International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:6, No:6, 2012

Critical Assessment of Scoring Schemes for Protein-Protein Docking Predictions

Dhananjay C. Joshi, Jung-Hsin Lin

***Abstract*—**Protein-protein interactions (PPI) play a crucial role in many biological processes such as cell signalling, transcription, translation, replication, signal transduction, and drug targeting, etc. Structural information about protein-protein interaction is essential for understanding the molecular mechanisms of these processes. Structures of protein-protein complexes are still difficult to obtain by biophysical methods such as NMR and X-ray crystallography, and therefore protein-protein docking computation is considered an important approach for understanding protein-protein interactions. However, reliable prediction of the protein-protein complexes is still under way. In the past decades, several grid-based docking algorithms based on the Katchalski-Katzir scoring scheme were developed, e.g., FTDock, ZDOCK, HADDOCK, RosettaDock, HEX, etc. However, the success rate of protein-protein docking prediction is still far from ideal. In this work, we first propose a more practical measure for evaluating the success of protein-protein docking predictions,the rate of first success (RFS), which is similar to the concept of mean first passage time (MFPT). Accordingly, we have assessed the ZDOCK bound and unbound benchmarks 2.0 and 3.0. We also createda new benchmark set for protein-protein docking predictions, in which the complexes have experimentally determined binding affinity data. We performed free energy calculation based on the solution of non-linear Poisson-Boltzmann equation (nlPBE) to improve the binding mode prediction. We used the well-studied thebarnase-barstarsystem to validate the parameters for free energy calculations. Besides,thenlPBE-based free energy calculations were conducted for the badly predicted cases by ZDOCK and ZRANK. We found that direct molecular mechanics energetics cannot be used to discriminate the native binding pose from the decoys.Our results indicate that nlPBE-based calculations appeared to be one of the promising approaches for improving the success rate of binding pose predictions.

***Keywords*—**protein-protein docking,protein-protein interaction, molecular mechanics energetics, Poisson-Boltzmann calculations

I. Introduction

BETTER understanding of machinery of life is achieved by in-depth studies of proteins. Although the functions of individual proteins are important for understanding this machinery, we usually also need to move up to a higher level, i.e., protein-protein interactions (PPI), which areindispensable in understanding almost all the cellular processes such as cell signaling, transcription, translation, replication, signal transduction, and drug targeting etc.

Dhananjay C. Joshi is with the Taiwan International Graduate Program for Chemical Biology and Molecular Biophysics, Academia Sinica, 115 Taiwan (phone: +886 2 2782 3212 ext 877; fax: +886 2 2782 3060; e-mail: joshidc@gmail.com).

Jung-Hsin Lin is with the Research Center for Applied Sciences & Institute of Biomedical Sciences, Academia Sinica, Taiwan; School of Pharmacy, National Taiwan University (e-mail: jhlin@gate.sinica.edu.tw, jlin@ntu.edu.tw).

Structural information about protein-protein interaction is essential for understanding the molecular mechanisms of these processes. However, structures of protein-protein complexes are still difficult to obtain by biophysical methods, such as NMR and X-ray crystallography. Therefore, computational protein-protein docking is considered as an importantapproach for elucidating protein-protein interactions. Another emerging field of biosciences is peptide-based drug designing or protein therapies. Engineered peptides could be used as inhibitor for some PPI. However, similar to protein-protein complexes, structural information of the protein-peptide complex is also sometimes difficult to obtain. Hence, it is preferred to carry out *in-silico* studies to reduce the efforts and optimize the biophysical studies.

Although*in-silico* prediction of such protein-protein and protein-small-peptide interaction is still challenging, recent progress in protein-protein docking studies suggested several directions towards future research. In the past two decades, many grid-based docking algorithms were developed. These algorithms employ efficient search and scoring schemes such as Fast Fourier Transform (FFT)(e.g., the Katchalski-Katzir scheme [1]) for correlation function evaluation, Monte-Carlo methods, geometric hashing, etc. Electrostatics, desolvation, and hydrophobic effect have been incorporated in several scoring functions. Despite of all the efforts, selecting the correct binding pose from the huge decoy data set, is still a big challenge.

Nevertheless, there are many famous docking suites and algorithms that have shown significant progress in predicting near-native binding poses by making better use of biophysical and biochemical information in combination with bioinformatics. The information such as protein-protein interaction data bases, alanine scanning, conserved sequence data bases etc., in combination with machine learning approach, helps to identify hot-spots for protein-protein interactions. Subsequently, there are several issues directly or indirectly related to protein-protein docking, e.g., protein conformational flexibility[2], interfacial water molecules[3], atomic radii optimization [4], implicit versus explicit solvent and water dielectric constant[5], etc.,which make the route towards reliable docking predictions more curved and rugged. Hence, predicting correct bio-molecular complex is still a formidable task.

The protein-protein docking procedure could be usually divided into two parts, rigid body docking and flexible docking. Most of the docking suites employ rigid body docking procedure as a first step. In rigid body docking, the protein is considered as non-flexible rigid body. The protein's coordinates are discretized into a three dimensionalCartesian grid. The grid cells are sorted out based on whether they belongto the surface or the core of a protein. Further, surface

grid cells of protein and ligand are used to score the degree of overlap between them for different orientations and relative positions[1]-[6]. This procedure involves translational and rotational search. The search is very slow and therefore accelerated by well-known Fast Fourier Transform (FFT) technique [1] to accelerate the translational search. The methods of predicting near-native complex vary depending upon various strategies. On the other hand, in the flexible docking part, the flexibility of side chain, backbone, etc., were also taken care in various ways. Well-known software that implements varieties of these strategies includesFTDock[7],[8],[9]; HADDOCK[10],[11][12]; ICM[13];RosettaDock, [14], [15], [16];HEX[17], [18],[19], [20];ZDOCK[21], [22],[23], [24], [25]; etc.

FTDock (Fourier Transform Dock) is the docking algorithm that uses simplified grid representation[7]-[9]. It implements the Fourier correlation theory based on Katchalski-Katziralgorithm[1] plus an electrostatics function amenable to Fourier correlation. It outputs multiple predictions that can be screened using biochemical information. Furthermore, RPScore (Residue-level Pair potential Score) uses a single distance constraint empirically derived pair potential to screen the output from FTDock. The algorithms are fully embedded in 3D-DOCK docking suite. As an extension to this approach, MultiDock (Multiple copy side-chain refinement Dock) is also available to improve quality of prediction.

HADDOCK (High Ambiguity Driven protein-protein Docking) approach makes use of biochemical and/or biophysical interaction data such as, for example, chemical shift perturbation data obtained from NMR titration experiments or mutagenesis data. The information on the interacting residues is introduced as ambiguous interaction restraints (AIRs) to drive the docking. The predicted complex structures are ranked according to their intermolecular energy, i.e., sum of electrostatic, van der Waals, and AIR energy terms. Thus, this approach demonstrates the usefulness of AIR. However, new version HADDOCK2.0 has been modified to support docking of proteins, DNA,RNA, oligosaccharides, and small ligand, up to a total of six separate molecules (or domains) per docking. The new version allows the inclusion of anisotropy restraints from NMR (both residual dipolar couplings and relaxation data) and supports solvated docking, that is, allowing the explicit inclusion of interfacial water molecules in the docking process[10].

ICM is another software suite that is facilitated with many tools. The basic algorithm includes ODA (Optimal Docking Areas) method that predicts protein-protein interaction sites on protein surfaces. This calculation involves desolvation energy.It identifies optimal surface patches with the lowest docking desolvation energy values as calculated by atomic solvation parameters (ASP) derived from octanol/water transfer experiments and adjusted for protein-protein docking. First, correct solution with lowest energy confirmation is found by docking rigid ligand (all-atom) molecule to a set of soft receptor. The potentials are pre-calculated on a 0.5 Å grid from realistic solvent-corrected force-field energies. The inclusion ofthe induced changes, as well as the optimization of the interface side-chains of up to 400 best solutions, take place in second step.The third step is the filtering step, in which information available from the experiment is implemented. However, the algorithm is less successful if the backbone undergoes large scale rearrangements [13].

RosettaDock is the software suite that provides online server facility for protein-protein docking. The docking algorithm mimics the physical process of docking, i.e., it contains a low-resolution recognition stage and a high-resolution binding stage. The high-resolution refinement simultaneously optimizes the rigid-body displacement and the side-chain conformations. In this suite the Rosetta techniques were adapted and expanded for docking problems. The algorithm includes a fast search using low-resolution potentials followed by an atomic-scale refinement step incorporating simultaneous optimization of side-chain positions and rigid-body displacement. The process mimics the steps involved in a diffusional encounter between two macromolecules, although the treatment is certainly not a rigorous physical simulation. Scoring functions include both physical and physically inspired statistical potentials derived from structures in the Protein Data Bank (PDB)[26]. Small-perturbation studies are employed to examine the quality of the scoring function [27]-[28].

HEX software suite handles the docking problem in a little different way. FFT based algorithms can speed up the calculationstremendously. However, it is not readily feasible to incorporate the prior knowledge about complex and focus on them. HEX uses closed-form 6 D Spherical Polar Fourier (SPF) correlation expressions, from which arbitrary multi-dimensionalmulti-propertymulti-resolution FFT correlations can be generated. The approach is demonstrated by calculating 1D, 3D and 5D rotationalcorrelations of 3D shape and electrostatic expansions up topolynomial order L=30 on a 2 GB personal computer. The SPF approach provides a natural way to define one or two simple angular constraints with which to focus docking searches around known or hypothesized binding sites. This accelerates the calculation and can significantly reduce the number of false-positive predictions. The approach provides a practicaland fast tool for rigid body protein-protein docking, especiallywhen prior knowledge about one or both binding sites is available. With online HexServer facility, recently, HEX have implemented Graphical Processing Unit (GPU) version to accelerate the calculations[17].

ZDOCK is oneof the successful suites that has shown great prediction abilities in Critical Assessment of PRedicted Interactions (CAPRI) [29].ZDOCK uses a fast Fourier transform to search all possible binding modes for the proteins, evaluating based on shape complementarity, desolvation energy, and electrostatics. The top 2000 predictions from ZDOCK are then given to RDOCK where they are minimized by CHARMM to improve the energies and eliminate clashes, and then the electrostatic and desolvation energies are recomputed by RDOCK (in a more detailed fashion than the calculations performed by ZDOCK). However, RDOCK approach is very time consuming as it involves molecular force-field based energy minimization of macromolecules. In the new protocol of ZDOCK, the rescoring scheme, ZRANK

[23]has been introduced. It utilizes detailed electrostatics, van der Waals, and desolvation to rescore initial-stage docking predictions.Weights for the scoring terms were optimized for a set of test cases, and this optimized function was then tested on an independent set of non-redundant cases[22].

Scoring functions bearing the information of binding affinity isgenerally more useful because they can be used to discriminate the binding complexes from non-binding complexes, especially when the shape complementarity is the only factor contributing to the binding. Many research groups are concerned about experimental data to include in scoring functions. Dissociation constant or binding affinity is one of those experimental data's that is considered as an important criterion for many new benchmarks [30]. However, poor correlation between experimental and calculated binding affinity indicates the need to improve the scoring functions [31]. It has been shown that the combination of rescoring and refinement significantly improves the protein docking performance [25].Also, molecular recognition and binding are the two main steps involved in protein complex formation. The recognition step must depend upon long range forces. Electrostatic complementarities are responsible for recognizing correct binding pose and the other factors such as short ranges van der Waals and hydrophobic interactions are involved stabilization of the formed complex.

In this study, we first propose a new benchmark set, in which experimentally determined binding affinity data are available. Specifically, we have assessed the performance of ZDOCK and ZRANK benchmark on Benchmark 2.0 and 3.0 data sets,as well as our new benchmark data set. Further, we studies a new scoring scheme based on the solution of non-linear Poisson-Boltzmann equation. We used the well-studied barnase-barstar complex to address some of the issues related to optimization ofthe parameters for free energy calculations.

II. Methods

A. ZDOCK and ZRANK assessment

Benchmark 2.0 [32]and 3.0 [33]are used for assessment. The benchmarks were obtained from thehttp://zlab.umassmed.edu/zdock/benchmark.shtmlwebpage. The set consists of bound and unbound cases. Initially, the bound and unbound cases are separated out and the surfaces were marked using the mark_surf script. Some of the residues and ligands (non-protein molecules) were excluded. Further, ZDOCK is run in two sets. In the first set,the lower sampling density with resolution of $15^0$was used, whichyield 3600 top predictions. In second set, the higher density sampling with 54000 predictions with resolution of 6^0 were obtained only forbarnase-barstar case (PDB ID 1BRS). All the 3600 predictions were re-ranked using ZRANK[23].In contrast to other assessment criteria, here we propose a more practical measure for the success of the docking predictions, the rate of first success (RFS), which is similar to the concept of mean first passage time (MFPT), and the quantity is normalized by the total number of predictions so that predictions with different density of sampling can be compared. This measure was proposed in view of the fact that the top predictions of most protein-protein docking algorithms or ranking schemes are usually not to correct binding poses.

The first rate of success would indicate the average number of experiments (mutagenesis, chemical-crosslinking, etc.,) needs to be conducted in order to find out the correct protein-protein binding pose. Here, thepredicted complexes were scanned from the top in the 3600 ranked prediction listuntil the root-mean-squared prediction (RMSD)is smaller than the defined threshold (in this study 3.0 Å). Plots were made using matplotlib module of python. For benchmark 3.0, the plots for ZDOCK and ZRANK bound and unbound cases appears to be a single red line, in fact, red and blue lines are overlapping.

B. A new benchmark for protein-protein docking

A newbenchmark set for protein-protein docking prediction, with binding affinity information, designated as PPIbind, was extracted from thePDBbind[34] with the three criteria: First, only two chains are present in the protein-protein complex. Second, no small molecule or chemical compound is present in the complex. Third, there should only be one biological assembly in this PDB entry.A total of 62 complexes were included in thePPIbindbenchmark set. Fig.1 shows the distribution of binding affinity of 1371 protein-protein complexes out of 1441 from PDBbind(released on September 22, 2011) and the 62 complexes in PPIbind having K_d/K_i values. It can be seen that the binding free energy distribution of PPIbind stays in the central region of the binding affinity distribution of PDBbind, which indicates that PPIbind, although a much smaller dataset, can well represent diverse protein-protein interactions.

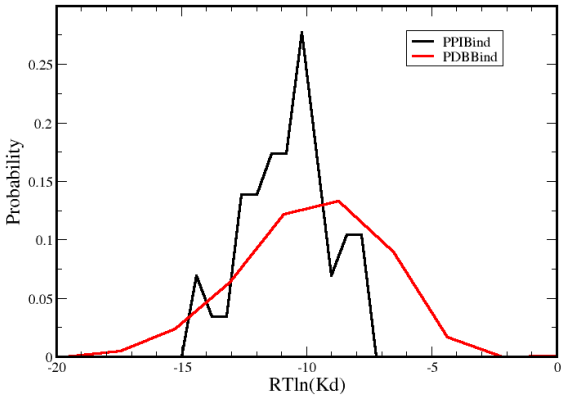

Fig. 1 Distribution of binding free energies of the 1371 protein-protein complexes from PDBbind data set (released on September 22, 2011) and PPIbind. The unit of the x-axis is kcal/mol

C. AMBER Energetics

$$\Delta E_{ele} = E_{ele,complex} - E_{ele,receptor} - E_{ele,ligand}$$
$$\Delta E_{vdw} = E_{vdw,complex} - E_{vdw,receptor} - E_{vdw,ligand}$$

The amber energies were obtained from 0 step MD simulations using AMBER Molecular Dynamics package [35].

D. APBS Energetics

The adaptive Poisson-Boltzmann solver(APBS) is used to calculate the electrostatic contribution of free energies[36]. In general, calculating binding free energies divides the binding process up into desolvation and Coulombic components:

$$\Delta\Delta_{bind}G = \Delta\Delta_{solv}G + \Delta\Delta_{coul}G$$

where,

$$\Delta\Delta_{solv}G = \Delta_4 G + \Delta_2 G$$
$$= \Delta_{solv}G_{complex} - \Delta_{solv}G_{mol1} - \Delta_{solv}G_{mol2}$$

$$\Delta\Delta_{coul}G = -\Delta_1 G$$
$$= \Delta_{coul}G_{complex} - \Delta_{coul}G_{mol1} - \Delta_{coul}G_{mol2}$$

Where $\Delta\Delta_{solv}G$ is the solvation energy and $\Delta\Delta_{coul}G$ is the electrostatic energy of complex formation (refer Fig. 4 for more details).

E. Barnase-Barstar analysis

The barnase-barstar protein complex (PDB ID: 1BRS)was used for in depth studies. The biological unit was obtained from the Protein Data Bank [26].

III. Results and Discussions

A. Assessment of Various Benchmarks

Predicting correct binding pose in protein-protein docking is indeed a big challenge. In order to address this problem, we performed critical assessment of the benchmarks using the well-known ZDOCK software suite and the ZRANK re-ranking scheme. We used benchmark 2.0 and 3.0 as well as our own K_d based benchmark from PDBbinddataset for the assessment.

The assessments onZDOCK benchmark 2.0 and 3.0 setswere carries out as described in the Methods. The data sets have bound and unbound cases. In benchmark 2.0 there are 84 sets of complexes [32] whereas in benchmark 3.0 there are 40 complexes more in addition to 2.0 [33]. Thus total of 124 complexes were used for the assessment.

Initially, the surface of the ligand and the receptors were marked and subjected to ZDOCK. Total of 3600 predictions were obtained and the RMSD (root-mean-square-deviation) is calculated from the native complex. As mentioned in Methods, the rate of first success was recorded for each case. Further, the predictions were re-ranked using ZRANK scheme and the new rate was determined for each data set.

We used a stringent criterion to evaluate the correct binding pose. If the RMSD < 3.0 Å the complex is considered to be near native prediction. The plots for the rate of first success(Fig. 2) show two curves, where the blue curve was the ranking by the ZDOCK score,and the red curveby the ZRANK score. The area under the curve decides the extent of success of docking. Larger the area under the curve, less the top ranked complexes needed to scan, thus higher success rate. Less area under the curve implies more scanning and more success.

In bound cases the plots for benchmark 2.0 and 3.0,Fig. 2, (a) and (b), shows that the re-scoring scheme helps to improve the success rate. In the unbound cases, theplot of the rate of first success shows the similar trend as that of the bound case (Fig 2. (c) and (d)).

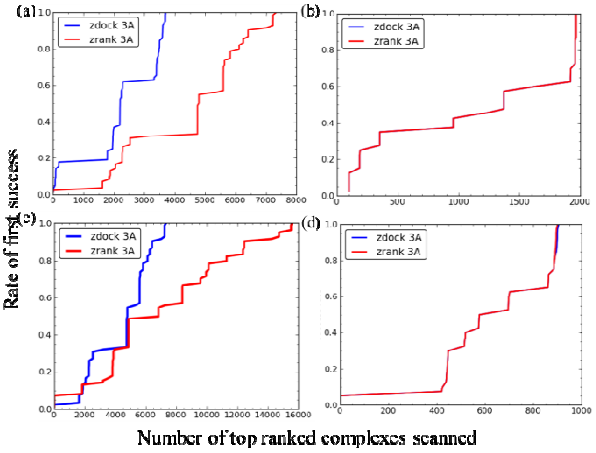

Fig. 2 Comparison of the rate of first successwith ZDOCK benchmark data set: ZDOCK and ZRANK comparison for bound (a) 2.0, (b) 3.0 and unbound (c) 2.0, (d) 3.0 benchmarks. (In 3.0 bound and unbound cases, the red and blue lines are almost perfectly overlapped)

On comparing the performance of ZDOCK, the unbound cases are difficult to rescore than that of bound cases (Fig. 2, blue plots). Same is with ZRANK, as the area under the curve in the unbound set is more than that of the bound set (Fig. 2 red plots). In bound 2.0, ZRANK rescoring scheme appears to be effective, which is not same in unbound cases, however. Both ZDOCK and ZRANK in benchmark 2.0 and 3.0 are overlapping suggesting that the re-ranking is not improving the ZDOCK prediction any more.

B. Assessment on PPIbind

Similar to above ZDOCK benchmark 2.0 and 3.0, we performed critical assessment for our benchmark. This benchmark is essentially a bound benchmarkwith entries extracted from thePDBbind database [34] and the three criteria mentioned in the Method section. Basically, PDBbind database provides a comprehensive collection of the experimentally measured binding affinity data for all types of bio-molecular complexes deposited in the Protein Data Bank (PDB). As we are interested only in protein-protein interactions in this study, we choose protein-protein binding affinity dataset. There are 1441 entries were found according to out criterion. The binding free energy distribution is shown in Fig. 1. Most of the binding free energyvalues spans between -20 to -2 kcal/mol, suggesting a large diversity in protein-protein interactions. However, 1441 is a significantly larger number than the number of complexes in ZDOCK benchmark sets, and it is often difficult to judge which biological unit of a given PDB entry represents the true binding scenario. On the other hand, in some protein-protein interactions, small chemical molecules could play some roles in facilitating or inhibiting the binding. Finally, to simplify the scenario for protein-protein docking, only binary complexes are considered. Nevertheless, the binding free energy distribution of PPIbind overlaps with the central region of the binding free energy distribution of the protein-protein interactions in PDBbind.

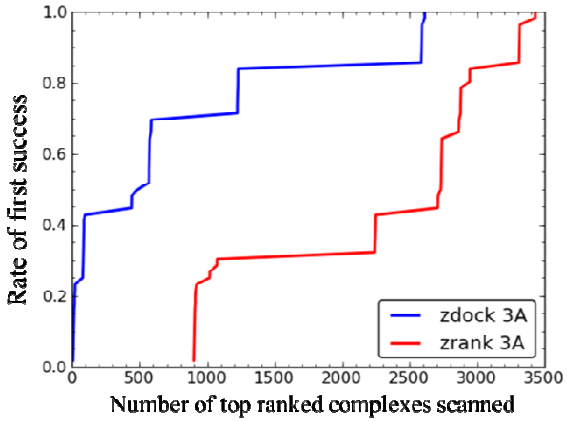

Fig. 3 The rate of first success of ZDOCK and ZRANK on thePPIbind benchmark set, which can be considered an external test set

The PDB IDs were extracted from the PDB data bank based on the criterion as described in Methods. The dataset is further subdivided based on number of biological units available in the PDB data base. We found that out of 64 cases, one has 3; seven have 2 and rest have one biological unit each. Each complex has experimentally determined K_d (dissociation constant) or Ki(inhibitor constant) values. The protein-protein complexes were separated in to subunit according to chain ID. The larger subunit is treated as the receptor and the smaller treated as the ligand.

We performed ZDOCK and ZRANK prediction and found that the plots of the rate of first successfollow similar pattern as that of Benchmark 2.0 and 3.0 (Fig. 3). Similar to ZDOCK benchmarks, ZDOCK has larger area under the curve as compare to ZRANK suggesting that the ZRANK re-ranking does not help to improve the prediction according to RFS definition.

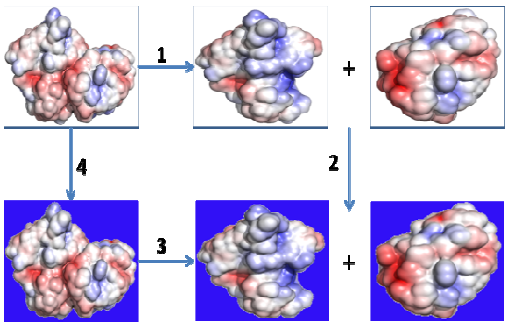

Fig. 4 Thermodynamic cycle for binding free energy calculations. The whitebackground represents the vacuum environment, and the blue background represents the aqueous environment

C.AMBER and APBS based energetics

From all above assessment, it has been clear that there is lots of scope to develop energy functions in order to predict correct binding poses. Although, ZRANK is performing on its level best, there is a need to design new scoring functions that could perform better than the present one. On these grounds, we proceed with some basic tests to define new energetics. We used two approaches, the simple molecular mechanics energetics and a free energy model (Fig. 4) based on solutions of non-linear Poisson-Boltzmann equation (*APBS Energetics*). The aim of these studies is to check whether simple or bit complicated energetics can distinguish between near-native and far-native predictions. For this purpose, we used the barnase-barstar protein complex with two models, one is in vacuum and the other in implicit solvent (Generalized Born). The 0 step simulation is performedand the electrostatic and Van der Waal components were extracted. These components were used for define energetics. Thebarnase-barstar complex is first separated into receptor and ligand. Using ZDOCK 54000(6^o sampling) complexes were obtained.

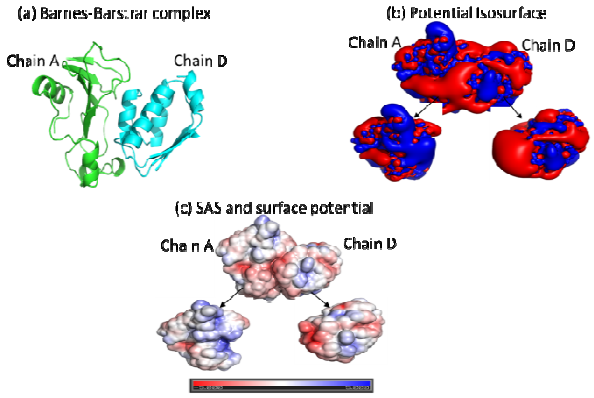

Fig. 5 (a) The barnase-barstar biological subunit with chains A and D. (b) APBS generated potential isosurface and (c) SAS and surface potential

The barnase-barstar system is awell-studied protein-protein complex [37]. The potential isosurface and solvent accessible surface area (SAS) is shown in Fig. 5. The interface surface shows electrostatic complementarity. 1BRS is a protein-protein complex, having totalof 196 residues, with remarkably high binding affinity, i.e. $K_d \approx$ 13 fM. This is an excellent complex system for energetic studies.

We chose 10 top ranked predictions by ZDOCK and one far-native prediction for the energetic studies as shown in Table I. The native structure has lowest energy and as we move away from the native conformation (i.e. from native complex to complex 10) the energy increases. For the randomly chosen far-native conformation (complex 399119), the energy is found to be lowest in the table. Thus, in this preliminary analysis, we found that the AMBER energetics, as a general trend,may be useful for distinguishing the conformations. Hence we performed calculations for all 54000 predictions. On analyzing the energetics, we found that there are huge number of energetically decoy poses do exists (Fig. 6). This suggests that some refinement in the criteria need to be included for recognizing the decoys.

TABLE I
AMBER ENERGETICS

Complex	ΔE_{ele}	ΔE_{ele}	ΔE_{vdw}	ΔE_{vdw}
	vac	GB	vac	GB

	kcal/mol			
Native	-538.57	-603.09	-66.87	-85.27
1	-513.82	-573.41	-51.00	-77.44
2	-506.68	-591.04	-63.10	-77.17
3	-504.52	-602.79	-55.16	-66.20
4	-494.50	-562.54	-61.63	-79.78
5	-496.47	-588.27	-53.57	-75.76
6	-527.24	-557.30	-52.17	-76.73
7	-525.03	-580.24	-44.59	-71.45
8	-528.12	-601.52	-68.37	-84.57
9	-495.10	-578.43	-49.03	-75.38
10	-550.46	-647.64	-46.93	-55.47
39119	-26.31	-91.00	27.10	8.76

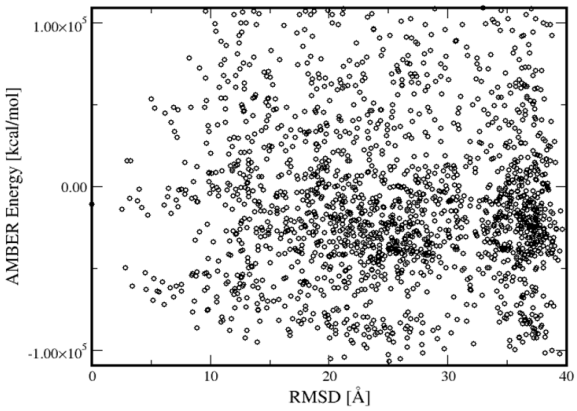

Fig. 6 AMBER Energetics. Energetically decoy data points

We also took another strategy based on the solutions of non-linear Poisson-Boltzmann equation (nPBE), calculated bythe APBS software suit[36]. We used this method to evaluate the binding free energy of the barnase-barstar system. From the dissociation constant, the free energy is estimated to be ΔG_{RT} = RT ln (K_d) = -19.18 kcal/mol. However, for the native complex, the calculated free energy is still far away from the experimental value, as shown in Table II. We performed calculations at 7 different dielectric constants and found that no calculation converges to the experimental values. Thus, it is not straightforward to apply this approach for discriminating the decoys from the native poses and more investigations (different atom radii set for PB calculations, different molecular boundary definitions, optimizing the contacts of interfaces, etc.) are needed.

TABLE II
APBS ENERGETICS FOR BARNASE-BARSTAR

DIELECTRIC CONSTANT	ΔΔG (KCAL/MOL)
1	-5231.41
2	-3050.59
4	-1887.09
10	-1109.14
12	-1013.51
20	-252.37
40	-645.85

The overall assessment of the ZDOCK and ZRANK help us to understand the challenge in prediction of correct binding pose and *insilico* identification of atomistic level protein-protein interactions. The molecular mechanics energeticscan be considered as useful tool for ruling out many decoys. More comprehensive free energy approaches based on non-linear Poisson-Boltzmann equation, which is more realistic to *invivo* or *invitro* protein system, is still a big challenge.It is worthwhile to note that recent study indicate the scope to achieve reliable electrostatics for protein-protein interaction by altering the electrostatic properties of proteins[38].

IV. Conclusion

The well-known docking suite, ZDOCK and the rescoring scheme, ZRANK, do not always generate correct binding modes of the protein-protein interacting partners, especially in unbound cases. APBS based free energy calculation based analysis for Barnase-Barstar model suggests the possible used of non-linear PBE based free energy calculation in scoring scheme. The simple molecular mechanics energetics with the AMBER force fieldmay be used to combine with PB-based energetics for better prediction of binding poses. However, it is found that AMBER energetics shows huge number decoys, thus other criteria need to be introduced for better prediction. The difference between calculated and experimental binding energy, corresponds to lack of proper weighting factors as well as important free energetic components, such as entropic term, hydrophobic interaction term, etc. New datasets based on experimental binding affinity or dissociation constant may help in improving theoretical predictions.

Acknowledgment

J.H.L thanks Dr. Alexandre M.J.J. Bonvin of Bijvoet Center for Biomolecular Research, Utrecht University, and Dr. Renxiao Wang of Shanghai Institute of Organic Chemistry, Chinese Academy of Science for the inspiring discussion. J.H.L was funded by National Science Council of Taiwan grant no. NSC 100-2627-B-001-001, 99-2627-B-001-002, 98-2627-B-001-002, 98-2323-B-002-001, 98-2323-B-077-001, 97-2323-B-002-015,and 97-2923-M-001-001-MY3. Supported from Research Center for Applied Sciences, Academia Sinica was also greatly acknowledged.

References

[1] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, I.A. Vakser, Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, Proc Natl Acad Sci U S A 89 (1992) 2195-2199.
[2] B. Honig, A. Ray, C. Levinthal, Conformational flexibility and protein folding: rigid structural fragments connected by flexible joints in subtilisin BPN, Proc Natl Acad Sci U S A 73 (1976) 1974-1978.

[3] C. Barillari, J. Taylor, R. Viner, J.W. Essex, Classification of water molecules in protein binding sites, Journal of the American Chemical Society 129 (2007) 2577-2587.
[4] J.M.J. Swanson, S.A. Adcock, J.A. McCammon, Optimized radii for Poisson-Boltzmann calculations with the AMBER force field, Journal of Chemical Theory and Computation 1 (2005) 484-493.
[5] J.M.J. Swanson, J.A. Wagoner, N.A. Baker, J.A. McCammon, Optimizing the Poisson dielectric boundary with explicit solvent forces and energies: Lessons learned with atom-centered dielectric functions, Journal of Chemical Theory and Computation 3 (2007) 170-183.
[6] R.W. Harrison, I.V. Kourinov, L.C. Andrews, The Fourier-Green's function and the rapid evaluation of molecular potentials, Protein Eng 7 (1994) 359-369.
[7] H.A. Gabb, R.M. Jackson, M.J.E. Sternberg, Modelling protein docking using shape complementarity, electrostatics and biochemical information, Journal of Molecular Biology 272 (1997) 106-120.
[8] G. Moont, H.A. Gabb, M.J.E. Sternberg, Use of pair potentials across protein interfaces in screening predicted docked complexes, Proteins-Structure Function and Genetics 35 (1999) 364-373.
[9] R.M. Jackson, H.A. Gabb, M.J.E. Sternberg, Rapid refinement of protein interfaces incorporating solvation: Application to the docking problem, Journal of Molecular Biology 276 (1998) 265-285.
[10] C. Dominguez, R. Boelens, A.M. Bonvin, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information, Journal of the American Chemical Society 125 (2003) 1731-1737.
[11] S.J. de Vries, A.D. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A.M. Bonvin, HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets, Proteins-Structure Function and Bioinformatics 69 (2007) 726-733.
[12] A.D. van Dijk, A.M. Bonvin, Solvated docking: introducing water into the modelling of biomolecular complexes, Bioinformatics 22 (2006) 2340-2347.
[13] J. Fernandez-Recio, M. Totrov, R. Abagyan, ICM-DISCO docking by global energy optimization with fully flexible side-chains, Proteins-Structure Function and Bioinformatics 52 (2003) 113-117.
[14] J.J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C.A. Rohl, D. Baker, Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations, Journal of Molecular Biology 331 (2003) 281-299.
[15] S. Chaudhury, A. Sircar, A. Sivasubramanian, M. Berrondo, J.J. Gray, Incorporating biochemical information and backbone flexibility in RosettaDock for CAPRI rounds 6-12, Proteins-Structure Function and Bioinformatics 69 (2007) 793-800.
[16] S. Chaudhury, J.J. Gray, Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles, Journal of Molecular Biology 381 (2008) 1068-1087.
[17] D.W. Ritchie, G.J.L. Kemp, Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces, Journal of Computational Chemistry 20 (1999) 383-395.
[18] D.W. Ritchie, G.J.L. Kemp, Protein docking using spherical polar Fourier correlations, Proteins-Structure Function and Genetics 39 (2000) 178-194.
[19] D.W. Ritchie, G. Macindoe, L. Mavridis, V. Venkatraman, M.D. Devignes, HexServer: an FFT-based protein docking server powered by graphics processors, Nucleic Acids Research 38 (2010) W445-W449.
[20] D.W. Ritchie, V. Venkatraman, Ultra-fast FFT protein docking on graphics processors, Bioinformatics 26 (2010) 2398-2405.
[21] R. Chen, Z.P. Weng, Docking unbound proteins using shape complementarity, desolvation, and electrostatics, Proteins-Structure Function and Genetics 47 (2002) 281-294.
[22] R. Chen, L. Li, Z.P. Weng, ZDOCK: An initial-stage protein-docking algorithm, Proteins-Structure Function and Genetics 52 (2003) 80-87.
[23] B. Pierce, Z. Weng, ZRANK: reranking protein docking predictions with an optimized energy function, Proteins-Structure Function and Genetics 67 (2007) 1078-1086.
[24] Z.P. Weng, J. Mintseris, B. Pierce, K. Wiehe, R. Anderson, R. Chen, Integrating statistical pair potentials into protein complex prediction, Proteins-Structure Function and Bioinformatics 69 (2007) 511-520.
[25] Z.P. Weng, B. Pierce, A combination of rescoring and refinement significantly improves protein docking performance, Proteins-Structure Function and Bioinformatics 72 (2008) 270-279.
[26] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Research 28 (2000) 235-242.
[27] M.D. Daily, D. Masica, A. Sivasubramanian, S. Somarouthu, J.J. Gray, CAPRI rounds 3-5 reveal promising successes and future challenges for RosettaDock, Proteins-Structure Function and Bioinformatics 60 (2005) 181-186.
[28] J.J. Gray, S.E. Moughon, T. Kortemme, O. Schueler-Furman, K.M. Misura, A.V. Morozov, D. Baker, Protein-protein docking predictions for the CAPRI experiment, Proteins-Structure Function and Bioinformatics 52 (2003) 118-122.
[29] H. Hwang, T. Vreven, B.G. Pierce, J.H. Hung, Z. Weng, Performance of ZDOCK and ZRANK in CAPRI rounds 13-19, Proteins-Structure Function and Genetics 78 (2010) 3104-3110.
[30] A.M.J.J. Bonvin, P.L. Kastritis, Are Scoring Functions in Protein-Protein Docking Ready To Predict Interactomes? Clues from a Novel Binding Affinity Benchmark, Journal of Proteome Research 9 (2010) 2216-2225.
[31] J. Janin, P.L. Kastritis, I.H. Moal, H. Hwang, Z.P. Weng, P.A. Bates, A.M.J.J. Bonvin, A structure-based benchmark for protein-protein binding affinity, Protein Science 20 (2011) 482-491.
[32] J. Mintseris, K. Wiehe, B. Pierce, R. Anderson, R. Chen, J. Janin, Z. Weng, Protein-Protein Docking Benchmark 2.0: an update, Proteins-Structure Function and Genetics 60 (2005) 214-216.
[33] H. Hwang, B. Pierce, J. Mintseris, J. Janin, Z. Weng, Protein-protein docking benchmark version 3.0, Proteins-Structure Function and Genetics 73 (2008) 705-709.
[34] R. Wang, Y. Lu, X. Fang, S. Wang, An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes, J Chem Inf Comput Sci 44 (2004) 2114-2125.
[35] D.A. Case, T.E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, Journal of Computational Chemistry 26 (2005) 1668-1688.
[36] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc Natl Acad Sci U S A 98 (2001) 10037-10041.
[37] A.M. Buckle, G. Schreiber, A.R. Fersht, Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution, Biochemistry 33 (1994) 8878-8889.
[38] R.D. Gorham, C.A. Kieslich, D. Morikis, Electrostatic Clustering and Free Energy Calculations Provide a Foundation for Protein Design and Optimization, Annals of Biomedical Engineering 39 (2011) 1252-1263.