

RESEARCH ARTICLE

Open Access



HawkRank: a new scoring function for protein–protein docking based on weighted energy terms

Ting Feng¹, Fu Chen¹, Yu Kang¹, Huiyong Sun¹, Hui Liu¹, Dan Li¹, Feng Zhu¹ and Tingjun Hou^{1,2*}

Abstract

Deciphering the structural determinants of protein–protein interactions (PPIs) is essential to gain a deep understanding of many important biological functions in the living cells. Computational approaches for the structural modeling of PPIs, such as protein–protein docking, are quite needed to complement existing experimental techniques. The reliability of a protein–protein docking method is dependent on the ability of the scoring function to accurately distinguish the near-native binding structures from a huge number of decoys. In this study, we developed HawkRank, a novel scoring function designed for the sampling stage of protein–protein docking by summing the contributions from several energy terms, including van der Waals potentials, electrostatic potentials and desolvation potentials. First, based on the solvation free energies predicted by the Generalized Born model for ~ 800 proteins, a SASA (solvent accessible surface area)-based solvation model was developed, which can give the aqueous solvation free energies for proteins by summing the contributions of 21 atom types. Then, the van der Waals potentials and electrostatic potentials based on the Amber ff14SB force field were computed. Finally, the HawkRank scoring function was derived by determining the most optimal weights for five energy terms based on the training set. Here, MSR (modified success rate), a novel protein–protein scoring quality index, was used to assess the performance of HawkRank and three other popular protein–protein scoring functions, including ZRANK, FireDock and dDFIRE. The results show that HawkRank outperformed the other three scoring functions according to the total number of hits and MSR. HawkRank is available at <http://cadd.zju.edu.cn/programs/hawkrank>.

Keywords: Protein–protein interaction, Docking, Scoring, HawkRank

Background

Protein–protein interactions (PPIs) are involved in a wide variety of biological processes, such as signal transduction [1, 2], transmembrane transport [3, 4], and antibody–antigen pairing [5, 6]. Deciphering structural and energetic determinants of PPIs is a prerequisite to understanding the PPIs-mediated functions in living cells. Unfortunately, only a tiny fraction of protein–protein complex structures have been characterized by high-resolution experimental techniques, such as X-ray crystallography, solution nuclear magnetic resonance

(NMR) spectroscopy and cryo-electron microscopy (cryo-EM), which cannot keep pace with the growing demand in structure-based interactome analysis. Moreover, many weak and/or transient PPIs that play essential roles in regulating dynamic networks in bio-systems cannot be easily captured by experiments due to their unstable nature. On that account, computational approaches, especially protein–protein docking, are expected to provide an alternative and efficient way based on the unbound protein structures for predicting the binding complexes and understanding the recognition mechanisms at the atomic level [7–9].

The ultimate goal of protein–protein docking is the prediction of a near-native structure of the complex from many docking decoys, which generally falls into two stages: sampling and refinement. In the sampling stage, a

*Correspondence: tingjunhou@zju.edu.cn; tingjunhou@hotmail.com

¹ College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China

Full list of author information is available at the end of the article

large number of docking poses are generated and scored by various scoring functions; and in the refinement stage, the top-hit poses (or decoys) given by the first stage are re-scored and re-ranked by more rigorous scoring functions. Apparently, the success of protein–protein docking is, to a large degree, dependent on the ability of the scoring function to score and rank the decoys accurately. So far a large number of scoring functions have been developed, ranging from force field-based scoring functions such as ZRANK and FireDock [10–13], to knowledge-based ones such as dDFIRE and InterEvScore [14–16] and machine-learning scoring functions [17, 18]. However, recognizing near-native structures from a huge pool of alternatives is still quite challenging because the accuracy of most scoring functions needs to be improved. Besides, the ease of use, efficiency and general utility of the scoring functions should also be taken into account. Since the establishment of the Critical Assessment of PRedicted Interactions (CAPRI) campaign [19] in 2001 offers a community-wide platform that assesses the accuracy of protein–protein docking approaches, all related scoring functions and algorithms can be evaluated by comparing the submitted structures with the unpublished crystal structures from wide range of participants including predictors, servers and scorers. In 2010, Kastiris and Bonvin assessed the performances of 9 commonly used scoring functions and a free energy prediction algorithm on their ability to predict the binding affinities for 81 complexes [20]. They found that all the tested scoring functions could not provide reliable predictions because they all failed to correlate the experimental binding affinities (pK_d) with the scores predicted by the corresponding scoring function, with the highest correlation of only -0.32 . Recently, our group analyzed the prediction results for the 24 targets tested from ROUND14 to ROUND 28 of CAPRI [21], and we found that, although the scorers perform better than the uploaders and predictors, they could give relatively high success rates ($> 50\%$) for only two targets. Therefore, more approaches should be explored in order to improve the prediction accuracy of scoring functions for more reliable protein–protein docking.

In the past decade, more theoretically rigorous free energy calculation methods, such as Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) and Molecular Mechanics/Generalized Born Surface Area (MM/GBSA), have been employed to predict binding affinities and identify correct binding structures for protein–protein systems [22–29]. For example, in our previous study [22], we evaluated the performances of MM/PBSA and MM/GBSA to predict the binding affinities and recognize the near-native binding structures for more than forty protein–protein complexes. The results

show that, compared with most scoring functions used in protein–protein docking, MM/GBSA achieved better accuracy to predict the correct binding modes and binding affinities for the studied protein–protein systems. Therefore, the desolvation energy, which is related to the leading role of solvent exclusion during the protein intermolecular assembly, is critical to identify these correct binding poses.

Although MM/GBSA is of more computational efficiency than other end-state free-energy calculation methods like thermodynamic integration (TI) and free energy perturbation (FEP), it is still much more time-consuming than the commonly used scoring functions in protein–protein docking, such as ZRANK, which treats the desolvation energy term with Atomic Contact Energy (ACE) model [30]. The computational cost in MM/GBSA is mainly attributed to the calculation of the polar desolvation energy term based on the GB model. In that regard, we developed HawkRank, a force field-based scoring function in which the energy terms are similar to those in MM/GBSA. Besides the frequently used van der Waals and electrostatic potentials, a simplified aqueous solvation model based on SASA (solvent accessible surface areas) was implemented into our scoring function. HawkRank is designed for the sampling stage of protein–protein docking and it can score a huge number of docked structures with low computational cost and high efficiency. We developed and benchmarked the present scoring function based on 176 high-resolution protein–protein complexes that are nonredundant at the family–family pair level. Compared with ZRANK, FireDock and dDFIRE, HawkRank performs consistently best on both the total number of hits and the (modified) success rate.

Methods

HawkRank was developed by combining the weighted van der Waals potentials, electrostatic potentials and desolvation potentials. The workflow of the development of HawkRank is discussed below in details.

Preparation of the protein–protein decoy dataset

More and more protein–protein complexes have been discovered, researchers classify protein–protein complexes based on various angles. The most common is that classify protein–protein complexes by protein family. The other researchers classify complexes as homo- and hetero-oligomeric complexes, non-obligate and obligate complexes and transient and permanent complexes, by the type of protein–protein interaction [31]. Besides, some researchers also excavate many effective statistical knowledge from the interface of the protein–protein interaction, such as the reported by Ref [32] and Ref [33]. Therefore, collecting a protein–protein complex database is a challenging

task, by reason that, comprehensive consideration including protein family, the type of protein–protein interaction or characteristics of interface is needed. However, there are still some databases pick out protein–protein complexes for theoretical research, such as protein–protein complexes in PDBbind [34], 2P2I-DB [35, 36], ZDOCK benchmark [37] and etc. The protein–protein complexes in ZDOCK benchmark 4.0 [38] were chosen to develop the HawkRank scoring function in our study. ZDOCK benchmark 4.0 provides 176 nonredundant protein–protein complexes with high-resolution X-ray or NMR structures in bound and unbound states at the family–family pair level, including 124 complexes in the previous version 3.0 plus 52 newly-added ones. Besides, the protein–protein complex dataset used in training ZARNK and FireDock is the same ZDOCK benchmark series. More than that, the structured files of predictions docked by the unbound receptor and ligand from the ZDOCK benchmark are available from Zlab official website (<https://zlab.umass-med.edu/zdock/benchmark.shtml>), which is convenient for the training of scoring function. Therefore, we choose ZDOCK benchmark 4.0 for its convenience and the rationality of comparing HawkRank with ZRANK and FireDock. In this study, the 124 complexes were used as the training set to develop HawkRank and the other 52 ones as the test set to validate the actual performance of HawkRank. It should be noted that the benchmark 4.0 only contains binary interactions, so HawkRank is not suitable for the interactions between more than two proteins. Besides, HawkRank is also not suitable for the interactions between protein and peptide.

ZDOCK (version 3.0) was used to generate the decoys for each complex. ZDOCK systematically evaluates a huge number of docked conformations on a grid by using a combination of shape complementarity, electrostatics and statistical potential terms for scoring [39], and the search process is accelerated by the Fast Fourier Transformation (FFT) algorithm [11]. Depending on the sampling density in the rotational space (15° or 6°), ZDOCK can output 3600 or 54,000 predictions for each system. Benchmark 4.0 was downloaded from Zlab official website. For each system, the unbound RCSB Protein Data Bank (PDB) files of the receptor and ligand are provided in Benchmark 4.0. Cases in Benchmark 4.0 have been docked using ZDOCK3.0 and the results depending on the sampling density in the 6° rotational space are deposited in decoys_bm4_zd3.0_6 deg package file which can be downloaded from Zlab official website. The missing hydrogen atoms in the unbound structures were added by using the *reduce* program (version 3.24) [40]. The decoys for each complex were generated by the Perl script in decoys_bm4_zd3.0_6 deg offered by ZDOCK, and a total of 54,000 decoys sorted by the ZDOCK scores

were generated for each system. It should be noted that for each system only the top scored 10,000 decoys were used in our analysis.

Criteria to evaluate the performance of protein–protein docking

Generally, in a protein–protein complex, the smaller protein is defined as the ligand protein and the larger one as the receptor protein. In our study, two types of root mean square deviations (RMSDs) between the predicted structure and the corresponding crystal structure, including ligand RMSD (L_RMSD) and interface RMSD (I_RMSD), were used as the criteria to evaluate the performance of protein–protein docking. L_RMSD, which is calculated over the C_α atoms of the ligand proteins when the receptors are superposed, was used to assess the global geometric fit between the predicted and native conformations [41]. I_RMSD, calculated over the C_α atoms of the interfacial residues when the predefined interfacial residues are superposed, was used to measure the geometric fit of the interface regions [41]. The interfacial residues in a protein–protein complex are defined as the residues within 10 \AA of any atom in another protein [42]. The L_RMSDs and I_RMSDs were calculated by using the *ProFit* program [43], which employs the McLachlan algorithm in fitting. The structures of the protein–protein complexes were predicted from the unbound proteins, and therefore the structures of the proteins in the crystal complexes and the predicted complexes may have obvious difference.

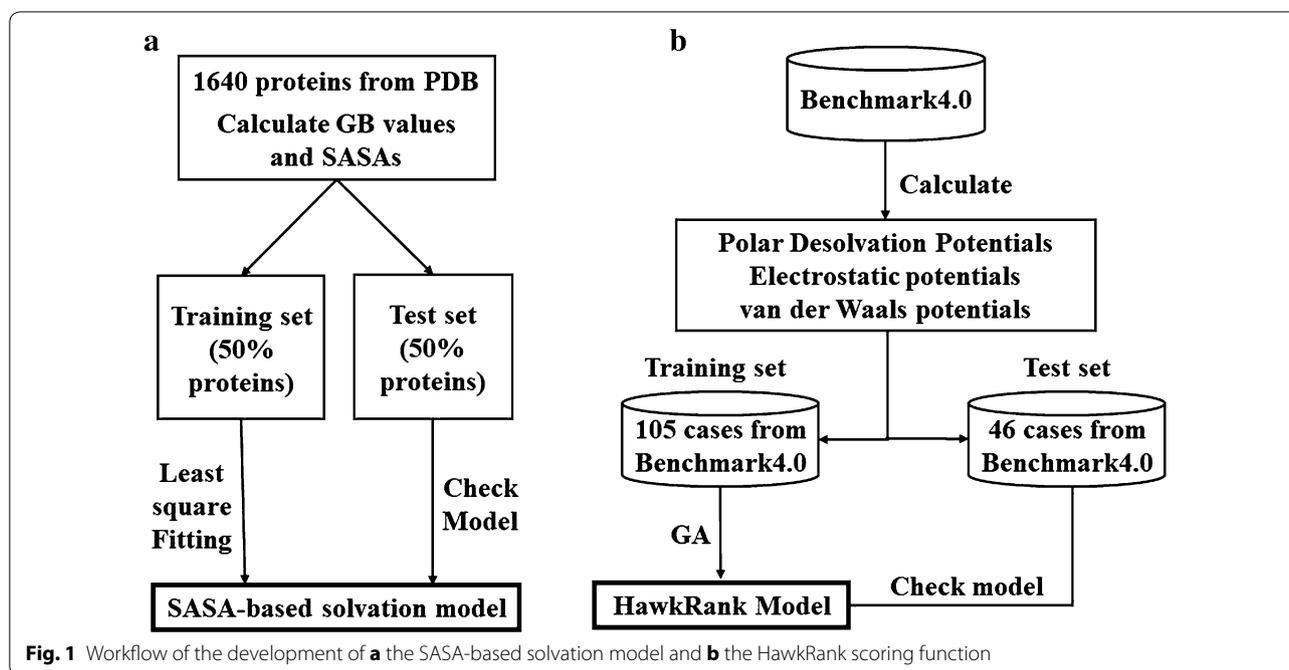
The criteria to evaluate the performance of protein–protein docking for Target 107 in Round 35 of the CAPRI campaign are summarized in Additional file 1: Table S1. Based on these criteria, the predictions can be classified into several categories: incorrect, acceptable, medium, and high quality predictions. In this study, the hits are the predictions with L_RMSD less than 10 \AA or I_RMSD less than 4 \AA , which follows the criteria used in CAPRI.

Parameterization of the SASA-based solvation model

Because of the high computational cost of MM/GBSA, particularly the GB calculation, the use of MM/GBSA as a scoring function to rank thousands to even millions of docked conformations is computationally unaffordable. In order to balance computational cost and prediction accuracy, a novel SASA-based solvation model was developed by fitting the solvation free energies of proteins predicted by GB, as shown in Fig. 1a.

Generation of the protein dataset for the GB calculation

A protein dataset was established for the GB calculations (referred to as the GB dataset). The 1640 proteins in this dataset were selected from PDB based on the following



criteria: (1). the structures are obtained by X-ray experiment, (2). the resolution should be lower than 2 Å, (3) the proteins do not contain any small molecule ligands, (4). the proteins are asymmetric, and (5) the proteins do not contain any modified residue.

Besides, for multiple structures whose sequences have at least 30% sequence identity, only a single structure is included in the dataset. Moreover, the protein structures with multiple conformations were eliminated.

Calculation of polar solvation free energies based on the GB model

The electrostatic/polar solvation free energies for the proteins in the GB dataset were computed by using the GB model implemented in Amber14. First, all missing heavy atoms and hydrogens of the proteins were added by using the *tleap* program in Amber14, and then the partial charges and force field parameters of the ff14SB force field were assigned. Subsequently, all proteins were optimized by 5000 cycles of minimization (2500 cycles of steepest descent and 2500 cycles of conjugate gradient). At last, the electrostatic solvation free energy for each protein was computed by using the modified GB model developed by Onufriev and colleagues (referred to as GB^{OBC1}) [44]. A value of 80 was used for the exterior dielectric constant, and 1 was used for the solute dielectric constant.

Definition of atom types

Based on the study reported by Hou et al. [45], all atoms in the 20 standard amino acids were classified into 21

atom types shown in Table 1. The atoms that have similar chemical environment are defined as the same atom type.

Parameterization of the SASA-based solvation model

For each protein in the GB dataset, the SASAs were calculated by the NSC function that is an implementation of the DCLM (Double Cubic Lattice Method) variant of the Shrake's and Rupley's algorithm for surface calculation [46]. The van der Waals radii for H, C, N, O and S are listed in Additional file 1: Table S2 [47], and the radius of the solvent probe was defined to 0.6 Å, which performed best to develop the solvation model based on SASAs [48]. Consequently, the SASAs for the 21 different atom types could be calculated, and the SASA-based solvation model given by Equation function (1) was developed.

$$psol_{SASA} = w_0 N_{res} + \sum_{i=1}^{21} w_i SASA_i + \text{constant} \quad (1)$$

$$i = 1, 2, 3 \dots 21$$

where $psol_{SASA}$ represents the polar component of solvation free energy; i is the IDs (Table 1) of the atom types for a given protein; w_i is the weight for atom type i ; $SASA_i$ is the sum of SASAs for atom type i in the given protein; N_{res} represents the total number of the residues in a given protein, and w_0 is the weight of N_{res} ; constant is the constant term in this formula. The whole GB dataset was randomly split into a training set (~ 800 proteins) with 50% of the proteins and a test set with the

Table 1 The definitions of the 21 atom types of proteins used in the SASA-based solvation model

ID	Weight	Feature	ID	Weight	Feature	ID	Weight	Feature
1	-0.96		8	0.77		15	-1.74	The rest of C atoms
2	-3.32		9	-0.05		16	7.81	
3	-0.55		10	-6.82		17	0.94	
4	0.33		11	2.69		18	-0.13	
5	9.34		12	-1.51		19	-0.23	
6	0.73		13	1.89		20	0.84	
7	-3.69		14	2.63		21	0.23	The rest of H atoms

(1) α represents C that is connected with only one O atom; (2) β represents C with sp³ hybridization; (3) X is O or S atom; (4) Y is N or C atom; (5) R is H or C atom; (6) Atoms at E1, E2 and E3 are N, O or S

remaining proteins. Least-square fitting was applied to optimize the weights of different atom types and N_{res} based on the training set, and then the reliability of the developed solvation model was evaluated by the consistency between the predictions given by the GB model and those given by the SASA-based model to the test set.

Development of the HawkRank scoring function

The HawkRank scoring function is a linear weighted sum of the van der Waals attractive and repulsive potentials, the electrostatic attractive and repulsive potentials, and the desolvation potentials. The core formula of HawkRank is shown in Equation function (2), where ΔE is the value of potentials and w is the weight of corresponding potentials:

$$\begin{aligned} \text{HawkRank Score} = & w_{vdW_{attr}} \Delta E_{vdW_{attr}} + w_{vdW_{repu}} \Delta E_{vdW_{repu}} \\ & + w_{elec_{attr}} \Delta E_{elec_{attr}} + w_{elec_{repu}} \Delta E_{elec_{repu}} \\ & + w_{pdsol} \Delta E_{pdsol} \end{aligned} \quad (2)$$

The details about the energy terms used in the scoring function are described below. The workflow of the development of HawkRank is shown in Fig. 1b.

Calculation of the van der Waals potentials

The definition of the van der Waals potentials (Equation function 3) is the same to that used in Kortemme's study [49]. The formula was modified from the classical 6-12 Lennard-Jones potential. As shown in Equation function (3), the van der Waals potentials are split into the attractive and repulsive parts. The repulsive part was calculated by the linear formula to reduce the potential local clashes caused by the discrete rotamers of side chains and fixed backbones [49], and the attractive part was defined as the traditional formula of the classical 6-12 Lennard-Jones potential. To improve computational efficiency, only the van der Waals interactions between two atoms with distance less than 12 Å were calculated.

$$\begin{aligned} E_{attr} = \varepsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right\} & \text{ for } 0.89\sigma_{ij} < r_{ij} < 12 \text{ \AA} \\ E_{repu} = 10.0 \left(1 - \frac{r_{ij}}{0.89\sigma_{ij}} \right) & \text{ for } 0.89\sigma_{ij} > r_{ij} \end{aligned} \quad (3)$$

where i and j are atoms i and j , r_{ij} is the interatomic distance between them, σ_{ij} is the sum of atomic radii, and ε_{ij} is the well depth derived from the Amber ff14SB force field [50].

Calculation of the electrostatic potentials

For the electrostatic potentials, the standard Coulomb equation is used (Equation function 4):

$$E_{elec} = 332 \frac{q_i q_j}{\epsilon r_{ij}} \quad (4)$$

where i and j are atoms i and j , r_{ij} is the interatomic distance between them, q_i and q_j are the atomic partial charges afforded by the Amber ff14SB force field [50]. The dielectric constant ϵ was set to 1. The electrostatic potentials are also split into an attractive and a repulsive parts. If the E_{ele} between two atoms is less than zero, it will be added into the electrostatic attractive potentials, and otherwise added into the electrostatic repulsive potentials.

Calculation of desolvation potentials

The SASA-based solvation model developed in this study was used to compute the solvation potentials of the unbound receptor, unbound ligand and complex for each system. Then, the desolvation (ΔE_{pdsol}) was calculated by taking the polar solvation free energy of complex ($psol_{SASA(com)}$) to subtract the sum of the polar solvation free energies of receptor ($psol_{SASA(rec)}$) and ligand ($psol_{SASA(lig)}$) (Equation function 5).

$$\Delta E_{pdsol} = psol_{SASA(com)} - (psol_{SASA(lig)} + psol_{SASA(rec)}) \quad (5)$$

Training the scoring function

For each case in ZDOCK benchmark 4.0, the I_RMSDs and L_RMSDs for the top 10,000 decoys were calculated. After eliminating those cases whose hits could not be found in the top 10,000 decoys, 151 cases out of 176 were included in the whole dataset. The 105 cases out of the 124 found in ZDOCK benchmark 3.0 were put into the training set and the remaining 46 cases into the test set. Then, based on the training set, genetic algorithm (GA) implemented in the R package (genalg) was used to determine the optimal set of the weights for the energy terms used in HawkRank. We set the maximum (3.0) and minimum (0.0) value for the weight of the van der Waals attractive potentials, electrostatic attractive potentials, electrostatic repulsive potentials and the polar desolvation potentials. However, we set the maximum (0.001) and minimum (0.0) value for the weight of the van der Waals repulsive potentials, on account of inherent defect in the equation function to calculate the van der Waals repulsive potentials and we want to reduce the impact of the van der Waals repulsive potentials on the whole scoring function. Besides, the population size was set as 200, the number of iterations was set as 1500 and the number of chromosomes that are kept into the next generation was about 20% of the population size.

Evaluation of the performance of HawkRank

The capability of HawkRank to recognize the near-native poses from the decoys was compared with those of two popular force field-based scoring functions used in protein–protein docking, ZRANK [11] and FireDock [12], and a knowledge-based scoring function named dDFIRE [15].

ZRANK is a force field-based scoring function that is a linear combination of atom-based potentials, including electrostatics, van der Waals, and desolvation potentials. Pairwise Atomic Contact Energy (ACE) model [30] is used to calculate the desolvation energy. Parameters used to calculate the van der Waals and desolvation potentials in the ZRANK scoring function are derived from the CHARMM 19 polar hydrogen force field.

FireDock is a method for the refinement and rescoring of rigid-body docking solutions. The function of FireDock includes ACE, softened van der Waals interactions, electrostatic interactions and internal energy. Moreover, hydrogen and disulfide bonds, π -stacking and aliphatic interactions are also considered [12]. Compared with ZRANK and HawkRank, FireDock includes more energy terms.

dDFIRE is an all-atom statistical and knowledge-based energy function. Each polar atom is treated as a dipole with a direction. The orientation of the dipole is defined by the bond vectors that connect the polar atoms with other heavy atoms, and the function of dDFIRE is extracted from protein structures based on the distance between two atoms and the three angles involved in dipole–dipole interactions [15]. Besides, the hydrogen bonding interactions are considered in dDFIRE via the physical dipole–dipole interactions.

It is well known that how to quantitatively evaluate the performance of a scoring function is quite essential. Traditionally, success rate (SR), the proportion of total cases with at least one hit in the top N predictions, has been widely used to evaluate the performance of a given scoring function [16]. However, SR has its own intrinsic deficiency. For example, for two different protein–protein complexes, one has nine hits in the top 100 predictions while the other has only one hit in the top 100 predictions. However, when we calculate SR for these two systems, the contributions of the predictions for these two complexes to SR are identical, but nonetheless, the capacity of the scoring function to correctly rank the predictions for these two complexes is quite different. Moreover, SR ignores the rank for each hit. The “top-ranking” method is also a popular way to evaluate the performance of a given scoring function by identifying the first hit in the ranked predictions and comparing their ranks. The two evaluation strategies mentioned above are relatively intuitional, but not quite reasonable.

In order to evaluate the ranking performance of a scoring function on the whole benchmark in a reasonable way, MSR (modified success rate) proposed by us (Equation function 6) [21] was used in this study.

$$Y = \sum F(x_i)/\text{size}(X) \quad (6)$$

$$F = \frac{\sum_{j=1}^N (1 + \text{top}_j)}{\text{HIT}} \quad (7)$$

$$\text{top}_j = \frac{1}{\text{rank}_j} \quad (8)$$

where x_i represents a protein–protein complex, $\text{size}(X)$ is the total number of the studied systems in a benchmark, HIT equals to the total number of the hits for a protein–protein complex, rank_j is the rank number of a hit, top_j is the reciprocal of rank_j , Y is the integrated value to evaluate the performance of the scoring function for multiple protein–protein complexes, and F is the value to evaluate the performance of the scoring function for a specific protein–protein complex.

Results and discussion

Performance of the SASA-based solvation model

On the whole, the SASA-based solvation model (Equation function 2) yields satisfactory results for the training set ($r = 0.96$). The plot of the polar solvation free energies predicted by the SASA-based model *versus* those predicted by the GB model is shown in Fig. 2a. Moreover, the model illustrates good predictive ability on the test set ($r = 0.93$, Fig. 2b), suggesting that the developed model is reliable and not overfitting. The weights (w_i) for the 21 atom types derived by fitting the GB polar solvation free energies for the training set are listed in Table 1. The w_0 and the constant term determined by fitting are -10.51 and 59.78 , respectively.

Performance of HawkRank

Based on the training set, the optimal weights for the five types of potentials were determined by GA as follows: 0.850 for the van der Waals attractive potentials, 0.0005 for the van der Waals repulsive potentials, 1.887 for the electrostatic attractive potentials, 1.853 for the electrostatic repulsive potentials, and 0.9303 for the polar desolvation potentials. Then, the trained HawkRank scoring function was used to score the decoys for the test set. Among the 46 cases in the test set, the correct binding structures for 2 cases (1CLV, 2AYO) are the best hits, those for 6 cases (1CLV, 1SYX, 2AYO, 2OUL, 3D5S, 4CPA) can be found in the top 5 ranked predictions and those for 13 cases can be found in the top 20 ranked predictions. Then, we calculated the SR values ($N = 10$,

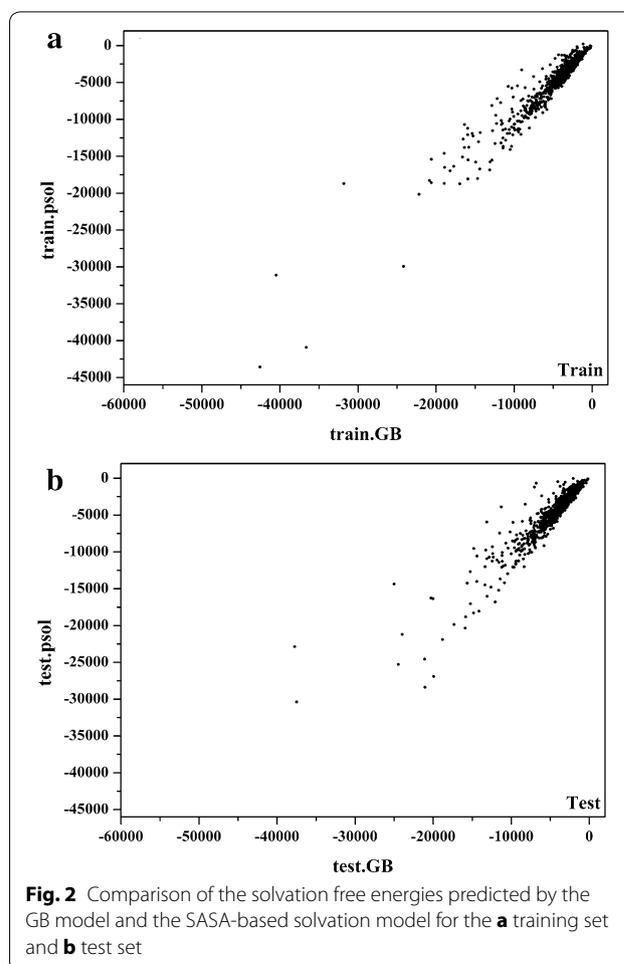
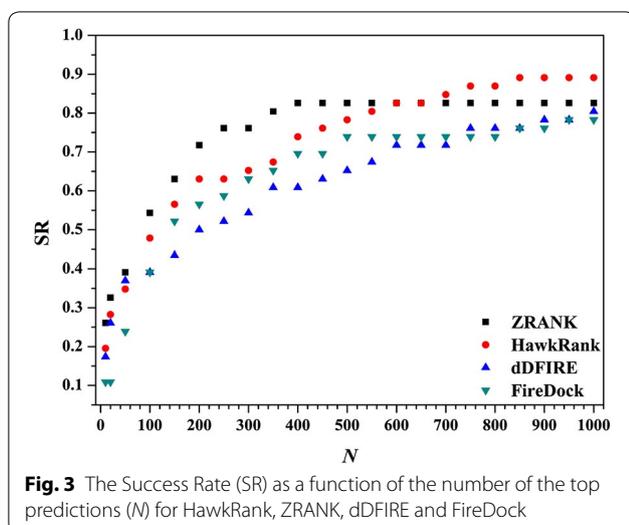


Fig. 2 Comparison of the solvation free energies predicted by the GB model and the SASA-based solvation model for the **a** training set and **b** test set

20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, and 1000) for the 46 cases in the test set. As shown in Fig. 3, HawkRank's SR is higher than 50% when N is larger than 150 while ZRANK's SR is higher than 50% when N is larger than 100. When N is larger than 600, HawkRank's SR value equals to or exceeds ZRANK's SR value. According to SR, the performances of HawkRank and ZRANK are comparative. However, HawkRank's SR values are always higher than those of dDFIRE and FireDock.

Next, we analyzed the performance of each individual energy term used in HawkRank from the view of MSR. The Y values of the five energy terms are shown in Additional file 1: Figure S1. The performances of the two repulsive potentials are not satisfactory, but the attractive ones perform much better, especially the electrostatic attractive potentials. When $N \leq 200$, the polar desolvation potentials show good performance, and its performance decreases gradually with the increase of N .

The performance of HawkRank was then compared with those of ZRANK, FireDock and dDFIRE. The Y values



(from $N = 20$ to $N = 2000$) were calculated for each scoring function. Overall, as shown in Fig. 4, HawkRank outperforms ZRANK, FireDock and dDFIRE for both of the training set and test set. In details, when N is small than 160, HawkRank shows comparative performance with ZRANK and dDFIRE, but when N becomes higher than 160 it performs better. With N increases, the Y values of HawkRank increase much faster than those of the other three scoring functions. Therefore the HawkRank scoring function can identify more hits with better ranking in the top $N = 160$ – $20,000$ than the other three scoring functions. Besides, the performance of FireDock performs apparently worse than the other three scoring functions (Fig. 4). In summary, we can make the following conclusion: for a force field-based scoring function, inclusion of more energy terms may not be necessary to achieve better ranking capabilities. The energy terms used in HawkRank are similar to those used in ZRANK while the major difference lies in the desolvation term. In HawkRank, a SASA-based solvation model was used to calculate the polar desolvation potentials, instead of the frequently-used ACE model used in ZRANK. The better performance of HawkRank over ZRANK was believed to be attributed to the better estimation of protein–protein desolvation free energy. Therefore we believe that, by fitting the solvation free energies predicted by GB, the SASA-based model should provide more reliable predictions than ACE.

Comparison of Pearson correlations between RMSDs and scores

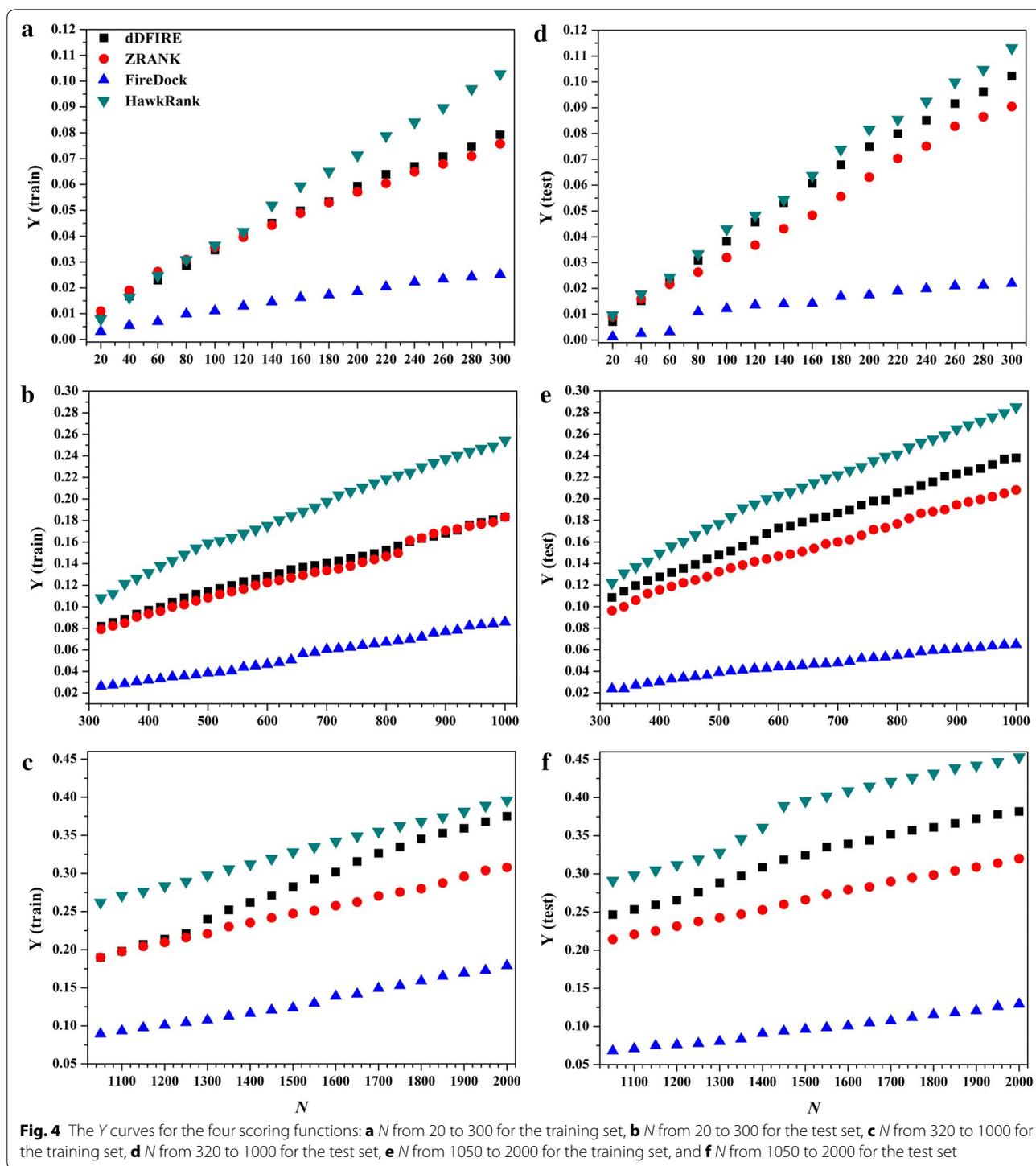
Generally, the smaller the RMSD between the predicted complex structure and its native conformation is, the more similar these two structures are. The calculation of RMSD requires a prior superimposition of the predicted

complex structure onto its native conformation, but the superimposition that minimizes the global RMSD of the predicted complex to the native conformation may be not necessarily the best solution: such superimposition is often compromised by a small number of significantly deviating fragments [51]. Therefore, we cannot exclude the possibility that a good docked structure with low energy has relatively larger I_RMSD or L_RMSD. It is expected that a lower RMSD is associated with a better docking score and vice versa. Therefore for a good scoring function which has lower scores for better predictions, a positive correlation between RMSDs and the docking scores should be observed. Then, the correlations between I/L_RMSDs and the docking scores predicted by ZRANK, FireDock, dDFIRE or HawkRank were analyzed for the 176 cases in the dataset. Here, r_I is referred to as the correlation between the scores and I_RMSDs and r_L as the correlation between the scores and L_RMSDs. Here, the r_I or r_L values were partitioned into the following four intervals: $0 < r_I/r_L \leq 0.2$, $0.2 < r_I/r_L \leq 0.4$, $0.4 < r_I/r_L \leq 0.6$, and $0.6 < r_I/r_L \leq 0.8$. The results are shown in Fig. 5. Moreover, the cumulative occurrences of the r_I/r_L values are listed Table 2. For HawkRank, r_L achieves 0.4 for 13 cases and r_I achieves 0.4 for 23 cases in the training set, and r_L achieves 0.4 for 9 cases and r_I achieves 0.4 for 11 cases in the test set. For dDFIRE, r_L achieves 0.4 for 3 cases and r_I achieves 0.4 for 3 cases in the training set, and r_L achieves 0.4 for 5 cases and r_I achieves 0.4 for 5 cases in the test set. ZRANK's r_I/r_L and FireDock's r_I/r_L even cannot achieve 0.4 for any case. Apparently, according to the cumulative occurrences of r_I/r_L shown in Table 2, HawkRank achieves much better performance than the other three scoring functions.

Three examples (1JIW, 2AYO, and 1SYX) with $r_I/r_L > 0.4$ and three examples (2FJU, 1PVH, and 1GXD) with $r_I/r_L < 0$ are shown in Figs. 6 and 7, respectively. In next section, we also found that 2AYO and 1SYX were predicted accurately while 2FJU and 1GXD were predicted inaccurately by HawkRank. We also calculated the correlations (r) between I/L_RMSDs and each energy term given by HawkRank, and the numbers of the cases in four intervals of r are also displayed in Fig. 8. It seems that the electrostatic attractive potentials have better correlation with I/L_RMSDs than the other energy terms, indicating that the electrostatic attractive potentials contribute more to the total scoring than any other energy term. In summary, the HawkRank scoring function provides a better funnel-shaped energy landscape than the other three scoring functions.

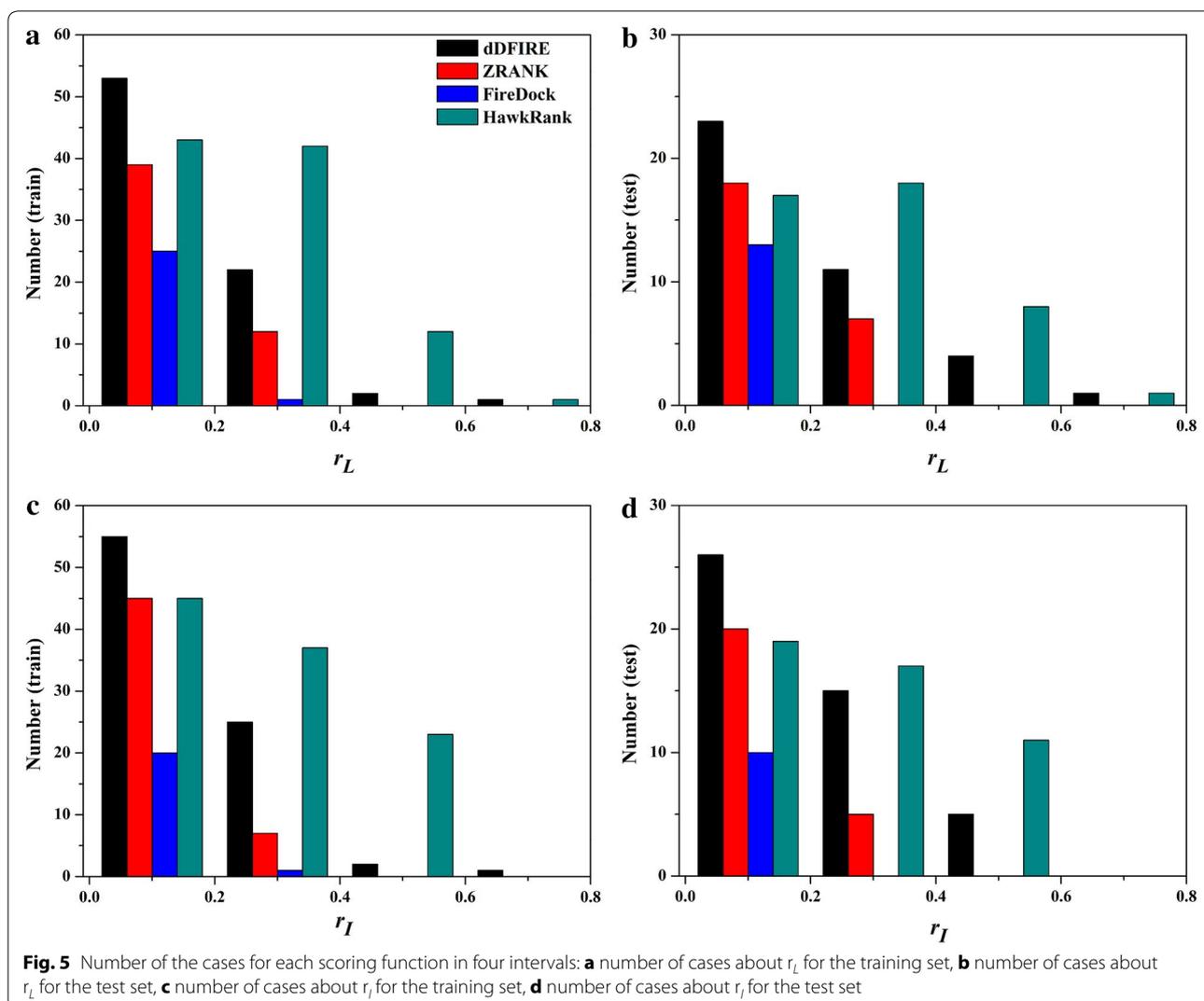
Limitations of HawkRank

Although HawkRank has better ranking capability than ZRANK, FireDock and dDFIRE, it still has some limitations. The $F(N = 500)$ values for the 46 tested cases with the



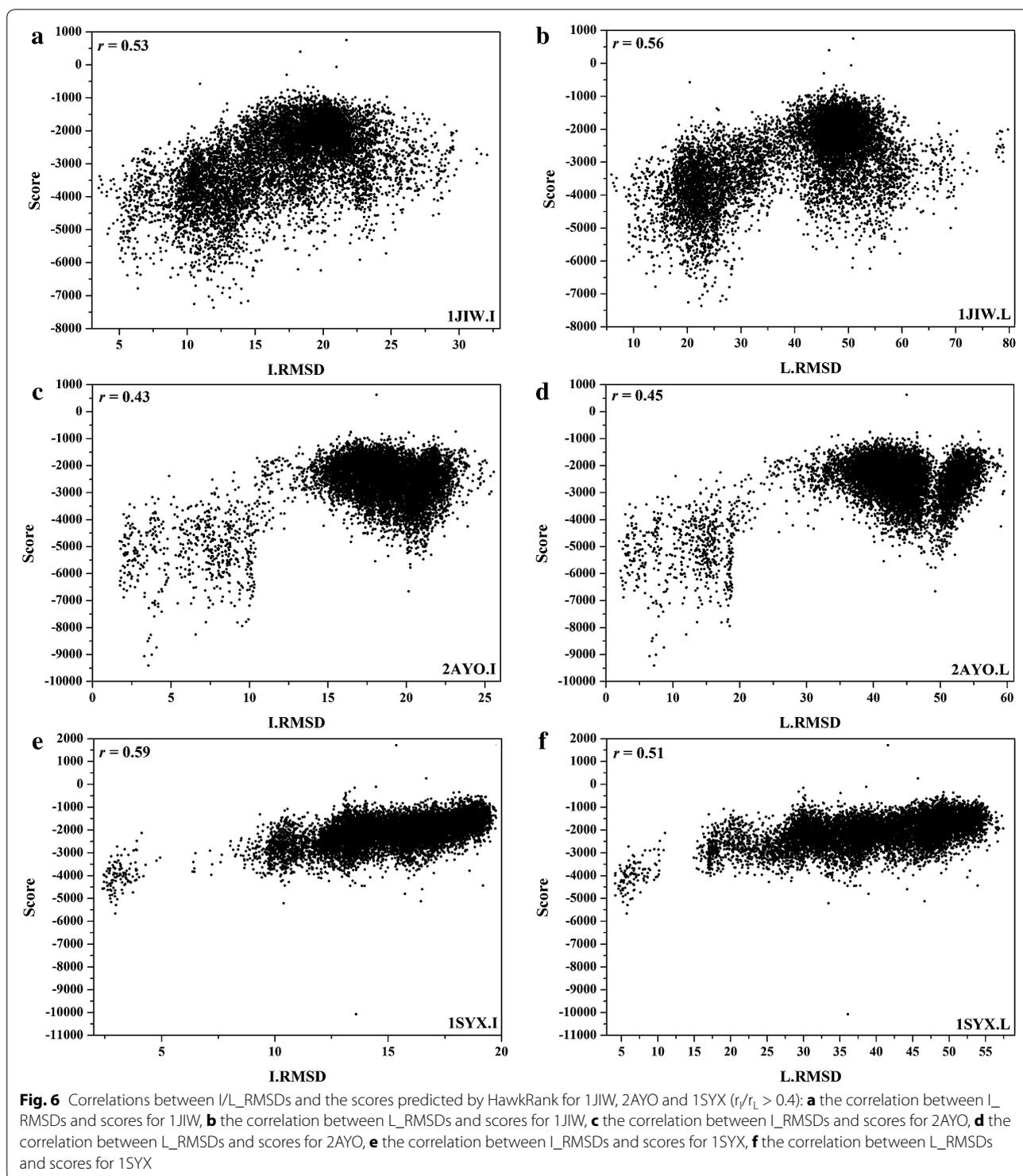
hits found in the top 10,000 decoys are listed in Additional file 1: Table S3. The 46 cases are ranked by the HawkRank's F values in ascending order. $F = 0$ means that the scoring function does not find any hit in the top 500 predictions. For dDFIRE, ZRANK, FireDock and HawkRank, the numbers

of the cases with $F = 0$ are 16, 8, 12 and 10, respectively. For 1XU1, 1ZHH, 2B4 J and 2OOR, the four scoring functions cannot find any hit in their top 500 predictions. For 1GXG, 1JZD, 1US7 and 2FJU, more than 60 hits docked by ZDOCK, but HawkRank performs poorly on them.

**Table 2** Number of the cases for each scoring function in different intervals of r^a

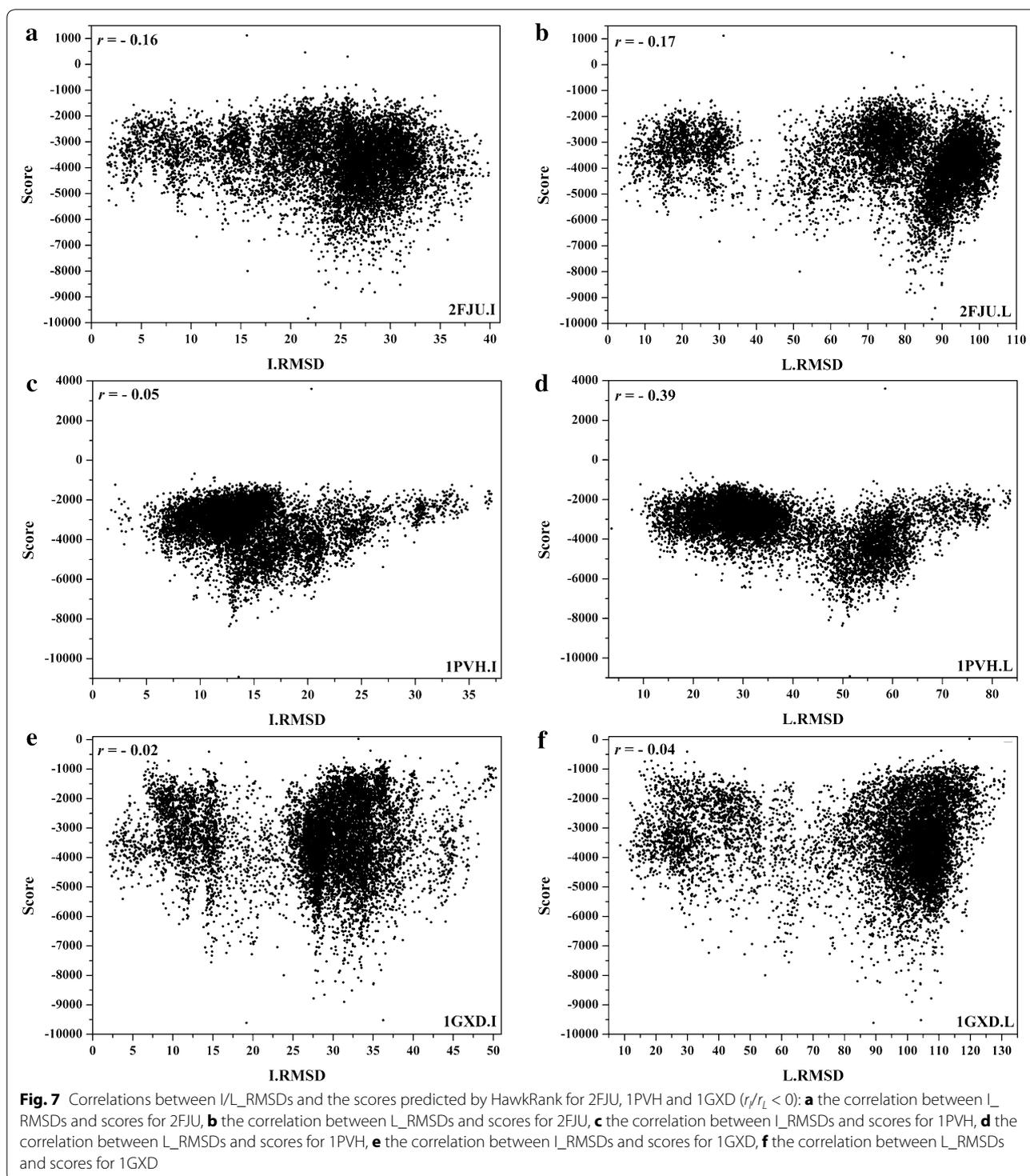
	dDFIRE	ZRANK	FireDock	HawkRank
<i>Training set (124 cases)</i>				
$0.6 < r \leq 0.8$	1 ^a	1 ^a	0	0
$0.4 < r \leq 0.8$	3	3	0	0
$0.2 < r \leq 0.8$	25	28	12	7
$0.0 < r \leq 0.8$	78	83	51	52
<i>Test set (52 cases)</i>				
$0.6 < r \leq 0.8$	1	0	0	0
$0.4 < r \leq 0.8$	5	5	0	0
$0.2 < r \leq 0.8$	16	20	7	5
$0.0 < r \leq 0.8$	39	46	25	25

^a For each scoring function, the first column is the number for r_L , and the second column is the number for r_I



The COCOMAPS online service [52] was used to analyze the protein–protein interfaces for the 46 cases in the test set. We found that the four cases (1GXDX, 1JZD, 1US7 and 2FJU) that cannot be successfully predicted by HawkRank have quite low percentages of near-native

interface areas relative to the total surface areas. The four cases (1SYX, 2AYO, 2OZA and 3D5S) that can be well predicted by HawkRank have relatively high percentages of near-native interface areas relative to the total surface areas. In other words, for the cases with



the higher percentages of near-native interface areas, HawkRank may provide better ranking capability. Then, we used the COCOMAPS online service to analyze the interface characteristics of the best solutions predicted by HawkRank for 1GXD, 1JZD, 1US7 and 2FJU. As shown in Table 3, the percentages of the interface areas

in the native complex structures for the four cases are much lower than those in the top ranked structures. Generally, for a protein–protein complex, larger interface area implies that the receptor and ligand can form more favorable surface complementary. Apparently, HawkRank is apt to recognize the decoys with larger

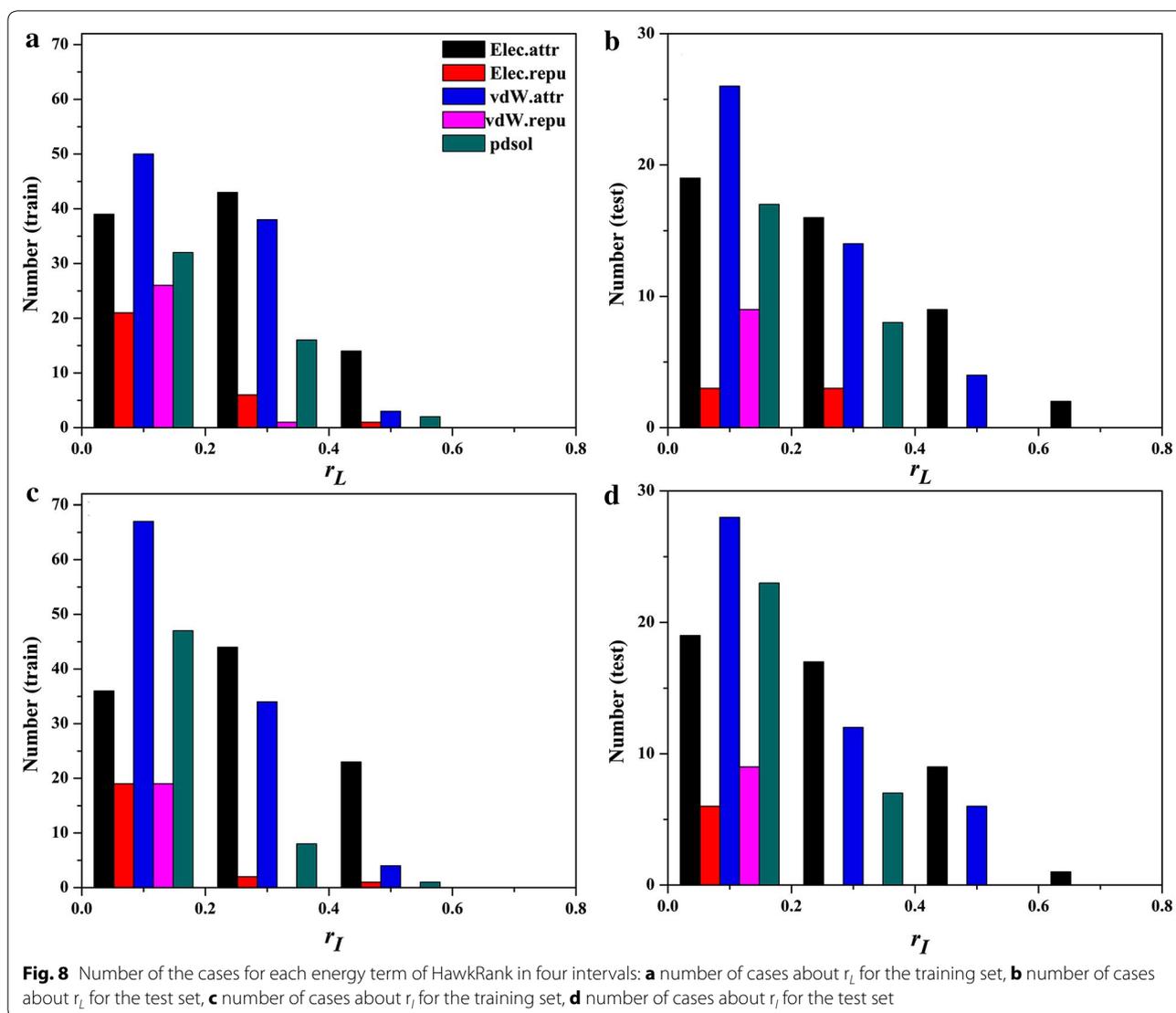


Table 3 Comparative analysis of the interface areas for 1GXD, 1JZD, 1US7 and 2FJU

	Native percentage (%) ^a	Spurious percentage (%) ^b
1GXD	2.885	5.665
1JZD	3.650	9.005
1US7	2.405	8.765
2FJU	1.611	4.470

^a The percentage of the near-native interface areas relative to the total surface areas of the native complex

^b The percentage of the interface areas relative to total surface areas of the best solutions predicted by HawkRank

interface area percentages in near-native structures. This is also the reason why HawkRank can find more hits in its top 500 predictions for these cases with larger

native interface areas such as 1SYX, 2AYO, 2OZA and 3D5S.

Computational cost of HawkRank

The arithmetic speed of HawkRank is closely related to the size of the protein–protein complex. The speed of HawkRank is fast enough to meet the requirements for scoring a large number of decoys in the sampling stage of protein–protein docking. The scoring of a complex with 500 residues needs less than 0.3 s on a core (Intel Xeon CPU E5-2692 v2 @2.20 GHz) with Linux operating system.

Conclusions

In this study, we developed a new scoring function named HawkRank by combining polar desolvation potentials, van der Waal potentials and electrostatic potentials.

HawkRank introduces a fast and effective way to calculate the desolvation potentials based on a SASA-based solvation model. Compared with ZRANK, FireDock and dDFIRE, HawkRank shows better ranking capabilities to the 46 cases in test set. Besides, the scores predicted by HawkRank have higher correlations with L/I_RMSDs than those predicted by ZRANK, FireDock and dDFIRE, suggesting that the HawkRank scoring gives a better funnel-shaped energy landscape than the other three scoring functions. Although its prediction accuracy still needs to be improved for some protein–protein complexes with small interface areas, HawkRank is efficient to meet current requirements for scoring a large number of decoys in the sampling stage of protein–protein docking. In the light of the above assessment and the conclusion in our previous study that MM/GBSA rescoring has good capability to distinguish the correct protein–protein binding structures from the decoys, it would suggest to be an efficient protocol of using HawkRank followed by the MM/GBSA rescoring to improve the predictions of protein–protein docking.

Additional file

Additional file 1: Contains the supplementary information of the manuscript. This includes **Table S1**. Criteria of assessing predictions in CAPRI (Å); **Table S2**. The van der Waals radii for H, C, N, O and S (Å); **Table S3**. The *F* values for the 46 cases in test set; **Figure S1**. The *Y* curves for the five energy terms in HawkRank: (A) *N* from 20 to 300 for the training set, (B) *N* from 20 to 300 for the test set, (C) *N* from 320 to 1000 for the training set, (D) *N* from 320 to 1000 for the test set, (E) *N* from 1050 to 2000 for the training set, and (F) *N* from 1050 to 2000 for the test set.

Authors' contributions

TF and TH designed and implement the algorithm. TF, FC and YK performed the simulations. TF, HS, HL, DL and FZ analyzed the data. TF and TH wrote the manuscript. All authors read and approved the final manuscript.

Author details

¹ College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China. ² State Key Lab of CAD&CG, Zhejiang University, Hangzhou 310058, Zhejiang, China.

Acknowledgements

We thank the National Supercomputer Center in Guangzhou (NSCC-GZ) for providing the computing resources.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This study was supported by the National Key R&D Program of China (2016YFA0501701), the National Science Foundation of China (81773632, 21575128, 81302679), and the Fundamental Research Funds for the Central Universities (2017QNA7032, 2017QNA7033, 2017QNA7034).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 September 2017 Accepted: 14 December 2017

Published online: 28 December 2017

References

- Arkin MR, Whitty A (2009) The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions. *Curr Opin Chem Biol* 13(3):284–290
- Pawson T, Nash P (2000) Protein-protein interactions define specificity in signal transduction. *Gene Dev* 14(9):1027–1047
- Hicke L, Dunn R (2003) Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. *Annu Rev Cell Dev Biol* 19:141–172
- Stone TA, Deber CM (2017) Therapeutic design of peptide modulators of protein-protein interactions in membranes. *Biochim Biophys Acta-Biomembr* 1859(4):577–585
- Peng HP, Lee KH, Jian JW, Yang AS (2014) Origins of specificity and affinity in antibody-protein interactions. *Proc Natl Acad Sci USA* 111(26):E2656–E2665
- Davies DR, Cohen GH (1996) Interactions of protein antigens with antibodies. *Proc Natl Acad Sci USA* 93(1):7–12
- Gong X, Wang P, Yang F, Chang S, Liu B, He H, Cao L, Xu X, Li C, Chen W, Wang C (2010) Protein-protein docking with binding site patch prediction and network-based terms enhanced combinatorial scoring. *Proteins* 78(15):3150–3155
- Moretti R, Fleishman SJ, Agius R, Torchala M, Bates PA, Kastriitis PL, Rodrigues JPGLM, Trellet M, Bonvin AMJJ, Cui M, Rooman M, Gillis D, Dehouck Y, Moal I, Romero-Durana M, Perez-Cano L, Pallara C, Jimenez B, Fernandez-Recio J, Flores S, Pacella M, Kilambi KP, Gray JJ, Popov P, Grudinin S, Esquivel-Rodriguez J, Kihara D, Zhao N, Korin D, Zhu X, Demerdash ONA, Mitchell JC, Kanamori E, Tsuchiya Y, Nakamura H, Lee H, Park H, Seok C, Sarmiento J, Liang S, Teraguchi S, Standley DM, Shimoyama H, Terashi G, Takeda-Shitaka M, Iwadate M, Umeyama H, Beglov D, Hall DR, Kozakov D, Vajda S, Pierce BG, Hwang H, Vreven T, Weng Z, Huang Y, Li H, Yang X, Ji X, Liu S, Xiao Y, Zacharias M, Qin S, Zhou H-X, Huang S-Y, Zou X, Velankar S, Janin J, Wodak SJ, Baker D (2013) Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. *Proteins* 81(11):1980–1987
- Huang S-Y, Yan C, Grinter SZ, Chang S, Jiang L, Zou X (2013) Inclusion of the orientational entropic effect and low-resolution experimental information for protein-protein docking in Critical Assessment of PRedicted Interactions (CAPRI). *Proteins* 81(12):2183–2191
- Camacho CJ, Zhang C (2005) FastContact: rapid estimate of contact and binding free energies. *Bioinformatics (Oxford, England)* 21(10):2534–2536
- Pierce B, Weng Z (2007) ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins* 67(4):1078–1086
- Andrusier N, Nussinov R, Wolfson HJ (2007) FireDock: fast interaction refinement in molecular docking. *Proteins* 69(1):139–159
- Cheng TM, Blundell TL, Fernandez-Recio J (2007) pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins* 68(2):503–515
- Zhou H, Zhou Y (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci Publ Protein Soci* 11(11):2714–2726
- Yang Y, Zhou Y (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins* 72(2):793–803
- Andreani J, Faure G, Guerois R (2013) InterEvScore: a novel coarse-grained interface scoring function using a multi-body statistical potential coupled to evolution. *Bioinformatics (Oxford, England)* 29(14):1742–1749
- Shih ES, Hwang MJ (2015) NPPD: a protein-protein docking scoring function based on dyadic differences in networks of hydrophobic and hydrophilic amino acid residues. *Biology* 4(2):282–297
- Khashan R, Zheng W, Tropsha A (2012) Scoring protein interaction decoys using exposed residues (SPIDER): a novel multibody interaction scoring

- function based on frequent geometric patterns of interfacial residues. *Proteins* 80(9):2207–2217
19. Janin J (2010) Protein-protein docking tested in blind predictions: the CAPRI experiment. *Mol Biosyst* 6(12):2351–2362
 20. Kastiris PL, Bonvin AM (2010) Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *J Proteome Res* 9(5):2216–2225
 21. Zhang Q, Feng T, Xu L, Sun H, Pan P, Li Y, Li D, Hou T (2016) Recent advances in protein-protein docking. *Curr Drug Targets* 17(14):1586–1594
 22. Chen F, Liu H, Sun H, Pan P, Li Y, Li D, Hou T (2016) Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. *Phys Chem Chem Phys* 18(32):22129–22139
 23. Gohlke H, Kiel C, Case DA (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaGDS complexes. *J Mol Biol* 330(4):891–913
 24. Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed Engl* 41(15):2644–2676
 25. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 51(1):69–82
 26. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J Comput Chem* 32(5):866–877
 27. Sun H, Li Y, Shen M, Tian S, Xu L, Pan P, Guan Y, Hou T (2014) Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys Chem Chem Phys* 16(40):22035–22045
 28. Sun H, Li Y, Tian S, Xu L, Hou T (2014) Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys Chem Chem Phys* 16(31):16719–16729
 29. Tian S, Li Y, Wang J, Xu X, Xu L, Wang X, Chen L, Hou T (2013) Drug-likeness analysis of traditional Chinese medicines: 2. Characterization of scaffold architectures for drug-like compounds, non-drug-like compounds, and natural compounds from traditional Chinese medicines. *J Cheminform* 5:5
 30. Zhang C, Vasmatzis G, Cornette JL, DeLisi C (1997) Determination of atomic desolvation energies from the structures of crystallized proteins. *J Mol Biol* 267(3):707–726
 31. Nooren IMA, Thornton JM (2003) Diversity of protein-protein interactions. *EMBO J* 22(14):3486–3492
 32. Janin J, Bahadur RP, Chakrabarti P (2008) Protein-protein interaction and quaternary structure. *Q Rev Biophys* 41(2):133–180
 33. Jubb H, Blundell TL, Ascher DB (2015) Flexibility and small pockets at protein-protein interfaces: new insights into druggability. *Prog Biophys Mol Biol* 119(1):2–9
 34. Liu Z, Li Y, Han L, Li J, Liu J, Zhao Z, Nie W, Liu Y, Wang R (2015) PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics* 31(3):405–412
 35. Basse MJ, Betzi S, Bourgeas R, Bouzidi S, Chetrit B, Hamon V, Morelli X, Roche P (2013) 2P2ldb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res* 41(D1):D824–D827
 36. Basse MJ, Betzi S, Morelli X, Roche P (2016) 2P2ldb v2: update of a structural database dedicated to orthosteric modulation of protein-protein interactions. *Database* 6:baw007
 37. Chen R, Mintseris J, Janin J, Weng ZP (2003) A protein-protein docking benchmark. *Proteins-Struct Funct Genet* 52(1):88–91
 38. Hwang H, Vreven T, Janin J, Weng Z (2010) Protein-protein docking benchmark version 4.0. *Proteins* 78(15):3111–3114
 39. Mintseris J, Pierce B, Wiehe K, Anderson R, Chen R, Weng Z (2007) Integrating statistical pair potentials into protein complex prediction. *Proteins* 69(3):511–520
 40. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 285(4):1735–1747
 41. Janin J (2005) Assessing predictions of protein-protein interaction: the CAPRI experiment. *Protein Sci Publ Protein Soc* 14(2):278–283
 42. Mendez R, Leplae R, Lensink MF, Wodak SJ (2005) Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. *Proteins* 60(2):150–169
 43. McLachlan A (1982) Rapid comparison of protein structures. *Acta Crystallogr A* 38(6):871–873
 44. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55(2):383–394
 45. Hou T, Zhang W, Huang Q, Xu X (2005) An extended aqueous solvation model based on atom-weighted solvent accessible surface areas: SAWSA v2.0 model. *J Mol Model* 11(1):26–40
 46. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) The double cubic lattice method—efficient approaches to numerical-integration of surface-area and volume and to dot surface contouring of molecular assemblies. *J Comput Chem* 16(3):273–284
 47. van der Bondi A (1964) Waals volumes and radii. *J Phys Chem* 68(3):441–451
 48. Wang JM, Wang W, Huo SH, Lee M, Kollman PA (2001) Solvation model based on weighted solvent accessible surface area. *J Phys Chem B* 105(21):5055–5067
 49. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci USA* 99(22):14116–14121
 50. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11(8):3696–3713
 51. Kufareva I, Abagyan R (2012) Methods of protein structure comparison. *Methods Mol Biol* 857:231–257
 52. Vangone A, Spinelli R, Scarano V, Cavallo L, Oliva R (2011) COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics (Oxford, England)* 27(20):2915–2916

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
