemistry* 39(49): 15044-54.
- Cowtan, K. (1994). *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31: 34-38.
- Cowtan, K. and Main, P. (1998). *Acta Crystallogr D Biol Crystallogr* 54(Pt 4): 487-93.
- Creighton, T. E. (1992). *Proteins: Structures and Molecular Properties*, W. H. Freeman.
- Crichton, R. (2001). *The Importance of Iron for Biological Systems. Inorganic Biochemistry of Iron Metabolism*. New York, John Wiley & Sons: 1-48.
- Cruickshank, D. W. J. (1999). *Acta Cryst. D* 55: 583-601.
- Daran, J. M., Dallies, N., Thines-Sempoux, D., Paquet, V. and Francois, J. (1995). *Eur J Biochem* 233(2): 520-30.
- Darnault, C., Volbeda, A., Kim, E. J., Legrand, P., Vernede, X., Lindahl, P. A. and Fontecilla-Camps, J. C. (2003). *Nat Struct Biol* 10(4): 271-9.
- Day, R., Beck, D. A., Armen, R. S. and Daggett, V. (2003). *Protein Sci* 12(10): 2150-60.
- de Vocht, M. L., Kooter, I. M., Bultink, Y. B. M., R., H. W. and Johnson, M. K. (1996). *J Am Chem Soc.* 118: 2766-2767.
- DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. San Carlos, CA, USA, DeLano Scientific.
- Dobbek, H., Svetlitchnyi, V., Gremer, L., Huber, R. and Meyer, O. (2001). *Science* 293(5533): 1281-5.
- Doukov, T. I., Iverson, T. M., Seravalli, J., Ragsdale, S. W. and Drennan, C. L. (2002). *Science* 298(5593): 567-72.
- Drennan, C. L., Heo, J., Sintchak, M. D., Schreiter, E. and Ludden, P. W. (2001). *Proc Natl Acad Sci U S A* 98(21): 11973-8.
- Drenth, J. (1999). *Principles of Protein X-Ray Crystallography*. New York, Springer.
- Dutta, T. K., Selifonov, S. A. and Gunsalus, I. C. (1998). *Appl Environ Microbiol* 64(5): 1884-9.
- Eimert, K., Villand, P., Kilian, A. and Kleczkowski, L. A. (1996). *Gene* 170(2): 227-32.
- Emsley, P. and Cowtan, K. (2004). *Acta Crystallogr D Biol Crystallogr* 60(Pt

12): 2126-32.

- Engh, R. A. and Huber, R. (1991). *Acta Cryst. A*(47): 392-400.
- EPA. (2004). "Federal Register Environmental Documents." Retrieved 13 June, 2007, from <http://www.epa.gov/fedrgstr/index.html>.
- Esnouf, R. M. (1999). *Acta Crystallogr D Biol Crystallogr* 55(Pt 4): 938-40.
- Fileenko, N., Spiro, S., Browning, D. F., Squire, D., Overton, T. W., Cole, J. and Constantinidou, C. (2007). *J Bacteriol*.
- Fileenko, N. A., Browning, D. F. and Cole, J. A. (2005). *Biochem Soc Trans* 33(Pt 1): 195-7.
- Fisher, R. A. (1912). *Messenger of Mathematics* 41; 155-160.
- Flatley, J., Barrett, J., Pullan, S. T., Hughes, M. N., Green, J. and Poole, R. K. (2005). *J Biol Chem* 280(11): 10065-72.
- Flores-Diaz, M., Alape-Giron, A., Persson, B., Pollesello, P., Moos, M., von Eichel-Streiber, C., Thelestam, M. and Florin, I. (1997). *J Biol Chem* 272(38): 23784-91.
- Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. and Chao-de, Z. (2000). *Acta Crystallogr D Biol Crystallogr* 56(Pt 9): 1137-47.
- Fraústo da Silva, J. J. R. and Williams, R. J. P. (1991). Non-Haem iron: redox reactions and controls. *The Biological Chemistry of the Elements: the Inorganic Chemistry of Life*. New York, Clarendon, Oxford University Press: 319-342.
- Fredrickson, J. K., Balkwill, D. L., Drake, G. R., Romine, M. F., Ringelberg, D. B. and White, D. C. (1995). *Appl Environ Microbiol* 61(5): 1917-22.
- Frey, P. A. (1996). *Faseb J* 10(4): 461-70.
- Gabrielsen, M., Bond, C. S., Hallyburton, I., Hecht, S., Bacher, A., Eisenreich, W., Rohdich, F. and Hunter, W. N. (2004). *J Biol Chem* 279(50): 52753-61.
- Giaever, H. M., Styrvold, O. B., Kaasen, I. and Strom, A. R. (1988). *J Bacteriol* 170(6): 2841-9.
- Greenwood, N. N. and Earnshaw, A. (1994). *Chemistry of the elements*. Great Britain, Pergamon Press.

H to L

- Hadley, C. and Jones, D. T. (1999). *Structure* 7(9): 1099-112.

- Hagen, W. R. (1989). *J Chem Soc Faraday Trans*(85): 4083-90.
- Harrison, P. M. and Arosio, P. (1996). *Biochim Biophys Acta* 1275(3): 161-203.
- He, Q., Huang, K. H., He, Z., Alm, E. J., Fields, M. W., Hazen, T. C., Arkin, A. P., Wall, J. D. and Zhou, J. (2006). *Appl Environ Microbiol* 72(6): 4370-81.
- Hengge-Aronis, R., Klein, W., Lange, R., Rimmele, M. and Boos, W. (1991). *J Bacteriol* 173(24): 7918-24.
- Heo, J., Wolfe, M. T., Staples, C. R. and Ludden, P. W. (2002). *J Bacteriol* 184(21): 5894-7.
- Higgins, D. G., Bleasby, A. J. and Fuchs, R. (1992). *Comput Appl Biosci* 8(2): 189-91.
- Holm, L. and Sander, C. (1993). *J Mol Biol* 233(1): 123-38.
- Holm, R. (1977). Model complexes and iron sulphur proteins. Biological aspects of inorganic chemistry. A. W. Addison, W. R. Cullen, D. Dolphin and B. R. James. New York, Wiley Interscience: 71-112.
- Hoof, R. W., Vriend, G., Sander, C. and Abola, E. E. (1996). *Nature* 381(6580): 272.
- Hossain, S. A., Tanizawa, K., Kazuta, Y. and Fukui, T. (1994). *J Biochem (Tokyo)* 115(5): 965-72.
- Jay, A. J., Colquhoun, I. J., Ridout, M. J., Brownsey, G. J., Morris, V. J., Fialho, A. M., Leitão, J. H. and Sá-Correia, I. (1988). *Carb Polym* 35: 179-188.
- Jelakovic, S. and Schulz, G. E. (2001). *J Mol Biol* 312(1): 143-55.
- Jin, X., Ballicora, M. A., Preiss, J. and Geiger, J. H. (2005). *Embo J* 24(4): 694-704.
- Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991). *Acta Crystallogr A* 47 (Pt 2): 110-9.
- Kabsch, W. and Sander, C. (1983). *Biopolymers* 22(12): 2577-637.
- Kabsch, W. J. (1993). *J. Appl. Crystallogr.* 26: 795-800.
- Kantardjieff, K. and Rupp, B. (2003). *Protein Sci.* 12(9): 1865-71.
- Kim, C. C., Monack, D. and Falkow, S. (2003). *Infect Immun* 71(6): 3196-205.
- Kim, E. J., Feng, J., Bramlett, M. R. and Lindahl, P. A. (2004). *Biochemistry* 43(19): 5728-34.
- Kim, H., Wu, C. A., Kim, D. Y., Han, Y. H., Ha, S. C., Kim, C. S., Suh, S. W. and Kim, K. K. (2004). *Acta Crystallogr D Biol Crystallogr* 60(Pt 8): 1447-9.
- Kleywegt, G. J., Harris, M. R., Zou, J. Y., Taylor, T. C., Wahlby, A. and

- Jones, T. A. (2004). *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1): 2240-9.
- Koropatkin, N. M., Cleland, W. W. and Holden, H. M. (2005). *J Biol Chem* 280(11): 10774-80.
- Koropatkin, N. M. and Holden, H. M. (2004). *J Biol Chem* 279(42): 44023-9.
- Krapp, S., Munster-Kuhnel, A. K., Kaiser, J. T., Huber, R., Tiralongo, J., Gerardy-Schahn, R. and Jacob, U. (2003). *J Mol Biol* 334(4): 625-37.
- Kraulis, J. P. (1991). *J. appl. Crystallogr* 24: 946–950.
- Krissinel, E. and Henrick, K. (2004). *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1): 2256-68.
- Krueger, R. J. and Siegel, L. M. (1982). *Biochemistry* 21(12): 2892-904.
- Kwak, B. Y., Zhang, Y. M., Yun, M., Heath, R. J., Rock, C. O., Jackowski, S. and Park, H. W. (2002). *J Biol Chem* 277(6): 4343-50.
- Lake, M. W., Temple, C. A., Rajagopalan, K. V. and Schindelin, H. (2000). *J Biol Chem* 275(51): 40211-7.
- Lamzin, V. S., Perrakis, A. and Wilson, K. S. (1999). *International Tables for Crystallography, Crystallography of Biological Macromolecules*. Dordrecht, Kluwer Academic Publishers.
- Lamzin, V. S. and Wilson, K. S. (1993). *Acta Crystallogr D Biol Crystallogr* 49(Pt 1): 129-47.
- Langermans, J. A., Hazenbos, W. L. and van Furth, R. (1994). *J Immunol Methods* 174(1-2): 185-94.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). *J. Appl. Cryst.* 26: 283-91.
- Leslie, A. G., Powell, H. R., Winter, G., Svensson, O., Spruce, D., McSweeney, S., Love, D., Kinder, S., Duke, E. and Nave, C. (2002). *Acta Crystallogr D Biol Crystallogr* 58(Pt 11): 1924-8.
- Leslie, A. G. W. (1992). *Protein Crystallogr.* 26.
- Lindahl, P. A. (2004). *J Biol Inorg Chem* 9(5): 516-24.
- Lindquist, L., Kaiser, R., Reeves, P. R. and Lindberg, A. A. (1993). *Eur J Biochem* 211(3): 763-70.
- Lindqvist, L., Kaiser, R., Reeves, P. R. and Lindberg, A. A. (1994). *J Biol Chem* 269(1): 122-6.
- Lu, W. P., Harder, S. R. and Ragsdale, S. W. (1990). *J Biol Chem* 265(6): 3124-33.

M to Z

- Macedo, S., Mitchell, E. P., Romão, C. V., Cooper, S. J., Coelho, R., Liu, M. Y., Xavier, A. V., LeGall, J., Bailey, S., Garner, D., Hagen, W. R., Teixeira, M., Carrondo, M. A. and Lindley, P. (2002). *J Biol Inorg Chem* 7(4): 514-25.
- Marques, A. R., Ferreira, P. B., Sá-Correia, I. and Fialho, A. M. (2003). *Mol Genet Genomics* 268(6): 816-24.
- Marques Pina, A. R. S. (2003). *Biossíntese e degradação de polissacáridos em Shingomonas paucimobilis ATCC 31461: clonagem dos genes *ugpG*, *bgl1* e *bgl3*, e sua análise computacional e funcional*. Instituto Superior Técnico. Lisboa, Universidade Técnica de Lisboa. Ph.D: 261.
- Martins, L. O. and Sá-Correia, I. (1991). *Biotechnol Appl Biochem* 14(3): 357-64.
- Matthews, B. (1968). *J Mol Biol.* 33(2): 491-7.
- McCoy, J. G., Bitto, E., Bingman, C. A., Wesenberg, G. E., Bannen, R. M., Kondrashov, D. A. and Phillips, G. N., Jr. (2007). *J Mol Biol* 366(3): 830-41.
- McRee, D. E. (1992). *J. Molecular Graphics* 10: 44-6.
- McRee, D. E. (1993). *Practical Protein Crystallography*. San Diego, Academic Press.
- Melo, A. and Glaser, L. (1965). *J Biol Chem* 240: 398-405.
- Merritt, E. A. and Bacon, D. J. (1997). *Methods in Enzymology* 277: 505-524.
- Meyer, O. and Rohde, M. (1984). *Microbial growth on C1 compounds. Proceedings of the 4th international symposium. American Society for Microbiology, Washington.*
- Mollerach, M., Lopez, R. and Garcia, E. (1998). *J Exp Med* 188(11): 2047-56.
- Mosimann, S. C., Gilbert, M., Dombrowski, D., To, R., Wakarchuk, W. and Strynadka, N. C. (2001). *J Biol Chem* 276(11): 8190-6.
- Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997). *Acta Crystallogr D Biol Crystallogr* 53(Pt 3): 240-55.
- Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995). *J Mol Biol* 247(4): 536-40.
- Navaza, J. (1993). *Acta Crystallogr D Biol Crystallogr* 49(Pt 6): 588-91.
- Navaza, J. (1994). *Acta Cryst. A* 50: 157-163.
- Nishikawa, S., Sonoki, T., Kasahara, T., Obi, T., Kubota, S., Kawai, S., Morohoshi, N. and Katayama, Y. (1998). *Appl Environ Microbiol* 64(3): 836-42.

- Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. and Thornton, J. M. (1997). *Structure* 5(8): 1093-108.
- Otwinowski, Z. and Minor, M. (1997). *Methods Enzymol.* 276: 307-326.
- Paule, M. R. and Preiss, J. (1971). *J Biol Chem* 246: 4602–4609.
- PDB. "Protein Data Bank." Retrieved 29 June, 2007, from http://www.pdb.org/pdb/static.do?p=general_information/pdb_statistics/index.html.
- Pearl, F. M., Bennett, C. F., Bray, J. E., Harrison, A. P., Martin, N., Shepherd, A., Sillitoe, I., Thornton, J. and Orengo, C. A. (2003). *Nucleic Acids Res* 31(1): 452-5.
- Peneff, C., Ferrari, P., Charrier, V., Taburet, Y., Monnier, C., Zamboni, V., Winter, J., Harnois, M., Fassy, F. and Bourne, Y. (2001). *Embo J* 20(22): 6191-202.
- Perrakis, A., Harkiolaki, M., Wilson, K. S. and Lamzin, V. S. (2001). *Acta Crystallogr D Biol Crystallogr* 57(Pt 10): 1445-50.
- Perrakis, A., Morris, R. and Lamzin, V. S. (1999). *Nat Struct Biol* 6(5): 458-63.
- Priebe, G. P., Dean, C. R., Zaidi, T., Meluleni, G. J., Coleman, F. T., Coutinho, Y. S., Noto, M. J., Urban, T. A., Pier, G. B. and Goldberg, J. B. (2004). *Infect Immun* 72(7): 4224-32.
- Ragsdale, S. W. and Wood, H. G. (1985). *J Biol Chem* 260(7): 3970-7.
- Ramakrishnan, C. and Ramachandran, G. N. (1965). *Biophys J* 5(6): 909-33.
- Read, R. J. (1986). *Acta Cryst. A* 42: 140-149.
- Richard, S. B., Bowman, M. E., Kwiatkowski, W., Kang, I., Chow, C., Lillo, A. M., Cane, D. E. and Noel, J. P. (2001). *Nat Struct Biol* 8(7): 641-8.
- Roeben, A., Plitzko, J. M., Korner, R., Bottcher, U. M., Siegers, K., Hayer-Hartl, M. and Bracher, A. (2006). *J Mol Biol* 364(4): 551-60.
- Ryter, A. (1985). *Comp Immunol Microbiol Infect Dis* 8(2): 119-33.
- Sá-Correia, I., Fialho, A. M., Videira, P., Moreira, L. M., Marques, A. R. and Albano, H. (2002). *J Ind Microbiol Biotechnol* 29(4): 170-6.
- Saitou, N. and Nei, M. (1987). *Mol Biol Evol* 4(4): 406-25.
- Sali, A. and Blundell, T. L. (1993). *J Mol Biol* 234(3): 779-815.
- Schirmer, F., Ehrt, S. and Hillen, W. (1997). *J Bacteriol* 179(4): 1329-36.
- Schmidt, H. A., Strimmer, K., Vingron, M. and von Haeseler, A. (2002). *Bioinformatics* 18(3): 502-4.
- Sheldrick, G. M. and Schneider, T. R. (1997). *SHELXL: High-Resolution Refinement. Macromolecular Crystallography, part B.* C. W. J. Carter and R. M. Sweet, Academic Press. 277 B: 319-343.

- Sheu, K. F., Richard, J. P. and Frey, P. A. (1979). *Biochemistry* 18(25): 5548-56.
- Siegel, L. M., Rueger, D. C., Barber, M. J., Krueger, R. J., Orme-Johnson, N. R. and Orme-Johnson, W. H. (1982). *J Biol Chem* 257(11): 6343-50.
- Silva, E., Marques, A. R., Fialho, A. M., Granja, A. T. and Sá-Correia, I. (2005). *Appl Environ Microbiol* 71(8): 4703-12.
- Sivaraman, J., Sauve, V., Matte, A. and Cygler, M. (2002). *J Biol Chem* 277(46): 44214-9.
- Steiner, T., Lamerz, A. C., Hess, P., Breithaupt, C., Krapp, S., Bourenkov, G., Huber, R., Gerardy-Schahn, R. and Jacob, U. (2007). *J Biol Chem* 282(17): 13003-10.
- Stokkermans, J. P., Pierik, A. J., Wolbert, R. B., Hagen, W. R., Van Dongen, W. M. and Veeger, C. (1992). *Eur J Biochem* 208(2): 435-42.
- Storoni, L. C., McCoy, A. J. and Read, R. J. (2004). *Acta Crystallogr D Biol Crystallogr* 60(Pt 3): 432-8.
- Streitwiser, A., Heathcock, C. H. and Kosower, E. M. (1992). *Nucleophilic Substitution. Introduction to Organic Chemistry*. New York, Macmillan Publishing Company: 171-303.
- Stryer, L. (1995). *Biochemistry*. 4th edition. New York, W. H. Freeman and Company.
- Sultana, Y., Aqil, M. and Ali, A. (2006). *Drug Deliv* 13(3): 215-9.
- Sutherland, I. W. (1998). *Trends Biotechnol* 16(1): 41-6.
- Terwilliger, T. C. (1994a). *Acta Crystallogr D Biol Crystallogr* 50(Pt 1): 11-6.
- Terwilliger, T. C. (1994b). *Acta Crystallogr D Biol Crystallogr* 50(Pt 1): 17-23.
- Terwilliger, T. C. (2000). *Acta Crystallogr D Biol Crystallogr* 56(Pt 8): 965-72.
- Terwilliger, T. C. and Berendzen, J. (1999). *Acta Crystallogr D Biol Crystallogr* 55(Pt 4): 849-61.
- Thoden, J. B. and Holden, H. M. (2007a). *Protein Sci* 16(3): 432-40.
- Thoden, J. B. and Holden, H. M. (2007b). *Protein Sci* 16(7): 1379-88.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). *Nucleic Acids Res* 25(24): 4876-82.
- Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). *Nucleic Acids Res* 22(22): 4673-80.
- Thorson, J. S., Kelly, T. M. and Liu, H. W. (1994). *J Bacteriol* 176(7): 1840-9.

- Tronrud, D. (1996). Macromolecular Refinement. Proceeding of CCP4 study weekend. Warrington, Daresbury Laboratory.
- Vagin, A. and Teplyakov, A. (1997). *J. Appl. Cryst.* 30: 1022.
- van den Berg, W. A., Hagen, W. R. and van Dongen, W. M. (2000). *Eur J Biochem* 267(3): 666-76.
- van den Berg, W. A. M., Hagen, W. R. and van Dongen, W. M. A. M. (2000). *Eur. J. Biochem.* 267: 666-676.
- Videira, P. A., Cortes, L. L., Fialho, A. M. and Sá-Correia, I. (2000). *Appl Environ Microbiol* 66(5): 2252-8.
- Volbeda, A. and Fontecilla-Camps, J. C. (2005). *Dalton Trans*(21): 3443-50.
- Vonrhein, C., Blanc, E., Roversi, P & Bricogne, G. (2005). Automated structure solution with autoSHARP. Totowa, NJ, Humana Press.
- Vriend, G. (1990). *J Mol Graph* 8(1): 52-6, 29.
- Weeks, C. M. and Miller, R. (1999). *Acta Crystallogr D Biol Crystallogr* 55(Pt 2): 492-500.
- Weissborn, A. C., Liu, Q., Rumley, M. K. and Kennedy, E. P. (1994). *J Bacteriol* 176(9): 2611-8.
- Wolfe, M. T., Heo, J., Garavelli, J. S. and Ludden, P. W. (2002). *J Bacteriol* 184(21): 5898-902.
- Yao, J. X. (2002). *Acta Crystallogr D Biol Crystallogr* 58(Pt 11): 1941-7.
- Zuccotti, S., Zanardi, D., Rosano, C., Sturla, L., Tonetti, M. and Bolognesi, M. (2001). *J Mol Biol* 313(4): 831-43.