GemAffinity: A Scoring Function for Predicting Binding Affinity and Virtual Screening

Kai-Cheng Hsu[1], Yen-Fu Chen[1], and Jinn-Moon Yang[123]*

[1] Institute of Bioinformatics, National Chiao Tung University, Hsinchu, 30050, Taiwan
[2] Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, 30050, Taiwan
[3] Core Facility for Structural Bioinformatics, National Chiao Tung University, Hsinchu, 30050, Taiwan
piki.bi96g@nctu.edu.tw
moon@faculty.nctu.edu.tw

*Corresponding author

Abstract—**Prediction of protein-ligand binding affinities is an important issue in molecular recognition and virtual screening. We have developed a scoring function, namely GemAffinity, to predict binding affinities by analyzing 88 descriptors derived from 891 protein-ligand structures selected from the Protein Data Bank (PDB). Based on these 88 descriptors, we derived GemAffinity using a stepwise regression method to identify five descriptors, including van der Waals contact; metal-ligand interactions; water effects; ligand deformation penalties; and highly conserved residues interacting to a bound ligand with hydrogen bonds. GemAffinity was evaluated on an independent set, and the correlation between predicted and experimental values is 0.572. GemAffinity is the best among 13 methods on this set. Our GemAffinity was then applied to virtual screening for thymidine kinase (TK), human carbonic anhydrase II (HCAII), estrogen receptor of antagonists (ER) and agonists (ERA). Experimental results indicate that GemAffinity is able to reduce the disadvantages (i.e. preferring highly polar or high molecular weight compounds) of energy-based scoring functions. In addition, GemAffinity easily combined with other scoring functions to enrich screening accuracies. We believe that GemAffinity is useful to predict binding affinity and virtual screening.**

Keywords-component; binding affinity prediction, scoring functions, protein-ligand interactions, structure-based drug design, virtual screening

I. INTRODUCTION

An important issue in computer-aided drug discovery is to predict binding affinities of compounds for target proteins. The compounds with high binding affinities for a target protein is often considered potential inhibitors that slow or block physiological, chemical, or enzymatic actions of the target protein. Binding affinities can be divided into computational and experimental methods. Experimental methods are often labor-intensive, time-consuming, and expensive. Therefore, many computational methods have been developed to identify lead compounds by predicting binding affinities of compounds for a target protein.

Virtual screening (VS) is effective for identifying lead compounds of a specific receptor from thousands of compounds [1]. Scoring methods of VS should be able to discriminate between correct binding states and non-native docked states during molecular docking phase, and differentiate a small number of active compounds from hundreds of thousands non-active compounds during post-docking analysis. Potential compounds identified by scoring methods will be selected for biological assays.

A scoring function performance is limited by an incomplete understanding of the complex issues involved in chemical interactions. The inaccuracy of scoring methods, *i.e.*, inadequately in predicting the actual binding affinity of a ligand for a receptor, is the primary weakness for identifying lead compounds in virtual screening. For instance, most scoring functions use simple models to handle metal-ligand and water-ligand interactions and do not highlight essential hydrogen-bonding interactions. In addition, scoring functions favor the selection of highly polar compounds because they measure binding affinities through summing up hydrogen-bonding interactions that are easily formed between polar atoms of compounds and residues [2].

Here, we addressed these issues by deriving 88 descriptors from 891 protein-ligand complexes. Based on these 88 descriptors, we derived GemAffinity using a stepwise regression method to identify five descriptors, including van der Waals contact; metal-ligand interactions; water effects; ligand deformation penalties; and highly conserved residues interacting to a bound ligand with hydrogen bonds. Experimental results show that GemAffinity is comparative to 12 methods on a test set. For virtual screening, GemAffinity is able to increase prediction accuracies of GEMDOCK [3] and GOLD [4] by reducing the ill effects (i.e. prefer highly polar and high molecular weight compounds) of energy-based scoring functions.

II. METHODS AND MATERIALS

Fig. 1 presents the overview of GemAffinity for predicting binding affinities.

A. Protein-ligand complex dataset

We collected 989 protein-ligand complexes with experimental binding affinities from the PDBbind [5]. These complexes were randomly divided into a training set (891 complexes) and a testing set (98 complexes). For each complex, we used the negative logarithm of K_d or K_i as the binding affinity.
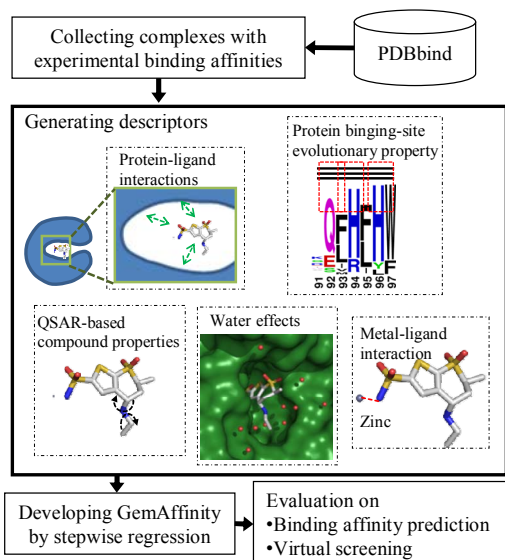
IEEE computer society

Figure 1. An overview of GemAffinity.

B. Descriptors

We derived 88 descriptors from 891 protein-ligand complexes using the following five dimensions: protein-ligand interactions; protein binding-site conserved properties; quantitative structure activity relationship (QSAR)-based compound properties; water effects; and metal-ligand interactions.

Protein-ligand interactions: We considered three protein-ligand interaction types frequently used in energy-based scoring functions—van der Waals (vdW) interactions, electrostatic interactions, and hydrogen-bonding interactions. Piecewise linear potentials (PLP) of GEMDOCK and potential functions (Lennard-Jones and hydrogen-bond potentials) of AutoDock [6] were used to generated descriptors. Furthermore, we also considered the numbers of protein-ligand interactions within different cutoff distances for each force type. For vdW force, the distance of an atom pair between a protein and its ligand was divided into 10 cutoff distances—2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, and 7.0 Å. The distance was divided into 10 cutoff distances (2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, and 3.4 Å) for hydrogen-bonding and electrostatic interactions, respectively. In total, 35 descriptors are extracted for measuring protein-ligand interactions.

Protein binding-site evolutionary property: Conserved residues interacting with bound compounds often play important roles for biological functions. In general, the catalytic residues (e.g., HIS94, HIS96, HIS119, and THR199 in Fig. 2) which polarize substrates and thereby stabilize transition states [7], are conserved. If a conserved residue mutates, the protein may lose its functions or execute different biological functions. Here, we used the number of highly conserved residues forming hydrogen bonds with ligands as a descriptor.

We developed a method, derived from our previous work on protein-protein interactions [8], to measure conversation of residues (Fig. 2) by using PSI-BLAST [9]. The E-value cutoff of PSI-BLAST searching on UniRef90 [10] was set to 10^{-5} and the iteration number was 3. Based on the multiple sequence alignment (Fig. 2a) and the

position-specific scoring matrix (PSSM) generated by PSI-BLAST, the conserved score of each residue is defined as

$$C_i = M_{ir} - K_{rr} \tag{1}$$

where M_{ir} is the value in the PSSM for residue type r at position i, and K_{rr} is the diagonal value of BLOSUM62 [11] for residue type r. The descriptor (N_{conHB}) is defined as

$$R_i = 1 \quad \text{if } C_i > 0 \text{ and number of hydrogen bonds} \geq 1$$

$$N_{conHB} = \sum_i R_i$$

where R_i is 1 when $C_i > 0$ and at least one hydrogen bond is formed at residue i between the protein and its bound ligand; and N_{conHB} is the number of highly conserved residues with hydrogen bonds (Fig. 2b).
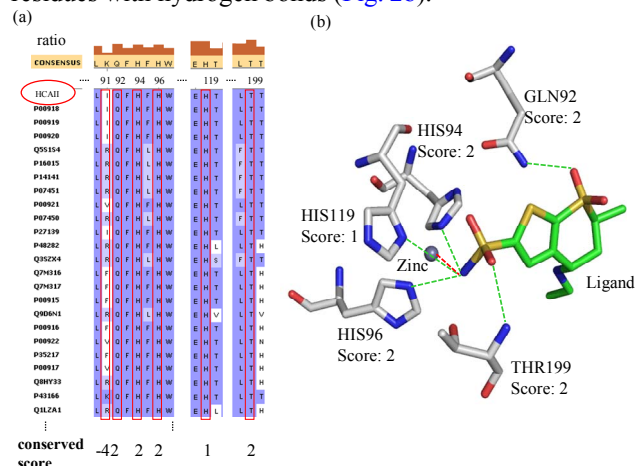

Figure 2. The metal-ligand bonds and highly conserved residues forming hydrogen bonds. (a) The multiple sequence alignment of HCAII. The conserved scores of part residues are given in the bottom. (b) The highly conserved residues form hydrogen bonds (PDB code 1cil). In this example, the number of key interactions is 5. Additionally, the ligand also forms a metal-ligand bond (red line) with Zinc (bond length is 1.96 Å).

QSAR-based compound properties: The QSAR methods [12] showed that physiochemical properties of compounds are useful for predicting the binding affinities. This study used the QSAR module in Cerius2 to generate 23 compound descriptors, such as spatial, conformational, electronic, structural, and thermodynamic terms.

Water effects: Water effects play a crucial role in that they mediate interactions between proteins and their bound ligands [13]. Here, we used the numbers of structural water molecules within specific distance cutoffs around a ligand as descriptors (N_{water}) to measure the water effects. The distance bins were classified into 16 cutoff distances—2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, and 4.0 Å.

Metal-ligand interactions: Metal atoms in active sites often play a key role of stabilizing ligands and reaction catalysis. For example, in HCAII, the ligand forms a metal-ligand bond that is one of the primary binding interactions [14] with Zinc and the bond length is 1.96 Å (Fig. 2b). However, in most scoring functions, metal-ligand interactions were modeled as hydrogen-bonding interactions. We separated metal-ligand interactions from hydrogen-bonding interactions. This work considered a metal-ligand interaction (V_{Metal}) when the distance between any atoms of a ligand and metal ion is less than a specific cutoff, and different types of metal ions were treated as identical. Here, we considered 13 cutoffs—2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, and 3.4 Å.

C. Stepwise regression analysis

Stepwise regression [15] was applied to select descriptors for developing GemAffinity. This method is simple and is able to avoid the adverse effects, such as over fitting and loss of biological/physical meanings of machine-learning approaches (*e.g.*, support vector machines, genetic algorithms, and neural networks). Stepwise regression first selects the descriptor with the highest Pearson correlation coefficient into the prediction model. Next, it sequentially adds other descriptors that increase the correlation between predicted binding affinities and experimental binding affinities. To reduce the adverse effects of overfitting data, the process was stopped when the increased correlation of the selected descriptor was <0.005.

D. Applications for virtual screening

A reliable scoring function can both reflect binding affinities and identify active compounds in virtual screening. In virtual screening, a scoring function should identify a small number of active compounds from hundreds of thousands of non-active compounds during post-docking analysis. This study used two docking programs, GEMDOCK (for GEMDOCK-PLP and GemAffinity) and GOLD (for GoldScore), generate docked poses for assessing the accuracies of three scoring functions for four protein targets. These four target proteins are thymidine kinase (TK, PDB code 1kim), estrogen-receptor antagonist (ER, PDB code 3ert), estrogen-receptor agonist (ERA, PDB code 1gwr), and human carbonic anhydrase II (HCAII, PDB code 1cil). The receptors for these screens cover different receptor types and, therefore, provide a reasonable test for measuring efficacy of scoring functions. The screening data sets and the table of analysis features are available at http://gemdock.life.nctu.edu.tw/dock/download.php.

TABLE I. THE SELECTED DESCRIPTORS IN GEMAFFINITY

Order	Descriptors	r^a	Description
1	f_{vdw}	0.497	Sum of Lennard-Jones potential
2	V_{Metal}	0.544	The distance between metal irons and a ligand is less than 2.2 Å
3	N_{rotBon}	0.579	Number of rotatable bonds of a ligand
4	N_{water}	0.594	Number of structural waters which are near to a ligand within 3.6 Å
5	N_{conHB}	0.600	Number of highly conserved residues forming hydrogen bonds

a. the correlation between predicted binding affinities and experimental binding affinities

III. RESULTS AND DISCUSSION

A. GemAffinity

Five descriptors of GemAffinity selected from stepwise regression are the sum of Lennard-Jones potential (f_{vdw}), metal-ligand interaction (V_{Metal}), number of rotatable bonds of a ligand ($N_{rotBond}$), number of structural waters that are within 3.6 Å of a ligand (N_{water}), and number of highly conserved residues forming hydrogen bonds between ligands and proteins (N_{conHB}) (TABLE I). The coefficient of each descriptor of GemAffinity is given as

$$-log(K_{d,pred}) = -0.081f_{vdw} + 1.329V_{Metal} - 0.0585N_{rotBond} - 0.1056N_{water} + 0.0847N_{conHB} + 3.331$$

f_{vdw} is negative and its coefficient (-0.081) is negative; therefore, the value $-0.081f_{vdw}$ is positive. Both V_{Metal} and N_{conHB} are positive and their respective coefficients are also positive values and positively enhance binding affinities. Conversely, N_{water} and $N_{rotBond}$ are positive but their respective coefficients are negative. These two terms adversely affect binding affinities.

B. Selected descriptors

The Lennard-Jones 12-6 potential is the first selected descriptor which yields the Pearson correlation coefficient is 0.497 (TABLE I). This result indicates that complementary shape between proteins and their ligands is critical for predicting binding affinities. The descriptor of metal-ligand interactions is the second selected term. In this study, a metal-ligand interaction is considered to be formed when an interaction distance is within a cutoff distance of $\leqq 2.2$ Å. Most metal-ligand interaction distances are less than 2.8 Å (Fig. 3); that is, the metal-ligand bonding distances are often shorter than the normal hydrogen-bonding distance. Therefore, considering metal-ligand interactions as hydrogen interactions may cause energy dramatic decreases in predicting binding affinities. Additionally, Fig. 3b shows the average affinity distribution of complexes in different distances of metal-ligand interactions. The average affinity of complexes that have metal-ligand interactions $\leqq$ 2.2 Å is 7.43, and the average affinity of complexes that have metal-ligand interactions between 2.2 and 3.6 Å is 5.42. The statistic results also indicate that the metal-ligand interactions strongly related binding affinities are often $\leqq$ 2.2 Å. As a result, GemAffinity highlights metal-ligand interactions as a special force instead of treating them as hydrogen-bonding interactions. For example, the binding affinity of the complex (PDB code 1ctu) with a metal-ligand interaction between zinc^{2+} and the bound ligand is 11.92. The binding affinity is 4.52 (Fig. 4) if the complex (PDB code 1ctt) lost this metal-ligand interaction.

The third selected descriptor is the number of rotatable bonds of a ligand. The number of rotatable bonds is measure of the effect of unfavorable torsional entropy loss on protein-ligand binding affinities. The fourth selected descriptor (the number of structural water molecules within a 3.6 Å distance from the bound ligand) is water effects. A large amount of water molecules around a ligand implies that a large volume of the ligand is exposed to the solvent. A ligand buried deeply within a protein usually has higher binding affinities than those bounded on the protein surface.

The final selected term selected for GemAffinity is the number of highly conserved residues interacting to ligands with hydrogen bond(s). A hydrogen bond formed between a highly conserved residue and a ligand often influences binding affinity markedly. In addition, this description is able to reduce the ill effects of (*i.e.,* prefer highly polar compounds) energy-based scoring functions. For example, the binding affinity of the beta-glucosidase protein (PDB code 1uz1 and 2j77) declines from 6.89 to 4.89 because of

the loss of a hydrogen bond between the ligand and a highly conserved residue (GLU166) (Fig. 5). Conversely, the binding affinity of the oligo-peptide binding protein (PDB codes 1b58 and 1b3h) reduced slightly from 6.58 (1b58) to 6.21 (1b3h) even though two hydrogen bonds lose between the ligand and non-conserved residues (ASN436 and TYR269) (Fig. 6).
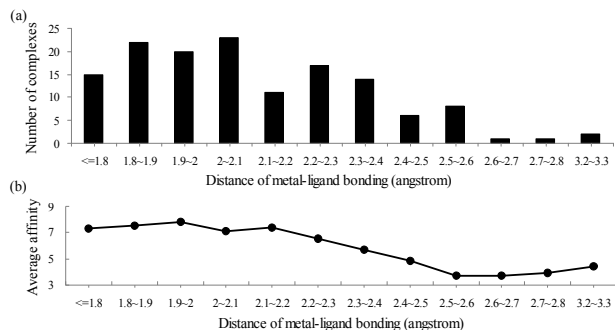

Figure 3. Metal-ligand interactions. (a) The distribution of metal-ligand interaction distances derived from 891 complexes in the training set. (b) Average binding affinity distribution of 891 complexes in different distances of metal-ligand interactions.
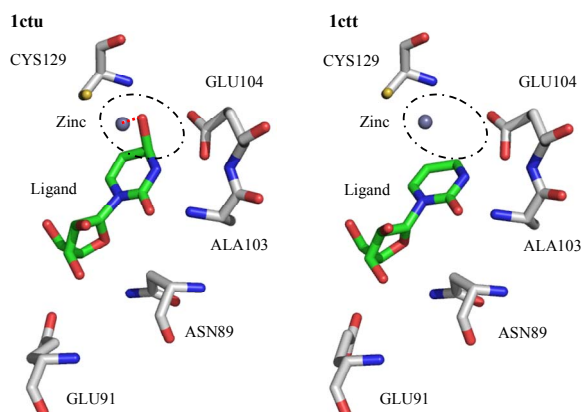
Figure 4. The effect of metal-ligand interactions. In the cytidine deaminase protein (PDB codes: 1ctu and 1ctt), the binding affinities of the complexes with a metal-ligand interaction (1ctu) and without a metal-ligand interaction (1ctt) are 11.92 and 4.52, respectively. The decrease in binding affinities of these two complexes is due to loss of the metal-ligand interaction.
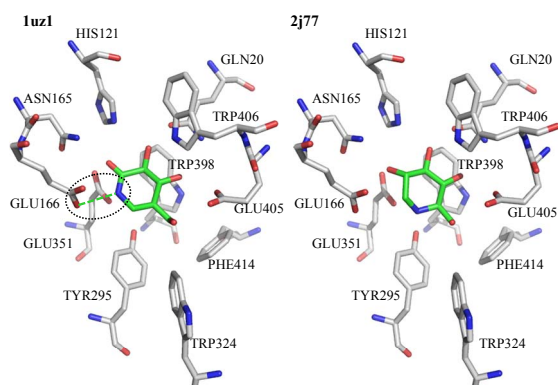

Figure 5. An example of a hydrogen bond lost in a highly conserved residue. The binding affinities of beta-glucosidase protein (PDB codes: 1uz1 and 2j77) are 6.89 and 4.89, respectively. For these two complexes, the bound-ligand structures and protein-ligand interactions are extremely similar; the only difference is the position of the nitrogen atom in the ligand. For the complex 1uz1, the nitrogen forms a hydrogen bond with GLU166; conversely, the nitrogen cannot form any hydrogen bond on complex 2j77. The conserved score of GLU166 is 2.
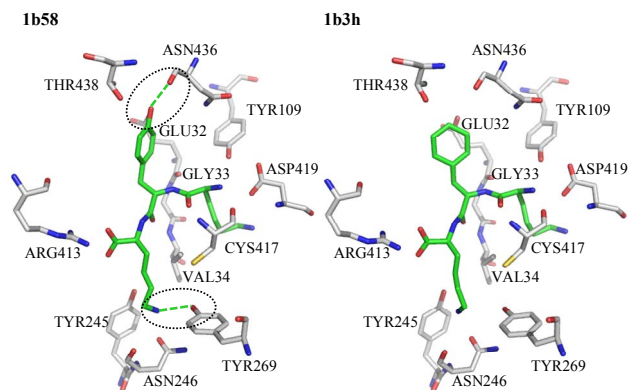

Figure 6. An example of hydrogen bonds lost in the low conserved residues. For the oligo-peptide binding protein (PDB code: 1b58 and 1b3h), the binding affinities of the complexes with hydrogen bonds (1b58) and without hydrogen bonds (1b3h) between bound ligands and residues (i.e., ASN436 and TYR269) are 6.58 and 6.21, respectively. The conserved scores of ASN436 and TYR269 are −3 and −6.

C. Binding affinity prediction

The accuracy of GemAffinity was evaluated for the independent test set, and its performance was compared with those of other 12 methods. Pearson correlation coefficients yielded by GemAffinity for the training set and independent test set are 0.600 and 0.572, respectively (Fig. 7). Based on this data set, GemAffinity is the best among these 12 scoring functions summarized from Wang et al. [16] (Fig. 8).

Most scoring functions consist of common types of protein-ligand interactions, such as vdW, hydrogen bonding, and electrostatic interactions. These 12 scoring functions are typically useful in predicting the affinities of most protein-ligand complexes. However, when some unusual factors (e.g., metal-ligand interactions) are necessary to the binding process, these scoring functions may fail to predict binding affinities. Conversely, GemAffinity is much better than these comparable functions when complexes have metal-ligand interactions or key interactions between conserved residues and their ligands (e.g., PDB codes 456c, 1r1j, 4tmn, 1ghy, 1ork, 1ttm, 1pyn, 1d4s, and 1o9d).
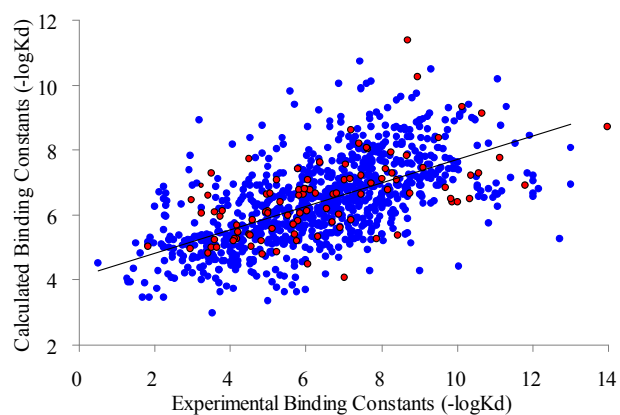

Figure 7. Correlations between experimental binding affinities and predicted binding affinities using GemAffinity on the training set (●) and testing set (●).
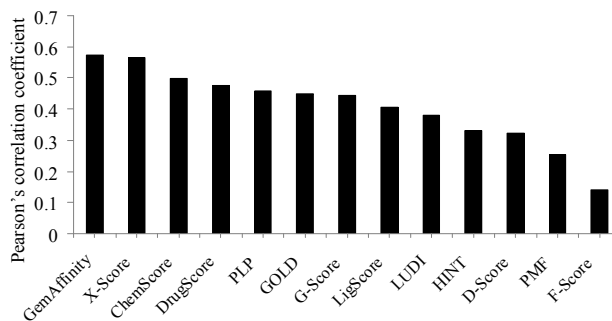
Figure 8. Comparisons of GemAffinity with 12 other scoring functions on 98 complexes in the independent set.

D. Virtual screening enrichment.

GemAffinity was evaluated and compared with other scoring functions on virtual screening. This study used two docking programs, GEMDOCK and GOLD, to assess the accuracies of three scoring functions (GEMDOCK-PLP, GoldScore, and GemAffinity) for four protein targets (TK, ER, ERA, and HCAII) based on the receiver operating characteristic (ROC) curve (Fig. 9). Note that during the docking process, docked protein-ligand complexes are not immersed in water by any programs for saving time; thus, N_{water} is 0 when GemAffinity calculates binding affinities. Pearson correlation coefficients for the training set and independent test set are 0.583 and 0.514, respectively, when GemAffinity ignores the water effect descriptor. Among these scoring methods, experimental results indicate that GemAffinity (red) is the best for targets ERA and HCAII, and is very comparable to other scoring methods for targets TK and ER.

For the target HCAII (Fig. 2), three metal binding residues (HIS94, HIS96 and HIS119), and three catalytic residues (HIS64, GLU106, and THR199) are highly conserved residues [17]. For the targets ER and ERA, the highly conserved residues are GLU353, ARG394, and HIS524. For the target TK, the highly conserved residues are GLU83, GLN125, ARG1635, ARG222, and GLU22. Most active compounds of the four targets have consistently hydrogen-bonding interactions with these highly conserved residues. For the target ER, GEMDOCK-PLP was slightly better than GemAffinity and GoldScore because two compounds (4-hydroxy tamoxifene and nafoxidene) of the ten ER inhibitors lack hydrogen bonds formed with the conserved residue (HIS524) on target ER.

In the target HCAII, GemAffinity outperforms GEMDOCK-PLP and GoldScore. GemAffinity treats metal-ligand interactions (V_{Metal}) as individual terms and divides the hydrogen bonds into conserved (N_{conHB}) and non-conserved interactions. Conversely, the scoring functions of GEMDOCK and GOLD consider metal-ligand interactions as one kind of hydrogen interactions and consider all hydrogen bonds the same.

GemAffinity can reduce the deleterious effects of screening ligand structures that are rich in charged or polar atoms. In general, energy-based scoring (*e.g.*, GEMDOCK-PLP and GoldScore) functions favor the selection of polar compounds that produce hydrogen-bonding and/or electrostatic potentials. GemAffinity uses N_{conHB} and V_{Metal} to measure essential interactions instead of other potential functions of hydrogen bonds that consider that all interactions contribute equally. The

average polar atom numbers of the top-ranking 50 compounds of the four targets selected from GemAffinity and GEMDOCK are 6.5 and 8.5, respectively (Fig. 10). These experimental results demonstrate that GemAffinity is able to increase the hit rate for virtual screening.
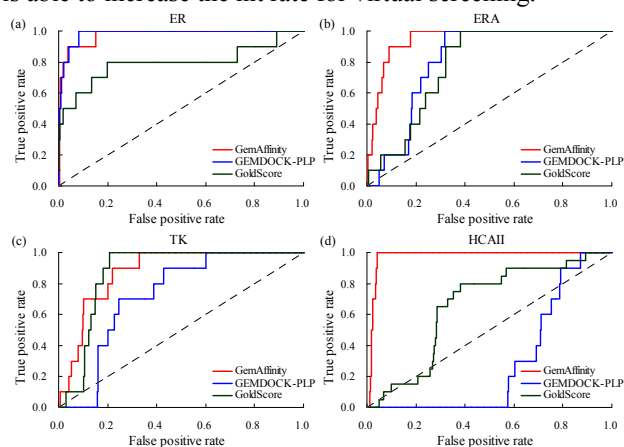

Figure 9. Comparisons of GemAffinity with GoldScore and GEMDOCK-PLP for virtual screening on four targets: (a) ER, (b) ERA, (c) TK, and (d) HCAII based on ROC curves. The curve (dashed line) of random selection is also plotted.
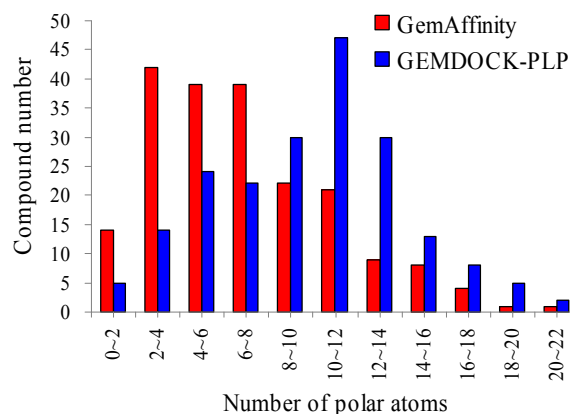

Figure 10. Distributions of polar atom number of the top-ranking 50 compounds of the four targets selected from GemAffinity and GEMDOCK-PLP.

IV. CONCLUSIONS

We used a simple stepwise regression to develop GemAffinity for predicting binding affinities and virtual screening. GemAffinity is able to reflect the experimental binding affinities and physicochemical meanings. Additionally, our GemAffinity reduces the deleterious effects of screening ligand structures that are rich in charged or polar atoms. Different from other scoring functions, GemAffinity highlights metal-ligand interactions and highly conserved residues with hydrogen bonds. For virtual screening, GemAffinity is very comparable to the GEMDOCK and GOLD scoring functions for four target proteins. These experimental results demonstrate that GemAffinity is useful for predicting binding affinities and can be combined with other scoring functions to improve prediction accuracy. We believe that the GemAffinity is useful for molecular recognition and virtual screening.

REFERENCES

[1] P. D. Lyne, "Structure-based virtual screening: an overview," *Drug Discov. Today*, vol. 7, pp. 1047-1055, 2002.

[2] J.-M. Yang and T.-W. Shen, "A pharmacophore-based evolutionary approach for screening selective estrogen receptor modulators," *Proteins*, vol. 59, pp. 205-220, 2005.

[3] J.-M. Yang and C.-C. Chen, "GEMDOCK: a generic evolutionary method for molecular docking," *Proteins*, vol. 55, pp. 288-304, 2004.

[4] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *J. Mol. Biol.*, vol. 267, pp. 727-748, 1997.

[5] R. Wang, X. Fang, Y. Lu, and S. Wang, "The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures," *J. Med. Chem.*, vol. 47, pp. 2977-2980, 2004.

[6] G. M. Morris, D. S. Goodsell, R. Huey, and A. J. Olson, "Distributed automated docking of flexible ligands to proteins: parallel applications of autodock 2.4," *J. Comput. Aided Mol. Des.*, vol. 10, pp. 293-304, 1996.

[7] A. Gutteridge and J. M. Thornton, "Understanding nature's catalytic toolkit," *Trends Biochem. Sci.*, vol. 30, pp. 622-629, 2005.

[8] Y. C. Chen, Y. S. Lo, W. C. Hsu, and J. M. Yang, "3D-partner: a web server to infer interacting partners and binding models," *Nucleic Acids Res.*, vol. 35, pp. W561-W567, 2007.

[9] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res.*, vol. 25, pp. 3389-3402, 1997.

[10] R. Apweiler, A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L. S. Yeh, "UniProt: the Universal Protein knowledgebase," *Nucleic Acids Res.*, vol. 32, pp. D115-D119, 2004.

[11] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 89, pp. 10915-10919, 1992.

[12] H. Kubinyi, "QSAR and 3-D QSAR in drug design. 1. Methodology," *Drug Discov. Today*, vol. 2, pp. 457-467, 1997.

[13] F. Osterberg, G. M. Morris, M. F. Sanner, A. J. Olson, and D. S. Goodsell, "Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in autodock," *Proteins*, vol. 46, pp. 34-40, 2002.

[14] G. M. Smith, R. S. Alexander, D. W. Christianson, B. M. McKeever, G. S. Ponticello, J. P. Springer, W. C. Randall, J. J. Baldwin, and C. N. Habecker, "Positions of His-64 and a bound water in human carbonic anhydrase II upon binding three structurally related inhibitors," *Protein Sci.*, vol. 3, pp. 118-125, 1994.

[15] N. Draper and H. Smith, *Applied regression analysis*, 2nd ed: John Wiley and Sons, 1981.

[16] R. Wang, Y. Lu, X. Fang, and S. Wang, "An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes," *J. Chem. Inf. Comput. Sci.*, vol. 44, pp. 2114-2125, 2004.

[17] J. Y. Liang and W. N. Lipscomb, "Hydration of CO2 by carbonic anhydrase: intramolecular proton transfer between Zn2+-bound H2O and histidine 64 in human carbonic anhydrase II," *Biochemistry*, vol. 27, pp. 8676-8682, 1988.