

GM-DockZn: A Geometry Matching based Docking Algorithm for Zinc Proteins

Author

Wang, Kai, Lyu, Nan, Diao, Hongjuan, Jin, Shujuan, Zeng, Tao, Zhou, Yaoqi, Wu, Ruibo

Published

2020

Journal Title

Bioinformatics

Version

Accepted Manuscript (AM)

DOI

[10.1093/bioinformatics/btaa292](https://doi.org/10.1093/bioinformatics/btaa292)

Rights statement

© 2020 Oxford University Press. This is a pre-copy-editing, author-produced PDF of an article accepted for publication in Bioinformatics following peer review. The definitive publisher-authenticated version GM-DockZn: A Geometry Matching based Docking Algorithm for Zinc Proteins, Bioinformatics, 2020 is available online at: <https://doi.org/10.1093/bioinformatics/btaa292>.

Downloaded from

<http://hdl.handle.net/10072/393814>

Griffith Research Online

<https://research-repository.griffith.edu.au>

Structural Bioinformatics

GM-Dock_{Zn}: A Geometry Matching based Docking Algorithm for Zinc Proteins

Kai Wang^{1,2,#}, Nan Lyu^{1,#}, Hongjuan Diao¹, Shujuan Jin^{3,4}, Tao Zeng¹, Yaoqi Zhou^{3,4,5*}, Ruibo Wu^{1,5,*}

¹ Guangdong Provincial Key Laboratory of New Drug Design and Evaluation, School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, P. R. China.

² School of agriculture and biology, Zhongkai University of Agriculture and Engineering, Guangzhou 510000, P. R. China

³ Peking University Shenzhen Graduate School, Shenzhen, 518055, P.R.China

⁴ Shenzhen Bay Laboratory, Shenzhen, 518055, P.R.China

⁵ Institute for Glycomics and School of Information and Communication Technology, Griffith University, Southport, QLD 4222, Australia

These authors contributed equally to this work.

* To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Molecular docking is a widely used technique for large-scale virtual screening of the interactions between small-molecule ligands and their target proteins. However, docking methods often perform poorly for metalloproteins due to additional complexity from the three-way interactions among amino acid residues, metal ions, and ligands. This is a significant problem because zinc proteins alone comprise about 10% of all available protein structures in the protein databank. Here, we developed GM-Dock_{Zn} that is dedicated for ligand docking to zinc proteins. Unlike the existing docking methods developed specifically for zinc proteins, GM-Dock_{Zn} samples ligand conformations directly using a geometric grid around the ideal zinc coordination positions of 7 discovered coordination motifs, which were found from the survey of known zinc proteins complexed with a single ligand.

Results: GM-Dock_{Zn} has the best performance in sampling near-native poses with correct coordination atoms and numbers within the top 50 and top 10 predictions when compared to several state-of-the-art techniques. This is true not only for a nonredundant dataset of zinc proteins but also for a homolog set of different ligand and zinc-coordination systems for the same zinc proteins. Similar superior performance of GM-Dock_{Zn} for near-native-pose sampling was also observed for docking to apo-structures and cross docking between different ligand complex structures of the same protein. The highest success rate for sampling nearest near-native poses within top 5 and top 1 was achieved by combining GM-Dock_{Zn} for conformational sampling with GOLD for ranking. The proposed geometry-based sampling technique will be useful for ligand docking to other metalloproteins.

Availability: GM-Dock_{Zn} is freely available at www.gmclab.com/ for academic users

Contact: wurb3@mail.sysu.edu.cn; yaoqi.zhou@griffith.edu.au

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

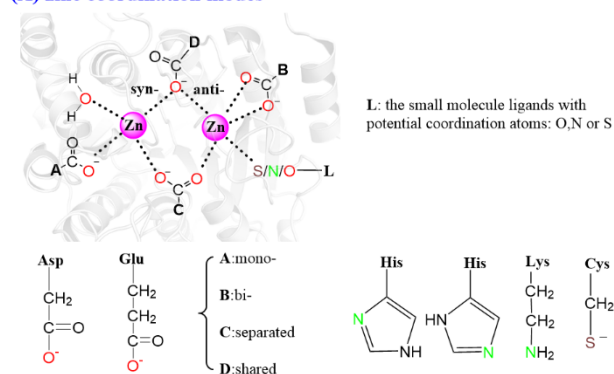
Zinc is the second most abundant trace transition metal found in living organisms. This is reflected from the fact that about 10% of the structures deposited in protein data bank (PDB: www.rcsb.org) are zinc metalloproteins (Berman, et al., 2000; Burley, et al., 2019). Zinc proteins have a multitude of essential functions, including catalysis, storage, transportation, transcription, and replication (Anzellotti and Farrell, 2008; Krezel and Maret, 2016; Maret, 2005; Maret, 2011; Maret, 2012; Parkin, 2004). Central to the functions of these zinc proteins is their zinc-coordination motifs. As shown in Fig. 1A, each zinc ion is located at a center coordinated with oxygen (O), nitrogen (N) or sulfur (S) atoms, which can be contributed by either a small molecule ligand, a water molecule, or an amino acid residue in the zinc protein. Typical zinc-binding amino-acid residues found in zinc proteins are cysteine, lysine, histidine, aspartate, and glutamic acid. Previous analysis (Andreini and Bertini, 2012; Auld, 2001; Maret and Li, 2009) indicates that zinc interacts with the thiolate group in cysteine, the amino group in lysine, one of the two nitrogen atoms of the imidazole ring in histidine, and one oxygen (syn- or anti-) or two oxygen (mono- or bi-) atoms of the carboxylate substituent in glutamate and aspartate whereas the Glu, Asp, and water ligands can bridge one zinc ions separately or share several together. Zinc coordination numbers (C.N.) range from 4 to 6 within the first zinc coordination shell in a tetrahedral, trigonal bipyramidal or octahedral geometry, respectively (Fig. 1B) (Andreini and Bertini, 2012; Auld, 2001; Harding, 2001; Koca, et al., 2003; Maret and Li, 2009; Roe and Pang, 1999). Several algorithms have been developed to predict the zinc sites based on the sequences or 3D structures of target proteins (Shu, et al., 2008; Zhao, et al., 2011). In addition, the coordination number may be dynamically varied between 4 and 5 or 5 and 6 depending on specific zinc proteins and the atoms in the coordination shells beyond the first shell according to the previous Quantum Mechanics/Molecular Mechanics (QM/MM) molecular dynamics (MD) simulations (Dudev and Lim, 2007; Wu, et al., 2009).

Many zinc proteins are established, or potential, drug targets (Anzellotti and Farrell, 2008; Krezel and Maret, 2016; Parkin, 2004). While progresses were made in molecular docking algorithms (Ballester and Mitchell, 2010; Boyles, et al., 2020; Cang and Wei, 2017; Johansson-Akhe, et al., 2020; Lu, et al., 2019; Schneider, et al., 2020; Velazquez-Libera, et al., 2020; Wang, et al., 2019; Zhang and Sanner, 2019) according to recent assessment (Li, et al., 2014; Su, et al., 2019), metalloproteins were found more challenging than nonmetalloproteins for docking because of additional interactions involving with metal ions. Hu et al (Hu, et al., 2004) showed that a correct zinc-coordination geometry is essential for the state-of-the-art docking software FlexX, Autodock, and GOLD (Jones, et al., 1995; Jones, et al., 1997; Kramer, et al., 1999; Morris, et al., 2009; Rarey, et al., 1996; Trott and Olson, 2010) to achieve a reasonable prediction. This leads to several zinc-protein-specific sampling techniques. FlexX (Kramer, et al., 1999; Rarey, et al., 1996) defines the interaction types and interaction geometry of a metal ion to score protein-ligand interactions in part based on the root-mean-squared deviation (RMSD) between the list of angles in the actual geometry and those in the ideal geometry of the same length. The ideal geometries in FlexX are trigonal bipyramidal, square-based pyramidal, tetrahedral and octahedral. A fragment-based approach is used for ligand docking. Glide XP (Friesner, et al., 2004; Friesner, et al., 2006; Halgren, et al., 2004) and GOLD (Jones, et al., 1995; Jones, et al., 1997), on the other hand, treat metal coordination interactions as special hydrogen bonds. Glide XP performs a grid-based docking conformational search in the functional pocket of the target protein whereas GOLD recognizes both tetrahedral and octahedral geometrical arrangements based on the angles between the metal ion and a pair of coordination positions. Ideal

coordination positions in the binding pocket were used to map the ligand acceptors to the coordination positions around the metal ion. AutoDock_{Zn} (Santos-Martins, et al., 2014) developed a zinc-specific potential to account for both the energetic and geometric (tetrahedral) components of zinc-associated interactions. More recently, force-field-based and knowledge-based scoring functions are combined to improve the ligand-binding ranking for zinc proteins in MpsDock_{Zn} (Bai, et al., 2015).

In this paper, we developed a new zinc-specific method denoted as GM-Dock_{Zn} for docking a small molecule ligand onto the zinc coordination shell of zinc proteins. Unlike previous methods where geometric models were employed as a docking filter, GM-Dock_{Zn} directly restricts potential coordination atoms in a ligand around ideal geometric models. Moreover, GM-Dock_{Zn} employed 7 ligand-coordination motifs (two in tetrahedral, three in trigonal bipyramidal and two in octahedral geometries) that were found in a survey of zinc protein structures. This new algorithm is shown to significantly improve over several docking programs in locating near-native conformations with correct poses and zinc coordination motifs within the top 10 or top 50 predictions. Its combination with GOLD yields the highest success rate in top 5 and top 1 predictions.

(A) zinc coordination modes



(B) ideal coordination models

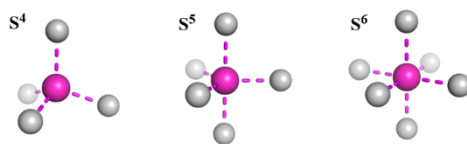


Fig. 1. (A) The representative zinc coordination shell in zinc proteins. The “ABCD” indicate the possible coordination modes of Asp/Glu. (B) The ideal zinc coordination models (S^4 , S^5 and S^6) refer to the standard tetrahedral, trigonal bipyramidal and octahedral geometries.

2 Methods

2.1 Datasets.

We obtained 9629 entries of zinc protein structures deposited in the PDB in October 2016. Excluding low-resolution structures ($> 2.5 \text{ \AA}$) led to 6553 zinc proteins with a total of 13,845 zinc coordination geometries because many proteins contain more than one zinc ion. We further separated these zinc geometries into amino-acid-only (11,589) (AA) and small-molecule-containing (2256) structures (SM, see Table 1). Amino-acid-only structures refer to those structures whose coordination positions are all occupied by amino acid residues. Small-molecule-containing structures contain at least one non-amino-acid atom in zinc coordination positions. These small molecules may be small molecule ligands or the molecules employed in the solution for crystallization such as water, SO_4^{2-} , PO_4^{3-} , and acetic acid. To avoid the complexity often associated

with dynamic solvent molecules, we further extract an SL set (685 structures) from the small-molecule set by limiting the structures with a single ligand only (no solvent ions) plus amino-acid residues in the first zinc coordination shell (SL). Here, we defined the first zinc coordination shell by the distance thresholds: 2.8 Å for Zn-S and 2.5 Å for zinc and the other coordination atoms as before (Andreini and Bertini, 2012; Auld, 2001).

Table 1. The datasets of zinc proteins (resolution < 2.5 Å) from PDB (Oct. 2016) along with the number of structures at different coordination numbers with different coordination motifs (ZN_{2,2}, ZN_{3,1}, ZN_{2,3}, ZN_{3,2}, ZN_{4,1}, ZN_{3,3} and ZN_{4,2}; ZN_{P,L}. “P” stands for the number of coordination atoms from amino acid residues and “L” stands for the number of coordination atoms from a ligand).

Set Name ^a	C.N.=4		C.N.=5			C.N.=6		Total
AA	10057		1064			468		11589
SM	1358		658			240		2256
SL	364		244			77		685
	10	354	1	178	65	12	65	
	ZN _{2,2}	ZN _{3,1}	ZN _{2,3}	ZN _{3,2}	ZN _{4,1}	ZN _{3,3}	ZN _{4,2}	
Test	40		45			23		108
NR	16		10			6		32
HOMO	20		15			2		37

^aAA: all coordination atoms are from amino acid residues; SM: at least one or more coordination atoms from small molecules; SL: only amino acid residues and a single ligand (not solvent molecules) contributing to coordination atoms; Test: a randomly selected set from the SL set; NR: a non-redundant set at 30% sequence identity cutoff for proteins from the Test set; HOMO: 3 proteins with 37 ligand-protein complexes from the Test set.

To compare the performance among the multiple methods on an equal footing, 108 zinc proteins were randomly selected from the SL set to serve as the test set (Test). The list of PDB IDs for the Test set along with the details on structural resolution and specific ligands is shown in Supplementary Table S1. This test set has 44 and 26 proteins in common with the test sets for FlexX (Rarey, et al., 1996) and MpsDock_{Zn} (Bai, et al., 2015), respectively. To remove potential biases due to homologous proteins in the test set, we obtained a non-redundant test set (NR) (32 proteins) by excluding proteins with sequence similarity >30% (calculated by CLUSTAL 2.1(Larkin, et al., 2007)) and randomly selecting a representative protein to represent homologous zinc proteins (Supplementary Table S2). We also examine the performance of a method for the same protein (with >90% sequence identity) with different ligands. This set (HOMO) has a total of 3 representative proteins of 37 ligand-protein complexes with different ligands and coordination numbers (Table 1). The list of PDB IDs for the HOMO set is shown in Supplementary Table S3.

2.2 The Deviation from Ideal Orientations: RMSD_{OR}.

We employ RMSD_{OR} to measure the orientational root-mean-squared deviations of observed geometries from the ideal geometries of tetrahedral, trigonal bipyramidal and octahedral models. It is defined as the minimum RMSD between the coordinates of unit vectors along the direction of Zn to a coordination atom in the observed structure and those in the ideal model. This definition allows us to focus on the orientational deviations by ignoring atomic distance fluctuations (e.g. the bond length

Zn-S is longer than zinc to other atoms). This is similar to the work by Seebeck, et al who calculated RMSD based on the angles between the vectors from the zinc to a coordination atom (Seebeck, et al., 2008).

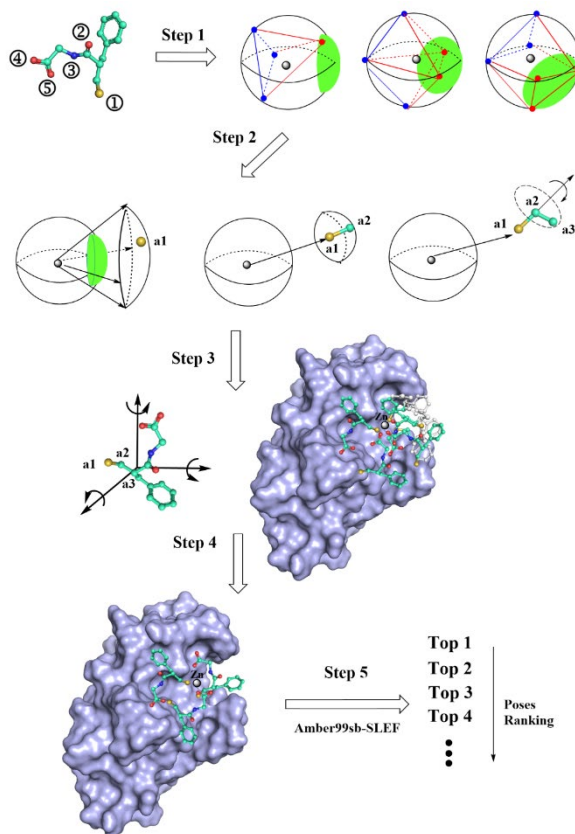


Fig. 2. The flow chart of the ligand-docking protocol of the GM-Dock_{Zn} in this work. Step 1: Identify potential coordinating atom(s) in a ligand and locate their potential positions according to RMSD_{OR} from ideal models (only three examples were shown); Step 2: Determine the initial ligand position; Step 3: Place ligand onto the target protein by rotation and translation to generate various binding poses; Step 4: Exclude poses with steric hindrance from amino-acid residues and the zinc center; Step 5: Rank poses by evaluating the zinc-ligand binding affinity based on the Amber99sb-SLEF force field.

2.3 Zn-ligand Coordination Pose Prediction

The schematic diagram for ligand docking is shown in Fig. 2. Before docking, all ligands and solvent molecules in the query PDB structure are removed. Then, the following procedure is employed. The first step is to locate all current coordination atoms in the query PDB structure according to the distance criterion and align these coordination atoms with all possible ideal models by minimizing RMSD_{OR}. The locations of any missing coordination atoms in an ideal model with RMSD_{OR} less than a cutoff value (0.25Å) are considered as the potential locations of ligand coordination atoms.

According to our statistical analysis of the SM set, the number of missing coordination atoms contributed by a single ligand can be between 1 and 3. Thus, we treat these three cases sequentially. We start with the case that the ligand provides a single atom for coordination. In this case, the possible positions of the ligand’s coordination atom a1 can be generated by using a grid of 0.1Å between a distance of 2 to 2.4 Å (or to 2.8 Å for S atom) and a step of 30° for angles (θ,φ) on the spherical surface centered at the zinc and around the ideal position (Fig 2). Only

those positions with $\text{RMSD}_{\text{OR}} < 0.25 \text{ \AA}$ are kept. Then, all O, N, and S atoms in a ligand are considered as atom a1 in turn as a potential coordination atom. Afterwards, possible coordinates of the nearest-connecting atom a2 are obtained by taking a1 as the center, the distance of a1-a2 as the radius and 30° for angles (θ, ϕ) on the spherical surface for sampling. Only those positions with their distances to zinc $(\mathbf{r}(\text{Zn}, \mathbf{a2})) > 2.5 \text{ \AA}$ are kept for a2. The positions of the third atom a3 are sampled by taking a1-a2 as the rotation axis with a 30-degree interval at a fixed angle a1-a2-a3 and the distance between a2 and a3. Only those positions with $\mathbf{r}(\text{Zn}, \mathbf{a3}) > 2.5 \text{ \AA}$ are kept for a3. Once the positions of a1, a2, and a3 are known, the whole ligand pose can be obtained by building on the xyz-coordinates of the a1, a2, and a3 atoms and the internal coordinate system of the ligand. Next, we examine the case that a ligand provides two atoms as coordination atoms. The possible positions of a1 are obtained as in the case of a single-coordination atom from the ligand. The possible positions of a2 are sampled by a 30-degree interval with the vector Zn-a1 as the rotation axis at the fixed ideal distance between a1 and a2 and the fixed ideal angle of Zn-a1-a2. Only those a1 and a2 positions with $\text{RMSD}_{\text{OR}} < 0.25 \text{ \AA}$ are kept. Once the positions of a1 and a2 are known, the third atomic positions and the entire ligand can be built as described before. Finally, in the case of a ligand providing three atoms as coordination atoms, the first two atoms (a1 and a2) are done as before. Then, the possible positions of a3 are sampled in a 30-degree interval with a1-a2 as the rotation axis and the fixed angle a1-a2-a3 and the distance between a2 and a3. Only those positions with $1.9 \text{ \AA} < \mathbf{r}(\text{Zn}, \mathbf{a3}) < 2.5 \text{ \AA}$ are kept for a3. For all possible three atomic positions as coordination atoms, only those with $\text{RMSD}_{\text{OR}} < 0.25 \text{ \AA}$ from ideal models are kept. Once three atomic positions are known, the conformation of the whole ligand can be obtained as before. To avoid steric clashes, all above-generated ligand poses are removed if any atoms of the ligand are within 2 \AA from any heavy atoms of the protein (cutoff) or 2.5 \AA from zinc.

Here, the threshold of $\text{RMSD}_{\text{OR}} (< 0.25 \text{ \AA})$ and the conformational step-size of ligand (0.1 \AA and 30°) are user-defined parameters. These default values employed herein are recommended to balance the efficiency and accuracy of docking.

2.4 Poses Scoring

To rank the ligand poses obtained above, we employed the Amberff99sb force field (Cornell, et al., 1996; Hornak, et al., 2006) to calculate the interactions between the small molecule ligand and the protein except the interactions associated with the zinc ion. The latter is described by using our previously developed non-bonded Short-Long-Effective-Function (SLEF) model. In this force field (Eqs. (1) and (2)), the total energy function of the SLEF model is the summation of electrostatic and vdW interaction terms (Gong, et al., 2015; Wu, et al., 2011). The vdW interactions between a zinc ion and any other atom are described by the traditional Lennard-Jones potential. The electrostatic interaction term (Eq. (2)), is expressed as the conventional Coulomb energy weighted by the sum of the short-range coefficient c_S and the long-range coefficient c_L , as shown in equation (3) and (4) below.

$$E_{\text{non-bond}}^{\text{ZN}} = E_{\text{es}}(\text{SLEF}) + E_{\text{vdw}} \quad (1)$$

$$E_{\text{es}}(\text{SLEF}) = E_S + E_L = (c_S + c_L) \times E_{\text{es}} = (c_S + c_L) \times \frac{1}{4\pi\epsilon_0} \times \frac{q_i q_j}{r_{ij}} \quad (2)$$

$$c_S = \frac{1}{1 + \alpha \times \frac{(|q_i| + |q_j|)^2}{(|R_i^*| + |R_j^*|)^2} \times \exp(\beta \times r_{ij}^2)} \quad (3)$$

$$c_L = \frac{1}{1 + \frac{(|q_i| + |q_j|)}{q_{\text{Zn}}} \times \exp(1 - \lambda \times r_{ij})} \quad (4)$$

All van der Waals parameters and partial charges for zinc interactions in Eqs (1)-(4) are obtained from the Amber99sb-SLEF force field. α , β , λ and R^* parameters in (3) and (4) optimized by QM/MM force are

$0.11(\text{\AA}^2/e^2)$, $0.81(\text{\AA}^{-2})$, $0.74(\text{\AA}^{-1})$, and $1.36(\text{\AA})$, respectively (Gong, et al., 2015).

2.5 Other Methods

GOLD version 4.1.2 was used for the redock experiments. Glide XP module was from Schrödinger (Schrödinger, LLC: New York, NY, 2015). AutoDock4_{Zn} was downloaded from its official website: autodock.scripps.edu. MpsDock_{Zn} was kindly provided by Dr. Honglin Li.

2.6 Docking Preparation

All apo-protein structures are generated by using the Molecule Operating Environment package ((MOE), 2013) to remove ligands from the native proteins. The hydrogen atoms of proteins are also added and pre-optimized by the MOE software suite. The general amber force field (GAFF) (Wang, et al., 2004) was applied for all small-molecule ligands and their atomic charges were assigned from restrained electrostatic potential (RESP) calculations at the HF/6-31G* theoretical level in Gaussian 09 package (M. J.Frisch, 2009).

2.7 Performance Measure

A ligand pose is evaluated according to the RMSD of the structure superposition between a predicted pose and the native structure based on all heavy atoms of the ligand in the presence of fixed zinc and protein positions. In addition to RMSD, docking performance is also measured by the reproducibility of the correct coordination number (C.N.) and the coordination atom in the native crystal structure because low RMSD conformations may be associated with an incorrect zinc coordination structure. Here, a successful re-docking is defined as $\text{RMSD} < 2.0 \text{ \AA}$ with the correct coordination number and coordination atoms. A RMSD cut-off value of 2.0 \AA was also used previously for evaluation (Bai, et al., 2015; Santos-Martins, et al., 2014).

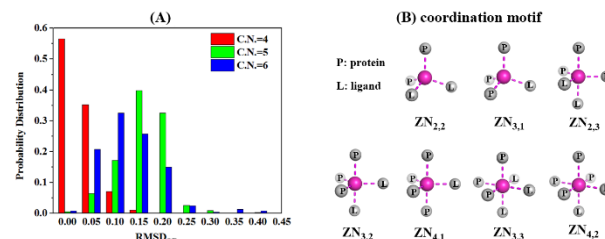


Fig. 3. (A) The distribution of the geometry matching parameter (RMSD_{OR}) generated from the Small-Molecule Set (SM). The majority (99.9% of 4-, 96.4% of 5- and 95.1% of 6-coordination zinc structures) have an RMSD_{OR} value of lower than 0.25 \AA . (B) All possible coordination motifs (seven) found in the PDB structures of the single ligand set (SL), denoted by $\text{Zn}_{P,L}$ with P, L are the number of atoms from proteins and a ligand, respectively.

3 Results

3.1 RMSD_{OR} Distribution

To choose a cut off for the deviation of a ligand-containing coordination system from ideal models, one needs to know the natural fluctuation around the ideal models in zinc proteins complexed with small molecules. Fig. 3A shows the distribution of RMSD_{OR} with 4, 5 and 6 coordination numbers, respectively, in the SM set that has 2256 zinc proteins complexed with one or more small molecules. The results show that over 95% of 2256 zinc coordination systems have an RMSD_{OR} value lower than 0.25 \AA . As a result, RMSD_{OR} of 0.25 \AA is used as a default cut off to remove those structures away from the ideal coordination models. Interestingly, the distribution of RMSD_{OR} for 5-coordination systems is

quite similar to those obtained from QM/MM MD calculations on ligand binding of HDAC (see Fig S1).

3.2 Possible Coordination Motifs

We examined the possible zinc-coordination motifs in the presence of a single ligand. Using the SL set (685 protein-ligand complexes), we found that there are only 7 coordination motifs (Fig 3B), which are annotated by $ZN_{P,L}$ with P, L are the number of atoms from proteins and a ligand, respectively. The occurrences of these motifs are listed in Table 1. A single ligand in the tetrahedral geometry can contribute one (354 structures) or two coordination atoms (10 structures). A single ligand in the trigonal bipyramidal geometry can contribute one on the triangle plane (65 structures), one on and one off the triangle plane (178 structures), and two on and one off the triangle plane (1 structure). A single ligand in the octahedral geometry can contribute two (65 structures) or three (12 structures) coordination atoms.

3.3 Docking Results

We examine the performance of GM-Dock_{Zn} by using a test set of 108 protein-ligand complexes. To make a comparison to other methods on an equal footing as much as we can, we obtain 50 top-ranked poses from all methods compared. Because the default output number of GOLD and AutoDock_{Zn} is less than 10, we modified the minimum-energy cut off so that we can obtain at least 100 docking poses to facilitate comparison. The resulting docking time for GOLD and AutoDock_{Zn} is about 10-times longer than the default.

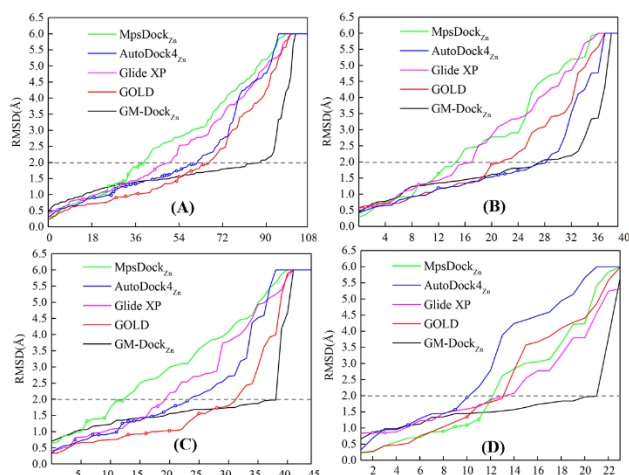


Fig. 4. (A) Method performance given by five docking algorithms as labelled according to the best binding pose (in RMSD, the y-axis) in the top 50 predictions for each target arranged in the increasing order (number of targets in the x-axis) for the whole test set. (B), (C) and (D): Same as (A) but the performance for 4-, 5-, and 6-coordination systems, respectively. An RMSD cut off of 2.0 Å is shown as a dashed line and >6.0 Å for truncation. Close and open circles are true and false-positive predictions, respectively. False-positive predictions are those predictions with $RMSD \leq 2.0$ Å but incorrectly predicted coordination numbers or coordination atoms.

Fig 4A summarizes the performance on the test set by five methods. The performance is measured by plotting the best RMSD value among 50 best poses predicted for each target from small to large. The lower the curve is, the better the performance is. For the first few best predictions, all methods have similar performance in term of RMSD with GOLD having a slight edge. However, all methods except GM-Dock_{Zn} made false-positive predictions that have small RMSD values but with either incorrectly predicted zinc coordination structures, or incorrect coordina-

tion atoms in the ligand, or both (shown as open circles). Moreover, GM-Dock_{Zn} has the highest number of the poses with $RMSD \leq 2\text{Å}$ (88/108 versus 67/108 by the next best (GOLD) and 60/108 after excluding those with incorrectly predicted coordination atoms or numbers, Table 2). That is, GM-Dock_{Zn} makes 47% (88/60) increase in success rate over the second-best GOLD in reproducing the correct binding pose and coordination number around the zinc ion.

Figs 4B-4D compare the performance for 4-, 5-, and 6-coordination systems, respectively. For the 4-coordination system, AutoDock_{4Zn} has essentially the same performance as GM-Dock_{Zn} in term of the number of the poses with $RMSD \leq 2\text{Å}$. Both have a much higher number of correctly predicted ligand binding poses than all other methods. However, AutoDock_{4Zn} made several false-positive predictions. For 5- and 6-coordination systems, GOLD is the second best although it contains false-positive predictions as well. GM-Dock_{Zn} is the only one having the highest number of binding poses with $RMSD \leq 2\text{Å}$ in the absence of any false positives. It is noted that there is a sudden increase RMSD at $RMSD > 2\text{Å}$ for 5- and 6-coordination systems. This is largely due to the finite grids we used in GM-Dock_{Zn} to map possible binding poses. If a near-native binding pose is not found, the next near-native binding pose will have a significantly different structure.

Table 2 summarizes the success rate of locating a correct pose within the top 50 predicted poses (Specific results for each structure by all methods are shown in Supplementary Table S4). This result is obtained after removing false-positive predictions by other methods. GM-Dock_{Zn} makes an absolute improvement over the next best AutoDock_{4Zn} by 4% in the 4-coordination system, 27% over the next best GOLD in the 5-coordination system, and 39% over the next best GOLD in the 6-coordination system. GM-Dock_{Zn} achieves 0% false-positive rates, compared to 10% by GOLD, 26% by AutoDock_{4Zn}, 8% by Glide XP, and 29% by MpsDock_{Zn}.

Table 3 further displays the performance of different methods for 7 possible coordination geometries along with mono-, bi- and tri-chelating ligands to the zinc. Except for some geometries with few cases (2 for $ZN_{2,2}$ and 1 for $ZN_{2,3}$), GM-Dock_{Zn} makes consistent improvement in all other geometries and all chelating possibilities. This indicates the robustness of performance improvement.

The above comparisons were based on the top 50 predictions. Table 4 further compares the success rate for the top 50, top 10, top 5 and top 1 predictions. GM-Dock_{Zn} improves over the second-best GOLD significantly at the top 50 and 10 predictions but is only comparable for the top 5 and worse for the top 1 prediction. This indicates that GM-Dock_{Zn} achieves the best in sampling but the force field employed in this work is not the best for ranking.

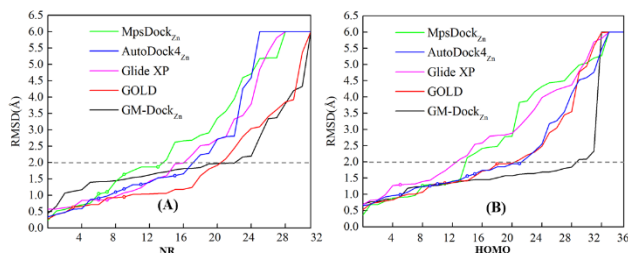
Table 2. The success rates of locating a correct binding pose within the top 50 predicted poses given by various docking methods for the Test set at different coordination numbers.

	All (108)	C.N.=4 (40)	C.N.=5 (45)	C.N.=6 (23)
MpsDock _{Zn}	28%	35%	20%	30%
AutoDock _{4Zn}	43%	66%	31%	30%
Glide XP	44%	43%	42%	48%
GOLD	56%	53%	60%	52%
^a GM-Dock _{Zn}	81%	70%	87%	91%

^aThis work

Table 3. The success rates of locating a correct binding pose within the top 50 predicted poses given by various docking methods for the Test set for different Zn-coordination and ligand-chelating motifs.

	ZN _{2,2}	ZN _{3,1}	ZN _{2,3}	ZN _{3,2}	ZN _{4,1}	ZN _{3,3}	ZN _{4,2}	mono(51)	bi-(50)	tri-(7)
	(2)	(38)	(1)	(31)	(13)	(6)	(17)			
MpsDock_{Zn}	0	37%	100%	19%	15%	33%	29%	31%	22%	43%
AutoDock4_{Zn}	50%	63%	100%	23%	54%	33%	29%	61%	26%	29%
Glide XP	50%	42%	0	42%	46%	33%	53%	43%	46%	29%
GOLD	50%	53%	100%	68%	46%	67%	53%	51%	62%	71%
^aGM-Dock_{Zn}	50%	71%	100%	94%	69%	83%	94%	71%	92%	86%

^aThis work**Fig. 5.** Same as Fig. 4 but for the performance on (A) the non-redundant set (NR) and (B) the homology set (HOMO), respectively.**Table 4.** The success rates of locating a correct binding pose within the top 50, top 10, top 5 and top 1 predicted poses given by various methods for the Test set.

	Top 50	Top 10	Top 5	Top 1
MpsDock_{Zn}	28%	9%	5%	2%
AutoDock4_{Zn}	43%	27%	20%	17%
Glide XP	44%	34%	17%	14%
GOLD	56%	47%	35%	26%
GM-Dock_{Zn}	81%	67%	34%	17%
GM-Dock_{Zn} + GOLD	72%	53%	47%	31%

Table 5. The success rates of locating a correct binding pose within the top 50, top 10, top 5 and top 1 predicted poses given by three methods for the set of 20 apo-proteins in the NR Test set.

	Top 50	Top 10	Top 5	Top 1
GOLD	30%	25%	20%	10%
GM-Dock_{Zn}	45%	30%	20%	10%
GM-Dock_{Zn} + GOLD	50%	30%	20%	15%

Table 6. The success rates of locating a correct binding pose within the top 50, top 10, top 5 and top 1 predicted poses given by three methods for all possible combinations of 452 cross docking results in the HOMO set.

	Top 50	Top 10	Top 5	Top 1
GOLD	61%	46%	36%	13%
GM-Dock_{Zn}	76%	48%	37%	14%
GM-Dock_{Zn} + GOLD	70%	63%	58%	35%

To further improve the usefulness of GM-Dock_{Zn} for docking, we examine the possibility of using GM-Dock_{Zn} for sampling and GOLD for scoring. The results of the combined method, which is labelled as GM-

Dock_{Zn}+GOLD, is shown in Table 4. We found that GM-Dock_{Zn}+GOLD substantially improves over GOLD and GM-Dock_{Zn} in Top 5 (>12%) and Top 1 prediction (>5%) in term of the success rate for sampling near native conformations. The results confirm the power of using a better scoring function in ranking the poses obtained by GM-Dock_{Zn}.

In the above method comparison for zinc-binding proteins (Cinaroglu and Timucin, 2019; Santos-Martins, et al., 2014), homologous proteins are often not excluded because even the same protein may bind different ligands differently in term of their poses, coordination numbers and coordination atoms. Nevertheless, it is necessary to examine the effect of binding to different proteins and binding to the same protein, separately. We have made non-redundant and homolog sets for this purpose (See Methods). As shown in Fig. 5, GM-Dock_{Zn} continues to have the best performance for both NR (Fig. 5A and Supplementary Table S5) and HOMO (Fig. 5B and Supplementary Table S6) sets. However, the improvement of success rate is smaller for the NR set (13% absolute improvement over the next best GOLD, compared to 25% in the whole test set) and larger for the HOMO set (30%). This indicates that NR provides a more realistic estimation of improvement without homology biases. On the other hand, the results on the HOMO dataset indicates that GM-Dock_{Zn} is more capable of handling different ligands docking into the same structure.

The above results were obtained by docking onto holo-structures. To investigate the effect of protein conformational changes on docking performance, we employed 20 proteins from the NR set with apo structures. The list of PDB IDs are shown in Supplementary Table S7. As shown in Table 5, GM-Dock_{Zn} continues to have the best performance for obtaining near native structures within Top 50 and Top 10. Interestingly, GM-Dock_{Zn} and GOLD have comparable performance for apo-docking. GM-Dock_{Zn} + GOLD, further improves over GM-Dock_{Zn} and GOLD in Top1 and Top50.

Another way to examine the effect of conformational transitions is to perform cross docking between different protein-ligand complexes of the same protein. We performed all possible combinations of cross-docking of the structures for the same protein in the HOMO set (listed in Supplementary Table S3). The corresponding results are summarized in Table 6. GM-Dock_{Zn} continues to have a significant improvement in Top 50 over GOLD. It also shows a slightly better performance than GOLD for Top 10, Top 5 and Top 1. Improved Top 1 performance relative to GOLD by GM-Dock_{Zn} suggests that GM-Dock_{Zn} is less affected by small conformational changes. The combination of two methods (GM-Dock_{Zn} + GOLD), although slightly worse than GM-Dock_{Zn} in Top 50, makes a significant improvement in Top 10, Top 5 and Top 1 (15%, 21% and 21% absolute improvement in success rate).

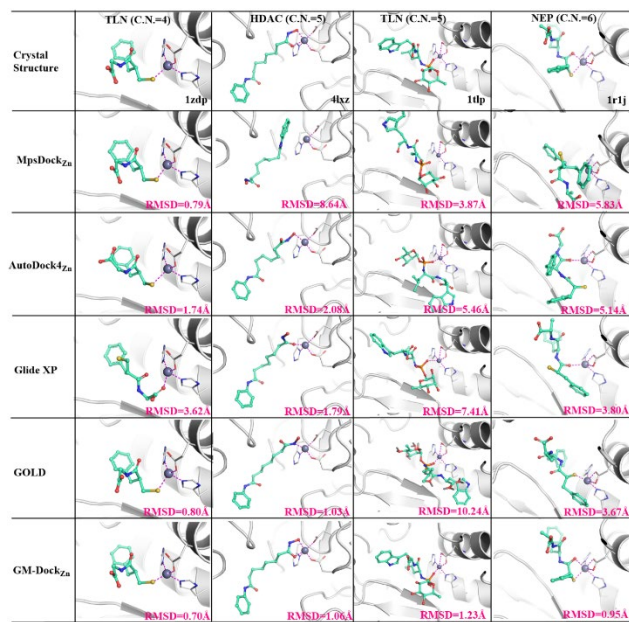


Fig. 6. The illustrative examples of re-docking results for four zinc coordination structures with the coordination number of 4, 5, 5, and 6, respectively. The best pose in the top 50 for each method is shown.

We illustrate the difference between GM-Dock_{Zn} and other methods in more details by using an example for coordination numbers of 4, 5, 5, and 6 (**Fig. 6**). The best poses in the top 50 for six methods are shown. For the coordination of four atoms around zinc, the crystal structure of Thermolysin (TLN, pdb code: 1ZDP) ([Bradner, et al., 2010](#)) was used as an illustration. Zinc is surrounded by three coordinative atoms from two HIS and one GLU residues and one sulfur atom from the ligand (2-mercaptomethyl-3-phenyl-propionyl-glycine). The best pose from Glide XP has a large RMSD of 3.6 Å with an incorrect coordination atom from the ligand (O instead of S). All other methods (GM-Dock_{Zn}, GOLD, AutoDock4_{Zn} and MpsDock_{Zn}) provide a reasonable prediction with the correct coordination atom and an RMSD value of < 2 Å.

For the coordination number of 5, two typical examples are shown with HDAC2 (4LXZ) ([Lauffer, et al., 2013](#)) and TLN (1TLP) ([Tronrud, et al., 1986](#)). In HDAC2, the zinc coordination shell is a square-pyramid geometry, with two ASP and one HIS residues providing three coordinative atoms while the ligand SAHA is bi-chelating to the zinc ion. For this example, GM-Dock_{Zn} is the only one successful to generate a near-native pose. As shown in **Fig 6**, Glide XP, GOLD, AutoDock4_{Zn} failed to produce the bidentate feature of SAHA. That is, they can only predict a mono-dentate coordination mode. MpsDock_{Zn} is unable to rank the coordination between zinc and SAHA within the top 50. For TLN (1TLP), the zinc coordination shell is a trigonal-bipyramidal geometry, made of one bidentate GLU, two HIS residues and a mono-chelating ligand. Similarly, GM-Dock_{Zn} is also the only method that correctly reproduces the native zinc coordination shell in the crystal structure. It should be noted that the RMSD values are reasonable for GOLD and Glide XP in HDAC2 but coordination details are incorrect.

For the 6-atom zinc coordination shell, the Neprilysin (NEP) protein bound with its inhibitor ORI (n-(3-phenyl-2-sulfanylpropanoyl) phenylalanylalanine) is selected as a representative case ([Oefner, et al., 2004](#)). The octahedral coordinative geometry consists of two His and one bidentate Glu residues, as well as a bidentate ligand (ORI). As shown in **Fig. 6**, the coordination number is only 5 for the best poses given by GOLD,

Glide XP, and AutoDock4_{Zn}, with the ligand in a mono-dentate pose instead of bi-dentate. There is no chelation interaction between the ligand and the zinc ion according to MpsDock_{Zn}. GM-Dock_{Zn} is the only method that successfully yields the correct binding mode and coordination motif.

Discussion

In this paper, we have developed a geometry-based docking technique for zinc proteins. In this technique, all potential coordination atoms in a ligand are placed near-ideal locations of 7 discovered ligand coordination motifs by a grid search. The method provides a substantially improved capability over several methods in sampling near-native poses that are not only low in RMSD but also correct in term of coordination numbers and coordination atoms. Many existing methods can yield low RMSD poses but with incorrectly predicted coordination atoms or coordination numbers. The results highlight the importance of using more than RMSD in docking assessment for metalloproteins.

Our systematic investigation of coordination motifs only found 7 possible geometries as shown in **Fig 3B**. That is, not all possible combinations of ligand positions are present. This may be interpreted by the requirement that the interactions between zinc and protein atoms and between zinc and ligand atoms have to be both strong. For example, a coordination motif in the tetrahedral geometry with three coordination atoms from a ligand will lead to only a single atom in the protein to interact with zinc. This interaction will likely be too weak for a protein to retain zinc and thus this coordination motif does not exist. On the other hand, two or three ligand atoms are needed in the octahedral geometry to retain the ligand. Otherwise, the zinc will be overwhelmed by interaction with 5 protein atoms.

GM-Dock_{Zn} makes a grid search on a spherical surface. In principle, we could further improve conformational sampling with finer grids. However, this would be at a significant cost in the computing time. With the current parameter sets for sampling top 50 poses, it takes about 16 CPU hrs for GM-Dock_{Zn} to complete the NR set on a single CPU of a personal computer (Intel Xeon E3-1240V5 Quad core 3.5GHz), compared to about 12 hrs by GOLD, 16 hrs by AutoDock_{Zn}, 20 hrs by Glide XP and 28 hrs by MpsDock_{Zn} on the same single CPU. Nevertheless, we have examined the effect of using different angle grids (10, 15, 20, 30 and 45) on the success rates of locating near-native poses by using the NR set. As shown in the Supplementary Table S8, we confirmed an improvement in the top 50 poses sampled at 20 and 15-degree grids but with 3 and 8 times increases in computing times, respectively. However, a further reduction to 10 degrees did not improve the sampling. This is mainly because more conformations increase the challenge for the scoring function to recognize the near-native structures. This is also reflected from the variation of the best grid for top 50, 10, 5, and 1 (15, 30, 20 and 15 degrees, respectively). The results highlight the critical need for improving the scoring function.

GM-Dock_{Zn} employs the AMBER force field (ff99SB) for protein-ligand interactions and the SLEF force field for the zinc-protein and zinc-ligand interactions. Although the entropic effect and the solvation free energy were not considered, the combination of the force fields has allowed a reasonable selection of the top 50 predictions. To examine the effect of the SLEF force field, we have re-performed scoring of the NR set in its absence. As shown in the supplementary Table S9. We found that including SLEF leads to 7% and 11% absolute increase in the success rate for sampling near-native conformations for top 50 and 10 conformations, respectively, but 8% and 2% decreases for top 5 and top 1

conformations, respectively. This suggests that further improvement of the SLEF force field is required for more accurate detection of near-native conformations.

The current implementation of GM-Dock_{Zn} has been focused on conformational sampling. We expect that GM-Dock_{Zn} could be further extended for other metalloproteins, if there is a sufficient number of known protein structures for characterization of the coordination motifs for other metalloproteins. For example, we can directly use the 6-coordination model for most magnesium metalloproteins but possible motifs with ligand atomic positions would require statistical analysis of their existing structures. Calcium ions, on the other hand, have a representative 8-coordination system and thus an 8-coordination model should be employed by a survey of typical protein complexes to characterize the calcium binding motifs. Once binding motifs with known positions of possible ligand atoms are resolved, it is straightforward to employ the approach developed here.

One weakness of the current implementation is that it does not sample all possible rotatable bonds in ligands. Another weakness is that it is unable to locate the best within the top 5 or top 1 as shown in Table 4. Despite of the weakness, we found that the performance of GM-Dock_{Zn} is more robust when docking onto apo-structures and when cross docking between the structures of the same protein complexed with different ligands (Tables 5 and 6). To make GM-Dock_{Zn} more practically useful, we have developed a combined method GM-Dock_{Zn} + GOLD: using the GOLD scoring function to score the conformations sampled by GM-Dock_{Zn}. Results on docking onto holo-structures, apo-structures and cross docking consistently showed that GM-Dock_{Zn} + GOLD provides a significant improvement in detecting near native poses within top 1 and top 5 by combining near native sampling by GM-Dock_{Zn} with accurate ranking capability of GOLD. We expect a similar superior performance of GM-Dock_{Zn} + GOLD for the structures from homology modeling because the grid-based sampling quality of GM-Dock_{Zn} is less sensitive to small conformational transitions near the metal active site. Including rotatable bonds in GM-Dock_{Zn} is working in progress.

The above usage of GOLD for scoring GM-Dock_{Zn} provides a practical solution to this challenging problem of metalloprotein docking. To go beyond GM-Dock_{Zn} + GOLD, it is necessary to develop a next-generation scoring function with improved characterization of metal-ligand and metal-protein interactions. In addition to traditional empirical and quantum-mechanical-based force fields employed here, machine-learning plays an increasingly important roles in docking scoring (Nguyen, et al., 2020; Nguyen and Wei, 2019). A machine-learning-based scoring function will be likely useful for ligand docking on metalloproteins as more and more structural data become available.

Acknowledgements

We thank the Guangzhou and Shenzhen Supercomputer Center for providing computational source. And we also thanks for the Three Big Constructions-Supercomputing Application Cultivation Projects from SYSU. The day of resubmission is also the 6th birthday of my (Ruibo Wu) son (Zhengyu Joey Wu), who gave me many inspirations on science, herein I want to wish him a happy childhood.

Funding

This work was supported by the National Natural Science Foundation of China (21773313, 21803079), the National Key R&D Program of China (2017YFB0202600), and Shenzhen Science and Technology Program (Grant No. KQTD20170330155106581).

Conflict of Interest: none declared.

References

- Andreini, C. and Bertini, I. A bioinformatics view of zinc enzymes. *J Inorg Biochem* 2012;111:150-156.
- Anzellotti, A.I. and Farrell, N.P. Zinc metalloproteins as medicinal targets. *Chem Soc Rev* 2008;37(8):1629-1651.
- Auld, D.S. Zinc coordination sphere in biochemical zinc sites. *Biomaterials* 2001;14(3-4):271-313.
- Bai, F., et al. An Accurate Metalloprotein-Specific Scoring Function and Molecular Docking Program Devised by a Dynamic Sampling and Iteration Optimization Strategy. *J Chem Inf Model* 2015;55(4):833-847.
- Ballester, P.J. and Mitchell, J.B. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* 2010;26(9):1169-1175.
- Berman, H.M., et al. The Protein Data Bank. *Nucleic Acids Res* 2000;28(1):235-242.
- Boyles, F., Deane, C.M. and Morris, G.M. Learning from the ligand: using ligand-based features to improve binding affinity prediction. *Bioinformatics* 2020;36(3):758-764.
- Bradner, J.E., et al. Chemical genetic strategy identifies histone deacetylase 1 (HDAC1) and HDAC2 as therapeutic targets in sickle cell disease. *Proc Natl Acad Sci U S A* 2010;107(28):12617-12622.
- Burley, S.K., et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res* 2019;47(D1):D464-D474.
- Cang, Z.X. and Wei, G.W. TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions. *Plos Comput Biol* 2017;13(7).
- Cinaroglu, S.S. and Timucin, E. Comparative Assessment of Seven Docking Programs on a Nonredundant Metalloprotein Subset of the PDBbind Refined. *J Chem Inf Model* 2019.
- Cornell, W.D., et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J Am Chem Soc* 1996;118(9):2309-2309.
- Dudev, T. and Lim, C. Effect of carboxylate-binding mode on metal binding/selectivity and function in proteins. *Accounts Chem Res* 2007;40(1):85-93.
- Friesner, R.A., et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004;47(7):1739-1749.
- Friesner, R.A., et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* 2006;49(21):6177-6196.
- Gong, W., Wu, R. and Zhang, Y. Thiol versus hydroxamate as zinc binding group in HDAC inhibition: An ab initio QM/MM molecular dynamics study. *J Comput Chem* 2015;36(30):2228-2235.
- Halgren, T.A., et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 2004;47(7):1750-1759.
- Harding, M.M. Geometry of metal-ligand interactions in proteins. *Acta Crystallogr D Biol Crystallogr* 2001;57(Pt 3):401-411.
- Hornak, V., et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 2006;65(3):712-725.
- Hu, X., Balaz, S. and Shelver, W.H. A practical approach to docking of zinc metalloproteinase inhibitors. *J Mol Graph Model* 2004;22(4):293-307.

GM-Dock_{Zn}: A Geometry Matching based Docking Algorithm for Zinc Proteins

- Johansson-Akhe, I., Mirabello, C. and Wallner, B. InterPep2: Global Peptide-Protein Docking using Interaction Surface Templates. *Bioinformatics* 2020.
- Jones, G., Willett, P. and Glen, R.C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol* 1995;245(1):43-53.
- Jones, G., *et al.* Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997;267(3):727-748.
- Koca, J., *et al.* Coordination number of zinc ions in the phosphotriesterase active site by molecular dynamics and quantum mechanics. *J Comput Chem* 2003;24(3):368-378.
- Kramer, B., Rarey, M. and Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* 1999;37(2):228-241.
- Krezel, A. and Maret, W. The biological inorganic chemistry of zinc ions. *Arch Biochem Biophys* 2016;611:3-19.
- Larkin, M.A., *et al.* Clustal W and clustal X version 2.0. *Bioinformatics* 2007;23(21):2947-2948.
- Lauffer, B.E., *et al.* Histone deacetylase (HDAC) inhibitor kinetic rate constants correlate with cellular histone acetylation but not transcription and cell viability. *J Biol Chem* 2013;288(37):26926-26943.
- Li, Y., *et al.* Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results. *J Chem Inf Model* 2014;54(6):1717-1736.
- Lu, J.N., *et al.* Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. *J Chem Inf Model* 2019;59(11):4540-4549.
- M. J.Frisch, G.W.T., H. B.Schlegel, G. E.Scuseria, M. A.Robb, J. R.Cheeseman, G.Scalmani, V.Barone, B.Mennucci, G. A.Petersson, H.Nakatsuji, M.Caricato, X.Li, H. P.Hratchian, A. F.Izmaylov, J.Bloino, G.Zheng, J. L.Sonnenberg, M.Hada, M.Ehara, K.Toyota, R.Fukuda, J.Hasegawa, M.Ishida, T.Nakajima, Y.Honda, O.Kitao, H.Nakai, T.Vreven, J. A.Montgomery Jr, J. E.Peralta, F.Ogliaro, M.Bearpark, J. J.Heyd, E.Brothers, K. N.Kudin, V. N.Staroverov, R.Kobayashi, J.Normand, K.Raghavachari, A.Rendell, J. C.Burant, S. S.Iyengar, J.Tomasi, M.Cossi, N.Regga, J. M.Millam, M.Klene, J. E.Knox, J. B.Cross, V.Bakken, C.Adamo, J.Jaramillo, R.Gomperts, R. E.Stratmann, O.Yazyev, A. J.Austin, R.Cammi, C.Pomelli, J. W.Ochterski, R. L.Martin, K.Morokuma, V. G.Zakrzewski, G. A.Voth, P.Salvador, J. J.Dannenberg, S.Dapprich, A. D.Daniels, Ö.Farkas, J. B.Foresman, J. V.Ortiz, J.Cioslowski and D. J.Fox. Gaussian 09. Gaussian 2009;Inc., Wallingford CT.
- Maret, W. Zinc coordination environments in proteins determine zinc functions. *J Trace Elem Med Biol* 2005;19(1):7-12.
- Maret, W. Metals on the move: zinc ions in cellular regulation and in the coordination dynamics of zinc proteins. *Biometals* 2011;24(3):411-418.
- Maret, W. New perspectives of zinc coordination environments in proteins. *J Inorg Biochem* 2012;111:110-116.
- Maret, W. and Li, Y. Coordination dynamics of zinc in proteins. *Chem Rev* 2009;109(10):4682-4707.
- Morris, G.M., *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry* 2009;30(16):2785-2791.
- Nguyen, D.D., *et al.* MathDL: mathematical deep learning for D3R Grand Challenge 4. *J Comput Aided Mol Des* 2020;34(2):131-147.
- Nguyen, D.D. and Wei, G.W. AGL-Score: Algebraic Graph Learning Score for Protein-Ligand Binding Scoring, Ranking, Docking, and Screening. *J Chem Inf Model* 2019;59(7):3291-3304.
- Oefner, C., *et al.* Structural analysis of neprilysin with various specific and potent inhibitors. *Acta Crystallogr D Biol Crystallogr* 2004;60(Pt 2):392-396.
- Parkin, G. Synthetic analogues relevant to the structure and function of zinc enzymes. *Chemical Reviews* 2004;104(2):699-767.
- Rarey, M., *et al.* A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology* 1996;261(3):470-489.
- Roe, R.R. and Pang, Y.P. Zinc's exclusive tetrahedral coordination governed by its electronic structure. *J Mol Model* 1999;5(7-8):134-140.
- Santos-Martins, D., *et al.* AutoDock4(Zn): An Improved Auto Dock Force Field for Small-Molecule Docking to Zinc Metalloproteins. *J Chem Inf Model* 2014;54(8):2371-2379.
- Schneider, M., *et al.* Towards accurate high-throughput ligand affinity prediction by exploiting structural ensembles, docking metrics and ligand similarity. *Bioinformatics* 2020;36(1):160-168.
- Seebeck, B., *et al.* Modeling of metal interaction geometries for protein-ligand docking. *Proteins* 2008;71(3):1237-1254.
- Shu, N., Zhou, T. and Hovmoller, S. Prediction of zinc-binding sites in proteins from sequence. *Bioinformatics* 2008;24(6):775-782.
- Su, M., *et al.* Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J Chem Inf Model* 2019;59(2):895-913.
- Tronrud, D.E., Monzinger, A.F. and Matthews, B.W. Crystallographic structural analysis of phosphoramidates as inhibitors and transition-state analogs of thermolysin. *Eur J Biochem* 1986;157(2):261-268.
- Trott, O. and Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;31(2):455-461.
- Velazquez-Libera, J.L., *et al.* LigRMSD: A web server for automatic structure matching and RMSD calculations among identical and similar compounds in protein-ligand docking. *Bioinformatics* 2020.
- Wang, J., *et al.* Development and testing of a general amber force field. *J Comput Chem* 2004;25(9):1157-1174.
- Wang, X., *et al.* Protein Docking Model Evaluation by 3D Deep Convolutional Neural Networks. *Bioinformatics* 2019.
- Wu, R., *et al.* Flexibility of Catalytic Zinc Coordination in Thermolysin and HDAC8: A Born-Oppenheimer ab initio QM/MM Molecular Dynamics Study. *J Chem Theory Comput* 2009;6(1):337.
- Wu, R., *et al.* A Transferable Non-bonded Pairwise Force Field to Model Zinc Interactions in Metalloproteins. *J Chem Theory Comput* 2011;7(2):433-443.
- Zhang, Y. and Sanner, M.F. AutoDock CrankPep: combining folding and docking to predict protein-peptide complexes. *Bioinformatics* 2019;35(24):5121-5127.
- Zhao, W., *et al.* Structure-based de novo prediction of zinc-binding sites in proteins of unknown function. *Bioinformatics* 2011;27(9):1262-1268.