*Supplementary information*

# Structure and metal-loading of a soluble periplasm cupro-protein

Kevin J. Waldron*, Susan J. Firbank*, Samantha J. Dainty, Mónica Pérez-Rama, Steve Tottey* and Nigel J. Robinson

*These authors made equal contributions.

Institute for Cell and Molecular Biosciences, University of Newcastle, Medical School, NE2 4HH, UK.

Correspondence should be addressed to NJR (n.j.robinson@ncl.ac.uk).

# Supplementary information

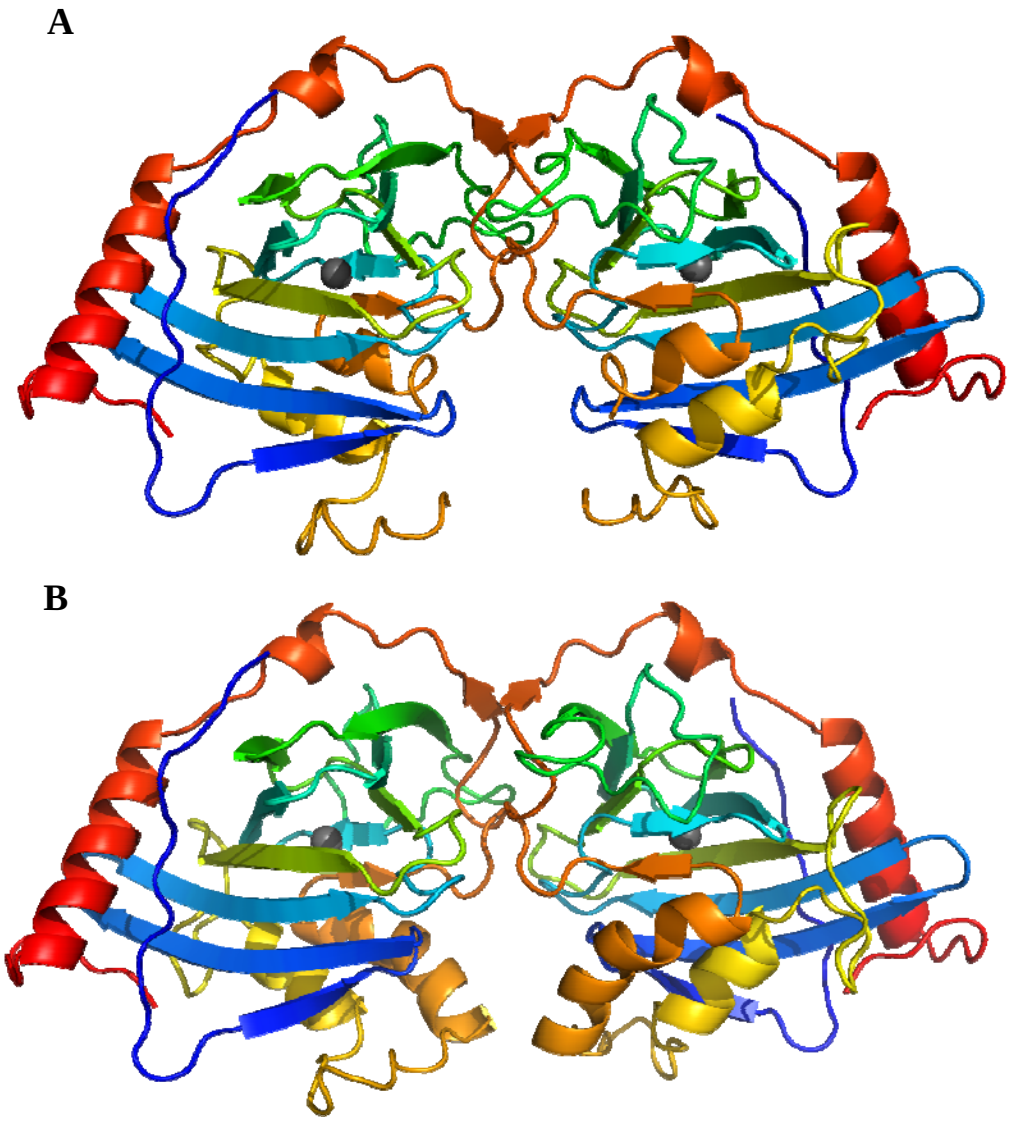
## Crystal structure determination

*Closed forms.* Different structures of the closed protein were obtained: $Cu^{2+}$-CucA, $Cu^{2+}$-CucA-SeMet, $Zn^{2+}$-CucA-SeMet and CucA-SeMet. The structures of the closed proteins were all highly similar (r.m.s.d. for all Cαs <0.5 Å for pairwise comparisons), with the majority of differences attributable to variations in crystal contacts or crystallisation liquor components. $Cu^{2+}$-CucA, crystallised in the absence of excess metal, contained copper exclusively at the metal binding site, whereas CucA crystallised in the presence of excess $Cu^{2+}$ or $Zn^{2+}$, exhibited an additional metal site at the surface of the protein forming crystal contacts. The sites were not the same and are not thought to be significant.

*$Cu^{2+}$-CucA – open native*. Despite being the highest resolution structure, several surface regions of the protein appeared mobile – either with two conformations for small loops required, or with residues 203-207 not satisfactorily buildable and absent from the final model, with the density suggesting at least two quite different conformations of this region and high flexibility. In addition, the unmodelled stretch is at the dimer interface with the corresponding region of a symmetry related molecule, thus complicating the correct assignment of the density. TLS refinement reduced the R and $R_{free}$ by 2 and 3 % respectively, with the final model containing 2 TLS groups: residues 175-210 forming one TLS group and the rest of the protein forming the other. This open form was also observed in some crystals formed from metal free protein (data not shown), though crystals of both the copper and apo forms often showed high mosaicity. At the copper binding site, the position of the water molecule is slightly different from the other copper and zinc structures, with water at a distance of 2.1Å commencing the string of molecules through the channel to the surface of the protein.

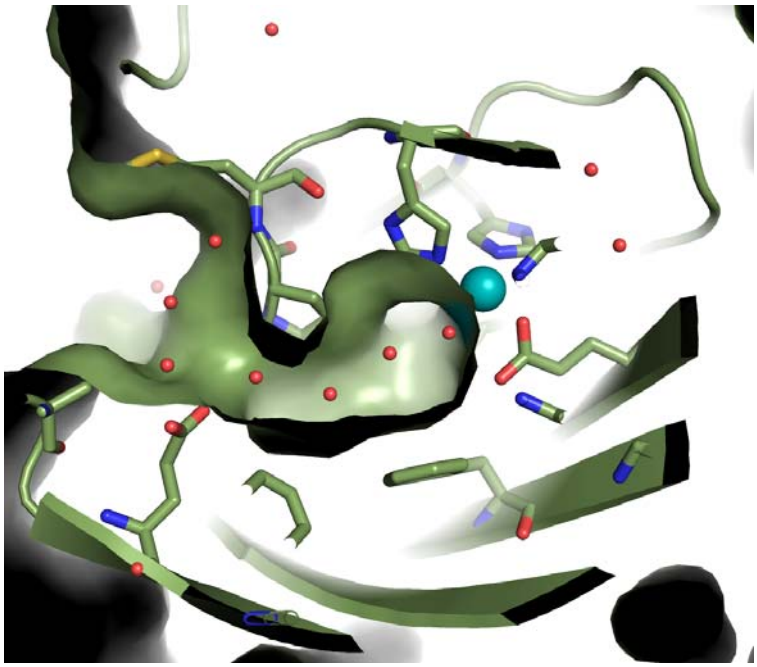
**Supplementary table 2 - X-ray diffraction data collection and refinement statistics**

| | Cu-Cuc closed SeMet | Cu-Cuc closed SeMet | Zn-CucA closed SeMet | Zn-CucA closed SeMet | Cu-CucA closed Native | CucA closed SeMet | Cu-CucA open Native |
|---|---|---|---|---|---|---|---|
| PDB code | 2XL7 | NA | 2XL9 | NA | 2XLA | 2XLF | 2XLG |
| Space group | $P6_422$ | $P6_422$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_1$ | C2 |
| Cell Dimensions | a=88.48 b=88.48 c=192.42 | A=89.2 b=89.2 c=195.0 | a=46.75 b=96.97 c=114.51 | a=46.45 b=96.58 c=114.40 | a=50.31 b=72.51 c=262.10 | a=57.43 b=142.47 c=63.50 β=95.14 | a=87.41 b=55.84 c=63.03 β=110.0 |
| Wavelength (Å) | 0.979 | 1.378 | 0.979 | 1.282 | 0.979 | 0.97 | 0.98 |
| Beamline | ESRF ID14.4 | ESRF ID29.1 | ESRF ID14.4 | ESRF ID29.1 | SRS 10.1 | DLS I03 | SRS 10.1 |
| Resolution (Å) | 60-2.4 (2.53-2.4) | 72-2.7 (2.85-2.7) | 57-2.06 (2.17-2.06) | 49-2.4 (2.53-2.4) | 37-1.93 (2.03-1.93) | 47-2.3 (2.42-2.3) | 46-1.8 (1.9-1.8) |
| Rmerge (%) | 10 (34) | 13.3 (29.0) | 9.7 (31.9) | 8.1 (25.2) | 9.2 (26.2) | 8.6 (31.2) | 5.8 (19.2) |
| Mean I/σI | 13.6 (4.9) | 18.5 (8.9) | 9.6 (2.5) | 14.1 (4.7) | 9.4 (4.1) | 8.4 (3.2) | 14.0 (4.1) |
| Completeness (%) | 99.7 (100) | 99.9 (99.9) | 96.3 (80.3) | 99.7 (100) | 96 (92) | 98.7 (99.1) | 99.9 (99.9) |
| Multiplicity (%) | 6.0 (6.1) | 10.1 (10.3) | 3.4 (2.9) | 4.5 (4.6) | 3.6 (3.3) | 2.5 (2.5) | 3.8 (3.8) |
| mols/a.s.u. | 1 | 1 | 2 | 2 | 4 | 4 | 1 |
| **Refinement** | | | | | | | |
| Resolution | 44-2.4 (2.46-2.4) | - | 42-2.06 (2.11-2.06) | - | 37-1.93 (1.98-19.3) | 41-2.3 (2.36-2.3) | 59-1.8 (1.85-1.8) |
| No. of reflections | 17157 (1247) | - | 20085 (1721) | - | 66822 (4845) | 42199 (3269) | 25190 (1893) |
| Rwork | 17.4 (19.2) | - | 19.2 (24.3) | - | 16.1 (17.0) | 18.45 (25.0) | 19.6 (24.6) |
| Rfree | 21.3 (23.8) | - | 23.7 (28.7) | - | 20.3 (20.4) | 22.9 (30.3) | 22.9 (29.4) |
| No. of atoms | 2057 | - | 3994 | - | 8409 | 7491 | 2021 |
| Bfactor | 31.3 | - | 39.3 | - | 13.7 | 51.5 | 35.17 |
| Rmsd bond lengths | 0.012 | - | 0.012 | - | 0.013 | 0.011 | 0.011 |
| Rmsd angles | 1.30 | - | 1.28 | - | 1.4 | 1.29 | 1.28 |



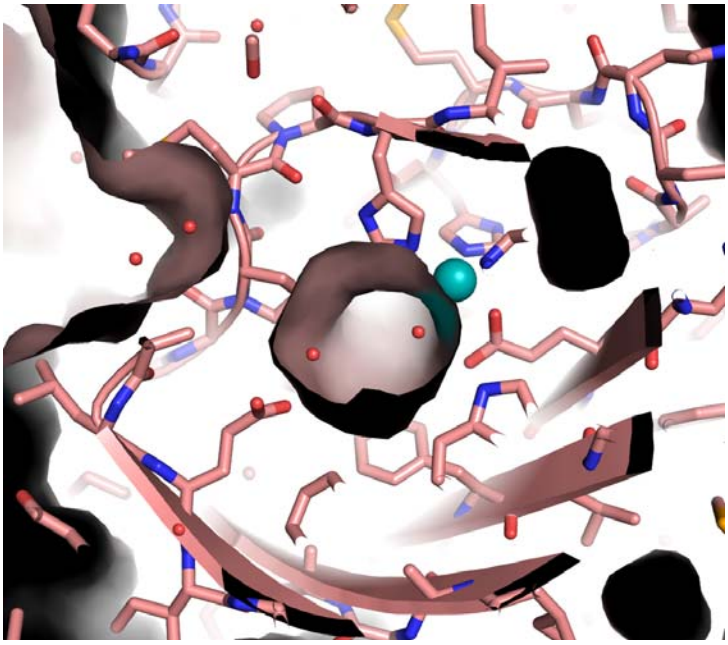

**Supplementary figure 1 - Dimerisation of CucA.**

All crystal forms of CucA contain a dimer, formed either by crystallographic or noncrystallographic symmetry molecules. That CucA appears to form dimers draws parallels with many other cupin proteins which exist as bicupins. A Dali search (Holm & Rosenström, 2010) indicates that the closest structures deposited in the PDB are those of bicupin quercetin dioxygenases. The dimer of CucA is present both in the open (A) and closed (B) forms of $Cu^{2+}$-CucA with the dimer interface not precluding the opening and closing of the protein. When the open and closed structures are compared, the unmodelled region of the open protein contains a section that is helical in the closed form. Whilst it is likely that the majority of this region maintains its helical structure, it is interesting to note that the C-terminal region of this helix is annotated as a $3_{10}$ helix in both forms, and the carbonyl oxygen of $Ala_{242}$ forms hydrogen bonds both with the N-H of the n+3 residue $Tyr_{245}$ and the n+5 residue $Ile_{247}$ creating a very stable entry point for the helix to the rest of the protein.



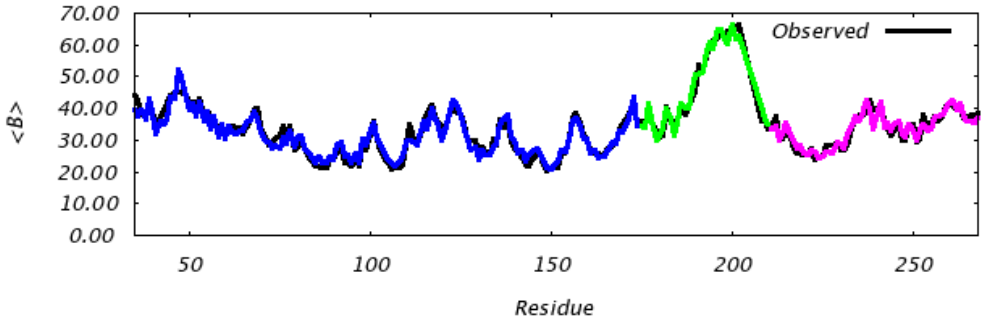
**Supplementary figure 2 - Solvent accessibility of the metal-binding site in the open crystal form.**

The open form of the protein reveals a channel to the metal binding site, also expected to act as the substrate entry channel. The entry to the pocket is lined by $Glu_{58}$, $Phe_{60}$, $Gly_{84}$, $Pro_{85}$ and $Phe_{238}$ with a single string of water molecules connecting the metal binding site to the surface. Although the substrate is not known, CucA exhibits quercetin dioxygenase activity, dependent upon $Cu^{2+}$ and not $Zn^{2+}$. Even when open, the entrance to the pocket is relatively narrow compared with that of quercetin dioxygenase (PDB ID = 1JUH; Fusetti *et al.*, 2002), and the pocket itself also smaller. Though the majority of the hydrophobic residues lining the pocket of CucA are structurally well conserved with those of quercetin dioxygenase (F151/F114, F97/F75, F185/F136, F60/Y35, P85/V63, M181/F132) some provide more bulky side chains (Y184/L135, W165/I127). More significantly, $Phe_{163}$ replaces a glycine (125) and, at the opposite side of the pocket, main chain differences result in the projection of $Gln_{220}$ and $Phe_{209}$ towards the pocket, also limiting the dimensions of space available for substrate. These restrictions suggest that the natural substrate would comprise only one or two rings, with perhaps additional aliphatic moieties.



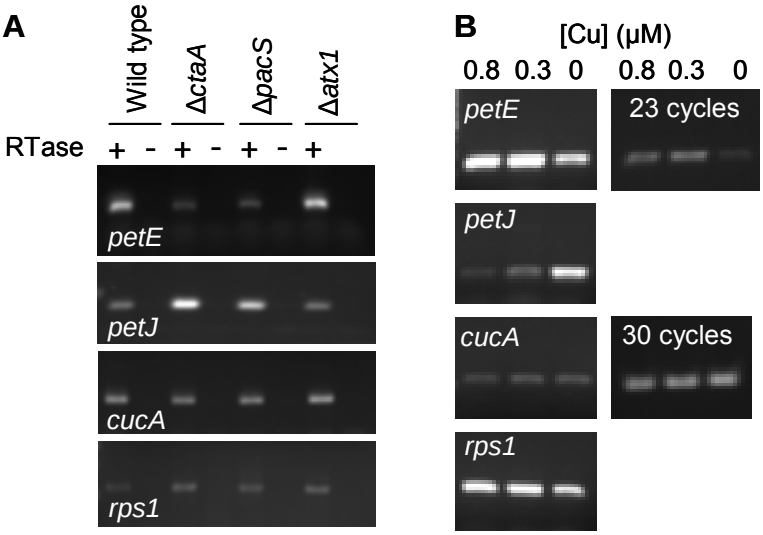
**Supplementary figure 3 - Solvent inaccessibility of the metal-binding site in the closed crystal form.**

In the closed form of the protein the channel is occluded by the movement of residues 52-60, resulting in $Glu_{58}$ and $Ala_{56}$ now projecting into the channel. The absence of a clear channel to the active site in the closed form suggests that while this opening and closing may play a role in selectivity of the metal, it is also relevant for presumed substrate binding and product release.




**Supplementary figure 4 –Observed and TLS-calculated mean B-factor per residue.**

Summary of B-factors per residue for the open form of $Cu^{2+}$-CucA observed and calculated using the online translation/libration/screw (TLS) tool (Painter & Merritt, 2006). Note that the residue numbers refer to those observed in the crystal structure.




**Supplementary figure 5 - Reverse transcriptase-dependence of RT-PCR amplification products & absence of Cu-dependent regulation of *cucA*.**

(A) DNase I-treated RNA was subjected to reverse transcription (RT) using ImPromII RT kit (Promega) in the presence (+) and absence (-) of reverse transcriptase enzyme using random hexamer primers. Aliquots of each RT reaction were subjected to polymerase chain reactions (PCR) with *Taq* polymerase (30 cycles) using primers specific for the *petE*, *petJ*, *cucA* or *rps1* gene, and amplification products analysed by agarose (1% w/v) gel electrophoresis stained with ethidium bromide. The results show that all amplification products were dependent on the presence of reverse transcriptase enzyme in the RT reaction. (B) RT-PCR was performed using RNA prepared from *Synechocystis* PCC 6803 cultured in BG11-C (BG11 medium lacking Cu) supplemented with 0, 0.3 or 0.8 μM $CuSO_4$. All panels show amplification after 25 PCR cycles with *Taq* DNA polymerase, except where labeled otherwise.

**Supplementary table 2**

Results of Mascot search with peptide masses detected after LC-MS/MS for trypsin-digested SDS-PAGE band from wild type and Δ*ctaA* periplasmic extract, depicted in figure 9. Ion scores > 37 and > 36 for wild type and Δ*ctaA*, respectively, are considered significant ($p < 0.05$). LC-MS/MS was performed essentially as described by Malik *et al.*, 2009, but used an LTQ-FT mass spectrometer (Thermo-Fisher). Protein identification was performed using mascot MS/MS search tool (Matrix Science, London).

| Sample | Protein | $M_r$ (expt) | $M_r$ (calc) | ppm | Score |
|---|---|---|---|---|---|
| Wild type | CucA | 890.45 | 890.45 | 3.41 | 17 |
| | | 1065.53 | 1065.52 | 1.64 | 78 |
| | | 1222.60 | 1222.60 | -1.35 | 49 |
| | | 1235.59 | 1235.60 | -7.41 | 51 |
| | | 1522.75 | 1522.74 | 1.36 | 9 |
| | | 1701.78 | 1700.78 | 586 | 45 |
| | | 1710.88 | 1710.89 | -6.19 | 54 |
| | | 1711.88 | 1710.89 | 579 | 39 |
| | | 1711.88 | 1710.89 | 581 | 41 |
| *MOWSE score: 232; total queries matched: 9* | | | | | |
| Wild type | General secretion protein G | 830.46 | 830.46 | -6.88 | 39 |
| | | 830.46 | 830.46 | -0.04 | 42 |
| | | 938.46 | 938.46 | -0.67 | 38 |
| | | 938.46 | 938.46 | -0.47 | 39 |
| | | 1364.56 | 1364.57 | -3.10 | 46 |
| | | 1374.71 | 1374.71 | -1.44 | 25 |
| | | 1380.56 | 1380.56 | -1.59 | 43 |
| *MOWSE score: 151; total queries matched: 7* | | | | | |
| Wild type | Hypothetical protein, Slr0924 | 1413.80 | 1413.80 | 3.16 | 57 |
| *MOWSE score: 57; total queries matched: 1* | | | | | |
| Δ*ctaA* | General secretion protein G | 830.46 | 830.46 | -1.00 | 46 |
| | | 938.47 | 938.46 | 3.43 | 24 |
| *MOWSE score: 70, total queries matched: 2* | | | | | |
| Δ*ctaA* | Hypothetical protein, Sll1891 | 1425.66 | 1424.65 | 709 | 56 |
| *MOWSE score: 53, total queries matched: 2* | | | | | |
| Δ*ctaA* | Hypothetical protein, Slr0924 | 857.46 | 857.46 | 1.49 | 14 |
| | | 1413.80 | 1413.80 | 2.81 | 39 |

## Supplementary References


Fusetti, F., Schröter, K.H., Steiner, R.A., van Noort, P.I., Pijning, T., Rozeboom, H.J., Kalk, K.H., Egmond, M.R. & Dijkstra, B.W. (2002) *Structure* **10**, 259-268.

Holm, L. & Rosenström, P. (2010) *Nucl. Acids Res.* **38**, 545-549.

Malik, R., Lenobel, R., Santamaria, A., Ries, A., Nigg, E.A. & Körner, R. (2009) *J. Proteome Res.* **8**, 4553-4563.

Painter, J. & Merritt, E.A. (2006), *Acta Cryst. D* **62**, 439-450.