BIOINFORMATICS ORIGINAL PAPER

Vol. 30 no. 12 2014, pages 1674–1680
doi:10.1093/bioinformatics/btu104

Structural bioinformatics

Advance Access publication February 21, 2014

Improved protein–ligand binding affinity prediction by using a curvature-dependent surface-area model

Yang Cao* and Lei Li

Center for Growth, Metabolism and Aging, Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, 610064, China

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Hydrophobic effect plays a pivotal role in most protein–ligand binding. State-of-the-art protein–ligand scoring methods usually treat hydrophobic free energy as surface tension, which is proportional to interfacial surface area for simplicity and efficiency. However, this treatment ignores the role of molecular shape, which has been found very important by either experimental or theoretical studies.

Results: We propose a new empirical scoring function, named Cyscore. Cyscore improves the prediction accuracy by using a novel curvature-dependent surface-area model, which is able to distinguish convex, planar and concave surface in hydrophobic free energy calculation. Benchmark tests show that this model significantly improves the protein–ligand scoring and Cyscore outperforms a variety of well established scoring functions using PDBbind benchmark sets for binding affinity correlation and ranking tests. We expect the curvature-dependent surface-area model and Cyscore would contribute to the study of protein–ligand interactions.

Availability: Cyscore is available to non-commercial users at http://clab.labshare.cn/software/cyscore.html.

Contact: cao@scu.edu.cn

Supplementary information: Supplementary information is available at *Bioinformatics* online.

Received on August 2, 2013; revised on January 28, 2014; accepted on February 16, 2014

1 INTRODUCTION

Scoring functions are approximate computational methods for predicting the binding affinity of two docked molecules, especially, a protein and a ligand. They are widely used in structure-based drug discovery and other molecular modeling applications, including virtual screening, de novo drug design and lead optimization (Blundell, 1996; Kuntz, 1992). In the past several decades, a variety of scoring functions have been developed. In general, they can be classified into three types, force-field-based (Ewing *et al.*, 2001; Morris *et al.*, 1996), knowledge-based (Gohlke *et al.*, 2000; Mooij and Verdonk, 2005; Muegge and Martin, 1999; Velec *et al.*, 2005; Zhou and Skolnick, 2011) and empirical scoring functions (Eldridge *et al.*, 1997; Friesner *et al.*, 2004; Koes *et al.*, 2013; Korb *et al.*, 2009; Tang and Marshall, 2011; Wang *et al.*, 2002). Force-field-based scoring functions are derived from first principle, which employs classical force fields to compute the direct interaction energies, e.g. van der Waals and electrostatic energies, between the ligand and its target protein. Sometimes these methods are combined with GB/SA or PB/SA methods to address the solvation effect (Liu and Zou, 2006; Moustakas *et al.*, 2006; Shoichet *et al.*, 1999; Wei *et al.*, 2002). Knowledge-based scoring functions are derived from statistical potentials, which predict the binding free energy using the probabilities of finding atom pairs at a given distance between proteins and ligands. They have obtained a high level of prediction accuracy according to some recent studies (Ballester and Mitchell, 2010; Huang and Zou, 2010; Zhang *et al.*, 2005; Zheng and Merz, 2013). However, knowledge-based scoring functions usually require large numbers of terms which are prone to over fitting by training, and can hardly provide immediate physical interpretation of results. Those weak points limit their applications. Empirical scoring functions incorporate the advantages of both force-field-based and knowledge-based scoring functions. They calculate the overall binding free energy as a sum of several physically meaningful terms, while their coefficients are derived from the regression analysis on experimental data. As they can achieve comparably accurate prediction and provide physical insights of protein–ligand binding as well, empirical scoring functions are widely used in structure-based drug-design applications.

In this article, we propose a new empirical scoring function, named Cyscore. It is composed of hydrophobic free energy, van der Waals interaction energy, hydrogen bond interaction energy and ligand's conformational entropy. We mainly focus on improving the prediction of hydrophobic free energy, because it is regarded to be dominant in most protein–ligand binding (Ball, 2007; Blokzijl and Engberts, 1993; Chandler, 2005; Meyer *et al.*, 2003; Setny *et al.*, 2009; Snyder *et al.*, 2011; Southall *et al.*, 2001). Hydrophobic effect is the observed tendency of nonpolar surfaces to association in an aqueous solution. Many theoretical approaches have been proposed to study its mechanism, such as 'iceberg' model (Frank and Evans, 1945; Kauzmann, 1959; Tanford, 1979), scaled-particle theory based 'void volume' models (Graziano and Lee, 2003; Lum *et al.*, 1999; Pratt and Chandler, 1977; Stillinger, 1973). Both experimental and theoretical studies have pointed out that hydrophobic effect depends not only on surface area, but also on shape (Cheng and Rossky, 1998; Fennell *et al.*, 2010; Rajamani *et al.*, 2005; Southall *et al.*, 2001). In the conventional protein–ligand scoring methods, hydrophobic free energy is usually treated as surface tension, which is proportional to interfacial surface area for simplicity and efficiency. Obviously, this treatment ignores the role of molecular shape and hinders the more accurate prediction

*To whom correspondence should be addressed.

of hydrophobic free energy. Inspired by the thermodynamic research of Tolman (1949) and the pioneer work of Nicholls *et al*. (1991; Sharp *et al*., 1991), we took curvature as the descriptor of shape and proposed a novel curvature-dependent surface-area model. Our analyses demonstrate that this model is superior to the conventional surface-area model significantly in protein–ligand scoring. Compared to a variety of well-established scoring functions, it achieves the top performance using the well known PDBbind benchmark sets (Cheng *et al*., 2009; Wang *et al*., 2004, 2005). The methods and results will be elaborated in the following sections.

2 METHODS

2.1 Scoring functions

We assume that the overall protein–ligand binding free energy can be decomposed into four terms:

$$\Delta G_{\text{bind}} = k_h \Delta G_{\text{hydrophobic}} + k_v \Delta G_{\text{vdw}} + k_b \Delta G_{\text{hbond}} + k_e \Delta G_{\text{entropy}} \quad (1)$$

$\Delta G_{\text{hydrophobic}}$ is the change of hydrophobic free energy from the free-state to the bound-state of the protein and ligand:

$$\Delta G_{\text{hydrophobic}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}}, \quad (2)$$

where G_{complex}, G_{protein} and G_{ligand} are the hydrophobic free energies of protein–ligand complex, individual protein and ligand in the solvent, respectively. Hydrophobic free energy is computed as the integration of surface curvature factor $\left(\frac{V_{\text{sphere}}}{V_{\text{out}}}\right)^2$ over the solvent accessible surface S:

$$G = \iint_s \left(\frac{V_{\text{sphere}}}{V_{\text{out}}}\right)^2 ds. \quad (3)$$

This is our curvature-dependent surface-area model. The curvature factor $\left(\frac{V_{\text{sphere}}}{V_{\text{out}}}\right)^2$ is defined by using a sphere centering at each sampling point on the solvent accessible surface (S) and calculating the square ratio of sphere volume (V_{sphere}) to volume out of the solvent accessible surface (V_{out}) (Fig. 1). Thus the convex, planar and concave surfaces are different in hydrophobic free energy prediction. The radius of the sphere is set to be 1.2Å according to the derivation of our model (Supplementary Material). The atom radii are from Chothia's (1976) work. To obtain the square ratio numerically, we first generate a set of points to represent the solvent accessible surface using the method developed by Shrake and Rupley (1973). Each of these points represents an integral element of surface area. Then we place all the points into a grid of 0.1Å spacing and compute the sphere volumes at each point by accumulating the grid volume. Finally we summarize all the products of square ratio and the surface element area. As the curvature factor requires a time consuming computation, we implemented the k-discrete oriented polytopes (k-DOPs) method to detect the overlap of spheres in this program for fast computation (Klosowski *et al*., 1998).

ΔG_{vdw} accounts for the van der Waals interaction and is calculated as the summation of Lennard–Jones 10-6 potentials between the protein and the ligand.

$$\Delta G_{\text{vdw}} = \sum_i^{\text{protein}} \sum_j^{\text{ligand}} \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] \quad (4)$$

r_{ij} is the distance between protein atom i and ligand atom j. And σ_{ij} is the sum of van der Waals radii of protein atom i and ligand atom j, ε_{ij} is the depth of the potential well, and is calculated as $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$. The van der Waals radii and ε_i, ε_j of each type of atoms are obtained from CHARMM22 parameters (Brooks *et al*., 1983). Hydrogen atoms are ignored in the calculation because their position is dependent on the methods of adding hydrogen and may result in atom clashes.

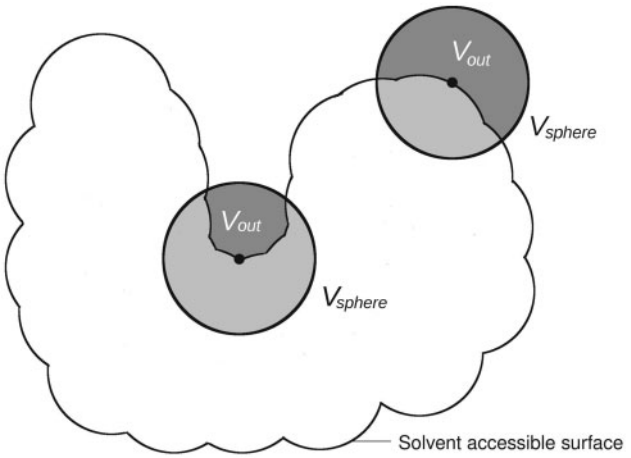

Fig. 1. A cross section for visualizing the curvature-dependent surface-area model at concave and convex surface. The white region surrounded by the solvent accessible surface represents a molecule, which is the protein or ligand. Dark grey region represents the volume of the sphere out of the solvent accessible surface (V_{out}). Light grey region represents the volume of the sphere buried in the solvent accessible surface ($V_{\text{sphere}} - V_{\text{out}}$). V_{out} can be a descriptor for the concave and convex surface

ΔG_{hbond} accounts for the hydrogen-bond interaction between the protein and ligand. It was adapted from that of Grid (Goodford, 1985):

$$\Delta G_{\text{hbond}} = \sum_i^{\text{protein}} \sum_j^{\text{ligand}} \left(\frac{C}{r_{ij}^6} - \frac{D}{r_{ij}^4}\right) \cos^4 \theta \quad (5)$$

where r_{ij} is the distance between heavy atoms of protein atom i and ligand atom j, θ is the angle formed by the donor atom, hydrogen and acceptor atom (Supplementary Fig. S2). If $\theta < 90°$, the hydrogen bond energy for the pair of atoms is set to be 0. C and D are the parameters obtained from Goodford' (1985) work.

$\Delta G_{\text{entropy}}$ accounts for the entropic loss effect of ligands upon binding process. We employed the classic method of counting the number of rotatable bonds (rotors). It is based on the assumption that each rotor in the unbound state associates with a discrete number of low-energy conformations but 'freezes' into a single conformation upon binding. Thus the number of rotatable bonds is proportional to a certain amount of conformational entropy (Bohm, 1994; Friesner *et al*., 2004; Salaniwal *et al*., 2007).

k_h, k_v, k_b and k_e are the coefficients of each term. They are obtained by the standard least-square multivariate linear regression of the training set ($k_h = -0.0082$, $k_v = -0.069$, $k_b = -0.0028$, $k_e = -0.042$).

In the following tests, we quantified the prediction accuracy through Pearson's correlation coefficient (R) and standard deviation (SD), which are the commonly used criterion for binding affinity prediction (Huang *et al*. 2010).

2.2 Datasets

We divided the datasets into a training set and two benchmark sets. They were selected from the database of PDBbind (Wang *et al*., 2004, 2005), which collects the experimentally measured binding affinity data for all the biomolecular complexes deposited in the Protein Data Bank (PDB).

The training set is used to obtain the coefficients before each term by linear regression. It was selected from the latest version (v2012) of PDBbind using the following criteria: (i) structure resolution < 1.8Å, (ii) binding affinity covers the range from 1 to 11 kcal/mol. Each complex

was double checked to ensure that its binding data was convincing and (iii) protein–ligand data should be different from the benchmark data. We checked the protein sequence similarity and ligand chemical composition of the complexes (protein sequence similarity should be <90% or ligands should be different in chemical formula). Thus there is no complex in the training set whose protein and ligand are the same as that in the benchmark data. Finally we obtained 247 protein–ligand data which can be found in Supplementary Table S1.

The benchmark sets are two 'core set' of PDBbind: the latest version (v2012) and a most widely used version (v2007). The 'core set' was selected by the author of PDBbind, which was aim to provide a set of high quality, non-redundant data with special focus on the diversity of structures and binding data. The two sets contains 201 (v2012) and 195 (v2007) protein–ligand data, respectively. Moreover, they can be clustered into 67 and 65 groups, each of which contains a high-affinity ligand, a medium-affinity ligand, and a low-affinity ligand bound to a common protein (Cheng *et al.*, 2009). The clustering of protein–ligand data can be used to test the ability to rank different ligands bound to the same protein when the correct binding poses of these ligands are known. The overlaps of the two benchmark sets are ~30%.

2.3 Software availability

Cyscore is implemented in C++ and has been complied on Linux system. It is now freely available to non-commercial users at http://clab.labshare.cn/software/cyscore.html. Valid input file formats are PDB for proteins and Tripos Mol2 for ligands. Typically this software is able to process around 4000 ligand molecules for a given protein target per hour on an Intel Core™ processor (3.0 GHz) using one CPU core.

3 RESULTS

In this section, we first give an overview of Cyscore. Then we investigate how the curvature-dependent surface-area model is superior to the conventional models in scoring protein–ligand interactions. Finally, to gain an overall performance of Cyscore, we compare it to a variety of well-established scoring functions using PDBbind benchmarks.

3.1 Overview of Cyscore

Cyscore is an empirical protein–ligand scoring function, composed of hydrophobic free energy, van der Waals interaction energy, hydrogen-bond interaction energy and the ligand's conformational entropy. Those terms follow the classic approaches except the hydrophobic free energy. Conventionally, hydrophobic free energy is treated as a form of surface tension by many researches (Sinanoglu, 1981; Tanford, 1979). Thus hydrophobic free energy is measured by the surface area of a molecule, including solvent accessible surface area, molecular surface area or some approximate methods such as atom contact numbers (Petrey and Honig, 2000). However, a thermodynamic research of Tolman (1949) shows that microscopic surface tension not only associates with surface area but also with curvature, that is:

$$\frac{\gamma(R)}{\gamma(\infty)} = \frac{1}{1 + 2\delta/R}, \tag{6}$$

where $\gamma(R)$ is the surface tension of spherical particle of radius R, $\gamma(\infty)$ is the macroscopic surface tension and δ is a constant. This equation in principle provides a means of obtaining the curvature-dependent hydrophobic free energy, but it is only applicable to spherical interfacial geometries. In order to derive a well-defined model, Nicholls *et al.* (1991) developed a curvature-corrected surface-area model, which calculates the curvature factor at the points distributed over the accessible surface of a molecule. The hydrophobic free energy is:

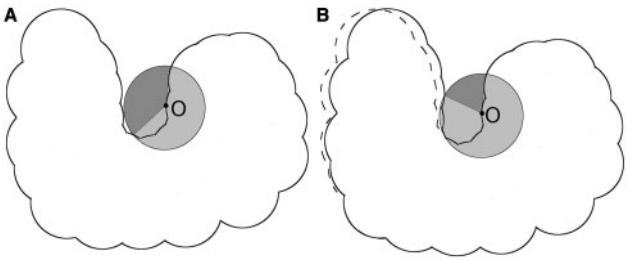

Fig. 2. An example for the distance sensitivity of Nicholls' model. Supposing for a protein, the experimentally determined or predicted structure (solid line in A but dash line in B) is slightly different from the real structure (solid line in B). To determine the curvature of the surface, a sphere is centered at the point O. Although the two structures are almost the same, the exposed surfaces (dark gray in A and B) of the sphere are quite different, resulting in the inaccurate value of curvature

$$G = \gamma(\infty)\sum S_i c_i, \tag{7}$$

where S_i is the area of the surface element, and c_i is the curvature factor associated with S_i. To determine c_i, they place a sphere, centering at each sample point of the solvent accessible surface, and calculate the fraction of sphere surface lies in the molecular solvent accessible surface, which is $c_i = \frac{2S_{\text{inner}}}{S_{\text{sphere}}}$. S_{inner} is the surface area of the sphere buried in the molecular solvent accessible surface, and S_{sphere} is the total surface area. In the planar surface, $c_i = 1.0$, which means that Equation (7) is equal to the conventional surface-area model. In the concave region, $c_i > 1.0$, which indicates a higher hydrophobic free energy, while in the convex region, vice versa. Originally, this model was used in analyzing hydrophobic effects in protein folding and association, etc. We tried to apply it to protein–ligand binding affinity scoring. However, it fails to show any improvements in our benchmark tests (Section 3.2). After a careful analysis, we found a possible reason is that c_i is too sensitive to the judgment on whether a surface element of the sphere lies in or out of the molecular solvent accessible surface. In other words, if a protein structure is not so precisely determined, the minor difference in structure may lead to incorrect value of S_{inner}. As illustrated in Figure 2, the real protein surface and the experimentally determined or predicted protein surface are almost the same but the values of buried surface area of the sphere centering at point O have great difference (dark grey region). In the case of protein–ligand scoring, this weak point is amplified because many of the atom distances in the binding pockets are around the size of a sphere, and thus errors will be accumulated, leading to bad scoring.

Inspired by the work of Nicholls *et al.* (1991), we developed a new curvature-dependent surface-area model using the sphere volume, instead of area, to evaluate the surface curvature. Our curvature factor is defined as the square ratio of sphere volume to the volume out of the molecular solvent accessible surface [Fig. 1 and Equation (3)]:

$$c = \left(\frac{V_{\text{sphere}}}{V_{\text{out}}}\right)^2. \tag{8}$$

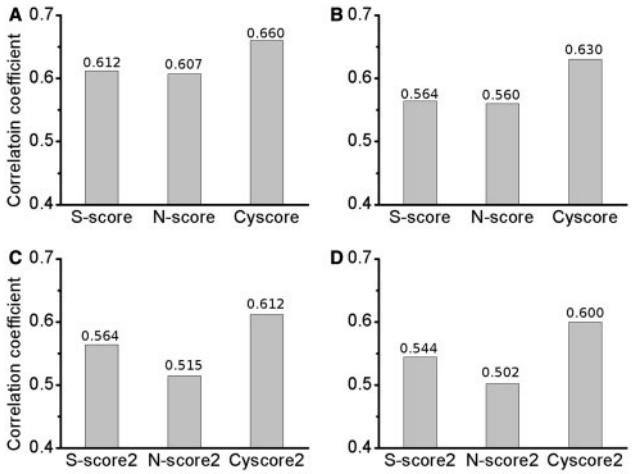

Fig. 3. Comparison of the curvature-dependent surface-area model with conventional and Nicholls' models in protein–ligand scoring by Pearson's correlation coefficient. (A and B) Results of four-term scoring functions on PDBbind core set of v2007 and v2012 respectively. (C and D) Results of two-term scoring functions on PDBbind core set of v2007 and v2012, respectively

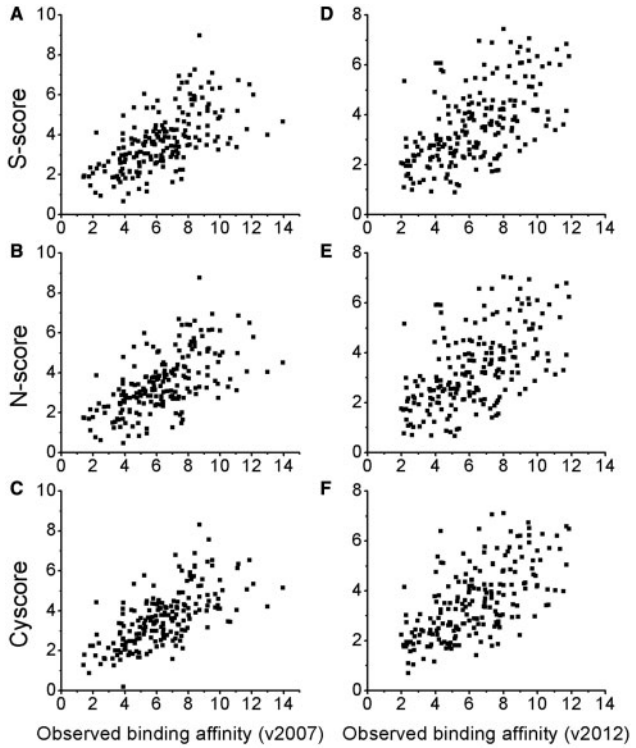

Fig. 4. Plots of S-, N-score and Cyscore predicted binding affinity versus experimentally observed binding affinity (in log K_d units). (A, B and C) The results on PDBbind core set of v2007 for S-, N-score and Cyscore, respectively. (C, D and E) The results on PDBbind core set of v2012

The use of volume can avoid the sensitivity in determining the buried or exposed surface area simply by the sphere radius. Moreover, derived from our geometry analysis, this factor is an approximate form of Tolman's Equation (6), which indicates the physical meaning of our model (Supplementary Material). We expected this new approach could provide robust estimation of surface curvature and contribute to protein–ligand binding affinity scoring.

3.2 Curvature-dependent surface-area model contributes to protein–ligand binding affinity prediction

To investigate the effectiveness of the new curvature-dependent surface-area model, we combined it with the van der Waals interaction energy, hydrogen-bond interaction energy and ligand's entropic term, which are defined in Section 2.1. This new scoring function is called Cyscore. For comparison, we replaced the curvature-dependent surface-area model of Cyscore by Nicholls' model [Equation (7)] and conventional surface-area model (hydrophobic free energy is proportional to the solvent accessible surface area), respectively. These two new scoring functions were named N- and S-score for short. Coefficients before each term of the three scoring functions are obtained from the training set using linear regression method individually. Then the three scoring functions were tested on the benchmark sets of PDBbind version 2007 and 2012 (Section 2.2). Results show that Cyscore achieves the highest Pearson's correlation coefficient in both sets, which is ~7–11% higher than the N- or S-score (Fig. 3A and B and Supplementary Tables S2 and S3). The correlation figures illustrate that our curvature-dependent surface-area model obviously improves prediction in low and medium affinity ligands (Fig. 4). A detailed analysis on individual testing data shows that Cyscore performs better in case of ligands binding in narrow pockets. Taking hydroxylated polychlorinated Biphenyl-4,4′-dihydroxy-3,3′,5,5′-tetrachlorobiphenyl binding with human transthyretin as an example (PDB code 2G5U, Supplementary Fig. S3), the binding pocket is deep but the interfacial surface is relatively small. Both N- and S-score underestimate their affinity except Cyscore which predicts much lower hydrophobic free energy. It is because the average curvature factor in this pocket weights 1.6 times higher than that at planar surface.

In the above test, Cyscore exhibits advantages in protein–ligand scoring. To more precisely identify the performance of the curvature-dependent surface-area model, we excluded the van der Waals interaction energy and hydrogen bond interaction terms from the scoring functions, as these contact-dependent terms may have overlaps with the hydrophobic term. Thus the three scoring functions are composed of only hydrophobic free energy and ligand's entropic terms (Their suffixes were set to be 2, as Cyscore2, N-score2 and S-score2 for short). We obtained the results by performing the same training and testing process as before. As illustrated in Figure 3C and D, Supplementary Tables S4 and S5, although those two-term scoring functions shows 7~15% decrease of the Pearson's correlation coefficients, Cyscore2 still performs the best in both sets, while N-score2 shows ~7% decrease compared to S-score2. This result implies that our curvature-dependent surface-area model contributes to protein–ligand scoring indeed.

3.3 Comparison of Cyscore with the well-established protein–ligand scoring functions

To gain an overall performance of Cyscore, we further compared it with a wide collection of scoring functions. Thank Cheng *et al.*,

Table 1. Correlations between the experimentally measured and the scoring-function-predicted binding affinities on the PDBbind core set (v2007)

Scoring function	N^a	R^b	SD[c]
Cyscore	195	0.660	1.79
X-Score::HMScore	195	0.644	1.83
DrugScore^{CSD}	195	0.569	1.96
SYBYL::ChemScore	195	0.555	1.98
DS::PLP1	195	0.545	2.00
GOLD::ASP	193	0.534	2.02
SYBYL::G-Score	195	0.492	2.08
DS::LUDI3	195	0.487	2.09
DS::LigScore2	193	0.464	2.12
GlideScore-XP	178	0.457	2.14
DS::PMF	193	0.445	2.14
GOLD::ChemScore	178	0.441	2.15
SYBYL::D-Score	195	0.392	2.19
DS::Jain	189	0.316	2.24
GOLD::GoldScore	169	0.295	2.29
SYBYL::PMF-Score	190	0.268	2.29
SYBYL::F-Score	185	0.216	2.35

Several scoring functions have different versions or multiple options. Only the results produced by the best version/option of a certain scoring function are listed (Cheng *et al.*, 2009). [a]Number of complexes receiving positive (favorable) binding scores by this scoring function. [b]Pearson correlation coefficients. [c]Standard deviations in linear correlation (in log K_d units).

Table 2. Success rates of scoring functions in 'ranking power' evaluation on the PDBbind core set (v2007)

Scoring function	Success rate (%)
X-Score::HSScore	58.5
Cyscore	53.8
DS::PLP2	53.8
DrugScore^{CSD}	52.3
SYBYL::ChemScore	47.7
SYBYL::D-Score	46.2
SYBYL::G-Score	46.2
GOLD::ASP	43.1
DS::LUDI3	43.1
DS::Jain	41.5
DS::PMF	41.5
SYBYL::PMF-Score	38.5
GOLD::ChemScore	36.9
DS::LigScore2	35.4
GlideScore-XP	33.8
SYBYL::F-Score	29.2
GOLD::GoldScore	23.1

who carefully selected a high-quality-benchmark set (PDBbind core set v2007) and assessed 16 well-known scoring functions, including five scoring functions [LigScore (Krammer *et al.*, 2005), PLP (Gehlhaar *et al.*, 1995), PMF (Muegge, 2006; Muegge and Martin, 1999), Jain (Jain, 1996), LUDI (Bohm, 1994, 1998)] in the Discovery Studio software version 2.0, five scoring functions (D-Score, PMF-Score, G-Score, ChemScore, F-Score) in the SYBYL software version 7.2, one scoring function [GlideScore (Friesner *et al.*, 2004, 2006)] in the Schrödinger software version 8.0, three scoring functions [GoldScore, ChemScore (Eldridge *et al.*, 1997), ASP (Mooij and Verdonk, 2005)] in the GOLD software version 3.2, two stand-alone scoring functions [DrugScore (Gohlke *et al.*, 2000; Velec *et al.*, 2005) and X-Score (Wang *et al.*, 2002)] released by academic groups. We can use Cyscore against the same benchmark set for a fair comparison. The test results for the 17 scoring functions are shown in Table 1. In terms of the Pearson's correlation coefficients (R) and the standard deviation (SD), Cyscore outperforms all of the tested scoring functions on this benchmark test. Moreover, we followed their newly defined 'ranking power' test, which benchmarks the ability to correctly rank the known ligands bound to a common target protein when native binding modes are known (Cheng *et al.*, 2009). Cheng *et al.* divide the above data set into 65 groups of protein–ligand complexes, each of which contains a high-affinity ligand, a medium-affinity ligand and a low-affinity ligand bound to a common protein. Scoring functions are evaluated by the success rate which counts for the ratio of correctly ranked groups in all the test groups. This test examines the ability of a scoring function to discriminate the structure changes of ligands, which is relevant to lead optimization and virtual screening studies because their aim is to rank molecules in the order of their binding affinities to the target protein. As shown in Table 2. Cyscore achieves a 53.8% success rate, which ranks the second best among the scoring functions.

Further more, we did another benchmark test using the most recent updated set of PDBbind (core set, version 2012). For comparison, two freely available stand-alone scoring functions are used, that are, X-Score (version 1.3) and DSX (DrugScore eXtended, a new improvement of DrugScore) (Neudert and Klebe, 2011). Their performances are among the top-scoring functions according to the above tests. Our results show that, the correlation coefficient of Cyscore is 0.630, while the values of other two scoring functions are less than 0.600 (Table 3, correlation Supplementary Fig. S4), and the success rate of Cyscore achieves 56.7% in 'ranking power' test, better than the other two scoring functions (Table 4). These results indicate that Cyscore has considerable and consistent performance.

The above benchmark tests illustrate the ability of scoring functions to evaluate binding affinities of diverse protein–ligand complexes. Generally, the inaccurate prediction of solvation energies and configurational entropies are the major limitations of present scoring functions (Cheng *et al.*, 2009; Huang *et al.*, 2010). Compared to other empirical scoring functions, such as X-Score, ChemScore, PLP, LUDI, LigScore and Jain, Cyscore has the similar or even simpler terms in composition. The main difference is in the method of calculating hydrophobic free energy, which is one of the main components of solvation energy. For example, X-Score implemented three types of hydrophobic terms, which were the conventional surface model, atom contact number based term and atom type-dependent contact term (Wang *et al.*, 2002). All the three methods ignored the effects of surface curvature on hydrophobicity. It is also a drawback of the solvation energy terms in most force-field-based

Table 3. Correlations between the experimentally measured and the scoring-function-predicted binding affinities on the PDBbind core set (v2012)

Scoring function	N^a	R^b	SD^c
Cyscore	201	0.630	1.89
X-Score::HMScore	201	0.597	1.95
DSX^{PDB}	201	0.573	2.00

The results produced by the best option of X-score are listed. [a]Number of complexes receiving positive (favorable) binding scores by this scoring function. [b]Pearson correlation coefficients. [c]Standard deviations in linear correlation (in log K_d units).

Table 4. Success rates of scoring functions in 'ranking power' evaluation on the PDBbind core set (v2012)

Scoring function	Success rate (%)
Cyscore	56.7
X-Score::HMScore	53.7
DSX^{PDB}	53.7

The results produced by the best option of X-score are listed.

scoring functions. As for the knowledge-based scoring functions, the curvature information may be obtained by the inverse Boltzmann relation, but the effectiveness is still practically unknown because it relates to the extension of pairwise potentials to many-body potentials (Huang *et al.*, 2010). Taken together, the encouraging performance of Cyscore suggests the significance of our work on designing the better algorithm for hydrophobic free energy.

4 DISCUSSION AND CONCLUSION

The hydrophobic effect plays a central role in bio-molecular recognition. It is estimated to provide, perhaps, 75% of the free energy of most binding or association events (Snyder *et al.*, 2013). Despite its importance, calculating hydrophobic energy is far from accuracy due to its complex origin in nature. In this article, we proposed a curvature-dependent surface-area model to estimate hydrophobic energy in protein–ligand binding. Compared to the conventional surface-area model, the new model is not only theoretically more exact, but also particularly more effective in the benchmark tests. Based on this discovery, we developed a new empirical protein–ligand scoring function, named Cyscore. In terms of binding affinity correlation, Cyscore outperforms a variety of well-established scoring functions in our tests. Of note, Cyscore performs better for ligands binding in narrow pockets, due to their larger contribution of the curvature factor to hydrophobic free energy. To our knowledge, the curvature-dependent surface tension has not been explicitly analyzed and taken into consideration in most of the state-of-the-art scoring functions. Thus, we believe our model will be useful to the study of protein–ligand interactions. In practice, Cyscore can be used in the molecular modeling applications whose accurate binding positions are accessible, such as lead optimization, binding-mode prediction (see additional experiments in Supplementary Table S6) and drug-resistant mutation prediction. It should be mentioned that, Cyscore has not been optimized for virtual screening, which strongly depends on the particular docking method.

In the present article, we mainly focus on identifying whether and how much the curvature-dependent surface-area model contributes to the protein–ligand binding affinity scoring. And in order to reduce the risk of overtraining, the composition of Cyscore is relatively simple in contrast to some other scoring functions. This limits the accuracy and universal applicability of Cyscore. For example, Cyscore does not count the water mediated protein–ligand interaction, the charge–charge interaction and the π–system interaction. Moreover, the calculation of entropy is a complex task. For simplicity and efficiency, Cyscore just counts the number of rotatable bonds of ligands to evaluate the entropy loss. In the future, we plan to adopt more detailed terms and larger training sets to achieve more accurate and general predictions of the protein–ligand binding affinity.

ACKNOWLEDGEMENT

We would like to thank Prof. Zhi-Xiong Jim Xiao and Prof. Taijiao Jiang for the stimulating discussion, Prof. Jiang Zhu, Prof. Qintong Li and Dr Shuang Chen for critical review of the article.

Funding: The Fundamental Research Funds for the Central Universities (2011SCU11104 to Y.C.).

Conflict of Interest: none declared.

REFERENCES

Ball,P. (2007) Water as an active constituent in cell biology. *Chem. Rev.*, **108**, 74–108.

Ballester,P.J. and Mitchell,J.B. (2010) A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics*, **26**, 1169–1175.

Blokzijl,W. and Engberts,J. (1993) Hydrophobic effects. Opinions and facts. *Angew. Chem. Int. Ed. Engl.*, **32**, 1545–1579.

Blundell,T.L. (1996) Structure-based drug design. *Nature*, **384(Suppl 6604)**, 23–26.

Bohm,H.J. (1998) Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from *de novo* design or 3D database search programs. *J. Comput. Aided Mol. Des.*, **12**, 309–323.

Bohm,H.J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.*, **8**, 243–256.

Brooks,B. *et al.* (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.

Chandler,D. (2005) Interfaces and the driving force of hydrophobic assembly. *Nature*, **437**, 640–647.

Cheng,Y.K. and Rossky,P.J. (1998) Surface topography dependence of biomolecular hydrophobic hydration. *Nature 1998*, **392**, 696–699.

Cheng,T. *et al.* (2009) Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model*, **49**, 1079–1093.

Chothia,C. (1976) The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.*, **105**, 1–12.

Eldridge,M.D. *et al.* (1997) 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.*, **11**, 425–445.

Ewing,T.J. *et al.* (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.*, **15**, 411–428.

Fennell,C.J. *et al.* (2010) Oil/water transfer is partly driven by molecular shape, not just size. *J. Am. Chem. Soc.*, **132**, 234–240.
Frank,H. and Evans,M. (1945) Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J. Chem. Phys.*, **13**, 507–532.
Friesner,R.A. *et al.* (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.*, **49**, 6177–6196.
Friesner,R.A. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, **47**, 1739–1749.
Gehlhaar,D.K. *et al.* (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.*, **2**, 317–324.
Gohlke,H. *et al.* (2000) Knowledge-based scoring function to predict protein–ligand interactions. *J. Mol. Biol.*, **295**, 337–356.
Goodford,P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.*, **28**, 849–857.
Graziano,G. and Lee,B. (2003) Entropy convergence in hydrophobic hydration: a scaled particle theory analysis. *Biophys. Chem.*, **105**, 241–250.
Huang,S.Y. and Zou,X. (2010) Inclusion of solvation and entropy in the knowledge-based scoring function for protein–ligand interactions. *J. Chem. Inf. Model*, **50**, 262–273.
Huang,S.-Y. *et al.* (2010) Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys. Chem.*, **12**, 12899–12908.
Jain,A.N. (1996) Scoring noncovalent protein–ligand interactions: a continuous differentiable function tuned to compute binding affinities. *J. Comput. Aided Mol. Des.*, **10**, 427–440.
Kauzmann,W. (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein chem.*, **14**, 1–63.
Klosowski,J.T. *et al.* (1998) Efficient collision detection using bounding volume hierarchies of k-DOPs. *IEEE Trans. Visual. Comput. Graph.*, **4**, 21–36.
Koes,D.R., Baumgartner,M.P. and Camacho,C.J. (2013) Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model*, **53**, 1893–1904.
Korb,O. *et al.* (2009) Empirical scoring functions for advanced protein–ligand docking with PLANTS. *J. Chem. Inf. Model*, **49**, 84–96.
Krammer,A. *et al.* (2005) LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graph Model*, **23**, 395–407.
Kuntz,I.D. (1992) Structure-based strategies for drug design and discovery. *Science*, **257**, 1078–1082.
Liu,H.Y. and Zou,X. (2006) Electrostatics of ligand binding: parametrization of the generalized Born model and comparison with the Poisson–Boltzmann approach. *J. Phys. Chem. B*, **110**, 9304–9313.
Lum,K. *et al.* (1999) Hydrophobicity at small and large length scales. *J. Phys. Chem. B*, **103**, 4570–4577.
Meyer,E. *et al.* (2003) Interactions with aromatic rings in chemical and biological recognition. *Angewandte Chemie International Edition*, **42**, 1210–1250.
Mooij,W.T. and Verdonk,M.L. (2005) General and targeted statistical potentials for protein–ligand interactions. *Proteins*, **61**, 272–287.
Morris,G.M. *et al.* (1996) Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J. Comput. Aided Mol. Des.*, **10**, 293–304.
Moustakas,D.T. *et al.* (2006) Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput. Aided Mol. Des.*, **20**, 601–619.
Muegge,I. (2006) PMF scoring revisited. *J. Med. Chem.*, **49**, 5895–5902.
Muegge,I. and Martin,Y.C. (1999) A general and fast scoring function for protein–ligand interactions: a simplified potential approach. *J. Med. Chem.*, **42**, 791–804.
Neudert,G. and Klebe,G. (2011) DSX: a knowledge-based scoring function for the assessment of protein–ligand complexes. *J. Chem. Inf. Model*, **51**, 2731–2745.
Nicholls,A. *et al.* (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
Petrey,D. and Honig,B. (2000) Free energy determinants of tertiary structure and the evaluation of protein models. *Protein Sci.*, **9**, 2181–2191.
Pratt,L. and Chandler,D. (1977) Theory of the hydrophobic effect. *J. Chem. Phys.*, **67**, 3683–3704.
Rajamani,S. *et al.* (2005) Hydrophobic hydration from small to large lengthscales: Understanding and manipulating the crossover. *Proc. Natl Acad. Sci. USA*, **102**, 9475–9480.
Salaniwal,S. *et al.* (2007) Critical evaluation of methods to incorporate entropy loss upon binding in high-throughput docking. *Proteins*, **66**, 422–435.
Setny,P. *et al.* (2009) Dewetting-controlled binding of ligands to hydrophobic pockets. *Phys. Rev. Lett.*, **103**, 187801.
Sharp,K.A. *et al.* (1991) Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science*, **252**, 106–109.
Shoichet,B.K. *et al.* (1999) Ligand solvation in molecular docking. *Proteins*, **34**, 4–16.
Shrake,A. and Rupley,J.A. (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.*, **79**, 351–371.
Sinanoglu,O. (1981) Microscopic surface tension down to molecular dimensions and microthermodynamic surface areas of molecules or clusters. *J. Chem. Phys.*, **75**, 463–468.
Snyder,P. *et al.* (2013) Is it the shape of the cavity, or the shape of the water in the cavity? *Eur. Phys. J. Spec. Top.*, 1–39.
Snyder,P.W. *et al.* (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl Acad. Sci. USA*, **108**, 17889–17894.
Southall,N. *et al.* (2001) A view of the hydrophobic effect. *J. Phys. Chem. B*, **106**, 521–533.
Stillinger,F. (1973) Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J. Sol. Chem.*, **2**, 141–158.
Tanford,C. (1979) Interfacial free energy and the hydrophobic effect. *Proc. Natl Acad. Sci. USA*, **76**, 4175–4176.
Tang,Y.T. and Marshall,G.R. (2011) PHOENIX: a scoring function for affinity prediction derived using high-resolution crystal structures and calorimetry measurements. *J. Chem. Inf. Model*, **51**, 214–228.
Tolman,R.C. (1949) The effect of droplet size on surface tension. *J. Chem. Phys.*, **17**, 5.
Velec,H.F. *et al.* (2005) DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.*, **48**, 6296–6303.
Wang,R. *et al.* (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.*, **16**, 11–26.
Wang,R. *et al.* (2004) The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures. *J. Med. Chem.*, **47**, 2977–2980.
Wang,R. *et al.* (2005) The PDBbind database: methodologies and updates. *J. Med. Chem.*, **48**, 4111–4119.
Wei,B.Q. *et al.* (2002) A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.*, **322**, 339–355.
Zhang,C. *et al.* (2005) A knowledge-based energy function for protein–ligand, protein–protein, and protein–DNA complexes. *J. Med. Chem.*, **48**, 2325–2335.
Zheng,Z. and Merz,K.M. Jr (2013) Development of the knowledge-based and empirical combined scoring algorithm (kecsa) to score protein–ligand interactions. *J. Chem. Inf. Model*, **53**, 1073–1083.
Zhou,H. and Skolnick,J. (2011) GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. *Biophys. J.*, **101**, 2043–2052.