

Benchmarking Zinc-Binding Site Predictors: A Comparative Analysis of Structure-Based Approaches

Cosimo Ciofalo, Vincenzo Laveglia, Claudia Andreini,* and Antonio Rosato*



Cite This: *J. Chem. Inf. Model.* 2025, 65, 5205–5215



Read Online

ACCESS |



Metrics & More

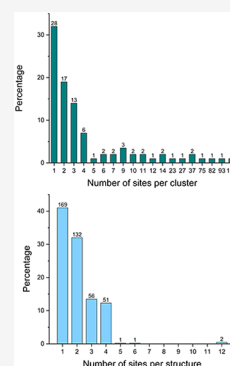


Article Recommendations



Supporting Information

ABSTRACT: Metalloproteins play crucial physiological roles across all domains of life, relying on metal ions for structural stability and catalytic activity. In recent years, computational approaches have emerged as powerful and increasingly reliable tools for predicting metal-binding sites in metalloproteins, enabling their application in the challenging field of metalloproteomics. Given the growing number of available tools, it is timely to design a reproducible approach to characterize their performance in specific usage scenarios. Thus, in this study, we selected some state-of-the-art structure-based predictors for zinc-binding sites and evaluated their performance on two data sets: experimental apoprotein structures and structural models generated by AlphaFold. Our results indicate that apoprotein structures pose significant challenges for predicting metal-binding sites. For these systems, the predictors achieved lower-than-expected performance due to the structural rearrangements occurring upon metalation. Conversely, predictions based on AlphaFold models yielded significantly better results, suggesting that they more closely resemble the holo forms of metalloproteins. Our findings highlight the great potential of metal-binding site predictions for advancing research in the field of metalloproteomics.



INTRODUCTION

Metalloproteins are a varied group of proteins that harbor metal ions as crucial structural components.¹ They are present in all forms of life and play a variety of physiological roles such as catalysis, electron transfer, oxygen transport, and gene regulation.² Metal ions can serve multiple functions in metalloproteins.³ They can be structural elements, stabilizing the protein, or catalytic components, activating substrates or stabilizing reaction intermediates. Metal ions can also transport electrons between redox-active sites, regulate protein activity, and transmit cellular signals.

Metalloproteomics is the study of metalloproteins at the whole organism or whole cell level.⁴ Due to the difficulties of experimental metalloproteomics, bioinformatics has emerged as an alternative method for mining metalloproteomes.⁵ In this context, 3D structure-based prediction of the presence of metal-binding sites, leveraging knowledge about the relative location in space of amino acids that may provide donor atoms for metal coordination, is an area of application that has received a lot of attention.⁶ The success of AlphaFold⁷ in the CASP programs⁸ has given a significant boost to such approaches, thanks to the extensive availability of viable 3D structural models for proteins not yet described experimentally. Notably, in 2024 the release of version 3 of AlphaFold included the possibility to predict the binding sites for some metal ions and metal-containing cofactors.⁹ Such predictors leverage the information on metal-binding sites extracted from the PDB. In principle, it would be relevant to ascertain whether any PDB site used for training is physiologically relevant or is an artifact due to experimental conditions, such as high metal concentration in the crystallization buffer. Previous analyses

have shown that nonphysiological sites tend to occur at the protein surface and to have metal coordination numbers lower than physiological sites.^{10–12} The number of protein residues in the second coordination sphere is also a useful indicator.¹¹ It is therefore possible to train predictors using data sets that have been cleansed of a significant share of nonphysiological sites; for the present work it is also relevant that the MetalPDB database provides extensive annotation regarding the physiological relevance of zinc sites.¹³ This has been indeed implemented in some 3D structure-based predictors of such sites.^{14,15}

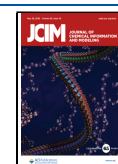
There are currently many structure-based predictors of metalloproteins available,^{14–24} with reported performance ranging from good to excellent. It is thus difficult for the user to determine which tool(s) are best suited to their needs. Consequently, we decided to benchmark some of the most recently published tools and use the results to provide an overview of the expected outcomes in typical application scenarios. To this end, we designed two different benchmarks. The first is based on the experimental structures of metalloproteins that lack their metal cofactor(s) (so-called apo-proteins). The second benchmark takes advantage of the high-quality structural models from the AlphaFold database.²⁵

Received: March 12, 2025

Revised: April 29, 2025

Accepted: May 8, 2025

Published: May 15, 2025



We focused on zinc-binding proteins, as most predictors have been trained on this subgroup of metalloproteins owing to their abundance in the Protein Data Bank. Our findings indicate that apoprotein structures are somewhat difficult for the predictors, who performed below expectations. When we used input structures from the AlphaFold database, however, we obtained excellent results for some predictors. This suggests that the identification of zinc metalloproteomes is indeed achievable with currently available tools.

METHODS

The methods section is divided into two parts to reflect the two distinct workflows employed in this study. The first part describes the benchmarking of state-of-the-art zinc-protein (ZnP hereafter) predictors using a data set of apo-proteins derived from MetalPDB.^{13,26} The second part details the application of these predictors, along with AlphaFold 3,⁹ for the prediction of zinc-binding sites in the *Saccharomyces cerevisiae* proteome.

Benchmarking of ZnP Predictors Based on Apo-Proteins. Data Set Creation. The predictors were tested using the apo structures of metalloproteins as input. An apoprotein is a protein that lacks its prosthetic group, such as a metal cofactor, while a holoprotein is a protein bound to its prosthetic group. All the data for constructing the data set for the benchmark was obtained from MetalPDB. This database groups collections of 3D templates with the same metal-binding site (MBS hereafter) into clusters. The MBS is defined by including any residue or chemical species having at least one atom within 5.0 Å of a metal ligand, where a metal ligand is a residue harboring any non-hydrogen atom within 3.0 Å from the metal. These clusters contain “equistructural” and “equivalent” sites. Two MBSs are defined as equivalent in MetalPDB if they satisfy all the following conditions: (i) they are found in PDB chains with the same structure (based on Pfam domain composition or on the sequence identity between the two chains being $\geq 50\%$); (ii) after structural superposition of the PDB chains, they are superimposed with the metal atoms in the same position; (iii) they contain the same metal(s). Condition (iii) is lifted in the definition of equistructural sites, consequently two equivalent sites are also equistructural but the vice versa may not be true, as two equistructural sites can harbor different metals.

We built the initial data set by including only clusters of equivalent mononuclear zinc sites (CLES) and their corresponding apo structures. From these clusters, we selected those that were physiologically relevant and contained only aspartate, cysteine, glutamate, and histidine as the metal-binding residues. The latter are the most common residues in zinc sites.^{26–28} The selection of physiological sites followed the criteria outlined by Laveglia et al.¹¹

The final data set for the benchmark included 87 CLES, containing 412 apo structures, for a total of 840 sites. The number of sites is greater than the number of apo structures because an apo structure can have multiple sites, with each site possibly belonging to a different CLES. This information is provided as [Supporting Table S1](#).

Predictors Used and Calculation Setup. We used several state-of-the-art metalloprotein predictors, such as BioMetAll,¹⁸ Metal1D (M1D),¹⁵ Metal3D (M3D),¹⁵ GASS,²¹ and Master of Metals (MoM).¹⁴ The latter was developed in our lab. Each of these predictors was selected based on their demonstrated effectiveness in previous studies and their ability to handle the

specific requirements of ZnP prediction. All are structure-based predictors, using homology to known structures or distance features to infer the location of metals. BioMetAll is a geometrical predictor that identifies MBSs based on backbone preorganization, it operates under the assumption that the geometric patterns of a protein's backbone encode sufficient information about its preorganization to coordinate metal ions. M1D is a distance-based predictor that infers the location of MBSs by utilizing coordination motifs extracted from the Protein Data Bank.²⁹ M3D is a deep learning-based predictor that operates on a voxelized representation of the protein environment and predicts metal density on a per-residue basis. GASS uses a parallel genetic algorithm to find candidate MBSs that are structurally similar to curated templates from the M-CSA³⁰ and MetalPDB²⁶ databases. MoM employs a neural network followed by a filter that compares the network output against the local structures of all known sites. This filter operates by comparing the distance matrices of the α and β atoms within the sites.

The parameters of each metalloprotein predictor were set as specified in their respective original articles to ensure the best possible output quality. For each predictor, we selected the top five outputs for every input based on its own ranking criteria. If the authors suggested a threshold for the score associated with their predictor's output, we discarded any predictions falling below that threshold. As a result, in some cases, less than five predictions were selected. This is further detailed below.

BioMetAll predicts MBSs by embedding a protein in a grid of fictitious metal probes, which are evaluated using geometric features. For each probe, the program identifies nearby amino acids that meet a set of predefined geometric criteria. For each output, we select the 5 predictions associated with the highest number of probes.

M1D scans all residues in the protein for compatibility with a computed probability map. Metals are positioned at the geometric center of residues that have high scores according to the map. The final ranking of the potential MBSs is then established using these probability scores. Thus, we selected the 5 predictions associated with the highest number of probability scores.

M3D processes a protein structure and a set of residues as input, converting the environment around each residue into a voxelized format. It then predicts the metal density for each residue individually. These per-residue predictions are averaged to produce an overall zinc density (ρ) for the entire protein. We retain the best five outputs with a ρ greater than or equal to 0.75.

GASS identifies MBS on protein structures using a parallel genetic algorithm and templates from the M-CSA and MetalPDB databases. GASS evaluates the effectiveness of potential binding sites using a fitness value, which measures the match between the candidate sites and the reference templates based on structural data. We selected the 5 predictions associated with the lowest value of fitness, where a lower fitness indicates a better result.

MoM identifies triads or quadruplets of amino acids with appropriate relative spatial arrangements from the PDB input using a machine learning algorithm. These identified structures are then ranked based on their structural similarity to a library of templates extracted from the MetalPDB database. The top 5 outputs with the best similarity scores values were selected.

Benchmarking Procedure. The benchmarking procedure involved using the apo structures as input and the

corresponding holo structures to determine the accuracy of the MBS predictions.

All procedures and data manipulation were conducted using homemade Python scripts. The jobs for all predictors were launched on the local cluster. M1D and M3D predict the binding location, so their output consists of the coordinates of the metal ion itself. In contrast, BioMetAll, MoM and GASS are binding site predictors and their output indicates the residues that bind the ion. In this benchmark, we used as the criterion for the validation of the output the ability to correctly predict the residues binding to the metal ion (so-called metal ligands). Therefore, to validate the outputs of M1D and M3D, we considered the residues within 3.8 Å of the metal position. This distance was chosen based on the statistics available for zinc sites in MetalPDB. The prediction of an individual metal site composed of n residues was considered a true positive (TP) if at least $n-1$ residues were correctly predicted; otherwise, it was considered a false positive (FP).

We conducted a second type of analysis to assess the capability of the predictors to globally predict a cluster of structures (i.e., a CLES). In this context, a CLES was considered identified if at least 50% of the sites within it were correctly predicted. This type of analysis aims to determine the sensitivity of the predictors to the rearrangement of side chains that can occur in apo sites. As extensively reported,³¹ donor atoms can experience significant changes in their relative positions when the metal is absent. To quantify how much this phenomenon is present in each CLES we defined a donor atom or donor atom proxy for each residue of the CHED group: for aspartate, we chose the $C\gamma$ atom; for cysteine, $S\gamma$; for glutamate, $C\delta$; and for histidine, $C\epsilon 1$. Once the "donors" were defined, we calculated the distance matrix for each apo site in a CLES using the distances between the donor atoms. All matrices are symmetric with zeros on the diagonal; for the same CLES, they all have the same $n \times n$ dimension, where n is the number of ligands in the site. These matrices were organized into a tensor of dimensions $N \times n \times n$, where N is the number of apo sites in the CLES. Considering only the upper part of the tensor, we calculated the standard deviation for each position on the vector along the N dimension. Then, by taking all the vectors for every position, we created a violin plot where each violin represents the distribution of distances between two donors within the CLES. Subsequently, we explored whether significant rearrangements of the side chains of individual residues could co-occur with similar shifts in other residues within the same site. For this investigation, we selected matrices from the tensor that exhibited at least two elements with values exceeding two standard deviations above the difference from the mean matrix. We then analyzed these identified sites to find connections between the side chain movements of residues within each site.

Evaluation of the Best Predictors on the *S. cerevisiae* Proteome. Data Set Creation. In this benchmark, we selected proteins from the *S. cerevisiae* proteome that do not have homologous metalloproteins with a deposited 3D structure in the PDB. To identify homologs, the BLAST software was used with an E-value threshold of 0.01.³² This approach allowed us to evaluate the performance of the algorithms, considering the structural diversity expected across a whole proteome. The selection of proteins without a structurally characterized homolog reduced the risk of introducing bias and provided a greater challenge in the prediction process, because all the predicted MBSs were in principle unprecedented sites. All the

input structures were obtained from the AlphaFold 2 database,²⁵ with at least 90% of their residues having a pLDDT > 0.7. Following the procedures outlined in our previous work,¹⁴ we further filtered the data set, resulting in 67 structures used to test whether the selected predictors agreed with each other in the identification of novel zinc-proteins. The UniProt identifiers of these proteins are given in [Supporting Table S2](#).

Predictors Used. In this second part of the work, MoM, M3D, and GASS, the three predictors that performed best in the first part of the benchmark, were used with the same settings already described, alongside AlphaFold 3 (AF3).⁹ AF3 is a deep-learning model designed to predict the 3D structure of biomolecular complexes, including proteins, nucleic acids, ligands, and metal ions. The results of AF3 were obtained using the proprietary web server.

Benchmarking Procedure. Given the lack of a reference holo structure to evaluate the predictions, the following approach was used: a protein was considered a ZnP if at least two predictors identified the same site. Notably, AF3 requires the identity and number of metals to predict as an input parameter. Since the number of requested sites can alter the protein structure output by AF3, the initial number of sites for each structure was set to 1. If the number of identified sites, based also on the output of the other three predictors, was greater than one, the AF3 prediction was repeated using as input the total number of predicted sites. The combined number of sites identified by all predictors in a single protein was three at most. For the construction of the confusion matrix, outputs from AF3 with a pLDDT < 0.7 for the predicted MBS, M3D with a ρ score < 0.75, and MoM with a d_{\min} > 0.35 were considered and classified as negative predictions, further categorized as either true negatives or false negatives. Conversely, outputs related to the identified sites in the input structure, ranked among the top five predictions from each predictor, were classified as true positives or false positives.

RESULTS

Description of the Benchmark. The first benchmarking procedure involved using apo structures as input and their corresponding holo structures to evaluate the accuracy of metal site predictions. The predictors used were BioMetAll, GASS,²¹ M1D,¹⁵ M3D,¹⁵ and MoM.¹⁸ The data set comprises 412 structures containing 840 zinc(II) sites in their apo form, organized into 87 clusters (CLESs). The number of sites exceeds the number of apo structures because a single structure can include multiple sites, potentially belonging to different CLESs. The methodology used to construct the clusters is detailed in the [Methods Section](#). [Figure 1](#) shows the distribution of the number of sites across the structures and the distribution of the number of sites within the CLESs. 73% of the apo structures in the data set contain either one or two sites; 48% percent of the total sites belong to the four largest CLESs.

Given the use of experimental apo structures, a qualitative investigation was conducted to assess the potential spread of distances between donor atoms of the side chains in equivalent sites across different structures. To visualize this, we generated a violin plot for each CLES. The graphical results indicate that this phenomenon is present in many CLESs, generally becoming more pronounced as the number of sites within the CLES increases. The four largest CLESs show the most

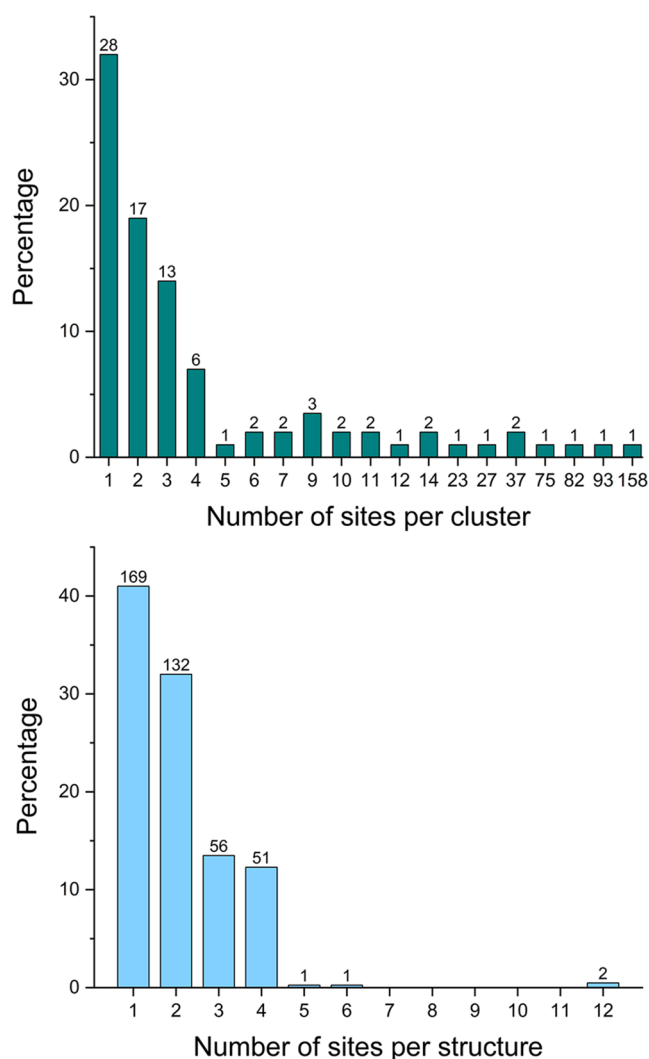


Figure 1. Percentage distribution of sites across CLEs and structures. (Top) The histogram shows the percentage distribution of the number of sites across CLEs. The number above each bar indicates how many CLEs have the corresponding number of sites. (Bottom) The histogram shows the percentage distribution of the number of sites per structure. The number above each bar indicates how many structures harbor the corresponding number of sites.

appreciable spread. We then explored whether significant rearrangements of the side chains of individual residues were associated with shifts in other residues within the same site. Our analysis revealed a notable co-occurrence between the shifts of the side chains within a site, suggesting that the rearrangement of one residue often aligns with coordinated movements of other residues in the same binding site. For example, in a site composed of three residues, if one residue deviates significantly from the average distance, it naturally causes two distances to fall outside the average, however, we frequently observed that all three residues deviate together, resulting in all distances being outside the average.

Details on how the values for the plots were calculated are provided in the [Methods Section](#). [Figure 2](#) shows the violin plot for a CLEs composed of 31 sites with a Cys-Asp-His binding pattern. The corresponding plots for all CLEs are provided as [Supporting Figures S1–S15](#).

Evaluation of the Predictions. [Table 1](#) presents the results of the benchmark based on the number of predictions

considered for each predictor for each input structure. Each predictor was evaluated based on its top 1, top 3, top 5, and top 10 outputs. The number of sites used to compute the recall was determined by the number of available binding sites in each structure. For instance, for the evaluation of the top 1 outputs, only one site per structure was considered; for the top 3, a maximum of three sites were considered in structures with at least that many known sites, and so on. Hereafter, we will describe the results obtained by considering the top 5 outputs from each predictor for every input apo structure, this approach provides the optimal balance between the total number of sites and the number of predictions evaluated for each structure.

Before going into the detailed analysis of the performance of the tools, it is important to point out that only GASS consistently provided an output for 411 input structures, with only input structure causing the program to crash. At the other extreme, M3D and MoM provided an output for only 239 and 238 structures, respectively, corresponding to 54% of the data set. This is due to either all the candidate sites following below the set thresholds or, in a limited number of instances, the programs crashing. Since the GASS authors did not specifically suggest a threshold for its outputs, we gathered all of them. [Figure 3](#) shows the percentages of true positives (TP), redundant true positives (RTP), false positives (FP) and the Recall for all predictors. The percentages of TP, RTP, and FP are calculated with respect to the total number of site predictions, while the Recall is based on the total number of sites in the data set. BioMetAll and M1D exhibit a high number of false positives, and they provide correct predictions for less than 20% of the sites in the database. GASS, M3D and MoM generally achieve better results. GASS achieves the best result in terms of Recall, due to its ability to produce outputs for all input structures as noted above, but generates many more false positives than M3D and MoM. The latter two predictors are generally more balanced than GASS. Among them, MoM achieves the highest percentage of true positives, even when excluding the RTP. Thus, when looking at all output predictions, MoM and M3D have comparable precision, between 50% and 60%, somewhat better than the 30% achieved by GASS.

In order to understand the results of the tools at the level of protein families rather than individual MBS, [Figure 4](#) shows the recall of the predictors across all CLEs. GASS, M3D, and MoM are the top performers. GASS achieves the best results overall (59% of CLEs identified), which may be due to the high number of predictions it generates compared to the other two predictors. Importantly, the recall of MoM is relatively close (54%), owing to its significantly higher precision.

Impact of Structural Diversity within Families. We selected the top-performing predictors from the previous evaluation, MoM, M3D, and GASS, to investigate their ability to cope with the spread in side chain orientation ([Figure 2](#)) within a protein family. By focusing on the CLEs that, based on the violin plots, appear to be particularly affected by this phenomenon, we observed that MoM and GASS are quite tolerant toward structural rearrangements, whereas M3D is impacted by the phenomenon to a larger extent ([Figure 5A](#)). [Figure 5B,C](#) illustrates how for an apo-site with four Cys residues, these side chains can be spatially close in some structures but point away from each other in other members of the CLEs. For this example, the results of GASS and MoM are indeed far better than M3D ([Figure 5A](#), second cluster).

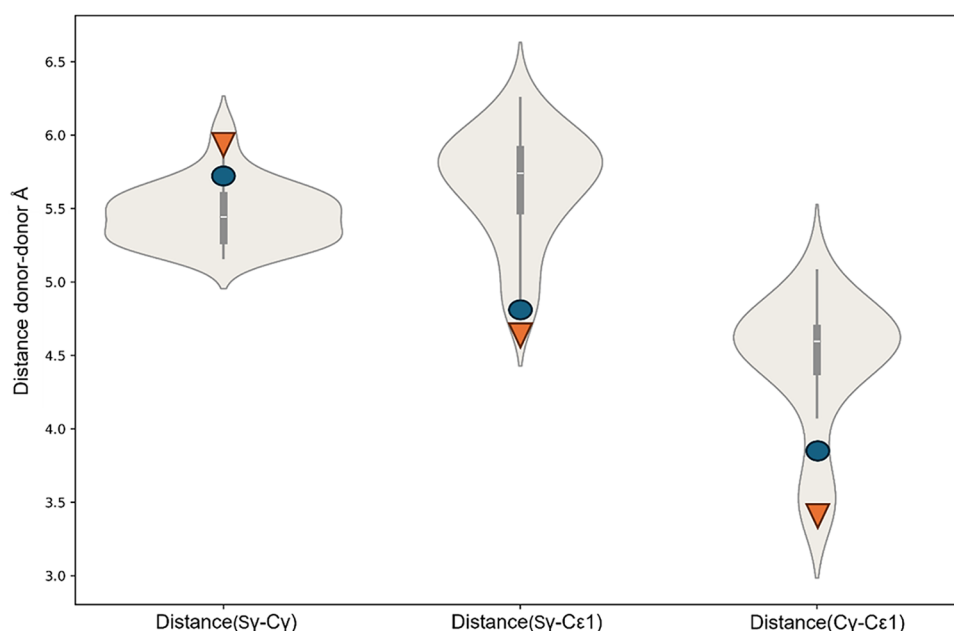


Figure 2. Violin plot for CLES 22022, composed of 31 sites. Each violin represents the distribution of the distance between two selected donor atoms in the binding site. The white line at the center of each violin indicates the median value, while the bars around it mark the first and third quartiles. The circles and the triangles of the same color correspond to the sites whose distance matrices contain at least two elements with values beyond two standard deviations from the mean. The site of this cluster consists of 1 cysteine ($S\gamma$), 1 aspartate ($C\gamma$) and 1 histidine ($C\epsilon 1$), where $S\gamma$, $C\gamma$ and $C\epsilon 1$ are the donor atom or donor atom proxy.

Table 1. Benchmark Results for Each Predictor^a

evaluated predictions	predictor	N. outputs	N. true positives	N. redundant	N. false positives	precision (%)	recall (%)
top 1	M1D	385	91	0	294	23.6	21.6
	BioMetAll	338	52	0	286	15.4	12.1
	GASS	411	143	0	268	34.8	34.5
	MoM	238	221	0	17	92.8	53.4
	M3D	239	154	0	85	64.4	36.9
top 3	M1D	759	126	2	633	16.6	16.2
	BioMetAll	1013	159	57	854	15.7	13.3
	GASS	1206	416	106	790	34.5	40.5
	MoM	521	467	134	54	89.6	43.5
	M3D	455	260	26	195	57.1	30.5
top 5	M1D	844	129	2	715	15.3	15.4
	BioMetAll	1687	260	116	1427	15.4	17.5
	GASS	1981	626	236	1355	31.6	47.3
	MoM	693	609	237	84	87.9	45.1
	M3D	503	272	27	231	54.1	29.7
top 10	M1D	854	129	2	725	15.1	15.2
	BioMetAll	4030	654	381	3376	16.2	32.7
	GASS	3897	1061	576	2836	27.2	58.0
	MoM	868	726	331	142	83.6	47.2
	M3D	527	279	28	248	52.9	30.0

^aThe columns in the table, from left to right, report the number of predictions included in the analysis for each predictor, the name of the predictor used (Predictor), the total number of outputs generated by the predictor for the benchmark (N. Outputs), the number of correct predictions (N. True Positives), the number of repeated TP predictions for the same site (N. Redundant), the number of incorrect predictions (N. False Positives), the percentage of TPs over the total predictions including redundant sites (Precision %), and the percentage of TPs over the total known binding sites for that structure (Recall %)

Impact of the Presence of Multiple Metal-Binding Sites and of Redundancy in the Predictions. A specific aspect of metalloproteins is that each structure may contain multiple true MBSs. In fact, our benchmark contained on average 2 sites per structure, going from 1 to 12; 38% of the structures contained a single site (Figure 1). In real life the number of MBS is not known for a novel protein, and thus for

an user it would be relevant to understand the behavior of the tools when taking into account a fixed number of predictions for each structure in order to assess whether the predictions of multiple sites in a structure are all equally reliable. We performed this analysis only for the three best predictors (GASS, M3D, MoM). Importantly, if the fixed number of predictions analyzed is lower than the number of real sites in

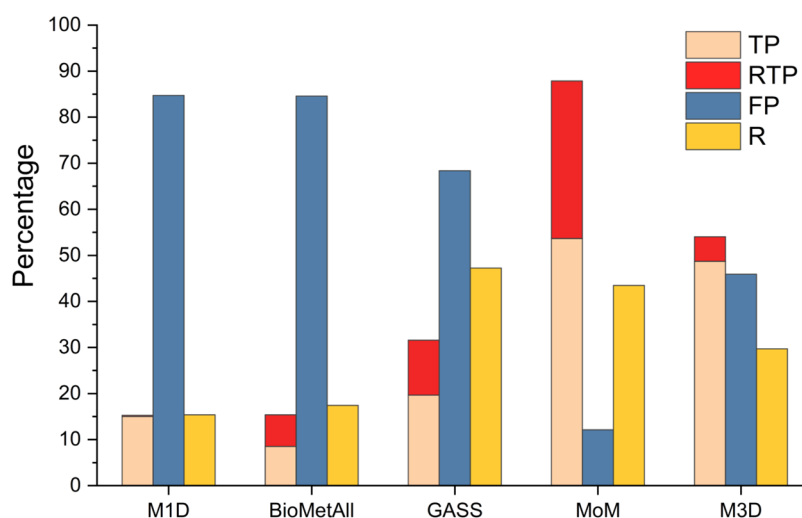


Figure 3. Performance of the predictors, based on the top 5 outputs. True positive (TP), redundant true positive (RTP), false positive (FP) and recall (R) values are displayed for all the predictors.

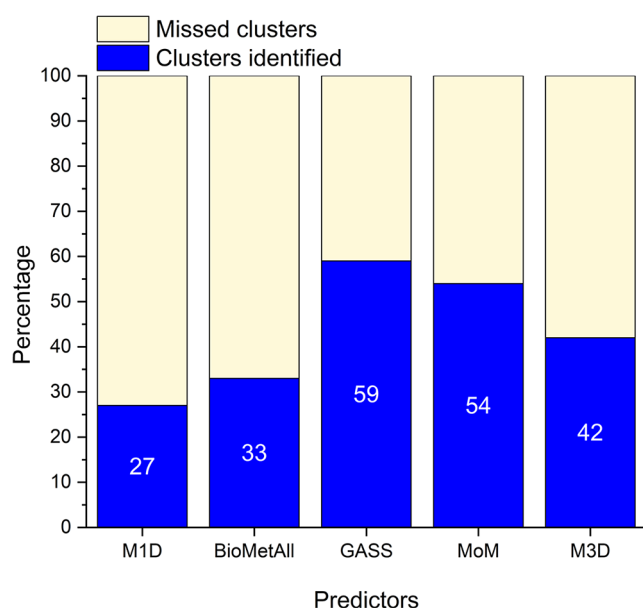


Figure 4. Percentage of CLEs identified by the predictors.

the structure, then the traditional recall (as computed for Figure 3) will always be lower than 100%, introducing a biased view. To provide a fairer evaluation of recall in light of these considerations, we introduced the following formulas

$$R(\text{top1}) = \frac{TP}{N_1^{\text{sites}} + \frac{1}{2}N_2^{\text{sites}} + \frac{1}{3}N_3^{\text{sites}} + \frac{1}{4}N_4^{\text{sites}} + \dots} \quad (1)$$

$$R(\text{top2}) = \frac{TP}{N_1^{\text{sites}} + N_2^{\text{sites}} + \frac{2}{3}N_3^{\text{sites}} + \frac{2}{4}N_4^{\text{sites}} + \dots} \quad (2)$$

$$R(\text{top3}) = \frac{TP}{N_1^{\text{sites}} + N_2^{\text{sites}} + N_3^{\text{sites}} + \frac{3}{4}N_4^{\text{sites}} + \dots} \quad (3)$$

where N_1 is the number of sites in structures with only one site, N_2 is the number of sites in structures with two sites, etc. In practice the denominator of the fractions in eqs 1–3

corresponds to the total number of sites that can be identified given the number of predictions examined.

To define these, we reasoned that if only the prediction with the highest score is considered (Top1), the structures harboring multiple sites would be penalized by the usual Recall equation, because only one site could be predicted. Thus, in Top1 we consider the prediction for a multisite structure successful when any of its sites is correctly identified. With this approach, the number of predictable sites for a two-site structure is effectively halved, whereas for a three-site structure the number of predictable sites is one-third, and so on (eq 1). When computing Top2, all the sites in mono- and two-site structures are predictable, whereas this is true only for two-thirds of the sites in three-site structures (eq 2). This reasoning can be extended to the analysis of any number of output predictions. Note that with monosite structures, it is possible to generate only one correct prediction. An ideal tool thus should output only one prediction with a score better than the threshold. Any additional nonredundant prediction is a FP by definition and thus contributes to decreasing Precision. With the definition of eqs 1–3, the denominator of the fraction increases with increasing number of predictions evaluated. This makes it possible for Recall(Topm) to decrease upon inclusion of additional predictions in the evaluation, contrary to its commonly observed behavior in binary classification applications. On the other hand, Precision can be computed according to the usual formula, as it corresponds to the percentage of predictions that are correct. For this parameter, predictions that are redundant with respect to a TP can be regarded as TPs as well. The results obtained are shown in Figure 6.

The Rtop measure ranges between 43 and 53% for MoM, and between 30 and 37% for M3D, with a similar trend as a function of the number of predictions used. The trend is completely different for GASS, whose Rtop increases almost linearly with increasing number of predictions, going from slightly above 34% up to 58% for ten predictions. The R top(5) value is similar for MoM and GASS, in agreement with Figure 3. For all predictors, precision decreases by approximately 10% when moving from the top 1 to the top 10 outputs; however MoM had a precision significantly higher than the other programs, with a maximum value of 92% when considering

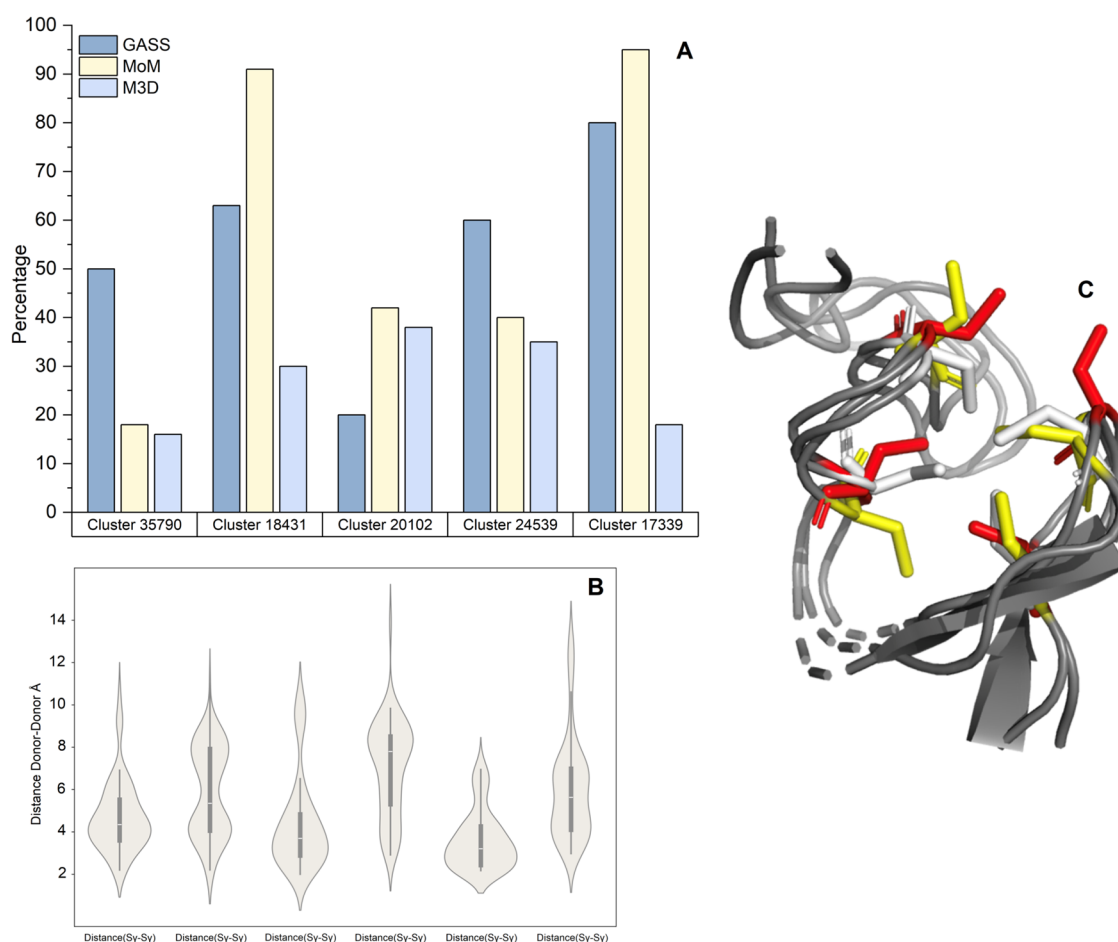


Figure 5. (A) Percentage of correctly predicted sites in CLEs featuring large rearrangements of the MBS. The performance of the three best predictors, each shown in a different color, is computed as the percentage of correctly identified sites over the structures in each CLE. Five different CLEs characterized by a high spread of side chains were selected. (B) Violin plot of CLES 18431. The site of this cluster consists of 4 cysteines, whose donor atoms are the Sy atoms. (C) View of the 3D structure of the sites in CLES 18431. This panel shows the spread of side chains for three different apo-sites belonging to CLES 18431. Each apo-site is represented by a different color.

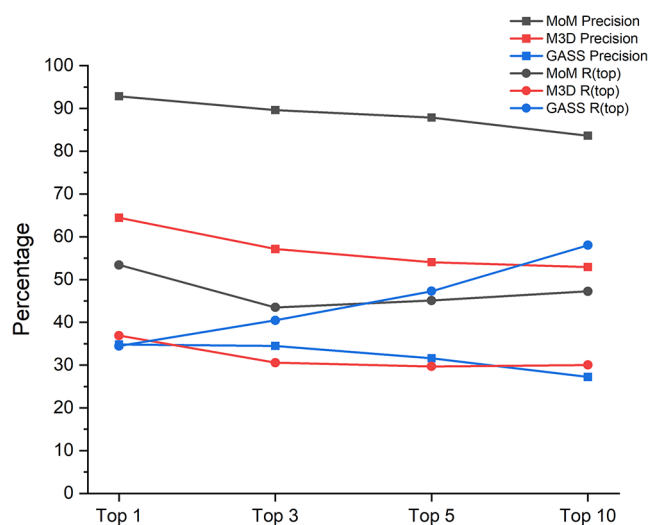


Figure 6. Precision and recall (R(top)) of the selected predictors as a function of the number of outputs considered.

only the first output with respect to about 65% for M3D and 35% for GASS.

In this work, it became apparent that the various predictors could output, albeit to different extents, a number of slightly

different predictions for the same site in each structure. To assess the impact of this, we repeated the above analysis by removing all redundant predictions. To do so, for all the correct predictions with $n-1$ ligands in common (where n is the number of ligands in the known holo-site), which are redundant TPs, we kept only the best-scoring one. The results are shown in Figure 7.

This procedure mimics a user who would consider all the predictions involving the same residues but one as identifying the same site. With this approach, the Rtop measure ranges between 45 and 55% for MoM, and between 30 and 40% for M3D, with a similar trend as a function of the number of predictions used. GASS maintains a significantly different trend with Rtop increasing almost linearly from 35% up to almost 60% when including the best 10 predictions in the analysis. In terms of precision, while the situation for the best prediction is unchanged (as there are no redundancies), we can observe a somewhat sharper decline when redundant predictions are excluded for MoM as well as GASS, whereas this change is less pronounced for M3D. In fact, the latter tool is one that produces the lowest number of redundant predictions on average.

The three tools on which we focused featured somewhat different values and trends of precision and recall as a function of the number of predictions included in the analysis. To

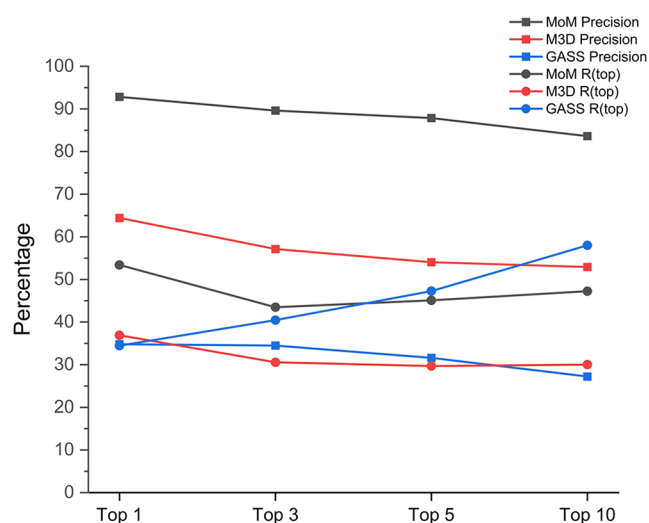


Figure 7. Precision and recall (R (top)) of the selected predictors as a function of the number of nonredundant outputs.

simplify the overall view of these results, the F-score can be used. Here, we used the F1-score, which is the harmonic mean of precision and recall, to represent the results after removing redundancies (Figure 8).

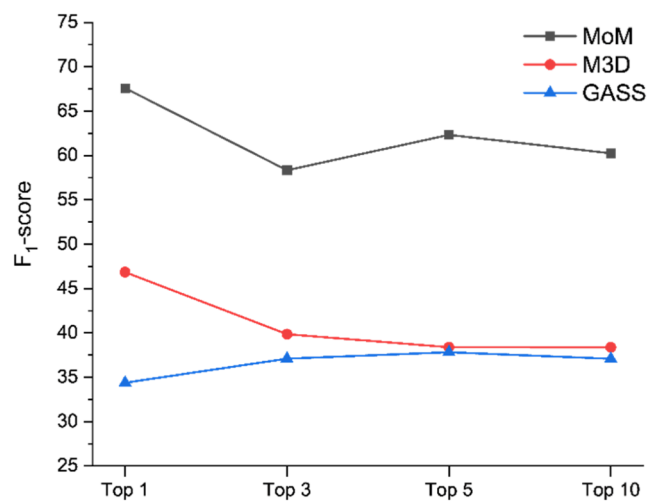


Figure 8. F1-score of the selected predictors as a function of the number of considered outputs. The F1-score has been calculated using the data of Figure 7.

It can be seen that the three programs have similar trends, as the F1-score decreases with increasing number of predictions analyzed. For MoM and M3D, which are the two best-performing tools based on the F1-score, there is a larger drop when extending from 1 to 3 predictions due to the larger number of true sites that ideally should be identified, which adversely affects both their precision and recall. The F1-score of GASS is systematically lower than the other two tools. The F1-score of MoM goes from 68 to 46% with increasing number of predictions, whereas M3D goes from 57 to 38%.

Predictions for AF Models. In the second benchmark, we selected as input all the proteins from the *S. cerevisiae* proteome that do not have homologous proteins with a deposited 3D structure in the PDB. For this benchmarking procedure we lack a reference holo structure to evaluate the

predictions. Thus, an input protein was considered a metalloprotein if at least two predictors identified the same site. The predictors used in this analysis were MoM, M3D, and GASS, along with AlphaFold 3. For these proteins, model structures were downloaded from the AlphaFold database and filtered on the basis of the average pLDDT over the entire model. A total of 62 sites were identified in the 67 target structures. It is important to note that, unlike the previous benchmark, this analysis also includes true negatives (TN), as some structures were found not to be metalloproteins according to the criterion described above. Of these 62 sites, 27 (44%) were identified by all four predictors, whereas 21 (34%) were predicted by three tools (Figure 9). Eight (13%) sites were identified by only MoM and GASS. The agreement between pairs of tools, excluding AF3, was in the range 69–75%, hence similar for all possible pairs.

From the confusion matrix (Table 2) and from Figure 10, we can observe a very satisfactory outcome for all predictors, with M3D and MoM achieving the highest precision (around 90%), whereas GASS and AF3 had similar results (around 70%). GASS, on the other hand, had the highest recall at about 89%, followed by MoM (84%), M3D (81%) and last AF3 at only 68%. As can be seen from the previously shown results, M3D and MoM achieve the best performance in this case as well, with M3D standing out from the others by achieving the highest precision value.

DISCUSSION

As the first step of our work, we benchmarked a number of structure-based predictors of MBS against experimental structures of metalloproteins in their apo form. Structural rearrangements occur upon metalation in apo vs holo proteins, leading to potential differences in the orientation of the side chains at the metal site.³¹ In a cluster (CLES) of apo structures of the same MBS, multiple local energy minima may arise during the crystallization process. This is typically not observed in holo sites, as metal binding imposes order through the geometric constraints of coordination.³³ The analysis of the CLEs that feature pronounced structural rearrangements revealed that the rearrangement of one residue often is accompanied by movements of other residues within the same binding site. In CLEs with enough sites (typically ≥ 4), this phenomenon becomes quite evident, so that many of the sites exhibiting significant rearrangement can be grouped into subclusters based on the distances between donor atoms. These subclusters represent some of the various relative energy minima that may arise during the crystallization process.

Our results (Figure 3) show that the performance observed here is worse than that reported in the predictors' respective articles. We can ascribe this to apo structures being a particularly challenging input, owing to their aforementioned structural variations with respect to the holo structure. This observation aligns with the per-CLES performance analysis of Figure 4, which shows that the three best-performing tools can identify only 40 to 60% of the CLEs. The difficulty of this data set is indirectly confirmed by the results obtained in the second part of this study (Figure 9), where all input structures were generated by AlphaFold 2. In this case, the performance was very good and consistent with the results reported in the literature. This is likely because AlphaFold 2 tends to orient the residues in the sites in a configuration relatively close to that of the holo-site,³⁴ thus reducing the inherent challenges of apo structures. Notably, in the related task of predicting drug

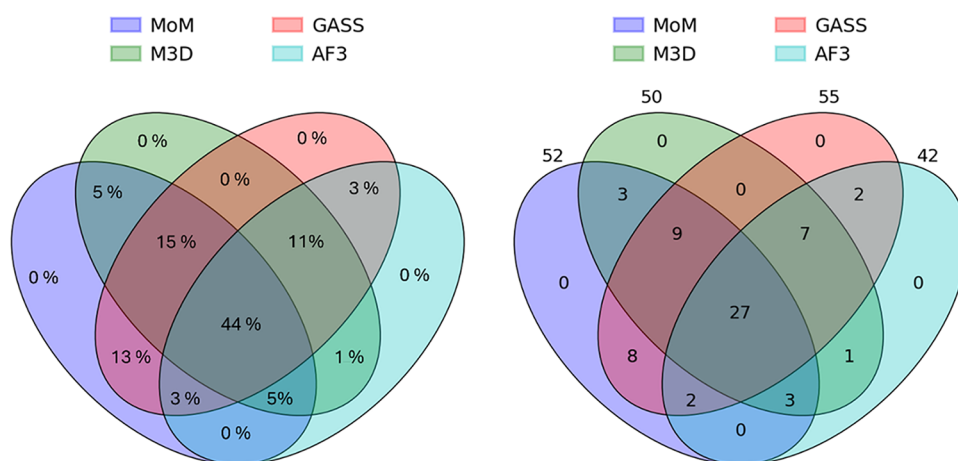


Figure 9. Venn diagram representing the agreement of the predictors for the second benchmark. The diagrams represent the intersections of the true positives (TP) identified by the four tools. The percentages in the left panel were computed with respect to the total of 62 sites identified by at least two predictors.

Table 2. Confusion Matrix for All Predictors for AF Models

predictor	GASS	MoM	M3D	AF3
TP	55	52	50	42
FP	24	7	4	17
TN	0	12	14	8
FN	0	8	13	12
recall (%)	88.7	83.8	80.6	67.7
precision (%)	69.6	88.1	92.5	71.1

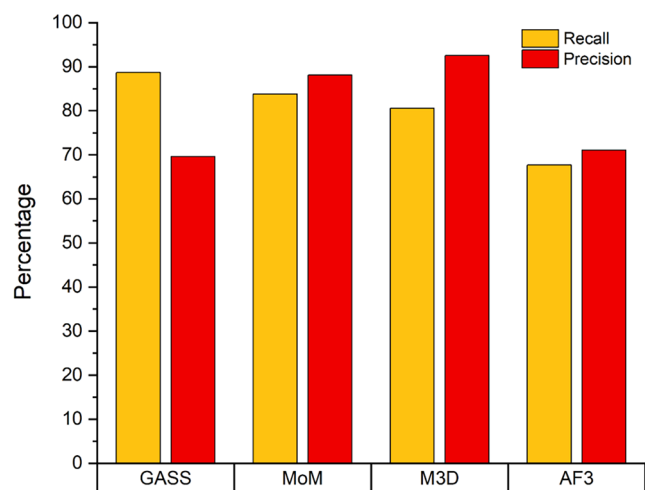


Figure 10. Recall compared with precision for all the predictors. Recall is defined as the ratio between TP and the total number of expected sites in the data set.

binding modes, AlphaFold models performed comparably to traditional homology models despite their closer similarity to experimental structures.³⁵ It is thus possible that zinc(II) ions constitute a particularly favorable example.

All the predictors are structure-based predictors, but each employs a different criterion for identifying potential MBS within an input structure (see the [Methods Section](#)). We observed (Figure 5) that M3D has limitations in accurately identifying sites within all the subclusters of the CLEs that feature a pronounced spread of side chain orientations. MoM and GASS have a better performance under these challenging conditions. The different performance of the predictors can be

attributed to the strategy they use for site prediction. MoM and GASS rely on the distance matrix between the α and β carbons of the residues in the potential MBS as their primary predictive parameter. This approach makes them less sensitive to variations in side chain orientation. In contrast, M3D considers the positions of all the atoms within the side chains, which makes it more susceptible to the effects of differing side chain orientations within a potential MBS. It is evident that a large variability between the apo sites within a CLES necessarily implies that certain subsets of distances will deviate significantly from those observed in the metalated form of the site. On the other hand, the global performance of M3D is intermediate between MoM and GASS, indicating that M3D performs very well when the distances and orientations of the residues are close to the holo form.

MoM and M3D tend to generate far fewer predictions than GASS for each input structure, based on the settings specified in the respective articles. Each predictor employs a specific metric to assess the reliability of its outputs; therefore, we investigated whether such a metric can be used meaningfully to rank the results, i.e., as a scoring function. With this approach, we observed that the performance of M3D and MoM is relatively independent of the number of predictions used in the analysis, whereas GASS featured a very strong increase of its recall (R (top) values) with increasing number of predictions analyzed. When only the best prediction is taken into account, GASS is outperformed by the M3D and MoM in terms of both recall and precision. When including three predictions in the analysis, the R (top) of GASS improves significantly, whereas it drops for the other tools. This can be ascribed to the scoring function of GASS being poorly informative, so that the correct predictions are not necessarily ranked best. This trend becomes more evident for the analysis of up to 10 predictions, when GASS achieves the best R (top) of all tools, along with a very poor precision. A combined view of these two parameters is afforded by the F1-score. The latter indicates that providing a limited number of predictions ranked by a reliable scoring function, as done by M3D and MoM, is the best approach. This is indeed also the situation that a user is most likely to appreciate, as it reduces the effort required to interpret the output predictions.

Utilizing computational models of the 3D structure as input is a likely scenario for the use of the predictors discussed here,

especially considering AlphaFold's relatively recent success. We therefore addressed this aspect by investigating a selection of 67 proteins from yeast that have no homologs with an experimental structure deposited in the PDB. As we do not know whether these proteins harbor an MBS, a consensus approach was taken to identify the "true" metalloproteins in this group, leading to an ensemble of 52 proposed ZnPs. In this analysis, AlphaFold 3 was used as well, taking advantage of its ability to predict metal sites, which was not possible with AlphaFold 2. 78% of the sites have been identified by at least three predictors, which appears to be a good consistency. Based on the false positive rates reported in the articles of GASS, M3D and MoM, which are all in the range 10–20%, we would expect that predictions where only two tools are in agreement would have an error rate of less than 5%, whereas the agreement of three tools would correspond to an error rate of about 1%. The pairwise agreement between pairs of predictors was fairly similar, close to or slightly above 70%, suggesting that indeed their radically different designs make the consensus approach meaningful. In practice, based on our second benchmark M3D and MoM have similar performance; the performance of GASS was slightly inferior, especially in terms of precision, whereas that of AF3 was even lower. Notably, M3D and MoM obtained recall and precision values in line with those reported in the respective publications, suggesting that AlphaFold models computed in the absence of the cofactor are a better representation of the holo form of a metalloprotein than the experimental structure of the apo form. Although this is a preliminary observation, due to the small size of the present data set, it is in line with previous observations e.g., for the bulkier iron–sulfur cofactors.³⁶ In fact, for the latter systems the effect might be even more dramatic, owing to the large structural changes and increased protein dynamics upon loss of the cofactor.^{37,38}

When using AlphaFold models as input, M3D and MoM reach precision values around 90%, which indicates that a computational pipeline that integrates high-quality modeling by AlphaFold with specialized predictors can enable the prediction of entire metalloproteomes. It is also noteworthy that M3D and MoM perform similarly despite the fundamentally different computational approaches they implemented. The use of consensus approaches involving these tools (and others of similar quality) should therefore further increase the reliability of the results. All the tools analyzed in this work are freely downloadable and can be installed with relative ease; a web interface is also available for all of them, except for MoM.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.5c00549>.

List of the apo structures and the corresponding sites used for the benchmark, grouped by CLES number (Table S1); complete list of the selected proteins from the *S. cerevisiae* proteome that do not have homologous metalloproteins with a deposited 3D structure in the PDB, constituting the second benchmark (Table S2); violin plots of the 15 most populated CLESs (Figures S1–S15) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Claudia Andreini – Department of Chemistry, University of Florence, Sesto Fiorentino 50019, Italy; Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino 50019, Italy; Consorzio Interuniversitario di Risonanze Magnetiche di Metallo Proteine, Sesto Fiorentino 50019, Italy; Email: andreini@cerm.unifi.it

Antonio Rosato – Department of Chemistry, University of Florence, Sesto Fiorentino 50019, Italy; Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino 50019, Italy; Consorzio Interuniversitario di Risonanze Magnetiche di Metallo Proteine, Sesto Fiorentino 50019, Italy; orcid.org/0000-0001-6172-0368; Email: antonio.rosato@unifi.it

Authors

Cosimo Ciofalo – Department of Chemistry, University of Florence, Sesto Fiorentino 50019, Italy; Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino 50019, Italy; orcid.org/0009-0006-0371-1280

Vincenzo Laveglia – Department of Chemistry, University of Florence, Sesto Fiorentino 50019, Italy

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jcim.5c00549>

Author Contributions

C.C., C.A., and A.R. designed the research. C.C., V.L., and C.A. curated the data. C.C. carried out the research and analyzed the results. C.C., V.L., C.A., and A.R. wrote, reviewed, and edited the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support from the projects Potentiating the Italian Capacity for Structural Biology Services in Instruct-ERIC, Acronym "ITACA.SB" (Project no. IR0000009) within the call MUR 3264/2021 PNRR M4/C2/L3.1.1 funded by the European Union—NextGenerationEU, and W-BioCat funded by the European Commission no. 101129798 (to C.A.) is gratefully acknowledged. We thank Dr. Sandro Izidoro for helping us with GASS calculations.

■ REFERENCES

- (1) Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J. Metalloproteins and Metal Sensing. *Nature* **2009**, *460* (7257), 823–830.
- (2) Andreini, C.; Banci, L.; Bertini, I.; Rosato, A. Zinc through the Three Domains of Life. *J. Proteome Res.* **2006**, *5* (11), 3173–3178.
- (3) Harding, M. M. The Architecture of Metal Coordination Groups in Proteins. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60* (5), 849–859.
- (4) Jakubowski, N.; Lobinski, R.; Moens, L. Metallobiomolecules. The Basis of Life, the Challenge of Atomic Spectroscopy. *J. Anal. At. Spectrom.* **2004**, *19*, 1–4.
- (5) Zhang, Y.; Gladyshev, V. N. Comparative Genomics of Trace Element Dependence in Biology. *J. Biol. Chem.* **2011**, *286* (27), 23623–23629.
- (6) Andreini, C.; Rosato, A. Structural Bioinformatics and Deep Learning of Metalloproteins: Recent Advances and Applications. *Int. J. Mol. Sci.* **2022**, *23* (14), No. 7684.
- (7) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A. W. R.; Bridgland, A.;

- Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706–710.
- (8) Kryshchuk, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moulton, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIII. *Protein: Struct., Funct., Bioinf.* **2019**, *87* (12), 1011–1020.
- (9) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Židek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500.
- (10) Laitaoja, M.; Valjakka, J.; Jänis, J. Zinc Coordination Spheres in Protein Structures. *Inorg. Chem.* **2013**, *52* (19), 10983–10991.
- (11) Laveglia, V.; Giachetti, A.; Sala, D.; Andreini, C.; Rosato, A. Learning to Identify Physiological and Adventitious Metal-Binding Sites in the Three-Dimensional Structures of Proteins by Following the Hints of a Deep Neural Network. *J. Chem. Inf. Model.* **2022**, *62* (12), 2951–2960.
- (12) Yao, S.; Moseley, H. N. B. Finding High-Quality Metal Ion-Centric Regions Across the Worldwide Protein Data Bank. *Molecules* **2019**, *24* (17), No. 3179.
- (13) 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* (D1), D459–D464.
- (14) Laveglia, V.; Bazayeva, M.; Andreini, C.; Rosato, A. Hunting down Zinc(II)-Binding Sites in Proteins with Distance Matrices. *Bioinformatics* **2023**, *39* (11), No. btad653.
- (15) Dürr, S. L.; Levy, A.; Rothlisberger, U. Metal3D: A General Deep Learning Framework for Accurate Metal Ion Location Prediction in Proteins. *Nat. Commun.* **2023**, *14* (1), No. 2713.
- (16) Hu, X.; Dong, Q.; Yang, J.; Zhang, Y. Recognizing Metal and Acid Radical Ion-Binding Sites by Integrating *Ab Initio* Modeling with Template-Based Transfers. *Bioinformatics* **2016**, *32* (21), 3260–3269.
- (17) 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.
- (18) Sánchez-Aparicio, J.-E.; Tiessler-Sala, L.; Velasco-Carneros, L.; Roldán-Martín, L.; Sciortino, G.; Maréchal, J.-D. BioMetAll: Identifying Metal-Binding Sites in Proteins from Backbone Preorganization. *J. Chem. Inf. Model.* **2021**, *61* (1), 311–323.
- (19) Ireland, S. M.; Martin, A. C. R. Zincbindpredict—Prediction of Zinc Binding Sites in Proteins. *Molecules* **2021**, *26* (4), No. 966.
- (20) Lu, C.-H.; Chen, C.-C.; Yu, C.-S.; Liu, Y.-Y.; Liu, J.-J.; Wei, S.-T.; Lin, Y.-F. MIB2: Metal Ion-Binding Site Prediction and Modeling Server. *Bioinformatics* **2022**, *38* (18), 4428–4429.
- (21) Paiva, V. A.; Mendonça, M. V.; Silveira, S. A.; Ascher, D. B.; Pires, D. E. V.; Izidoro, S. C. GASS-Metal: Identifying Metal-Binding Sites on Protein Structures Using Genetic Algorithms. *Briefings Bioinf.* **2022**, *23* (5), No. bbac178.
- (22) Krapp, L. F.; Abriata, L. A.; Cortés Rodríguez, F.; Dal Peraro, M. PeSTo: Parameter-Free Geometric Deep Learning for Accurate Prediction of Protein Binding Interfaces. *Nat. Commun.* **2023**, *14* (1), No. 2175.
- (23) Essien, C.; Wang, N.; Yu, Y.; Alqarghuli, S.; Qin, Y.; Manshour, N.; He, F.; Xu, D. Predicting the Location of Coordinated Metal Ion-Ligand Binding Sites Using Geometry-Aware Graph Neural Networks. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 137–148.
- (24) Dai, X.; Henderson, M.; Yoo, S.; Liu, Q. Predicting Metal-Binding Proteins and Structures Through Integration of Evolutionary-Scale and Physics-Based Modeling. *J. Mol. Biol.* **2025**, *437* (6), No. 168962.
- (25) Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.; Yeo, J.; Kovalevskiy, O.; Tunyasuvunakool, K.; Laydon, A.; Židek, A.; Tomlinson, H.; Hariharan, D.; Abrahamson, J.; Green, T.; Jumper, J.; Birney, E.; Steinegger, M.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database in 2024: Providing Structure Coverage for over 214 Million Protein Sequences. *Nucleic Acids Res.* **2024**, *52* (D1), D368–D375.
- (26) Andreini, C.; Cavallaro, G.; Lorenzini, S.; Rosato, A. MetalPDB: A Database of Metal Sites in Biological Macromolecular Structures. *Nucleic Acids Res.* **2012**, *41* (D1), D312–D319.
- (27) Patel, K.; Kumar, A.; Durani, S. Analysis of the Structural Consensus of the Zinc Coordination Centers of Metalloprotein Structures. *Biochim. Biophys. Acta, Proteins Proteomics* **2007**, *1774* (10), 1247–1253.
- (28) Auld, D. S. Zinc Coordination Sphere in Biochemical Zinc Sites. In *Zinc Biochemistry, Physiology, and Homeostasis*; Maret, W., Ed.; Springer Netherlands: Dordrecht, 2001; pp 85–127.
- (29) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (30) Ribeiro, A. J. M.; Holliday, G. L.; Furnham, N.; Tyzack, J. D.; Ferris, K.; Thornton, J. M. Mechanism and Catalytic Site Atlas (M-CSA): A Database of Enzyme Reaction Mechanisms and Active Sites. *Nucleic Acids Res.* **2018**, *46* (D1), D618–D623.
- (31) Bazayeva, M.; Laveglia, V.; Andreini, C.; Rosato, A. Metal-Induced Structural Variability of Mononuclear Metal-Binding Sites from a Database Perspective. *J. Inorg. Biochem.* **2023**, *238*, No. 112025.
- (32) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.
- (33) Wilson, C. J.; Apiyo, D.; Wittung-Stafshede, P. Role of Cofactors in Metalloprotein Folding. *Q. Rev. Biophys.* **2004**, *37* (3–4), 285–314.
- (34) Lyu, J.; Kopolka, N.; Gumpfer, R.; Alon, A.; Wang, L.; Jain, M. K.; Barros-Alvarez, X.; Sakamoto, K.; Kim, Y.; DiBerto, J.; Kim, K.; Glenn, I. S.; Tummino, T. A.; Huang, S.; Irwin, J. J.; Tarkhanova, O. O.; Moroz, Y.; Skiniotis, G.; Kruse, A. C.; Shoichet, B. K.; Roth, B. L. AlphaFold2 Structures Guide Prospective Ligand Discovery. *Science* **2024**, *384* (6702), No. eadn6354.
- (35) Karelina, M.; Noh, J. J.; Dror, R. O. How Accurately Can One Predict Drug Binding Modes Using AlphaFold Models? *eLife* **2023**, *12*, No. RP89386.
- (36) Wehrspan, Z. J.; McDonnell, R. T.; Elcock, A. H. Identification of Iron-Sulfur (Fe-S) Cluster and Zinc (Zn) Binding Sites Within Proteomes Predicted by DeepMind's AlphaFold2 Program Dramatically Expands the Metalloproteome. *J. Mol. Biol.* **2022**, *434* (2), No. 167377.
- (37) Bentrop, D.; Bertini, I.; Iacoviello, R.; Luchinat, C.; Niikura, Y.; Piccioli, M.; Presenti, C.; Rosato, A. Structural and Dynamical Properties of a Partially Unfolded Fe₄ S₄ Protein: Role of the Cofactor in Protein Folding. *Biochemistry* **1999**, *38* (15), 4669–4680.
- (38) Cowley, A. B.; Rivera, M.; Benson, D. R. Stabilizing Roles of Residual Structure in the Empty Heme Binding Pockets and Unfolded States of Microsomal and Mitochondrial Apocytochrome b₅. *Protein Sci.* **2004**, *13* (9), 2316–2329.