

**Structural and biochemical characterization of diheme
c-type cytochromes**

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August-Universität zu Göttingen

vorgelegt von

Daniel Heitmann

aus Paderborn

Göttingen 2008

D7

Referent:

Prof. Dr. Oliver Einsle

Korreferent:

Prof. Dr. Ralf Ficner

Tag der mündlichen Prüfung:

24.10.2008

Table of contents

1	Zusammenfassung	8
2	Summary	10
3	Diheme <i>c</i> -type cytochrome DHC2	12
3.1	Introduction	12
3.1.1	Principles of respiration	12
3.1.2	The heme group	13
3.1.3	Cytochrome characteristics.....	14
3.1.4	Cytochrome <i>c</i> formation.....	16
3.1.5	<i>C</i> -type cytochrome heme configuration	17
3.1.6	Cytochrome <i>c</i> maturation systems	18
3.1.7	Consequences of covalent heme group attachment.....	21
3.1.8	Multi-heme cytochromes.....	22
3.1.9	Structural heme packing motifs	23
3.1.10	<i>Geobacter sulfurreducens</i>	24
3.2	Aim of the present work	26
3.3	Materials and Methods.....	27
3.3.1	Materials.....	27
3.2.1.1	Chemicals	27
3.2.1.2	DNA and Protein Weight marker	27
3.2.1.3	Enzymes and Proteins	27
3.2.1.4	Bacterial strains	27
3.2.1.5	DNA Oligonucleotides.....	28
3.2.1.6	Plasmids	28
3.2.1.6.a	pEC86.....	28
3.2.1.6.b	pET22b(+)	29
3.2.1.7	Growth media	29
3.3.2	Molecular biology methods.....	30
3.3.2.1	PCR-Primer design	30
3.3.2.2	PCR-Amplification	30
3.3.2.3	Agarose gel electrophoresis.....	31
3.3.2.4	Transformation of <i>Escherichia coli</i> competent cells.....	32
3.3.2.5	DNA restriction and ligation reaction	32
3.3.3	Protein biochemistry methods	34

3.3.3.1	Induction test and expression of DHC2 in <i>Escherichia coli</i>	34
3.3.3.2	Preparation of the soluble cell fraction.....	34
3.3.3.3	Affinity chromatography	35
3.3.3.4	Size exclusion chromatography	35
3.3.3.5	SDS-PAGE.....	36
3.3.3.6	Protein concentration determination	37
3.3.3.6.a	Bicinchoninic acid method (BCA).....	37
3.3.3.6.b	Lyophilization.....	37
3.3.4	Protein spectroscopy techniques.....	38
3.3.4.1	UV/Vis absorption spectroscopy	38
3.3.4.2	Electron paramagnetic resonance spectroscopy (EPR spectroscopy)	38
3.3.5	Electrochemical methods.....	41
3.3.5.1	Redox titrations.....	41
3.3.6	Crystallographic methods.....	43
3.3.6.1	Protein crystallization.....	43
3.3.6.2	X-ray diffraction experiments.....	45
3.3.6.3	Cryocrystallography.....	46
3.3.6.4	Data collection.....	47
3.3.6.4.a	General considerations on data collection	47
3.3.6.4.b	Multiple-wavelength anomalous dispersion experiment	48
3.3.6.5	Data processing.....	50
3.3.6.5.a	Indexing.....	50
3.3.6.5.b	Integration.....	51
3.3.6.5.c	Scaling and Merging	51
3.3.6.5.d	Data quality indicators.....	51
3.4	Results.....	55
3.4.1	Recombinant production of DHC2 in <i>Escherichia coli</i>	55
3.4.2	Purification of DHC2	56
3.4.3	UV/Vis absorption spectrum	57
3.4.4	Electron paramagnetic resonance (EPR) spectrum	58
3.4.5	Redox titrations.....	59
3.4.6	Crystallization and data collection.....	60
3.4.7	Structure determination and refinement	62
3.4.8	Protein structure of DHC2.....	63
3.4.9	Folding properties	64
3.4.10	Heme group configuration.....	65

3.4.11	Protein surface	66
3.5	Discussion	68
3.5.1	Refinement process of DHC2.....	68
3.5.2	Significance of DHC2 for the family of multi-heme cytochromes.....	68
3.5.3	Conservation of structural heme packing motifs	69
3.5.4	DHC2 structure in comparison to NapB	70
3.5.5	Biochemical properties of DHC2.....	71
3.5.6	Structural arrangement of DHC2	72
3.5.7	Heme group characteristics of DHC2	73
3.6	Summary	75
4	Twinning in crystals of DHC2	76
4.1	Introduction.....	76
4.1.1	Crystal twinning.....	76
4.1.2	Twinning in monoclinic crystal systems	78
4.1.3	Twinning in biological X-ray crystallography.....	79
4.1.4	The Order-disorder (OD) theory.....	80
4.2	Aim of the present work	84
4.3	Methods	85
4.3.1	Data intensity statistic plots.....	85
4.3.1.1	Cumulative intensity distribution.....	85
4.3.1.2	Twin law dependent twinning tests.....	86
4.3.1.3	Statistics of local intensity differences - L-test.....	88
4.3.1.4	Twin refinement by SHELXL	89
4.3	Results.....	91
4.3.1	Data indexing and merging statistics	91
4.3.2	Cumulative intensity distribution.....	93
4.3.3	Data intensity statistics.....	94
4.3.4	Detection of twinning using the L-test.....	95
4.3.5	Twin law dependent twinning tests.....	96
4.3.5.1	H-test.....	97
4.3.5.2	Britton plot	98
4.3.6	Self rotation function.....	98
4.3.7	Twin refinement using SHELXL.....	99
4.4	Discussion	102
4.4.1	Analysis of twinning	102
4.4.2	Consequence of non crystallographic symmetry for twinning	103

4.4.3	Consequences of NCS for intensity statistics	105
4.4.4	Implications for crystal packing.....	105
4.4.5	Analysis of DHC2 crystal packing using the OD-terminology	107
4.5	Summary	109
5	Elimination of twinning in crystals of DHC2 Δ A81	110
5.1	Introduction.....	110
5.1.1	DHC2 - Mutational studies.....	110
5.1.2	Aim of the present work.....	111
5.2	Material and Methods.....	112
5.2.1	DNA Oligonucleotides.....	112
5.2.2	Site directed PCR mutagenesis	112
5.3	Results.....	113
5.3.1	Crystallization and data collection of DHC2 Δ A81	113
5.3.2	Data processing.....	114
5.3.3	Cumulative intensity distribution.....	115
5.3.4	Data intensity statistics.....	116
5.3.5	Twin law-dependent twinning tests	117
5.3.6	Detection of pseudo-translational symmetry	118
5.3.7	Twinning test for local intensities.....	119
5.3.8	Structure solution by single wavelength anomalous dispersion (SAD)	120
5.3.9	Model building and structure refinement of DHC2 Δ A81	121
5.3.10	Protein structure of DHC2 Δ A81	123
5.3.11	Heme group configuration of DHC2 Δ A81.....	124
5.4	Discussion	125
5.4.1	Comparison of DHC2 Δ A81 and wildtype DHC2 structures.....	125
5.4.2	Comparison of B factor distribution	126
5.4.3	Implications from symmetry	127
5.4.4	Structural reasons for elimination of twinning	128
5.4.5	Changes of packing in the terminology of order disorder	129
5.5	Conclusions and future perspectives.....	131
6	Danksagung.....	132
7	Appendix.....	133
8	References.....	140
9	Publications.....	152
10	<i>Curriculum vitae</i>	153

1 Zusammenfassung

Cytochrome sind eine weit verbreitete Proteinfamilie die Häm als prosthetische Gruppe beinhalten und in Vorgängen der Elektronenleitung oder Redoxreaktionen involviert sind. Das Hämmolekül besteht aus einem aromatischen Porphyrinring und einem zentral gebundenem Eisenatom. Cytochrome des *c*-Typs zeichnen sich durch die kovalente Bindung des Häms an das Proteinrückgrat aus. Ein auf Aminosäureebene konserviertes *c*-Typ Cytochrom Proteinbindemotiv stellt mit zwei Cysteinen, Sulfhydrylgruppen zur Verfügung, an die das Häm mittels zweier Thioetherbindungen angeheftet werden kann. Diese kovalente Bindung an das Protein ermöglicht den *c*-Typ Cytochromen einzigartige Eigenschaften, z.B. eine enge Anordnung von einer Vielzahl an Hämgruppen innerhalb des Proteins. Röntgenkristallographische Strukturaufklärung enthüllte das Vorhandensein zwei verschiedener Hämgruppenanordnungen. Ein paralleles und ein senkrecht Packungsmotiv sind strukturell konserviert. Untersuchungen dieser Packungsmotive werden durch eine große Anzahl von Hämgruppen in Multihämcytochromen erschwert, da sich biochemische und spektroskopische Charakteristika mehrerer Hämgruppen aufsummieren und nicht unterscheidbar sind. Das rekombinante *c*-Typ Dihämcytochrom DHC2 aus dem anaeroben Mikroorganismus *Geobacter sulfurreducens* wurde in dem Bakterium *Escherichia coli* überexprimiert und mittels Affinitätschromatographie bzw. Gelfiltrationschromatographie bis zur Homogenität aufgereinigt. Das Protein wurde mittels UV/Vis-Spektralphotometrie, Elektronenspinresonanzspektroskopie und einer Redox Titration charakterisiert. DHC2 wurde kristallisiert und mittels eines röntgenkristallographischen Beugungsexperiments strukturell charakterisiert. Die Struktur zeigt das Vorhandensein von zwei Monomeren in der asymmetrischen Einheit, sowie zwei kovalent gebundene Hämgruppen in jedem Monomer, die in einem konservierten parallelen Packungsmotiv angeordnet sind. Die strukturelle Anordnung von Aminosäuren an den Hämgruppen spiegeln die Beobachtungen aus der Elektronenspinresonanzspektroskopie wieder, da die Imidazolringebenen der Histidinliganden einen kleinen dihedralen Winkel zueinander aufweisen. Die beobachteten Mittelpunktspotentiale von -135 mV und -289 mV werden durch die spezifische Geometrie der Hämgruppen und ihrer Liganden erreicht. Während der Strukturverfeinerung von DHC2 wurden ungewöhnlich hohe R-Faktoren für den Grad der Verfeinerung festgestellt, dies deutete auf ein Problem innerhalb der DHC2

Kristalle hin. Daher wurde eine intensive Analyse der kristallographischen Daten durchgeführt. Das Ergebnis dieser Untersuchung zeigte, dass die Kristalle von DHC2 pseudo-merohedral verzwilligt sind. Der Kristall besteht aus zwei Zwillingsdomänen in einem Verhältnis von 2:1 mit dem Verzwilligungsgesetz $l, -k, h$. Die pseudo-merohedrale Verzwilligung wird durch eine pseudo-orthorhombische nicht-kristallographische Symmetrie ermöglicht. Der Kristall besteht der Ordnungs- und Disordnungstheorie (OD) zufolge aus Schichten, die durch Translation den Kristall aufbauen. Schwache Bindungen der einzelnen Proteinmoleküle führen zu einer uneindeutigen Translation dieser Schichten. Die Verzwilligung ist somit ein Resultat der Proteinkontakte innerhalb des Kristalls. Eine Mutationsstudie an DHC2 wurde mit dem Zweck durchgeführt diese Verzwilligung zu beseitigen. Die strukturell charakterisierte Mutante DHC2 Δ A81 zeigt dabei ein verbessertes Streuvermögen, eine höhere Symmetrie der Kristallstruktur und eine Beseitigung des Verzwilligungsphänomens. Die strukturelle Anordnung des Proteins unterscheidet sich nur im Bereich der Mutation vom Wildtyp. Das Vorhandensein nicht-kristallographischer Translationssymmetrie, repräsentiert den einzigen Nachteil dieser Mutante, da sie sich negativ auf die Intensitäten der gemessenen Beugungsreflexe auswirkt.

2 Summary

Cytochromes are an abundant family of proteins incorporating heme as a prosthetic group and are involved in processes of electron conduction or redox reactions. Heme consists of an aromatic porphyrin macrocycle and the central coordinated iron ligand. Cytochromes of the *c*-type bind the heme group covalently to the protein backbone. A conserved amino acid heme binding motif is therefore providing two cysteine sulfhydryl groups, to which the heme group is attached via two thioether bonds. The covalent attachment allows for unique properties, such as massive heme group clustering within a protein. X-ray crystallography experiments revealed the presence of a perpendicular and a parallel structural conserved heme group packing motif. Investigations on these motifs are often complicated by the sheer number of heme groups in multi heme cytochromes, as biochemical and spectroscopic data is summing up and therefore indistinguishable. The recombinant diheme *c*-type cytochrome DHC2 from the anaerobic microorganism *Geobacter sulfurreducens* was overexpressed in *Escherichia coli* and purified to homogeneity using affinity- and gel filtration chromatography. The protein was characterized using UV/Vis absorption spectroscopy, electron paramagnetic resonance spectroscopy and redox titrations. DHC2 was crystallized and its structure solved by an X-ray diffraction experiment. The structure is showing two monomers in the asymmetric unit, as well as two covalently bound heme groups per monomer, showing a structural conserved heme packing motif. The structural arrangement of amino acids at the heme groups is reflected by results from the electron paramagnetic resonance spectroscopy, as histidine imidazole planes show a small dihedral angle in mutual plane arrangement. The observed mid-point potentials of -135 mV and -289 mV are a result of the specific geometry of heme groups and their ligands respectively. During the structure refinement process unusual high refinement R-factors were recognized, which were indicating a problem with DHC2 crystals. For this reason, an analysis of diffraction data was carried out, showing the presence of pseudo-merohedral twinning in crystals of DHC2. Two twin domains build up the crystal in a ratio of 2:1 using the twin law $l, -k, h$. The twinning by pseudo-merohedry is allowed due to pseudo-orthorhombic non-crystallographic symmetry. The order disorder (OD) theory describes the crystal as being composed of layers, forming the crystal lattice by translation. The weak binding energies connecting these layers are reflecting weak protein interactions, and are therefore responsible for the ambiguous translation

operation, forming the crystal. These weak contacts seem to be responsible for twinning in the case of DHC2. To eliminate twinning a site directed mutagenesis study was carried out. The structural characterized mutant DHC2 Δ A81 is showing improved diffraction behaviour, higher crystallographic symmetry and no presence of twinning. The structural model is differing from the wild-type model in the region, where the mutagenesis was carried out. The only drawback from this mutation is non-crystallographic translation symmetry having a negative effect on reflection intensities.

3 Diheme *c*-type cytochrome DHC2

3.1 Introduction

3.1.1 Principles of respiration

Life is defined through metabolic activity. The interchange and conversion of matter and energy of cells with the surrounding system characterizes a metabolism. Sustaining metabolic activity requires a source for free energy, which is either energy-rich electromagnetic radiation for photosynthesis or energetically-rich chemical compounds for respiration. The classification 'phototroph' characterizes organisms being able to fuel their need for free energy directly by absorption of sun radiation energy. Organisms being characterized as 'chemotroph' on the other hand make use of chemical energy rich compounds that are either of inorganic (lithotrophs) or organic (organotrophs) source. Compartmentation is a key concept in the upkeep of physiological metabolic processes. Two dimensional layers of bipolarly oriented fatty acids provide a relatively constrictive barrier for hydrophilic, freely moving chemical compounds. In addition to their barrier function, the distinct orientation of lipid bilayers provides a scaffold for insertion and assembly of membrane protein clusters. Mutual organization of protein amino acids and accessory cofactors allows for arrangement of massive redox cluster systems, capable of generating energy for the metabolic cell processes. Examples for this are the photosystems of higher plants or the respiratory chain of mitochondria. Electrons and protons as the smallest elementary particles are utilized in these systems as carriers for free energy. Excitation of electrons to a higher energy state is either achieved by utilizing light radiation or conversion of chemical compounds to a lower energy state. Reactions where electrons are absorbed (reduction) and released (oxidation) are called redox reactions. The redox potential is a currency given in Volt which is related to the free energy of the electron with respect to a reference system by the Nernst-equation, with each compound being of a defined potential. Spatial arrangements of cofactors with decreasing redox potential are able to perform the conduction of electrons, converting the free energy or potential of the electrons into a gradient of protons for example in higher plant photosystems or mitochondria of eukaryotes. The transport of protons from the inner side to the outer side of the membrane is coupled to a decrease of electron free energy. Equilibrium of proton concentrations at the inner and outer

membrane would be energetically favored and is therefore a storage for potential energy over the membrane, which is released in the case of respiration by forming an energy-rich phosphoanhydridic bond between adenosin diphosphate and an inorganic phosphate. The ATP molecule is representing the universal energy currency of the cell. Cofactors of organic and inorganic kind are used in these membrane bound electron conduction chains. But also mobile carriers are present, allowing for transport of electrons at a defined potential and bridging space between various redox cluster assemblies. A prominent example is the soluble and mobile electron carrier cytochrome *c* appearing in the respiratory chain of mitochondria and connecting the cytochrome *bc₁* complex (complex III) to the cytochrome *c* oxidase (complex IV). Cytochrome *c* is a small protein providing the transport of one electron and contains the prosthetic group heme.

3.1.2 The heme group

Iron protoporphyrin IX, or heme, is a very abundant prosthetic group in all taxonomic kingdoms. Heme is characterized by a large organic heterocycle called Proto-porphyrin IX and a central, coordinated iron atom (Gouterman, 1978; Pettigrew & Moore, 1987; Moore & Pettigrew, 1990). Porphyrins are formed by four pyrrole rings interconnected by carbon atoms bridging the rings via a methine group (=CH-). Porphyrin is an aromatic compound with a strongly conjugated bond system, absorbing light with certain maxima and allowing for spectroscopic characterization. Tetrapyrrole closing to a porphyrin ring is synthesized from δ -aminolevulinic acid, which in eukaryotes (except plants and algae) and α -proteobacteria is formed from glycine and succinyl-coenzyme A. Other organisms use the C₅-pathway, where glutamate acts as a precursor for δ -aminolevulinic acid via glutamyl-tRNA (Labbe-Bois *et al.*, 1990; Jahn *et al.*, 1991; Thöny-Meyer, 1997). Ferrochelatase catalyzes the last step of heme synthesis by inserting the central ferrous iron atom into the porphyrin macrocycle. Nitrogen atoms of each pyrrole ring are building the square planar, equatorial ligand field, leaving the two axial coordination sites of the octahedral iron atom ligand field unoccupied. Product of this process is heme (Figure 1). Different types of hemes are formed by chemical modification of the porphyrin macrocycle. Heme a, for example, present in the cytochrome *c* oxidase, differs from the precursor heme by a replaced vinyl side chain at

position 2 against an isoprenoid chain and a methyl side chain by a formyl group at ring position 8.

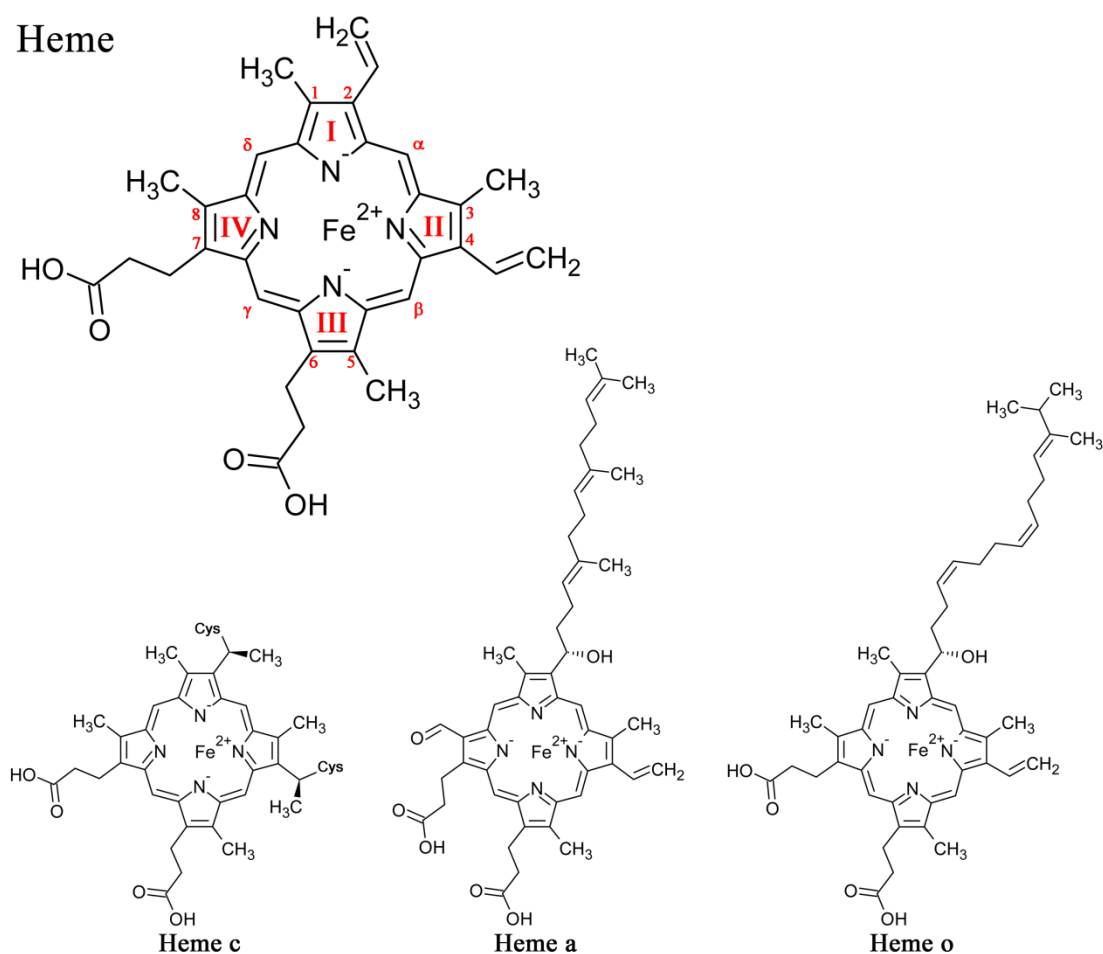


Figure 1: Structure of selected heme group types. Heme groups of type a, b, c and o are shown giving porphyrin nomenclature in red characters, where roman numerals indicate the pyrrole ring. Heme represents the unmodified proto-porphyrin IX, heme of type a, c and o exhibits chemical modified porphyrin macrocycles.

3.1.3 Cytochrome characteristics

Cytochromes are proteins incorporating heme as a prosthetic group and showing a deep red color (greek *chromos* 'colored' and *kytos* 'cell'). Hemoproteins like cytochromes are involved in processes of gas transport, gas detection, electron transfer and enzymatic catalysis. Photosynthetic and respiratory electron transport chains of eukaryotes and also prokaryotic anaerobic metabolisms incorporate a vast number of various cytochromes to ensure energy supply of the cells. Variability of function is not only originating from different types of attached heme groups, but also from the mode of

heme incorporation, properties of surrounding protein environment and the apparent protein ligand field. As the fifth and sixth coordination site of the central iron atom are unoccupied by porphyrin atoms, axial coordination sites can be occupied by amino acid residues or different chemical compounds. Depending on the nature of these axial ligands, the central iron atom can reach oxidation states of +II (ferrous), +III (ferric) and +IV (ferryl), where the electron configuration of this transition metal can be in a low-spin or high-spin state as is explained by the ligand field theory (LFT). In the octahedral ligand field the energy difference Δ_O between orbitals d_{xz} , d_{xy} , d_{yz} (t_{2g} set) and d_z^2 , $d_{x^2-y^2}$ (e_g set) is the reason for different occupation of iron orbitals and is dependent on the attached ligand characteristics, as the energy resulting from splitting of these orbitals can be higher or lower than the energy needed to spin pair an electron (Figure 2). Switching of redox-states allows for redox-activity in enzymatic reactions or conduction of electrons. Redox potentials of cytochromes depend strongly on the properties of the protein environment, on the nature of ligands, heme modifications and ruffling of the porphyrin macrocycle (Walker *et al.*, 1986; Walker, 2004). Cytochromes exhibit complex absorption spectra in the visible and ultraviolet range of light (UV/Vis), as various electronic transitions are responsible for absorption of photons in this range of radiation energy and are resulting from overlaps of the conjugated porphyrin orbitals with iron atom orbitals (Suslick *et al.*, 1992). Prominent bands in UV/VIS spectra represent $\pi \rightarrow \pi^*$ transitions are therefore a useful tool for the characterization of cytochromes, providing useful information about protein concentration, oxidation states, ligand field geometry and heme group surrounding.

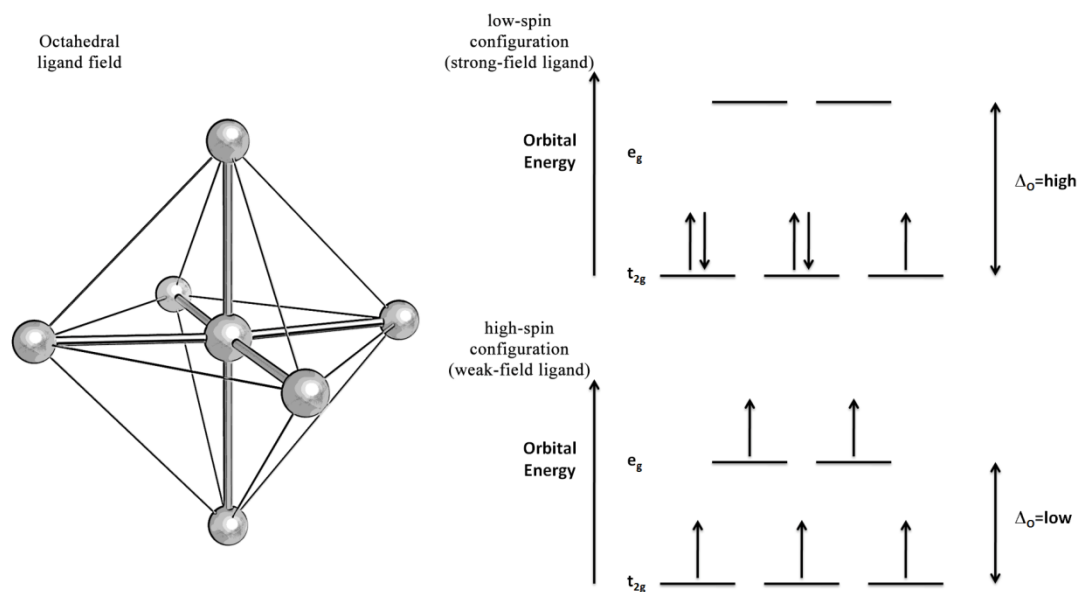


Figure 2: Octahedral ligand field. On the left side, octahedral coordination of a central atom is shown, with an iron atom in central position and four nitrogen atoms forming a common plane, representing the situation of hemes. The axial ligands would reside on the opposite position of this plane, forming the octahedron shape. On the right side two orbital diagrams of iron in the oxidation state of +III present the effect of strong- and weak-field ligands on electron orbital occupancy. Electrons are represented by arrows, indicating the sign of the spin quantum number $1/2$ by pointing into two different directions. If the energy difference Δ_O is higher than the energy needed to fill an already singly occupied t_{2g} orbital, Hund's rule is broken and energetically more favored orbitals are occupied twice before occupying e_g orbitals, this is true for strong-field ligands that induce higher splitting energies.

3.1.4 Cytochrome *c* formation

Heme as an aromatic compound, with only the propionate carboxyl groups exhibiting a hydrophilic moment, can attach to proteins that provide a hydrophobic binding pocket. Number and kind of amino acid residues building such a binding pocket are therefore limiting the quantity and mutual orientation of heme groups in the cytochrome. Free abundance of heme is relative unlikely, as heme is toxic for the cell and tends to aggregate at neutral pH. Heme carrier proteins and pathways facilitating transport inside the cell and over the inner membrane are still unrevealed (Thöny-Meyer, 1997). Cytochromes of the *c*-type represent an exception to most of the former, above points. Heme groups of *c*-type cytochromes are covalently attached to the protein backbone via two thioether bonds, formed by sulfhydryl groups of protein cysteine residues and vinyl groups of the porphyrin macrocycle. Stereospecifically correct attachment of the heme group to the protein moiety is achieved by a complex maturation system that catalyzes

the electrophilic addition of a thiol group to a vinyl group of the porphyrin macrocycle following Markovnikov's rule (Figure 3). Position of the thioether formation is therefore determined by protonation of the less alkylated carbon atom of the vinyl group (Kojo & Sano, 1981; Pettigrew & Moore, 1987). In vivo assembly of *c*-type cytochromes requires a complex maturation system that in contrast to the abundance of *c*-type cytochromes is not ubiquitous in nature.

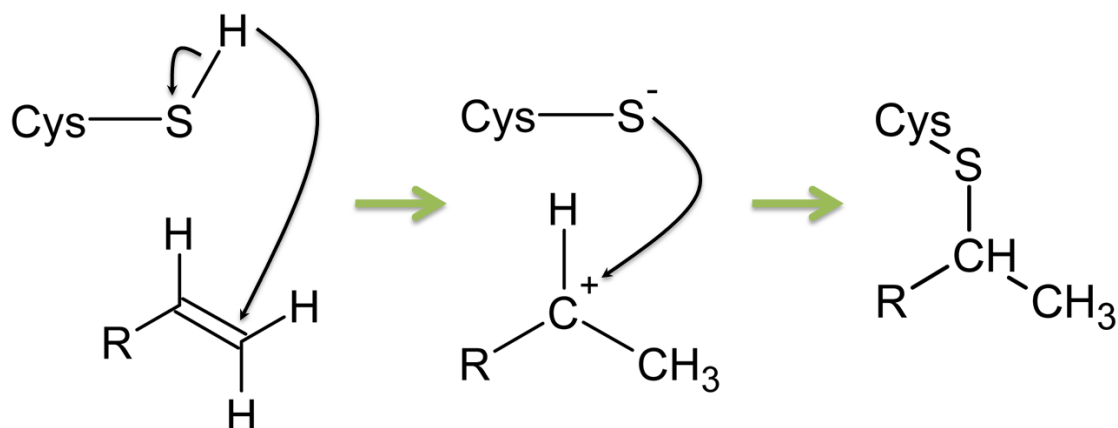


Figure 3: Electrophilic addition of cysteine thiol groups to the vinyl group of a porphyrin resulting in formation of a thioether bond.

3.1.5 *C*-type cytochrome heme configuration

Specificity of heme group attachment to the protein moiety is achieved by recognizing conserved cytochrome *c* binding motifs, normally consisting of an amino acid sequence of CXXCH, where X can be nearly all amino acids (Stevens *et al.*, 2004). Cysteines of this amino acid motif provide the thiol groups for formation of the covalent thioether bonds with the porphyrin vinyl groups. N^ε of the fifth histidine imidazole ring commonly represents the ligand for the fifth coordination site of the iron atom and is called the 'proximal ligand', in accordance with calling the ligand occupying the sixth coordination site the 'distal ligand', as it normally does not originate from a sequence motif of close proximity. Coordination of heme iron by histidine as proximal and histidine as distal or histidine as proximal and methionine as distal ligand are the most common ligand motifs (Barker & Ferguson, 1999; Bushnell *et al.*, 1990). According to the functional needs of the specific protein, these attachment motifs can considerably vary, by introducing different proximal ligands and spacing between the cysteine residues (Figure 4). For example in variants of naturally occurring or even artificial

CXXXCH and CXXXXCH motifs, which are known to have heme attached. Structural data of bacterial cytochrome *c* nitrite reductases show an attachment of four bis-histidinyl coordinated heme groups via the classical CXXCH motif and incorporation of the active center heme via a Cys-Trp-(Ser/Thr/Asn)-Cys-Lys motif, with lysine being an unusual proximal ligand and the sixth coordination site of the high-spin configured iron being free for substrate binding (Einsle *et al.*, 1999; Einsle *et al.*, 2000; Einsle *et al.*, 2001; Bamford *et al.*, 2002). Mitochondria of some kinetoplastid species have only one thioether bond binding the heme to the protein, as the attachment is mediated through a (Phe/Ala)-X₁-X₂-Cys-His motif (Pettigrew *et al.*, 1975).

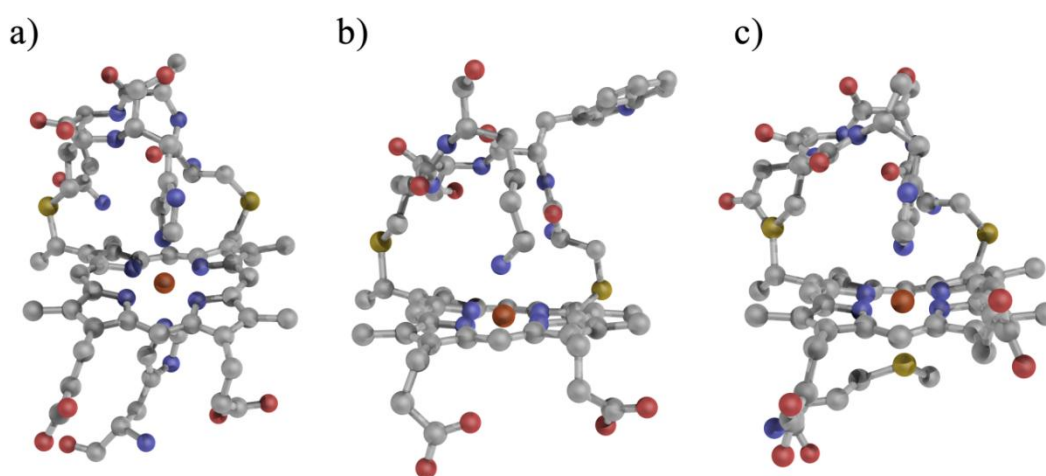


Figure 4: Heme group configuration of different *c*-type cytochrome hemes are shown in ball and stick representation, with carbon atoms colored grey, oxygens red, sulfur yellow, nitrogen blue and iron brown. (a) Bis-histidinyl coordinated heme group III of the cytochrome *c* nitrite reductase NrfA from *Sulfurospirillum deleyianum* (pdb-code 1QDB). (b) Lysine coordinated active center heme group of the cytochrome *c* nitrite reductase NrfA from *Sulfurospirillum deleyianum* (pdb-code 1QDB). (c) Histidine and methionine coordinated heme group in horse heart cytochrome *c* (pdb-code 1HRC).

3.1.6 Cytochrome *c* maturation systems

Presence of *c*-type cytochromes is limited to the outside of the inner membrane, already indicating that the maturation machinery is not located in the cytoplasm or as in the case of mitochondria the matrix. In fact, the cytochrome *c* maturation machinery is an inner membrane bound complex exposed to the outside of the cell or compartment and facilitating the covalent attachment of heme to an apocytochrome in a stereospecifically distinct manner. It is thereby also ensuring a correct folding of the holocytochrome.

Heme itself is synthesized in the cytoplasm or the matrix and has to overcome the inner membrane to reach the point of attachment. Unfortunately an active transporter of heme could not be identified so far. The reaction mechanism of heme group attachment by a part of the complexes (heme lyases) is still elusive. In general there are four types of *c*-type cytochrome maturation systems known and characterized as maturation systems of type I, II, III and IV (Kranz *et al.*, 1998; Thöny-Meyer, 2002; Allen *et al.*, 2003).

- System I

Cytochrome *c* maturation system of type I represent one of the most complex and abundant systems in nature and is widely used in α - and γ -proteobacteria, archea and mitochondria of plants or protozoa. The cytochrome *c* maturation system of *Escherichia coli* is formed by the gene products of *ccmABCDEFGH* and represents the so far best characterized member of this family. Physiologically expressed only under anaerobic conditions in the presence of nitrate, nitrite or trimethylamine N-oxide (Iobbi-Nivol *et al.*, 1994), the membrane bound cytochrome *c* maturation system facilitates the formation of covalent bonds between the apoprotein and the heme group. Transport of the unfolded apoprotein is carried out by the secretory pathway (Duong *et al.*, 1997; Pohlschröder *et al.*, 1997), used for protein export to the periplasm or integration of proteins into the inner membrane space. Although *ccmAB* encodes for an ABC-transporter, it was shown not to be involved in transporting heme across the cytoplasmic membrane. After transport of heme groups via an unknown pathway to the periplasm, CcmC transports, mediated by CcmD, the heme group to the heme chaperone CcmE (Schulz *et al.*, 1998; Ahuja & Thöny-Meyer, 2005). Cysteines of the apoprotein CXXCH motif are reduced by CcmH, which itself is reduced by CcmG and DsbD, before the heme lyase CcmF can facilitate the heme ligation process, removing the heme group from CcmE and attaching it to the CXXCH motif of the apoprotein, forming a holocytochrome *c*. Modular composition of this maturation machinery does also allow for incorporation of heme to unusual binding motifs, like for example in the active site heme group of cytochrome *c* nitrite reductase, which is directly bound to the motif CXXCK, mediated by *ccmFH* paralogs *nrfEG* in *Escherichia coli* (Eaves *et al.*, 1998) or

nrfl in *Sulfurospirillum deleyianum* and *Wolinella succinogenes* (Pisa *et al.*, 2002).

- System II

Cytochrome *c* maturation systems of the second kind are used by Gram-positive bacteria, cyanobacteria, some members of the β -, δ - and ϵ -proteobacteria, as well as in chloroplasts of higher plants and algae. Apoprotein transported by the secretory pathway incorporates heme with the help of the four proteins CcsA, Ccs1, ResA and CcdA, which are seen to form a functional complex, as fusion of orthologous genes in *Helicobacter* implies (Beckman & Kranz, 1993).

- System III

Maturation of cytochrome *c* and *c*₁ in mitochondria of fungi, invertebrates and vertebrates, uses a very simple system, as only one protein namely the cytochrome *c* heme-lyase CCHL is responsible for attachment of heme to apocytochrome *c* and CC₁HL to apocytochrome *c*₁ respectively. Both proteins alone seem to interact directly with the heme group and the apoprotein, forming the thioether bond for this small electron carrier and parts of the cytochrome bc₁ complex (Kranz *et al.*, 1998).

- System IV

The oxygenic photosynthesis process relies deeply on the cytochrome *b₆f* complex, which is containing an unusual, covalently bound heme *c'* comprising a high-spin penta-coordinated iron atom and no bound axial amino-acid ligands. Maturation of cytochrome *b₆f* seems to be dependent on four CCB proteins, which are facilitating the integration of the unusual heme *c'* to the protein. This system differs significantly from system II maturation system and is therefore denoted as type IV (Kuras *et al.*, 2007).

3.1.7 Consequences of covalent heme group attachment

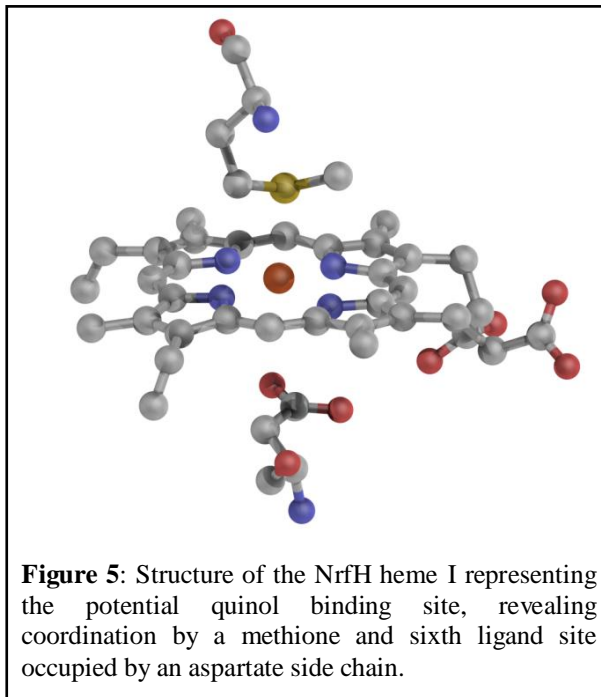
Structural covalent attachment of heme groups via two thioether bonds discriminates *c*-type cytochromes from other cytochrome types that incorporate their heme groups by providing a specific hydrophobic binding pocket. Existence of complex maturation machineries for cytochrome *c* formation raises the question for the overall benefit of organisms attaching heme groups in this manner. Several points may help to elucidate the advantage of maintaining an expensive assembly pathway for *c*-type cytochromes (Barker & Ferguson, 1995).

Cytochromes that are not of the *c*-type are associated to the protein by providing a scaffold of an amino acid environment that matches the hydrophobic nature of the porphyrin macrocycle. The amino acids forming this binding pocket provide a kinetically and thermodynamically stable surrounding, preventing diffusional loss of the heme groups. In contrast to this, heme groups in *c*-type cytochromes are already tightly bound by covalent attachment. In this case amino acid residues form a minimized hydrophobic core giving the heme group a proper functional orientation, reducing the overall protein to heme ratio and is therefore allowing for smaller proteins containing the same amount of heme groups (Einsle *et al.*, 2000).

Without the need for a specific binding pocket and a decreased heme to amino acid ratio, the following properties of *c*-type cytochromes are unique: Folding of *c*-type cytochromes is not dependent on building huge amino acid binding pockets, so covalent attached heme groups can provide a folding core themselves, shaping the overall fold of a protein and therefore can be considered as a kind of secondary structure element. The rate of effective funneling of electrons between redox centers is strongly depending on close distances. Burying of heme groups and surrounding it with a huge number of amino acid residues, like in *b*-type cytochromes, increases the space that has to be bridged between these redox centers and decreases the effectiveness of electron transport processes, as well as potential coupling of heme group energy states. *C*-type cytochromes allow also for closer surface proximity of redox active hemes, making electron transport between proteins more efficient (Page *et al.* 1999).

3.1.8 Multi-heme cytochromes

Direct consequence of all described properties that arise from a covalent attachment in *c*-type cytochrome formation, is the possibility of an arrangement of a huge amount of covalently bound heme groups residing close in protein space. A striking example for massive clustering of heme groups is the multi-heme cytochrome *c* nitrite reductase complex NrfHA from *Desulfovibrio vulgaris* (Figure 6), where 28 heme groups are clustered in the biological relevant asymmetric $\alpha_4\beta_2$ (NrfA/NrfH) complex (Rodrigues *et al.*, 2006). Cytochrome *c* nitrite reductase complex is facilitating the six electron reduction of nitrite directly to ammonia. Electrons used for the reduction process have to be transported from membrane bound quinol molecules to the active site heme group, where the high-spin iron is coordinated by a lysine at position five, leaving the sixth coordination site open for substrate binding and the following reduction process. An unusually coordinated heme is located at the membrane interface of NrfH, and is the probable binding site for quinol to the complex (Figure 5). Although this heme is bound via the CXXCH motif, the proximal ligand is Met49, two amino acids upstream of His47 as seen in protein sequence CXXCHXM, and Asp 89 occupying the sixth iron coordination site. As in the octaheme tetrathionate reductase this represents a heme group setup, where the histidine of the CXXCH motif is not the proximal ligand for the central iron atom (Mowat *et al.*, 2004). It seems imperative that massive heme group clustering is directly related to functional properties in the reduction process of nitrite to ammonia.



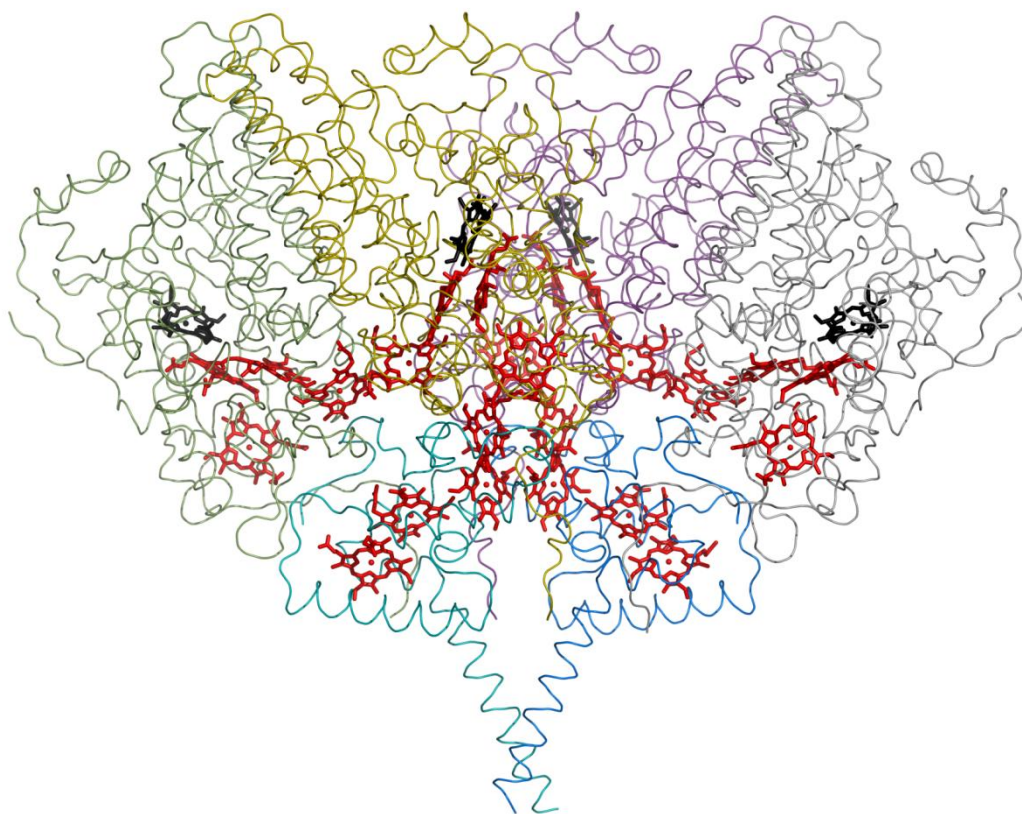


Figure 6: Structure of the cytochrome *c* nitrite reductase complex NrfHA from *Desulfovibrio vulgaris* (pdb-code 2J7A). The protein backbone is shown in ribbon representation, where NrfH has the colors blue and cyan, NrfA is colored in green, yellow, pink and grey. The 28 heme groups are shown in stick representation, with heme groups involved in electron conduction are colored in red and active site heme groups are colored in black.

3.1.9 Structural heme packing motifs

Structural analysis of several multi-heme *c*-type cytochromes such as the hydroxylamine oxidoreductase from *Nitrosomonas europaea* (Igarashi *et al.*, 1997), the tetraheme cytochrome *c*₅₅₄ from *Nitrosomonas europaea* (Iverson *et al.*, 1998), the cytochrome *c* nitrite reductase from *Sulfurspirillum delyianum* (Einsle *et al.*, 1999), the cytochrome *c* nitrite reductase from *Desulfovibrio desulfuricans* ATCC27774 (Cunha *et al.*, 2003), the tetraheme cytochrome *c*₃ from *Desulfovibrio desulfuricans* Essex 6 (Einsle *et al.* 2001) or the octaheme tetrathionite reductase from *Shewanella oneidensis* (Mowat *et al.*, 2004), reveal a strong structural homology in heme group arrangement and relative positioning, while not sharing any remarkable degree of sequence homology. Two relative heme to heme packing motifs were observed in all of the structures, depending on the mutual positioning of the porphyrin planes and thus

showing either a perpendicular or parallel packing arrangement (Figure 7). Functional meaning of heme organization in these motifs, building massive redox clusters, as observed in several multi-heme oxidoreductases, is still unclear. The high degree of structural motif conservation must therefore provide functional advantages, as evolutionary pressure led to only two distinctly conserved motifs (Einsle *et al.* 2001). First speculations about heme packing motif conservation have arisen from structural investigations of the tetraheme cytochrome *c*₅₅₄ (Iverson *et al.*, 1998).

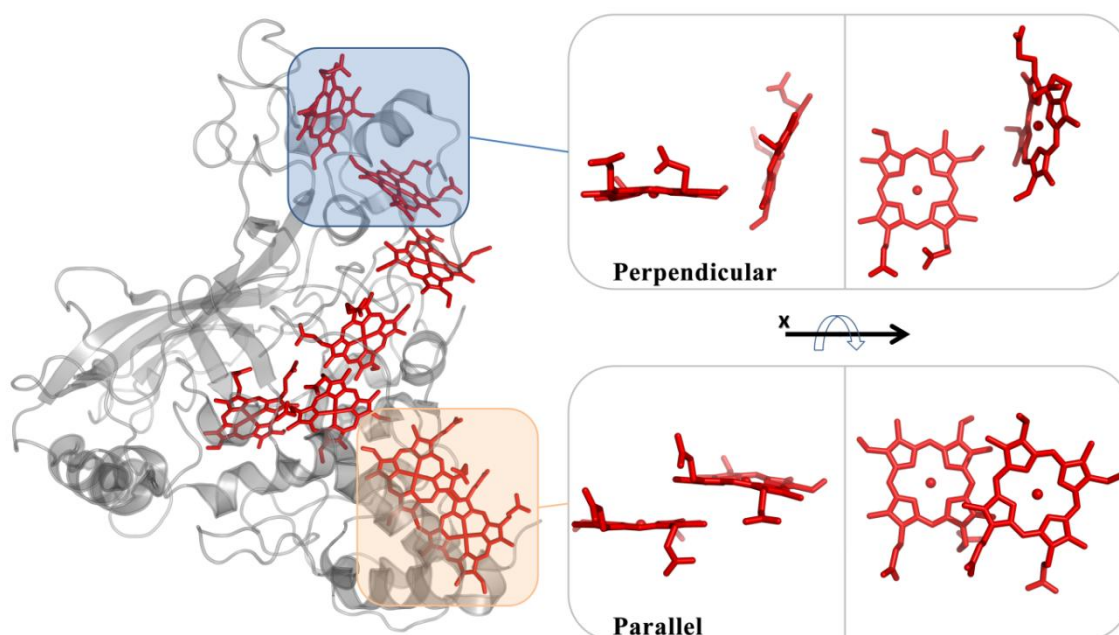


Figure 7: Structural arrangement of heme packing motifs in octaheme tetrathionite reductase from *Shewanella oneidensis* (Mowat *et al.*, 2004). Two different arrangements of heme packing can be observed. In the upper part a perpendicular heme motif is shown in stick representation, whereas the lower part shows a parallel heme arrangement.

3.1.10 *Geobacter sulfurreducens*

Geobacter sulfurreducens is an anaerobically growing soil bacterium capable of reducing metals as terminal electron acceptors, while oxidizing organic compounds (Caccavo *et al.*, 1994). Rod-shaped non-pathogenic *Geobacter sulfurreducens* is belonging to the class of δ -proteobacteria and is of

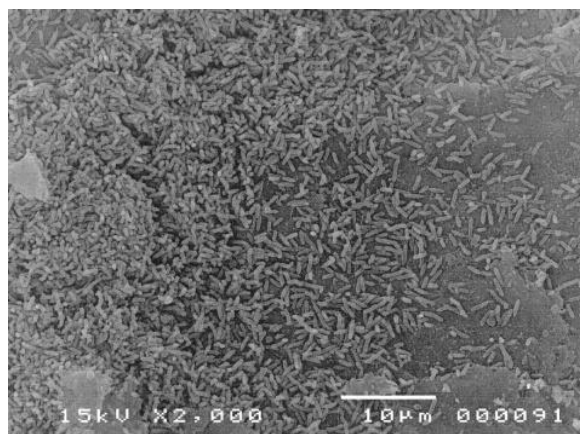


Figure 8: Electron microscopy picture of *Geobacter sulfurreducens* bacteria coating an electrode surface (Bond & Lovley, 2003)

biotechnological relevance, in processes like bioremediation of insoluble metals and generation of electricity by attachment to electrodes (Bond & Lovely, 2003) (Figure 8). The genome comprises about 3.8 million base pairs including at least 111 genes containing one or more cytochrome *c* binding motifs. This finding already indicates the relevance of electron conduction for this species (Methé *et al.*, 2003). Gene sequences having more than two heme binding sites number up to a total of 73, making this organism an ideal source for DNA encoding multi-heme *c*-type cytochromes.

3.2 Aim of the present work

Functional analysis of structural heme packing motif conservation by biochemical methods is often obscured due to the large number of closed packed metal centers. Signals from heme group redox centers merge to an indistinguishable sum, not bearing any information of the specific heme group interactions. Recombinant production of large *c*-type cytochromes on the other hand is not trivial and excludes large scale site directed mutagenesis studies on these huge multi-heme cytochromes, limiting the so far existing work of this kind to a few studies (Wood *et al.*, 1988; Mus-Veteau *et al.*, 1992). Simplifying the system of interest is therefore a crucial strategy in characterizing a *c*-type cytochrome containing two heme groups that can be described by structural and spectroscopic methods at the same time. Obtaining good yields of this *c*-type cytochrome for experimental procedures requires an efficient system for recombinant production in a system such as the ϵ -proteobacteria *Escherichia coli*. This organism also allows for straightforward protein engineering by site directed mutagenesis. Structural data showing one of the two conserved heme packing motifs would be an essential result for the first characterization of such a model system. In the present work, structural and biochemical characterization of a diheme *c*-type cytochrome, named DHC2, from the metal-reducing δ -proteobacteria *Geobacter sulfurreducens* is described. Physiological role of this *c*-type cytochrome is unknown and represents one out of 11 diheme cytochromes encoded in the genome of *Geobacter sulfurreducens*, which is showing similar codon usage to *Escherichia coli*.

3.3 Materials and Methods

3.3.1 Materials

3.2.1.1 Chemicals

All chemicals used were of analytical purity grade (p.a.) and have been obtained from the companies Applichem (Darmstadt, Germany), Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany) and Sigma-Aldrich (Deisenhofen, Germany).

3.2.1.2 DNA and Protein Weight marker

Size determination of DNA fragments was achieved by using the 1kb and 100 bp DNA ladder and for protein size determination the Broad Range protein marker (New England Biolabs, USA).

3.2.1.3 Enzymes and Proteins

Restriction endonucleases NdeI and XhoI, *Pfu* DNA polymerase and T4 DNA ligase were obtained from New England Biolabs (Ipswich, USA).

3.2.1.4 Bacterial strains

XL10-Gold (Stratagene, USA):

*Tetr*Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1*

BL21(DE3) (Novagen, Darmstadt, Germany):

[*F*- *ompT hsdSB (rB- mB-) gal dcm (DE3)*]

3.2.1.5 DNA Oligonucleotides

DNA-oligonucleotides were synthesized and purified by Eurofins MWG Operon (Ebersberg, Germany).

dhc2_flp_NcoI 5'- G GCC ATG GTC TCC GGT GAA GTA AGG A -3'

dhc2_rH_XhoI 5'- CC CTC GAG CTT GAA CTT GTG GCA G -3'

3.2.1.6 Plasmids

3.2.1.6.a pEC86

The *pEC86* plasmid encodes for the cytochrome *c* maturation system (*ccm*) of *Escherichia coli*, which is physiologically active only under anaerobic conditions and now under control of a constitutive *tet* promoter (Figure 9), allowing on one hand for cytochrome *c* expression under aerobic conditions and effective cell growth of *Escherichia coli* on the other hand (Thöny-Meyer *et al.*, 1995). Besides the *ccm* system, it encodes for a chloramphenicol resistance marker. The *pEC86* plasmid was derived from the *pACYC184* plasmid and was a generous gift from Prof. Dr. Linda Thöny-Meyer (EMPA, St. Gallen).

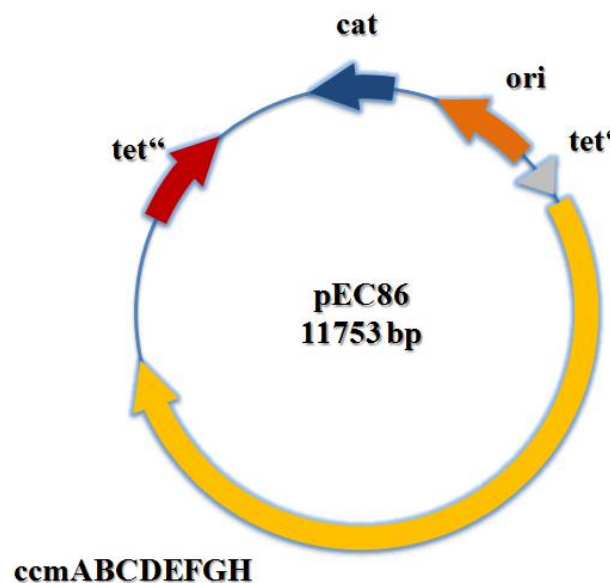


Figure 9: Schematic representation of *pEC86* plasmid organisation. Different colored arrows represent organisation elements of *pEC86*, with *tet'* as the *tet* promoter, *tet''* as a tetracycline resistance marker, *cat* as a chloramphenicol resistance marker, *ori* as the origin of replication region and *ccmABCDEFGHIH* showing the cytochrome *c* maturation genes of *Escherichia coli*.

3.2.1.6.b pET22b(+)

The plasmid *pET22b(+)* was obtained from Novagen (Darmstadt, Germany) and used because of its N-terminal *pelB* leader sequence, which is a signal peptide for the secretory pathway of *Escherichia coli* and is cleaved off after protein precursor transport to the periplasm by a signal peptidase (Figure 10). By elimination of the *dhc2* stop codon, the C-terminal hexa-histidine tag is fused to the target gene product.



Figure 10: Schematic representation of the pET22b(+) open reading frame. The N-terminal *pelB* leader sequence allows for efficient periplasmatic transport, *dhc2* was cloned using NcoI and XhoI, omitting the endogenous *Geobacter sulfurreducens* leader sequence and the stop codon for use of the hexahistidine affinity tag.

Omission of the *pelB* leader sequence by cloning with NcoI led to two artificial N-terminal amino acids namely Met23 and Val24.

3.2.1.7 Growth media

Luria-Bertani (LB)-Medium

For Luria-Bertani medium 10 g of Tryptone, 5 g of Yeast extract and 10 g NaCl were dissolved in dH₂O to a total volume of 1 litre. Medium was sterilised by autoclaving at 121°C for 20 min.

3.3.2 Molecular biology methods

3.3.2.1 PCR-Primer design

Primers for PCR amplification of the *dhc2* gene (annotated gene GSU2927; <http://tigr.org>) were designed by elimination of the signal peptide sequence for the secretory pathway of *Geobacter sulfurreducens* from the PCR product and providing a *Nco*I DNA restriction enzyme recognition site at the 5'- and a *Xho*I site at the 3'-end of the PCR product. Additional to this modification, the stop codon was omitted in the 3'-Primer.

3.3.2.2 PCR-Amplification

Polymerase chain reaction (PCR) is an in-vitro technique for exponential amplification of defined DNA-fragments (Mullis and Faloona, 1987; Saiki *et al.*, 1988). Synthesized DNA oligonucleotides serve as priming DNA fragments for a thermo-stable DNA polymerase, such as *Taq* polymerase from *Thermus aquaticus* or *Pfu* polymerase from *Pyrococcus furiosus*, elongating primer DNA depending on bound template DNA. Amplification of distinct DNA fragments makes use of binding primers to the forward (5'->3') and reverse (3'->5') strand of DNA, forming a boundary for the polymerase reaction and leading to the predominant amplification of desired DNA fragments. PCR thermocyclers sequentially heat and cool the PCR-reaction depending on the experimental settings, to ensure a serial process of DNA denaturation, primer annealing and primer elongation. Amplification of PCR products identical to template DNA sequence requires use of DNA polymerases like *Pfu* that provide a 3'->5' proof reading activity. PCR reaction using *Pfu* polymerase (Stratagene) was controlled with a Biometra thermocycler (Biometra, Göttingen). As a template genomic DNA was used, extracted from *Geobacter sulfurreducens* cells (Ausubel *et al.*, 1990).

No.	Temp. °C	Time sec
	95	300
	95	45
1x	64	45
	72	240
	95	45
1x	62	45
	72	240
	95	45
1x	60	45
	72	240
	95	45
20x	52	45
	72	240
	72	300
15	4	end

Table 1 PCR thermocycler program for specific amplification of *dhc2* DNA

Volume (μl)	
2.5	dNTP's
1	5' - Primer 100 μM
1	3' - Primer 100 μM
0.5	genomic DNA
2.5	Pfu buffer (Stratagene)
0.5	Pfu DNA polymerase (1500 U/ml)
17.5	ddH ₂ O

Table 2 PCR reaction mixture for specific amplification of *dhc2* DNA

3.3.2.3 Agarose gel electrophoresis

Agarose gel electrophoresis in a defined buffer system was used for separating DNA fragments by their different electrophoretic mobility u in an electric field given by:

$$u = \frac{v}{E} = \frac{ze}{f}$$

where v is the velocity of the molecule, E is the electric field strength, z the number of charges, e the elemental charge and f the coefficient of friction with:

$$f = 6\pi\eta r$$

where η is the viscosity of the medium (agarose + buffer) and r is the Stokes radius. Particles are accelerated in the direction of the anode for negatively charged molecules. Polymerisation of agarose in a defined buffer system provides a matrix of defined average pore size depending on the concentration of agarose. This allows for separation of fragments reflecting their mutual sizes, as smaller molecules migrate faster through the gel matrix than larger molecules do. DNA markers allow for size estimation of observed bands. Gels were made using TBE buffer with 1.1% agarose, samples were

mixed with 6x loading dye (Fermentas) and fluorescence dye SYBR green (Molecular Probes) for band detection. Electrophoresis procedure was carried out in a horizontal gel system (Bio-Rad) at 50 V. UV-fluorescence gel pictures of bands stained with SYBR green were taken with a Geldoc EQ system (Bio-Rad).

TBE buffer		DNA sample		6x loading dye	
Konz. (mM)		Vol. (µl)		Konz.	
89	Tris base	3	DNA sample	10 mM	Tris-HCl (pH 7.6)
89	boric acide	1	6x loading dye	0.03%	bromophenol blue
2	EDTA	1	ddH2O	0.03%	xylene cyanol FF
		1	SYBR green	60%	glycerol
				60 mM	EDTA

Table 3: Composition of TBE buffer, DNA sample mixture and 6x loading dye.

3.3.2.4 Transformation of *Escherichia coli* competent cells

Chemically competent cells of *Escherichia coli* were transformed with vector DNA, by adding 1µl volume of plasmid (100-400 µg/ml) to cells refrigerating on ice. Cells and plasmid were incubated for 30 min and heat shocked at 37°C for 10 min, before chilling cells 10 min on ice. By adding 200 µl of LB and incubation at 37°C cells could express antibiotic resistance markers encoded by plasmid or genomic DNA.

3.3.2.5 DNA restriction and ligation reaction

For ligation of *dhc2* PCR products into the pET22b(+) vector system (Novagen), vector and insert were cut with restriction enzymes *NcoI* and *XhoI* at 37°C for two hours. Vector pET22b(+) is coding for an N-terminal signal peptide sequence, that can be ligated to the target gene for periplasmic export of recombinant protein via the secretory pathway of *Escherichia coli*. Omission of the stop codon in PCR amplification, allows for use of the pET22b(+) encoded C-terminal hexa-histidine affinity tag. After heat inactivation of restriction enzymes at 65°C for 10 min, vector and insert were ligated to each other in a ratio of 1:1 and 1:3 at 4°C over night and making use of sticky DNA ends produced in the process of DNA digestion. Ligation reactions were transformed

into *Escherichia coli* strain XL10(Gold) (Novagen) and plated onto LB agar plates containing ampicillin (Amp) and kanamycin (Kan) antibiotics, thus selecting for *Escherichia coli* cells containing the pET22b(+):dhc2 construct. Single cell colonies were picked after incubation at 37°C over night and transferred to 5ml LB culture containing Amp and Kan. Plasmid DNA was isolated after incubation over night at 37°C using the plasmid purification kit NucleoSpinPlasmid (Macherey-Nagel). Correctness of pET22b(+):dhc2 sequence was proved by agarose gel electrophoresis of restriction digestion controls and DNA sequence reaction (Sanger *et al.*, 1977).

<u>Insert</u>		<u>Vector</u>	
<u>Vol. (µl)</u>		<u>Vol. (µl)</u>	
1	NEB buffer # 4	1	NEB buffer # 4
1	Nco I	1	Nco I
1	Xho I	1	Xho I
7	PCR product	7	Vector pET-22b(+)

Table 4: Digestion reaction mixture for insert and vector DNA.

<u>1:3</u>		<u>1:1</u>	
<u>Vol.</u>		<u>Vol.</u>	
1 µl	T4 ligase buffer	1 µl	T4 ligase buffer
1 µl	T4 ligase	1 µl	T4 ligase
2 µl	Vector pET-22b(+)	2 µl	Vector pET-22b(+)
6 µl	Insert	2 µl	Insert
		4µl	ddH ₂ O

Table 5: Ligation reaction mixture for vector and insert DNA ratios of 1:3 and 1:1.

3.3.3 Protein biochemistry methods

3.3.3.1 Induction test and expression of DHC2 in *Escherichia coli*

Large yields in protein expression of pET22b(+):dhc2 gene product DHC2 in *Escherichia coli* cells are necessary for large scale protein purification. Therefore pET22(+):dhc2 was transformed into *Escherichia coli* strain BL21(DE3), already containing the accessory plasmid pEC86 (Thöny-Meyer *et al.*, 1995). Induction test of the Lac-promoter controlled expression of pET22(+):dhc2 was performed using 0, 0.1, 0.5 and 1 mM of Isopropyl β -D-1-thiogalactopyranoside (IPTG) in 5 ml LB, supplemented with chloramphenicol and ampicillin. Test expression cultures were grown over night at 30°C. Upscaling of this test expression was achieved by growth of BL21(DE3) transformed with pEC86 and pET22b(+):dhc2 in 6 litres of LB medium at 30°C over night without any further addition of IPTG.

3.3.3.2 Preparation of the soluble cell fraction

Cells grown in expression cultures were harvested using an Avanti J-20XP centrifuge (Beckman Coulter) at 4,000g, for 10 min and at 4°C. After washing the cell pellet in 40 ml of ice-cold PBS, cells were resuspended in 48 ml of purification buffer containing 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl. Disruption of cells was carried out using an M-110S laboratory microfluidizer (Microfluidics). This method utilizes high-pressure streams that collide at high-velocities in defined microchannels, with cells getting sheared and are colliding, which is resulting in a very fine dispersion of cell debris and soluble cell fractions. Centrifugation at 20,000g for 15 min at 4°C was used to remove cell debris and an ultracentrifugation step at 100,000g for 60 min at 4°C was carried out for removal of the insoluble membrane fraction, as the inner-membrane bound cytochrome *c* maturation system could still bear unprocessed precursors of recombinant DHC2.

3.3.3.3 Affinity chromatography

Purification of DHC2 by affinity chromatography was carried out at 20°C, using an Äkta prime chromatography system (GE Healthcare) and the progress was followed by plotting UV absorption at 280 nm against the retention time. The soluble cell fraction was loaded to a nickel-loaded HiTrap HP 5ml column (GE Healthcare) pre-equilibrated with purification buffer, where the hexa-histidine affinity tag of recombinant DHC2 was supposed to chelate sepharose-matrix coordinated Ni²⁺ ions (Figure 11). Removal of unspecifically bound protein from the HiTrap column material was performed by a stepwise increase of the imidazole concentration in the purification buffer to 10 mM, thus competing for Ni²⁺ binding sites and displacing unspecific interacting material (Graves & Wu, 1974). After adequate washing of the column with imidazole-containing buffer, an increase to 100 mM of imidazole concentration was used for elution of red protein from the column material. Fractions containing red protein were pooled and concentrated to a final volume of 1.5 ml by using an ultrafiltration Vivaspin concentrator (Sartorius) with a molecular weight cut-off of 10,000 Da.

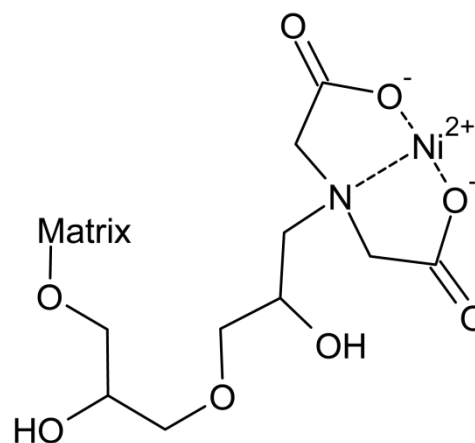


Figure 11: Sepharose matrix bound chelated Ni²⁺ ion.

3.3.3.4 Size exclusion chromatography

The method of size exclusion chromatography separates different molecules depending on their hydrodynamic volume, which represents the volume of particles in liquid solution. Polymeric gel material (Superdex 75) building a matrix of defined pore size diameters, that allows for differential flow-rates of soluble molecules. The probability distribution of molecules in this matrix volume is reciprocally proportional to their own volume. Depending on their size, molecules can access different total volumes of column material during the chromatography run. When applied in a sufficiently small injection volume, molecules of same size will elute at the same time from the column, with all effects of peak broadening possible. Large molecules will therefore elute faster

from the column than smaller molecules. Molecules having a hydrodynamic volume larger than what is accessible for the polymer matrix, will elute from the column at the same time independently from their mutual sizes, in what is called the exclusion volume.

The concentrated protein fraction was applied to a Superdex 75 HiLoad 26/60 gelfiltration column (GE Healthcare) equilibrated with purification buffer at a flow rate of 1 ml/min and followed by plotting UV absorption at 280 nm against the retention time. Red coloured protein fractions were pooled and concentrated to a volume of about 1 ml. The concentrated protein was applied to a PD-10 gravity flow desalting column, packed with Sephadex G-25 material (GE Healthcare) and equilibrated with water for removal of any soluble buffer and salt molecules.

3.3.3.5 SDS-PAGE

Polyacrylamide gel electrophoresis was performed for size analysis of protein samples, separating molecules in an electric field depending on relative molecular size (Laemmli, 1970). Proteins were boiled in buffer containing the relatively harsh detergent sodium dodecyl sulfate (SDS) which leads to unfolding of the protein by binding to the hydrophobic protein parts and introducing additional negative charge through its polar head groups. Therefore also uncharged or positively charged proteins become negatively charged, migrating to the anode of the electric field. A matrix of defined average pore size, polymerizing from a certain ratio of acrylamide and crosslinking bis-acrylamide, causes size separation of proteins during migration through the electric field. Besides SDS used in standard loading buffers, β -mercaptoethanol or other reducing reagents (e.g. dithiothreitol) would be useful for breaking of tertiary structure stabilizing disulfide bridges. The nature of heme group attachment in *c*-type cytochromes forbids use of reducing agents in this case in order to avoid loss of heme groups prior or during the SDS-PAGE experiment. Gels were directly casted in the Hoefer miniVE SDS-PAGE running chamber (GE Healthcare) using 12.5 % of acrylamide/bisacrylamide resolving gel and a 5 % stacking gel. Samples were mixed with SDS loading buffer and boiled at 95°C before being applied to the gel. Protein molecular weight marker was applied as a size standard. The running chamber was attached to a power supply unit at 25 mA and 300 V. Polyacrylamide gels were treated

after running procedure using a silver stain method for band detection (Heukešhoven & Dernick, 1988).

Resolving gel 1x [ml] :	12.50%	Stacking gel 1x [ml] :	5%
bisacrylamide 30%	3	Bis-acrylamid 30%	0.5
1.88 M Tris pH 8.8	1.8	0.625 M Tris pH 6.8	0.6
0.5 % SDS	1.8	0.5 % SDS	0.6
dH2O	2.4	dH2O	1.3
TEMED [μ l]	8	TEMED [μ l]	3
10% APS [μ l]	45	10% APS [μ l]	15

Table 6: SDS-PAGE gel composition.

3.3.3.6 Protein concentration determination

3.3.3.6.a Bicinchoninic acid method (BCA)

Concentration of purified DHC2 was determined in solution with a BCA protein concentration determination kit (Pierce Biotechnology; Smith *et al.*, 1985).

3.3.3.6.b Lyophilization

Because of inaccuracies of dye based protein determination methods for small proteins and *c*-type cytochromes, an eppendorf cup of defined weight was filled with 350 μ l of DHC2 protein and lyophilized over night at 30°C in a Concentrator *plus* (Eppendorf, Hamburg). Sample was weighted after lyophilization and concentration calculated using the molar mass of 11,310 g/mol.

3.3.4 Protein spectroscopy techniques

3.3.4.1 UV/Vis absorption spectroscopy

Ultraviolet-visible absorption spectroscopy is a method utilizing the absorption of photons ranging in energy from the near ultraviolet ($33,000\text{ cm}^{-1}$) over blue light ($21,000\text{ cm}^{-1}$) to near infrared ($14,000\text{ cm}^{-1}$). Radiation of this energy range is causing outer shell electron transitions in target molecules or atoms, for example in d-orbitals of transition metals like iron or in aromatic compounds like porphyrin. The absorption spectrum of heme involves peaks deriving from electron transitions within the porphyrin macrocycle (π to π^* -transitions), ligand-to-metal or metal-to-ligand charge transfer, as well as d-orbital to d-orbital transitions (Gouterman, 1978; Henderson, 1990). Beer-Lambert's law with

$$A = \log\left(\frac{I_0}{I}\right) = c \times d \times \varepsilon$$

relates the absorption of light to the concentration of dissolved molecules, where I_0 is the light intensity of reference sample, I is the intensity of the sample, c the concentration of the dissolved molecules, d the cuvette width and ε the experimental determined molar extinction coefficient, that has to be determined experimentally. This empirical law only holds for diluted solutions.

UV-Vis spectra of DHC2 were recorded with a concentration of $70\text{ }\mu\text{g/ml}$ protein, buffered with 100 mM HEPES-NaOH (pH 7.5) on an Ultrospec 2100 pro instrument (GE Healthcare). Reduction of DHC2 was achieved by stepwise addition of $10\text{ }\mu\text{l}$ Ti(III)-citrate with a concentration of approximately 83 mM , until reduction of DHC2 was complete. Oxidized and excess Ti(III)-citrate was removed by a Nap5 Sephadex G-25 column (GE Healthcare) equilibrated with anaerobic buffer (see above).

3.3.4.2 Electron paramagnetic resonance spectroscopy (EPR spectroscopy)

EPR spectroscopy is a method utilizing the absorption of electromagnetic radiation by unpaired (paramagnetic) electrons at wavelengths in the microwave range (1m-1mm). Every electron is defined by a set of four quantum numbers. The secondary spin quantum number m_s describes the angular moment of an electron with relation to an

external magnetic field. Electrons are fermions having a spin quantum number of $S=1/2$ and therefore two possible angular moments relative to a magnetic field, parallel $m_s=+\frac{1}{2}$ or antiparallel $m_s=-\frac{1}{2}$ orientations are observed. In the absence of a magnetic field both states are degenerated, meaning none is energetically preferred. Applying an external magnetic field abolishes degeneration of both states (Zeeman effect), as parallel alignment to the external magnetic field is lower in energy than the antiparallel state. Energy difference of both states is direct proportional to the magnetic field strength B_0 given by

$$\Delta E = g_e \mu_B B_0$$

with g_e representing the magnetic moment of a free electron, μ_B is the Bohr magneton and ΔE being the resonance energy, where electromagnetic radiation with a wavelength λ of

$$\lambda = \frac{hc}{\Delta E}$$

with h representing Planck's constant, c representing light speed, would excite an electron from the ground state to the excited state (Figure 12). Absorption spectra are recorded by applying a constant microwave wavelength and modulating the magnetic field strength. Limitations of instrumentation to create high magnetic field strength can lead to a decrease in signal to noise ratio, especially when considering signal saturation depending on temperature, as the probability for occupancy of the higher energy state increases with rising temperature. Relative numbers of population in excited $N_{exc.}$ and ground state N_{ground} are given by the Maxwell-Boltzmann distribution, where k_B is the Boltzmann constant and T the temperature:

$$\frac{N_{exc.}}{N_{ground}} = \exp\left(-\frac{g_e \mu_B B_0}{k_B T}\right)$$

The magnetic moment of a free electron g_e represents a reference for peaks in the observed spectra, as unpaired electrons in the sample condition are not only exposed to an external magnetic field, but also to local magnetic fields originating from the nature of the paramagnetic center surrounding. Derivatives of the absorption spectra can therefore give information about the organization of redox-active metal centers, as

unpaired electrons are normally not present for amino acids of proteins.

The EPR spectrum for isolated DHC2 was recorded in perpendicular X-band mode on a Bruker Eleksys 500 instrument with an ER 049 X microwave bridge (Bruker), equipped with an Oxford Instruments ESR 900 helium cryostat controlled by the ITC 503 temperature device. The modulation frequency was 100 kHz and the modulation amplitude was typically 0.5 mT. Measurements were performed with a Bruker 4122 SHQE cavity at ≈ 9.38 GHz. The sample tubes were Suprasil quartz tubes (705-PQ-9.50, Wilmad) with an \varnothing_{out} of 4 mm, and the sample volume was 250 μL . For spin quantitation, a copper standard was measured [anhydrous Cu_2SO_4 in 2 M $\text{NaClO}_4\text{-HCl}$ (pH 1.5)] containing 50 and 200 μM Cu^{2+} and the results were averaged. The concentration of EPR detectable copper was calculated according to the method of Aasa and Vännngård (Aasa & Vännngård, 1975) and for spin quantitation, double-integrated spectra of the reference and sample were compared. Spectra were recorded at different microwave powers to exclude the possibility of saturation. The EPR experiments were performed by Dr. Marc Rudolf and Prof. Peter Kroneck (Universität Konstanz).

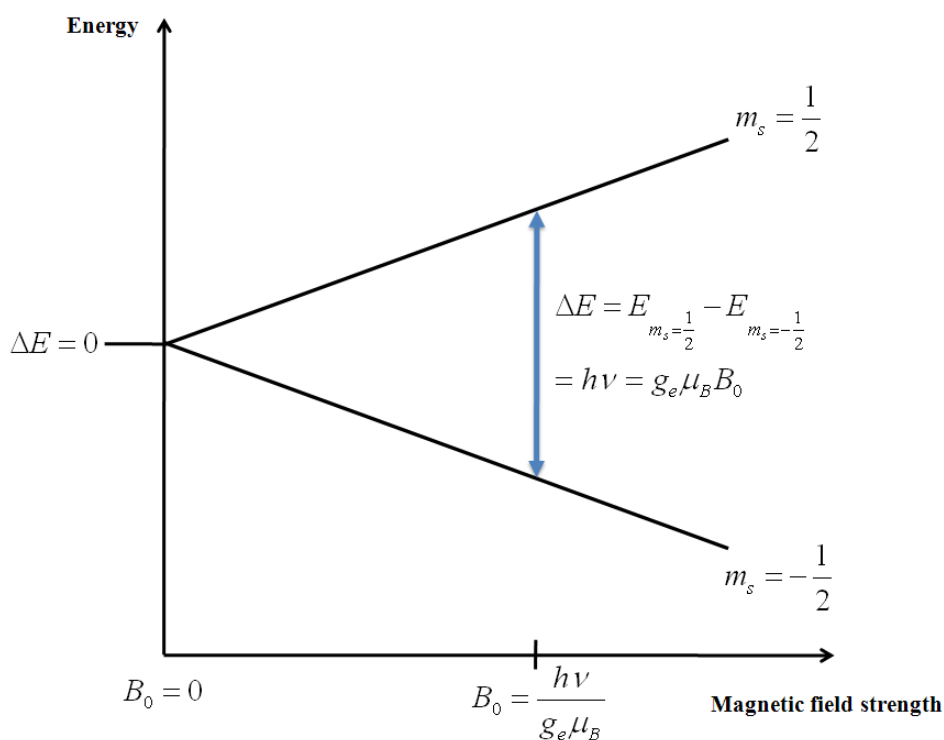


Figure 12: Energy diagram of electron spin states in EPR spectroscopy. Degeneration of spin angular momentum energy states is eliminated by an external magnetic field (Zeeman effect). Electrons of the ground state $m_s = -1/2$ absorb electromagnetic radiation of defined energy $h\nu$ when the resonance condition holds, at B_0 strength of the shifting magnetic field. Depending on the chemical surrounding of the paramagnetic electron, g is shifted with respect to theoretical free electron magnetic moment g_e .

3.3.5 Electrochemical methods

3.3.5.1 Redox titrations

Redox active proteins can facilitate the absorption and release of electrons depending on their redox potential. The redox potential is the relative tendency of a system to release (oxidation) or absorb (reduction) electrons in a chemical reaction. Free enthalpy of such a reaction at standard conditions can be directly related to an electrochemical potential:

$$\Delta E_0 = -\frac{\Delta G}{nF}$$

where ΔE_0 is the standard potential of the reaction, ΔG is the Gibbs energy, n the number of electrons and F the Faraday constant. The electrochemical potential of a redox reaction is given by the Nernst equation with respect to the activity a of the reductant and the oxidant:

$$\Delta E = \Delta E_0 + \frac{RT}{nF} \times \ln\left(\frac{a_{Ox}}{a_{Red}}\right)$$

where ΔE is the electrochemical potential of the reaction, R is the ideal gas constant, n the number of involved electrons, T the absolute temperature and F the Faraday constant. The reaction potential is equal to the standard reaction potential in an equilibrium condition and is called mid-point potential at pH 7. Measurements of mid-point potentials gives insight to the functionality of biologically relevant redox systems, as they are involved in the conduction of electrons, thus giving a direction and potential difference, or can give insight to redox enzyme mechanisms facilitating the oxidation or reduction of a specific substrate. Identification of mid-point potentials is therefore crucial for the understanding of electron conduction chains, energy differences in redox reactions and mechanisms of redox enzymes.

Redox titrations were performed according to the protocol of Dutton (Dutton, 1978), where 2 ml of 12 μ M DHC2 in 100 mM HEPES-NaOH pH 7.5 was placed into a modified QS 109.904 cuvette (Hellma), which was attached to a branching glass component (Ochs, Bovenden) (Figure 13). To ensure efficient electron conduction from/to reductant/oxidant to the protein/ electrode, the following mediators were used to a total concentration of 1 μ M: methyl viologen (-436 mV), neutral red (-325 mV), safranin (-280 mV), anthraquinone-2-sulfonic acid (-225mV), 2-hydroxy-1,4-

naphthoquinone (-145 mV), phenazine(-125 mV), indigo carmine (-110 mV), resorufin (-51mV), duroquinone (5 mV), methylene blue (11 mV), juglone (30 mV), 1,4-naphthoquinone (60 mV), N-methylphenazoniummethyl sulfate (80 mV), and 1,2-naphthoquinone (118mV). Values in parentheses indicate the respective midpoint potentials. For recording of potentials an EMC 50 Ag/AgCl electrode (Sensortechnik, Meinsberg, Germany) was dipped into protein solution, after equilibration with quinone/hydroquinone and titration of $K_3[Fe(CN)_6]$ against $K_2[Fe(CN)_6]$ and vice versa. The redox cuvette was sealed with gas tight rubber septums and anaerobized, using an anaerobic line with 8 sequential steps of vacuum and dinitrogen gassing. UV/Vis absorption was monitored with a S2000 fiber optic spectrometer (Ocean Optics) containing a deuterium-tungsten halogen light source (Micropack). The amount of reduced DHC2 was followed by recording of absorption at a wavelength of 552 nm, setting absorption intensity of fully reduced DHC2 as 100% reduction and absorption intensity of fully oxidized DHC2 as 0% reduction. Decrease of total absorption intensity by precipitation during the experiment was corrected for by linear scaling of intensities using isosbestic points at 542 and 561 nm. Reduction of the protein sample was achieved by stepwise addition of 20 mM dithionite until the spectrum indicated full reduction. Oxidation was achieved by sequential addition of 20 mM $K_3[Fe(CN)_6]$ until the spectrum indicated full oxidation. Addition of reductant/oxidant and mixing of the sample solution was performed in a mixing chamber connected to the redox cuvette by a cannula. Between each reduction/oxidation step, the potential of the protein solution was recorded together with absorption peak height at 552 nm, after letting the system equilibrate for 10 minutes. The measured potentials were corrected with respect to the normal hydrogen electrode (NHE). All steps in the redox titration experiment were performed at a temperature of 298 K.

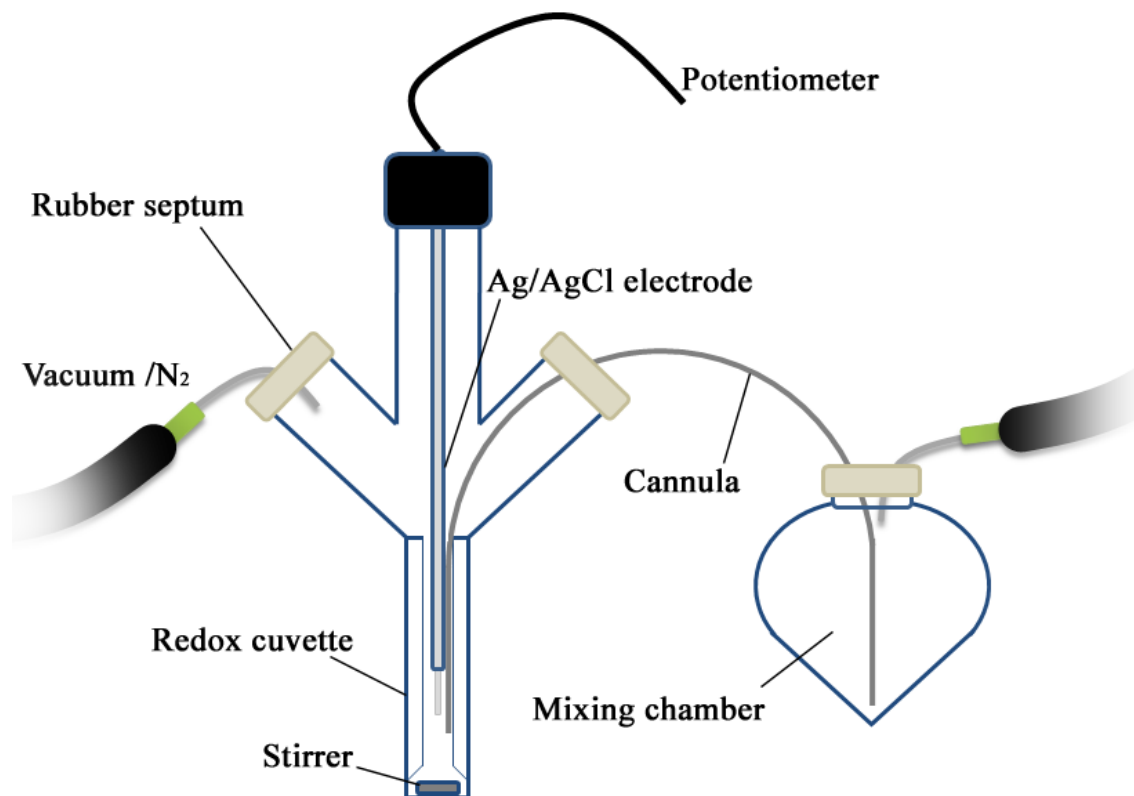


Figure 13: Schematic representation of the experimental setup for redox titration.

3.3.6 Crystallographic methods

3.3.6.1 Protein crystallization

Crystallization is a process in which soluble molecules form a three dimensional macromolecular solid of repetitive order. The process of crystallization is thermodynamically allowed when a gain of entropy, by releasing of a hydration shell, is bigger than the loss of entropy through crystal formation, which means a lesser degree of freedom for these molecules. Therefore undersaturated solutions can not promote crystal growth, as the ratio of solvent molecules to solvated molecules is too big. Supersaturated solutions are thermodynamically instable, as the ratio of solvent molecules to solvated molecules is too small. Undersaturation of a solution can be transformed to saturation, by the addition of molecules that decrease the solubility of the protein or increase the protein concentration itself. Both approaches are used for protein crystallization experiments, as supersaturation of protein solutions would allow for the formation of crystals (Figure 14). Unfortunately kinetics of protein to protein interactions are crucial for the appearance of ordered aggregates, able to grow into

crystals. Nature and concentration of the added molecules and all kind of physical parameters can therefore influence parameters of crystal nucleation and growth, like temperature, sample homogeneity, air pressure, solution viscosity and so forth. Empirical identification of conditions, where the protein is in a state of supersaturation, not precipitating and being able to perform the committing step from crystal nucleation to growth, is a key element in the determination of the three dimensional structure of a protein by X-ray crystallography.

Initial crystallization experiments were performed with DHC2 protein concentrated to 10 mg/ml. In the vapor diffusion method, using sitting drop Cryschem crystallization plates (Hampton Research, USA), 1 μ l of protein solution was mixed with 1 μ l of precipitant solution equilibrated against 300 μ l of reservoir solution, sealed with crystal clear tape (Hampton Research, USA) and stored at 20°C. Testing of a broad spectrum of precipitation conditions was performed by use of premade standardized crystallization screens: CrystalScreen I, CrystalScreen II, Natrix Screen, PEG/Ion Screen, Crystal Screen Lite (all Hampton Research), Footprint I, II, III (Stura *et al.*, 1994), JB Screen 2,3,6 (Jena Bioscience), Magic Screen I, II, III, IV. (BioGenova).

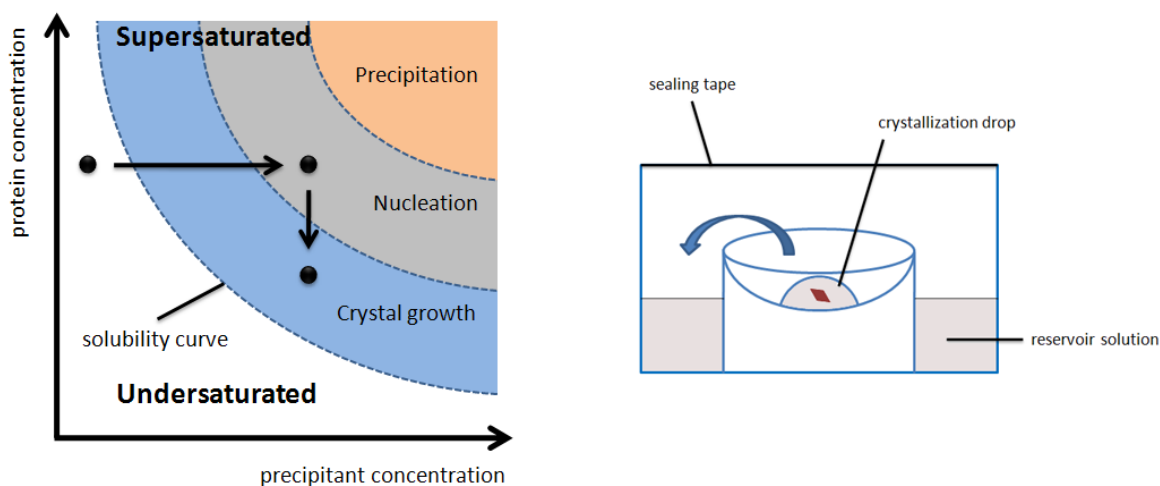


Figure 14: (Left) Schematic representation of a two dimensional solubility diagram, illustrating processes in crystal nucleation and growth. (Right) Schematic representation of the vapor diffusion method in a sitting drop crystallization experiment. The blue arrow indicates diffusion of water from the crystallization drop, which has a higher vapor pressure than the reservoir solution.

3.3.6.2 X-ray diffraction experiments

X-rays are electromagnetic radiations with wavelengths ranging from 10^{-8} to 10^{-11} m. Diffraction behavior is most pronounced at objects of a size equaling the wavelength of the propagating electromagnetic radiation. Electrons therefore represent an obstacle for the propagating X-ray, which is resulting in strong diffraction of the incident wave. In theory crystals are not necessary for X-ray diffraction experiments, but without signal amplification through constructive interference of diffracted X-rays at ordered lattice planes, no X-ray source could generate X-ray beams of the desired power, that would unfortunately destroy the sample. X-rays diffracted by electrons of a single crystal give rise to a diffraction pattern reflecting structural properties of the smallest individual unit forming the crystal lattice. Diffraction patterns of crystals are following Bragg's law for constructive interference:

$$n\lambda = 2d\sin\theta$$

where n is an integer, λ is the wavelength of the x-ray beam, d is the distance between the scattering lattice planes h , k , l and θ is the angle of the incident wave with respect to the lattice plane. Measuring of diffraction intensities in an X-ray crystallography experiment gives the structure factor amplitude of a point h , k , l in the reciprocal lattice, which is a fourier transform of atom scattering factors in the real crystal lattice:

$$F(h, k, l) = \sum_j f_j e^{2\pi i(hx+ky+lz)}$$

where F is the structur factor, f is the scattering contribution of a single atom j , h , k , l are the Miller indices, and x , y , z are real space coordinates. The Fourier transform of atomic scattering factors is a wave function or complex number with a magnitude $|F(h,k,l)|$ and a phase angle α denoted as the structure factor F :

$$F(h, k, l) = |F(h, k, l)|e^{-i\alpha(h,k,l)}$$

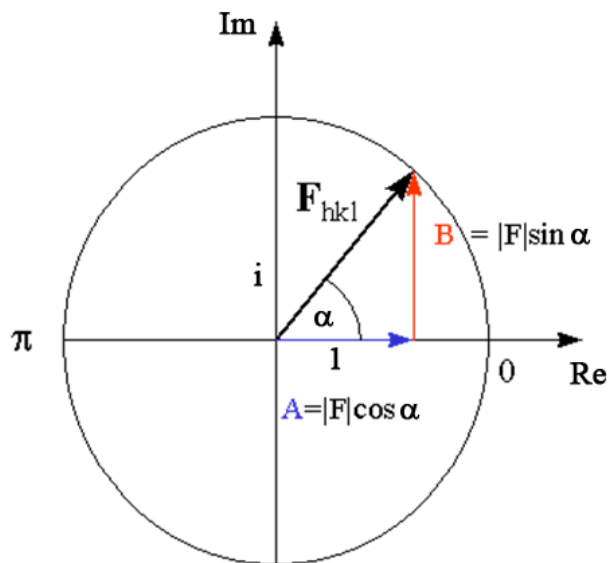


Figure 15: Argand diagram of the Gaussian plane showing the wave function of the structure factor.

With the structure factor amplitude proportional to the measured diffraction spot intensity I :

$$I(h, k, l) \propto |F(h, k, l)|^2$$

only the phase is missing for a back fourier transform into real space, calculating the electron density $\rho(x, y, z)$ given by the electron density function:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h, k, l} F(h, k, l) e^{-2\pi i(hx + ky + lz)}$$

The phase angle α is the only variable that is not directly measurable by experimental means. Missing information about the phase angle is a crucial obstacle in the process of structure determination by X-ray crystallography and is therefore denoted as the 'phase-problem' in X-ray crystallography.

3.3.6.3 Cryocrystallography

X-ray radiation acts ionising when interacting with matter, therefore protein crystals measured by this technique are damaged by the emergence of free radicals (Garman & Owen, 2006). Diffusion of these compounds damages the crystals depending on the

mobility of solvent molecules and is greatly reduced at lower temperatures. Cryo-cooling of protein crystals in a liquid nitrogen stream at 100 K is therefore indispensable when exposing them to X-ray radiation generated by modern synchrotrons. A disadvantage of this method is the need for a proper cryo-buffer for the crystal prior to freezing it in liquid nitrogen. Water molecules in the drop can form ordered structures during cryo-cooling, giving rise to diffuse diffraction patterns interfering with experimental data, or in the worst case destroying the protein crystal. The addition of cryo-protectants to the drop solution leads to vitrification of the crystal drop, being free from any crystalline structure beside the protein crystals. Transfer of crystals from the crystallization drop to the cryo-condition can often lead to damaging of the crystal lattice due to rapid hydration or dehydration events. Empirical determination of minimal cryo-protectant concentration is therefore inevitable, as vitrification of the drop solution can be dependent on loop size, crystal shape and viscosity of the drop solution.

3.3.6.4 Data collection

3.3.6.4.a General considerations on data collection

Exposing the protein crystal to an X-ray beam and collecting a set of emerging diffraction patterns with an attached X-ray detector is the last step in experimental procedures determining the three-dimensional structure of a protein. A setup where the crystal is rotating around a spindle axis allows for bringing the crystal into different diffraction conditions with respect to the crystal lattice and the inflicting X-ray beam. A data collection experiment is therefore separated into several steps, consisting of crystal rotation and irradiation, detector-readout and -erasure. Increments in rotation angle are chosen to maximize the number of individual detectable reflections (completeness of the experiment) and to minimize exposure time of the crystal to X-radiation. The choice of suitable sets of experimental parameters is often complicated due to imperfections of protein crystals, X-ray equipment and crystal handling. After inspection of diffraction images at angles differing in 90°, a strategy for collecting a set of reflection images, including parameters such as the rotation angle, exposure time, X-ray wavelength (modulation only possible at Synchrotron sources), detector distance and total rotation degree, represents a compromise in maximizing parameters of overall data quality with

maximal resolution, completeness, redundancy, signal to noise ratio and data comparability within the whole X-ray diffraction experiment.

3.3.6.4.b Multiple-wavelength anomalous dispersion experiment

Data collection strategies maximizing the anomalous diffraction behavior of an incident X-ray beam can be used to overcome the phase-problem in the determination of an overall protein structure. When the energy of an X-ray photon matches or is higher than the energy of exciting an inner shell electron to a higher energy state (see energy range of x-ray-photons), the photon is either absorbed and excitation energy is released at a longer wavelength, which is termed X-ray fluorescence, or the photon is scattered at the electron, contributing to the overall diffraction behavior of the crystal. Atoms with inner shell electrons corresponding to X-ray photon energy are normally heavy metal atoms like Fe, Cu, Au, Pt, Hg and so forth. Scattering contribution of a single atom in the unit cell is given by the atomic scattering factor f . Scattering of X-ray photons by electrons of the same energy or oscillation wavelength alters the scattering behavior of the atom introducing the real atomic scattering correction term f' and the imaginary correction term f'' :

$$f(\lambda) = f_0 + f'(\lambda) + if''(\lambda)$$

where f_0 is the Thomson scattering factor and correction terms are dependent on the wavelength λ . Depending on the wavelength the correlation of atomic scattering factor and structure factor changes to:

$$\begin{aligned} F_{PH}(h) &= \sum_j^P f_j e^{2\pi i(h \cdot r_P)} + \sum_k^H (f_{0H} + f'_H + if''_H) e^{2\pi i(h \cdot r_H)} \\ &= F_P(h) + F_{0H}(h) + F'_H(h) + iF''_H(h) = F_P(h) + F_H(h) \end{aligned}$$

where h is the reciprocal space vector h , k , l , P indicates atoms not contributing to anomalous scattering or structure factors of the remote data set, r_P and r_H are real space vectors of non anomalous scattering and anomalous scattering atoms respectively, H indicates atoms contributing to anomalous scattering. With $iF''(h)$ being always positive and perpendicular to F_{0H} and F'_H Friedels law with:

$$|F_{hkl}| = |F_{\bar{h}\bar{k}\bar{l}}| \text{ and } \alpha_{hkl} = -\alpha_{\bar{h}\bar{k}\bar{l}}$$

is broken as amplitudes of inverse symmetric related reflexes are not equal and not inverse in phase angle. The problem of phase determination can therefore be divided in substructures of atoms not contributing to anomalous scattering and anomalous scattering atoms, which are normally few in number.

The aim of a MAD data collection experiment (Hendrickson *et al.*, 1985) is maximization of the anomalous signal from a certain fraction of atoms in the protein crystal, which in the case of DHC2 is the central heme iron atom. When f'' is at a maximum the difference of amplitudes of Friedel pairs is also at a maximum. So modulation of X-ray beam wavelengths is inevitable to measure data at elemental absorption edges. For these experiments a synchrotron radiation facility is needed. X-ray fluorescence scan of protein crystal samples reveals the exact wavelength of the absorption edge and therefore the maximum in f'' , as local chemical surroundings can shift the theoretical absorption edges (Cromer & Liberman, 1970). The imaginary part of the scattering correction factor, f'' , is related to the real part, f' , via the Kramers-Kronig relationship:

$$f'(\omega) = \frac{\omega' f''(\omega') \partial \omega'}{\omega^2 - \omega'^2}$$

For obtaining the substructure of anomalous scattering atoms the Patterson function is applied (Patterson, 1934). The Patterson function P is a fourier transform of structure factor amplitudes with u, v, w representing interatomic vectors in the patterson space:

$$P(u, v, w) = \sum_{h,k,l} |F(h, k, l)|^2 e^{-2\pi i(hu+kv+lw)}$$

Calculation of the Patterson function with an amplitude of $|F_P(h,k,l) - F_{PH}(h,k,l)|$ gives peaks of interatomic distance vectors, only contributed by anomalous scattering atoms:

$$P(u, v, w) = \sum_{h,k,l} |F_P(h, k, l) - F_{PH}(h, k, l)|^2 e^{-2\pi i(hu+kv+lw)}$$

From these interatomic distances real space vectors of the heavy atoms can be derived and structure factors F_H for the heavy atoms can be calculated. Using the Harker

construction phase angles for remote data set structure factors can be estimated (Harker, 1956).

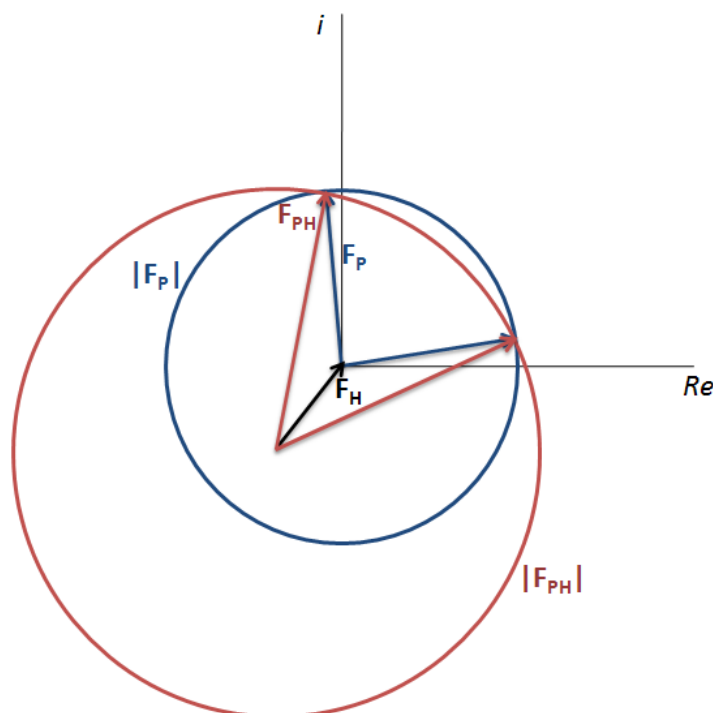


Figure 16: Harker construction for structure factor estimation of remote data set phase angles.

3.3.6.5 Data processing

3.3.6.5.a Indexing

Assigning Miller indices h , k , l , representing individual lattice planes of the crystal corresponding to Bragg's law, to each individual reflection spot is denoted as the process of indexing. As the crystal is not perfectly aligned to the spindle axis an algorithm has to predict the relative orientation of the crystal inside of the loop. A unit cell described by the length of its basic vectors \mathbf{a} , \mathbf{b} , \mathbf{c} and the angles α , β , γ between them is fitted in the process of indexing, representing the smallest repetitive unit building the crystal by three dimensional translation operations. Quality of the fitting procedure compared to conditions of unit cell parameters to all 14 Bravais lattices gives an indication for crystal symmetry, meaning molecules residing in a unit cell could be related by distinct crystallographic symmetry operations.

3.3.6.5.b Integration

Reflections picked from a raw-crystal image provide intensities and their mean deviation as they are estimated in the process of integration. Profile fitting is used as a method where the background profile of the detector is subtracted from the profile of reflections to obtain the corresponding intensity. Pattern of peaks distributing on a reflection image is predicted by parameters derived from the indexing process.

3.3.6.5.c Scaling and Merging

Reflection intensities collected during an X-ray crystallography experiment depend strongly on characteristics of experimental setup and crystal properties. Intensities of the same Miller indices h , k , l can therefore deviate depending on fluctuations in the beam intensity (especially at a synchrotron sources), imperfect spindle axis, crystal shape, crystal decay and so forth. In the procedure of scaling redundant reflections are used to bring reflection intensities to a reasonable level. This method is also used when combining reflection intensities of diffraction experiments obtained under different data collection conditions. Isomorphism is a key prerequisite for applying this method, as scaling of data from anisomorphic crystals leads to data with no reasonable structural reference. Reflection intensities of same Miller indices h , k , l would be merged to a single intensity with resulting in a mean intensity that is actually not representing any of the two used crystals. Merging is the process where all multiply measured reflection intensities of same h , k , l are merged into a single reflection with averaged intensity, whereby crystal symmetry can greatly decrease the total number of unique reflections.

3.3.6.5.d Data quality indicators

High quality structural model data is dependent on good quality of diffraction data. Several statistical indicators therefore provide information about the goodness of intensities collected during an x-ray crystallographic experiment, like resolution, completeness of the set of observed reflections and I/σ , which is the effective signal to noise ratio. Statistics on merging equivalent reflections is also providing information on data quality, especially for the whole range of the experiment. Commonly used in this

case is the conventional symmetry R-factor R_{sym} or merging R-factor R_{merge} , summing differences of individual measurements from the mean of the measurement divided by the total sum of measurements:

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)}$$

where i is the number of observation for reflection h, k, l , $\overline{I(hkl)}$ is the mean intensity of reflection h, k, l . When the number of total observations in a data set increases the R_{merge} will also increase pretending a decrease in data quality. But in fact the mean of observation gets sharpened by an increase of observations, reflecting better data quality than is indicated by the merging R factor R_{merge} . Therefore R_{merge} is a relatively weak indicator for data of high redundancy or in comparing data quality of different redundancies, for example different crystal systems. Therefore the redundancy independent R factor $R_{\text{r.i.m.}}$ can be applied (Weiss & Hilgenfeld, 1997), which is a data quality indicator independent of the number of observations:

$$R_{\text{r.i.m.}} = \frac{\sum_{hkl} \sqrt{\frac{N}{N-1}} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)}$$

where N is the redundancy of observations, which gives now higher R factors for data of lower redundancy than R_{merge} , but is better comparable to data of higher redundancy. Precision of the averaged measurement used in later processes of structure determination and refinement is indicated by the precision indicating merging R factor $R_{\text{p.i.m.}}$ (Weiss & Hilgenfeld, 1997) and gives the best estimate for the reasonable use of reflection data:

$$R_{\text{p.i.m.}} = \frac{\sum_{hkl} \sqrt{\frac{1}{N-1}} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)}$$

3.3.6.7 Model building and structure refinement

The result of an x-ray crystallography experiment is not the electron density, but a structural model of the protein in real space, building up the macromolecular crystal.

Interpretation of the experimental electron density distribution is an iterative process of model building and subsequent model refinement. A protein model consists of atoms with distinct position in real space given by the basis vector $\mathbf{r} = (x, y, z)$ and normally one isotropic temperature factor B indicating the spatial displacement of an atom around the position given by its coordinates and thus describing imperfections in crystal packing and/or flexibility of protein residues directly affecting the scattering contribution of the individual atom by:

$$f = f_0 e^{-B \left(\frac{\sin \theta}{\lambda} \right)^2}$$

with B being the temperature-, B - or Debye-Waller factor with:

$$B = 8\pi^2 \langle u \rangle^2$$

where $\langle u \rangle^2$ is the mean square positional displacement given in \AA^2 , which is sufficient for structural models to a certain resolution and has to be replaced for high resolution structures by an anisotropic 3x3 tensor.

When a structural model is built, resembling the visible features of observed density, to a certain degree, structure factors and therefore structure factor amplitudes $|F_{\text{calc}}|$ can be calculated from the atomic model as indicated before and compared to observed structure factor amplitudes $|F_{\text{obs}}|$. In the process of refinement, optimization of the model is carried out in an iterative way, bringing $|F_{\text{obs}}|$ and $|F_{\text{calc}}|$ to an absolute minimum of divergence. Model parameters refined are the coordinates and the isotropic temperature factors B . The refinement process is not free in terms of maximizing the fit between $|F_{\text{obs}}|$ and $|F_{\text{calc}}|$ with respect to chemical reasonability of a structure. Therefore constraints and restraints are used that limit the independent optimization of parameters to an absolute or approximate value respectively. A problem in refinement processes are local minima in the parameter optimization process that have to be overcome or else an absolute minimum for the structural model will not be achieved. Three different methods for structure refinement are utilized in commonly employed refinement programs, which are all using different approaches and algorithms for parameter optimization, being maximum likelihood (REFMAC 5; Murshudov *et al.*, 1997), least-squares fit (SHELXL; Sheldrick, 2008) and simulated annealing (CNS; Brünger *et al.*, 1998). After each cycle of refinement, phase angles of the models are combined with experimental structure factor amplitudes, where the resulting model can be further used

for model building and subsequent refinement. A quality indicator for a structural model whether it matches the real three dimensional arrangement of atoms in the crystal is the refinement R-factor R with:

$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

In contrast to other methods in structure determination, like for example in cryo electron microscopy, where overfitting of parameters during a refinement process is not perceivable from the refinement process, a cross-validation statistical technique is implemented into the refinement process in X-ray crystallography. Therefore reflexes h , k , l are divided into different subsets used for refining or validation, a working set and a validation set. Only the working set is used for refinement of structural model parameters minimizing the deviation of observed and calculated data. For both sets an R factor can be calculated, yielding R_{crist} and R_{free} (Brünger, 1993). During the refinement process both values should decrease in a similar manner. If R_{free} should be significantly higher than the R_{crist} , then overfitting of data in the refinement process is indicated, by trying to fit a poor structural model, resulting in a good correlation for the working set reflections, but not for the test set reflections.

3.4 Results

3.4.1 Recombinant production of DHC2 in *Escherichia coli*

The *dhc2* gene encodes for a protein of 94 amino acids, including a periplasmic export signal for the secretory pathway, consists of residues 1-27 according to SignalP (Bendtsen *et al.*, 2004). Mature DHC2 therefore consists of 67 amino acids with a calculated pI of 9.8 (Bjellqvist *et al.*, 1993) and two cytochrome *c* binding motifs CXXCH (Figure 17). Exchange of the *Geobacter sulfurreducens* export signal for the *pelB* leader sequence from *Escherichia coli* before residue Ser25 and fusion to a C-terminal hexa-histidine affinity tag resulted in a size of 78 amino acids for the recombinant protein with Met23 and Val24 being cloning artifacts of the *NcoI* restriction site.

```

1  MKLAKKDWFF  IVLIVVVVGV  FWAISGEVRT  KKVPLDTNHK  RFYDAFAQGA  50

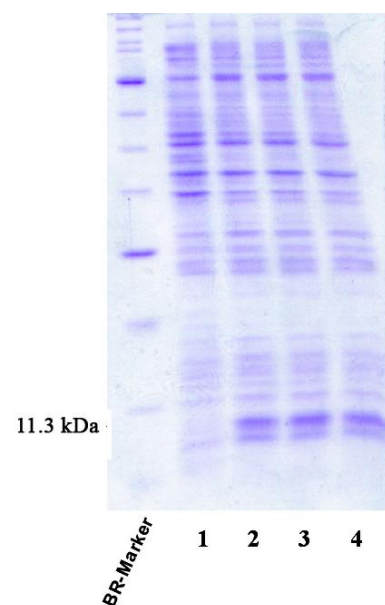
51  GKLDLDRQCV  ECHHEKPGGI  PFPKNHPVKP  ADGPMRCLFC  HKFK  94

```

Figure 17: Protein sequence of DHC2 from *Geobacter sulfurreducens*, as annotated in the genome sequence (GSU2927). The N-terminal leader sequence is colored in gray and heme binding motifs are colored in cyan.

The expression of *c*-type cytochrome DHC2 together with the accessory pEC86 plasmid (Thöny-Meyer, 1995) in *Escherichia coli* could be followed by the red color of cell suspension or pellet. The protein is visible as a red band in the SDS-PAGE. Induction tests showed an increased expression of DHC2 at concentrations of 0.1, 0.5 and 1 mM of IPTG (Figure 18), but cell yields were decreased dramatically, as cytochrome *c* expression seems to heavily stress the cells. Therefore expression without induction was preferred over induction by IPTG.

Figure 18: Polyacrylamide gel of DHC2 induction test with (1) 0 mM (2) 0.1 mM (3) 0.5 mM and (4) 1 mM of IPTG. Size of DHC2 is indicated by 11.3 kDa size.



3.4.2 Purification of DHC2

Soluble, reddish cell fraction retained after cell breakage and centrifugation indicated the soluble expression of DHC2 and was subjected to HiTrap affinity chromatography. Red coloured protein bound to the column and could be eluted with a concentration of 100 mM imidazole. A silver-stained SDS-gel indicated no impurities after the HiTrap column (Figure 19). However, to remove unspecifically aggregated DHC2 and to obtain information about the physiological oligomerization state of DHC2, the sample was applied to size exclusion chromatography. DHC2 eluted as a single peak with a retention time indicating a monomeric state, while a small peak around the void volume likely represented oligomerized DHC2. Only the monomeric form was used for further experiments, SDS-PAGE looked as after the HiTrap purification, implying that the oligomeric state can be dissolved by boiling with SDS. The purification process yielded 12 mg of protein from 6 L of cell culture.

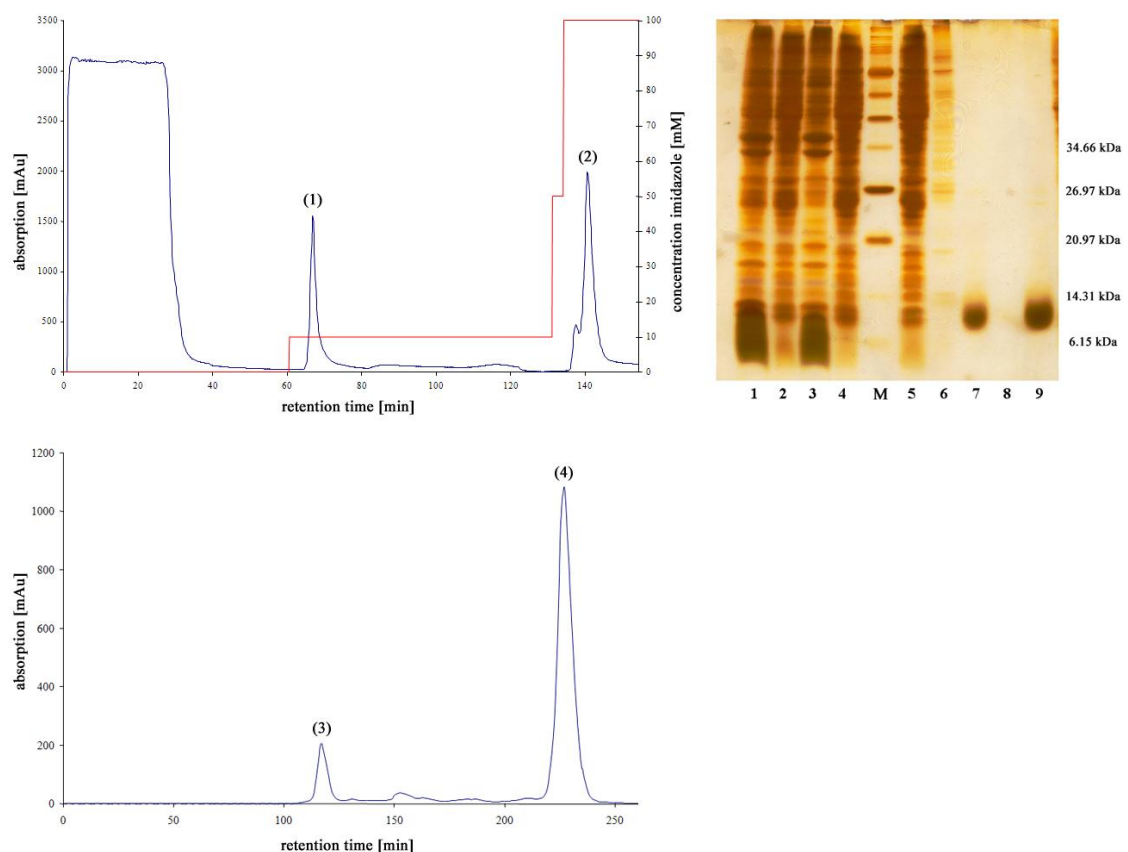


Figure 19: Purification of DHC2. (Upper-left) Purification chromatogram of HiTrap affinity column run with (1) the wash peak and (2) the elution peak at a concentration of 100 mM imidazole are shown. (Lower-left) Purification chromatogram of size exclusion chromatography on a Superdex S75 HiLoad 26/60 with (3) oligomerized DHC2 and (4) DHC2 monomer are shown. (Upper-right) Silver stained 12.5 % polyacrylamide gel showing purification progress of DHC2: 1 - pellet 20.000xg, 2 - supernatant 20.000xg, 3 - pellet 100.000xg, 4 - supernatant 100.000xg, M - NEB BR protein weight marker, 5 - flow-

through HiTrap, 6 - 10 mM imidazole wash step HiTrap, 8 - 1 M imidazole cleaning step HiTrap, 9 - monomeric peak fraction S75.

3.4.3 UV/Vis absorption spectrum

The extinction coefficient of DHC2 was calculated using the peak absorption height at 408 nm for an oxidized DHC2 solution with concentration derived from weighing lyophilized DHC2. Maxima of the UV/Vis absorption spectrum for oxidized DHC2 were at 408 nm ($214\,000\text{ M}^{-1}\text{ cm}^{-1}$) and 530 nm ($18\,800\text{ M}^{-1}\text{ cm}^{-1}$) (Figure 20). The absence of a peak or shoulder at around 610 nm indicated a low-spin heme Fe (III) state, commonly observed for bis-histidinyl coordinated *c*-type cytochromes (Gouterman, 1978; Pettigrew & Moore, 1987). Reduction of DHC2 solution with a 10-fold molar excess of sodium dithionite led to absorption maxima at 419 nm ($353\,600\text{ M}^{-1}\text{ cm}^{-1}$, γ -band), 523 nm ($30\,200\text{ M}^{-1}\text{ cm}^{-1}$, β -band), and 552 nm ($53\,400\text{ M}^{-1}\text{ cm}^{-1}$, α -band). An overlay of the oxidized and reduced spectra indicated isosbestic points at 412 nm, 432 nm, 508 nm, 531 nm, 542 nm and 561 nm. The observed features are typical for *c*-type cytochromes and no independent absorption features of the individual heme groups could be observed during the process of reduction, using reduction agents such as dithionite or Ti(III)citrate and oxidation by ferricyanide or oxygen. Responsible for the absence of absorption maxima at 280 nm is the low abundance of aromatic amino acids in the protein sequence of DHC2.

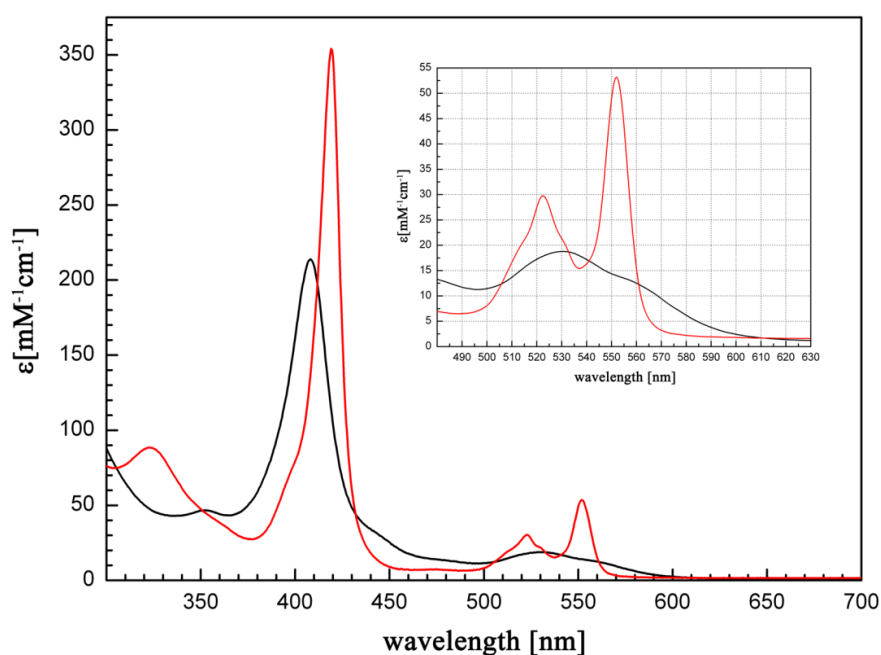


Figure 20: UV/Vis absorption spectra of oxidized (black) and dithionite-reduced DHC2 (red). The oxidized form shows absorption maxima at 408 nm and 530 nm, the reduced form at 419 nm, (γ -band), 523 nm (β -band) and 552 nm (α -band).

Reoxidation of dithionite-reduced DHC2 with air showed no individual peak separations during the whole procedure (Figure 21), being indicative for a shared potentiometric state of both heme groups.

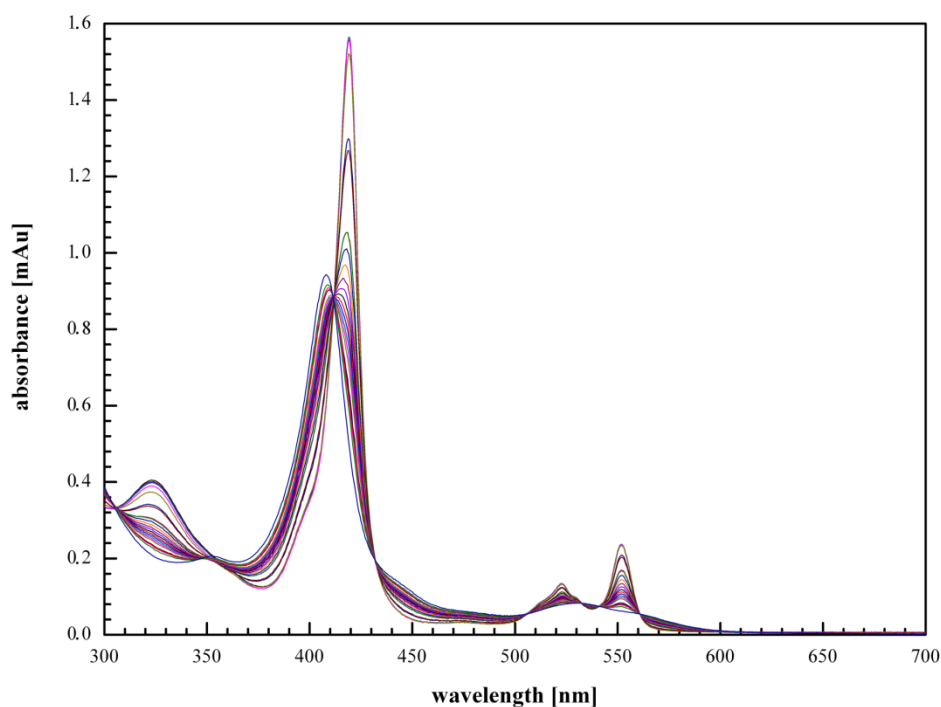


Figure 21: Reoxidation spectra of diheme cytochrome DHC2 from *Geobacter sulfurreducens* exhibit no visible separation of individual heme peaks during an oxidation process with air.

3.4.4 Electron paramagnetic resonance (EPR) spectrum

Samples of DHC2 exhibited g -values of 2.96, 2.26 and 1.50 in the X-band EPR spectrum, which is indicative for low spin-spin Fe (III) and similar to values observed for other multiheme *c*-type cytochromes (Einsle *et al.*, 2001) (Figure 22). Relatively weak signals at 4.3 and 2.06 point to minor impurities from Fe(III) and Cu(II) in the sample. In contrast to single heme spectra, the peaks seem to be relatively broadened, which could be related to magnetic coupling of both heme groups. A low value for g_{\max} of 2.96 indicates the presence of bis-histidinyl-ligated heme groups as axial ligands (Walker, 2004), both comprising a small dihedral angle of the histidine imidazole planes. Spin quantitation of the EPR spectrum compared to the quantitation standard

yielded 2.0 ± 0.2 spins pro monomer, indicating that both heme groups contribute fully to the EPR spectrum, as Fe(III) in low-spin configuration has one unpaired electron.

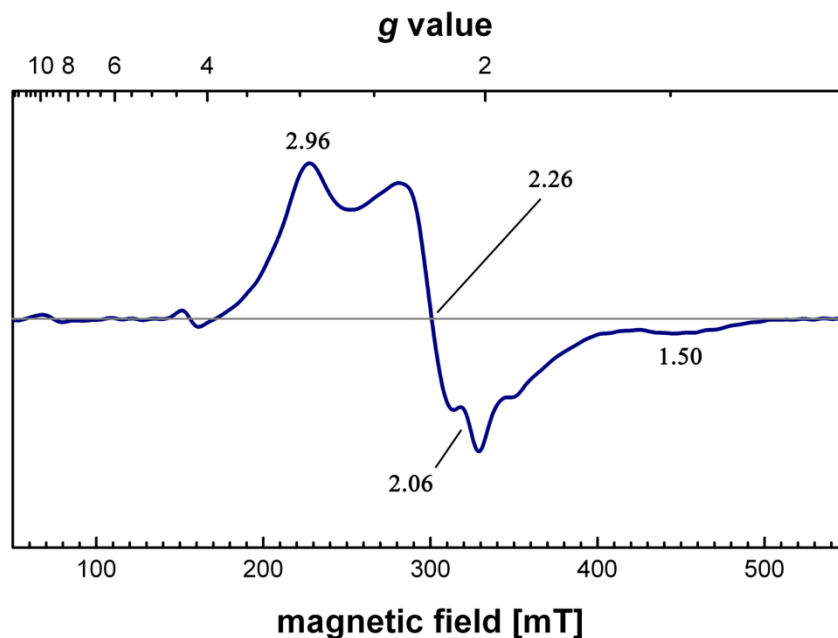


Figure 22: X-band EPR spectrum of DHC2 at 10 K and a microwave power of 20 mW recorded by Dr. Marc Rudolf and Prof. Peter Kroneck (Universität Konstanz). The apparent g values at 2.96, 2.26 and 1.50 are typical for bishistidinyl-liganded *c*-type cytochromes with Fe(III) in a low-spin configuration. Relatively small g_{\max} values point to small dihedral angles for both heme groups with respect to the axial ligand imidazole planes. Slightly broadened peaks could be an indication for magnetic coupling of the individual heme groups (Heitmann & Einsle, 2005).

3.4.5 Redox titrations

Potentiometric redox titrations of DHC2 examined by UV/Vis absorption spectroscopy showed reversible redox behaviour. Identical curves were obtained for reduction by dithionite and oxidation with $K_3[Fe(III)(CN)_6]$ indicating a reliable behaviour of the investigated redox system (Figure 23). Absorption of redox mediators did not intervene with absorption of the cytochrome spectra, as was checked by recording individual spectra of each redox mediator in a reduced and oxidized form. Mid-point potentials of the DHC2 redox system were obtained by fitting two Nernst equations with separately variable mid-point potentials and a factor representing the relative contribution of each heme group to the total potential, to the observed potential curve:

$$\%red = \frac{b}{1 + 10^{\frac{E-E_1}{59\text{ mV}}} + 10^{\frac{E_2-E}{59\text{ mV}}}}$$

where b is the individual heme contribution factor, E is the measured potential, E_1 and E_2 are the potentials of sequential reduction steps. The contribution factor was fitted to values of 0.71/0.29 for the ratio of DHC2 heme groups and mid-point potentials for the two heme groups to -135 ± 1 and -289 ± 4 mV with respect to the normal hydrogen electrode (NHE) using Microcal Origin.

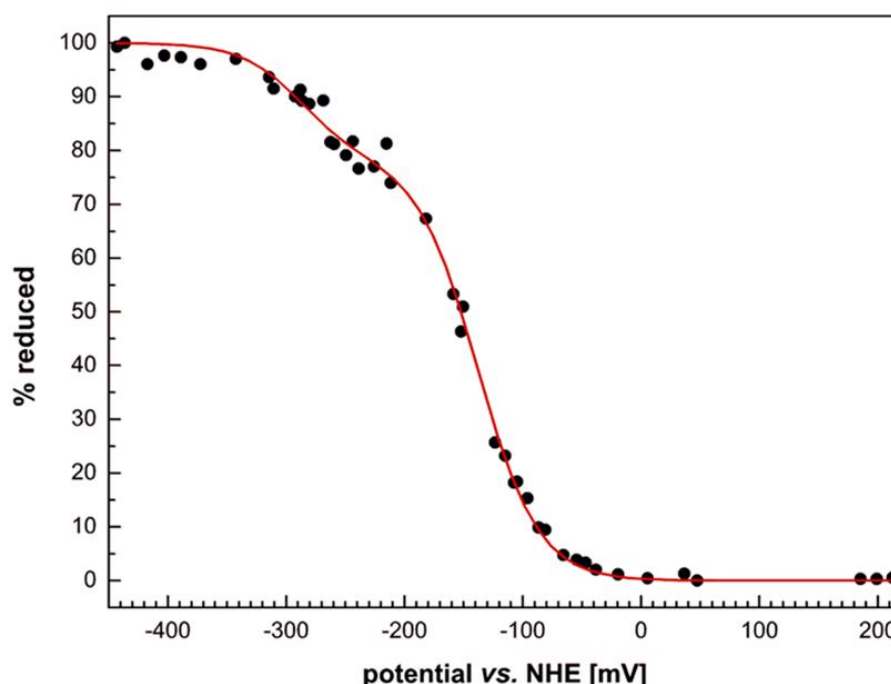


Figure 23: Redox titration of DHC2 at pH 7.5 followed by UV/Vis absorption spectroscopy at 552 nm. Black points indicate single measurements, where the red line is an approximation of data points to fit two Nernst equations with a relative heme group contribution of 0.71/0.29 and mid-point potentials of -135 ± 1 and -289 ± 4 mV.

3.4.6 Crystallization and data collection

Small, dagger shaped crystals were observed after two days in crystal screen I, condition 30 and 31, containing 0.2 M ammonium-sulfate and 30% PEG 4000 or PEG 8000 respectively (Figure 24), but were irreproducible by screening around the condition with freshly prepared PEG, until checking for the pH in original crystal screen conditions. Using citrate as a pH buffer around 3.5, crystals could be obtained again and improved in size and be detached from other crystals. Further improvement of crystals

could be achieved by including 0.1 M LiCl as was indicated by examination of the additive screen conditions (Hampton). Increase of PEG 4000 concentration led to growth of crystals to their maximum size and most regular shape. Final crystallization conditions using 10 mg/ml DHC2 protein were 32.5% PEG 4000, 0.1 M LiCl and 0.1 M Citrate buffer pH 3.5.

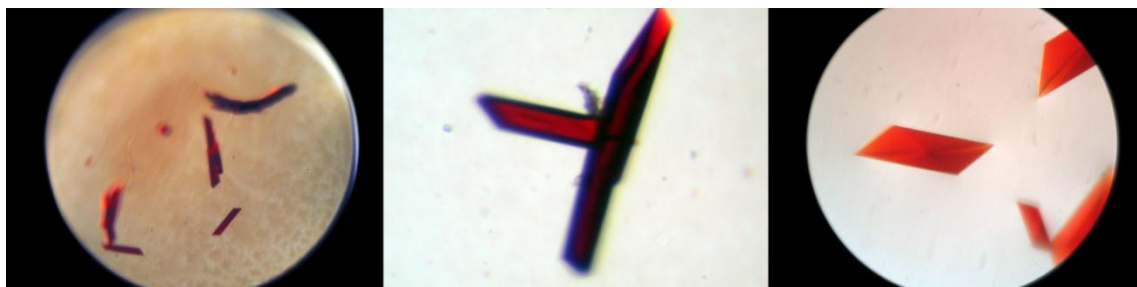


Figure 24: Crystal pictures of DHC2 at different stages of crystallization optimization. (Left) Initial crystal form as observed in crystal screen I condition 30. (Middle) Crystal form after increasing PEG 4000 concentration to 32.5 % and addition of 0.1M LiCl to the crystallization drop. (Right) Crystal form after increasing the protein concentration to 10 mg/ml.

Unfortunately all crystals were sticking to the ground of the sitting drop plate and had to be loosened prior to X-ray crystallography experiments. For measuring crystals of DHC2, loosened crystals were taken from the drop solution using a nylon loop and transferred to a cryo-protection buffer containing 35% PEG 4000, 0.1 M LiCl and 0.1 M citrate buffer pH 3.5 and flash-frozen in liquid nitrogen. Crystals were tested using a rotating anode generator (Rigaku) providing Cu-K α radiation and a mar345dtb image plate detector (MAR Research). Crystals of suitable quality were measured at the tunable beamline BW7A at the EMBL outstation at DESY (Hamburg). Three isomorphous data sets of a single DHC2 crystal were collected at two different wavelengths at the iron K-edge (a) maximizing f'' contribution, (b) at the minimum of f' and at a remote wavelength, giving the maximum intensity of synchrotron radiation. Additionally, two different data sets were collected from a second crystal, which showed diffraction behavior superior to the crystal used for structure solution by MAD (Hendrickson *et al.*, 1985). Data was indexed, integrated, scaled and merged using XDISP, DENZO and SCALEPACK (Otwinowski & Minor, 1997). The space group was determined to be P2₁ with unit cell parameters of $a = 39.64 \text{ \AA}$, $b = 55.67 \text{ \AA}$, $c =$

39.63 Å and $\beta=105.91^\circ$. The asymmetric unit contained two monomers with a molecular mass of 11.31 kDa each.

	remote 1	peak (f')	inflection (f')	remote 2
wavelength (Å)	0.9941	1.7394	1.7414	0.9941
resolution range (Å)	50.0-1.6 (1.64-1.60)	50.0-2.25 (2.30-2.25)	50.0-2.25 (2.30-2.25)	50.0-1.5 (1.53-1.50)
no. of reflections	302826	69359	75505	516137
no. of unique reflections	21838	6516	6515	26375
completeness (%)	95.7 (95.9)	81.3 (86.2)	81.2 (83.2)	97.9 (95.1)
$I/\sigma(I)$	10.6 (1.9)	9.0 (6.3)	9.2 (5.6)	10.3 (1.6)
R_{sym}	0.040 (0.47)	0.062 (0.186)	0.046 (0.194)	0.049 (0.559)

Table 7: Data collection statistics for DHC2 crystals. Data sets for remote 1, peak and inflection were taken from the same crystal to obtain isomorphous data for structure solution by MAD. Data set remote 2 was collected from a second crystal intended for refinement against high resolution data. Values in parentheses are given for the highest resolution shell.

3.4.7 Structure determination and refinement

The position of four iron atom sites in the asymmetric unit were identified using SHELXD (Schneider & Sheldrick, 2002). Phase calculations and further density improvements were carried out in SHELXE (Sheldrick, 2002). An initial structural model of DHC2 was fitted into the experimental electron density by using the baton mode in O (Jones *et al.*, 1991), building the C- α backbone trace of the protein. Including several amino acid residue side chains and four heme groups, based on the visible electron density, constituted an initial starting point for DHC2 model refinement against the second data set remote 2. Refinement was carried out in REFMAC 5 (Collaborative Computational Project, Number 4, 1994; Murshudov *et al.*, 1997) using data to 1.5 Å resolution and an isotropic B factor model. The refined model of DHC2 contained 137 amino acids, four heme groups, four sulfate ions, two sodium ions and 153 water molecules. The refinement was carried out to a total crystallographic R-factor R_{cryst} of 0.221 and an R_{free} of 0.272 (Brünger, 1993). Correctness of the structural model was verified using PROCHECK (Laskowski *et al.*, 1993). In a Ramachandran plot, torsion angles of 85.7 % of the residues occupied the most favored regions, 14.3 % the allowed regions and no residues the disallowed regions (Ramachandran *et al.*, 1963).

no. of protein atoms	1118
no. of heteroatoms	22
no. of heme atoms	172
no. of water molecules	153
R_{cryst}	0.221
R_{free}	0.272
resolution range (Å)	50.0-1.5
mean B value (Å ²)	29.3
DPI (R_{cryst} -based) (Å)	0.101
DPI (R_{free} -based) (Å)	0.107
rms deviation from ideality	
bond lengths (Å)	0.014
bond angles (deg)	1.581

Table 8: Refinement statistics of DHC2, where DPI is the diffraction precision index introduced by Cruickshank (Cruickshank, 1999) and rms is the root mean square.

3.4.8 Protein structure of DHC2

The structural model of DHC2 consists of two monomers in the asymmetric unit, each containing two covalently bound heme groups. The recombinant character of DHC2 is expressed in ordered amino acid residues, representing part of the hexa-histidine affinity tag and the linker region with Leu95 and Glu96. Monomer A therefore consists of amino acids number 28-97 and monomer B of amino acids 28-96 with respect to the numbering of the unmodified *dhc2* gene. The atom positions in superimposed monomers differ by root-mean-square of 0.25 Å. For further analysis therefore only structural features of monomer A will be discussed. Secondary structure elements are of α -helical type only and are ranging from Lys40 to Ala47 for helix I, from Lys52 to Cys59 for helix II and a further short, α -helical segment is built from residues Cys87 to Cys90, representing the only secondary structure in the C-terminal region of DHC2. The heme groups of DHC2 are bound via two thioether bonds to cysteine side chains in a classical *c*-type cytochrome binding motif C-X-X-C-H, with the heme group I binding motif consisting of the sequence C-V-E-C-H ranging from Cys59 to Cys63 and the heme group II binding motif of the sequence C-L-F-C-H ranging from Cys87 to Cys91. Heme groups are, as already indicated by spectroscopical data, bis-histidinyl liganded, resulting in the observed low-spin Fe(III) configuration. Histidine ligands not originating from the *c*-type cytochrome binding motif, occupying the sixth ligand positions are His39 for heme group I and His76 for heme group II.

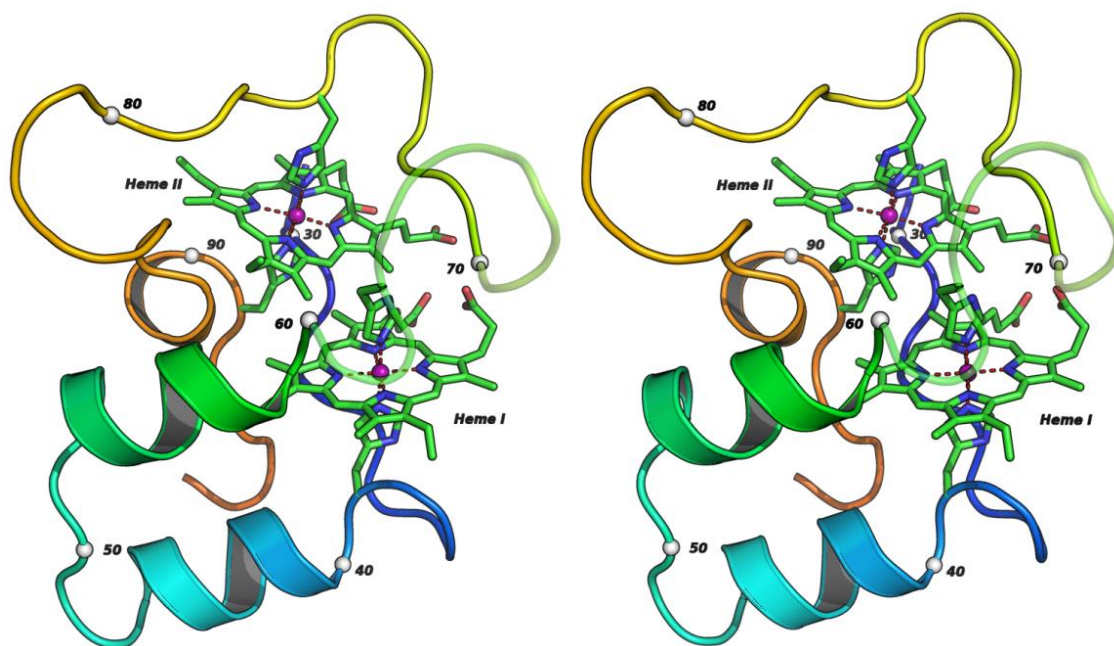


Figure 25: Stereo representation of DHC2 monomer from *Geobacter sulfurreducens*. White spheres indicate every tenth C- α trace atom, sticks are used to draw heme groups and their axial ligands, as cartoon representation in a continuous color gradient from blue to red shows the secondary structure elements of DHC2 as observed from the N- to the C-terminus.

3.4.9 Folding properties

In addition to its axial ligand function, His76 seems to play a crucial role in anchoring parts of the N-terminal loop region to the protein, as this flexible region has only a weak association to the folding core (Figure 26). The residues being cloning artifacts or part of the affinity tag are not involved in folding or the overall organization of the protein core, as the N- and C-termini are pointing towards the solvent space. Despite the small number of hydrophobic amino acids in DHC2, Val33, Phe42, Ile70, Phe72 and Leu88 are building a central hydrophobic core between the heme groups and α -helices I and II, stabilizing the overall fold of DHC2. The rigidity of this folding core is indicated by B-values well below the mean of 29.3 \AA^2 . Direct interactions of hydrophobic amino acid residues with side chains with aromatic porphyrin macrocycles are present for Val33, Phe42, Ile70, Phe72 and Leu88 with distances below 3.7 \AA . In contrast to the well ordered folding core, the C-terminal loop region shows a poor degree of order, indicated by the worse quality of the electron density maps and higher B-factors.

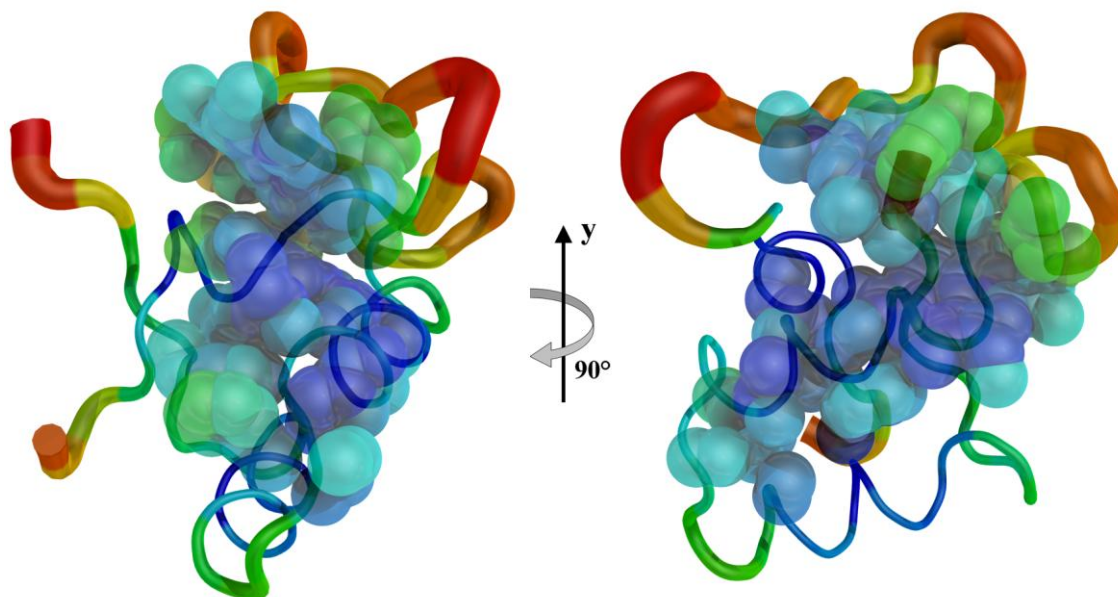


Figure 26: Representation of DHC2 folding and flexibility properties, color from blue to red indicates a gradient of rising temperature B factor indicating rigidity and flexibility of protein regions. The protein backbone is represented in b factor putty representation giving a relative estimate for atom position displacements by the thickness in the representation. Spheres showed in half transparent representation indicate amino acid residue side chains of hydrophobic amino acids in DHC2 and heme groups, with spheres scaled to the Van-der-Waals radii of the respective elements. This illustration clearly emphasizes the presence of a hydrophobic folding core build up between helix I, II and the adjacent heme groups.

3.4.10 Heme group configuration

Folding properties already revealed a central positioning of the heme groups in the overall structure of DHC2. The heme groups are therefore flanked by helix I and II at one side and several loop residues at the opposite site. Porphyrin macrocycle atoms exhibit a minimum edge-to-edge distance of 4.3 Å, whereas the iron atoms have a distance of 9.4 Å. The mutual heme group arrangement is characterized by nearly parallel porphyrin planes, with their relative plane normals at 5° to each other. Deviation from an ideal, parallel configuration is indicated by a minimum intersection angle of plane normals by 17°. The dihedral angle of the axial ligands imidazole planes (His39 and His63) of heme I is 36°. An almost coplanar arrangement is found for the imidazole planes of residues His76 and His91, comprising a dihedral angle of only 2° at heme group II. Hydrogen bonding networks might be involved in the positioning of histidine imidazole planes, as N^δ atoms of both axial ligands in heme group II are H-bonded to backbone carbonyl oxygens at a distance of 2.7 Å. Whereas His39 of heme group I forms a longer H-bond to the carboxy group of Asp36 with a distance of 2.75 Å.

In contrast to this, His63 is bonded to a water molecule at a distance of 2.8 Å only. Planarity of the porphyrin macrocycle is almost perfect for heme group I, with only pyrrole ring B showing a slight out-of-plane arrangement. In contrast, heme group II shows a strong distortion of the porphyrin macrocycle from ideal planarity, attaining a ruffled conformation, where two opposing meso-carbon atoms are either shifted below or above an imaginary heme plane.

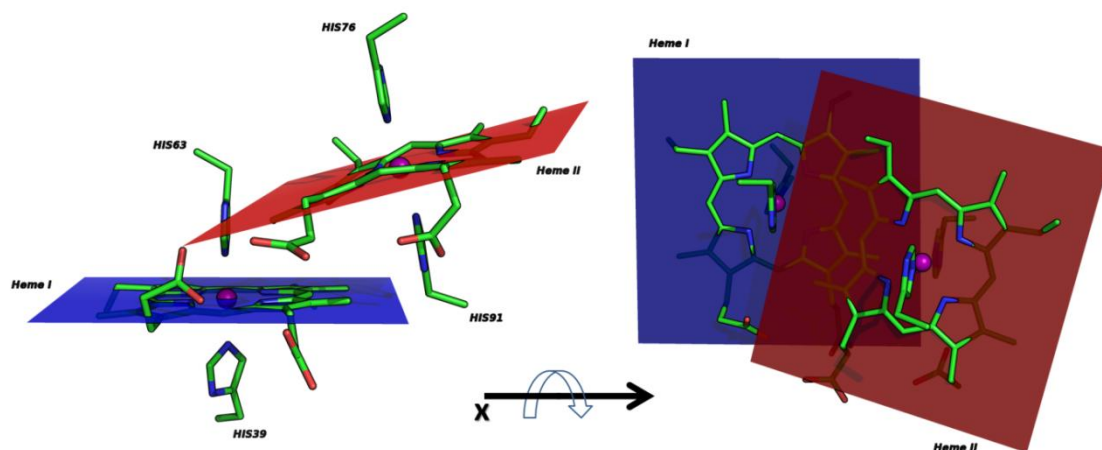


Figure 27: Heme group organization of di-heme *c*-type cytochrome DHC2, with sticks representing the heme groups and the axial ligands, where Fe(III) is shown in sphere representation. Planes were approximated to fit pyrrole nitrogen atoms and are indicated the distortion of heme groups from an ideal planar arrangement, as well as the mutual geometrical arrangement.

3.4.11 Protein surface

The total solvent accessible surface of DHC2 calculated using MSMS (Sanner *et al.*, 1996) comprises 4201 Å², whereby amino acids contribute 87.5 % and heme groups 12.5 % to the overall surface area. The individual contributions of the heme groups are 5.8% for heme group I and 6.7 % for heme group II, respectively. As the porphyrin macrocycle does only carry charge at the carboxyl-groups of the heme propionates, these heme areas are solvent exposed. Electrostatic potential maps, calculated with APBS (Baker *et al.*, 2001) indicate negative charged surfaces for the propionates and surrounding area. Positive surface areas were observed for the C-terminal loop regions and the second heme binding motif, forming an α -helical like structure (Figure 28).

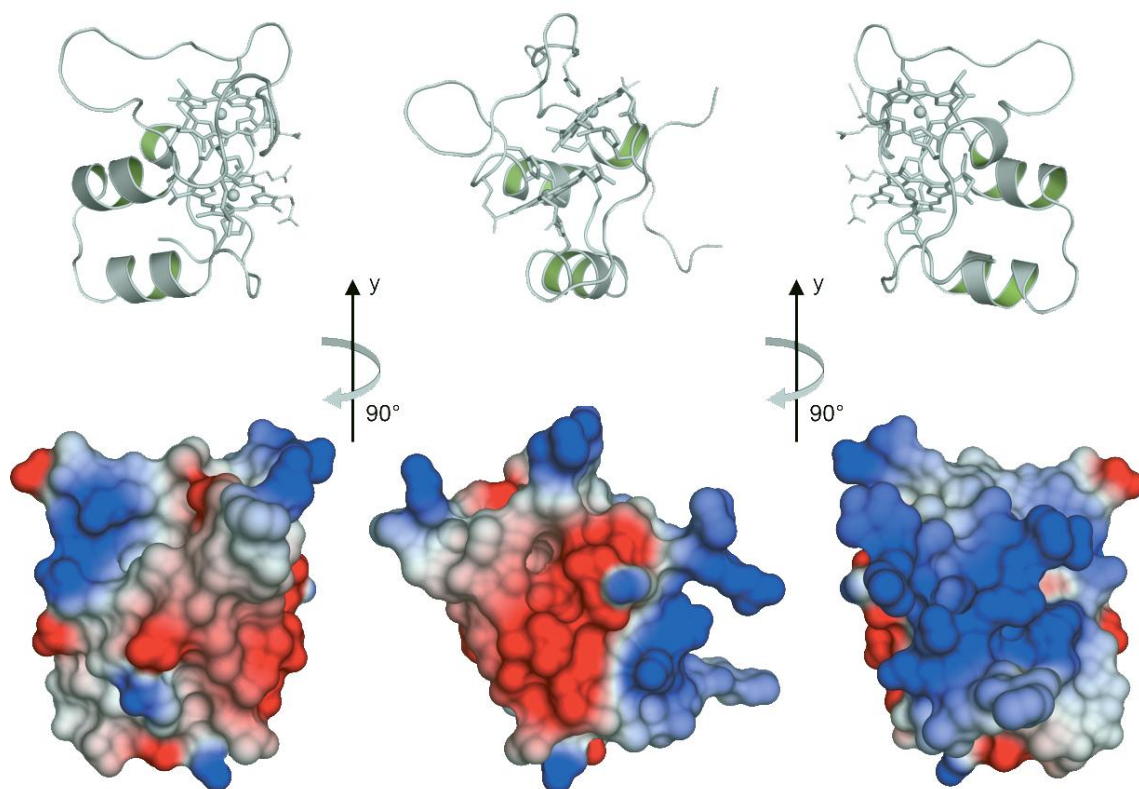


Figure 28: Surface representations of the diheme *c*-type cytochrome DHC2 from *Geobacter sulfurreducens*, where negative charges are indicated by red color, positive charges by blue color and uncharged surface areas are represented by white color. Orientation of the protein surface is indicated by a structural model depending on rotation around axis *y*.

3.5 Discussion

3.5.1 Refinement process of DHC2

In the structural refinement process of DHC2 model against diffraction data, R_{cryst} and R_{free} were refined to values of 22.1 % and 27.2 %, respectively. With experimental X-ray crystallographic data from DHC2 being of reasonable quality as indicated by several data quality indicators, these values for R_{cryst} and R_{free} seem to be too high for data of this resolution. Despite this fact, no further improvement of the R factors could be achieved by altering any parts of the protein model. Also no clue for further missing parts in the model could be drawn from the electron density map, indicating that the actual model represented a state of convergence in refinement. On the other hand, the DHC2 electron density map is well shaped and quality indicators do not point to any problem with structural data, despite a slight disorder of N-terminal loop regions.

3.5.2 Significance of DHC2 for the family of multi-heme cytochromes

Multi-heme *c*-type cytochromes are of high importance for respiratory and enzymatic reactions in bacterial metabolisms. The functionality of this protein family is gained by specifically adjusting the arrangement and potentials of multiple redox centers within a protein scaffold. The occurrence of highly conserved parallel and perpendicular structural heme packing motifs has to be connected to a general mechanism for how the physiological redox-reactions of these proteins work (Einsle et al, 2001). Biochemical and structural data indicate that DHC2 is a novel member of this family, with no other homolog found in the databases so far, except for a paralog from *Geobacter sulfurreducens* itself, which was denoted DHC1 (GSU2767). Heme binding motifs encoded by the *dhc2* sequence showed covalent attachment of functional heme groups to the protein moiety and a parallel packing motif. Despite the presumed involvement of DHC2 in an electron conduction process, no further information about its physiological relevance has been obtained. With its small size and basic characteristics DHC2 - and probably DHC1 - are soluble electron carriers residing in the periplasmic space of *Geobacter sulfurreducens*.

3.5.3 Conservation of structural heme packing motifs

Structural organization of heme packing motifs as observed in DHC2 structure is used in a broad set of redox active proteins, such as the cytochrome *c* nitrite reductase from *Sulfurospirillum deleyianum* (Einsle *et al.*, 1999), *Wolinella succinogenes* (Einsle *et al.*, 2000), *Escherichia coli* (Bamford *et al.*, 2002), *Desulfovibrio desulfuricans* ATCC 27774 (Cunha *et al.*, 2003) and *Desulfovibrio vulgaris* (Rodrigues *et al.*, 2006), the hydroxylamine oxidoreductase and its physiological redox partner cytochrome *c*₅₅₄ from *Nitrosomonas europaea* (Igarashi *et al.*, 1997; Iverson *et al.*, 1998), the split-Soret cytochrome *c* from *Desulfovibrio desulfuricans* ATCC 27774 (Abreu *et al.*, 2003), flavocytochrome *c*₃ fumarate reductase from *Shewanella putrefaciens* (Leys *et al.*, 1999) or *Shewanella frigidimarina* (Bamford *et al.*, 1999; Taylor *et al.*, 1999), the octaheme tetrathionate reductase from *Shewanella oneidensis* MR-1 (Mowat *et al.*, 2004) and the small subunit NapB of the heterodimeric nitrate reductase from *Rhodobacter sphaeroides* (Arnoux *et al.*, 2003). Superimposition of parallel, bis-histidinyl liganded heme packing motifs observed in these proteins to the parallel motif obtained by structural investigation on DHC2 reveals an overall spatial conservation of the porphyrin macrocycle positions. The positions of Fe and porphyrin planes seen from the direction of the plane normal are very similar, with very slight deviations in plane normal angles. Observation of the heme planes in a perpendicular view to the plane normal on the other hand reveals divergent kinds of bending of porphyrin macrocycles with respect to an ideal plane, which seems to be a relatively weakly conserved arrangement parameter. A comparison of the relative axial ligand positions reveals a good alignment of histidine imidazole planes of the C-X-X-C-H binding motif, whereas imidazole planes of the distal ligand show large mutual divergence as well in relative arrangement of imidazole plane to the porphyrin macrocycle as for imidazole plane dihedral angles with the proximal ligand plane. The propionate groups of the porphyrin macrocycle show no indication of a common arrangement, as they tend to be very flexible and their arrangement is rather dependent on the surrounding solvent. Two conclusions can be drawn from these observations:

The porphyrin plane parameters are conserved to a high degree. This is true for the proximal heme ligand conserved in imidazole plane positioning as well as in a relative tilting degree. Both findings seem to be a direct consequence of the tight structural alignment induced by the C-X-X-C-H binding motif. On the other hand

distortion of the heme plane from ideality and distal heme ligand seems to provide a relatively, freely tunable system for altering redox potentials and protein physiological functionality.

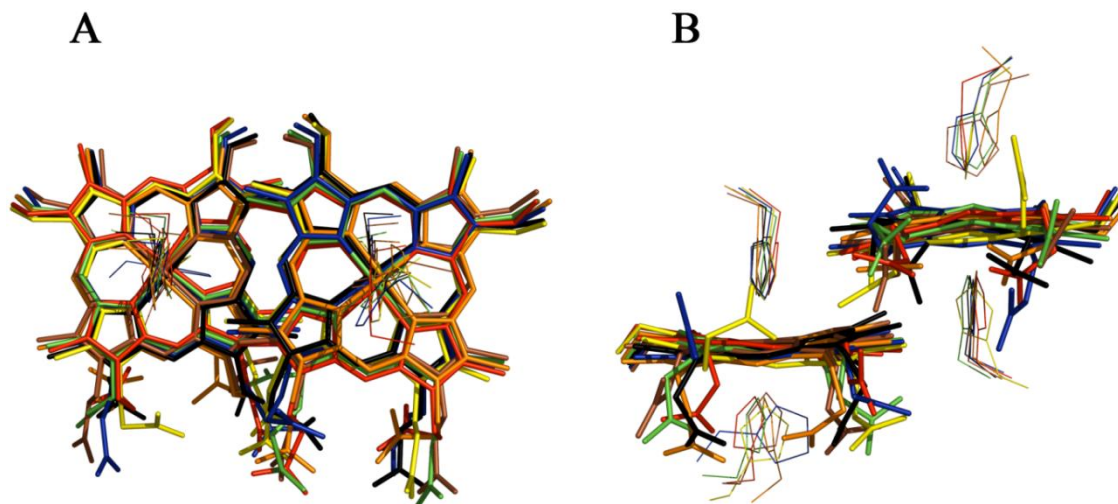


Figure 29: Superposition of parallel heme packing motifs from *Geobacter sulfurreducens* DHC 2 (red), *Nitrosomonas europaea* *c*₅₅₄ (blue), *Shewnella putrefaciens* fumarate reductase (brown), *Desulfovibrio desulfuricans* split-Soret cytochrome (black), *Haemophilus influenzae* NapB (orange) and *Sulfurospirillum deleyianum* ccNiR (green). The top view (A) and the side view (B) show similar heme group arrangement and positioning, but strong differences in imidazole plane angles, distances and porphyrin distortion (Heitmann & Einsle, 2005).

3.5.4 DHC2 structure in comparison to NapB

Despite showing a similar structural arrangement of heme groups on one hand, the above systems are not applicable by recombinant techniques and their analysis is complicated by the presence of more than two heme groups, rendering the observation of sole heme group signals impossible. Only one *c*-type cytochrome seems to exhibit similar properties to DHC2, the small subunit of the periplasmic nitrate reductase, NapB, from *Haemophilus influenzae* (Brigé *et al.*, 2005), with a molecular weight of 14.7 kDa and a comparable, parallel heme packing motif. However, structural similarities are limited to these points only (Figure 30). Conspicuous is the total absence of structure homology beside the two heme binding motifs and the large structural deviations from the heterodimeric periplasmic nitrate reductase complex of *Rhodobacter sphaeroides* (Arnoux *et al.*, 2003), with the only common structural arrangement being the heme group arrangement in NapB. This could be an indication for a non physiological conformation of NapB when crystallized without its

physiological interaction partner and could also indicate that the heme groups have a stronger association behavior than a presumable protein folding core.

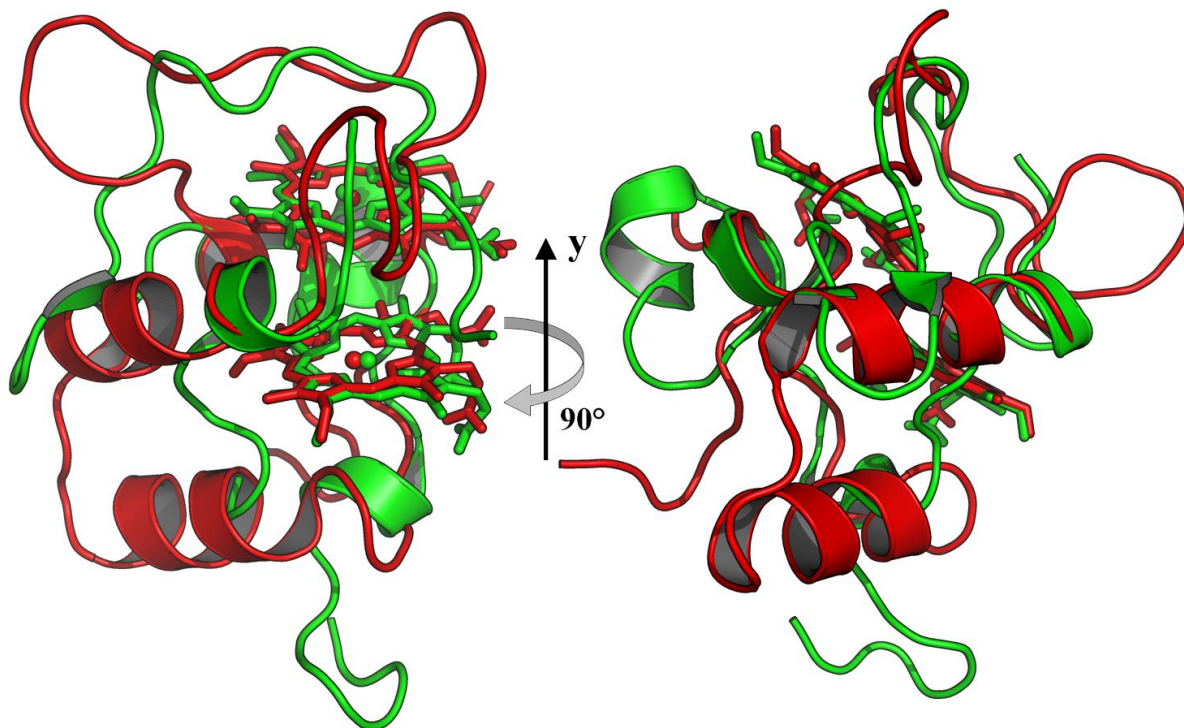


Figure 30: Superimposition of DHC2 from *Geobacter sulfurreducens* and the small subunit of the periplasmic nitrate reductase complex NapB from *Haemophilus influenzae*. It exhibits only structural similarity for heme group arrangement and their α -helical like binding motifs. In contrast to DHC2, the structure of NapB is lacking well defined secondary structure elements.

3.5.5 Biochemical properties of DHC2

UV/Vis absorption spectra of DHC2 are typical for bis-histidinyl-liganded *c*-type cytochromes (Pettigrew & Moore, 1987) and the spectra of the individual heme groups are indistinguishable. This finding is supported by a reoxidation experiment that shows a continuum of spectra during the oxidation process. No separation of peaks or prominent shoulders could be observed during this experiment, in contrast to diheme peroxidases where heme groups seem to be independent (Hoffmann, 2007). Together with the relative broadening of EPR spectra and the short iron to iron distance of 9.4 Å, as well as an edge to edge distance of porphyrin macrocycles about 4.3 Å, DHC2 hemes couple in an effective way, leading to spatial delocalization of electrons conducted to this redox centers. The mid-point potentials of -135 mV and -289 mV reveal a consecutive reduction process, where the uptake of one electron increases the reduction

potential of the whole system so that the second electron has a much higher reduction potential itself. Spatial arrangement of heme groups in large cluster sizes could therefore ensure sequential reduction by raising a certain potential of the overall cluster to higher values, than would be possible for single hemes alone. Mid-point potentials for parallel packing heme motifs have also been obtained for NapB from *Haemophilus influenzae* yielding values of -25 and -172 mV using cyclic voltametry (Brigé *et al.*, 2001) and -60 and -205 mV for the *Rhodobacter sphaeroides* protein using optical redox titrations (Arnoux *et al.*, 2003). For cytochrome *c*₅₅₄ from *Nitrosomonas europaea* mid-point potentials of their parallel packing heme motif have been determined to be -147 mV and 47 mV (Arciero *et al.*, 1991; Upadhyay *et al.*, 2003). The large diversity of mid-point potentials indicate that beyond the general structural conservation of heme group packing motifs, proteins utilizing this device have a very flexible way to fine tune the redox potential for their needs. This finding was already indicated by the structural alignment of heme groups, showing large differences in the configuration of histidines and the distortion of the porphyrin macrocycle.

3.5.6 Structural arrangement of DHC2

Secondary structure elements building up DHC2 are of α -helical type only, as was observed for many *c*-type cytochromes. Stabilization of the protein is achieved by a rigid hydrophobic folding core made up by α -helix I, II and the heme groups as well as hydrophobic amino acids of the second heme binding motif (C-L-F-C-H), leaving the opposite C-terminus and the propionates of hemes relatively exposed to the surrounding solvent (Martin *et al.*, 1998). This observation is emphasized by the distribution of the overall atomic displacement temperature factor, being low for the folding core, the heme groups and the heme binding motifs, but high for the C-terminal loop region ranging from residue 72 to 86, that is only loosely associated to the rest of the protein by a histidine ligand anchor formed by His76 and heme group II and is comparable to the situation in the split-Soret cytochrome *c* (Abreu *et al.*, 2003). Loose arrangement and hydrophobic residues pointing to the outer solvent make this region a putative interaction site for a physiological redox partner as heme to heme interactions either in a parallel or perpendicular mode would be possible in terms of spacing, as can be seen from the low covering of heme groups by protein amino acid residues.

3.5.7 Heme group characteristics of DHC2

The superimposition of heme groups (Figure 29) showed large deviations of the porphyrin macrocycles from planarity that is not observed in free solution or most model complexes (Jentzen *et al.*, 1995; Shelnut *et al.*, 1998). This distortion is termed puckering and is imposed by the protein itself, as it can shape the porphyrin macrocycle by providing a special surrounding of amino acid residues. On the other hand positioning of distal histidine ligand imidazole planes is very flexible, in contrast to the proximal histidine ligands. Both features of heme group arrangement should have a prominent influence on mid-point potentials and the physiological function of the proteins (Barkigia *et al.*, 1988; Kadish *et al.*, 1995; Ravikanth & Chandrashekar, 1995; Ma *et al.*, 1998). DHC2 itself shows several kinds of this fine tuning mechanisms, as imidazole planes of histidine ligands of heme group II are nearly coplanar and ligands of heme group I comprise a dihedral angle of 36° . These features were already indicated without knowledge of the crystal structure from rhombic EPR spectra, showing g values of 2.96 for the highest g value (called g_{\max}) (Walker, 2004). For nearly perpendicular arranged imidazole planes ($>70^\circ$) of bis-histidinyl ligands, g_{\max} values above 3.0 are common and therefore denoted as "large g_{\max} " spectra. A value below 3.0 in maximal g value, points to a structural arrangement as observed in this case. Having such a consequence for energy levels of electrons as observed for EPR spectra this emphasizes the importance of imidazole plane arrangement for the fine tuning of redox potentials (Walker, 2004). Structural data of DHC2 also shows two different modes of porphyrin macrocycle distortion. Heme group I is nearly planar, where heme group II is strongly distorted, with one set of opposite meso-carbons lying above and the other below an imaginary ideal heme plane. This arrangement is termed ruffled and it was discussed that it forces bis-histidinyl imidazole planes into a mutually perpendicular arrangement (Yatsunyk *et al.*, 2003). However the contrary was observed for DHC2, as the imidazole plane of His76 is directly pointing to the meso-carbons distorted in its direction. Either the effect of ruffling on imidazole plane arrangements is small or the specific protein surrounding forces the imidazole plane in this distinct conformation, negating the energetically unfavored arrangement. Two structural properties could therefore force the imidazole planes in this arrangement. His76 could be stacked by the relative bulky aromatic side chains of Phe72 and Pro77, forbidding out of plane rotations and

effectively preventing a perpendicular imidazole plane arrangement. Hydrogen bonds of N^{δ} atoms of axial ligand imidazole planes to the protein might also be involved in positioning, as observed in the case of DHC2.

3.6 Summary

Geobacter sulfurreducens as a source for *c*-type cytochromes was used for the identification and biochemical and structural characterization of a diheme *c*-type cytochrome that shows a heme packing motif that is conserved in members of the multi-heme *c*-type cytochrome family, such as the cytochrome *c* nitrite reductase, the hydroxylamine reductase or cytochrome *c*₅₅₄. DHC2 shows typical UV/Vis absorption and EPR spectra indicating that heme groups act as a coupled pair and already indicate coordination geometry at the central iron atoms as was confirmed by structural data. Redox titrations exhibited mid-point potentials of -135 and -289 mV, differing from potentials of other parallel heme packing motifs and indicating mechanisms for freely fine tuning the redox potentials by the surrounding protein. Structural properties of imidazole plane arrangements and porphyrin macrocycle distortion, according to EPR spectrum properties, were identified to vary by a large degree and therefore are key elements in fine tuning redox potentials. With its small size, good expression yields, easy applicable gene sequence, stable folding, diverse structural optimizations for fine tuning the redox potential of a structural highly conserved parallel heme packing motif, DHC2 represents an ideal model system for the investigation of conserved heme packing motif properties. Mutational studies on DHC2 will be carried out, with the aim of specifically changing functional features such as the axial heme ligands and residues in the immediate surroundings, but also introducing structural changes forcing a different heme group arrangement itself. Mutational studies are then accompanied by biochemical and structural analysis and can be related to the overall meaning of conserving diheme *c*-type packing motifs on a structural scale. A significant drawback represents the structural refinement process, which seems to be obscured by crystal properties that cannot be derived from the structural model itself or the quality of diffraction data. This seems to imply a more complicated crystallographic problem that has to be accounted further.

4 Twinning in crystals of DHC2

4.1 Introduction

4.1.1 Crystal twinning

The phenomenon of crystal twinning describes the geometrically oriented intergrowth of two or more individuals belonging to the same crystalline phase (Friedel, 1926; Giacovazzo, 2002; Yeates, 1997; Parsons, 2003) and forming a twinned crystal. Mutual orientations of the individuals forming the crystal lattice are described by twin operations. The twin operation is a symmetry operation that relates individuals in the twinned crystal and belongs to a crystallographic point group, but does not belong to the holohedral point group of specialized metric. The fractional contribution of each individual forming a crystal lattice is described by the twin fraction, which therefore is an indicator for the degree of twinning (Herbst-Irmer & Sheldrick, 1998). Individuals contribute to the diffraction pattern of an X-ray crystallography experiment by relating reflections depending on the twin law and intensities depending on the twin fraction. Reflection intensities of different protein individuals are therefore related by rotation operations only. Classifications of twinning types have been made according to relations of the twin lattice and the lattice of the individual. These are twinning by merohedry, pseudo-merohedry, reticular merohedry and non-merohedry (Herbst-Irmer & Sheldrick, 1998).

- Twinning by merohedry is the simplest case of twinning, where two or more individuals are related by a twin operation that is not belonging to the point group of the crystal pattern, but is of the same crystal system. Crystal lattices of individuals are thus overlapping perfectly, forming a twin lattice with a diffraction pattern that shows no signs for the presence of twinning. Only intensity statistics, can reveal the presence of twinning in these cases. Protein crystals only allow for the presence of holohedral twins, where the twin operator does not belong to the Laue group of the crystal pattern. Occurrence of merohedral twinning is therefore limited to tetragonal, trigonal, hexagonal and cubic crystal systems (Liang *et al.*, 1996; Yeates & Fam, 1999; Dauter, 2003).

- Twinning by pseudo-merohedry describes a twin operation relating two or more individuals by a symmetry operation that is of a higher crystal system and corresponds to the present crystal metric, but is not part of the holohedral point group of specialized metric (the symmetry of the individual structure). Crystal lattices of the twin individuals approximate a common twin lattice, but are not perfectly overlapping. Diffraction patterns arising from pseudo-merohedral twinned crystals exhibit splitting or broadening of reflections when analyzed at high resolution (Herbst-Irmer & Sheldrick, 1998). Pseudo-merohedral twinning occurs when the metric of the symmetry is that of higher crystal symmetry, for example when β is close to 90° in a monoclinic system an orthorhombic crystal system (holohedry) can be emulated (Donnay, 1940; Yeates & Fam 1999, Dauter, 2003, Parsons, 2003).
- Twinning by reticular merohedry can occur when a crystal belongs to a holohedral point group and its lattice does not have a specialized metric. As the point group of the lattice and the crystal patterns coincide, no merohedral twinning is possible. If a sublattice has a different point group than the original lattice and the translation group is a subgroup of the original lattice, then twinning by reticular merohedry is present, with a twin operator relating the different orientations of the individuals. Such twins are classified by an additional number, the twin index, which is a ratio of lattice nodes of the individual to the number of lattice nodes restored by the twin operation. Obverse/reverse twinning in rhombohedral space group is a typical example of twinning by reticular merohedry. The reciprocal lattices from different domains superimpose exactly, but the systematically absent reflections for one domain are present for the other domain and *vice versa* (Herbst-Irmer & Sheldrick, 1998).
- Twinning by non-merohedry (NMT) is characterized by an arbitrary twin law and represents a combination of reticular and pseudo-merohedral twinning. Depending on the twin axis there are existing certain sets of exactly overlapping lattice nodes that are visible as single reflections (EOR), close lattice nodes that

can be perceived as split reflections (POR) and individual lattice nodes, visible as non-overlapping reflections (NOR). The presence of NMT can normally be observed from inspection of the diffraction pattern. NMT can occur in every space group and the twin operation can be any n -fold rotation (Herbst-Irmer & Sheldrick, 1998).

4.1.2 Twinning in monoclinic crystal systems

The definition of twinning by pseudo-merohedry comprises the emulation of a higher crystal system by fortuitously matching unit cell parameters of lower crystal symmetry. Monoclinic crystal systems therefore can be for example pseudo-orthorombic with their unit cell parameters being equivalent to a unit cell setup of the higher crystal system and their reflection data being related by a symmetry element (in this case twinning) pointing to a false symmetry. Several ways of unit cell parameter setups have been observed and are well described in literature.

- The first possibility arises when an angle β in the monoclinic system is close to 90° and therefore conveys orthorombic symmetry. The primitive cell setup in this case is equivalent to the cell setup in orthorombic symmetry. The unit cell parameter setup adds to the monoclinic twofold operator $-h, k, -l$ the possibility of twofold rotation around the basis vectors \mathbf{a} or \mathbf{c} . In this case the twin individuals are related by the twin law $-h, -k, l$ or $h, -k, -l$ transforming one individual into another. With twin fraction α being close to 0.5, data can be easily merged in space group $P2_12_12_1$, as observed e.g. for the cocaine hydrolytic antibody 15A10 (Larsen *et al.*, 2002).
- In the second setup, unit cell parameters fulfill the condition $c \cos \beta = -a/2$ constructing a minimal supergroup in an orthorombic system of the primitive monoclinic cell using the latter condition. The twin law is a rotation around the \mathbf{c}^* basis vector of the orthorombic cell, which is the vector normal to the \mathbf{ab} (110) plane. The twin individuals thus share the same basis vectors \mathbf{a} and \mathbf{b} , but mutual orientations are of opposite directions, resulting data can be indexed in space group $C222_1$ and the corresponding twin law is $-h, -k, h + l$. Cases of this type of twinning were for example depicted in structure solution of the monoclinic crystal form of human peroxiredoxin 5 (Declercq & Evrard, 2001) or

in the monoclinic crystal form of the $\chi\delta$ T-cell ligand T10 (Rudolph *et al.*, 2004).

- A third case of twinning in monoclinic setups requires unit cell parameters of $a \approx c$, with the primitive cell being a subgroup of a conventional cell in a c-centered orthorombic setup, with $a^* = 2a \cos(\beta/2)$, $b^* = 2a \sin(\beta/2)$ and $c^* = b$. This setup allows for superimposition of the cells by a rotation around the **ac** diagonal emulating orthorombic symmetry. Data can be indexed in space group $C222_1$ and the corresponding twin law is $l, -k, h$. Pseudo-merohedral twinning of this kind was for example observed in the structure of the capsid-stabilizing protein of bacteriophage λ gpD (Yang *et al.* 2000) and the human orotidine-5'-monophosphate decarboxylase (Wittmann & Rudolph, 2007).
- Recently a fourth case was reported (Sultana *et al.*, 2007), where two-domain twinning of the third case was present in crystals of FAD-dependent aclinomycin oxidoreductase (AknOx) from *Streptomyces galileus*, with $l, -k, h$ as a twin operator, but also included a feature so far only observed in small molecule crystallography, which cannot be denoted twinning, as three domains are related by two symmetry operations of a hexagonal crystal system and was therefore referred to as *Drilling* (German for tripling; Sheldrick, 1997). Fortuitous unit cell parameters of $a \approx c$ and β being close to 120° allowed for this pseudo-hexagonal setup, with symmetry operators relating the twin individuals by $l, k, -h -l$ and $-h -l, k, h$.

4.1.3 Twinning in biological X-ray crystallography

Two reflections observed for a twinned crystal and related by a twin operation, have intensities I_1 and I_2 depending on intensities J_1 and J_2 from reflections of an untwinned crystal weighted by the twin fraction α according to:

$$I_1 = \alpha J_1 + (1 - \alpha) J_2$$

$$I_2 = (1 - \alpha) J_1 + \alpha J_2$$

With reflection intensities altered and introduction of additional symmetry into the diffraction pattern, data analysis can be obscured, especially when the twin fraction approximates 0.5. When the fractional contribution of both individuals to the twin lattice is equal, one speaks of perfect twinning, which can render structure solution impossible. In case of twinning by pseudo-merohedry a higher crystal symmetry might be chosen by auto-indexing programs than would be chosen for untwinned crystals. Differences in the merging R factor of the pseudo higher crystal system setup and the real crystal system get less pronounced with rising twin fraction α and are even not existent in the presence of perfect twinning. Merging of data in an actual higher space group than existing, renders structure solution impossible, as symmetry simulated by twinning is in fact not present. Data of twinned crystals that have been merged in the correct symmetry and show a moderate twin fraction, allow for structure solution by standard methods (Yang *et al.*, 2000). Not being aware of twinning in those cases will obscure the refinement process of the structure, as structure factor amplitudes $|F_{\text{calc}}|$ calculated from the model atom positions and B-factors will never converge to $|F_{\text{obs}}|$ in a satisfying way, as one is only refining the contribution of the (normally) strongest contributing twin individual, neglecting contribution to intensities from the other twin individuals. The case is even worse when refining the model of the weaker contributing twin individuals. Only with knowledge of the apparent twin law a refinement of twins to a reasonable R factor is therefore possible.

4.1.4 The Order-disorder (OD) theory

The order-disorder (OD) terminology was developed to describe disorder phenomena observed in crystal lattices. The basic concept of this theory introduced by Dornberger-Schiff in the 1950s is the existence of layer slabs through a crystal forming the overall crystal lattice (Dornberger-Schiff, 1956; Dornberger-Schiff & Grell-Niemann, 1961), with a set of recurring symmetry operations converting one layer into the adjacent layer (layer group). This special transformation of adjacent layers by this symmetry operation is only existent for a subspace of the whole crystal and can therefore be denoted as local symmetry. This symmetry is not part of the structure symmetry and does not appear in its space group. The periodicity of this set of local symmetry operations indicate whether the crystal is ordered (periodic) or disordered (non-periodic). Slabs in such

crystals are referred to as OD-layers forming the crystal. Symmetry operations of the whole crystal are summarized in the space group, whereas the addition of local symmetry to the set of crystal symmetry operations describes the space groupoid (Brandt, 1927; Dornberger-Schiff, 1964).

Most important for a clear definition of layer slabs in this issue is the fulfilling of so called vicinity conditions (VC), which define geometrically equivalent sets of paired layers. Three vicinity conditions can be postulated (Dornberger-Schiff & Fichtner, 1972; Dornberger-Schiff, 1979):

1. Vicinity condition α : VC layers are either geometrically equivalent or very few in kind.
2. Vicinity condition β : Sets of all translation operations (translation group) of all VC layers are either equal or have a common subgroup.
3. Vicinity condition γ : Equivalent layers face equivalent adjacent layers by a equivalent interface, so that the resulting pair is equivalent.

The position of a layer only defined by the positions of the adjacent layer and fulfilling the VC are termed fully ordered. If more than one position is allowed fulfilling the VC conditions, the structure is termed an OD structure and the layers are OD layers.

Symmetry in OD structures is either referring to a layer itself or it builds the adjacent layer and is classified as follows (Dornberger-Schiff, 1956; Āuroviĉ, 1997):

- λ -symmetry: Is defined as the symmetry proper of an individual layer.
- σ -symmetry: Is the symmetry of a layer pair.

Symmetry operations arising from these symmetry conditions depending on the kind of symmetry operation are:

- λ -operations: Symmetry operations transforming a layer into itself, the set of λ -operations add to a diperiodic group (layer group).
- σ -operations: Coincidence operations transforming a layer into the adjacent one.

- τ -operations: Symmetry or coincidence operations which do not invert the sign of adjacent layer orientation.
- ρ -operations: Symmetry or coincidence operations which do invert the sign of the coordinate stacking direction of a layer and thus turn the adjacent layer upside down.

Choice of the OD-layers is crucial for the correct description of the symmetry in an OD-structure as the concept of OD-layer choice is not an absolute one, but helps only to describe the observed architecture of a crystal.

Classification of OD-structures

Classification of OD-structures helps to compare reasons and consequences of OD arrangement in a more distinct manner (Dornberger-Schiff, 1956).

- OD structures of the first kind are consisting of layers with planes being parallel to each other, which either form a set of translatable layers or belong to several such layers. Further classification can be made on the set of layers by:

Type A: All layers belong to a set of translatable layers.

Type B: All layers are substantially equal, but belong to two or more different interleaved sets of translatable layers. Adjacent layers can belong to different sets and are transformed by σ -operations.

- OD structures of the second kind consist of rod-like structures with one dimensional periodicity and can exhibit disorder effects as well. Further classification has not been introduced so far.

OD twinning

The presence of OD-structures can explain the occurrence of twinning, as symmetry operations that are not part of the holohedral point group of specialized metric or the symmetry of the whole crystal lattice are only valid for a subspace of the crystal. OD

twins therefore consist of periodic OD-layers and the twin operation relating twin individuals is isomorphous to λ - and σ -operations described by the OD groupoid (Nespolo *et al.*, 2004). The process leading to twinning is the ambiguity in layer stacking described by the OD theory, as boundary interactions of adjacent layers are equal in energy for two or more different orientations.

4.2 Aim of the present work

Biochemical and structural characterization of the diheme *c*-type cytochrome DHC2 from *Geobacter sulfurreducens* have been carried out, obtaining structural data indicating the presence of a conserved heme packing motif and several structural features for fine tuning of the observed mid-point potentials of -139 mV and - 289 mV. Indicators for data and model quality indicate reliable experimental and structural model data. Furthermore, the model seems to be in line with other multi-heme *c*-type cytochrome structures and no evidence can be drawn, that the current model does not represent the state of DHC2 in its crystalline form. However, R factors indicated a problem in structure refinement, as the process converged at disagreeably high R factors for data of this quality and resolution, pointing to additional features of DHC2 crystals that were not detected during the process of structure solution and refinement in the first place. Reassessing diffraction data from the point of data merging will be described in the following, looking for signs revealing the presence of twinning, using a variety of intensity statistics methods, with a final refinement using the information gained by these methods. Structural and thermodynamic reasons for the presence of pseudo-merohedral twinning in crystals of DHC2 will be discussed on the basis of non crystallographic symmetry relations and protein crystal packing behavior as described by the order-disorder (OD) terminology.

4.3 Methods

4.3.1 Data intensity statistic plots

4.3.1.1 Cumulative intensity distribution

Cumulative distribution of intensities is a method describing the distribution of weak and strong reflections observed for a diffraction data set (Rees, 1980). Plotting of Z with:

$$Z = \frac{I}{\langle I \rangle}$$

where I is the intensity of a reflection h, k, l against the number of observations $N(Z)$ of this intensity, results in a distribution that shows characteristic behavior for different pathological cases. It can be used for detection of twinning, pseudo-translation or general problems in data collection or processing, indicated by an unusual distribution of intensities. The plot discriminates between intensities of centric and acentric reflections. Centric reflections are Friedel pairs $h, k, l = -h, -k, -l$ and despite measuring anomalous signal having equal intensities $I(h, k, l) = I(-h, -k, -l)$ or reflections related by symmetry operations of the actual space group being of equal intensity as well. For both reflection types a theoretical distribution for acentric and centric reflections is given to estimate, whether the intensities in the actual case differs from these data. Two prominent pathological cases can be directly assessed using this distribution.

- Twinning, especially at high twin fractions, shows a distribution strongly differing from theoretical data, as intensity from twin related reflections are averaged, without any loss of overall intensities. The probability of a weak reflection being twin related to a weak reflection is much lower than being related to an average or strong reflection, therefore the overall differences in intensities are averaged out by twinning and the distribution of intensities is much closer to the mean intensity than in the untwinned case. One speaks of a sharpened intensity distribution. It was demonstrated recently, that choosing a reasonable resolution cutoff is essential for the detection of twinning, as the

signal to noise ratio has a severe effect on this twin law independent twinning test (Lebedev *et al.*, 2006). Also the fractional contribution of twin domains to the overall crystal lattice cannot be assessed with good estimates using this method.

- In the case of pseudo-translation, a non-crystallographic symmetry (NCS) operator is close to a unity matrix with its rotational part and one, two or all components of the translational part being close to or equal to a half axis length (Chook *et al.*, 1998). In this case extinction of reflection intensities along these axes occurs, leading to an overall decrease in intensity and to the opposite effect as observed for twinning. A set of reflection is weakened and another set of reflections remains that is unaffected by this phenomenon. A bimodal distribution of intensities is the consequence, leading to a much higher variance in intensities around the mean (Zwart *et al.*, 2008).

Taking the effects of both pathological cases on the cumulative intensity distribution into account, it already implies a drawback of this method. If both phenomena coexist in a single crystal pattern, the phenomena tend to extinguish their effects on the cumulative intensity distribution and no real evidence for the presence of any of the two phenomena can be easily drawn. Also the presence of non crystallographic symmetry parallel to the twin axis can mask the presence of twinning.

4.3.1.2 Twin law dependent twinning tests

Traditional twinning tests other than the cumulative intensity distribution are in need for a potential twin law judging whether twinning is present in the actual case. Twin law dependent twinning tests have the big advantage that they can give a relatively accurate estimate for the twin fraction α , by describing the fractional contribution of the individual twin domains to the overall crystal lattice. Two twin law dependent twinning tests are commonly used in this case.

The H-test (Yeates, 1997) compares twin related reflection intensities with:

$$H = \frac{|I_1 - I_2|}{(I_1 + I_2)}$$

where I_1 and I_2 are twin related reflections. A plot of the cumulative distribution of H with $S(H)$ being:

$$S(H) = \frac{\left[\frac{1+H}{1-2\alpha} \right]}{2}$$

therefore indicates twinning by comparing the distribution to a distribution as it would be observed for twin fractions of 0, 0.1, 0.2, 0.3 and 0.4, where 0.5 would be a vertical line. The initial linear part of the distribution can be extrapolated to estimate the twin fraction α . For untwinned data no relation of I_1 and I_2 exists and the plot shows just a diagonal line. Unfortunately this twin test can be obscured by the presence of pseudo-translation, anisotropic scattering and the presence of non crystallographic symmetry parallel to the twin axis. This leads either to estimates for a twin fraction that is lower than the real one or in case of NCS parallel to the twin axis to a large difference of the distribution from linearity, so that only a crude estimate of initial slope linearity can give a reliable estimate for the twin fraction α (Lebedev *et al.*, 2006).

The Britton plot (Britton, 1972) is differing systematically from the method used by the H-test and therefore represents an independent test, helping to judge a twin fraction α . Detwinning of reflection data is the principle of the Britton plot using:

$$I_{s1} = \frac{[(1-\alpha)I_{t1} - \alpha I_{t2}]}{1-2\alpha}$$

$$I_{s2} = \frac{[(1-\alpha)I_{t2} - \alpha I_{t1}]}{1-2\alpha}$$

with I_{s1} and I_{s2} being detwinned intensities of reflections $s1$, $s2$, and I_{t1} , I_{t2} being the intensities of the twin affected reflections $t1$ and $t2$ equivalent to $s1$ and $s2$. For estimation of the correct twin fraction, a plot of twin fraction against negative intensities is used. When the twin fraction α is too small, no negative intensities will be observed, as the intensities resulting from a subtraction of twin related intensities will be equal or larger than zero. If the twin fraction α is overestimated, more intensity will be subtracted from reflection intensities than is actually present and will result in negative intensity. The number of reflections showing a negative intensity correlates linearly with the degree of overestimation for the twin fraction α . Therefore, extrapolating the distribution curve of negative intensities gives an estimate for the present twin fraction α .

Twin fractions estimated by these methods tend to be accurate enough for structural refinement programs that cannot refine the twin fraction α together with atom positions and temperature B factors, as for example CNS (Brünger *et al.*, 1998). Indistinct signs for the presence of twinning or the exact estimate of twin fraction can arise from these methods. Choice of intensities directly based on the twin law can mislead the method by occurrence of pseudo-translation, anisotropic scattering, NCS parallel to the twin axis or a twin fraction approaching values close or equal to 0.5 (Rudolph *et al.*, 2004). In these cases the twin fraction is unreliable and twin fractions should be refined in SHELXL (Sheldrick, 2008), if the data is showing better resolution than 2.5 Å. Twin law dependent tests for a given twin law, used on data that is of real higher symmetry will always result in a twin fraction close to 0.5, as the test will mistake centric reflections as acentric, twin related reflections. As all intensities of these wrong twin-related reflections are equal, it will see them as if they were a result of perfect twinning (Zwart *et al.*, 2008).

4.3.1.3 Statistics of local intensity differences - L-test

For dealing with pathological cases that can mislead data intensity statistics in the case of the latter twinning tests, a twin law independent twinning test comparing local intensity differences was introduced (Padilla & Yeates, 2003). The L-test should provide robustness against pseudo-centering and anisotropic diffraction and can elucidate the presence of twinning in ambiguous cases. Reflections under consideration by the L-test are not twin related but are proximal in reciprocal space, assuming that two of these reflections will have a similar expected value and probability distribution. If the two reflections are chosen so that the differences between their corresponding indices are all even, then they show the same parity and intensity values will be similar for pseudo-centering operations. The distribution of $|L|$ with:

$$L = \frac{I(h_1) - I(h_2)}{I(h_1) + I(h_2)}$$

where L can take values of $-1 \leq L \leq 1$ and h_1, h_2 are reflections residing close in reciprocal space, against $N(|L|)$ therefore indicates the presence of twinning by

following the distribution for untwinned data or being equal to the twinned distribution and lying in between the twinned and untwinned distribution. Twin fractions derived from this method are much less reliable than ones obtained from the H-test or the Britton plot, but as they are independent of a potential twin law gives a good indication for the presence of twinning, not being affected by altering of intensity statistics as observed in the described, pathological cases.

4.3.1.4 Twin refinement by SHELXL

SHELXL is a refinement program intended for the refinement of small molecules, but can be also applied to macromolecular structures with a resolution of about maximally 2.5 Å (Sheldrick, 2008). Instead of using a fast fourier transformation approach, SHELXL uses conventional structure factor summation, which although being slower is more accurate. SHELXL refines against F^2 rather than F 's as they tend to give more experimental information and help to prevent the refinement processes from getting arrested in local minima. As data of weak reflection intensities are pronounced in pseudo-symmetry, a stronger contribution from these reflections is desired in order to find a correctly converging solution. The method of least squares refinement is implemented for SHELXL and is representing an iterative process where model parameters are shifted to minimize the target function of observed minus calculated data given in a very simple way by:

$$Q = \sum_{hkl} w(hkl)(|F_{obs}(hkl)| - |F_{calc}(hkl)|)$$

where w is the weight given to an observation.

Twin refinement in SHELXL is implemented using the method of Pratt, Coyle & Ibers (1971) and Schneider, Dubler & Oswald (1982) (Pratt *et al.*, 1971; Jameson *et al.*, 1982). F_c^2 values are calculated by:

$$(F_c^2) = osf^2 \sum_{m=1}^n k_m F_{c_m}^2$$

where osf is the overall scale factor, k_m is the fractional contribution of twin domain m and F_{cm} is the calculated structure factor of twin domain m . Fractional contributions of k_m from all twin domains has to be one and therefore the twin fraction k_1 can be refined and is calculated by:

$$k_1 = 1 - \sum_{m=2}^n k_m$$

Refinement of merohedral and pseudo-merohedral twins is straightforward in SHELXL. As the lattices coincide nearly perfectly, refinement on a standard data set can be carried out. In the case of reticular merohedry and non-merohedral twinning, special software for indexing and integration has to be used, as lattices of different twin domains have to be indexed and integrated separately. Twin refinement in SHELXL can be carried out by giving a twin law in a fractional matrix using the keyword TWIN and an initial estimate for the twin fraction α using the keyword BASF. However, twin refinements tend to be less stable and have a lower data to parameter ratio than normal refinement methods. Also automatic model and water building programs should be avoided as they tend to introduce model bias into the refinement process. Whether a crystal is twinned can be perceived by a significant drop in R_{cryst} and R_{free} , depending on the real twin fraction α . Perfect twinning, not handled by other twin refinement methods can be refined using SHELXL. On the other hand great care has to be taken selecting reflections for the validation data set, as twin related reflections being in the working and validation set, will introduce significant bias to the R_{free} , as the model in effect is also refined against intensities from the validation set.

4.3 Results

4.3.1 Data indexing and merging statistics

Using DENZO and SCALEPACK (Otwinowski & Minor, 1997), the data could be indexed and merged in space groups $P2_1$ and $C222_1$, whereby space group $P2_1$ was preferred over $C222_1$ by the autoindexing software. Unit cell parameters of the $P2_1$ space group setup were $a = 39.65 \text{ \AA}$, $b = 55.68 \text{ \AA}$, $c = 39.64 \text{ \AA}$ and $\beta = 105.91^\circ$. Cell parameters for space group $C222_1$ were determined to be $a = 47.76 \text{ \AA}$, $b = 63.27 \text{ \AA}$ and $c = 55.67 \text{ \AA}$. With unit cell parameters a and c being almost equal, construction of a $C222_1$ cell being a minimal non-isomorphic supergroup of $P2_1$ was possible using the conditions $a' = 2a \cos(\beta/2)$ and $b' = 2a \sin(\beta/2)$ and $c' = b$ (Figure 31). With $168,962 \text{ \AA}^3$ the orthorhombic cell has exactly the double volume of the monoclinic cell and contains one monomer in the asymmetric unit (regarding a Matthews coefficient of $1.86 \text{ \AA}^3/\text{Da}$).

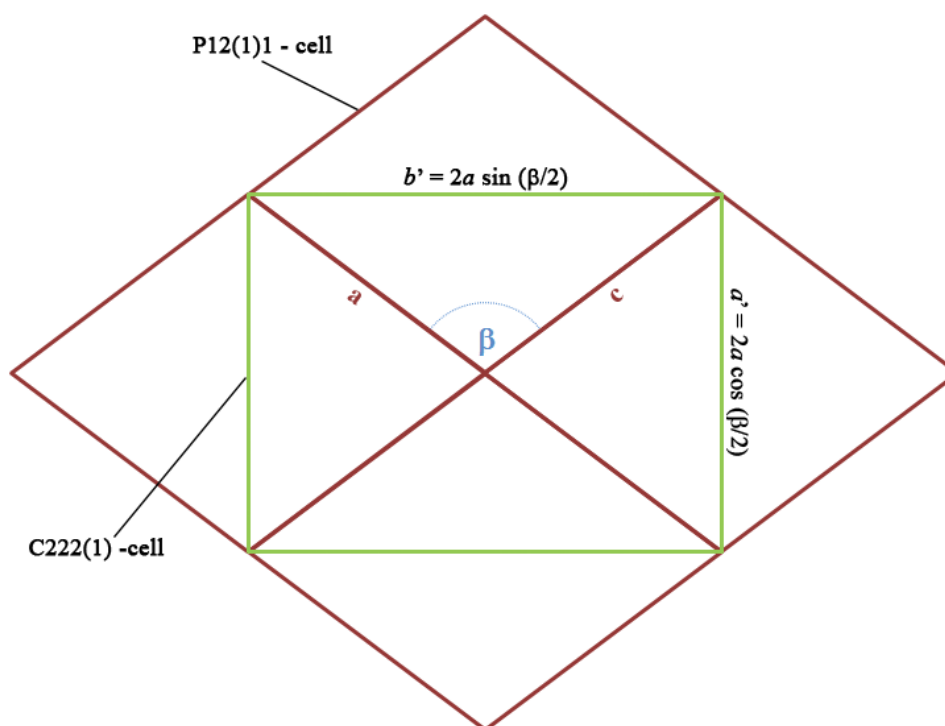


Figure 31: Construction of an orthorhombic unit cell from a monoclinic cell with unit cell parameters a and c being equal, building half-diagonals in the new cell setup. Unit cell parameters are related by $a' = 2a \cos(\beta/2)$ and $b' = 2a \sin(\beta/2)$ and $c' = b$.

Space group choice in this case can be justified by merging statistics (see 3.4.6 and Table 9). If $C222_1$ represents the correct space group choice, then merging R factors should be very similar to merging R factors in $P2_1$. Merging of reflection intensities in a higher symmetry space group leads necessarily to bias in the linear merging R factor R_{merge} , as the redundancy of data is doubled in the case of $C222_1$ in contrast to $P2_1$. As the linear merging R factor R_{merge} does not account for higher redundancy, data merging statistics of different space group redundancies are not comparable, therefore the redundancy-independent merging R factor $R_{\text{r.i.m.}}$ was used to compare both space group setups. Inspection of unmerged data from SCALEPACK (Otwinowski & Minor, 1997) with the program RMERGE (Weiss & Hilgenfeld, 1999; Weiss, 2001) revealed a redundancy-independent merging factor $R_{\text{r.i.m.}}$ about 0.18 for space group $C222_1$ in contrast to 0.06 for $P2_1$, clearly showing that intensities merged for orthorhombic symmetry were deviating much stronger than in the case of the monoclinic space group setup. This clearly supports that data in space group $P2_1$ was misinterpreted as a pseudo-orthorhombic setup by fortuitous unit cell parameters. Also the merging R factor R_{merge} of 0.17 in the case of $C222_1$ and 0.05 in the case of $P2_1$ or the precision indicating R factor $R_{\text{p.i.m.}}$ of 0.05 for $C222_1$ and 0.02 for $P2_1$ support this assumption. Nonetheless, merging R factors for $C222_1$ are much lower than would be expected for a totally wrong space group choice. Two more or less interconnected phenomena could be the reason for this behavior. On one hand, the presence of twinning with a significant twin fraction α could average intensities using a twin law that is a symmetry operation of a higher crystal system. Therefore merging in this space group will go well for reflections that in this case are twin related, showing averaged and identical intensities for perfect twinning. On the other hand, if non-crystallographic symmetry is closely missing real crystallographic symmetry, intensities are also strongly related.

Data collection		
Wavelength (Å)	0.994	
Resolution range (Å)	50.0-1.5 (1.52-1.50)	
Merging		
Space group	C222 ₁	P2 ₁
R _{merge} †	0.17 (0.59)	0.05 (0.56)
R _{r.i.m.} ‡	0.18 (0.64)	0.06 (0.62)
R _{p.i.m.} §	0.05 (0.24)	0.02 (0.31)

Table 9: Data collection and merging statistics for space group setups P2₁ and C222₁. Values in parentheses are given for the highest resolution shell. † R_{merge} is the merging R-factor. ‡ R_{r.i.m.} is the redundancy independent merging R-factor. § R_{p.i.m.} is the precision-indicating merging R-factor (Weiss & Hilgenfeld, 1997; Weiss, 2001). Data was collected at beamline BW7A at the EMBL outstation at DESY (Hamburg).

4.3.2 Cumulative intensity distribution

Calculation of the cumulative distribution of intensities (Rees, 1980) with Z against $N(Z)$ is implemented as a standard procedure in TRUNCATE (French & Wilson, 1978) which is part of the CCP4-suite (Collaborative Computational Project, Number 4, 1994) and widely used for conversion of intensities to structure factor amplitudes. The distribution calculated for DHC2, using data to a resolution of 2.0 Å, reveals a clear deviation of acentric reflection intensities from the theoretical distribution, indicated by shifting of the observed distribution to a distribution expected for perfect twinning (Figure 32). Unfortunately these indicators become much less pronounced at higher resolution, because accuracy of intensities is a very crucial criterion and in principle data should be rejected for this kind of tests when the completeness for reflections with $I/\sigma > 3$ falls below 85 %. On the other hand, non-crystallographic symmetry parallel to a potential twin axis can shift the distribution in the direction of (theoretically) untwinned data (Lebedev *et al.*, 2006; Zwart *et al.* 2008). This effect is diminished at higher resolution, but resolution of the data in this case was not high enough. For DHC2 twinning was not supposed to be a serious problem in the first place, as the intensity distribution calculated with data to 1.5 Å shows much stronger scattering around the mean intensity than for higher resolution, as weaker intensities tend to deviate more around a mean, than strong data does. But at 2.0 Å resolution data cutoff a clear indication for the presence of twinning can be drawn.

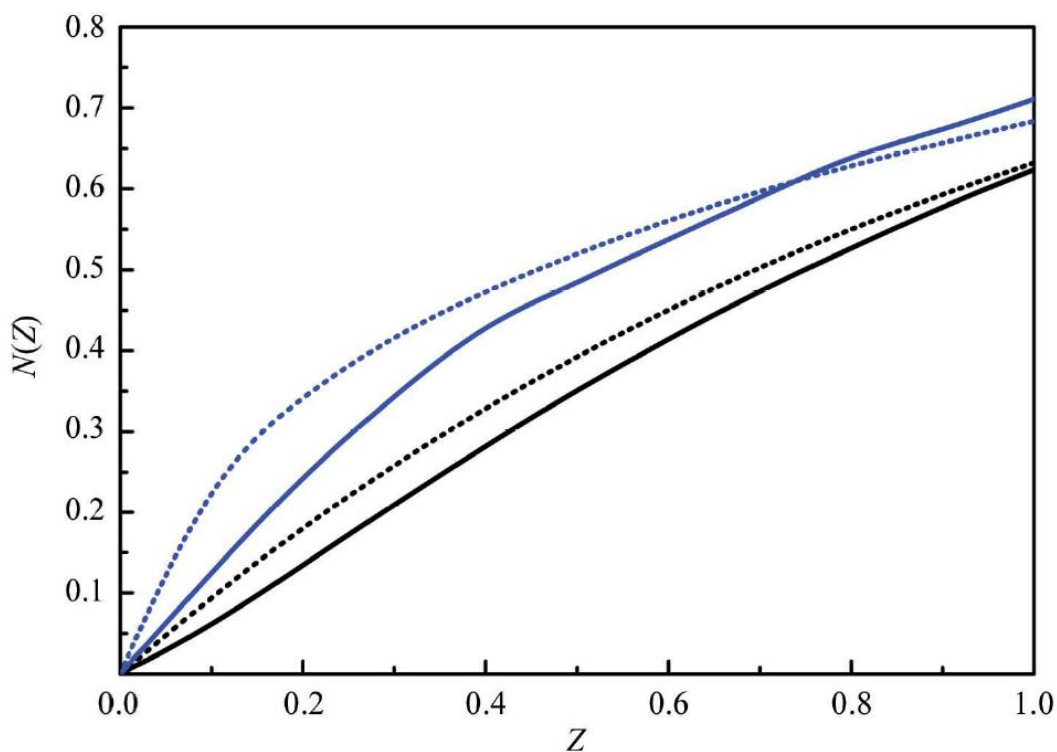


Figure 32: Cumulative distribution of $Z = I / \langle I \rangle$ for observed acentric reflections (black line) and centric reflections (blue line) of DHC2, calculated with data to 2.0 Å resolution. The theoretical distribution of untwinned data for acentric and centric reflections are shown as black and blue dotted lines, respectively. The deviation from the theoretical intensity distribution, which is shifted to the right as expected for perfectly twinned data.

4.3.3 Data intensity statistics

There are several data intensity statistics that indicate the presence of twinning by showing significant deviation from theoretical values for untwinned data. These intensity statistics are not applied as a standard technique during structure solution, but could give a hint whether the possibility of twinning can be considered. First, the distribution $|E^2 - 1|$ for all acentric reflections was calculated using the program *phenix.xtriage* (Adams *et al.*, 2002). With a value of 0.678 DHC2 data clearly differs from theoretical, untwinned data with 0.736 for acentric reflections, but lies above a value of 0.541 which would indicate the presence of perfect twinning, thus clearly exhibiting the presence of twinning with a twin fraction differing substantially from non-twinned or perfectly twinned data (Herbst-Irmer & Sheldrick, 1998). The second moments of normalized intensity distribution,

$$\frac{\langle I^2 \rangle}{\langle I \rangle^2}$$

and the analogous ratio of amplitudes,

$$\frac{\langle F^2 \rangle}{\langle F \rangle^2}$$

were examined by using CNS (Brünger *et al.*, 1998). For untwinned data the ratio of intensities should be 2.0, decreasing to 1.5 for perfectly twinned data, with 0.785 and 0.865 for the ratio of amplitudes respectively. The observed ratio of intensities for DHC2 was 1.84 and 0.82 for amplitudes (Table 10). Both values again reside between the values for theoretical untwinned and perfectly twinned data, again suggesting the presence of twinning with a moderate twin fraction α of $0 \leq \alpha \leq 0.5$.

	Observed	Untwinned	Perfect twin
Wilson ratio and moments			
$\langle I^2 \rangle / \langle I \rangle^2$	1.84	2	1.5
$\langle F^2 \rangle / \langle F \rangle^2$	0.82	0.785	0.885
$ E^2 - 1 $	0.678	0.736	0.541
Yeates [†] statistics (H-test)			
$\langle H \rangle$	0.169	0.5	0
$\langle H^2 \rangle$	0.045	0.33	0
Padilla & Yeates [‡] statistics (L-test)			
$\langle L \rangle$	0.449	0.5	0.375
$\langle L^2 \rangle$	0.276	0.333	0.2

Table 10: Statistics of acentric data intensities for DHC2. [†] According to Yeates (1997). [‡] According to Padilla & Yeates (2003).

4.3.4 Detection of twinning using the L-test

The L-test proved to be robust against anisotropic data and pseudo-centering phenomena in the process of twin detection (Padilla & Yeates, 2003). Distribution of local intensities $|L|$ against $N(|L|)$ clearly indicates the presence of twinning, as differing from untwinned data, but not reaching perfectly twinned data. Therefore a moderate

twin fraction between 0 and 0.5 seems to be most probable (Table 10). Twin fractions estimated by this method seem to be unreliable and were therefore not included in further considerations.

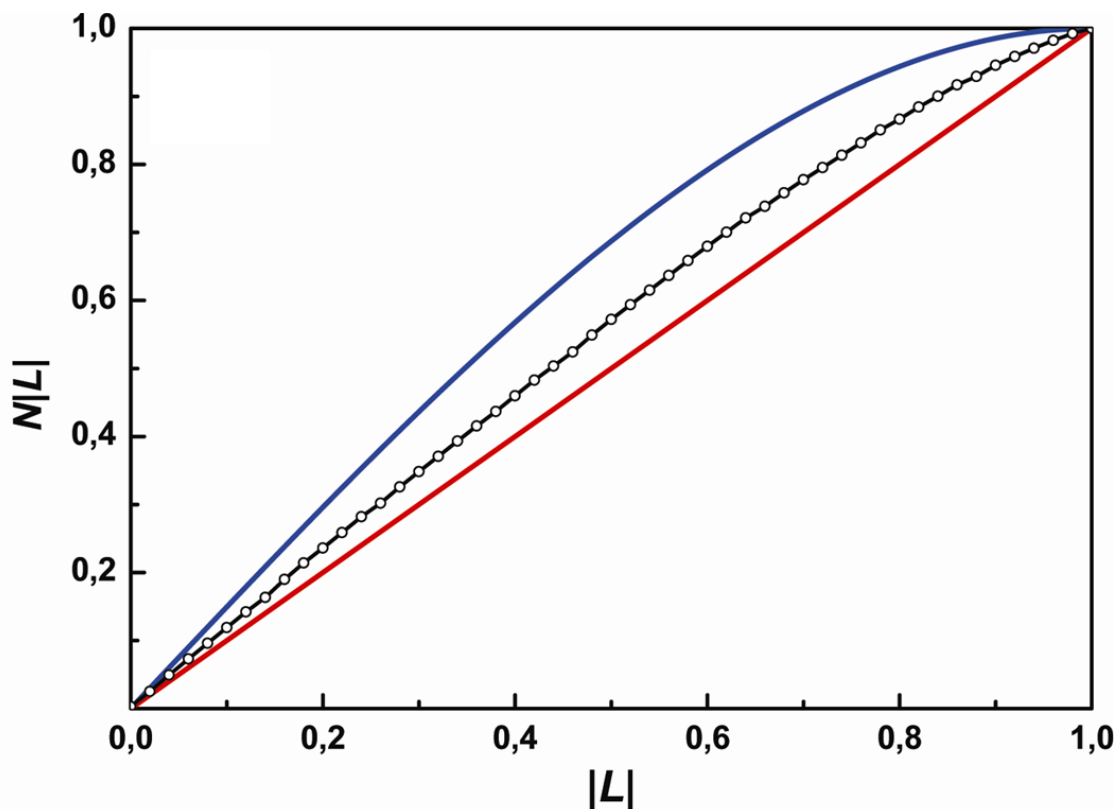


Figure 33: The distribution of local statistics $L = I_1 - I_2 / (I_1 + I_2)$ for DHC2 data, where I_1 and I_2 represent the intensities of unrelated reflections residing close in reciprocal space, is a test for twinning (L-test) (Padilla & Yeates, 2003, Heitmann & Einsle, 2008).

4.3.5 Twin law dependent twinning tests

Intensity statistics are the only possibility to recognize signs for pseudo-merohedral twinning, as this kind of twinning results in nearly perfect overlap of lattice points and renders detection from observation of diffraction patterns nearly impossible. While a variety of twin law-independent intensity statistics were able to recognize the general presence of twinning in crystals of DHC2, only twin law-dependent intensity statistic twinning tests are able to reveal the correct twin operation and a reliable estimate for the twin fraction α . Therefore the H-test and the Britton plot were used to elucidate the correct twin law and twin fraction. With unit cell parameters a being equal to c , only

one potential twin law remains from the so far discussed possibilities, where a rotation around the diagonal spanned between a and c relates reflections by $l, -k, h$.

4.3.5.1 H-test

Data of the H-test (Yeates, 1997) (Table 10 and Figure 34), which is implemented in the program DETWIN (Collaborative Computational Project, Number 4, 1994), clearly shows the presence of twinning using the predicted twin law $l, -k, h$. The initial slope can be extrapolated to a potential twin fraction of about 0.33. In contrast to theoretical distributions, H in the case of DHC2 shows a very pronounced curvature that could arise from the presence of non crystallographic symmetry parallel to the potential twin axis. The clear deviation from a distribution equal or near to a twin fraction of 0.5, excludes the possibility of real higher symmetry in $C222_1$ as well.

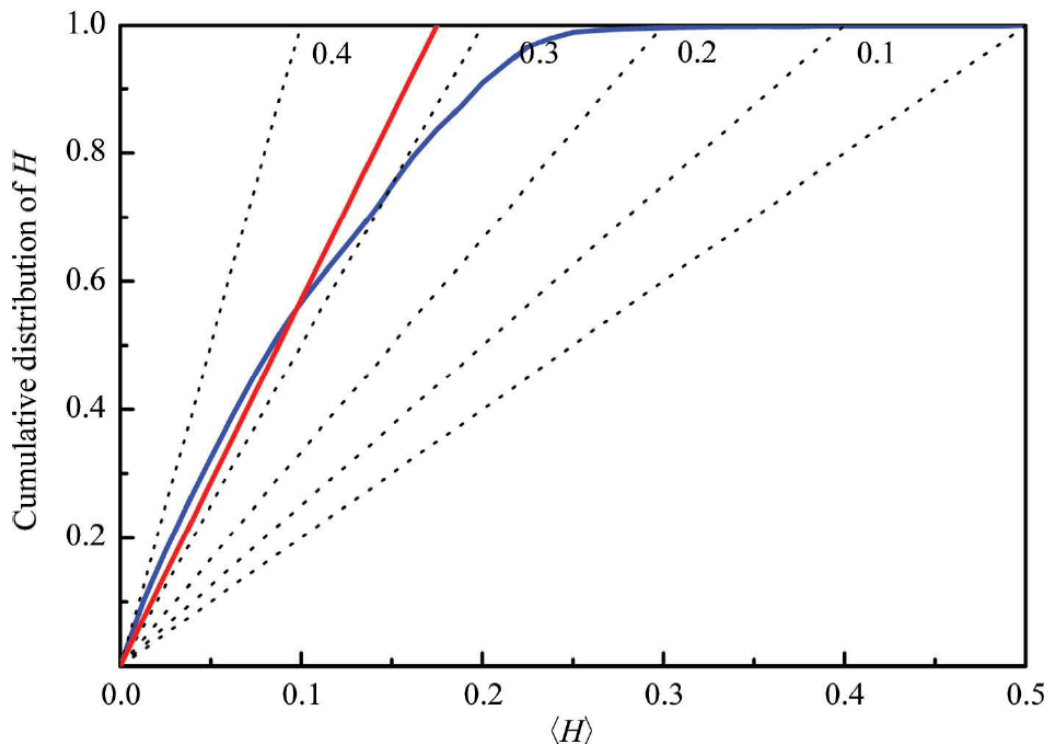


Figure 34: Plot of the cumulative distribution of H for acentric reflections, with $H = |I_1 - I_2| / (I_1 + I_2)$ and intensities of reflections related by the twin law $l, -k, h$ (H-test) (Yeates, 1997). Comparison of the observed distribution (blue line, with linear regression in red) with theoretical distributions of H , with a twin fraction increment of 0.1, yields a twin fraction of $a \approx 0.33$ (Heitmann & Einsle, 2008).

4.3.5.2 Britton plot

With the H-test showing a clear indication for twinning and a relation of the twin domains by the twin law $l, -k, h$, the Britton plot (Britton, 1972) was applied to data of DHC2 to receive a method-independent confirmation of twin law and twin fraction. Extrapolation from a plot of the twin fraction α against negative intensities results in a twin fraction of 0.31, a value very similar to the twin fraction obtained by the H-test and confirming the presence of pseudo-merohedral twinning (Figure 35).

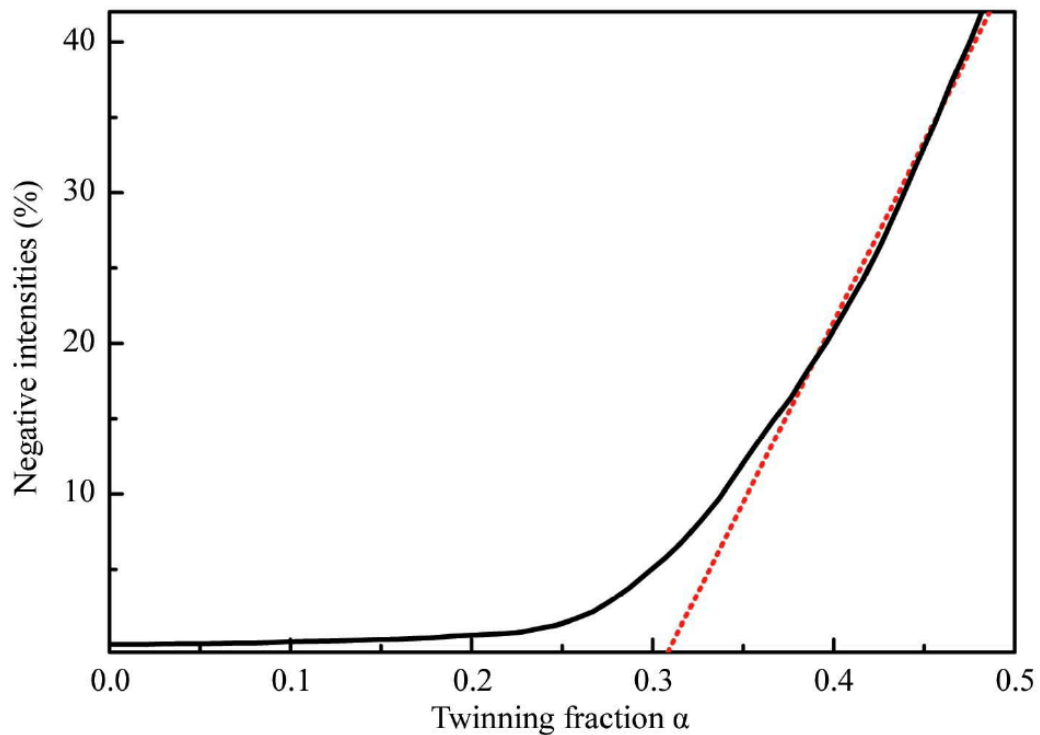


Figure 35: Twin-fraction estimation by the Britton plot. Extrapolation of negative intensities against the potential twin fraction α (black line) yields a twin fraction of $\alpha \approx 0.31$ (red) (Heitmann & Einsle, 2008).

4.3.6 Self rotation function

Further investigation of DHC2 was carried out in order to identify elements of pseudo-symmetry that provide the structural basis for twinning in the case of DHC2. A self rotation function for the data of DHC2 was calculated using the program GLRF (Tong & Rossmann, 1990) for $\kappa = 180^\circ$ and revealed maxima orthogonal to \mathbf{b} with $\varphi = 0^\circ$, $\psi = -53^\circ$ and $\varphi = 0^\circ$, $\psi = 37^\circ$ sharing a relative peak height of 77.4 % to the origin peak

(Figure 36). These off origin peaks represent rotational non crystallographic symmetry axes parallel to the **ac** plane and perpendicular to each other, emulating a higher crystal system that clearly provides a structural basis for the presence of twinning as it emulates orthorhombic symmetry.

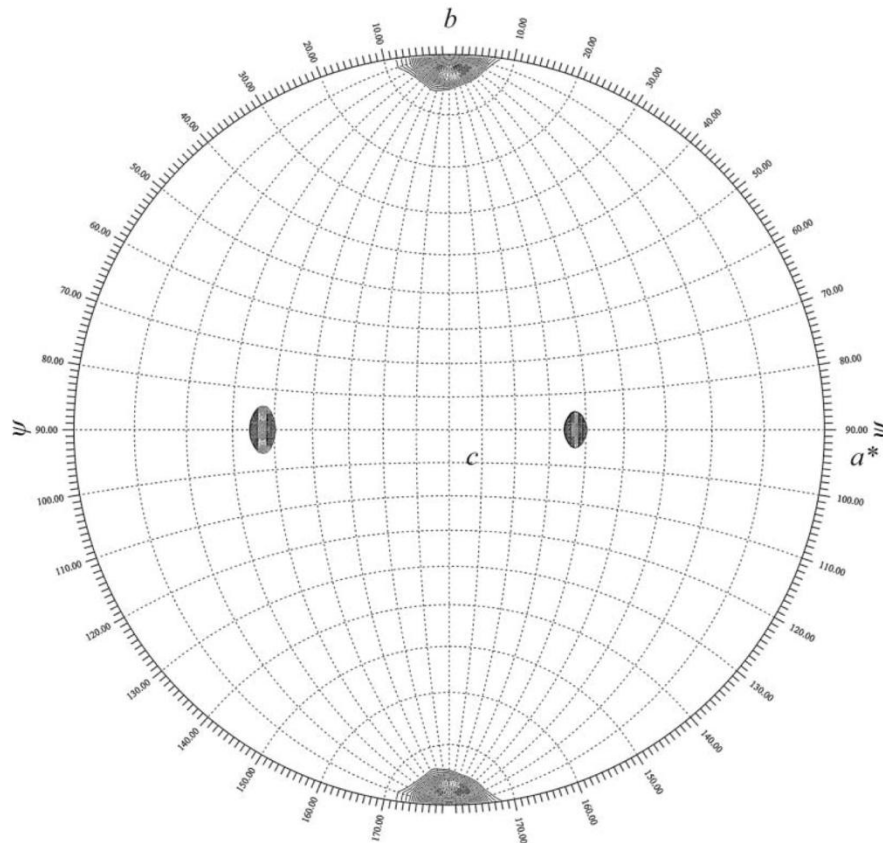


Figure 36: Self-rotation function depicting the non-crystallographic symmetry perpendicular to **b** in crystals of DHC2. The plot was generated with $\kappa = 180^\circ$ and reveals two prominent peaks perpendicular to **b**, with $\varphi = 0^\circ$, $\psi = -53^\circ$ and $\varphi = 0^\circ$, $\psi = 37^\circ$ at a relative peak height of 77.4 % of the origin peak. The peaks are related to each other by a 90° rotation around **b** (Heitmann & Einsle, 2008).

4.3.7 Twin refinement using SHELXL

In summary intensity statistics, twinning tests and the self-rotation function clearly indicate the presence of pseudo-merohedral twinning in crystals of DHC2. A moderate twin fraction of about 0.32 relates the two individual twin domains by a twin law of $h, -k, l$. This operation is a rotation around the diagonal of the **ac** plane (101) and originates from pseudo-orthorhombic symmetry elements in the asymmetric unit of DHC2.

Confirmation of these assumptions can be made with SHELXL, the only refinement program able to refine atom positions, B factors and a twin fraction at the same time (Sheldrick, 2008). As choice of reflections for the validation set is crucial for preventing bias on R_{free} , test set reflections were chosen by putting reflections related by the twin law $l, -k, h$ into the same set using CNS (Brünger *et al.*, 1998), such that the simultaneous appearance of twin related reflections in working and test set is excluded. Furthermore, the initial model for structure solution was chosen on the basis of bringing the least possible model bias into the refinement process. An initial model was used that was the first model built from experimental electron density maps obtained with SHELXE (Sheldrick, 2002) and was never refined against a data set. After an initial round of rigid body refinement in space group $P2_1$ with two monomers in the asymmetric unit, further residues were fitted into the electron density map and waters included to a restrained refinement process, not accounting for the presence of twinning. After several iterative rounds of model building and refinement the R factors converged to values of R_{cryst} of 0.223 and an R_{free} of 0.278 being in good accordance to refinement statistics obtained for the structure refinement process in REFMAC 5 (Murshudov *et al.*, 1997). Twin refinement was introduced by setting the keywords TWIN 0 0 1 0 -1 0 1 0 0, with the twin law $h, -k, l$ given in a fractional rotation matrix, and an initial twin fraction of 0.1, given by the keyword BASF. The choice of initial twin fraction was made to prevent bias by already giving the twin fraction obtained by the H-test and Britton plot. The refined twin fraction then would be an independent cross validation for the twin fractions obtained by the intensity twinning tests. Twin refinement led to a drop in R factors about 4.5 % to an R_{cryst} of 0.176 and an R_{free} of 0.227 (Table 11). The twin fraction α was refined to a value of 0.322 corresponding very well to values estimated by the H-test and the Britton plot. The model integrity of twin-refined DHC2 structure was assured by PROCHECK (Lakowski *et al.*, 1993). The Ramachandran plot shows no difference for refinement of structural model data not considering and considering the presence of twinning. Quality indicators for bond lengths and bond angles were getting worse by an r.m.s.d (root mean square deviation) of 0.001 Å and 0.1° respectively, when compared to refinement without the twin law. Atom positions of the protein backbone show an r.m.s.d. of 0.15 Å. This already indicates that the original structural model was correct, although twinning was not taken into account. This observation may be dependent on the degree of twinning, as the twin domain with a fractional

contribution of about 0.668 was still sufficient for structure solution and model refinement.

Refinement statistics		
Twinning operation ¶	$l, -k, h$	
Resolution range (Å)	50.0-1.5	
Refined twin fraction	0.322	
	Untwinned	Twinned
R_{cryst} †	0.223	0.176
R_{free} †	0.278	0.227
R.m.s.d. from ideality ‡		
Bond lengths (Å)	0.008	0.009
Bond angles (°)	2.2	2.3
Ramachandran plot ‡		
Core (%)	89.2	89.2
Allowed (%)	10.8	10.8
Disallowed (%)	0	0

Table 11: Twin refinement statistics of DHC2 (Heitmann & Einsle, 2008). ¶ The twin operator corresponds to the twin law relating individual twin domains by a fractional ratio given by the refined twin fraction (SHELXL (Sheldrick, 2008)). † The refinement R factor $R_{\text{cryst}} = \sum_{hkl} |F_{\text{obs}}(hkl)| - k|F_{\text{calc}}| / |F_{\text{obs}}(hkl)|$; R_{free} is calculated for a test set including 5% of the overall reflections and selected by CNS. ‡ The r.m.s.d. from ideality and Ramachandran plot were calculated using PROCHECK (Laskowski *et al.*, 1993).

4.4 Discussion

4.4.1 Analysis of twinning

The presence of pseudo-merohedral twinning in crystals of DHC2 was confirmed by analyzing intensity statistics and performing a twin refinement using the twin law $h, -k, l$ in SHELXL. Independently refined twin fraction and intensity statistics-based estimations for the twin fraction α are in good agreement. Pseudo-symmetry, the basis for the presence of twinning, was revealed during the process of merging and by calculation of a self rotation function. The presence of two peaks in the self rotation function at the described position indicates a non crystallographic symmetry rotation axis perpendicular to \mathbf{b} and automatically generates a second rotation axis perpendicular to the first one and \mathbf{b} due to the twofold crystallographic symmetry. A setup of three twofold rotation axes with angles of 90° between them is indicative for orthorhombic crystallographic point-group symmetry. Closer investigation of the non crystallographic symmetry operation, by superimposition of monomers in the asymmetric unit of the DHC2 structure using COOT (Emsley & Cowtan, 2004; Krissinel & Henrick, 2004) yields the exact operator:

$$R = \begin{pmatrix} -0.220 & 0.006 & -0.994 \\ 0.014 & -1.000 & -0.008 \\ -1.006 & -0.027 & 0.022 \end{pmatrix}, T = \begin{pmatrix} 0.46 \\ -0.03 \\ 0.54 \end{pmatrix}$$

with R being the rotational matrix of the symmetry operation and T being the translational vector. The NCS symmetry operator is closely missing real crystallographic symmetry as a combination of symmetry operator of $P2_1$ with $(-x, y+\frac{1}{2}, -z)$ and the idealized NCS operator $(-z+\frac{1}{2}, -y, -x+\frac{1}{2})$ would result, together with the orthogonal operations $(z+\frac{1}{2}, -y+\frac{1}{2}, x+\frac{1}{2})$, in a $C222_1$ space group setup with \mathbf{b} and \mathbf{c} exchanged due to space group nomenclature conventions. An estimate for the difference of real crystallographic symmetry and pseudo-symmetry can be given using the program PDBSET (Collaborative Computational Project, Number 4, 1994) for transformation of a DHC2 monomer using ideal symmetry operators on an equivalent, NCS-translated monomer and comparing the r.m.s.d. of atomic positions to the real,

non-crystallographic symmetry arrangement. For this case an r.m.s.d. of 3.41 Å makes the difference to real crystallographic symmetry perceivable.

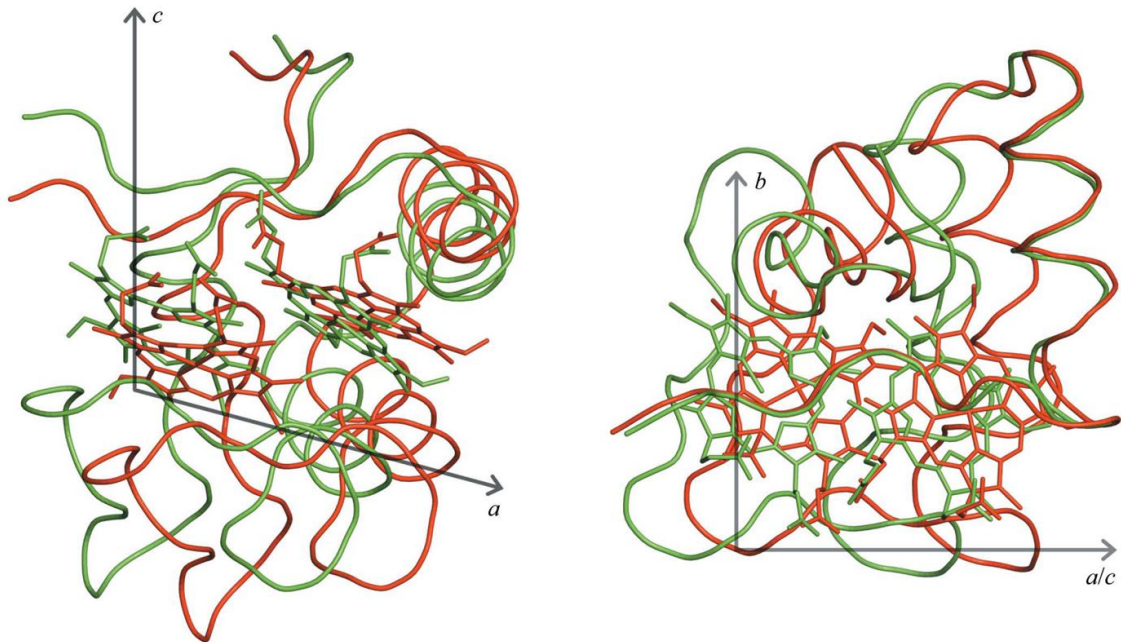


Figure 37: Superposition of $P2_1$ DHC2 monomer (red) in ribbon representation and the corresponding $C222_1$ symmetry monomer, created by combination of the idealized NCS operation $(-z+\frac{1}{2}, -y, -x+\frac{1}{2})$ with the second monomer in the asymmetric unit. The r.m.s.d. of 3.14 Å is a direct indication for the discrepancy of apparent pseudo-symmetry from higher symmetry (Heitmann & Einsle, 2008).

4.4.2 Consequence of non crystallographic symmetry for twinning

The presence of twinning by pseudo-merohedry in the case of DHC2 can directly be related to the occurrence of pseudo-symmetry. A primitive monoclinic unit cell with axes **a** and **c** almost equal can emulate a pseudo C-centered orthorhombic cell with unit cell parameter relations of $a' = 2a \cos(\beta/2)$ and $b' = 2a \sin(\beta/2)$ and $c' = b$ (Figure 38a). This fact alone is not sufficient to promote twinning by pseudo-merohedry, as relations of monomers in the asymmetric unit can provide relative distinct crystal growth directions. Therefore, pseudo-symmetry emulating an orthorhombic crystal setup is the crucial prerequisite for the occurrence of pseudo-merohedral twinning, as now proteins in the **ac** layer show almost identical orientation when pointing in one or the opposite direction (Figure 38b). For crystals of DHC2 the packing of molecules in the $P2_1$ cell necessitates an origin shift of $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ in order to make the NCS axes coincide with the crystallographic axes of the $C222_1$ cell setup (Figure 38c). This setup demonstrates the

divergence of NCS operations in DHC2 and those of the higher symmetry space group that allows for twinning using the symmetry elements of the $C222_1$ setup as twin operators with a twin law of $l, -k, h$.

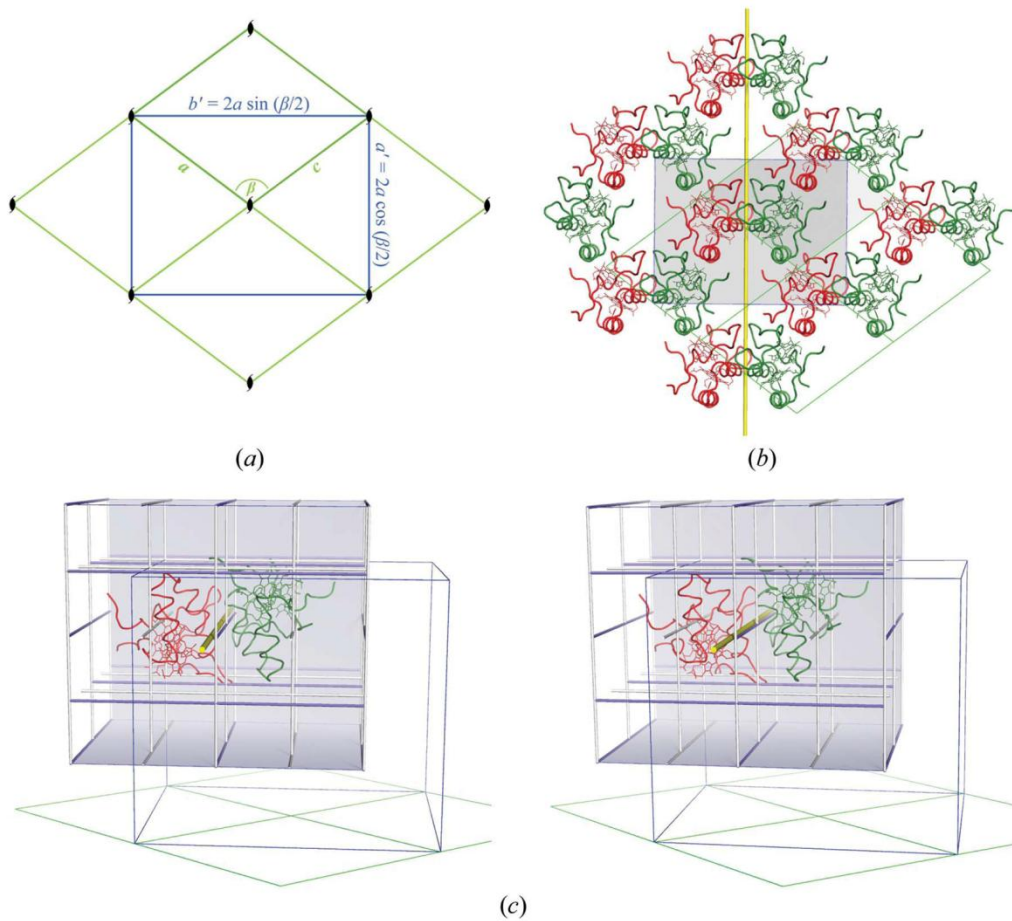


Figure 38: Noncrystallographic symmetry in DHC2 crystals. (a) A primitive monoclinic lattice with a almost equal to c allows for a C-centered orthorhombic setting with dimensions of $a' = 2a \cos(\beta/2)$ and $b' = 2a \sin(\beta/2)$ and $c' = b$. The diagram shows a view of the ac (010) plane with the (unchanging) basis vector b pointing towards the viewer. (b) Crystal packing in a single layer of DHC2, showing the two molecules in the asymmetric unit of the $P2_1$ cell in red and green, respectively. The noncrystallographic symmetry axis connecting two monomers (yellow) almost satisfying a C-centered orthorhombic setup, albeit with an origin shift of $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. (c) Stereo representation of the $C222_1$ unit cell including ribbon representation of the DHC2 crystallographic dimer from space group $P2_1$ closely missing higher symmetry. Twofold rotation axes are shown in blue and twofold screw axes in white. The noncrystallographic twofold of the $P2_1$ cell (yellow) closely misses the axis of the higher symmetry cell (Heitmann & Einsle, 2008).

4.4.3 Consequences of NCS for intensity statistics

Although DHC2 has been characterized as being twinned by pseudo-merohedry, the cumulative intensity distribution and the H-test showed significant differences from other pathological examples of twinning. The reason for this behavior is the presence of NCS almost parallel to the twin axis setup, as showed above (Zwart *et al.*, 2008). Therefore the sigmoidal shape and shift of the cumulative intensity distribution of acentric reflection intensities was not well pronounced. The H-test on the other hand was affected by revealing a strongly curved shape, what normally would be linear. Therefore only the initial slope was used for extrapolating the twin fraction α . At higher resolutions this effect should diminish, but the weaker signal intensity can disturb this test even stronger, therefore no improvements are visible when decreasing the resolution data cut-off. This effect therefore counteracts the improvement of separating effects of twinning and NCS (Lebedev *et al.*, 2006).

4.4.4 Implications for crystal packing

Two possible mechanisms for the occurrence of twins can be considered. As twins have symmetry elements of higher crystal systems (pseudo-merohedry) or the same crystal system but a different point group (merohedry), one could consider that these crystals were in fact of different symmetry before they were transformed. This could occur through mechanisms such as temperature change, manual manipulating, hydration, dehydration, flash-freezing, stream freezing, annealing and so forth (Parsons, 2003; Lebedev *et al.*, 2006). One could consider that higher crystallographic order in these cases was destroyed and that symmetry elements of the former higher order, now act as twin elements in the transformed crystal form. The second mechanism of twin formation is introduced by crystal growth itself, either by intergrowth of crystals, resulting for example in non-merohedral twinning, if the mutual growth directions are totally random. Or by ordered growth where a dominant crystal interface is not present and a more or less sequential growth of crystal layers can occur, with the relatively invariant crystal packing interfaces being a twin element around which a twin operation is possible. The latter cases are called statistical twins where the packing behavior in two dimensions is relatively good, because strong interactions were formed, but crystal contacts in the direction of the plane normal are relatively weak (Bragg & Howells,

1954; Cochran & Howells, 1954). If these layer or layer pairs contain an internal twofold symmetry axis close to, or perpendicular to the plane normal then plane surfaces on both sides are equal and therefore can form indifferent stacks, where more than one interaction can support crystal growth. Occurrence of twinning and the mutual ratio of twin domains in these cases is then more a question of kinetics than an ordered and directed growth. Decrease of symmetry in DHC2 crystals from higher symmetry to the observed monoclinic space group setup is a very uncertain explanation, as no DHC2 crystal was ever measured being of space group $C222_1$. In fact features of indifferent growth behavior can be observed for DHC2 crystals. They tend to be plate-like and thin, already indicating a three-dimensional growth defect in one certain direction. The macromolecular crystal appearance could therefore be explained by crystal contacts derived from structural data, as crystal contacts building an **ac** (101) layer are very strong and a tight packing can be observed. On the other hand, crystal packing in the direction of **b**, which is pointing in the direction of the plane normal, shows significantly less crystal contacts, only formed by poorly ordered sidechains, and could be responsible for poor growth in one dimension and the plate like crystal shape. This finding is also supported by the relative distribution of the protein contact surface areas. Each monomer has direct contact to three different monomers within the **ac** layer comprising surface contact areas of 930.9, 1008.6 and 1121.5 Å², but in the direction of **b** only one crystal contact with a surface area of 561.8 Å² is formed (Figure 39). Adding this poorly ordered arrangement of loop sidechains and surfaces to the fact of non crystallographic symmetry, then both sides of the layer are relative similar and flexible in terms of contacts. This would allow packing layers in more than one possible mutual orientation and would promote the existence of twinning in the case of DHC2

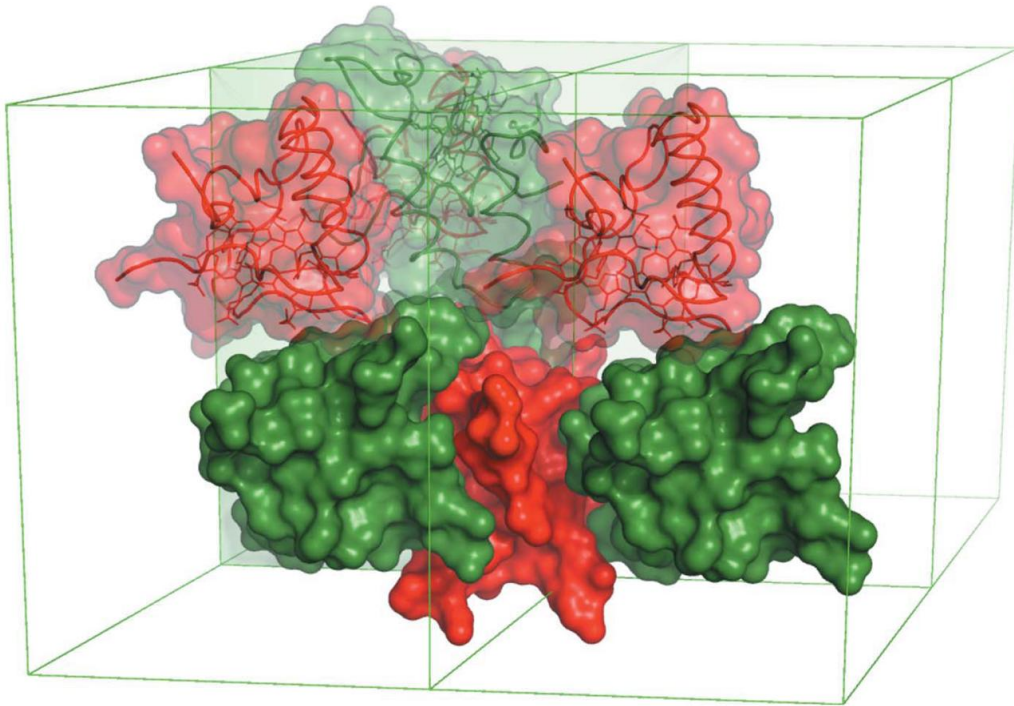


Figure 39: Crystal packing in DHC2. Crystal-packing interactions are significantly stronger in layers perpendicular to the *b* axis of the P21 cell than between these layers. This is reflected in an approximately doubled area of the interacting protein surfaces within the layers. All protein interactions between these layers are mediated by the same parts of the protein, allowing for a rotation of the entire layer to create twinned domains (Heitmann & Einsle, 2008).

4.4.5 Analysis of DHC2 crystal packing using the OD-terminology

Crystals in the OD terminology are composed of a distinct set of parallel layers forming the overall lattice by subsequent translation operations. If only one set of symmetry operations is present forming the adjacent layer from the present layer, the crystal is fully ordered. If there is more than one operation possible and not valid for the whole crystal, but for a crystal subspace instead, then an OD structure is present (Dornberger-Schiff, 1956). DHC2 seems to be arranged of layers in a comparable manner. Proteins forming strong crystal contacts in the *ac* plane can be seen as layers building the crystal by subsequent symmetry operations in the direction of *b*. In the case of twinning two possible situations can arise for forming the adjacent layer, with respect to the origin layer. In the first situation, a symmetry operation is applied that is part of the crystallographic symmetry and thus forms the adjacent layer. In the second situation the layer is rotated by the twin law $l, -k, h$ and thus represents a symmetry operation that is not valid for the whole crystal, but for a subset of the crystal space. Therefore the

definition of an OD structure is fulfilled as there is more than one possible arrangement of layers fulfilling the vicinity conditions (Dornberger-Schiff & Fichtner, 1972; Dornberger-Schiff 1964). As only one twin law is present and no other indifferent behavior for layer arrangement could be observed, the subset of symmetry operations is periodic, so the crystal does not show total disorder. Ambiguity of layer arrangements in the case of DHC2 represents the structural basis for twinning and directly arises from the pseudo-symmetry within the layer subspace. Therefore DHC2 clearly fulfills the classification for being an OD-twin (Nespolo *et al.*, 2004). DHC2 consists of OD-layers and the twin operation $l, -k, h$ is isomorphous to the σ -operation transforming a layer into an adjacent one, but being only valid for a subset of the crystal. Ambiguity in boundary layer energies is the reason for this packing behavior (Dornberger-Schiff & Grell-Niemann, 1961). DHC2 can be classified as part of OD structures of type I, as it consists of parallel layers in the **ac** (101) plane building the crystal by a symmetry operation in the direction of **b**. Classification can be further specified to type B, as there exists more than one set of symmetry operations relating the adjacent layers to each other (Dornberger-Schiff, 1956). Twinning by this type of layer arrangement seems to happen quite frequently and applies to several structures in the PDB, for example for the crosslinked-WGA3/GlcNAc β 1,6Gal complex (Muraki *et al.*, 2002) or for the ferrochelatase Hemh-1 from *Bacillus anthracis* (Muller *et al.*, to be published).

4.5 Summary

Twinning is a well known phenomenon in small molecule X-ray crystallography, but the opposite is true for protein crystallography, where twinning can be a serious obstacle in structure determination and refinement. Standard methods in dealing with this problem are therefore either trying to find an untwinned crystal or just completely ignoring the presence of twinning, as structures with a twin fraction about 0.3 can still be refined to a reasonable degree. Recent statistical analysis of the *protein data bank* (PDB) (Lebedev *et al.*, 2006) showed several cases, where the presence of twinning was overlooked or ignored. With ongoing automatization of crystal structure solution, mainly forced by structural genomic projects, the probability for twin affected structures being deposited in the databases is rising, as examination of data statistics for twinning is still no standard method.

The analysis of reflection data for DHC2 crystals clearly indicated the presence of pseudo-merohedral twinning in space group $P2_1$. Due to fortuitous unit cell parameters and non crystallographic symmetry being close to a symmetry operator of a higher crystal system, twinning with a twin law of $l, -k, h$ was observed, with a refined twin fraction of 0.32, being in good agreement with values from intensity statistic based methods for estimation of the twin fraction. Weak boundary interactions between adjacent layers in the **ac** plane and non crystallographic twofold rotation within a layer led to a relative ambiguous packing interaction in the direction of **b**. DHC2 is therefore an OD-twin and the twin fraction is mainly based on kinetics of the growth condition. R factors for refinement are now in a reasonable range, although no significant change in structural model has been recognized. As the boundary layer of **ac** planes consists only of poorly ordered loops, a stabilization of these regions by site directed mutagenesis can be thought of, to remove the ambiguity in crystal packing interactions. Invariant packing would eliminate local symmetry operators from the overall crystal setup and boundary layers would show clear deviations in interaction energies, thus clearly preferring an ordered crystal growth. As a result there would be no structural basis for the occurrence of twinning and a better crystal growth behavior could be achieved. A higher degree of crystal order could also promote diffraction to higher resolution limits.

5 Elimination of twinning in crystals of DHC2 Δ A81

5.1 Introduction

5.1.1 DHC2 - Mutational studies

This work covers only the DHC2 deletion mutant DHC2 Δ A81, but additional mutants were prepared (see Appendix) and will be shortly discussed:

Exchange of the distal ligands of heme groups was performed to alter spectroscopic and electrochemical properties of DHC2. Instead of a histidine, three different amino acids were chosen, which were methionine, lysine and leucine. Methionine and lysine are potential ligands for the central heme group irons, as it is observed for other proteins (for example methionine in horse heart cytochrome *c*). Leucine on the other hand is no potential ligand for a heme group and was intended to eliminate the binding of amino acids as a sixth iron ligand. Single and double mutants were prepared, changing His39 and/or His76. Site directed mutagenesis failed only for the double mutant DHC2H39KH76K. From the prepared mutants, only DHC2H39M could be purified and characterized by UV/Vis absorption spectroscopy. The spectrum showed an additional shoulder (data not shown) around 610 nm indicating a change in the heme ligand field. No crystallization of this DHC2 variant was achieved.

The amino acid residue Phe89 was exchanged to alanine in order to prove, whether the phenylalanine residing in the second heme binding motif has an effect on the ruffled conformation of heme group II (see 3.5.7). The mutant could be expressed, purified and crystallized. The exchange of phenylalanine to alanine was visible in the electron density, but had no effect on the ruffling of heme group II or the overall structure. Differences to the wildtype in UV/Vis absorption spectroscopy were not observed for DHC2F89A.

The relative arrangement of His76 imidazole plane was discussed to be fixed by stacking between residues Phe72 and Pro77 (see 3.5.7). Therefore the phenylalanine was exchanged to a leucine, in order to remove this stacking effect. DHC2F72L was expressed, purified and crystallized. The electron density maps were showing the exchange of phenylalanine to leucine, but no effect on His76 imidazole plane position was visible. Also UV/Vis absorption spectra were unchanged.

The N-terminal loop region of DHC2 was poorly ordered, as was seen

from electron density (see 3.4.9 and Figure 29). Elevated B factors in this loop region indicate an increased flexibility that could prevent a more ordered crystal packing and therefore decrease the resolution in diffraction experiments. Deletion of residues inside this region could result in a better crystal packing and improved resolution. Two loop regions were targeted for this experiment, including residues Lys74, Asn75 on one side and Ala81, Asp82 on the other side. The mutants DHC2 Δ K74 and DHC2 Δ K74N75 were expressed, purified, but failed to crystallize. The mutant DHC2 Δ A81D82 was expressed, purified and crystallized, showing similar properties to DHC Δ A81, but no crystals were found showing the same diffraction quality. The mutant DHC2 Δ A81 will be further discussed and is the subject of this work.

5.1.2 Aim of the present work

In the present work the structural analysis of the deletion mutant DHC2 Δ A81 generated by site directed mutagenesis is presented and reveals an overall altered packing behavior, the elimination of twinning, a higher crystallographic symmetry and improved data quality over the DHC2 wildtype.

5.2 Material and Methods

5.2.1 DNA Oligonucleotides

DNA-oligonucleotides were synthesized and purified by Eurofins MWG Operon (Ebersberg, Germany) and designed to eliminate residue Ala81 from the protein sequence encode by pET22b(+):dhc2. Therefore the DNA base triplet 'GCG' was removed from the oligonucleotides and flanked by DNA bases to ensure specific binding of oligonucleotides to template DNA:

dhc2ΔA81_fw 5'-C CAC CCC GTG AAA CCG GAC GGC CCC ATG CGC TGC-3'

dhc2ΔA81_rv 5'-GCA GCG CAT GGG GCC GTC CGG TTT CAC GGG GTG G-3'

DNA base-triplets coding for amino acids Pro80 and Asp82 flanking residue Ala81 in the wild-type protein sequence are underlined.

5.2.2 Site directed PCR mutagenesis

DNA site directed mutagenesis was done using the Quickchange site directed mutagenesis protocol (Stratagene).

5.3 Results

5.3.1 Crystallization and data collection of DHC2ΔA81

Sequences of DHC2ΔA81 showed the deletion of the base triplet GCG from the pET22b::dhc2 sequence. DHC2ΔA81 was expressed in *Escherichia coli* strain BL21(DE3) with the accessory plasmid pEC86 as described before (see 3.3.3.1). Protein was purified according to wild-type DHC2 (see 3.3.3.2 - 3.3.3.6). No differences in UV/Vis absorption spectroscopy were observed (see 3.3.4.1). Crystallization trials of DHC2ΔA81 were prepared using variations of initial screens (see 3.3.6.1) at protein concentrations of about 10 mg/ml. Initial crystallization hits were observed in a Matrix Screen (Hampton) condition containing 2.4 M of ammonium sulfate and a pH buffer of 0.1 M citrate at pH 4.0 and 20°C, thus differing from the wild-type crystallization condition. After further rounds of improvement crystals grew after two days and reached their final size after seven days, with crystallization conditions of 2.2 M ammonium sulfate and 0.1 M citrate buffer pH 4.2 at 20°C. Crystal morphology was very different from crystals of wild type DHC2, which showed plate like crystal growth (see 3.4.6). In the case of DHC2ΔA81 crystals were three-dimensional in shape, with no visible crystal growth defect in one of the space directions (Figure 40). Crystals tended to grow to large size (600-800 μm) and were of deep red color. For X-ray diffraction experiments crystals were transferred to a cryo-protection buffer containing 2.4 M LiSO₄ and 0.1 M citrate buffer at pH 4.2 (Rubinson *et al.*, 2000). Crystals were tested using a rotating anode X-ray generator (Rigaku) providing Cu-K α radiation and collecting diffraction images with a mar345dtb image plate detector (MAR Research). X-ray diffraction of DHC2ΔA81 was exceeding the detector resolution limits. Crystals of suitable quality were therefore measured at the beamline BW7A at the EMBL outstation at DESY (Hamburg). Low and high resolution data sets were collected at a single wavelength of 1.000 Å. Indexing of diffraction data revealed a space group of P2₁2₁2₁ with unit cell parameters of $a = 47.30$ Å, $b = 52.30$ Å and $c = 61.09$ Å.

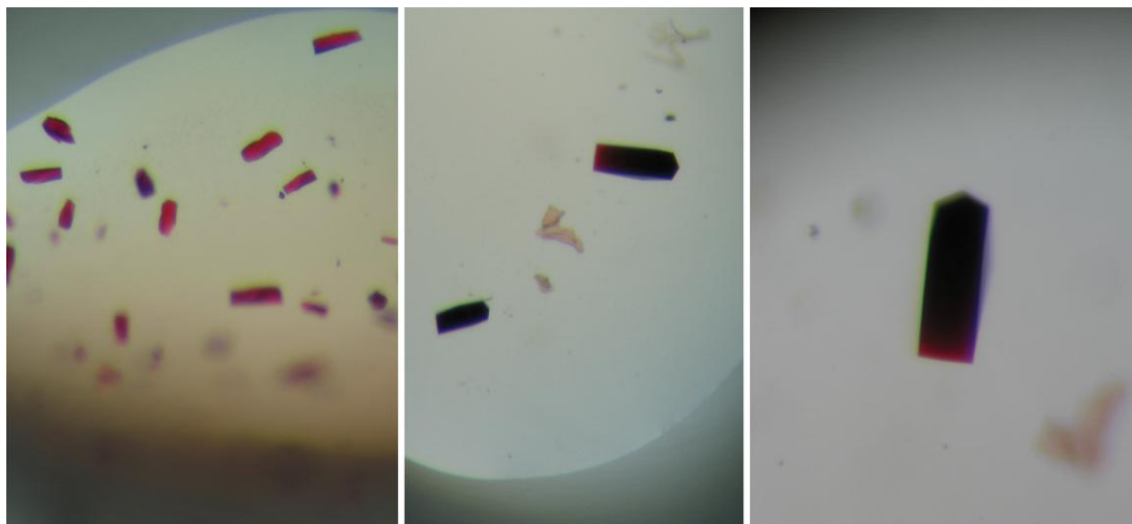


Figure 40: Picture of DHC2ΔA81 crystals, with initial hits in Natrix screen condition with 2.4 M of ammonium sulfate and a pH buffer of 0.1 M citrate at pH 4.0 (left). And the middle and right condition present the fully refined crystal condition at 2.4 M LiSO₄ and 0.1 M citrate buffer at pH 4.2.

5.3.2 Data processing

Data sets were integrated, scaled and merged using the programs DENZO and SCALEPACK (Otwinowski & Minor, 1997). To account for possible pseudo-symmetry as in the case of wild-type DHC2, where monoclinic data exhibited pseudo-orthorombic symmetry, data was merged in space groups $P2_1$ and $P2_12_12_1$ (Rudolph *et al.*, 2004; Zwart *et al.*, 2008). If unit cell parameters in a monoclinic crystal system show an angle of β close to 90° , then twinning by pseudo-merohedry with a rotation around the axis **a** or **c** can be present, described by a twin law of $h, -k, -l$ or $-h, -k, l$ (Larsen *et al.*, 2002). This data then can be merged in the space group $P2_12_12_1$ setup with better merging R factors for a twin fraction getting closer to perfect twinning. To exclude the possibility of DHCΔA81 being pseudo-orthorombic, merging statistics were analyzed to detect possible signs of pseudo-symmetry. With a volume of $151,123 \text{ \AA}^3$ the unit cell of DHC2ΔA81 is $17,839 \text{ \AA}^3$ smaller than the pseudo-orthorombic space group $C222_1$ for wildtype DHC2 (see 4.3.1). DHC2ΔA81 contains two monomers in the asymmetric unit for space group $P2_12_12_1$ and four in the case of space group $P2_1$. The redundancy-independent merging R factor $R_{\text{i.m.}}$ (Weiss & Hilgenfeld, 1997; Weiss, 2001) with 0.05 shows equal values for space group $P2_1$ and $P2_12_12_1$, giving no clear indication for $P2_12_12_1$ being pseudo-orthorombic (Table 12). This is also true for R_{merge} with values of 0.07 for space group $P2_1$ and 0.08 for space group $P2_12_12_1$, and the precision indicating

R factor $R_{p.i.m.}$ with values of 0.02 for both space group setups. No indications for the presence of pseudo-symmetry could be drawn from these statistics. Nonetheless data quality indicators for both space group setups allow for using data up to a resolution of 1.10 Å resolution, representing a drastic improvement in overall diffraction quality in contrast to wild type DHC2.

Data collection and merging statistics of DHC2 Δ A81		
Data collection		
Wavelength (Å)	1.000	
Resolution range (Å)	47.17-1.10 (1.14-1.10)	
No. of reflections	991770	
Merging		
Space group	P2 ₁	P2 ₁ 2 ₁ 2 ₁
No. of unique reflections	113269	60308
Completeness(%)	94.7(91.2)	97.8(95.4)
Redundancy	8.3(6.8)	16.1(13.2)
I/σ	11.5 (1.5)	15.8 (2.1)
R_{merge}^{\dagger}	0.07 (0.77)	0.08 (0.81)
$R_{r.i.m.}^{\ddagger}$	0.07 (0.75)	0.07 (0.74)
$R_{p.i.m.}^{\S}$	0.02 (0.30)	0.02 (0.22)

Table 12: Data collection and merging statistics of DHC2 Δ A81. Values in parentheses are given for the highest resolution shell. $\dagger R_{merge}$ is the merging R -factor. $\ddagger R_{r.i.m.}$ is the redundancy independent merging R -factor. $\S R_{p.i.m.}$ is the precision-indicating merging R -factor (Weiss & Hilgenfeld, 1997; Weiss, 2001).

5.3.3 Cumulative intensity distribution

Calculation of the cumulative distribution of intensities with Z against $N(Z)$ for acentric reflection intensities (Rees, 1980), shows an unusual distribution (Figure 41). But in contrast to wild type DHC2 the distribution is shifted in the direction of centric reflections. This observation implies a bigger deviation from mean intensities and can be explained by splitting of intensities into a bimodal distribution. This effect on cumulative intensity distribution is known for data where pseudo-translation affects the overall data statistics. Fractional half translation symmetry elements in one or more directions of the basis vector \mathbf{a} , \mathbf{b} and \mathbf{c} lead to extinction of reflections along the axes, such that the distribution of intensities is split into weak reflection intensities (affected by extinction) and strong reflection intensities (unaffected by extinction) (Chook *et al.*,

1998). This leads to a bimodal distribution of reflection data intensities and would explain the distribution of intensities in the case of DHC2 Δ A81. As the effect of pseudo-translation would overlap with twinning, no indication for the presence of twinning can be drawn from this test (Rudolph *et al.*, 2004).

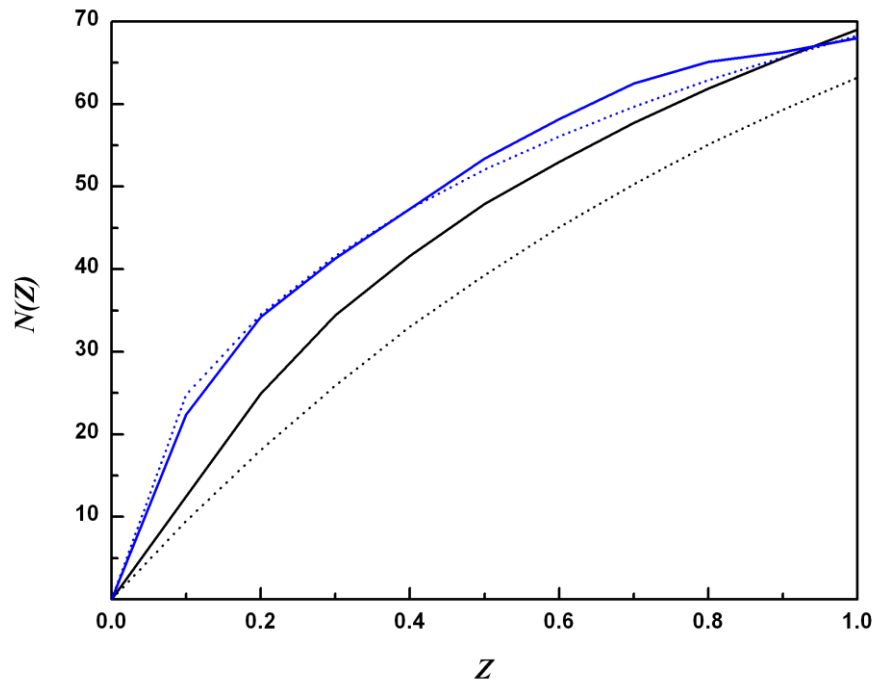


Figure 41: Cumulative distribution of intensities, showing theoretical untwinned acentric data as a black dotted line and centric data as a blue dotted line. Observed data is shown in blue for acentric and black for centric reflections. Significant deviations from theoretical data can be observed for acentric reflections

5.3.4 Data intensity statistics

Intensity statistics calculated with the program *phenix.xtriage* (Adams *et al.*, 2002) show similar characteristics as already observed for the cumulative intensity distribution. Statistics of $|E^2-1|$ for acentric reflections show a value of 0.81 and therefore lie far above theoretical, untwinned value, (0.736) and the one for twinned data (0.541) (Table 13). The same observation holds true for the second moment of normalized intensities, with an observed value of 2.47 in contrast to the theoretical untwinned data value of 2.0 and a value for perfectly twinned data of 1.5. Similar observations can be made for the analogous distribution of structure factor amplitudes, where the observed value of 0.75 is below the value for theoretical untwinned data of

0.785 and far below the value for perfectly twinned data with 0.885. All these values indicate an effect on intensity statistics that has exactly the opposite effect on the distribution of intensities as twinning would have. Pseudo-translation therefore seems to be the most probable explanation for the occurrence of unusual behaving intensity statistics (Chook *et al.*, 1998; Rudolph *et al.*, 1994; Zwart *et al.*, 2008).

	Observed	Untwinned	Perfect twin
Wilson ratio and moments			
$\langle I^2 \rangle / \langle I \rangle^2$	2.470	2.000	1.500
$\langle F^2 \rangle / \langle F \rangle^2$	0.750	0.785	0.885
$ E^2 - 1 $	0.810	0.736	0.541
Yeates [†] statistics (for $h, -k, -l$) (H-test)			
$\langle H \rangle$	0.020	0.500	0.000
$\langle H^2 \rangle$	0.001	0.330	0.000
Padilla & Yeates [‡] statistics (L-test)			
$\langle L \rangle$	0.494	0.500	0.375
$\langle L^2 \rangle$	0.325	0.333	0.200

Table 13: Statistics of acentric data intensities for DHC2 Δ A81. [†] According to Yeates (1997). [‡] According to Padilla & Yeates (2003).

5.3.5 Twin law-dependent twinning tests

To prove the presence of pseudo-merohedral twinning in space group $P2_1$ twin law dependent twinning tests were done using the program DETWIN (CCP4, 1994). H-test (Yeates, 1997) and Britton (Britton, 1972) plot using a twin law of $h, -k, -l$ are showing twin fractions near to 0.5 indicating perfect twinning, as the H-test reveals twinning with a twin fraction of 0.48 and the Britton plot about 0.46 (Figure 42). For approaching perfect twinning, twin law dependent twinning tests tend to become kind of uncertain (Rudolph *et al.*, 2004). On the other hand, twinning tests will always indicate perfect twinning when being applied to data that was merged in a too low symmetry, as the twin law used in these plots is part of real crystallographic symmetry. Therefore these plots do not represent any indication for the presence of twinning in the case of DHC2.

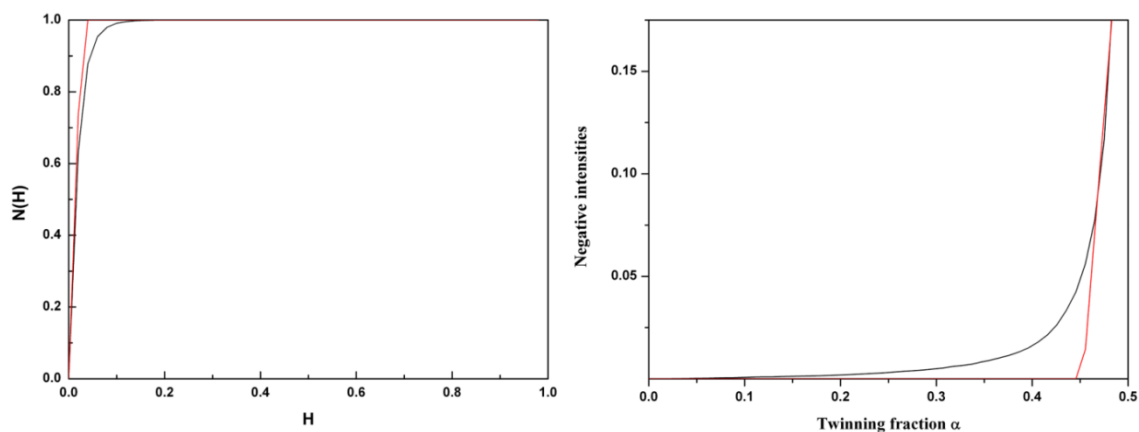


Figure 42: (left) H-test and (right) Britton plot for DHC2 Δ A81 merged in space group P2₁, showing either perfect twinning or merging in a too low symmetry space group.

5.3.6 Detection of pseudo-translational symmetry

A native Patterson map can help to elucidate presence of pseudo-translation vectors, which could be the reason for the altered intensity statistics. With the program XPREP (Bruker AXS) a native Patterson map was calculated for $\nu=0$ using a maximum resolution of 2.5 Å (Patterson, 1934). Two peaks are observed in the representation having a significant height of 44 % of the origin peak (Figure 43). The coordinates of the peaks are $u = 0.424$, $v = 0.000$, $w = 0.500$ and $u = 0.576$, $v = 0.000$ and $w = 0.500$. This result indeed shows the presence of pseudo-translational symmetry which is close to an ideal centering operation of $(\frac{1}{2}, 0, \frac{1}{2})$. Intensity statistics are very strongly affected by this finding, as translations of about half axis lengths lead to destructive intensity interference and therefore to extinction of reflexes and changing intensity distribution into a bimodal shape, decreasing overall data intensity. As most twinning tests are dependent on a sharpening of the intensity distribution, their results are not reliable anymore. Additionally the correct setup of the space group screw axis is based on systematic absences, but as they are being affected by pseudo-translational symmetry, choice of screw axis setup has to be justified by other criteria.

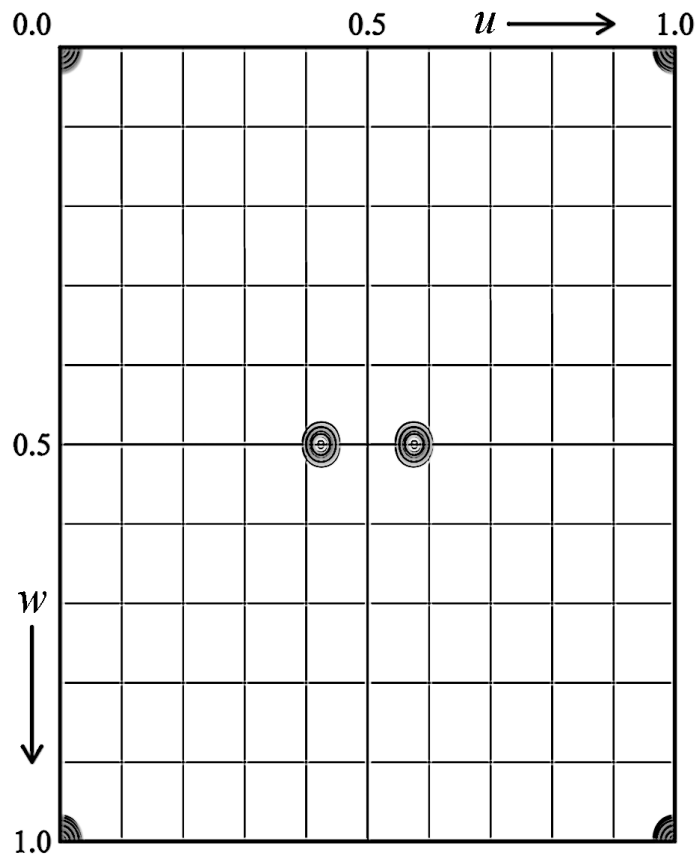


Figure 43: Native Patterson map for the detection of pseudo-translational symmetry calculated for DHC2ΔA81 using XPREP (Bruker AXS). Plot data shows a clear off-origin peak at $u = 0.42$, $v = 0.0$, $w = 0.5$ and a corresponding second peak, sharing 44 % height of the origin peak. Translational symmetry operation nearly matches a single face centering operation of space group $C222_1$ with a rotational unity operator and a translational component of $(\frac{1}{2}, 0, \frac{1}{2})$.

5.3.7 Twinning test for local intensities

As commonly used intensity statistics for the detection of twinning are heavily affected by the presence of pseudo-translational symmetry, the twinning test of local intensities was applied, as it is a robust test for the detection of twinning when pseudo-centering and anisotropic diffraction are present (Padilla & Yeates, 2003). The L-test indicated (Adams *et al.*, 2002) no twinning, as the distribution of local intensities followed exactly the expected distribution for theoretical untwinned data (Figure 44). Therefore data of DHC2ΔA81 is in a setting where intensities are only affected by pseudo-translation, but not by twinning. A primitive orthorhombic spacegroup is thus the right choice of symmetry for DHC2ΔA81 crystals.

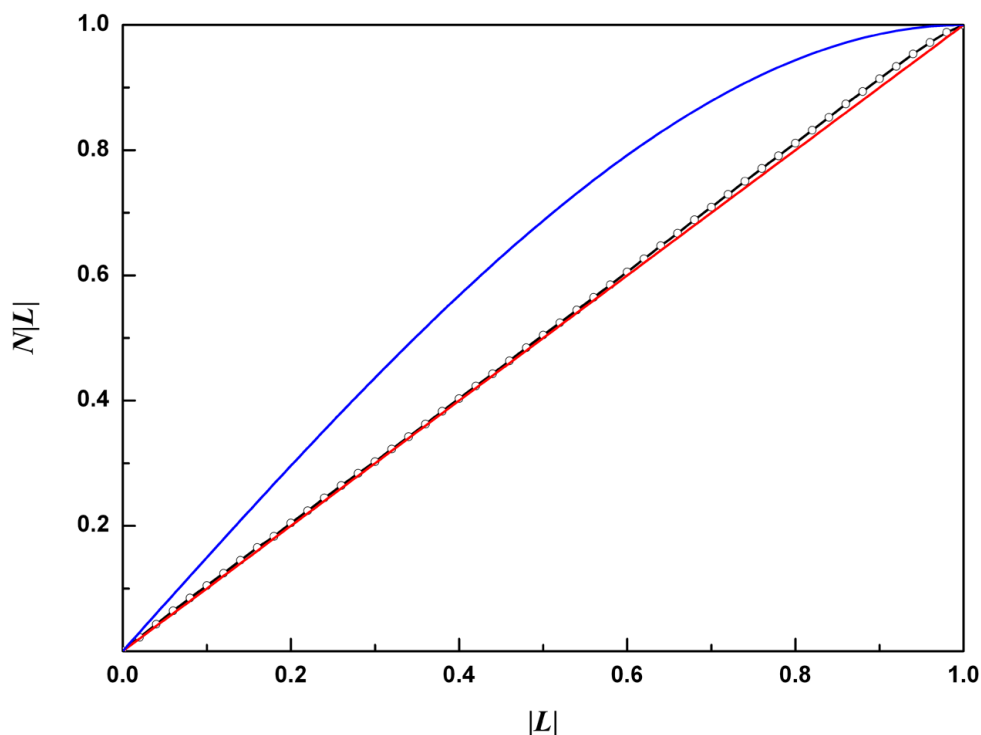


Figure 44: Distribution of local intensities $|L|$ against $N(|L|)$ indicates the total absence of twinning in the case of DHC2ΔA81, with $L = I_1 - I_2 / (I_1 + I_2)$, where I_1 and I_2 are intensities of reflections residing close in reciprocal space.

5.3.8 Structure solution by single wavelength anomalous dispersion (SAD)

Standard structure solution for an already known similar structure, especially in cases of a single mutation, would imply a molecular replacement approach. As this technique is based on Patterson function search algorithms, it can get easily obscured by pseudo-translational symmetry. Unfortunately the model used for molecular replacement can introduce model bias to the initial model in structure refinement. Therefore a direct method approach was chosen over molecular replacement, to obtain an initial model for DHC2ΔA81 and to establish the correct screw axis setup for DHC2ΔA81, as this choice being made on the basis of systematic absences can be strongly influenced by pseudo-translational symmetry. Although the wavelength at which diffraction data for DHC2DA81 was collected, is not near an absorption edge of the transition metal iron, a strong anomalous signal can still be observed, as the data is of very high accuracy. Therefore structure solution was attempted in all eight screw axis setups possible for primitive orthorhombic symmetry, namely P222, P2(1)22, P22(1)2, P222(1), P2(1)2(1)2, P2(1)22(1), P22(1)2(1) and P2(1)2(1)2(1), using the anomalous iron signal to a

resolution of 1.1 Å in the program SHELXD (Schneider & Sheldrick, 2002) and searching for four iron atoms in the asymmetric unit (Figure 45a). With an overall correlation coefficient of 45.1 the solution for space group $P2_12_12_1$ by far exceeded all other space group solutions and after 200 cycles in SHELXE (Sheldrick, 2002) shows very well defined electron densities after solvent flattening and phase improvement. The pseudo-free correlation coefficient for the original hand was 80.1 %, with a connectivity of 0.928 and a contrast of 0.559. The inverted hand showed a pseudo-free correlation coefficient of 63.5 %, a connectivity of 0.876 and a contrast of 0.458. Only the original hand yielded a reasonable electron density map. Experimental electron density showed already holes in aromatic rings, well defined waters and several amino acid sidechain arrangements (Figure 45b). All other space group setups were not yielding interpretable electron density maps leaving $P2_12_12_1$ as the correct space group choice. This method already indicates the high quality of the collected data, as without any model phase information electron densities can be generated.

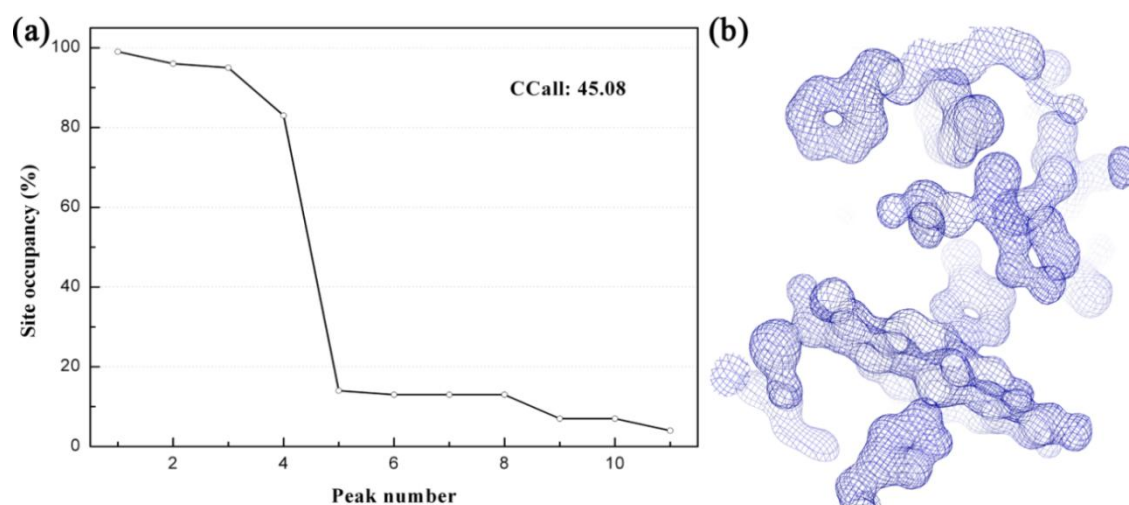


Figure 45: (a) Patterson peak search reveals four iron atom sites that seem to be fully occupied, indicated by a sharp drop in relative site occupancy after peak number 4. The overall correlation coefficient CCall shows a value of 45.08 and represents a good correlation. (b) Experimental electron density map derived from Fe-SAD experiment using SHELXD and SHELXE (Schneider & Sheldrick, 2002; Sheldrick, 2002). Electron density map contoured at 1.5σ shows the overall data quality and the right space group choice, as atomic resolution is indicated by structural visible features, being well defined density for phenylalanine, histidine, proline, leucine and heme groups. Also density for solvent molecules could be found.

5.3.9 Model building and structure refinement of DHC2ΔA81

The structural model of wild-type DHC2 was fitted into the experimental electron density map, residue Ala81, water molecules, sulfate ions and sodium ions were removed and the following residue numbers decreased by one using the program COOT (Emsley & Cowtan, 2004). Atom positions of the structural model were refined against the experimental electron density using COOT and amino acid residues with undefined density were deleted from the model, to avoid introduction of model bias into the refinement process. As twinning was not suspected, no special care was taken for picking reflections for the validation set, which contained 5 % of unique reflections in space group $P2_12_12_1$. Refinement was carried out using SHELXL (Sheldrick, 2008) and building in COOT. After some iterative cycles of building and refinement with isotropic B-factors, an anisotropic B-factor model was introduced, leading to an overall drop in R_{cryst} and R_{free} of about 5 % (Cruickshank, 1956; Dunitz *et al.*, 1988). Overall refinement R factors of 15.5 % for R_{cryst} and 21.6 % for R_{free} (Table 14) were reached after building two monomers into the asymmetric unit containing together 141 amino acid residues, 4 heme groups, 7 sulfate ions, one citrate molecule and 264 water molecules, using data to a resolution of 1.1 Å. Structure quality indicators calculated using the program WHAT_CHECK (Vriend, 1990; Hoof *et al.*, 1996) showed an r.m.s.d. of 0.013 in bond lengths and 2.569° in bond angles with respect to theoretical ideal data (Engh & Huber, 1991). A Ramachandran plot calculated with the program PROCHECK (Laskowski *et al.*, 1993), shows that 83.2 % residues lay in the core region and 16.8 % of the residues in the allowed region, with no residues in the generously allowed or disallowed regions. Refinement against intensities with $I/\sigma > 4$ shows an R_{cryst} of 13.7 % and an R_{free} of 20.7 % indicating a difference between strong and weak reflections that might be a consequence of pseudo-translational symmetry.

Refinement statistics	
$R_{\text{cryst}} (\sigma) / (4\sigma)$	0.154 / 0.137
$R_{\text{free}} (\sigma) / (4\sigma)$	0.219 / 0.207
r.m.s.d. from ideality†	
Bond lengths (Å)	0.013
Bond angles (°)	2.569
Ramachandran plot‡	
Core (%)	83.2
Allowed (%)	16.8
Dissallowed (%)	0

Table 14: Refinement statistics of DHC2ΔA81. The refinement R factor $R_{cryst} = \sum_{hkl} ||F_{obs}(hkl)| - k|F_{calc}|| / |F_{obs}(hkl)|$; R_{free} is calculated for a test set including 5% of the overall reflections. † The r.m.s.d. from ideality was calculated using WHATCHECK (Hooft *et al.*, 1996) ‡ The Ramachandran plot was calculated using PROCHECK (Laskowski *et al.*, 1993).

5.3.10 Protein structure of DHC2ΔA81

The structural model of DHC2ΔA81 contains two monomers in the asymmetric unit, with monomer A built from amino acid residues Arg29 to His96 and monomer B from Thr30 to His96 and shows the same heme group and secondary structure arrangement as wild-type DHC2 (see 3.4.8, Figure 28 and 46). Monomer A and B differ by an r.m.s.d. of 0.486 Å in main chain atom positions, possibly due to a flexible loop region ranging from His64 to Gly69 in monomer A, which shows very high B factors and poor electron density, indicating disorder of this loop region. Monomer B in contrast seems to be well-ordered indicated by overall low B factors in all parts of the structural model and does not exhibit any disorder or absent electron density.

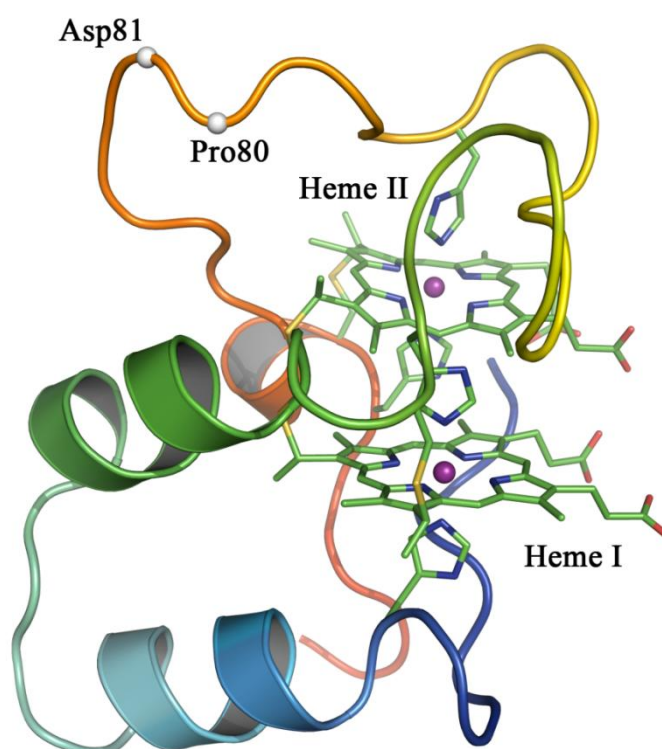


Figure 46: Structural model representation of *c*-type cytochrome DHC2ΔA81 from *Geobacter sulfurreducens*. Protein backbone is shown in cartoon representation, heme binding motifs and heme itself are shown in stick representation. Position of residues Asp81 and Pro80, which flanked Ala81 in wildtype DHC2, are indicated by a white sphere.

5.3.11 Heme group configuration of DHC2ΔA81

The configuration of heme groups and their ligands is slightly shifted with respect to wildtype DHC2. Where the proximal ligands His63 and His 90 provided by the C-X-X-C-H binding motif seem to be relatively congruent with wild type DHC2 histidines, distal ligand histidine positions with His39 and His76 have slightly or significantly changed. His39 and His63 show a dihedral angle of imidazole planes about 31° , whereas the former coplanar arrangement of His76 and His90 now comprises an dihedral angle of 33° . Additionally to this observation, the whole position of His76 residing in the C-terminal loop region of DHC2ΔA81 is shifted by almost 1 Å and therefore also the tilt with respect to the heme plane. Relative geometrical properties of the porphyrin macrocycles stay unchanged, as heme group I is still planar and heme group II shows a ruffled conformation.

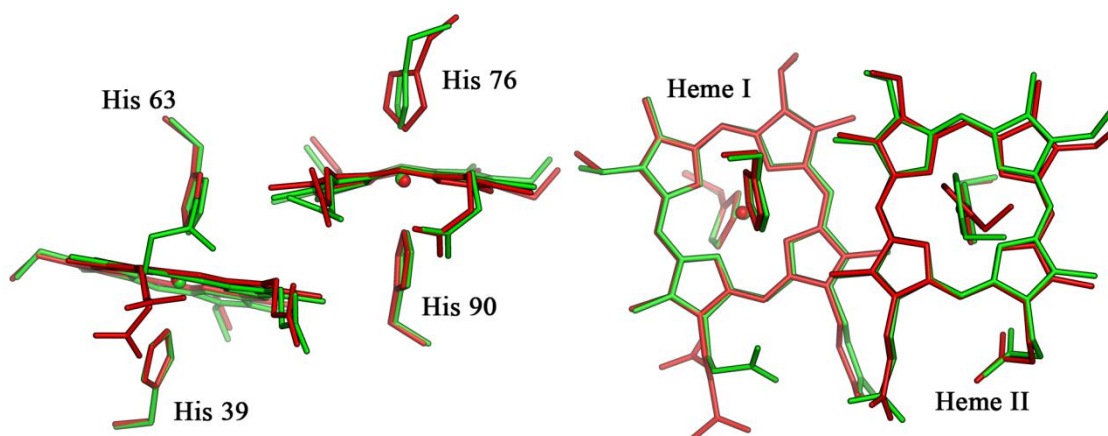


Figure 47: Heme group configuration of DHC2ΔA81 (red) monomer B and wild-type DHC2 (green), with sticks representing the heme groups and the axial ligands, Fe(III) is shown in sphere representation. This picture indicates the alternative arrangement of His 76 as a consequence of sited directed mutagenesis.

5.4 Discussion

5.4.1 Comparison of DHC2 Δ A81 and wildtype DHC2 structures

Superimposition of wildtype DHC2 (see 3.4.8) and DHC2 Δ A81 structures using PYMOL shows an almost identical assembly of secondary structure elements and heme groups. Large differences of structures are mainly present in the C-terminal loop region ranging from Phe72 to Arg86 for wildtype DHC2, which shows huge differences in protein backbone atom positions, resulting in an r.m.s.d. of 2.7 Å for C-alpha atoms. Also the distal ligand of heme group II His76 is affected by this different positioning and is shifted as a consequence of rearrangement of surrounding amino acid residues.

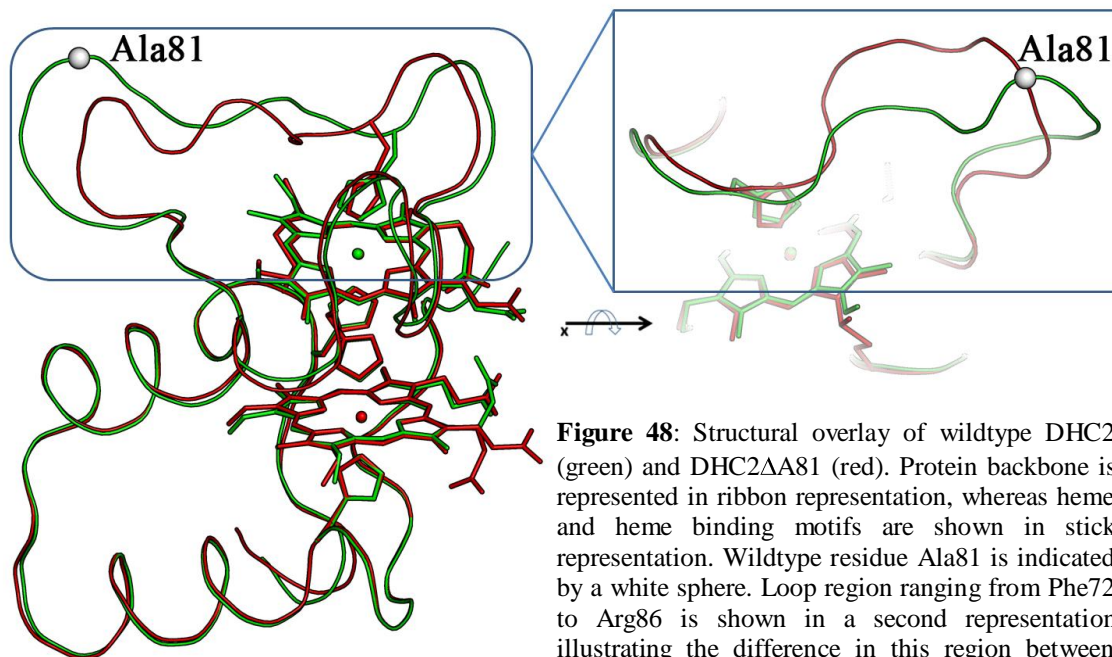


Figure 48: Structural overlay of wildtype DHC2 (green) and DHC2 Δ A81 (red). Protein backbone is represented in ribbon representation, whereas heme and heme binding motifs are shown in stick representation. Wildtype residue Ala81 is indicated by a white sphere. Loop region ranging from Phe72 to Arg86 is shown in a second representation illustrating the difference in this region between wildtype DHC2 and DHC2 Δ A81.

The position of the Ala81 loop region shows the largest deviation from positions in DHC2 Δ A81. Especially positions of and around Pro83/84 are shifted and can be described as inverted in contrast to wild type DHC2. An explanation for this finding might be the decrease in degrees of freedom for torsion angles, as prolines can't occupy the same set of torsion angles as other amino acid residues. The deletion of Ala81 might have introduced a strong bending to this loop region, as might be indicated by the relative planar orientation of amino acids between Pro80 and Pro83. This bending

seems to have influence on the overall structural fold of this loop region as atom positions exactly around this amino acid motif are altered. At the same time this arrangement seems to be of much higher stability, as is indicated by the decrease in loop region temperature B factors and reduction of the overall volume of this region.

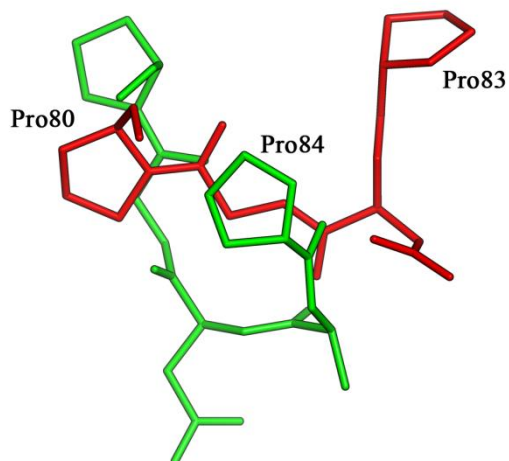


Figure 49: Stick representation of loop region Pro80 to Pro83/84 of wildtype DHC2 (green) and DHC2 Δ A81 (red) given in stick representation and showing the enormous difference of atom positions in this loop region.

5.4.2 Comparison of B factor distribution

The B factor distributions show a clear improvement of the overall protein order in contrast to wildtype DHC2, which results especially in the C-terminal loop region in a strong decrease of B factors, only the ends of N- and C-terminus in DHC2 Δ A81 are showing disorder and missing electron density (Figure 50). Improvement of crystal order in the case of DHC2 Δ A81 is a direct consequence of elimination of Ala81 from wildtype DHC2 in combination with the newly found crystallization condition.

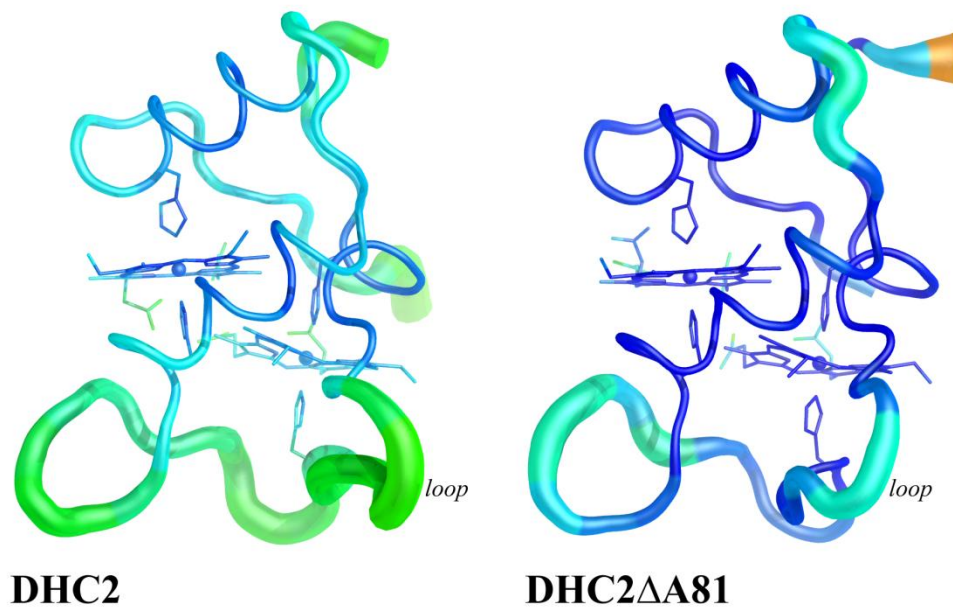


Figure 50: Representation of B factor values of protein regions in wildtype DHC2 (left) and DHC2 Δ A81 (right) as indicated by a B factor putty representation, where blue colors indicate rigidity and red colors a higher flexibility, the thickness of the representation should visualize relative flexibility about the mean atom positions.

5.4.3 Implications from symmetry

Wildtype DHC2 crystals were twinned by pseudo-merohedry, with two monomers in the asymmetric unit related by non crystallographic symmetry close to a real orthorhombic crystal system operator, leaving monoclinic DHC2 crystals as pseudo-orthorhombic. The minimal supergroup $C222_1$ of $P2_1$ exhibits unit cell parameters of $a = 47.76 \text{ \AA}$, $b = 63.27 \text{ \AA}$ and $c = 55.67 \text{ \AA}$ for wildtype DHC2 crystals and are close to the observed unit cell parameters of $a = 47.30 \text{ \AA}$, $b = 52.30 \text{ \AA}$ and $c = 61.09 \text{ \AA}$ as observed for DHC2 Δ A81, with **b** and **c** being interchanged because of space group nomenclature conventions. Cell volumes changed from $168,222 \text{ \AA}^3$ in the case of DHC2 to $151,123 \text{ \AA}^3$ for DHC2 Δ A81 already indicating a tighter crystal packing. This is also reflected by a decrease of the Matthews coefficient from $1.86 \text{ \AA}^3/\text{Da}$ to $1.68 \text{ \AA}^3/\text{Da}$ and a decrease of the solvent content from 33.8 % to 26.98 %. Changes introduced on packing therefore can be directly related to the deletion of Ala81 and the newly applied crystallization condition.

5.4.4 Structural reasons for elimination of twinning

Shifting of the loop region Phe72 to Arg85 seems to have introduced a totally new arrangement of interactions into the former **ac** (101) and now **ab** (110) layer. Where non-crystallographic symmetry was present in DHC2, real crystallographic symmetry is now present and *vice versa*. Composition of the **ab** layer in DHC2ΔA81 does not show the regular arrangement observed for DHC2 where monomers were aligned in a horizontal and vertical direction, leading directly to the ambiguity of overall layer arrangement as such a regular surface also gives no real differentiation in binding energies, like a brick with no distinguishable surface (Parsons, 2003). In DHC2ΔA81 these layers seem to be shifted with respect to each other, as vertical layer rows are set in different mutual arrangement than is visible for DHC2, induced by the deletion of Ala81. Altering of the relative spatial positions of layers in the **ab** plane of DHC2ΔA81 leads on one hand to real crystallographic symmetry, as one monomer in one layer is connected to a monomer in a different layer by a distinct interface and not a loose arrangement. This correlates to an overall increase in charged surface contacts, especially in the direction of layer stacking. Interestingly both crystal setups can be connected via the space group $C222_1$, as the non crystallographic symmetry in $P2_1$ is closely missing a space group operator of $C222_1$ makes this setup pseudo-orthorombic and promotes the presence of twinning by pseudo-merohedry. The centering operator in this case is a crystallographic translation of the monomer to the center of the $C222_1$ cell. For DHC2ΔA81 a different setup emerges. The former non-crystallographic symmetry turns into real crystallographic symmetry in space group $P2_12_12_1$, but the centering operation already observed for DHC2 still does not hold and closely misses a centering operation of $C222_1$, which would transform the $P2_12_12_1$ setup into this space group $C222_1$ as well. Still a higher degree of order was achieved, as is indicated by the high resolution data and overall structural quality indicators. And it is exactly the elimination of residue Ala81 that is responsible for different monomer arrangement, a tighter packing, invariant crystal interface and a higher degree of order, with the only drawback of a pseudo-centering operation of (0.42, 0, 0.5) that affects data intensity statistics and reduces the overall intensity of the data set (Chook *et al.*, 1998; Rudolph *et al.*, 2004; Zwart *et al.*, 2008), by systematically weakening or extinguishing reflections that would be absent in the presence of space group $C222_1$. As the condition does not hold only

overall intensity is decreased, as symmetry information does not increase the overall data to parameter ratio.

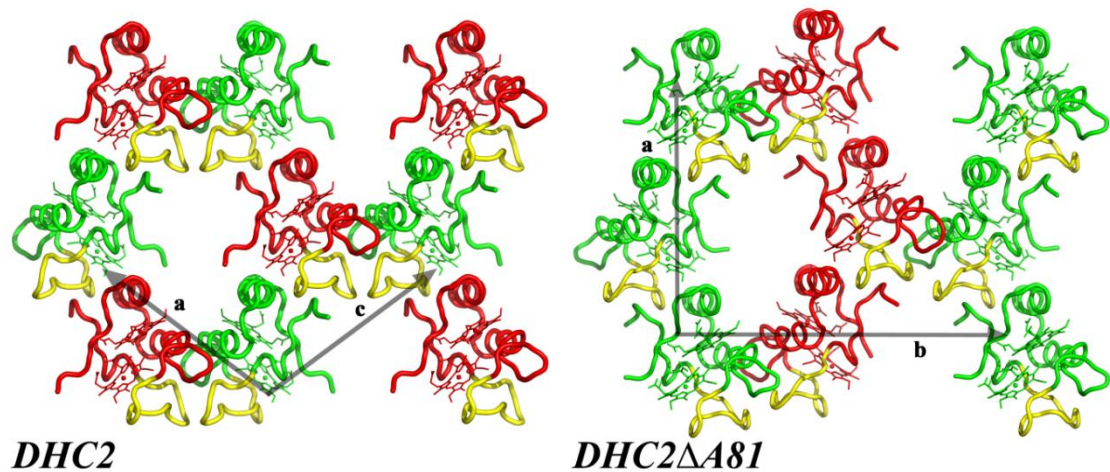


Figure 51: Crystal packing properties of DHC2 and DHC2 Δ A81 in the ac (101) ab (110) respectively. Protein backbones are shown in ribbon, heme groups in stick and Fe (III) atoms in sphere representation. Symmetry mates were generated using PYMOL indicating crystal structure packing. Coloring in green and red indicates symmetry related monomers of the asymmetric unit. Yellow color indicates the loop region Phe72 to Arg85/86 that is affected by the deletion of Ala81 in a prominent way. Differences in the layer arrangement become obvious as regular horizontal arrangement of DHC2 molecules is altered in contrast to DHC2 Δ A81, where regular horizontal layer are not formed.

5.4.5 Changes of packing in the terminology of order disorder

Order disorder terminology is able to explain the observed packing behavior in a convenient way and shows why layers of DHC2 Δ A81 molecules do not fall under this term anymore (Dornberger-Schiff, 1956). In the case of DHC2, two different sets of σ -symmetry operations were able to form adjacent ac layer planes in the direction of the plane normal. These were on one hand the set of crystallographic symmetry operators and on the other hand the set of symmetry operations generated by the twin law $l, -k, h$. Both symmetry operations are possible because of the ambiguous packing possibilities of layer planes, being a result of poorly ordered loop regions facing the layer surface and pseudo-orthorombic non-crystallographic symmetry arrangement (Dornberger-Schiff & Grell-Niemann, 1961). For DHC2 Δ A81 this looks totally different. There is only one set of σ -symmetry operations left, transforming one layer into the adjacent one. Ambiguities in layer contacts have been eliminated by mutating Ala81 out of the

structure and a new crystallization condition. By shifting of lattice rows within the **ac** (101) / **ab** (110) layer the ambiguity of the interface is eliminated as a sole rotation could not fit a layer stack in a thermodynamically and kinetically preferred manner. The structure is fully ordered and layers cannot be denoted as OD-layers anymore, as there is no local symmetry differing from the overall crystal symmetry (Ďurovič, 1997; Nespolo *et al.*, 2004).

5.5 Conclusions and future perspectives

Aim of this work was converting a twinned crystal into an untwinned crystal, with an increase in diffraction power and order. Elimination of Ala81 from the structure of DHC2 led to an already macroscopically visible improvement in crystal growth behavior. Crystals were of space group $P2_12_12_1$ and showed aberrant intensity statistics that could obscure twinning detection. The reason for these statistics is a pseudo-translation symmetry operation of (0.42, 0, 0.5) relating the two monomers of the asymmetric unit and being close to a centering operation of space group $C222_1$ with (0.5, 0, 0.5). This generally broadens scattering of intensities about the mean intensity and thus affects intensity statistics. Only the statistics of local intensities were therefore able to show that crystals of DHC Δ A81 were real orthorhombic with space group $P2_12_12_1$ and no twinning is present. The structure could be solved by a Fe-SAD experiment and refined with SHELXL to reasonable R factors, taking into account the presence of intensities strongly affected by pseudo-translational symmetry. Structural models of DHC2 and DHC2 Δ A81 are only differing in their C-terminal loop regions, where elimination of Ala81 seems to have led to a different loop arrangement that is also involving the positioning of His76, which shows an altered behavior in heme group configuration and has to be further characterized by biochemical investigations like EPR and redox titrations. Bringing Pro80 and Pro83/84 closer together may have led to a decrease in torsion angle degree of freedom and led to this loop arrangement. Overall packing and quality of data is improved for Ala81, which can be related to a more distinct order in mutual layer orientations during crystallization. The aims set for this experiment have been completely fulfilled, as twinning is eliminated, resolution of data has increased and the structure is of higher quality. The only drawback is the presence of pseudo-translational symmetry affecting the overall intensity quality.

6 Danksagung

Die vorliegende Arbeit entstand im Zeitraum von Februar 2005 bis September 2008 in der Abteilung für Molekulare Strukturbiologie an der Georg-August-Universität zu Göttingen.

Meinem Doktorvater Prof. Dr. Oliver Einsle möchte ich für die reizvolle Aufgabenstellung, die hervorragende Betreuung, die angenehme und fordernde Arbeitsatmosphäre, die stetige Unterstützung, sowie die vielen Diskussionen über Röntgenkristallographie danken. Ganz besonders möchte ich mich für das in mich und meine Projekte gesetzte Vertrauen bedanken.

Herrn Prof. Dr. Ralf Ficner danke ich für die Übernahme des Korreferats.

Dr. Markus Rudolph danke ich für die fruchtbaren Ratschläge zum Thema Verzwilligung.

Dr. Susana Andrade danke ich für die Unterstützung und die vielen Ratschläge zum Thema der Membranproteinkristallographie.

Allen meinen jetzigen und ehemaligen Kollegen im Labor Maren, Antje, Anja, Peer, Wei, Juan, Claudia, Haitham und Daniel für die nette Arbeitsatmosphäre und die vielen und lustigen außeruniversitären Aktivitäten.

Dr. Julia Wittman und Dr. Dirk Röser danke ich für die vielen Tipps und die nette Atmosphäre im Schreibraum.

Dr. Rupert Lang danke ich für die Möglichkeit meine Pymol Kenntnisse auch einmal außerhalb der Wissenschaft anwenden zu können.

Bedanken möchte ich mich auch bei allen Mitgliedern der Abteilung für Molekulare Strukturbiologie für die gute Atmosphäre und für das ein oder andere Grillfest.

Ich danke allen meinen Freunden und besonders meiner Familie für den Rückhalt den sie mir geben.

Mein ganz besonderer Dank gilt Ulrike, die mich immer unterstützt und verstanden hat.

7 Appendix

Abbreviations

Å	Ångström ($1 \text{ Å} = 10^{-10} \text{ m}$)
APS	Ammonium persulfate
Au	gold (Aurum)
BCA	bicinchoninic acid
bp	base pairs
ccm	cytochrome c maturation
Cu	copper
Da	Dalton; $1 \text{ Da} = 1 \text{ g}\cdot\text{mol}^{-1}$
DESY	Deutsches Elektronen-Synchrotron
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxy-nucleoside-triphosphate
DPI	diffraction-component precision index
EDTA	(ethylenedinitrilo)tetraacetic acid
EMBL	European Molecular Biology Laboratory
EOR	exactly overlapping reflections
EPR	electron paramagnetic resonance
Fe	iron (Ferrum)
g	earth's gravitational acceleration
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
Hg	Mercury (Hydrargyrum)
IPTG	Isopropyl β -D-thiogalactoside
K	Kelvin
LB	Luria-Bertani medium
MAD	multiwavelength anomalous dispersion
NHE	normal hydrogen electrode
NMT	non-merohedral twinning

NOR	non overlapping reflections
OD	order - disorder
POR	partially overlapping reflections
Pt	platinum
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	RCSB Protein Data Bank
PEG	polyethylene glycol
r.m.s.d.	root-mean-square deviation
SDS	sodium dodecyl sulfate
TAE	tris-acetate-EDTA
TEMED	tetramethylethylenediamin
Ti	titanium
Tris	tris(hydroxymethyl)aminomethane
UV/Vis	ultraviolet and visible light range
VC	vicinity condition

Amino acids

A	Ala	alanine	M	Met	methionine
C	Cys	cysteine	N	Asn	asparagine
D	Asp	aspartate	P	Pro	proline
E	Glu	glutamine	Q	Gln	glutamine
F	Phe	phenylalanine	R	Arg	arginine
G	Gly	glycine	S	Ser	serine
H	His	histidine	T	Thr	threonine
I	Ile	isoleucine	V	Val	valine
K	Lys	lysine	W	Trp	tryptophane
L	Leu	leucine	Y	Tyr	tyrosine

DNA and amino acid sequence of pET22b(+):dhc2 and mutants (GSU2927)

Wild-type DHC2

```

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
                                     M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
E V R T K K V P L D T N H K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC CAC CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
F P K N H P V K P A D G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

```

DHC2-H39M

```

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
                                     M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT ATG AAG CGA
E V R T K K V P L D T N M K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC CAC CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
F P K N H P V K P A D G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

```

DHC2-H76M

```

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
                                     M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
E V R T K K V P L D T N H K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC ATG CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
F P K N M P V K P A D G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

```

DHC2-H39M-H76M

```

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
                                     M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT ATG AAG CGA
E V R T K K V P L D T N M K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC ATG CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC

```

F P K N **V** P V K P A D G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

DHC2-H39V

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT **GTG** AAG CGA
E V R T K K V P L D T N **V** K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC CAC CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
F P K N H P V K P A D G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

DHC2-H76V

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
E V R T K K V P L D T N H K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC **GTG** CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
F P K N **V** P V K P A D G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

DHC2-H39V-H76V

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT **GTG** AAG CGA
E V R T K K V P L D T N **V** K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC **GTG** CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
F P K N **V** P V K P A D G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

DHC2-H39K

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT **AAG** AAG CGA
E V R T K K V P L D T N **K** K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC

F Y D A F A Q G A G K L D L D
 CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
 R Q C V E C H H E K P G G I P
 TTC CCA AAG AAC CAC CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
 F P K N H P V K P A D G P M R
 TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
 C L F C H K F K L E H H H H H
 CAC TGA
 H ST

DHC2-H76K

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
 M A M V S G
 GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
 E V R T K K V P L D T N H K R
 TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
 F Y D A F A Q G A G K L D L D
 CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
 R Q C V E C H H E K P G G I P
 TTC CCA AAG AAC AAG CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
 F P K N K P V K P A D G P M R
 TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
 C L F C H K F K L E H H H H H
 CAC TGA
 H ST

DHC2-H39K-H76K

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
 M A M V S G
 GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT AAG AAG CGA
 E V R T K K V P L D T N K K R
 TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
 F Y D A F A Q G A G K L D L D
 CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
 R Q C V E C H H E K P G G I P
 TTC CCA AAG AAC AAG CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
 F P K N K P V K P A D G P M R
 TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
 C L F C H K F K L E H H H H H
 CAC TGA
 H ST

DHC2-F72L

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
 M A M V S G
 GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
 E V R T K K V P L D T N H K R
 TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
 F Y D A F A Q G A G K L D L D
 CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
 R Q C V E C H H E K P G G I P
 CTG CCA AAG AAC CAC CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
 L P K N H P V K P A D G P M R
 TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
 C L F C H K F K L E H H H H H
 CAC TGA
 H ST

DHC2-F89A

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT

```

M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
E V R T K K V P L D T N H K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC CAC CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
F P K N H P V K P A D G P M R
TGC CTT TTC TGC CAC AAG GCG AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K A K L E H H H H H
CAC TGA
H ST

```

DHC2-ΔA81

```

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
E V R T K K V P L D T N H K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC CAC CCC GTG AAA CCG GCG GGC CCC ATG CGC
F P K N H P V K P D G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

```

DHC2-ΔA81-D82

```

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
E V R T K K V P L D T N H K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAG AAC CAC CCC GTG AAA CCG GCG GGC CCC ATG CGC
F P K N H P V K P G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

```

DHC2-ΔK74

```

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
M A M V S G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
E V R T K K V P L D T N H K R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F Y D A F A Q G A G K L D L D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R Q C V E C H H E K P G G I P
TTC CCA AAC CAC CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
F P N H P V K P A D G P M R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC CAC
C L F C H K F K L E H H H H H
CAC TGA
H ST

```

DHC2-ΔK74-N75

```

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
                                M  A  M  V  S  G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
E   V   R   T   K   K   V   P   L   D   T   N   H   K   R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F   Y   D   A   F   A   Q   G   A   G   K   L   D   L   D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R   Q   C   V   E   C   H   H   E   K   P   G   G   I   P
TTC CCA CAC CCC GTG AAA CCG GCG GAC GGC CCC ATG CGC
F   P   H   P   V   K   P   A   D   G   P   M   R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC
C   L   F   C   H   K   F   K   L   E   H   H   H   H   H
CAC TGA
H   ST

```

DHC2- ΔA81-D82-ΔK74-N75

```

CTG CTG CTC CTC GCT GCC CAG CCG GCG ATG GCC ATG GTC TCC GGT
                                M  A  M  V  S  G
GAA GTA AGG ACA AAA AAG GTT CCC CTC GAC ACC AAT CAC AAG CGA
E   V   R   T   K   K   V   P   L   D   T   N   H   K   R
TTC TAC GAC GCA TTC GCC CAA GGA GCG GGC AAG CTG GAT CTG GAC
F   Y   D   A   F   A   Q   G   A   G   K   L   D   L   D
CGC CAG TGC GTG GAA TGC CAC CAC GAA AAG CCC GGC GGC ATC CCG
R   Q   C   V   E   C   H   H   E   K   P   G   G   I   P
TTC CCA CAC CCC GTG AAA CCG GGC CCC ATG CGC
F   P   H   P   V   K   P   G   P   M   R
TGC CTT TTC TGC CAC AAG TTC AAG CTC GAG CAC CAC CAC CAC
C   L   F   C   H   K   F   K   L   E   H   H   H   H   H
CAC TGA
H   ST

```

Red colored sequence indicates the pelB leader sequence; green underlayed sequence indicates point of site directed mutagenesis.

8 References

- Aasa, R., & Vännngård, T.** (1975). Signal intensity and powder shapes: A reexamination. *J. Magn. Reson.* **19**, 308-315.
- Abreu, I., Lourenco, A., Xavier, A., LeGall, J., Coelho, A., Matias, P., Pinto, D., Carrondo, M., Teixeira, M., & Saraiva, L.** (2003). A novel iron centre in the split-soret cytochrome *c* from *Desulfovibrio desulfuricans* ATCC 27774. *J. Biol. Chem.* **8**, 360-370.
- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C.** (2002) *PHENIX*: building new software for automated crystallographic structure determination. *Acta Cryst.* **D58**, 1948-1954.
- Ahuja, U. & Thöny-Meyer, L.** (2005). CcmD is involved in complex formation between CcmC and the heme chaperone CcmE during cytochrome *c* maturation. *J Biol Chem.* **280**. 236-43.
- Allen, J., Daltrop, O., Stevens, J., & Ferguson, S.** (2003). C-type cytochromes: Diverse structures and biogenesis systems pose evolutionary problems. *Philos. Trans. R. Soc. London, Ser. B* **358**, 255-266.
- Arciero, D. M., Collins, M. J., Haladjian, J., Bianco, P., & Hooper, A. B.** (1991). Resolution of the 4 hemes of cytochrome *c*₅₅₄ from *Nitrosomonas europaea* by redox potentiometry and optical spectroscopy. *Biochemistry* **30**, 11459-11465.
- Arnoux, P., Sabaty, M., Alric, J., Frangioni, B., Guigliarelli, B., Adriano, J. M., & Pignol, D.** (2003). Structural and redox plasticity in the heterodimeric periplasmic nitrate reductase. *Nat. Struct. Biol.* **10**, 928-934.
- Arslan, E., Schulz, H., Zufferey, R., Kunzler, P. & Thöny-Meyer, L.** (1998). Overproduction of the *Bradyrhizobium japonicum* c-type cytochrome subunits of the *cbb*₃ oxidase in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **251**, 744-747.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K.** (1990). *Current protocols in molecular biology*. Greene Publishing Associates and Wiley-Interscience, New York.
- Baker, N., Sept, D., Joseph, S., Holst, M., & McCammon, J.** (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037-10041.
- Bamford, V., Dobbin, P., Richardson, D., & Hemmings, A.** (1999). Open conformation of a flavocytochrome *c*₃ fumarate reductase. *Nat. Struct. Biol.* **6**, 1104-1107.

- Bamford, V. A., Angove, H. C., Seward, H. E., Thomson, A. J., Cole, J. A., Butt, J. N., Hemmings, A. M., & Richardson, D. J.** (2002). Structure and spectroscopy of the periplasmic cytochrome *c* nitrite reductase from *Escherichia coli*. *Biochemistry* **41**, 2921-2931.
- Barends, T. R., de Jong, R. M., van Straaten, K. E., Thunnissen, A. M. & Dijkstra, B. W.** (2005). *Escherichia coli* MltA: MAD phasing and refinement of a tetartohedrally twinned protein crystal structure. *Acta Cryst.* **D61**, 613–621.
- Barker, P. D., & Ferguson, S. J.** (1999). Still a puzzle: Why is haem covalently attached in *c*-type cytochromes? *Structure* **7**, 281-290.
- Barkigia, K., Chantranupong, L., Smith, K., & Fajer, J.** (1988). Structural and theoretical models of photosynthetic chromophores: Implications for redox, light absorption properties and vectorial electron flow. *J. Am. Chem. Soc.* **110**, 7566-7567.
- Beckman, D. L. & Kranz R. G.** (1993). Cytochromes *c* biogenesis in a photosynthetic bacterium requires a periplasmic thioredoxin-like protein. *Proc. Natl. Acad. Sci. U.S.A.* **90**(6), 2179-2183.
- Bendtsen, J., Nielsen, H., von Heijne, G., & Brunak, S.** (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783-795.
- Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J., Frutiger, S. & Hochstrasser, D. F.** (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* **14**, 1023-1031.
- Bond, D. R. & Lovley, D. R.** (2003). Electricity production by *Geobacter sulfurreducens* attached to electrodes. *Appl. Environ. Microbiol.* **69**, 1548-55.
- Bragg, W. L. & Howells, E. R.** (1954). X-ray Diffraction by imidazole methaemoglobin. *Acta Cryst.* **7**, 409–411.
- Brandt, H.** (1927). Über eine Verallgemeinerung des Gruppenbegriffes. *Mathematische Annalen* **96**, 360-366.
- Brigé, A., Cole, J. A., Hagen, W. R., Guisez, Y., & Van Beeumen, J. J.** (2001). Overproduction, purification and novel redox properties of the dihaem cytochrome *c*, NapB, from *Haemophilus influenzae*. *Biochem. J.* **356**, 851-858.
- Brigé, A., Leys, D., Meyer, T. E., Cusanovich, M. A., & Van Beeumen, J. J.** (2002). The 1.25 Å resolution structure of the diheme NapB subunit of soluble nitrate reductase reveals a novel cytochrome *c* fold with a stacked heme arrangement. *Biochemistry* **41**, 4827-4836.
- Britton, D.** (1972). Estimation of twinning parameter for twins with exactly superimposed reciprocal lattices. *Acta Cryst.* **A28**, 296–297.

- Brünger, A. T.** (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Cryst.* **D49**, 24-36.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L.** (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst.* **D54**, 905-921.
- Bushnell, G. W., Louie, G. V. & Brayer, G. D.** (1990). High-resolution three-dimensional structure of horse heart cytochrome *c*. *J. Mol. Biol.* **214**, 585-95.
- Caccavo, F., Lonergan, D. J., Lovley, D. R., Davis, M., Stolz, J. F., & McInerney, M. J.** (1994). *Geobacter sulfurreducens* sp.nov., a hydrogen-oxidizing and acetate-oxidizing dissimilatory metal-reducing microorganism. *Appl. Environ. Microbiol.* **60**, 3752-3759.
- Cochran, W. & Howells, E. R.** (1954). X-ray diffraction by a layer structure containing random displacements. *Acta Cryst.* **7**, 412-415.
- Collaborative Computational Project No. 4.** (1994). The CCP4 Suite: Programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.
- Chook, Y. M., Lipscomb, W. N. & Ke, H.** (1998). Detection and use of pseudo-translation in determination of protein structures. *Acta Cryst.* **D54**, 822-827.
- Cromer, D. T. & Liberman, D.** (1970). Relativistic calculation of anomalous scattering factors for X-rays. *J. Chem. Phys.* **53**, 1891-1898.
- Cruickshank, D.** (1956). The analysis of the anisotropic thermal motion of molecules in crystals. *Acta Cryst.* **9**, 754-756.
- Cruickshank, D.** (1999). Remarks about protein structure precision. *Acta Cryst.* **D55**, 1108-1108.
- Cunha, C. A., Macieira, S., Dias, J. M., Almeida, G., Goncalves, L. L., Costa, C., Lampreia, J., Huber, R., Moura, J. J. G., Moura, I., & Romao, M. J.** (2003). Cytochrome *c* nitrite reductase from *Desulfovibrio desulfuricans* ATCC 27774: The relevance of the two calcium sites in the structure of the catalytic subunit (NrfA). *J. Biol. Chem.* **278**, 17455-17465.
- Daily, H. A.** (1997). Enzymes of heme biosynthesis. *JBIC.* **2**, 411-417.
- Dauter, Z.** (2003). Twinned crystals and anomalous phasing. *Acta Cryst.* **D59**, 2004-2016.
- Declercq, J. P. & Evrard, C.** (2001). A twinned monoclinic crystal form of human peroxiredoxin 5 with eight molecules in the asymmetric unit. *Acta Cryst.* **D57**, 1829-1835.

- DeLano, W. L.** (2002) PyMol, DeLano Scientific, San Carlos, CA.
- Donnay, G.** (1940). Width of albite-twinning lamellae. *Am. Mineral.*, **25**, 578-586.
- Dornberger-Schiff, K.** (1956). On Order-Disorder Structures (OD-Structures). *Acta Cryst.* **9**, 593-601.
- Dornberger-Schiff K & Grell-Nieman, H.** (1961). On the theory of Order-Disorder (OD) Structures. *Acta Cryst.* **14**, 167-177.
- Dornberger-Schiff, K.** (1964). Grundzüge einer Theorie von OD-Strukturen aus Schichten. *Abh. dtsh. Akad. Wiss. Berlin, Kl. f. Chem.* **3**, 107.
- Dornberger-Schiff K, & Fichtner, K.** (1972). On the Symmetry of OD-Structures Consisting of Equivalent Layers. *Kristall und Technik* **9**, 1035-1056
- Dornberger-Schiff, K.** (1979). OD structures - a game and a bit more. *Kristall und Technik* **14**, 1027-1045.
- Duong, F., Eichler, J., Price, A., Leonard, M. R. & Wickner, W.** (1997). Biogenesis of the Gram-negative bacterial envelope. *Cell* **91**, 567-573.
- Dunitz, J. D., Maverick, E. F., & Trueblood, K. N.** (1988). Atomic motions in molecular crystals from diffraction measurements. *Angew. Chem. Int. Ed. Engl.* **27**, 880-895.
- Đurovič, S.** (1997) Fundamentals of OD theory. *Modular aspects of Minerals* (Ed. S. Merlino), Eötvös University Press, Budapest.
- Dutton, P.** (1978). Redox potentiometry: Determination of midpoint potentials of oxidation-reduction components of biological electron-transfer systems, *Methods Enzymol.* **54**, 411-435.
- Eaves, D. J., Grove, J., Staudenmann, W., James, P., Poole, R. K., White, S. A., Griffiths, I. & Cole, J. A.** (1998). Involvement of products of the *nrfEFG* genes in the covalent attachment of haem *c* to a novel cysteine-lysine motif in the cytochrome *c*₅₅₂ nitrite reductase from *Escherichia coli*. *Mol. Microbiol.* **28**, 205-216.
- Einsle, O., Messerschmidt, A., Stach, P., Bourenkov, G. P., Bartunik, H. D., Huber, R., & Kroneck, P. M. H.** (1999). Structure of cytochrome *c* nitrite reductase. *Nature* **400**, 476-480.
- Einsle, O., Stach, P., Messerschmidt, A., Simon, J., Kroger, A., Huber, R., & Kroneck, P. M. H.** (2000). Cytochrome *c* nitrite reductase from *Wolinella succinogenes*: Structure at 1.6 Å resolution, inhibitor binding, and heme-packing motifs. *J. Biol. Chem.* **275**, 39608-39616.
- Einsle, O., Foerster, S., Mann, K., Fritz, G., Messerschmidt, A., & Kroneck, P.** (2001). Spectroscopic investigation and determination of reactivity and structure

- of the tetraheme cytochrome *c*₃ from *Desulfovibrio desulfuricans* Essex 6. *Eur. J. Biochem.* **268**, 3028-3035.
- Einsle, O., Messerschmidt, A., Huber, R., Poulos, T., & Wieghardt, K.** (2001). *Handbook of Metalloproteins*, Wiley, New York.
- Emsley, P. & Cowtan K.** (2004). Coot: model-building tools for molecular graphics. *Acta Cryst.* **D60**, 2126-2132.
- Engh, R. A. & Huber, R.** (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* **D4**, 392-400.
- French, G. S. & Wilson, K. S.** (1978). On the treatment of negative intensity observations. *Acta Cryst.* **A34**, 517.
- Friedel, G.** (1926). *Leçons de Cristallographie*. Nancy/Paris/Strasbourg: Berger Leirault.
- Garman, E. & Owen, R. L.** (2006). Cryocrystallography of macromolecules: practice and optimization. *Methods Mol. Biol.* **364**, 1-18.
- Giacovazzo, C.** (1992). *Fundamentals of Crystallography*. Oxford University Press. Oxford.
- Gouterman, M.** (1978). *The Porphyrins, Vol. III*, Academic Press, New York.
- Graves, D.J. & Wu Y.T.** (1974). On predicting the results of affinity procedures, *Methods Enzymol.* **34**:140-63.
- Harker, D.** (1956). The determination of the phases of the structure factors on noncentrosymmetric crystals by the method of double isomorphous replacement. *Acta Cryst.* **9**, 1-9.
- Heitmann, D. & Einsle, O.** (2005). Structural and biochemical characterization of DHC2, a novel diheme cytochrome *c* from *Geobacter sulfurreducens*. *Biochemistry* **44**, 12411-12419.
- Heitmann, D. & Einsle, O.** (2008). Pseudo-merohedral twinning in crystals of the dihaem *c*-type cytochrome DHC2 from *Geobacter sulfurreducens*. *Acta Cryst.* **D64**, 993-999.
- Heitmann, D. & Einsle, O.** Atomic resolution structure and elimination of pseudo-merohedral twinning in dihaem *c*-type cytochrome DHC2 from *Geobacter sulfurreducens*. (In preparation).
- Henderson, G.** (1990). A new look at carbonyl electronic transitions. *J. Chem. Educ.* **67**, 392.
- Hendrickson, W. A., Smith, J. L., & Sheriff, S.** (1985). Direct phase determination based on anomalous scattering. *Methods Enzymol.* **115**, 41-55.

- Herbst-Irmer, R. & Sheldrick, G. M.** (1998). Refinement of twinned structures with SHELXL97. *Acta Cryst.* **B54**, 443-449.
- Heukeshoven, J. & Dernick, R.** (1988). Improved silver staining procedure for fast staining in PhastSystem development unit. I. Staining of sodium dodecyl-sulfate gels. *Electrophoresis* **9**(1), 28-32.
- Hoffmann, M.** (2007). Crystallographic studies on diheme cytochrome *c* enzymes. Dissertation, Universität Göttingen.
- Hooft, R. W. W., Vriend, G., Sander, C., & Abola, E.E.** (1996). Errors in protein structures. *Nature* **381**, 272-272.
- Igarashi, N., Moriyama, H., Fujiwara, T., Fukumori, Y., & Tanaka, N.** (1997). The 2.8 Å structure of hydroxylamine oxidoreductase from a nitrifying chemoautotrophic bacterium, *Nitrosomonas europaea*. *Nat. Struct. Biol.* **4**, 276-284.
- Ibbi-Nivol, C., Crooke, H., Griffiths, L., Grove, J., Hussain, H., Pommier, J., Mejean, V. & Cole, J. A.** (1994). A reassessment of the range of c-type cytochromes synthesized by *Escherichia coli* K-12. *FEMS Microbiol Lett.* **119**, 89-94.
- Iverson, T. M., Arciero, D. M., Hsu, B. T., Logan, M. S. P., Hooper, A. B., & Rees, D. C.** (1998). Heme packing motifs revealed by the crystal structure of the tetra-heme cytochrome *c*₅₅₄ from *Nitrosomonas europaea*. *Nat. Struct. Biol.* **5**, 1005-1012.
- Jahn, D., Verkamp, E. & Söll, D.** (1991). Glutamyl transfer RNA: A precursor of heme and chlorophyll biosynthesis. *Trends Biochem. Sci.* **17**, 215-218.
- Jameson, G. B., Schneider, R., Dubler, E. & Oswald, H. R.** (1982). Tris(ethylenediamine)nickel(II) sulphate at 110 K. Structure determination in the presence of twinning. *Acta Cryst.* **B38**, 3016-3020.
- Jentzen, W., Simpson, M., Hobbs, J., Song, X., Ema, T., Nelson, N., Medforth, C., Smith, K., Veyrat, M., Mazzanti, M., Ramasseul, R., Marchon, J., Takeuchi, T., Goddard, W., & Shelnut, J.** (1995). Ruffling in a series of nickel(II) meso-tetrasubstituted porphyrins as a model for the conserved ruffling of the heme of cytochromes *c*. *J. Am. Chem. Soc.* **117**, 11085-11097.
- Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M.** (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr.* **A47**, 110-119.
- Kadish, K., Vancaemelbecke, E., Dsouza, F., Medforth, C., Smith, K., Tabard, A., & Guilard, R.** (1995). Electrochemistry and spectroelectrochemistry of σ -bonded iron(III) porphyrins with nonplanar porphyrin rings. Reactions of (OETPP)Fe(R) and (OETPP)FeCl, where R = C₆H₅, C₆F₄H, or C₆F₅ and OETPP

- is the dianion of 2,3,7,8,12,13,17,18-octaethyl-5,10,15,20-tetraphenylporphyrin. *Inorg. Chem.* **34**, 2984-2989.
- Kojo, S. & Sano, S.** (1981). Mechanism of a novel synthesis of hemin-c from protohemin and L-cysteine - A Markownikoff-type radical-addition reaction. *Perkin Transactions* **1(11)**: 2864-2870.
- Kranz, R., Lill, R., Goldman, B., Bonnard, G. & Merchant, S.** (1998). Molecular mechanisms of cytochrome c biogenesis: three distinct systems. *Mol. Microbiol.* **29**. 383-96.
- Krissinel, E. & Henrick, K.** (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst.* **D60**, 2256–2268.
- Kuras, R., Saint-Marcoux, D., Wollman, F.A. & de Vitry, C.** (2007). A specific *c*-type cytochrome maturation system is required for oxygenic photosynthesis, *Proc. Natl. Acad. Sci.U.S.A.* **104(23)**. 9906–9910.
- Labbe-Bois, R., Labbe, P., & Dailey, H.A.** (1990). Biosynthesis of heme and chlorophylls. McGraw-Hill, New York.
- Laemmli, U. K.** (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**. 680-685.
- Larsen, N. A., Heine, A., de Prada, P., el Redwan, R., Yeates, T. O., Landry, D. W. & Wilson, I. A.** (2002). Structure determination of a cocaine hydrolytic antibody from a pseudomerohedrally twinned crystal. *Acta Cryst.* **D58**, 2055–2059.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M.** (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
- Lebedev, A. A., Vagin, A. A. & Murshudov, G. N.** (2006). Intensity statistics in twinned crystals with examples from the PDB. *Acta Cryst.* **D62**, 83-95.
- Leys, D., Tsapin, A. S., Nealson, K. H., Meyer, T. E., Cusanovich, M. A., & Van Beeumen, J. J.** (1999). Structure and mechanism of the flavocytochrome *c* fumarate reductase of *Shewanella putrefaciens* MR-1. *Nat. Struct. Biol.* **6**, 1113-1117.
- Liang, J., Ealick, S., Nielsen, S., Schreiber, S. L. & Clardy, J.** (1996). Crystallization and preliminary X-ray analysis of twinned crystals of a chimeric FK506 binding protein 12 and 13 complexed with FK506. *Acta Cryst.* **D52**, 207–210.
- Ma, J. G., Zhang, J., Franco, R., Jia, S. L., Moura, I., Moura, J. J. G., Kroneck, P. M. H., & Shelnutt, J. A.** (1998). The structural origin of nonplanar heme distortions in tetraheme ferricytochromes *c*₃. *Biochemistry* **37**, 12431-12442.

- Martin, A., Orengo, C., Hutchinson, E., Jones, S., Karmirantzou, M., Laskowski, R., Mitchell, J., Taroni, C., & Thornton, J.** (1998). Protein folds and functions. *Structure* **6**, 875-884.
- Matthews, B. W.** (1968). Estimation of the solvent content in proteins. *J. Mol. Biol.* **33**, 491-497.
- Methé, B. A., Nelson, K. E., Eisen, J. A., Paulsen, I. T., Nelson, W., Heidelberg, J. F., Wu, D., Wu, M., Ward, N., Beanan, M. J., Dodson, R. J., Madupu, R., Brinkac, L. M., Daugherty, S. C., DeBoy, R. T., Durkin, A. S., Gwinn, M., Kolonay, J. F., Sullivan, S. A., Haft, D. H., Selengut, J., Davidsen, T. M., Zafar, N., White, O., Tran, B., Romero, C., Forberger, H. A., Weidman, J., Khouri, H., Feldblyum, T. V., Utterback, T. R., Van Aken, S. E., Lovley, D. R., & Fraser, C. M.** (2003). Genome of *Geobacter sulfurreducens*: Metal reduction in subsurface environments. *Science* **302**, 1967-1969.
- Moore, G. R., & Pettigrew, G. W.** (1990). Cytochromes *c*. Evolutionary, structural and physicochemical aspects. Springer-Verlag KG, Berlin.
- Mowat, C. G., Rothery, E., Miles, C. S., McIver, L., Doherty, M. K., Drewette, K., Taylor, P., Walkinshaw, M. D., Chapman, S. K., & Reid, G. A.** (2004). Octaheme tetrathionate reductase is a respiratory enzyme with novel heme ligation. *Nat. Struct. Biol.* **11**, 1023-1024.
- Muller, A., Lebedev, A.A., Moroz, O.V., Blagova, E.V., Levnikov, V.M., Fogg, M.J., Brannigan, J.A., Wilkinson, A.J., & Wilson, K.S.** Crystal Structure of Ferrochelatase Hemh-1 from *Bacillus Anthracis, strain Ames*. (to be published).
- Mullis, K. B. & Faloona, F. A.** (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155**, 335-50.
- Muraki, M., Ishimura, M. & Harata, K.** (2002). Interactions of wheat-germ agglutinin with GlcNAc β 1,6Gal sequence. *Biochim. Biophys. Acta*, **1569**, 10-20.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J.** (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst.* **D53**, 240-255.
- Mus-Veteau, I., Dolla, A., Guerlesquin, F., Payan, F., Czjzek, M., Haser, R., Bianco, P., Haladjian, J., Rapp-Giles, B., Wall, J., Voordouw, G., & Bruschi, M.** (1992). Site-directed mutagenesis of tetraheme cytochrome *c*₃. Modification of oxidoreduction potentials after heme axial ligand replacement. *J. Biol. Chem.* **267**, 16851-16858.
- Nespolo, M., Ferraris, G., Durovic, S. & Takeuchi, Y.** (2004). Twins vs. modular crystal structures. *Z. Kristallogr.* **219**, 773-778.
- Otwinowski, Z., & Minor, W.** (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

- Padilla, J. E. & Yeates, T. O.** (2003). A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Cryst.* **D59**, 1124–1130.
- Page, C. C., Moser, C. C., Chen, X. X., & Dutton, P. L.** (1999). Natural engineering principles of electron tunneling in biological oxidation-reduction. *Nature* **402**, 47-52.
- Parsons, S.** (2003). Introduction to twinning. *Acta Cryst.* **D59**, 1995–2003.
- Patterson, A. L.** (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. *Phys. Rev.* **46**, 372-376.
- Pettigrew, G. W., & Moore, G. R.** (1987). Cytochromes *c*. Biological aspects. Springer-Verlag KG, Berlin.
- Pisa, R., Stein, T., Eichler, R., Gross, R. & Simon, J.** (2002). The *nrfI* gene is essential for the attachment of the active site haem group of *Wolinella succinogenes* cytochrome *c* nitrite reductase. *Mol. Microbiol.* **43**, 763–770.
- Pohlschröder, M., Prinz, W. A., Hartmann, E. & Beckwith, J.** (1997). Protein translocation in the three domains of life: variations on a theme. *Cell* **91**, 563-566.
- Pratt, C. S., Coyle, B. A. & Ibers J. A.** (1971). Redetermination of the structure of nitrosylpenta-amminecobalt(III) dichloride. *J. Chem. Soc.*, 2146 - 2151.
- Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V.** (1963). Stereochemistry of polypeptide chain configurations. *J Mol Biol.* **7**. 95-99.
- Ravikanth, M., & Chandrashekar, T.** (1995). Nonplanar porphyrins and their biological relevance: Ground and excited-state dynamics. *Coord. Chem. Rev.* **82**, 105-188.
- Rees, D. C.** (1980). The influence of twinning by merohedry on intensity statistics. *Acta Cryst.* **A36**, 578-581.
- Rodrigues, M.L., Oliveira, T.F., Pereira, I.A., & Archer, M.** (2006). X-ray structure of the membrane-bound cytochrome *c* quinol dehydrogenase NrfH reveals novel haem coordination. *Embo J.* **25**, 5951-5960.
- Rubinson, K. A., Ladner, J. E., Tordova M. & Gilliland G. L.** (2000). Cryosalts: Suppression of ice formation in macromolecular crystallography. *Acta Cryst.* **D56**, 996-1001.
- Rudolph, M. G., Wingren, C., Crowley, M. P., Chien, Y. H. & Wilson, I. A.** (2004). Combined pseudo-merohedral twinning, non-crystallographic symmetry and pseudo-translation in a monoclinic crystal form of the $\gamma\delta$ T-cell ligand T10. *Acta Cryst.* **D60**, 656–664.

- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-91.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-7.
- Sanner, M. F., Olson, A. J., & Spehner, J. C. (1996). Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **38**, 305-320.
- Schneider, T. R., & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Cryst.* **D58**, 1772-1779.
- Schulz, H., Hennecke, H., & Thöny-Meyer, L. (1998). Prototype of a heme chaperone essential for cytochrome *c* maturation. *Science* **281**, 1197-1200.
- Sheldrick, G. M. (1997). *The SHELX-97 Manual*, ch. 6.4, pp. 43–44. <http://shelx.uni-ac.gwdg.de/SHELX/>
- Sheldrick, G. M. (2002). Macromolecular phasing with SHELXE. *Z. Kristallogr.* **217**, 644-650.
- Sheldrick, G. M. (2008). A short history of SHELX. *Acta Cryst.* **A64**, 112-122.
- Shelnutt, J., Song, X., Ma, J., Jia, S., Jentzen, W., & Medforth, C. (1998). Nonplanar porphyrins and their significance in proteins. *Chem. Soc. Rev.* **27**, 31-41.
- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., & Klenk, D. C. (1985). Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76-85.
- Stevens, J., Daltrop, O., Allen, J., & Ferguson, S. (2004). C-type cytochrome formation: Chemical and biological enigmas. *Acc. Chem. Res.* **37**, 999-1007.
- Stura, E. A., Satterthwait, A. C., Calvo, J. C., Kaslow, D. C. & Wilson, I. A. (1994). Reverse screening. *Acta Cryst.* **D50**, 448-455.
- Sultana A., Alexeev I., Kursula I., Mäntsälä P., Niemi J. & Schneider, G. (2007). Structure determination by multiwavelength anomalous diffraction of aclacinomycin oxidoreductase: indications of multidomain pseudomerohedral twinning. *Acta Cryst.* **D63**, 149-159.
- Suslick, K. S. & R. A. Watson (1992). The photochemistry of chromium, manganese and iron porphyrin complexes. *New Journal of Chemistry* **16(5)**: 633-642.
- Taylor, P., Pealing, S., Reid, G., Chapman, S., & Walkinshaw, M. (1999). Structural and mechanistic mapping of a unique fumarate reductase, *Nat. Struct. Biol.* **6**, 1108-1112.

- Thöny-Meyer, L., Fischer, F., Kunzler, P., Ritz, D., & Hennecke, H.** (1995). *Escherichia coli* genes required for cytochrome *c* maturation. *J. Bacteriol.* **177**, 4321-4326.
- Thöny-Meyer, L.** (1997). Biogenesis of respiratory cytochromes in bacteria. *Microbiol. Mol. Biol. Rev.* **61**, 337-376.
- Thöny-Meyer, L.** (2002). Cytochrome *c* maturation: A complex pathway for a simple task? *Biochem. Soc. Trans.* **30**, 633-638.
- Tong, L. & Rossmann, M. G.** (1990). The locked rotation function. *Acta Cryst.* **A46**, 783-792.
- Upadhyay, A. K., Petasis, D. T., Arciero, D. M., Hooper, A. B., & Hendrich, M. P.** (2003). Spectroscopic characterization and assignment of reduction potentials in the tetraheme cytochrome *c*₅₅₄ from *Nitrosomonas europaea*. *J. Am. Chem. Soc.* **125**, 1738-1747.
- Vriend, G.** (1990). WHAT IF: A molecular modelling and drug design program. *J. Mol. Graph.* **8**, 52-56.
- Walker, F. A., Huynh, B. H., Scheidt, W. R., & Osvath, S. R.** (1986). Models of the cytochromes *b*: Effect of axial ligand plane orientation on the electron-paramagnetic-res and Mössbauer-spectra of low-spin ferrihemes. *J. Am. Chem. Soc.* **108**, 5288-5297.
- Walker, F. A.** (2004). Models of the bis-histidine-ligated electrontransferring cytochromes. Comparative geometric and electronic structure of low-spin ferro- and ferrihemes. *Chem. Rev.* **104**, 589-615.
- Weiss, M. S. & Hilgenfeld, R.** (1997). On the use of the merging R factor as a quality indicator for X-ray data. *J. Appl. Cryst.* **30**, 203-205.
- Weiss, M. S.** (2001). Global indicators of X-ray data quality. *J. Appl. Cryst.* **34**, 130-135.
- Wittmann, J. G. & Rudolph, M. G.** (2007). Pseudo-merohedral twinning in monoclinic crystals of human orotidine-5'-monophosphate decarboxylase. *Acta Cryst.* **D63**, 744-749.
- Wood, L., Muthukrishnan, K., White, T., Ramdas, L., & Nall, B.** (1988). Construction and characterization of mutant iso-2-cytochromes-*c* with replacement of conserved prolines. *Biochemistry* **27**, 8554-8561.
- Yang, F., Dauter, Z. & Wlodawer, A.** (2000). Effects of crystal twinning on the ability to solve a macromolecular structure using multiwavelength anomalous diffraction. *Acta Cryst.* **D56**, 959-964.

-
- Yeates, T. O.** (1997). Detecting and overcoming crystal twinning. *Methods Enzymol.* **276**, 344–358.
- Yeates, T. O. & Fam, B. C.** (1999). Protein crystals and their evil twins. *Structure Fold. Des.* **7**, R25-R29.
- Yatsunyk, L., Carducci, M., & Walker, F.** (2003). Low-spin ferriheme models of the cytochromes: Correlation of molecular structure with EPR spectral type. *J. Am. Chem. Soc.* **125**, 15986-16005.
- Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N. & Adams, P. D.** (2008). Surprises and pitfalls arising from (pseudo)symmetry. *Acta. Cryst.* **D64**, 99-107.

9 Publications

Publications

Heitmann, D. & Einsle, O. (2005). Structural and biochemical characterization of DHC2, a novel diheme cytochrome *c* from *Geobacter sulfurreducens*. *Biochemistry* **44**, 12411-12419.

Heitmann, D. & Einsle, O. (2008). Pseudo-merohedral twinning in crystals of the dihaem *c*-type cytochrome DHC2 from *Geobacter sulfurreducens*. *Acta Cryst.* **D64**, 993-999.

Heitmann, D. & Einsle, O. Atomic resolution structure and elimination of pseudo-merohedral twinning in dihaem *c*-type cytochrome DHC2 from *Geobacter sulfurreducens*. (In preparation).

Conference Presentations

Heitmann, D. & Einsle O. (2004). A novel type of cytochrome *c* from *Geobacter sulfurreducens*. Poster presentation at the 7th European Biological Inorganic Chemistry Conference (EUROBIC 7), Garmisch-Partenkirchen, Germany.

Heitmann, D. & Einsle O. (2005). DHC2 - A novel diheme cytochrome *c* from *Geobacter sulfurreducens*. Poster presentation at the VAAM Jahrestagung, Göttingen, Germany.

Heitmann, D. & Einsle O. (2005). DHC2 – A multiheme cytochrome *c* model system. Oral presentation at the 8th Heart of Europe Crystallography meeting (HEC 2005), Karlovy Vary, Czech Republic.

Heitmann, D. & Einsle O. (2007). Engineering the diheme cytochrome DHC2 into a model system for heme stacking interactions. Poster presentation at the VAAM Jahrestagung, Osnabrück, Germany.

Heitmann, D., Lü, W., Andrade, S., & Einsle, O. (2007). Structural and functional analysis of MFS transporters of the NNP family. Poser presentation at the New Methods in Membrane Protein Research EMBO workshop, Stockholm, Sweden.

10 *Curriculum vitae*

Personal data

Name	Daniel Heitmann
Address	Friedländer Weg 17 37085 Göttingen Germany
Birth date	13.03.1980
Birthplace	Paderborn, Germany
Nationality	German
Marital status	Single

Education

2005-2008	Ph.D. studies at the Department of Molecular Structural biology, Institute of Microbiology and Genetics, Georg-August-Universität of Göttingen, Germany
2004-2005	Diploma Thesis 'Präparation und Charakterisierung neuartiger <i>c</i> -Typ Cytochrome' at the Department of Molecular Structural biology, Institute of Microbiology and Genetics, Georg-August-Universität of Göttingen, Germany
2000-2004	Studies of Biology at Georg-August-Universität of Göttingen, Germany
1999-2000	Civilian service at the Bureau of Deutsche Pfadfinderschaft St. Georg Diözesanverband, Paderborn
1999	Abitur, Gymnasium Theodorianum, Paderborn