BioMetAll: Identifying Metal-Binding Sites in Proteins from Backbone Preorganization

José-Emilio Sánchez-Aparicio[a], Laura Tiessler-Sala[a], Lorea Velasco-Carneros[b],
Lorena Roldán-Martín[a], Giuseppe Sciortino[a,c] and Jean-Didier Maréchal*[a]

[a] *Insilichem, Departament de Química, Universitat Autònoma de Barcelona, Edifici C.n., 08193 Cerdanyola del Vallés, Barcelona, Spain.*

[b] *Biofisika Institute (UPV/EHU, CSIC) and Department of Biochemistry and Molecular Biology, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Barrio Sarriena s/n, 48940 Leioa, Spain.*

[c] *Dipartimento di Chimica e Farmacia, Università di Sassari, Via Vienna 2, I-07100 Sassari, Italy.*

Abstract

With a large amount of research dedicated to decoding how metallic species bind to protein, in silico methods are interesting allies for experimental procedures. To date, computational predictors mostly work by identifying the best possible sequence or structural match of the target protein with metal binding templates. These approaches are fundamentally focused on the first coordination sphere of the metal. Here, we present the BioMetAll predictor that is based on a different postulate: the formation of a potential metal-binding site is related to the geometric organization of the protein backbone. We first report the set of convenient geometric descriptors of the backbone needed for the algorithm and their parametrization from a statistical analysis. Then, the successful benchmark of BioMetAll on a set of more than 50 metal-binding X-Ray structures is presented. Because BioMetAll allows structural predictions regardless of the exact geometry of the side chains, it appears extremely valuable for systems which structures (either experimental or theoretical) are not optimal for metal binding sites. We report here its application on three different challenging cases i) the modulation of metal-binding sites during conformational transition in human serum albumin, ii) the identification of possible routes of metal migration in hemocyanins, and iii) the prediction of mutations to generate convenient metal-binding sites for de novo biocatalysts. This study shows that BioMetAll offers a versatile platform for numerous fields of research at the interface between inorganic chemistry and biology, and allows to highlight the role of the preorganization of the protein backbone as a marker for metal binding.

Introduction

Metal ions, including transition ones, play a fundamental role in living organisms. Present in about 30% of the known genome,[1] their impacts on molecular and cellular processes are countless. Since biometallic complementarities sustain many mechanisms in living systems, they are essential for their adaptation to the environment and key to their health. Understanding how metallic moieties interact with biological systems is not just a health-related question. It is also a source of inspiration for the development of new synthetic routes, including environmentally friendly ones. As such, chemists have long intended to mimic biometallic synergies for industrial purposes. In the form of small amenable chemical systems (biomimetics) or complex *de novo* enzymes, a substantial part of chemical efforts focuses on reproducing or even extending Nature's bioinorganic repertoire.

The specific problem of predicting where and how metallic moieties bind to proteins is of crucial importance in numerous areas of chemistry and its closest interfaces.[2] After decades, researchers in bioinorganics still intensively work on decoding the exact weight of the different molecular variables that are involved in the construction of biometallic edifices (charges of the metal ions and their amino acid environment versus electric field of the entire protein, redox properties of the metal, etc.). *In silico* approaches offer an excellent option to get a better insight into metal-protein interactions and help in establishing predictive patterns. To date, though, only a limited ensemble of computational tools is available. Those methods are either based on sequence or structural analyses.[3–5]

On one side, computational predictors of metal-binding sites built on sequence analyses are mostly based on scanning the sequence of a target protein to identify those regions where amino acids patterns match a metal-binding site fingerprint. Some examples are MetalDetector,[6] that identifies potential transition metals binding sites involving Cys and His side chains as donors; MetalPredator,[7] that predicts potential Fe-S motifs in proteins; or ZincFinder,[8] which is aimed at predicting zinc-binding motives. On the other, structure-

based predictors mostly focus on identifying specific structural metal-binding motifs derived from databases of existing protein-bound to metal structures that match the best with the atomic organization of the target protein. Some of these approaches are metal-specific, like TEMPS and FETURE, that are dedicated to identifying Zinc binding motifs.[9,10] Others consider more extensive lists of metals that typically include Ca, Cu, Fe, Mg, Mn, and Zn.[11–17] Both families of methods could actually be used conjointly.

Metal-binding site predictors are successful in finding sites that match well-known patterns. However, the identification of those sites in systems that do not display an optimal sequence or which structure does not present the sufficient amount of preorganization can be challenging. Such situations are frequent in modern inorganic biochemistry. For example, de novo design of metalloenzymes frequently require mutations of the protein to generate a convenient coordination center. Moreover, experimental structures may only provide with metal-free (apo) form without convenient organization of the protein side chain for the metal to bind including transient geometries that require conformational changes or modification of the coordination environment (vacant sites, presence of labile ligands etc…) for the metal to bind.

In recent years, our group has actively worked in areas where it was essential to predict with confidence those amino acids that could coordinate metal compounds either for proteins[18] or peptides.[19] In those works, we frequently confront with new-to-Nature metal-binding sites, X-ray structures of proteins without a well-defined binding site[20,21] or even putative binding sites of metallodrugs.[22–29] Part of our efforts allowed us to develop a series of docking approaches[30–32] that can well reproduce low energy poses of metal-protein complexes. However, such approaches are only viable if good guesses for metal-binding regions are accessible on the first hand. For this aspect, in a first statistical analysis about ten years ago,[20] we observed that i) any amino acid with its α-carbon located within a sphere up to 7-9 Å could potentially be considered as a metal ligand (obviously considering possible coordination by residues with long side chains like lysine or tyrosine) and ii) the center of mass of the coordinating residues generally does not stand far from the exact position of the metal ions observed in the X-ray structure. In a subsequent study,[32] we preselected docking areas by filtering the distance from the metal to the β-carbon of the amino acids Asp, His, Glu and Cys, using a range from 2.5 to 5.0 Å. Both works showed that few geometric descriptors related to the backbone of the protein afford with convenient information to predict metal-binding sites rather than focusing only on the first coordination sphere donors (mainly side chains) that coordinate the metal.

BioMetAll represents a major step forward in how to predict metal-binding sites in proteins. The approach stands on the geometric information extracted from a collection of about 171000 biological complexes to infer a combination of geometric rules that are used to evaluate the regions of a protein for metal binding. The program accurately reproduces the experimental metal-binding motifs on a benchmark of 53 structures. Furthermore, we present three specific case studies to illustrate its potential. BioMetAll is efficient to i) predict metal-binding sites with particular amino acid motifs even for non-canonical metal coordinating amino acids and backbone atoms; ii) determine transient metal-binding sites in

structures that need conformational changes to reach the complete coordination sphere, which could be related to metal diffusion pathways; and even iii) predict potential mutations of an existing protein to generate convenient metal-binding sites. This new approach can make a significant contribution to future development in fields like metalloenzyme and metallodrug design.

Computational methodology

The concept

BioMetAll relies on the assumption that the geometric patterns of the backbone of a given protein environment encode sufficient information on its preorganization to coordinate metal ions. By doing so, predictions are less conditioned by finding perfect matches and so could detect those binding sites that are partially organized (e.g. missing one or several well-positioned side chains) or even help in designing new metal coordination environment throughout mutations (e.g. for application in enzyme design).

BioMetAll prediction works by embedding a protein in a grid of fictitious metal probes, which are further evaluated using geometric features obtained for each amino acid by statistical analysis. The entire set of geometric parameters is divided in two: one for coordinating atoms being part of the side chains and the other for atoms being part of the backbone. For side chains, the geometric features are (Figure 1A): Ma, the distance from the metal to the α-carbon; Mb, the distance from the metal to the β-carbon; and Mab, the angle between the metal, the α-carbon and the β-carbon. For backbone oxygen atoms (Figure 1B): MO, the distance from the metal to the backbone oxygen; and MOC, the angle between the metal, the backbone oxygen and the backbone carbon.

Statistical analysis

We analyzed the entire bunch of the structures available in the MetalPDB, a publicly accessible database of preselected metal-containing proteins from the Protein Data Bank (PDB).[33,34] This analysis was carried out thanks to an in-house PyChimera[35,36] script applied on ca. 171000 metal-binding sites. It ended up with a list of ca. 500000 entries (Table S1) after cleaning the database from metal-donor distances lower than 1.5 Å or greater than 3.0 Å.[37]
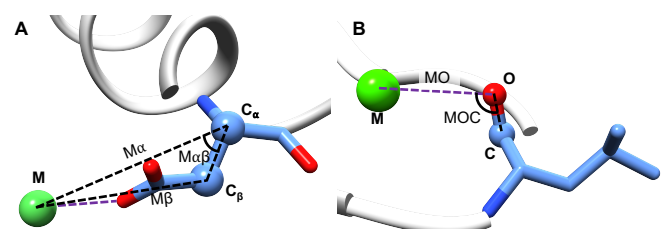

Figure 1. Geometric features analyzed for all the structures of the MetalPDB. **(A)** Geometric features considered for coordination bonds with a side chain donor. The metal (M) is depicted in a green ball, α-carbon (C_α) and β-carbon (C_β) are depicted in blue balls. Distance from the metal to the α-carbon (Mα) and distance from the metal to the β-carbon (Mβ) are represented in black dashed lines. The angle between metal, α-carbon and β-carbon is labelled as Mαβ **(B)** Geometric features considered for coordination bonds with a backbone oxygen donor. The metal (M) is depicted in a green ball, the backbone oxygen atom (O) is depicted as a red ball, and the backbone carbon (C) is depicted in a blue ball. Distance from the metal to the backbone oxygen (MO) is represented in a purple dashed line. The angle between metal, backbone oxygen and backbone carbon is labelled as MOC.

From the 500000 entries, a range of values was obtained for each geometric feature Ma, Mb, Mab (Table 1, Figure 2). For histidine, the subranges obtained for His$_\delta$ and His$_\varepsilon$ tautomers were merged. Although the analysis of coordination by backbone atoms was undertaken for each amino acid, the final results (Table S2) presented a large overlap. In consequence, a global range for MO and MOC was considered for all the entries: MO, distance metal - backbone oxygen from 1.809 Å to 3.000 Å; and MOC, angle metal - backbone oxygen - backbone carbon from 1.774 Rad to 3.139 Rad.

Further analyses were conducted to provide a series of default parameters for a prototypical prediction. It first emerges that 80% of donor atoms are part of the side chains (Table S3), and 19% are part of the backbone (Table S4). Default calculations in BioMetAll were therefore set to look for side chains primarily and left backbone coordination as optional. For side chains, the 89% of coordinating atoms comes from what we called *canonical* metal coordinating amino acids namely Asp, His, Glu and Cys while a

Table 1. Ranges (excluding outliers) for each one of the geometric features analyzed, grouped by type of amino acid. Coordinations with side-chain donors were considered in this analysis.

Amino acid	Percentage of entries[a]	Distance Mα (Å)	Distance Mβ (Å)	Angle M$\alpha\beta$ (Rad)
Asp	28.63%	[3.907-6.192]	[3.658-5.052]	[0.003-1.871]
His[b]	23.36%	[3.076-7.098]	[3.047-6.073]	[0.001-2.018]
Glu	18.96%	[3.591-8.303]	[4.123-6.108]	[0.000-2.305]
Cys	18.01%	[3.248-5.451]	[2.794-3.829]	[0.006-1.734]
Asn	3.80%	[4.048-5.701]	[3.866-5.250]	[0.682-1.697]
Thr	1.99%	[3.313-5.096]	[2.530-4.371]	[0.184-1.647]
Ser	1.81%	[3.384-5.171]	[2.587-4.353]	[0.163-1.521]
Gln	1.26%	[3.217-7.835]	[3.512-6.565]	[0.012-2.268]
Met	0.72%	[4.426-6.706]	[3.654-5.094]	[0.016-1.228]
Tyr	0.71%	[6.896-9.016]	[6.367-8.087]	[0.501-1.433]
Lys	0.34%	[5.134-9.280]	[4.358-7.997]	[0.029-1.918]
Arg	0.30%	[3.409-10.162]	[2.913-9.042]	[0.024-1.991]
Leu	0.03%	[3.784-4.678]	[4.353-5.622]	[1.464-2.470]
Trp	0.02%	[3.912-7.478]	[4.244-6.573]	[0.471-2.703]
Ile	0.02%	[4.236-4.817]	[5.039-5.702]	[1.826-2.169]
Val	0.01%	[3.517-5.003]	[4.358-5.722]	[1.639-2.166]
Ala	0.01%	[3.402-5.126]	[4.254-5.991]	[1.584-2.368]
Phe	0.00%	[3.499-4.922]	[4.289-5.590]	[1.666-2.300]
Pro	0.00%	[3.646-4.529]	[4.860-5.246]	[1.746-2.678]

[a] Percentage of bonds corresponding to each residue of the total of 400077 entries analyzed. [b] Ranges for His were obtained from merging the subranges calculated for Hisδ1 and Hisε2-H tautomers.
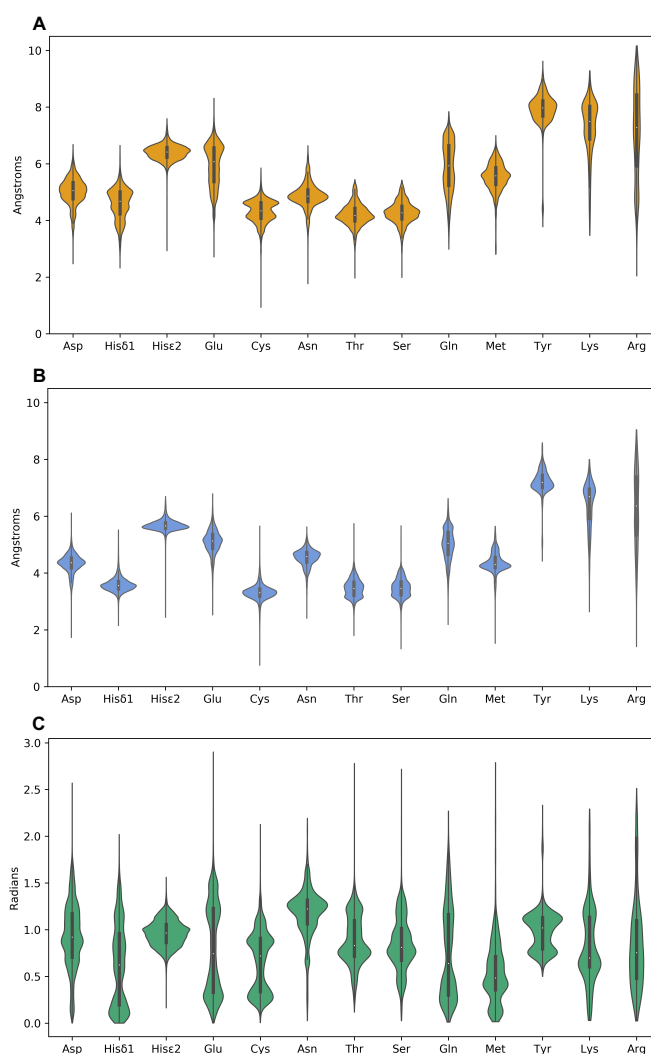
Figure 2. Violin plots for the different features studied in the statistical analysis. Coordination bonds between the metal and a side-chain atom are considered in all cases. **(A)** Distance from metal to α-carbon of the coordinating amino acid, for canonical and *non-canonical* amino acids. **(B)** Distance from metal to β-carbon of the coordinating amino acid, for canonical and *non-canonical* amino acids. **(C)** Angle between metal, α-carbon and β-carbon of the coordinating amino acid, for canonical and *non-canonical* amino acids.

10.9% belongs to *non-canonical* amino acids (Asn, Thr, Ser, Gln, Met, Tyr, Lys and Arg). Only the last small fraction (<0.1%) are coordinations with a carboxylate terminal amino acid (Table S3). Finally, most common metal-binding sites account with three coordinating amino acids on average (with an interquartile range of 2). On those grounds, BioMetAll works with default parameters that involved canonical amino acids side chains with a number of amino acids per metal-binding sites of three. The user can configure any other possible situation in a user-friendly manner.

It is to notice that, besides standard amino acids, exogenous ligands (e.g. water or small molecules) can participate in the coordination of the metal. Although metrics for BioMetAll parameters have been obtained purely on amino acid geometric magnitudes, exogenous ligands could participate in the coordination of the metal on the analyzed set. The calculations we performed so far show that BioMetAll implicitly takes into account the changes in

the metal position that those exogenous ligands could cause (as showed latter in the manuscript).

Overview of BioMetAll Workflow

The BioMetAll workflow (Figure 3) is divided into three sequential steps.

i) Generation of a probes grid. An input *.pdb* file (that can be downloaded from the same server of the Protein Data Bank[38] -PDB-) is parsed to retain only the cartesian coordinates of α-carbons, β-carbons, backbone oxygens and backbone carbons. Subsequently, a spherical grid of equidistributed points (i.e. probes) is generated which by default, embeds the entire protein with an extra margin of 8 Å (to account for superficial amino acids). The user can customize this sphere if interested in a portion of the system only. Probes ubicated less than 1 Å from backbone atoms are discarded.

ii) Checking geometric criteria. The second stage of the process checks the suitability of the probes. Each of them is paired with the list of amino acids that match all the required geometric and user-specified constraints. The following parameters can be selected: *residues*, *motif*, *backbone* and *mutations* (Table 2). They allow for restraining the search to a specific list of amino acids (e.g. limiting *residues* to the canonical list), consider or not coordinations with the backbone oxygens, search for a particular structural motif, or even output feasible mutations which will be suitable to complete that motif.

iii) Clustering of possible binding areas and output. Once a list of possible coordinating residues for every probe is obtained, all the probes with the same coordination environment are grouped. Those clusters of probes represent the predicted site with metal binding capacities. The result can be reported into a *.pdb* format that can further be visualized with any visualization program. The resulting coordination area/s are ordered by decreasing number of probes and saved in an additional output text file. It is important to bear in mind that BioMetAll does not return a unique spot but groups of probes that constitute the areas where a metal could bind.

Availability of the code

BioMetAll is an open-source command-line application implemented in Python 3.7 language with NumPy[39] and psutil are the only required dependencies. User manual and tutorials are available at https://github.com/insilichem/biometall. It can be run as usual with any Python script, once the dependencies mentioned above are installed on the system. Pre-built executables (generated with PyInstaller[40]) for Linux, Mac OS and Windows 10 are available for direct use of BioMetAll without the need for any installation process.
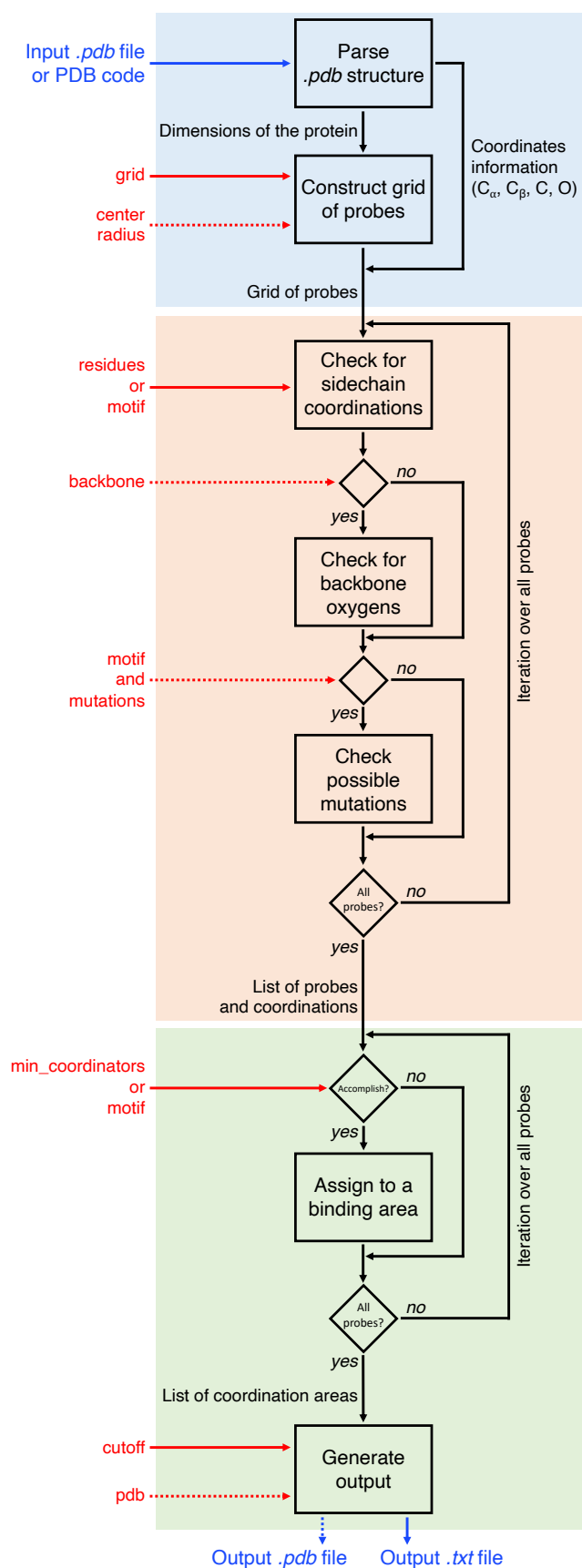

Figure 3. BioMetAll workflow. User configurable parameters are colored in red (optional parameters have a dashed arrow). Input and outputs are colored in blue. The three main steps of the algorithm are highlighted in blue, orange and green background, respectively.

Table 2. User-configurable parameters for a BioMetAll calculation

Parameter	Default value	Utility
grid	1.0	Distance (in Å) between probes in the grid.
center	None	Optional. Coordinates of the sphere center, when the user wants a custom search zone.
radius	None	Optional. Radius (in Å) of the search sphere, when the user wants a custom search zone.
residues	[Asp, His, Glu, Cys]	List of residues that will be considered as potential side-chain donors.
backbone	None	List of residues that will be considered as potential backbone donors.
motif	None	Optional. Concrete structural motif to search.
mutations	None	Optional. Propose mutations to complete a motif.
min_coordinators	3	Minimum number of amino acids that a coordination environment must have to be valid.
cutoff	0	Cut the solutions proposed by the software to a given threshold of the number of probes.
pdb	No	Generates a *.pdb* file with the proposed binding areas ready to open with a visualization program.

Results and discussion

Motif search benchmark

The predictiveness of BioMetAll was tested for the detection of the two-histidine one-carboxylate motifs (FTM) in a set of 53 crystallographic structure from the Protein Data Bank. The FTM was chosen because it is a coordination sphere motif present in the active site of many metalloenzymes, like carboxypeptidase A or naphthalenedioxygenase.[41–43] Thus, it is interesting to benchmark BioMetAll predictions against a set of experimental structures containing this motif.[44] This set appeared very convenient because it shows structures with a wide range in size (from 108 to 1074 residues) and great diversity of metals (Cd, Co, Cu, Fe, Hg, Mg, Mn, Ni, Ru and Zn).

In average, BioMetAll returns ten predicted metal-binding sites for each crystallographic structure. Those solutions have a number of probes of ca. 25 on average with a deviation of about 10. In 100% of the set, the FTM motif of the crystal structure is found in a high-populated cluster of probes (93% in one of the first five solutions, see Table S5). In 75% of the cases, the metal-binding site is well characterized in the most populated solution, although not necessarily with the complete set of donors displayed in the X-ray structure. BioMetAll results are therefore particularly promising. It is important to highlight that BioMetAll not only predicts the X-ray structure metal modes, it also predicts some that could be different from the experimental structures (from missing one of the amino acids in the coordination sphere to being far from the crystallographic site). In fact, reaching the maximum exigence of accuracy, that would correspond to the prediction of the X-ray structures to be found in the most populated solution only with the complete set of amino acids, success rate falls to 40%. Obviously, like for other techniques based on empirical scores, the predictiveness of the method cannot be limited to the best solution only and those secondary metal binding sites that BioMetAll detects are not necessarily false positives but transient metallic environments accessible to the protein (see our first application case on HSA). Energetic terms and other physico-chemical magnitudes that are not present in BioMetAll, could dictate the fine structure of the metal-binding site. Posterior filtering could be necessary depending on the objective of its use.

The accuracy of the predicted binding area was further assessed by measuring the distance from the experimental metal position to the nearest computed probe. On average a difference between calculated and experimental positions of the metal is about 0.56 ± 0.19 Å. Importantly, such quality is not influenced by the presence of donors coordinated to the metal provided by exogenous compounds (e.g. phosphate or water molecules). In these cases, we generally observe a displacement to the limits of the predicted region (Figure 4), suggesting that the exogenous donors only slightly modify the position of the metal and the organization of the side chains.

Because calculations are performed in a blind manner, meaning that all the system is screened for predicting metal-binding sites, these results are quite remarkable. Regarding the computational cost of the calculations, they range from 3s to 245s, with a median of 18s (with the size of the system highly correlating with the computational time r=0.89, p<0.0001).

The benchmark shows that BioMetAll is a suitable software to predict metal-binding sites. One can readily perform screening of protein structures in a rational enzyme design protocol
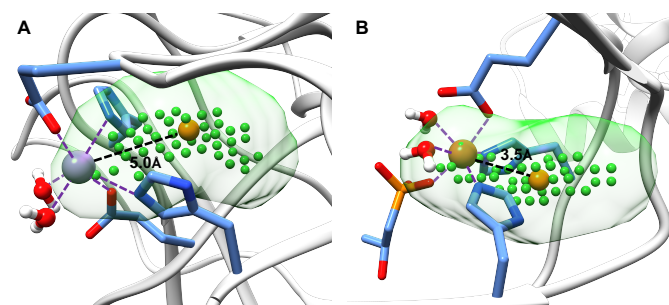

Figure 4. Detail of the binding site for two cases presenting exogenous ligands besides the standard amino acids. In both examples, the center of the predicted area is represented by an orange ball, the probes that form the binding area are represented by smaller green balls surrounded by surface visualization, and the coordination bonds are shown in dashed purple lines. **(A)** Binding of Mn in PDB 2hxg. The HIS-HIS-GLU-GLU motif is depicted in blue sticks, and two water molecules are highlighted in ball and sticks. **(B)** Binding of Fe in PDB 1zz7. The HIS-HIS-GLU motif and an exogenous ligand are depicted in blue sticks. Two water molecules are highlighted in ball and sticks.

or testing a set of conformations extracted from an MD simulation. With such a tool in hand, a large panel of applications is accessible.

Case study (1): Human Serum Albumin

Human Serum Albumin (HSA) is the most abundant protein in the blood plasma and, together with human serum transferrin, is the first carrier for most of the essential metal ions.[45] Two main metal-binding sites have been characterized in HSA: the NTS or ATCUN motif located at the N-Terminus,[46] and the multi-metal binding site (MBS) at the interface of subdomains IA and IIA.[47] While the NTS (with coordination set $NH_{2,Asp1};N^-_{Asp1};N^-_{Ala2};N_{His3}$) is preferential for species able to deprotonate amidic hydrogens like Cu^{2+} and Ni^{2+}, the MBS one (with the binding $N_{His67};N_{His247};O_{Asn99};COO^-_{Asp249}$) is the primary site for Zn^{2+} and other bivalent cations.[48,49] One interesting feature of HSA emerges when comparing its structure in the presence of medium- and long-chain fatty acids (FAs) with those of the defatted state. The binding of FAs to site FA2, which is one of the strongest FAs sites located at the same region of MBS, causes the mutual rotation of subdomains IA and IIA, separating the Zn^{2+} coordinating amino acids and moving part of Zn^{2+} to NTS with possible coordination $N_{His3/His9};COO^-_{Asp13/Glu6}$.[50] This conformational change hinders the metal coordination at MBS and has been pointed out in the literature as a regulatory mechanism of zinc homeostasis (Figure 5). [45,48,51–55]

Predicting metal-binding sites in front of conformational changes is an interesting challenge for structure-based predictors. Indeed, sequence-based ones would not consider such modification of the protein structure. To ascertain the discriminative capabilities of BioMetAll, both fatted (HSA[f]) and defatted (HSA[d]) X-Ray structures of HSA (PDB codes 1bj5 and 5ijf, respectively) were probed for zones containing a minimum of three coordinator amino acids, whose types should be a combination of His, Asp and Glu. Those are amino acids which side chains are those with the highest affinity toward Zn^{2+}. A second prediction was performed with a more specific motif similar to the MBS site: two histidine residues and either an aspartate or a glutamate.

As a result of the first analysis, both known primary binding sites (MBS and NTS) were detected by BioMetAll for HSA[d] with His-His-Asp/Glu along with other secondary sites that satisfy patterns like $N_{His}/COO^-_{Asp/Glu};COO^-_{Glu};COO^-_{Asp/Glu}$ (Table S6) and
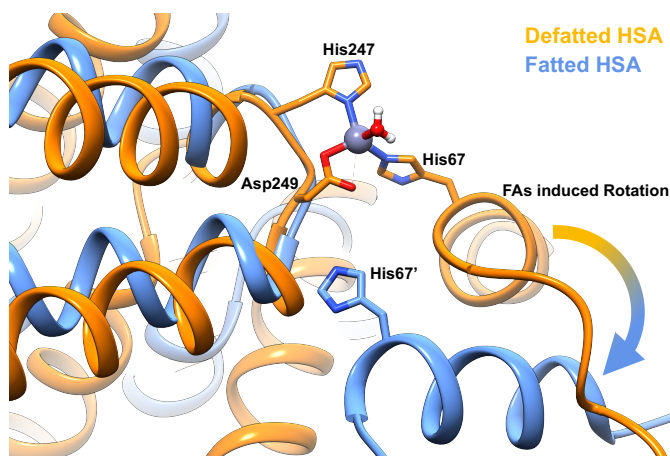

Figure 5. FAs induced rotation, highlighted by the superposition of IA/IIA interface of HSA in HSA[f] (in blue) and HSA[d] (in orange) X-Ray structures.
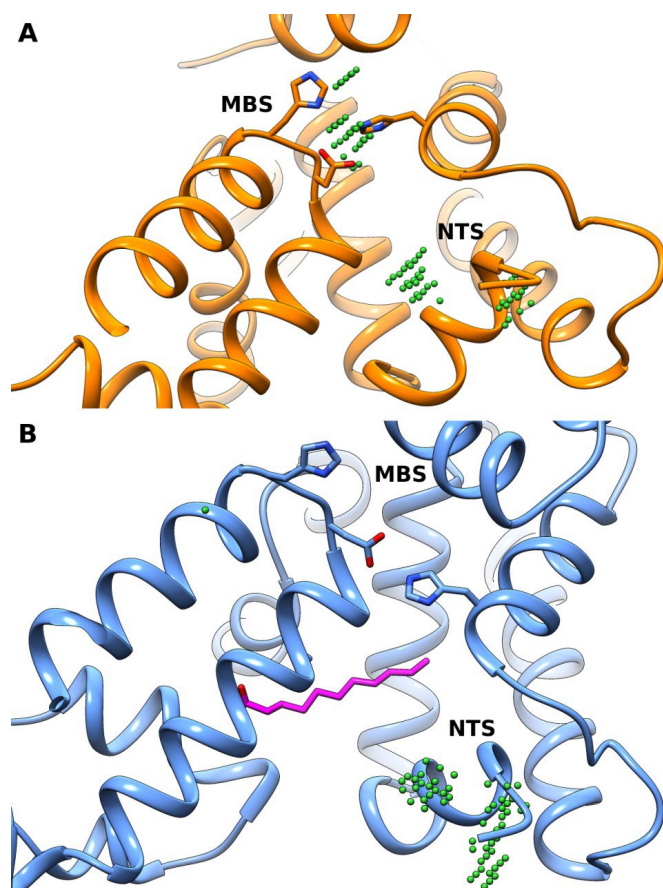

Figure 6. Probes (green dots) predicted by BioMetAll by looking for the motif His-His-Asp/Glu. **(A)** Defatted conformation, in which the experimental binding site is predicted with high precision. **(B)** Fatted conformation, where the experimental binding site is no longer predicted. FA bound to FA2 is depicted in purple.

that have also been reported in the literature.[50] However, for HSA[f], only the NTS (His3-His9-Glu6) together with the same secondary sites than HSA[d] are identified (Table S7). By looking for the specific motif His-His-Asp/Glu, the MBS and NTS sites were uniquely identified for HSA[d] (Figure 6A) while only the NTS binding site was detected for HSA[f] (Figure 6B). Hence, BioMetAll differentiates the metal-binding profile between both conformational states of HSA.

This case study points out that BioMetAll is not only able to predict with accuracy primary and secondary metal-binding sites but also to account for the impact of conformational changes of the protein and how it affects its preorganization for metal binding.

Case study (2): Hemocyanins

Our second illustrative case focuses on copper-binding proteins. From hemocyanin to blue copper protein, Nature has intensively adapted its toolbox to incorporate this metal. Interestingly, copper is a very selective transition metal in terms of amino acids to which it could bind. Statistics (Table S3) show that copper highly prefers histidine (70.8%). Other amino acids like cysteines (16.3%) or methionines (5.6%) are found with far lower frequencies. Glutamates (3.5%), as well as aspartates (2.3%), end the list of main coordinating amino acids to copper. One could therefore hypothesize that a way to detect copper-binding sites consists in identifying those histidine residues that are well pre-organized for metal binding.

We performed BioMetAll runs on a series of hemocyanin proteins –a prototypical family of di-copper binding oxygen carriers– that coordinate copper throughout motifs of three histidine amino acids. After removal of the metal ions from the .pdb files, a set of experimentally characterized structures were screened to look for sites with a minimum of three histidine residues able to coordinate a metal ion. The search led to a 100% prediction for the di-copper binding sites for all the hemocyanins analyzed at the exact location of the X-ray structure (Table S8). One example is given in Figure 7 with Manduca sexta prophenoloxidase, a heterodimer consisting of two homologous polypeptide chains.[56]

One of the advantages of BioMetAll is to provide the user to control the specificities of the metal-binding search pattern. Even those that are not the "best-case scenario" can be evaluated based, for example, on chemical hypothesis. With such an idea in mind, we performed a second series of calculations releasing the restraint of the search motif from three histidine residues to only two with eventually a third amino acids being cysteine, methionine, aspartate or glutamate. The results provide with several putative metal-binding sites spreading to a broader part of the protein but, to our surprise, those were not randomly scattered inside the scaffold (example for Manduca sexta prophenoloxidase shown in Figure 8). Instead, a continuum pattern of adjacent sites that evolves from solvent-exposed locations to the buried di-copper site was observed. This observation, combined with the presence of water molecules along those channels in several crystal structures, strongly suggest the possibility that BioMetAll allowed the detection of metal diffusion pathways. It is interesting to comment that our intents to use
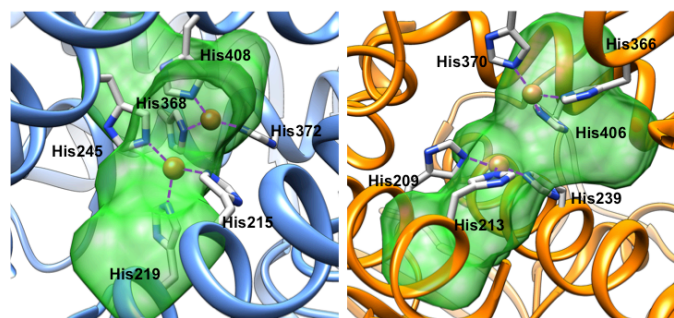

Figure 7. Di-copper binding sites for the two monomers (in blue and orange ribbons) of the Manduca sexta prophenoloxidase (PDB code 3hhs). Probes predicted by BioMetAll are shown in green surface and amino acids participating in copper binding are highlighted in white sticks.
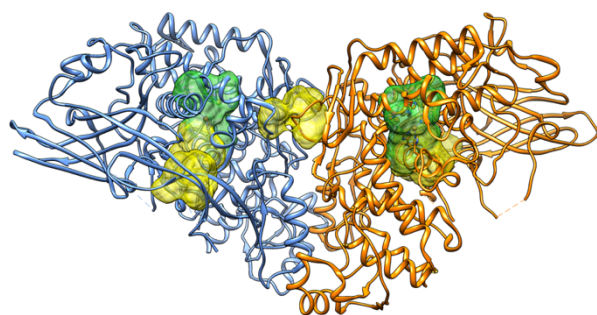

Figure 8. Overall view of the structure of Manduca sexta prophenoloxidase, with the results of BioMetAll calculations overlapped. Probes for the three-histidine calculation are shown in green surface and probes for the two-histidine calculation are shown in yellow surface. Continuous paths from the solvent to the di-copper binding sites are observed.

traditional approaches for channeling predictions, like those based on calculating solvent excluded surface calculations, did not lead to such clear prediction of metal compatible pathways neither a map of histidine onto the crystallographic structure would provide with such information (although being obviously related). Although falling far for the main objectives of this work, such observation could be particularly interesting to investigate the properties of metal diffusion in hemocyanins and their relationship to oligomeric states. The versatility of BioMetAll could lead to interesting outcoming when looking for diffusion pathways of metallic species in proteins.

Case study (3): Metalloenzyme design

Because BioMetAll is mostly based on backbone patterns, prediction of mutations to generate metal-binding sites is probably of the most interesting applications. For our last illustrative case, we aimed at showing the interest of this approach for metalloenzyme design. To do so, the repurposing study of a hydrolase to a peroxidase by Ward et al.[57] was revisited.

The original work by Ward and coworkers started with a screening of the PDB[38] for FTM compatible motifs including possible mutations from Asn/Gln to Asp/Glu. With an initial screening of the entire Protein Data Bank with the software STAMPS,[58] six structures were identified as interesting candidates and followingly tested for peroxidase activity. Only 6-phosphogluconolactonase from mycobacterium smegmatis (which structure is accessible at the PDB with the code 3oc6) with an Asn131Asp mutation showed catalysis in the presence of CuSO$_4$. Interestingly, X-ray crystallography of the resulting system enlightened that only the His67-His104 tandem with histidine *trans* to each other coordinates the copper ion; the predicted de novo aspartate at the 131 position was not able to coordinate.

To show the interest of BioMetAll in an artificial metalloenzyme framework, we started by looking for the His-His-Asp/Glu motif in the 6-PGLac structure. In agreement with the initial study, no positive outcomes came out. The search was therefore expanded to His-His-Asn/Gln motifs obtaining the His67-His104-Asn131 triad found by Ward *et al*. The subsequent analysis was searching for any mutation that could complete the His67-His104 pair with a Glu/Asp. Unlike other structural prediction programs that need the exact sequence to be present in the target protein, BioMetAll directly outputs a list of mutations.

This run of BioMetAll looking for possible sites allowing a mutation to either Asp or Glu provided with an extensive list of possible solutions including Asn131 (Table 3). The following step was a post-process by filtering the solutions that show collapse between the probes and any amino acid side chain or backbone from close neighbors of the putative metal-binding areas. This filtering process together with a visual inspection diminished the number of possibilities to five candidates (Table 3 and Figure 9A).

To further assess the predictiveness of BioMetAll under a metalloenzyme design framework, the predicted systems were used as a starting point for docking experiments. Calculations were carried out with the protein-ligand docking software GOLD software[59–61] and repeated inside our GaudiMM platform (to provide with a freeware alternative).[32,62] Both calculations were performed

Table 3. Summary of mutations proposed by BioMetAll to achieve a His-His-Glu/Asp in 6-PGLac structure, starting from the already present His67-His104.[a]

Number	Mutated a.a.	Asp[b]	Glu[d]
1	**Tyr69**	**18 (17)**	**34 (27)**
2	**Thr38**	**11 (5)**	**31 (13)**
3	**Asn131**	**8 (0)**	**29 (9)**
4	Ile68[e]	3 (1)	18 (8)
5	Ile39[e]	2 (0)	9 (5)
6	**Val101**	**2 (2)**	**18 (13)**
7	Val40[e]	2 (1)	10 (4)
8	**Asn102**	**1 (1)**	**19 (14)**
9	Leu127[e]	0	7 (3)
10	Leu128	0	2 (0)
11	Phe139	0	2 (0)
12	Val103	0	2 (0)
13	Leu41	0	1 (0)
14	Pro99	0	1 (0)

[a] In bold the possibilities took into account in the docking assay. [b] Number of probes for the binding area when mutating the a.a. to Asp is reported, in parenthesis the number of probes after filtering. [c] Number of probes for the binding area when mutating the a.a. to Glu is reported, in parenthesis the number of probes after filtering. [e] Solutions excluded due to visual inspection (see Figure 9A).

following our protocols optimized for metalloligands binding to proteins[30,31] (see Supplementary Information for further details). The list of FTM tested includes therefore His67-His104-Asp/Glu69, His67-His104-Asp/Glu38, His67-His104-Glu131, His67-His104-Asp/Glu101, and His67-His104-Asp/Glu102. For the sake of the comparison with the initial work and the consequent X-ray structure, His67-His104-Asp131 was also analyzed. From all the calculations performed, solutions presenting convenient metal-binding modes were only observed for the mutation Tyr69Asp/Glu even though the Asp mutation reaches better coordination geometry in the docking assay. In this case, the FTM motive appears excellently reproduced with the side chains displaying geometries that allow the aspartate to bind the metal facially (Figure 9B). It must be highlighted that only those mutations with about 10 probes and more appear as viable for coordination, with the mutation at the position 69 being outstanding for both Asp and Glu mutations. This finding suggests that the number of probes could be used as criteria to prioritize possible proposed solutions. Strikingly, no docking calculation allows producing a convenient binding mode of the FTM motive for the Asn131Asp mutation with no rearrangement of the His67-His104-Asp131 side chains to coordinate with the metal. This result sustains the BioMetAll prediction regarding the low potential of such mutant to coordinate convenient the metal as well as with the experimental X-ray structure.
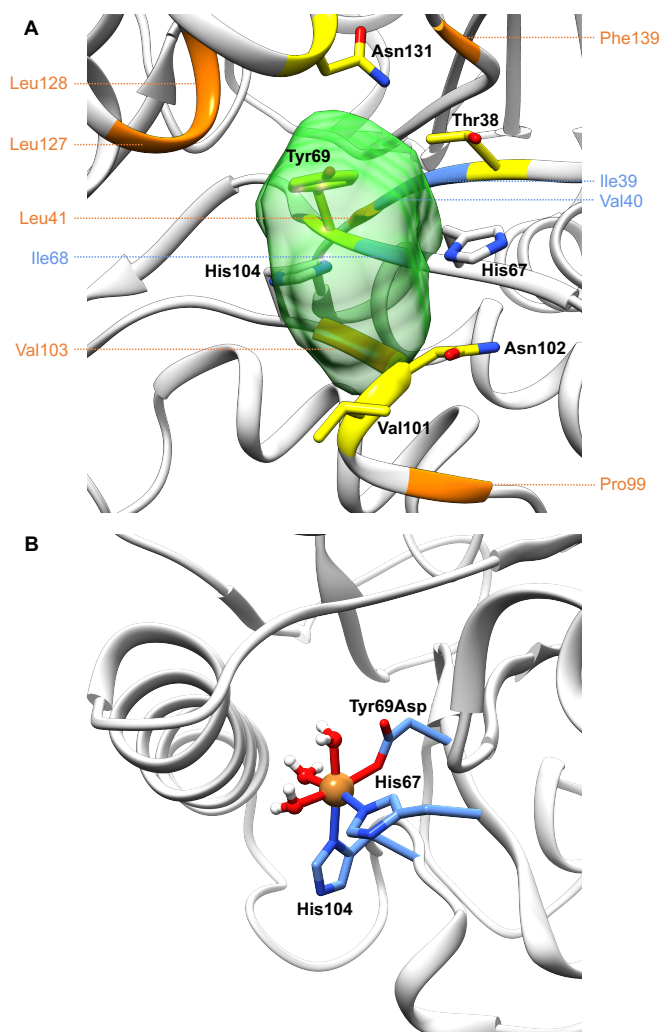
Figure 9. (A) Set of feasible mutations predicted by BioMetAll to add an Asp/Glu to the His67-His104 motif. The five options considered in further docking analysis (Tyr69, Thr38, Asn131, Val101 and Asn102) are depicted in yellow sticks. The discarded mutations are highlighted in orange (options for Glu) or blue (options for Asp/Glu) ribbon. The probes area predicted as feasible for metal binding is shown in green surface. **(B)** Best docking pose obtained for Tyr69Asp mutation, where the copper is facially coordinated by the His67-Asp69-His104 triad.

From this part of the study, BioMetAll and its combination with docking experiments optimized for metal binding appear to predict convenient side-chain rearrangements. For the purpose of this enzyme design exercise should only be viable for the Tyr69Asp and to less extent Tyr69Glu.

Conclusions

The large number of experimental structures available in the Protein Data Bank offers today a unique landscape to understand the molecular rules that govern the binding of metallic moieties to proteins. In this work, we show that it is possible to build an efficient predictor of metal-binding sites based on few geometric descriptors of the conformation of the backbone. Our intent, BioMetAll, gives a major weight of the prediction to the preorganization capacities of the protein instead of a perfect matching with known geometries of the first coordination sphere of the metal.

The predictiveness of BioMetAll is excellent and compares very well with other algorithms when it comes to determine metal-binding sites as well as the exact list of the amino acids involved in the coordination motifs. Those predictions are of good quality even when exogenous compounds are involved. BioMetAll tests on several key systems show that i) it efficiently predicts primary as well as secondary metal-binding sites, ii) it can easily account on how the conformational changes of the protein could alter the formation of metal-binding sites, iii) it allows to work with common metal-binding templates as well as non-standard ones, iv) it identifies sites even with incomplete coordination spheres, and v) it allows to predict mutations that could be necessary to generate new metal-binding sites in a protein for example for building new biocatalysts.

BioMetAll, and its ground on backbone preorganization, appears as a very interesting option in fields related to biometallic interactions. It is fundamental to highlight that BioMetAll does not consider energetic terms. Posterior refinements might be necessary depending on the object of the study. At the moment, we put a large amount of effort to bridge BioMetAll to other methods like those allowing extensive conformational sampling such as Molecular Dynamics and protein-ligand dockings. We believe that this kind of combination would open new horizons for research in chemistry and its interface with biology.

References

(1) Rehder, D. *Bioinorganic Chemistry: An Introduction*; OUP Oxford, 2014.

(2) Reilley, D. J.; Hennefarth, M. R.; Alexandrova, A. N. The Case for Enzymatic Competitive Metal Affinity Methods. *ACS Catal.* **2020**, *10* (3), 2298–2307.

(3) Riccardi, L.; Genna, V.; De Vivo, M. Metal–Ligand Interactions in Drug Design. *Nat. Rev. Chem.* **2018**, *2* (7), 100–112.

(4) Mallick, M.; Sharan Vidyarthi, A. Tools for Predicting Metal Binding Sites in Protein: A Review. *Curr. Bioinform.* **2011**, *6* (4), 444–449.

(5) Akcapinar, G. B.; Sezerman, O. U. Computational Approaches for de Novo Design and Redesign of Metal-Binding Sites on Proteins. *Biosci. Rep.* **2017**, *37* (2), BSR20160179.

(6) Passerini, A.; Lippi, M.; Frasconi, P. MetalDetector v2.0: Predicting the Geometry of Metal Binding Sites from Protein Sequence. *Nucleic Acids Res.* **2011**, *39* (Web Server issue), W288–W292.

(7) Valasatava, Y.; Rosato, A.; Banci, L.; Andreini, C. MetalPredator: A Web Server to Predict Iron-Sulfur Cluster Binding Proteomes. *Bioinformatics* **2016**, *32* (18), 2850–2852.

(8) Passerini, A.; Andreini, C.; Menchetti, S.; Rosato, A.; Frasconi, P. Predicting Zinc Binding at the Proteome Level. *BMC Bioinformatics* **2007**, *8*, 39.

(9) Ebert, J. C.; Altman, R. B. Robust Recognition of Zinc Binding Sites in Proteins. *Protein Sci.* **2008**, *17* (1), 54–65.

(10) Zhao, W.; Xu, M.; Liang, Z.; Ding, B.; Niu, L.; Liu, H.; Teng, M. Structure-Based de Novo Prediction of Zinc-Binding Sites in Proteins of Unknown Function. *Bioinformatics* **2011**, *27* (9), 1262–1268.

(11) Lin, Y.-F.; Cheng, C.-W.; Shih, C.-S.; Hwang, J.-K.; Yu, C.-S.; Lu, C.-H. MIB: Metal Ion-Binding Site Prediction and Docking Server. *J. Chem. Inf. Model.* **2016**, *56* (12), 2287–2291.

(12) Brylinski, M.; Skolnick, J. FINDSITE-Metal: Integrating Evolutionary Information and Machine Learning for Structure-Based Metal-Binding Site Prediction at the Proteome Level. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (3), 735–751.

(13) He, W.; Liang, Z.; Teng, M.; Niu, L. MFASD: A Structure-Based Algorithm for Discriminating Different Types of Metal-Binding Sites. *Bioinformatics* **2015**, *31* (12), 1938–1944.

(14) Sodhi, J. S.; Bryson, K.; McGuffin, L. J.; Ward, J. J.; Wernisch, L.; Jones, D. T. Predicting Metal-Binding Site Residues in Low-Resolution Structural Models. *J. Mol. Biol.* **2004**, *342* (1), 307–320.

(15) Wass, M. N.; Kelley, L. A.; Sternberg, M. J. E. 3DLigandSite: Predicting Ligand-Binding Sites Using Similar Structures. *Nucleic Acids Res.* **2010**, *38* (suppl_2), W469–W473.

(16) Lu, C.-H.; Lin, Y.-F.; Lin, J.-J.; Yu, C.-S. Prediction of Metal Ion-Binding Sites in Proteins Using the Fragment Transformation Method. *PLoS One* **2012**, *7* (6), e39252–e39252.

(17) Hu, X.; Dong, Q.; Yang, J.; Zhang, Y. Recognizing Metal and Acid Radical Ion-Binding Sites by Integrating Ab Initio Modeling with Template-Based Transferals. *Bioinformatics* **2016**, *32* (21), 3260–3269.

(18) Allard, M.; Dupont, C.; Muñoz Robles, V.; Doucet, N.;

Lledós, A.; Maréchal, J.-D.; Urvoas, A.; Mahy, J.-P.; Ricoux, R. Incorporation of Manganese Complexes into Xylanase: New Artificial Metalloenzymes for Enantioselective Epoxidation. *ChemBioChem* **2012**, *13* (2), 240–251.

(19) Mujika, J. I.; Rodríguez-Guerra Pedregal, J.; Lopez, X.; Ugalde, J. M.; Rodríguez-Santiago, L.; Sodupe, M.; Maréchal, J.-D. Elucidating the 3D Structures of Al(Iii)–Aβ Complexes: A Template Free Strategy Based on the Pre-Organization Hypothesis. *Chem. Sci.* **2017**, *8* (7), 5041–5049.

(20) Robles, V. M.; Ortega-Carrasco, E.; Fuentes, E. G.; Lledos, A.; Marechal, J.-D. What Can Molecular Modelling Bring to the Design of Artificial Inorganic Cofactors? *Faraday Discuss.* **2011**, *148*, 137–159.

(21) Alonso-Cotchico, L.; Rodríguez-Guerra, J.; Lledós, A.; Maréchal, J.-D. Molecular Modeling for Artificial Metalloenzyme Design and Optimization. *Acc. Chem. Res.* **2020**, *53* (4), 896–905.

(22) Sciortino, G.; Sanna, D.; Ugone, V.; Micera, G.; Lledós, A.; Maréchal, J.-D.; Garribba, E. Elucidation of Binding Site and Chiral Specificity of Oxidovanadium Drugs with Lysozyme through Theoretical Calculations. *Inorg. Chem.* **2017**, *56* (21), 12938–12951.

(23) Sciortino, G.; Sanna, D.; Ugone, V.; Lledós, A.; Maréchal, J.-D.; Garribba, E. Decoding Surface Interaction of VIVO Metallodrug Candidates with Lysozyme. *Inorg. Chem.* **2018**, *57* (8), 4456–4469.

(24) Sciortino, G.; Sanna, D.; Ugone, V.; Maréchal, J.-D.; Garribba, E. Integrated ESI-MS/EPR/Computational Characterization of the Binding of Metal Species to Proteins: Vanadium Drug–Myoglobin Application. *Inorg. Chem. Front.* **2019**, *6* (6), 1561–1578.

(25) Sciortino, G.; Sanna, D.; Ugone, V.; Maréchal, J.-D.; Alemany-Chavarria, M.; Garribba, E. Effect of Secondary Interactions, Steric Hindrance and Electric Charge on the Interaction of VIVO Species with Proteins. *New J. Chem.* **2019**, *43* (45), 17647–17660.

(26) Ugone, V.; Sanna, D.; Sciortino, G.; Maréchal, J.-D.; Garribba, E. Interaction of Vanadium(IV) Species with Ubiquitin: A Combined Instrumental and Computational Approach. *Inorg. Chem.* **2019**, *58* (12), 8064–8078.

(27) Sciortino, G.; Sánchez-Aparicio, J.-E.; Rodríguez-Guerra Pedregal, J.; Garribba, E.; Maréchal, J.-D. Computational Insight into the Interaction of Oxaliplatin with Insulin. *Metallomics* **2019**, *11* (4), 765–773.

(28) Sanna, D.; Ugone, V.; Sciortino, G.; Buglyó, P.; Bihari, Z.; Parajdi-Losonczi, P. L.; Garribba, E. VIVO Complexes with Antibacterial Quinolone Ligands and Their Interaction with Serum Proteins. *Dalt. Trans.* **2018**, *47* (7), 2164–2182.

(29) Sciortino, G.; Ugone, V.; Sanna, D.; Lubinu, G.; Ruggiu, S.; Maréchal, J.-D.; Garribba, E. Biospeciation of Potential Vanadium Drugs of Acetylacetonate in the Presence of Proteins. *Front. Chem.* **2020**, *8* (345), 1–11.

(30) Sciortino, G.; Garribba, E.; Maréchal, J.-D. Validation and Applications of Protein–Ligand Docking Approaches Improved for Metalloligands with Multiple Vacant Sites. *Inorg. Chem.* **2019**, *58* (1), 294–306.

(31) Sciortino, G.; Rodríguez-Guerra Pedregal, J.; Lledós, A.; Garribba, E.; Maréchal, J.-D. Prediction of the Interaction of Metallic Moieties with Proteins: An Update for Protein-Ligand Docking Techniques. *J. Comput. Chem.* **2018**, *39* (1), 42–51.

(32) Sciortino, G.; Garribba, E.; Rodríguez-Guerra Pedregal, J.; Maréchal, J. D. Simple Coordination Geometry Descriptors Allow to Accurately Predict Metal-Binding Sites in Proteins. *ACS Omega* **2019**, *4* (2), 3726–3731.

(33) Putignano, V.; Rosato, A.; Banci, L.; Andreini, C. MetalPDB in 2018: A Database of Metal Sites in Biological Macromolecular Structures. *Nucleic Acids Res.* **2018**, *46* (Database issue), D459–D464.

(34) Andreini, C.; Cavallaro, G.; Lorenzini, S.; Rosato, A. MetalPDB: A Database of Metal Sites in Biological Macromolecular Structures. *Nucleic Acids Res.* **2013**, *41* (Database issue), D312–D319.

(35) Rodríguez-Guerra Pedregal, J.; Maréchal, J. D. PyChimera: Use UCSF Chimera Modules in Any Python 2.7 Project. *Bioinformatics* **2018**, *34* (10), 1784–1785.

(36) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.

(37) Orpen, A. G.; Brammer, L.; Allen, F. H.; Kennard, O.; Watson, D. G.; Taylor, R. Supplement. Tables of Bond Lengths Determined by X-Ray and Neutron Diffraction. Part 2. Organometallic Compounds and Co-Ordination Complexes of the d- and f-Block Metals. *J. Chem. Soc. Dalt. Trans.* **1989**, *12*, S1–S83.

(38) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.

(39) Van Der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13* (2), 22–30.

(40) PyInstaller https://www.pyinstaller.org (accessed May 10, 2020).

(41) Koehntop, K. D.; Emerson, J. P.; Que, L. The 2-His-1-Carboxylate Facial Triad: A Versatile Platform for Dioxygen Activation by Mononuclear Non-Heme Iron(II) Enzymes. *JBIC J. Biol. Inorg. Chem.* **2005**, *10* (2), 87–93.

(42) Bruijnincx, P. C. A.; van Koten, G.; Klein Gebbink, R. J. M. Mononuclear Non-Heme Iron Enzymes with the 2-His-1-Carboxylate Facial Triad: Recent Developments in Enzymology and Modeling Studies. *Chem. Soc. Rev.* **2008**, *37* (12), 2716–2744.

(43) Parkin, G. Synthetic Analogues Relevant to the Structure and Function of Zinc Enzymes. *Chem. Rev.* **2004**, *104* (2), 699–768.

(44) Amrein, B.; Schmid, M.; Collet, G.; Cuniasse, P.; Gilardoni, F.; Seebeck, F. P.; Ward, T. R. Identification of Two-Histidines One-Carboxylate Binding Motifs in Proteins Amenable to Facial Coordination to Metals. *Metallomics* **2012**, *4*, 379–388.

(45) Bal, W.; Sokołowska, M.; Kurowska, E.; Faller, P. Binding of Transition Metal Ions to Albumin: Sites, Affinities and Rates. *Biochim. Biophys. Acta - Gen. Subj.* **2013**, *1830* (12), 5444–5455.

(46) Harford, C.; Sarkar, B. Amino Terminal Cu(II)- and Ni(II)-Binding (ATCUN) Motif of Proteins and Peptides: Metal Binding, DNA Cleavage, and Other Properties. *Acc. Chem. Res.* **1997**, *30* (3), 123–130.

(47) Bal, W.; Christodoulou, J.; Sadler, P. J.; Tucker, A. Multi-Metal Binding Site of Serum Albumin. *J. Inorg. Biochem.* **1998**, *70* (1), 33–39.

(48) Stewart, A. J.; Blindauer, C. A.; Berezenko, S.; Sleep, D.; Sadler, P. J. Interdomain Zinc Site on Human Albumin. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (7), 3701–3706.

(49) Garribba, E.; Sciortino, G.; Sanna, D.; Lubinu, G.; Maréchal, J.-D. Unveiling VIVO2+ Binding Modes to Human Serum Albumins by an Integrated Spectroscopic-Computational Approach. *Chem. – A Eur. J.* **2020**, *n/a* (n/a).

(50) Handing, K. B.; Shabalin, I. G.; Kassaar, O.; Khazaipoul, S.; Blindauer, C. A.; Stewart, A. J.; Chruszcz, M.; Minor, W. Circulatory Zinc Transport Is Controlled by Distinct Interdomain Sites on Mammalian Albumins. *Chem. Sci.* **2016**, *7* (11), 6635–6648.

(51) Bhattacharya, A. A.; Grüne, T.; Curry, S. Crystallographic Analysis Reveals Common Modes of Binding of Medium and Long-Chain Fatty Acids to Human Serum Albumin11Edited by R. Huber. *J. Mol. Biol.* **2000**, *303* (5), 721–732.

(52) Curry, S.; Brick, P.; Franks, N. P. Fatty Acid Binding to Human Serum Albumin: New Insights from Crystallographic Studies. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1999**, *1441* (2), 131–140.

(53) Lu, J.; Stewart, A. J.; Sadler, P. J.; Pinheiro, T. J. T.; Blindauer, C. A. Albumin as a Zinc Carrier: Properties of Its High-Affinity Zinc-Binding Site. *Biochem. Soc. Trans.* **2008**, *36* (6), 1317–1321.

(54) Blindauer, C. A.; Harvey, I.; Bunyan, K. E.; Stewart, A. J.; Sleep, D.; Harrison, D. J.; Berezenko, S.; Sadler, P. J. Structure, Properties, and Engineering of the Major Zinc Binding Site on Human Albumin. *J. Biol. Chem.* **2009**, *284* (34), 23116–23124.

(55) Coverdale, J. P. C.; Khazaipoul, S.; Arya, S.; Stewart, A. J.; Blindauer, C. A. Crosstalk between Zinc and Free Fatty Acids in Plasma. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **2019**, *1864* (4), 532–542.

(56) Li, Y.; Wang, Y.; Jiang, H.; Deng, J. Crystal Structure of Manduca Sexta Prophenoloxidase Provides Insights into the Mechanism of Type 3 Copper Enzymes. *Proc. Natl. Acad. Sci.* **2009**, *106* (40), 17002–17006.

(57) Fujieda, N.; Schätti, J.; Stuttfeld, E.; Ohkubo, K.; Maier, T.; Fukuzumi, S.; Ward, T. R. Enzyme Repurposing of a Hydrolase as an Emergent Peroxidase upon Metal Binding. *Chem. Sci.* **2015**, *6* (7), 4060–4065.

(58) Debret, G.; Martel, A.; Cuniasse, P. RASMOT-3D PRO: A 3D Motif Search Webserver. *Nucleic Acids Res.* **2009**, *37* (SUPPL. 2), W459–W464.

(59) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.

(60) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein–Ligand Docking Using GOLD. *Proteins Struct. Funct. Bioinforma.* **2003**, *52* (4), 609–623.

(61) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245* (1), 43–53.

(62) Rodríguez-Guerra Pedregal, J.; Sciortino, G.; Guasp, J.; Municoy, M.; Maréchal, J.-D. GaudiMM: A Modular Multi-Objective Platform for Molecular Modeling. *J. Comput. Chem.* **2017**, *38* (24), 2118–2126.