JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

An Accurate Metalloprotein-Specific Scoring Function and Molecular Docking Program Devised by a Dynamic Sampling and Iteration Optimization Strategy

Fang Bai,[†,‡,⊥] Sha Liao,[§,⊥] Junfeng Gu,[†] Hualiang Jiang,[∥] Xicheng Wang,[†] and Honglin Li*[,§]

[†]Department of Engineering Mechanics, State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian, Liaoning 116023, China

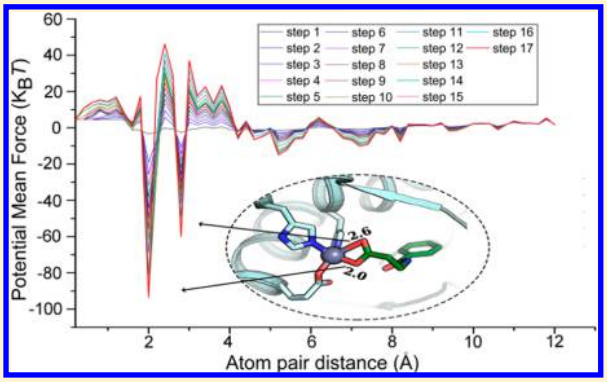
[‡]Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States

[§]State Key Laboratory of Bioreactor Engineering, Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China

[∥]Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

Supporting Information

ABSTRACT: Metalloproteins, particularly zinc metalloproteins, are promising therapeutic targets, and recent efforts have focused on the identification of potent and selective inhibitors of these proteins. However, the ability of current drug discovery and design technologies, such as molecular docking and molecular dynamics simulations, to probe metal−ligand interactions remains limited because of their complicated coordination geometries and rough treatment in current force fields. Herein we introduce a robust, multiobjective optimization algorithm-driven metalloprotein-specific docking program named MpSDock, which runs on a scheme similar to consensus scoring consisting of a force-field-based scoring function and a knowledge-based scoring function. For this purpose, in this study, an effective knowledge-based zinc metalloprotein-specific scoring function based on the inverse Boltzmann law was designed and optimized using a dynamic sampling and iteration optimization strategy. This optimization strategy can dynamically sample and regenerate decoy poses used in each iteration step of refining the scoring function, thus dramatically improving both the effectiveness of the exploration of the binding conformational space and the sensitivity of the ranking of the native binding poses. To validate the zinc metalloprotein-specific scoring function and its special built-in docking program, denoted $MpSDock_{Zn}$, an extensive comparison was performed against six universal, popular docking programs: Glide XP mode, Glide SP mode, Gold, AutoDock, $AutoDock4_{Zn}$, and EADock DSS. The zinc metalloprotein-specific knowledge-based scoring function exhibited prominent performance in accurately describing the geometries and interactions of the coordination bonds between the zinc ions and chelating agents of the ligands. In addition, $MpSDock_{Zn}$ had a competitive ability to sample and identify native binding poses with a higher success rate than the other six docking programs.



INTRODUCTION

Metalloproteins that bind one or more metal ions are involved in various important biological processes, such as signal transcription, enzymatic catalysis, and transcriptional regulation. Zinc metalloproteins constitute approximately 10% of the total human proteome and are particularly attractive as drug targets.[1] To date, numerous inhibitors against angiotensin-converting enzyme (ACE), carbonic anhydrases (CAs), matrix metalloproteinases (MMPs), TNF-α converting enzyme (TACE), histone deacetylases (HDACs), and farnesyltransferase have been reported.[2−8]

A general understanding of zinc metalloproteins can be obtained by systematically analyzing zinc metalloprotein−ligand complexes available in the Protein Data Bank (PDB).[9] Like other metal ions, the zinc ion in zinc metalloproteins is usually coordinated by the nitrogen, oxygen, or sulfur atoms of histidine, glutamic acid, aspartic acid, and cysteine residues, water, or other small molecules. These heteroatoms, which have isolated electron pairs, can form coordination bonds with the zinc ion within certain distances. Groups that bind zinc ion are diverse and include not only the frequently reported carboxylate, sulfonamide, hydroxamate, and phosphonate/phosphate[10] but also hydroxyl, sulfamate, thiol, and imidazole, as reported by Chaskar et al.[11] Moreover, the zinc ion can form various coordination geometries in enzymes, including

Received: October 27, 2014
Published: March 6, 2015

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

bidentate or tetra-, penta- or hexacoordinated geometries, among which the tetracoordinated geometry is dominant.[12] However, the evaluation of metal−ligand interactions using current drug discovery and design technologies, such as molecular docking and molecular dynamics simulation, is hampered by complicated coordination geometries and the limitations of current force field approaches.[13]

Zinc ion−ligand coordination interactions have traditionally been treated using three types of models: bonded, nonbonded, and semibonded models. The bonded model approaches[14−18] treat the interactions between metal ions and coordinating atoms of residues or ligands as "pseudobonds" or "dummy cations" and have been frequently used in molecular dynamics simulations.[19,20] The nonbonded model approaches[21−24] adopt nonbonded interaction terms, such as Lennard-Jones and electrical interaction terms, to simulate the coordination bonds by taking the charges of zinc ions and ligand atoms as constants. By contrast, the semibonded model[25] symmetrically places four cationic dummy atoms to impose orientation requirements for the coordination of the zinc ion during the simulation. However, these methods have certain limitations with respect to the inherent flexibility of zinc ion coordination; hence, they are restricted to broad applications.[26] For example, the bonded model methods cannot represent changes in metal coordination types,[27] the nonbonded model method has limited accuracy in treating low-coordination-number systems, and the semibonded model methods have been successfully applied only in tetracoordinated systems. Recently, Wu et al.[26] developed a transferable pairwise zinc force field based on AMBER ff99SB to describe charge interactions between the zinc ion and other atoms and considered the polarization and charge transfer effects by using a short−long effective function (SLEF). In this function, the first energy term mainly describes the short-range interaction, in which the coordination interaction is expected to be dominant, whereas the second term emphasizes the long-range interaction, which is relatively flat at short range (<4.5 Å) but becomes $1/r$ at long range. This force field has been successfully applied to molecular dynamics simulations of seven zinc enzymes and is described by the following expression (eq 1):

$$E^{\mathrm{es}}_{\mathrm{Zn},j,\mathrm{SLEF}}(r_{\mathrm{Zn},j}) = \frac{1}{4\pi\varepsilon_0}\left\{\frac{q_{\mathrm{Zn}}q_j}{\sqrt{{r_{\mathrm{Zn},j}}^2 + \alpha\frac{{q_j}^2}{(R^*_{\mathrm{Zn}} + R^*_j)}\exp(\beta {r_{\mathrm{Zn},j}}^2)}} + \frac{1}{1+\exp\left[-2\left(\frac{2r_{\mathrm{Zn},j}}{3} - 1.0\right)\right]} \times \frac{q_{\mathrm{Zn}}q_j}{r_{\mathrm{Zn},j}}\right\} \quad (1)$$

where q_{Zn} is the charge of the zinc ion, which has a fixed value of +2.0; q_j is the partial charge of any other atom j of the ligand; R^*_{Zn} and R^*_j represent the van der Waals (vdW) radii of the zinc ion and atom j, respectively; α and β are two positive parameters designed for this function; $r_{\mathrm{Zn},j}$ is the distance between the zinc ion and atom j; and ε_0 is the vacuum permittivity.

In addition to the electrostatic interaction term, Wu et al. modified the conventional Lennard-Jones 12−6 function term by redefining the values of the vdW radius of the zinc ion and corresponding well depth to describe the vdW interaction between the zinc ion and any other atom j in the SLEF force field.

However, effective zinc metalloprotein-specific drug discovery methods, including molecular docking or virtual screening tools, are lacking. Although some general docking programs, such as Gold and Glide, can accommodate coordination bonds by treating them as special hydrogen bonds or covalent bonds,[28−39] these scenarios frequently result in inaccurate calculations.

Molecular docking has become one of the most popular methods in drug discovery and design. Docking approaches generally consist of two crucial processes: sampling of the binding poses of a ligand at the active site of a protein and scoring of the sampled binding poses. Conformational sampling methods have evolved to relative maturity, while scoring functions remain subject to certain limitations, particularly for specific enzyme or protein systems such as metalloproteins.

Generally speaking, scoring functions can be divided into three types: (1) force-field-based scoring functions that use the vdW, electrostatic, and other interaction energy terms in molecular force fields to evaluate the binding affinity of a ligand to a protein; (2) empirical scoring functions that empirically decompose the binding free energy into different energy terms with coefficients determined by regression analysis of the experimental data for the ligand−protein affinities, such that these functions are strongly dependent on the training set and are far from universal for all target proteins; and (3) knowledge-based scoring functions that are derived from statistical analysis of available protein−ligand complexes, such as eq 2:

$$A(r) = -k_{\mathrm{B}}T\ \ln[\rho(r)/\rho^*(r)] \quad (2)$$

where $A(r)$ denotes the atom-pair potential; k_{B} is the Boltzmann constant; T denotes the absolute temperature; and $\rho(r)$ and $\rho^*(r)$ are the atom-pair densities of the protein and ligand at a distance r observed in available protein−ligand structures and in a so-called reference state, respectively. According to the inverse Boltzmann law, atom-pair potentials are constructed from different atom pairs and implicitly include solvation and entropic effects. Many studies have reported the development of knowledge-based scoring functions, illustrating the effectiveness of this strategy.[40−46]

However, the reference state for deriving knowledge-based scoring functions remains a challenge, and no achievable accurate reference state is available.[47] To address this problem, Huang and Zou developed an iteration method to obtain relatively convincing knowledge-based scoring functions by circumventing the calculation of the reference state, and this method has been successfully used to develop scoring functions for protein−protein[41] and general protein−ligand[42] systems. In general, this method iteratively adjusts the atom-pair potentials $A(r)$ until they meet the convergence criterion, that is, they reproduce the experimentally determined atom-pair distribution function in the training set so that the resulting potentials can distinguish native binding conformations from decoy poses of ligands with proteins.[47] Notably, the decoy poses used for iterations in these studies are preprepared before iterating and are unchanged throughout the iteration process. Therefore, the conformational space covered by the limited decoy poses in the iteration process is quite limited and cannot represent the real binding conformational space for searching by docking. In this situation, the sensitivity and accuracy of the obtained atom-pair

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

potentials would be restricted to identifying the active binding pose of a ligand from the entire conformational space. However, as noted by Kang et al.,[48] a good scoring function should not only be sufficiently sensitive to distinguish the active binding pose from decoy poses but also sufficiently efficient to accelerate the convergence of the conformational searching process (docking optimization process). However, previously reported iteration methods have not met these criteria in adjusting the atom-pair potentials for the scoring function.

In consideration of these challenges, a dynamical conformation sampling and iteration optimization strategy is proposed in this study to devise and obtain a zinc metalloprotein-specific knowledge-based scoring function dedicated for our newly designed zinc metalloprotein-specific molecular docking program, MpSDock$_{Zn}$. Thus, we sought to improve the accuracy of binding pose sampling, avoid unexpected computational bias, and resolve limitations in computational accuracy when MpSDock$_{Zn}$ runs the solo zinc metalloprotein-specific knowledge-based scoring function, which cannot describe the repulsive force because of the lack of an effective short-distance repulsive component. Meanwhile, since other pairwise atomic force fields may be inherently unsuitable for describing flexible zinc coordination,[26] the SLEF force field, which adopts the nonstandard function shown in eq 1 to describe charge interactions between the zinc ion and all other atoms in the short-range regime, was integrated into MpSDock$_{Zn}$ as a second scoring function. Thus, MpSDock$_{Zn}$ was designed to run on a scheme similar to consensus scoring consisting of a knowledge-based scoring function and a force-field-based scoring function.

METHODS

1. Design of MpSDockZn. *Multiobjective Optimization Model and Algorithm for MpSDock$_{Zn}$.* To properly address the unit difference between the two scoring functions intended to be integrated into MpSDock$_{Zn}$ and to distinguish the importance of the intramolecular interactions of the binding partners themselves, we designed a multiobjective molecular docking optimization model based on a multiobjective optimization algorithm (nonsorting genetic algorithm II (NSGA II))[49] consisting of five objective functions $f_1(\mathbf{x})$–$f_5(\mathbf{x})$ derived from the two types of zinc metalloprotein-specific scoring functions (i.e., a force-field-based scoring function designed from the SLEF force field and a knowledge-based scoring function developed in this study) and a set of decision variables ($\mathbf{x}$) subject to the conformational space of $\boldsymbol{S}$ as follows:

$$\begin{aligned} &\text{min} && \mathbf{y} = f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x}), f_4(\mathbf{x}), f_5(\mathbf{x})) \\ &\text{subject to} && \boldsymbol{e}(\mathbf{x}) = (e_1(\mathbf{x}), e_2(\mathbf{x}), \cdots, e_k(\mathbf{x})) \le 0 \\ &\text{where} && \mathbf{x} = \{x_1, x_2, \cdots, x_m\}^{\mathrm{T}} \in \boldsymbol{S} \end{aligned}$$

where $\mathbf{x} = \{x_1, x_2, ..., x_m\}^{\mathrm{T}} = \{T_x, T_y, T_z, R_x, R_y, R_z, R_{b1}, ..., R_{bn}\}^{\mathrm{T}}$, in which (T_x, T_y, T_z) and (R_x, R_y, R_z) are the state variables of translation and rotation, respectively, of the entire ligand for the orientation search and $(R_{b1}, ..., R_{bn})$ are the torsion angles of the n rotatable bonds of the ligand for the conformational search. The constraints of these decision variables ($\boldsymbol{e}(\mathbf{x})$s) are as follows:

$$X_{\mathrm{low}} \le T_x \le X_{\mathrm{up}}$$

$$Y_{\mathrm{low}} \le T_y \le Y_{\mathrm{up}}$$

$$Z_{\mathrm{low}} \le T_z \le Z_{\mathrm{up}}$$

$$0 \le R_x, R_y, R_z, R_{b1}, \cdots, R_{bn} \le 2\pi$$

where X_{up} (X_{low}), Y_{up} (Y_{low}), and Z_{up} (Z_{low}) are the upper (lower) bounds of the translational motion of the ligand. In the current version of MpSDock$_{Zn}$, the geometries of the protein and zinc ions are fixed during the docking process.

Two Types of Scoring Functions Designed for MpSDock$_{Zn}$. The force-field-based scoring function is described as follows:

$$E = E_{\mathrm{pro-lig}} + E_{\mathrm{Zn-lig}} + E_{\mathrm{lig}} \tag{3}$$

where $E_{pro-lig}$ and E_{Zn-lig} are the protein–ligand and zinc ion–ligand intermolecular interaction energies, respectively, and E_{lig} is the intramolecular ligand conformational energy. The energy terms used in eq 3 are given by eqs 4–6 and are designed as the objective functions $f_1(\mathbf{x})$–$f_3(\mathbf{x})$ of the multiobjective optimization model of MpSDock$_{Zn}$:

$$f_1(\mathbf{x}) = E_{\mathrm{pro-lig}} = \sum_{k}^{\mathrm{pro}} \sum_{j}^{\mathrm{lig}} E^{\mathrm{vdW}}_{k,j,\mathrm{AMBER}}(r) + \sum_{k}^{\mathrm{pro}} \sum_{j}^{\mathrm{lig}} E^{\mathrm{es}}_{k,j,\mathrm{AMBER}}(r) \tag{4}$$

$$f_2(\mathbf{x}) = E_{\mathrm{Zn-lig}} = \sum_{j}^{\mathrm{lig}} E^{\mathrm{es}}_{\mathrm{Zn},j,\mathrm{SLEF}}(r) + \sum_{j}^{\mathrm{lig}} E^{\mathrm{vdW}}_{\mathrm{Zn},j,\mathrm{SLEF}}(r) \tag{5}$$

$$f_3(\mathbf{x}) = E_{\mathrm{lig}} = \sum_{i}^{\mathrm{lig}} \sum_{l \ne i}^{\mathrm{lig}} E^{\mathrm{vdW}}_{i,l,\mathrm{AMBER}}(r) + \sum_{i}^{\mathrm{lig}} \sum_{l \ne i}^{\mathrm{lig}} E^{\mathrm{es}}_{i,l,\mathrm{AMBER}}(r) \tag{6}$$

where $E^{vdW}_{k,j,AMBER}(r)$ and $E^{es}_{k,j,AMBER}(r)$ in eq 4 are the vdW and electrostatic interactions, respectively, between protein atom k and ligand atom j at a distance r as evaluated using the AMBER ff99SB force field; $E^{es}_{Zn,j,SLEF}(r)$ and $E^{vdW}_{Zn,j,SLEF}(r)$ in eq 5 are the vdW and electrostatic interactions between the zinc ion and ligand atom j at a distance r as defined by the SLEF force field; and $E^{vdW}_{i,l,AMBER}(r)$ and $E^{es}_{i,l,AMBER}(r)$ in eq 6 represent the internal vdW and electrostatic interactions of the nonbonded atom pair i, l of the ligand at a distance r as evaluated using the AMBER ff99SB force field.

The knowledge-based scoring function is given by eq 7:

$$U = U_{\mathrm{pro-lig}} + U_{\mathrm{Zn-lig}} \tag{7}$$

where $U_{pro-lig}$ and U_{Zn-lig} are the protein–ligand and zinc ion–ligand intermolecular interactions, respectively, defined as the sum of all weighted protein–ligand and zinc ion–ligand atom-pair potentials $A_{k,j}(r)$ and $A_{Zn,j}(r)$ at a distance r within the given distance r_{cutoff}, separately described by eqs 8 and 9 and designed as the objective functions $f_4(\mathbf{x})$ and $f_5(\mathbf{x})$:

$$f_4(\mathbf{x}) = U_{\mathrm{pro-lig}} = \sum_{k}^{\mathrm{pro}} \sum_{j}^{\mathrm{lig}} c_{k,j} A_{k,j}(r) \qquad r \le r_{\mathrm{cutoff}} \tag{8}$$

$$f_5(\mathbf{x}) = U_{\mathrm{Zn-lig}} = \sum_{j}^{\mathrm{lig}} c_{\mathrm{Zn},j} A_{\mathrm{Zn},j}(r) \qquad r \le r_{\mathrm{cutoff}} \tag{9}$$

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833–847

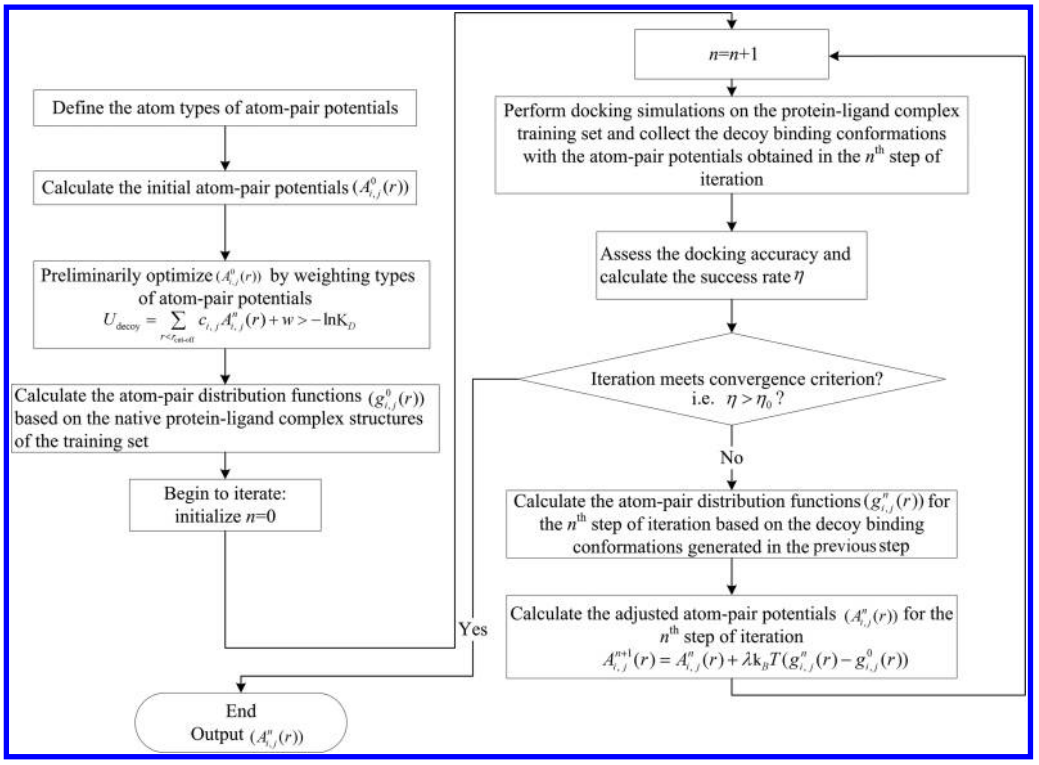


Figure 1. Flowchart to obtain the zinc metalloprotein-specific atom-pair potentials.

where $A_{k,j}(r)$ in eq 8 and $A_{Zn,j}(r)$ in eq 9 represent the atom-pair potential terms between ligand atom j and protein atom k or the zinc ion, respectively, which have been developed in this study on the basis of the conventional inverse Boltzmann relation and steps of iterative adjustment and optimization. In addition, $c_{k,j}$ and $c_{Zn,j}$ are the optimized weights for the atom-pair potentials between ligand atom j and protein atom k or the zinc ion, respectively, which will be described later in this study.

2. Preparation of the Data Sets for Training the Scoring Function of MpSDock$_{Zn}$ and Assessing the Performance of MpSDock$_{Zn}$. All of the available zinc metalloprotein−ligand complexes were collected from the PDB,[34] and only those with relatively high resolution complex structures (resolution less than 3.0 Å) were retained as the training and test sets. The accuracy of the knowledge-based scoring function relies heavily on the quality and quantity of the experimental structures in the training set. Strictly speaking, structures with a resolution of approximately 3.0 Å cannot be considered as "high-resolution" structures. However, to accumulate sufficient data for statistics, the resolution threshold was set as 3.0 Å in the current study. A more reasonable potential could be obtained in the future when high-quality X-ray structures are sufficiently abundant. A total of 540 complexes were obtained and randomly separated into a training set for calculating the initial potentials (434 complexes; PDB codes listed in Table S1 in the Supporting Information) and a test set for validating the scoring function (106 complexes; PDB codes listed in Table S2 in the Supporting Information). Additionally, the reported experimental binding affinities of these complexes (142 complexes; PDB codes listed in Table S3 in the Supporting Information), which were used to weight the different types of potentials between the protein or zinc ion and the ligand, were obtained from other publicly available databases, such as PDBBind,[50,51] PDTD,[52] AffinDB,[53] and BindingDB.[54]

3. Obtaining the Zinc Metalloprotein-Specific Knowledge-Based Scoring Function. To obtain and optimize the zinc metalloprotein-specific knowledge-based scoring function, we designed a workflow that consists of several calculation and optimization steps, as shown in Figure 1.

Atom-Type Definitions for the Knowledge-Based Scoring Function. The atom-type definitions for our zinc metalloprotein-specific knowledge-based scoring function were derived from KScore,[55] a potential of mean force (PMF) scoring function previously developed by our group. The metal atom type (MET) in KScore was specified as a zinc ion in this study.

Calculation of the Initial Atom-Pair Potentials. Similar to KScore and other knowledge-based scoring functions, both the protein−ligand atom-pair potentials $A^0_{k,j}(r)$ and the zinc ion−ligand atom-pair potentials $A^0_{Zn,j}(r)$ were calculated as follows (for clarity, both $A^0_{k,j}(r)$ and $A^0_{Zn,j}(r)$ are represented by the united form $A^0_{i,j}(r)$):

$$A^0_{i,j}(r) = -k_B T \ln\left[f^j_{\text{vol-corr}}(r)\frac{\rho^0_{i,j,\text{seg}}(r)}{\rho^0_{i,j,\text{bulk}}}\right] \quad (10)$$

where $f^j_{\text{vol-corr}}(r)$ is the ligand volume correction factor, $\rho^0_{i,j,\text{seg}}(r)$ is the number density of atom pair (i,j) that occurs in a spherical shell between r and $r + \Delta r$, and $\rho^0_{i,j,\text{bulk}}$ is the number density of atom pair (i,j) that occurs in a reference sphere with radius R. In this study, we set R = 12.0 Å and Δr = 0.2 Å. In detail, $\rho^0_{i,j,\text{seg}}(r)$ and $\rho^0_{i,j,\text{bulk}}$ can be defined as follows:

$$\rho^0_{i,j,\text{seg}}(r) = \sum_m \frac{n^m_{i,j}(r)}{4\pi r^2 \Delta r} \quad (11)$$

$$\rho^0_{i,j,\text{bulk}} = \sum_m \frac{n^m_{i,j,\text{bulk}}}{4\pi R^3/3} \quad (12)$$

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

where $n^m_{i,j}(r)$ and $n^m_{i,j,\mathrm{bulk}}(r)$ are the numbers of atom pairs of types i and j in the spherical shell and reference sphere, respectively, for the mth native complex. Importantly, if less than 400 atom pairs occurred in a 12.0 Å radius, statistical significance was not reached, and the statistical count was discarded from the atom-pair potential calculation. Additionally, to further ensure statistical significance and avoid missing atom-pair potentials between crucial but infrequent atoms and metal ions, ligand atom types that coordinated with zinc ions were combined according to their hybridization types. For example, sp^2-hybridized nitrogen (N.2), aromatic nitrogen (N.ar), and amide nitrogen (N.am) were merged into N.2 to represent all types of sp^2-hybridized nitrogens, and sp^3-hybridized nitrogen (N.3), trigonal-planar nitrogen (N.pl3), and positively charged sp^3-hybridized nitrogen (N.4) were merged into N.3 to represent all types of sp^3-hybridized nitrogens. Thus, all of the nitrogen atoms were classified into three types (N.1, N.2, and N.3) according to their hybridization type.

Here the ligand volume correction factor was introduced to ignore the ligand−ligand interactions and treat desolvation and entropic effects implicitly.[56] The relevant formulas (eqs 13−15) can be written as follows:

$$\hat{f}^{j}_{\text{vol-corr}}(r) = \frac{\hat{\rho}^{0}_{i,j,\mathrm{bulk}}\rho^{0}_{i,j}(r)}{\hat{\rho}^{0}_{i,j}(r)\rho^{0}_{i,j,\mathrm{bulk}}} \tag{13}$$

$$\hat{\rho}^{0}_{i,j}(r) = \rho^{0}_{i,j}(r)\frac{\rho^{0}_{k,j}(r)}{\rho^{0}_{k,j}(r) + \rho^{0}_{l,j}(r)} \tag{14}$$

$$\hat{\rho}^{0}_{i,j,\mathrm{bulk}}(r) = \rho^{0}_{i,j,\mathrm{bulk}}\frac{\rho^{0}_{k,j,\mathrm{bulk}}}{\rho^{0}_{k,j,\mathrm{bulk}} + \rho^{0}_{l,j,\mathrm{bulk}}} \tag{15}$$

where $\rho^0_{k,j}(r)$ and $\rho^0_{k,j,\mathrm{bulk}}$ represent the number densities of protein atom k around ligand atom j in a spherical shell ranging from r to $r + \Delta r$ and a reference sphere with radius R, respectively. Likewise, $\rho^0_{l,j}(r)$ and $\rho^0_{l,j,\mathrm{bulk}}$ represent the number density of ligand atom l around ligand atom j. Furthermore, to smooth the potential curves, the sphere with radius R was separated into N_{seg} spherical shells (segments), where $N_{\mathrm{seg}} = R/\Delta r$, and $\hat{f}^{j}_{\text{vol-corr}}(r)$ was further delaminated according to this set of spherical shells.[40] The ligand volume correction factor was then calculated as follows:

$$f^{j}_{\text{vol-corr}}(r) = \begin{cases} \hat{f}^{j}_{\text{vol-corr}}(r) & s \le m \text{ or } s \ge N_{\mathrm{seg}} - m \\ \dfrac{1}{2m+1}\displaystyle\sum_{s-m}^{s+m}\hat{f}^{j}_{\text{vol-corr}}(r) & m < s < N_{\mathrm{seg}} - m \end{cases} \tag{16}$$

where r lies in the sth shell (s = 1, 2, ..., N_{seg}) and m can be adjusted on the basis of the size of the training set or required precision (m was set to 8 in this study).

Inequality-Constrained Optimization Model for the Initial Potentials. To characterize different types of atom-pair potentials and highlight the potentials of zinc−ligand atom pairs, the initial atom-pair potentials were optimized by weighting types of potentials on the basis of the correlation between the predicted and experimental binding energies of ligands to zinc metalloproteins, as shown in eq 17:

$$U = \sum c_{i,j}A_{i,j}(r) + w \tag{17}$$

where U is the total energy of a protein−ligand interaction calculated using the atom-pair potentials; i indicates a protein atom type or the zinc ion; j represents a ligand atom type; $c_{i,j}$ is the optimized weight coefficient for the potential of specific atom pair (i, j); and w is the designated weight. All of the atom pairs except the metal−ligand atom pair were integrated into one unit with a common weight representing all normal nonbonded atom pairs to highlight the coordination bond strength.

On the basis of the assumption that any decoy binding pose of a protein−ligand complex should not have a binding energy less than its native binding conformation, an inequality-constrained optimization model to obtain the weights of the types of atom-pair potentials was developed, as indicated in eq 18:

$$U_{\mathrm{decoy}} > -\ln K_{\mathrm{D}} \tag{18}$$

where K_{D} is the experimental dissociation constant and U_{decoy} generally represents the binding energy of a decoy conformation calculated using eq 17. To construct the optimization model, decoy poses for 142 protein−ligand complexes with available experimental binding affinity data, including K_{D}, inhibition constant K_i, and binding free energies, were generated using $\mathrm{MpSDock}_{\mathrm{Zn}}$ to perform docking simulations on each protein−ligand complex structure after supplying the initial potentials. The binding pocket of the target protein for docking was defined as the active site, including the area that covered protein residues within a sphere of radius 20.0 Å centered on the center of mass of the ligand in the complex. For each optimization run in the docking simulation, 600 generations were performed on a population of 3000 individuals, and the operator weights for crossover and mutation were set to 0.9 and 0.2, respectively. Through sampling, a series of binding conformations (nondominated solutions) for each ligand in complexes of the set were obtained, and only those binding conformations whose root-mean-square deviations (RMSDs) with respect to the corresponding X-ray conformations in the complex structures were less than 3.0 Å were retained as decoy poses. A total of 1164 decoy poses were ultimately collected, and 1164 corresponding inequalities based on eqs 17 and 18 were then established.

We used a nonlinear optimization algorithm named BPMPD, implemented in the NEOS server,[57] to solve this inequality-constrained optimization problem. The set of weight coefficients 3.75, 1.49, 2.46, 1.60, 2.01, 2.99, 0.94, 1.89, 0.12, 0.11, and 123 were obtained for the atom pairs Zn−C.3 (C.3 = sp^3-hybridized carbon atom), Zn−C.2 (C.2 = sp^2-hybridized carbon atom), Zn−C.ar (C.ar = aromatic carbon atom), Zn−N.3, Zn−N.2, Zn−O.3 (O.3 = sp^3-hybridized oxygen atom), Zn−O.2 (O.2 = sp^2-hybridized oxygen atom), Zn−O.co2 (O.co2 = oxygen atom in a carboxylate or phosphate group), Zn−H, other protein−ligand atoms, and a constant, respectively.

Improvement and Refinement of the Potentials by the Dynamic Sampling and Iteration Optimization Strategy. Generally speaking, the scores calculated using the initial atom-pair potentials were unlikely to generate sufficiently accurate predictions. Therefore, we refined the above optimized potentials using our iteration optimization strategy until they could correctly discriminate native binding poses (defined as poses whose RMSDs with respect to the corresponding X-ray binding poses were less than 2.0 Å) from the decoy binding

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

conformations. In contrast to the conventional iteration strategy,[42] in our refining iteration process, the decoy poses used in each iteration run were dynamically generated and updated by MpSDock$_{Zn}$ with synchronous updates to the knowledge-based scoring function. The same set of protein−ligand complexes used previously to obtain the initial potentials was still used in this iteration optimization process. For clarity, we use a superscript n on the following variables to indicate the iteration step. Iteration began with a set of initial atom-pair potentials $A^0_{i,j}(r)$ from eq 10 but included weights derived through restrictive constraint optimization. That is, every binding pose, including each orientation and conformation of ligand to protein regenerated by MpSDock$_{Zn}$ at each iteration step, was evaluated with the adjusted potentials from the previous step, weighted as shown in eq 19 with the nth-iteration potentials $A^n_{i,j}(r)$:

$$U = \sum_{r<r_{\text{cutoff}}} c_{i,j}A^n_{i,j}(r) + w \tag{19}$$

If the best-scoring binding pose of each complex (evaluated only on the basis of the knowledge-based scoring function), called the "Score Best" binding pose, is close to the X-ray binding pose in the crystal structure of the complex (i.e., if the RMSD between the best-scoring binding pose and its X-ray binding pose is less than 2.0 Å), then the binding pose of the complex can be successfully predicted using the current potentials. In this work, the convergence parameter η (eq 20) represents the success rate for the identification of native binding poses:

$$\eta = \frac{1}{M}\sum_{m}^{\text{RMSD}_m<2} l \tag{20}$$

where M is the number of protein complexes in the training set; RMSD$_m$ is the root-mean-square deviation between the best-scoring binding pose and the X-ray binding pose in the crystal structure for the mth complex in the training set; and l is a Boolean parameter that is equal to 1 if the complex is successfully predicted (i.e., if RMSD$_m$ < 2 Å) and 0 otherwise. The convergence criterion was set according to the following equation:

$$\eta > \eta_0$$

in which the convergence threshold η_0 was set to 99%. If the iteration satisfied the convergence criterion, the iteration procedure was terminated, and the refinement was regarded as a success. The current potentials were then exported for use in our knowledge-based scoring function. Otherwise, the current atom-pair distribution functions $g^n_{i,j}(r)$ of the decoy poses were calculated using the Boltzmann-weighted averaging method as follows:

$$g^n_{i,j}(r) = \hat{f}^j_{\text{vol-corr}}(r)\frac{\rho^n_{i,j}(r)}{\rho^n_{i,j,\text{bulk}}} = \hat{f}^j_{\text{vol-corr}}\frac{\sum_{m,l}\frac{n^{m,l}_{i,j}(r)\exp(-U_{m,l}/k_BT)}{4\pi r^2\Delta r}}{\sum_{m,l}\frac{N^{m,l}_{i,j}\exp(-U_{m,l}/k_BT)}{4\pi R^3/3}} \tag{21}$$

where $\rho^n_{i,j}$ and $\rho^n_{i,j,\text{bulk}}$ are the number densities of atom pair (i,j) occurring in a spherical shell of radius from $r - \Delta r/2$ to $r + \Delta r/2$ and a reference sphere with radius R at the nth iteration step, respectively; m is the number of protein−ligand complexes; l is the number of ligand binding conformations; and $n^{m,l}_{i,j}(r)$ and $N^{m,l}_{i,j}$ represent the numbers of atom-pair (i,j) in the spherical shell and the reference sphere for the lth decoy binding conformation of the mth complex, respectively.

Here, to correct the trial potentials, the current potentials $A^n_{i,j}(r)$ were modified by adding a correction term to obtain improved potentials $A^{n+1}_{i,j}(r)$, and the iteration step n was increased by 1 to yield the following:

$$A^{n+1}_{i,j}(r) = A^n_{i,j}(r) + \lambda k_BT(g^n_{i,j}(r) - g^0_{i,j}(r)) \tag{22}$$

where λ is a parameter to control the speed of convergence that was set to 0.5[58] and $g^0_{i,j}(r)$ is the experimentally observed atom-pair distribution function for the X-ray binding poses, which was calculated in a similar manner as $g^n_{i,j}(r)$:

$$g^0_{i,j}(r) = \hat{f}^j_{\text{vol-corr}}(r)\frac{\rho^0_{i,j}(r)}{\rho^0_{i,j,\text{bulk}}} = \hat{f}^j_{\text{vol-corr}}\frac{\sum_m \frac{n^m_{i,j}(r)}{4\pi r^2\Delta r}}{\sum_m \frac{N^m_{i,j}}{4\pi R^3/3}} \tag{23}$$

where $n^m_{i,j}(r)$ and $N^m_{i,j}$ are the numbers of atom-pair i, j in the spherical shell and the reference sphere for the mth complex, respectively.

For the next iteration step, an ensemble of decoy poses was regenerated, and their energy scores were calculated using the improved potentials $A^{n+1}_{i,j}(r)$. Thus, the iteration terminated when the binding poses predicted for the training set by using the potentials obtained in the nth iteration step were close to the native ones.

4. Validation of MpSDock$_{Zn}$. To validate the docking accuracy of MpSDock$_{Zn}$, we compared its performance with six other popular docking programs: AutoDock[59] (version 4.2), AutoDock4$_{Zn}$,[60] Glide SP (version 4),[61] Glide XP (version 4),[62] Gold (version 4.1.2),[63] and EADock DSS.[11] In addition, to specifically evaluate the zinc metalloprotein-specific knowledge-based scoring function developed for MpSDock$_{Zn}$, the other two reported general knowledge-based scoring functions (KScore and PMF04[64,65]) were separately considered as substitutes for the knowledge-based scoring function of MpSDock$_{zn}$ and used to run MpSDock as MpSDock$_{KScore}$ and MpSDock$_{PMF04}$, respectively. The results were compared using the same parameter settings as for MpSDock$_{Zn}$.

AutoDock uses a force-field-based empirical free energy scoring function. Because zinc parameters were not available in AutoDock, we referred to the work by Hu and Shelver[66] (r = 0.87 Å, ε = 0.35 kcal/mol, and a charge of +0.95e) to set the optimized zinc parameters. The Lamarckian genetic algorithm (LGA) was used as the conformational search algorithm. The binding pocket, defined as a three-dimensional grid with dimensions of 70 × 70 × 70 points along the x, y, and z axes, was centered on the ligand in the experimental complex with a grid spacing of 0.375 Å. A population with a size of 150 and a maximum number of 27000 LGA generations was used for 30 independent searching and optimization runs. For the other options, the default values were retained, and the top 30 ranked binding poses for each ligand were reserved.

Additionally, AutoDock4$_{Zn}$,[60] a recently reported improved AutoDock force field for small-molecule docking to zinc metalloproteins, was also used as a docking approach. In the AutoDock4$_{Zn}$ computations, the charge of the zinc ion was assigned as zero, and a pseudoatom TZ was added to mediate the attractive component of the force field using a Python script obtained from the author of AutoDock4$_{Zn}$. The other

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

parameters were identical to those used for the AutoDock computation above.

Glide is another empirical scoring function-based docking program that can define metal constraints if a particular metal−ligand interaction in the active site must be considered. Two docking algorithms, the standard precision (SP) model and the extra precision (XP) model, are available in Glide, and both were used in this study. The settings for the binding pocket, including its size and center, were identical to those used in $MpSDock_{Zn}$. Default settings were used for the other parameters, and the top 30 ranked binding poses for each ligand were saved.

Gold, which uses a genetic algorithm to explore ligand binding poses, was also used in this study. It treats coordination bonds as special hydrogen bonds and can automatically recognize the coordination geometry of a particular metal atom on the basis of the environment of the protein binding site. The settings for the active site were similar to those used in the above methods. For each genetic algorithm run, a population size of 100 and a maximum number of 100000 generations were used for 30 independent searching and optimization runs. In addition, Gold offers four fitness functions, and the Astex statistical potential (ASP) was chosen in this study because of its superior performance.[67] The "early termination" option was deselected, and the other options were set to their default values. The top 30 ranked solutions for each ligand were saved.

EADock DSS, a docking approach based on the CHARMM22 force field[68] that has been released as a public-accessible Web server named SwissDock,[69] can be used to calculate the interactions of ligands with zinc metalloproteins. Moreover, its computational accuracy can be significantly improved by running it with the hybrid quantum mechanical/molecular (QM/MM) scoring function developed by Röhrig's research group.[11] However, this scoring function is still not integrated into the publicly accessible EADock DSS (SwissDock server). Thus, the general EADock DSS was used as another docking approach for our comparison study, and the "accurate" model available on the server was selected. The center and size of the binding pocket were set to the same values as in $MpSDock_{Zn}$, and the prepared protein and ligand used in the other docking methods were uploaded and reprepared automatically by the server itself. The top 30 ranked binding poses were selected to perform the comparison study.

RESULTS AND DISCUSSION

1. Extraction of Atom-Pair Potentials. General force fields or scoring functions may not be applicable to the types of coordination bonds and various coordination geometries in metalloproteins. To address this problem and adequately evaluate the different types of atom-pair interactions, such as coordination and hydrogen bonds and electrostatic interactions, for a ligand binding to its receptor, i.e., zinc metalloprotein, the initial atom-pair potentials derived from the original complexes were optimized by weighting types of atom-pair potentials on the basis of the correlation between the predicted and experimental binding energies of the ligand to zinc metalloproteins. The obtained weights of the atom-pair potentials for zinc and other ligand atoms are larger than those of other protein−ligand atom pairs, indicating that the zinc ion greatly influences the binding affinities of metalloproteins.

In addition to obtaining the weights of the types of atom-pair potentials, a dynamic sampling and iteration optimization strategy was designed to further refine the atom-pair potentials. In each iteration step, decoy poses for all of the complexes in the training set were regenerated and updated using $MpSDock_{Zn}$ according to the corrected potentials obtained from the previous iteration step. This strategy set an even higher demand on the docking method. In the first 16 iteration steps, the ligands were specified as completely flexible; however, in the 17th iteration step, the binding poses of some ligands in those complexes that frequently failed to be retrieved (reproduced) in the previous 16 iteration steps were constrained, and only six degrees of spatial freedom (variables) for determining the binding orientations of ligands were allowed. We made this modification because the success rate of "Score Best" increased slightly (from 68.4% to 71.5%; Figure 2) and

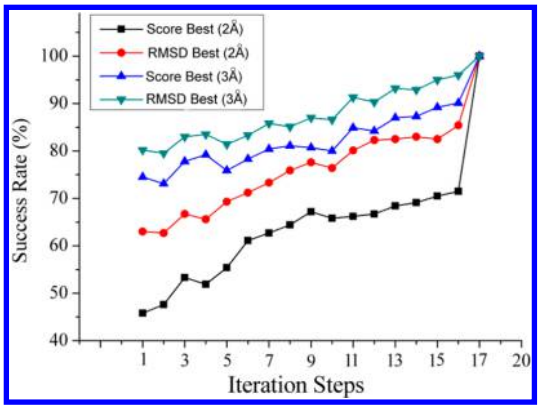


Figure 2. Progress of the convergence of the iteration optimization process.

because the success rate of "RMSD Best" (the docked binding pose whose structure is closest to the X-ray binding pose, i.e., the RMSD between it and the X-ray binding pose is the smallest compared with the other sampled binding poses) decreased slightly after the 13th iteration step (83.0% in the 14th iteration step but 82.5% in the 15th iteration step; Figure 2). Therefore, it could be artificially considered converged, and the iteration process was terminated at the 17th step, although the success rate did not meet our predefined convergence threshold. Analysis of the ligands that failed to be reproduced revealed that they were extraordinarily large and flexible, as illustrated in Figure S1 in the Supporting Information; therefore, they were difficult to retrieve. By adoption of this modified strategy in the 17th iteration step, a 100% success rate was achieved, demonstrating that our optimized atom-pair potentials can successfully identify the active binding poses for most ligands to their binding proteins, at least guaranteeing the correct prediction of the binding orientations of those large and flexible ligands to their binding targets. Finally, the Savitzky−Golay smoothing algorithm[70] was adopted to remove the noise points of the adjusted atom-pair potentials obtained in the 17th iteration step.

Previously reported iteration processes nearly converge at approximately the 10th iteration step.[42,46] However, in this study, the process did not achieve satisfactory convergence before the 16th iteration step. Thus, compared with the iteration methods reported in previous studies,[42,46] our dynamic sampling and iteration optimization process was much more computationally time-consuming and more difficult

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

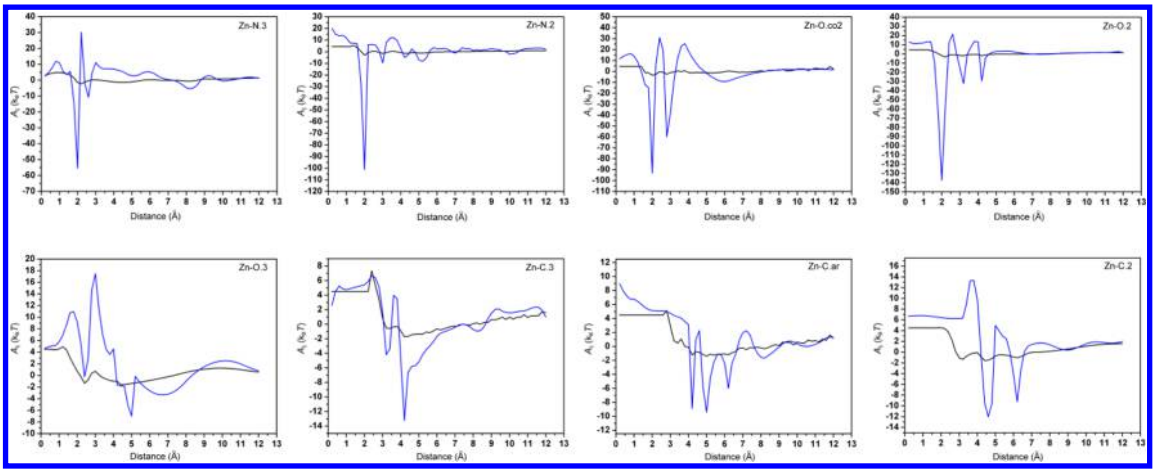


Figure 3. Eight representative zinc ion-involved atom-pair potentials of our zinc metalloprotein-specific scoring function built into $MpSDock_{Zn}$. The first atom type in the label of the upper right corner of each subgraph refers to the zinc ion, and the second atom type is the ligand atom. Black lines represent the primary potentials without optimization, and the blue lines provide the eventual potentials after the optimization. The atom-type definitions for $MpSDock_{Zn}$, excluding the zinc ion, originated from KScore. Zn represents zinc ion; N.2 and N.3 represent sp^2- and sp^3-hybrydized nitrogens, respectively; O.2, O.3, and O.co2 represent an sp^2-hybridized oxygen, an sp^3-hybridized oxygen, and an oxygen in a carboxylate or phosphate group, respectively; and C.2, C.3, and C.ar represent sp^2-hybridized, sp^3-hybridized, and aromatic carbons, respectively. A detailed atom-type definition was derived from ref 55.

to converge. However, our process is more reasonable for obtaining the atom-pair potentials. Theoretically, a correct scoring function should not only distinguish native binding poses from decoy poses but also guide the conformational search algorithm to explore the conformational and orientation space to identify native binding poses. However, the iteration processes of previous studies only focused on optimizing atom-pair potentials to distinguish X-ray binding poses from one preprepared ensemble of decoy poses that covered only a limited binding conformational space and remained unchanged throughout the iteration process. This methodology greatly reduces the computational complexity and accelerates the convergence of the iteration process but does not consider the ability of the scoring function to perform conformational searching. By contrast, the dynamic sampling and iteration optimization strategy used in our study regenerated binding poses for every complex in the training set to remake the ensemble of decoy poses at each iteration step according to the potentials adjusted in the last step. Thus, our iteration process can simultaneously optimize the robustness of guiding conformational searching and the reliability of identifying native binding poses for atom-pair potentials.

2. Properties of the Optimized Atom-Pair Potentials. To observe the change in atom-pair potentials due to the iteration optimization, eight representative atom pairs and their refined potentials were selected for comparison with their primary potentials, as illustrated in Figure 3. The four pairs of atom-pair potentials in the top panel of Figure 3 are for atom pairs that can form coordination bonds. Compared with the primary potentials (black lines), the iteration optimization strategy resulted in extraordinarily sharp potentials for the atom pairs that can form coordination bonds at the corresponding distance, approximately 2.0 Å (blue lines). In general, the potential wells of Zn−N.2 (where Zn represents the zinc ion and N.2 represents an sp^2-hybridized nitrogen atom), Zn−O.co2 (where O.co2 represents an oxygen in a carboxylate or phosphate group), and Zn−O.2 (where O.2 represents an sp^2-hybridized oxygen atom) are obviously deeper than the potential well of Zn−N.3 (where N.3 represents an sp^3-hybridized nitrogen atom). That is, the coordination bonds of Zn−O.2, Zn−O.co2, and Zn−N.2 are much more energetically favorable. In particular, the potential of Zn−O.co2 presents bidentate potential wells with varying depths at distances of approximately 2.0 and 2.8 Å. This result confirms the observation that the geometry of the carboxylate anions is particularly suitable for the formation of complexes in which the carboxylate ligand is coordinated to the metal ion in a bridging mode (Figure 4A) or even a bidentate asymmetrical coordination mode (Figure 4B) or a monodentate coordination mode (Figure 4C). By contrast, the geometry is not well suited for a bidentate symmetrical coordination mode (Figure 4D).[71] For the potential of a Zn−O.2 atom pair, the two relatively minor, equivalently deep dentate potential wells located at distances of approximately 3.0 and 4.0 Å likely reflect the

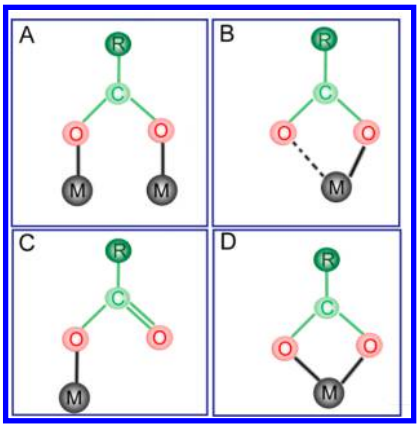


Figure 4. Coordination bonding modes of a metal ion for a carboxylate ligand:[71] (A) bridging coordination mode; (B) bidentate asymmetrical coordination mode; (C) monodentate coordination mode; (D) bidentate symmetrical coordination mode. The gray spheres labeled as "M" represent metal ions, the light-red spheres labeled as "O" represent the oxygen atoms of the carboxylate group, the light-green spheres labeled as "C" represent the carbon atoms that connect with oxygen atoms, and the green spheres labeled as "R" generally refer to the other parts of the molecule.

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

binding modes of sulfonamide ligands to their zinc metalloproteins; the experimental binding mode of dorzolamide to the human carbonic anhydrase II (PDB code 1CIL) presents an example (Figure 5). The nitrogen atom of the sulfonamide group of dorzolamide forms a coordination bond with a length of 1.95 Å with the zinc ion, and its two oxygen atoms are located 2.94 and 4.10 Å from the zinc ion and form hydrogen bonds with the residues of the members of the coordination bond network (H119 and H94) and T199 separately. Herein the binding mode in the complex of 1CIL validates the potential characteristic of Zn−O.2 and is consistent with the preference of sulfonamides to favor the monodentate coordination mode.[10] These results indicate that the potentials obtained in this work reasonably reflect the energies and geometries of coordination bonds.

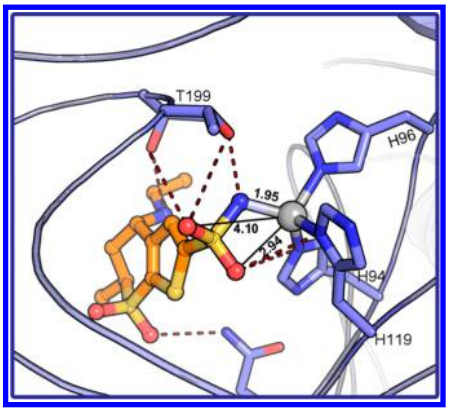


Figure 5. Experimental binding mode of dorzolamide to the human carbonic anhydrase II (PDB code 1CIL). The lilac cartoon mode represents carbonic anhydrase II, lilac stick modes represent the focused residues in the active site, the gray sphere represents the zinc ion, black lines label the distances between the zinc ion and specific atoms of dorzolamide, the orange ball-and-stick mode is dorzolamide, and the red dashed lines represent the hydrogen bonds formed between the residues of the protein and dorzolamide.

For the potential of a Zn−O.3 pair (where O.3 represents an sp^3-hybridized oxygen atom), illustrated in the subgraph of the lower-left corner of Figure 3, a relatively shallow well can be found at approximately the distance of 2.0 Å, but it did not further deepen through the iteration process, most likely because of the insufficient occurrence of Zn−O.3 in the training set. Deeper potential wells at distances between 5.0 and 8.0 Å characterize the electrical force of atom pairs of a zinc ion and atoms of type O.3. The lower-middle and lower-right subgraphs in Figure 3 illustrate the potentials of the zinc ion and three types of carbon atoms (C.3, C.ar, and C.2). Compared with the potentials of atom pairs that can form coordination bonds, the potentials of zinc ions and carbon atoms are much weaker.

Although most of the strengths and bond lengths of the coordination bonds between zinc ion and donor atoms, including types of nitrogen atoms and oxygen atoms, were successfully interpreted in this study, the coordination bonds between zinc ion and types of sulfur atoms failed to be described by the potentials of these atom pairs because of their insufficient occurrence in our training set. Because sulfur belongs to the oxygen family, we used the potentials of zinc ion and oxygen atom pairs also for zinc ion and sulfur atom pairs.

In Figure 6, the potentials of some atom pairs of PMF04[64,65] (orange lines) were used as a reference for comparison with our obtained potentials and are illustrated to characterize the particular interactions. Our iteration optimization process clearly highlighted the potential wells to characterize the critical interactions, including hydrophobic interactions (cF−C.2 and cF−C.ar), hydrogen-bonding interactions (ND−O.2, and OD−N.2), and electrostatic interactions (NC−O.co2 and OC−N.4). However, the hydrogen-bonding interaction of OA−N.3 did not accurately characterize the performance of a relatively shallow potential well at a distance of 2.8 Å.

3. Validation of $MpSDock_{Zn}$. As mentioned above in Preparation of the Data Sets, a test set of 106 zinc metalloprotein complexes was used to evaluate the docking accuracy of $MpSDock_{Zn}$ regarding its robustness in generating and identifying native docking poses by comparison with six

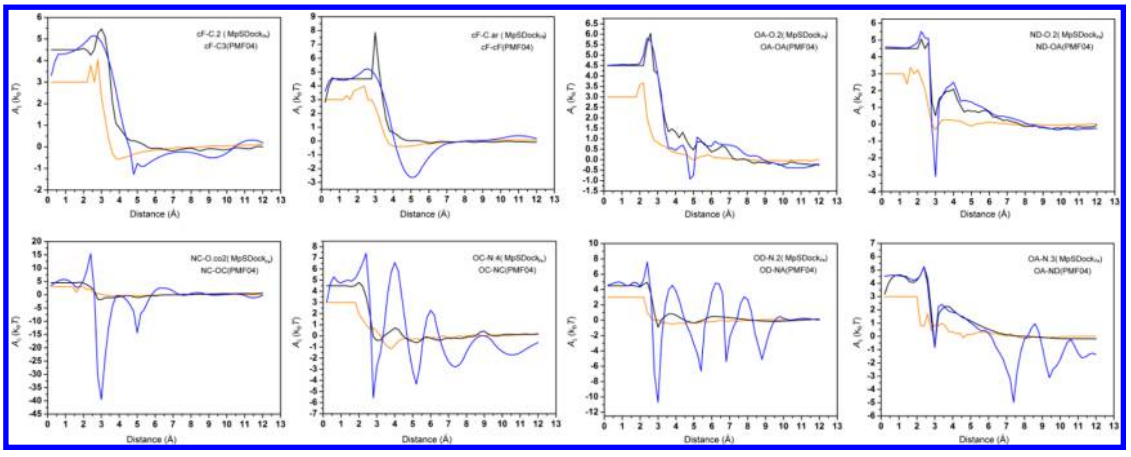


Figure 6. Eight selected ordinary (zinc ion not involved) atom-pair potentials of our zinc metalloprotein-specific scoring function integrated into $MpSDock_{Zn}$. The atom-type definitions for $MpSDock_{Zn}$ originated from KScore. The first atom label refers to the type of the protein atom (cF, nonpolar aromatic carbon; OD, oxygen as hydrogen-bond donor; OA, oxygen as hydrogen-bond acceptor; OC, negatively charged oxygen; ND, nitrogen as hydrogen-bond donor; NA, nitrogen as hydrogen-bond acceptor; NC, positively charged nitrogen), and the second refers to the type of ligand atom (C.2 and C.ar, sp^2-hybridized and aromatic carbons, respectively; N.2, N.3, and N.4, sp^2-hybridized, sp^3-hybridized, and positively charged sp^3-hybridized nitrogens, respectively; O.2 and O.co2, sp^2-hybridized oxygen and oxygen in a carboxylate or phosphate group, respectively). Black lines represent the primary potentials without refinement, the blue lines provide the eventual potentials after the optimization, and the orange lines are the potentials of PMF04.

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

popular docking programs: AutoDock, AutoDock4$_{Zn}$, Glide XP, Glide SP, Gold, and EADock DSS. Thus, two evaluation criteria were considered in this study: the success rate of predicted lowest-RMSD poses with an RMSD of less than 2.0 Å among protein−ligand complexes in the test set (the "RMSD Best" criterion) and the success rate of predicted top-scored poses with an RMSD of less than 2.0 Å among protein−ligand complexes (the "Score Best" criterion), which was mainly used to assess the sensitivity of the scoring function for predicting and ranking the docking poses. Because the units of the knowledge-based and force-field-based scoring functions built into MpSDock$_{Zn}$ differ, the docked poses were ranked only on the basis of our zinc metalloprotein-specific knowledge-based scoring function in the final docking step.

The following parameters were used for docking: the active site included protein residues within a sphere with a radius of 20.0 Å centered on the center of mass of the ligand in the experimental complex. For each NSGA II optimization run, 600 generations were performed on an initial population with a size of 3000, and the operator weights for crossover and mutation were set to 0.9 and 0.2, respectively.

Figure 7 summarizes the comparative success rates under the two evaluation criteria of MpSDock$_{Zn}$ and the other docking

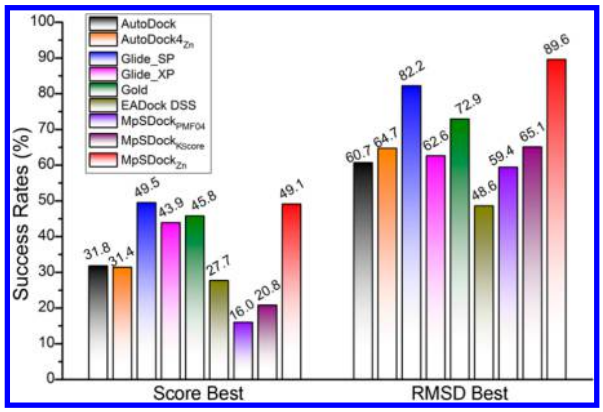


Figure 7. Comparison of the performances of MpSDock$_{Zn}$ and six other popular docking programs (AutoDock, AutoDock4$_{Zn}$, Glide XP mode, Glide SP mode, Gold, and EADock DSS) as well as the MpSDock-based docking strategies in which the zinc metalloprotein-specific knowledge-based scoring function was replaced by the general knowledge-based scoring functions KScore and PMF04 (denoted as MpSDock$_{KScore}$ and MpSDock$_{PMF04}$, respectively, for discrimination from MpSDock$_{Zn}$) in predicting and ranking native docking poses on the test set of 106 zinc metalloproteins.

approaches in docking evaluation studies. The detailed docking results for MpSDock$_{Zn}$ for the test set are presented in Table S4 in the Supporting Information. For the evaluation criterion of "RMSD Best", MpSDock$_{Zn}$ yielded the highest success rate of 89.6% in reproducing the native binding poses of ligands in the test set compared with the other docking approaches. Glide XP also performed well and was ranked second, with a success rate of 82.2%, followed by Gold (72.9%), MpSDock$_{KScore}$ (65.1%), AutoDock4$_{Zn}$ (64.7%), Glide XP (62.6%), AutoDock (60.7%), MpSDock$_{PMF04}$ (59.4%), and EADock DSS (48.6%). For the evaluation criterion of "Score Best", MpSDock$_{Zn}$ exhibited competitive performance in identifying and ranking native binding poses (the highest-ranked docking pose had an RMSD of <2.0 Å compared with the X-ray binding pose) of the ligand from other decoy poses with a success rate of 49.1% (when X-ray binding poses were excluded, the rate increased to 56.6%; the X-ray binding poses were included solely for scoring), compared with Glide SP (49.5%). Gold and Glide XP yielded comparable results (45.8% and 43.9%), while AutoDock (31.8%), EADock DSS (27.7%), MpSDock$_{KScore}$ (20.8%), and MpSDock$_{PMF04}$ (16.0%) yielded relatively lower success rates. These results indicate that the zinc metalloprotein-specific knowledge-based scoring function and MpSDock$_{Zn}$ developed in this study are a promising strategy for treating zinc metalloproteins.

4. Performance of MpSDock$_{Zn}$ for Different Coordination Types. Because of the complicated coordination modes of zinc metalloproteins, the test set was clustered into subgroups according to coordination bond mode for further analysis. Among the 106 zinc metalloproteins in our test set, tetra- and pentacoordinated geometries were the major constituents: 51 were tetracoordinated and 41 were pentacoordinated. Figure 8 compares the performances of

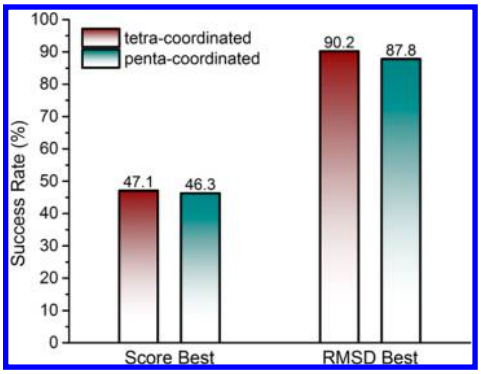


Figure 8. Comparison of the performances of MpSDock$_{Zn}$ in working with tetracoordinated and pentacoordinated zinc metalloproteins.

MpSDock$_{Zn}$ in reproducing the native binding poses of ligands in tetracoordinate and pentacoordinate complexes. The zinc metalloprotein-specific scoring function in MpSDock$_{Zn}$ achieved competitive performance in ranking native binding poses for both tetracoordinated and pentacoordinated geometries. The accuracy of MpSDock$_{Zn}$ was higher for computing tetracoordinated zinc metalloproteins than pentacoordinated zinc metalloproteins.

To further evaluate the performance of MpSDock$_{Zn}$, a series of tetra- and pentacoordinated complexes with different zinc ion-coordinated groups were selected from the test set as examples, and their "RMSD Best" binding poses are illustrated with their experimental binding poses in Figures 9 and 10. Figure 9 presents MpSDock$_{Zn}$ docking results for four tetracoordinate complexes: carbonic anhydrase IV in complex with thioether benzenesulfonamide (Figure 9A, PDB code 3F7B), fructose-1,6-bisphosphatase cocrystallized with fructose-2,6-diphosphate (Figure 9B, PDB code 2QVV), serine protease in complex with bis(5-amidino-2-benzimidazolyl) methane ketone (Figure 9C, PDB code 1C2E), and the complex of carboxypeptidase II with (2*S*)-2-{[*S*]-(2-carboxyethyl)-(hydroxy)phosphoryl]methyl}pentanedioic acid (Figure 9D, PDB code 3BHX). In these four tetracoordinated zinc metalloproteins, the zinc ion separately coordinates with the nitrogen of the sulfonamide group (N.3), the oxygen of the 1-hydroxyl group (O.3), one of nitrogens from each of the two benzimidazole groups (N.2), and an oxygen of the phosphonate group (O.co2) of the corresponding ligands.

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

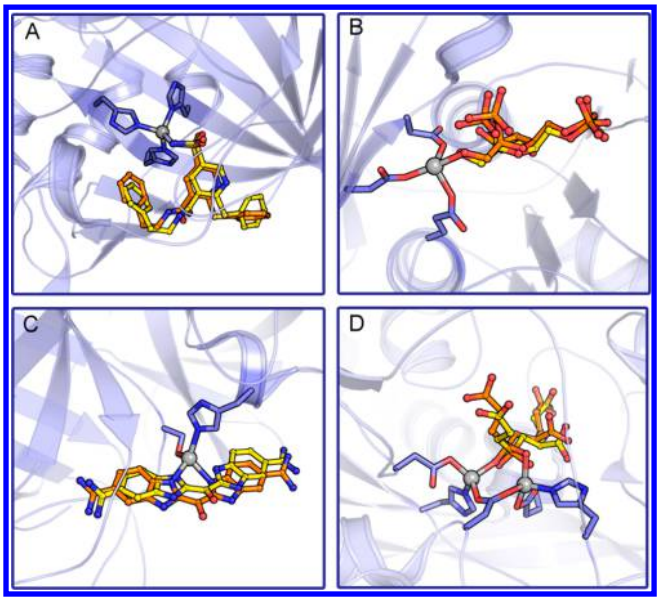


Figure 9. Docking result examples of calculated tetracoordinate zinc metalloproteins in which zinc ions coordinate with different functional groups of ligands: (A) Zn−N.3 coordinated case (PDB code 3F7B); (B) Zn−O.3 coordinated case (PDB code 2QVV); (C) bidentate Zn−N.2 coordinated case (PDB code 1C2E); (D) Zn−O.co2 coordinated case (PDB code 3BHX). The lilac cartoon modes represent the receptors, the violet stick modes represent the focused residues in the active sites, the gray spheres are the zinc ions, the orange ball-and-stick modes represent the experimentally determined binding poses of the ligands, and the yellow ball-and-stick modes represent the best predicted binding poses of the ligands.

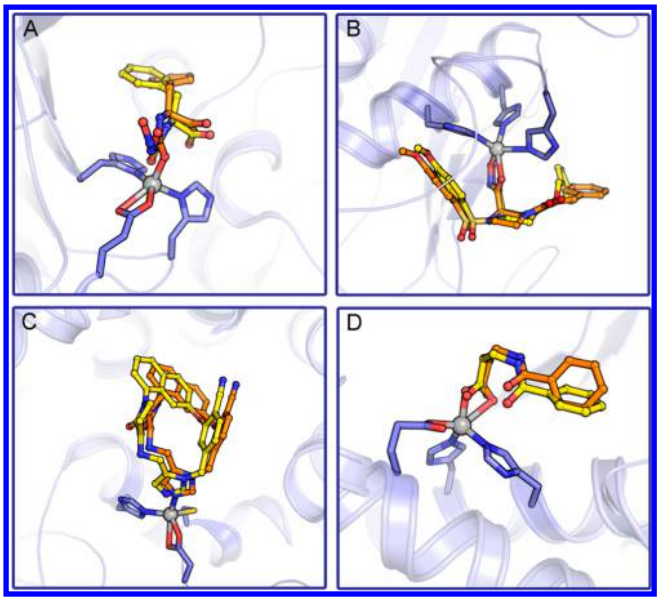


Figure 10. Docking result examples of calculated pentacoordinated cases in which zinc ions are coordinated with different functional groups of ligands: (A) Zn−O.2 coordinated case (PDB code 1HEE); (B) bidentate Zn−O.3 and Zn−O.2 coordinated case (PDB code 1D8F); (C) Zn−N.2 coordinated case (PDB code 1LD8); and (D) bidentate Zn−O.co2 coordinated case (PDB code 3FGD). The lilac cartoon modes represent the receptors, the violet stick modes represent the focused residues in the active sites, the gray spheres represent the zinc ions, the orange ball-and-stick modes represents the experimentally determined binding poses of the ligands, and the yellow ball-and-stick modes represent the best predicted binding poses of the ligands.

Similarly, Figure 10 presents the $MpSDock_{Zn}$ docking results for four representative pentacoordinated zinc metalloproteins: carboxypeptidase A with *N*-(hydroxyaminocarbonyl)-phenylalanine (Figure 10A, PDB code 1HEE), matrix metalloproteinase III with *N*-hydroxy-1-(4-methoxyphenyl)sulfonyl-4-benzyloxycarbonylpiperazine-2-carboxamide (Figure 10B, PDB code 1D8F), farnesyltransferase with (20*S*)-19,20,21,22-tetrahydro-19-oxo-5*H*-18,20-ethano-12,14-etheno-6,10-methenobenz[*d*]imidazo[4,3-*k*][1,6,9,12]-oxatriazacyclooctadecosine-9-carbonitile (Figure 10C, PDB code 1LD8), and thermolysin with *N*-(phenylcarbonyl)-*β*-alanine (Figure 10D, PDB code 3FGD). In these four pentacoordinated zinc metalloproteins, the zinc ions form coordination bonds with the oxygen of the amido group (O.2), the oxygen in the *N*-hydroxyl group (O.3) and the oxygen of the amido group (O.2), the nitrogen in the imidazole group (N.2), and the two oxygen atoms of the carboxyl group (O.co2) of the corresponding ligands. The ligands in Figures 9C and 10B,D form coordination bonds with zinc ions in a bidentate coordination mode. All of the predicted binding poses illustrated in Figures 9 and 10 match well with the experimental binding modes in the crystal complexes. These results again validate the efficiency of our zinc metalloprotein-specific potential and the performance of $MpSDock_{Zn}$.

As mentioned above in Properties of the Optimized Atom-Pair Potentials, the potentials of atom pairs of zinc ion and types of sulfur atoms failed to be characterized in this study, and the corresponding potentials of zinc ions and oxygen atoms were compensatorily adopted. To validate the feasibility of this compensation, the docking results for the complexes with coordination bonds between the zinc ion and sulfur atoms of ligands were isolated from the test set for further study. There were 15 zinc metalloproteins complexed with sulfur ligands in our test set, including seven sulfonamide and eight thiol ligands (Table S4 in the Supporting Information). The docking results for the cocrystal complex of testicular angiotensin I-converting enzyme with 1-(3-mercapto-2-methylpropionyl)pyrrolidine-2-carboxylic acid (PDB code 1UZF) and the complex of metallo-*β*-lactamases with *N*-(3-mercaptopropanoyl)-D-alanine (PDB code 2QDT), shown in Figure 11A,B, respectively, illustrate the excellent agreement achieved. These results demonstrate that $MpSDock_{Zn}$ can effectively sample their binding poses;

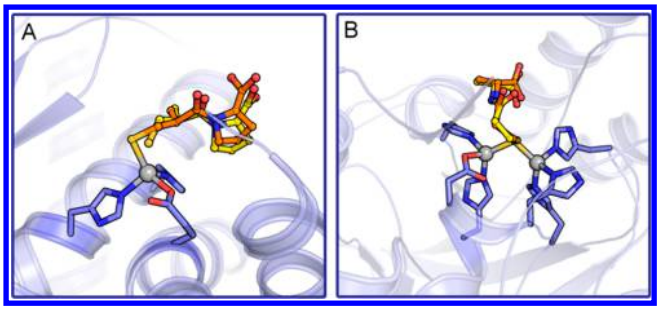


Figure 11. Docking result examples of sulfur ligands in participating coordination cases: (A) PDB code 1UZF; (B) PDB code 2QDT. The lilac cartoon modes represent the receptors, the violet stick modes represent the focused residues in the active sites, the gray spheres are the zinc ions, the orange ball-and-stick modes represent the experimentally determined binding poses of the ligands, and the yellow ball-and-stick modes represent the "RMSD Best" binding poses of the ligands.

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

however, the accuracy and sensitivity are not comparable to those for the binding pose sampling and still need to be improved in future studies.

5. Validation of Zinc Metalloprotein-Specific Atom-Pair Potentials. To further validate our zinc metalloprotein-specific atom-pair potentials, additional comparative docking tests were performed among $MpSDock_{Zn}$, $MpSDock_{KScore}$, and $MpSDock_{PMF04}$ against the two test sets used in the KScore validation study:[55] 15 zinc metalloproteins (PDB codes are provided in Table S5 in the Supporting Information) and 286 ordinary complexes (without metal ions; PDB codes are provided in Table S6 in the Supporting Information). The same settings for the parameters used in the test set of 106 zinc metalloproteins were used in this validation. The assessment was performed on the basis of the "RMSD Best" criterion, and the comparison results are illustrated in Figure 12. Our zinc

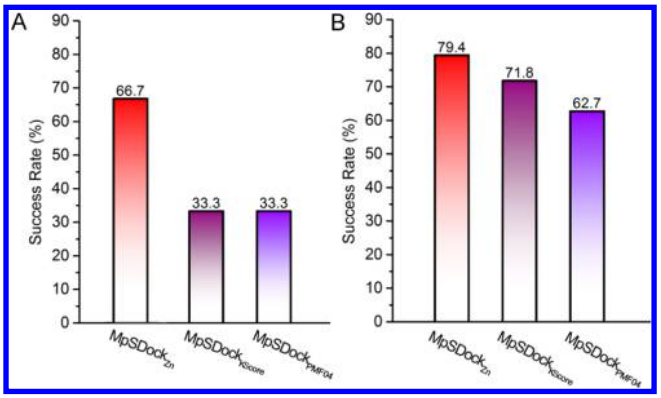


Figure 12. Comparison of the performances of MpSDock run with different knowledge-based scoring functions (our optimized zinc metalloprotein-specific scoring function, KScore, and PMF04, specifically noted as $MpSDock_{Zn}$, $MpSDock_{KScore}$, and $MpSDock_{PMF04}$, respectively) against (A) the test set of the 15 zinc metalloproteins used in the KScore validation[55] and (B) the test set of 286 ordinary complexes used in the KScore validation.

metalloprotein-specific knowledge-based scoring function more efficiently enabled MpSDock to reproduce the binding poses of ligands to zinc metalloproteins (Figure 12A). Figure 12B illustrates that our zinc metalloprotein-specific potential is also competent to calculate ordinary proteins or enzymes due to the better performance of $MpSDock_{Zn}$ compared with $MpSDcok_{KScore}$ and $MpSDcok_{PMF04}$. In other words, our zinc metalloprotein-specific knowledge-based scoring function is not only sufficiently sensitive and effective to calculate zinc metalloproteins but is also reasonable for other nonbonded atom-pair potentials, indicating its general usefulness in docking tools. These results demonstrate that the optimized atom-pair potentials are reasonable and effective.

6. Comparison of Zinc Metalloprotein-Specific Atom-Pair Potentials before and after Optimization. To determine the effect of the dynamic sampling and iteration optimization strategy on our knowledge-based zinc metalloprotein-specific atom-pair potentials, the atom-pair potentials before iteration were applied as the knowledge-based scoring function in MpSDock (marked as $MpSDock_{Zn(before\ optimization)}$), and a similar docking test was performed on the test set of 106 zinc metalloproteins. The parameters were the same as above, and the docking results were also assessed under the two evaluation criteria, as illustrated in Figure 13. Remarkably, compared with $MpSDock_{Zn(before\ optimzation)}$, $MpSDock_{Zn}$ ex-

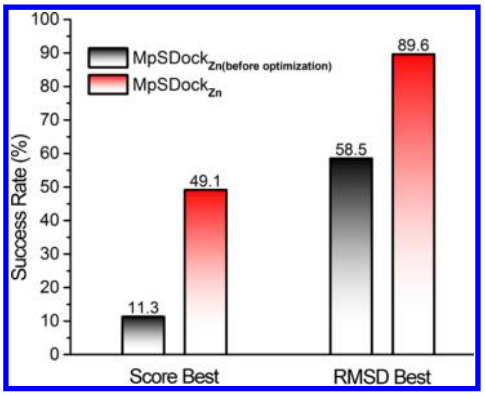


Figure 13. Investigation of the improvement of our zinc metalloprotein-specific potentials in predicting and ranking native binding poses before and after optimization.

hibited a significant improvement in predicting native binding poses: (1) the sensitivity of the scoring function was remarkably increased, and the success rate under the "Score Best" criterion was increased 4.3-fold (when the X-ray poses were excluded, the increase was 2.4-fold; the X-ray poses were included only for scoring) after the built-in zinc metalloprotein-specific atom-pair potentials of $MpSDock_{Zn}$ were optimized by iteration adjustment; (2) the ability of $MpSDock_{Zn}$ to sample the binding poses and predict the native binding poses was significantly improved (the success rate under the "RMSD Best" criterion was increased 1.5-fold) with the optimized atom-pair potentials.

CONCLUSIONS

In the present study, a metalloprotein-specific docking program named MpSDock was designed and introduced. This program employs NSGA II, a multiobjective optimization algorithm, to identify the binding poses of a ligand for its metalloprotein receptor. We designed a zinc metalloprotein-specific force-field-based scoring function based on the recently reported zinc metalloprotein-specific, nonbonded, pairwise SLEF force field and developed a new zinc metalloprotein-specific knowledge-based scoring function based on the inverse Boltzmann law, nonlinear constrained optimization, and a designed dynamic sampling and iteration optimization strategy. These zinc metalloprotein-specific atom-pair potentials can reasonably and quantitatively describe the characteristics of different interactions between different atom pairs, providing insights into the functions of these atom-pair potentials and highlighting important atom-pair interactions, such as coordination bonds, hydrogen bonds, and electrostatic interactions, thereby enhancing the ability to identify true-positive molecules and avoid false-positive results.

On the basis of these two scoring functions, a scheme similar to consensus scoring was designed, and a zinc metalloprotein-specific docking program, $MpSDock_{Zn}$, was specifically developed. An extensive comparative validation study of $MpSDock_{Zn}$ and six other popular docking programs (AutoDock, $AutoDock4_{Zn}$, Glide XP mode, Glide SP mode, Gold, and EADock DSS) were performed using a test set of 106 zinc metalloprotein complexes. The success rates under the two evaluation criteria ("Score Best" and "RMSD Best") for sampling and ranking of native binding poses of $MpSDock_{Zn}$ were clearly competitive with those of the other docking programs, increasing from a success rate of 49.1% ("Score

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

Best", ranking native binding poses) to 89.6% ("RMSD Best", sampling native binding poses). The Glide XP mode, the next-best-performing docking program, achieved competitive results with success rates of 49.5% and 82.2% under the two evaluation criteria, respectively. Thus, we conclude that to address the number and complexity of zinc metalloproteins, the zinc metalloprotein-specific knowledge-based scoring function designed in this study is a reliable means of scoring ligands to zinc metalloproteins. The performance of $MpSDock_{Zn}$ in both predicting and ranking native binding poses indicates that this scoring function is a valuable approach for drug discovery and design targeting zinc metalloproteins. The detailed analysis of complex-by-complex docking results indicated that the types of zinc ion-coordinated groups (e.g., zinc−carboxylate coordination, zinc−hydroxamate coordination, zinc−sulfamate coordination, zinc−sulfonamide coordination, and zinc−sulfoxylate coordination) were well-predicted by $MpSDock_{Zn}$, further establishing and validating this method. The executable program of $MpSDock_{zn}$ can be provided upon request, and the knowledge-based zinc metalloprotein-specific scoring function of $MpSDock_{Zn}$ is available at http://lilab.ecust.edu.cn/home/resource.html.

■ ASSOCIATED CONTENT

Supporting Information

All of the PDB information on the training sets and test sets used in this study and the detailed docking results of $MpSDock_{Zn}$ on the test set of 106 zinc metalloproteins. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Address: School of Pharmacy, East China University of Science and Technology 130 Mei Long Road, Shanghai 200237, China. Phone: +86-21-64250213. Fax: +86-21-64250213. E-mail: hlli@ecust.edu.cn.

Author Contributions

⊥F.B. and S.L. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We dedicate this paper to the memory of Prof. Xicheng Wang, who unfortunately passed away before the paper was submitted for publication. This work was supported by the 863 Hi-Tech Program of China (Grant 2012AA020308), the National Natural Science Foundation of China (Grants 81222046, 81230076, 11202049, and 11072048), and the Fundamental Research Funds for the Central Universities. H.L. was also sponsored by the Shanghai Rising-Star Tracking Program (Grant 13QH1401100), the Innovation Program of the Shanghai Municipal Education Commission (Grant 13SG32) and the Fok Ying Tung Education Foundation (Grant 141035).

■ REFERENCES

(1) Andreini, C.; Banci, L.; Bertini, I.; Rosato, A. Counting the Zinc-Proteins Encoded in the Human Genome. *J. Proteome Res.* **2006**, *5*, 196−201.

(2) Kamei, N.; Tanaka, T.; Kawai, K.; Miyawaki, K.; Okuyama, A.; Murakami, Y.; Arakawa, Y.; Haino, M.; Harada, T.; Shimano, M. Reverse Hydroxamate-Based Selective TACE Inhibitors. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2897−2900.

(3) Sheppeck, J. E., II; Gilmore, J. L.; Yang, A.; Chen, X. T.; Xue, C. B.; Roderick, J.; Liu, R. Q.; Covington, M. B.; Decicco, C. P.; Duan, J. J. Discovery of Novel Hydantoins as Selective Non-Hydroxamate Inhibitors of Tumor Necrosis Factor-α Converting Enzyme (TACE). *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1413−1417.

(4) Attolino, E.; Calderone, V.; Dragoni, E.; Fragai, M.; Richichi, B.; Luchinat, C.; Nativi, C. Structure-Based Approach to Nanomolar, Water Soluble Matrix Metalloproteinases Inhibitors (MMPIs). *Eur. J. Med. Chem.* **2010**, *45*, 5919−5925.

(5) Conti, P.; Tamborini, L.; Pinto, A.; Sola, L.; Ettari, R.; Mercurio, C.; De Micheli, C. Design and Synthesis of Novel Isoxazole-Based HDAC Inhibitors. *Eur. J. Med. Chem.* **2010**, *45*, 4331−4338.

(6) Duez, S.; Coudray, L.; Mouray, E.; Grellier, P.; Dubois, J. Towards the Synthesis of Bisubstrate Inhibitors of Protein Farnesyltransferase: Synthesis and Biological Evaluation of New Farnesylpyrophosphate Analogues. *Bioorg. Med. Chem.* **2010**, *18*, 543−556.

(7) Brzozowski, Z.; Slawinski, J.; Gdaniec, M.; Innocenti, A.; Supuran, C. T. Carbonic Anhydrase Inhibitors. Synthesis, Molecular Structures, and Inhibition of the Human Cytosolic Isozymes I and II and Transmembrane Isozymes IX, XII (Cancer-Associated) and XIV with Novel 3-Pyridinesulfonamide Derivatives. *Eur. J. Med. Chem.* **2011**, *46*, 4403−4410.

(8) Wang, Z. L.; Zhang, S. S.; Jin, H. W.; Wang, W.; Huo, J. X.; Zhou, L. S.; Wang, Y. F.; Feng, F. Q.; Zhang, L. R. Angiotensin-I-Converting Enzyme Inhibitory Peptides: Chemical Feature Based Pharmacophore Generation. *Eur. J. Med. Chem.* **2011**, *46*, 3428−3433.

(9) Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **2003**, *10*, 980.

(10) Kawai, K.; Nagata, N. Metal−Ligand interactions: An Analysis of Zinc Binding Groups Using the Protein Data Bank. *Eur. J. Med. Chem.* **2012**, *51*, 271−276.

(11) Chaskar, P.; Zoete, V.; Röhrig, U. F. Toward On-the-Fly Quantum Mechanical/Molecular Mechanical (QM/MM) Docking: Development and Benchmark of a Scoring Function. *J. Chem. Inf. Model.* **2014**, *54*, 3137−3152.

(12) Patel, K.; Kumar, A.; Durani, S. Analysis of the Structural Consensus of the Zinc Coordination Centers of Metalloprotein Structures. *Biochim. Biophys. Acta* **2007**, *1774*, 1247−1253.

(13) Donini, O. A.; Kollman, P. A. Calculation and Prediction of Binding Free Energies for the Matrix Metalloproteinases. *J. Med. Chem.* **2000**, *43*, 4180−4188.

(14) Chakravorty, D. K.; Wang, B.; Lee, C. W.; Giedroc, D. P.; Merz, K. M., Jr. Simulations of Allosteric Motions in the Zinc Sensor CzrA. *J. Am. Chem. Soc.* **2012**, *134*, 3367−3376.

(15) Yang, Y.; Chakravorty, D. K.; Merz, K. M., Jr. Finding a Needle in the Haystack: Computational Modeling of Mg^{2+} Binding in the Active Site of Protein Farnesyltransferase. *Biochemistry* **2010**, *49*, 9658−9666.

(16) Lee, C. W.; Chakravorty, D. K.; Chang, F.-M. J.; Reyes-Caballero, H.; Ye, Y. Z.; Merz, K. M., Jr.; Giedroc, D. P. Solution Structure of *Mycobacterium tuberculosis* NmtR in the Apo State: Insights into Ni(II)-Mediated Allostery. *Biochemistry* **2012**, *51*, 2619−2629.

(17) Vedani, A.; Huhta, D. W. A New Force Field for Modeling Metalloproteins. *J. Am. Chem. Soc.* **1990**, *112*, 4759−4767.

(18) Hoops, S. C.; Anderson, K. W.; Merz, K. M. Force Field Design for Metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262−8270.

(19) Lin, F.; Wang, R. X. Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *J. Chem. Theory Comput.* **2010**, *6*, 1852−1870.

(20) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M., Jr. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6*, 2935−2947.

(21) Deeth, R. J.; Anastasi, A.; Diedrich, C.; Randell, K. Molecular Modelling for Transition Metal Complexes: Dealing with d-Electron Effects. *Coord. Chem. Rev.* **2009**, *253*, 795−816.

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

(22) Stote, R. H.; States, D. J.; Karplus, M. On the Treatment of Electrostatic Interactions in Biomolecular Simulation. *J. Chim. Phys.* **1991**, *88*, 2419−2433.
(23) Wu, J. C.; Piquemal, J.-P.; Chaudret, R.; Reinhardt, P.; Ren, P. Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 2059−2070.
(24) Ponomarev, S. Y.; Click, T. H.; Kaminski, G. A. Electrostatic Polarization Is Crucial in Reproducing Cu(I) Interaction Energies and Hydration. *J. Phys. Chem. B* **2011**, *115*, 10079−10085.
(25) Pang, Y. P. Successful Molecular Dynamics Simulation of Two Zinc Complexes Bridged by a Hydroxide in Phosphotriesterase Using the Cationic Dummy Atom Method. *Proteins: Struct., Funct., Bioinf.* **2001**, *45*, 183−189.
(26) Wu, R. B.; Lu, Z. Y.; Cao, Z. X.; Zhang, Y. K. A Transferable Nonbonded Pairwise Force Field To Model Zinc Interactions in Metalloproteins. *J. Chem. Theory Comput.* **2011**, *7*, 433−443.
(27) Chakravorty, D.; Wang, B.; Lee, C.; Guerra, A.; Giedroc, D.; Merz, K. M., Jr. Solution NMR Refinement of a Metal Ion Bound Protein Using Metal Ion Inclusive Restrained Molecular Dynamics Methods. *J. Biomol. NMR* **2013**, *56*, 125−137.
(28) Wang, R. W.; Liu, L.; Lai, L. H.; Tang, Y. Q. SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein−Ligand Complex. *J. Mol. Med.* **1998**, *4*, 379−394.
(29) Han, L. Y.; Lin, H. H.; Li, Z. R.; Zheng, C. J.; Cao, Z. W.; Xie, B.; Chen, Y. Z. PEARLS: Program for Energetic Analysis of Receptor−Ligand System. *J. Chem. Inf. Model.* **2006**, *46*, 445−450.
(30) Kortemme, T.; Morozov, A. V.; Baker, D. An Orientation-Dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein−Protein Complexes. *J. Mol. Biol.* **2003**, *326*, 1239−1259.
(31) Chen, Y.; Kortemme, T.; Robertson, T.; Baker, D.; Varani, G. A New Hydrogen-Bonding Potential for the Design of Protein−RNA Interactions Predicts Specific Contacts and Discriminates Decoys. *Nucleic Acids Res.* **2004**, *32*, 5147−5162.
(32) Schiffmann, R.; Neugebauer, A.; Klein, C. D. Metal-Mediated Inhibition of *Escherichia coli* methionine Aminopeptidase: Structure−Activity Relationships and Development of a Novel Scoring Function for Metal−Ligand Interactions. *J. Med. Chem.* **2006**, *49*, 511−522.
(33) Tietze, S.; Apostolakis, J. GlamDock: Development and Validation of a New Docking Tool on Several Thousand Protein−Ligand Complexes. *J. Chem. Inf. Model.* **2007**, *47*, 1657−1672.
(34) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function To Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(35) Korb, O.; Stutzle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84−96.
(36) Morozov, A. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. Close Agreement between the Orientation Dependence of Hydrogen Bonds Observed in Protein Structures and Quantum Mechanical Calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6946−6951.
(37) Raub, S.; Steffen, A.; Kamper, A.; Marian, C. M. AIScore Chemically Diverse Empirical Scoring Function Employing Quantum Chemical Binding Energies of Hydrogen-Bonded Complexes. *J. Chem. Inf. Model.* **2008**, *48*, 1492−1510.
(38) Artemenko, N. Distance Dependent Scoring Function for Describing Protein−Ligand Intermolecular Interactions. *J. Chem. Inf. Model.* **2008**, *48*, 569−574.
(39) Stroganov, O. V.; Novikov, F. N.; Stroylov, V. S.; Kulkov, V.; Chilov, G. G. Lead Finder: An Approach To Improve Accuracy of Protein−Ligand Docking, Binding Energy Estimation, and Virtual Screening. *J. Chem. Inf. Model.* **2008**, *48*, 2371−2385.
(40) Huang, S. Y.; Zou, X. Q. An Iterative Knowledge-Based Scoring Function To Predict Protein−Ligand Interactions: II. Validation of the Scoring Function. *J. Comput. Chem.* **2006**, *27*, 1876−1882.
(41) Huang, S. Y.; Zou, X. Q. An Iterative Knowledge-Based Scoring Function for Protein−Protein Recognition. *Proteins: Struct., Funct., Bioinf.* **2008**, *72*, 557−579.
(42) Huang, S. Y.; Zou, X. Q. An Iterative Knowledge-Based Scoring Function To Predict Protein−Ligand Interactions: I. Derivation of Interaction Potentials. *J. Comput. Chem.* **2006**, *27*, 1866−1875.
(43) Huang, S. Y.; Zou, X. Q. Inclusion of Solvation and Entropy in the Knowledge-Based Scoring Function for Protein−Ligand Interactions. *J. Chem. Inf. Model.* **2010**, *50*, 262−273.
(44) Xue, M. Z.; Zheng, M. Y.; Xiong, B.; Li, Y. L.; Jiang, H. L.; Shen, J. K. Knowledge-Based Scoring Functions in Drug Design. 1. Developing a Target-Specific Method for Kinase−Ligand Interactions. *J. Chem. Inf. Model.* **2010**, *50*, 1378−1386.
(45) Zheng, M. Z.; Xiong, B.; Luo, C.; Li, S. S.; Liu, X.; Shen, Q. C.; Li, J.; Zhu, W. L.; Luo, X. M.; Jiang, H. L. Knowledge-Based Scoring Functions in Drug Design: 3. A Two-Dimensional Knowledge-Based Hydrogen-Bonding Potential for the Prediction of Protein−Ligand Interactions. *J. Chem. Inf. Model.* **2011**, *51*, 2994−3004.
(46) Shen, Q. C.; Xiong, B.; Zheng, M. Y.; Luo, X. M.; Luo, C.; Liu, X.; Du, Y.; Li, J.; Zhu, W. L.; Shen, J. K.; Jiang, H. L. Knowledge-Based Scoring Functions in Drug Design: 2. Can the Knowledge Base Be Enriched? *J. Chem. Inf. Model.* **2011**, *51*, 386−397.
(47) Thomas, P. D.; Dill, K. A. Statistical Potentials Extracted from Protein Structures: How Accurate Are They? *J. Mol. Biol.* **1996**, *257*, 457−469.
(48) Kang, L.; Li, H. L.; Jiang, H. L.; Wang, X. C. An Improved Adaptive Genetic Algorithm for Protein−Ligand Docking. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 1−12.
(49) Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182−197.
(50) Wang, R. X.; Fang, X. L.; Lu, Y. P.; Wang, S. M. The PDBbind Database: Collection of Binding Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.
(51) Wang, R. X.; Fang, X. L.; Lu, Y. P.; Yang, C. Y.; Wang, S. M. The PDBbind Database: Methodologies and Updates. *J. Med. Chem.* **2005**, *48*, 4111−4119.
(52) Gao, Z. T.; Li, H. L.; Zhang, H. L.; Liu, X. F.; Kang, L.; Luo, X. M.; Zhu, W. L.; Chen, K. X.; Wang, X. C.; Jiang, H. L. PDTD: A Web-Accessible Protein Database for Drug Target Identification. *BMC Bioinf.* **2008**, *9*, No. 104.
(53) Block, P.; Sotriffer, C. A.; Dramburg, I.; Klebe, G. AffinDB: A Freely Accessible Database of Affinities for Protein−Ligand Complexes from the PDB. *Nucleic Acids Res.* **2006**, *34*, D522−D526.
(54) Liu, T. Q.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: A Web-Accessible Database of Experimentally Determined Protein−Ligand Binding Affinities. *Nucleic Acids Res.* **2007**, *35*, D198−D201.
(55) Zhao, X. Y.; Liu, X. F.; Wang, Y. Y.; Chen, Z.; Kang, L.; Zhang, H. L.; Luo, X. M.; Zhu, W. L.; Chen, K. X.; Li, H. L.; Wang, X. C.; Jiang, H. L. An Improved PMF Scoring Function for Universally Predicting the Interactions of a Ligand with Protein, DNA, and RNA. *J. Chem. Inf. Model.* **2008**, *48*, 1438−1447.
(56) Muegge, I. Effect of Ligand Volume Correction on PMF Scoring. *J. Comput. Chem.* **2001**, *22*, 418−425.
(57) Czyzyk, J.; Mesnier, M. P.; Mor, J. J. The NEOS Server. *IEEE Comput. Sci. Eng.* **1998**, *5*, 68−75.
(58) Almarza, N. G.; Lomba, E. Determination of the Interaction Potential from the Pair Distribution Function: An Inverse Monte Carlo Technique. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2003**, *68*, No. 011202.
(59) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785−2791.
(60) Santos-Martins, D.; Forli, S.; Ramos, M. J.; Olson, A. J. AutoDock4$_{Zn}$: An Improved AutoDock Force Field for Small-

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847

Molecule Docking to Zinc Metalloproteins. *J. Chem. Inf. Model.* **2014**, *54*, 2371−2379.
(61) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(62) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177−6196.
(63) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein−Ligand Docking Using GOLD. *Proteins: Struct., Funct., Bioinf.* **2003**, *52*, 609−623.
(64) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791−804.
(65) Muegge, I. PMF Scoring Revisited. *J. Med. Chem.* **2006**, *49*, 5895−5902.
(66) Hu, X.; Shelver, W. H. Docking Studies of Matrix Metalloproteinase Inhibitors: Zinc Parameter Optimization To Improve the Binding Free Energy Prediction. *J. Mol. Graphics Modell.* **2003**, *22*, 115−126.
(67) Cheng, T. J.; Li, X.; Li, Y.; Liu, Z. H.; Wang, R. X. Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Inf. Model.* **2009**, *49*, 1079−1093.
(68) Mackerell, A. D., Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25*, 1584−1604.
(69) Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a Protein−Small Molecule Docking Web Service Based on EADock DSS. *Nucleic Acids Res.* **2011**, *39*, W270−W277.
(70) Vivó-Truyols, G.; Schoenmakers, P. J. Automatic Selection of Optimal Savitzky−Golay Smoothing. *Anal. Chem.* **2006**, *78*, 4598−4608.
(71) Mehrotra, R. C.; Bohra, R. *Metal Carboxylates*; Academic Press: New York, 1983.

DOI: 10.1021/ci500647f
J. Chem. Inf. Model. 2015, 55, 833−847