J Struct Func Genom (2006) 7:37–50
DOI 10.1007/s10969-006-9008-x

ORIGINAL PAPER

Crystal structure of the vitamin B_{12} biosynthetic cobaltochelatase, $CbiX^S$, from *Archaeoglobus fulgidus*

Jiang Yin · Linda X. Xu · Maia M. Cherney · Evelyne Raux-Deery · Amanda A. Bindley · Alexei Savchenko · John R. Walker · Marianne E. Cuff · Martin J. Warren · Michael N. G. James

Received: 10 February 2006 / Accepted: 20 April 2006 / Published online: 12 July 2006

Abstract The *Archaeoglobus fulgidus* gene af0721 encodes $CbiX^S$, a small cobaltochelatase associated with the anaerobic biosynthesis of vitamin B_{12} (cobalamin). The protein was shown to have activity both in vivo and in vitro, catalyzing the insertion of Co^{2+} into sirohydrochlorin. The structure of $CbiX^S$ was determined in two different crystal forms and was shown to consist of a central mixed β-sheet flanked by four α-helices, one of which originates in the C-terminus of a neighboring molecule. $CbiX^S$ is about half the size of other Class II tetrapyrrole chelatases. The overall topography of $CbiX^S$ exhibits substantial resemblance to both the N- and C-terminal regions of several members of the Class II metal chelatases involved in tetrapyrrole biosynthesis. Two histidines (His10 and His74), are in similar positions as the catalytic histidine residues in the anaerobic cobaltochelatase CbiK (His145 and His207). In light of the hypothesis that suggests the larger chelatases evolved via gene duplication and fusion from a $CbiX^S$-like enzyme, the structure of AF0721 may represent that of an ''ancestral'' precursor of class II metal chelatases.

Keywords Cobalamin (vitamin B12) biosynthesis · Tetrapyrrole · CbiK · CbiX · Cobaltochelatase · Protein structure evolution

Electronic Supplementary Material Supplementary material is available to authorised users in the online version of this article at http://www.dx.doi.org/10.1007/s10969-006-9008-x.

J. Yin · M. M. Cherney · M. N. G. James (✉)
Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, AB, Canada, T6G 2H7
e-mail: jyin@ualberta.ca

L. X. Xu · A. Savchenko · J. R. Walker
Ontario Centre for Structural Proteomics, C.H. Best Institute, University of Toronto, Toronto, ON, Canada, M5G 1L6

E. Raux-Deery · A. A. Bindley · M. J. Warren
Department of Biosciences, University of Kent, Canterbury CT2 7NJ, United Kingdom

M. E. Cuff
Structural Biology Center & Midwest Center for Structural Genomics, Biosciences Division Argonne National Laboratory, Argonne, IL 60439, USA

M. N. G. James
Alberta Synchrotron Institute, University of Alberta, Edmonton, AB, Canada, T6G 2E1

Introduction

Modified tetrapyrroles, in the form of metallo-prosthetic groups, play a variety of roles in such essential biochemical processes as photosynthesis, oxygen and electron transport, methanogenesis, complex molecular rearrangements and nitrogen and sulfur assimilation. Chlorophyll, heme, coenzyme F_{430}, cobalamin (vitamin B_{12}), and siroheme are all end products of a multi-branched biosynthetic pathway, along which their macrocyclic framework is modified and chelated with a distinct metal ion (Fig. 1). In vivo, the insertion of the proper metal ion into the corresponding tetrapyrrole is catalyzed by a branch-specific metal-chelating enzyme [1, 2].

In the anaerobic biosynthesis of cobalamin (vitamin B_{12}), cobalt is inserted into sirohydrochlorin to give cobalt–sirohydrochlorin by a Class II chelatase that

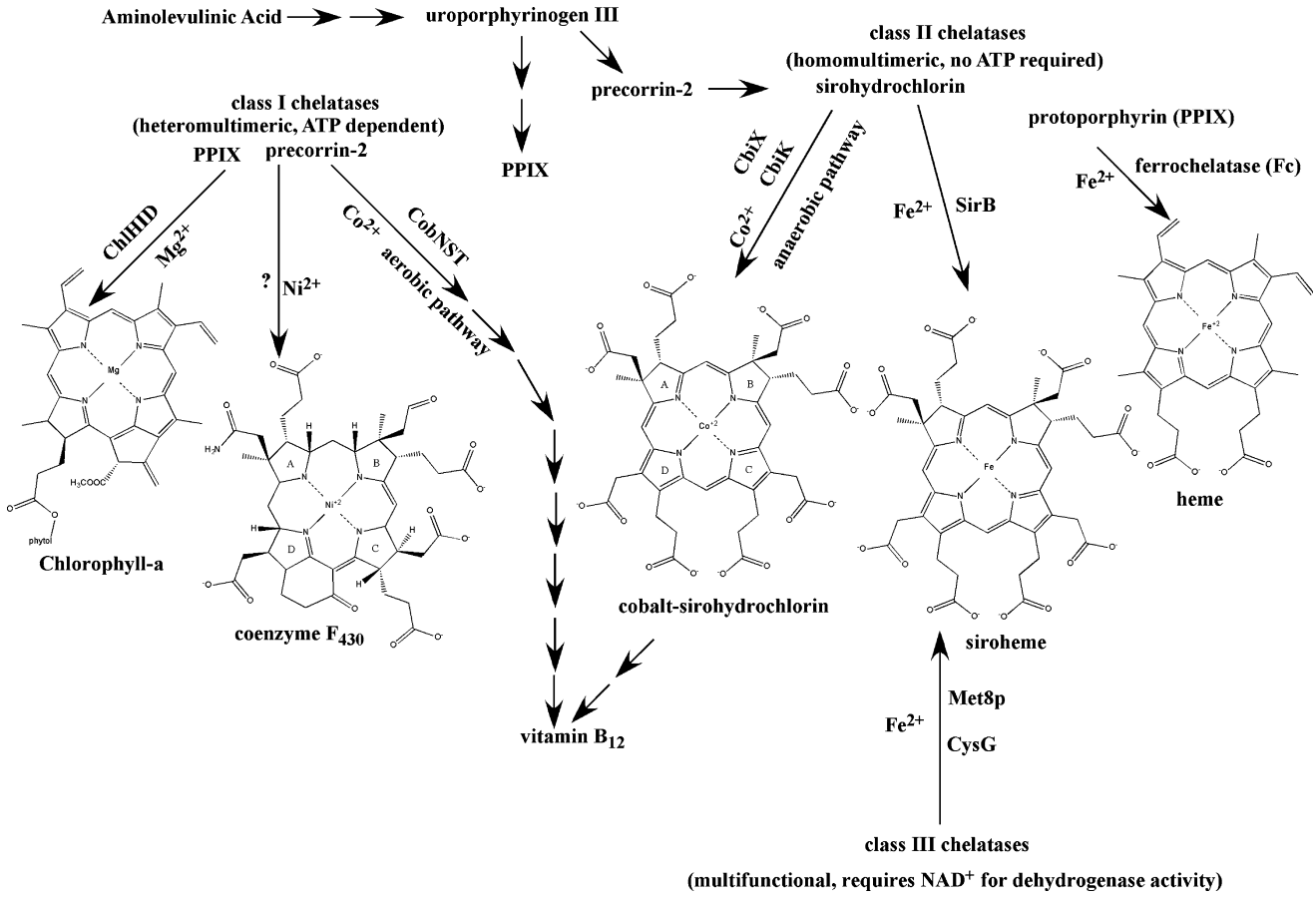

Fig. 1 A schematic view of the multi-branched biosynthetic pathways of tetrapyrrole molecules

was identified initially as CbiK [3]. The subsequent structure determination of CbiK revealed that it has a topology resembling that of the ferrochelatase (HemH) of heme biosynthesis [4]. The similarity in three dimensional structure between these two enzymes was not reflected in their amino acid sequences, as the two proteins have only approximately 10% sequence identity. However, it seems that the two enzymes may have been derived from a common ancestor and that Nature used the same structural blueprint in different branches of the tetrapyrrole biosynthetic pathway. Further studies have identified several other Class II chelatases including $CbiX^L$ (long version of CbiX) and $CbiX^S$ (short version of CbiX), cobaltochelatases of vitamin B_{12} biosynthesis, as well as the sirohydrochlorin ferrochelatase SirB of siroheme synthesis. These proteins display a low but significant level of similarity with CbiK, again indicating a likely common evolutionary origin. $CbiX^L$ and SirB share approximately 60% similarity but the $CbiX^L$ proteins contain a C-terminal extension that includes both a histidine-rich region and an Fe–S center. The length of $CbiX^S$ proteins typically ranges from ~100 to 130 amino acids, which is about half the size of the long versions of CbiX. Significantly, the primary structures of $CbiX^S$s align to both the N- and C-terminal regions of $CbiX^L$, SirB and, to a lesser extend, CbiK, indicating that the longer proteins are likely to have arisen from gene duplication and fusion events involving *cbiX*^{S} [5]. Some, but not all, $CbiX^S$ proteins contain a histidine-rich region but none are thought to harbor any Fe–S centers.

Class II chelatases are defined as metal-inserting enzymes that are homomeric and do not require ATP for activity. The metal chelation reactions catalyzed by these enzymes all encompass the following basic steps: the puckering of the pyrrole ring(s), metal ion insertion, and deprotonation of the pyrrole nitrogens. The structures of several chelatases of this class have been reported, including the protoporphyrin ferrochelatase (HemH) from *Bacillus substilis* (bFc), human (hFc) and yeast (yFc), as well as the cobaltochelatase (CbiK) from *Salmonella enterica* [4, 6–8]. As alluded to earlier, the sequence identity among these chelatases is low, yet they share extensive structural similarity and spatial arrangement of active site residues. The CbiX

family of proteins make up a large portion of Class II chelatases identified to date and are widely found in bacteria and archaea; their coding sequences usually abut on or lie within the *cob* (cobalamin biosynthesis) operon.

In this paper we report the structure and enzymatic activity of a $CbiX^S$ (AF0721) from *Archaeoglobus fulgidus* (*A. fulgidus*), a sulfur-reducing archeon whose ability to carry out anaerobic biosynthesis of tetrapyrroles is likely to be fundamental to its livelihood. Although AF0721 shares only ~20% identity in primary structure to any known or predicted CbiX protein, our biochemical studies have shown that AF0721 can function as a metal chelatase both in vivo and in vitro. We have also solved the structure of the *A. fulgidus* $CbiX^S$ by X-ray crystallography and suggest that this single-domain structure represents the ''original'' chelatase module.

Materials and methods

Crystallization sample preparation

The ORF of af0721 was amplified by PCR from *Archaeoglobus fulgidus* genomic DNA. The PCR product was cloned into the *Nde*I and *Bam*HI sites of a modified form of pET15b (Novagen) in which a TEV protease cleavage site replaced the thrombin cleavage site and a double stop codon was introduced downstream from the *Bam*HI site. This construct provides for an N-terminal hexahistidine tag separated from the protein by a TEV protease cleavage site (ENLYFQ↓G). The fusion protein was overexpressed in *E. coli* BL21-Star(DE3) (Stratagene) harboring an extra plasmid encoding three rare tRNAs (AGG and AGA for Arg, and ATA for Ile). The cells were grown in LB at 37°C to an OD_{600}~0.6 and the expression induced with 0.4 mM IPTG. After addition of IPTG, the cells were incubated with shaking at 15°C overnight. The cells were harvested, resuspended in binding buffer (500 mM NaCl, 5% Glycerol, 50 mM HEPES pH 7.5, 5 mM imidazole), flash-frozen in liquid N_2 and stored at –70°C. Thawed cells were lysed by sonication in the presence of 0.5% NP-40 and 1 mM each of PMSF and benzamidine. The lysate was clarified via centrifugation (30 min at 17,000 rpm, Beckman Coulter Avanti J-25 centrifuge) and passing through a DE52 column pre-equilibrated in binding buffer. The flow-through fraction was then applied to a metal chelating affinity-column charged with Ni^{2+}. After the column was washed, the protein was eluted from the column in elution buffer (binding buffer with 500 mM imidazole).

The AF0721 protein was dialyzed against 10 mM HEPES pH 7.5, 500 mM NaCl, and concentrated using a BioMax concentrator (Millipore). Before crystallization, any particulate matter was removed from the sample by passage through a 0.2 μm Ultrafree-MC centrifugal filtration device (Millipore).

Selenomethionine(SeMet)-substituted AF0721 was expressed using the same vector and host strain but in supplemented M9 media [9]. The sample was prepared under the same conditions as the native protein except for the addition of 5 mM β-mercaptoethanol to the purification buffers.

Subcloning the *A. fulgidus* $cbiX^S$ gene into pET3a

The modified pET15baf0721 construct was digested with *Nde*I and *Bam*HI and the resulting 399 bp gene fragment was ligated into pET3a (Novagen), which was also digested with *Nde*I and *Bam*HI.

In vivo complementation assay

For complementation studies, the *A. fulgidus* $cbiX^S$ was cloned into a plasmid containing a *tac* promoter. This was achieved by subcloning the af0721 gene from the modified pET15b construct, into pETac that had been linearized with *Nco*I and *Hind*III so that the gene was under the control of the *tac* instead of the T7 promoter. Complementation studies were performed in strain ER247, *E. coli* 302Δa (pCIQ-Ptac-*Pseudomonas denitrificans cobA*), which was transformed with each of the 3 different plasmids, pETac (negative control), pETac-AF0721 and pETac-Sc*MET8* (positive control).

Cobalt binding of pET3aaf0721 protein product

The non His-tagged protein was overexpressed in the same cell line and under the same conditions as the fusion protein. The cells were harvested and resuspended in binding buffer (500 mM NaCl, 5% glycerol, 50 mM HEPES pH 8, 5 mM imidazole). Cells were lysed by sonication and the cleared lysate was applied to a chelating Sepharose-column charged with Co^{2+} (AmershamBiosciences). After the column was washed, the protein was eluted in binding buffer containing 250 mM imidazole.

In vitro assay of chelatase activity

Chelatase assays were performed as described previously [10] by monitoring the conversion of sirohydrochlorin into cobalt–sirohydrochlorin. Sirohydrochlorin

was generated in situ under an atmosphere of nitrogen in a glove box, with less than 2 ppm oxygen. The *A. fulgidus* CbiX^S was buffer exchanged via a PD10 column (Amersham Biosciences) into anaerobic 50 mM Tris–HCl pH 8.0 inside the glove box. The chelatase activity was measured with sirohydrochlorin (2.5 μM), Co^{2+} (20 μM) and 5 μg of *A. fulgidus* CbiX^S in an 1 ml reaction volume in 50 mM Tris–HCl pH 8.0, at 37°C. Initial rates were recorded on a Hewlett-Packard 8452A photodiode array spectrophotometer, and the assays were performed in triplicate.

Crystallization, data collection, phasing and refinement

The primary crystallization condition was determined by a sparse crystallization matrix (Hampton Research, Crystal Screen 1 and PEG/Ion Screen), at room temperature using the sitting drop vapor diffusion technique in 96-well plates. Each condition was modified by varying the pH and the concentration of the solutes. The best condition was obtained using hanging drops (2 μl protein: 2 μl precipitant ratio) in crystallization condition: 0.1 M HEPES pH 6.8, 1.9 M ammonium sulfate, 4% methylpentanediol (MPD) in 2–5 days at room temperature. For diffraction studies, the crystals were flash-cooled with the crystallization buffer plus 40% MPD as the cryoprotectant. Both the wild type and SeMet-labeled proteins were crystallized under the same condition.

Diffraction data were collected at 100 K at the beam line 19BM of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory (Table 1). The absorption peak and the rising inflection point were determined by calculating and plotting f′ and f′′ values against energy from the fluorescence spectrum [11]. The three-wavelength MAD data were collected from a SeMet substituted protein crystal using an inverse-beam strategy. All crystallographic data were measured with the custom-built 3 × 3 tiled CCD (charge-coupled device) detector [12] with a 210 × 210 mm^2 active area and fast duty cycle (~1.7 s). The experiment, data collection, and visualization were controlled with d*TREK [13] and all data were integrated and scaled with the program package HKL2000 [14].

Program SOLVE was used to locate the positions of the selenium atoms using the scaled data collected at Beamline 19BM [15]. The results from the lower resolution dataset (30–2.4 Å) showed 3 of the 4 Se atoms with occupancies greater than 60% (*Z*-score 22.77). The calculated phases were solvent flattened with RESOLVE and the resulting map was used for model building by arp/WARP [16]. The model of the entire polypeptide chain of Gly-AF0721 (as a result of TEV cleavage), with the exception of the N-terminal additional glycine and the seven C-terminal residues (126–132), was further refined through manual tracing and refinement cycles using programs XFit [17] and REFMAC5 [18]. The coordinates were then refined against the higher resolution data (20–2.0 Å) using REFMAC5 to obtain the final model (Table 1).

A second crystal form of AF0721 was obtained from a mother liquor containing 1.9 M ammonium phosphate pH 6.8 and 4% 1,6-hexanediol. The diffraction data from these new crystals were collected at the Beamline 8.3.1 at the Advanced Light Source (Berkeley, California). Processing of the data was performed using program package HKL2000 [14]. Molecular replacement using the structural model solved from datasets of the original crystal form was applied to obtain an initial model for further refinement. The final model was obtained after cycles of adjustment to the model in XFit [17], addition of waters and restrained refinement using the CNS package and REFMAC5 [18, 19]. The density for the additional glycine at the N-terminus of the protein is visible, consequently, this extra residue is built into the final model (Table 1).

Structural analysis and figure generation

Pair-wise structural alignment was carried out using program ALIGN [20]. Sequence alignment with accompanying secondary structure was generated using program ESPript [21]. Model graphics were made with program Pymol (www.pymol.org).

Structural modeling

Automated docking was performed using the coordinates of siroheme (PDB code, 1AOP), the *N*-methylmesoporphyrin (NMeP) molecule bound to bFc (PDB code, 1C1H) and the AF0721 dimer (crystal form II) as input for the program Autodock [22]. In all solutions, the tetrapyrrole macroring in siroheme or NMeP shifted toward the bottom of the substrate-binding cleft. The structure with the lowest overall energy was chosen as the representative solution.

Results and discussion

Protein purification and metal binding

The His-tagged version of *A. fulgidus* CbiX^S was purified to homogeneity by metal chelation chromatography.

Table 1 Crystallographic statistics of data collection and structure refinement

Crystal (PDB code)	Form I	Form I (1TJN)	Form II (2DJ5)
Space group	$I4_122$	$I4_122$	$P3_121$
a (Å)	102.28	102.46	62.72
b (Å)	102.28	102.46	62.72
c (Å)	60.0	59.84	150.61
α (°)	90	90	90
β (°)	90	90	90
γ (°)	90	90	120
No. molecule/asymmetric unit	1	1	2
Vm (Matthews' Coefficient)	2.6	2.6	2.8
Data collection	APS 19BM	APS 19BM	ALS Beamline 8.3.1
Wavelength (Å)	0.979240	0.979238	1.019943
Resolution range (Å)	30.0–2.40 (2.49–2.40)[a]	30.0–2.01 (2.08–2.01)	60.0–2.50 (2.59–2.50)
Unique reflections	6470 (629)	10043 (1030)	8637 (602)
I/σ (I)	8.0 (5.9)	6.1 (2.5)	12.8 (1.9)
Data completeness (%)	98.2 (94.2)	86.9 (88.7)	69.0 (48.9)
R_{merge}	0.07 (0.119)	0.078 (0.301)	0.06 (0.265)
Phasing			
Figure of merit (SOLVE)	0.58	0.36	N/A
Molecular replacement			
R_{free} (%)/Correlation (%)	N/A	N/A	45.5/63.2
Refinement	N/A		
No. reflections used		9264 (448)	7925 (409)
Resolution range		19.46–2.01 (2.06–2.01)	54.3–2.55 (2.62–2.55)
Free set size (%)		4.8	4.9
No. atoms		1083	2121
No. waters		75	55
R_{factor} (%)		21.7 (23.9)	21.2 (26.1)
R_{free} (%)		26.0 (24.0)	26.7 (30.6)
Mean B value		29.9	50.7
r.m.s.d. from ideal geometry	N/A		
Bond length (Å)		0.010	0.014
Bond angle (°)		1.299	1.479
Ramachandran plot (%)		91.5/8.5[b]	86.4/13.6[b]

$R_{merge}=\Sigma_h\Sigma_j|I_{hj}-<I_h>|/\Sigma_h\Sigma_j I_{hj}$, where $<I_h>$ is the weighted mean intensity of the symmetry-related reflections I_{hj}

$R_{working}=\Sigma_h||F_o|-|F_c||/\Sigma_h F_o$, where $|F_o|$ and $|F_c|$ represent the observed and calculated structure factor amplitudes, respectively. R_{free} is $R_{working}$ calculated with the reference set

[a]Parentheses indicate values for the highest resolution shell

[b]Shown in the order of most favored region/additionally allowed region

Analytical gel filtration HPLC studies suggested that the protein (before or after TEV cleavage) exists primarily as multimers (data not shown). Significantly, the non-tagged version of the *A. fulgidus* $CbiX^S$, produced from the pET3aaf0721 construct, could also be purified by metal chelation chromatography. The *A. fulgidus* $CbiX^S$ amino acid sequence contains four histidine residues, two of which (His10 and His74) align with the proposed catalytic histidine residues for other $CbiX^S$ proteins. The non-tagged protein also showed a strong tendency to aggregate in solution.

In vitro chelatase activity of the *A. fulgidus* $CbiX^S$

Both the tagged and non-tagged versions of $CbiX^S$ were analyzed for cobaltochelatase activity in vitro and both were found to be capable of catalyzing the insertion of cobalt into sirohydrochlorin, yielding cobalt–sirohydrochlorin. The His-tagged protein was found to have a specific activity of 17 nmol/min/mg, whereas the non His-tagged protein had activity just above background at 4 nmol/min/mg. The activity is much lower than that recorded for the $CbiX^S$ from *Methanosarcina barkeri*, however, it is similar to that reported for the $CbiX^S$ from *Methanobacter thermoautotrophicum* [5]. The discrepancy between the two *A. fulgidus* $CbiX^S$ protein products may well be related to the predominant aggregated state of the non-His-tagged version.

In vivo chelatase activity

Other cobaltochelatases have been observed to complement an *E. coli* strain deficient in sirohydrochlorin ferrochelatase activity, where they accept ferrous iron

as their metal substrate instead of cobalt. The ability of CbiX^S to act as such a chelatase in vivo was therefore investigated. Strain ER247, *E. coli* 302Δ a (pCIQ-Ptac-*Pseudomonas denitrificans cobA*) was transformed with 3 different plasmids, pETac (negative control), pETac-AF0721 and pETac-ScMET8 (positive control). In order to study whether the cysteine auxotrophy due to the lack of siroheme biosynthesis can be overcome by adding a sirohydrochlorin–ferrochelatase, the new strains were restreaked on minimum media with and without cysteine. The strain with pETac grew only on medium containing cysteine whereas the MET8 construct grew well in the presence and absence of cysteine. Finally, the construct harboring the *A. fulgidus cbiX*^{S} produced only about 20 colonies without cysteine, in comparison to the 2,000 or so colonies that were observed in the presence of cysteine. The revertant-like colonies on the cysteine-minus plate were not observed in any of the control experiments. The plasmid DNA from these revertant-like colonies was extracted, sequenced and found to contain the AF0721 coding sequence identical to that in the input plasmid. When these retrieved plasmids were used to transform strain ER247, the same complementation pattern was observed in that only comparatively few colonies appeared to grow in the absence of cysteine. However, when the revertant-like colonies were replated on cysteine-minus plates, a normal number (i.e., in excess of 2,000) of colonies were observed, indicating full recovery of siroheme production in the bacteria. These data suggest that the *A. fulgidus cbiX*^{S} can minimally complement the *E. coli* deficiency for a chelatase and that some genomic change might have taken place to allow for successful complementation. Such a change could perhaps involve mutations in the proteins involved in the delivery of iron to the cobaltochelatase.

Monomer structure

The monomer structure of AF0721 contains four parallel β-strands, one anti-parallel β-strand and four α-helices (Fig. 2A). The parallel β-strands (β1, β2, β3 and β5) display an average twist of ~60 degrees in the direction of polypeptide chain, forming a half β-barrel together with the anti-parallel strand β4. This mixed sheet is arranged in the sequence of β2-β1-β3-β5-β4, with the α-helices alternating with the β-strands and one (α4) following β5. The C-terminal helix α4 is composed of mainly hydrophobic residues (9 aliphatic/aromatic out of 12 non-charged) (Fig. 2C). Like an ''arm'' in a striking pose, α4 points away from the convex side of the β-sheet and forms little contact with the rest of the protein (Figs. 2, 3). The overall topology of AF0721 can be viewed as a special case of a Rossmann fold in which the typical six-stranded parallel β-sheet (β3-β2-β1-β4-β5-β6) is downsized to four parallel β-strands (β2-β1-β3-β5) with β4 being a variation on the connection between β3 and β5 [23]. All enzymes having Rossmann fold-like structures bind their ligands through the residues in the loops linking the α-helices and β-strands along the C-terminal edge of the β-sheet [24]. Two histidine residues (His10 and His74, in AF0721 numbering), invariant among CbiX, SirB and CbiK chelatases, are located in this region of the AF0721 structure (Figs. 1A, 2).

Structurally similar proteins

The AF0721 structures were compared to other protein structures using the Dali server [25]. The top hits include the protoporphyrin ferrochelatase from *Bacillus subtilis* (bFc, PDB codes 1AK1, *Z*-score 9.9, r.m.s.d. 2.8 Å over 114 C^α pairs, related PDB entries, 1C1H, 1ID3, 1DOZ, 1C9E, 1N0I), the anaerobic cobaltochelatase from *Salmonella enterica* (CbiK, PDB code 1QGO, *Z*-score 9.6, r.m.s.d. 3.3 Å over 112 C^α pairs), and the human protoporphyrin ferrochelatase (hFc, PDB code 1HRK, *Z*-score 9.0, r.m.s.d. 3.0 Å over 113 C^α pairs).

The larger chelatases contain two structurally homologous domains related by a two-fold pseudo-symmetry as shown by the alignments between these domains (Table 2). A rather large translation vector is needed in order to superimpose the two terminal domains in CbiK (16.3 Å in CbiK vs. <0.5 Å in ferrochelatases). This may not be a unique feature of the cobaltochelatases, since the two CbiX^S molecules forming the dimer observed in crystal form II are related by an almost perfect two-fold symmetry. Af0721 aligns well to both the N- and C-terminal domains of known class II metal chelatases, implying the evolutionary ancestry for *A. fulgidus*-like CbiX^S proteins (Table 2).

The anti-parallel β-strand β4 in CbiX^S has no counterpart in other known class II chelatase structures. β4, together with the connecting loop region from α3, interacts with α3 and β5 via several hydrogen bonds and a number of hydrophobic interactions. This likely amounts to an increased stability of the core β-sheet, which is achieved in CbiK, bFc, hFc and yFc by having longer α3 helices that form more contacts with their corresponding β5 strands (Fig. 3). Alternatively, this region of AF0721 could engage in protein–protein interaction. In the alignment between *A. fulgidus* CbiX^S and the N-terminal domains of eukaryotic protoporphyrin ferrochelatases, this region of CbiX^S structurally overlaps with the C-terminal tails in hFc

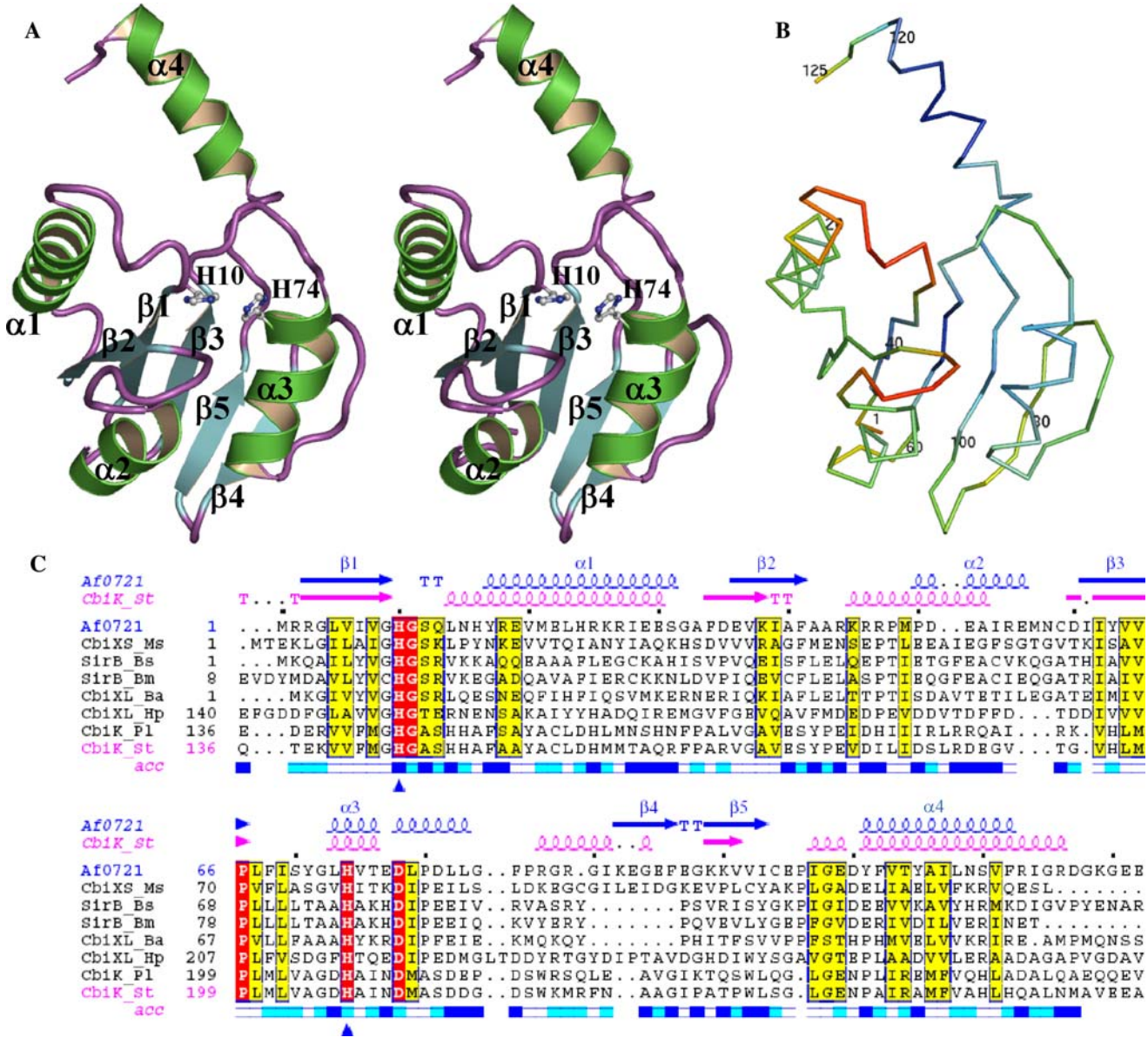

Fig. 2 Monomer structure of AF0721 and its alignment to several class II chelatases. (**A**) A stereo ribbon diagram of AF0721 monomer, the *β*-strands and helices are colored cyan and green, respectively. (**B**) The C^{α} trace of an AF0721 monomer colored in spectral colors according to the temperature factors of the C^{α} atoms. Dark blue and red represent the lowest and the highest B factors, respectively. (**C**) Sequence and structural alignment of AF0721 with CbiK, SirB, $CbiX^{S}$ and $CbiX^{L}$. The secondary structure of AF0721 and that of the C-terminal domain of CbiK are shown above in blue and magenta, respectively. Labels of secondary structure in AF0721 are the same as in **A**. Yellow background highlights similar sequences, whereas the red color denotes invariant residues. Blue triangles indicate the catalytic residues. acc, solvent accessibility calculated with CbiK coordinates, darker color meaning more exposed to solvent; T, turn. Ba, *Bacillus anthracis*; Bs, *Bacillus subtilis*; Hp, *Helicobacter pylori*; Ms, *Methanosarcina acetivorans*; Pl, *Photorhabdus luminescens*; St, *Salmonella typhimurium*

and yFc, which were also implicated to play a role in the dimerization of these enzymes [6, 8].

Quaternary structures and functional chelatase assembly

All previously known structures of class II metal chelatases are assembled from two similar modules not very different from the $CbiX^{S}$ basic design. The active site, where metallation occurs, is located at the interface between the two domains in a cleft composed of the residues along the C-terminal edges of the two opposing *β*-sheets. Following this stoichiometry, it is likely that the *A. fulgidus cbiX*^{S}, and possibly other $CbiX^{S}$ proteins, form their active site clefts via dimerization. This is vividly observed in Form I $CbiX^{S}$

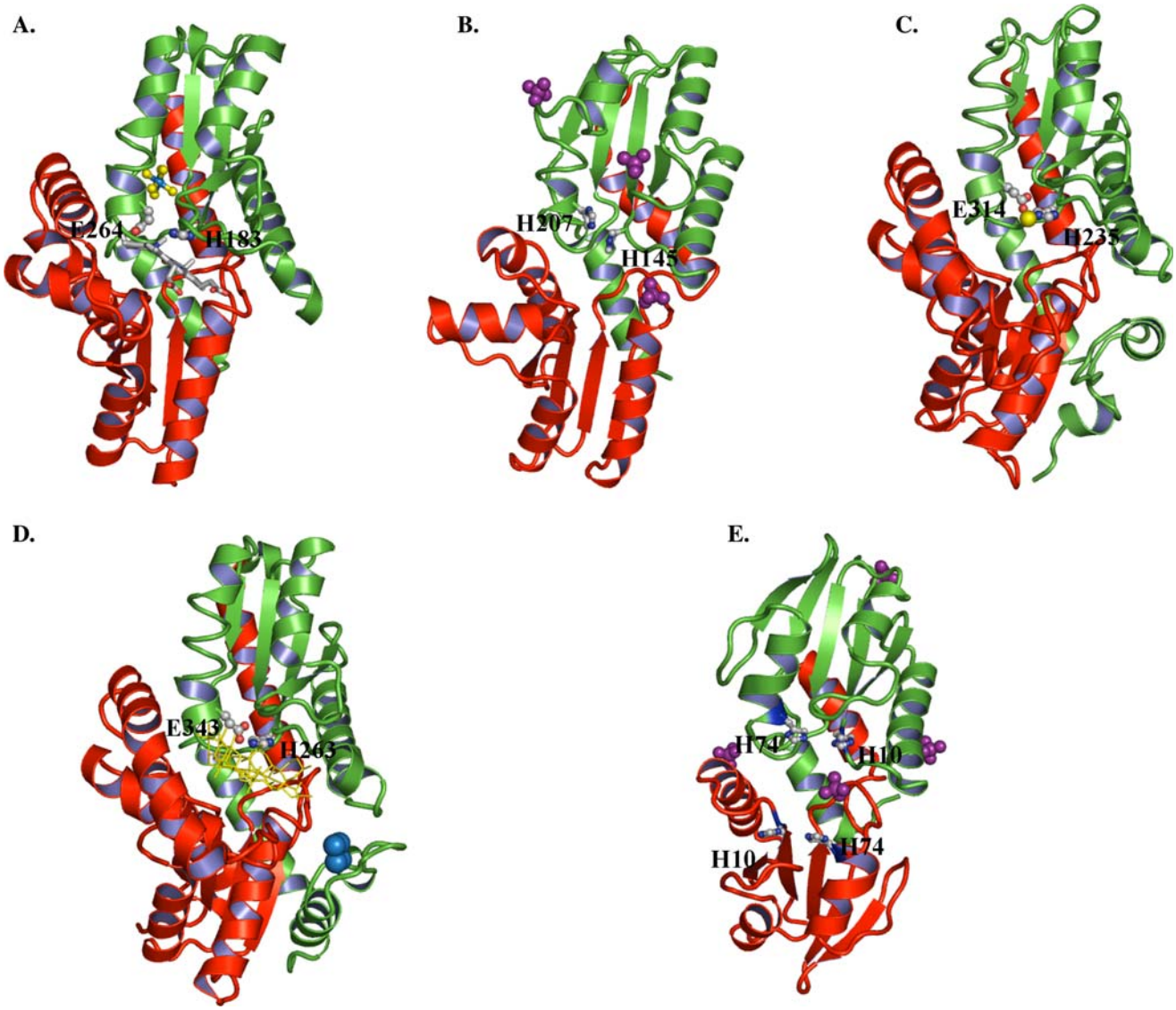

Fig. 3 Similar fold design among class II metal chelatases. The N- and C-terminal domains are shown in red and green, respectively. The catalytic residues are represented in ball and stick. (**A**) bFc. Gray sticks show a molecule of *N*-methylmesoporphyrin bound at the active site. Marine and yellow spheres depict the magnesium ion (Mg^{2+}) and 6 coordinating water oxygen atoms, respectively. (**B**) CbiK. Purple spheres represent three bound sulfate ions (SO_4^{2-}). (**C**) yFc. A cobalt ion bound to the catalytic residue His235 is shown in yellow. (**D**) hFc. The three cholate molecules bound in the substrate-binding cleft are shown in yellow lines. The C-terminal [Fe–S] center is shown in marine spheres. (**E**) AF0721 dimer in crystal form II. Four bound phosphate ions (HPO_4^{2-}) are shown as purple spheres

crystals, where two molecules of $CbiX^S$ interact with each other intimately over a crystallographic two-fold axis. The interface buries ~2049.9 $Å^2$ of accessible surface area per monomer (25% of the total). In this dimeric structure, α4, the hydrophobic residue-rich C-terminal helix, is swapped from a neighboring molecule to form extensive interactions with the dimer related helices α4, α1, and the ''back'' of the parallel β-sheet (Fig. 3). Hydrophobic interactions are the driving force behind dimerization; this avoids the otherwise energetically unfavorable exposure of many hydrophobic residues to solvent. In the Form II crystals, there are two $CbiX^S$ molecules per asymmetric unit. These two $CbiX^S$ molecules form a dimer in the same fashion as the crystallographic dimer observed in the $I4_122$ crystal form. Although the resolutions of the two crystal forms are quite different (2.0 Å and 2.6 Å for forms I and II, respectively), most of the inter-chain hydrogen bonds and hydrophobic interactions observed in the dimeric structure of crystal form II are also present in the crystallographic dimer structure of crystal form I (Table 3).

Table 2 Statistics of pair-wise alignment between the domains in class II metal chelatases and AF0721 (two crystal forms). The alignment was performed for all main chain atoms (N–C^{α}–C–O)

Fixed molecule	Aligned molecule	r.m.s.d./No. pairs	Rotation (°)	Translation (Å)
AF0721 ($I4_122$)	AF0721_A ($P3_121$)	0.55/459	N/A	N/A
AF0721 ($I4_122$)	AF0721_B ($P3_121$)	0.52/459	N/A	N/A
AF0721 ($I4_122$)	CbiK C-terminal	2.17/401	N/A	N/A
AF0721 ($I4_122$)	CbiK N-terminal	2.52/419	N/A	N/A
AF0721 ($I4_122$)	hFc C-terminal	2.76/433	N/A	N/A
AF0721 ($I4_122$)	hFc N-terminal	2.04/415	N/A	N/A
AF0721 ($I4_122$)	bFc C-terminal	2.29/414	N/A	N/A
AF0721 ($I4_122$)	bFc N-terminal	2.78/442	N/A	N/A
AF0721 ($I4_122$)	yFc C-terminal	2.33/408	N/A	N/A
AF0721 ($I4_122$)	yFc N-terminal	2.19/421	N/A	N/A
AF0721_A ($P3_121$)	AF0721_B ($P3_121$)	0.11/496	180	0.03
bFc N-terminal	bFc C-terminal	2.02/472	180	0.39
yFc N-terminal	yFc C-terminal	2.96/464	170	0.07
hFc N-terminal	hFc C-terminal	2.65/482	178	0.14
CbiK N-terminal	CbiK C-terminal	3.15/382	178	16.3

Table 3 Interactions in AF0721 quaternary structures. (A) Number of interactions. (B) Hydrogen bonds at the crystallographic dimer–dimer interfaces of three AF0721 tetramers

(A)

	Crystal	Total No. of interactions	No. of hydrogen bonds
AF0721 dimer	Form I	506	232
AF0721 dimer	Form II	446	219
Back-to-back tetramer	Form I	1432	468
Side-by-side tetramer	Form I	1250	474
AF0721 tetramer	Form II	1208	442

(B)

Back-to-back tetramer (crystal form I)	A:Y115O$^{\eta}$–D:N119O$^{\delta1}$[a]	2.77[b]
	B:Y115O$^{\eta}$–C:N119O$^{\delta1}$	2.77
	C:Y115O$^{\eta}$–B:N119O$^{\delta1}$	2.77
	D:Y115O$^{\eta}$–A:N119O$^{\delta1}$	2.77
Side-by-side tetramer (crystal form I)	A:R87N$^{\eta1}$–C:E56O$^{\varepsilon1}$	3.07
	A:R87N$^{\eta2}$–C:E56O$^{\varepsilon1}$	3.06
	B:R46N$^{\varepsilon}$–C:Q13O	2.82
	B:R47N$^{\eta1}$–D:78O$^{\delta1}$	2.93
	B:R47N$^{\eta2}$–D:78DO$^{\delta2}$	2.73
	C:R47N$^{\eta1}$–A:D78O$^{\delta1}$	2.93
	C:R47N$^{\eta2}$–A:D78O$^{\delta2}$	2.73
	C:R46N$^{\varepsilon}$–B:Q13O	2.82
	D:R87N$^{\eta1}$–B:E56O$^{\varepsilon1}$	3.07
	D:R87N$^{\eta2}$–B:E56O$^{\varepsilon1}$	3.06
af0721 tetramer (crystal from II)	A:Y115O$^{\eta}$–C:N119O$^{\delta1}$	2.57
	B:Y115O$^{\eta}$–D:N119O$^{\delta1}$	2.55
	C:Y115O$^{\eta}$–A:N119O$^{\delta1}$	2.57
	D:Y115O$^{\eta}$–B:N119O$^{\delta1}$	2.55

[a]Chain:ResidueNumerDonatorAtom–Chain:ResidueNumerAcceptorAtom

[b]Distances in Å

The $CbiX^S$ may form a higher level of quaternary structure. In Form I $CbiX^S$ crystals, two adjacent dimers form a tetramer in a ''back-to-back'' style, resulting in a further reduction in accessible surface area of 754.3 $Å^2$/monomer (Fig. 4C). Within the tetramer, the interfaces are formed chiefly by packing the four C-terminal helices α4 into a helix bundle, plus contributions from helices α1 and the residues in the loop immediately following α1. Most of the interactions are hydrophobic, especially those involving residues in helix α4, with Asn119 ($O^{\delta1}$) and Tyr115 (O^{η}) also adding four fairly strong hydrogen bonds (2.8 Å) between subunits B–C and A–D. An alternative way of forming a tetramer is the ''side-by-side'' arrangement

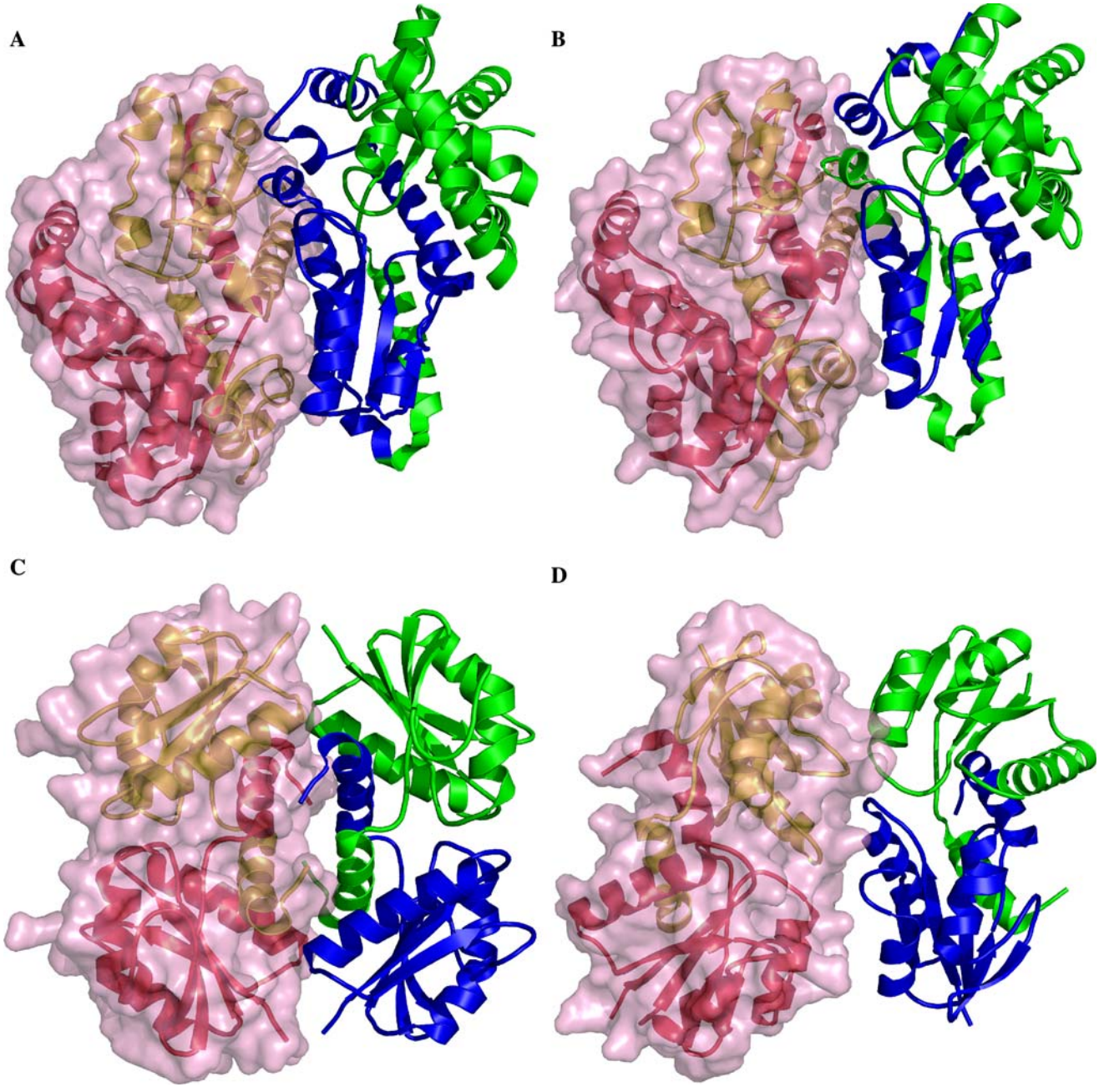

Fig. 4 Comparison between tetrameric structures of AF0721 and the quaternary structures observed in human and yeast ferrochelatase crystals. (**A**) and (**B**) hFc and yFc dimers, respectively. Red, N-terminal domain of molecule **A**; brown, C-terminal domain of molecule **A**; green, N-terminal domain of molecule **B**; blue, C-terminal domain of molecule **B**. Crystallographic tetramer of AF0721, ''back-to-back'' assembly of dimers in **C** and ''side-by-side'' assembly in **D**. The colors red, yellow, green, blue, were used to show molecules **A**, **B**, **C** and **D** of the tetramer, respectively

of two AF0721 dimers (Fig. 4D). This tetrameric association mirrors the dimeric structures observed in hFc and yFc protein crystals (one dimer per asymmetric unit) [6, 8]. The additionally buried accessible surface area in this form of tetramer is 176.2 $Å^2$ for molecule A or D and 526.1 $Å^2$ for molecule B or C, with the total reduction less than half of that for the ''back-to-back'' tetramer. There are only three quaternary interfaces in this tetrameric structure, because molecule A and C are too far apart from each other to form any contacts. The interfaces between molecules A, D and molecules B, C are symmetrically established through the interactions between their $\alpha 2$ and $\alpha 3$ helices. Molecules B and D mainly form contacts via the loop regions connecting $\beta 1$ to $\alpha 1$ and $\beta 2$ to $\alpha 2$. Although the ''side-by-side'' tetramer contains 6 more

inter-chain hydrogen bonds, the ''back-to-back'' tetramer is stabilized by over 170 more van der Waals interactions than the ''side-by-side'' tetramer. This suggests that the ''back-to-back'' tetramer may be the more stable quaternary structure in nature. In accordance to that, only the ''back-to-back'' tetramer is observed (along a crystallographic two-fold axis) in crystal form II (Table 3).

Chelatase function: substrates binding prospects

A significant difference between the structures of the two AF0721 crystal forms is the placement of the two loop regions that form part of the substrate-binding cleft: those loops connecting β1 to α1 and β2 to α2. When aligned to the monomeric structure in crystal form I, these loops in the dimeric structure of crystal form II are the regions showing the greatest variation in atom-to-atom distances, indicating higher inherent flexibility. This is further evidenced by the relatively larger temperature factors associated with the residues in these loop regions (Fig. 2B and supplemental Fig. 1). The variation in these regions causes a widening of the outer rims of the substrate-binding cleft in the AF0721 dimeric structure. This is meaningful, since local structural rearrangements seem inevitable in the crystallographic dimer of crystal form I in order to accommodate the sirohydrochlorin substrate. In that dimeric structure, the side chains of the two Leu14 residues effectively block access of the tetrapyrrole substrate to the catalytic histidine residues due to the inherent two-fold symmetry of the active site (Fig. 5A). This structural flexibility near the substrate-binding cleft may represent a mechanism to regulate the enzymatic activity of CbiX^S.

Because our efforts to obtain a crystal of AF0721-substrate/ion complex came to no avail, we used computer docking tools to analyze the potential interactions between an AF0721 dimer and a corresponding substrate. The coordinates of crystal form II (amino acid coordinates only) and siroheme (PDB code, 1AOP, the ferrous ion was stripped to mimic the substrate sirohydrochlorin instead of the product siroheme) were input into the program AutoDock to obtain possible interaction modes. The best solutions (lowest overall energy) cluster in a single conformational species in which the C and D rings of sirohydrochlorin are buried deep inside the substrate binding cleft, which agrees nicely with the fact that both the A and B rings contain a bulkier 4° substituted carbon atom (see supplemental Table S1, Figures S2 and S3). This is contrary to what was observed in the crystal structure of bFc with a substrate-analog inhibitor *N*-methylprotoporphyrin (NMeP) bound, in which rings C and D, the only two with charged sidegroups, are protruding from the cleft into solvent [26]. The modeled AF0721–sirohydrochlorin complex is stabilized by a large number of polar and van der Waals interactions, comparable to those observed in bFc–NMeP crystal structure (Fig. 5A and Table 4A). His74 and the residues in its vicinity, Arg44, Ile69, Ser70, Gly72, and Val75, provide the major polar interactions to the carboxyl oxygens of sirohydrochlorin. Most of the hydrogen bonds are donated by main chain nitrogen atoms, with the O^{γ} atom of Ser70 also participating. Additionally, these residues, together with residues Gly11, Ser12, Leu14, and Phe68, form a number of van der Waals interactions with the modeled substrate. Docking a sirohydrochlorin molecule into the active site in CbiK resulted in a similar solution (i.e., the orientation of sirohydrochlorin with respect to the active site) (data not shown), although the details of interactions between the enzyme and the modeled substrate were not necessarily reproduced.

In the modeled AF0721–sirohydrochlorin complex structure, the interactions involving the carboxyl group of ring D-derived propionic acid are very similar to those observed between the same residues and a phosphate ion bound in the substrate-binding cleft of the AF0721 dimer structure (crystal form II) (Fig. 5B and Table 4B). Coincidentally, a sulfate ion was observed in the crystal structure of CbiK occupying a similar site, where the oxygen atoms of the sulfate ion form hydrogen bonds to the residues equivalent to those of AF0721 that interact with the carboxyl group of the C-ring propionic acid of the sirohydrochlorin in the docked complex model (Fig. 5B and Table 4B). The interactions between these residues and the pyrrole ring-derived propionic acid were indeed observed in the modeled CbiK–sirohydrochlorin complex. Of note is that we obtained very similar docking results using NMeP instead of sirohydrochlorin as the ligand, suggesting a fundamental difference between bFc and cobaltochelatases in their ways of binding substrates.

The catalytic residues in bFc (Glu264 and His183) and yFc (His235) have been implicated to play major roles in metal binding and metal insertion [27, 28]. Structural superimpositions indicate that the catalytic residues in CbiK, His145 and His207, are the structurally equivalent residues to those in ferrochelatases. Strong structural conservation suggests common catalytic function; this is substantiated by the observations that mutations at these residues were often detrimental to the enzymatic activity in various class II chelatases [4, 27]. His74 and His10 in *A. fulgidus* CbiX^S, show a spatial orientation with respect to the core mixed

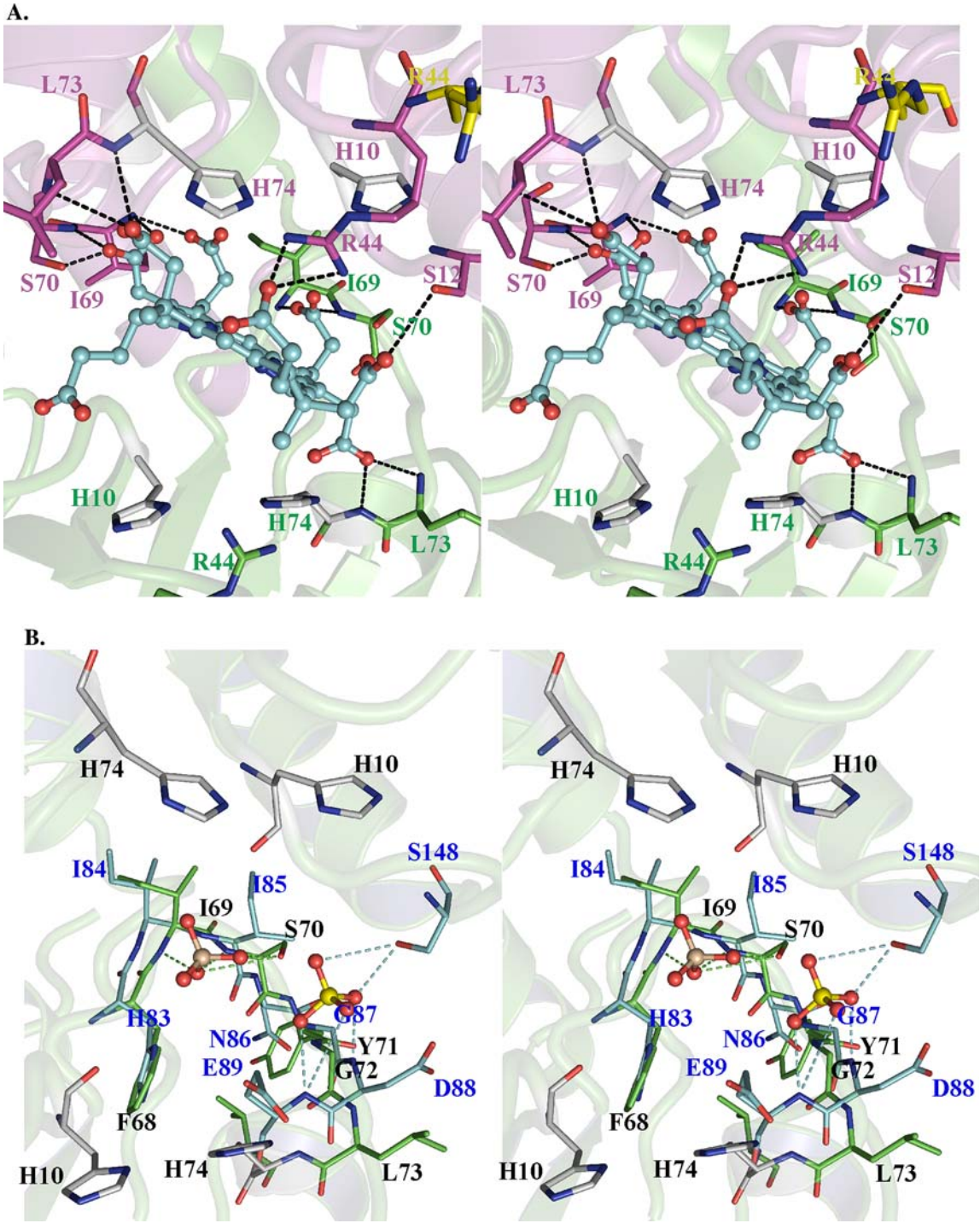

Fig. 5 Enzyme–substrate models of AF0721. (**A**) A modeled structure of AF0721 dimer–sirohydrochlorin complex. The components are distinguished according to the color of carbon atoms: purple and green, molecules **A** and **B** of AF0721 dimer (crystal form II), respectively; cyan, sirohydrochlorin; yellow, AF0721 (crystal form I). Residue labels are colored accordingly. For clarity, only the proposed hydrogen bonds between AF0721 and the carboxyl groups of sirohydrochlorin are shown as black dashed lines. The side chain of residue Arg44 from AF0721 (crystal form I) is also shown for comparison. (**B**) Aligned partial structures of AF0721 and CbiK showing similar interactions with bound sulfate/phosphate ions. The carbon atoms are shown in green (molecule **B**) and cyan for AF0721 dimer and CbiK, respectively. Phosphorous atom is shown in wheat and sulfur atom shown in yellow. Residues of AF0721 are labeled in black and those of CbiK in blue. Hydrogen bonds in AF0721 structure are shown in green and those in CbiK in cyan

Table 4 Tabulation of the interactions involving ligands in chelatase structures. (A) The interactions observed between NMeP and bFc (PDB code 1C1H) and those proposed between sirohydrochlorin (SRH) and AF0721 dimer (modeled complex). (B) The interactions between a bound HPO_4^{2-} ion and AF0721 are compared with those between a bound SO_4^{2-} ion and CbiK (PDB code 1AK1)

(A)

bFc–NMeP complex (1C1H)			AF0721–SRH complex (modeled)		
Hydrogen bond[a]	vdW interaction		Hydrogen bond	vdW interaction	
(B) N2–Y13O$^{\eta}$	Y13	30	(A) O3–A:S12O$^{\gamma}$	A:H10[b]	1
(C) O31–R31N$^{\eta 2}$	I29	10	(A) O1–A:R44N$^{\eta 1}$	A:S12	2
(C) O32–K188N$^{\zeta}$	R30	5	(A) O1–A:R44N$^{\eta 2}$	A:Y17	4
(D) O41–R33N$^{\eta 1}$	R31	2	(B) O4–A:H74N	A:R44	9
(D) O42–R33N$^{\eta 1}$	R33	5	(B) O4–A:L73N	A:L67	3
(D) O42–R33N$^{\eta 2}$	K87	2	(C) O1–A:I69N	A:F68	13
	F120	11	(C) O3–A:I69N	A:I69	15
	S121	4	(C) O4–A:S70N	A:S70	8
	H183	20	(C) O4–A:S70O$^{\gamma}$	A:G72	5
	L185	2	(D) O2–B:I69N	A:L73	6
	P186	2	(D) O2–B:S70N	A:H74	22
	K188	2	(D) O3–B:L73N	B:G11	12
	S222	1	(D) O3–B:H74N	B:S12	11
	G224	1		B:Q13	2
	T226	1		B:L14	7
	W230	2		B:Y17	3
	L263	1		B:R44	1
	E264	4		B:F68	2
				B:I69	7
				B:S70	8
				B:G72	3
				B:L73	3
				B:H74	19
	Total	105		Total	170

(B)

CbiK-SO_4^{2-} (PDB code 1QGO)			AF0721-HPO_4^{2-} (crystal form II)		
Hydrogen bond (Å)	vdW interaction		Hydrogen bond (Å)	vdW interaction	
O1–S148O$^{\gamma}$ (3.37)	I85	1	O1–B:I69N (2.55)	A:Y17	1
O3–D88N (3.27)	G87	3	O1–B:S70N (2.85)	B:F68	7
O3–E89N (2.89)	D88	9	O1–B:S70O$^{\gamma}$ (3.47)	B:I69	8
O3–D88N (2.78)	E89	4	O4–B:S70O$^{\gamma}$ (2.72)	B:S70	5
O4–S148O$^{\gamma}$ (2.79)	S148	5			
	H150	3			
	A151	1			
	Total	26		Total	21

[a]Letters in brackets are the names of the rings in tetrapyrrole molecules. O3 and O4 refer to the oxygens of the carboxylate group of the propionic group, whereas O1 and O2 refer to those of the acetate group. Hydrogen bonds are shown in the order of donator atom-acceptor atom

[b]In AF0721 dimer, chainName:ResidueNo

β-sheet structure that is very similar to that of Glu264 and His183 in bFc (Fig. 3). Moreover, the side chains of both His74 and His10 are very superimposable with those of their equivalent catalytic residues in other Class II chelatases, respectively (data not shown). Therefore, His74 and His10 of *A. fulgidus* CbiX^{S} are likely to perform functions in the metallation of tetrapyrrole molecules similar to those observed for Glu264 and His183 in bFc. Mutational analysis of the residues in the active site of AF0721 would not only identify the residues involved in metal chelation but also enhance our understanding of how the CbiX^{S}-like ''ancestor'' chelatase domains evolved into their modern descendants each performing a similar but highly specialized metal chelation reaction in vivo.

Acknowledgements Use of the Argonne National Laboratory Structural Biology Center beamlines (Beamline 19BM) at the Advanced Photon Source was supported by the US Department of Energy, Office of Energy Research, under Contract No. W-31-109-ENG-38. We wish to thank all members of the SBC at ANL for their help in conducting experiments. This work was supported by National Institutes of Health Grants GM62414-01, the

Ontario Research and Development Challenge Fund, grants from the Canadian Institutes of Health Research (CIHR), and the Biotechnology and Biological Sciences Research Council (BBSRC). X-ray diffraction data for crystal form II were collected at the beamline 8.3.1 of the Advanced Light Source (ALS) at Lawrence Berkeley Lab, under an agreement with the Alberta Synchrotron Institute (ASI). The ALS is operated by the Department of Energy and supported by the National Institute of Health. Beamline 8.3.1 is funded by the National Science Foundation, the University of California and Henry Wheeler. The ASI synchrotron access program is supported by grants from the Alberta Science and Research Authority (ASRA), the Alberta Heritage Foundation for Medical Research (AHFMR) and Western Economic Diversification (WED) of the Canadian Government. The authors would like to thank Drs. Ernst Bergmann (ASI), Jonathan Parish (ASI), and James Holton (ALS) for their help in data collection. MNGJ holds a Canada Research Chair in Protein Structure and Function. JY is a recipient of the Izaak Walton Killam Memorial postdoctoral fellowship at the University of Alberta.

References

1. Raux E, McVeigh T, Peters SE, Leustek T, Warren MJ (1999) Biochem J 338(Pt 3):701–708
2. Walker CJ, Willows RD (1997) Biochem J 327(Pt 2):321–333
3. Raux E, Thermes C, Heathcote P, Rambach A, Warren MJ (1997) J Bacteriol 179(10):3202–3212
4. Schubert HL, Raux E, Wilson KS, Warren MJ (1999) Biochemistry 38(33):10660–10669
5. Brindley AA, Raux E, Leech HK, Schubert HL, Warren MJ (2003) J Biol Chem 278(25): 22388–22395
6. Wu CK, Dailey HA, Rose JP, Burden A, Sellers VM, Wang BC (2001) Nat Struct Biol 8(2):156–160
7. Al-Karadaghi S, Hansson M, Nikonov S, Jonsson B, Hederstedt L (1997) Structure 5(11):1501–1510
8. Karlberg T, Lecerof D, Gora M, Silvegren G, Labbe-Bois R, Hansson M, Al-Karadaghi S (2002) Biochemistry 41(46): 13499–13506
9. van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) J Mol Biol 229(1):105–124
10. Leech HK, Raux E, McLean KJ, Munro AW, Robinson NJ, Borrelly GP, Malten M, Jahn D, Rigby SE, Heathcote P, Warren MJ (2003) J Biol Chem 278(43):41900–41907
11. Evans G, Pettifer RF (2001) J Appl Cryst 34:82–86
12. Westbrook EM, Naday I (1997) Methods Enzymol 276:244–268
13. Pflugrath JW (1999) Acta Crystallogr D Biol Crystallogr 55:1718–1725
14. Otwinowski Z, Minor W (1997) Methods Enzymol 276:307–326
15. Terwilliger TC, Berendzen J (1999) Acta Crystallogr D Biol Crystallogr 55(Pt 4):849–861
16. Morris RJ, Perrakis A, Lamzin VS (2003) Methods Enzymol 374:229–244
17. McRee DE (1999) J Struct Biol 125(2–3):156–165
18. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999) Acta Crystallogr D Biol Crystallogr 55(Pt 1):247–255
19. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Acta Crystallogr D Biol Crystallogr 54:905–921
20. Cohen GH (1997) J Appl Cryst 30:1160–1161
21. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) Bioinformatics 15(4):305–308
22. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) J Comp Chem 19:1639–1662
23. Rossmann MG, Moras D, Olsen KW (1974) Nature 250(463):194–199
24. Branden C, Tooze J (1991) In: Introduction to protein structure. Garland Pub., New York
25. Holm L, Sander C (1996) Science 273(5275):595–603
26. Lecerof D, Fodje M, Hansson A, Hansson M, Al-Karadaghi S (2000) J Mol Biol 297(1):221–232
27. Gora M, Grzybowska E, Rytka J, Labbe-Bois R (1996) J Biol Chem 271(20):11810–11816
28. Lecerof D, Fodje MN, Alvarez Leon R, Olsson U, Hansson A, Sigfridsson E, Ryde U, Hansson M, Al-Karadaghi S (2003) J Biol Inorg Chem 8(4):452–458