PROTEINS: Structure, Function, and Bioinformatics 67:1167–1178 (2007)

Computational Protocol for Predicting the Binding Affinities of Zinc Containing Metalloprotein–Ligand Complexes

Tarun Jain[1,2] **and B. Jayaram**[1,2]*
[1]*Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India*
[2]*Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India*

ABSTRACT **Zinc is one of the most important metal ions found in proteins performing specific functions associated with life processes. Coordination geometry of the zinc ion in the active site of the metalloprotein–ligand complexes poses a challenge in determining ligand binding affinities accurately in structure-based drug design. We report here an all atom force field based computational protocol for estimating rapidly the binding affinities of zinc containing metalloprotein–ligand complexes, considering electrostatics, van der Waals, hydrophobicity, and loss in conformational entropy of protein side chains upon ligand binding along with a nonbonded approach to model the interactions of the zinc ion with all the other atoms of the complex. We examined the sensitivity of the binding affinity predictions to the choice of Lennard-Jones parameters, partial atomic charges, and dielectric treatments adopted for system preparation and scoring. The highest correlation obtained was R^2 = 0.77 (r = 0.88) for the predicted binding affinity against the experiment on a heterogenous dataset of 90 zinc containing metalloprotein–ligand complexes consisting of five unique protein targets. Model validation and parameter analysis studies underscore the robustness and predictive ability of the scoring function. The high correlation obtained suggests the potential applicability of the methodology in designing novel ligands for zinc–metalloproteins. The scoring function has been web enabled for free access at www.scfbio-iitd.res.in/software/drugdesign/bapplz.jsp as BAPPL-Z server (Binding Affinity Prediction of Protein–Ligand complexes containing Zinc metal ions). Proteins 2007;67:1167–1178.**

Key words: scoring function; binding affinity; metalloenzyme; zinc ion; structure-based drug design

INTRODUCTION

Metal ions (Zn^{2+}, Mg^{2+}, Ca^{2+}, Mn^{2+}, Na^{+}, K^{+}, etc.) perform a wide variety of specific functions associated with life processes. They are involved in respiration, triggering cellular responses, electron transfer, catalytic reactions and stabilizing the structure of folded proteins,[1,2] and so forth. One of the most common transition metals found in enzymes is zinc. Zinc containing proteins play a key role in the biosynthesis and metabolism of a number of bioactive peptides and have been implicated in a variety of disease states such as cancer, arthritis, and multiple sclerosis.[3] Carbonic anhydrase (CA),[4] carboxypeptidase A (CPA),[5] alcohol dehydrogenase (AD),[6] matrix metalloproteinase (MMP),[7] and thermolysin (TL)[8] are some of the zinc–metalloproteinases[9] investigated thoroughly for their role in various biological processes. As promising therapeutic drug targets, zinc–metalloproteinases have attracted much interest in recent years.

Regardless of the metal and its precise pattern of ligation to the protein, there is a common qualitative feature to the binding site: the metal is ligated by a shell of hydrophilic atomic groups (containing oxygen, nitrogen, or sulfur atoms) and this hydrophilic shell is embedded within a larger shell of hydrophobic atomic groups (containing carbon atoms).[10,11] Zn^{2+} generally exists in 4-, 5-, or 6- coordinate geometry, with ligands such as, His, Cys, Asp, Glu, substrate/inhibitor, and water molecules.[12] Modeling of the ligand binding to the zinc ion is quite problematic because of polarization, charge transfer, multiple coordination geometries, and lack of accurate force fields.[13–16] These limitations have hindered structure-based drug design studies[17,18] aimed at designing novel potent and selective ligands for zinc–metalloproteins. There are two basic ways to model the zinc ion using a purely classical potential function: the bonded model and the nonbonded model. In the bonded models, explicit bond and angle terms are introduced into the potential energy function to account for interactions between the metal and the protein and ligand atoms.[19–22]

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsors: Department of Biotechnology, Government of India and Dabur Research Foundation.

*Correspondence to: B. Jayaram, Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India. E-mail: bjayaram@chemistry.iitd.ac.in

Received 8 June 2006; Revised 12 November 2006; Accepted 15 November 2006

Published online 22 March 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21332

TABLE I. Some Methodologies Reported in the Literature for Estimating the Binding Affinities of Zinc Containing Metalloprotein–Ligand Complexes

S. no.	Contributing group	Method	Metalloprotein studied	Training set	Test set	R^2
1.	Donini et al.[45]	MM-PBSA	MMP	—	6	
2.	Raha et al.[46]	QM	CA and CPA	—	23	0.69
3.	Toba et al.[47]	FEP	MMP	—	2	—
4.	Hou et al.[48]	LIE	MMP	—	15	0.85
5.	Hu et al.[49]	Force field	MMP	—	14	0.50
6.	Rizzo et al.[50]	MM-GBSA	MMP	—	6	0.74
7.	Khandelwal et al.[51]	QM/MM	MMP	—	28	0.90
8.	*Present work*	*Force field/empirical*	*CA, CPA, MMP, AD, and TL*	*40*	*50*	*0.77*

R^2 is the correlation coefficient obtained on the test set for the predicted binding affinities against the experiment.

This model suffers from certain limitations viz. the overall conformational flexibility of the metal binding site gets restricted and the multiplicity of zinc binding modes becomes difficult to treat. In the alternative nonbonded model, only the van der Waals and electrostatic terms are included for the zinc ion.[23] This avoids the conformational restrictions of the bonded model, but encounters difficulty in treating the strong electrostatic interactions of the divalent zinc ion.[24] Several examples exist in the literature where bonded or nonbonded models have been used for zinc protein simulations.[25,26] Besides the aforementioned approaches, quantum mechanics (QM)/molecular mechanics methods have also been used to study the structural and functional roles of zinc bound to protein residues.[27–29]

Computational approaches that "dock" small molecules into the structures of macromolecular targets and "score" their potential complementarity to binding sites are widely used in hit identification and lead optimization.[30] Computational protocols that utilize the receptor structure information for estimating binding affinities[31,32] can be classified into five major classes with respect to their methodological background: (1) molecular simulation based approaches;[33,34] (2) empirical/force field/additivity based approaches;[35,36] (3) knowledge-based approaches;[37] (4) QM-based scoring functions[38]; and (5) hybrid approaches.[39,40] Comparative evaluations of different docking programs[41] in combination with various scoring functions for their applications in virtual screening have been carried out and results show that many of the popular scoring functions are able to select correct docked from misdocked structures, but correlation with experimental binding affinities particularly of metalloprotein–ligand complexes still remains a major limiting factor in virtual screening for drug discovery.[42–44] Despite the hurdles, several groups have reported methodologies to predict the binding affinities of zinc containing metalloprotein–ligand complexes (Table I). Hu et al. have carried out docking and scoring studies on 40 zinc–metalloproteinase complexes using the most popular available scoring functions and found $R^2 = 0.58$ ($r = 0.76$) to be the best correlation obtained between experimental and predicted binding affinities.[52]

In this study, we report a computational protocol for a reliable prediction of the binding affinities of zinc containing metalloprotein–ligand complexes. We have examined 2 different charge derivation methods for the ligand and the zinc ion, 2 different van der Waals parameters for the zinc ion, 3 different dielectric treatments in the energy minimization for complex preparation, and 9 different electrostatic treatments in the scoring function, adding to a total of 108 different protocols and arrive at the most theoretically consistent computational protocol that gives the best correlation for the predicted binding affinities of zinc containing metalloprotein–ligand complexes against experiment. The protocol presented has been validated on a heterogenous dataset of 90 complexes comprising five unique targets and is fast enough to be used in virtual screening protocols.

MATERIALS AND METHODS

Scoring Function

The scoring function employed here considers the nonbonded energy of a protein–ligand complex as a sum of electrostatics, van der Waals, hydrophobicity, and loss in conformational entropy of protein side chains and has already been validated on a dataset of 161 nonmetallo protein–ligand complexes giving a correlation of $R^2 = 0.85$ ($r = 0.92$) between the experimental and the predicted binding affinities.[53,54]

$$\Delta G^0 = \alpha(E_{\mathrm{Tel}}) + \beta(E_{\mathrm{Tvdw}}) + \sum_{A=1}^{22} \sigma_A \Delta A_{\mathrm{LSA}} + \lambda(\Delta S_{\mathrm{CR}}) + \delta \quad (1)$$

A zinc containing metalloprotein–ligand complex is divided into three parts; the zinc ion (z), the protein (p), and the ligand (l). The ligand (l) here refers to the small molecule in the active site of the metalloprotein, which is in noncovalent interaction with the protein residues and forms one or two coordinate bonds with the zinc ion. The total electrostatics (E_{Tel}) and total van der Waals (E_{Tvdw}) are computed as a combination of [I] + [II] described later. Electrostatic energy is computed via Coulomb's law using a dielectric function, while the van

der Waals energy is computed using a (12, 6) Lennard-Jones potential.

[I] (p − l) is the energy between the protein (p) and the ligand (l) atoms excluding the zinc ion (z). These are denoted as electrostatics I (el_{p-l}) and van der Waals I (vdw_{p-l}).

[II] (z − pl) is the energy between the zinc ion (z) and the rest of the complex, that is protein and ligand (pl). These are denoted as electrostatics II (el_{z-pl}) and van der Waals II (vdw_{z-pl}). The subscript pl here refers to all those atoms of the protein (p) and the ligand (l) that are not coordinately bonded to the zinc ion, that is all the protein and ligand atoms which are at a distance >2.7 Å from the zinc ion. We have adopted the nonbonded model from the work of Stote and Karplus[23] to model the interactions (electrostatics and van der Waals) of the zinc ion with the protein and the ligand atoms (z − pl).

$$E_{\mathrm{Tel}} = \sum \mathrm{el}_{p-l} + \mathrm{el}_{z-pl}$$

$$E_{\mathrm{Tvdw}} = \sum \mathrm{vdw}_{p-l} + \mathrm{vdw}_{z-pl}$$

The hydrophobic contribution to binding is computed using a modified version of Eisenberg-Mclachlan[55] model. The limitation of five basic atom types in the Eisenberg-Mclachlan model is overcome by combining the atom types in AMBER force field[56] for proteins/nucleic acids with the atom types in GAFF[57] force field for small molecules. This gives a common set of 22 atom types (Table VI) with the advantage that any atom of the protein or ligand can be defined using this set ensuring transferability of derived parameters for organic and biological molecules. For the present study, we have introduced a new atom type for the zinc ion with its cavity radius adopted from the work of Rashin and Honig.[58] The zinc ion is considered as a part of the protein during the surface area calculations. Also in our approach, we consider the loss in surface area of the individual atoms upon binding, reflecting the changes in binding process. Total surface area of an atom type is obtained by summing up all the contributions from that atom type. The net loss in surface area of an atom type upon binding is computed as:

$$\Delta A_{\mathrm{LSA}} = \sum A_{\mathrm{complex}} - \sum A_{\mathrm{protein}} - \sum A_{\mathrm{ligand}}$$

where, ΔA_{LSA} is the net loss in surface area of an atom type A. A_{complex}, A_{protein}, and A_{ligand} are the total surface areas of atom type A in the complex, the protein, and the ligand respectively.

ΔS_{CR} is the energy contribution due to loss in conformational entropy of protein side chains[59] upon ligand binding. We have utilized an empirical scale of side chain conformational entropy ($-T\Delta S$) developed by Pickett and Sternberg[60] and relative accessibility as a measure of the loss in conformational entropy of protein side chains upon protein–ligand binding. Computational details for hydrophobicity and ΔS_{CR} have been explained earlier.[53]

TABLE II. Some Physicochemical Properties With Their Observed Limits in the 90 Zinc Containing Metalloprotein–Ligand Complex Dataset Considered in This Study

S. no.	Descriptor/physicochemical property	Limits
Ligand		
1.	Number of rotatable bonds	1–17
2.	Hydrogen bond donors	0–8
3.	Hydrogen bond acceptors	1–10
4.	Net charge on ligand	0 – (−)2
5.	Molecular weight (Daltons)	74–1989
6.	Total number of atoms	9–71
Protein		
7.	Number of unique proteins	5
8.	Number of residues	242–374
Zinc		
9.	Net charge on zinc binding motif	(−)1 – (+)2
10.	Net charge on zinc	(+)0.19 – (+)1.53
11.	Number of atoms coordinate bond to zinc	3–5
Complex		
12.	Experimental binding affinity (kcal/mol)	(−)19.03 – (−)2.88
13.	Resolution (Å)	1.4–2.9

ΔG^0 in Eq. (1) is the standard free energy of binding and α, β, σ, and λ are the regression coefficients for the electrostatics, van der Waals, hydrophobicity, and entropy terms respectively obtained via a multiple linear regression analysis with the scoring function. E_{Tel}, E_{Tvdw}, ΔA_{LSA}, and ΔS_{CR} serve as independent variables and the experimental binding free energy (ΔG^0) serves as a dependent variable.

Dataset Description

There are about 3500 proteins, complexed with ligands, substrate, prosthetic groups, and metal ions in the protein databank (RCSB).[61] For the present study, we focused on the zinc containing metalloprotein–ligand complexes for which experimental binding free energies are available in the public domain databases such as LPDB,[62] PLD,[63] AffinDB,[64] and PDBbind[65] and prepared a dataset of 90 complexes (Tables IA and IB, supplementary information) as described in the dataset preparation section. A description of the dataset with the observed limits for various descriptors/physicochemical properties is given in Table II. Table II shows that the dataset in consideration is heterogeneous enough with respect to the ligand, protein, and complex descriptors/physicochemical properties to facilitate a rigorous evaluation of the performance of the proposed protocol and its extensions to other systems.

We have analyzed 90 zinc containing metalloprotein–ligand complexes, comprising five unique targets: 35 CA, 7 CPA, 32 MMP, 11 TL, and 5 AD systems. In each complex, zinc is coordinately bonded to amino acid residues and ligand with a distinct coordination geometry, called here, the zinc binding motif (ZBM) (see Fig. 1). The Zn

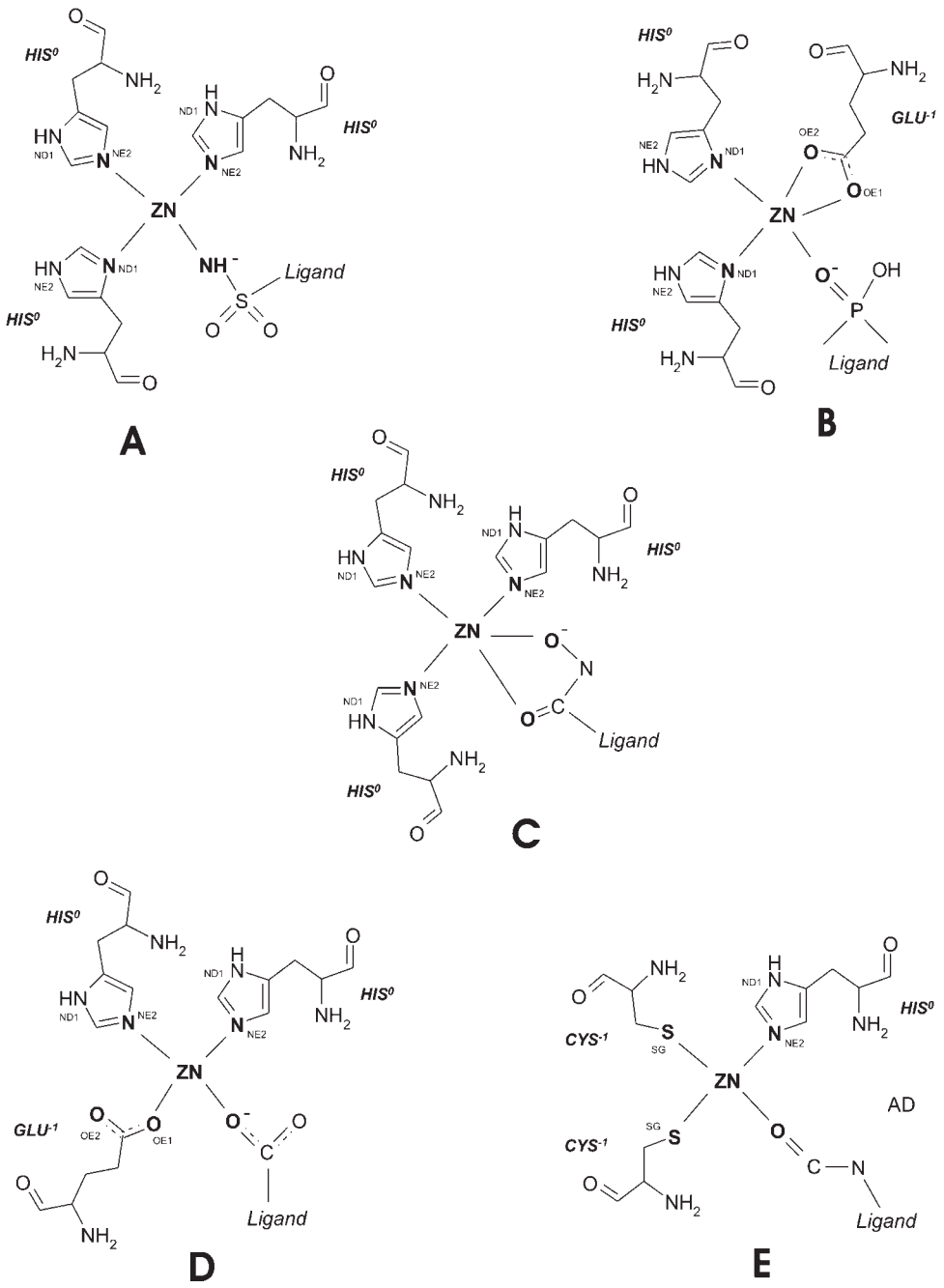

Fig. 1. ZBMs for the five categories of metalloproteins studied. **A**: CA (carbonic anhydrase). **B**: CPA (carboxypeptidase A. **C**: MMP (matrix metalloproteinase. **D**: TL (thermolysin). **E**: AD (alcohol dehydrogenase). The functional group/atom of the ligand that coordinates to the zinc ion is highlighted.

binding motif for CA is Zn-HIS-HIS-HIS-DRG, for CPA it is Zn-HIS-HIS-GLU-DRG, for MMP it is Zn-HIS-HIS-HIS-DRG, for TL it is Zn-HIS-HIS-GLU-DRG, and for AD it is Zn-HIS-CYS-CYS-DRG, where DRG refers to the ligand. Although (CA and MMP) and (CPA and TL) have similar ZBMs in terms of amino acid residues, their atomic positions of HIS (ND1/NE2) where the zinc coordination takes place makes them distinct. In Table III, we summarize the ZBM for all the five categories of proteins in terms of amino acid residues, their atomic positions (NE2/ND1 atom of HIS, OE1/OE2 atom of GLU, and SG atom of CYS), net charge on each residue, and the ligand atoms that are in coordinate bond with the zinc ion. HIS with its dual protonation states is the most abundant amino acid found in ZBM when compared to GLU and CYS. Zinc occurs in four or five coor-

TABLE III. A Description of the Zinc Binding Motif (ZBM) for the Five Categories of Zinc Containing Metalloprotein–Ligand Complexes Studied

	Coordinating residue 1			Coordinating residue 2			Coordinating residue 3			Coordinating residue 4 (ligand)
Protein	Amino acid	Atom type	Net charge	Amino acid	Atom type	Net charge	Amino acid	Atom type	Net charge	Atoms
CA	HIS	NE2		HIS	NE2		HIS	ND1		N
CPA	HIS	ND1	0	HIS	ND1	0	GLU	OE1/OE2/ OE1 and OE2	−1	N/O
MMP	HIS	NE2	0	HIS	NE2	0	HIS	NE2	0	O1/O2/O1 and O2/N/S
TL	HIS	NE2	0	HIS	NE2	0	GLU	OE1/OE2	−1	O1/O2/O1 and O2/S/S and O
AD	HIS	NE2	0	CYS	SG	−1	CYS	SG	−1	O

dinate geometry with the protein and ligand atoms N, O, and S within a distance of 1.8–2.7 Å. In zinc containing metalloprotein–ligand complexes, ligand replaces water molecule and coordinate bonds to zinc with a specific functional group in a monodentate or bidentate fashion with one or two contacts. Generally, these functional groups are hydroxamate, carboxylate, hydrazide, sulfonamide, phosphinate, thiol, sulfodiimines, and so forth (see Fig. 1). Apart from the coordinate bonding with the zinc ion, there occurs a strong hydrogen bond interaction of these functional groups with the amino acids in the ZBM and neighboring residues.

Dataset Preparation

Figure 2 describes a general protocol for the preparation of a metalloprotein–ligand complex in a force field compatible manner. The protocol is divided into the following steps:

1. *Selection of the complex*: X-ray coordinates of the complex are extracted from the RCSB[61] and crystallographic water molecules are removed.
2. *Derivation of partial atomic charges for the ligand and the zinc ion*: In all the zinc containing metalloprotein–ligand complexes, the ligand is coordinately bonded to the zinc ion with one or two coordinate bonds. Zinc ion has a formal charge of +2, but due to the charge transfer from the amino acids and ligand to the zinc cation, net charge on the zinc ion is always less than +2 as shown.[21] The protein residues within the coordinate bond distance ($\leq$2.7 Å) of the zinc ion in each complex are identified. To the resulting amino acid residues, ligand, and the zinc ion motif (ZBM) hydrogen atoms are added. Atoms in the motif, which are within 2.7 Å distance from the zinc ion, are kept deprotonated. If an atom has two hydrogens connected to it, such as NH_2, then the hydrogen that is nearer to the zinc ion is removed. To determine the net charge on the ZBM, we adopted a simple formula:

$$\text{Net charge on the ZBM} = \text{formal charge on each amino acid residue} + \text{formal charge on the ligand} + 2(+) \text{ charge on the zinc ion.}$$

   Two types of charge derivation protocols have been tested in this work. In the first protocol, HF/6-31G* ab initio level calculations are performed on the motif using GAMESS[66] to obtain the Mulliken charges (MULL) for the ligand and the zinc ion. In the second protocol, RESP fitting is applied on the electrostatic potentials obtained from the HF/6-31G* ab initio calculations to obtain the equivalent partial atomic charges for the ligand atoms and the zinc ion.
3. *Parameter assignment for the protein, the ligand, and the zinc ion*: van der Waals parameters for the zinc ion are adopted from the work of Stote and Karplus[23] [σ = 1.95 Å, ε = 0.25 (kcal/mol)] and Hoops et al.[21] [σ = 1.1 Å, ε = 0.0125 (kcal/mol)] and for the ligand atoms from the GAFF force field.[57] Bonded parameters for the ligand are also adopted from the GAFF force field. AntechAMBER[67] module of AMBER 8[68] is used to assign the bonded and the nonbonded parameters to the ligand atoms. Assignment of force field parameters for the protein atoms (RESP derived partial atomic charges, van der Waals, and bonded parameters) is carried out using the AMBER force field.[56] Hydrogen atoms are added to the protein and the protonation states of the charged residues inside the active site are fixed based on the literature for each complex. The protonation state information was also verified from LIGPLOT[69] of the noncovalent interactions involving ligand (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/).
4. *Energy minimization of the complex:* Energy minimization of only the hydrogen atom positions using dielectrics of 1, 80, and sigmoidal is performed with AMBER 8[68] suite of programs to remove any clashes from the structure, since an all atom energy minimization resulted in the loss of coordination geometry of the ZBM in the active site. For

Crystal Structure of the Zinc Containing Metalloprotein-Ligand Complex
RCSB (http://www.rcsb.org/pdb/)

↓

Parameterization of the Ligand and the Zinc
Building Zinc Binding Motif (ZBM)
Assignment of Ionization States of Amino Acid Residues & Ligand
Hydrogen Atom Addition
HF/6-31G*/RESP Charge Derivation
Force Field Parameter Assignment

↓

Parameterization of Protein
Assignment of Protonation States
Hydrogen Atom Addition
Force Field Parameter Assignment

↓

Energy Minimization of the Protein-Ligand-Zinc Complex

↓

Estimation of Binding Affinity
(www.scfbio-iitd.res.in/software/drugdesign/bapplz.jsp)

Fig. 2. A computational flowchart adopted for estimating the binding affinities of zinc containing metalloprotein–ligand complexes.

minimizations, 1000 steps of steepest descent and 1500 steps of conjugate gradient are carried out.

An organizational flowchart is shown in Figure 3, which describes the different force field parameters and dielectric treatments tested in this work. The ligand and the zinc ion in the complex are assigned RESP or Mulliken derived partial atomic charges, resulting in two replicas with different charges for each complex. Protein atoms are assigned RESP charges from the AMBER force field. Each complex is then assigned Hoops et al. or Stote and Karplus van der Waals parameters for the zinc ion, resulting in four replicas for each complex. Protein atoms are assigned AMBER and ligand atoms GAFF force field van der Waals parameters. The four replicas generated for each complex differ in the force field parameters assigned. All the four replicas of each complex are now energy minimized with three different dielectric models (1, 80, and sigmoidal), resulting in 12 different structures for each complex ($1 \times 2 \times 2 \times 3 = 12$). These structures act as an input for the scoring function for binding affinity estimations.

The scoring function has two electrostatics terms, electrostatics I (p − l) and electrostatics II (z − pl), and each term incorporates a dielectric function to calculate the electrostatic energy. Three different dielectric models (1, 80, and sigmoidal) have been tested for electrostatics I and electrostatics II, resulting in nine different models of the scoring function. Binding affinity calculations of 12 energy minimized structures with 9 different models of the scoring function gives a total of ($12 \times 9 = 108$) 108 different computational protocols for binding affinity estimations for each zinc containing metalloprotein–ligand complex (system). This process was repeated for all the 90 systems considered, giving 108 different sets of binding affinities for each system. This enabled identification of the most suitable and theoretically consistent computational protocol for predicting the binding affinities of zinc containing metalloprotein–ligand complexes.

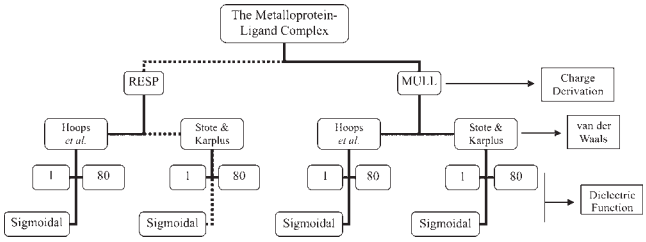

Fig. 3. An organizational flowchart showing the different force field parameters and dielectric functions employed in preparing the zinc containing metalloprotein–ligand complex. The highlighted dotted line path shows the most successful protocol.

TABLE IV. Correlation Coefficient (r) Obtained Between the Predicted and Experimental Binding Affinity for the 16 Different Sets of Methods Tested on the 90 Complex Dataset

Method	Charge derivation	Zn Van der Waals parameters	Minimization dielectric	Electrostatic I dielectric	Electrostatic II dielectric	r
I	RESP	Hoops et al.	Sigmoidal	Sigmoidal	Sigmoidal	0.80
II	RESP	Stote and Karplus	Sigmoidal	Sigmoidal	Sigmoidal	0.88
III	RESP	Hoops et al.	1	Sigmoidal	Sigmoidal	0.80
IV	RESP	Stote and Karplus	1	Sigmoidal	Sigmoidal	0.86
V	RESP	Hoops et al.	Sigmoidal	Sigmoidal	80	0.81
VI	RESP	Stote and Karplus	Sigmoidal	Sigmoidal	80	0.84
VII	RESP	Hoops et al.	1	Sigmoidal	80	0.82
VIII	RESP	Stote and Karplus	1	Sigmoidal	80	0.84
IX	Mulliken	Hoops et al.	Sigmoidal	Sigmoidal	Sigmoidal	0.78
X	Mulliken	Stote and Karplus	Sigmoidal	Sigmoidal	Sigmoidal	0.83
XI	Mulliken	Hoops et al.	1	Sigmoidal	Sigmoidal	0.80
XII	Mulliken	Stote and Karplus	1	Sigmoidal	Sigmoidal	0.74
XIII	Mulliken	Hoops et al.	Sigmoidal	Sigmoidal	80	0.85
XIV	Mulliken	Stote and Karplus	Sigmoidal	Sigmoidal	80	0.87
XV	Mulliken	Hoops et al.	1	Sigmoidal	80	0.85
XVI	Mulliken	Stote and Karplus	1	Sigmoidal	80	0.91

RESULTS AND DISCUSSION

Validation of the Scoring Function and the Computational Protocol

A multiple linear regression analysis was performed on the 108 different sets of binding affinities and each set contains information on all the 90 complexes considered.

1. The complexes energy minimized with a dielectric of 80 (36 sets) had a correlation (r) < 0.5 for the predicted against the experimental binding affinity.
2. The complexes for which the binding affinity was computed using a dielectric function of 1 or 80 for the electrostatics I (p − l) (48 sets) had a correlation (r) < 0.6 for the predicted against the experimental binding affinity.
3. The complexes for which the binding affinity was computed using a dielectric function of 1 for the electrostatics II (z − pl) (eight sets) had a correlation (r) < 0.7 for the predicted against the experimental binding affinity.

After removing all the above 92 sets which had a correlation (r) < 0.7 for the predicted against the experimental binding affinity, we were left with 16 sets of binding affinities having correlation (r) > 0.7. The results of these 16 sets are presented in Table IV, showing the effects of variation in the computational protocol and the scoring function on the correlation (r). Table IV shows that, RESP and Mulliken (MULL) charges for the ligand and the zinc ion perform equally well in terms of the correlation (r) obtained. However, employing RESP charges for the ligand and the zinc ion makes the methodology more consistent because RESP charges are used for the protein atoms. The van der Waals parameters from Stote and Karplus perform better when compared to Hoops et al. parameters. We found that dielectric of 1 or sigmoidal in minimization, sigmoidal in electrostatics I (p − l), and sigmoidal or 80 in electrostatics II (z − pl) when employed results in correlation (r) > 0.8.

Out of these 16 sets, methods II and XVI yield the highest correlation (r) for the predicted against the experimental binding affinities for the 90 zinc containing metalloprotein–ligand complexes out of the 108 different computational protocols tested in this work. The highest correlation $R^2 = 0.83$ ($r = 0.91$) was obtained for the method XVI where Mulliken charges were used for the ligand and the zinc ion, RESP charges for the protein atoms, Stote and Karplus van der Waals parameters for the zinc ion, dielectric of 1 (vacuum) in energy minimization, sigmoidal dielectric for electrostatic I (p − l), and dielectric of 80 for electrostatic II (z − pl). Despite the high correlation, this methodology is not consistent in terms of the partial atomic charges and dielectric function employed for electrostatic calculations.

The most consistent methodology is seen to be method II, which yields a good correlation of $R^2 = 0.77$ ($r = 0.88$). Method II employs RESP charges for the protein, ligand, and the zinc ion, Stote and Karplus van der Waals parameters for the zinc ion and sigmoidal dielectric for energy minimization, electrostatics I and electrostatics II. In this work, we propose method II as a theoretically consistent protocol for predicting the binding affinities of zinc containing metalloprotein–ligand complexes.

Model validation is a crucial aspect of any model development technique and establishes its predictive power. Recent studies[70,71] have shown that, in addition to leave-one-out (LOO) cross-validation (q^2) procedure, validation of the model using an external test set of compounds is necessary. For a robust validation, the training and test sets must have a uniform distribution of

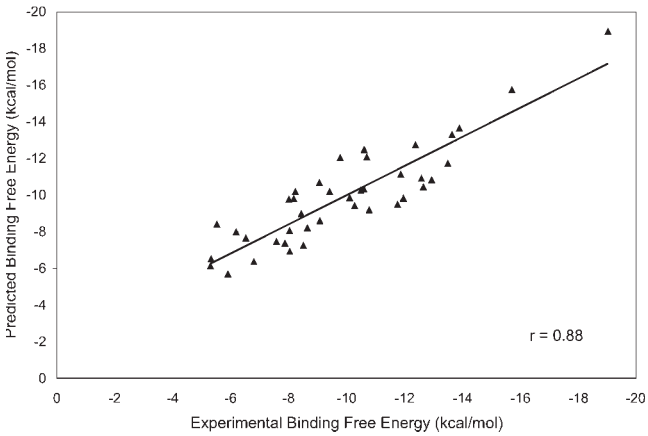

Fig. 4. Correlation between the predicted and experimental binding free energies for 40 zinc containing metalloprotein–ligand complexes for method II (training set).

the representative points in the multidimensional descriptor space. In addition, the model should also satisfy the following conditions:

1. $q^2 > 0.5$
2. $R^2 > 0.6$
3. $\frac{(R^2-R_0^2)}{R^2} < 0.1$ and $0.85 \leq K \leq 1.15$
4. $\frac{(R^2-R_0'^2)}{R^2} < 0.1$ and $0.85 \leq K' \leq 1.15$
5. $|R_0^2 - R_0'^2| < 0.3$

All the above equations have been explained in Table II of the supplementary information.

Keeping these issues in consideration, we started with the LOO cross validation procedure to make the training and test sets of the prepared zinc containing metalloprotein–ligand complexes via method II. We used the experimental binding free energy of the complexes as a descriptor for their uniform distribution across multidimensional descriptor space in the training and test sets. The dataset is separated into a training set of 40 complexes and a test set of 50 complexes (Tables IA and IB, Supplemantary Information). The training set of 40 zinc containing metalloprotein–ligand complexes for method II gave a correlation coefficient $R^2 = 0.77$ ($r = 0.88$) for the predicted against the experimental binding affinities (see Fig. 4). A graphical residual analysis plot [Fig. S(I) of the supplementary information] of the standardized residuals against the experimental binding affinities for the training set shows a uniform distribution of the points above and below the base line, suggesting that the model fits the data well. The five statistical tests defined earlier in addition to S_{PRESS} and RMS_{error} were then performed on the training set of method II. The results shown in Table V indicate that the scoring function passes all the validation tests. The final validation was performed on the external test set of 50 complexes using the parameters obtained from the training set (Table VI). A correlation coefficient of $R^2 = 0.77$ ($r = 0.88$) for method II was obtained on the test set (see Fig. 5) between the experimental and the predicted binding free energies, indicating the robustness of the computational protocol, the scoring function, and the regression coefficients obtained in predicting the binding affinities of zinc containing metalloprotein–ligand complexes. We further tested the ability of the scoring function in the prediction of relative binding affinities of a series of ligands against the same protein target. From the 50 test set, we selected CA and MMP, which have more than six distinct ligands. Individual correlation studies on these groups of complexes show an average correlation coefficient of $R^2 = 0.77$ ($r = 0.88$). The training and the test set PDB IDs of the complexes for method II along with their experimental and predicted binding free energies and component-wise separation of the energetics are provided in Tables IA and IB of the supplementary information.

TABLE V. Statistical Tests and the Values for the Corresponding Indices for the Training Set for Method II

Statistical test	Method II
q^2	0.77
R^2	0.77
$\frac{(R^2-R_0^2)}{R^2}$	−0.24
$\frac{(R^2-R_0'^2)}{R^2}$	−0.24
K	1
K'	0.98
$\lvert R_0^2 - R_0'^2 \rvert$	0.0025
S_{PRESS} (kcal/mol)	2.34
RMS error (kcal/mol)	±63.1

Empirical Parameter (Regression Coefficient) Analysis

The empirical scoring function for predicting the binding affinities of zinc metalloproteinases has 25 independent variables (electrostatics, van der Waals, loss in conformational entropy, and 22 atom types for hydrophobicity corresponding to a combined GAFF[57] and AMBER force field[56]) and therefore 25 empirical parameters (Table VI). Figure S(II) (supplementary information) gives a percentage wise occurrence of each variable in the 90 dataset. Of the 21 atom types (excluding zinc) C3, C4, C5, H1, H2, H3, N1, N2, O1, and O2 occur in more than 90% of the complexes. HL, N5, N6, and O3 are present in less than 50% of the complexes. C2, N3, and P are present in very few complexes (less than 10%).

Although the coefficients are empirical (Table VI), the model is phenomenological and is in accord with the thermodynamics of protein–ligand binding. The calculated E_{Tel} and E_{Tvdw} have negative signs and their regression coefficients α and β have positive signs respectively, indicating a net favorable contribution of electrostatics and van der Waals towards binding. The calculated ΔS_{CR} has a positive sign as expected and its regression coefficient has negative sign. The result is that the net loss in conformational entropy of protein side chains is either zero or marginally favorable to

TABLE VI. A Description of the 22 Derived Atom Types With Their Atomic Desolvation Parameters kcal/mol/Å^2 (ADP)

S. No.	Atom type symbol	Description	Parameters method II
1.	C1	sp^2 carbonyl	0.1224
2.	C2	sp carbon	0.0583
3.	C3	sp^2 carbon aliphatic	0.0168
4.	C4	sp^2 carbon aromatic	−0.0074
5.	C5	sp^3 carbon	−0.2971
6.	HL	Halogens (Fl, Cl, Br, I)	0.0018
7.	H1	Hydrogen bonded to aliphatic carbon	0.0008
8.	H2	Hydrogen bonded to aromatic carbon	0.0030
9.	H3	Hydrogen bonded to nitrogen	−0.0289
10.	H4	Hydroxyl group	−0.0072
11.	N1	sp^2 nitrogen in amide groups	0.0113
12.	N2	sp^2 nitrogen in aliphatic systems	0.0300
13.	N3	sp^2 nitrogen in aromatic systems	0.0037
14.	N4	sp nitrogen	−0.0056
15.	N5	sp^3 nitrogen	0.0288
16.	N6	Amine nitrogen connected to one or more aromatic rings	0.0111
17.	O1	Oxygen with one connected atom	0.0007
18.	O2	Oxygen in hydroxyl group	−0.0088
19.	O3	Ether and ester oxygen	0.0204
20.	P	Phosphate	0.6386
21.	S	Sulphur	0.0081
22.	Zn	Zinc	0.0139
23.	α	Empirical coefficient for electrostatics	0.2197
24.	β	Empirical coefficient for van der Waals	0.1984
25.	λ	Empirical coefficient for conformational entropy	−0.2916
26.	δ	Constant (Intercept)	0.3292

Empirical regression coefficients for electrostatics (α), van der Waals (β), conformational entropy (λ), and the regression constant (δ) are also given for method II.

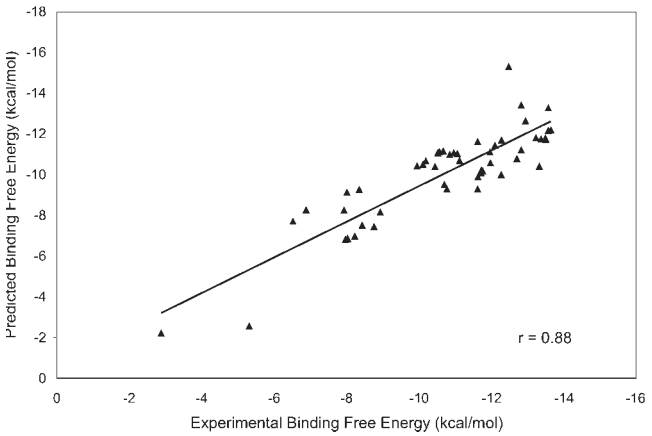

Fig. 5. Correlation between the predicted and experimental binding free energies for 50 zinc containing metalloprotein–ligand complexes for method II (test set).

binding. The loss in surface area of all the 22 atom types has a negative sign, indicating that the net loss in surface area is favorable for binding. However, the atomic desolvation parameters (regression coefficients) for these atom types have different contributions. Aromatic sp^2 carbon, sp^3 carbon, hydrogen bonded to nitrogen atom, hydroxyl group hydrogen, sp nitrogen, and oxygen in hydroxyl group have a negative desolvation parameter and all the remaining atom types have a positive desolvation parameter, which shows that their desolvation is favorable for binding. Nitrogen and phosphorous have a negative sign for the desolvation parameters, indicating that their desolvation is unfavorable for binding.

Component-Wise Analysis

A component-wise analysis of the predicted binding affinities obtained using method II was performed for all the 90 zinc containing metalloprotein–ligand complex dataset [Fig. S(III), supplementary information]. Figure S(III) shows the contribution of each component and their additive sums and their effect on the correlation. Individually, van der Waals term shows the maximum correlation of $R^2 = 0.45$ ($r = 0.67$) against the experimental binding free energies. This suggests that structural complementarity/packing in particular is an essential prerequisite for specific binding. Adding electrostatic contribution to van der Waals component further increases the correlation to $R^2 = 0.64$ ($r = 0.8$), suggesting the importance of hydrogen bonding/ionic interactions in providing specificity to the complex formation Adding hydrophobicity contribution to this further increases the correlation to $R^2 = 0.74$ ($r = 0.86$), reflect-

ing the importance of hydration effects in protein–ligand binding. Adding the loss in conformational entropy increases the correlation to $R^2 = 0.77$ ($r = 0.88$), reflecting the contribution of loss of protein side chain conformation upon ligand binding. The above results show the order of importance of each component to the binding affinity.

Web Server

The empirical/force field based scoring function (1) together with method II described and analyzed above has been web enabled for free access at www.scfbio-iitd.res.in/software/drugdesign/bapplz.jsp as Binding Affinity Prediction of Protein–Ligand complex containing Zinc (BAPPL-Z) server, to aid in designing novel ligands for zinc–metalloproteinases. Input to the server is an energy minimized zinc containing metalloprotein–ligand complex, minimized using a sigmoidal dielectric. The input complex has hydrogens added and protonation states assigned. RESP charges are assigned to the protein, ligand and the zinc ion. Stote and Karplus van der Waals parameters are assigned to the zinc ion. AMBER and GAFF force field van der Waals parameters are assigned to the protein and the ligand atoms respectively. The server directly computes the binding affinity of the input complex. Further work on automation of the structure preparation and parameter assignment is in progress. The energy minimized atomic coordinates of 90 zinc containing metalloprotein–ligand complexes prepared using method II along with all the parameters for binding affinity estimates are also made accessible at the website at www.scfbio-iitd.res.in/software/drugdesign/bapplzdataset.jsp.

CONCLUSIONS

Often metalloproteinases are included with nonmetalloprotein complexes in the validation of scoring functions, but have generally shown a poor agreement with experiment. Designing a methodology that yields a good correlation of the predicted binding affinity with the experiment on a large dataset of zinc containing metalloprotein–ligand complexes has been a challenging task. We propose here an empirical free energy function comprising contributions from electrostatics, van der Waals, hydrophobicity, and loss in conformational entropy of protein side chains. We have examined the sensitivity of the predicted binding affinities to different choices of nonbonded van der Waals parameters for zinc ion (Stote and Karplus, Hoops et al.), partial atomic charges for the ligand and the zinc ion (Mulliken, RESP), and dielectric treatments (1, 80, and sigmoidal) in energy minimization and in calculating the electrostatic contribution to the binding free energy. Based upon the theoretical consistency of the protocol and the accuracy of the results, we propose in this work a computational methodology for predicting binding affinities of zinc containing metalloprotein–ligand complexes. The proposed method gives a correlation coefficient $R^2 = 0.77$ ($r = 0.88$) for the predicted binding affinities against the experiment. Heterogenity of the dataset on which the protocols have been validated and parameters obtained promises transferability to systems from different families of zinc metalloproteins, with distinct zinc coordination geometries, different active sites, and a variety of ligand architectures. An average correlation coefficient of $R^2 = 0.77$ ($r = 0.88$) for different ligands against the same target indicates the ability of the protocol and the scoring function in predicting relative binding affinities of ligands. The results suggest that, building ZBM to derive partial atomic charges for the ligand and the zinc ion, the nonbonded approach to model zinc ion, correct protonation states for ligand and protein residues in the active site, compatibility between the parameters obtained from GAFF force field for ligand, Stote and Karplus van der Waals parameters for zinc ion and AMBER force field for proteins, the dielectric function employed, the desolvation parameters for each atom type, and the energy minimization protocol are some of the important issues which have strengthened the protocol and the scoring function in obtaining a good correlation between the experimental and the predicted binding affinities of zinc–metalloproteinase complexes from "single-point" calculations. Model validation, various statistical tests, and parameter analysis studies underscore the predictive ability of the scoring function. The proposed computational methodology is simple and fast and can be easily implemented in structure-based drug design to design novel ligands binding to zinc metalloproteins.

ACKNOWLEDGMENT

The authors thank Mr. Shailesh in preparing the web tool.

REFERENCES

1. Frausto da Silva JJR, Williams RJP. The biological chemistry of the elements. Oxford: Oxford University Press; 1991.
2. Vallee BL, Auld DS. Functional zinc-binding motifs in exzymes and DNA-binding proteins. Faraday Discuss 1992;93:47–65.
3. Overall CM, Lopez-Otin C. Strategies for MMP inhibition in cancer: innovations for the post-trial era. Nat Rev Cancer 2002;2:657–672.
4. Christianson DW, Fierke CA. Carbonic anhydrase: evolution of the zinc binding site by nature and by design. Acc Chem Res 1996;29:331–339.
5. Quiocho FA, Lipscomb WN. Carboxypeptidase A: a protein and an enzyme. Adv Protein Chem 1971;25:1–49.
6. Branden CI, Eklund H, Nordstrom B, Boiwe T, Soderlund G, Zeppezauer E, Ohlsson I, Åkeson Å. Structure of liver alcohol dehydrogenase at 2.9Å resolution. Proc Natl Acad Sci USA 1973;70:2439–2442.
7. Rao GB. Recent developments in the design of specific matrix metalloproteinase inhibitors aided by structural and computational studies. Curr Pharm Des 2005;11:295–322.
8. Matthews BW. Structural basis of the action of thermolysin and related zinc peptidases. Acc Chem Res 1988;21:333–340.
9. Lipscomb WN, Strater N. Recent advances in zinc enzymology. Chem Rev 1996;96:2375–2434.
10. Yamashita MM, Wesson L, Eisenman G, Eisenberg D. Where metal ions bind in proteins. Proc Natl Acad Sci USA 1990;87:5648–5652.

11. Dudev T, Lim C. Principles governing Mg, Ca and Zn binding and selectivity in proteins. Chem Rev 2003;103:773–787.
12. Alberts IL, Nadassay K, Wodak SJ. Analysis of zinc binding sites in protein crystal structures. Protein Sci 1998;7:1700–1716.
13. van der Vaart A, Merz KM, Jr. The role of polarization and charge transfer in the salvation of biomolecules. J Am Chem Soc 1999;121:9182–9190.
14. Sakharov D, Lim C. Zn protein simulations including charge transfer and local polarization effects. J Am Chem Soc 2005; 127:4921–4929.
15. Banci L. Molecular dynamics simulations of metalloproteins. Curr Opin Chem Biol 2003;7:143–149.
16. Brothers EN, Suarez D, Deerfield DW, II, Merz KM, Jr. PM3-compatible zinc parameters optimized for metalloenzyme active sites. J Comput Chem 2004;25:1677–1692.
17. Kuntz ID. Structure-based strategies for drug design and discovery. Science 1992;257:1073–1082.
18. Latha N, Jayaram B. A binding affinity based computational pathway for active-site directed lead molecule design: some promises and perspectives. Drug Design Rev Online 2004;2: 145–165.
19. Vedani A, Huhta DW. A new force field for modeling metalloproteins. J Am Chem Soc 1990;112:4759–4767.
20. Vedani A, Dobler M, Dunitz JD. An empirical potential function for metal centers: application to molecular mechanics calculations on metalloproteins. J Comput Chem 1986;7:701–710.
21. Hoops SC, Anderson KW, Merz KM, Jr. Force field design for metalloproteins. J Am Chem Soc 1991;113:8262–8270.
22. Ryde U. Molecular dynamics simulations of alcohol dehydrogenase with a four- or five-coordinate catalytic zinc ion. Proteins: Struct Funct Genet 1995;21:40–56.
23. Stote RH, Karplus M. Zinc binding in proteins and solution: a simple but accurate nonbonded representation. Proteins: Struct Funct Genet 1995;23:12–31.
24. Makinen MW, Troyer JN, van der Werff H, Berendsen JC, van Gunsteren WF. Dynamical structure of carboxypeptidase A. J Mol Biol 1989;207:210–216.
25. Bredenberg J, Nilsson L. Modeling zinc sulfhydryl bonds in zinc fingers. Int J Quantum Chem 2001;83:230–244.
26. Pang YP. Novel zinc protein molecular dynamics simulations: steps towards antiangiogenesis for cancer treatment. J Mol Model 1999;5:196–202.
27. Dudev T, Lim C. Metal selectivity in metalloproteins: Zn^{2+} vs. Mg^{2+}. J Phys Chem B 2001;105:4446–4452.
28. Lin Y, Lim C. Factors governing the protonation state of Zn-bound histidine in proteins: a DFT/CDM study. J Am Chem Soc 2004;126:2602–2612.
29. Garmer DR, Gresh N, Roques BP. Modeling of inhibitor-metalloenzyme interactions and selectivity using molecular mechanics grounded in quantum chemistry. Proteins: Struct Funct Genet 1998;31:42–60.
30. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004;3:935–949.
31. Ajay, Murcko MA. Computational methods to predict binding free energy in ligand–receptor complexes. J Med Chem 1995;38: 4953–4967.
32. Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptor. Angew Chem Int Ed 2002;41:2644–2676.
33. Brandsdal BO, Osterberg F, Almlof M, Feierberg I, Luzhkov VB, Åqvist J. Free energy calculations and ligand binding. Adv Protein Chem 2003;66:123–158.
34. Wang W, Donini O, Reyes CM, Kollman PA. Biomolecular simulations: recent developments in force fields, simulation of enzyme catalysis, protein–ligand, protein–protein and protein–nucleic acid noncovalent interactions. Annu Rev Biophys Biomol Struct 2001;30:211–243.
35. Williams DH, Cox JPL, Doig AJ, Gardner M, Gerhard U, et al. Toward the semiquantitative estimation of binding constants. Guides for peptide-peptide binding in aqueous solution. J Am Chem Soc 1991;113:7020–7030.
36. Vajda S, Weng Z, Rosenfeld R, DeLisi C. Effect of conformational flexibility and solvation on receptor–ligand binding free energies. Biochemistry 1994;33:13977–13988.
37. Gohlke H, Klebe G. Statistical potential and scoring functions applied to protein–ligand binding. Curr Opin Struct Biol 2001;11: 231–235.
38. Raha K, Merz KM, Jr. Large scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein–ligand complexes. J Med Chem 2005;48:4558–4575.
39. Åqvist J, Luzhkov VB, Brandsdal BO. Ligand binding affinities from MD simulations. Acc Chem Res 2002;35:358–365.
40. Kalra P, Reddy TV, Jayaram B. Free energy component analysis for drug design: a case study of HIV-I protease inhibitor binding. J Med Chem 2001;44:4325–4338.
41. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Ann Rev Biophys Biomol Struct 2003;32:335–373.
42. Wang R, Lu Y, Wang S. Comparitive evaluation of 11 scoring functions for molecular docking. J Med Chem 2003;46:2287–2303.
43. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL, III. Assessing scoring functions for protein–ligand interactions. J Med Chem 2004;47:3032–3047.
44. Perola E, Walters PW, Charifson PS. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins: Struct Funct Genet 2004;56:235–249.
45. Donini OAT, Kollman PA. Calculation and prediction of binding free energies for matrix metalloproteinases. J Med Chem 2000; 43:4180–4188.
46. Raha K, Merz KM, Jr. A quantum mechanics based scoring function: study of zinc ion-mediated ligand binding. J Am Chem Soc 2004;126:1020–1021.
47. Toba S, Damodaran KV, Merz KM, Jr. Binding preferences of hydroxamate inhibitors of the matrix metalloproteinase human fibroblast collagenase. J Med Chem 1999;42:1225–1234.
48. Hou T, Zhang W, Xu XJ. Binding affinities for a series of selective inhibitors of gelatinase-A using molecular dynamics with a linear interaction energy approach. J Phys Chem B 2001;105: 5304–5315.
49. Hu X, Shelver WH. Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. J Mol Graph Model 2003;22: 115–126.
50. Rizzo RC, Toba S, Kuntz ID. A molecular basis for the selectivity of thiadiazole urea inhibitors with stromelysin-1 and gelatinase-A from generalized born molecular dynamics simulations. J Med Chem 2004;47:3065–3074.
51. Khandelwal A, Lukacova V, Comez D, Kroll DM, Raha S, Balaz S. A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. J Med Chem 2005;48:5437–5447.
52. Hu X, Balaz S, Shelver WH. A practical approach to docking of zinc metalloproteinase inhibitors. J Mol Graph Model 2004;22: 293–307.
53. Jain T, Jayaram B. An all atom energy based computational protocol for predicting binding affinities of protein–ligand complexes. FEBS Lett 2005;579:6659–6666.
54. Arora N, Jayaram B. Energetics of base pairs in B-DNA in solution: an appraisal of potential functions and dielectric treatments. J Phys Chem 1998;102:6139–6144.
55. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
56. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, et al. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. J Am Chem Soc 1995;117: 5179–5197.
57. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general AMBER force field. J Comput Chem 2004;25:1157–1174.
58. Rashin AA, Honig B. Reevaluation of the Born model of ion hydration. J Phys Chem 1985;89:5588–5593.
59. Lee KH, Xie D, Freire E, Amzel LM. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. Proteins: Struct Funct Genet 1994;20:68–84.
60. Pickett SD, Sternberg MJE. Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol 1993;231:825–839.

61. Berman HM, et al. The protein data bank. Nucleic Acids Res 2000;28:235–242.
62. Roche O, Kiyama R, Brooks CL, III. Ligand–protein database: linking protein–ligand complex structures to binding data. J Med Chem 2001;44:3592–3598.
63. Puvanendrampillai D, Mitchell JBO. Protein ligand database (PLD): additional understanding of the nature and specificity of protein–ligand complexes. Bioinformatics 2003;19:1856–1857.
64. Block P, Sotriffer CA, Dramburg I, Klebe G, Affin DB. A freely accessible database of affinities of protein–ligand complexes from the PDB. Nucleic Acids Res 2006;34:D522–D526.
65. Wang R, Fang X, Lu Y, Yang CY, Wang S. The PDBbind database: methodologies and updates. J Med Chem 2005;48:4111–4119.
66. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, et al. General atomic and molecular electronic structure system. J Comput Chem 1993;14:1347–1363.
67. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model 2006;25:247–260.
68. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheathem JE, III, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun 1995;91:1–41.
69. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. Prot Eng 1995;8:127–134.
70. Golbraikh A, Tropsha A. Beware of q^2! J Mol Graph Model 2002;20:269–276.
71. Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. J Comput-Aided Mol Des 2003;17:241–253.