

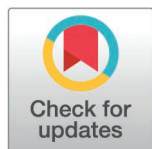
RESEARCH ARTICLE

Predicting unknown binding sites for transition-metal-based compounds in proteins

Andrea Levy¹ , Ursula Rothlisberger¹ *

Laboratory of Computational Chemistry and Biochemistry, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

* ursula.roethlisberger@epfl.ch



OPEN ACCESS

Citation: Levy A, Rothlisberger U (2026) Predicting unknown binding sites for transition-metal-based compounds in proteins. PLoS One 21(6): e0349622. <https://doi.org/10.1371/journal.pone.0349622>

Editor: Soumendranath Bhakat, AlloTec Bio, UNITED STATES OF AMERICA

Received: February 19, 2026

Accepted: May 1, 2026

Published: June 9, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0349622>

Copyright: © 2026 Levy, Rothlisberger. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

Abstract

Transition-metal-based compounds are promising therapeutic agents, particularly in cancer treatment. However, predicting the binding sites of such compounds remains a major challenge. In this work, we investigate the applicability of two tools, Metal3D and Metal1D, for this purpose. Although originally trained to predict zinc ion binding sites only, both predictors correctly identify at least one of the experimentally observed binding sites for transition metal complexes in each of the apo protein structures tested. At the same time, we highlight current limitations, such as the sensitivity to side-chain conformations, and discuss possible strategies for improvement. This work provides a first step toward establishing a robust computational pipeline in which rapid and low-cost predictors are able to identify putative hotspots for transition metal binding, which can then be refined using more accurate but computationally demanding methods.

Introduction

Compounds based on transition metals (TMs) are promising candidates for biomedical applications, as the combination of organic ligands with TM centers enhances key properties such as stability, solubility, and bioavailability [1,2]. However, some TMs pose considerable health hazards, and TM-based drugs often show severe side effects. This is the case for cisplatin, which, despite its several toxic side effects, remains a chemotherapeutic workhorse for various tumors [3]. Significant efforts have been devoted to the search for novel chemotherapeutic strategies to overcome these shortcomings, often relying on alternative TM centers with lower toxicity [1].

From a theoretical point of view, modeling TM compounds is particularly challenging due to some practical considerations. Drug design often relies on molecular docking [4,5], where different binding poses are evaluated with some loss function to identify candidates showing high affinity. However, TM-based compounds can bind biomolecules through the formation of one or more covalent bonds, and in docking, accurately predicting the binding pose of covalent compounds is

provided the original author and source are credited.

Data availability statement: All the predictions for the different protein structures are deposited on Zenodo under <https://zenodo.org/records/18416988>, including VMD visualization states to reproduce the figures presented in the paper. In the same repository, we also included a Jupyter Notebook to perform all predictions with Metal3D and Metal1D, as well as generate the new probability maps for Metal1D.

Funding: U.R. acknowledges funding from the Swiss National Science Foundation (grant 200020-185092 and 200020-219440). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

particularly challenging [6,7]. In the investigation of drug candidates interacting with specific protein targets, often molecular dynamics (MD) simulations based on classical force fields (FFs) are performed to evaluate the energetics and dynamics of the compound at different putative binding pockets. However, in the case of TMs, classical FFs often lack accuracy and are not always able to reproduce, e.g., the correct coordination geometry [8]. In addition, standard FFs are not capable of describing the chemical reactions that lead to the formation of covalently or coordinatively bound ligands.

One possible solution to overcome these limitations is to use hybrid quantum mechanics/molecular mechanics (QM/MM) methods, which are able to accurately describe bond formation events by taking the electronic structure of the TM-based complex and its protein environment explicitly into account [9–11]. However, such methods are orders of magnitude more computationally expensive than molecular docking or even classical FFs approaches. This makes them impractical for high-throughput studies, where various putative sites must be assessed. For this reason, fast but accurate predictors to highlight potential ‘hotspots’ for the binding of TM complexes are fundamental to focus the computational effort where it is really needed.

We recently introduced two prediction tools to locate the binding site of zinc ions in proteins: Metal3D, based on a 3D convolutional neural network (3DCNN), and Metal1D, based on geometric criteria and a probability map associated with the metal under study [12]. Albeit solely trained on zinc, Metal3D is able to accurately locate TM ions different from zinc. This is particularly interesting for TM binding site location, since the training of 3DCNNs and machine learning (ML) models in general is computationally expensive and often requires a large amount of data. Metal1D showed similar precision in predictions for zinc and other TM ions, i.e., most predictions made are correct, but lower recall, i.e., a larger number of binding sites are not predicted in the case of TM ions different from zinc. However, since Metal1D relies on probability maps and geometric rules only, it can be expected to be less data-demanding than Metal3D, as well as more easily adaptable to different metals with low computational cost for generating a new probability map.

The ability of Metal3D and Metal1D to predict the binding site of other TM ions triggered the question of whether these tools are also capable of providing useful predictions for the location of binding sites for TM-based compounds. Such compounds present different organic ligands that coordinate the metal center, introducing steric constraints on the binding site accessibility. In addition, the presence of ligands coordinating the metal center limits the number of possible bonds that it can form with protein ligands, typically resulting in one or two coordinating amino acids only. However, binding sites for such compounds and for bare TM ions share several similarities, particularly in the types of amino acid residues involved.

The focus of this paper is to probe the performance of Metal3D and Metal1D for the identification of binding sites for TM-based agents. In the case of Metal1D, we also evaluated binding site predictions performed with two additional probability maps generated for this study, i.e., for Pt- and Ru-based compounds. In the following, we

first provide an overview of the two predictors and the protein systems used in this study, followed by the presentation of the results and a discussion of the main conclusions and future outlook.

Methods

Metal3D and Metal1D predictors

A detailed description of Metal3D and Metal1D can be found in the original publication by Dürr et al. [12] In the same work, the dataset composition is analysed in detail, including the distribution of coordination motifs. Here, we provide a short overview of the main features of the two predictors and highlight their differences and similarities. A schematic representation of their training and inference mode to predict metal ion positions is provided in Fig 1.

Metal3D is based on voxelization of input structures into three-dimensional grids, which are then fed to a 3DCNN that uses eight different input channels to describe the molecular environment in a voxel, i.e., aromaticity, hydrophobicity, positive/negative charge, hydrogen bond donor/acceptor, occupancy, and metal ion binding side chain. The model uses boxes of size 16 Å centered on the C α atom of different residues, with a grid resolution of 0.5 Å, to generate a probability density for metal ion location at each grid point, computed from the atomic coordinates within the input box. This is done through a series of convolutional layers with filter size 1.5 Å, except for the fifth layer where a larger filter (8 Å) is used to capture interactions on a longer range. The probability density from boxes within 0.25 Å of each grid point is then combined by averaging the probability box densities to generate a global probability for the entire protein, which can be used to place metal ions. However, for this project, we focused on the total probability density itself rather than on the exact placement of the metal. This allowed us to systematically explore a broader range of probability values, including low-probability sites that might otherwise be overlooked when Metal3D does not place a metal ion.

On the other hand, predictions performed with Metal1D are based on a probability map derived from the LINK records in the protein structures from the RCSB Protein Data Bank (PDB) [13], which specify the connectivity between a ligand (the metal ion in this case) and the amino acids of a protein. These probability maps represent empirical coordination preferences derived from the frequency with which specific amino acids are observed to coordinate a given metal ion in experimentally determined structures in the PDB. To construct a probability map, all relevant LINK records, including the metal of interest and protein atoms, are extracted. Next, the coordinating residues for each metal are converted into a unique coordination environment by associating one letter code and alphabetically sorting this code to group equivalent environments, e.g., CCH or CHC, etc. Duplicate PDB entries are removed to avoid over-representing frequently crystallized proteins, retaining the protein structure with the best resolution. The counts of each unique coordination environment across the filtered dataset are then normalized, producing a probability map that reflects the relative frequency of each observed coordination motif. To perform a prediction, a geometrical search is performed around each amino acid within a search radius, starting from a reference point corresponding to the atom most likely to coordinate a metal. Based on the surrounding amino acids within the search radius, a score is assigned to the reference point by summing the probabilities in the probability map for coordinations compatible with the one observed. In this way, higher scores correspond to environments that more closely resemble frequently observed coordination motifs for the metal considered. The resulting scores indicate how consistent a local amino-acid environment is with known metal coordination patterns. In the limiting case where the local environment is compatible with all coordination motifs present in the probability map, the total probability sums up to 1. In practice, typical scores lie between 0 and <1

After each amino acid in the protein is scored with this approach, the putative metal sites are predicted by grouping the highest-scored amino acids in clusters based on distance and locating a putative site as the weighted average between the coordinates of the reference point of each amino acid, using the amino acid score as a weighting factor. Each site is then re-scored by performing a geometrical search centered on the site position and assigning a probability score based on the probability map for the given metal, in the same way as the amino acid scores are computed.

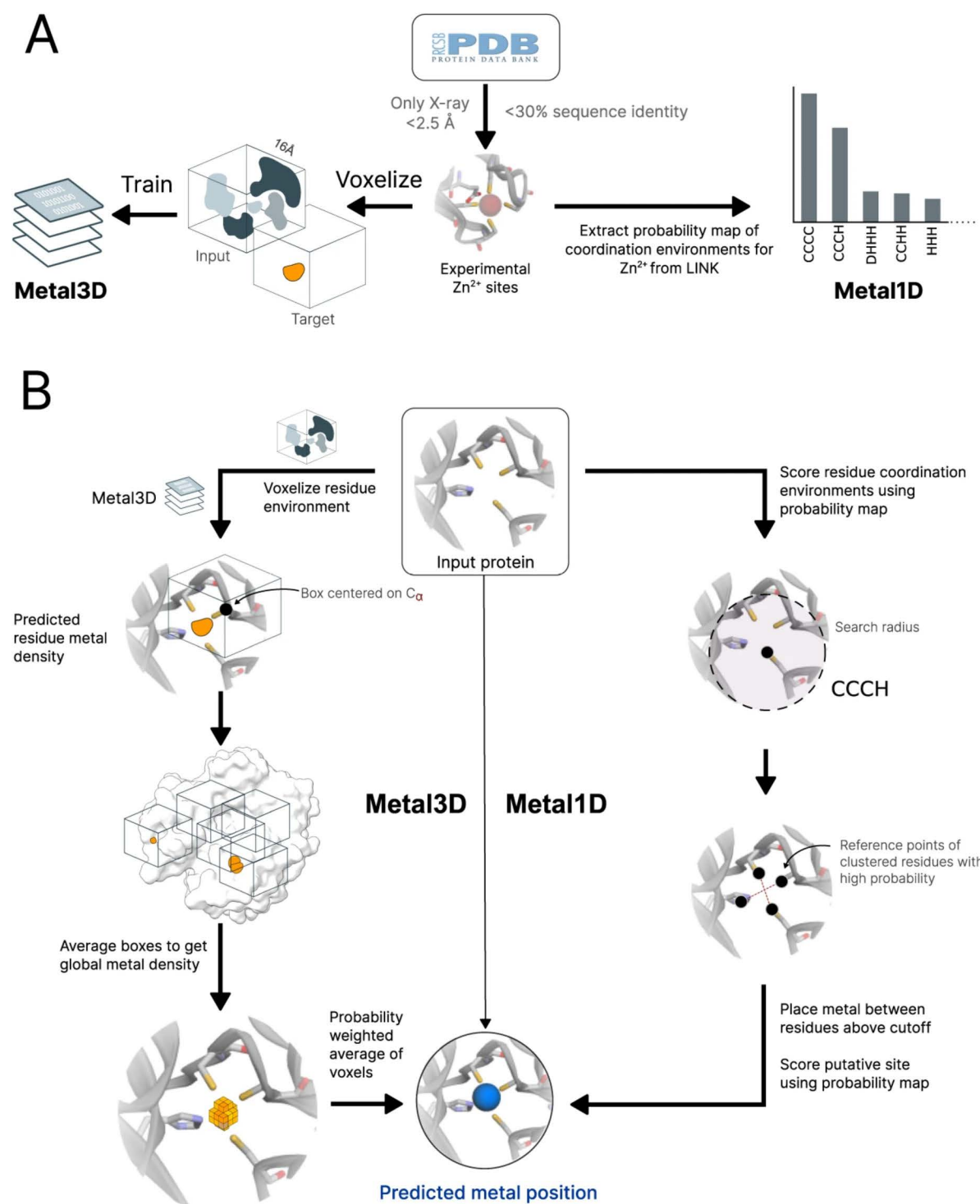


Fig 1. Schematic representation of the training (A) and the inference mode (B) for Metal3D and Metal1D predictors. Reprinted from Ref. [12] (Dürr et al., *Nature Communications*, Springer Nature) under CC BY 4.0 license, original copyright 2023.

<https://doi.org/10.1371/journal.pone.0349622.g001>

Metal binding site predictions

In this study, we chose two systems to test the prediction ability of Metal3D and Metal1D (v0.2) for TM-complexes: hen egg-white lysozyme C (HEWLC, UniProt ID P00698) and bovine pancreatic ribonuclease A (RNaseA, UniProt ID P61823), for both of which experimentally determined binding sites for Pt- and Ru-based compounds are known. For each protein, we selected different crystal structures to create a representative set that includes both the apo protein and holo structures containing Pt- and Ru-based agents. The selection of multiple structures for the same protein serves a dual purpose. First, it enables the identification of experimentally known binding sites to use as a reference. Additionally, it allowed us to evaluate the stability of the predictors with respect to small changes in the input structure: albeit referring to the same protein, different experimental structures can show some variations, e.g., in the side chain positioning or rotameric state, and ideally, a good predictor should have limited sensitivity to such changes. In particular, for the HEWLC enzyme, we selected PDB IDs 194L [14,15], 2I6Z [16,17], 5II3 [18,19], 6QEA [20,21], 5V4G [22,23], and 6WGO [24,25], while for the RNaseA enzyme we selected PDB IDs 1RPH [26,27], 4S0Q [28,29], 4S18 [29,30], and 5JLG [31,32]. More details about the similarity between the different protein structures are reported in the corresponding sections, as well as in the Supporting Information (SI). In particular, S2 and S3 Tables report the resolution of all the PDB structures considered, showing that the structures in this dataset have uniformly resolutions below the original threshold of 2.5 Å used to select crystal structures for the probability map generation of Metal1D and the training of Metal3D [12].

Due to the limited number of available protein structures in our dataset, a fully quantitative statistical comparison is not feasible. In particular, classifying a prediction as a false positive (FP) can be misleading, because unobserved sites in this small number of crystal structures may still be biologically relevant or occupied by a different TM complex. In contrast, true positives (TPs) and false negatives (FNs) can be defined for the experimentally known sites in our dataset, and hence the recall, defined as $TP/(TP + FN)$. The results of this analysis are reported in S1 Table. To provide a meaningful comparison, in the case of Metal1D, we classified a prediction as TP or FN if it was located within 5 Å of the experimentally observed metal center in at least one of the TM complexes. For Metal3D, the same procedure was applied to probability density regions within 5 Å of an experimental metal center, testing three different probability thresholds per protein based on the minimum isovalue identified for the system. Specifically, we used low, medium, and high thresholds of $p \geq 0.04$, 0.09, and 0.14 for HEWLC, and 0.06, 0.11, and 0.16 for RNaseA. This approach allows a fair comparison of the different predictors while acknowledging the limitations of the dataset and the uncertainty associated with unobserved sites.

To test the sensitivity with respect to different rotamers in the case of a bidentate TM-based compound, i.e., binding to two amino acids in the protein, we introduced a third test system extracted from the crystal structure of a nucleosome core particle (NCP) in which the TM complex binding induces a conformational change. We investigated the binding site composed of Glu61 and Glu64 in the H2A histone protein, named GLU site [33]. In particular, we considered the GLU site from the apo structure (PDB ID 3REH [34,35]) and a second structure in which it is occupied by a bidentate Ru-based compound (PDB ID 5XF3 [19,36], more details in the corresponding section and in the SI). Upon binding of the metal compound, the side chain conformations of the Glu61 and Glu64 are different. This feature is consistently observed in several crystal structures, e.g., the conformation of the apo structure is the same as in other crystal structures in which the site is not occupied, e.g., in PDB IDs 4J8U [37,38] and 4J8W [39,38], and analogously, the observed rotamer change in the holo structure is shared by PDB ID 5DNN [40,41], where a different Ru-based compound is co-crystallized (more details in S8 Fig).

In the case of predictions performed with Metal3D, we analyzed the global probability density file generated using VMD [42] (v1.9.3), which has also been used to compare the RMSD of the different structures. Since we are interested in exploring the possibility of using it for predicting TM-based compounds, and we expect an associated probability significantly lower than the one for a typical zinc binding site, we analysed regions with relatively low probability values. By visual inspection of the isosurfaces of the probability density, we empirically determined a minimal isovalue that enabled localizing distinct sites around each of the proteins tested ($p = 0.04$ for HEWLC and $p = 0.06$ for RNaseA).

In the case of Metal1D, there exists the possibility of selecting a threshold value to exclude predictions that have too low probability scores. To maximize the number of predictions and avoid overlooking any possible site, even with low probability, we retained all sites with a score of at least 1% of the maximum site score (Metal1D threshold parameter of 0.99). Metal1D then provides the coordinates of putative binding sites and a probability score associated with them. We investigated the site location by visual inspection using VMD [42] (v1.9.3). In addition to predictions performed with the original probability map for zinc ions [12], we also introduced two additional probability maps for Pt- and Ru-based compounds. Another parameter specific to Metal1D is the search radius, defined as twice the distance between the metal ion and the coordinating atom of the amino acid. By default, Metal1D uses a search radius of 5.5 Å. The same value was used for all Metal1D predictions in this study, as it exceeds the ideal coordination distance, thereby accounting for deviations from ideal coordination geometries. A detailed explanation of which PDB structures have been used to generate these maps is provided in the Results section. We deposited the new probability maps and a Jupyter notebook for their generation at <https://zenodo.org/records/18416988>.

It is important to note that both tools base their predictions only on the coordinates of the heavy atoms in the protein crystal structure. The hydrogenation or protonation state of the different amino acids is not taken into account explicitly, so the absence or presence of hydrogen atoms does not affect the prediction. Moreover, even if the crystal structure includes non-protein atoms, such as ions, water molecules, or other ligands, they are not considered by Metal1D and Metal3D when performing a prediction, even if they are present in the input structure.

Rotamer sampling

We explicitly investigated the effect of different rotameric states on the predictions by Metal1D and Metal3D. To generate such conformations, we used an in-house Jupyter Notebook available on Zenodo (<https://zenodo.org/records/18416988>). To sample different rotamer conformations, we used a similar approach as in the EVOLVE package [43]: rotamers are derived from the Richardson rotamer library [44], which corresponds to the most commonly observed side chain conformations for the naturally occurring amino acids in the PDB, i.e., low-energy conformers of the side chain torsion angles $\{\chi_i\}$. For each sampled rotamer, the backbone atoms are aligned to those in the original structure, and we performed a separate Metal3D or Metal1D prediction on the resulting conformation.

Results and discussion

Differences in data availability

In the case of Metal1D, we generated two additional probability maps to compare them with the results from the original one based on zinc, using PDB structures containing Pt or Ru ions, as well as complexes containing such TMs. The generation of the probability map is analogous to the one for zinc described in Ref. [12], and is based on the LINK records in the PDB structures. In the case of Pt, fewer than 250 crystal structures in the PDB database contain a Pt-based compound covalently bound to a biomolecule. Among those structures, we identified only 81 X-ray structures with resolution higher than 2.5 Å corresponding to unique proteins, containing 244 binding sites, i.e., where Pt ions (or Pt atoms in Pt-based complexes) contain a LINK to at least one amino acid. In the case of Ru, these numbers are even more reduced: about 160 crystal structures with covalently bound Ru-based complexes are present in the PDB, and only 28 unique protein structures, corresponding to 47 Ru binding sites. The number of available structures is drastically low in comparison with what was used to train the 3DCNN model for Metal3D, where more than 2 000 crystal structures corresponding to unique proteins with high resolution have been used, featuring ~15 600 zinc binding sites. This comparison highlights how challenging the problem of designing a predictor for some TM binding sites can be due to the limited amount of data for training, especially in the case of 3DCNNs, and ML models in general.

To compare the binding propensity to different amino acids, we report in Fig 2 the frequency of binding for the different amino acids extracted from the LINK analysis for the Zn(II) ion used in the original training of Metal1D, as well as the ones resulting from generating the probability maps for Pt- and Ru-based compounds (including ions and metal-based compounds). In the case of Zn(II), the preferential binding amino acids are histidines, followed by cysteines and, at lower frequencies, aspartates and glutamates. This is in line with previous studies in which the PDB has been analyzed to determine the TM binding selectivity in proteins. [45–48] From this simple comparison, it is possible to observe metal-specific characteristics in the amino acid binding patterns, which will be reflected in the corresponding probability maps for Metal1D. This is particularly evident in the case of Pt-based compounds, where methionine binding is preferential and presents even higher probability than other typical metal-binding amino acids.

Lysozyme C TM binding site prediction

The hen egg-white lysozyme C (HEWLC) protein was the first enzyme for which a three-dimensional structure has been resolved [49,50] and, to date, is among the enzymes with more crystal structures available thanks to its propensity to easily crystallize in various crystal forms, yielding high-resolution X-ray data [51]. Our test set (Fig 3) contains an apo HEWLC

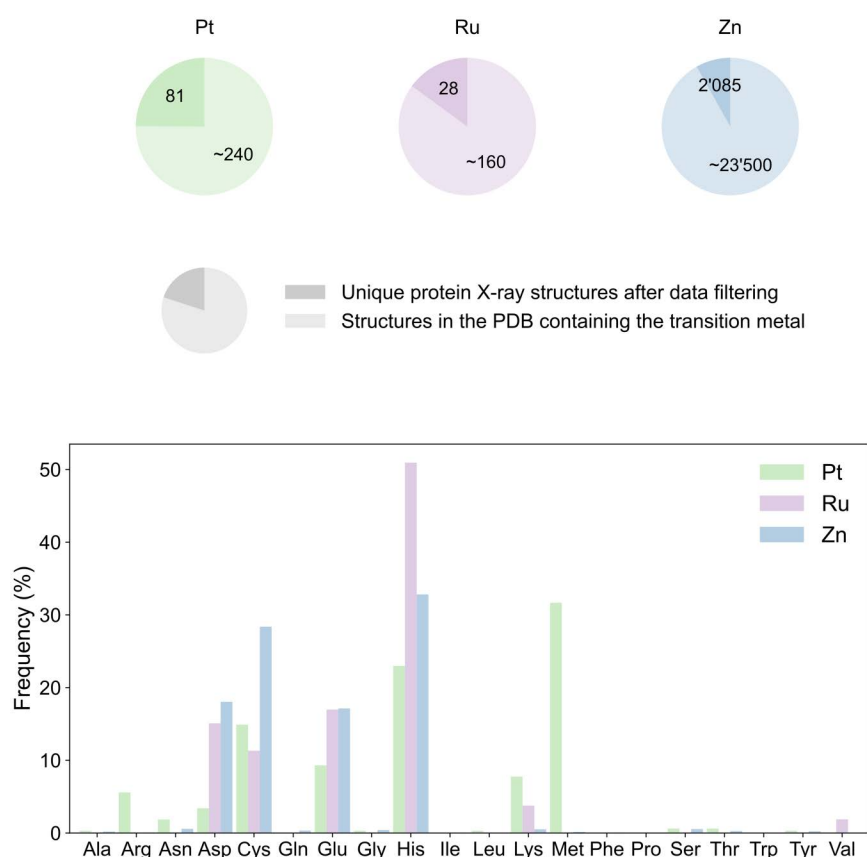


Fig 2. Dataset statistics on the PDB structures used to generate Metal1D probability maps for Zn(II) ions and for Pt and Ru (ions and metallo-compounds). The pie charts at the top show the difference in the available number of experimental structures from the PDB for the different cases, and their further reduction after filtering to generate the probability maps for Metal1D. The histograms at the bottom show the binding preference to different amino acids, based on the analysis of the LINK records. Note that for each case, the histogram is generated from the unique protein X-ray structures after data filtering.

<https://doi.org/10.1371/journal.pone.0349622.g002>

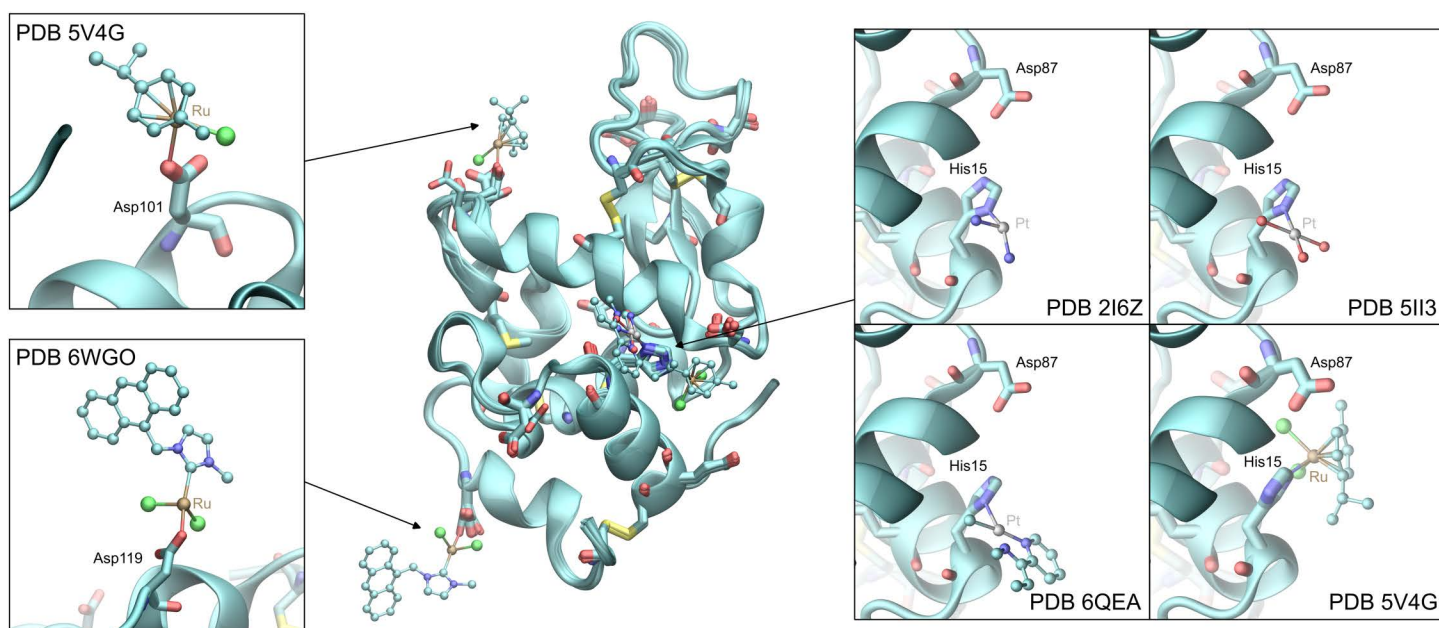


Fig 3. PDB structures for the HEWLC enzyme used as test set. All the PDB structures have been aligned, and the metal binding amino acids are represented in licorice representation. The binding sites for each TM compound are reported in different inserts.

<https://doi.org/10.1371/journal.pone.0349622.g003>

structure [14], one in complex with cisplatin [16], two with different Pt(II)-based compounds binding the protein in a planar geometry [18,20], and two different Ru(II)-based compounds at different binding sites [22,24]. All the TM compounds considered form monodentate adducts, binding a single amino acid.

All the holo structures containing TM complexes deviate only slightly from the apo structure with RMSD lower than ~ 0.3 Å for the backbone atoms, and ~ 0.9 Å when considering the metal binding amino acids (Asp, Cys, Gly, His, and Met), including their side chains. In particular, focusing on the amino acids of the three binding sites considered, the binding of a TM-based compound at His15 is associated with a different rotamer, while in the case of Asp119, only a different positioning of the side chain is observed, without a change of the rotameric state. The binding site at Asp101 appears to be more mobile, with different side chain positioning and rotamer changes observed among the different structures, even in the absence of a binding TM compound. More details about the rotameric states, as well as the RMSD values, are reported in the SI (S2 Table).

Starting with the discussion of the results from Metal3D, we notice that the predictions for the different structures are quite consistent, with small changes in the regions with larger probability. The predictions for the apo structure are reported in Fig 4, while the predictions for all the other structures are reported in S1 Fig. In particular, the region around His15, i.e., the binding site for the Pt-based compounds and one of the Ru-based ones, corresponds to the highest probability region predicted by Metal3D, with values ≥ 0.10 – 0.15 depending on the structure. This result is not affected by the change in conformation observed for some of the structures, which corresponds to a rotation of the imidazole ring, and corresponds to a large number of TP predictions for this site for all structures (in all the structures when considering a low or medium probability threshold, S1 Table). This might also be due to the proximity of Asp87, which makes the region between the two amino acids favorable for the binding of TMs. It is important to notice that the probability values are significantly lower than what Metal3D predicts for, e.g., Zn(II) binding sites. This can be attributed to different factors: zinc ions often bind with tetracoordinated geometries with at least four binding amino acids, resulting in higher probability

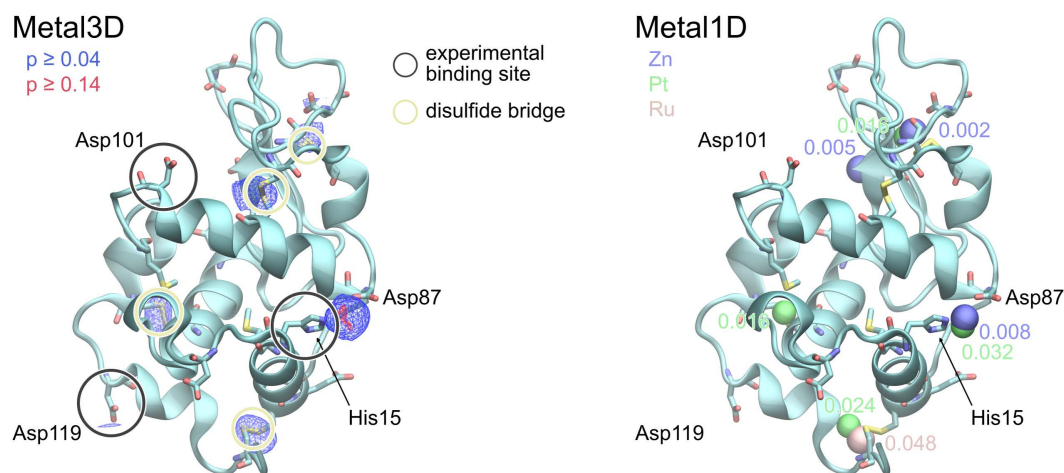


Fig 4. Metal3D (left) and Metal1D (right) predictions for the apo structure of HEWLC (PDB 194L). The experimental binding sites and disulfide bridges are highlighted with circles in the structure on the left. In the test set considered, Ru-based compounds bind all three sites, while only Pt-based compounds bind at His15. For Metal3D, two probability density contours are represented in different colors as isosurfaces with a wireframe representation: lower probability isovalues of 0.04 in blue and higher ones (0.14) in red. For Metal1D, the results obtained with different probability maps are reported: zinc (blue), platinum (green), and ruthenium (pink), and for each prediction, the probability associated with each site is indicated in the figure.

<https://doi.org/10.1371/journal.pone.0349622.g004>

values compared to the Pt and Ru complexes forming mono- or bi-dentate adducts with proteins. Nevertheless, even if with low probability in absolute terms, Metal3D is able to identify the binding site at His15.

Concerning the two other binding sites for Ru-based compounds, both corresponding to a single aspartate residue (Asp101 or Asp119), they are not successfully predicted by Metal3D: some very low-probability region is identified in some of the cases ($p \sim 0.04$), but this prediction is not consistently found in all the structures (corresponding to a recall value of 0.5 for a low probability threshold for both sites, [S1 Table](#)). This is true even in the case of the holo structures, for which only extremely low-probability regions, or no regions at all, are identified at Asp101 or Asp119. This is interesting since it can be assumed that the amino acid side chain in structures containing a TM complex is in ‘optimal’ conformation for binding. However, both sites present a single metal binding amino acid, which is a really different environment from a typical Zn(II) ion binding site, on which Metal3D was trained. Other Metal3D predictions, consistent among all the structures, correspond to the four disulfide bridges present in the protein structure. Considering that cysteines are the second most likely amino acids to bind Zn(II) ([Fig 2](#)), it is not surprising that such regions are identified as probable for metal binding. In practice, when investigating binding sites for TMs compounds, predictions associated with disulfide bridges can be identified by visual inspection or automatically flagged in a post-processing step detecting cysteine pairs forming disulfide bonds, e.g., by geometrical criteria or by using the information contained in the `SSBOND` section of PDB files. Predicted sites located within a short distance of the corresponding sulfur atoms can therefore be excluded. For example, applying this criterion to all Metal1D predictions shows that none of the putative sites in the proximity of a disulfide bridge would be retained when discarding sites located within 3.0 Å of the sulfur atoms involved in the disulfide bond (maximum distance for the predictions 2.87 Å).

Also in the case of Metal1D (predictions for the apo structure in [Fig 4](#) and for all other structures in [S2 Fig](#)), the His15 site is identified consistently in all the structures with the highest probability values using the original probability map from Zn(II) structures, and a recall value of 0.83 ([S1 Table](#)). This is also the case when using the probability map generated for Pt-based compounds, which, in general, identifies similar sites as the zinc one, including some of the cysteines forming disulfide bridges, similarly to what has been observed in the Metal3D predictions. Moreover, predictions performed with the Pt-based probability map present larger absolute values than the ones from the Zn-based probability map, e.g., with

a probability associated with the site at His15 of 0.032 for the Pt map, and 0.008 for the Zn map (in both cases, this is the highest probability site predicted). These values are significantly lower than predictions of Zn(II) tetracoordinated binding sites, but the same considerations as the ones for Metal3D predictions apply. No Metal1D prediction is located in proximity of the two binding sites at Asp101 and Asp119, suggesting that Metal1D, as Metal3D, fails to predict such sites containing a single amino acid ligand. Concerning the prediction performed with the probability map extracted for Ru-based compounds, in all structures, only one or two disulfide bridges are identified, and even if histidines are the most frequent Ru-binding amino acids in the probability map for ruthenium (Fig 2), no prediction occurs for the His15 binding site.

Ribonuclease A TM binding site predictions

Bovine pancreatic ribonuclease A (RNaseA) is another extensively studied protein, which was the first enzyme to have its complete amino acid sequence determined [52,53]. Our test set (Fig 5) contains an apo RNaseA structure [26], two with different Pt(II)-based compounds forming mono- and bi-dentate adducts at different binding sites [28,30], and one with a Ru(II)-based monodentate compound [31]. This second set of protein structures allows us to extend the test of Metal3D and Metal1D in the prediction of binding sites for a larger variety of Pt and Ru compounds, also including a bidentate site. From a practical point of view, such sites are closer to ideal binding sites of TM ions, potentially increasing the possibility of predictions for Metal3D and Metal1D.

Similar to HEWLC, the holo structures in the test set are similar to the apo one, enabling us to probe the sensitivity of the predictions to small changes in the side chain conformation. In particular, all the structures deviate from the apo structure by less than ~ 0.5 Å RMSD for the backbone atoms, and ~ 0.7 Å for the binding amino acids (Asp, Cys, Glu, His, and Met), including their side chains. The binding sites at His105 and His119 do not present a rotamer change with respect to the apo structure when a TM-based compound is bound, while the site at Asp14–Met29 presents a change in the rotameric state of the methionine when the monodentate compound binds to it. In the case of the bidentate compound, both

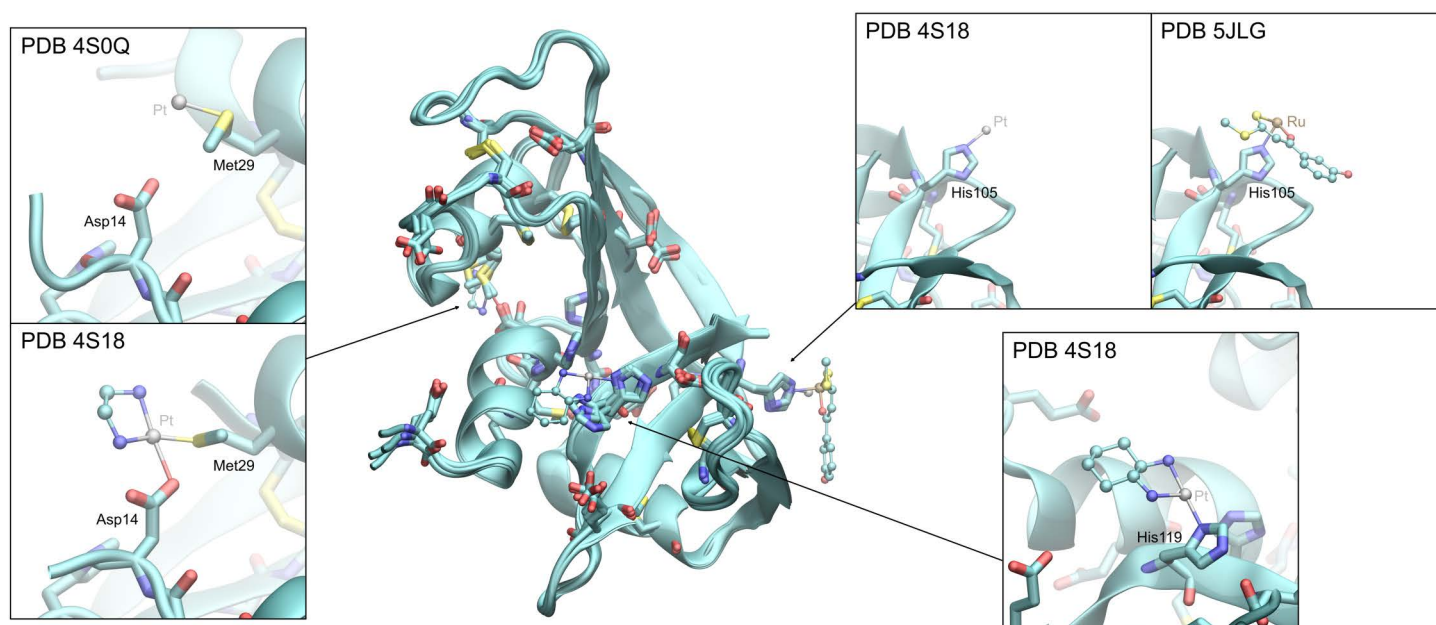


Fig 5. PDB structures for RNaseA used as test set. All the PDB structures have been aligned, and the metal binding amino acids are represented in licorice representation. The binding sites for each TM compound are reported in different inserts.

<https://doi.org/10.1371/journal.pone.0349622.g005>

Asp14 and Met29 change the rotameric state upon binding. More details about the rotameric states, as well as the RMSD values, are reported in the SI (S3 Table).

In the case of Metal3D, three regions with high probability are found consistently in all the structures: two of them in proximity of His119 with a corresponding recall of 1.0 even for high probability thresholds (S1 Table) either in the direction of His12 or Asp121, respectively, and one corresponding to a disulfide bridge between two cysteine residues (predictions for the apo structure in Fig 6 and for all other structures in S4 Fig). Similar to what has been observed in the case of HEWLC, all four disulfide bridges in the structure are also identified by Metal3D, with a lower associated probability. The binding site at His105 is not predicted, with the exception of the PDB structure 4S18, in which this site is occupied by a Pt ion from oxaliplatin. The probability density at this site, however, has a low probability ($p \sim 0.1$) and is not consistent between the other structures, even if the side chain positioning of His105 varies only slightly among all structures without a change in the rotameric state. In particular, His105 presents an all-atom RMSD of ≤ 0.3 Å (a detailed comparison is given in S4 Table and S3 Fig). Furthermore, Metal3D does not predict one of the known binding sites involving Asp14 and Met29 (PDB 4S18 and 4S0Q). The reason for that might be the presence of a methionine at the site, which is an amino acid commonly bound by platinum, but not zinc (Fig 2), on which Metal3D was trained. However, a low-probability region between Asp14 and His48 is identified in some of the structures, in close proximity to Met29.

In the case of Metal1D, the predictions based on the zinc probability map are in partial agreement with Metal3D: the His119 site is consistently identified in all the structures, except for one (PDB 5JLG), in which this residue assumes a different conformation, corresponding to a recall value of 0.75 (S1 Table). Predictions for the apo structure in Fig 6 and for all other structures in S5 Fig. Other consistent predictions among all of the structures correspond to the region with the disulfide bridge and to His48, in close proximity to Asp14, which together with Met29 is part of the bidentate binding site of oxaliplatin (PDB 4S18). However, also in this case, His105 is not identified as a possible binding site. In the case of the probability map generated from Pt-based compounds, the predictions partially agree with those based on zinc. Moreover, Metal1D predicts some disulfide bridges, but only in the case of the apo structure, and His48 in proximity to the

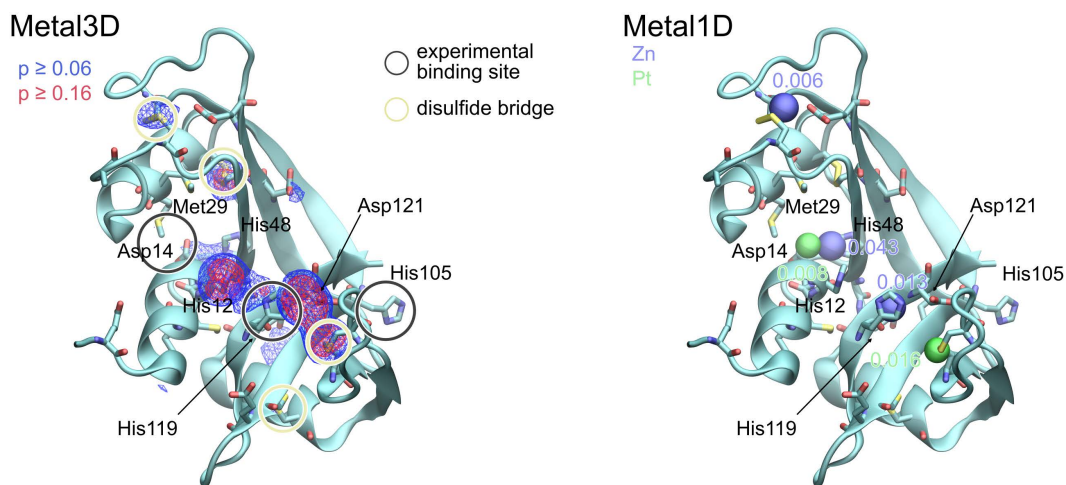


Fig 6. Metal3D (left) and Metal1D (right) predictions for RNaseA apo structure (PDB 1RPH). The experimental binding sites and disulfide bridges are highlighted with circles in the structure on the left. In the test set considered, Pt-based compounds bind all three sites, while the Ru-based compound binds at His105. For Metal3D, two probability density contours are represented in different colors as isosurfaces with a wireframe representation, in blue the one with lower probability and in red the one with higher probability (isovalues of 0.06 and 0.16, respectively). For Metal1D, the results obtained with different probability maps are reported: zinc (blue) and platinum (green), and for each prediction, the probability associated with each site is indicated in the figure. For this structure, Metal1D did not predict any site for ruthenium binding.

<https://doi.org/10.1371/journal.pone.0349622.g006>

Asp14–Met29 binding site is identified, similarly to what was observed for Metal3D. In the case of the Ru-based probability map, only some of the disulfide bridges are identified inconsistently between the different structures.

Dependence of the predictions on rotameric variations

Monodentate histidine-containing sites. Based on the results for HEWLC and RNaseA discussed so far, it is possible to observe some dependence of the prediction on the rotamer conformations of the side chains of the metal binding amino acids. For both predictors, this is reasonable, considering that the local environment processed by the 3DCNN for Metal3D, and the geometric criteria for Metal1D, will be optimal when the conformation is close to that of a metal binding site. However, a large degree of dependency on the rotamer conformations is detrimental for the prediction of unknown binding sites starting from an apo structure, since it is possible to overlook some binding sites simply because the amino acids are in a different conformation.

To elucidate the dependency on rotamer state, we selected some of the binding sites in our test enzymes and probed the effect on the predictions upon variations of the rotamer state of one of the amino acids (Fig 7). In particular, in the apo structures, we selected His15 in HEWLC and His105 in RNaseA as representative cases of a successful and unsuccessful prediction, respectively. For both structures, we investigated the influence of sampling different rotamer conformations on the prediction probabilities for Metal3D and Metal1D.

In the case of His15, the holo structure presents a different rotamer for the imidazole ring, but not a change in the orientation of the side chain. This site was already predicted with sufficient confidence in the apo structure, but the sampling of further rotamers results in a decrease of the Metal3D probability around the amino acid. In particular, the region with high probability identified in the apo structure ($p \sim 0.14$) disappears for all of the other rotamers, while the region with lower probability ($p \sim 0.04$) remains present, even if its shape and size change. In the case of the predictions with Metal1D, all the rotamers sampled lead to a prediction with the zinc and platinum probability maps in proximity to His15, in analogy to what was observed in the apo structure. The probability value is similar to that of the original structure, especially in the case of the predictions with the probability map for platinum. Similarly to what has been observed in the apo structure, no prediction is made with the probability map extracted from the Ru-based compounds. In the case of His105, which was not predicted for the apo structure, the sampling of different rotamers leads Metal3D to identify low-probability regions ($p \sim 0.004$) for some of the sampled rotamers. In contrast, Metal1D does not make any prediction in the proximity of His105 for any of the sampled configurations.

This comparison highlights some differences between the two predictors: on one hand, Metal3D seems to be more sensitive to the side chain positioning and the rotameric state, with significant differences in the probability regions generated when different rotamers are considered. However, none of the rotamers tested for His15 led to a significant increase in the probability with respect to the apo structure, in which a high probability region was already identified. On the other hand, in the case of Metal1D, the dependence on the rotameric state of the side chain seems to be less pronounced, and the probability values associated with the predicted metal site are more consistent in the case of the probability map extracted from Pt-based compounds. This observation is in line with what was observed in the original publication by Dürer et al. [12], where Metal3D presented similar precision and recall for TM ions different from zinc, while Metal1D showed similar precision, but lower recall.

Comparison with cavity-detection predictors

A natural question arising from the analysis above is whether a general cavity-detection algorithm can identify the same metal-binding regions, or whether specialized predictors, such as Metal1D and Metal3D, provide added value. Over the years, many tools have been developed to detect binding pockets in protein structures [54,55], making use of different approaches to detect regions in the protein that can accommodate small molecules. In general, they primarily rely on geometric features of the protein surface, without explicitly taking into account the chemical nature of the residues involved.

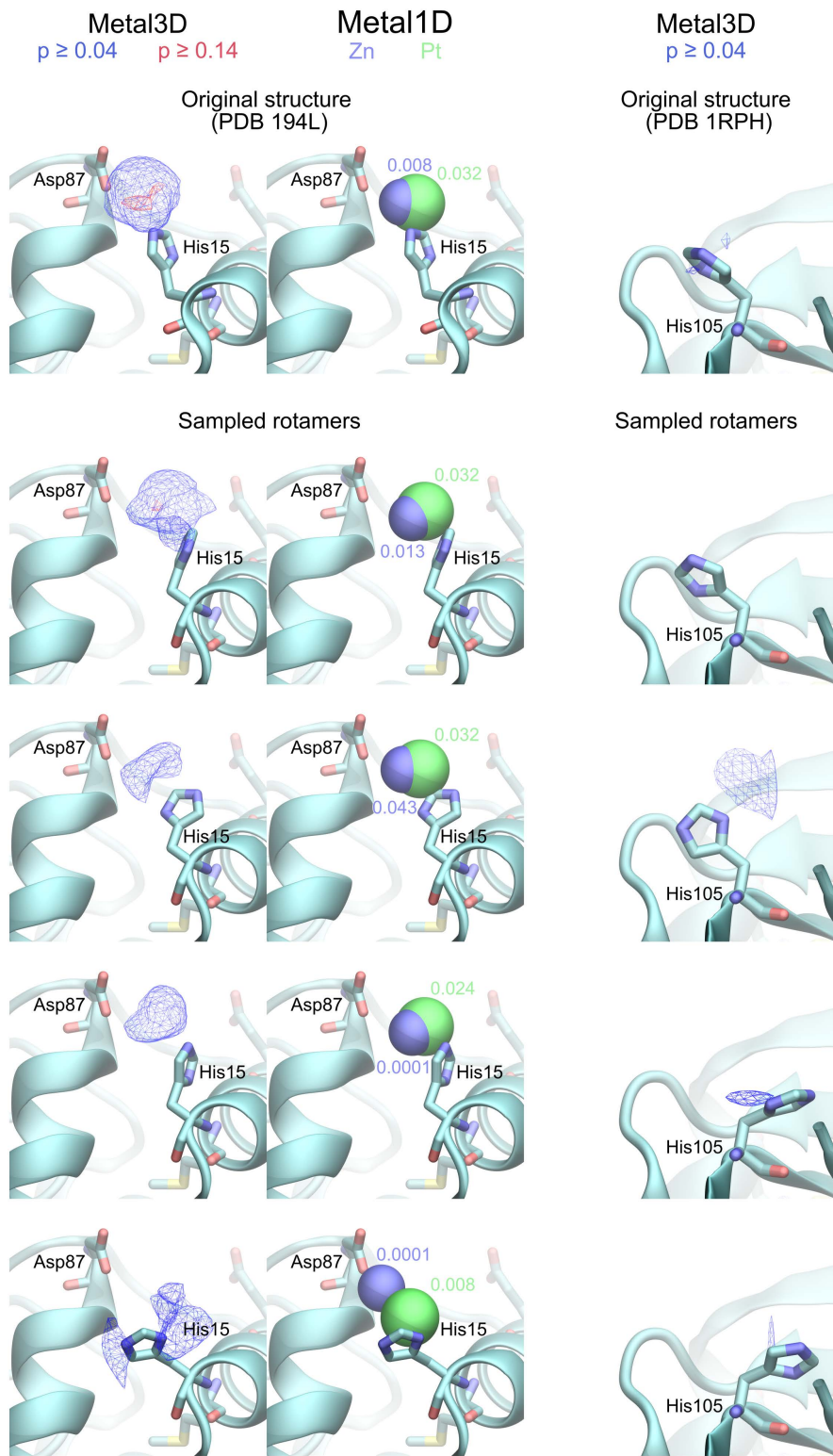


Fig 7. Metal3D and Metal1D predictions for different rotamers of His15 in HEWLC and His105 in RNaseA. In both cases, the initial structure is the apo structure, and only a single rotamer is changed. In the case of His105, Metal1D does not predict any site for the different rotamers, hence only Metal3D predictions are reported.

<https://doi.org/10.1371/journal.pone.0349622.g007>

In contrast, TM binding sites are often characterized by specific amino acid preferences and coordination geometries, and the local chemical environment plays a crucial role.

Nevertheless, we tested the detection ability of a cavity predictor to investigate whether such tools provide similar information as Metal1D and Metal3D, or complementary data. For this comparison, we used Fpocket [56,57], a popular cavity detection predictor which is based on Voronoi tessellation, publicly available through a web server. Fpocket relies on the concept of *alpha spheres*, defined as spheres that contact four atoms with no internal atoms [58], whose centers correspond to Voronoi vertices. By filtering alpha spheres according to their radii, Fpocket identifies potential cavities.

Fig 8 shows a comparison of the apo structure of the two proteins analysed in the previous sections, HEWLC and RNaseA. In the SI, we report a detailed comparison for all the structures in our dataset (S6 and S7 Figs), and we compute the TP, FN, and recall for the experimentally-known binding sites using similar criteria as those applied in the case of Metal1D and Metal3D (S5 Table). Specifically, we classified a pocket prediction as TP if any of the alpha spheres were located within 5 Å of an experimental metal center.

Across different structures, Fpocket identifies a relatively large number of pockets. Notably, the result is sensitive to the input structure: for example, in the apo structure of HEWLC (PDB 194L), seven pockets are identified, whereas for one of the holo structures in which the site at His15 is occupied (PDB 2I6Z), only one pocket is detected. In contrast, Metal1D and Metal3D predictions show greater consistency in identifying high-probability metal-binding regions across different structures. In the case of the apo structure of HEWLC, some alpha spheres are placed in proximity to all three binding sites, but this is not consistently observed for the holo structures of the same protein. For RNaseA, no pocket is located near the binding sites, except in a single holo structure (PDB 6JGL). When considering TP, FN, and recall values for the experimentally known sites, the results vary substantially depending on the protein and the site, with only a limited number of successful predictions for RNaseA.

Overall, this comparison highlights that cavity predictors alone are not sufficient to reliably identify the binding sites of TM-based compounds. Nevertheless, their combination with metal-site predictors, such as Metal1D or Metal3D, could provide complementary information, particularly regarding the available space within the pocket to accommodate the ligands coordinating the metal center in TM-based compounds.

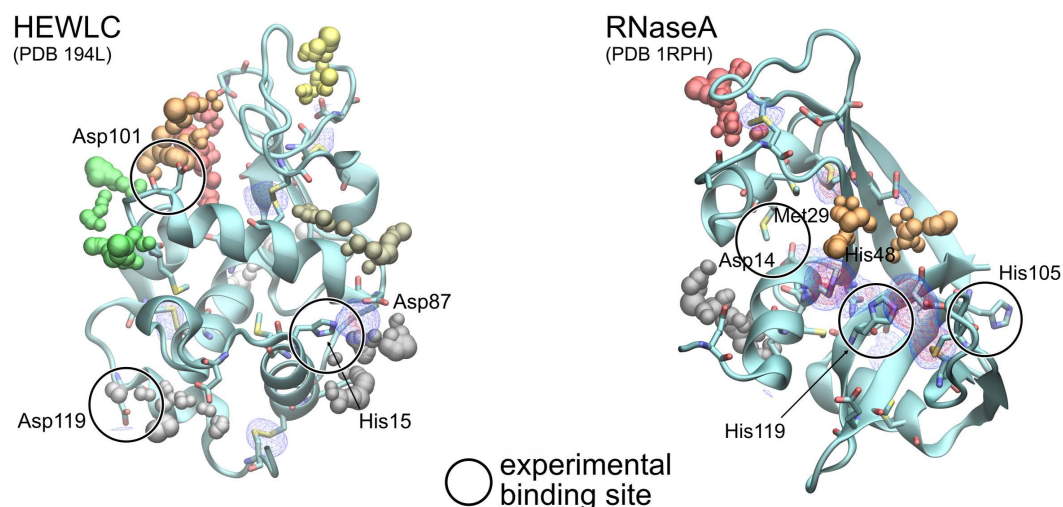


Fig 8. Cavity predictions with Fpocket for the apo structures of HEWLC (left, PDB 194L) and RNaseA (right, PDB 1RPH). For each prediction, alpha spheres identified by Fpocket are represented as colored spheres. The experimental binding sites are highlighted with circles. Metal3D predictions with low (blue) and high (red) probability are also represented as isosurfaces with wireframe representation.

<https://doi.org/10.1371/journal.pone.0349622.g008>

Bidentate glutamate-containing site. To further test the sensitivity to the rotamer conformations and the prediction ability of Metal1D and Metal3D, we also performed additional predictions for a known bidentate TM binding site in NCPs containing two glutamates (GLU site). For our work, this represents an interesting system considering that the two glutamates significantly change their rotameric state when a metal compound binds. Fig 9 shows the GLU site from the apo and holo structures, from PDB ID 3REH [34] and 5XF3 [36], respectively, as well as the predictions from Metal3D and Metal1D. Only in the case of the holo structure, in which the glutamates are in the conformation corresponding to the binding of the metal compound, the predictors identify the site. In the case of Metal3D, a high probability region with $p \sim 0.2$ can clearly be identified, i.e., with a significantly larger probability than for the sites tested so far, possibly thanks to the bidentate nature of the site. For Metal1D, the predictions with the three probability maps (zinc, platinum, and ruthenium) are all able to locate this binding site. However, in the apo structure, no prediction is made by either Metal3D or Metal1D.

When testing the influence of different rotamer conformations for the two glutamates in the GLU site, in the case of Glu61, none of the tested rotamers leads to a favorable conformation for metal binding (S6 Fig). As a result, neither Metal3D nor Metal1D identifies a binding region between Glu61 and Glu64. However, in the case of Glu64, some of the rotamers assume a 'favorable' conformation, pointing in the direction of Glu61. In particular, we report in Fig 9 the corresponding predictions for one of such rotamers, which lead to similar predictions as the holo structure with the bound drug. Similarly to what is observed for the holo structure, Metal3D identifies a high probability region ($p \sim 0.2$), and Metal1D locates a metal ion with all three probability maps tested.

This last comparison highlights the influence of rotamer sampling for bidentate sites, in which it is reasonable to assume that sampling different rotameric states from the apo structure is more likely to lead to a conformation in which

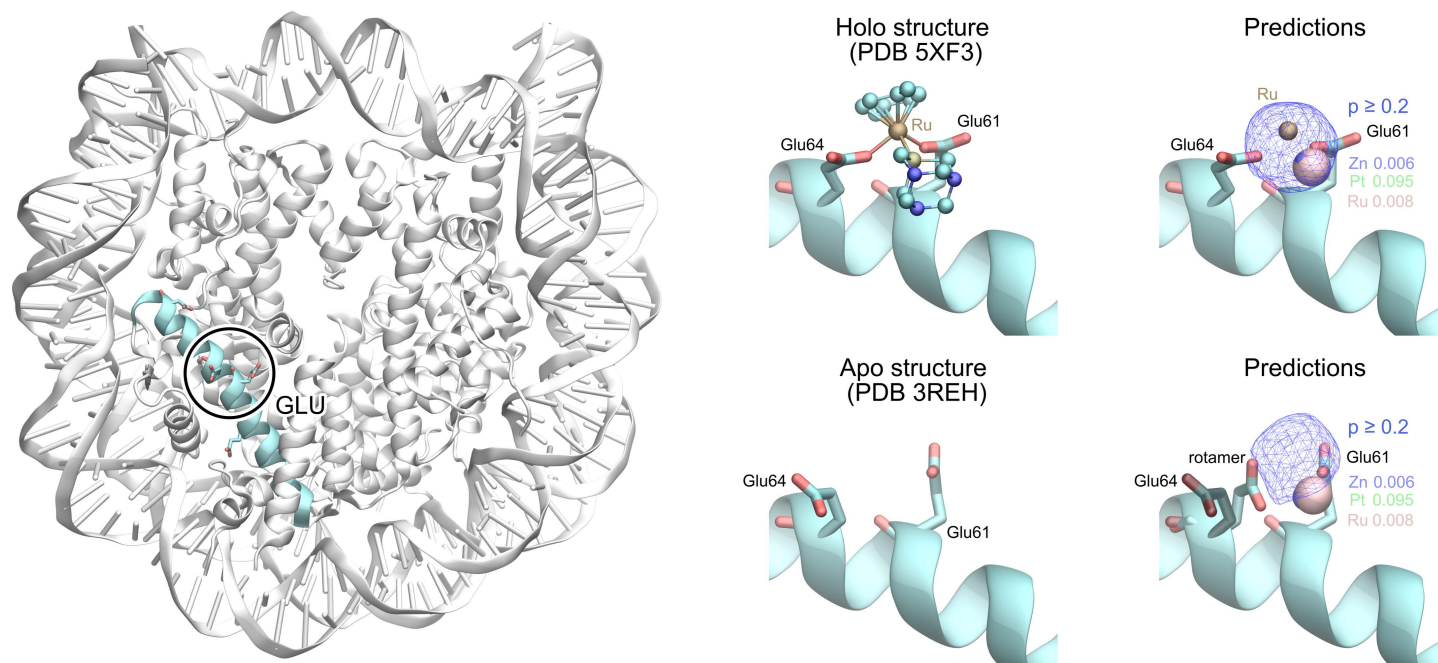


Fig 9. Metal3D and Metal1D predictions for the GLU site in the case of the holo (top) and apo (bottom) structures. For the latter, we also report the successful prediction after sampling a different rotamer for Glu64. In the case of Metal1D predictions, the metal ions located using the three different probability maps overlap, and only the one corresponding to ruthenium is visible, but the corresponding probability values for each probability map are indicated.

<https://doi.org/10.1371/journal.pone.0349622.g009>

two (or more) amino acids are in a favorable conformation for metal site binding, which is not necessarily favorable in the absence of a TM ion.

Conclusion

In this work, we explored the applicability of the Metal3D and Metal1D predictors [12] to identify the binding sites of TM-based compounds. Although they were originally designed to predict binding sites of zinc ions, our results demonstrate a surprising degree of predictive power for more complex TM binding sites. Both methods consistently identify some of the drug binding sites across the tested structures, enabling the identification of the primary binding sites from the apo conformations. However, their performance is limited in cases where the binding site involves only a single solvent-exposed residue.

Both predictors, particularly Metal3D, show sensitivity to the side chain positioning and rotameric state of the amino acids at the binding site. This dependence suggests that binding sites may be overlooked in apo structures if the key amino acids are not 'pre-organized' in a favorable orientation. To investigate this, we sampled alternative rotameric states for selected residues. For monodentate sites, where the TM-based compound binds a single amino acid, these variations had little impact on prediction outcomes, regardless of whether the original prediction was correct or incorrect. In contrast, for a bidentate site, introducing alternative rotamers enabled the correct identification of the binding site directly from the apo structure.

Overall, our findings suggest that Metal3D and Metal1D represent rapid, low-cost pre-screening tools to highlight putative 'hotspot' regions for the binding of TM-based compounds. These predictions can then be refined by more accurate methods, such as QM/MM simulations. To design a robust pipeline for predicting unknown binding sites in apo structures, however, several limitations must be addressed. In particular, a clear difference between a TM ion and a TM-based compound is the additional steric volume occupied by the ligands coordinating the metal center. This could be taken into account, e.g., in Metal3D by retraining the 3DCNN on datasets of TM complexes to properly capture the influence of the coordinating ligands, although the limited availability of such structures currently represents a major limitation to this strategy. For this purpose, the size of the boxes centered on the $C\alpha$ atoms may need to be increased, as the 16 Å cubic box used for zinc ions might be too small to fully capture the steric environment of the coordinating ligands. In the case of Metal1D, steric effects could be incorporated by extending the geometric criteria used in the scoring procedure. An alternative approach would be to combine TM-site predictions with cavity-detection algorithms. In contrast to specific predictors for TMs, such as Metal1D and Metal3D, which take into account the local chemical environment, cavity detectors primarily rely on geometric features of the protein surface. As shown in our comparison with Fpocket, the two types of tools tend to identify different regions of the protein structure, indicating that cavity detection alone is not sufficient to locate TM binding sites. Nevertheless, integrating the information from both approaches could be valuable, as cavity predictors can provide complementary information about the available space within a pocket. In this way, cavity detection could help assess whether a putative TM binding region identified by a dedicated predictor offers sufficient space to accommodate the additional ligands coordinating the metal center in a TM-based compound.

Moreover, in this study, we highlighted the dependence of the predictions on the side chain orientation by sampling alternative conformations for a single amino acid using the Richardson rotamer library [44]. This limitation is partly inherent to any approach based on single static structures. Possible improvements could involve extending the analysis to different rotamer libraries [59,60] and integrating the predictors with more extensive conformational sampling strategies, such as Monte Carlo methods or genetic algorithms optimization. Both approaches pose non-trivial challenges, as in Monte Carlo, suitable trial moves need to be established. Such moves must balance the increase in predictor probability with maintaining physically realistic conformations, avoiding steric clashes, and accounting for the energy cost of rotamer transitions, potentially including solvent effects. In the case of genetic algorithms, a possible solution includes the use of

multi-objective optimization [43] by designing a fitness function based on the probability from the predictors and combining it with one for the stability of the resulting conformation. This second fitness function should describe the protein stability, taking into account the protein–metal interaction, but this cannot be easily evaluated, e.g., with classical FFs, due to their limitations in describing TMs.

Another crucial limitation is the dataset availability. For Metal1D, we generated probability maps from Pt- and Ru-containing structures, but their number is extremely small compared, e.g., to Zn. We observed some advantages in the case of the Pt-based probability map, but the Ru-based one was unsuccessful, possibly due to the extremely limited set of structures available for its generation, and we discourage its use in the current form. Expanding the dataset, potentially by combining related transition metals, could improve prediction accuracy and may eventually provide sufficient data to train ML models such as Metal3D. For such ML-based approaches, the scarcity of experimentally resolved structures of TM-based compounds remains a major challenge, as it strongly limits the size and diversity of available training datasets.

In conclusion, while not yet fully general, Metal3D and Metal1D represent a valuable first step toward efficient prediction of TM compound binding sites. Further work will focus on addressing the limitations identified, with the ultimate goal of developing robust predictors capable of identifying unknown binding sites for TM-based compounds in proteins.

Supporting information

S1 Table. Summary of the true positive (TP), false negative (FN), and recall values for the different predictors and probability thresholds/maps. For each case, the values reported are: TP/FN (recall). For Metal3D, the thresholds for p(low), p(medium), and p(high) are 0.04, 0.09, and 0.14 for HEWLC, and 0.06, 0.11, and 0.16 for RNaseA.
(PDF)

S2 Table. RMSD values for the different X-ray structures considered for the HEWLC. All structures have been aligned to the backbone of the apo structure (PDB ID 194L). The RMSD corresponding to an occupied binding site is indicated in bold, and an asterisk indicates a rotamer change with respect to the reference (apo) structure.
(PDF)

S3 Table. RMSD values for the different X-ray structures considered for the RNaseA enzyme. All structures have been aligned to the backbone of the apo structure (PDB ID 19PH). The RMSD corresponding to an occupied binding site is indicated in bold, and an asterisk indicates a rotamer change with respect to the reference (apo) structure.
(PDF)

S4 Table. MSD values for the His105 residue in the different X-ray structures considered for the RNaseA enzyme. All structures have been aligned to the backbone of the apo structure (PDB ID 1RPH) and the RMSD is computed for all the atoms of His105 using the structure of PDB 4S18 as reference, for which Metal3D made a successful prediction.
(PDF)

S5 Table. Summary of the true positive (TP), false negative (FN), and recall values for the sites identified with the cavity-detector predictor Fpocket. For each case, the values reported are: TP/FN (recall).
(PDF)

S1 Fig. Metal3D predictions for HEWLC made on the different crystal structures. Metal3D probability densities are represented in different colors as isosurfaces with a wireframe representation. The isovalue for the low probability predictions (blue) is 0.04, while the one for higher probability predictions (red) is 0.14. In the cases where no region had probability higher than 0.14, a lower isovalue is used (0.09, green).
(PNG)

S2 Fig. Metal1D predictions for HEWLC made on the different crystal structures with different probability maps. Zinc (blue), platinum (green), and ruthenium (pink). For each prediction, the probability associated with each site is indicated with the same color as the corresponding probability map.

(PNG)

S3 Fig. Comparison of the environment around the His105 residue. All structures have been aligned to the backbone of the apo structure (PDB ID 19PH), and the probability density generated from Metal3D for PDB 4S18 is also represented in blue.

(PNG)

S4 Fig. Metal3D predictions for RNaseA made on the different crystal structures. Metal3D probability densities are represented in different colors as isosurfaces with a wireframe representation. The isovalue for the low probability predictions (blue) is 0.06, while the one for higher probability predictions (red) is 0.16.

(PNG)

S5 Fig. Metal1D predictions for RNaseA made on the different crystal structures with different probability maps. Zinc (blue), platinum (green), and ruthenium (pink). For each prediction, the probability associated with each site is indicated with the same color as the corresponding probability map.

(PNG)

S6 Fig. Cavity predictions with Fpocket for HEWLC made on the different crystal structures. For each prediction, alpha spheres are represented as colored spheres. Metal3D predictions with low (blue) and high (red) probability are also represented as isosurfaces with wireframe representation.

(PNG)

S7 Fig. Cavity predictions with Fpocket for RNaseA made on the different crystal structures. For each prediction, alpha spheres are represented as colored spheres. Metal3D predictions with low (blue) and high (red) probability are also represented as isosurfaces with wireframe representation.

(PNG)

S8 Fig. GLU site conformations for the apo and holo protein from different PDB structures. For the apo structure (PDB ID 3REH), the different sampled rotamers for Glu61 and Glu61 are also reported.

(PNG)

Author contributions

Conceptualization: Andrea Levy, Ursula Rothlisberger.

Data curation: Andrea Levy.

Formal analysis: Andrea Levy.

Funding acquisition: Ursula Rothlisberger.

Investigation: Andrea Levy.

Methodology: Andrea Levy, Ursula Rothlisberger.

Project administration: Ursula Rothlisberger.

Resources: Ursula Rothlisberger.

Software: Andrea Levy.

Supervision: Ursula Rothlisberger.

Validation: Andrea Levy.

Visualization: Andrea Levy.

Writing – original draft: Andrea Levy.

Writing – review & editing: Andrea Levy, Ursula Rothlisberger.

References

- Palermo G, Magistrato A, Riedel T, von Erlach T, Davey CA, Dyson PJ, et al. Fighting Cancer with Transition Metal Complexes: From Naked DNA to Protein and Chromatin Targeting Strategies. *ChemMedChem*. 2016;11(12):1199–210. <https://doi.org/10.1002/cmdc.201500478> PMID: 26634638
- Abdolmaleki S, Aliabadi A, Khaksar S. Bridging the gap between theory and treatment: Transition metal complexes as successful candidates in medicine. *Coord Chem Rev*. 2025;531:216477. <https://doi.org/10.1016/j.ccr.2025.216477>
- Ghosh S. Cisplatin: The first metal based anticancer drug. *Bioorg Chem*. 2019;88:102925. <https://doi.org/10.1016/j.bioorg.2019.102925> PMID: 31003078
- Morris GM, Lim-Wilby M. Molecular Docking. In: Kukol A, editor. *Molecular Docking*. Totowa (NJ): Humana Press; 2008. p. 365–82.
- Paggi JM, Pandit A, Dror RO. The Art and Science of Molecular Docking. *Annu Rev Biochem*. 2024;93(1):389–410. <https://doi.org/10.1146/annurev-biochem-030222-120000>
- Kumalo HM, Bhakat S, Soliman MES. Theory and applications of covalent docking in drug discovery: merits and pitfalls. *Molecules*. 2015;20(2):1984–2000. <https://doi.org/10.3390/molecules20021984> PMID: 25633330
- Oyedele A-QK, Ogunlana AT, Boyenle ID, Adeyemi AO, Rita TO, Adelusi TI, et al. Docking covalent targets for drug discovery: stimulating the computer-aided drug design community of possible pitfalls and erroneous practices. *Mol Divers*. 2023;27(4):1879–903. <https://doi.org/10.1007/s11030-022-10523-4> PMID: 36057867
- Li P, Merz Jr KM. Metal Ion Modeling Using Classical Mechanics. *Chem Rev*. 2017;117(3):1564–686. <https://doi.org/10.1021/acs.chemrev.6b00440>
- Warshel A, Levitt M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol*. 1976;103(2):227–49. [https://doi.org/10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9) PMID: 985660
- Brunk E, Rothlisberger U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem Rev*. 2015;115(12):6217–63. <https://doi.org/10.1021/cr500628b>
- Lipparini F, Mennucci B. Hybrid QM/classical models: Methodological advances and new applications. *Chem Phys Rev*. 2021;2(4). <https://doi.org/10.1063/5.0064075>
- Dürr SL, Levy A, Rothlisberger U. Metal3D: a general deep learning framework for accurate metal ion location prediction in proteins. *Nat Commun*. 2023;14(1):2713. <https://doi.org/10.1038/s41467-023-37870-6> PMID: 37169763
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235–42. <https://doi.org/10.1093/nar/28.1.235> PMID: 10592235
- Vaney M, Maignan S, Ries-Kautt M, Ducruix A. The 1.40 Å structure of SpaceHab-01 Hen Egg White Lysozyme. *Protein Data Bank*. 1995. <https://doi.org/10.2210/pdb194/pdb>
- Hase WL. Simulations of Gas-Phase Chemical Reactions: Applications to SN2 Nucleophilic Substitution. *Science*. 1994;266(5187):998–1002. <https://doi.org/10.1126/science.266.5187.998> PMID: 17779941
- Temperini C, Casini A, Messori L. X-ray diffraction studies of adducts between anticancer platinum drugs and hen egg white lysozyme. *Protein Data Bank*. 2006. <https://doi.org/10.2210/pdb2i6z/pdb>
- Casini A, Mastrobuoni G, Temperini C, Gabbiani C, Francese S, Moneti G, et al. ESI mass spectrometry and X-ray diffraction studies of adducts between anticancer platinum drugs and hen egg white lysozyme. *Chem Commun (Camb)*. 2006;(2):156–8. <https://doi.org/10.1039/b611122j> PMID: 17180231
- Ferraro G, Merlino A. The X-ray structure of the adduct formed in the reaction between hen egg white lysozyme and compound 3, a platin(II) compound containing a O, S bidentate ligand. 2016. <https://doi.org/10.2210/pdb5ii3/pdb>
- Hildebrandt J, Häfner N, Görls H, Kritsch D, Ferraro G, Dürst M, et al. Platinum(II) O,S complexes as potential metallodrugs against Cisplatin resistance. *Dalton Trans*. 2016;45(47):18876–91. <https://doi.org/10.1039/c6dt01388k> PMID: 27897281
- Merlino A, Ferraro G. The X-ray structure of the adduct formed in the reaction between hen egg white lysozyme and complex I, a pentacoordinate Pt(II) compound containing 2,9-dimethyl-1,10-phenanthroline, dimethylfumarate, methyl and iodine as ligands. 2019. <https://doi.org/10.2210/pdb6qea/pdb>
- Ferraro G, Marzo T, Cucciolito ME, Ruffo F, Messori L, Merlino A. Reaction with Proteins of a Five-Coordinate Platinum(II) Compound. *Int J Mol Sci*. 2019;20(3):520. <https://doi.org/10.3390/ijms20030520> PMID: 30691130
- Sullivan MP, Hartinger CG, Goldstone DC. Ruthenium(II)(cymene)(chlorido)2-lysozyme adduct with two binding sites. 2017. <https://doi.org/10.2210/pdb5v4g/pdb>

23. Sullivan MP, Groessl M, Meier SM, Kingston RL, Goldstone DC, Hartinger CG. The metalation of hen egg white lysozyme impacts protein stability as shown by ion mobility mass spectrometry, differential scanning calorimetry, and X-ray crystallography. *Chem Commun (Camb)*. 2017;53(30):4246–9. <https://doi.org/10.1039/c6cc10150j> PMID: 28361137
24. Sullivan MP, Hartinger CG, Goldstone DC. The interaction of dichlorido(3-(anthracen-9-ylmethyl)-1-methylimidazol-2-ylidene)(η^6 -p-cymene) ruthenium(II) with HEWL after 1 week. 2020. <https://doi.org/10.2210/pdb6wgo/pdb>
25. Lee BYT, Sullivan MP, Yano E, Tong KKH, Hanif M, Kawakubo-Yasukochi T, et al. Anthracenyl Functionalization of Half-Sandwich Carbene Complexes: In Vitro Anticancer Activity and Reactions with Biomolecules. *Inorg Chem*. 2021;60(19):14636–44. <https://doi.org/10.1021/acs.inorgchem.1c01675> PMID: 34528438
26. Zegers I, Wyns L, Palmer R. Structures of RNase A complexed with 3'-CMP and d(CpA): active site conformation and conserved water molecules. Protein Data Bank. 1994. <https://doi.org/10.2210/pdb1rph/pdb>
27. Zegers I, Maes D, Dao-Thi MH, Poortmans F, Palmer R, Wyns L. The structures of RNase A complexed with 3'-CMP and d(CpA): active site conformation and conserved water molecules. *Protein Sci*. 1994;3(12):2322–39. <https://doi.org/10.1002/pro.5560031217> PMID: 7756988
28. Merlino A. The X-ray structure of the adduct formed in the reaction between bovine pancreatic ribonuclease and carboplatin. Protein Data Bank; 2015. <https://doi.org/10.2210/pdb4s0q/pdb>
29. Messori L, Marzo T, Merlino A. Interactions of carboplatin and oxaliplatin with proteins: Insights from X-ray structures and mass spectrometry studies of their ribonuclease A adducts. *J Inorg Biochem*. 2015;153:136–42. <https://doi.org/10.1016/j.jinorgbio.2015.07.011> PMID: 26239545
30. Merlino A. The X-ray structure of the adduct formed in the reaction between bovine pancreatic ribonuclease and oxaliplatin. Protein Data Bank; 2015. <https://doi.org/10.2210/pdb4s18/pdb>
31. Ferraro G, Merlino A. The X-ray structure of the adduct formed in the reaction between bovine pancreatic ribonuclease and compound I, a piano-stool organometallic Ru(II) arene compound containing an O,S-chelating ligand. Protein Data Bank; 2016. <https://doi.org/10.2210/pdb5jlg/pdb>
32. Hildebrandt J, Görls H, Häfner N, Ferraro G, Dürst M, Runnebaum IB, et al. Unusual mode of protein binding by a cytotoxic π -arene ruthenium(ii) piano-stool compound containing an O,S-chelating ligand. *Dalton Trans*. 2016;45(31):12283–7. <https://doi.org/10.1039/c6dt02380k> PMID: 27427335
33. Levy A, Adhikarsan Z, von Erlach T, Palermo G, Nazarov AA, Hartinger CG, et al. Differential targeting of the nucleosome surface and superhelical crevice sites with Ru and Os organometallic agents. *bioRxiv*. 2025. <https://doi.org/10.1101/2025.11.13.688318>
34. Wu B, Davey CA. 2.5 angstrom crystal structure of the nucleosome core particle assembled with a 145 bp alpha-satellite DNA (NCP145). Protein Data Bank; 2011. <https://doi.org/10.2210/pdb3reh/pdb>
35. Wu B, Davey GE, Nazarov AA, Dyson PJ, Davey CA. Specific DNA structural attributes modulate platinum anticancer drug site selection and cross-link generation. *Nucleic Acids Res*. 2011;39(18):8200–12. <https://doi.org/10.1093/nar/gkr491> PMID: 21724603
36. Ma Z, Adhikarsan Z, Murray BS, Dyson PJ, Davey CA. 2.5 Angstrom Crystal Structure of the Nucleosome Core Particle Assembled with a 145 bp Alpha-Satellite DNA (NCP145). Protein Data Bank; 2017. <https://doi.org/10.2210/pdb5xf3/pdb>
37. Adhikarsan Z, Davey CA. X-ray structure of NCP145 with chlorido(eta-6-p-cymene)(N-phenyl-2-pyridinecarbothioamide)osmium(II). Protein Data Bank; 2013. <https://doi.org/10.2210/pdb4j8u/pdb>
38. Meier SM, Hanif M, Adhikarsan Z, Pichler V, Novak M, Jirkovsky E. Novel metal(ii) arene 2-pyridinecarbothioamides: a rationale to orally active organometallic anticancer agents. *Chem Sci*. 2013;4(4):1837–46. <https://doi.org/10.1039/C3SC22294B>
39. Adhikarsan Z, Davey CA. X-ray structure of NCP145 with chlorido(eta-6-p-cymene)(N-fluorophenyl-2-pyridinecarbothioamide)osmium(II). Protein Data Bank; 2013. <https://doi.org/10.2210/pdb4j8w/pdb>
40. Adhikarsan Z, Ma Z, Davey CA. Nucleosome core particle containing adducts of gold(I)-triethylphosphane and ruthenium(II)-toluene PTA complexes. Protein Data Bank; 2015. <https://doi.org/10.2210/pdb5dnn/pdb>
41. Adhikarsan Z, Palermo G, Riedel T, Ma Z, Muhammad R, Rothlisberger U, et al. Allosteric cross-talk in chromatin can mediate drug-drug synergy. *Nat Comm*. 2017;8:14860. <https://doi.org/10.1038/ncomms14860> PMID: 28358030
42. Humphrey W, Dalke A, Schulten K. VMD – Visual Molecular Dynamics. *J Mol Graphics*. 1996;14:33–8. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
43. Browning NJ, Perez MAS, Brunk E, Rothlisberger U. EVOLVE: An Evolutionary Toolbox for the Design of Peptides and Proteins. *ChemRxiv*. 2023. <https://doi.org/10.26434/chemrxiv-2023-gb366>
44. Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. *Proteins*. 2000;40(3):389–408. [https://doi.org/10.1002/1097-0134\(20000815\)40:3<389::aid-prot50>3.0.co;2-2](https://doi.org/10.1002/1097-0134(20000815)40:3<389::aid-prot50>3.0.co;2-2) PMID: 10861930
45. Harding MM. The architecture of metal coordination groups in proteins. *Acta Crystallogr D Biol Crystallogr*. 2004;60(Pt 5):849–59. <https://doi.org/10.1107/S09074444904004081> PMID: 15103130
46. Zheng H, Chruszcz M, Lasota P, Lebioda L, Minor W. Data mining of metal ion environments present in protein structures. *J Inorg Biochem*. 2008;102(9):1765–76. <https://doi.org/10.1016/j.jinorgbio.2008.05.006> PMID: 18614239

47. Dokmanić I, Sikić M, Tomić S. Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination. *Acta Crystallogr D Biol Crystallogr*. 2008;64(Pt 3):257–63. <https://doi.org/10.1107/S090744490706595X> PMID: [18323620](#)
48. Barber-Zucker S, Shaanan B, Zarivach R. Transition metal binding selectivity in proteins and its correlation with the phylogenomic classification of the cation diffusion facilitator protein family. *Sci Rep*. 2017;7(1):16381. <https://doi.org/10.1038/s41598-017-16777-5> PMID: [29180655](#)
49. Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature*. 1965;206(4986):757–61. <https://doi.org/10.1038/206757a0> PMID: [5891407](#)
50. Blake CC, Mair GA, North AC, Phillips DC, Sarma VR. On the conformation of the hen egg-white lysozyme molecule. *Proc R Soc Lond B Biol Sci*. 1967;167(1009):365–77. <https://doi.org/10.1098/rspb.1967.0034> PMID: [4382800](#)
51. Weiss MS, Palm GJ, Hilgenfeld R. Crystallization, structure solution and refinement of hen egg-white lysozyme at pH 8.0 in the presence of MPD. *Acta Crystallogr D Biol Crystallogr*. 2000;56(Pt 8):952–8. <https://doi.org/10.1107/s0907444900006685> PMID: [10944331](#)
52. HIRS CH, MOORE S, STEIN WH. The sequence of the amino acid residues in performic acid-oxidized ribonuclease. *J Biol Chem*. 1960;235:633–47. [https://doi.org/10.1016/s0021-9258\(19\)67917-2](https://doi.org/10.1016/s0021-9258(19)67917-2) PMID: [14401957](#)
53. Smyth DG, Stein WH, Moore S. The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations. *J Biol Chem*. 1963;238:227–34. [https://doi.org/10.1016/s0021-9258\(19\)83984-4](https://doi.org/10.1016/s0021-9258(19)83984-4) PMID: [13989651](#)
54. Kawabata T, Go N. Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins*. 2007;68(2):516–29. <https://doi.org/10.1002/prot.21283> PMID: [17444522](#)
55. Zheng X, Gan L, Wang E, Wang J. Pocket-based drug design: exploring pocket space. *AAPS J*. 2013;15(1):228–41. <https://doi.org/10.1208/s12248-012-9426-6> PMID: [23180158](#)
56. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*. 2009;10:168. <https://doi.org/10.1186/1471-2105-10-168> PMID: [19486540](#)
57. Schmidtke P, Le Guilloux V, Maupetit J, Tuffery P. fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res*. 2010;38(Web Server issue):W582–9. <https://doi.org/10.1093/nar/gkq383> PMID: [20478829](#)
58. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci*. 1998;7(9):1884–97. <https://doi.org/10.1002/pro.5560070905> PMID: [9761470](#)
59. Dunbrack RL Jr, Karplus M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J Mol Biol*. 1993;230(2):543–74. <https://doi.org/10.1006/jmbi.1993.1170> PMID: [8464064](#)
60. Shapovalov MV, Dunbrack RL. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*. 2011;19(6):844–58. <https://doi.org/10.1016/j.str.2011.03.019> PMID: [21645855](#)