Centro Nacional de Biotecnología

UAB Universitat Autònoma de Barcelona

MSc Bioinformatics

Universitat Autònoma de Barcelona

MetalPlacer: a Scipion-integrated pipeline for metal-ion identification and placement in protein structures

Author: Eduardo Rivas Tortuero

Supervised by

Dr. Carlos Óscar Sorzano

Academic tutor

Dr. Alex Perálvarez-Marín

2025-2026

Abstract

Metal ions are essential for the function of approximately one-third of all proteins, yet computationally predicted structures (including AlphaFold models) systematically lack metal coordinates. Several tools exist to screen for metal-binding sites, identify metal types, or place ions in protein structures, but they operate independently and require manual intervention to connect their outputs. Here we present MetalPlacer, a Scipion plugin that integrates metal-ion identification and placement into a single reproducible workflow. The pipeline comprises two modules. Path 1 (ProtMetalPlacer), the main contribution, infers the metal type and ion count from a homologous experimental structure retrieved via the RCSB PDB API and places the ions in the query structure using MetalKB statistical potentials, separating metal identity inference from coordinate calculation. Path 0 (ProtMetalScreener) is an exploratory structural screening module that evaluates whether the input protein is likely to be a metalloprotein using 62 interpretable structural descriptors and a calibrated Random Forest classifier. Validation of Path 1 on 14 pseudo-apo structures covering seven metals (Zn, Ca, Fe, Cu, Mn, Mg, Ni) showed successful placement in 12 of 14 cases, with sub-angstrom mean per-ion distance in 11 of the 12 successful cases. The two Mg structures were the main failure cases, as Mg frequently coordinates with water molecules and cofactors rather than with protein residues alone, which falls outside the scope of the statistical potentials used by MetalKB. Path 0 achieved a site-level ROC-AUC of 0.791 and a protein-level ROC-AUC of 0.777 on an independent test set partitioned by CD-HIT clustering at 40% sequence identity. Both modules are integrated into Scipion as independent protocols, enabling traceable, reproducible execution and compatibility with existing structural biology pipelines. The output is a pseudo-holo PDB file ready for downstream analysis. MetalPlacer is available as an open-source Scipion plugin at https://github.com/edurivvass/metalplacer.

Contents

1 Introduction

1.1 Biological relevance of metal ions in proteins

Metal ions are essential cofactors in a substantial fraction of all proteins. Current estimates indicate that approximately 30–40% of proteins require one or more metal ions to perform their biological function (1).

These ions fulfill diverse roles, including structural stabilization (as in zinc-finger domains), enzymatic catalysis (as in metalloenzymes), electron transfer, and signaling.

The identity of the coordinating metal, the geometry of the binding site, and the composition of the local protein environment are tightly coupled to protein function (2, 3), making the accurate characterization of metal-binding sites a central problem in structural biology.

1.2 The missing metal problem in predicted structures

Experimental determination of metal-binding sites relies primarily on X-ray crystallography and, increasingly, on cryo-electron microscopy. However, a growing number of protein structures are now generated computationally. AlphaFold has demonstrated remarkable accuracy in predicting protein backbone and side-chain conformations (4), yet its models systematically lack coordinates for metal ions, cofactors, and other non-protein components essential for biological function (5). More recently, AlphaFold 3 can predict protein-ion complexes as part of its output (6), but its accuracy for metal placement has not been independently tested, and the model is not freely available for local use, which limits its application in automated pipelines.

Zinc-finger motifs are predicted without zinc; metalloproteases lack the catalytic metal; and ATPases contain no bound nucleotide. This gap between predicted folds and functional completeness has created an urgent need for computational tools capable of identifying and placing metal ions in protein structures.

1.3 Current computational approaches

Computational approaches to address the absence of metal ions in protein structures can be organized around three successive questions:

- Is this protein likely to bind a metal ion?
- Which metal, and how many ions?
- Where exactly should the ion be placed in 3D space?

A variety of tools have been developed to answer each of these questions, but they operate independently and address different aspects of the problem (Figure 1).

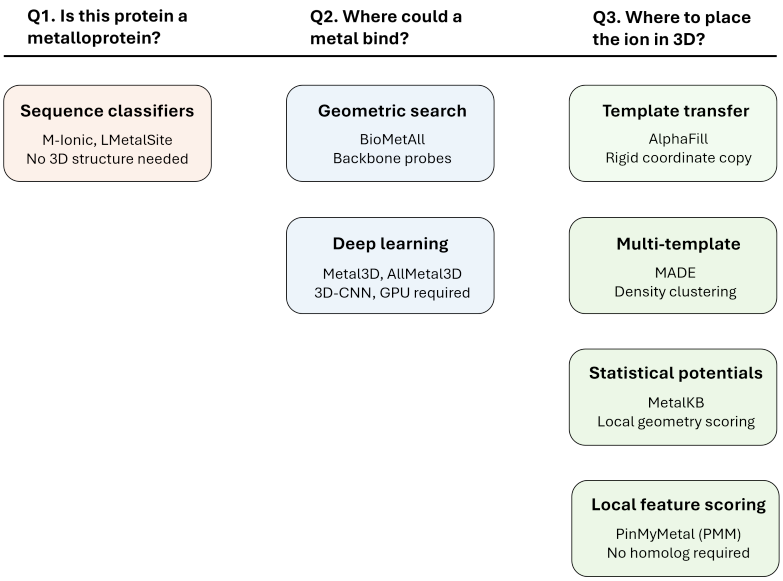

Figure 1: Overview of existing computational approaches for metal-ion prediction in proteins, organized by the three successive questions they address.

1.3.1 Is this protein likely to bind a metal?

Sequence-based classifiers can determine whether a protein is likely to bind metal ions without requiring a 3D structure:

- **Language model classifiers.** M-Ionic uses numerical representations learned by a protein language model (ESM-2) to classify metalloproteins from sequence alone, reporting an ROC-AUC of 0.83 (7). LMetalSite demonstrated that pretrained language models can surpass previous structure-based methods for predicting metal-binding residues by learning structural context implicitly from sequence (8).

These methods are fast and effective, but their predictions reflect the average binding potential encoded in evolutionary data, not the actual geometry of a specific conformer. A protein may adopt different conformations depending on its environment, crystallization conditions, or oligomeric state, and whether a metal-binding site is geometrically preorganized in a given conformer can only be assessed from its 3D coordinates.

1.3.2 Where could a metal bind?

Two main strategies exist for identifying candidate metal-binding sites in protein structures:

- **Geometric search.** BioMetAll identifies sites based on the preorganization of the protein backbone, placing probes on a regular grid wherever potential donor atoms are arranged at appropriate distances for metal coordination (9). Candidate clusters are ranked by the number of probes they contain, which serves as a rough confidence indicator, but BioMetAll does not provide a calibrated probability or energy score for each site, and its high recall comes at the cost of low precision, generating many false positive candidates that require further filtering.

- **Deep learning.** Metal3D uses a 3D convolutional neural network trained on grid-based representations of protein environments to predict the spatial probability density of zinc ions, achieving predictions within 0.70 ± 0.64 Å of experimental positions (10). Its extension AllMetal3D generalizes this approach to multiple metal types (11). These methods provide high spatial precision and a calibrated confidence metric, but require GPU computation.

Both approaches can localize potential binding sites, but neither identifies the metal type from the protein context nor connects the prediction to a downstream placement step.

1.3.3 Where should the ion be placed in 3D space?

For the actual placement of metal ions in a structure, four main strategies have been proposed:

- **Template-based coordinate transfer.** AlphaFill transfers ion coordinates from homologous experimental structures to predicted models via structural superposition (5). This approach is effective when a close homolog is available, but it transfers coordinates rigidly and depends on a minimum sequence identity threshold (25% in AlphaFill).
- **Multi-template density clustering.** MADE superimposes multiple homologs simultaneously and identifies high-density clusters of metal positions, which provides statistical robustness but requires a sufficient number of homologous structures (12).
- **Statistical potential scoring.** MetalKB scores candidate ion positions using metal-specific statistical potentials derived from the MESPEUS database of metal coordination environments (13, 14). Unlike the previous approaches, MetalKB evaluates positions based on the local coordination geometry of the query structure itself, rather than transferring coordinates from a template, and can predict both the 3D coordinates and the coordinating residues for multiple metal types.
- **Hybrid local-feature scoring.** PinMyMetal (PMM) uses local structural and physicochemical features to predict transition metal locations without requiring homologs, achieving median deviations of 0.19 Å for structural sites, 0.33 Å for catalytic sites, and 0.36 Å for regulatory sites (15). PMM assigns a certainty score to each predicted site and provides interactive validation through the CheckMyMetal server. Like MetalKB, it is lightweight and does not require GPU computation.

1.4 The integration gap

Despite the availability of these tools, a significant gap remains: **no existing workflow connects the three questions (screening, identification, and placement) into a single integrated and reproducible pipeline**. Each tool addresses one step in isolation, and using them together requires manual intervention at every stage: running a classifier, interpreting its output, selecting a homolog, extracting the metal annotation, executing the placement tool, and assembling the final structure. This fragmentation has several practical consequences:

- There is no automated path from an apo or predicted structure to a pseudo-holo structure with placed metal ions.

- The metal type and count must typically be specified manually by the user, rather than inferred automatically from available experimental evidence.
- Results are difficult to reproduce because each step uses different software, input formats, and parameter conventions, with no shared record of the decisions made along the way.

Furthermore, none of these tools is currently integrated into a structural biology workflow framework that would provide traceability, reproducibility, and interoperability between steps.

Scipion is an open-source software framework originally developed for cryo-electron microscopy image processing that provides workflow-based execution of reusable, standardized, traceable, and reproducible protocols (16). In Scipion, each processing step is encapsulated as a protocol that receives defined inputs, exposes configurable parameters, executes a series of computational steps, and registers outputs that can be directly connected to subsequent protocols within the same project. This architecture ensures full traceability of parameters and software versions, enables reproducibility of results, and provides interoperability between tools from different developers through a unified graphical interface (17). Integrating a new tool into Scipion as a protocol, rather than distributing it as a standalone script, provides immediate benefits: standardized input/output handling, parameter logging, compatibility with existing structural biology pipelines, and accessibility for users without command-line expertise.

1.5 Objectives and scope of this work

MetalPlacer was developed to address this gap. It is a Scipion plugin composed of two modules: Path 1 (ProtMetalPlacer), the main workflow, which infers the metal type and count from a homologous experimental structure and places the ions using MetalKB statistical potentials; and Path 0 (ProtMetalScreener), an exploratory structural screening module that evaluates whether the input structure is likely to be a metalloprotein, acting as an optional pre-filter. Both modules are integrated into Scipion as independent protocols. The output of the complete pipeline is a pseudo-holo PDB file ready for downstream analysis.

The specific objectives of this work are:

- To develop and validate a homology-based workflow (Path 1) that automatically identifies the metal type and count from a homologous structure and places the ions in the query geometry using MetalKB.
- To design and evaluate an exploratory structural screening module (Path 0) based on interpretable geometric descriptors, as an optional pre-filter for the placement workflow.
- To integrate both modules into the Scipion framework as reproducible, interoperable protocols.

The following sections describe the implementation of each module (Section 2), present the validation results (Section 3), and discuss the capabilities, limitations, and future directions of the pipeline (Section 4).

2 Methods

2.1 Pipeline architecture overview

MetalPlacer is a Scipion plugin designed to support the prediction and placement of metal ions in protein structures. The pipeline is organized into two complementary and independently executable modules, referred to as Path 0 and Path 1.

Path 0 (ProtMetalScreener) is a lightweight structural screening module that evaluates whether an input protein structure is likely to be a metalloprotein. It does not attempt to identify the metal type or place ions; instead, it acts as a preliminary filter to prioritize structures that warrant further analysis. Starting from candidate metal-binding sites generated by BioMetAll, Path 0 extracts 62 local structural descriptors per site, applies a Platt-calibrated Random Forest classifier, and aggregates the site-level scores into a single protein-level probability using the 95th percentile. Three configurable operating points (permissive, moderate, and high-specificity) allow the user to adjust the sensitivity–specificity trade-off depending on the application context.

Path 1 (ProtMetalPlacer) is the main workflow of MetalPlacer. It performs homology-based metal identification and placement: the sequence of the query structure is extracted and used to search the RCSB PDB for homologous experimentally characterized structures. From the selected homolog, the type and number of biologically relevant metal ions are inferred. The ions are then placed in the query structure using the statistical potentials of MetalKB, which evaluate candidate positions using statistical potentials (described in Section 2.2.1), rather than transferring coordinates directly from the template. The output is a pseudo-holo PDB file containing the predicted metal HETATM records, together with a JSON metadata file reporting the selected homolog, sequence identity, metal type, ion count, and predicted coordinates.

Both modules can be executed sequentially or independently, when the user already knows that the query protein is a metalloprotein and wishes to proceed directly to metal placement. The two modules are integrated into Scipion as independent protocols, enabling reproducible execution, parameter tracking, and compatibility with other structural biology tools available in the framework. Figure 2 shows the general architecture of the pipeline.

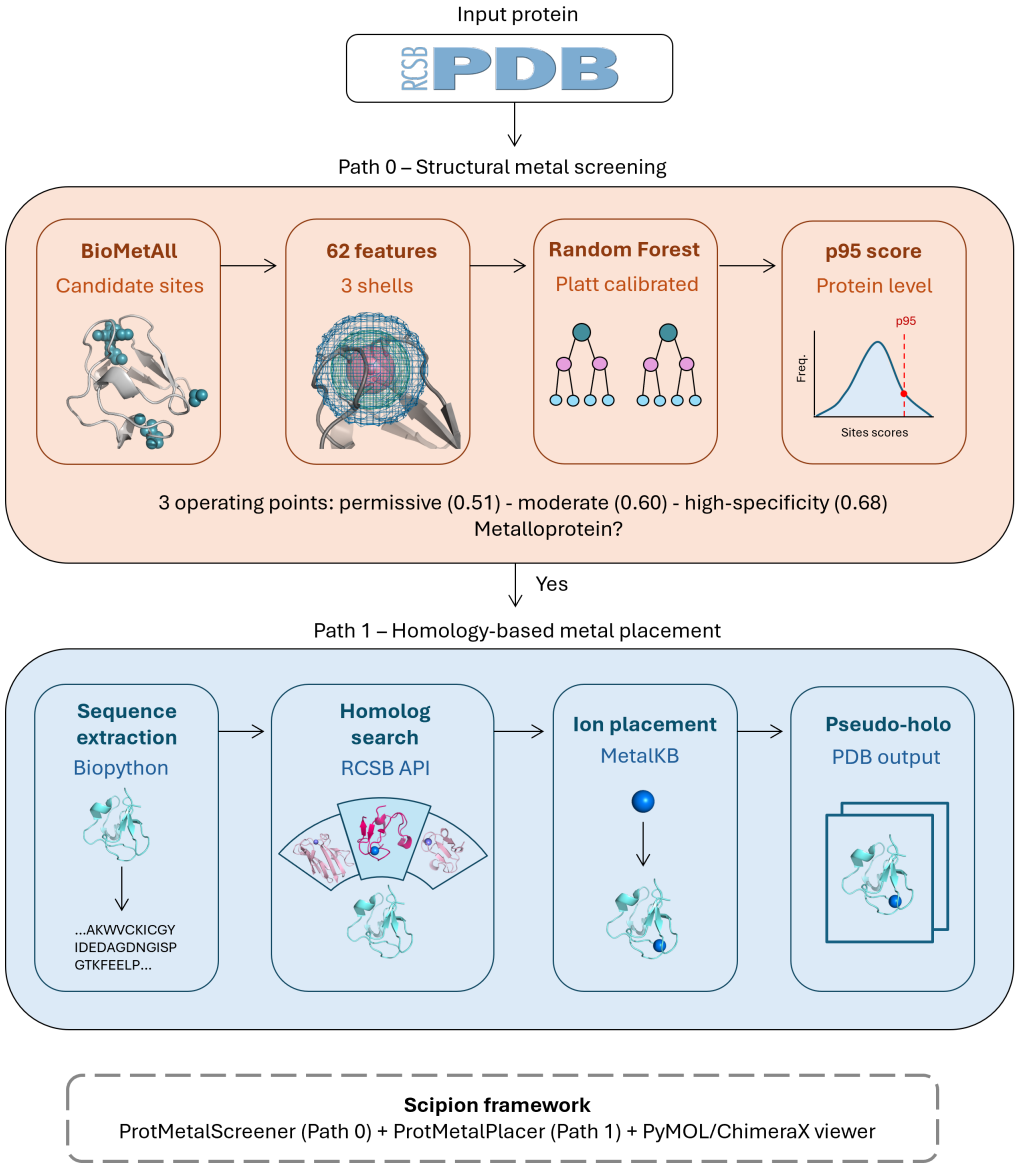

Figure 2: Overview of the MetalPlacer pipeline. Path 0 screens input protein structures for metal-binding likelihood using BioMetAll candidate sites, 62 structural features extracted from three concentric shells, a Platt-calibrated Random Forest, and p95 site-to-protein aggregation at three configurable operating points. Path 1 identifies metal type and count from a sequence homolog retrieved via the RCSB API and places ions using MetalKB statistical potentials. Both modules are integrated as Scipion protocols.

The following sections describe each module in detail: Section 2.2 covers Path 1 (homology-based identification and placement) and Section 2.3 covers Path 0 (structural screening).

2.2 Path 1: homology-based identification and placement

This section describes the technical implementation of the Path 1 approach for the identification and localization of metal ions in protein structures based on sequence homology. The pipeline components, control parameters, and validation strategy are detailed below.

2.2.1 Biological basis and rationale for the approach

A well-established principle in structural biology is that proteins sharing $\geq$30% sequence identity generally adopt similar three-dimensional folds and retain functionally relevant residues (18, 19). In metalloproteins, this structural conservation is especially relevant because metal-binding sites are often defined by conserved local arrangements of coordinating residues, such as histidines, cysteines, aspartates or glutamates. This provides the basis for Path 1: the type and number of metal ions in a protein of interest can be inferred from an experimentally characterized homologous structure, provided that the metal-coordinating environment is conserved (20, 21).

The key difference from direct coordinate transfer methods such as AlphaFill (5) lies in the separation of metal identification from metal positioning. Path 1 uses the homologous template to determine which metal ions are likely to be present (type and number), delegating the calculation of spatial positions to MetalKB, a knowledge-based tool that detects clusters of potential coordinating atoms in the query protein and scores candidate ion positions using metal-specific statistical potentials derived from the MESPEUS database (13). This approach is applicable to any input structure: experimental apo forms, pseudo-apo structures generated by removing HETATM records from holo structures deposited in the PDB, and de novo predicted models (e.g., AlphaFold structures) (4).

2.2.2 Protocol description in four steps

The Path 1 pipeline is integrated into the Scipion framework as a series of four sequential steps that process an input PDB structure and generate predictions of metal-binding sites. Figure 3 provides a schematic overview of the protocol's entire workflow.

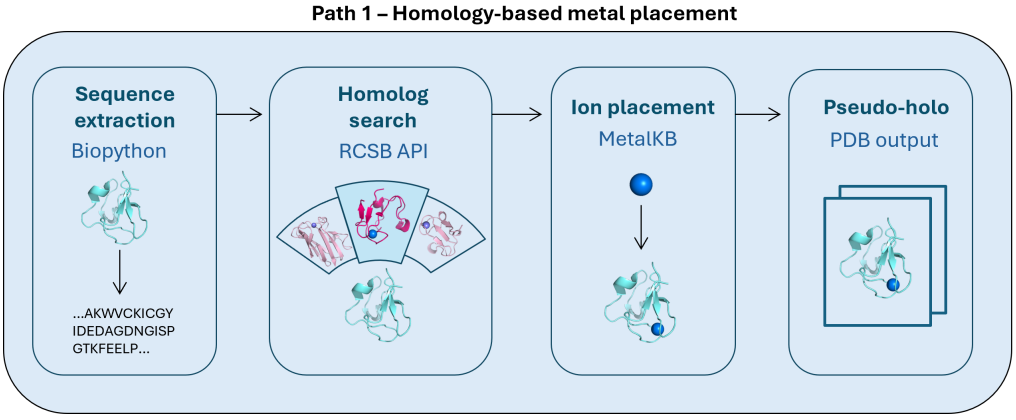

Figure 3: Overview of the Path 1 workflow. 1) The amino acid sequence of the query protein (cyan) is extracted from the input PDB file. 2) Homologous structures (pink) containing a metal ion (spheres) are searched in the RCSB PDB. 3) For each identified metal, MetalKB calculates the optimal 3D coordinates using metal-specific statistical potentials. 4) The protocol generates a pseudo-holo PDB file with the original query coordinates and the placed metal ion(s).

Step 1: Sequence extraction

The amino acid sequence of the query protein was extracted from the ATOM records of the input PDB file using Biopython (22). Standard amino acid residues, including selenomethionine (MSE) and common

histidine variants (HSD/HSE/HSP) were included in the query sequence used for homolog search. By default, chain A was selected, and if absent, the first available chain was used instead.

Step 2: Homolog identification via the RCSB PDB search and data APIs

The extracted sequence was sent to the RCSB PDB (23) sequence search endpoint, which internally uses MMseqs2 (24) for fast sequence alignment. The RCSB search infrastructure is described in (25, 26).

The search applied two criteria simultaneously: (i) a minimum sequence identity of 30% (fold conservation threshold) (18, 19), and (ii) a maximum e-value of 1×10^{-5} (exclusion of random alignments). To compensate for subsequent losses due to metal filtering, the search retrieved three times the number of requested results.

For each candidate, a second query was made to the RCSB GraphQL endpoint to confirm the presence of at least one metal ion of biological interest: Zn, Ca, Fe, Cu, Mn, Ni or Mg. The best result that satisfied both criteria was selected as the reference homolog. A parameter controlled whether any metal found in the homolog was accepted, or whether the user could restrict the search to a specific metal.

If no qualified homolog was found, the protocol stopped and issued a warning. This intentional signal indicated that no experimentally characterized metalloprotein satisfied the defined criteria in the RCSB PDB, and therefore the query required alternative analysis.

Step 3: Metal ion placement with MetalKB

Once the metals were identified, the protocol executed MetalKB (13) for each type of metal. For example, if the homolog had Zn and Cu, MetalKB was executed once for Zn and another for Cu.

Before executing MetalKB, the input PDB was copied to a temporary directory with the name “input.pdb” to avoid problems with long paths. Then, the program was launched with three arguments: the PDB file, the metal type, and the energy threshold. The energy threshold was set to -1.7 kcal/mol, the default value recommended by MetalKB (13).

MetalKB returned many position candidates, organized by energy. The best ones were selected by the protocol until the number of ions specified by the homolog was reached.

Deduplication was also considered as a feature in this process. If a position for an ion was already accepted, a new position closer than 1.5 Å was ruled out, to avoid two different metals in the same position. This threshold is much smaller than the metal-metal distances in real binuclear sites (usually $\geq$3 Å), so true multi-metal sites are not affected (3). Moreover, for each accepted position the coordinates, the metal and the number of found sites and their coordinating residues were saved for the later viewer.

Step 4: Pseudo-holo output generation

The output of this protocol is “pseudo_holo.pdb”, a file that retains the original coordinates of the query protein but adds a final line HETATM for every metal ion placed. Also, it generates “placer_result.json” that summarizes the prediction, including the method used, the PDB of the homolog, the sequence identity and a complete list of placements. In addition, for each metal, the number of expected sites, the number of sites placed, their coordinates, the assigned chain and the coordinating residues were recorded.

2.2.3 Validation

The comparison between predicted and crystallographic positions was performed exclusively on pseudo-apo structures (generated by removing the metal ions from the original holo PDB entry). Because each pseudo-apo structure and the original holo structure share the same protein coordinates, any deviation between the predicted and the real crystallographic position of the metal can be attributed solely to the accuracy of MetalKB, not to the conformation change between both apo/holo structures.

A validation set of 14 experimentally determined holo PDB structures was used to validate this section of the pipeline, with two representatives for each of the seven metals of interest (Zn, Ca, Fe, Cu, Mn, Mg, Ni) representing distinct biological functions. All entries were filtered by crystallographic resolution ($\leq$2.0 Å) to provide reliable metal positions.

Position accuracy was quantified as the Euclidean distance between each predicted metal ion and the nearest crystallographic ion of the same type. For structures with multiple ions, the mean and maximum of these per-ion distances are reported.

Table 1: Validation dataset used for Path 1 metal placement. The table lists the 14 holo structures, including two examples per metal type, the coordinated metal, the number of crystallographic ions, and the experimental resolution.

PDB ID	Protein	Metal	Number of ions	Resolution (Å)
1CA2	Carbonic anhydrase II	ZN	1	2.00
2CTC	Carboxypeptidase A and L-phenyl lactate	ZN	1	1.40
1CLL	Calmodulin	CA	4	1.70
1K9U	Calcium-binding pollen allergen	CA	4	1.75
1BRF	Rubredoxin	FE	1	0.95
1DQI	Superoxide reductase	FE	4	1.70
4AZU	Oxidized *Pseudomonas* azurin	CU	4	1.90
1PLC	Poplar plastocyanin	CU	1	1.33
5A9G	Mn superoxide dismutase	MN	2	1.35
1D3V	Binuclear Mn metalloenzyme arginase	MN	4	1.70
1L3R	cAMP dependent protein kinase	MG	2	2.00
1PHP	Phosphoglycerate kinase	MG	1	1.65
1Q0M	Ni superoxide dismutase	NI	6	1.68
5I8T	Acireductone dioxygenase	NI	1	1.75

2.3 Path 0: structural screening module

Path 0 was designed as an exploratory structural screening module to address a question complementary to the main MetalPlacer workflow: to what extent can lightweight, interpretable structural descriptors identify proteins likely to contain metal-binding sites without relying on GPU-based deep learning methods? Currently, there are cutting-edge methods such as AllMetal3D (11) that use 3D neural networks to address these types of questions. Path 0 does not compete with these methods in terms of metrics but rather provides a lightweight structural approach for a rapid pre-selection process within the MetalPlacer workflow.

The following subsections describe the implementation of each component outlined in Section 2.1.

2.3.1 Candidate site generation with BioMetAll

Candidate sites for metal binding were generated using BioMetAll (9), a tool that identifies sites in a protein structure that are geometrically compatible with metal coordination. BioMetAll creates a regular three-dimensional grid and assesses whether potential donor atoms are arranged at appropriate distances to coordinate a metal ion. The candidate sites were later used as starting points for extracting features for the model.

2.3.2 Feature extraction

For each candidate site, a vector of 62 structural features was computed exclusively from the ATOM records of the structure, organized into four groups. A complete list of feature names is provided in Supplementary Table S1.

Shell composition (42 features). Three concentric spherical shells were defined around each candidate site, with radii 0.0–2.5 Å, 2.5–4.0 Å, and 4.0–5.5 Å, reflecting the multi-layer organization of metal-binding environments (27, 28). Each shell contributes 14 features about residue composition: counts of amino-acid residues grouped by physicochemical type (polar, apolar, positively charged, negatively charged, aromatic…), counts of potential donor atoms by type (2, 29), the number of unique residues within the shell, and the mean and standard deviation of Cα–centroid distances.

Coordination geometry (12 features). Calculated over the atoms in a 3.5 Å radius of the site centroid (3, 27). Donor atom positions were used to characterize the local geometry: the number of donor atoms, distance statistics (minimum, mean, standard deviation), angle statistics for donor–centroid–donor angles (mean, standard deviation, and minimum), the radius of gyration and maximum pairwise distance of donor atoms (30). Two additional binary features indicate whether the donor arrangement is compatible with linear or tetrahedral coordination geometry.

Physicochemical environment (5 features). These features summarize the local chemical environment around each candidate site. The burial proxy estimates how surrounded the candidate site is by nearby residues (31). Local charge was calculated from the balance between positively and negatively charged residues in the innermost shell (28). Polarity was represented as the fraction of polar residues in the two innermost shells, and hydrophobicity as the mean Kyte–Doolittle score (32) of residues in the innermost shell. Together, these descriptors capture whether the surrounding environment is chemically compatible with metal coordination (2, 33).

BioMetAll probe density (3 features). These features summarize the local support provided by BioMetAll (9) for each candidate site. They include the number of BioMetAll probes located within 3 Å and 5 Å of the candidate centroid, and a binary feature indicating whether the candidate coincides with a BioMetAll cluster centroid.

2.3.3 Dataset construction and leakage prevention

Positive examples were taken from MetalPDB (34), filtered to six biologically prevalent metals (Zn, Ca, Fe, Mn, Cu, Ni) and structures with crystallographic resolution $\leq$ 2.5 Å, a standard quality threshold for structure-based metal-binding prediction (1). Mg was excluded from the Path 0 training set because its coordination chemistry is poorly suited to the structural descriptors used by the classifier. Unlike the six included metals, which are predominantly coordinated by protein side-chain atoms (His, Cys, Asp, Glu), Mg binding sites frequently depend on water molecules and cofactors such as ATP for a substantial fraction of their coordination sphere (3). Three types of negative examples were constructed to cover different failure modes of the classifier:

Type A (internal negatives): random Cα positions at least 10 Å from any known metal in metalloprotein structures, up to three per protein. These represent easily distinguishable negatives from the same proteins as the positives.

The non-metalloprotein set used for Type B and Type C negatives was assembled from a combination of manually curated proteins from families annotated as non-metal-binding in UniProtKB and additional structures retrieved via an automated RCSB query (X-ray diffraction, resolution $\leq$ 2.5 Å, at least one protein entity, and zero non-polymer entities). This curated component ensures that the negative set includes biologically diverse proteins, not only clearly distinguishable cases. All entries were verified as metal-free by automated HETATM scanning. After feature extraction, 272 structures were retained and partitioned into train (206), validation (34), and test (32) splits by CD-HIT clustering at 40% sequence identity.

Type B (external negatives): three random Cα positions were sampled from each of the 272 non-metalloprotein structures, yielding 816 sites.

Type C (hard negatives): BioMetAll (9) was executed on the same 272 structures. Cluster centroids were retained as candidate sites, up to five per protein. BioMetAll identified candidates in 267 of the 272 structures, producing 1,325 sites. These are the most challenging negatives because they are geometrically compatible with metal coordination yet come from proteins verified to not contain metal ions.

Table 2: Path 0 training dataset composition by site type.

Type	Description	Sites	Fraction of negatives (approx.)
Positive	MetalPDB sites (ZN, CA, FE, MN, CU, NI; $\leq$2.5 Å)	8,348	–
Neg A (internal)	Random Cα $\geq$10 Å from metal in metalloproteins	5,358	70%
Neg B (external)	Random Cα in verified non-metalloproteins	816	10%
Neg C (hard)	BioMetAll cluster centroids on non-metalloproteins	1,325	20%
Total		**15,847**	

To prevent data leakage, CD-HIT (35) was used with a 40% identity threshold (1) and grouped similar pro-

teins into clusters so that homologous proteins would not appear in different partitions (train/val/test) (36), ensuring a realistic evaluation. The training partition was used to fit the Random Forest model, the validation partition was used for probability calibration and threshold selection, and the test partition was used for the final evaluation of Path 0.

Table 3: Train/validation/test partitioning of the Path 0 dataset.

Partition	Cluster fraction	Number of sites	Role
Training	70%	11,242	Model fitting
Validation	15%	2,062	Model tuning, calibration
Test	15%	2,543	Final evaluation

2.3.4 Binary model training

The binary classifier was trained using the 62-feature matrix generated for Path 0. In this matrix, each row corresponds to a candidate site and each column to one structural feature.

A Random Forest (37) classifier was selected because it is well suited for heterogeneous tabular data without requiring feature normalization (38). In addition, it allows estimation of how important the features are, which facilitates interpretation of how much the structural descriptors contribute to the classification. The model therefore assigns each candidate site a probability of being compatible with metal binding.

Before applying the classifier, median imputation was used to handle missing values. This allows all candidate sites to remain in the dataset without removing examples due to isolated errors in the calculation of certain features. The median was calculated on the training set and then applied to the validation and test sets, preventing information leakage.

Class weights were used to compensate for the slight imbalance between positive and negative values without oversampling or undersampling. This prevents the most frequent class from having a greater influence simply because it appears more often (39, 40).

2.3.5 Hyperparameter optimization

The hyperparameters of the Random Forest (number of trees, maximum depth, etc.) were optimized with RandomizedSearchCV (30 random combinations) (41). Each combination was evaluated in 5-fold cross-validation, using ROC-AUC as the scoring metric.

The cross-validation used GroupKFold with the CD-HIT cluster identifiers as groups, ensuring that homologous proteins were not separated between folds (42). This follows the same anti-leakage logic as in the main train/validation/test split. This optimization was performed only on the training partition. A fixed random seed, 42, was used to ensure reproducibility.

Table 4: Random Forest hyperparameter search space used for Path 0.

Hyperparameter	Candidate values	Description	Selected
`n_estimators`	100, 200, 300, 500	Number of trees in the forest	500
`max_depth`	None, 10, 20, 30	Maximum depth of each tree	20
`min_samples_split`	2, 5, 10	Minimum samples required to split a node	2
`min_samples_leaf`	1, 2, 4	Minimum samples required at each leaf	1
`max_features`	sqrt, log2	Number of features considered at each split	"sqrt"

The final model used the best hyperparameter combination selected by the cross-validation procedure, achieving a cross-validated ROC-AUC of 0.797.

2.3.6 Probability calibration

Although Random Forest produces a probability based on the proportion of trees that vote "yes," these probabilities may not be well calibrated as Random Forest tends to avoid extreme predictions near 0 or 1 (43). Therefore, post-calibration was applied to make the scores more interpretable as probabilities and to facilitate threshold selection.

Calibration was performed using Platt scaling (44), by fitting a univariate logistic regression on the Random Forest scores in the validation set. This sigmoid transformation converts the raw score into a calibrated probability. The test set remained completely separate and was not used to fit either the model or the calibrator.

$$P(y = 1|s) = \sigma(As + B)$$

Where s is the raw Random Forest score, A and B are the fitted calibration parameters, and σ is the sigmoid function. Calibration quality was assessed using the Brier score (45) and Expected Calibration Error (46).

Platt scaling reduced the expected calibration error from 0.053 to 0.036 (32% relative reduction), confirming that the calibrated probabilities better reflect the true positive rates and are suitable for threshold-based decision-making.

2.3.7 Threshold selection

Path 0 produces scores that can be evaluated at two different levels: at the site level, determining whether each candidate resembles a metal-binding site, and at the protein level, combining all the scores into a single protein-level score.

Each candidate site received a score between 0 and 1, representing how closely that site resembles an actual metal-binding site. Since this probability had to be converted into a binary decision, a site-level threshold called τ_{site} (0.76) was used.

This threshold was selected on the validation set (47) by maximizing MCC, which considers all four terms of the confusion matrix and is more informative than accuracy when classes are not perfectly balanced (48).

τ_{site} measures how the model performs on a site-by-site basis, but it does not filter out sites before calculating the protein-level score. Therefore, it does not affect the pipeline's final decision.

To answer whether the query protein is likely to be a metalloprotein, the scores of all candidate sites were combined into a single protein-level score. To determine this, the 95th percentile was used, avoiding the use of the maximum alone, which could be affected by an outlier, or the mean, because many low-scoring candidate sites could mask the presence of a real metal-binding region.

A separate protein-level threshold, τ_{protein}, was created so that this preliminary filter would not be excessively strict. For this reason and to create flexibility, different operating points were characterized on the validation set: permissive (0.51), moderate (0.60), and high-specificity (0.68). The moderate threshold ($\tau = 0.60$) was selected as the default operating point. The rationale for this choice is discussed in Section 4.2.

See Table 8 in Results for more details.

2.3.8 Evaluation metrics

Path 0 was evaluated at both the candidate site level and the full-protein level. The site-level evaluation measures whether the classifier correctly distinguishes actual metal-binding sites from negative sites. In contrast, the full-protein-level evaluation measures whether, after aggregating the scores of all candidate sites using p95, the full protein is correctly classified as a probable metalloprotein or a non-metalloprotein.

All thresholds were selected exclusively from the validation set. The test set was kept separate and was used only once for the final evaluation of the pre-determined operating points (47).

To analyze performance, both ranking-based and threshold-based metrics were used. ROC-AUC and PR-AUC (49) assess the model's ability to rank positives above negatives without relying on a single threshold (50). Sensitivity, specificity, precision, F1 score, MCC, balanced accuracy, and confusion matrices were used to evaluate the selected decision thresholds (48).

2.4 Scipion integration

MetalPlacer was integrated into the Scipion framework (16) as a plugin composed of two main protocols, one for Path 0 and another for Path 1. The objective of this integration was not only to run the different scripts of the pipeline, but also to make the workflow easier to use, reproduce and connect with other structural biology tools.

In Scipion, a protocol is a module that receives an input, allows the user to define some parameters, runs a series of steps and saves the corresponding outputs inside the project. For this reason, integrating MetalPlacer required adapting the Path 0 and Path 1 workflows to the Scipion structure. This included defining the input protein structure, exposing the main parameters to the user, executing the external

tools and Python scripts, and registering the final outputs in a format that could be used by other Scipion protocols.

This design allows Path 0 and Path 1 to be used together or independently. In the complete workflow, Path 0 can first screen the protein and, if the result is positive, pass the structure directly to Path 1. However, Path 1 can also be launched directly when the user already wants to perform the homology-based metal placement.

The Path 0 protocol, ProtMetalScreener, takes a PDB or PDBx/mmCIF protein structure as input and allows the user to modify the protein-level decision threshold. It returns the protein-level probability (pMetal), the binary decision (isMetal) and the number of candidate sites scored (nSitesScored). When the prediction is positive, the input structure is also registered as an output object that can be directly connected to Path 1. The Path 1 protocol, ProtMetalPlacer, takes an apo or pseudo-apo structure as input, performs the homology-based identification and MetalKB-based placement described in Section 2.2, and generates a pseudo-holo PDB file together with a JSON metadata file containing the selected homolog, sequence identity, metal type, ion count and predicted coordinates. Therefore, the Scipion implementation connects the screening, identification and placement steps into a single traceable workflow while still preserving the possibility of running each protocol separately.

The Path 1 protocol includes a dedicated viewer that opens the pseudo-holo structure in PyMOL or ChimeraX, displaying each placed ion as a colored sphere and the coordinating residues as sticks with distance lines to the relevant sidechain atoms.

3 Results

3.1 Path 1. Homology-based identification and placement

To verify that the integrated Path 1 workflow correctly executes the entire process, it was applied to a set of 14 experimental holo structures. The set includes proteins for each of the seven metals of interest (Zn, Ca, Fe, Cu, Mn, Mg, Ni), selected based on a crystallographic resolution of 2.0 Å or less, and within each metal, ensuring a diversity of coordination geometries.

For each structure, the metal ions were removed to generate a pseudo-apo form. The predicted ion positions were compared with the original crystallographic positions only to verify that the integrated workflow produced structurally consistent outputs, this was done by measuring the Euclidean distance between each placed ion and the nearest crystallographic ion of the same type.

Table 5 summarizes the results obtained for the 14 structures. In all 14 cases, the workflow successfully retrieved a homolog with 100% sequence identity, indicating that the homology search was correctly executed in this validation set. Five of the selected homologs correspond to the query structure itself (1K9U, 1D3V, 5A9G, 1PLC, 1PHP), which is methodologically valid given that the objective is to verify the workflow's functionality. Of the 14 structures, 12 resulted in the placement of at least one ion, while the remaining 2 (both Mg) did not lead to any placements.

Among the 12 structures with successful placement, 11 had a mean distance of less than 0.6 Å, confirming sub-angstrom accuracy in most cases. The lowest per-ion distances were observed for Zn and Fe examples, with values of 0.060 Å for 1CA2 and 0.044 Å for 1BRF, respectively. The Cu and Mn sites exhibited somewhat higher distances (range 0.258–0.714 Å). The highest per-ion distance among the sub-angstrom cases was observed for 1D3V (Mn), with a mean of 0.572 Å and a maximum of 0.714 Å. 1Q0M (Ni) showed the highest mean per-ion distance of all the proteins, at 2.193 Å.

In most structures, the number of ions placed matched the number of crystallographic ions present in the processed chain. In 1DQI (2/4) and 4AZU (1/4), the workflow placed fewer ions than the total number present in the asymmetric unit, in both cases, the placed ions correspond to those in chain A. In 1Q0M, the workflow placed 3 of the 6 crystallographic ions, making it the only case of partial placement among the structures that produced results.

The two Mg structures (1PHP and 1L3R) did not yield any assigned ions. In both cases, the error occurred in the MetalKB step (Step 3), not in the homolog search (Step 2), which worked correctly. For 1PHP, the workflow correctly identified the metal (Mg) based on a self-hit, but MetalKB did not find any coordination sites above the energy threshold of -1.7 kcal/mol. For 1L3R, the retrieved homolog (1ATP) contains Mn instead of Mg, which resulted in an incorrect identification of the metal type. MetalKB searched for Mn sites and also found no candidates above the threshold.

Table 5: Functional validation of Path 1 on the 14 pseudo-apo structures. For each query structure, the table reports the metal type, selected homolog, sequence identity, number of ions placed relative to the crystallographic reference, per-ion distance values, mean distance and maximum distance. Dashes indicate cases where no ion was placed. Exp refers to the number of crystallographic ions of the corresponding metal in the query structure.

PDB	Metal	Homolog	Identity	Ions placed/Exp	Per-ion distance (Å)	Mean distance (Å)	Max distance (Å)
1CA2	ZN	1A42	100%	1/1	0.060	0.060	0.060
2CTC	ZN	1HDQ	100%	1/1	0.099	0.099	0.099
1CLL	CA	12DK	100%	4/4	0.117, 0.331, 0.175, 0.116	0.185	0.331
1K9U	CA	1K9U	100%	4/4	0.029, 0.160, 0.094, 0.150	0.108	0.160
1BRF	FE	1BQ8	100%	1/1	0.044	0.044	0.044
1DQI	FE	1DO6	100%	2/4	0.189, 0.349	0.269	0.349
4AZU	CU	1AZU	100%	1/4	0.312	0.312	0.312
1PLC	CU	1PLC	100%	1/1	0.258	0.258	0.258
5A9G	MN	5A9G	100%	2/2	0.377, 0.510	0.444	0.510
1D3V	MN	1D3V	100%	4/4	0.714, 0.536, 0.532, 0.506	0.572	0.714
1PHP	MG	1PHP	100%	0/1	–	–	–
1L3R	MG	1ATP	100%	0/2	–	–	–
1Q0M	NI	1Q0G	100%	3/6	2.169, 2.221, 2.188	2.193	2.221
5I8T	NI	1VR3	100%	1/1	0.112	0.112	0.112

Figure 4 shows three representative examples of the workflow. Panels A and B illustrate the overlap between the predicted positions (orange spheres) and the crystallographic positions (teal spheres) for 1CA2 (Zn, one ion, per-ion distance 0.060 Å) and 1K9U (Ca, four ions, mean per-ion distance 0.108 Å), where both sets of positions are virtually indistinguishable. Panel C shows 1Q0M (Ni-SOD), where sites without placement (teal spheres without an adjacent orange sphere) are visible, along with a magnified view of one of the placed ions, which is displaced by 2.19 Å from the crystallographic position.

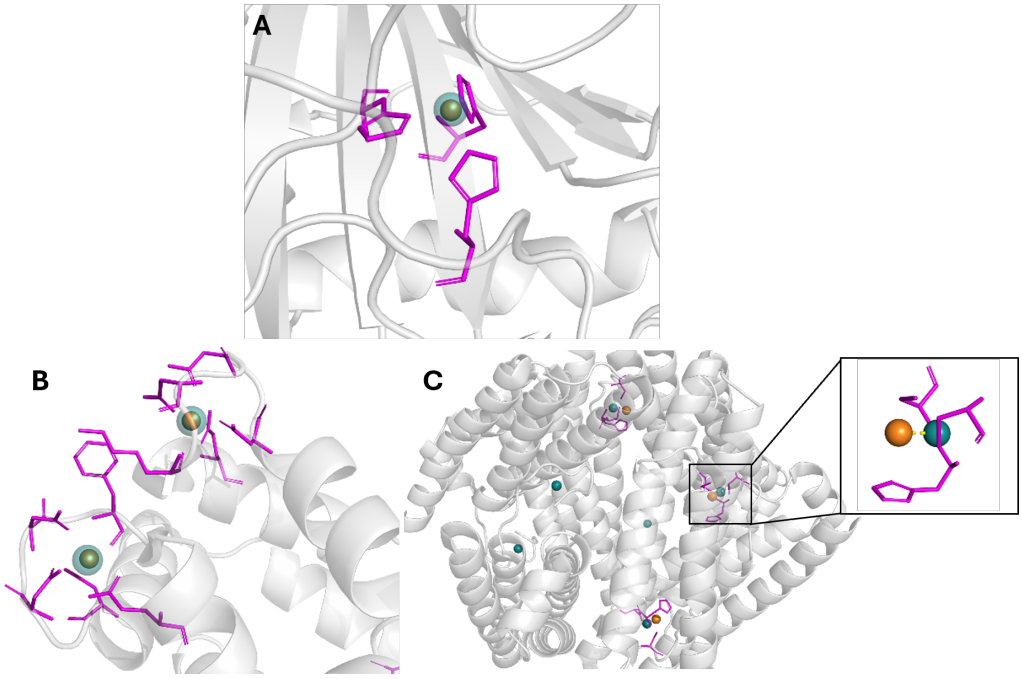

Figure 4: Representative outputs of the integrated Path 1 workflow. Predicted ions are shown as orange spheres and crystallographic reference ions as teal spheres. Coordinating residues identified by the workflow are shown as magenta sticks, with the protein backbone represented as a semitransparent cartoon. (A) 1CA2 shows an excellent Zn placement with a per-ion distance of 0.060 Å. (B) 1K9U illustrates a multi-ion Ca case (4/4 ions placed, mean per-ion distance 0.108 Å), two of the four placed ions are shown in detail. (C) 1Q0M (Ni): only 3 of 6 crystallographic ions were placed. Inset: detail of one placed ion showing a displacement of 2.19 Å from the crystallographic position.

Taken together, these results show that Path 1 produced at least one metal placement in 12 of the 14 evaluated structures, with sub-angstrom mean per-ion distance in most successful cases. The two Mg failures and the partial placement of 1Q0M define the system's current boundaries.

3.2 Path 0. Overall classifier performance

Table 6 summarizes the ranking and calibration metrics of the Path 0 classifier across the three data partitions.

The training partition showed a ROC-AUC of 1.000. The validation and test partitions showed ROC-AUC values of 0.790 and 0.791, respectively. The Brier score was 0.181 in validation and 0.184 in test, and the ECE was 0.036 in validation and 0.050 in test. Detailed threshold-based metrics at the site and protein levels are presented in Sections 3.3 and 3.4.

Table 6: Ranking and calibration metrics of the Path 0 classifier across data partitions (site level, post-Platt scaling). Partitions were defined by CD-HIT clustering at 40% sequence identity.

Split	n (sites)	n (proteins)	ROC-AUC	PR-AUC	Brier	ECE
Train	11,242	1,430	1.000	1.000	–	–
Val	2,062	303	0.790	0.802	0.181	0.036
Test	2,543	325	0.791	0.838	0.184	0.050

3.3 Path 0. Site-level performance

Figure 5 shows the ROC and precision-recall curves for the site-level classifier on the test set, with a ROC-AUC of 0.791 and a PR-AUC of 0.838.

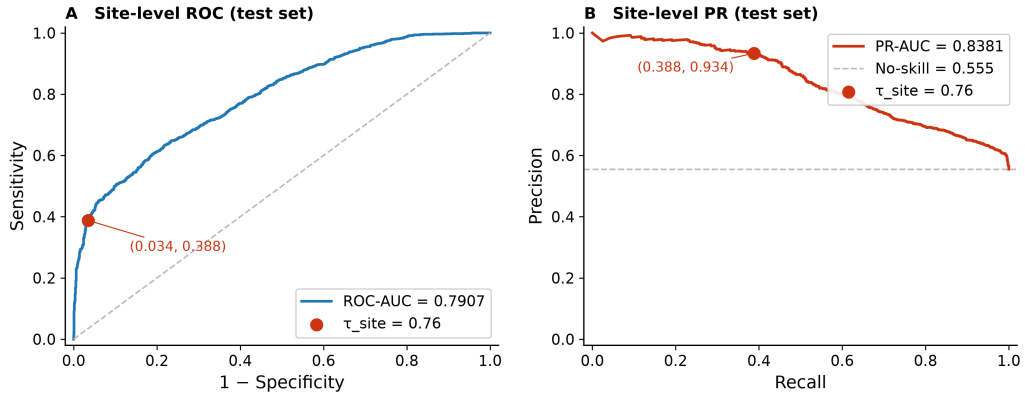

Figure 5: Site-level ROC (A) and precision-recall (B) curves on the test set. The selected operating point ($\tau_{\text{site}} = 0.76$) is marked in both panels.

At the selected operating point ($\tau_{\text{site}} = 0.76$), the model reached a precision of 0.934 and a specificity of 0.966, with a sensitivity of 0.388 (Table 7).

Table 7: Site-level classification metrics on the test set at $\tau_{\text{site}} = 0.76$ (n=2543 sites, 1412 positive, 1131 negative). Sens: sensitivity; Spec: specificity; Prec: precision; MCC: Matthews correlation coefficient; balanced acc.: balanced accuracy.

Dataset	Threshold	TP	TN	FP	FN	Sens	Spec	Prec	F1	MCC	Balanced acc.	ROC-AUC	PR-AUC
Test	0.76	548	1092	39	864	0.388	0.966	0.934	0.548	0.417	0.677	0.791	0.838

3.4 Path 0. Protein-level performance

After aggregating the scores from the individual sites at the protein level using the 95th percentile, the classifier achieved a ROC-AUC of 0.777 on the test set (Figure 6). Three operating points on the ROC curve were evaluated: permissive ($\tau = 0.51$), moderate ($\tau = 0.60$, default), and high-specificity ($\tau = 0.68$).

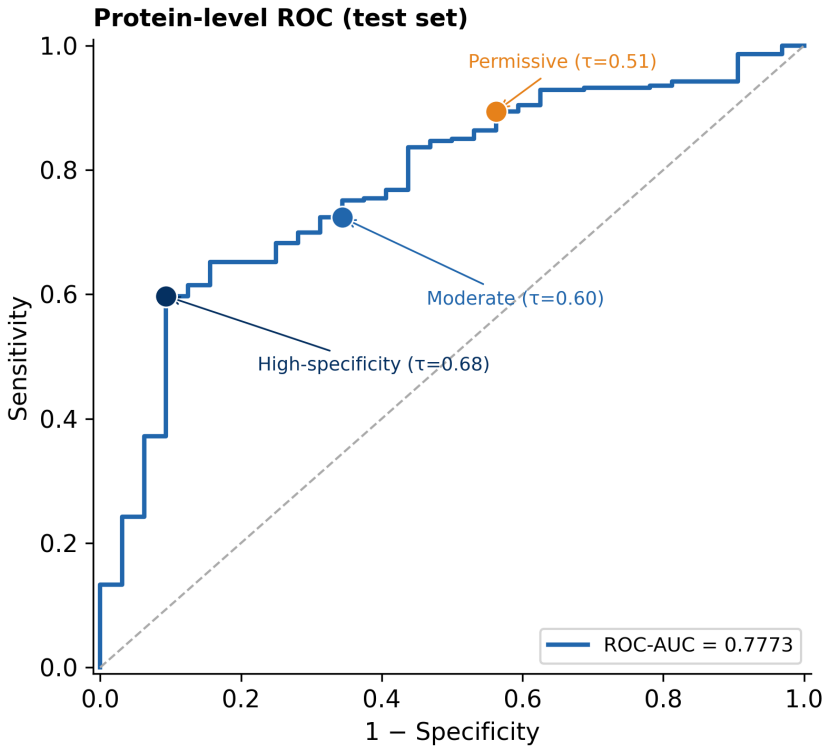

Figure 6: Protein-level ROC curve on the test set (ROC-AUC = 0.777). Three operating points are marked: permissive ($\tau = 0.51$), moderate ($\tau = 0.60$, default), and high-specificity ($\tau = 0.68$).

Table 8 summarizes the metrics for the three operating points. The permissive setting achieved the highest sensitivity (0.894) with a specificity of 0.438. The moderate setting had a sensitivity of 0.727 and a specificity of 0.656. The high-specificity setting achieved 0.906 specificity with a sensitivity of 0.590.

Table 8: Protein-level classification metrics at three operating points on the test set (n = 325 proteins; 293 metalloproteins, 32 non-metalloproteins). Thresholds were selected on the validation set and applied unchanged to test.

Operating point	Threshold	TP	TN	FP	FN	Sens	Spec	Prec	F1	MCC	Balanced acc.
Permissive	0.51	262	14	18	31	0.894	0.438	0.936	0.915	0.286	0.666
Moderate (default)	0.60	213	21	11	80	0.727	0.656	0.951	0.824	0.247	0.692
High-specificity	0.68	173	29	3	120	0.590	0.906	0.983	0.738	0.297	0.748

Figure 7 shows the evolution of sensitivity, specificity, and MCC as a function of τ_{protein}. Sensitivity decreases progressively as the threshold increases, while specificity rises in steps. The MCC reached its highest value in the operating point “high-specificity” (MCC = 0.297).

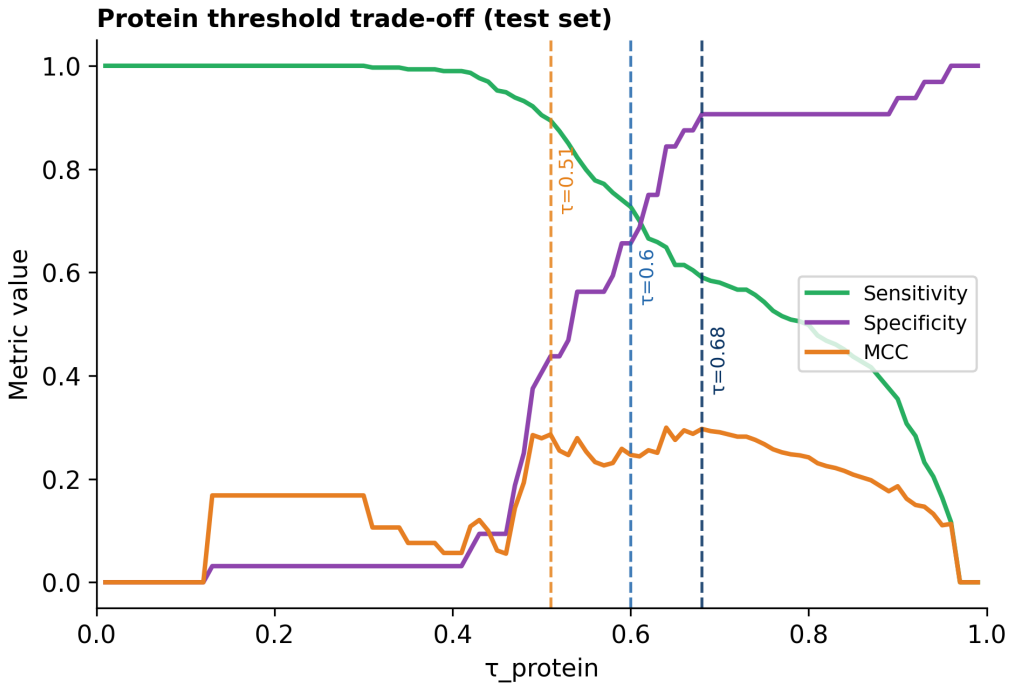

Figure 7: Sensitivity, specificity, and MCC as a function of τ_{protein} on the test set. Dashed vertical lines indicate the three operating points.

Figure 8 shows the confusion matrix at the protein level for the default operating point ($\tau = 0.60$). The model correctly classified 213 metalloproteins (TP) and 21 non-metalloproteins (TN), with 11 false positives and 80 false negatives.

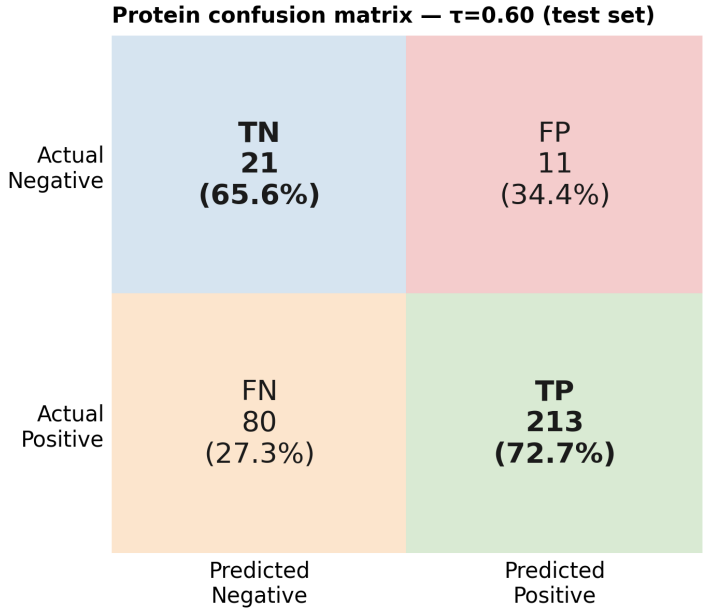

Figure 8: Protein-level confusion matrix at the default operating point ($\tau = 0.60$) on the test set.

3.5 Feature importance

To evaluate which variables contribute most to the model's performance, permutation importance was calculated on the test set by measuring the reduction in ROC-AUC when each feature was randomly shuffled (n = 10 repetitions).

Figure 9 shows the 20 most informative features. The features with the greatest impact were `coord_min_dist` (coordination geometry) and `bm_probe_density_3A` (BioMetAll), both with an average reduction of 0.011 in ROC-AUC. All four feature groups (coordination geometry, shell composition, BioMetAll, and physicochemistry) were represented in the top 20 (Supplementary Table S2).

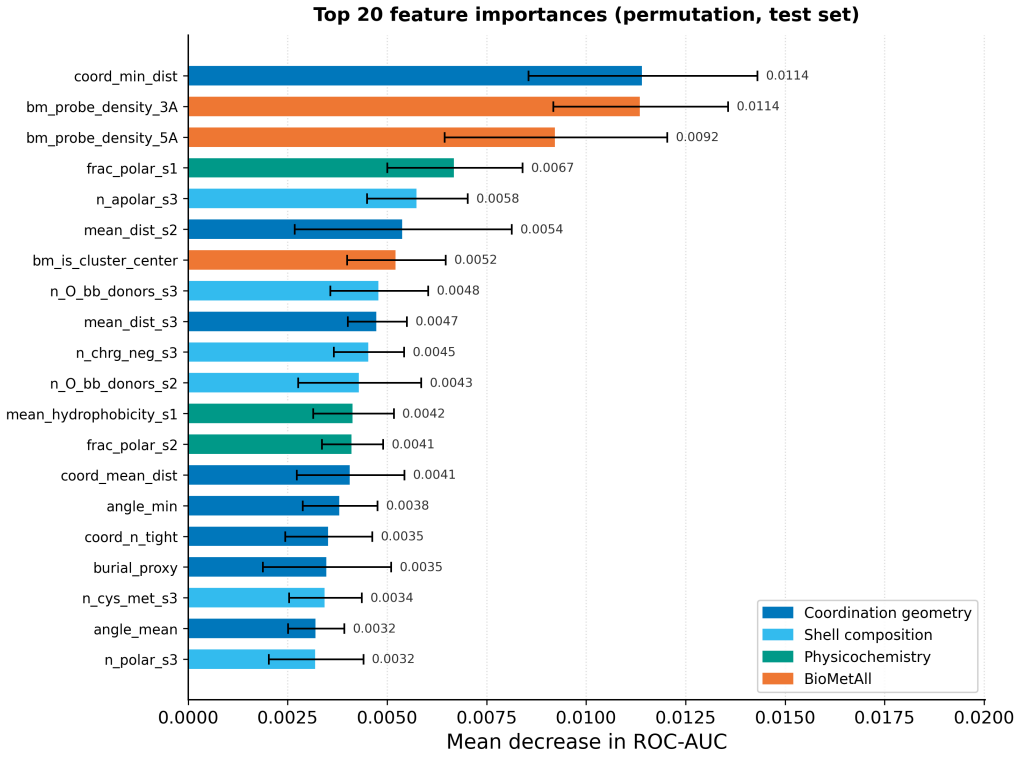

Figure 9: Top 20 features by permutation importance (mean decrease in ROC-AUC, test set, n_repeats = 10). Bars are colored by feature group; error bars show standard deviation across repetitions.

3.6 Scipion integration

The two modules developed (ProtMetalScreener and ProtMetalPlacer) were integrated as Scipion protocols, allowing the entire pipeline to be run from the framework's graphical interface.

Figure 10 shows the execution tree of a project with the three chained protocols (import, screening, and placement), the integrated PyMOL viewer displaying the placement result, and the summary automatically generated by the protocol, which includes the selected homolog, the sequence identity, the coordinates of the placed ion, and the identified coordinating residues.

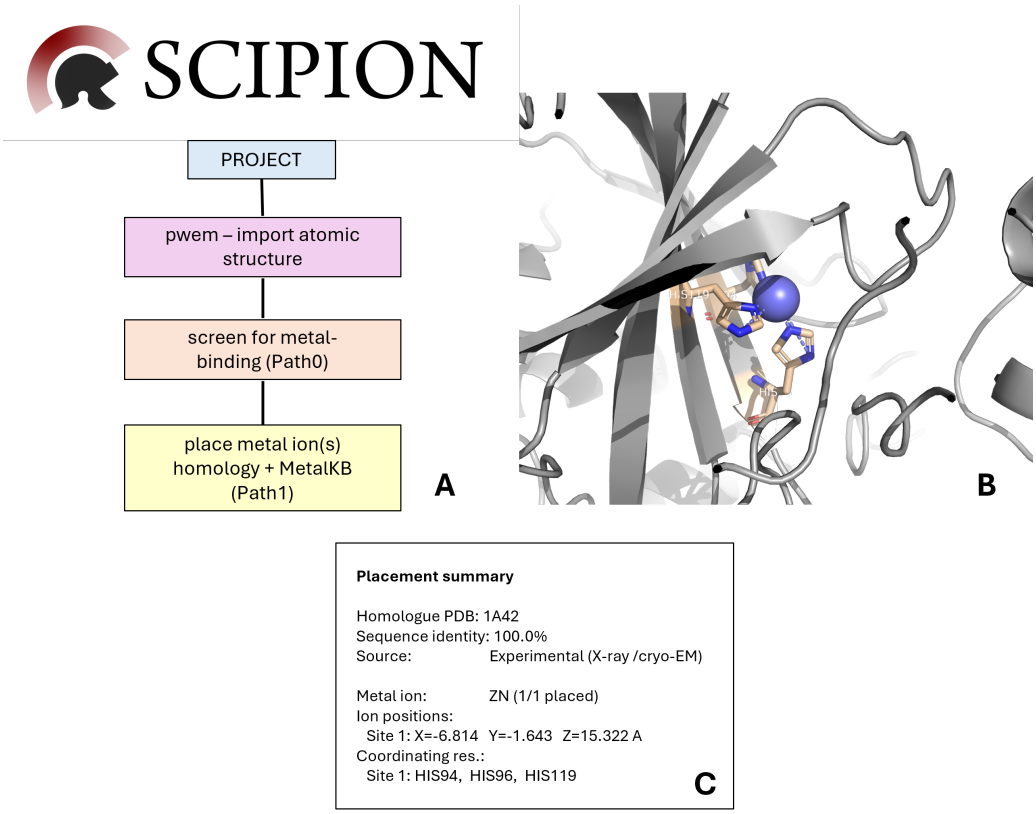

Figure 10: MetalPlacer pipeline in Scipion. (A) Protocol execution tree. (B) Built-in PyMOL viewer showing the placed Zn ion (blue sphere) in carbonic anhydrase II (1CA2). (C) Placement summary output.

4 Discussion

4.1 Path 1: Capabilities and limitations of homology-based placement

The Path 1 workflow successfully placed at least one metal ion in 12 of the 14 structures evaluated, with an average per-ion distance of less than 0.6 Å in 11 of the 12 successful cases (Table 5). The lowest per-ion distance values were observed for Zn and Fe (0.044–0.099 Å), metals that tend to have rigid and geometrically well-defined coordination sites, for which MetalKB reports the best precision and recall values in its multimetal evaluation (Figure 5c from MetalKB (13)). The Cu and Mn sites showed somewhat higher distances (0.258–0.714 Å), which is consistent with the greater geometric diversity of these coordination environments (3) and with the slightly lower performance that MetalKB achieves for these metals in the same benchmark. It is to be expected that MetalPlacer's per-ion distances are generally low, given that validation was performed using 100% identity homologs, which eliminates conformational variability between the query and the template. However, the agreement between the metal-specific pattern observed in MetalPlacer and that reported by MetalKB confirms that integration into the Scipion workflow preserves the behavior of the underlying tool without introducing additional deviations.

Systematic failure of Mg. The two Mg structures (1PHP and 1L3R) represent the main failure mode observed, and both failed in the MetalKB step, not in the homolog search. In 1PHP, the workflow correctly identified the metal (Mg) based on a self-hit, but MetalKB did not find any coordination sites above the energy threshold of -1.7 kcal/mol. This result can be explained in light of a limitation acknowledged by the authors of MetalKB themselves: the statistical potentials were derived exclusively from metal-protein interactions, without including small molecules, and cases with a protein-metal coordination number of less than three were not specifically handled (13). Mg frequently coordinates with water molecules and cofactors such as ATP, so its binding sites often have fewer than three direct protein ligands, falling outside the representativeness domain of the MESPEUS potentials (14). In 1L3R, the problem predated the placement: the retrieved homolog (1ATP) is annotated with Mn instead of Mg, which led to an erroneous identification of the metal type. This case reveals an additional limitation of the homology-based approach, as proteins with high sequence identity can coordinate different metals depending on experimental conditions.

The difficulty with Mg is not unique to MetalPlacer. In the multi-metal evaluation of MetalKB (Figure 5c in MetalKB (13)), non-transition metals (Na, K, and Mg) perform the worst in both MetalKB and Metal3D, consistently ranking below transition metals in terms of precision and recall (10, 13). This background places the failure of Mg in MetalPlacer within a recognized limitation of the field, rather than as a defect specific to the pipeline.

Partial placement and non-canonical coordination (1Q0M). The 1Q0M (Ni-SOD) structure was the most problematic case in the validation, with two overlapping issues: partial placement (3 out of 6 ions) and high per-ion distance for the placed ions (mean 2.193 Å). Both problems share the same root cause: the atypical coordination of Ni in Ni-SOD. At this site, known as the "Ni-hook," (51) the metal is coordinated by the free amino nitrogen at the N-terminus (His-1), the amide nitrogen of the Cys-2 peptide bond, and the thiolates of Cys-2 and Cys-6, in a square-planar arrangement. These types of donors (amines and amides from the peptide backbone) are underrepresented in the MESPEUS statistical database (14), which has two direct consequences: the statistical potential does not distinguish these sites well relative to the energy threshold (-1.7 kcal/mol), and when it does accept them, the energy minimum is

displaced from the actual crystallographic position. The six sites of the homohexamer are chemically identical, but small numerical asymmetries in the potential evaluation cause only three to exceed the threshold. An additional factor is that the homolog used (1Q0G) contains two hexamers in its asymmetric unit, which leads the workflow to estimate 12 expected ions instead of 6, exacerbating the appearance of incompleteness in the result.

In contrast, the second Ni structure evaluated (5I8T, per-ion distance 0.112 Å) exhibits a conventional His_3-Asp coordination, for which MESPEUS has abundant training data, and the placement was accurate. This comparison defines the scope of MetalKB's applicability: placement works well with canonical geometries (tetrahedral, octahedral), but performance degrades for non-standard coordinations that are underrepresented in its statistical potentials (13).

Stoichiometry and global count of the homolog. The workflow infers both the type and number of ions to be placed based on the selected homolog, extracting the count from the homolog's complete asymmetric unit. This approach can lead to discrepancies in either direction depending on the relationship between the homolog and the query.

When the recovered homolog is a version with fewer chains than the query, the ion count is lower than expected. In 4AZU (tetramer, 4 Cu), the recovered homolog was 1AZU (monomer, 1 Cu), so the workflow attempted to place only one ion and did so correctly (per-ion distance 0.312 Å). Similarly, in 1DQI (4 Fe, tetramer), the homolog 1DO6 corresponds to a dimer with 2 Fe, and the workflow placed exactly those 2 ions. In both cases, the placement was successful for the requested ions; the discrepancy does not reflect a placement failure but rather an inherent limitation of homolog-based counting, which does not necessarily reproduce the query's oligomeric state.

Conversely, when the homolog contains more ions than the query, the count may overestimate the metal content. In 1CLL, the homolog (12DK) contains 8 Ca ions in its asymmetric unit compared to the 4 in the query's biological chain; the workflow attempted to place all 8, although only 4 found positions with crystallographic meaning.

These observations reveal that ion counting is a source of error independent of placement accuracy. In cases of underestimation, the workflow could adjust the count to match the query's oligomeric state; in cases of overestimation, post-placement filtering based on energy or spatial proximity between placed ions would allow the excess ions to be discarded. Both improvements are proposed as future work.

Validation limitation: 100% identity. All homologs retrieved during validation showed 100% sequence identity, which represents a favorable scenario that does not reflect the pipeline's typical operating conditions. This choice was deliberate: by using identical homologs, any deviation in the per-ion distance is attributed exclusively to the placement step (MetalKB), isolating it from the uncertainty introduced by sequence divergence. However, future validation with homologs in the 30–70% identity range would be necessary to evaluate the workflow's performance under more realistic conditions, where conformational variability between the query and homolog could affect both metal identification and placement accuracy.

4.2 Path 0: Performance, justifications, and limitations

Overall model performance. The Path 0 classifier achieved an ROC-AUC of 0.791 at the site level and 0.777 at the protein level (Figures 5 and 6). The model discriminates above chance, but does not match

current sequence-based classifiers designed for the same binary task. For example, M-Ionic reports an ROC-AUC of 0.83 using numerical representations learned by a protein language model from sequence alone (7), and LMetalSite showed that pretrained language models can learn structural context implicitly from sequence, surpassing even some structure-based methods (8). These figures are not directly comparable, as they were obtained on different datasets and tasks, they are cited only to contextualize the expected performance range. This gap is expected, language models trained on hundreds of millions of sequences capture evolutionary and structural signals (52) that are not available to Path 0, which relies on 62 structural descriptors from a single 3D structure. Path 0 does not aim to match these methods in classification accuracy. Instead, it offers three complementary advantages:

- **Interpretability:** the contribution of each structural descriptor is directly quantifiable (Figure 9), whereas language model representations are difficult to interpret at the individual feature level.
- **Conformation-specificity:** Path 0 evaluates the actual 3D geometry of the input structure, which is relevant when assessing a specific conformer from molecular dynamics.
- **Pipeline integration:** Path 0 functions as a lightweight, GPU-independent pre-screening module that passes positive results directly to Path 1 for metal placement.

A detailed comparison with Metal3D, which addresses a different task (metal ion localization), is provided in Section 4.3.

Justification for the "moderate" threshold as the default. The high-specificity operating point had the highest MCC (0.297 compared to 0.247 for moderate, Table 8). However, Path 0 functions as a pre-filter for Path 1: a false negative means that a metalloprotein never reaches the placement module and is permanently lost, whereas a false positive only generates additional computational work in Path 1. In a screening context, prioritizing sensitivity (0.727 for "moderate" versus 0.590 for "high-specificity") is more appropriate than maximizing the MCC, since the asymmetric cost of errors favors minimizing false negatives.

Limitation of the negative test set. Protein-level specificity was calculated based on only 32 non-metalloproteins in the test set (TN + FP = 32 in Table 8). With such a small sample size, ± 1 misclassified protein alters specificity by approximately $\pm 3\%$, and the MCC values are inherently unstable. This imbalance (293 metalloproteins versus 32 non-metalloproteins) reflects the natural composition of the PDB, where metalloproteins are overrepresented among the deposited structures, but it limits confidence in the specificity estimates (34).

Beyond sample size, the composition of the negative set may influence the reported metrics. Approximately 70% of negative examples are Type A (random Cα positions in metalloproteins), which are structurally far from any metal-binding site and easy to classify. Only 20% are Type C hard negatives (BioMet-All candidates on verified non-metalloproteins). This distribution was not a deliberate design choice but a consequence of the available data: the non-metalloprotein set was limited to 272 structures after processing failures, which restricted the number of Type B and Type C sites that could be generated. In contrast, Type A negatives are abundant because they are sampled from the same metalloproteins used for positives, where each structure can contribute up to three random sites. As a result, the reported ROC-AUC of 0.791 may be higher than what would be obtained with a more challenging negative set.

Generalization: train/val/test stability. The ROC-AUC on the training set was 1.000, an expected result for a Random Forest model without strict depth constraints, which tends to memorize the training data (37). However, the validation (0.790) and test (0.791) ROC-AUCs are virtually identical, confirming that no overfitting to the validation set occurred during the selection of hyperparameters and thresholds. This stability can be attributed to the 40% cluster partitioning of CD-HIT using GroupKFold, which ensures that homologous proteins do not appear in different partitions (35, 42). The calibration remained reasonably stable across partitions, from validation to test (ECE 0.036 to 0.050; Brier 0.181 to 0.184), with a slight increase on the test set as expected for a calibrator fitted on a separate partition.

Feature importance: what the model learns. The two most important features were `coord_min_dist` (coordination geometry) and `bm_probe_density_3A` (BioMetAll probe density), both of which resulted in an average reduction of 0.011 in ROC-AUC (Figure 9). The fact that all four feature groups (coordination geometry, shell composition, BioMetAll, and physicochemistry) are represented in the top 20 indicates that the model integrates diverse and complementary signals. The presence of BioMetAll features among the most informative ones confirms that probe density provides information complementary to the pure geometry of the donor atoms, validating the design decision to include these BioMetAll-derived features (9). No single feature dominates overwhelmingly (maximum drops of $\sim$0.011), which is typical behavior for a Random Forest with correlated features and suggests that the model does not critically depend on any single descriptor (37).

A per-metal breakdown of the test set (Supplementary Table S3) shows that performance varies across metal types: Cu achieved the highest protein-level sensitivity (0.867), followed by Zn (0.778) and Ca (0.772), while Ni was the lowest (0.574). These differences are consistent with the coordination chemistry of each metal. Cu and Zn coordinate through geometrically well-defined side-chain donors that are well captured by the structural descriptors, whereas Ni exhibits more diverse coordination environments.

4.3 Comparison with the state of the art

Comparison with Metal3D. Metal3D is currently the most accurate predictor for zinc ion localization, with predictions within 0.70 $\pm$ 0.64 Å of the experimental positions (10). Its multi-metal extension (AllMetal3D) extends this capability to multiple types of ions (11). MetalPlacer does not compete with these methods in terms of pure predictive performance. Its contribution is of a different nature: it provides a pipeline integrated into Scipion that combines structural screening (Path 0) with homology-based identification and placement based on statistical potentials (Path 1), without requiring a GPU and offering three configurable operating modes. Furthermore, while Metal3D requires specifying the metal type a priori, MetalPlacer's Path 1 automatically infers the type and number of ions from the homolog.

Comparison with AlphaFill. AlphaFill is conceptually the method most similar to Path 1: it transfers ligands and ions from homologous experimental structures to AlphaFold models via structural superposition (5). The main difference is that AlphaFill directly transfers the ion's coordinates from the template via superposition, whereas Path 1 separates the inference of the metal's identity/number (from the homolog) from the calculation of its spatial position (using MetalKB in the query geometry). This separation allows, in principle, for the placement to adapt to local conformational differences between the template and the query, although this theoretical advantage has not been experimentally evaluated

in the present work. Furthermore, MetalPlacer is integrated into Scipion as a reproducible workflow, whereas AlphaFill operates as a standalone web service.

Comparison with AlphaFold 3. AlphaFold 3 can predict some metal ions together with the protein structure (6), but it requires the user to specify the ion type as input and its metal placement accuracy has not been systematically compared against experimental positions. In addition, AlphaFold 3 is not available for local execution, which prevents its integration into reproducible workflows such as Scipion. A direct comparison with MetalPlacer is therefore not possible at this time.

Comparison with MADE. The MADE approach also uses homologous structures to identify metal ion positions through superposition and density clustering (12). Unlike MetalPlacer, MADE overlays multiple homologs simultaneously and searches for high-density clusters, which gives it statistical robustness but requires enough homologs. MetalPlacer uses a single selected homolog and delegates fine-tuning to MetalKB, making it operational even with only a single available template.

Comparison with BioMetAll. BioMetAll generates candidates that are geometrically compatible with metal coordination based on the pre-organization of the protein backbone, but it does not classify sites or place ions (9). MetalPlacer uses it as a candidate-generation component in Path 0 and adds classification (Random Forest) and placement (MetalKB). The relationship is one of pipeline extension, not competition.

Comparison with PinMyMetal (PMM). PMM predicts transition metal locations and coordinating residues directly from the input structure without requiring a homologous template (15). MetalKB was compared against PMM in its original publication, showing comparable spatial accuracy and a higher F1 score on the TEMSP zinc benchmark (13). The two methods address different aspects of the problem: PMM localizes metal sites from the structure alone but does not determine the specific metal identity, whereas MetalPlacer infers the metal type and count from a homolog and delegates placement to MetalKB. These approaches are therefore complementary rather than alternatives.

4.4 Scipion integration

One of the central objectives of this work was to integrate MetalPlacer into the Scipion framework as a plugin consisting of two protocols (ProtMetalScreener and ProtMetalPlacer) (16). This integration is not merely a bundling of scripts; rather, it offers several concrete advantages. First, Scipion provides complete traceability: each run is logged along with its parameters, software versions, and input/output files, allowing any result to be reproduced exactly (17). Second, interoperability with other Scipion protocols allows MetalPlacer to be chained with cryo-EM processing, atomic refinement, or structural validation tools within the same project. Third, Scipion's graphical interface lowers the barrier to entry for users who are not command-line experts by displaying only the relevant parameters (such as the Path 0 decision threshold) and hiding the complexity of the underlying pipeline.

Integration into a workflow framework is particularly relevant in the current context of structural biology, where the proliferation of models predicted by AlphaFold that lack metal ions and cofactors creates a growing need for automated tools for the preparation of pseudo-holo structures (5). Having a Scipion protocol that performs this preparation in a standardized and reproducible manner represents a practical contribution to the computational structural biology tool ecosystem.

A minor visual bug was identified in the viewer: the label for Ca (calcium) ions is confused with Cα backbone atoms. This does not affect the placement calculation or output files but should be corrected in a future release.

4.5 General limitations

Beyond the metal-specific cases discussed in Section 4.1, two broader limitations should be noted:

- **Validation scope.** The current validation of Path 1 covers only 14 structures with 100% identity homologs and seven metal types. This is sufficient to verify the workflow's functionality but does not constitute a large-scale benchmark comparable to those used by MetalKB (13) or Metal3D (10). The performance of the pipeline under realistic conditions (lower identity homologs, diverse protein families, predicted structures) remains to be established.

The set of 14 structures was chosen to cover all seven target metals with two representatives each, prioritizing diversity of coordination environments over sample size.

- **Dependence on homolog availability.** Path 1 requires at least one experimentally characterized homolog with an annotated metal ion in the RCSB PDB. For proteins with no detectable homolog above 30% identity, or for metals not annotated in the PDB entry, the workflow cannot operate. This is an inherent constraint of any homology-based approach and defines the boundary between MetalPlacer and de novo prediction methods such as Metal3D.

4.6 Future work

The results obtained suggest several avenues for future development:

1. Larger-scale validation of Path 1. The current validation uses 14 structures with 100% identity homologs, with two examples per metal, which was sufficient to verify the workflow but does not represent a large-scale benchmark. Future work should test the pipeline on a larger and more diverse set of structures, including homologs with 30–70% identity, to evaluate performance under more realistic conditions.

2. Expansion of the negative test set. The small number of non-metalloproteins in the Path 0 test set (n = 32) limits the reliability of specificity estimates. A more balanced set would allow for a more confident assessment of the false-positive rate.

3. Multi-chain processing. Extend the workflow to process all chains of the biological unit, with a post-placement filtering step that removes redundant ions or those outside the processed chain.

4. Application to AlphaFold models. Models generated by AlphaFold lack metal ions but reasonably reproduce the geometry of the coordination sites (4). Since MetalPlacer is designed to operate on apo structures, its application to predicted models is a natural extension of the pipeline. However, MetalKB's statistical potentials were derived from experimental structures, so it would be necessary to specifically validate their accuracy on predicted geometries.

5. **Exploration of more powerful models for Path 0.** The current Random Forest could be replaced or supplemented with more expressive architectures, such as coordination graph neural networks (GNNs) or protein language model embeddings like ESM (53), which could capture more complex patterns without requiring manual feature engineering.

6. **Extension to additional metals.** Include other biologically relevant metals, such as Co and Mo among the transition metals, and evaluate the feasibility of incorporating Na and K, whose more flexible and less directional coordination poses additional challenges for both Path 0 training and Path 1 validation.

5 Conclusions

This work defined three objectives: develop a homology-based metal placement workflow (Path 1), build an exploratory structural screening module (Path 0), and integrate both into Scipion as reproducible protocols. The following conclusions address each of them:

- **Path 1.** The homology based placement achieved sub-angstrom per-ion distance in 11 of 12 successful cases across six metals (Zn, Ca, Fe, Cu, Mn, Ni), consistent with MetalKB's reported benchmarks. Mg and non-canonical Ni coordination define the current boundaries of the approach.
- **Path 0.** The structural screening module (ROC-AUC 0.791) provides an interpretable, conformation-specific, GPU-independent pre-filter whose value lies in its integration within the pipeline rather than in competing with sequence-based classifiers.
- **Scipion integration.** Both modules were encapsulated as Scipion protocols, connecting screening, identification, and placement into a single traceable workflow that produces pseudo-holo PDB files from apo or predicted structures without manual intervention between steps.

Future priorities include validation with 30–70% identity homologs, multi-chain processing, application to AlphaFold models, and expansion of the negative test set for Path 0.

6 References

[1] N. Ye, F. Zhou, X. Liang, H. Chai, J. Fan, B. Li, et al. A comprehensive review of computation-based metal-binding prediction approaches at the residue level. *Biomed Res Int*, 2022:8965712, 2022.

[2] L. Rulísek and J. Vondrásek. Coordination geometries of selected transition metal ions (co^{2+}, ni^{2+}, cu^{2+}, zn^{2+}, cd^{2+}, and hg^{2+}) in metalloproteins. *J Inorg Biochem*, 71(3–4):115–127, 1998.

[3] H. Zheng, M. Chruszcz, P. Lasota, L. Lebioda, and W. Minor. Data mining of metal ion environments present in protein structures. *Journal of Inorganic Biochemistry*, 102(9):1765–1776, 2008.

[4] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[5] M. L. Hekkelman, I. de Vries, R. P. Joosten, and A. Perrakis. Alphafill: enriching AlphaFold models with ligands and cofactors. *Nat Methods*, 20(2):205–213, 2023.

[6] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024.

[7] A. Shenoy, Y. Kalakoti, D. Sundar, and A. Elofsson. M-Ionic: prediction of metal-ion-binding sites from sequence using residue embeddings. *Bioinformatics*, 40(1):btad782, 2024.

[8] Q. Yuan, S. Chen, Y. Wang, H. Zhao, and Y. Yang. Alignment-free metal ion-binding site prediction from protein sequence through pretrained language model and multi-task learning. *Brief Bioinform*, 23(6):bbac444, 2022.

[9] J. E. Sánchez-Aparicio, L. Tiessler-Sala, L. Velasco-Carneros, L. Roldán-Martín, G. Sciortino, and J. D. Maréchal. BioMetAll: Identifying metal-binding sites in proteins from backbone preorganization. *J Chem Inf Model*, 61(1):311–323, 2021.

[10] S. L. Dürr, A. Levy, and U. Rothlisberger. Metal3D: a general deep learning framework for accurate metal ion location prediction in proteins. *Nat Commun*, 14:2713, 2023.

[11] S. L. Dürr and U. Rothlisberger. AllMetal3D: joint prediction of localization, identity and coordination geometry of common metal ions in proteins. bioRxiv, 2025.

[12] V. Ravnik, M. Jukič, and U. Bren. Identifying metal binding sites in proteins using homologous structures, the made approach. *J Chem Inf Model*, 63(16):5204–5219, 2023.

[13] X. Zhao, H. Li, and S. Y. Huang. MetalKB: Predicting metal binding sites on proteins with a knowledge-based graph framework. *J Chem Inf Model*, 66(7):4322–4334, 2026.

[14] G. Y. Lin, Y. C. Su, Y. L. Huang, and K. Y. Hsin. MESPEUS: a database of metal coordination groups in proteins. *Nucleic Acids Res*, 52(D1):D483–D493, 2024.

[15] H. Zhang, J. Zhong, M. Gucwa, Y. Zhang, H. Ma, L. Deng, et al. PinMyMetal: a hybrid learning system to accurately model transition metal binding sites in macromolecules. *Nat Commun*, 16(1):3043, 2025.

[16] J. M. de la Rosa-Trevín, A. Quintana, L. Del Cano, A. Zaldívar, I. Foche, J. Gutiérrez, et al. Scipion: A software framework toward integration, reproducibility and validation in 3d electron microscopy. *J Struct Biol*, 195(1):93–99, 2016.

[17] J. Jiménez de la Morena, P. Conesa, Y. C. Fonseca, F. P. de Isidro-Gómez, D. Herreros, E. Fernández-Giménez, et al. Scipiontomo: Towards cryo-electron tomography software integration, reproducibility, and validation. *J Struct Biol*, 214(3):107872, 2022.

[18] C. Chothia and A. M. Lesk. The relation between the divergence of sequence and structure in proteins. *EMBO J*, 5(4):823–826, 1986.

[19] B. Rost. Twilight zone of protein sequence alignments. *Protein Eng*, 12(2):85–94, 1999.

[20] R. Levy, M. Edelman, and V. Sobolev. Prediction of 3d metal binding sites from translated gene sequences based on remote-homology templates. *Proteins*, 76(2):365–374, 2009.

[21] I. N. Kasampalidis, I. Pitas, and K. Lyroudia. Conservation of metal-coordinating residues. *Proteins*, 68(1):123–130, 2007.

[22] P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009.

[23] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, et al. The protein data bank. *Nucleic Acids Res*, 28(1):235–242, 2000.

[24] M. Steinegger and J. Söding. Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*, 35(11):1026–1028, 2017.

[25] Y. Rose, J. M. Duarte, R. Lowe, J. Segura, C. Bi, C. Bhikadiya, et al. Rcsb protein data bank: Architectural advances towards integrated searching and efficient access to macromolecular structure data from the pdb archive. *Journal of Molecular Biology*, 433(11):166704, 2021.

[26] S. Bittrich, C. Bhikadiya, C. Bi, H. Chao, J. M. Duarte, S. Dutta, et al. Rcsb protein data bank: Efficient searching and simultaneous access to one million computed structure models alongside the pdb structures enabled by architectural advances. *Journal of Molecular Biology*, 435(14):167994, 2023.

[27] S. Karlin, Z. Y. Zhu, and K. D. Karlin. The extended environment of mononuclear metal centers in protein structures. *Proc Natl Acad Sci U S A*, 94(26):14225–14230, 1997.

[28] T. Dudev, Lin, M. Dudev, and C. Lim. First–second shell interactions in metal binding sites in proteins: A pdb survey and dft/cdm calculations. *J Am Chem Soc*, 125(10):3168–3180, 2003.

[29] S. Barber-Zucker, B. Shaanan, and R. Zarivach. Transition metal binding selectivity in proteins and its correlation with the phylogenomic classification of the cation diffusion facilitator protein family. *Sci Rep*, 7:16381, 2017.

[30] G. Sciortino, E. Garribba, J. Rodríguez-Guerra Pedregal, and J. D. Maréchal. Simple coordination geometry descriptors allow to accurately predict metal-binding sites in proteins. *ACS Omega*, 4(2):3726–3731, 2019.

[31] M. M. Yamashita, L. Wesson, G. Eisenman, and D. Eisenberg. Where metal ions bind in proteins. *Proc Natl Acad Sci U S A*, 87(15):5648–5652, 1990.

[32] J. Kyte and R. F. Doolittle. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, 157(1):105–132, 1982.

[33] C. Andreini, G. Cavallaro, and S. Lorenzini. Findgeo: a tool for determining metal coordination geometry. *Bioinformatics*, 28(12):1658–1660, 2012.

[34] V. Putignano, A. Rosato, L. Banci, and C. Andreini. Metalpdb in 2018: a database of metal sites in biological macromolecular structures. *Nucleic Acids Res*, 46(D1):D459–D464, 2018.

[35] L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152, 2012.

[36] Q. Hou, K. Waury, D. Gogishvili, and K. A. Feenstra. Ten quick tips for sequence-based prediction of protein properties using machine learning. *PLoS Comput Biol*, 18(12):e1010669, 2022.

[37] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[38] W. G. Touw, J. R. Bayjanov, L. Overmars, L. Backus, J. Boekhorst, M. Wels, et al. Data mining in the life sciences with random forest: a walk in the park or lost in the jungle? *Brief Bioinform*, 14(3):315–326, 2013.

[39] W. J. Lin and J. J. Chen. Class-imbalanced classifiers for high-dimensional data. *Brief Bioinform*, 14(1):13–26, 2013.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res*, 12:2825–2830, 2011.

[41] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *J Mach Learn Res*, 13:281–305, 2012.

[42] J. Bernett, D. B. Blumenthal, D. G. Grimm, F. Haselbeck, R. Joeres, O. V. Kalinina, et al. Guiding questions to avoid data leakage in biological machine learning applications. *Nat Methods*, 21(8):1444–1453, 2024.

[43] A. Niculescu-Mizil and R. Caruana. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, pages 625–632, New York, NY, USA, 2005. Association for Computing Machinery.

[44] J. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, 1999.

[45] Y. Huang, W. Li, F. Macheret, R. A. Gabriel, and L. Ohno-Machado. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inform Assoc*, 27(4):621–633, 2020.

[46] M. P. Naeini, G. F. Cooper, and M. Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. *Proc AAAI Conf Artif Intell*, 2015:2901–2907, 2015.

[47] Y. Liu, P. H. C. Chen, J. Krause, and L. Peng. How to read articles that use machine learning: Users' guides to the medical literature. *JAMA*, 322(18):1806–1816, 2019.

[48] D. Chicco and G. Jurman. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1):6, 2020.

[49] T. Saito and M. Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3):e0118432, 2015.

[50] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[51] J. Shearer. Insight into the structure and mechanism of nickel-containing superoxide dismutase derived from peptide-based mimics. *Acc Chem Res*, 47(8):2332–2341, 2014.

[52] H. Gamouh, M. Novotný, and D. Hoksza. Hybrid protein-ligand binding residue prediction with protein language models: does the structure matter? *Bioinformatics*, 41(8):btaf431, 2025.

[53] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

7 Supplementary material

Table S1: Complete list of the 62 features used by the Path 0 binary screener.

#	Group	Feature	Description	Unit
1–3	Shell composition	`n_polar_{s1,s2,s3}`	Polar residues count (Ser, Thr, Asn, Gln, Tyr)	count
4–6	Shell composition	`n_apolar_{s1,s2,s3}`	Apolar residues count (Ala, Val, Leu, Ile, Pro, Phe, Trp, Met)	count
7–9	Shell composition	`n_chrg_pos_{s1,s2,s3}`	Positively charged residues count (Arg, Lys, His)	count
10–12	Shell composition	`n_chrg_neg_{s1,s2,s3}`	Negatively charged residues count (Asp, Glu)	count
13–15	Shell composition	`n_aromatic_{s1,s2,s3}`	Aromatic residues count (Phe, Tyr, Trp, His)	count
16–18	Shell composition	`n_cys_met_{s1,s2,s3}`	Cys + Met count (sulfur donors)	count
19–21	Shell composition	`n_his_{s1,s2,s3}`	His count (imidazole N donor)	count
22–24	Shell composition	`n_N_donors_{s1,s2,s3}`	Side-chain nitrogen donor atoms count	count
25–27	Shell composition	`n_S_donors_{s1,s2,s3}`	Side-chain sulfur donor atoms count	count
28–30	Shell composition	`n_O_sc_donors_{s1,s2,s3}`	Side-chain oxygen donor atoms count	count
31–33	Shell composition	`n_O_bb_donors_{s1,s2,s3}`	Backbone oxygen donor atoms count	count
34–36	Shell composition	`n_residues_{s1,s2,s3}`	Unique residues count in the shell	count
37–39	Shell composition	`mean_dist_{s1,s2,s3}`	Mean Cα–centroid distances	Å
40–42	Shell composition	`std_dist_{s1,s2,s3}`	Standard deviation of Cα–centroid distances	Å
43	Coordination geometry	`coord_n_atoms`	Total atom count within the coordination sphere	count
44	Coordination geometry	`coord_min_dist`	Minimum atom-to-centroid distance	Å
45	Coordination geometry	`coord_mean_dist`	Mean atom-to-centroid distance	Å
46	Coordination geometry	`coord_std_dist`	Standard deviation of atom-to-centroid distances	Å
47	Coordination geometry	`coord_n_tight`	Count of atoms within 2.5 Å of the centroid	count
48	Coordination geometry	`coord_radius_of_gyration`	Radius of gyration of coordinating atoms	Å
49	Coordination geometry	`coord_span`	Maximum pairwise distance between coordinating atoms	Å
50	Coordination geometry	`angle_mean`	Mean atom–centroid–atom angle	°
51	Coordination geometry	`angle_std`	Standard deviation of atom–centroid–atom angles	°
52	Coordination geometry	`angle_min`	Minimum atom–centroid–atom angle	°
53	Coordination geometry	`geom_linear`	Binary indicator of linear donor arrangement (angle $\geq 160^\circ$)	binary

Table S1 – continued from previous page

#	Group	Feature	Description	Unit
54	Coordination geometry	geom_tetrahedral	Binary indicator of tetrahedral donor arrangement (angles 109.5° ± 15°)	binary
55	Physicochemistry	burial_proxy	Burial estimate: n_residues_s3 / (n_residues_s1 + 1)	ratio
56	Physicochemistry	local_charge_s1	Net formal charge in s1 (positive minus negative residues, at pH 7)	charge
57	Physicochemistry	frac_polar_s1	Fraction of polar heavy atoms (N + O) in s1	fraction
58	Physicochemistry	frac_polar_s2	Fraction of polar heavy atoms (N + O) in s2	fraction
59	Physicochemistry	mean_hydrophobicity_s1	Mean Kyte–Doolittle hydrophobicity in s1	scale value
60	BioMetAll probe density	bm_probe_density_3A	Fraction of BioMetAll probes within 3 Å of the centroid	fraction
61	BioMetAll probe density	bm_probe_density_5A	Fraction of BioMetAll probes within 5 Å of the centroid	fraction
62	BioMetAll probe density	bm_is_cluster_center	Binary indicator: site lies within 1.5 Å of a BioMetAll cluster centroid	binary

Table S2: Permutation importance of all 62 features in the Path 0 model, measured as the mean decrease in ROC-AUC when each feature is permuted (test set, n_repeats = 10, random_state = 42). Features are ranked in descending order of importance.

Rank	Feature name	Group	Mean ROC-AUC decrease	Std
1	coord_min_dist	Coordination geometry	0.01142	0.00288
2	bm_probe_density_3A	BioMetAll	0.01137	0.00220
3	bm_probe_density_5A	BioMetAll	0.00924	0.00280
4	frac_polar_s1	Physicochemistry	0.00670	0.00170
5	n_apolar_s3	Shell composition	0.00575	0.00127
6	mean_dist_s2	Coordination geometry	0.00540	0.00273
7	bm_is_cluster_center	BioMetAll	0.00523	0.00124
8	n_O_bb_donors_s3	Shell composition	0.00480	0.00123
9	mean_dist_s3	Coordination geometry	0.00475	0.00074
10	n_chrg_neg_s3	Shell composition	0.00454	0.00088
11	n_O_bb_donors_s2	Shell composition	0.00431	0.00155
12	mean_hydrophobicity_s1	Physicochemistry	0.00415	0.00101
13	frac_polar_s2	Physicochemistry	0.00413	0.00077
14	coord_mean_dist	Coordination geometry	0.00408	0.00135
15	angle_min	Coordination geometry	0.00382	0.00094
16	coord_n_tight	Coordination geometry	0.00353	0.00110
17	burial_proxy	Physicochemistry	0.00349	0.00161
18	n_cys_met_s3	Shell composition	0.00345	0.00091
19	angle_mean	Coordination geometry	0.00322	0.00071
20	n_polar_s3	Shell composition	0.00321	0.00119
21	coord_span	Coordination geometry	0.00297	0.00124
22	n_chrg_pos_s3	Shell composition	0.00296	0.00110
23	n_aromatic_s3	Shell composition	0.00293	0.00231
24	n_S_donors_s2	Shell composition	0.00292	0.00021
25	coord_n_atoms	Coordination geometry	0.00275	0.00225
26	coord_radius_of_gyration	Coordination geometry	0.00274	0.00141
27	n_chrg_neg_s2	Shell composition	0.00237	0.00111

Table S2 – continued from previous page

Rank	Feature name	Group	Mean ROC-AUC decrease	Std
28	n_O_sc_donors_s3	Shell composition	0.00237	0.00051
29	n_residues_s3	Shell composition	0.00235	0.00041
30	n_apolar_s2	Shell composition	0.00205	0.00056
31	n_N_donors_s3	Shell composition	0.00189	0.00029
32	n_aromatic_s2	Shell composition	0.00182	0.00057
33	angle_std	Coordination geometry	0.00179	0.00123
34	coord_std_dist	Coordination geometry	0.00162	0.00145
35	local_charge_s1	Physicochemistry	0.00150	0.00025
36	std_dist_s3	Coordination geometry	0.00149	0.00084
37	n_chrg_pos_s2	Shell composition	0.00148	0.00096
38	mean_dist_s1	Coordination geometry	0.00145	0.00146
39	n_polar_s2	Shell composition	0.00128	0.00144
40	n_his_s3	Shell composition	0.00124	0.00067
41	n_chrg_neg_s1	Shell composition	0.00124	0.00053
42	n_his_s2	Shell composition	0.00120	0.00056
43	n_O_sc_donors_s1	Shell composition	0.00114	0.00046
44	n_O_sc_donors_s2	Shell composition	0.00109	0.00049
45	n_chrg_pos_s1	Shell composition	0.00106	0.00040
46	std_dist_s2	Coordination geometry	0.00101	0.00167
47	n_his_s1	Shell composition	0.00100	0.00016
48	geom_linear	Coordination geometry	0.00100	0.00015
49	n_S_donors_s1	Shell composition	0.00077	0.00034
50	n_apolar_s1	Shell composition	0.00065	0.00019
51	n_cys_met_s2	Shell composition	0.00064	0.00069
52	geom_tetrahedral	Coordination geometry	0.00062	0.00087
53	n_N_donors_s1	Shell composition	0.00059	0.00025
54	n_N_donors_s2	Shell composition	0.00055	0.00031
55	n_cys_met_s1	Shell composition	0.00032	0.00041
56	n_aromatic_s1	Shell composition	0.00031	0.00040
57	std_dist_s1	Coordination geometry	0.00031	0.00043
58	n_residues_s1	Shell composition	0.00024	0.00029
59	n_residues_s2	Shell composition	0.00015	0.00088
60	n_S_donors_s3	Shell composition	0.00008	0.00023
61	n_O_bb_donors_s1	Shell composition	0.00002	0.00043
62	n_polar_s1	Shell composition	−0.00007	0.00048

Table S3: Per-metal performance of Path 0 on the test set. For each metal, the table reports the number of proteins and sites, protein-level and site-level sensitivity at the default operating point ($\tau_{\text{protein}} = 0.60$, $\tau_{\text{site}} = 0.76$), mean p95 score, and mean site score.

Metal	n proteins	Sens (protein)	n sites	Sens (site)	Mean p95	Mean site score
Zn	90	0.778	470	0.334	0.764	0.608
Ca	57	0.772	199	0.437	0.784	0.668
Fe	10	0.600	62	0.387	0.727	0.619
Mn	52	0.635	153	0.392	0.701	0.622
Cu	83	0.867	345	0.490	0.825	0.688
Ni	61	0.574	183	0.279	0.671	0.559