

MetalPDB in 2018: a database of metal sites in biological macromolecular structures

Valeria Putignano¹, Antonio Rosato^{1,2}, Lucia Banci^{1,2} and Claudia Andreini^{1,2,*}

¹Magnetic Resonance Center (CERM)—University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy and

²Department of Chemistry—University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Received September 14, 2017; Revised October 10, 2017; Editorial Decision October 11, 2017; Accepted October 20, 2017

ABSTRACT

MetalPDB (<http://metalweb.cerm.unifi.it/>) is a database providing information on metal-binding sites detected in the three-dimensional (3D) structures of biological macromolecules. MetalPDB represents such sites as 3D templates, called Minimal Functional Sites (MFSs), which describe the local environment around the metal(s) independently of the larger context of the macromolecular structure. The 2018 update of MetalPDB includes new contents and tools. A major extension is the inclusion of proteins whose structures do not contain metal ions although their sequences potentially contain a known MFS. In addition, MetalPDB now provides extensive statistical analyses addressing several aspects of general metal usage within the PDB, across protein families and in catalysis. Users can also query MetalPDB to extract statistical information on structural aspects associated with individual metals, such as preferred coordination geometries or aminoacidic environment. A further major improvement is the functional annotation of MFSs; the annotation is manually performed via a password-protected annotator interface. At present, ~50% of all MFSs have such a functional annotation. Other noteworthy improvements are bulk query functionality, through the upload of a list of PDB identifiers, and ftp access to MetalPDB contents, allowing users to carry out in-depth analyses on their own computational infrastructure.

INTRODUCTION

For the large majority of organisms, 30–40% of proteins require one or more metal ions to perform their biological function in cells (1;2). Additionally, metal ions play a decisive role in stabilizing the structure of nucleic acids (3). MetalPDB (4) is a resource derived from the automated analysis of all the three-dimensional (3D) structures of the

adducts between biological macromolecules and metal ions or metal-containing cofactors available from the Protein Data Bank (PDB, <http://www.wwpdb.org/>) (5). MetalPDB stores the metal sites observed in PDB structures in the form of Minimal Functional Sites (MFSs) (6;7). Each MFS is the ensemble of atoms of the metal cofactor, the metal ligands and any other residue or chemical species within 5 Å from a ligand. The MFS describes the local 3D environment around the cofactor, independently of the larger context of the macromolecular structure in which it is embedded. The usefulness of the MFS concept has its chemico-physical foundation in the fact that the local environment of the metal has a determinant role in tuning its properties and thus its chemical reactivity. Consequently, MFSs can provide an unbiased insight into the function or mechanism of action of a metalloprotein (i.e. a protein that binds at least one metal ion or metal-containing cofactor) (6;8). The structural comparison of MFSs is useful also to predict function from 3D structure in the absence of experimental biochemical data. MetalS³ tool is designed to search MetalPDB for all those sites that have a similar local structure with a query site (9).

Since its first release, in 2012, MetalPDB has been widely exploited by the scientific community. In the last 12 months, there have been on average 1450 unique IPs contacting the database each month, corresponding on average to almost 4000 visits (a new visit is counted if the same IP makes requests at half-hour intervals or longer). The current release includes 287 122 sites from 50 797 structures. It was 175 115 in the first release of MetalPDB (64% growth in 6 years). MetalPDB is updated monthly in an automated manner.

In the current update of MetalPDB, we extended its contents to include various new features and expanded the information available via the web interface. A number of improvements were made to the usability of the web interface, including bulk query functionality and faster visualization of pages. As a major upgrade, we specifically addressed the identification of potential MFSs in 3D structures lacking the metal cofactor. In addition, statistical analyses on the MetalPDB contents are now available on the web site, in order to provide a better understanding of the diversity of the biochemical roles of metals.

*To whom correspondence should be addressed. Tel: +39 55 4574267; Email: andreini@cerm.unifi.it

New contents of MetalPDB

We added secondary structure and solvent accessibility to the precomputed analyses of the structural properties of MFSs. For each metalloprotein, we used ProMotif (<http://www.img.bio.uni-goettingen.de/ms-www/internal/manuals/promotif/promotif.html>) (10) to calculate the secondary structure elements of the entire 3D structure and then linked this information to the MFSs within the structure. The same procedure was applied with the program NACCESS (<http://wolf.bms.umist.ac.uk/naccess/>) to compute the solvent accessibility of the metal-binding residues in each MFS. For the calculation of solvent accessibility, each chain in the structure was considered individually and the steric hindrance of the metal neglected.

We introduced functional annotations for MFSs. All equivalent sites (i.e. MFSs that occur at the same position within a conserved protein fold, as observed in the structural alignment of all the chains of the superfamily, and bind the same metal ions) share the same functional annotation so the clustering procedure is critical for the quality of annotation. To improve the homogeneity of groups of equivalent sites we revised our previous procedure (4) (see point 7 of the Section Database Construction) by using exclusively the Pfam domain classification (11) as the criterion to create protein superfamilies. Functional annotations are manually curated via a dedicated, password-protected annotator interface. This interface uses drop-down menus and a guided annotation procedure in order to minimize clerical errors. At the top level, we annotate the physiological relevance of each MFS by assigning it to one of these classes: ‘Physiological site’, ‘Modified Physiological site’, ‘Not physiological site’ and ‘Unknown’ (for a description, see http://metalweb.cerm.unifi.it/help/functional_annotation/). A MFS is considered to be physiological only if all the metal ions identified in the structure correspond to those required for the system to function in the cell (native metal ions), and all and only the required metals are present. In a modified physiological site, at least one metal ion has been removed, added or substituted by another metal with respect to the physiological site. A not physiological site is one that is known to be not relevant *in vivo*. When a metal ion in a structure has no donor atoms in its first coordination sphere it is automatically annotated as ‘Not physiological’; this can happen e.g. if a water molecule in the crystal structure was incorrectly assigned as a metal by the depositors. Each physiological site has one or more associated functions among ‘Catalytic’, ‘Structural’, ‘Transport’, ‘Electron transfer’, ‘Regulatory’, ‘Substrate’ and ‘Protection’ (see http://metalweb.cerm.unifi.it/help/functional_annotation/) (12). Some of these terms have a further level of annotation to improve the information content of the record. At present, a functional information is available for the majority of the sites binding iron or copper (Table 1).

A commonly asked question is what the structural impact of metal-binding is at the local and/or global structural level. To address this the 3D structures of the same protein with and without the cofactor needs to be compared. We therefore implemented a protocol to identify protein structures related to a structurally characterized MFS avail-

Table 1. Percentage of annotated MFSs, grouped by metal. Data are shown only for essential metals (18)

| Metal ion | Percentage of annotated sites |
|-----------|-------------------------------|
| Cu | 90% |
| Fe | 86% |
| Mg | 70% |
| Ni | 35% |
| Mn | 34% |
| K | 32% |
| Na | 29% |
| Mo | 22% |
| Co | 21% |
| Zn | 17% |
| W | 12% |
| Ca | 12% |
| V | 3% |

This percentage reports on the number of MFSs with a functional annotation of any type with respect to the total number of MFSs in MetalPDB.

able in MetalPDB but devoid of the metal cofactor (apo-structures). To this end, we generated a multiple sequence alignment between all chains that bind equistructural MFSs (i.e. MFSs that occur at the same position within a conserved protein fold, regardless of the chemical identity of the bound metal) and the chains of apo-structures that have at least 50% identity with at least one of them. Potential MFSs in apo-structures are then identified based on the conservation of all metal-binding residues in this alignment. This procedure identifies apo-structures with the metal-binding pattern (Figure 1). Chains lacking one or more of the metal-binding residues probably have lost or significantly changed their interaction with the metal cofactor, and are listed separately as apo-structures without the metal-binding pattern (Figure 1). This provides the user with an innovative structural perspective on apo-structures, enabling the systematic analysis of the structural impact of metal binding and providing hints on the possible evolution of the MFS itself. In implementing this protocol, we realized that distinct groups of equistructural sites sometimes have some or even all metal ligands in common in the protein sequence alignment. Different groups of equistructural MFSs are created when structures with the same metal-binding protein domain have MFSs in different relative positions within the structural alignment of all the chains (4). However, the present sequence alignments reveal that this can happen while maintaining some metal ligands from the protein unchanged, i.e. the spatial shift of the MFS can be a result of structural rearrangements or flexibility rather than of evolutionary changes altering the sequence. We thus decided to dub sites that belong to different equistructural groups but share at least a protein ligand in the sequence alignment as ‘related sites’.

The MetalPDB interface

The 2018 version of MetalPDB features an additional mode of querying the database, i.e. by providing a list of PDB identifiers. The interface analyses the list to separate entries corresponding to metal-containing, apo- or not-metal-binding structures, and then allows the user to select specific MFSs from each metal-containing entry. In this way one

Site 12ca_2

Download Alignment (FASTA format)

Equivalent Site(s) Equistructural Site(s) Ligands residues Neighbouring residues
 Related Site(s)
 Apo-structures with the metal-binding pattern Apo-structures without the metal-binding pattern
 H-Bonded Residues M = Main Chain S = Side Chain
Only H-bonds involving metal-binding ligands are displayed

| Code | Sequence(s) [click on sequence area then use arrow keys to slide] | Metal(s) in site |
|--------------------------------|---|------------------|
| 12ca_A | -----WG Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| 1bnm_1 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| 1am6_2 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| 1avn_2 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| 1bn3_1 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| 1bn1_2 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| 1a42_1 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| 1bn4_2 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| 1azm_1 | -----PDW Y DD-KNGPEQWSKL-----YPIA-NGNNQSPVDIKTSETKH-----DTS-----LKI | ZN |
| 1bic_2 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| 1bcd_1 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN |
| Load More Equivalent Sites | | |
| 1fsr_1 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | CU |
| 1fsr_2 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | CU |
| 1fr4_1 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | CU |
| 1crm_1 | -----W Y DD-KNGPEQWSKL-----YPIA-NGNNQSPVDIKTSETKH-----DTS-----LKI | HG |
| 1can_1 | -----SHHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | HG |
| 1fqr_1 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | CO |
| 1fsq_2 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | CO |
| 1fsq_1 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | CO |
| 3koi_1 | -----W Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | CO |
| 1cah_1 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | CO |
| Load More Equistructural Sites | | |
| 5brv_1 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN IR |
| 4lp6_4 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN ZN |
| 4lp6_9 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN ZN |
| 2foy_4 | -----W Y DD-KNGPEQWSKL-----YPIA-NGNNQSPVDIKTSETKH-----DTS-----LKI | ZN CU |
| 3zp9_1 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN IR |
| 2foy_1 | -----D W YDD-KNGPEQWSKL-----YPIA-NGNNQSPVDIKTSETKH-----DTS-----LKI | ZN CU |
| 2fov_3 | -----SHHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN CU |
| 3ca2_1 | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | HG ZN HG |
| 1lug_2 | -----SHHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN HG |
| 3pyk_1 | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ZN RU |
| Load More Related Sites | | |
| 3d93_A | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ? |
| 1fsn_A | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ? |
| 1fsn_B | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ? |
| 4knm_B | -----SW Y RE-HNGPIHWKEF-----FPIA-DGDQQSPVIEIKTKEVKY-----DSS-----LRI | ? |
| 4knm_A | -----LSW Y RE-HNGPIHWKEF-----FPIA-DGDQQSPVIEIKTKEVKY-----DSS-----LRI | ? |
| 4kp8_A | -----KW T YFG-PDGENSWSKK-----YPS-CGGLLQSPIDLHSDILQY-----DAS-----LTI | ? |
| 1zsa_A | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ? |
| 2cbe_A | -----HHW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ? |
| 4knj_A | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ? |
| 1fqj_A | -----HW Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | ? |
| Load More Apo Sites | | |
| 1cnb_A | -----W Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | None |
| 1hva_A | -----W Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | None |
| 1cvf_A | -----W Y GK-HNGPEHWHKD-----FPIA-KGERQSPVDIDHTAKY-----DPS-----LKI | None |

Figure 1. The Sequence tab for entry 12ca.2 (16). The new Sequence tab displays the sequence alignment of all proteins in the same superfamily. The proteins are grouped based on the relationship of their MFSS to the query MFS (equivalent or equistructural sites), whereas for proteins lacking any metal in the site the grouping is based on the conservation of the metal ligands (apo-structures with or without the metal-binding pattern). The metal ligands have a yellow background; the residues belonging to the MFS have a cyan background. H-bonded residues are highlighted in red.

Summary Coordination Sphere Equivalent Sites Equistructural Sites

Information on the PDB Chain(s) containing the Site

| PDB Chain | Molecule Name | Organism Name | UniProt Id | EC Number |
|-----------|---------------|-------------------------------|------------|-----------|
| 1joi_A | Azurin | Pseudomonas fluorescens bv. A | | |

Based on literature: pc is a small blue copper protein belonging to a family of monomeric cupredoxins, which contains a type-1 copper center with four coordination ligands

Information on the Site

| Site Name | Nuclearity | Location | Physiological Relevance | Site Image |
|-----------|-------------|----------------|-------------------------|------------|
| 1joi_1 | Mononuclear | Within a Chain | Physiological | |

Information on the Function(s) of the Site

| Function | Function Details | Metal(s) performing the functions | Reliability |
|-------------------|------------------|-----------------------------------|---|
| Electron Transfer | - | Cu_129(A)_CU | literature (the paper does not refer to a specific structure) |

single t1 cu centers are encountered in small proteins such as plastocyanin and azurin that shuttle electrons between various donors and acceptors

Figure 2. The summary page of 1joi.1 (17). The new *Information on the site* table reports the *Physiological relevance* of the site (highlighted with a red circle). Each physiological site has an associated function that is detailed in a further new table (*Information on the function(s) of the site*, also highlighted with a red circle) When an annotation is based on the literature, it is possible to display the sentence of the article that supports the functional annotation by hovering the mouse on the book icon. The book icon links to the article entry on PubMed.

can select, for example, only physiologically relevant sites or only a given site in a family of metalloproteins containing multiple MFSs. After completing the selection, it is possible to create a personalized report on the properties of all the selected MFSs. For each MFS, the report can include features of the site (CATH (13)/SCOP (14)/Pfam (11)) domain containing the site, number of ligands, EC number for metalloenzymes), of the metal (coordination geometry, coordination number, metal-binding pattern) and of the ligands (donor atoms, metal-to-donor distances). The report can be downloaded as a csv file.

To facilitate the analysis of the entire MetalPDB contents, we implemented two new options for large data download: an ftp interface providing access to all the MFSs, grouped by the bound metal (each group is available as a compressed tar file), and a link to a flat file version of the database.

MetalPDB returns results on a per-MFS basis, i.e. the result page shows the information contained in the database for an individual MFS. The information is distributed under different tabs within the page. Below we report the modified or the new tabs of the current version of MetalPDB:

- **Summary tab:** the table ‘Information on the Site’ now reports, when available, the physiological relevance of the site. When a site is ‘Physiological’, it also has an associated function, which is reported in the ‘Function Details’ table below. By hovering the mouse over the book icons, a sentence of the article supporting the annotation appears in a box (Figure 2). For Modified Physiological MFSs, we additionally provide a description of the changes with respect to the physiological site in a separate ‘Site Modification’ tab (see below).
- **Coordination Sphere tab:** each ligand is now associated with a relative solvent accessibility and with a secondary structure element.
- **Sequence tab:** this tab was not present in the previous version. It displays the sequence alignment of all the mem-

bers of the protein superfamily of the query MFS (Figure 1). These include: (i) sequences harbouring equivalent sites (white) and (ii) sequences harbouring equistructural sites (blue); (iii) sequences with ‘related sites’ (light green), (iv) sequences of apo-structures that conserve all the metal-binding residues of the query MFS (dark green) and (v) sequences of apo-structures which have lost at least one metal-binding ligand with respect to the query MFS (grey). A structural superposition of the putative sites in the apo-structures with the metal-binding pattern to the query MFS can be downloaded. It is also possible to download the alignment of all the sequences. The user can move along the alignment by shifting it right or left to inspect its different regions, or by showing more or less sequences. A color code highlights the protein residues forming the MFS as well as the position and interactions of the metal-binding residues.

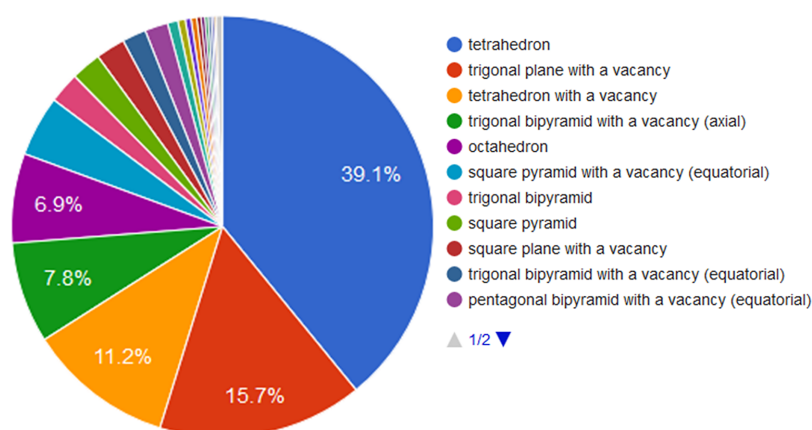
- **Site Modification tab:** this new tab is present when the query MFS is annotated as a ‘Modified Physiological Site’. It details the modifications of the query MFS with respect to the physiologically relevant site(s).

Statistics pages

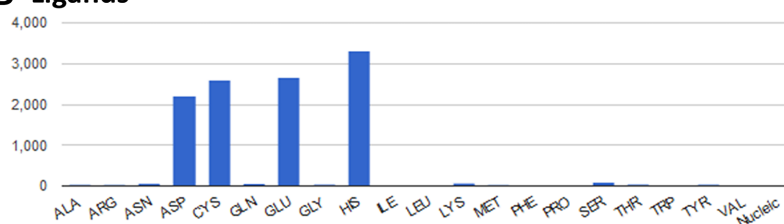
We have extended the previous version of MetalPDB to provide extensive statistics on its contents, providing both structural and functional information. Several different pages, which can be accessed via the *Statistics* drop-down menu of the navigation bar, are available:

- **Summary,** which lists the number of sites, atoms and PDB structures contained in MetalPDB on a per-metal basis;
- **Metals in PDB,** which provides an overview of the fractional occurrence of metal-binding structures in different repositories or for different macromolecule types, a histogram of the number of MFS with a given nuclearity (number of metal ions per site), and a statistics of the most common coordination geometries observed in MetalPDB;

A Geometries distribution



B Ligands



C Coordination distances



Figure 3. Example of statistics for zinc coordination spheres in the PDB. The information is accessible from the ‘Per metal’ statistics menu. (A) Pie chart displaying the coordination geometries of zinc sites; (B) histogram reporting the occurrence of residues in the first coordination sphere of zinc ions; (C) distances between zinc ions and different donor atoms.

- *Per Geometry*, which provides statistics per each coordination geometry defined in FindGeo (15). By clicking on the geometry of interest, the user enters a page describing which metals were assigned that geometry in MetalPDB and how many different metal-binding patterns adopted that geometry for each metal;
- *Metal domains*, which provides an overview of the fractional occurrence of metal-binding domains in domain databases, in total and on a per-metal basis. For the SCOP and CATH databases, the per-metal statistics is further subdivided by domain class;
- *Per metal*, which enables two different kinds of analyses: coordination geometries or metal ligand distributions. In this page, the users selects one specific metal ion for which s/he wants to obtain statistics; then the desired analysis is selected by pressing a button at the bottom of the page. In the Geometries section, MetalPDB reports the occurrence of all regular coordination geometries (Figure 3A), the distribution of aminoacidic ligands for each geometry, and the number of different metal-binding patterns observed for the selected metal as a function of the coordination geometry. In the Ligands section, MetalPDB reports the statistics on the presence of aminoacidic or nucleic ligands in the coordination sphere of the selected metal (Figure 3B), the distribution of metal to donor atom distances (Figure 3C), and data on non-bonded interactions between aminoacidic ligands and other aminoacids of the protein (so-called second-sphere interactions);
- *Metals in enzymes*, which reports on the presence of metal sites in enzymes as well as on the occurrence of the different metal ions among the six EC classes and on the distribution of the six EC classes among metalloproteins on a per-metal basis (note that we include both catalytic and non-catalytic MFSs for any protein that has a EC number associated);
- *Metal substitutions in sites*, which reports on the distribution of the different metal ions replacing any given metal in all the sites (for example showing that the most common replacement for Ni is Zn, whereas for Mg it is Ca);

this statistics is derived from the comparison of equistructural groups.

All these pages are updated every time the database content is updated to the newest PDB release. Several of the statistics listed above, in particular those involving ligands and metal-binding patterns, address only metalloproteins.

CONCLUSIONS AND PERSPECTIVES

The number of structurally characterized metal-binding sites in biological macromolecules is still experiencing a significant growth. We have coped with this growth (64% in 6 years) by reviewing and improving the protocols for the construction of MetalPDB contents. In parallel, we expanded the options available to users for interacting with MetalPDB as well as the amount and complexity of pre-computed structural and functional information displayed in the pages of MetalPDB. In the next releases of MetalPDB, we will continue to improve the functional information, also by enabling queries and statistics that target functional aspects directly. An important advancement is the functional annotation of individual MFSSs, which is only partial at present. In the future development of MetalPDB we will work on increasing the coverage of annotated MFSSs.

ACKNOWLEDGEMENTS

We gratefully acknowledge the technical help of Enrico Morelli. Maria Laura Vieri is thanked for the functional annotation of some copper-binding sites. We thank Dr Serena Lorenzini for an initial version of some of the statistics pages.

FUNDING

European Commission through the H2020 projects West-Life [675858]; EGI-ENGAGE [654142]. Funding for open access charge: European Commission [H2020].
Conflict of interest statement. None declared.

REFERENCES

- Andreini, C., Bertini, I. and Rosato, A. (2009) Metalloproteomes: a bioinformatic approach. *Acc. Chem. Res.*, **42**, 1471–1479.
- Andreini, C., Bertini, I., Cavallaro, G., Holliday, G.L. and Thornton, J.M. (2008) Metal ions in biological catalysis: from enzyme databases to general principles. *J. Biol. Inorg. Chem.*, **13**, 1205–1218.
- Pechlaner, M. and Sigel, R.K.O. (2012) Characterization of metal ion-nucleic acid interactions in solution. *Met. Ions. Life Sci.*, **10**, 1–42.
- Andreini, C., Cavallaro, G., Lorenzini, S. and Rosato, A. (2013) MetalPDB: a database of metal sites in biological macromolecular structures. *Nucleic Acids Res.*, **41**, D312–D319.
- Rose, P.W., Prlic, A., Altunkaya, A., Bi, C., Bradley, A.R., Christie, C.H., Costanzo, L.D., Duarte, J.M., Dutta, S., Feng, Z. *et al.* (2017) The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.*, **45**, D271–D281.
- Andreini, C., Bertini, I. and Cavallaro, G. (2011) Minimal functional sites allow a classification of zinc sites in proteins. *Plos ONE*, **10**, e26325.
- Rosato, A., Valasatava, Y. and Andreini, C. (2016) Minimal functional sites in metalloproteins and their usage in structural bioinformatics. *Int. J. Mol. Sci.*, **17**, 671.
- Valasatava, Y., Andreini, C. and Rosato, A. (2015) Hidden relationship between metalloproteins unveiled by structural comparison of their metal sites. *Sci. Rep.*, **5**, 9486.
- Valasatava, Y., Rosato, A., Cavallaro, G. and Andreini, C. (2014) MetalS³, a database-mining tool for the identification of structurally similar metal sites. *J. Biol. Inorg. Chem.*, **19**, 937–945.
- Hutchinson, E.G. and Thornton, J.M. (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.*, **5**, 212–220.
- Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
- Maret, W. (2010) Metalloproteomics, metalloproteomes, and the annotation of metalloproteins. *Metallomics*, **2**, 117–125.
- Dawson, N.L., Lewis, T.E., Das, S., Lees, J.G., Lee, D., Ashford, P., Orengo, C.A. and Sillitoe, I. (2017) CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.*, **45**, D289–D295.
- Chandonia, J.M., Fox, N.K. and Brenner, S.E. (2017) SCOPe: manual curation and artifact removal in the structural classification of proteins - extended database. *J. Mol. Biol.*, **429**, 348–355.
- Andreini, C., Cavallaro, G. and Lorenzini, S. (2012) FindGeo: a tool for determining metal coordination geometry. *Bioinformatics*, **28**, 1658–1660.
- Nair, S.K., Calderone, T.L., Christianson, D.W. and Fierke, C.A. (1991) Altering the mouth of a hydrophobic pocket. Structure and kinetics of human carbonic anhydrase II mutants at residue Val-121. *J. Biol. Chem.*, **266**, 17320–17325.
- Zhu, D.W., Dahms, T., Willis, K., Szabo, A.G. and Lee, X. (1994) Crystallization and preliminary crystallographic studies of the crystals of the azurin *Pseudomonas fluorescens*. *Arch. Biochem. Biophys.*, **308**, 469–470.
- Frausto da Silva, J.J.R. and Williams, R.J.P. (2001) *The Biological Chemistry of the Elements: the Inorganic Chemistry of Life*. Oxford University Press, NY.