

# Protein Similarity Networks Reveal Relationships among Sequence, Structure, and Function within the Cupin Superfamily

Richard Uberto, Ellen W. Moomaw\*

Department of Chemistry and Biochemistry, Kennesaw State University, Kennesaw, Georgia, United States of America

## Abstract

The cupin superfamily is extremely diverse and includes catalytically inactive seed storage proteins, sugar-binding metal-independent epimerases, and metal-dependent enzymes possessing dioxygenase, decarboxylase, and other activities. Although numerous proteins of this superfamily have been structurally characterized, the functions of many of them have not been experimentally determined. We report the first use of protein similarity networks (PSNs) to visualize trends of sequence and structure in order to make functional inferences in this remarkably diverse superfamily. PSNs provide a way to visualize relatedness of structure and sequence among a given set of proteins. Structure- and sequence-based clustering of cupin members reflects functional clustering. Networks based only on cupin domains and networks based on the whole proteins provide complementary information. Domain-clustering supports phylogenetic conclusions that the N- and C-terminal domains of bicupin proteins evolved independently. Interestingly, although many functionally similar enzymatic cupin members bind the same active site metal ion, the structure and sequence clustering does not correlate with the identity of the bound metal. It is anticipated that the application of PSNs to this superfamily will inform experimental work and influence the functional annotation of databases.

**Citation:** Uberto R, Moomaw EW (2013) Protein Similarity Networks Reveal Relationships among Sequence, Structure, and Function within the Cupin Superfamily. PLoS ONE 8(9): e74477. doi:10.1371/journal.pone.0074477

**Editor:** Pratul K. Agarwal, Oak Ridge National Laboratory, United States of America

**Received:** June 27, 2013; **Accepted:** August 1, 2013; **Published:** September 6, 2013

**Copyright:** © 2013 Uberto, Moomaw. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the National Science Foundation (MCB-1041912) to EWM (<http://www.nsf.gov>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

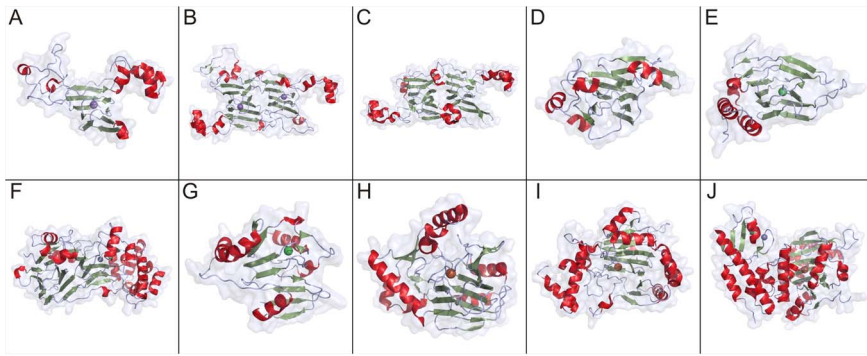
\* E-mail: [emoomaw@kennesaw.edu](mailto:emoomaw@kennesaw.edu)

## Introduction

The cupin superfamily of proteins possesses remarkable functional diversity with representatives found in Archaea, Eubacteria, and Eukaryota [1,2,3,4,5]. The identification of the cupin superfamily was originally based on the recognition that the wheat protein germin shared a nine amino acid sequence with another protein, spherulin, produced by the slime mold *Physarum polycephalum* during starvation [3]. This sequence similarity was also observed in a number of seed storage proteins called germin-like proteins (GLPs). Knowledge of the three dimensional structures of these proteins led to the collective name “cupin” on the basis of their  $\beta$ -barrel shape (“cupa” means “small barrel” in Latin) [4]. Characteristic features of proteins with this fold include high thermal stability and resistance to proteases. These features are consistent with their high degree of subunit contacts, hydrophobic interactions, and short loops. The cupin domain was originally described as two conserved motifs, each composed of two  $\beta$ -strands [2,5]. Motif 1 was designated as G(X)<sub>5</sub>HXXH(X)<sub>3,4</sub>E(X)<sub>6</sub>G and Motif 2 as G(X)<sub>5</sub>PXG(X)<sub>2</sub>H(X)<sub>3</sub>N. The two motifs are separated by an intermotif region which ranges from 15 to 50 amino acids long. With more sequences analyzed, it became clear that the primary sequence of the two motifs is not as highly conserved as previously thought [1,6].

Genomic enzymology; large-scale sequence, functional, and structural databases; and inferences from molecular evolution have informed family, superfamily, and suprafamily designations

[7,8,9,10]. Enzyme families are relatively recently-diverged groups of enzymes that share similar three-dimensional structures and functions. Enzyme superfamilies consist of enzymes that diverged earlier and possess fewer common elements but share a common mechanistic attribute, while suprafamilies share conserved residues but do not share a common mechanistic attribute [8,11]. Inconsistencies exist in the usage of the term ‘cupin’. According to the Structural Classification of Proteins (SCOP) database [12,13,14], cupin proteins are members of the ‘RmlC-like Cupins’ superfamily within the double-stranded  $\beta$ -helix (DSBH) multi-catalytic fold [15]. The term ‘cupin superfamily’ has often been used to refer to those proteins defined by the SCOP database as well as the 2-oxoglutarate-Fe<sup>2+</sup>-dependent dioxygenase superfamily that also possesses the DSBH fold [1,5]. However defined, the cupin superfamily is extremely diverse and includes catalytically inactive seed storage and sugar-binding metal-independent proteins as well as metal-dependent enzymes possessing dioxygenase, decarboxylase, and other activities [10]. Figure 1 (A–J) shows the structures of 10 representative proteins designated as cupins in the Pfam database. Although the majority of enzymatic cupins contain iron as an active site metal, other members may contain nickel, zinc, manganese, cobalt, or copper. Each cofactor allows a different type of chemistry to occur within the conserved tertiary structure. Proposed reaction mechanisms of the metal-dependent cupins generally involve sequential binding of the substrate and dioxygen to the catalytic metal cation [16]. Figure 2 (A–J) shows



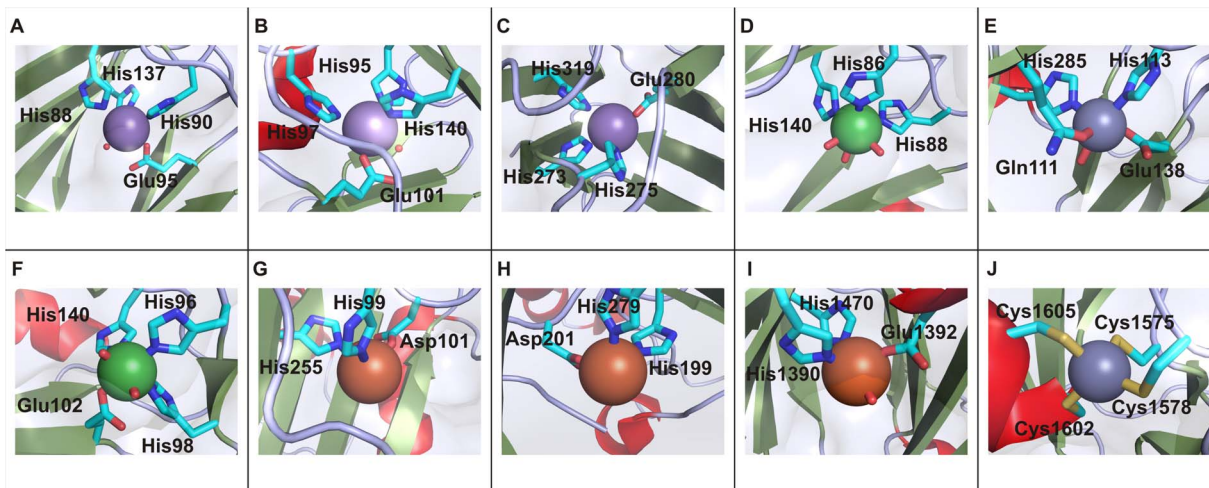
**Figure 1. Structures of representative members of the cupin superfamily.** A. oxalate oxidase (PDB code: 2et1) [72], B. oxalate decarboxylase (PDB code: 1uw8) [43], C. seed storage protein Ara h (PDB code: 3s7i) [39], D. NovW, a 4-keto-6-deoxy sugar epimerase (PDB code: 2c0z) [71], E. cysteine dioxygenase (PDB code: 2q4s) [48], F. phosphomannose isomerase (PDB code: 1pmi) [53] G. acireductone dioxygenase (PDB code: 1zrr) [62] H. taurine/alpha-ketoglutarate dioxygenase (PDB code: 1os7) [73], I. hypoxia-inducible factor 1-alpha inhibitor (PDB code: 2y0i) [74], J. lysine-specific demethylase 6B (PDB code: 2xue) [75].  $\beta$ -sheets are shown in green,  $\alpha$ -helices are shown in red, and random coils are shown in grey. Spheres represent bound metal ions. Figures were generated using Pymol (The PyMOL Molecular Graphics System, Schrödinger, LLC). doi:10.1371/journal.pone.0074477.g001

the metal-binding sites of 8 representative members of the enzymatic cupins.

The DSBH (also referred to as jelly roll) topology is most often composed of eight  $\beta$ -strands that form a  $\beta$ -sandwich structure comprised of two four-stranded antiparallel  $\beta$ -sheets. Although four forms are possible (two hands of the helix and two directions to trace the structure), only one form (right-handed class I) is prevalent in nature [17]. Ancestral cupins can be evolutionarily reconstructed as simple, small molecule-binding domains that likely bound sugars and cyclic nucleotides [5,7,18]. These sugar-binding domains later gave rise to sugar-modifying domains such as isomerases and epimerases [19]. Analyses of the evolution of the

fold suggest that a set of conserved histidine residues employed in sugar-binding in the ancestral non-enzymatic domain evolved into the metal-coordinating histidine residues observed in oxalate oxidase (Figures 1A and 2A) [19] and oxalate decarboxylase (Figures 1B, 2B, and 2C) [20]. Another lineage of DSBH domains acquired a new set of conserved residues with the ability to bind 2-oxoglutarate which gave rise to the 2-oxoglutarate-Fe<sup>2+</sup>-dependent dioxygenases [7,21].

The exponential growth of structural information for proteins provides abundant material for the analysis of how protein structure informs biological function and chemical reactivity. Babbitt *et al.* recognized the need to associate structure and



**Figure 2. Metal-binding sites of representative members of the cupin superfamily.** A. Mn ion of oxalate oxidase coordinated by His88, His90, Glu95, and His137 (PDB code: 2et1) [72], B. N-terminal Mn ion of oxalate decarboxylase coordinated by His95, His97, Glu101, and His140 (PDB code: 1uw8) [43], C. C-terminal Mn ion of oxalate decarboxylase coordinated by His273, His275, Glu280, and His319 (PDB code: 1uw8) [43], D. Ni ion of cysteine dioxygenase coordinated by His86, His88, and His140 (PDB code: 2q4s) [48], E. Ni ion of phosphomannose isomerase coordinated by Gln111, His113, Glu138, and His285 (PDB code: 1pmi) [53] F. Ni ion of acireductone dioxygenase coordinated by His96, His98, Glu102, and His140 (PDB code: 1zrr) [62], G. Fe ion of taurine/alpha-ketoglutarate dioxygenase coordinated by His99, Asp101, and His255 (PDB code: 1os7) [73], H. Fe ion of hypoxia-inducible factor 1-alpha inhibitor coordinated by His199, Asp201, and His279 (PDB code: 2y0i) [74], I. Fe ion of lysine-specific demethylase 6B coordinated by His1390, Glu1392, and His1470 (PDB code: 2xue) [75], J. Zn ion of lysine-specific demethylase 6B coordinated by Cys1575, Cys1578, Cys1602, and Cys1605 (not part of the cupin domain) (PDB code: 2xue) [75].  $\beta$ -sheets are shown in green,  $\alpha$ -helices are shown in red, and random coils are shown in grey. Spheres represent bound metal ions. Figures were generated using Pymol (The PyMOL Molecular Graphics System, Schrödinger, LLC). doi:10.1371/journal.pone.0074477.g002

sequence information with biological function in ways that are accessible to both experimental and computational biologists. They presented protein similarity networks (PSNs) to fulfill this need [22]. PSNs have contributed to our understanding of a number of large groups of proteins including the enolase superfamily [23], the ePK-like superfamily [24], glutathione transferases [25,26], strictosidine synthase-like proteins [27], cysteine peptidases [28], and proteins used in algal metal transport [29]. These studies have yielded meaningful insights, validated PSN methodology, and provided an understanding of the caveats and limitations of PSNs. PSNs are complementary to phylogenetic studies and provide different and new information compared to other methods relating structures and sequences. It has been noted that protein similarity networks are most compelling when painted with functional or structural information that is orthogonal to the data used to generate the networks [22].

The Pfam database [30] lists 112,082 cupin sequences represented in 6529 species and 945 associated protein structures. This represents a greater than 10-fold increase in the number of sequences in only four years [15]. Protein similarity networks are a way to visualize large-scale computational analyses of sequence and structure among a given set of proteins [22,24] and have been used to guide experimental design and data interpretation [27,31]. In this work we describe the first visualization of sequence and structure data of the cupin superfamily using PSNs. The PSNs were built using open source software programs Pythoscape [32] and Cytoscape [33]. In these networks the protein structures (or sequences) are represented as nodes, and their similarities to each other are represented as edges. A value of this work lies in displaying nodes by particular attributes. Node attributes that can be gathered automatically through Pythoscape include sequence, organism, description, and identification codes. Other manually-retrieved attributes may be inputted, such as the presence of a catalytic motif or reaction intermediates. Overlaying attribute information onto clusters allows visualization of the relationship between structure and function. There exists a large amount of structure and sequence information for cupin superfamily members for which there has been no experimental work conducted. By facilitating the prediction of the functions of these proteins, PSNs may be used to guide experimental inquiry into the cupin superfamily.

## Results and Discussion

The Pfam database [30] contains 484 unique cupin structures. For visual simplicity, only a single representative structure of a protein with multiple structures was used in the networks. This reduced the networks to 183 structures, which are listed in Supporting Information (Tables S1–S7 in File S1). Of the 183 PDB structures representing unique members of the cupin superfamily, 76 bind no metal, 49 bind iron, 18 bind nickel, 16 bind zinc, 10 bind manganese, 1 binds cadmium (protein yhhW, PDB code: 1tq5) [34], 1 binds copper (quercetin dioxygenase from *Aspergillus japonicus*, PDB code: 1juh) [35], 1 binds mercury (protein YhcH, PDB code: 1jop) [36], and 11 bind multiple metals.

Figure 3 shows pairwise similarities for this non-redundant set of structures calculated using TM-align [37] (See Methods) at two different stringency thresholds. This permits alternative views of the same structural relationships. These networks are painted according to the identity of the bound metal in the structure. In the less stringent network (Figure 3A), all but four structures have connections to a single large constellation of structures. This constellation partitions off into smaller clusters of structures in the more stringent network (Figure 3B). It can be observed that often

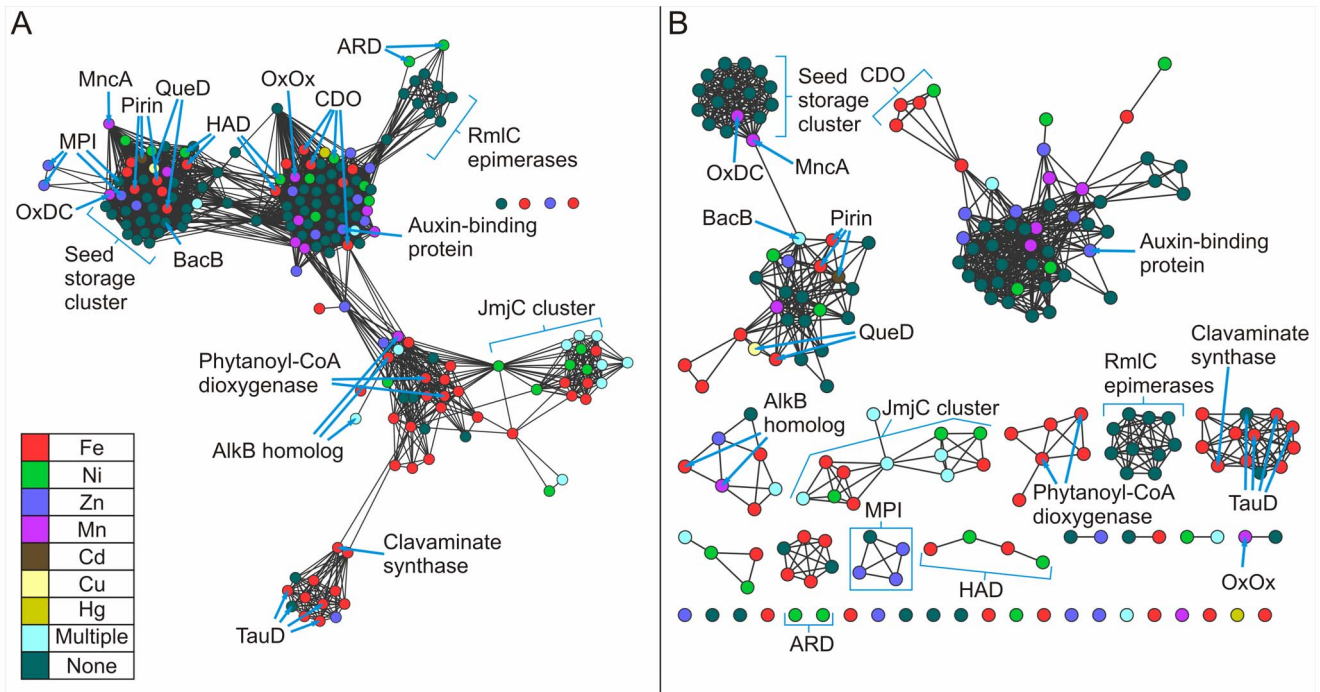
proteins of the same function cluster together. For example, at the higher stringency the catalytically inactive seed storage proteins cluster together with two metal-binding cupins, oxalate decarboxylase and the MncA protein. Other clusters analyzed in this work include the cysteine dioxygenases (CDOs), the Jumonji C domain-containing proteins (JmjC), and the RmlC epimerases.

### Seed Storage Proteins Cluster Together

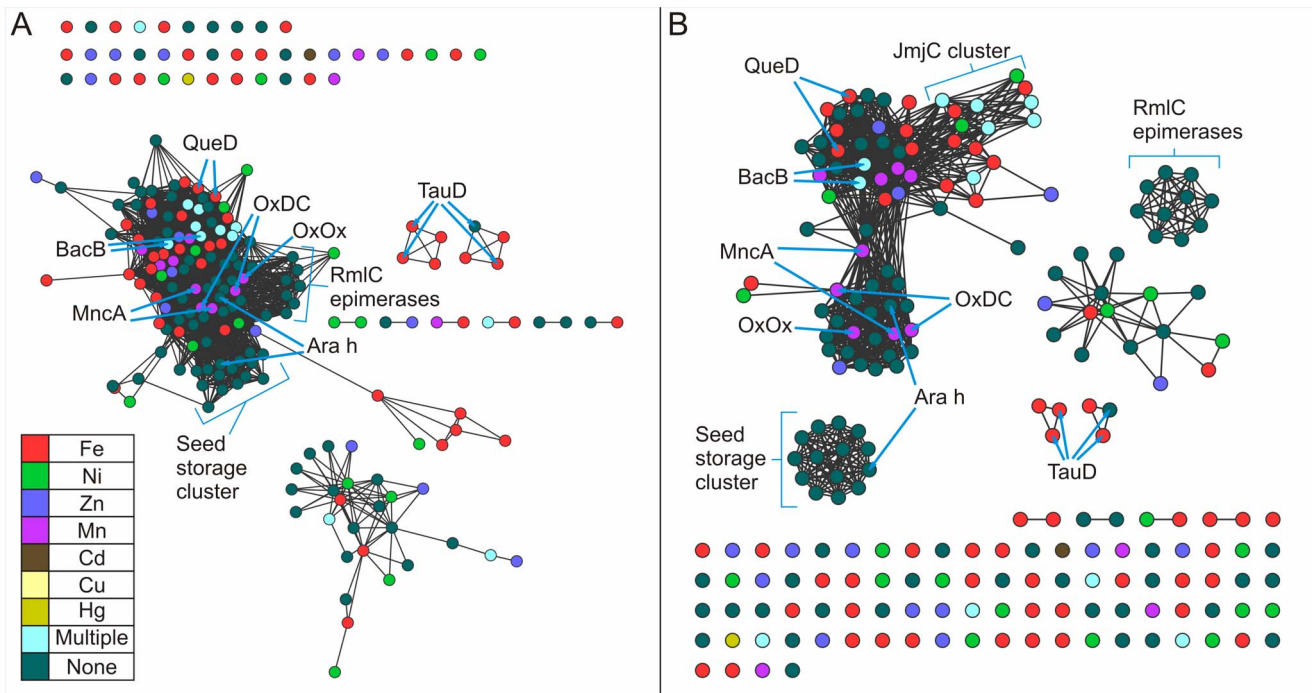
The seed storage proteins cluster together at the lower threshold (Figure 3A), but share edges with other protein structures. At the higher threshold (Figure 3B), however, these proteins partition exclusively. Members of this cluster contain two cupin domains and are, therefore, classified as bicupins. The majority of the members of this group contain no metal and include procruciferin (PDB code: 3kgl) [38], Ara h (PDB code: 3s7i) [39], phaseolin (PDB code: 2phl) [40], beta-conglycinin (PDB code: 1uij) [41], and proglycinin (PDB code: 1fxz) [42]. The two metal-containing members of this cluster are oxalate decarboxylase (PDB code: 1uw8) [43] and MncA (PDB code: 2vqa) [44] with both incorporating manganese. The architecture of this group is represented in Figure 1B by oxalate decarboxylase (OxDC) and in Figure 1C by Ara h, of interest as a major peanut allergen. The two nearly identical Mn-binding sites of OxDC are shown in Figures 2A and 2B. Previous work with protein similarity networks in general had indicated that the clustering of proteins does not change dramatically whether the domain in common is isolated or is a component of a larger multi-domain complex [22], and this is our observation also. However, we were concerned that monocupins were generally segregated from bicupins in the network built using entire structures. In an effort to explore this possible limitation, we constructed networks of isolated cupin domains for the structures used in Figure 3. Figures 4A and 4B show the cupin domain networks at the same stringencies as those in Figures 3A and 3B, respectively, also colored by metal ligand. At both stringencies, monocupin oxalate oxidase (OxOx) (Figures 1A and 2A) clusters with both the N-terminal and C-terminal cupin domains of bicupin OxDC. At the higher stringency (Figure 4B), OxDC (both domains), MncA (both domains), and OxOx partition away from the group labeled “seed storage cluster.” Interestingly, many of the seed storage proteins such as Ara h split one domain to the “seed storage cluster” and one domain remains in the group containing OxDC, MncA, and OxOx. This observation is consistent with the conclusions from phylogenetic analyses that the N- and C-terminal domains of many bicupins arose through independent evolutionary events [15] whereas others such as OxDC arose through gene duplication events [2].

Figure 5 reproduces the same network as Figure 3B but is painted by kingdom (Figure 5A) and function (Figure 5B). Figure 5A illustrates that OxDC and MncA are bacterial proteins and the only non-plant members of the seed storage cluster. OxDC is a Mn-dependent enzyme that catalyzes the carbon-carbon bond cleavage of oxalate to yield carbon dioxide and formate in a reaction with no net change in oxidation state (the only lyase represented in Figure 5B) [45]. MncA is the most abundant Mn-containing protein in cyanobacteria *Synechocystis* PCC 6803 and played a key role in a study that elucidated a mechanism whereby the compartment in which a protein is folded overrides its binding preference to control its metal content [44]. MncA is first neighbors with (shares edges) in the structure networks (Figures 3A, 3B, 5A, and 5B) with all members of the seed storage cluster as well as BacB (PDB code: 3h7j) [46] in an adjacent cluster. BacB binds both Fe and Co, has been shown to play a role in the synthesis of bacilysin, and clusters with pirins and quercetin dioxygenases (see below). Networks based on sequence (Figures 6A

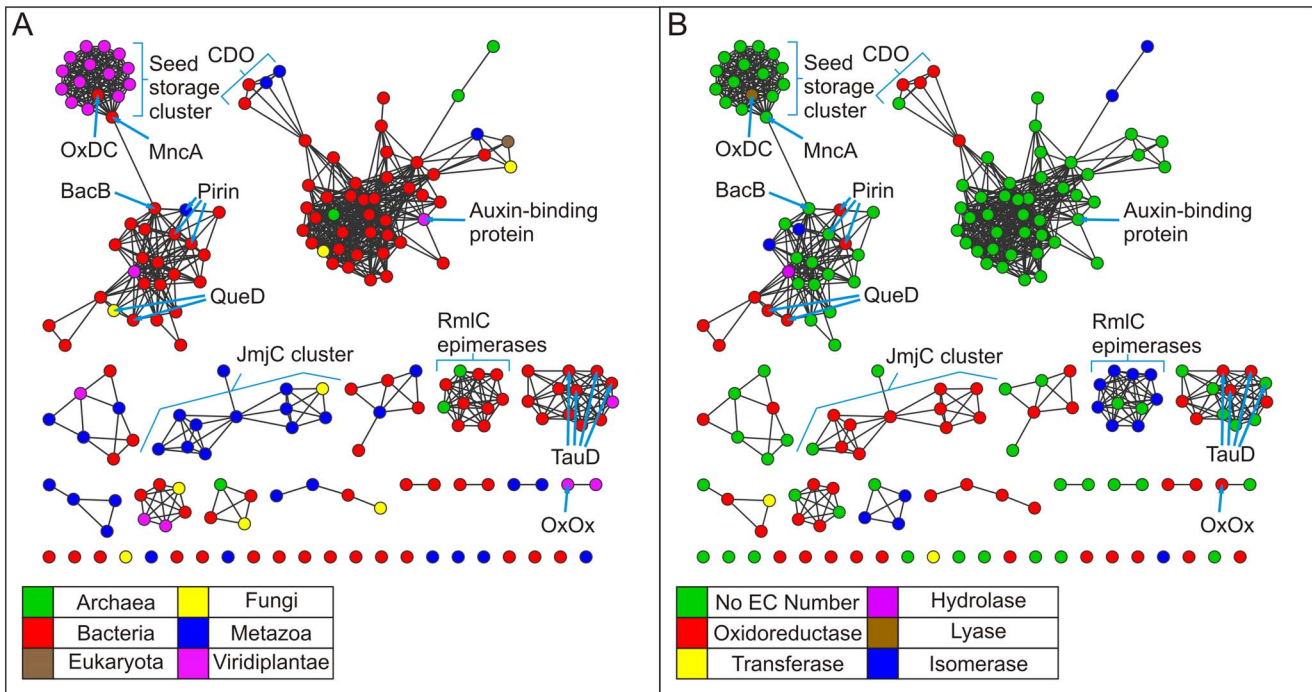




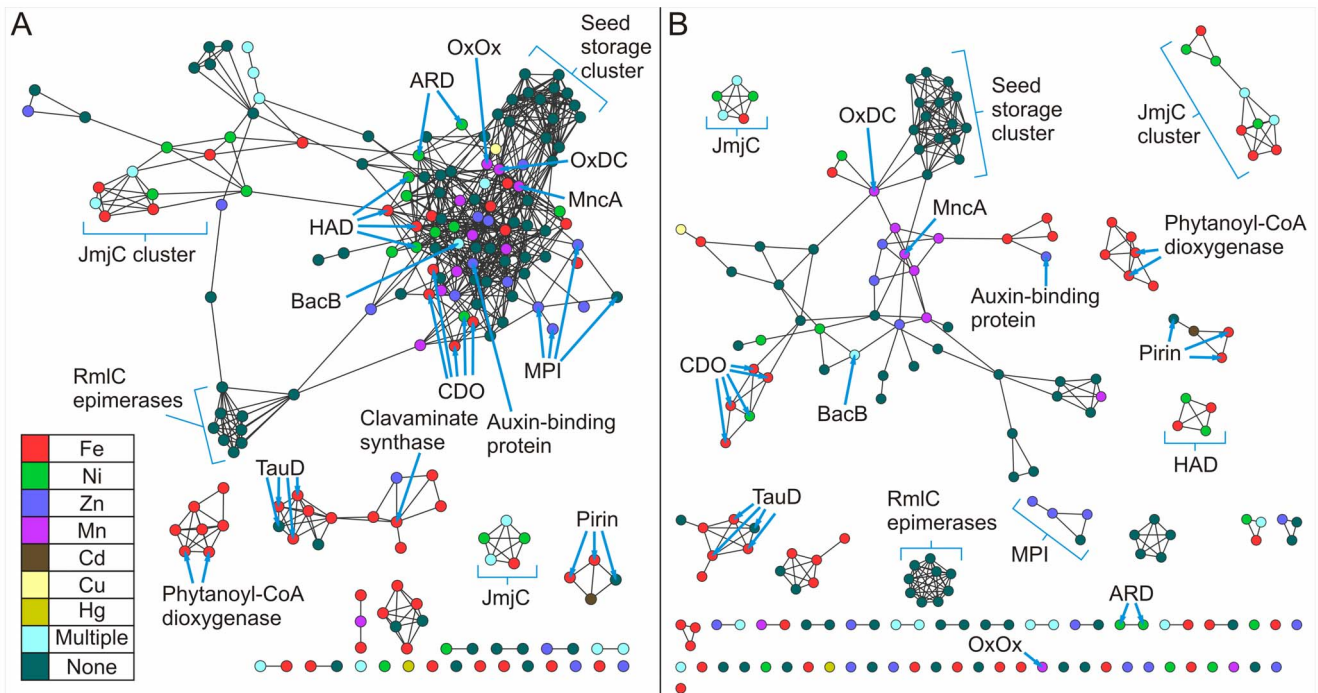
**Figure 3. Structure similarity networks of the cupin protein structures colored by metal ligand.** Pairwise similarities for a non-redundant set of 183 structures from the Pfam cupin clan (CL0029) were calculated using TM-align [37]. Each node represents a structure. Nodes were arranged using the yfiles organic layout of Cytoscape version 3.0. A. Edges between nodes were drawn only if the average TM-score  $>0.53$  for that edge. At this cutoff, the average r.m.s.d. is 2.91 Å with an average of 158.0 C $\alpha$  atoms aligned. B. Edges between nodes were drawn only if the average TM-score  $>0.65$  for that edge. At this cutoff, the average r.m.s.d. is 2.44 Å with an average of 185.4 C $\alpha$  atoms aligned. doi:10.1371/journal.pone.0074477.g003



**Figure 4. Structure similarity networks of cupin domains colored by metal ligand.** Pairwise similarities for a non-redundant set of 213 domains from the Pfam cupin clan (CL0029) were calculated using TM-align [37]. Each node represents a domain. Nodes were arranged using the yfiles organic layout of Cytoscape version 3.0. A. Edges between nodes were drawn only if the average TM-score  $>0.53$  for that edge. At this cutoff, the average r.m.s.d. is 1.73 Å with 74.2 C $\alpha$  atoms aligned. B. Edges between nodes were drawn only if the average TM-score  $>0.65$  for that edge. At this cutoff, the average r.m.s.d. is 1.42 Å with 80.1 C $\alpha$  atoms aligned. doi:10.1371/journal.pone.0074477.g004



**Figure 5. Structure similarity networks colored by species and function.** This is the same network as in 3B but colored according to A. species and B. function.  
doi:10.1371/journal.pone.0074477.g005



**Figure 6. Sequence similarity networks colored by metal ligand.** Networks were generated by all-by-all BLAST comparisons of the 183 sequences corresponding to the unique cupin structures shown in Figure 3. Nodes were arranged using the yfiles organic layout of Cytoscape version 3.0. A. Edges between nodes are drawn only if the E-value is better than of  $1E-3.5$ . At this cutoff, edges at this threshold represent alignments with a median 32.1% identity over 93 residues. B. Edges between nodes are drawn only if the E-value is better than of  $1E-6.0$ . At this cutoff, edges at this threshold represent alignments with a median 36.2% identity over 185 residues.  
doi:10.1371/journal.pone.0074477.g006

**Table 1.** JmjC protein clusters.

Protein	Species	PDB code	Pfam family	Metal ion(s)	Ref.
Lysine-specific demethylase 6B	<i>Homo sapiens</i>	2 xue	JmjC	Fe, Zn	[75]
JmjC domain-containing histone demethylation protein 1 <sup>a</sup>	<i>Homo sapiens</i>	2yu2	JmjC	Fe	
Lysine-specific demethylase 6A	<i>Homo sapiens</i>	3 avr	JmjC	Ni, Zn	[76]
catalytic domain of PHD finger protein 8	<i>Homo sapiens</i>	3k3n	JmjC	Fe	[77]
Lysine-specific demethylase 7	<i>Caenorhabditis elegans</i>	3n9l	JmjC	Fe, Zn	[78]
PHD finger protein 2	<i>Homo sapiens</i>	3pu8	JmjC	Fe	[79]
Lysine-specific demethylase 7	<i>Homo sapiens</i>	3u78	JmjC	Ni	[80]
Jumonji domain-containing protein 2A	<i>Homo sapiens</i>	2gp3	JmjC	Fe, Ni	[81]
Lysine-specific demethylase 4E	<i>Homo sapiens</i>	2w2i	JmjC	Ni	
JMJD2C catalytic domain	<i>Homo sapiens</i>	2 xml	JmjC	Ni, Zn	
JmjC domain-containing histone demethylation protein 3D	<i>Homo sapiens</i>	3 dxu	JmjC	Fe	
DNA damage-responsive transcriptional repressor RPH1	<i>Saccharomyces cerevisiae</i>	3 opt	JmjC	Ni	[82]
Bifunctional arginine demethylase and lysyl-hydroxylase JMJD6	<i>Homo sapiens</i>	3ld8	cupin 8	Fe, Hg	[83]
Lysine-specific demethylase 8	<i>Homo sapiens</i>	4 aap	cupin 8	Ni	
JmjC domain-containing protein C2orf60	<i>Homo sapiens</i>	3al6	cupin 8	Ni	[84]
factor inhibiting HIF1	<i>Caenorhabditis elegans</i>	2y0i	cupin 8	Fe	[74]
Putative asparaginyl hydroxylase	<i>Homo sapiens</i>	1 vrb	cupin 4	Fe	
Lysine-specific demethylase NO66	<i>Homo sapiens</i>	4 diq	cupin 4	Ni	
MYC-induced nuclear antigen	<i>Homo sapiens</i>	2 xdv	cupin 4	Cd, Ni, Mn	

doi:10.1371/journal.pone.0074477.t001

and 6B) similarly cluster the seed storage proteins which share edges with OxDC and MncA.

### Enzymes Cluster Based on Function

The structure network at a more permissive threshold (Figure 3A) clearly shows clustering correlated with function. For example, the four cysteine dioxygenases, which catalyzes the oxidation of the L-cysteine to its sulfenic acid by incorporating both of the oxygen atoms from O<sub>2</sub> into the product, are part of a larger cluster also containing oxalate oxidase (OxOx) and auxin-binding protein at the lower stringency (Figure 3A). In the more stringent network (Figure 3B), the CDO structures (*Homo sapiens*, PDB code: 2ic1 [47]; *Ralstonia eutropha*, PDB code: 2gm6; *Mus musculus* PDB code: 2q4s [48]; and *Pseudomonas aeruginosa* PDB code: 3 uss) share first neighbor status with each other only and are connected to the larger cluster through the putative CDO from *Bacillus subtilis* (PDB code: 3 eqe). These monocupins are represented in Figure 1E by the *Mus musculus* (mouse) enzyme. The metal ion in this structure is Ni(II) and is coordinated by three conserved residues (His 86, His88, and His140) as shown in Figure 2D. Biochemical characterization of this protein has shown that iron is required for catalytic activity [16]. CDOs cluster by function when the entire structures are compared (Figure 3), but remain separate due to the smaller region of alignment when compared at the domain level at the same stringency (Figure 4, data not indicated with arrows). The majority of cupins for which there is structural information are from bacterial organisms. The CDOs cluster by function in the sequence networks (Figure 6). These networks confirm and extend previous observations that members of the cupin superfamily cluster by function when the function is defined at a fine level, such as the oxidation of cysteine but do not do so when function is defined more broadly [15].

For example, the Fe-containing 2-oxoglutarate-dependent dioxygenases such as phytanoyl-CoA dioxygenase, taurine dioxygenase (TauD) (Figures 1H and 2G), and lysine-specific demethylases (JmjC cluster, see below) are separated in both the structure and sequence networks. Furthermore, the networks painted by metal ion (Figures 3, 4, and 6) show that neither structure-based nor sequence-based clustering correlates with the identity of the active site metal.

Structures are available for 3-hydroxyanthranilate-3,4-dioxygenase (HAD) from four species. HAD from *Cupriavidus metallidurans* (formerly *Ralstonia metallidurans*) (PDB code:1 yfu) [49,50] and *Bos taurus* (PDB code: 3fe5) [51] contains iron, while HAD from *Saccharomyces cerevisiae* (PDB code:1 zvf) and *Homo sapiens* (PDB code: 2 qnk) contains nickel. These four structures group together by function in the network generated using the whole protein (Figure 3) but not when the cupin domain alone is used to generate the network (Figure 4). The HADs cluster by function in the sequence networks (Figure 6). Similarly, four structures are available for mannose-6-phosphate isomerases (MPIs) (Figures 1F and 2E). These structures segregate exclusively at the more stringent thresholds of the whole protein structure network (Figure 3B) and sequence network (Figure 6B), but not in the domain networks (Figure 4). Of these, three contain Zn (*Bacillus subtilis*, PDB code: 1 qwr; *Salmonella typhimurium*, PDB code: 3h1m [52]; and *Candida albicans*, PDB code: 1 pmi [53]). One putative PMI was crystallized with no metal present (*Archaeoglobus fulgidus*, PDB code:1zx5).

Quercetin 2,3-dioxygenase (QueD) catalyzes the insertion of molecular oxygen into polyphenolic flavonols. Structural information is available for quercetin 2,3-dioxygenases from *Aspergillus japonicus* (PDB code: 1 juh, [35]) and *Bacillus subtilis*, (PDB code: 1y3t [54]). QueD is able to employ a variety of metals in catalysis



**Table 2.** The RmlC epimerase cluster.

Protein	Species	PDB code	# of chains	Metal ion	Ref.
dTDP-4-keto-6-deoxy-d-hexulose 3, 5-epimerase	<i>Methanobacterium thermoautotrophicum</i>	1ep0	1	none	[85]
4-keto-6-deoxy sugar epimerase (NovW)	<i>Streptomyces spheroides</i>	2c0z	1	none	[86,87]
dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase	<i>Pseudomonas aeruginosa</i>	1 rtv	1	none	[88]
dTDP-4-dehydrorhamnose 3,5 epimerase	<i>Mycobacterium tuberculosis</i>	1 upi	1	none	[88,89]
dTDP-4-keto-6-deoxy-glucose-5 epimerase EvaD	<i>Amycolatopsis orientalis</i>	1 ofn	2	none	[90]
dTDP-4-dehydrorhamnose 3,5 epimerase	<i>Sulfolobus tokodaii</i>	1 wlt	2	none	
dTDP-4-dehydrorhamnose 3,5 epimerase	<i>Streptococcus suis</i>	1 nxm	2	none	[91]
dTDP-4-dehydrorhamnose 3,5 epimerase	<i>Salmonella typhimurium</i>	1 dzr	2	none	[92]
dTDP-4-dehydrorhamnose 3,5 epimerase	<i>Bacillus anthracis</i>	3 ryk	2	none	
dTDP-6-deoxy-3,4-keto-hexulose isomerase	<i>Aneurinibacillus thermoaerophilus</i>	2pa7	2	none	[63]

doi:10.1371/journal.pone.0074477.t002

[54,55,56]. Incubation of apoprotein with transition metal ions has been done to examine the effects of different metal ions on enzymatic activity. Results yielded an activity profile with trends that were consistent with the Irving-Williams metal ion series based on the stability of metal ion complexes [57]. Data suggest that Mn(II) is the preferred cofactor for the enzyme [55]. Pirins have been shown to possess QueD activity [34], and structural information is available for pirins from *Escherichia coli* (PDB code: 1tq5) [34], *Homo sapiens* (PDB code: 1j11), and *Geobacillus kaustophilus* (PDB code: 2p17). The QueDs and pirins are members of the same larger cluster in the structure-based networks (Figure 3). Consistent with previous analyses and observations [15], the pirin N- and C- terminal domains do not share edges with each other in the domain-based networks, suggesting an independent evolution of these domains in bicupins. This is in contrast to the N- and C-terminal domain clustering of bicupins such as OxDC, QueD, and BacB from *Bacillus subtilis* where gene duplication events are proposed to play a large role in i) increasing the size of the genome and ii) producing bicupin architectures [2].

Acireductone dioxygenase (ARD) is a particularly interesting example of how the cupin scaffold can be used to catalyze different reactions in that this enzyme catalyzes different reactions depending on the type of metal ion bound in the active site [58,59]. ARD serves as a branch point in the methionine salvage pathway. This pathway returns the  $\gamma$ -thiomethyl group of methylthioadenosine to methionine [60]. The overexpression of the ARD gene in *E. coli* yields two enzymes with different activities that are able to be separated by conventional chromatographic techniques of ion exchange and hydrophobic interaction resins. One enzyme catalyzes the on-pathway oxidation of acireductone to ketoacid and formate; the other enzyme catalyzes an off-pathway oxidation with formation of CO. The only difference between the two enzymes (Fe-ARD and Ni-ARD) was the metal bound in the active sites [58]. There is structural information available for ARD from *Mus musculus* (PDB code: 1vr3) [61] and *Klebsiella pneumonia* (PDB code: 1 zrr) (Figures 1G and 2F) [62]. In the more permissive structure network (Figure 3A) both ARD structures, like the RmlC epimerases, are connected to the larger cluster through dTDP-6-deoxy-3,4-keto-hexulose (PDB code: 2pa7) [63] but lose all edges in the more stringent network (Figure 3B). Similarly, in the more permissive sequence network (Figure 6A) the ARD structures share an edge with each other as well as other proteins but partition off into an independent doublet (sharing a single edge) under a more stringent threshold (Figure 6B).

That members of the cupin superfamily cluster by function when the function is defined at a fine level allows one to make

functional inferences which may provide a starting point for biochemical investigations. For example, nearest neighbor analysis of the protein product of AT3G21360 of *Arabidopsis thaliana* (PDB code: 1y0z) [64] in the more stringent structure network (Figure 3B, not shown with arrows) suggests that the protein product of AT3G21360 may be an alpha ketoglutarate-dependent TauD. This inference is further supported by the exclusive partitioning of this protein with PDB codes “3v15” and “10s7” (both confirmed TauDs) in the more stringent domain-based network (Figure 4B). Similarly, the uncharacterized *Escherichia coli* protein yeaR (PDB code: 3bb6) shares an edge exclusively with the *Vibrio fischeri* tellurite resistance protein B (PDB code: 3dl3) in both the more stringent structure network (Figure 3B) and the less stringent domain-based network (Figure 4A) (not shown with arrows in either network).

### JmjC Proteins

The Jumonji C domain-containing proteins are a subfamily of the Fe(II)/2-oxoglutarate-dependent oxygenases. Sequence similarity between JmjC and cupin metalloenzyme domains allowed the prediction of active-site residues in JmjC domains and provided the insights that guided the experimental determination of the reactions catalyzed by these enzymes [65]. The JmjC domain-containing factor inhibiting hypoxia (FIH) (Figures 1I and 2H) was identified as an asparaginyl hydroxylase that transcriptionally regulates the activity of hypoxia-inducible factor (HIF) [66]. JmjC domain-containing proteins were proposed to function as histone demethylases in regulating chromatin structure [67] and many of the JmjC domain-containing enzymes have been shown to comprise the largest class of histone demethylases that catalyze lysine demethylation of mono-, di-, and trimethylated lysine residues through the formation of a hemiaminal intermediate that yields the demethylated product and formaldehyde [68,69]. The lysine demethylases are represented by lysine-specific demethylase 6B in Figures 1J, 2I, and 2J. Additionally, JmjC domain-containing enzymes have been identified that have RNA hydroxylase activity [21,70].

At the time of this writing, the cupin clan of the Pfam database contains structures of 13 unique JmjC domain-containing proteins (Table 1). Most (12) of these are given the Pfam family designation ‘JmjC’. These 12 structures form an exclusive cluster together in Figure 3B. The other JmjC domain-containing protein structure (PDB code: 3ld8) is designated ‘cupin 8’ in Pfam along with an additional three proteins that have structural information available (Table 1). These four structures form an exclusive cluster in Figure 3B (not labeled under the AlkB-containing cluster). Finally,

the three proteins with structures available designated ‘cupin 4’ in Pfam have been included in Table 1 for comparison. In Figure 3B these structures exist as a pair of nodes and as a single node. In the more stringent sequence similarity network (Figure 6B), six JmjC domain-containing protein structures and the ‘cupin 8’ structures cluster together (PDB codes: 2yu2, 3k3n, 3n9l, 3pu8, 3u78, 3ld8, 4 aap, 3a16, 2y0i), while five JmjC domain-containing protein structures form an exclusive cluster (PDB codes: 2gp3, 2w2i, 2 xml, 3 dxu, 3 opt). The three ‘cupin 4’ structures partition into an exclusive triad. These networks capture a current snapshot of relationships within this subfamily and can be used to update relationships and guide experimental design as new structures become available.

### RmlC Epimerases

The RmlC epimerases in Figure 3A cluster together but share edges with other protein structures. At a more stringent threshold (Figure 3B), however, these same 10 structures (Table 2) cluster independently from other structures. The epimerases are monocupins. Members of this group do not bind a metal ion and are represented in Figure 1D by NovW, a 4-keto-6-deoxy sugar epimerase (PDB code: 2c0z) [71]. This grouping is in agreement with a published structure-based phylogenetic analysis of the cupin superfamily generated using a structure dissimilarity matrix through pairwise structure-based alignment of 52 cupin proteins [15]. Our analysis includes 2 additional proteins whose structures were solved after the phylogenetic analysis was published. The network in Figure 3B clearly groups the functionally similar RmlC epimerases together as does the phylogenetic analysis, providing further validation that PSNs recapitulate much of the information present in phylogenetic trees [22]. When the network is constructed of isolated domains, the 10 monocupin RmlC epimerase domains form the same cluster as in the whole protein network (Figure 4B). Furthermore, a sequence-based network clusters 9 of the 10 epimerases together only when edges between nodes are drawn if the E-value is better than of  $1E-6.0$  (Figure 6B). In this case the enzyme from *Aneurinibacillus thermoaerophilus* is excluded from the cluster.

### Conclusions

Protein similarity networks (structure and sequence) of the cupin superfamily recapitulate and complement phylogenetic studies. Structure- and sequence-based clustering of cupin members reflects functional clustering. Networks based only on cupin domains and networks based on the whole proteins provide complementary information. Domain-clustering supports phylogenetic conclusions that the N- and C- terminal domains of bicupin proteins evolved independently. Interestingly, although many functionally similar enzymatic cupin members bind the same active site metal ion, the structure and sequence clustering does not correlate with the identity of the bound metal. PSNs are expected to be a valuable tool for directing experimental work and for predicting the functions of uncharacterized members of the cupin superfamily.

### Methods

#### Dataset Curation

All structures associated with the PFAM cupin clan (CL0029) were downloaded 31 March 2013. There were 945 chains from 484 different structures (many structures had multiple chains). However, many structures were that of the same protein, a problem that could cause an undue distortion of network topology.

Duplicates were manually removed to avoid this issue, reducing the set of proteins to 183 unique structures. The structure used to represent duplicate structures was selected arbitrarily. Initially the structures of individual chains of whole protein structures were compared to each other. This approach, however, also resulted in the unequal representation of the 183 structures in the network. Therefore we treated the overall quaternary organization of proteins with multiple chains as a single structure. Biological, ligand, and domain information were obtained using the RESTful web service interface provided by the RCSB. Only the biologically-significant transition metals were considered when painting the networks by bound metal. The Taxonomy Database from the NCBI was used to classify species into their respective domains and phyla. For the sequence similarity network, UniProt was used to map the PDB IDs used in the structure similarity network to their respective sequences in the UniProtKB database.

### Building the PSNs

Pythoscape was integral in the construction of the network. Pythoscape (1) imported sequences and structures into a database, (2) deployed TM-align and BLASTp for edge calculations, and (3) exported the completed database as a Cytoscape network file. Because TM-Align edge scores are directional (i.e., the score for the edge from A to B might be different than the score from B to A), edges were filtered based on the average of the two scores. Cytoscape 3.0 was used for network visualization in the as well as edge filtering. The Organic Layout was used as described in Atkinson, Babbitt et al. To construct the structure network composed only of domains, a script was written for PyMOL to extract and save the parts of the PDB files that had a domain as defined as by the RCSB. Domains outside of the PFAM cupin clan were manually removed from the network. Networks were visualized using two thresholds in order to illustrate effects edge stringency had on certain clusters. TM-scores of 0.53 and 0.65 were used as thresholds for the structure networks, and E-values of  $1 \times 10^{-3.5}$  and  $1 \times 10^{-6.0}$  were used as thresholds for the sequence networks. These values were selected based on overall visual appeal of the resulting layout in Cytoscape.

### Supporting Information

**File S1 Contains:** Table S1. Iron-containing structures used in the networks. Table S2. Nickel-containing structures used in the networks. Table S3. Zinc-containing structures used in the networks. Table S4. Manganese-containing structures used in the networks. Table S5. Structures used in the networks which contain copper, mercury, and cadmium, respectively. Table S6. Structures used in the networks that contain multiple metals. Table S7. Structures used in the networks that contain no transition metal. (DOCX)

### Acknowledgments

The authors wish to thank Dr. Patrick Frantom for introducing us to the usefulness of protein similarity networks and for his generous help in getting us started on this project.

### Author Contributions

Conceived and designed the experiments: RU EWM. Performed the experiments: RU EWM. Analyzed the data: RU EWM. Contributed reagents/materials/analysis tools: EWM. Wrote the paper: RU EWM.



## References

- Dunwell JM, Purvis A, Khuri S (2004) Cupins: the most functionally diverse protein superfamily? *Phytochemistry* 65: 7–17.
- Dunwell JM, Khuri S, Gane PJ (2000) Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. *Microbiol Mol Biol Rev* 64: 153–179.
- Dunwell JM, Gane PJ (1998) Microbial relatives of seed storage proteins: conservation of motifs in a functionally diverse superfamily of enzymes. *J Mol Evol* 46: 147–154.
- Dunwell JM (1998) Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins. *Biotechnol Genet Eng Rev* 15: 1–32.
- Dunwell JM, Culham A, Carter CE, Sosa-Aguirre CR, Goodenough PW (2001) Evolution of functional diversity in the cupin superfamily. *Trends Biochem Sci* 26: 740–746.
- Woo EJ, Dunwell JM, Goodenough PW, Marvier AC, Pickersgill RW (2000) Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nat Struct Biol* 7: 1036–1040.
- Anantharaman V, Aravind L, Koonin EV (2003) Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr Opin Chem Biol* 7: 12–20.
- Gerlt JA, Babbitt PC (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* 70: 209–246.
- Glasner ME, Gerlt JA, Babbitt PC (2006) Evolution of enzyme superfamilies. *Curr Opin Chem Biol* 10: 492–497.
- Galperin MY, Koonin EV (2012) Divergence and convergence in enzyme evolution. *J Biol Chem* 287: 21–28.
- Allewell NM (2012) Thematic minireview series on enzyme evolution in the post-genomic era. *J Biol Chem* 287: 1–2.
- Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247: 536–540.
- Lo Conte L, Brenner SE, Hubbard TJ, Chothia C, Murzin AG (2002) SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res* 30: 264–267.
- Andreva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, et al. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res* 32: D226–229.
- Agarwal G, Rajavel M, Gopal B, Srinivasan N (2009) Structure-based phylogeny as a diagnostic for functional characterization of proteins with a cupin fold. *PLoS One* 4: e5736.
- McCoy JG, Bailey LJ, Bitto E, Bingman CA, Aceti DJ, et al. (2006) Structure and mechanism of mouse cysteine dioxygenase. *Proc Natl Acad Sci U S A* 103: 3084–3089.
- Stirk HJ, Woolfson DN, Hutchinson EG, Thornton JM (1992) Depicting topology and handedness in jellyroll structures. *FEBS Lett* 308: 1–3.
- Anantharaman V, Koonin EV, Aravind L (2001) Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *J Mol Biol* 307: 1271–1292.
- Woo EJ, Baully J, Chen JG, Marshall J, Macdonald H, et al. (2000) Crystallization and preliminary X-ray analysis of the auxin receptor ABP1. *Acta Crystallogr D Biol Crystallogr* 56 (Pt 11): 1476–1478.
- Anand R, Dorrestein PC, Kinsland C, Begley TP, Ealick SE (2002) Structure of oxalate decarboxylase from *Bacillus subtilis* at 1.75 angstrom resolution. *Biochem* 41: 7659–7669.
- Iyer LM, Abhiman S, de Souza RF, Aravind L (2010) Origin and evolution of peptide-modifying dioxygenases and identification of the wybutosine hydroxylase/hydroperoxidase. *Nucleic Acids Res* 38: 5261–5279.
- Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* 4: e4345.
- Gerlt JA, Babbitt PC, Jacobson MP, Almo SC (2012) Divergent evolution in enolase superfamily: strategies for assigning functions. *J Biol Chem* 287: 29–34.
- Brown SD, Babbitt PC (2012) Inference of functional properties from large-scale analysis of enzyme superfamilies. *J Biol Chem* 287: 35–42.
- Atkinson HJ, Babbitt PC (2009) Glutathione transferases are structural and functional outliers in the thioredoxin fold. *Biochemistry* 48: 11108–11116.
- Atkinson HJ, Babbitt PC (2009) An atlas of the thioredoxin fold class reveals the complexity of function-enabling adaptations. *PLoS Comput Biol* 5: e1000541.
- Hicks MA, Barber AE, 2nd, Giddings LA, Caldwell J, O'Connor SE, et al. (2011) The evolution of function in strictosidine synthase-like proteins. *Proteins* 79: 3082–3098.
- Atkinson HJ, Babbitt PC, Sajid M (2009) The global cysteine peptidase landscape in parasites. *Trends Parasitol* 25: 573–581.
- Blaby-Haas CE, Merchant SS (2012) The ins and outs of algal metal transport. *Biochim Biophys Acta* 1823: 1531–1552.
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290–301.
- Pieper U, Chiang R, Seffernick JJ, Brown SD, Glasner ME, et al. (2009) Target selection and annotation for the structural genomics of the amidohydrolase and enolase superfamilies. *J Struct Funct Genomics* 10: 107–125.
- Barber AE, 2nd, Babbitt PC (2012) Pythoscape: a framework for generation of large protein similarity networks. *Bioinformatics* 28: 2845–2846.
- Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431–432.
- Adams M, Jia Z (2005) Structural and biochemical analysis reveal pirins to possess quercetinase activity. *J Biol Chem* 280: 28675–28682.
- Fuseti F, Schroter KH, Steiner RA, van Noort PI, Pijning T, et al. (2002) Crystal structure of the copper-containing quercetin 2,3-dioxygenase from *Aspergillus japonicus*. *Structure* 10: 259–268.
- Teplyakov A, Obmolova G, Toedt J, Galperin MY, Gilliland GL (2005) Crystal structure of the bacterial YhcH protein indicates a role in sialic acid catabolism. *J Bacteriol* 187: 5520–5527.
- Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33: 2302–2309.
- Tandang-Silvas MR, Fukuda T, Fukuda C, Prak K, Cabanos C, et al. (2010) Conservation and divergence on plant seed 11 S globulins based on crystal structures. *Biochim Biophys Acta* 1804: 1432–1442.
- Chruszcz M, Maleki SJ, Majorek KA, Demas M, Bublin M, et al. (2011) Structural and immunologic characterization of Ara h 1, a major peanut allergen. *J Biol Chem* 286: 39318–39327.
- Lawrence MC, Izard T, Beuchat M, Blagrove RJ, Colman PM (1994) Structure of phaseolin at 2.2 Å resolution. Implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins. *J Mol Biol* 238: 748–776.
- Maruyama N, Maruyama Y, Tsuruki T, Okuda E, Yoshikawa M, et al. (2003) Creation of soybean beta-conglycinin beta with strong phagocytosis-stimulating activity. *Biochim Biophys Acta* 1648: 99–104.
- Adachi M, Takenaka Y, Gidamis AB, Mikami B, Utsumi S (2001) Crystal structure of soybean proglycinin A1aB1b homotrimer. *J Mol Biol* 305: 291–305.
- Just VJ, Stevenson CE, Bowater L, Tanner A, Lawson DM, et al. (2004) A closed conformation of *Bacillus subtilis* oxalate decarboxylase OxdC provides evidence for the true identity of the active site. *J Biol Chem* 279: 19867–19874.
- Tottey S, Waldron KJ, Firkbank SJ, Reale B, Bessant C, et al. (2008) Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature* 455: 1138–1142.
- Tanner A, Bowater L, Fairhurst SA, Bornemann S (2001) Oxalate decarboxylase requires manganese and dioxygen for activity - Overexpression and characterization of *Bacillus subtilis* YvrK and YoaN. *J Biol Chem* 276: 43627–43634.
- Rajavel M, Mitra A, Gopal B (2009) Role of *Bacillus subtilis* BacB in the synthesis of bacilysin. *J Biol Chem* 284: 31882–31892.
- Ye S, Wu X, Wei L, Tang D, Sun P, et al. (2007) An insight into the mechanism of human cysteine dioxygenase. Key roles of the thioether-bonded tyrosine-cysteine cofactor. *J Biol Chem* 282: 3391–3402.
- Levin EJ, Kondrashov DA, Wesenberg GE, Phillips GN, Jr. (2007) Ensemble refinement of protein crystal structures: validation and application. *Structure* 15: 1040–1052.
- Zhang Y, Colabroy KL, Begley TP, Ealick SE (2005) Structural studies on 3-hydroxyanthranilate-3,4-dioxygenase: the catalytic mechanism of a complex oxidation involved in NAD biosynthesis. *Biochemistry* 44: 7632–7643.
- Colabroy KL, Zhai H, Li T, Ge Y, Zhang Y, et al. (2005) The mechanism of inactivation of 3-hydroxyanthranilate-3,4-dioxygenase by 4-chloro-3-hydroxyanthranilate. *Biochemistry* 44: 7623–7631.
- Dilovic I, Gliubich F, Malpeli G, Zanotti G, Matkovic-Calogovic D (2009) Crystal structure of bovine 3-hydroxyanthranilate 3,4-dioxygenase. *Biopolymers* 91: 1189–1195.
- Sagurthi SR, Gowda G, Savithri HS, Murthy MR (2009) Structures of mannose-6-phosphate isomerase from *Salmonella typhimurium* bound to metal atoms and substrate: implications for catalytic mechanism. *Acta Crystallogr D Biol Crystallogr* 65: 724–732.
- Cleasby A, Wonacott A, Skarzynski T, Hubbard RE, Davies GJ, et al. (1996) The x-ray crystal structure of phosphomannose isomerase from *Candida albicans* at 1.7 angstrom resolution. *Nat Struct Biol* 3: 470–479.
- Gopal B, Madan LL, Betz SF, Kossiakoff AA (2005) The crystal structure of a quercetin 2,3-dioxygenase from *Bacillus subtilis* suggests modulation of enzyme activity by a change in the metal ion at the active site(s). *Biochem* 44: 193–201.
- Schaab MR, Barney BM, Francisco WA (2006) Kinetic and spectroscopic studies on the quercetin 2,3-dioxygenase from *Bacillus subtilis*. *Biochemistry* 45: 1009–1016.
- Bowater L, Fairhurst SA, Just VJ, Bornemann S (2004) *Bacillus subtilis* YxaG is a novel Fe-containing quercetin 2,3-dioxygenase. *FEBS Lett* 557: 45–48.
- Irving H, Williams RJP (1953) The Stability of Transition-Metal Complexes. *Journal of the Chemical Society*: 3192–3210.
- Dai Y, Wensink PC, Abeles RH (1999) One protein, two enzymes. *J Biol Chem* 274: 1193–1195.
- Dai Y, Pochapsky TC, Abeles RH (2001) Mechanistic studies of two dioxygenases in the methionine salvage pathway of *Klebsiella pneumoniae*. *Biochemistry* 40: 6379–6387.

60. Pochapsky TC, Pochapsky SS, Ju T, Mo H, Al-Mjeni F, et al. (2002) Modeling and experiment yields the structure of acireductone dioxygenase from *Klebsiella pneumoniae*. *Nat Struct Biol* 9: 966–972.
61. Xu Q, Schwarzenbacher R, Krishna SS, McMullan D, Agarwalla S, et al. (2006) Crystal structure of acireductone dioxygenase (ARD) from *Mus musculus* at 2.06 angstrom resolution. *Proteins* 64: 808–813.
62. Pochapsky TC, Pochapsky SS, Ju T, Hoefler C, Liang J (2006) A refined model for the structure of acireductone dioxygenase from *Klebsiella* ATCC 8724 incorporating residual dipolar couplings. *Journal of Biomolecular NMR* 34: 117–127.
63. Davis ML, Thoden JB, Holden HM (2007) The x-ray structure of dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase. *J Biol Chem* 282: 19227–19236.
64. Bitto E, Bingman CA, Allard ST, Wesenberg GE, Aceti DJ, et al. (2005) The structure at 2.4 Å resolution of the protein from gene locus At3g21360, a putative Fe(II)/2-oxoglutarate-dependent enzyme from *Arabidopsis thaliana*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 61: 469–472.
65. Clissold PM, Ponting CP (2001) JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2beta. *Trends Biochem Sci* 26: 7–9.
66. Lando D, Peet DJ, Gorman JJ, Whelan DA, Whitelaw ML, et al. (2002) FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev* 16: 1466–1471.
67. Treweek SC, McLaughlin PJ, Allshire RC (2005) Methylation: lost in hydroxylation? *EMBO Rep* 6: 315–320.
68. Klose RJ, Kallin EM, Zhang Y (2006) JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* 7: 715–727.
69. Loenarz C, Schofield CJ (2011) Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci* 36: 7–18.
70. Noma A, Ishitani R, Kato M, Nagao A, Nureki O, et al. (2010) Expanding role of the jumonji C domain as an RNA hydroxylase. *J Biol Chem* 285: 34503–34507.
71. Jakimowicz P, Tello M, Meyers CL, Walsh CT, Buttner MJ, et al. (2006) The 1.6-Å resolution crystal structure of NovW: a 4-keto-6-deoxy sugar epimerase from the novobiocin biosynthetic gene cluster of *Streptomyces spheroides*. *Proteins* 63: 261–265.
72. Opaleye O, Rose RS, Whittaker MM, Woo EJ, Whittaker JW, et al. (2006) Structural and spectroscopic studies shed light on the mechanism of oxalate oxidase. *J Biol Chem* 281: 6428–6433.
73. O'Brien JR, Schuller DJ, Yang VS, Dillard BD, Lanzilotta WN (2003) Substrate-induced conformational changes in *Escherichia coli* taurine/alpha-ketoglutarate dioxygenase and insight into the oligomeric structure. *Biochemistry* 42: 5547–5554.
74. Yang M, Chowdhury R, Ge W, Hamed RB, McDonough MA, et al. (2011) Factor-inhibiting hypoxia-inducible factor (FIH) catalyses the post-translational hydroxylation of histidiny residues within ankyrin repeat domains. *FEBS J* 278: 1086–1097.
75. Kruidenier L, Chung CW, Cheng Z, Liddle J, Che K, et al. (2012) A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* 488: 404–408.
76. Sengoku T, Yokoyama S (2011) Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A. *Genes Dev* 25: 2266–2277.
77. Yu L, Wang Y, Huang S, Wang J, Deng Z, et al. (2010) Structural insights into a novel histone demethylase PHF8. *Cell Res* 20: 166–173.
78. Yang Y, Hu L, Wang P, Hou H, Lin Y, et al. (2010) Structural insights into a dual-specificity histone demethylase ceKDM7A from *Caenorhabditis elegans*. *Cell Res* 20: 886–898.
79. Horton JR, Upadhyay AK, Hashimoto H, Zhang X, Cheng X (2011) Structural basis for human PHF2 Jumonji domain interaction with metal ions. *J Mol Biol* 406: 1–8.
80. Upadhyay AK, Rotili D, Han JW, Hu R, Chang Y, et al. (2012) An analog of BIX-01294 selectively inhibits a family of histone H3 lysine 9 Jumonji demethylases. *J Mol Biol* 416: 319–327.
81. Chen Z, Zang J, Whetstone J, Hong X, Davrazou F, et al. (2006) Structural insights into histone demethylation by JMJD2 family members. *Cell* 125: 691–702.
82. Chang Y, Wu J, Tong XJ, Zhou JQ, Ding J (2011) Crystal structure of the catalytic core of *Saccharomyces cerevisiae* histone demethylase Rph1: insights into the substrate specificity and catalytic mechanism. *Biochem J* 433: 295–302.
83. Hong X, Zang J, White J, Wang C, Pan CH, et al. (2010) Interaction of JMJD6 with single-stranded RNA. *Proc Natl Acad Sci U S A* 107: 14568–14572.
84. Kato M, Arais Y, Noma A, Nagao A, Suzuki T, et al. (2011) Crystal structure of a novel JmjC-domain-containing protein, TYW5, involved in tRNA modification. *Nucleic Acids Res* 39: 1576–1585.
85. Christendat D, Saridakis V, Dharamsi A, Bochkarev A, Pai EF, et al. (2000) Crystal structure of dTDP-4-keto-6-deoxy-D-hexulose 3,5-epimerase from *Methanobacterium thermoautotrophicum* complexed with dTDP. *J Biol Chem* 275: 24608–24612.
86. Jakimowicz P, Freil Meyers CL, Walsh CT, Buttner MJ, Lawson DM (2003) Crystallization and preliminary X-ray studies on the putative dTDP sugar epimerase NovW from the novobiocin biosynthetic cluster of *Streptomyces spheroides*. *Acta Crystallogr D Biol Crystallogr* 59: 1507–1509.
87. Jakimowicz P, Tello M, Freil Meyers CL, Walsh CT, Buttner MJ, et al. (2006) The 1.6-Å resolution crystal structure of NovW: A 4-keto-6-deoxy sugar epimerase from the novobiocin biosynthetic gene cluster of *Streptomyces spheroides*. *Proteins* 63: 261–265.
88. Dong C, Major LL, Srikannathasan V, Errey JC, Giraud MF, et al. (2007) RmlC, a C3' and C5' carbohydrate epimerase, appears to operate via an intermediate with an unusual twist boat conformation. *J Mol Biol* 365: 146–159.
89. Kantardjiev KA, Kim CY, Naranjo C, Waldo GS, Lakin T, et al. (2004) *Mycobacterium tuberculosis* RmlC epimerase (Rv3465): a promising drug-target structure in the rhamnose pathway. *Acta Crystallogr D Biol Crystallogr* 60: 895–902.
90. Merkel AB, Temple GK, Burkart MD, Losey HC, Beis K, et al. (2002) Purification, crystallization and preliminary structural studies of dTDP-4-keto-6-deoxy-glucose-5-epimerase (EvaD) from *Amycolatopsis orientalis*, the fourth enzyme in the dTDP-L-epivancosamine biosynthetic pathway. *Acta Crystallogr D Biol Crystallogr* 58: 1226–1228.
91. Dong C, Major LL, Allen A, Blankenfeldt W, Maskell D, et al. (2003) High-resolution structures of RmlC from *Streptococcus suis* in complex with substrate analogs locate the active site of this class of enzyme. *Structure* 11: 715–723.
92. Giraud MF, Leonard GA, Field RA, Berling C, Naismith JH (2000) RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. *Nat Struct Biol* 7: 398–402.