Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

The 1.4 Å resolution structure of *Paracoccus pantotrophus* pseudoazurin

Shabir Najmudin,*‡ Sofia R. Pauleta,* Isabel Moura and Maria J. Romão

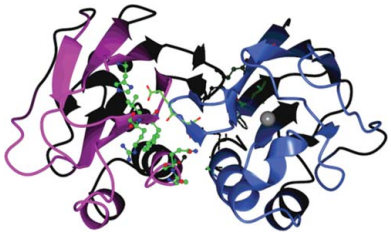
REQUIMTE, Centro de Química Fina e Biotecnologia, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

‡ Present address: CIISA, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal.

Correspondence e-mail: shabir@fmv.utl.pt, srp@dq.fct.unl.pt

Received 14 January 2010
Accepted 15 April 2010

PDB Reference: pseudoazurin, 3erx.



Pseudoazurins are small type 1 copper proteins that are involved in the flow of electrons between various electron donors and acceptors in the bacterial periplasm, mostly under denitrifying conditions. The previously determined structure of *Paracoccus pantotrophus* pseudoazurin in the oxidized form was improved to a nominal resolution of 1.4 Å, with R and R_{free} values of 0.188 and 0.206, respectively. This high-resolution structure makes it possible to analyze the interactions between the monomers and the solvent structure in detail. Analysis of the high-resolution structure revealed the structural regions that are responsible for monomer–monomer recognition during dimer formation and for protein–protein interaction and that are important for partner recognition. The pseudoazurin structure was compared with other structures of various type 1 copper proteins and these were grouped into families according to similarities in their secondary structure; this may be useful in the annotation of copper proteins in newly sequenced genomes and in the identification of novel copper proteins.

1. Introduction

The cupredoxins are small blue copper proteins (10–22 kDa) that are involved in various biochemical processes, usually involving single-electron transport, in a variety of plants and bacteria. These proteins have been classified as type 1 copper proteins on the basis of their spectroscopic properties and contain a single Cu atom as the only redox-active group, which cycles between cupric and cuprous states (Adman, 1991; Lappin, 1981; Han *et al.*, 1991).

In their oxidized state, type 1 copper proteins are characterized by an intense (3–5 mM^{-1} cm^{-1}) optical absorption at 590–630 nm, a distinctive small hyperfine coupling constant (Solomon, 2006; Canters & Gilardi, 1993) and an unusually high redox potential [184 mV for stellacyanin (Romero *et al.*, 1993) and 680 mV for rusticyanin (Sailasuta *et al.*, 1979)] compared with small cupric complexes (Kaim & Rall, 1996). Based on their EPR spectra, it is possible to divide type 1 blue copper proteins into two groups: the rhombic group (perturbed blue copper site) and the axial group (classic blue copper site). The rhombicity observed is also related to an increase in the 460 nm band in their visible spectrum (Solomon, 2006).

This class of copper proteins includes proteins isolated from plant and bacterial sources and can be divided into several groups according to sequence homology (De Rienzo *et al.*, 2000): (i) the plastocyanin family [plastocyanin (Guss *et al.*, 1992), amicyanin (Durley *et al.*, 1993), pseudoazurin (Williams *et al.*, 1995) and halocyanin (Scharf & Engelhard, 1993)]; (ii) azurin (Adman & Jensen, 1981); (iii) the phytocyanin family [stellacyanin group: classic stellacyanin (Hart *et al.*, 1996), umecyanin (Koch *et al.*, 2005) and mavicyanin (Xie *et al.*, 2005); plantacyanin group: basic blue protein (Guss *et al.*, 1996) and uclacyanin (Giri *et al.*, 2004; Nersissian *et al.*, 1998)]; (iv) auracyanin (Bond *et al.*, 2001) and (v) rusticyanin (Botuyan *et al.*, 1996; Walter *et al.*, 1996). Although there is an appreciable amount of divergence in the sequences of all these proteins, the copper ligand residues are conserved.

The three-dimensional structures of most of these type 1 blue copper proteins, with the exceptions of halocyanin and uclacyanin, have been determined using NMR and/or X-ray crystallography,

showing that the polypeptide chain shares a common fold of seven or eight core β-strands forming two β-sheets arranged in a complex double-wound Greek-key motif, with the type 1 copper centre buried at one end of this β-barrel. The Cu atom is coordinated in a distorted tetrahedral geometry by strong ligands that comprise the thiolate S atom of a cysteine residue and the imidazole N atoms of two histidine residues (Guss *et al.*, 1992). The fourth ligand is typically the S atom of a methionine residue, but other axial ligands have also been found. For example, in stellacyanins this position is occupied by the side chain of a glutamine (Hart *et al.*, 1996). The azurins are an exception since their copper centres adopt a trigonal bipyramidal geometry with a second weak axial ligand being provided by the backbone carbonyl O atom of a glycine residue (Adman & Jensen, 1981; Nar *et al.*, 1991).

Another common characteristic of the type 1 copper proteins is the location of three of the copper ligands (a histidine, a cysteine and a methionine/glutamine) on a single loop and of the N-terminal histidine in a loop anchored in the β-sandwich (Adman, 1991; Guss *et al.*, 1992). The C-terminal histidine ligand, which is located in the middle of the loop and on the surface of the protein, provides electronic coupling with the Cu atom (Canters *et al.*, 2000).

The pseudoazurin isolated from *Paracoccus pantotrophus* is a dimeric protein that is located in the periplasm of this organism and functions as an electron donor to several enzymes of the denitrification pathway, such as nitrite reductase (Sam *et al.*, 2008), nitrous oxide reductase (Berks *et al.*, 1993) and nitric oxide reductase (Thorndycroft *et al.*, 2007). This small cupredoxin can also donate electrons to other enzymes, such as cytochrome *c* peroxidase (Pauleta, Cooper *et al.*, 2004; Pauleta, Guerlesquin *et al.*, 2004), an enzyme that reduces hydrogen peroxide to water. The electron donor of pseudoazurin is believed to be the cytochrome bc_1 complex, but most electron-transfer studies have disregarded the electron donors.

The ability of this small redox protein to interact with a multitude of structurally different partners has been attributed to the hydrophobic character of the binding surface, which is centred at the electron-entry/exit point. This characteristic has been designated as pseudospecificity (Williams *et al.*, 1995) and has also been described for other periplasmic electron shuttles. Hydrophobic interactions would be involved in these complexes, as they are less specific than those involving salt bridges or hydrogen bonds. Nevertheless, the presence of charged side chains and even a dipole moment can play a role in the orientation of the two redox partners prior to collision and can also explain the different binding specificities of azurins and pseudoazurins, which have similar electron acceptors: blue and green nitrite reductases (Dodd *et al.*, 1998; Murphy *et al.*, 2002).

The structure of *P. pantotrophus* pseudoazurin has been determined at a resolution of 2.5 Å for the oxidized form by X-ray crystallography (Williams *et al.*, 1995; PDB code 1adw), while NMR spectroscopy was used to determine the structure of the reduced form (Thompson *et al.*, 2000).

Here, we report the structure of oxidized *P. pantotrophus* pseudoazurin at high resolution (1.4 Å). This structure was compared with those of all of the different type 1 copper proteins and the various members were subsequently grouped into families according to similarities in their secondary structure.

2. Materials and methods

2.1. Protein isolation and purification

P. pantotrophus pseudoazurin was heterologously expressed, isolated and purified as described previously (Pauleta, Guerlesquin *et al.*, 2004). The purified pseudoazurin had a purity ratio ($A_{590\,\mathrm{nm}}$/$A_{280\,\mathrm{nm}}$) of 0.6 and was stored in 10 m*M* HEPES, 10 m*M* NaCl pH 7.0 at 193 K until further use. The concentration of the protein was determined spectrophotometrically using the extinction coefficient at 590 nm of ε = 3.00 mM^{-1} cm^{-1} for the fully oxidized pseudoazurin. The pseudoazurin samples used in the crystallization assays were in the oxidized state.

Table 1
X-ray crystallography data-collection statistics.

Values in parentheses are for the lowest/highest (54.85–4.43/1.48–1.4) resolution shells. Note: although the data are limited to 1.4 Å resolution to reflect the true nature of the data, they extend to beyond 1.25 Å at the corners. These additional data have been included because they add to the overall quality of the electron-density maps and aid the refinement.

X-ray source	ID23-EH1, ESRF	
Wavelength (Å)	0.95370	
Space group	*C*2	
Unit-cell parameters		
a (Å)	108.82	
b (Å)	55.99	
c (Å)	67.98	
β (°)	126.22	
Resolution limits (Å)	54.88–1.4	54.88–1.25
No. of observations	213846 (6103/21105)	240220 (8730/8914)
No. of unique observations	62484 (2028/7108)	73293 (2863/4172)
Multiplicity	3.4 (3.0/3.0)	3.3 (3.0/2.1)
Completeness (%)	94.7 (93.8/74.2)	79.2 (94.6/31.2)
$\langle I/\sigma(I)\rangle$	12.6 (16.8/5.1)	16.3 (28.8/2.2)
Wilson *B* factor (Å^2)	14.5	13.2
R_{merge}†	6.3 (7.1/17.0)	5.6 (5.6/28.0)
$R_{\mathrm{p.i.m.}}$‡	4.2 (5.2/11.4)	4.3 (4.9/24.8)
PDB code	3erx	

† $R_{\mathrm{merge}} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. ‡ $R_{\mathrm{p.i.m.}} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ and is a measure of the quality of the data after averaging the multiple measurements.

2.2. Crystallization and data collection

Crystals of pseudoazurin were obtained using the conditions described by Williams *et al.* (1995), namely 3.0–3.2 *M* ammonium sulfate and 50 m*M* potassium phosphate pH 7.5, using the hanging-drop vapour-diffusion technique at 293 K. Monoclinic crystals appeared within 1–2 d with dimensions of 0.3–0.4 mm. Single crystals were only obtained by the preparation of a very large batch of crystals, most of which were twinned. In the course of the preparation of these crystals it was observed that the protein needed to be freshly prepared and to have a high purity index. In addition, all of the buffers were prepared in Milli-Q water and filtered using a 0.2 µm membrane to remove any small particles or dust. The crystals were prepared using fully oxidized protein at room temperature. The crystals were passed through the cryoprotectant Paratone oil before cryocooling in liquid dinitrogen. Diffraction data were collected on beamline ID23-EH1 at the ESRF (Grenoble, France) using a Quantum 4 charge-coupled device detector (ADSC Q315) from a crystal cooled to 100 K using a Cryostream (Oxford Cryosystems Ltd). Although complete data were only collected to 1.4 Å resolution, the crystals diffracted to beyond 1.25 Å resolution in the corners of the square detector. The data were processed using the programs *MOSFLM* (Leslie, 1992) and *SCALA* (Evans, 1997) from the *CCP*4 suite (Collaborative Computational Project 4, Number 4, 1994). The crystal belonged to space group *C*2. Calculation of the Matthews coefficient indicated the presence of a dimer in the asymmetric unit, with a V_{M} of 3.07 Å^3 Da^{-1} and 61% solvent content (Matthews, 1968). Data-collection statistics are given in Table 1.

2.3. Structure determination and refinement

The program *Phaser* (McCoy *et al.*, 2005) was used to solve the structure by molecular replacement using the previously determined structure of the protein from *P. pantotrophus* (PDB code 1adw, excluding the Cu^{2+} ions and water molecules; Williams *et al.*, 1995). Although the space groups of the crystals were identical, the unit-cell parameters were significantly different (by more than 1 Å in each dimension). *Phaser* found both monomers with a final log-likelihood gain (LLG) score of 2159, with *Z* scores of 14.6 and 12.2 after the rotation search and 15.9 and 33.0 after the translation search for the first and second monomer, respectively. Automated model building using *ARP*/*wARP* (Cohen *et al.*, 2004) with the solutions from *Phaser* and the X-ray data located 234 of 246 amino-acid residues, with a sequence coverage of 99% and an estimated correctness of the model of 95%. Refinement was carried out with the program *REFMAC*5 (Murshudov *et al.*, 1997) and rebuilding was performed using the program *Coot* (Emsley & Cowtan, 2004). A free *R* factor calculated from 5% of reflections that were set aside at the outset was used to monitor the progress of refinement. Thorough stereochemical checks were performed after each round using the validation tools in *Coot*. Refinement was carried out to 1.25 Å resolution as the data were good [$I/\sigma(I) > 2.0$] but incomplete (∼30% completeness in the outer shell). This still amounted to over 10 000 reflections between 1.25 and 1.4 Å resolution, adding to the overall quality of the electron-density maps during refinement. The *R* and $R_{\rm free}$ factors were 0.25 and 0.26, respectively, after the first round of refinement. The electron-density maps were of excellent quality and the whole polypeptide chain could be identified for both monomers. Several amino-acid residues from the flexible loops (between Phe*A*18 and Ala*A*21, Ile*A*48 and Glu*A*51, Phe*B*18 and Ala*B*21, and Asp*B*94 and Glu*B*97) and the N- and the C-termini of both monomers were added at this stage. Two Cu^{2+} ions, three SO_4^{2-} ions and 485 water molecules were also identified. Alternative positions for 16 amino-acid side chains were defined during the course of refinement (Met*A*7, Lys*A*10, Ile31, Gln*A*47, Ser*A*55, Lys*A*59, Met*A*84, Lys*A*103, Lys*A*110, Lys*B*10, Lys*B*38, Lys*B*46, Glu*B*62, Lys*A*106, Lys*B*109 and Lys*B*110). His6 has also multiple conformations, which are linked to its interaction with Glu4, but these were difficult to define. The final round of refinement was performed using the TLS/restrained refinement procedure using 20 segments for each monomer as determined by the *TLSMD* server (Painter & Merritt, 2006), which improved the *R* and $R_{\rm free}$ factors from 0.204 and 0.220 to 0.188 and 0.206, respectively. The final model was verified with *PROCHECK* (Laskowski *et al.*, 1998) and checked and validated during submission to the PDB. The r.m.s. deviations of the bond lengths, bond angles and torsion angles from standard values were 0.007 Å, 1.2° and 5.9°, respectively. The estimated overall coordinate error based on the $R/R_{\rm free}$ values was 0.053 Å. The following surface amino-acid residues have highly flexible side chains and have little or no electron-density definition in the final map: Glu*A*4, Glu*A*54, Asp*A*94, Asp*A*100, Glu*B*47, Glu*B*54, Asp*B*94 and Asp*B*100. Final refinement statistics are given in Table 2.

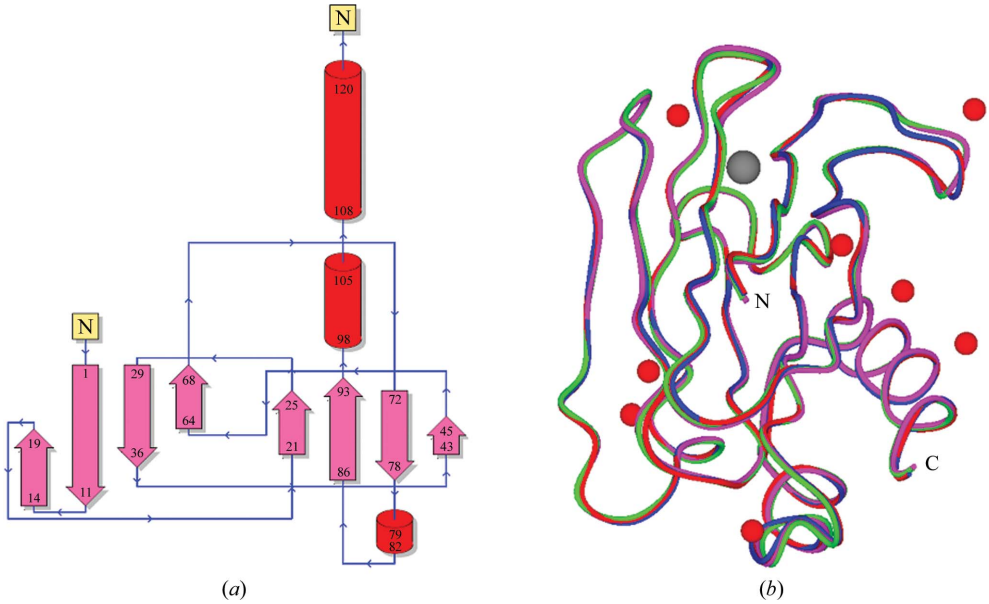
Coordinates and observed structure-factor amplitudes have been deposited in the Protein Data Bank under accession code 3erx.

3. Results and discussion

3.1. Overall structure

Pseudoazurin has a cupredoxin-like structural fold belonging to the azurin/plastocyanin family (Adman, 1991; Han *et al.*, 1991). Two mixed β-sheets, each consisting of four β-strands, form a β-sandwich, with two α-helices located in the carboxy-terminus (Fig. 1*a*). The highly conserved *cis*-proline residue (Pro20 in 3erx) provides the twist separating the two β-sheets (see below).

The *P. pantotrophus* pseudoazurin crystallizes with a dimer in the asymmetric unit. The last helix (H3) at the carboxy-terminus forms part of the dimer interface. In the 2.5 Å resolution structure 1adw the monomers were refined with strict noncrystallographic constraints until the final round of refinement (Williams *et al.*, 1995) and are to all purposes virtually identical [the r.m.s. deviation between the two monomers on superpositioning was 0.04 Å for all main-chain atoms using *SSM* (Krissinel & Henrick, 2004), see Fig. 1*b*].



(*a*) (*b*)

Figure 1
(*a*) Topological diagram of the *P. pantotrophus* pseudoazurin structure. (*b*) Worm trace showing the superimposition of subunits *A* (magenta) and *B* (green) from 3erx and subunits *A* (blue) and *B* (red) from 1adw using *SSM* (Krissinel & Henrick, 2004). The Cu^{2+} ion is depicted as a grey sphere and the eight common waters found in all subunits are depicted as red spheres. The picture was prepared using *CCP*4*mg* (Potterton *et al.*, 2004).

Table 2
Structure-refinement statistics.

Data set	3erx
Resolution (Å)	1.25
Reflections used	69604
R_{work}† (%)	18.8
R_{free}‡ (%)	20.6
No. of metals	2 Cu
No. of waters	485
No. of ligands	3 SO_4^{2-}
Mean B factors (Å^2)	
Protein main-chain atoms	10.0
Protein side-chain atoms	16.8
Coppers	11.6
Waters	28.0

† $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. ‡ R_{free} was calculated for a randomly chosen 5% of the reflections.

Table 3
Comparison of the geometry at the Cu metal-binding site for the low-resolution (1adw; 2.5 Å) and high-resolution (3erx; 1.4 Å) structures.

	1adw		3erx	
	Chain *A*	Chain *B*	Chain *A*	Chain *B*
Distances (Å)				
Cu—His40 $N^{\delta1}$	2.01	2.04	2.13	2.12
Cu—Cys78 S^{γ}	2.22	2.22	2.13	2.14
Cu—His81 $N^{\delta1}$	2.06	2.07	2.11	2.14
Cu—Met86 S^{δ}	2.63	2.57	2.75	2.70
Angles (°)				
His40 $N^{\delta1}$—Cu—Cys78 S^{γ}	134	135	136	135
His40 $N^{\delta1}$—Cu—His81 $N^{\delta1}$	102	99	106	110
His40 $N^{\delta1}$—Cu—Met86 S^{δ}	90	89	82	83
Met86 S^{δ}—Cu—His81 $N^{\delta1}$	109	114	108	110
Met86 S^{δ}—Cu—Cys78 S^{γ}	109	110	113	114
His81 $N^{\delta1}$—Cu—Cys78 S^{γ}	111	108	111	111

In our work, no constraints were used in the refinement of the structure at any stage. The r.m.s.d. between the two monomers in the 1.4 Å resolution structure (3erx) is 0.26 Å for all main-chain atoms and requires a rotation of 180° along the (0.8, 0, 0.6) axis and a translation of 91 Å for fitting. The following regions show the greatest deviations (r.m.s.d. > 0.4 Å): residues 33–38 and residues 62–68 (Fig. 1*b*).

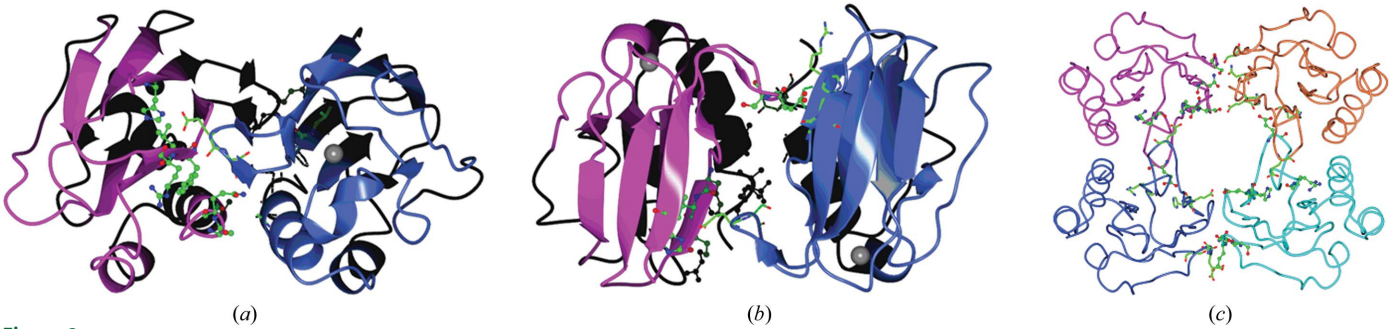
Chains *A* and *B* of 1adw can be superposed onto chains *A* and *B* of 3erx, respectively, with an r.m.s.d. of 0.41. A rotation of 9° along the (−0.3, 0.9, 0.4) axis and a translation of 3.25 Å is required to fit the second subunits onto each other after superimposition of the first subunits (Fig. 1*b*). This reflects the difference in the crystallographic packing between the 3erx and 1adw dimers. The following regions show the greatest deviations (r.m.s.d. > 0.5 Å): *A*10–*A*15, *A*28, *A*35–*A*36, *A*52, *A*62–*A*63, *A*69, *A*106–*A*107, *A*120–*A*123, *B*4, *B*9–*B*15, *B*35, *B*94–*B*100, *B*106–*B*107 and *B*122–*B*123. In structural terms, the noticeable differences are in the hairpin between the β1 and β2 strands, the loop/turn regions and at the carboxy-terminus.

Oxidized *P. pantotrophus* pseudoazurin has been shown to undergo an ionic strength-dependent monomer–dimer equilibrium in solution, which has been studied by size-exclusion chromatography (Pauleta, Guerlesquin *et al.*, 2004). Moreover, analytical ultracentrifugation (sedimentation-equilibrium) experiments demonstrated that there is also a monomer–dimer–tetramer equilibrium that is dependent on the ionic strength (Williams *et al.*, 1995; Pauleta *et al.*, unpublished work). Indeed, the dimer found in the asymmetric unit can be paired up in the crystal lattice to give a tetramer (Fig. 2*c*).

An analysis of the interfaces has been carried out using the *Protein Interfaces, Surfaces and Assemblies* service *PISA* at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html; Krissinel & Henrick, 2004). The dimer interface within the asymmetric unit involves residues from the first β-hairpin (Glu12 and Ser13), the β3 strand (Pro20, Phe22, Val23 and Arg24) and the carboxy-terminal helix (Ala117, Ala120 and Gln121) (Figs. 2*a* and 2*b*). In this interface there are also two pairs of hydrogen bonds formed between His3 $N^{\delta1}$ of one subunit and Glu12 $O^{\varepsilon2}$ of the other and between Glu12 $O^{\varepsilon1}$ of one subunit and the peptide N atom of Arg24 of the other. An additional polar contact is formed between the carbonyl O atom of Ser13 and Gln121 $O^{\varepsilon2}$ when compared with the 1adw dimer interface. The buried surface area per subunit at this interface is 463 Å^2 (Figs. 2*a* and 2*b*).

Comparison of the residues found at the dimer interface of 3erx with those of other pseudoazurins shows that the variability between them is too large to expect similar dimer formation. In fact, with the exception of *Hyphomicrobium denitrificans* pseudoazurin, all other pseudoazurins are monomers in the crystal. *H. denitrificans* pseudoazurin forms a trimer, in which one of the protein–protein interactions involves the same surface on the *A* subunit interacting with the displaced loop region between the β5 and β6 strands of the *B* subunit. *H. denitrificans* pseudoazurin also forms a head-to-head cupredoxin dimer involving the *A* and *C* subunits, particularly involving residues Pro81 and His82.

In *P. pantotrophus* pseudoazurin a tetramer can be defined by a crystallographic symmetry operation $-x, y, -z + 1$ in which a tighter interface is formed between the *A* and #*B* (or *B* and #*A*, where # indicates the crystallographic symmetry-related chain) subunits, with 496 Å^2 of surface area buried between each *A*–#*B* and *B*–#*A* pair (Fig. 2*c*). This interaction involves a contribution from the residues in β-strands β1, β4 (and the subsequent β-turn) and β5 (and the preceding β-turn). Nine hydrogen bonds contribute to this interface



(*a*) (*b*) (*c*)

Figure 2
Oligomerization of *P. pantotrophus* pseudoazurin into dimers (*a*, *b*) and tetramers (*c*). The residues involved in interchain contacts are shown in ball-and-stick representation for chain *A* of 3erx and as fat bonds for chain *B* of 3erx. These figures were prepared with *CCP4mg*. The following nine hydrogen bonds contribute to this interface: Glu*A*4 $O^{\varepsilon1}\cdots$Glu#*B*4 $O^{\varepsilon2}$ (3.01 Å), Glu*A*4 $O^{\varepsilon2}\cdots$Glu#*B*4 $O^{\varepsilon1}$ (3.17 Å)/Glu#*B*4 $O^{\varepsilon2}$ (2.69 Å), Asn*A*32 $O^{\delta1}\cdots$Pro#*B*35 O (3.14 Å), Thr*A*36 O$\cdots$Thr#*B*2 $O^{\gamma1}$ (3.26 Å), Asn*A*61 O$\cdots$Asn#*B*32 $O^{\delta1}$ (2.92 Å), Ser*A*63 N$\cdots$Asn#*B*32 $N^{\delta2}$ (2.90 Å), Ser*A*63 $O^{\gamma}\cdots$Ser#*B*63 O^{γ} (2.76 Å) and Thr*A*65 $O^{\gamma1}\cdots$Ser#*B*63 O^{γ} (2.65 Å).

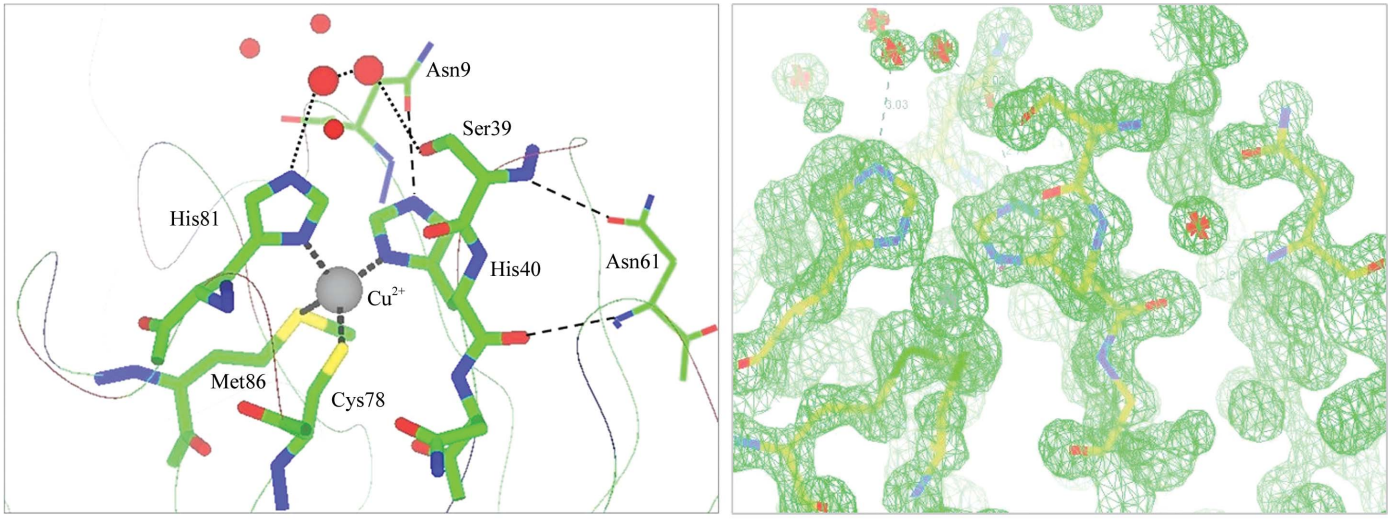


Figure 3
(*a*) The copper-binding site, depicting the residues that coordinate the copper ion: His40, Cys78, Met86 and His81. The two waters bridging His81 $N^{\varepsilon 2}$ to Ser39 O^{γ} are depicted as large red spheres. Only three of the six common waters between subunits *A* and *B* of 3erx are depicted as small red spheres. The overall fold is shown as a thin ribbon tracing. (*b*) The copper-binding site shown with electron density (contoured at 0.8 e $Å^{-3}$) drawn in *Coot*.

(Fig. 2*c*) together with residues involved in hydrophobic interactions (His6, Val30, Val34, Asp37, Lys38 and Tyr64 on each monomer). The hydrogen bond between the Glu4 and His6 side chains in each subunit may play a role in stabilizing this intermolecular interaction. These residues are conserved in the *Achromobacter cycloclastes* and *Alcaligenes faecalis* pseudoazurins, but not in the *Methylobacterium extorquens* and *H. denitrificans* pseudoazurins.

Two other smaller regions on the surface of each pseudoazurin monomer are also involved in crystal lattice contacts. Contributions are made by residues in the loop regions between β4 and β5, β6 and β7, and β8 and helix H2 on one subunit, and those between H1 and β8 and between H2 and H3 on the other. These include polar contacts between the pairs Glu70 and Glu97 with Gly#83 and Lys#109 (and *vice versa*) and lead to the burial of 255 $Å^2$ of the surface of the *A* subunit and 225 $Å^2$ of the *B* subunit.

Thus, in summary, four regions of each pseudoazurin subunit are involved in protein–protein association. The first is that involved in the formation of the dimer interface in the asymmetric unit, with His3, Glu12, Ser13, Phe22, Val23, Arg24, Ala117, Ala120 and Gln121 as the key residues. The second extends the dimer to a crystallographic tetramer and involves the following: Glu4, His6, Val30, Asn32, Val34, Pro35, Asp37, Lys38, Asn61, Ser63, Tyr64 and Thr65. The third and fourth regions extend the tetramer into the crystallographic lattice, with Glu51, Glu70, Glu97 and Pro71 interacting with the symmetry-related Met#84, Pro#108, Lys#109 and Gly#83 (and *vice versa*).

3.2. Metal-binding site

Pseudoazurin contains a type 1 copper site with the side chains of His40, His81, Cys78 and Met86 as ligands of the Cu^{2+} ion (Fig. 3). The geometry of the metal-binding site in 3erx is better defined than that in 1adw. The metal centre displays a tetrahedral distorted geometry in 3erx, as in other pseudoazurins (Table 3, Fig. 3). The geometry of the metal-binding site imposes a strained left-handed conformation on Ser39, which is presumably stabilized by a hydrogen bond between the peptide N atom and Asn61 $O^{\delta 1}$. The O^{γ} and the carbonyl O atoms of Ser39 are hydrogen bonded to Asn9 *via* water molecules. This residue is usually a Gly or a Pro in other family members (marked by a # in Fig. 4).

Three of the ligand-binding amino-acid residues are located in the β-hairpin between strands β7 and β8, which adopts a short 3_{10}-helical (H1) conformation for Cys78, Thr79, Pro80 and His81. This may maintain the stability of the loop despite losing an extra hydrogen bond owing to the presence of the Pro. This is the case for all of the other pseudoazurins and plastacyanins, which also have the consensus sequence **Cys**-*X*-Pro-**His**-*X*-*X*-*X*-Gly **Met**, where *X* can be any type of amino acid (Fig. 4). However, all of the other cupredoxins have a **Cys**-*X*-*X*-*X*-*X*-**His**-*X*-*X*-(*X*)-*X*-**Met** consensus, except for the amicyanins, which have the shortest β-hairpin comprising **Cys**-*X*-Pro-**His**-Pro-Phe-**Met** (Dennison, 2005*a*; Velarde *et al.*, 2007; Fig. 4 and Figs. S1 and S2 in supplementary material[1]).

Despite the variation in the length of the loop in this ligand-binding motif, the geometry of the copper binding is highly conserved (with differences in ligand distances to the Cu atom of less than 0.1 Å between any two cupredoxins). This difference in the loop length may well reflect the different properties required by each member of the subfamily for their specific interactions (Dennison, 2005*b*; Velarde *et al.*, 2007).

3.3. Water structure

In the medium-resolution 1adw structure only 56 water molecules were allocated compared with 485 in the high-resolution 3erx structure (232 of them were closer to the *A* subunit and 253 were closer to the *B* subunit). The *A* and *B* subunits of 1adw and the *B* subunit of 3erx were superimposed onto the *A* subunit of 3erx to analyze the conservation of the water structure.

A water molecule was deemed to be conserved if it was within 1.0 Å of another water molecule from a different subunit after superpositioning. Only eight water molecules were absolutely conserved between all four subunits (Fig. 1*b*). They were all on the surface and formed at least two hydrogen bonds to the main-chain

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: GX5158). Services for accessing this material are described at the back of the journal.

amide or carbonyl groups. Furthermore, 15 water molecules from 1adw were also conserved in one of the subunits of 3erx.

There are 140 water molecules in common positions in both 3erx subunits. Two of these form a hydrogen-bond network linking His*A*81 $N^{\varepsilon 2}$–HOH (*A*227/*B*242)–HOH (*A*194/*B*178)–Ser*A*39 O^{γ} and involve a further six conserved water molecules in the region (Fig. 1*b*). There are some 30 waters in common in the dimer interface and ten between the other subunits in the tetramer.

The analysis of the conservation of the water structure was extended to include pseudoazurins from other organisms using the program 3*dSS* (Sumathi *et al.*, 2006), but with a cutoff limit of 1.8 Å to allow for the sequence and structural differences. We used the highest resolution structures from each of the following species: *Ac. cycloclastes* (PDB code 1bqk, 121 waters, 1.35 Å resolution; Inoue *et al.*, 1999), the same protein with an engineered shorter amicyanin loop between His81 and Met86 (PDB code 2ux6, 115 waters, 1.30 Å resolution; Velarde *et al.*, 2007), *Al. faecalis* (PDB code 1paz, 93 waters, 1.55 Å resolution; Petratos *et al.*, 1988), *M. extorquens* (PDB code 1pmy, 132 waters, 1.50 Å resolution; Inoue *et al.*, 1994) and *H. denitrificans* (PDB code 3ef4, 500 waters, 1.18 Å resolution; Hira *et al.*, 2009). Seven waters were found in a common position in all these psuedoazurins. One of these (*A*136/*B*148, 3erx numbering) was near the metal-binding site, forming hydrogen bonds to the main-chain amide of His40 and the carbonyl of Asp37 and to the side chain of Asn61, and may play in an important role in maintaining the metal-binding site and in electron coupling. The others (*A*127/*B*127, *A*155/*B*187, *A*163/*B*173, *A*255/*B*274, *A*337/*B*216 and *A*345/*B*194) form at least two hydrogen bonds to the main-chain amide or carbonyl groups and presumably add to the overall structural stability of the proteins. *H. denitrificans* pseudoazurin is a trimer

```
                                 β1          β2  $  β3        β4        # *   β5
3ERX PAZ  ------------ATHEVHMLNKG-ESGAMVFEPAFVRAEPGDVINFVPTD-K-S-HN-VEAIKEILP-E-G-V-------E-------
1BQK PAZ  ------------ADFEVHMLNKG-KDGAMVFEPASLKVAPGDTVTFIPTD-K-G-HN-VETIKGMIP-D-G-A-------E-------
2UX6 PAM  ------------ADFEVHMLNKG-KDGAMVFEPASLKVAPGDTVTFIPTD-K-G-HN-VETIKGMIP-D-G-A-------E-------
1PAZ PAZ  ------------ENIEVHMLNKG-AEGAMVFEPAYIKANPGDTVTFIPVD-K-G-HN-VESIKDMIP-E-G-A-------E-------
1PMY PAZ  ------------DEVAVKMLNSG-PGGMMVFDPALVRLKPGDSIKFLPTD-K-G-HN-VETIKGMAP-D-G-A-------D-------
3EF4 PAZ  ------------AEHIVEMRNKDDAGNTMVFQPGFVKVEAGDTVKFVPTD-K-S-HN-AESVREVWP-E-G-VA--------------
1PCS PLA  --------------ANATVKMGS-DSGALVFEPSTVTIKAGEEVKWVNNK-L-SPHN-IVFDADGVP-A-D-T-AAKLSHKGLL----
1M9W PLA  --------------ANATVKMGS-DSGALVFEPSTVTIKAGEEVKWVNNK-L-SPHNIVFAADG-VD-A-D-T-AAKLSHKGL-----
2PLT PLA  ---------------DATVKLGA-DSGALEFVPKTLTIKSGETVNFVNNA-GFP-HN-IVFDEDAIP-S-G-VNADAISRDDYL----
1IUZ PLA  ---------------AQIVKLGG-DDGSLAFVPSKISVAAGEAIEFVNNA-GFP-HN-IVFDEDAVP-A-G-VDADAISYDDYL----
7PCY PLA  ---------------AAIVKLGG-DDGSLAFVPNNITVGAGESIEFINNA-GFP-HN-IVFDEDAVP-A-G-VDADAISAEDYL----
1PLA PLA  ----------------AEVKLGS-DDGGLVFSPSSFTVAAGEKITFKNNA-GFP-HN-IVFDEDEVP-A-G-VNAEKISQPEYL----
1BXU PLA  ---------------QTVAIKMGA-DNGMLAFEPSTIEIQAGDTVQWVNNK-L-APHN-VVVE--GQP-ELS--------HKDLA----
1B3I PLA  ---------------ASVQIKMGT-DKYAPLYEPKALSISAGDTVEFVMNK-V-GPHN-VIFD--KVP-A-G-E----SAPAL----SN
9PCY PLA  ----------------LEVLLGSG-D-GSLVFVPSEFSVPSGEKIVFKNNA-GFP-HN-VVFDEDEIP-A-G-VDAVKIS-------MP
1AG6 PLA  ----------------VEVLLGG-DDGSLAFLPGDFSVASGEEIVFKNNA-GFP-HN-VVFDEDEIP-S-G-VDAAKIS-------MS
1BYO PLA  ----------------AEVLLGS-SDGGLAFVPSDLSIASGEKITFKNNA-GFP-HN-DLFDEDEVP-A-G-VDVTKIS-------MP
1AAC AMI  -DKATIPSESPFAAAEVADGAIVV-DIAKMKYETPELHVKVGDTVTWINRE-AMP-HN-VHFVAGVLG-E-A-A------LK------G
1PCY PLA  ----------------IDVLLGA-DDGSLAFVPSEFSISPGEKIVFKNNA-GFP-HN-IVFDEDSIP-S-G-VDASKIS-------MS
1BAW PLA  ---------------ETFTVKMGA-DSGLLQFEPANVTVHPGDTVKWVNNKLP-P-HN-ILFDDKQVP-G-A-SKELADKLSHSQLMFS
1ID2 AMI  QDKITVTSEKPVAAADVPADAVVV-GIEKMKYLTPEVTIKAGETVYWVNGEVM-P-HN-VAFKKGIV--G-E—DA-----FR-GE-MMT
1FA4 PLA  ----------------ETYTVKLGSDKGLLVFEPAKLTIKPGDTVEFLNNKVP-P-HN-VVFDAALNP-A---KSAD-L-AK-SL---S

                                 β6                β7 *H1*     *  β8          H2          H3
3ERX PAZ  SFK-SK-I-NE-SYTLTVTE--P-----GLYGVKCTPH-FGMGMVGLVQVGDAPENLDAAKTAKMPKKARERMDAELAQVN-
1BQK PAZ  AFK-SK-I-NE-NYKVTFTA--P-----GVYGVKCTPH-YGMGMVGVVQVGDAPANLEAVKGAKNPKKAQERLDAALAALGN
2UX6 PAM  AFK-SK-I-NE-NYKVTFTA--P-----GVYGVKCTPHPF---MVGVVQVGDAPANLEAVKGAKNPKKAQERLDAALAALGN
1PAZ PAZ  KFK-SK-I-NE-NYVLTVTQ--P-----GAYLVKCTPH-YAMGMIALIAVGDSPANLDQIVSAKKPKIVQERLEKVIA----
1PMY PAZ  YVK-TT-V-GQ-EAVVKFDK--E-----GVYGFKCAPH-YMMGMVALVVVGDKRDNLEAAKSVQHNKLTQKRLDPLFAQIQ-
3EF4 PAZ  PVK-GG-F-SK-EVVFNAEK--E-----GLYVLKCAPH-YGMGMVVLVQVGKPVNLDQIKEYKATGLAKKRLDGEIAKVVQ-
1PCS PLA  -FAA-----GE-SFTSTFTE--P-----GTYTYYCEPH-RGAGMVGKVVVE----------------------------------
1M9W PLA  AFAA-----GE-SFTSTFTE--P-----GTYTYYCEPH-RGAGMVGKVVVE----------------------------------
2PLT PLA  -----NAP-GE-TYSVKLTA--A-----GEYGYYCEPH-QGAGMVGKIIVQ----------------------------------
1IUZ PLA  -----NSK-GE-TVVRKLST--P-----GVYGVYCEPH-AGAGMKMTITVQ----------------------------------
7PCY PLA  -----NSK-GQ-TVVRKLTT--P-----GTYGVYCDPH-SGAGMKMTITVQ----------------------------------
1PLA PLA  -----NGA-GE-TYEVTLTE--K-----GTYKFYCEPH-AGAGMKGEVTVN----------------------------------
1BXU PLA  -----FSP-GE-TFEATFSE--P-----GTYTYYCEPH-RGAGMVGKIVVQ----------------------------------
1B3I PLA  TKLA-IAP-G-SFYSVTLGT--P-----GTYSFYCTPH-RGAGMVGTITVE----------------------------------
9PCY PLA  EEELLNAP-GE-TYVVTLDT--K-----GTYSFYCSPH-QGAGMVGKVTVN----------------------------------
1AG6 PLA  EEDLLNAP-GE-TYKVTLTE--K-----GTYKFYCSPH-QGAGMVGKVTVN----------------------------------
1BYO PLA  EEDLLNAP-GE-EYSVTLTE--K-----GTYKFYCAPH-AGAGMVGKVTVN----------------------------------
1AAC AMI  PM--MK-K--EQAYSLTFTE--A-----GTYDYHCTPHPF---MRGKVVVE----------------------------------
1PCY PLA  EEDLLNAK-GE-TFEVALSN--K-----GEYSFYCSPH-QGAGMVGKVTVN----------------------------------
1BAW PLA  P--------GE-SYEITFSSDFPA----GTYTYYCAPH-RGAGMVGKITVEG---------------------------------
1ID2 AMI  K---------DQAYAITFNE--A-----GSYDYFCTPHPF---MRGKVIVE----------------------------------
1FA4 PLA  HKQLLMSP----GQSTSTTF--PADAPAGEYTFYCEPH-RGAGMVGKITVAG---------------------------------
```

Figure 4
Structure-based sequence alignment of pseudoazurins (PAZ) from different organisms with some plastocyanins (PLA) and amicyanins (AMI), which are other type 1 copper proteins, based on *SAS* in *PDBSUM* (Laskowski *et al.*, 2005). The copper-coordinating residues are marked by asterisks. The highly conserved *cis*-proline residue that provides the twist separating the two *β*-sheets (Pro20 in 3erx) is marked by a $. The symbol # identifies Ser39 (in *P. pantotrophus*), which presents a strained left-handed conformation and is hydrogen bonded through its O^{γ} and carbonyl O atoms to Asn9 *via* water molecules. Details and statistics of the structures used in the comparison are given in Table S1. Residue colours for secondary-structure definitions: red, helix; blue, strand; green, turn; black, coil. PAM, pseudoazurin with an engineered amicyanin loop.

in the asymmetric unit and has 60 waters in common between each subunit. However, the *A* and *C* subunits make contacts near the metal-binding site and this leads to the exclusion of some waters, mimicking the interaction with the redox partner. Thus, although the other pseudoazurins have these water molecules linking His*A*81 $N^{\varepsilon 2}$ to Ser*A*39 O^{γ}, in the case of *H. denitrificans* pseudoazurin these are absent.

The conserved water molecules near His81, which is one of the copper ligands and is also the proposed entry/exit point for electrons (Dennison, 2005*a*; Pauleta, Guerlesquin *et al.*, 2004), might have an important role in both facilitating electron transfer and in the formation of the electron-transfer complex. Indeed, formation of the electron-transfer complex between pseudoazurin and cytochrome *c* peroxidase has been shown to be entropically driven, with waters being released from the interface upon complex formation (Pauleta, Guerlesquin *et al.*, 2004).

3.4. The cupredoxin structural family

Over 25 NMR/X-ray crystal structures of pseudoazurins from five different species have been reported: those from *P. pantotrophus* (3erx and 1adw; Williams *et al.*, 1995), *Ac. cycloclastes* (1bqk; Inoue *et al.*, 1999), *Al. faecalis* (1paz; Petratos *et al.*, 1988), *M. extorquens* (1pmy; Inoue *et al.*, 1994) and *H. denitrificans* (3ef4; Hira *et al.*, 2009) (Table S1). A comparison of these structures using *SSM* (Krissinel & Henrick, 2004) showed that they are highly homologous, with an r.m.s.d. of around 1.0 Å between them. The largest deviations between them are found in the loop regions and particularly in the C-terminal helical regions. *H. denitrificans* pseudoazurin (Hira *et al.*, 2009) has a single insertion at position 15 and shows the largest deviation in the hairpin between the β1 and β2 strands and a deletion at position 95 in the loop between the β7 strand and helix H2. The presence of the bulky tryptophan residue instead of Leu/Ile/Ala just before the conserved Pro50 (in the loop region between the β5 and β6 strands) also shifts this region by more than 2.0 Å compared with the other pseudoazurins. However, the overall β-barrel shape and the Cu-ligand site are structurally highly conserved in all the pseudoazurins.

Recently, the conformation of *Al. faecalis* pseudoazurin in an electron-transfer complex with a nitrite reductase was determined by paramagnetic NMR spectroscopy (Vlasie *et al.*, 2008). The inter-

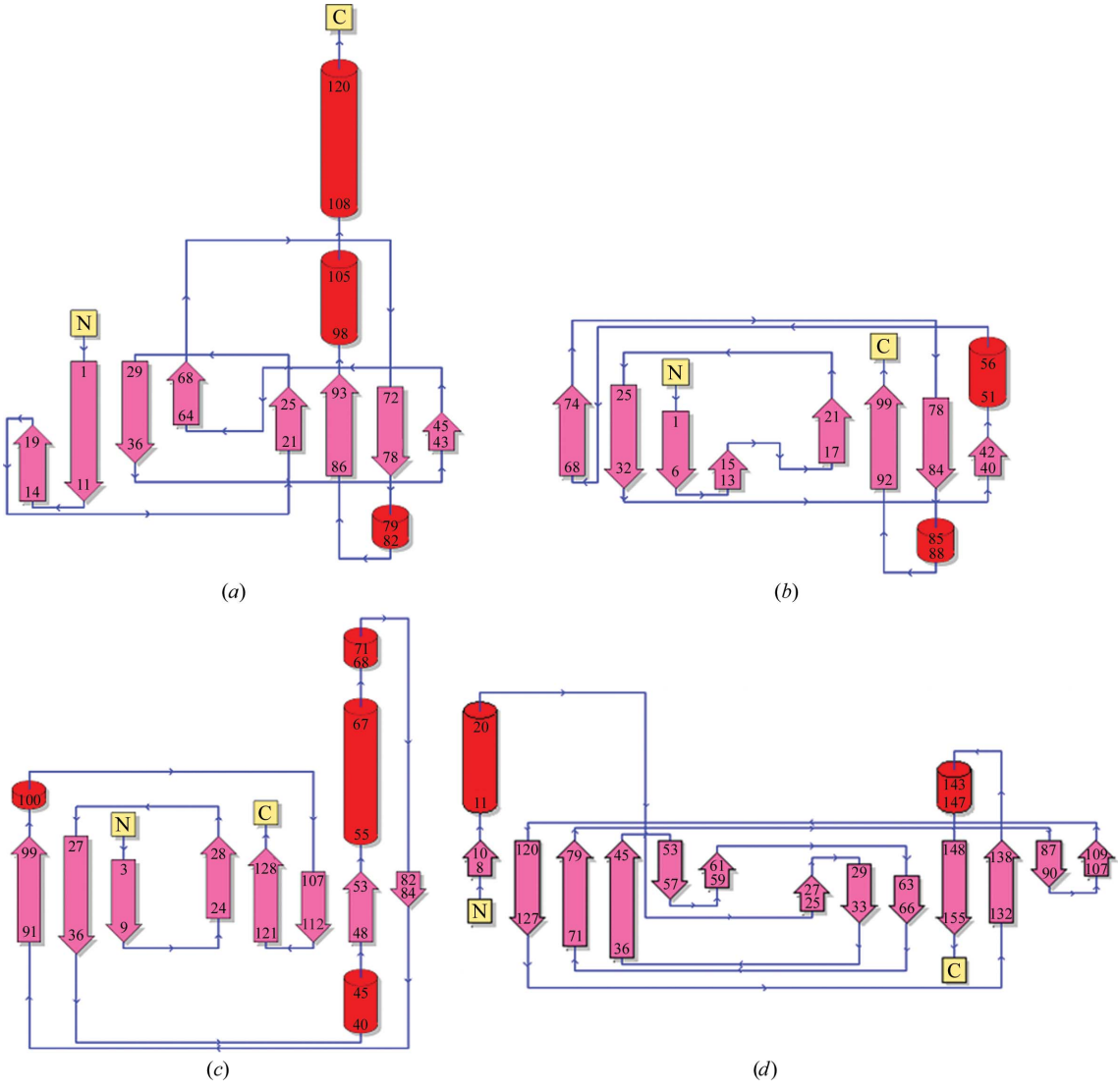


Figure 5
Two-dimensional topological diagram for the (*a*) *P. pantotrophus* pseudoazurin (PDB code 3erx), (*b*) *Populus nigra* plastocyanin (1pcy; Guss *et al.*, 1992), (*c*) *Al. denitrificans* azurin (2aza; Baker, 1988) and (*d*) *Thiobacillus ferrooxidans* rusticyanin (2cak; Barrett *et al.*, 2006) structures, highlighting the common two-sheet β-sandwich fold (in pink) and emphasizing the highly varied connectivity, loop and helical (red) nature of the cupredoxin family.

protein distance between the two type 1 copper sites was ~16 Å. The binding interface consists of polar and nonpolar residues surrounded by charged residues, with 16 residues contributed from the nitrite reductase and 13 residues from the pseudoazurin, with burial of over 700 Å^2 of interface area. Most of the 13 residues involved in the binding to nitrite reductase (highlighted in yellow for 1paz in Fig. 4) are highly conserved in other pseudoazurins. The major differences are at position 13, where a Glu is substituted by a Ser, and at 110, where Ile is substituted by a Lys in 3erx. Indeed, the same regions that are involved in the binding interface of *Al. faecalis* pseudoazurin with nitrite reductase have been identified to be part of the binding of *P. pantotrophus* pseudoazurin to cytochrome *c* peroxidase (Pauleta, Guerlesquin *et al.*, 2004).

Currently, there are over 200 structures of type 1 copper proteins from more than 30 different organisms in the PDB. This family can be further subdivided into azurins, plastocyanins, amicyanins, pseudoazurins, rusticyanins, auricyanins and phytocyanins (stellacyanin, umecyanin, mavicyanin and plantacyanin) based on our structural and sequence-alignment analysis of this family carried out using both the *SAS* facility in *PDBSUM* (Laskowski *et al.*, 2005) and *SSM* (Krissinel & Henrick, 2004) (Figs. 4, S1 and S2).

Although the sequence identity can be as low as 10% between any two sequences, the r.m.s.d. of the structure is less than 3.0 Å and is less than 1.0 Å between members of the same subfamily. For convenience, the sequence alignments based on pseudoazurins, plastocyanins and amicyanin (Fig. 4), those based on azurins and auracyanins (Fig. S1) and that based on phytocyanins (Fig. S2) have been depicted separately. Analysis of these sequence alignments enabled the identification of differences in the position and motifs of the secondary structure that can be used to classify an uncharacterized protein into one of the type 1 copper subfamilies.

Although all of the members of this family have the common mixed β-stranded two-sheet β-sandwich fold, there is considerable variation in the connectivity between the β-strands and the number and positions of α-helices (Fig. 5).

The pseudoazurins and plastocyanins have a similar copper-binding site and core fold, but with a striking difference in that the pseudoazurins have a C-terminal extension of some 20 amino-acid residues which form two helices. In the case of amicyanins there is an extra N-terminal extension compared with pseudoazurins, but they lack the C-terminal α-helical region (Fig. 4).

The azurins lack both an N- and C-terminal region, but have α-helices before and after the β4 strand. The β2 strand of pseudoazurin and the β5 strand of azurins do not have a direct structural alignment, whereas the rest of the strands do. The auracyanins have a common fold with azurins, but present an extra N-terminal extension compared with the azurins (Figure S1).

The nonchloroplast phytocyanins can be grouped as a subgroup (Fig. S2) as they have seven β-strands instead of eight as in other type 1 blue copper proteins. In addition, cucumber stellacyanin, horseradish umecyanin and zucchini mavicyanin have a glutamine residue replacing the methionine in the copper-binding site, but still retain the copper-binding geometry (Fig. S2).

The rusticyanins are the most divergent members of this family. They share about 28% sequence identity with the auracyanins. They are larger proteins with 13 β-strands and an additional α-helix at the N-terminus (Fig. 5).

In conclusion, this structure-based sequence alignment enabled the classification of the type 1 copper proteins into five families that differ from the sequenced-based families mentioned in §1: pseudoazurin/amicyanins, plastocyanins, azurin/auracyanins, phytocyanins and rusticyanins.

This classification was applied to three genes annotated as coding for blue copper proteins (Table S4). Their secondary structure, as predicted using *Jpred* 3 (Cole *et al.*, 2008), demonstrates that these genes are wrongly annotated and should have been identified as pseudoazurin precursors (Fig. S3). It is clear that they will have eight β-strands and two helices at the C-terminus, as is characteristic of pseudoazurins. Indeed, the structure of the abovementioned blue copper protein from *H. denitrificans* (PDB code 3ef4) has been determined and confirmed its predicted secondary structure and its classification as a pseudoazurin (Hira *et al.*, 2009).

4. Conclusions

The structure of *P. pantotrophus* pseudoazurin was solved at high resolution, which revealed the presence of several water molecules at the surface. Comparison of the water structure with those of other pseudoazurins identified conserved waters. These water molecules have been shown to be important in the molecular recognition of the partners (Chandler, 2005), as proposed in the study of the entropy-driven complex formed with cytochrome *c* peroxidase (Pauleta, Cooper *et al.*, 2004; Pauleta, Guerlesquin *et al.*, 2004). These water molecules may also be instrumental in the rapid electron transfer between the partners.

Type 1 blue copper proteins have been the focus of several studies. These studies have been based on several criteria, such as sequence, identification of specific signatures of each copper-binding site, study of the surface properties or correlation of the redox properties with their active site. In the present work, we have analyzed the structures of several type 1 copper proteins and have identified specific signatures that enable the separation of these proteins into different subfamilies. With the recent improvements in bioinformatics analysis and prediction of secondary structure, the analysis performed here can help in the classification of novel copper-binding proteins into a specific subfamily of the type 1 copper proteins. This will be useful in improving the identification and annotation of this type of protein in the increasing number of sequenced genomes, a task that is usually based on sequence homology. The identification of a copper-binding protein as belonging to one of these subfamilies has the additional advantage that some of its spectroscopic and functional properties can also be predicted.

This work was supported in part by Fundação para a Ciência e a Tecnologia (Lisbon, Portugal) through project grant POCTI/QUI/64638/2006 and individual grant SFRH/BPD/20357/2004 (SN). The authors would like to thank Dr Cecília Bonifácio for help with crystallization and Dr José Trincão for help with data collection.

References

Adman, E. T. (1991). *Adv. Protein Chem.* **42**, 145–197.
Adman, E. T. & Jensen, L. H. (1981). *Isr. J. Chem.* **21**, 8–12.
Baker, E. N. (1988). *J. Mol. Biol.* **203**, 1071–1095.
Barrett, M. L., Harvey, I., Sundararajan, M., Surendran, R., Hall, J. F., Ellis, M. J., Hough, M. A., Strange, R. W., Hillier, I. H. & Hasnain, S. S. (2006). *Biochemistry*, **45**, 2927–2939.
Berks, B. C., Baratta, D., Richardson, J. & Ferguson, S. J. (1993). *Eur. J. Biochem.* **212**, 467–476.
Bond, C. S., Blankenship, R. E., Freeman, H. C., Guss, J. M., Maher, M. J., Selvaraj, F. M., Wilce, M. C. & Willingham, K. M. (2001). *J. Mol. Biol.* **306**, 47–67.
Botuyan, M. V., Toy-Palmer, A., Chung, J., Blake, R. C. II, Beroza, P., Case, D. A. & Dyson, H. J. (1996). *J. Mol. Biol.* **263**, 752–767.
Canters, G. W. & Gilardi, G. (1993). *FEBS Lett.* **325**, 39–48.

Canters, G. W., Kolczak, U., Armstrong, F., Jeuken, L. J. C., Camba, R. & Sola, M. (2000). *Faraday Discuss.* **116**, 205–220.
Chandler, D. (2005). *Nature (London)*, **437**, 640–647.
Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J. & Perrakis, A. (2004). *Acta Cryst.* D**60**, 2222–2229.
Cole, C., Barber, J. D. & Barton, G. J. (2008). *Nucleic Acids Res.* **36**, W197–W201.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
De Rienzo, F., Gabdoulline, R. R., Menziani, M. C. & Wade, R. C. (2000). *Protein Sci.* **9**, 1439–1454.
Dennison, C. (2005*a*). *Coord. Chem. Rev.* **249**, 3025–3054.
Dennison, C. (2005*b*). *Dalton Trans.*, pp. 3436–3442.
Dodd, F. E., Van Beeumen, J., Eady, R. R. & Hasnain, S. S. (1998). *J. Mol. Biol.* **282**, 369–382.
Durley, R., Chen, L., Lim, L. W., Mathews, F. S. & Davidson, V. L. (1993). *Protein Sci.* **2**, 739–752.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (1997). *CCP4 Newsl. Protein Crystallogr.* **33**, 22–24.
Giri, A. V., Anishetty, S. & Gautam, P. (2004). *BMC Bioinformatics*, **5**, 127.
Guss, J. M., Bartunik, H. D. & Freeman, H. C. (1992). *Acta Cryst.* B**48**, 790–811.
Guss, J. M., Merritt, E. A., Phizackerley, R. P. & Freeman, H. C. (1996). *J. Mol. Biol.* **262**, 686–705.
Han, J., Adman, E. T., Beppu, T., Codd, R., Freeman, H. C., Huq, L. L., Loehr, T. M. & Sanders-Loehr, J. (1991). *Biochemistry*, **30**, 10904–10913.
Hart, P. J., Nersissian, A. M., Herrmann, R. G., Nalbandyan, R. M., Valentine, J. S. & Eisenberg, D. (1996). *Protein Sci.* **5**, 2175–2183.
Hira, D., Nojiri, M. & Suzuki, S. (2009). *Acta Cryst.* D**65**, 85–92.
Inoue, T., Kai, Y., Harada, S., Kasai, N., Ohshiro, Y., Suzuki, S., Kohzuma, T. & Tobari, J. (1994). *Acta Cryst.* D**50**, 317–328.
Inoue, T., Nishio, N., Suzuki, S., Kataoka, K., Kohzuma, T. & Kai, Y. (1999). *J. Biol. Chem.* **274**, 17845–17852.
Kaim, W. & Rall, J. (1996). *Angew Chem. Int. Ed. Engl.* **35**, 43–60.
Koch, M., Velarde, M., Harrison, M. D., Echt, S., Fischer, M., Messerschmidt, A. & Dennison, C. (2005). *J. Am. Chem. Soc.* **127**, 158–166.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Lappin, A. G. (1981). *Metal Ions in Biological Systems*, edited by H. Sigel, Vol. 13, pp. 15–64. New York: Marcel Dekker.
Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, D266–D268.
Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. (1998). *Curr. Opin. Struct. Biol.* **8**, 631–639.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Murphy, L. M., Dodd, F. E., Yousafzai, F. K., Eady, R. R. & Hasnain, S. S. (2002). *J. Mol. Biol.* **315**, 859–871.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M. & Canters, G. W. (1991). *J. Mol. Biol.* **221**, 765–772.
Nersissian, A. M., Immoos, C., Hill, M. G., Hart, P. J., Williams, G., Herrmann, R. G. & Valentine, J. S. (1998). *Protein Sci.* **7**, 1915–1929.
Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.
Pauleta, S. R., Cooper, A., Nutley, M., Errington, N., Harding, S., Guerlesquin, F., Goodhew, C. F., Moura, I., Moura, J. J. & Pettigrew, G. W. (2004). *Biochemistry*, **43**, 14566–14576.
Pauleta, S. R., Guerlesquin, F., Goodhew, C. F., Devreese, B., Van Beeumen, J., Pereira, A. S., Moura, I. & Pettigrew, G. W. (2004). *Biochemistry*, **43**, 11214–11225.
Petratos, K., Dauter, Z. & Wilson, K. S. (1988). *Acta Cryst.* B**44**, 628–636.
Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A. & Noble, M. (2004). *Acta Cryst.* D**60**, 2288–2294.
Romero, A., Hoitink, C. W., Nar, H., Huber, R., Messerschmidt, A. & Canters, G. W. (1993). *J. Mol. Biol.* **229**, 1007–1021.
Sailasuta, N., Anson, F. C. & Gray, H. B. (1979). *J. Am. Chem. Soc.* **101**, 455–458.
Sam, K. A., Fairhurst, S. A., Thorneley, R. N., Allen, J. W. & Ferguson, S. J. (2008). *J. Biol. Chem.* **283**, 12555–12563.
Scharf, B. & Engelhard, M. (1993). *Biochemistry*, **32**, 12894–12900.
Solomon, E. I. (2006). *Inorg Chem.* **45**, 8012–8025.
Sumathi, K., Ananthalakshmi, P., Roshan, M. N. & Sekar, K. (2006). *Nucleic Acids Res.* **34**, W128–W132.
Thompson, G. S., Leung, Y. C., Ferguson, S. J., Radford, S. E. & Redfield, C. (2000). *Protein Sci.* **9**, 846–858.
Thorndycroft, F. H., Butland, G., Richardson, D. J. & Watmough, N. J. (2007). *Biochem. J.* **401**, 111–119.
Velarde, M., Huber, R., Yanagisawa, S., Dennison, C. & Messerschmidt, A. (2007). *Biochemistry*, **46**, 9981–9991.
Vlasie, M. D., Fernandez-Busnadiego, R., Prudencio, M. & Ubbink, M. (2008). *J. Mol. Biol.* **375**, 1405–1415.
Walter, R. L., Ealick, S. E., Friedman, A. M., Blake, R. C. II, Proctor, P. & Shoham, M. (1996). *J. Mol. Biol.* **263**, 730–751.
Williams, P. A., Fülöp, V., Leung, Y. C., Chan, C., Moir, J. W., Howlett, G., Ferguson, S. J., Radford, S. E. & Hajdu, J. (1995). *Nature Struct. Biol.* **2**, 975–982.
Xie, Y., Inoue, T., Miyamoto, Y., Matsumura, H., Kataoka, K., Yamaguchi, K., Nojini, M., Suzuki, S. & Kai, Y. (2005). *J. Biochem.* **137**, 455–461.