J Biol Inorg Chem (2014) 19:937–945
DOI 10.1007/s00775-014-1128-3

ORIGINAL PAPER

MetalS^3, a database-mining tool for the identification of structurally similar metal sites

Yana Valasatava · Antonio Rosato · Gabriele Cavallaro · Claudia Andreini

Received: 23 January 2014 / Accepted: 19 March 2014 / Published online: 4 April 2014

Abstract We have developed a database search tool to identify metal sites having structural similarity to a query metal site structure within the MetalPDB database of minimal functional sites (MFSs) contained in metal-binding biological macromolecules. MFSs describe the local environment around the metal(s) independently of the larger context of the macromolecular structure. Such a local environment has a determinant role in tuning the chemical reactivity of the metal, ultimately contributing to the functional properties of the whole system. The database search tool, which we called MetalS^3 (Metal Sites Similarity Search), can be accessed through a Web interface at http://metalweb.cerm.unifi.it/tools/metals3/. MetalS^3 uses a suitably adapted version of an algorithm that we previously developed to systematically compare the structure of the query metal site with each MFS in MetalPDB. For each MFS, the best superposition is kept. All these superpositions are then ranked according to the MetalS^3 scoring function and are presented to the user in tabular form. The user can interact with the output Web page to visualize the structural alignment or the sequence alignment derived from it. Options to filter the results are available. Test calculations show that the MetalS^3 output correlates well with expectations from protein homology considerations. Furthermore, we describe some usage scenarios that highlight the usefulness of MetalS^3 to obtain mechanistic and functional hints regardless of homology.

Keywords Metalloprotein · Metalloenzyme · Bioinorganic chemistry · Structural biology · Zinc · Iron

Electronic supplementary material The online version of this article (doi:10.1007/s00775-014-1128-3) contains supplementary material, which is available to authorized users.

Y. Valasatava · A. Rosato · G. Cavallaro · C. Andreini (✉)
Magnetic Resonance Center (CERM), University of Florence,
Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy
e-mail: andreini@cerm.unifi.it

A. Rosato · C. Andreini
Department of Chemistry, University of Florence,
Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Introduction

Bioinorganic or biological inorganic chemistry is the discipline dealing with the interaction between inorganic substances and molecules of biological interest [1–3]. It is a wide scientific field that addresses the role, uptake, and fate of elements essential for life, the response of living organisms to toxic inorganic substances, the function of metal-based drugs, the synthetic production of functional models, and so on. The interaction between metal ions or metal-containing cofactors and biological macromolecules can be studied in atomic detail through 3D structural studies, thus providing a connection between bioinorganic chemistry and structural biology [4].

Metal ions are bound to biological macromolecules via coordination bonds. The bonds are made by so-called donor atoms that can belong to either the polymer (protein or nucleic acid) backbone or side chains/bases. Additional donor atoms may belong to nonmacromolecular ligands, such as oligopeptides, small organic molecules, anions, and water molecules. The ensemble comprising a metal ion (or cluster of metal ions) together with its donor atoms defines the metal-binding site. Metal-binding sites are occasionally extended to include all of the atoms in the donor amino acid or nucleotide. Databases reporting on the geometric properties of metal-binding sites in proteins [5] or nucleic acids [6] are available. They are derived from the coordinate files deposited in the Protein Data Bank (PDB) [7]. Metal-binding sites have been shown to be useful for the bioinformatic analysis of metal-binding proteins (metalloproteins) and, in particular, for the prediction of metalloproteins from genome sequences [8–10]. We have described how the inclusion of the surroundings of the metal-binding site in structure-based analyses strengthens the relationship of the sites with functional properties [11, 12]. This larger ensemble can be thought of as the minimal environment determining metal function, which in previous work we dubbed the “minimal functional site” (MFS). In practice, we defined an MFS in a metal–macromolecule adduct as the ensemble of atoms containing the metal ion or cofactor, all its ligands, and any other atom belonging to a chemical species within 5 Å from a ligand [11, 13] (Fig. S1). The MFS describes the local 3D environment around the cofactor, independently of the larger context of the protein fold in which it is embedded. The usefulness of the MFS concept outlined above has its chemicophysical foundation in the fact that the local environment of the metal has a determinant role in tuning its properties and thus its chemical reactivity [14, 15]. Instead, the macromolecular matrix is instrumental in determining, e.g., substrate selection [16] or partner recognition [17].

To make MFS analyses available to the scientific community, we developed two different resources: (1) MetalPDB [18], a database of all MFSs contained in the PDB, which is automatically updated, providing access to structural and functional information, including atomic coordinates, for each MFS in any metal-binding macromolecule of known 3D structure; (2) MetalS^2 (Metal Sites Superposition) [12], a tool for the metal-centered superposition of MFS pairs, applicable to structures already in the PDB or to structural files belonging to the user. In the present work, we present MetalS^3 (Metal Sites Similarity Search), a new tool that bridges the two aforementioned resources by allowing researchers to input the coordinates of one MFS and perform a systematic search of the entire MetalPDB database to identify structurally similar sites, regardless of overall fold similarity or protein homology. MetalS^3 is based on the same conceptual approach of MetalS^2, with some minor modifications. However, its implementation as a tool for a database search makes possible a completely different usage scenario, with a main focus on knowledge discovery through the unbiased exploration of the structural space of metal sites.

Methods

The MetalS^3 algorithm

MetalS^2 performs the superposition of two MFSs by performing the following steps [12]: (1) computing and overlapping the geometric centers of the metal atoms contained in each MFS; (2) systematically computing a set of initial configurations (poses), in each of which the geometric centers of the metals and two different pairs of donor atoms from the two sites are used to superimpose the MFSs (Fig. S2); (3) ranking all the poses on the basis of a specifically designed scoring function; (4) optimizing a subgroup of the poses (by default, those in the best 40 % of the entire score range) by allowing the geometric centers and the ligands to be displaced with respect to one another. The MetalS^2 score consists of three terms that account, respectively, for the biochemical similarity of the amino acids put in correspondence (sequence similarity term), the ratio between the total length of the sequence alignment and the length of the smallest site (i.e., the fractional coverage of the smaller site) (fractional coverage term), and the number and length of consecutive sequence segments in the superposition (fragmentation term). Amino acid correspondences are established on the basis of Cα–Cα and Cβ–Cβ distances. In step 4 of the procedure, the root mean square deviation (RMSD) of the coordinates in the superposition is optimized and amino acid correspondences are reevaluated. Note that atoms from exogenous (i.e., nonprotein, non-nucleic acid) ligands are not included in the computation neither of the RMSD nor of the score. The reason for this is that, especially in the context of MetalS^3, we want to identify and quantify similarities among the macromolecular components of the MFSs. Exogenous ligands contribute to the definition of each MFS geometry as well as to the calculation of the set of initial poses, which is based purely on geometrical considerations. Thereafter, and especially for the purpose of scoring the solutions, such ligands are no longer taken into account. This makes the final ranking dependent only on the similarities between the macromolecular structures, as desired, and avoids possible biases due to common arrangements of the ligands around the metal ion, e.g., as for chelators such as hydroxamic acid derivatives in zinc enzymes, which maintain a fixed geometry in most or all structures.

For the present work, we implemented a new Web interface, MetalS^3, that allows a user to upload a metal-containing macromolecular structure (or select it from the MetalPDB database) in PDB format, select any MFS (automatically detected) contained in it, and systematically compare it against all MFSs in MetalPDB using the MetalS^2 algorithm. A list of hits is returned by MetalS^3, sorted by the corresponding score. We introduced some minor modifications to the MetalS^2 procedure and scoring function described in the previous paragraph. In MetalS^3, the fractional coverage term always refers to the input (query) MFS rather than to the smallest site of the pair being superposed. In addition, the optimization step is iterated as long as the superposition score keeps decreasing.

To reduce the computational effort, we imposed some limitations on the difference in the number of donor atoms between the query MFS and any MFS from MetalPDB, which are recapitulated by the following formula:

$$\begin{cases} a = \frac{N}{4},\ \text{if},\ \frac{N}{4} > 2,\ \text{else } 2 \\ b = 4N,\ \text{if } 4N < 20,\ \text{else } 20 \end{cases} \qquad (1)$$

where a and b are, respectively, the smallest and largest number of donor atoms that an MFS from the database can have for it to be included in the search set and N is the number of donor atoms in the query. In practice, any MFS in MetalPDB with a number of donor atoms outside the [a; b] range is excluded from the search. For example, a query MFS with four donor atoms will be compared only with MFSs from MetalPDB having between two and 16 donors. We believe that the application of the above-mentioned restriction does not reduce the usefulness of the results, as it seems reasonable to assume that any structural similarity between MFSs with a disparity in the number of donor atoms beyond the limits imposed by Eq. 1 does not have functional relevance.

Implementation of MetalS^3

All back-end scripts are implemented in Python 2.6.6 (http://www.python.org/) on a Linux platform. The front end was implemented using Mako, a template library written in Python included by default with the Pylons Web application framework, JavaScript, and Cascading Style Sheets. By using the Python language, we could also exploit the following resources: SciPy 0.7.2, a library of scientific and numerical routines; NumPy 1.4.1, a language extension that adds support for large and fast, multidimensional arrays and matrices; and p3d [19], a Python module for structural bioinformatics. The MetalS^3 server is currently hosted on a 24-CPU (AMD Opteron$^{\text{TM}}$ 6234) server.

The MetalS^3 Web interface

The Web interface of MetalS^3 allows the user to run queries against all representative MFSs of the equistructural MFS clusters defined in MetalPDB. Each of these MFSs represents a group of sites that are found in proteins with the same fold, as judged from sequence similarity and Pfam [20] domain assignments, and occur at the same spatial location within that fold. For example, a single representative MFS represents all the sites of rubredoxins from various organisms and with different metalation. MetalPDB currently contains 17,936 clusters of equistructural MFSs. As mentioned previously, the dataset of representative MFSs against which the query is actually compared is the subgroup of all 17,936 sites that satisfies Eq. 1. Thus, the size and the characteristics of the subgroup depend on the input query MFS, and particularly on the number of donor atoms it contains. In turn, this influences the overall calculation time.

After a calculation is finished, the user is presented with a list of hits having structural similarity to the query, ordered by the total MetalS^3 score (the list can be resorted according to different parameters, such as individual score components). It is then possible to select a specific hit, i.e., a specific representative MFS, and run a refinement calculation in which the query is compared with each individual site in the corresponding equistructural MFS cluster. A link to the results of the search is e-mailed to the user at the end of each of these two stages.

Results

A brief description of the input and output interfaces of MetalS^3 is available as electronic supplementary material (text and Figs. S3 and S4). We conducted various experiments to assess our implementation of MetalS^3 with respect to its capability to identify relevant hits within the MetalPDB database as well as with respect to the typical times required to obtain the results of a calculation.

Because MetalS^3 searches are initially performed only against representative MFSs and not the entire content of the MetalPDB database, it is important to assess whether this approach consistently returns relevant functional information. To do this, we used an example dataset of 100 different MFSs randomly picked from deposited PDB structures (Table S1). These examples, which differed in metal content as well as coordination number and geometry, were used as input queries to MetalS^3. Crucially, the examples were selected in order to avoid including any representative MFS as defined in MetalPDB. In this way, we could straightforwardly classify the output of MetalS^3 depending on whether the best-scoring hit corresponded to the representative of the cluster to which each query MFS was known to belong. In fact, even though the clustering procedure implemented in MetalPDB does not directly compare the structure of the different MFSs assigned to a cluster, in the large majority of cases the MFSs within a cluster should be similar to each other because the proteins in the cluster can be assumed to be homologous. In 75 % of cases, this was indeed observed. Notably, if we optimize all the poses, instead of a well-scoring subgroup, the above-mentioned result increases only to 76 %. We then analyzed manually the 25 cases for which the best hit identified by MetalS^3 was not the representative MFS of the cluster to which the query belongs in the MetalPDB database. For 20 of them we observed that the result obtained depended on the clustering within MetalPDB being incomplete, i.e., failing to group together MFSs that indeed are bound to homologous proteins. In turn, this is due to missing Pfam assignments or, often, to a given protein superfamily being mapped to multiple Pfam domains [21]. Instead, in five cases MetalS^3 identified a structural similarity between a pair of MFSs (the query and the returned hit) that was higher than that between the query and the representative MFS of its equistructural cluster in MetalPDB. These are cases where either highly similar MFSs are embedded in different folds (three) or the MFS representative does not adequately represent the cluster (two). The representative MFS of a cluster is chosen solely on the basis of the resolution (i.e., quality) of the corresponding 3D structure [18]. Consequently, the representative MFS cannot be regarded as a sort of "average" MFS, and there is no specific property regarding its structural similarity to the other MFSs in the cluster. A third option is that the assignment of the query MFS to the MetalPDB cluster, which was performed automatically, did not reflect the large structural variability of the MFSs within the cluster. This was not observed here. An additional consideration is that, because of the way the score is constructed, smaller query MFSs tend to be less discriminative and therefore may more easily provide high-scoring hits also to MFSs not closely related (but still structurally similar).

If one looks at the five best scoring hits, then in only ten cases from the 100 examples run was the MFS representative of the cluster of the query site not included. As already mentioned, in two instances we observed that the specific representative MFS did not reflect the "consensus" coordination geometry of its cluster. However, in most cases, the reason for the observed behavior was an incomplete clustering of the structures, in turn typically resulting from problems in the mapping of Pfam domains. This caused structures highly similar to the query not to be included in the same equistructural cluster.

The calculation times are dependent on the number of donor atoms (N) in the query MFS, as the number of poses that need to computed and compared scales with $N(N-1)$ [12] (Fig. 1). For a given number of atoms, calculations are faster the higher the number of donor atoms from exogenous ligands (such as small metal-binding molecules or ions) because these are not considered in amino acid matching and RMSD computations (see "Methods"). The calculation times are less than 2 h for sites with up to four protein donor atoms, whereas, owing to the parabolic increase of calculation times, they are as long as 10 h for sites with nine donor atoms (if all are from protein ligands) and within 24 h for multinuclear sites with 12 donor atoms from the protein moiety. Of all representative MFS sites collected in MetalPDB, 95.1 % have nine donor atoms or fewer. Under the assumption that MetalPDB adequately describes the diversity of MFSs occurring in nature, the data given above may suggest that users will most often submit queries that can be dealt with in 10 h or less. In any case, results are always sent to the users via e-mail, as even the simplest calculations require at least a few minutes.

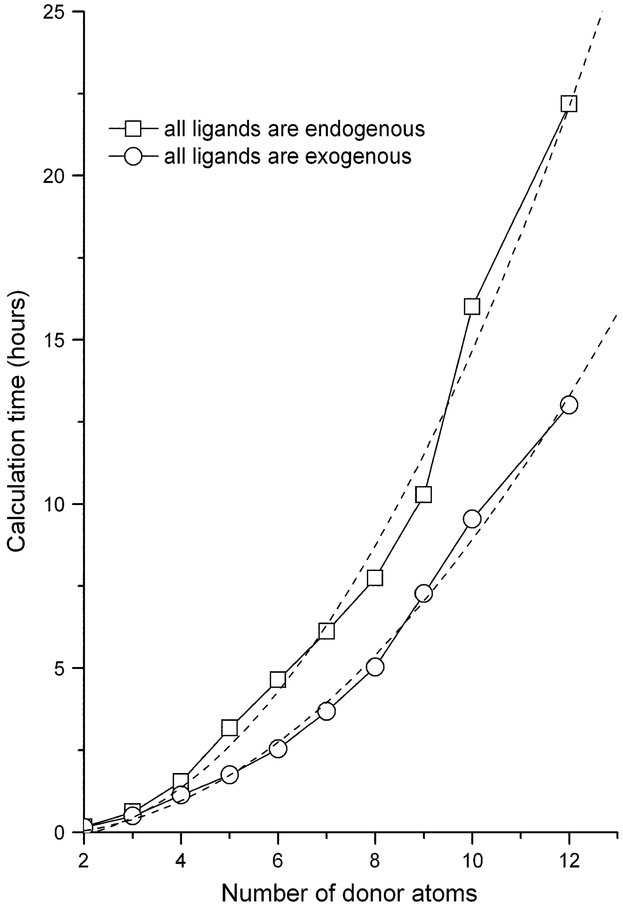


Fig. 1 Calculation times for MetalS^3 queries as a function of the number and type of donor atoms. *Dashed lines* are the best fit to a second-order polynomial

Discussion

MetalS^3 is a Web interface that allows the user to systematically compare an MFS of interest (query) with the contents of the MetalPDB database [18], i.e., with an ensemble representing the diversity of known MFSs. This is achieved through a suitably modified implementation of the MetalS^2 algorithm [12]. Typically, the hits returned for a query will comprise sites that are contained within a protein homologous to the protein containing the query MFS as well as sites from unrelated proteins. The presence in the output page of one or the other type of hit, as well as their relative abundance, will depend on the cutoffs defined to exclude hits from the visualization (Fig. S3). The cutoffs can be adjusted also after the calculation has finished, through the “Filter Results” button on the output page (Fig. S4). Increasing the cutoff values will result in a longer list of hits being displayed.

Our test calculations show that the top position in the list of the hits is highly likely to be occupied by an MFS contained within a homolog of the query protein; when the top five hits are considered, this is verified for as many as 90 % of the examples that we run. According to the definition of the MetalPDB database, on which MetalS^3 builds, this situation corresponds to the query and hit MFSs belonging to the same equistructural cluster. For MFSs in MetalPDB to be clustered, it is actually requested that the sites occupy the same position within the fold after the entire protein structure has been superimposed, and the structures of the MFSs belonging to a given cluster are not compared with one another. The approach of MetalS^3 is entirely different, as it operates only on the MFSs, disregarding the rest of the protein structure. The very good correlation between the fold-based clustering results and the MetalS^3 output points to the high similarity of the local 3D structure around the metal site being a possible indicator of metalloprotein homology. This is supported also by the fact that in 20 of the examples, MetalS^3 indicated that the clustering within MetalPDB was incomplete. Incomplete clustering typically results from the homology relationship between metalloproteins bearing structurally similar MFSs being hidden by the fact that the Pfam domain assignments we use in the definition of equistructural clusters are fine-grained and may occasionally separate a single superfamily into multiple domain definitions. To address this issue, the user can verify if the Pfam domains of interest belong to the same Pfam clan [21]. One can possibly further speculate that if the MFS properties must be defined tightly to make possible the correct protein function [i.e., to correctly define the reactivity of the metal ion(s) in the MFS], then conservation of the 3D structure of the MFS will be particularly strict among homologous proteins. Consequently, the intracluster variability of the MFS structure may be informative on the requirements imposed by the catalysis on the MFS features or, in other words, on how the functional and mechanistic properties of the system are encoded in the structure.

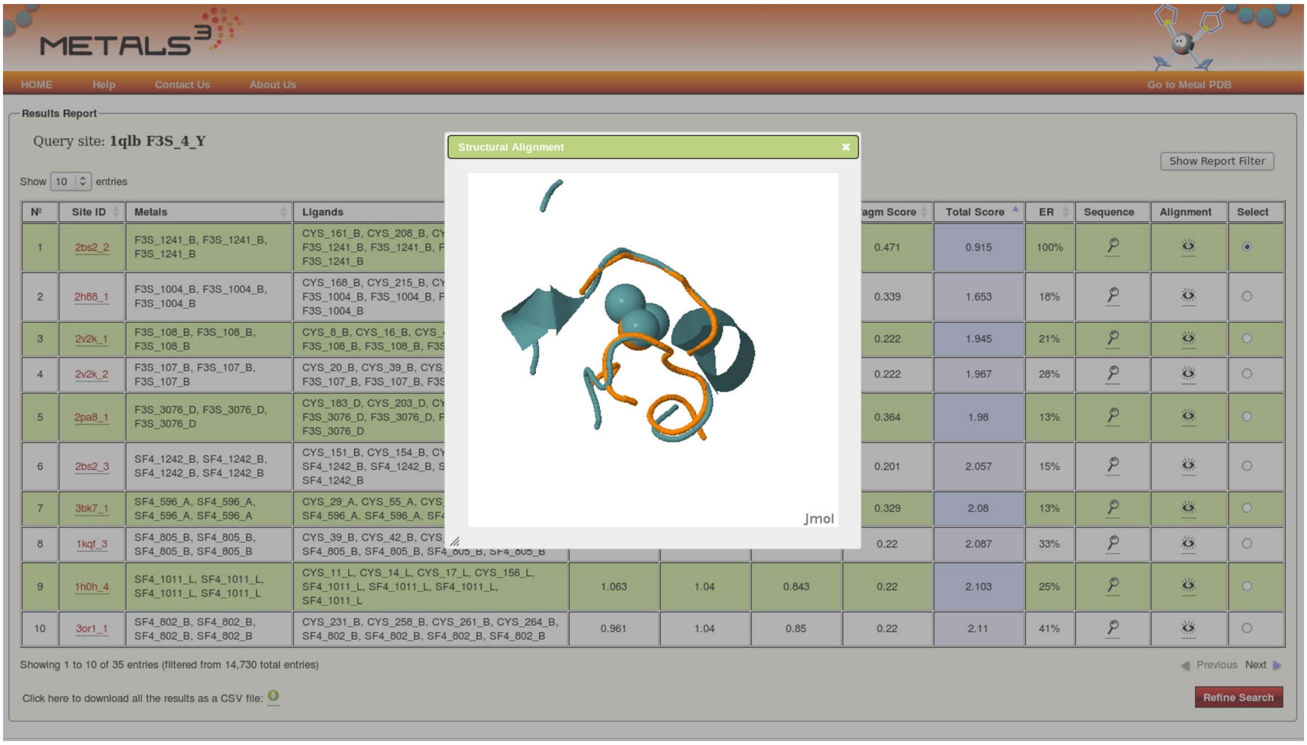
A practical application of MetalS^3 is to detect MFS structural similarities that are not associated with a homology relationship among the proteins harboring the MFSs (indicated by the MFS mapping to a shared Pfam domain or domain clan). These situations may be indicative of the occurrence of common functional properties that are endowed by the MFS itself. Such observations can provide useful hints for experimental work. In this usage scenario, the best hit returned by MetalS^3 is often uninteresting (i.e., when it is bound to a protein with the same domain composition as the protein containing the query MFS), and one should focus on worse-scoring hits. Operatively, the domain composition of a hit MFS can be immediately obtained by looking up that MFS in the MetalPDB database [18]. Below, we briefly discuss some examples not included in the 100 test dataset.

As a first example, we took one of the two equivalent Fe_3S_4 clusters in the PDB structure of fumarate reductase from *Wolinella succinogenes* (PDB ID 1QLB [22]), which is identified as site 1qlb_4 in MetalPDB (hereafter, we will use the PDB code in lowercase letters followed by an underscore and a number to indicate a specific MFS within the MetalPDB database, whereas we will use the PDB code in uppercase letters to indicate the PDB entry). This site is located with a ferredoxin-type domain, and it is likely to be part of the electron transfer pathway. MetalS^3 returns as the fifth hit, with a total score of 1.98, a site harboring an Fe_3S_4 cluster in the D subunit of the structure of the DNA-directed RNA polymerase from *Sulfolobus solfataricus* P2 (PDB ID 2PA8 [23]). Despite a sequence identity between these two MFSs of only 13 % over 15 amino acids, the superposition is good (RMSD 0.799 Å) (Fig. 2).

The latter cluster, which is possibly an Fe_4S_4 cluster in vivo, is found in the corresponding subunits of the polymerases from various species of Archaea and Eukarya, but not of Bacteria [24]. The domain containing the MFS within subunit D is not present in all archaeal RNA polymerases, but it is actually characteristics of a specific evolutionary lineage of Archaea. Here we observed that the binding mode of the Fe_3S_4 cluster within subunit D of *S. solfataricus* P2 polymerase actually bears some similarity to an unrelated episilonproteobacterial system.

A second example is provided by the MFS containing the magnesium(II) ion identified as residue 9,018 (MetalPDB entry 1g0u_1) within the structure of the core particle of the yeast proteasome (PDB ID 1G0U [25]). This MFS is interfacial, as it contains protein ligands from subunits I and Y. MetalS^3 returns hits also to sites containing metal ions other than magnesium. One of these is the MFS defined around the calcium(II) ion identified as residue 501 in the structure of human calcium and integrin binding protein 1 (PDB ID 1Y1A [26]), with a total score of 2.427 and, in particular, a sequence identity of 0 % (Fig. 3). This MFS is located within an EF-hand motif. Such a structural similarity would be extremely hard to identify by any other method, especially a sequence-based method. Magnesium(II) and calcium(II) are known to compete for binding in EF-hand sites [27]. The similarity between the two MFSs may thus underlie commonalities in the atomic mechanism by which the metal affinity is tuned.

3ZFJ is a recently solved NMR structure of a PhtD domain from *Streptococcus pneumoniae* that binds a single zinc(II) ion [28]. At the time of writing, it is not yet



Query site: **1qlb F3S_4_Y**

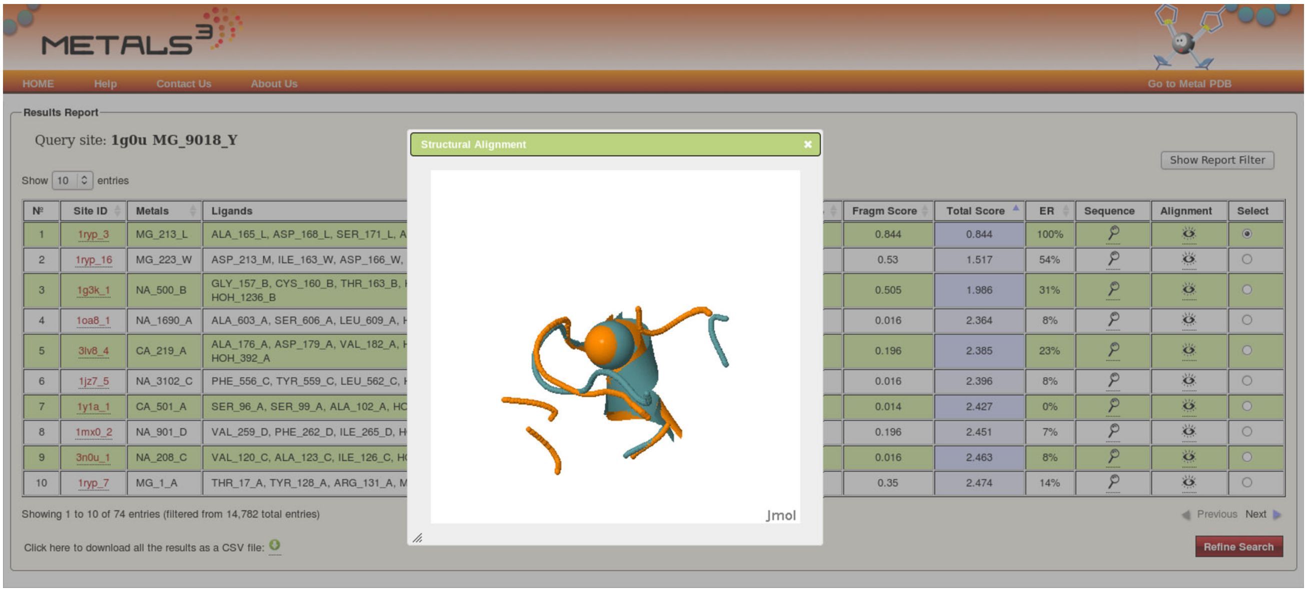
Show 10 entries

№	Site ID	Metals	Ligands				agm Score	Total Score	ER	Sequence	Alignment	Select
1	2bs2_2	F3S_1241_B, F3S_1241_B, F3S_1241_B	CYS_161_B, CYS_208_B, CY F3S_1241_B, F3S_1241_B, F F3S_1241_B				0.471	0.915	100%	🔍	👁	●
2	2h88_1	F3S_1004_B, F3S_1004_B, F3S_1004_B	CYS_168_B, CYS_215_B, CY F3S_1004_B, F3S_1004_B, F F3S_1004_B				0.339	1.653	18%	🔍	👁	○
3	2v2k_1	F3S_108_B, F3S_108_B, F3S_108_B	CYS_8_B, CYS_16_B, CYS_ F3S_108_B, F3S_108_B, F3S				0.222	1.945	21%	🔍	👁	○
4	2v2k_2	F3S_107_B, F3S_107_B, F3S_107_B	CYS_20_B, CYS_39_B, CYS F3S_107_B, F3S_107_B, F3S				0.222	1.967	28%	🔍	👁	○
5	2pa8_1	F3S_3076_D, F3S_3076_D, F3S_3076_D	CYS_183_D, CYS_203_D, CY F3S_3076_D, F3S_3076_D, F F3S_3076_D				0.364	1.98	13%	🔍	👁	○
6	2bs2_3	SF4_1242_B, SF4_1242_B, SF4_1242_B, SF4_1242_B	CYS_151_B, CYS_154_B, CY SF4_1242_B, SF4_1242_B, S SF4_1242_B				0.201	2.057	15%	🔍	👁	○
7	3bk7_1	SF4_596_A, SF4_596_A, SF4_596_A, SF4_596_A	CYS_29_A, CYS_55_A, CYS SF4_596_A, SF4_596_A, SF4				0.329	2.08	13%	🔍	👁	○
8	1kqf_3	SF4_805_B, SF4_805_B, SF4_805_B, SF4_805_B	CYS_39_B, CYS_42_B, CYS SF4_805_B, SF4_805_B, SF4_805_B, SF4_805_B				0.22	2.087	33%	🔍	👁	○
9	1h0h_4	SF4_1011_L, SF4_1011_L, SF4_1011_L, SF4_1011_L	CYS_11_L, CYS_14_L, CYS_17_L, CYS_156_L, SF4_1011_L, SF4_1011_L, SF4_1011_L, SF4_1011_L	1.063	1.04	0.843	0.22	2.103	25%	🔍	👁	○
10	3or1_1	SF4_802_B, SF4_802_B, SF4_802_B, SF4_802_B	CYS_231_B, CYS_258_B, CYS_261_B, CYS_264_B, SF4_802_B, SF4_802_B, SF4_802_B, SF4_802_B	0.961	1.04	0.85	0.22	2.11	41%	🔍	👁	○

Showing 1 to 10 of 35 entries (filtered from 14,730 total entries)

Click here to download all the results as a CSV file:

Fig. 2 Output result page for a calculation performed using the 1qlb_4 site as the query. The *inset* shows the structural alignment to the fifth hit, 2pa8_1



Query site: **1g0u MG_9018_Y**

Show 10 entries

№	Site ID	Metals	Ligands		Fragm Score	Total Score	ER	Sequence	Alignment	Select
1	1ryp_3	MG_213_L	ALA_165_L, ASP_168_L, SER_171_L, A		0.844	0.844	100%	🔍	👁	●
2	1ryp_16	MG_223_W	ASP_213_M, ILE_163_W, ASP_166_W,		0.53	1.517	54%	🔍	👁	○
3	1g3k_1	NA_500_B	GLY_157_B, CYS_160_B, THR_163_B, I HOH_1236_B		0.505	1.986	31%	🔍	👁	○
4	1oa8_1	NA_1690_A	ALA_603_A, SER_606_A, LEU_609_A, H		0.016	2.364	8%	🔍	👁	○
5	3lv8_4	CA_219_A	ALA_176_A, ASP_179_A, VAL_182_A, H HOH_392_A		0.196	2.385	23%	🔍	👁	○
6	1jz7_5	NA_3102_C	PHE_556_C, TYR_559_C, LEU_562_C, I		0.016	2.396	8%	🔍	👁	○
7	1y1a_1	CA_501_A	SER_96_A, SER_99_A, ALA_102_A, HC		0.014	2.427	0%	🔍	👁	○
8	1mx0_2	NA_901_D	VAL_259_D, PHE_262_D, ILE_265_D, H		0.196	2.451	7%	🔍	👁	○
9	3n0u_1	NA_208_C	VAL_120_C, ALA_123_C, ILE_126_C, HC		0.016	2.463	8%	🔍	👁	○
10	1ryp_7	MG_1_A	THR_17_A, TYR_128_A, ARG_131_A, M		0.35	2.474	14%	🔍	👁	○

Showing 1 to 10 of 74 entries (filtered from 14,782 total entries)

Click here to download all the results as a CSV file:

Fig. 3 Output result page for a calculation performed using the 1g0u_1 site as the query. The *inset* shows the structural alignment to the seventh hit, 1y1a_1

included in the MetalPDB database and therefore simulates well the situation of a real user. MetalS^3 identifies the 2CS7 structure [29] as the second best hit. In fact, both proteins contain the Pfam domain "Strep_his_triad," and have 23 % sequence identity. This is a case where the next update of MetalPDB would put the two in the same equistructural cluster. The above-mentioned proteins have a role in the uptake of zinc(II), by scavenging zinc(II) ions and then providing them to the extracellular membrane-anchored AdcAII transporter at the surface of *S.*

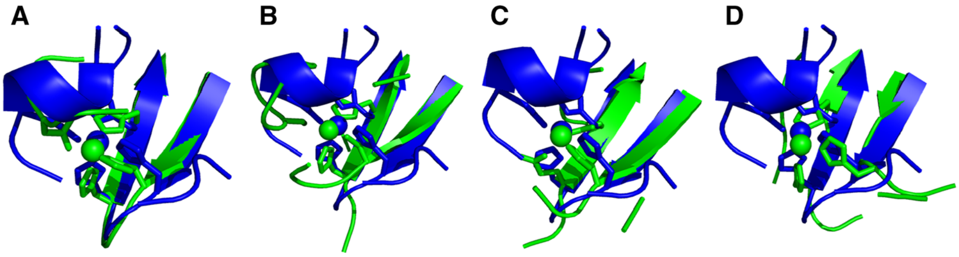


Fig. 4 Selected high-scoring zinc sites among the search results for a zinc-containing minimal functional site (MFS) from 3ZFJ. The 3ZFJ query structure is always in *blue* and in the same orientation. The superpositions to the sites **a** 2cs7_1, **b** 4hhj_1, **c** 2e26_5, and **d** 1txl_1 are displayed. Only protein ligands are shown; ZN(603) in 2E26 is additionally coordinated by two water molecules; ZN(216) in 1TXL is additionally coordinated by a water molecule

pneumoniae. The first hit is an iron-binding MFS from *Escherichia coli* galactose 1-phosphate uridylyltransferase (structure 1GUP [30]). This iron ion plays a structural role and is not essential to the enzyme activity [31]. It is useful to compare the hits returned by using either 3ZFJ or 2CS7 as queries. Among the shared top-scoring zinc proteins, one finds an MFS from structure 4HHJ [32], identified by the zinc ion with residue number 1,001. This ion has been proposed to have a structural and/or regulatory role for the activity of this RNA-dependent RNA polymerase [33]. Another common hit is from PDB entry 2E26 [34], identified by the zinc ion with residue number 603, which describes the structure of mouse reelin, a secreted glycoprotein. This ion is observed in the structures of both reelin alone and reelin in complex with apolipoprotein E receptor 2 [35], where it has fractional occupancy. Finally, MetalS^3 identifies the zinc-containing MFS of the ZinT protein (PDB ID 1TXL; S. Eswaramoorthy and S. Swaminathan, unpublished) as a further hit to the MFS in 2CS7; the MFSs of 3ZFJ and 1TXL also display good structural similarity (Fig. 4). ZinT is a periplasmic zinc transporter that facilitates metal recruitment during zinc shortage by binding zinc(II) with high affinity and subsequently transferring it to the ZnuA component of the ZnuABC membrane transporter [36, 37]. Intriguingly, in the zinc(II)-specific ABC uptake system AdcABC of *S. pneumoniae*, the AdcA protein, which does not interact with PhtD domains (see above), is a fusion between a ZnuA-like protein and a ZinT-like protein [38]. In summary, the present MetalS^3 analysis identified a minimal zinc-binding structure as being associated with reversible metal ion binding in zinc(II) transport, where different protein systems for zinc(II) uptake contain structurally similar MFSs, and in (hypothesized) zinc(II)-dependent regulation of intermolecular interactions.

An additional example is provided by the 4NAO structure, a homodimer that contains a single iron(II) ion per subunit [39], which was released in the PDB on January 15, 2014, and is not yet included in MetalPDB. This enzyme is an iron(II)/2-ketoglutarate-dependent dioxygenase that hydroxylates an *N*-(D-lysergyl-aminoacyl) lactam in the ergot fungus *Claviceps purpurea*. MetalS^3 identifies similarities to various other dioxygenases that are active against different substrates. In particular, the best hit is the iron(II) site of the 2CSG structure, an uncharacterized protein addressed by the Midwest Center for Structural Genomics, with 17 % sequence identity between the sites. Both structures feature organic ligands (2-ketoglutarate for 4NAO; succinate, which is a reaction product, and isocitrate for 2CSG) bound to the metal ion in corresponding positions (Fig. 5a). The second hit is a isopenicillin N synthase from *Emericella nidulans* (PDB ID 1ODM) [40]. This site has lower RMSD and higher sequence similarity to the query, and also features an organic ligand chelating the iron(II) ion in a manner relatively similar to that of 2-ketoglutarate of 4NAO (Fig. 5b). Notably, isopenicillin N synthase is not dependent on 2-ketoglutarate, whose functional role is performed by the tripeptide substrate [41]. The third hit contains a group of dioxygenases more closely related to 4NAO, which includes human phytanoyl-CoA dioxygenase (PhyH; PDB ID 2A1X). The article describing 4NAO provides a detailed comparison with PhyH and its homolog PhyHD1, which are actually the best results returned by a Dali [42] search based on the entire structure [39]. The 2-ketoglutarate molecules present in the 4NAO and PhyH structures chelate the metal ion in a closely similar manner (Fig. 5c). Finally, the fourth hit is a manganese(II) site in the 2-ketoglutarate-dependent dioxygenase AlkB (PDB ID 4JHT) [43] (Fig. 5d). AlkB is an iron(II)/2-ketoglutarate-dependent dioxygenase that catalyzes the oxidative demethylation of nucleic acids and histones [44]. It can bind manganese(II) in its catalytic site, yielding an inactive enzyme. Indeed, the aforementioned 4jht_1 site is the representative of a relatively large equistructural cluster in MetalPDB that contains the other structurally characterized AlkB MFSs. The cluster contains, for example, also the 3O1T structure [45], where the iron(II) ion is chelated by succinate, again in a position close to that of 2-ketoglutarate in 4NAO. The systems described in this paragraph map to three different, but

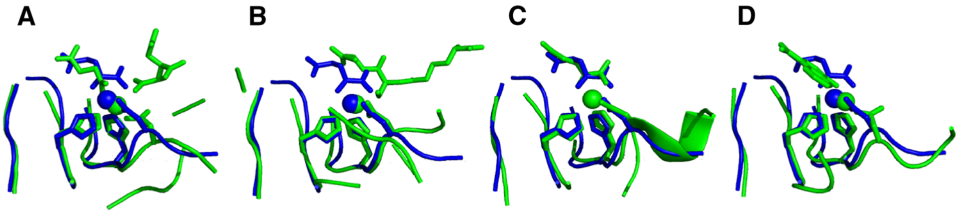


Fig. 5 The four top-scoring sites among the search results for the iron(II)-containing MFS in the A chain of 4NAO. The 4NAO query structure is always in *blue* and in the same orientation. The superpositions to the sites **a** 2csg_1, **b** 1odm_1, **c** 2a1x_1, and **d** 4jht_1 are displayed. The organic iron(II) ligands present in the various MFSs are shown as *sticks*. Water molecules are not shown

related to the same superfamily, Pfam domains: DUF1479 (2CSG), 2OG-FeII_Oxy (1ODB, 4JHT), and PhyH (4NAO, 2A1X). The results include also a case of a system where the physiological iron(II) ion was substituted in vitro. Thus, even for a large and widely studied protein superfamily such as that of iron(II)/2-ketoglutarate-dependent dioxygenases, MetalS^3 proves useful in the analysis of a newly solved structure to identify relationships across different subgroups in a manner that is independent of overall fold similarity.

Concluding remarks

MFSs in metal-binding biological macromolecules constitute a novel viewpoint for the elucidation of the mechanisms of function in these systems [11]. In this frame, we have developed the MetalPDB database [18]. MetalPDB contains a systematic analysis of all known MFSs. In particular, within the database all MFSs were grouped into so-called equistructural clusters. Each cluster contains all MFSs located at corresponding positions within the fold of homologous proteins. Recently, we developed the MetalS^2 program and Web server to perform pairwise structural superpositions of MFSs, providing a ground for the quantitative evaluation of MFS similarity [12]. MetalS^3, which is described in this work, is a Web-based tool (http://metalweb.cerm.unifi.it/tools/metals3/) that adopts the MetalS^2 algorithm to perform searches in the MetalPDB database. This is implemented as a first coarse-grained search against the ensemble of the MFSs representing MetalPDB equistructural clusters, followed by a refinement step in which the query MFS is compared with all the MFSs in a user-selected cluster. Although algorithmically very similar, MetalS^2 and MetalS^3 have somewhat different usage scenarios and make possible access to distinct information. MetalS^2 requires the user to have prior knowledge of the structures to be compared, either a pair or a group of related metalloproteins. In contrast, MetalS^3 constitutes an unbiased approach to seeking structural similarities between metal sites, independently of the user's prior knowledge. The hits returned by MetalS^3 can be a combination of relatively obvious ones (e.g., homologs of the query metalloprotein) and unexpected ones. The latter can be identified only through the present approach, whereas MetalS^2 is a tool to quantify structural similarities within groups of sites already familiar to the user.

The MetalS^3 approach may help researchers in the field of bioinorganic chemistry to assess the relationships or evaluate possible evolutionary links between different groups of metalloproteins and may help guide experimentalists' work in understanding the function of uncharacterized metalloproteins. Overall, this contributes to achieving a better comprehension of the role of metal ions in living systems.

Acknowledgments This work was supported by the Ministero Italiano dell'Università e della Ricerca through the FIRB project RBFR08WGXT and by the European Commission through the BioMedBridges project (grant no. 284209). We gratefully acknowledge the technical help of Enrico Morelli.

References

1. Frausto da Silva JJR, Williams RJP (2001) The biological chemistry of the elements: the inorganic chemistry of life. Oxford University Press, New York
2. Bertini I, Sigel A, Sigel H (2001) Handbook on metalloproteins. Dekker, New York
3. Bertini I, Gray HB, Stiefel EI, Valentine JS (2006) Biological inorganic chemistry. University Science Books, Sausalito
4. Bertini I, Rosato A (2003) Proc Natl Acad Sci USA 100:3601–3604
5. Hsin K, Sheng Y, Harding MM, Taylor P, Walkinshaw MD (2008) J Appl Crystallogr 41:963–968
6. Schnabl J, Suter P, Sigel RKO (2012) Nucleic Acids Res 40:D434–D438
7. Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, Prlic A, Quesada M, Quinn GB, Westbrook JD, Young J, Yukich B, Zardecki C, Berman HM, Bourne PE (2011) Nucleic Acids Res 39:D392–D401
8. Andreini C, Bertini I, Rosato A (2009) Acc Chem Res 42:1471–1479
9. Andreini C, Bertini I, Rosato A (2004) Bioinformatics 20:1373–1380
10. Shu N, Zhou T, Hovmoller S (2008) Bioinformatics 24:775–782
11. Andreini C, Bertini I, Cavallaro G (2011) PLoS ONE 10:e26325

12. Andreini C, Cavallaro G, Rosato A, Valasatava Y (2013) J Chem Inf Model 53:3064–3075
13. Andreini C, Bertini I, Cavallaro G, Najmanovich RJ, Thornton JM (2009) J Mol Biol 388:356–380
14. Maret W, Li Y (2009) Chem Rev 109:4682–4707
15. Choi M, Davidson VL (2011) Metallomics 3:140–151
16. Banci L, Bertini I, Calderone V, Della Malva N, Felli IC, Neri S, Pavelkova A, Rosato A (2009) Biochem J 422:37–42
17. Bertini I, Fragai M, Luchinat C, Melikian M, Venturi C (2009) Chem Eur J 15:7842–7845
18. Andreini C, Cavallaro G, Lorenzini S, Rosato A (2013) Nucleic Acids Res 41:D312–D319
19. Fufezan C, Specht M (2009) BMC Bioinform 10:258
20. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) Nucleic Acids Res 40:D290–D301
21. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A (2006) Nucleic Acids Res 34:D247–D251
22. Lancaster CR, Kroger A, Auer M, Michel H (1999) Nature 402:377–385
23. Hirata A, Klein BJ, Murakami KS (2008) Nature 451:851–854
24. Hirata A, Murakami KS (2009) Curr Opin Struct Biol 19:724–731
25. Groll M, Bajorek M, Kohler A, Moroder L, Rubin DM, Huber R, Glickman MH, Finley D (2000) Nat Struct Biol 7:1062–1067
26. Blamey CJ, Ceccarelli C, Naik UP, Bahnson BJ (2005) Protein Sci 14:1214–1221
27. Malmendal A, Linse S, Evenas J, Forsen S, Drakenberg T (1999) Biochemistry 38:11844–11850
28. Hastie KM, Kimberlin CR, Zandonatti MA, MacRae IJ, Saphire EO (2011) Proc Natl Acad Sci USA 108:2396–2401
29. Riboldi-Tunnicliffe A, Isaacs NW, Mitchell TJ (2005) FEBS Lett 579:5353–5360
30. Thoden JB, Ruzicka FJ, Frey PA, Rayment I, Holden HM (1997) Biochemistry 36:1212–1222
31. Geeganage S, Frey PA (1999) Biochemistry 38:13398–13406
32. Noble CG, Lim SP, Chen YL, Liew CW, Yap L, Lescar J, Shi PY (2013) J Virol 87:5291–5295
33. Yap TL, Xu T, Chen YL, Malet H, Egloff MP, Canard B, Vasudevan SG, Lescar J (2007) J Virol 81:4753–4765
34. Yasui N, Nogi T, Kitao T, Nakano Y, Hattori M, Takagi J (2007) Proc Natl Acad Sci USA 104:9988–9993
35. Yasui N, Nogi T, Takagi J (2010) Structure 18:320–331
36. Petrarca P, Ammendola S, Pasquali P, Battistoni A (2010) J Bacteriol 192:1553–1564
37. Ilari A, Alaleona F, Tria G, Petrarca P, Battistoni A, Zamparelli C, Verzili D, Falconi M, Chiancone E (2014) Biochim Biophys Acta 1840:535–544
38. David G, Blondeau K, Schiltz M, Penel S, Lewit-Bentley A (2003) J Biol Chem 278:43728–43735
39. Havemann J, Vogel D, Loll B, Keller U (2014) Chem Biol 21:146–155
40. Elkins JM, Rutledge PJ, Burzlaff NI, Clifton IJ, Adlington RM, Roach PL, Baldwin JE (2003) Org Biomol Chem 1:1455–1460
41. Roach PL, Clifton IJ, Fulop V, Harlos K, Barton GJ, Hajdu J, Andersson I, Schofield CJ, Baldwin JE (1995) Nature 375:700–704
42. Holm L, Sander C (1995) Trends Biochem Sci 20:478–480
43. Hopkinson RJ, Tumber A, Yapp C, Chowdhury R, Aik W, Che KH, Li XS, Kristensen JBL, King ONF, Chan MC, Yeoh KK, Choi H, Walport LJ, Thinnes CC, Bush JT, Lejeune C, Rydzik AM, Rose NR, Bagg EA, McDonough MA, Krojer TJ, Yue WW, Ng SS, Olsen L, Brennan PE, Oppermann U, Muller S, Klose RJ, Ratcliffe PJ, Schofield CJ, Kawamura A (2013) Chem Sci 4:3110–3117
44. Yu B, Edstrom WC, Benach J, Hamuro Y, Weber PC, Gibney BR, Hunt JF (2006) Nature 439:879–884
45. Yi C, Jia G, Hou G, Dai Q, Zhang W, Zheng G, Jian X, Yang CG, Cui Q, He C (2010) Nature 468:330–333