HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2017 January 24.

Published in final edited form as:

Biochemistry. 2016 April 19; 55(15): 2278–2290. doi:10.1021/acs.biochem.6b00175.

The CopC Family: Structural and Bioinformatic Insights into a Diverse Group of Periplasmic Copper Binding Proteins

Thomas J. Lawton, **Grace E. Kenney**, **Joseph D. Hurley**, and **Amy C. Rosenzweig***

Departments of Molecular Biosciences and of Chemistry, Northwestern University, Evanston, Illinois 60208, United States

Abstract

The CopC proteins are periplasmic copper binding proteins believed to play a role in bacterial copper homeostasis. Previous studies have focused on CopCs that are part of seven-protein Cop or Pco systems involved in copper resistance. These canonical CopCs contain distinct Cu(I) and Cu(II) binding sites. Mounting evidence suggests that CopCs are more widely distributed, often present only with the CopD inner membrane protein, frequently as a fusion protein, and that the CopC and CopD proteins together function in the uptake of copper to the cytoplasm. In the methanotroph *Methylosinus trichosporium* OB3b, genes encoding a CopCD pair are located adjacent to the particulate methane monooxygenase (pMMO) operon. The CopC from this organism (*Mst*-CopC) was expressed, purified, and structurally characterized. The 1.46 Å resolution crystal structure of *Mst*-CopC reveals a single Cu(II) binding site with coordination somewhat different from that in canonical CopCs, and the absence of a Cu(I) binding site. Extensive bioinformatic analyses indicate that the majority of CopCs in fact contain only a Cu(II) site, with just 10% of sequences corresponding to the canonical two-site CopC. Accordingly, a

***Corresponding Author:** amyr@northwestern.edu. Telephone: (847) 467-5301.

S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.6b00175.

Heatmap derived from hierarchical clustering of copC genes and traits in the unpruned data set (Figure S1) (PDF)

Fasta-formatted file, including all CopC sequences (ZIP, Supplemental File 1)

Excel file containing metadata on all CopC sequences (XLSX, Supplemental File 2)

Fasta-formatted file, including the pruned CopC data set (ZIP, Supplemental File 3)

Excel file containing metadata on the pruned CopC data set (XLSX, Supplemental File 4)

HMM-aligned CopC subgroup C_{0-0}, Stockholm-formatted (ZIP, Supplemental File 5)

HMM-aligned CopC subgroup C_{0-1}, Stockholm-formatted (ZIP, Supplemental File 6)

HMM-aligned CopC subgroup C_{0-2}, Stockholm-formatted (ZIP, Supplemental File 7)

HMM-aligned CopC subgroup C_{1-0}, Stockholm-formatted (ZIP, Supplemental File 8)

HMM-aligned CopC subgroup C_{1-1}, Stockholm-formatted (ZIP, Supplemental File 9)

Raw hierarchical clustering data, pruned CopC data set (XLSX, Supplemental File 10)

Raw hierarchical clustering data, unpruned CopC data set (ZIP, Supplemental File 11)

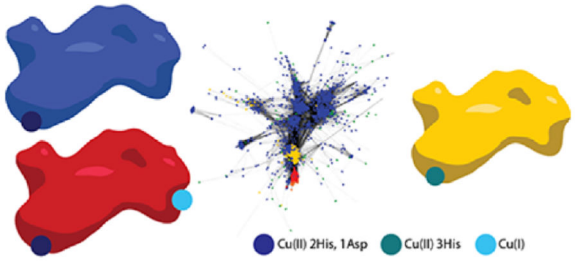
Accession Codes

Protein Data Bank entry 5ICU.

The authors declare no competing financial interest.

new classification scheme for CopCs was developed, and detailed analyses of the sequences and their genomic neighborhoods reveal new proteins potentially involved in copper homeostasis, providing a framework for expanded models of CopCD function.

Graphical abstract



Copper is an essential but toxic metal in biology.[1,2] It is a critical cofactor in proteins such as cytochrome c oxidase, superoxide dismutase, and numerous oxygenases, but excess copper causes oxidative damage and may disrupt iron–sulfur clusters.[1–4] To satisfy intracellular copper requirements and avoid toxicity, bacteria have evolved multiple copper homeostasis pathways. Strikingly, more than 44% of the copper proteome is dedicated to copper homeostasis.[5] In prokaryotes, there are at least three systems that impart copper resistance: the Cue system (Cu efflux), the Cus system (Cu sensing), and the Cop/Pco system (copper resistance, or plasmid-borne copper resistance).[6–9]

The Cue and Cus systems are generally believed to function in copper export and detoxification.[6,7] The Cue system typically comprises three proteins: a P_{1B}-type ATPase (often annotated as CopA and herein termed CopA*) that pumps Cu(I) from the cytoplasm to the periplasm, the periplasmic multicopper oxidase CueO, which couples the oxidation of Cu(I) to the reduction of molecular oxygen, and the cytoplasmic copper binding regulatory protein CueR.[10–12] In contrast to the Cue system, the Cus system does not rely on the reduction of oxygen and is therefore capable of functioning under anaerobic conditions. This system includes six proteins. The CusABC complex spans the entire periplasmic envelope and effluxes Cu(I) ions, driven by a proton gradient.[13] CusF delivers Cu(I) ions to the CusABC complex,[14] and expression of the Cus proteins is regulated by the CusR/CusS two-component system.[15]

While the Cue and Cus systems are strongly implicated in copper efflux, it is not yet clear how the Cop/Pco system imparts copper resistance to the cell. It has been suggested that the Cop proteins export copper,[9] import copper to the cytoplasm,[2,16] and/or sequester copper in the periplasm.[17,18] The possibility of import has been relatively underinvestigated, likely due to the fact that there is no known function for copper in the bacterial cytoplasm beyond its involvement in copper homeostasis and gene regulation.[2] The traditional Cop/Pco system is composed of seven proteins, CopABCDERS. CopD and CopB are inner and outer membrane proteins, respectively (CopB should not be confused with the identically named copper-transporting P_{1B}-type ATPase, not traditionally part of the *cop* operon).[8] CopA is a periplasmic multicopper oxidase that can substitute for CueO.[19] CopE is a soluble protein that binds Cu(I) and is hypothesized to sequester copper in the periplasm.[20] CopC is

believed to be a periplasmic copper chaperone.[21] The PcoABCD proteins are homologous to the CopABCD proteins but were originally identified on plasmids; annotations may show either set of names. Like the *cus* operon, the canonical *cop* operon is regulated by a two-component system, the CopRS proteins.[19]

The proposed physiological functions of the Cop/Pco systems are based largely on phenotypic characterization of strains expressing different combinations of the cop genes and on biophysical characterization of CopC/PcoC.[8,22,23] Knockouts of the individual *cop/pco* genes result in increased copper sensitivity, suggesting that all four proteins work in concert to impart copper resistance.[23,24] Counterintuitively, expression of *copCD* without *copAB* leads to copper hypersensitivity, suggesting that CopD imports copper and that CopD-mediated copper import is facilitated by CopC.[25] A definitive model for the functions and interactions of the Pco/Cop proteins remains elusive, however. Interpretation of phenotypic data has been further complicated by biophysical studies of traditional CopC/PcoC proteins. Until recently, CopCs were generally thought to contain both a Cu(I) binding site and a Cu(II) binding site.[21,26–28] Structural studies show that the Cu(I) ion binds in a methionine rich loop, and that the Cu(II) ion is coordinated by the amino terminus, two histidines, and a water molecule. One of the coordinating histidines is the N-terminal residue, which is generated by cleavage of a signal peptide, while the other histidine is part of a DXH motif in which the aspartic acid hydrogen bonds to the coordinating water molecule.[21,26,29,30] Biochemical studies with PcoC indicate that the Cu(I) and Cu(II) sites are capable of intermolecular copper transfer and that PcoA and CueO are capable of oxidizing the Cu(I) site.[21,28,31]

Over the years, evidence that Cop/Pco systems are significantly more diverse and less similar to the traditional *cue* and *cus* copper efflux systems than was believed originally has accumulated. First, a CopC lacking the Cu(I) binding site, but with an unusually high affinity for Cu(II), has been characterized.[32] The higher affinity is believed to derive from an additional coordinating histidine, which is present in the third position of the mature peptide. Second, proteins that appear to be fusions of CopC and CopD have been reported in the genomes of numerous organisms, and there is strong biochemical evidence that these Cop-like systems are involved in the import of copper into the cytoplasm.[16,33,34] Finally, operons containing *copC* and *copD* without *copA*, *copB*, or *copE* have been observed in multiple organisms.[18,33,35] For example, in methanotrophic bacteria, *copC* and *copD* genes are located adjacent to the operon encoding particulate methane monooxygenase (pMMO), a copper-containing enzyme that catalyzes the first step of methane metabolism.[36] Notably, the *copD* gene in a mutant strain of *Methylosinus trichosporium* OB3b that is impaired in cytoplasmic copper uptake and regulation contains a frameshift deletion, consistent with CopC and CopD functioning in copper import in this organism.[37]

To further understand the methanotroph CopCD system, a CopC from *M. trichosporium* OB3b (*Mst*-CopC) was structurally characterized, and an extensive bioinformatics study was undertaken to understand *Mst*-CopC in the broader context of the CopC protein family. A total of 8636 CopC homologues were examined, providing an overview of the diversity of CopC sequences and operons. We find that CopCs containing two copper binding sites (named Type A by Wijekoon et al.[32]) are relatively rare, accounting for approximately 10%

of CopCs. The vast majority of CopCs contain only a Cu(II) site similar to that in either Mst-CopC or the recently characterized higher-affinity *Pseudomonas fluorescens* SBW25 CopC (*Psf*-CopC, named Type B by Wijekoon et al.).[32] These data, in conjunction with the crystal structure of the newly identified CopC type, warrant the creation of a new system for classifying CopC-like proteins. Moreover, when combined with a growing body of evidence of CopCD-mediated copper uptake, these data necessitate a re-examination of the existing models of CopCD function.

MATERIALS AND METHODS

Materials

Unless otherwise stated, all chemicals were purchased from VWR, Research Products International, or Sigma-Aldrich and were not purified further before being used.

Construction of the Expression Plasmid

The *Mst-copC* gene (IMG gene ID 2507406428, locus tag Mettr-DRAFT_0380) was codon-optimized using *Escherichia coli* codon frequencies, synthesized by GenScript, and subcloned into pE-SUMO (Life Sensors) using the BsaI restriction site. This sequence contains a C-terminal tag encoding a TEV restriction site, a sequence corresponding to β-strand 11 of green fluorescent protein,[38] and a *Strep*-Tactin (IBA) affinity tag in that order. In place of the native N-terminal signal peptide sequence, an N-terminally encoded His-SUMO tag was introduced, which allows for generation of the authentic N-terminus (His 23) after the protein is incubated with SUMO protease.[39]

Protein Expression and Purification

The resultant vector was transformed into chemically competent BL21 Star (DE3) (Life Technologies) cells, and one colony was used to inoculate a 1 L culture of autoinduction medium.[40] Cultures were shaken at 37 °C until they reached an OD_{600} of ~0.5, at which point the temperature of the incubator was lowered to 20 °C. Following overnight growth, cells were harvested by centrifugation for 15 min at 7100g and flash-frozen for later use.

Cells were suspended in Tris buffer [50 mM Tris and 500 mM NaCl (pH 8.0)] and lysed by sonication on ice for 8 min using 1 s pulses with 3 s rests. Lysates were centrifuged for 1 h at 186000g and applied to a 15 mL *Strep*-Tactin affinity column (IBA) that had been pre-equilibrated with equilibration buffer [50 mM Tris and 500 mM NaCl (pH 8.0)]. Following extensive washing with equilibration buffer, *Mst*-CopC was eluted from the *Strep*-Tactin column with elution buffer [50 mM Tris, 500 mM NaCl, and 2 mM desthiobiotin (pH 8.0)].

To cleave the SUMO tag, fractions containing *Mst*-CopC were incubated in TCEP-containing Tris buffer [50 mM Tris, 500 mM NaCl, and 2 mM TCEP (pH 8.0)] at 37 °C overnight with excess TEV and SUMO proteases that were prepared using previously published protocols.[39,41] Cleaved *Mst*-CopC containing the native amino-terminal histidine residue was then applied to a Cu-loaded 5 mL HiTrap Chelating HP column (GE) to remove the SUMO protein and the His-tagged proteases. *Mst*-CopC was then exchanged into 100 mM $NaPO_4$ and 500 mM NaCl (pH 7.0) and purified using a 24 mL Superdex 200 column

(GE). Protein purity was assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis using a 15% gel, and protein concentrations were measured using the calculated molar extinction coefficient at 280 nm (5960 M^{-1} cm^{-1}).[42] Protein yields of 10–50 mg/L of culture were obtained.

Copper Loading and Metal Analysis

Purified *Mst*-CopC in 100 mM $NaPO_4$ and 500 mM NaCl (pH 7.0) was incubated with 10 equiv of $CuSO_4$ at room temperature for several minutes. Samples were then centrifuged to remove precipitated copper(II) phosphate, and unbound copper was removed by buffer exchange [into 100 mM $NaPO_4$ and 500 mM NaCl (pH 7.0)] on a 24 mL Superdex 200 column (GE). Metal content was measured by inductively coupled plasma mass spectrometry (ICP-MS, Thermo iCAP Q). The concentrations of iron, cobalt, nickel, copper, and zinc were measured using a multielement standard (Inorganic Ventures). ICP-MS samples were prepared in 5% nitric acid with 5 ppb indium, lithium, scandium, and yttrium internal standards (Inorganic Ventures).

Crystallization and Structure Determination

Mst-CopC was crystallized by sitting drop vapor diffusion at room temperature by mixing 1 μL of 10 mg/mL protein with 1 μL of a well solution containing 0.1 M HEPES sodium salt, 0.8 M sodium phosphate, and 0.8 M potassium phosphate (pH 7.5). Football-shaped crystals grew within 1 week. All crystals were cryoprotected in a well solution containing 50% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected at LS-CAT Sector 21 and GM/CA Sector 23 of the Advanced Photon Source at Argonne National Laboratory (Argonne, IL) at an energy of 12 keV. HKL-2000 was used to index, integrate, and scale all data sets.[43] Molecular replacement using *E. coli* PcoC [*Esc*-PcoC, Protein Data Bank (PDB) entry 1LYQ][26] as a starting model was used to determine phases. Molecular replacement solutions were calculated using Phaser, and structures were refined using REFMAC as implemented in the CCP4 program suite (Table 1).[44–46] During the final stages of refinement, TLS and anisotropic refinement parameters for the copper ion were added to the model.

Identification of CopC Homologues in Bacterial Genomes

All protein sequences matching the CopC PFAM profile hidden Markov model (HMM) (PF04234) and associated metadata were obtained from the JGI-IMG database (http://img.jgi.doe.gov/w/, September 17, 2014) (Supplemental Files 1 and 2). To avoid the loss of potentially relevant minor differences at copper binding sites, the data set was analyzed with and without pruning on the basis of sequence similarity. In the pruned data set (Supplemental Files 3 and 4), sequences were clustered according to similarity using UClust,[47] and in groups with >95% similarity, a centroid sequence was chosen to represent the cluster, resulting in a smaller data set of 2550 sequences. Unless otherwise specified, this is the data set discussed in the body of this paper, and discussion of the auxiliary unpruned data is relegated to the Supporting Information. All sequences were aligned against the HMM via hmmalign (Supplemental Files 5–9),[48] and additional domains not matching the HMM were trimmed for further sequence-based analyses.

Sequences were classified into one of several categories depending on the conservation of their Cu(I) and Cu(II) sites (Table 2). The Cu(I) site was considered to be present if, after alignment with the CopC HMM, residues corresponding to at least four of the five residues, M63, M66, M69, H72, and M75 (using *Esc*-PcoC residue numbering, IMG entry 640940805), were conserved. In the case of H69 and M75, the presence of an M or H within one residue of that position was considered sufficient for conservation of the residue. The Cu(II) site was considered to be present if, after alignment with the CopC HMM, residues corresponding to H24, D113, and H115 were conserved. Sequences containing a Cu(II) site with an additional conserved histidine at H26 were assigned to a separate category, higher-affinity Cu(II) binding sequences (using *Esc*-PcoC residue numbering, IMG entry 640940805). Systemic misalignments appear to be minimal or due to the presence of significant insertions, deletions, or multiple-residue differences in sequence with the potential to alter the structure or function of the resulting proteins. For the generation of consensus sequences in Jalview,[49] only residues fitting the HMM were used [along with residues or gaps in the alignment corresponding to *Esc*-PcoC M51, which is part of the Cu(I) binding site].

Network Analysis of CopC Sequences

Sequence similarity networks were generated as described previously.[50] Nodes represent the trimmed CopC amino acid sequences, and edge values were generated via an all-by-all BLAST+ (NCBI) against a database consisting of either the 8636 raw or the 2500 pruned CopC sequences (Supplemental Files 2–6), trimmed but unaligned. Initial networks were generated using the Blast2Sim plug-in[51] in Cytoscape 2.8.3[52] with an initial BLAST threshold of 100, a coverage factor of 15, and a “sum of all hits” similarity function. Edges corresponding to low similarity were removed according to a numeric filter; filter values were set below a level that would result in >1% of the nodes dissociating from the main cluster. Transitivity clustering was applied via the TransClust plugin, using a threshold of 35 with the Blast2SimGraph_sim score as an edge value and similarity as an edge weight determiner.[51] In the final cluster analysis, edges with a value below 35 were deleted and the network was organized according to the edge-weighted, force-directed Cytoscape layout.

A tab-delimited text file identifying characteristics of interest for each node was loaded and overlaid onto the network via the Cytoscape VizMapper plug-in. Copper site assignment was calculated as described above. The genus and gene length were obtained from exported JGI metadata for each gene, while the phylum was determined using the NCBI Taxonomy Database on the basis of genus (http://www.ncbi.nlm.nih.gov/taxonomy). Operon categories were assigned as described below.

Hierarchical Clustering of CopC Genome Neighborhoods

In the JGI-IMG database, each gene has a gene object id (gene_oids). On a DNA contig, scaffold, or chromosome, neighboring ORFs are assigned contiguous numeric gene oids. On the basis of this information, metadata were obtained for all genes within five ORFs of *copC* (i.e., all genes with gene_oids ±5 from the gene_oid of the relevant *copC*). Neighboring genes were classified by annotated Pfam[53] and TIGRFAM[54] profile HMMs present in the gene, and for genes observed near >1% of *copCs* in a given position, the distance in number

of genes from *copC* was recorded, from 0 (indicating fusion with *copC*) to 5. A binary value was assigned to indicate the presence or absence of a given trait at a given distance from the *copC* gene; in a separate set of analyses, a binary value was assigned to the presence or absence of those traits in the genomic neighborhood of *copC* regardless of their relative positions. Several other traits were also assigned binary numeric values for each gene: presence of a given type of copper site, membership in one of the 15 most common genera in this data set, and membership in the most common phyla. Only gene presence or presence/distance values and copper site values were used for hierarchical clustering.

The gene IDs and their associated traits were subjected to hierarchical clustering in R.[55] For both genes and traits, the distance matrix was calculated using the Manhattan method (absolute distance) in the *dist* function, and the clustering was determined using the Ward.D2 method[56] in the *hclust* function (Supplemental File 10). A heatmap and dendrograms were generated from the hierarchical clustering data via the heatmap.2 function in gplots.[57] Co-clustered sets of traits were presumed to represent potential operons, and five randomly selected *copCs* from each of the 20 largest operon families were checked to visually confirm the clustering results. Similar sets of analyses were performed using the unpruned CopC data set (Figure S1, SI Supplemental File 11). Only genes with unidentified roles or with roles plausibly related to copper homeostasis were considered to be potential operon members, although all gene neighbors are represented in the raw data in the Supporting Information (Supplemental File 5). In some clusters with more morphological variation, only the core operon structure observed in the significant majority of the cluster is depicted.

RESULTS AND DISCUSSION

CopC from *M. trichosporium* OB3b

The cloned *copC* gene used in this work is located downstream of the *pmoCAB* operon, which encodes the three subunits of pMMO, between a gene termed *pmoD* and the *copD* gene.[58] The *copD* gene is followed by a gene encoding a putative member of the DUF461 family, which includes the proteins known as PCu_AC, PcuC, and EcuC.[35,59–61] This entire gene cluster forms an operon that is coregulated in response to copper.[37] The predicted *Mst*-CopC protein includes an N-terminal signal peptide for export to the periplasm. Previous studies of CopC proteins indicate that cleavage of the signal peptide results in an N-terminal histidine, which coordinates a Cu(II) ion with both the N-terminal amino group and the imidazole side chain.[21] Accordingly, the N-terminal histidine in *Mst*-CopC was generated by genetically inserting a SUMO tag before the *copC* gene. Cleavage of this tag yields the native N-terminus, and ICP-MS analysis of the resulting protein indicates that 1.0 ± 0.3 (n = 4) copper ions bind per monomer after metal loading.

The crystal structure of *Mst*-CopC was determined to 1.46 Å resolution in space group $P2_12_12_1$ with a single molecule in the asymmetric unit (Table 1 and Figure 1). There is no evidence of biologically relevant oligomerization in the crystal packing or for interference of crystal lattice contacts with the metal center. All regions of the protein are well-ordered, and all of the residues following the signal peptide (residues 23–120) in addition to a C-terminal TEV cleavage site were modeled. Similar to previously characterized CopCs, *Mst*-CopC

forms a seven-stranded β-barrel, albeit with a slightly different arrangement of β-strands. The crystal structure reveals a single copper ion, modeled at an occupancy of 1.0.

Two CopCs have previously been structurally characterized: PcoC[62] encoded on plasmid PRJ1004[23,63] found in an *E. coli* strain isolated from a piggery (*Esc*-PcoC)[26] and CopC[64] from *Pseudomonas syringae* pv. tomato (*Pss*-CopC).[21,24,29,30] These CopCs contain Cu(I) and Cu(II) binding sites, located at opposite ends of the molecule. In *Pss*-CopC, the Cu(I) ion is coordinated by methionine(s) and a histidine (Figures 1 and 2), with variable coordination observed in crystal structures obtained at different pH values and in different crystallographic space groups.[21] The Cu(II) ion is coordinated in square planar geometry by two histidines, the amino terminus, and a water molecule, with an aspartic acid within hydrogen bonding distance of the ligands (Figure 2B). At low pH values in a different space group, an altered coordination involving a histidine from a second molecule is also observed. Copper coordination was not observed in the *Esc*-PcoC structure because the recombinantly produced protein used for crystallography did not include authentic N-terminal histidine, which is found in the native protein.

Mst-CopC contains only the Cu(II) binding site and completely lacks the conserved residues necessary for Cu(I) binding (Figure 1). The residues near the Cu(II) site, H23, D105, H107, and the amino terminus, are similar to those observed in *Pss*-CopC[21] but coordinate Cu(II) differently (Figures 2 and 3A). In *Mst*-CopC, Cu(II) is bound in a distorted square pyramidal geometry with H23, H107, D105, and the amino terminus comprising the planar ligands and a water molecule serving as the axial ligand (Figure 3B). The D105 coordinating distance is 2.30 Å, whereas the histidine nitrogen to copper distances range from 1.96 to 2.06 Å. The 2.30 Å distance is likely due to anisotropy along the D105 H23 imidazole axis, as evidenced by egg-shaped electron density in the copper omit map (Figure 3B). The distance between the copper ion and water molecule is 2.37 Å. This elongation is likely a result of Jahn-Teller distortion. Similar to *Met*-CopC, *P. fluorescens* SBW25 CopC (*Psf*-CopC) also contains only a Cu(II) binding site, which was characterized biochemically and spectroscopically. The *Psf*-CopC Cu(II) site has a Cu(II) affinity 2 orders of magnitude higher than that of Pss-CopC, which was attributed to the presence of an additional histidine ligand in the third sequence position of the mature peptide (Figure 4).[32] Notably, this residue is absent in *Mst*-CopC. Thus, *Mst*-CopC appears to be a Cu(II)-only chaperone distinct from the two types of CopCs previously characterized: those that bind both Cu(I) and Cu(II) (e.g., *Pss*-CopC and *Esc*-PcoC) and those that bind only Cu(II) with an additional histidine ligand (e.g., *Psf*-CopC).

The fold of *Mst*-CopC is the same as that of *Pss*-CopC and *Esc*-PcoC, but the β-strands adopt a slightly different arrangement (Figures 1 and 2). In particular, a “dog leg”[65] is introduced into β-strand 4, resulting in secondary structure contacts with β-strands 3 and 5 and the formation of a tighter loop between β-strands 3 and 4 (Figure 1). The contacts between β-strand 3 and the “dog leg” consist of backbone hydrogen bonds from a single pair of amino acids (labeled β-strand 4B in Figures 1 and 2), whereas four pairs of amino acids mediate the interaction between the “dog leg” and β-strand 5 (labeled β-strand 4A in Figures 1 and 2). By contrast, β-strand 4 of *Pss*-CopC contacts only β-strand 5, resulting in a more extensive loop that also includes the Cu(I) binding residues. It has been suggested that the Cu(I) bound in this loop can be oxidized by periplasmic oxidases,[28] and the extended loop

may play a role in protein-protein interactions between Cu(I) binding CopCs and cognate oxidases or other partner proteins. Although the overall structure of all three CopCs is similar, the differences in metal binding site arrangement and ligands are consistent with *Mst*-CopC representing a new type of CopC.

Bioinformatic Classification of CopC Subfamilies

To determine if the *Mst*-CopC type is unique to methanotrophs or more prevalent, we performed an extensive bioinformatics study. At the time of analysis, the JGI-IMG database contained 8636 genes predicted to include a CopC-like domain. After alignment against the CopC profile HMM (PF04234), these genes were categorized on the basis of the composition of the copper sites, considering each of the two sites independently. This classification scheme is more comprehensive than previous analyses, in which CopCs were identified on the basis of the presence or absence of the Cu(I) binding site.[32] We have assigned each sequence according to whether the Cu(I) site is present in the canonical form represented by the *Pss*-CopC and *Esc*-PcoC ligand set (1) or absent (0) and defined three categories for the Cu(II) site: absent (0), present in the canonical/*Mst*-CopC form (1), or present in the higher-affinity form with an additional histidine ligand (2) (Table 2 and Figure 4). Using this nomenclature, we have given each CopC class a designation consisting of "C" followed by two subscripted digits, of which the first identifies the Cu(I) site type and the second identifies the Cu(II) site type, resulting in five categories: C_{0-1}, C_{0-2}, C_{1-1}, C_{0-0}, and C_{1-0}, in order of prevalence (no C_{1-2} CopCs were identified). The sequences were also grouped into CopC supersets, including combinations of multiple types (Table 2). Surprisingly, this analysis indicates that C_{0-1} CopCs, of which *Mst*-CopC is the founding member, are the most abundant (71.8%), followed closely by members of the C_{0-2} class (13.0%). Thus, the overwhelming majority of CopCs contain only the Cu(II) site, and only 5.4% of the sequences belong to the C_{1-1} class, upon which most previous studies and functional hypotheses have been based. A total of 9.3% of the sequences lack one or more ligands to either copper site (C_{0-0}), and a very few (0.4%) appear to have only the Cu(I) site (C_{1-0}).

To improve our understanding of this newly discovered CopC diversity, similarity network analysis was performed using nodes composed of trimmed CopC sequences and edge values generated by an all-by-all BLAST. The resulting clusters were examined for the following traits: CopC type, source organism phylum, and length as a proxy for detecting fusions to other proteins (Figure 5). The results indicate that type C_{0-1} is the most widely distributed, appearing in the actinobacteria, proteobacteria, and firmicutes; it is clearly not limited to methanotrophs. By contrast, C_{0-2} and C_{1-1} CopCs are mostly confined to three clusters within the Proteobacteria phylum (Figure 5A, B). The C_{0-2} CopCs are found primarily in gamma-proteobacterial genera like *Escherichia* and *Pseudomonas*.[32]

In terms of predicted final polypeptide length, the C_{1-1} and C_{0-2} CopCs are typically 90–100 amino acids and do not contain additional domains (Figure 5C). Although some C_{0-1} CopC sequences, including the *Mst*-CopC sequence, are similar in length and composition to the C_{1-1} and C_{0-2} CopCs, a significant subset of these sequences are fused to CopD homologues. These CopCD fusions are especially common in Gram-positive bacteria. The

YcnJ CopCD fusion protein from *Bacillus subtilis* has been investigated by gene disruption and copper regulation studies as well as whole cell copper uptake experiments.[16,33] Some of the fused genes encode additional C-terminal domains, in particular the functionally uncharacterized YktA domain (PF05256); both the fused CopC and the C-terminal domains are predicted to be periplasmic in Gram-negative bacteria and external to the cell in Gram-positive bacteria. In Gram-positive bacteria, C_{0-1} CopC sequences not fused to CopD are frequently longer than CopCs in other classes likely due to the need for an N-terminal transmembrane region that tethers the protein to the external surface of the Gram-positive cell membrane (Figure 5C).

Genomic Neighborhood Analysis

To gain insight into CopC functional diversity, we analyzed the genomic neighborhoods of all the *copC* genes. For each *copC* gene, predicted genes within a five-gene radius were identified via profile hidden Markov models (TIGRFAMs and PFAMs), as were copper site types and the presence of any fusion proteins. These traits and the CopCs associated with them were submitted to hierarchical clustering to identify conserved gene groups and sets of CopCs with similar genomic neighborhoods (Figure 6). In addition, investigation of particularly common clusters allowed us to eliminate genes likely to be irrelevant to copper homeostasis on the basis of predicted function, relative distance and direction, and membership in an overrepresented genus. Examination of the 21 remaining common operon types highlights possible functional differences between CopC subfamilies as well as genes encoding several previously uninvestigated proteins that may be involved in copper homeostasis (Figures 5D and 6).

The C_{1-1} *copC* genes are almost invariably found preceding *copD* and are frequently found in operons containing other *cop* machinery (Figures 5 and 6). Common examples include *copABCDRS*, *copBARSA*CD*, and *copAB(DUF411) (DUF2933)A*CD*. Proximity to *czc* and *cus* metal homeostasis operons is common, as is the existence on a plasmid rather than a chromosomal operon. Of special note in these putative *cop* operons are *DUF411* and *DUF2933* gene sequences. DUF411 has occasionally been conflated with several different regulatory proteins (including the CopY family repressor CopR[66] and the characterized repressor CopG)[67,68] and as a potential exporter;[69] the sequences of DUF411 proteins support a third role, namely a periplasmic copper binding protein.[70,71] It has been hypothesized that DUF411 is involved in copper resistance,[70–72] whereas DUF2933 (occasionally annotated as CopO) is completely uncharacterized. Both hypothetical proteins also appear in the proximity of *cus* operons. The 11 type C_{1-0} *copC* genes are found in operons that are similar to C_{1-1} *copC*-containing operons. It is not clear that C_{1-0} CopCs are a significant and distinct group of CopCs; rather, they seem to be the result of mutations to the Cu(II) site in otherwise C_{1-1} CopCs.

The C_{0-2} *copC* genes, which are frequently (although not invariably) annotated as *yobA*,[6] occur almost exclusively before separate *copD* genes (often annotated as *yebZ*). Sometimes these two genes comprise the entire operon, but in many cases, the *copD* homologue is followed by an uncharacterized *DUF2511* gene sequence (sometimes annotated as *yebY*) (Figure S1 and Figure 6). DUF2511 is predicted to be a small periplasmic protein, and the

proximity of its gene to those encoding CopC and CopD homologues suggests that it too may be involved in copper homeostasis. However, unlike the C_{1-1} *copC* genes, C_{0-2} *copC* genes do not appear in the genomic proximity of other *cop* genes, and they are broadly chromosomal rather than plasmid-based.

The genome neighborhoods of the C_{0-1} *copC* genes, which include *Mst-CopC*, are quite varied. This diversity is consistent with the range of species containing these genes (Figure 6). DUF1775 sequences (PF07987, sometimes annotated *ycnI*) are frequently adjacent to *copC* or *copD* genes, and this protein is produced in response to copper in some species, though the function remains unknown.[73] Two other members of the ycn operon are sometimes observed, although primarily in *Bacillus* species and related genera; examples include the DeoR family regulator *ycnK* (PF08220) and a reductase or disulfide isomerase sometimes annotated as *ycnL*. Other frequent neighbors across species and genera include genes encoding DyP-type peroxidases (PF04261), periplasmic copper binding proteins belonging to the DUF461 family (PF04314, also known as PCu_AC, PcuC, and EcuC),[61] and genes encoding other periplasmic copper chaperones resembling Sco1 and SenC (PF02630).[74] Perhaps because of their wide distribution across many species, likely aided by horizontal gene transfer, C_{0-1} *copC* genes occur without a neighboring *copD* gene at a frequency higher than those of other *copC* types (Figure 5D), although even these isolated *copC* genes frequently occur in the genomic vicinity of genes encoding DyP peroxidases, DUF461 proteins, Sco1 homologues, and YcnI homologues.

Unlike the type C_{1-0} CopCs, many type C_{0-0} CopCs appear to be distinct CopC variants. They lack one or more metal binding residues at the copper binding sites or have regions of divergent length or sequence composition that render comparison to the standard HMM difficult to interpret. Several clusters are seen within groups of related species (Figure 5), suggesting that these sequences have been maintained despite likely alterations in copper binding. The genes encoding C_{0-0} CopCs are found next to or fused with *copD* a little less than half the time. Given that many of CopCs encoded by these genes are unlikely to bind copper or to have Cu(I) or Cu(II) sites resembling those of established CopC families, their function remains ambiguous.

Functional Implications

The crystal structure of *Mst*-CopC and accompanying bioinformatic analysis reveal a new CopC subfamily, the C_{0-1} CopCs, characterized by the presence of a single Cu(II) binding site similar to those found in the previously characterized C_{1-1} CopCs.[21,26,29,30] This subfamily is broadly distributed throughout the microbial world and is in fact the most common form of CopC (Table 2). Combined with the higher-affinity Cu(II) C_{0-2} subfamily, CopCs that bind only Cu(II) comprise approximately 85% of all CopC family proteins. Thus, previous functional roles suggested for CopC, including delivery of Cu(I) to outer membrane components or to the multicopper oxidase CopA as a source of reducing equivalents,[75] as well as models involving intra- or intermolecular transfer of copper ions between the two sites in CopC or to other partners such as CopA, CopB, and CopD,[21,30] may not be applicable to the broader CopC family. Whereas the C_{1-1} *copC* genes are typically found in operons with these previously proposed partner proteins, the genomic

neighborhoods of C_{0-1} *copC* genes are variable. The *copD* gene is usually present, however, and in many instances, the encoded CopCD pairs or CopCD fusion proteins appear to be functional units by themselves. This observation is particularly notable given the increasingly strong evidence of the role of CopD as a copper importer, despite the lack of a well-understood nonregulatory role for copper in the cytoplasm of microorganisms.[16,18,25,33,37]

It is likely that the C_{0-1} and C_{0-2} CopCs are periplasmic Cu(II) chaperones for a Cu(II) import system that utilizes CopD for transport across the inner membrane followed by reduction to Cu(I) in the cytoplasm.[2,16] These CopCs could also donate Cu(II) to other proteins in the periplasm, including Cop proteins. In this scenario, the C_{1-1} CopCs would function similarly and may also participate in Cu(I) detoxification. As noted previously, the presence of the two distinct sites for Cu(I) and Cu(II) is unusual,[30] and Cu(I) binding could serve a specialized function in some organisms. Alternatively, the methionine rich Cu(I) binding site could function in protein-protein interactions,[26] not necessarily always binding Cu(I) in vivo. Methionine residues are often found in patches at protein interaction interfaces.[76,77]

The full range of functional roles for CopC and CopD remains unclear. In some systems, copper imported by CopD clearly interacts with cytoplasmic regulatory proteins. For example, copper imported by B. *subtilis* YcnJ, a CopCD fusion protein, binds the copper-dependent regulator CsoR, resulting in expression of the *copZA* genes encoding a cytoplasmic copper efflux system, which pumps copper to the extracellular matrix.[33] In *P. fluorescens* SBW25, the Cop and Cue systems work in concert to regulate copper levels.[18] In support of a similar model in methanotrophs, a frameshift/deletion in the *copD* gene adjacent to the gene encoding *Mst*-CopC results in a copper-starved phenotype, suggesting that copper is not reaching cytoplasmic regulators.[37] This type of model likely applies to other systems and diverse cytoplasmic regulators. For example, in *E. coli*, copper imported by CopC and CopD could activate the Cue system as well as the outer membrane protein ComC, which reduces the permeability of the outer membrane to copper. These systems are regulated by the copper binding transcription factors CueR[78] and ComR,[79] respectively.

Beyond import for regulation, it is also possible that CopC and CopD import copper into the cytoplasm for protein assembly. In *Rhodobacter capsulatus*, assembly of active cbb_3-type cytochrome *c* oxidase (COX) depends on CcoA, a member of the major facilitator superfamily involved in the uptake of copper to the cytoplasm.[80,81] The frequent appearance of periplasmic copper proteins implicated in COX assembly, including DUF461 and Sco1 homologues,[35,59–61,82–84] in the genomic neighborhood of C_{0-1} *copC* and *copD* genes suggests that CopD could function analogously to CcoA. In support of this notion, at least 15% of *copD* gene sequences contain a C-terminally fused domain annotated as *ctaG* (not to be confused with the COX11/CtaG protein family), which is a transmembrane protein implicated in cytochrome *c* biogenesis in *Bacillus* species.[37,85,86] The presence of multiple *copC* genes in 37.6% of CopC-producing species may also support functional roles of CopC and CopD other than regulation (SI File 4). Copper imported by the CopCD system might also be used to metalate fully folded copper-containing proteins that are exported from the cytoplasm via the twin arginine translocation (TAT) pathway. Disruption of the TAT

secretion pathway affects metal homeostasis in *E. coli*,[87] and numerous copper proteins are exported by the TAT pathway, including multicopper oxidases, nitric oxide reductases, and nitrite reductases.[8] This mechanism of metal loading has been previously proposed for CueO[6] and PcoA.[28] Finally, in the case of the relatively abundant *copCD* mini-operons, it seems equally probable they could import copper for regulation, for loading, or for both functions.

In this study, structural biology and bioinformatics have been used to significantly expand and thoroughly categorize the CopC family of periplasmic copper binding proteins. Contrary to the established model, the CopC family is dominated by proteins that bind only Cu(II), namely, the recently discovered C_{0-2} CopCs and the newly identified C_{0-1} CopCs. The vast majority of these genes are not found in canonical *cop* operons, and the composition of their genomic neighborhoods depends in part on the identity of their copper sites. Several conserved and uncharacterized proteins potentially associated with copper homeostasis can be identified on the basis of their genomic association with *copCD* pairs. Taken together, these data indicate that truly representative cop operons and CopC copper binding properties cannot be defined. Exploring the range of potential CopC functions in the context of this previously unappreciated diversity will form the basis for further studies of the widespread yet poorly understood CopCD copper homeostasis system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Experiments performed at the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Argonne National Laboratory, were supported by the U.S. Department of Energy under Contract DE-AC02-06CH11357. Use of LS-CAT Sector 21 was supported by the Michigan Economic Development Corp. and Michigan Technology Tri-Corridor Grant 085P1000817. GM/CA-CAT has been funded in whole or in part with Federal funds from National Institutes of Health Grant Y1-CO-1020 from the National Cancer Institute and Grant Y1-GM-1104 from the National Institute of General Medical Sciences. ICP-MS experiments were performed at the the Quantitative Bio-element Imaging Center at Northwestern University.

Funding

This work was funded by National Institutes of Health Grant GM58518 (A.C.R.), American Heart Association Predoctoral Fellowship 14PRE20460104 (G.E.K.), and a Northwestern University Undergraduate Research Grant (J.D.H.).

ABBREVIATIONS

COX	cytochrome *c* oxidase
***Esc*-PcoC**	*E. coli* PcoC
HMM	hidden Markov model
***Mst*-CopC**	*M. trichosporium* OB3b CopC
***Psf*-CopC**	*P. fluorescens* SBW25 CopC
***Pss*-CopC**	*P. syringae* pv tomato

TAT twin arginine translocation

REFERENCES

1. Festa RA, Thiele DJ. Copper: An essential metal in biology. Curr. Biol. 2011; 21:R877–R883. [PubMed: 22075424]
2. Rensing, C., McDevitt, SF. The copper metallome in prokaryotic cells. In: Banci, L., editor. Metallomics and the Cell. Dordrecht, The Netherlands: Springer; 2013. p. 417-450.
3. Macomber L, Imlay JA. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:8344–8349. [PubMed: 19416816]
4. Gaetke LM, Chow-Johnson HS, Chow CK. Copper: toxicological relevance and mechanisms. Arch. Toxicol. 2014; 88:1929–1938. [PubMed: 25199685]
5. Andreini C, Bertini I, Rosato A. Metalloproteomes: a bioinformatic approach. Acc. Chem. Res. 2009; 42:1471–1479. [PubMed: 19697929]
6. Rensing C, Grass G. *Escherichia coli* mechanisms of copper homeostasis in a changing environment. FEMS Microbiol. Rev. 2003; 27:197–213. [PubMed: 12829268]
7. Samanovic MI, Ding C, Thiele DJ, Darwin KH. Copper in microbial pathogenesis: meddling with the metal. Cell Host Microbe. 2012; 11:106–115. [PubMed: 22341460]
8. Arguello JM, Raimunda D, Padilla-Benavides T. Mechanisms of copper homeostasis in bacteria. Front. Cell. Infect. Microbiol. 2013; 3:14. [PubMed: 23630667]
9. Ladomersky E, Petris MJ. Copper tolerance and virulence in bacteria. Metallomics. 2015; 7:957–964. [PubMed: 25652326]
10. Outten FW, Huffman DL, Hale JA, O'Halloran TV. The independent cue and cus systems confer copper tolerance during aerobic and anaerobic growth in *Escherichia coli* . J. Biol. Chem. 2001; 276:30670–30677. [PubMed: 11399769]
11. Rensing C, Fan B, Sharma R, Mitra B, Rosen BP. CopA: an *Escherichia coli* Cu(I)-translocating P-type ATPase. Proc. Natl. Acad. Sci. U. S. A. 2000; 97:652–656. [PubMed: 10639134]
12. Outten FW, Outten CE, Hale J, O'Halloran TV. Transcriptional activation of an *Escherichia coli* copper efflux regulon by the chromosomal MerR homologue, cueR. J. Biol. Chem. 2000; 275:31024–31029. [PubMed: 10915804]
13. Kim EH, Nies DH, McEvoy MM, Rensing C. Switch or funnel: how RND-type transport systems control periplasmic metal homeostasis. J. Bacteriol. 2011; 193:2381–2387. [PubMed: 21398536]
14. Chacón KN, Mealman TD, McEvoy MM, Blackburn NJ. Tracking metal ions through a Cu/Ag efflux pump assigns the functional roles of the periplasmic proteins. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:15373–15378. [PubMed: 25313055]
15. Gudipaty SA, McEvoy MM. The histidine kinase CusS senses silver ions through direct binding by its sensor domain. Biochim. Biophys. Acta, Proteins Proteomics. 2014; 1844:1656–1661.
16. Chillappagari S, Miethke M, Trip H, Kuipers OP, Marahiel MA. Copper acquisition is mediated by YcnJ and regulated by YcnK and CsoR in *Bacillus subtilis* . J. Bacteriol. 2009; 191:2362–2370. [PubMed: 19168619]
17. Cooksey DA. Molecular mechanisms of copper resistance and accumulation in bacteria. FEMS Microbiol. Rev. 1994; 14:381–386. [PubMed: 7917425]
18. Zhang XX, Rainey PB. Regulation of copper homeostasis in *Pseudomonas fluorescens* SBW25. Environ. Microbiol. 2008; 10:3284–3294. [PubMed: 18707611]
19. Grass G, Rensing C. CueO is a multicopper oxidase that confers copper tolerance in *Escherichia coli* . Biochem. Biophys. Res. Commun. 2001; 286:902–908. [PubMed: 11527384]
20. Zimmermann M, Udagedara SR, Sze CM, Ryan TM, Howlett GJ, Xiao Z, Wedd AG. PcoE — A metal sponge expressed to the periplasm of copper resistance *Escherichia coli*. Implication of its function role in copper resistance. J. Inorg. Biochem. 2012; 115:186–197. [PubMed: 22658755]
21. Zhang L, Koay M, Maher MJ, Xiao Z, Wedd AG. Intermolecular transfer of copper ions from the CopC protein of *Pseudomonas syringae*. Crystal structures of fully loaded Cu(I)Cu(II) forms. J. Am. Chem. Soc. 2006; 128:5834–5850. [PubMed: 16637653]

22. Cha J-S, Cooksey DA. Copper resistance in *Pseudomonas syringae* mediated by periplasmic and outer membrane proteins. Proc. Natl. Acad. Sci. U. S. A. 1991; 88:8915–8919. [PubMed: 1924351]
23. Brown NL, Barrett SR, Camakaris J, Lee BTO, Rouch DA. Molecular genetics and transport analysis of the copper-resistance determinant (pco) from *Escherichia coli* plasmid pRJ1004. Mol. Microbiol. 1995; 17:1153–1166. [PubMed: 8594334]
24. Mellano MA, Cooksey DA. Nucleotide sequence and organization of copper resistance genes from *Pseudomonas syringae* pv. *tomato* . J. Bacteriol. 1988; 170:2879–2883. [PubMed: 3372485]
25. Cha JS, Cooksey DA. Copper hypersensitivity and uptake in *Pseudomonas syringae* containing cloned components of the copper resistance operon. Appl. Environ. Microbiol. 1993; 59:1671–1674. [PubMed: 16348944]
26. Wernimont AK, Huffman DL, Finney LA, Demeler B, O'Halloran TV, Rosenzweig AC. Crystal structure and dimerization equilibria of PcoC, a methionine-rich copper resistance protein from *Escherichia coli* . JBIC, J. Biol. Inorg. Chem. 2003; 8:185–194. [PubMed: 12459914]
27. Huffman DL, Pearlso KL, Penner-Hahn JE, O'Halloran TV. The PcoC copper resistance protein coordinates Cu(I) via novel S-methionine interactions. J. Am. Chem. Soc. 2003; 125:342–343. [PubMed: 12517140]
28. Djoko KY, Xiao Z, Wedd AG. Copper resistance in *E. coli*: the multicopper oxidase PcoA catalyzes oxidation of copper(I) in Cu^{I}-Cu^{II}-PcoC. ChemBioChem. 2008; 9:1579–1582. [PubMed: 18536063]
29. Arnesano F, Banci L, Bertini I, Thompsett AR. Solution structure of CopC: a cupredoxin-like protein involved in copper homeostasis. Structure. 2002; 10:1337–1347. [PubMed: 12377120]
30. Arnesano F, Banci L, Bertini I, Mangani S, Thompsett AR. A redox switch in CopC: an intriguing copper trafficking protein that binds copper(I) and copper(II) at different sites. Proc. Natl. Acad. Sci. U. S. A. 2003; 100:3814–3819. [PubMed: 12651950]
31. Cortes L, Wedd AG, Xiao Z. The functional roles of the three copper sites associated with the methionine-rich insert in the multicopper oxidase CueO from *E. coli* . Metallomics. 2015; 7:776–785. [PubMed: 25679350]
32. Wijekoon CJK, Young TR, Wedd AG, Xiao Z. CopC protein from *Pseudomonas fluorescens* SBW25 features a conserved novel high-affinity Cu(II) binding site. Inorg. Chem. 2015; 54:2950–2959. [PubMed: 25710712]
33. Hirooka K, Edahiro T, Kimura K, Fujita Y. Direct and indirect regulation of the *ycnKJI* operon involved in copper uptake through two transcriptional repressors, YcnK and CsoR. Bacillus subtilis. J. Bacteriol. 2012; 194:5675–5687. [PubMed: 22904286]
34. Serventi F, Youard ZA, Murset V, Huwiler S, Buhler D, Richter M, Luchsinger R, Fischer HM, Brogioli R, Niederer M, Hennecke H. Copper starvation-inducible protein for cytochrome oxidase biogenesis in *Bradyrhizobium japonicum* . J. Biol. Chem. 2012; 287:38812–38823. [PubMed: 23012364]
35. Blundell KLIM, Hough MA, Vijgenboom E, Worrall JAR. Structural and mechanistic insights into an extracytoplasmic copper trafficking pathway in *Streptomyces lividans* . Biochem. J. 2014; 459:525–538. [PubMed: 24548299]
36. Sirajuddin S, Rosenzweig AC. Enzymatic oxidation of methane. Biochemistry. 2015; 54:2283–2294. [PubMed: 25806595]
37. Kenney GE, Sadek M, Rosenzweig AC. Copper-responsive gene expression in the methanotroph *Methylosinus trichosporium* OB3b. Metallomics. 2016 in press.
38. Cabantous S, Waldo GS. *In vivo* and *in vitro* protein solubility assays using split GFP. Nat. Methods. 2006; 3:845–854. [PubMed: 16990817]
39. Lima, CD., Mossessova, E. and Cornell Research Foundation. Rapidly cleavable sumo fusion protein expression system for difficult to express proteins. U.S. Patent. 20050069988 A1. 2005.
40. Studier FW. Stable expression clones and auto-induction for protein production in *E. coli* . Methods Mol. Biol. 2014; 1091:17–32. [PubMed: 24203322]
41. Tropea, JE., Cherry, S., Waugh, DS. Expression and purification of soluble His_6-tagged TEV protease. In: Doyle, SA., editor. High throughput protein expression and purification. Totowa, NJ: Humana Press; 2009.

42. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. Methods Mol. Biol. 1998; 112:531–552.
43. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.
44. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J. Appl. Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
45. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2011; 67:355–367. [PubMed: 21460454]
46. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2011; 67:235–242. [PubMed: 21460441]
47. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010; 26:2460–2461. [PubMed: 20709691]
48. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39:W29–W37. [PubMed: 21593126]
49. Clamp M, Cuff J, Searle SM, Barton GJ. The Jalview Java alignment editor. Bioinformatics. 2004; 20:426–427. [PubMed: 14960472]
50. Smith AT, Smith KP, Rosenzweig AC. Diversity of the metal-transporting P_{1B}-type ATPases. JBIC, J. Biol. Inorg. Chem. 2014; 19:947–960. [PubMed: 24729073]
51. Wittkop T, Emig D, Truss A, Albrecht M, Bocker S, Baumbach J. Comprehensive cluster analysis with Transitivity Clustering. Nat. Protoc. 2011; 6:285–295. [PubMed: 21372810]
52. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. Integration of biological networks and gene expression data using Cytoscape. Nat. Protoc. 2007; 2:2366–2382. [PubMed: 17947979]
53. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A. The Pfam protein families database. Nucleic Acids Res. 2010; 38:D211–D222. [PubMed: 19920124]
54. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and Genome Properties in 2013. Nucleic Acids Res. 2013; 41:D387–D395. [PubMed: 23197656]
55. Ihaka R, Gentleman R. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics. 1996; 5:299–314.
56. Ward JH. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 1963; 58:236–244.
57. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Martin Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B. gplots: Various R programming tools for plotting data. 2015 https://cran.r-project.org/web/packages/gplots/index.html.
58. El Sheikh AF, Poret-Peterson AT, Klotz MG. Characterization of two new genes, amoR and amoD, in the amo operon of the marine ammonia oxidizer *Nitrosococcus oceani* ATCC 19707. Appl. Environ. Microbiol. 2008; 74:312–318. [PubMed: 17993553]
59. Dash BP, Alles M, Bundschuh FA, Richter OM, Ludwig B. Protein chaperones mediating copper insertion into the Cu_A site of the *aa* $_3$-type cytochrome *c* oxidase of *Paracoccus denitrificans* . Biochim. Biophys. Acta, Bioenerg. 2015; 1847:202–211.
60. Banci L, Bertini I, Ciofi-Baffoni S, Katsari E, Katsaros N, Kubicek K, Mangani S. A copper(I) protein possibly involved in the assembly of Cu_A center of bacterial cytochrome c oxidase. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:3994–3999. [PubMed: 15753304]
61. Abriata LA, Banci L, Bertini I, Ciofi-Baffoni S, Gkazonis P, Spyroulias GA, Vila AJ, Wang S. Mechanism of Cu_A assembly. Nat. Chem. Biol. 2008; 4:599–601. [PubMed: 18758441]

62. Sequence numbering in the crystal structure is based on sequential order in the structure, not the actual physiological sequence. The physiological sequence of this PcoC has not been deposited in the JGI-IMG database or the NCBI nonredundant database. See ref 23 for the original sequence. Alternatively, the PcoC sequence from *E. coli* CATCC 8739 (IMG entry 641602826) is identical to the sequence encoded on plasmid pRJ1004 isolated from *E. coli* strain RJ92.

63. Tetaz TJ, Luke RKJ. Plasmid controlled resistance to copper in *Escherichia coli* . J. Bacteriol. 1983; 154:1263–1268. [PubMed: 6343346]

64. Sequence numbering in the crystal structure is based on sequential order in the structure, not the actual physiological sequence. The physiological sequence of this CopC has not been deposited in the JGI-IMG database or the NCBI nonredundant database. See ref 24 for the original sequence. Alternatively, the CopC sequence from *P. syringae* pv maculicola M4a (IMG entry 2648015009) is identical to the sequence from *P. syringae* pv tomato.

65. Murphy ME, Lindley PF, Adman ET. Structural comparison of cupredoxin domains: domain recycling to construct proteins with novel functions. Protein Sci. 1997; 6:761–770. [PubMed: 9098885]

66. Karpinets TV, Obraztsova AY, Wang Y, Schmoyer DD, Kora GH, Park BH, Serres MH, Romine MF, Land ML, Kothe TB, Fredrickson JK, Nealson KH, Uberbacher EC. Conserved synteny at the protein family level reveals genes underlying *Shewanella* species' cold tolerance and predicts their novel phenotypes. Funct. Integr. Genomics. 2010; 10:97–110. [PubMed: 19802638]

67. Costa M, Sola M, del Solar G, Eritja R, Hernandez-Arriaga AM, Espinosa M, Gomis-Rüth FX, Coll M. Plasmid transcriptional repressor CopG oligomerises to render helical superstructures unbound and in complexes with oligonucleotides. J. Mol. Biol. 2001; 310:403–417. [PubMed: 11428897]

68. del Solar G, Hernandez-Arriaga AM, Gomis-Ruth FX, Coll M, Espinosa M. A genetically economical family of plasmid-encoded transcriptional repressors involved in control of plasmid copy number. J. Bacteriol. 2002; 184:4943–4951. [PubMed: 12193609]

69. Behlau F, Canteros BI, Minsavage GV, Jones JB, Graham JH. Molecular characterization of copper resistance genes from *Xanthomonas citri* subsp. *citri* and *Xanthomonas alfalfae* subsp. *citrumelonis* . Appl. Environ. Microbiol. 2011; 77:4089–4096. [PubMed: 21515725]

70. Marrero K, Sánchez A, González LJ, Ledón T, Rodriguez-Ulloa A, Castellanos-Serra L, Pérez C, Fando R. Periplasmic proteins encoded by VCA0261-0260 and VC2216 genes together with *copA* and *cueR* products are required for copper tolerance but not for virulence in *Vibrio cholerae* . Microbiology. 2012; 158:2005–2016. [PubMed: 22653946]

71. Monchy S, Benotmane MA, Wattiez R, van Aelst S, Auquier V, Borremans B, Mergeay M, Taghavi S, van der Lelie D, Vallaeys T. Transcriptomic and proteomic analyses of the pMOL30-encoded copper resistance in *Cupriavidus metallidurans* strain CH34. Microbiology. 2006; 152:1765–1776. [PubMed: 16735739]

72. Rademacher C, Moser R, Lackmann JW, Klinkert B, Narberhaus F, Masepohl B. Transcriptional and posttranscriptional events control copper-responsive expression of a *Rhodobacter capsulatus* multicopper oxidase. J. Bacteriol. 2012; 194:1849–1859. [PubMed: 22287514]

73. Karlsen OA, Larsen O, Jensen HB. The copper responding surfaceome of *Methylococccus capsulatus* Bath. FEMS Microbiol. Lett. 2011; 323:97–104. [PubMed: 22092708]

74. Fujimoto M, Yamada A, Kurosawa J, Kawata A, Beppu T, Takano H, Ueda K. Pleiotropic role of the Sco1/SenC family copper chaperone in the physiology of *Streptomyces* . Microb. Biotechnol. 2012; 5:477–488. [PubMed: 22117562]

75. Huffman DL, Huyett J, Outten FW, Doan PE, Finney LA, Hoffman BM, O'Halloran TV. Spectroscopy of Cu(II)-PcoC and the multicopper oxidase function of PcoA, two essential components of *Escherichia coli* pco copper resistance operon. Biochemistry. 2002; 41:10046–10055. [PubMed: 12146969]

76. Wendt, MD. Protein-protein interactions as drug targets. In: Wendt, MD., editor. Topics in Medicinal Chemistry: Protein-protein interactions. Heidelberg, Germany: Springer; 2012.

77. Yuan T, Weljie AM, Vogel HJ. Tryptophan fluorescence quenching by methionine and selenomethionine residues of calmodulin: orientation of peptide and protein binding. Biochemistry. 1998; 37:3187–3195. [PubMed: 9485473]

78. Changela A, Chen K, Xue Y, Holschen J, Outten CE, O'Halloran TV, Mondragon A. Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR. Science. 2003; 301:1383–1387. [PubMed: 12958362]

79. Mermod M, Magnani D, Solioz M, Stoyanov JV. The copper-inducible ComR (YcfQ) repressor regulates expression of ComC (YcfR), which affects copper permeability of the outer membrane of *Escherichia coli* . BioMetals. 2012; 25:33–43. [PubMed: 22089859]

80. Ekici S, Turkarslan S, Pawlik G, Dancis A, Baliga NS, Koch HG, Daldal F. Intracytoplasmic copper homeostasis controls cytochrome *c* oxidase production. mBio. 2014; 5:e01055–e01113. [PubMed: 24425735]

81. Ekici S, Yang H, Koch HG, Daldal F. Novel transporter required for biogenesis of cbb_3-type cytochrome *c* oxidase in *Rhodobacter capsulatus* . mBio. 2012; 3:e00293–e00311. [PubMed: 22294680]

82. Trasnea PI, Utz M, Khalfaoui-Hassani B, Lagies S, Daldal F, Koch HG. Cooperation between two periplasmic copper chaperones is required for full activity of the cbb_3 -type cytochrome *c* oxidase and copper homeostasis in *Rhodobacter capsulatus* . Mol. Microbiol. 2015 n/a.

83. Abajian C, Rosenzweig AC. Crystal structure of yeast Sco1. JBIC, J. Biol. Inorg. Chem. 2006; 11:459–466. [PubMed: 16570183]

84. Banci L, Bertini I, Cavallaro G, Rosato A. The functions of Sco proteins from genome-based analysis. J. Proteome Res. 2007; 6:1568–1579. [PubMed: 17300187]

85. Bengtsson J, von Wachenfeldt C, Winstedt L, Nygaard P, Hederstedt L. CtaG is required for formation of active cytochrome *c* oxidase in *Bacillus subtilis* . Microbiology. 2004; 150:415–425. [PubMed: 14766920]

86. Greiner P, Hannappel A, Werner C, Ludwig B. Biogenesis of cytochrome *c* oxidase-*in vitro* approaches to study cofactor insertion into a bacterial subunit I. Biochim. Biophys. Acta, Bioenerg. 2008; 1777:904–911.

87. Ize B, Porcelli I, Lucchini S, Hinton JC, Berks BC, Palmer T. Novel phenotypes of *Escherichia coli* tat mutants revealed by global gene expression and phenotypic analysis. J. Biol. Chem. 2004; 279:47543–47554. [PubMed: 15347649]

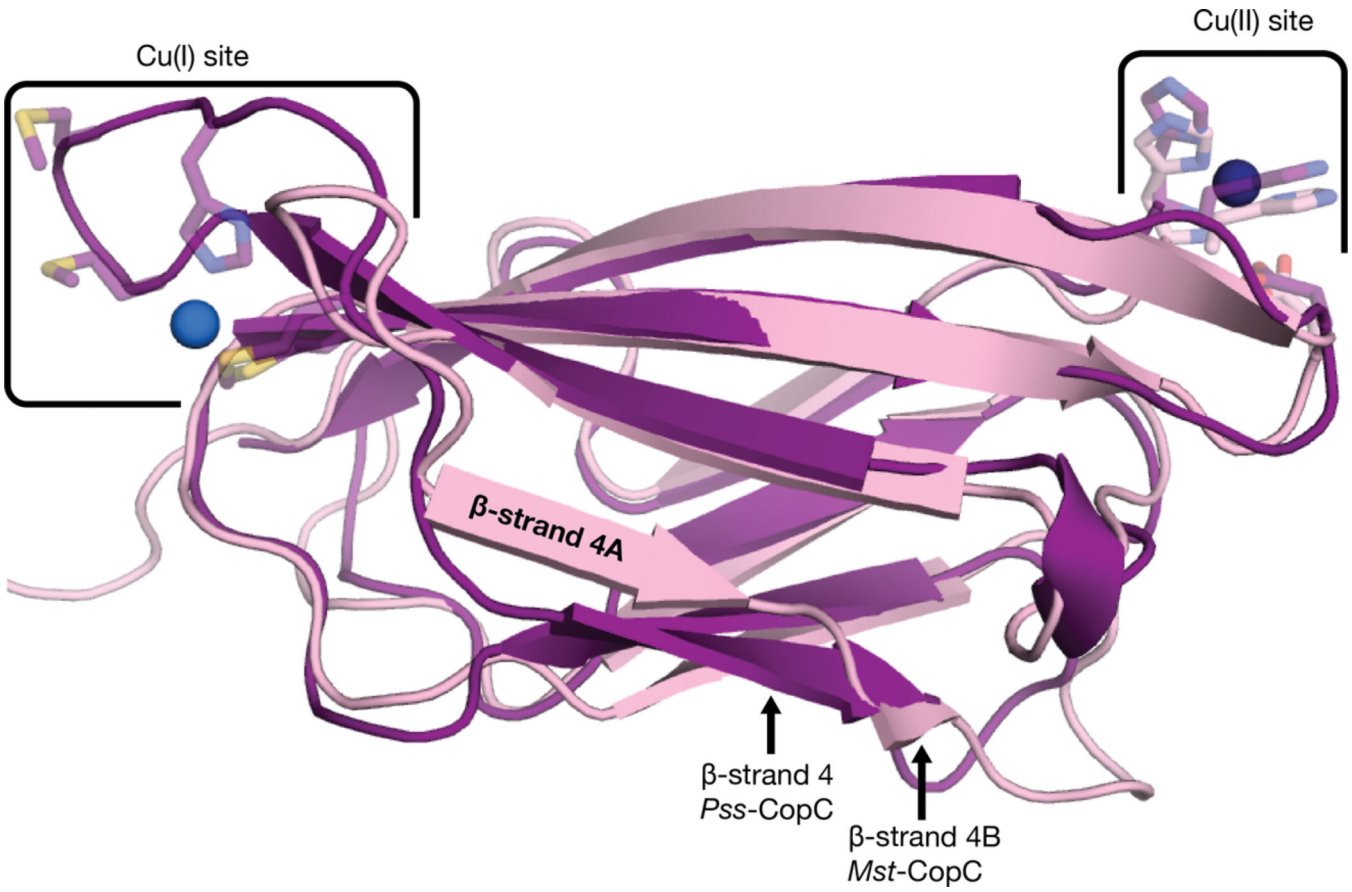


Figure 1.

Structural alignment of *Mst*-CopC (light pink) and *Pss*-CopC (purple, PDB entry 2C9Q). Copper ions are shown as spheres.

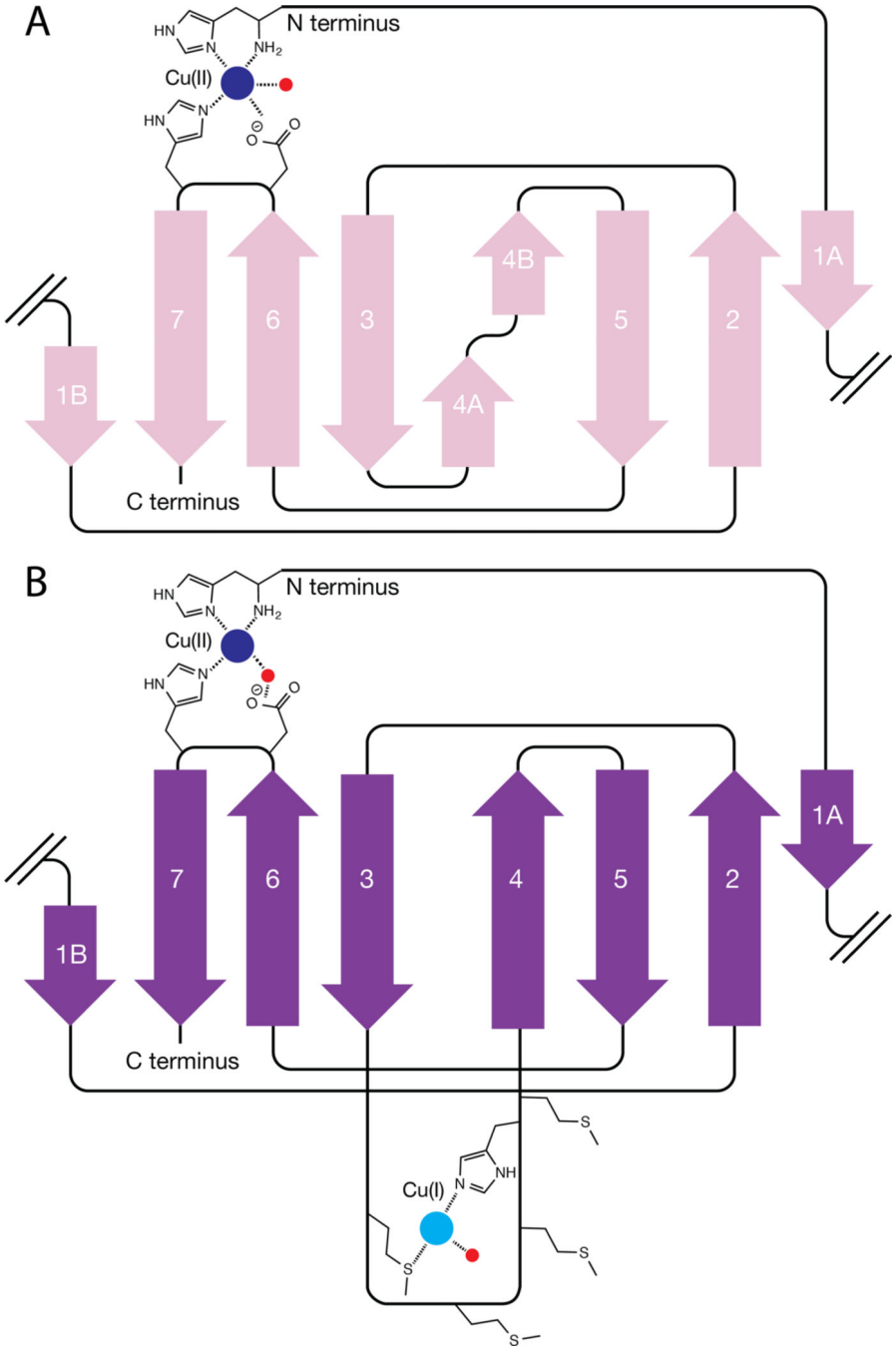


Figure 2.

Topology diagrams of (A) *Mst*-CopC and (B) *Pss*-CopC and *Esc*-PcoC. Cu(II) and Cu(I) ions are shown as dark and light blue spheres, respectively. Water is shown as small red spheres.

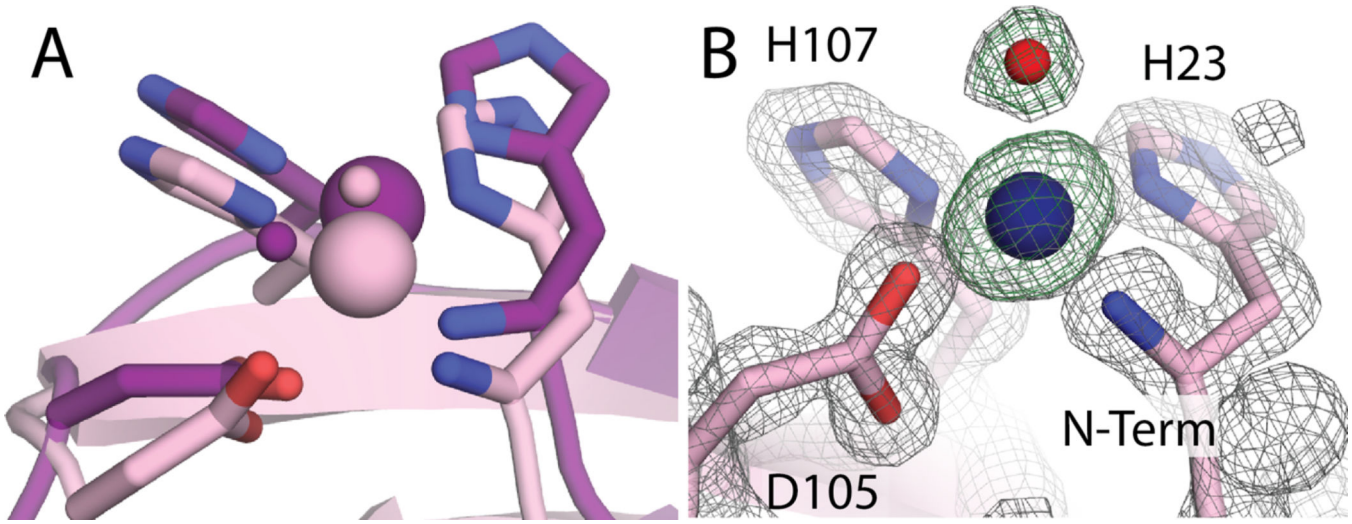


Figure 3.

Coordination environment of the Cu(II) site. (A) Structural alignment of the Cu(II) site in *Mst*-CopC (light pink) and *Pss*-CopC (purple, PDB entry 2C9Q). Copper ions are shown as large spheres, and water molecules are shown as small spheres. (B) Electron density map ($2F_o - F_c$ map contoured at 1.6σ, colored gray) and difference density map ($F_o - F_c$ map contoured at 4.5σ with positive density colored green, negative density not observed at this σ, and copper and water omitted) of the *Mst*-CopC Cu(II) site.

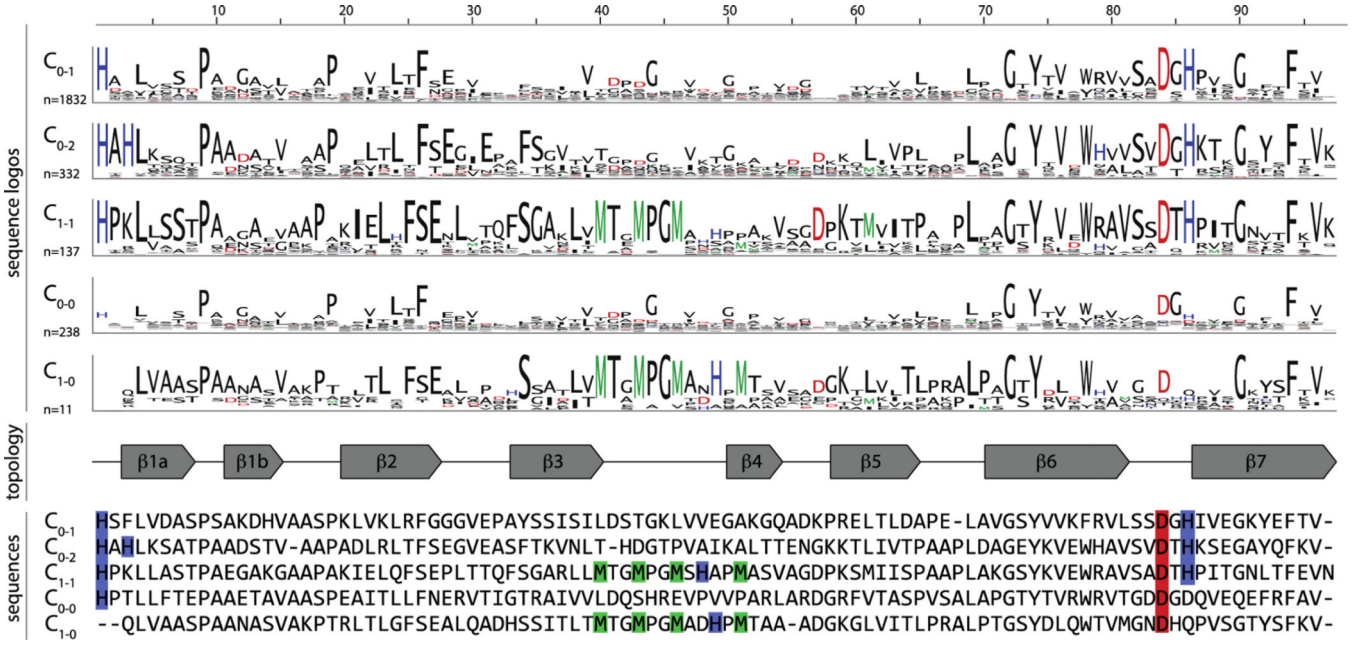


Figure 4.

CopC sequence logos and alignment. Residues that do not match a position in the profile HMM have been removed for the sake of clarity except for M72, which is a key Cu(I) binding residue; it should be noted that M72 and H75 are located in a somewhat variable loop region. Copper binding residues are highlighted in blue (histidine), red (aspartate), or green (methionine). Representative sequences are from the following organisms: C_{1-1}, *Pseudomonas* sp. GM17 (IMG ID 2511266057); C_{0-1}, *M. trichosporium* OB3b (IMG ID 2507406428); C_{0-2}, *P. fluorescens* R124 (IMG ID 2503652063); C_{0-0}, *Mycobacterium tusciae* JS617 (IMG ID 2508745664); and C_{1-0}, *Sphingobium yanoikuyae* XLDN2-5 (IMG ID 2549030586).

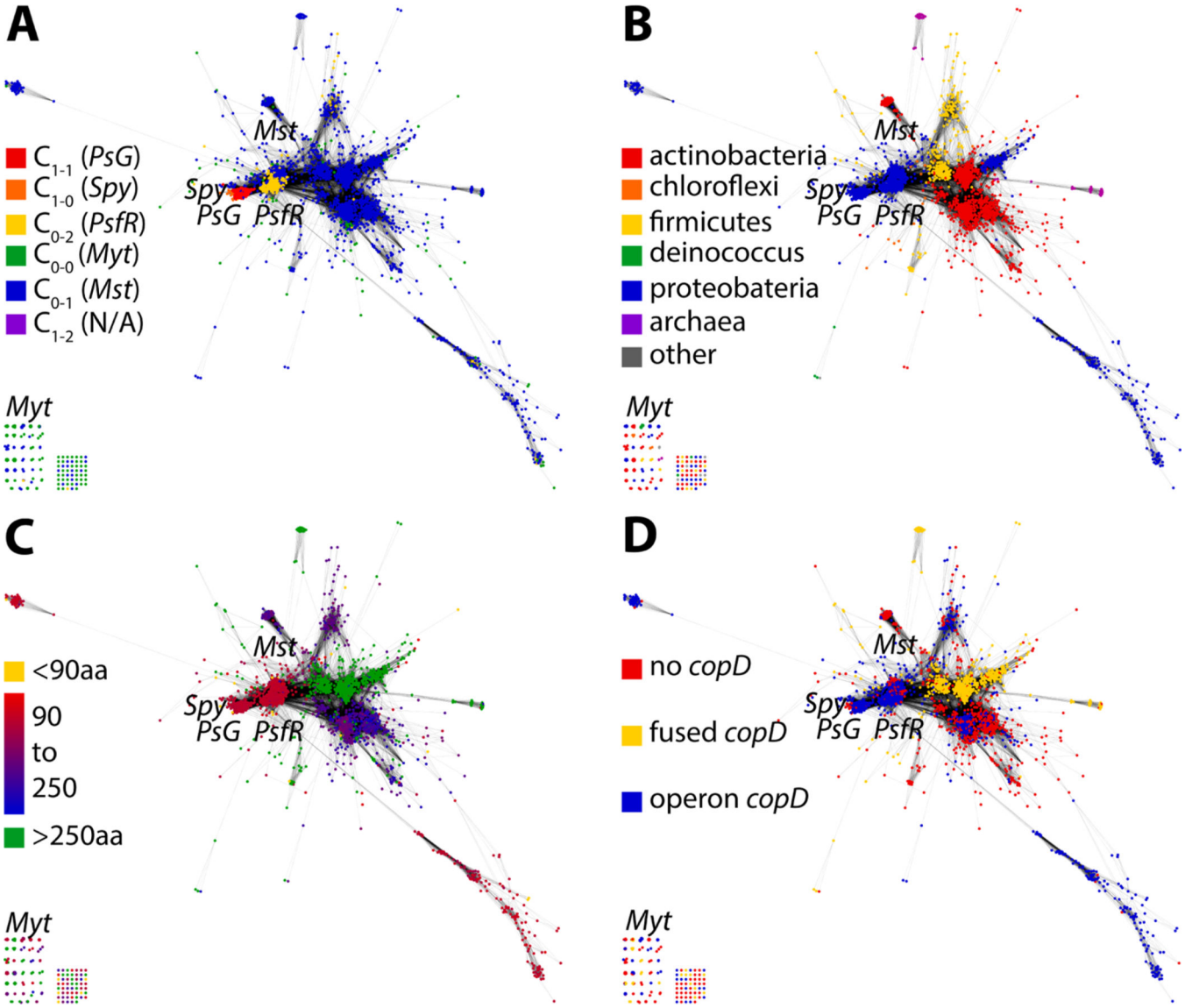


Figure 5.

CopC sequence similarity networks colored by (A) copper site, (B) phylum, (C) full sequence length in amino acids, and (D) CopD presence, absence, or fusion. Labeled variants correspond to CopC sequences highlighted in Figure 4 [*PsG, Ps.* sp. GM17; *Mst, M. trichosporium* OB3b (IMG ID 2507406428); *Psf R, P. fluorescens* R124 (IMG ID 649637371); *Myt, My. tusciae* JS617; and C_{1-0} *Spy., Sp. yanoikuyae* XLDN2-5].

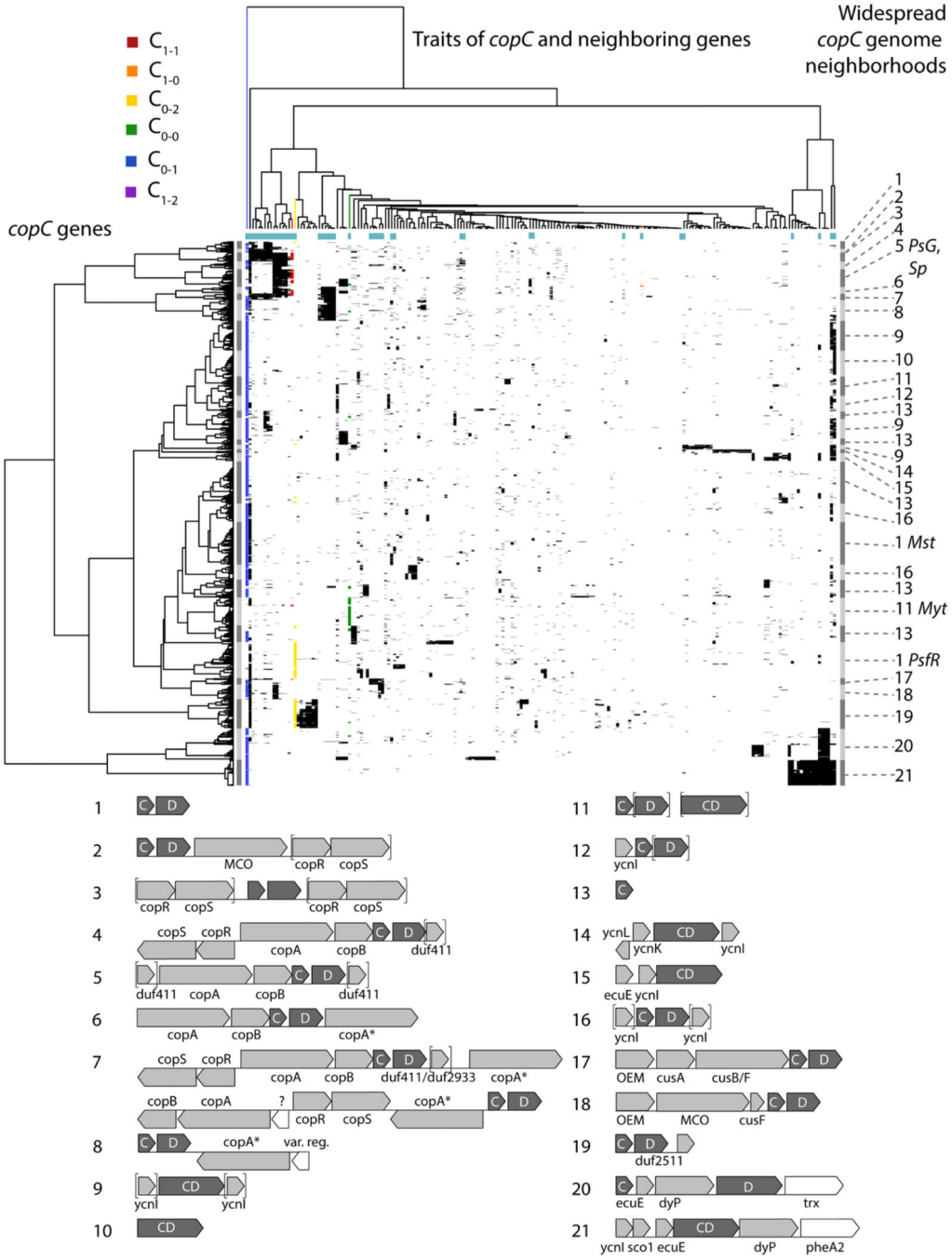


Figure 6.

Hierarchical clustering of *copC* genes and traits [including genome neighborhood, copper site identity, and the presence of additional domains within the *copC* gene (e.g., *copD, ytkA*)]. *copC* genes are clustered by which traits are observed in and around them (left), while traits are clustered by the *copC* genes to which they are connected (top); correlation between a given *copC* gene and trait is indicated by black (central heatmap). The 21 most widespread sets of neighboring genes are depicted (bottom). In some cases, multiple potential operon architectures are seen; the most common one is depicted. Neighboring

genes to either side with no predicted relationship to metal homeostasis have been omitted. The *copC* genes in operons 2–8 are often adjacent to a range of other copper-related genes and operons beyond the five-gene neighborhood monitored in this data set; these genes include *cusABCF*, other *cop* genes (*copE, copK*, and *copZ*), regulatory proteins (*cueR*, additional *cusRS*/*copRS*), and disulfide isomerases.

Table 1

Data Collection and Refinement Statistics of Cu- Loaded *Mst*-CopC

space group	$P2_12_12_1$
resolution (Å)	1.46 (1.50–1.46)
R_{sym}	0.048 (0.067)
$I/\sigma I$	41.9 (22.1)
completeness (%)	99.9 (99.9)
redundancy	8.6 (8.2)
no. of reflections	17434
data cutoff	$F > 0$
no. of atoms	876
R_{work}/R_{free} (%)	16.6/19.1
B factor (Å^2)	
average	8.5
protein	7.7
water	20.3
ligands/ions	18.4
root-mean-square deviation	
bond lengths (Å)	0.0141
bond angles (deg)	1.661

Table 2

Prevalence of CopC Types and Selected Supersets[a]

	Type	Cartoon of copper ions	Cu(I) ligand set (M40, M43, M46, H48, M51)	Cu(II) ligand set (H1, H3, D79, H81)	Number of Sequences (Unpruned)	Number of Sequences (Pruned)	Typical protein name	Classification in Wijekoon *et al.*[32]
CopC types	C_{0-1}		<4	3(H3 excluded)	3,807 (44.1 %)	1,832 (71.8%)	CopC, YcnJ	-
	C_{0-2}		<4	4(H3 included)	3,471 (40.2 %)	332 (13.0%)	YobA,CopC	Type B
	C_{1-1}		≥4	3(113 excluded)	930 (10.8%)	137 (5.4 %)	PcoC, CopC	Type A
	C_{0-0}		<4	<3(H3 excluded)	411 (4.8 %)	238 (9.3 %)	-	-
	C_{1-0}		≥4	<3(H3 excluded)	17 (0.2 %)	11 (0.4 %)	-	-
CopC supersets	No Cu(I)		<4	-	7,689 (89.0 %)	2,402 (94.2 %)	-	-
	Cu(I)		≥4	-	947 (11.0%)	148 (5.8 %)	-	-
	No Cu(II)		-	<3(H3 excluded)	428 (5.0 %)	249 (9.8 %)	-	-
	Cu(II) only		<4	3(H3 excluded)	7,278 (84.3 %)	2,164 (84.9 %)	-	-

Cu(I)

Cu(II)2 His, 1Asp

Cu(II) 3 His

[a]The “Type” column shows the CopC types defined herein with pictorial representations of each type shown in the adjacent column. The shaded columns indicate the ligand sets used to define each CopC type. The “Number of Sequences (Unpruned)” column shows the number of all known CopC sequences (8636 total) with the specified ligand set. The “Number of Sequences (Pruned)” column shows the

number of CopC sequences with the specified ligand set after highly similar sequences (>95%) have been removed (2550 total). Supersets include all CopCs with the listed sequence feature; e.g., "Cu(I)" includes any sequence containing a Cu(I) site, regardless of Cu(II) site residues.