
PrOMet: Accurate and Robust Metal-binding site and Type Prediction

Gangeun Park

Department of Chemistry, Seoul National University
gangeun@snu.ac.kr

Chaok Seok*

Department of Chemistry, Seoul National University
chaok@snu.ac.kr

Hahnbeom Park†

Biomedical research division, Korea Institute of Science and Technology
hahnbeom@kist.re.kr

Abstract

Metal binding to proteins is crucial for many biological functions, yet the prediction of metal-binding sites and metal types remains underexplored compared to other biomolecular interactions. Despite advances in deep learning, computational methods have achieved limited success in this area; even the best-reported approaches reach only moderate localization accuracy on high-resolution crystal structures, and even worse when applied to medium-resolution data. We hypothesize that the main challenge lies in the nature of metal binding: while the putative sites span the entire protein surface, actual binding requires the atomic-level coordination geometry. We present a structure-based metal binding site and type predictor, PrOMet, in this work, which leverages a Prune-and-Optimization strategy built upon two hierarchically connected machine learning architectures. This strategy allows the method to effectively narrow the search space while maintaining robustness to AlphaFold models and medium-resolution cryo-EM structures. PrOMet outperforms existing methods, including structure-based and language model-based predictors, achieving a PR-AUC of 0.739 for metal ion localization and an average F1-score of 0.743 for metal type classification—Zn, Ca, Mg, Mn, Fe, Cu, respectively. PrOMet can be an effective tool for allocating uncertain metal positions and types within protein structures, as well as for metalloenzyme design studies.

1 Introduction

Metal ions play essential roles in biology, including catalysis [1], structural stability [2], transport, electron transfer [3], and regulation, and are known to exist in over a third of Protein Data Bank (PDB) structures [4]. Yet, identifying metal binding sites and ion types remains challenging, especially in moderate or low-resolution maps, where metals are hard to distinguish from solvent or nearby atoms [5]. These challenges highlight the need for accurate computational methods to predict metal-binding sites from structure alone. Such computational tools should help understand metal-

*correspondence to

†correspondence to

mediated biological events, but also for optimizing metal-mediated protein functions, such as de novo metalloenzyme design [6] or metal-dependent activity engineering [7].

There are three groups of research directions regarding the metal site prediction: i) sequence-only prediction to estimate residue-level coordination preference [8, 9], ii) structure-based predictions leveraging 3D protein coordinates based on geometrical heuristics [10] or on deep-learning feature extraction [11], and finally iii) template-based methods ([12, 13]) using known metal-bound structural information. Each type of approach has its own advantages and limitations. Sequence-based methods scale well but lack 3-D context, making precise prediction difficult. Structure-based approaches have limited applicability to structures with errors or uncertainties. Template-based approaches can hardly be generalized to unseen proteins. Finally, two common limitations exist regardless of the approach: first, the majority of prediction methods just find sites without the metal type specification, and second, they cannot be utilized for those jointly coordinating with other biomolecules[14]. Given these limitations, a more universal metal prediction method is demanded that works robustly and accurately for broader types of problems.

In this work, we introduce PrOMet (Prune-and-Optimization for Metals), a two-stage structure-based metal binding site and type prediction framework. The main hypothesis in this work is that a structure-based method can provide a universal solution once it can address the mixed-scale nature of the task: scanning the whole protein surface while evaluating fine atomic-level coordination. A high-recall pruning step narrows the global search across the protein surface, and a fine-grained optimization step enforces coordination geometry and local chemistry to localize sites and assign metal types. PrOMet accepts ligand-bound inputs (small molecules, RNA, or DNA) and, to our knowledge, is among the first structure-based deep-learning methods to jointly predict both 3-D location and the metal ion identity while leveraging ligand context.

2 Results and Discussion

2.1 Overview of the work

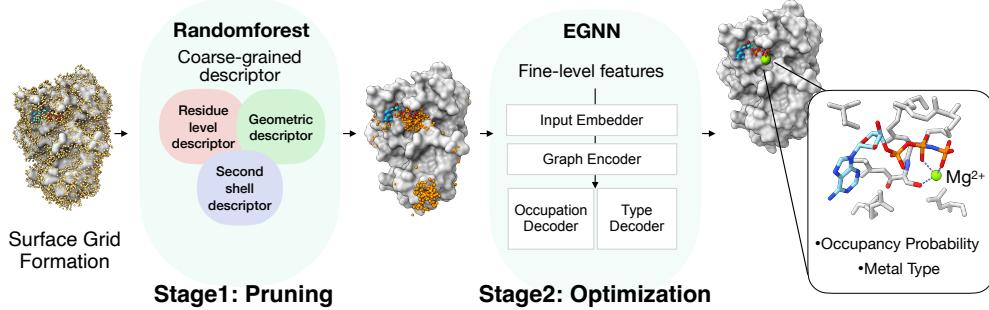


Figure 1: Framework of PrOMet.

An overview of the model workflow and architecture is presented in Figure 1. The framework operates in two sequential stages: the Pruning stage followed by the Optimization stage, which are designed for their own distinct purposes. Conceptually speaking, at the first Pruning stage, coarse-grained geometric filters are applied to remove unlikely patches in a protein surface. At the second Optimization stage, a fine-grained atomic-level score is applied to select and rank the most plausible binding location and types among selected patches. Conceptual overview of each model is provided below; details of input features and model architecture are reported in Methods.

In the first stage, a Random Forest (RF) classifier aims to reduce binding site candidates as much as possible. The classifier operates on solvent-accessible grid points on the protein surface, where coarse-grained features are collected from nearby atoms—such as the number of nearby donor atoms and core residues (Asp/Glu/Cys/His) [15, 16], ligand/donor presence, accessibility, and second-shell features (non-coordinating neighbors within $\sim 5 \text{ \AA}$) [16, 17]. Because the main purpose of this stage is to reduce the number of candidates with minimal loss of true points, the filtering criterion is chosen to maximize recall, yielding a recall of 98.8% with a precision of 0.27% on an independent test set.

In the subsequent optimization stage, the likelihood values of these candidate grid points with probability > 0.5 are clustered and re-ranked, along with the prediction of the binding metal type. The

binding site candidates were allocated to the probability-weighted cluster centers, and metal types are predicted through an independent head. The scoring network was implemented using the Equivariant Graph Neural Network (EGNN)[18].

Both the RF classifier and EGNN scorer were trained on 16,196 PDB entries encompassing protein-metal complexes determined either by X-ray crystallography or cryo-EM. In contrast to previous works in which only high-resolution structures were taken, we collected structures with resolutions up to 35 Å, and all ligands and nucleic acids except waters were retained. This choice was intended for two reasons: first, we sought a network generally applicable to medium-resolution or computationally modeled structures and in the presence of co-factor structures, and second, to extend the dataset as much as possible. To reduce non-biological metals[19], we applied filtering based on the number, types, and proximity of coordinating atoms. For a fair evaluation, we enforced chain-level non-overlap with a 30% sequence-identity clustering between training and held-out test sets, and constructed independent benchmark subsets (crystal, cryo-EM, AF3) from this split.

2.2 PrOMet accurately localizes metal-binding sites in experimental structures.

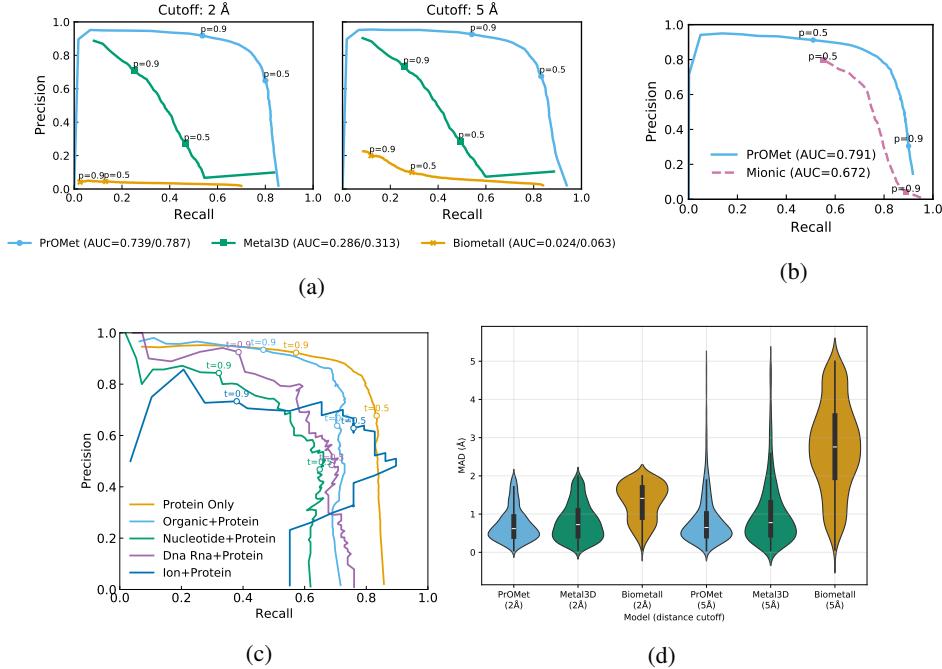


Figure 2: Performance comparison of metal-binding site prediction methods in precision–recall (PR) curves. (a) Metal-site–level predictions with distance cutoffs of 2 Å (left) and 5 Å (right). (b) Residue-level predictions. (c) Per-ligand type prediction of PrOMet. (d) Distribution of metal site positional errors among correctly predicted sites at 2 Å (left) and 5 Å (right).

We first benchmarked the metal binding site prediction on experimental protein structures, a metal-type agnostic task. Three approaches were selected for comparison, representing different types of approaches: Metal3D for a structure-based deep learning method, BioMetAll for a heuristic and geometry–driven method, and M-Ionic for a sequence-based method.[20]

In Figure 2a, we compared site-level precision–recall (PR) curves. A prediction is considered correct if the predicted positional deviation from the answer is within two cutoffs: 2 Å and 5 Å for precise and tolerable analyses. At both thresholds, PrOMet achieved the highest PR-AUC (area under curve) compared to all the other methods. Using a 2 Å cutoff, PrOMet achieved a recall of 80.0% and a precision of 65.0%, which is superior to others (the second best Metal3D, recall/precision = 46.7/27.0%). To compare against a residue-level predictor M-Ionic, we evaluated the precision and recall of coordinating residue prediction defined as < 3 Å to the metal (Figure 2b). Again, in this metric, our model achieved a higher PR-AUC (0.791 vs 0.672).

Decomposing the prediction accuracy by whether coordination involves molecules other than proteins, PrOMet broadly works for various molecular interfaces (Figure 2c). A slight drop in non-protein interfaces might be due to less training data; however, to our knowledge, unlike other published works, it is promising that we first showed PrOMet reasonably works to broader types of biomolecular interfaces.

The binding site prediction accuracy by PrOMet generally worked for all metal types. According to the analyses decomposed to metal types (Zn, Ca, Mg, Mn, Fe, and Cu, Figure S1, PrOMet achieved a good balance between precision and recall values across the 6 most common metals. Interestingly, although Metal3D was solely trained on Zn-only data, PrOMet’s performance was better for the Zn-binding site predictions (PR-AUC 0.494 vs 0.100).

Finally, we evaluated the structural accuracy of binding site coordinates (Figure 2d). This analysis measures how exactly the coordinate can be estimated when the binding site was predicted correctly. Under the probability thresholds suggested by methods (e.g. p=0.5 for ours), PrOMet achieved the lowest median error (0.66 Å at 2 Å cutoff) on 2,382 sites. Metal3D also showed comparable accuracy, but because of its poor recall, this level of coordinate estimation applied only to 1241 sites.

2.3 PrOMet reliably distinguishes between different types of bound metal ions.

Table 1: Comparison of PrOMet and M-Ionic across metal types.

Metal	Model	Precision	Recall	F1-score	N
ZN	PrOMet	0.938	0.793	0.859	1577
	M-Ionic	0.219	0.784	0.343	
CA	PrOMet	0.843	0.763	0.801	733
	M-Ionic	0.114	0.619	0.193	
MG	PrOMet	0.745	0.350	0.476	535
	M-Ionic	0.099	0.514	0.166	
MN	PrOMet	0.341	0.288	0.312	160
	M-Ionic	0.142	0.356	0.203	
FE	PrOMet	0.459	0.189	0.268	90
	M-Ionic	0.096	0.289	0.144	
CU	PrOMet	0.565	0.433	0.491	30
	M-Ionic	0.124	0.533	0.201	
Average	PrOMet	0.847	0.661	0.743	–
	M-Ionic	0.156	0.659	0.156	–

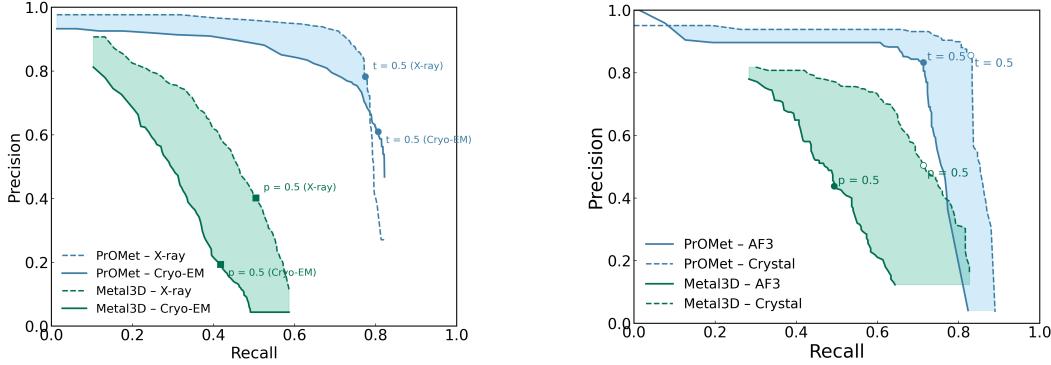
Next, models were benchmarked in their ability to jointly predict the binding site and metal type. The comparison could be made only between PrOMet and M-Ionic; BioMetAll and Metal3D provide site only. The prediction is regarded as correct if the site prediction is within 5 Å and the highest probability metal type matches the actual type. For the residue-level predictor M-Ionic, residue-level signals were first converted to sites by averaging the centers of coordinating atoms in predicted residues, then by assigning a metal type with the highest probability. In Table 1, per-metal type benchmark results on the test set are reported. PrOMet attained much higher precision compared to M-Ionic, resulting in a higher F1-score for every metal class in our evaluation. Note that Co and Ni had very few data and were ignored from the analysis. The gains were largest on well-represented ions: Zn improves from 0.343 (M-Ionic) to 0.859, and Ca from 0.193 to 0.801.

To better understand PrOMet’s behavior beyond statistics, we examined qualitative patterns found in successful and failed predictions. First, a relatively good performance on the sparse metal type Cu is attributed to unique local coordination geometry (e.g., His/Cys-enriched coordination and characteristic geometric distortions) [16, 21]. Second, certain prediction errors were still chemically consistent: a confusion analysis (Figure S2) shows predictions were often swapped more systematically (rather than randomly) between chemically similar ones, such as alkaline-earth cations (Ca and Mg) [22–24].

2.4 Robustness to low-resolution cryo-EM and AlphaFold model structures

One of the well-known weaknesses of a structure-based approach over a sequence- or a template-based approach, is its undesired sensitivity to structural noise. To investigate the tolerance of methods against structural uncertainty, we benchmarked methods in two ways. First, by conditioning on structure determination methods (X-ray crystallography versus cryo-EM), all three methods—PrOMet, Metal3D, and BioMetAll—show performance decreases on cryo-EM relative to crystal (Figure 3a). However, PrOMet’s drop was considerably smaller than Metal3D’s, indicating greater tolerance to coordinate noise.

Because predicted structures are increasingly used, we next asked whether PrOMet maintains accuracy on AlphaFold (AF) model inputs [25, 26]. We first analyzed the structural errors around metal



(a) **PR curves on X-ray vs Cryo-EM.** Shaded area indicates the drop from X-ray set to Cryo-EM set.

(b) **PR curves on AF3 model vs Crystal.** Shaded area indicates the drop from Crystal to AF3.

Figure 3: Precision–recall comparisons under different structure sources.

binding sites (residues around $\leq 5 \text{ \AA}$ the metal) in AF2 models (AFDB models were taken for sufficient statistics) [27], and found backbone and sidechain RMSD were small: $\leq 0.5 \text{ \AA}$ and $\leq 3 \text{ \AA}$, respectively (Figure S3). When metals were copied into the aligned AF2 models, donor–metal distance generally showed broader variance with respect to experimental structures in AF models (Figure S4).

We then benchmarked methods on the AF3 models (modeled without metal inputs) of a subset of the test set. The subset contained precise models of 200 targets, in which metals bind to a single chain only, having global backbone RMSD $\leq 5 \text{ \AA}$ and local side-chain RMSD $\leq 3 \text{ \AA}$ at the binding pocket. Evaluating site and type predictions, PrOMet again showed only a limited degradation relative to their crystal structures, whereas Metal3D dropped more sharply (Figure 3b). The metal type ranking was consistent between experimental and AF3 structures. (Figure S5).

Analyzing the cases when the prediction was wrong only for AF3 models, two local factors dominated: (i) the binding site side-chain accuracy, and (ii) the change in the coordination number (ΔCN). [28] Failures were dominant when either $RMSD > 2 \text{ \AA}$ or when CN became chemically implausible. Consistent with this, cases remained correct when several donor atoms were still present even with some local distortions. (Figure S6)

3 Discussions

3.1 Ablation study reveals the effectiveness of the two-stage protocol

We tested to what extent PrOMet benefits from a two-stage protocol. When fine-grained scorer was tested without the prior pruning stage, the PR-AUC value significantly dropped from 0.686 to 0.386 (Figure S7). The main reason was a large increase in the number of false positives, revealing two-level filters were necessary not only for efficiency but also for prediction accuracy. This ablation study strengthens our major hypothesis underlying this work.

3.2 Training Data curation improves robustness to structural errors

PrOMet’s robustness stems from the dataset composition and the label curation. First, training data include both crystal and cryo-EM structures spanning high to medium/low resolutions, repeatedly exposing the model to coordinate noise and partial distortions. Inclusion of medium to low resolution structures prevents overfitting to perfect geometry, guiding networks to learn tolerances to moderate structural noises. Second, the biological filtering process helped the model consistently learn chemical constraints (coordination number, donor preferences, distance/angle bounds) and to reject chemically implausible arrangements. Together, resolution diversity and high-quality annotations explain why PrOMet shows robustness to cryo-EM and AF3 structures.

References

- [1] Claudia Andreini, Ivano Bertini, and Antonio Rosato. Metalloproteomes: a bioinformatic approach. *Accounts of Chemical Research*, 42:1471–1479, 2009.
- [2] Maria E. Palm-Espling, M. S. Niemiec, and Pernilla Wittung-Stafshede. Role of metal in folding and stability of copper proteins in vitro. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1823:1594–1603, 2012.
- [3] Jay R. Winkler and Harry B. Gray. Electron flow through metalloproteins. *Chemical Reviews*, 114(7):3369–3380, 2014. doi: 10.1021/cr4004715.
- [4] John A. Tainer, Victoria A. Roberts, and Edward D. Getzoff. Metal-binding sites in proteins. *Current Opinion in Biotechnology*, 2:582–591, 1991.
- [5] Katarzyna B. Handing, Ewa Niedzialkowska, Ivan G. Shabalin, Misty L. Kuhn, Heping Zheng, and Wladek Minor. Characterizing metal-binding sites in proteins with x-ray crystallography. *Nature Protocols*, 13:1062–1090, 2018.
- [6] Tobias Vornholt, Florian Leiss-Maier, Woo Jae Jeong, Cathleen Zeymer, Woon Ju Song, Gerard Roelfes, and Thomas R. Ward. Artificial metalloenzymes. *Nature Reviews Chemistry*, 6:89–107, 2022.
- [7] Linda Leone, Flavia Nastri, and Angela Lombardi. Natural, designed and engineered metalloenzymes: Structure, catalytic mechanisms and applications. *Chemical Reviews*, 120:123–187, 2020.
- [8] Aditi Shenoy, Yogesh Kalakoti, Durai Sundar, and Arne Elofsson. M-ionic: prediction of metal-ion-binding sites from sequence using residue embeddings. *Bioinformatics*, 40(1):btad782, 2024. doi: 10.1093/bioinformatics/btad782.
- [9] N. Ye, F. Zhou, X. Liang, H. Chai, J. Fan, B. Li, and J. Zhang. A comprehensive review of computation-based metal-binding prediction approaches at the residue level. *BioMed Research International*, page 8965712, 2022. doi: 10.1155/2022/8965712.
- [10] José-Emilio Sánchez-Aparicio, Laura Tiessler-Sala, Lorea Velasco-Carneros, Lorena Roldán-Martín, Giuseppe Sciortino, and Jean-Didier Maréchal. Biometall: Identifying metal-binding sites in proteins from backbone preorganization. *Journal of Chemical Information and Modeling*, 62:1065–1077, 2022.
- [11] Simon L. Dürr, Andrea Levy, and Ursula Rothlisberger. Metal3d: a general deep learning framework for accurate metal ion location prediction in proteins. *Nature Communications*, 14:2713, 2023. doi: 10.1038/s41467-023-37870-6.
- [12] Yu-Feng Lin, Chih-Wen Cheng, Chung-Shiuan Shih, Jenn-Kang Hwang, Chin-Sheng Yu, and Chih-Hao Lu. Mib: Metal ion-binding site prediction and docking server. *Journal of Chemical Information and Modeling*, 56(12):2287–2291, 2016. doi: 10.1021/acs.jcim.6b00407.
- [13] Chih-Hao Lu, Chih-Chieh Chen, Chin-Sheng Yu, Yen-Yi Liu, Jia-Jun Liu, Sung-Tai Wei, Yu-Feng Lin, and Sheng Yu. Mib2: metal ion-binding site prediction and modeling server. *Bioinformatics*, 38(18):4428–4429, 2022. doi: 10.1093/bioinformatics/btac534.
- [14] Cosimo Ciofalo, Vincenzo Laveglia, Claudia Andreini, and Antonio Rosato. Benchmarking zinc-binding site predictors: A comparative analysis of structure-based approaches. *Journal of Chemical Information and Modeling*, 65(10):5205–5215, 2025. doi: 10.1021/acs.jcim.5c00549.
- [15] M. Bazayeva, C. Andreini, and A. Rosato. A database overview of metal-coordination distances in metalloproteins. *Acta Crystallographica Section D: Structural Biology*, 80(5):362–376, 2024. doi: 10.1107/S2059798324003152.
- [16] Shani Barber-Zucker, Boaz Shaanan, and Raz Zarivach. Transition metal binding selectivity in proteins and its correlation with the phylogenomic classification of the cation diffusion facilitator protein family. *Scientific Reports*, 7:16381, 2017. doi: 10.1038/s41598-017-16777-5.

- [17] T. Dudev and C. Lim. First–second shell interactions in metal binding sites in proteins: A pdb survey and dft/cdm calculations. *Journal of the American Chemical Society*, 130:9668–9676, 2008.
- [18] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *PMLR*, pages 9323–9332, 2021.
- [19] Katarzyna B. Handing, Ewa Niedzialkowska, Ivan G. Shabalin, Misty L. Kuhn, Heping Zheng, and Wladek Minor. Characterizing metal-binding sites in proteins with x-ray crystallography. *Nature Protocols*, 13:1062–1090, 2018. Duplicate of [5].
- [20] Zeming Lin, H. Akin, R. Rao, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.
- [21] Jeffrey T. Rubino. Coordination chemistry of copper proteins: How nature handles a toxic cargo for essential function. *Coordination Chemistry Reviews*, 256(15–16):1899–1916, 2012.
- [22] David Allouche, Joseph Parello, and Yves-Henri Sanejouand. Ca²⁺/mg²⁺ exchange in parvalbumin and other ef-hand proteins. a theoretical study. *Journal of Molecular Biology*, 285(2):857–873, 1999.
- [23] Miklós Nyitrai, Gabriella Hild, Zsuzsa Lakos, and Béla Somogyi. Effect of ca²⁺–mg²⁺ exchange on the flexibility and/or conformation of the small domain in monomeric actin. *Biophysical Journal*, 74(5):2474–2481, 1998.
- [24] A. R. Nayak, W. Rangubpit, A. H. Will, et al. Interplay between mg²⁺ and ca²⁺ at multiple sites of the ryanodine receptor. *Nature Communications*, 15:4115, 2024.
- [25] John Jumper, Richard Evans, Alexander Pritzel, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596:583–589, 2021.
- [26] Joel Abramson, Justin Adler, Johannes Dunger, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630:493–500, 2024.
- [27] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, et al. Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444, 2022.
- [28] Stephen J. Lippard and Jeremy M. Berg. *Principles of Bioinorganic Chemistry*. University Science Books, 1994.
- [29] Vincenzo Laveglia, Andrea Giachetti, Davide Sala, Claudia Andreini, and Antonio Rosato. 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. *Journal of Chemical Information and Modeling*, 62(12):2951–2960, 2022. doi: 10.1021/acs.jcim.2c00522.

Acknowledges

A Datasets

To construct a dataset for biologically relevant metal-binding site prediction, we utilized protein–metal complexes from the BioLiP database. Proteins containing Mg, Zn, Mn, Ca, Fe, Ni, Co, Cu, or K were included. Initially, 12 candidate metals with more than 500 associated PDB entries were considered: Mg, Zn, Ca, Fe, Na, Mn, K, Ni, Cu, Co, Cd, and Hg. Toxic metals (Cd and Hg) were excluded, as well as sodium (Na) to avoid artifacts from buffer effects. For each entry, ligands present in the PDB structure were retained, excluding only water molecules, to enable prediction of metal-binding in both apo-protein and ligand-bound complexes.

A.1 Determination of Biological vs. Non-Biological Sites

To establish a criterion for distinguishing biological from non-biological metal-binding sites, we analyzed the dataset provided in [29], which includes explicit annotations of binding sites as biological or non-biological. We examined the distribution of coordinating atoms (N, O, S, Se) within 3 Å for both categories and found that the two groups were most distinctly separated when at least three coordinating atoms were present. The reported performance values (1697/1921 correctly identified biological sites and 1770/2823 correctly identified non-biological sites) were obtained from this annotated dataset used for threshold validation. Based on this analysis, we adopted the ≥ 3 coordinating atom threshold to filter sites in our dataset, thereby reducing crystallographic artifacts and focusing on biologically relevant patterns.

A.2 Training, Validation, and Testsets

Metal-binding proteins were collected from BioLiP as of February 19, 2025. Entries available only in CIF format (without corresponding PDB files) and those determined by NMR spectroscopy were excluded. Sequence redundancy was reduced using MMseqs2 clustering at 30% sequence identity (`mmseqs2 easy-cluster <fasta> <Result> tmp -min-seq-id 0.3 -cov-mode 0 -c 0.8 -s 8 -cluster-mode 1`).

To prevent temporal and sequence redundancy biases, datasets were first split chronologically: training/validation cutoff at 2019-01-01, and validation/test cutoff at 2021-03-31. Then PDBs belonging to the same sequence cluster as training entries were excluded from the validation and test sets. This ensured strict separation of training and evaluation data.

A.3 The AlphaFold3 model dataset

To evaluate performance on predicted protein structures, we generated benchmark data using AlphaFold3 (AF3). Protein chains containing metal-binding sites were extracted (chain A only), where the selected metal-binding sites involved direct coordination with the protein rather than ligand-mediated interactions. The extracted chains were modeled with AF3, and the predicted structures were aligned to experimental PDBs. Entries were retained if the global backbone RMSD was ≤ 5 Å and the local RMSD within 5 Å of the metal site was ≤ 3 Å. This dataset enabled evaluation under realistic modeling scenarios.

B Model Architecture

B.1 Grid representation

We discretized the three-dimensional space surrounding protein complexes into grid points. These grid points were sampled from the coordination spheres of atoms, defined by their van der Waals radii with a slight expansion to capture regions where interactions and coordination can occur. Sterically clashing or redundant points were excluded to obtain a compact and non-overlapping set of nodes.

B.2 Random Forest for the pruning stage

Feature embedding. For the RF classifier, each grid point was embedded with structural descriptors designed to capture the coordination environment of potential metal-binding sites.

Coarse-level descriptors. At every grid point, three types of descriptors – residue-level, geometric, and second-shell descriptors – are collected from the surrounding environment. For the residue level descriptors, information from protein residues are summarized by counting nearby coordinating atoms (N, O, S, Se), with separate counts for core residues (His, Cys, Asp, Glu) and backbone oxygen atoms, and from cofactor ligands (if any exist nearby) the number of donor atoms, halide anions, and other heteroatoms are counted. For the geometric descriptors, steric and geometric accessibility descriptors reflected the openness and geometric suitability of the site, using measures such as distances to nearby atoms, relative surface accessibility, and geometric constraints on core residues. Finally, for the second-shell descriptors, stabilizing effects from residues within 3.2–5.0 Å are captured by counting the number of core residues and backbone oxygens in this region. Together, these feature groups provide an interpretable description of the local environment and enable the Random Forest to efficiently identify candidate metal-binding regions.

Training. The RF was trained using protein monomers from the training set. Multimeric assemblies were excluded due to a memory issue. To address the strong class imbalance between binding and non-binding grid points, we employed a balanced Random Forest strategy, which equalizes class representation during the bootstrap sampling. This ensured that the classifier maintained sensitivity to true binding regions while avoiding bias toward the abundant non-binding class. The classifier was implemented using the `BalancedRandomForestClassifier` (imblearn v0.12.3) with default settings, specifying only `random_state=42` and `n_jobs=-1`. Default parameters include 100 estimators, Gini impurity as the split criterion, unlimited tree depth, bootstrap sampling, and automatic undersampling of the majority class.

B.3 EGNN for the optimization stage

Feature embedding. In the second stage, we applied an Equivariant Graph Neural Network (EGNN) to refine predictions within the candidate regions identified by the Random Forest. The graph representation was constructed using grid points and protein atoms as nodes, with geometric and chemical relations encoded as edges. In addition to conventional descriptors (atom type, amino acid identity, solvent accessibility, inter-node distance, and bond connectivity), we introduced the *polarity vector angle* as a new edge feature. The angle θ between the polarity vector of a protein atom and the displacement vector toward a grid point was encoded as $\sin \theta$ and $\cos \theta$, allowing directional polarity information to be incorporated within the EGNN framework. This angular descriptor substantially improved performance, reducing validation loss compared to baseline models without polarity information.

Training. The EGNN model was trained on a system equipped with 24 physical CPU cores (48 logical CPUs) and 4 NVIDIA RTX A5000 GPUs. The total training runtime was 2 days, 21 minutes, and 40 seconds.

B.4 Clustering and site-level annotation

Grid-level predictions from the EGNN were aggregated into site-level annotations using clustering algorithms. Specifically, neighboring grid points with high binding probabilities were grouped into clusters, each corresponding to a predicted metal-binding site. We used DBSCAN with $\text{eps} = 2$ and $\text{min_samples} = 2$ for clustering. For each cluster, the overall binding probability was determined, and the most probable metal ion type was assigned based on the aggregated grid-level distributions. This post-processing step allowed grid-based predictions to be translated into biologically interpretable binding site annotations, enabling evaluation at the site level in downstream analyses.

C Supplement

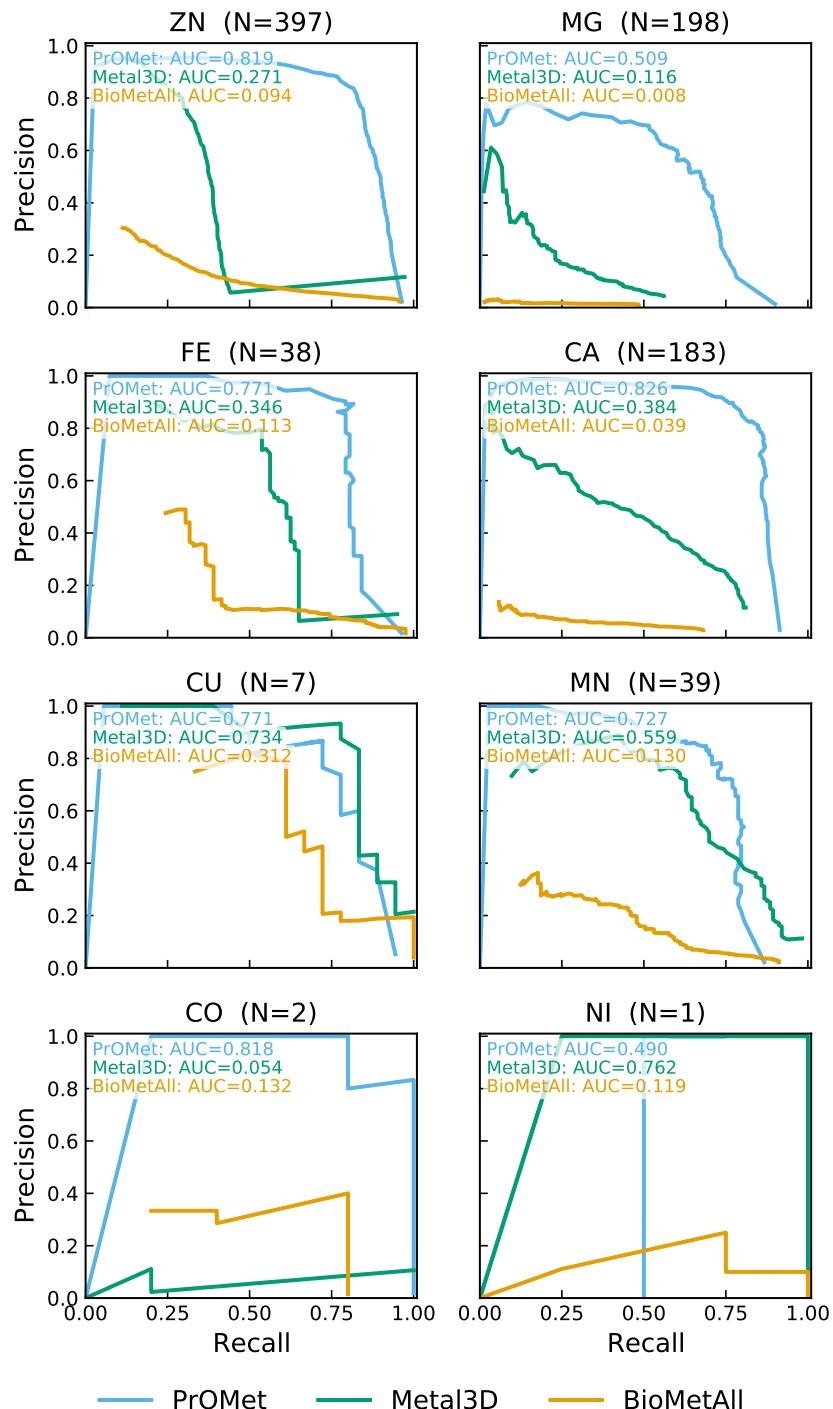


Figure S1: PR curve for localization performance of each metal type

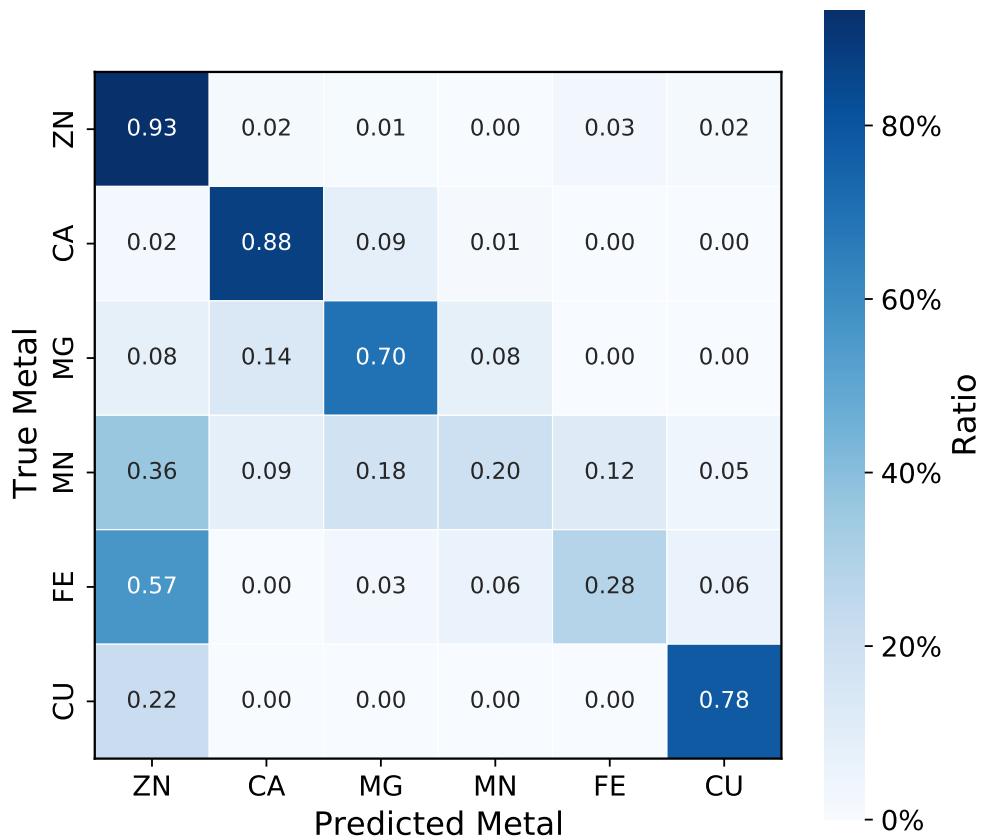


Figure S2: Confusion matrix of predicted metal and true metal type

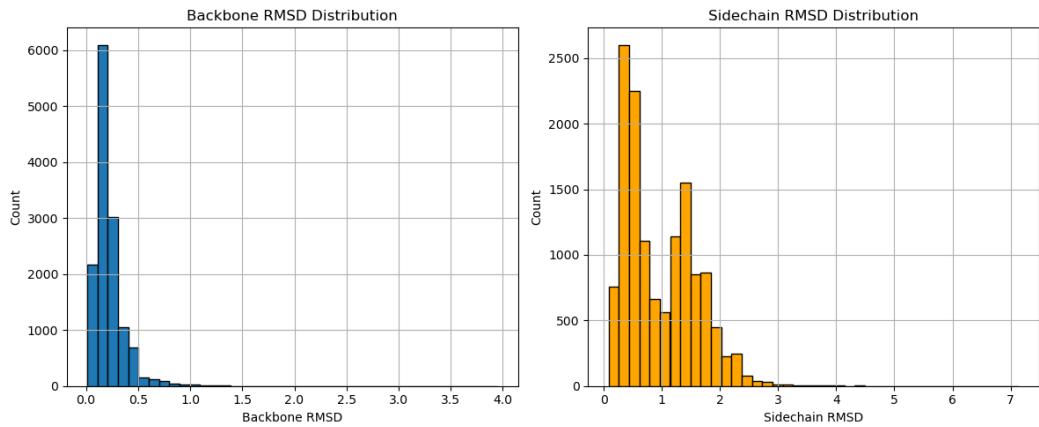


Figure S3: Backbone and sidechain rmsd of metal binding site between Alphafold2 models and crystal structures

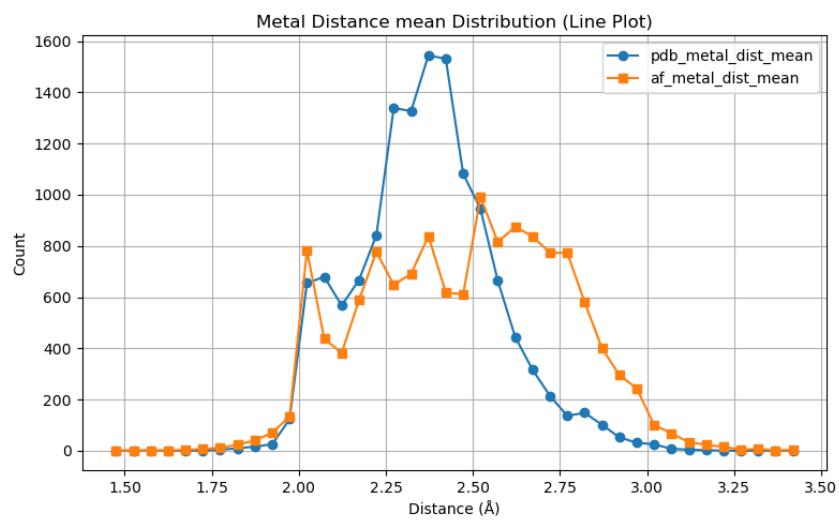
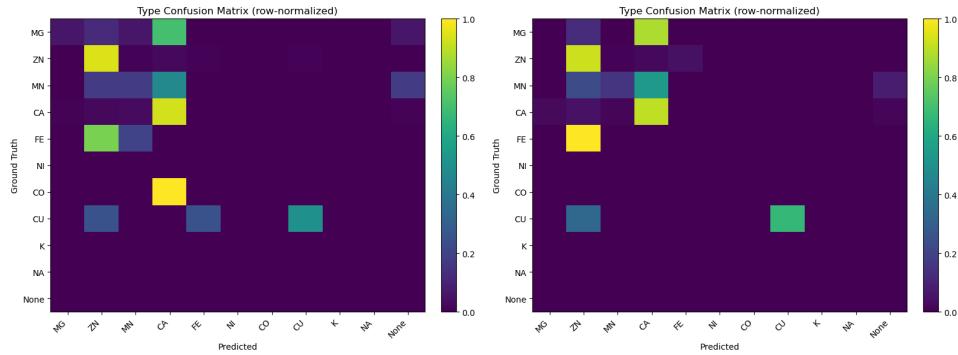
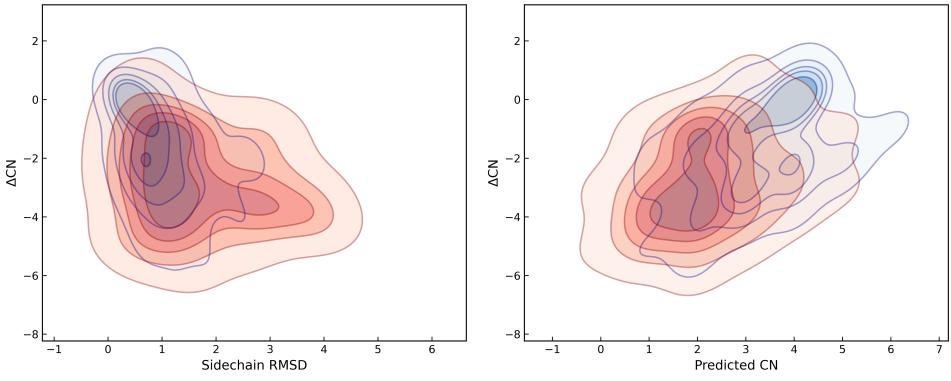


Figure S4: Distribution of mean metal–donor distances in experimental PDB structures (blue) and AlphaFold models (orange). While PDB structures show a sharp peak around 2.3–2.4 Å, AlphaFold models display broader and more variable donor–metal distances, reflecting larger deviations in modeled local geometries.



(a) Confusion matrix predicted with crystal structure.
(b) Confusion matrix predicted with AF3 model structure.

Figure S5: Confusion matrix with Crystal structure and AF3 model structure. They show similar patterns and no distinct difference between type prediction ordering.



(a) Sidechain RMSD vs ΔCN (crystal reference). (b) Predicted CN vs ΔCN (crystal reference). Fail-Robust cases cluster at low RMSD and near ures concentrate at lower predicted CN with negative $\Delta CN \approx 0$, whereas failures shift to larger RMSD and negative ΔCN , indicating under-coordination relative to ground truth.

Figure S6: 2D-KDE analysis separating AF3-robust predictions (blue; correct on crystal and AF3) from AF3-failed cases (red; correct on crystal, wrong on AF3).

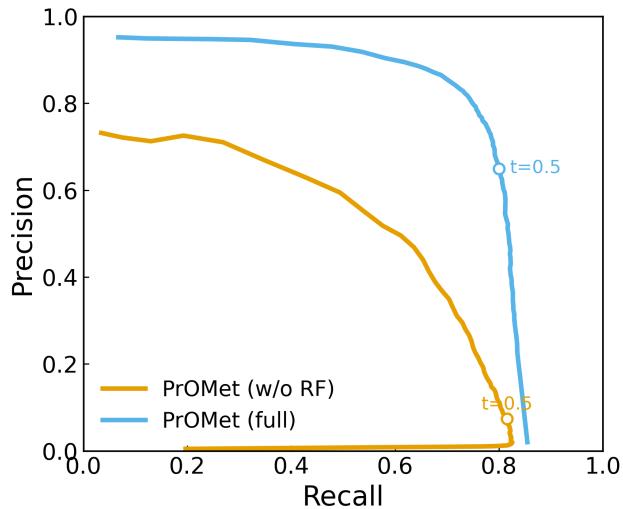


Figure S7: Precision–recall comparison of PrOMet with and without the Random Forest pruning stage. At threshold $t = 0.5$, the ablated model without Random Forest achieves low precision (7.75%) despite high recall (81.4%), whereas the full PrOMet maintains both high precision (65.6%) and recall (80.0%), highlighting the critical role of the pruning stage in improving prediction quality.