

Using residue propensities and tightness of fit to improve rigid-body protein-protein docking

Bingding Huang and Michael Schroeder

Bioinformatics Group, Biotechnological Centre, Technical University Dresden, Germany

Abstract: Protein-protein interactions are fundamental as many proteins mediate their biological function through protein interactions. Over the past 20 years there have been many computational approaches to dock proteins. These approaches are mostly based on the shape complementarity of structures and the physio-chemical properties of the interfaces. However, these docking approaches are far from perfect and there still remains potential space to improve.

We propose to use family-based residue interaction propensity as well as the tightness of fit between residues with high propensities as a scoring function to improve rigid body docking. This approach is evaluated on an established benchmark data set. Our scoring function improves the number of hits for enzyme-inhibitor complexes by a factor of 4-30 in comparison with shape complementarity alone and for antibody-antigen complexes by 4-11.

The supplementary data are available at <http://www.biotec.tu-dresden.de/~bhuang/bdock>.

1 Introduction

1.1 Protein Docking

Protein docking and scoring functions. Most processes in the living cell requires molecular recognition and formation of complexes, which may be stable or transient assemblies of two or more molecules with one molecule acting on the other, or promoting intra- and intercellular communication, or permanent oligomeric ensembles ([EKK04]). The rapid accumulation of data on protein-protein interactions, sequences, structures calls for the development of computational methods for protein docking. Typically docking methods are investigated which attempt to predict the complex structures given the structures of components. Most docking approaches follow two steps. First, a set of candidate conformations is generated based on shape complementarity of the rigid bodies. A fast method using fast Fourier transform (FFT) for this approach has been proposed by [KKSE⁺92] and further developed by [GJS97, SGJ98, RK00, TWV02, BZE03, CW03, EKK04, BKKE02]. A good review about this method can be found in ([EKK04]). In a second step, the set of conformations are re-ranked using various scoring functions, used either independently or in combination. The scoring functions generally include geometric and chemical

complementarities measures, electrostatics, hydrogen-bonding and van der Waals interaction energy and some empirical potential functions such as residue-residue pair potential ([GS99]). A number of algorithms and many different scoring functions have been developed in the last ten years as recently reviewed by ([HMWN02, SS02, VC04, EKK04]).

Benchmark for testing the docking algorithms. The docking problem can be divided into two classes depending on the input of components structures. If we separate the complex structure into two components and then try to dock them together, it is called bound-bound docking. This is quite successful with rigid body docking methods. For the unbound-unbound docking, the separately crystallized component structures are used as input for docking which is more challenging than the former. Since the component structures are slightly different from the subunits in the complex structures (RMSD 0.5-1.0 Å).

In order to test the performance of the new docking approaches, various benchmark datasets with bound and unbound structures have been proposed ([GJS97, RK00, JTA02, PKWM00, GNS04]). A widely used ([LCZ03, CGVC04, GMW⁺03, GNS04, DRK05]) benchmark data set has been defined by [CMJW03]. It contains 59 non-redundant protein complexes including 22 enzyme-inhibitor complexes, 19 antibody-antigen complexes, 11 other complexes and 7 difficult test cases.

To judge whether a docking algorithm is good or not, the docked complex structures are compared to the native complex structures. If near-native (RMSD below 3 Å) structures are found in the top 100 to 1000 solutions, the docking approach is considered useful. The number of hits in the top 100 to 1000 docked solutions, the ranking of the best hit and the RMSD value of the best hit are the three main parameters used for evaluation of docking algorithms.

Docking difficulty. Recently, investigations on the interfaces of known protein-protein complexes have revealed that enzyme-inhibitor, antibody-antigen and other complexes present important differences in the amino acid composition, hydrophobicity and electrostatics ([DTD⁺01, GSVBT01]). [Jac99] compared protein-protein interactions in different types of complexes and concluded that enzyme-inhibitor are more static and hence more easily predictable than antibody-antigen. This suggests that different filtering criteria should be applied to different types of complexes. [LMCW03] applied type-dependent filtering technique to docking algorithm and retained much more native-like structures and increased the successful probability of predicting complex structures.

[VC04] classified protein complexes based on docking difficulty. They claimed that enzyme-inhibitor complexes can be determined by current docking methods with reasonable accuracy - possibly to within a few alternative structures. Results for antigen-antibody pairs are less predictable and data for small signaling complexes are generally poor. Transient complexes with large interface areas undergo substantial conformational changes and are beyond the reach of current docking methods. Moreover, based on measurements of conformational change, interface area and hydrophobicity, they defined five types of protein-protein complexes to characterize the expected level of docking difficulty.

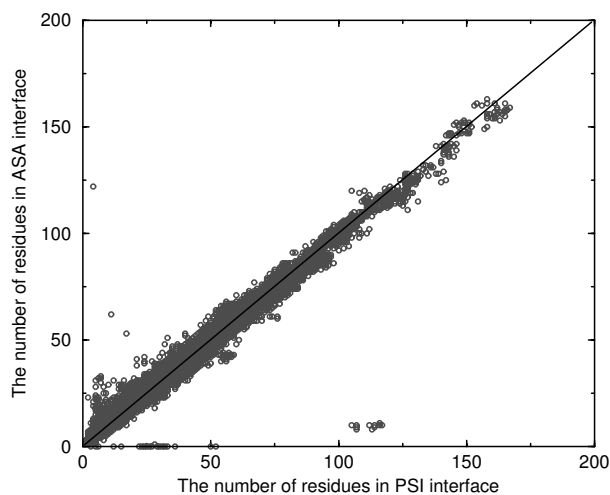


Figure 1: Scatterplot of the number of residues in the interface based on distance (PSIMAP) and based on solvent accessible surface area.

1.2 Predicting Interaction Interfaces

The core idea of this paper is the use of interaction interface predictions in the docking scoring functions to further improve the filtering.

Geometrical and physico-chemical approaches to interface prediction. In blind protein-protein docking approaches, it is of great importance that the binding sites are predicted correctly in the first step. Knowing where the binding sites are located on the protein surface can limit the conformational search space and reduce computational time. In the last 10 years, there have been many efforts to predict the protein-protein interaction binding sites based on the analysis of the protein surface properties ([JT97a, JT97b, NRS04, BW05, ZS01]). Jones and Thornton [JT97a, JT97b] analyzed the surface patches using six parameters: solvation potential, residue interface propensity, hydrophobicity, planarity, protrusion and solvation accessible surface area (ASA). The six parameters were then combined into a global score that gave the probability of a surface patch forming protein-protein interaction. Bradford [BW05] trained a support vector machine (SVM) to distinguish interacting and non-interacting surface patches using the six surface properties surface shape, hydrophobicity, conservation, electrostatic potential, residue interface propensity and solvation accessible surface area. Using this method, they were able to predict the location of the binding sites on 76% of the 180 protein data set using a leave-one-out validation procedure. This method was shown to be applicable to both obligate and transient binding sites.

Although many binding sites prediction methods have been developed, only one group

integrated it into docking. Recently, [GNS04] predicted protein-protein binding sites first using their own prediction program: ProMate ([NRS04]), of which the success rate was about 70%, and then they used these predicted binding sites to calculate the tightness of fit of the two docked proteins. A linear relation between this score and the RMSD relative to the true structure is found in most of the cases they evaluated. Their results encourage us that using predicted interaction sites can improve protein docking. However, we will try a different prediction methods and develop different scoring functions based on these methods. All scoring functions will be integrated together to improve docking.

Family-based residue interaction propensities. With the growth of the Protein Data-bank PDB ([BWF⁺00]), more and more complex structures are available. In this paper, we propose to calculate residue interaction propensities for families of the structural classification of proteins SCOP ([AHB⁺04]). For residues with high propensities, we calculate the tightness of fit between these residues. Overall, this novel scoring function shows a good correlation with the RMSD and therefore can improve docking. Its only limiting factor is the availability of structural data to compute the family-based interaction propensities.

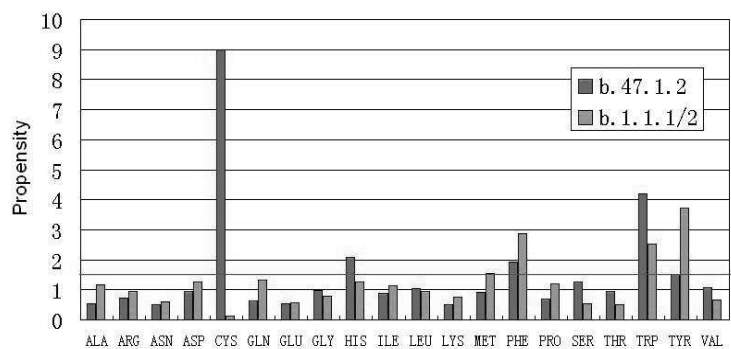
2 Materials and Methods

2.1 FFT docking method

Our proposed scoring function is included into BDOCK, an implementation of the FFT docking method ([KKSE⁺92]) using the BALL library ([KL00]). BDOCK also includes scoring by residue pair potential and desolvation energy. BDOCK scans the protein surface at a rotational angle of 10° using a grid size of 1.0 \AA and a surface thickness of 2.0 \AA .

2.2 PSIMAP, the Protein Structure Interaction Map

To compute the residue interaction propensities we use PSIMAP, a database with over 40,000 structural interaction interfaces ([PLT01, DBG⁺04]) for over 8000 PDB structures. PSIMAP denotes two domains as interacting with each other if at least 5 residue pairs are within 5 Angstroms (the 5-5 rule). The 5-5 rule correlated very well with defining the interface based solvent accessible surface area (ASA) as shown in Fig. 1. Fig. 2 shows a screenshot of PSIMAP with the family members aligned by sequence and interacting residues highlighted. Non-redundant family-members are used to compute residue interaction propensities are described below.



Residue	b.47.1.2	b.1.1.1/2	Residue	b.47.1.2	b.1.1.1/2
ALA	0.53	1.16	LEU	1.05	0.96
ARG	0.72	0.94	LYS	0.52	0.75
ASN	0.51	0.59	MET	0.92	1.54
ASP	0.95	1.27	PHE	1.94	2.86
CYS	8.99	0.13	PRO	0.69	1.19
GLN	0.64	1.13	SER	1.25	0.55
GLU	0.55	0.58	THR	0.94	0.51
GLY	0.97	0.80	TRP	4.18	2.52
HIS	2.07	1.26	TYR	1.50	3.71
ILE	0.88	1.13	VAL	1.07	0.65

Figure 3: The residue propensities of Trypsin-like serine protease (SCOP ID: b.47.1.2, calculated from 747 interactions) and V set domains (antibody variable domain-like)/C1 set domains (antibody constant domain-like) (SCOP ID b.1.1.1/2, calculated from 620 interactions). TRP and TYR have high preference in the interface in both families. CYS has highest preference in b.47.1.2 but it has very low propensity in b.1.1.1/2. The right table shows the propensity values for each residue. Those residues with propensity ≥ 1.5 are regarded as predicted interface residues, i.e. for b.47.1.2 they are CYS, HIS, PHE, TRP and TYR; MET, PHE, TRP and TYR for b.1.1.1/2.

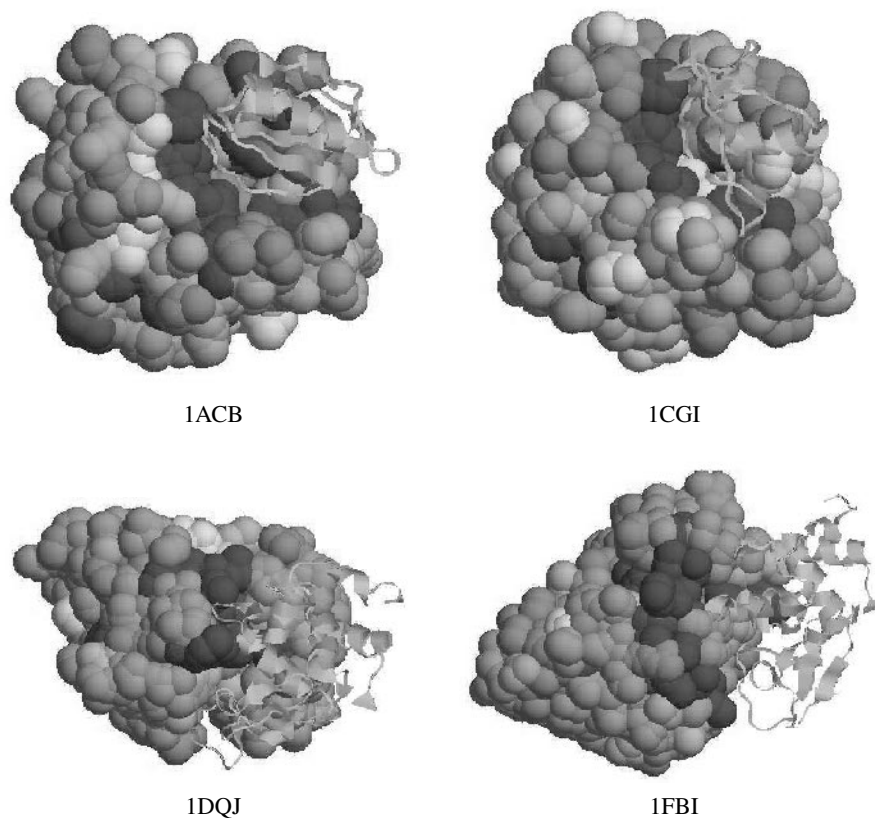


Figure 4: The residue interface propensities on the protein surface of the receptors. The gray one is its binding partner. Red: high propensity, green: low propensity (The color scheme is referred to the values from Fig 3). The receptor of 1ACB and 1CGI is Trypsin-like serine protease (b.47.1.2). The high propensity residues for this family are CYS, HIS, PHE, TRP and TYR. The receptor of 1DQJ and 1FBI is V set domains (antibody variable domain-like)/ C1 set domains (antibody constant domain-like) (b.1.1.1/2). The high propensity residues for this two families are MET, PHE, TRP and TYR. These figures indicate that the high propensity residues accumulate around the real binding sites.

sities at the family-level rather than globally. In Fig 4 the propensities of Fig. 3 have represented as colors (red = high propensity, green = low) of the surface residues. This example shows that the propensity in the interface is higher than in the rest of the surface residues thus encouraging the overall approach.

The reason for using residue interface propensity as a scoring function is that the near-native docked complex structures should have highest propensity value if we calculate the average residue interaction propensity (IP) for those interface residues of receptor as bellow:

$$IP = \frac{1}{N} \sum_i^N Pro_i \quad (2)$$

where

$$Pro_i = \begin{cases} 10 * Propen_i & \text{if } Propen_i \geq 1.5 \\ Propen_i & \text{else} \end{cases}$$

Here $Propen_i$ is the precalculated propensity for the residue i in the receptor from docked structure using equation 1 and N is the number of interface residue. If $Propen_i$ is ≥ 1.5 , it is made 10 times bigger to ensure that those docked complexes having more predicted interface residues have higher propensity score. This propensity score can measure how likely the interface of docked solutions trend to be native interface. Using this scoring function to score the docked solutions is very fast since we only care about the interface residues.

As shown for an enzyme-inhibitor example in Fig. 5.c-e scoring with desolvation energy, residue pair potential and shape complementarity does not correlate well with RMSD. The residue interaction propensity in Fig. 5.a is already much better, as it singles out near-native structures together with a few very poor predictions. To further reduce such false positives we combine the residue interaction propensities with the tightness of fit (ToF) proposed by [GNS04]. For ToF we only consider residues with ≥ 1.5 propensity. ToF is calculated according to:

$$ToF = \frac{d_{inter} - d_{all}}{d_{all}} \quad (3)$$

where

$$d_{inter} = \frac{1}{n} \sum_{i=1}^n \frac{D_{inter; i}}{Propen_i} \quad \text{and} \quad d_{all} = \frac{1}{m} \sum_{j=1}^m \frac{D_{all; j}}{Propen_j} \quad (4)$$

D_{inter} is the minimum distance of the C_α of residue i with propensity ≥ 1.5 (high probability to be in interface) of receptor to any atom of ligand. D_{all} is the minimum distance of C_α atom of surface residue j of receptor to any atom of ligand. There are n residues with ≥ 1.5 propensity and m surface residues. As Fig. 5.b shows for the enzyme-inhibitor example the tightness of fit of high propensity residues correlates very well with the RMSD of near-native complexes.

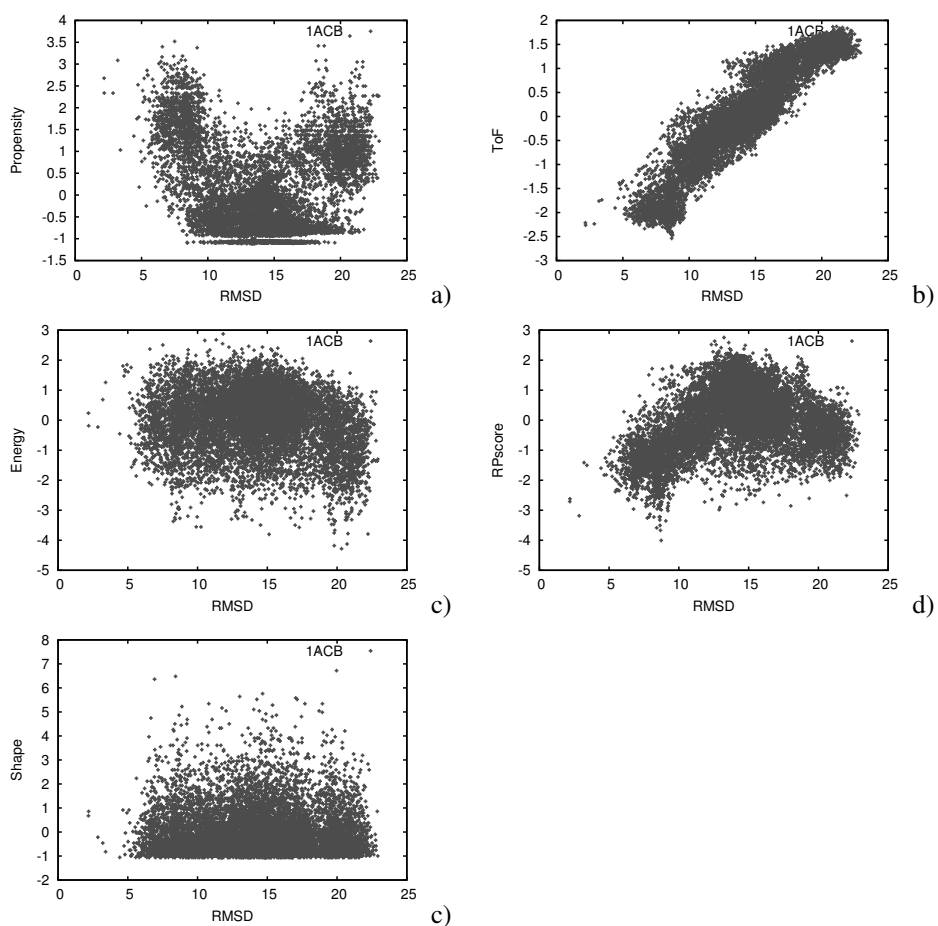


Figure 5: Different Z-scores vs RMSD of the enzyme-inhibitor complex 1ACB. a). Residue interaction propensity b). Tightness of fit c). Desolvation energy d). Residue pair potential e). Shape complementarity. While c, d, and e are not well correlated with the RMSD, the propensity can single the best RMSD solutions together with some poor solutions. These false positives are completely absent from the tightness of fit score, which correlates very well with the RMSD. The scatterplots of Z-scores vs RMSD for all the complexes are available at www.biotec.tu-dresden.de/~bhuang/bdock.

3 Results

To evaluate the above scoring functions we considered 24 examples from the Chen data set (12 enzyme-inhibitor and 12 antibody-antigen). For each complex in the data set, we keep 10000 docking solutions and score each by atom contact energy, residue pair potential, propensity score, tightness of fit. Next, we compute the Z-score for each scoring function:

$$Z\text{-score}_i = \frac{X_i - \bar{X}}{\sigma} \quad (5)$$

where X_i is the score of i solution, \bar{X} is the mean score of total 10000 solutions and σ is the standard deviation. The benefit of using Z-score to re-rank the docked solutions is that the scores of different scoring function for different complex are comparable.

Here we define a docking solution as near-native structure (hit) if the RMSD between it and the native complex is below 4.5\AA . It is obviously that when we applied filters to the complex structures pool, some near-native structures are also filtered out together with non-native structures. To see the improvement after filter, the improvement factor IF is calculated according to

$$IF = \frac{hits_{af}/N_{af}}{hits_{bf}/N_{bf}} \quad (6)$$

where N_{bf} , N_{af} are the number of the complex structures; $hits_{bf}$ and $hits_{af}$ are the number of near-native structures ($\text{RMSD} \leq 4.5\text{\AA}$) in the pool, before filter and after filter.

Enzyme-inhibitor complexes The receptors in these complexes are in the same family (Trypsin-like serine protease, SCOP ID b.47.1.2) which interact with different inhibitors. They are 1ACB, 1AVW, 1BRC, 1BTH, 1CGI, 1CHO, 1PPE, 1TAB, 1TGS, 2KAI, 2PTC, 4HTC. For this family, CYS, HIS, PHE, TRP and TYR have propensity ≥ 1.5 and are regarded as predicted interface residues (see Fig 3). The docking and filter results are showed in table I. Based on shape complementarity, BDOCK generates some near-native structures for all 12 E/I complexes, ranging from 6 to 981, in 10000 docking results. The best RMSD for these complexes is below 3\AA except for 1BTH among these 10000 solutions.

When filtering these 10000 solutions by propensity score, the number of complexes in the pool reduced to one thousand to two thousand but most of the near-native structures are still remaining in the pool. The improvement factor (IF) after filter by propensity score ≥ 1.0 is greater than 2 for all the complexes. Filter by ToF (≤ -1.5) can remove more false positive solutions except for 1BTH and 4HTC. For 1BTH, there is a very deep pocket in the enzyme surface where the inhibitor fits into tightly, while 4HTC has a long tail of the inhibitor involved in interaction (see Fig 6). This is the reason why ToF fails to rank near-native structures high for them.

Using both filter criteria, the number of docking candidates reduces to less than one thousand and the best result is still remained in this pool except for 1CHO and 1PPE. The improve factors become better in all cases, ranging from 4 to 30.

Table I. The docking results for enzyme-inhibitor complexes

Complex	$hits_{bf}^a$	RMSD ^b	Filter by IP ^c		Filter by ToF ^d		Filter by IP and ToF		
			$hits_{af}/N_{af}^e$	IF ^f	$hits_{af}/N_{af}$	IF	$hits_{af}/N_{af}$	IF	RMSD ^g
1ACB	6	2.17	6/2049	4.88	6/1068	9.35	6/910	11.00	2.17
1AVW	11	1.38	9/1906	4.30	11/954	10.50	9/502	16.30	1.38
1BRC	42	2.21	24/1933	2.96	9/624	3.43	8/451	4.22	2.21
1BTH	12	3.85	12/1462	6.84	0/112	0	0/3	0	8.04
1CGI	52	1.30	51/1807	5.43	20/130	29.59	20/130	29.59	1.30
1CHO	38	0.90	34/1767	5.06	27/692	10.27	26/562	12.17	0.95
1PPE	981	0.29	653/1930	2	510/881	6.0	417/554	7.67	0.38
1TAB	62	0.71	26/1801	2.33	39/837	7.52	21/262	12.93	0.71
1TGS	106	0.69	70/1859	3.55	88/987	8.41	69/654	10.55	0.69
2KAI	114	1.43	82/1809	3.98	96/135	6.20	66/654	8.85	1.44
2PTC	52	1.66	46/1798	4.90	52/916	10.20	46/559	15.82	1.66
4HTC	41	2.20	33/1449	5.55	0/121	0	0/6	0	10.07

^aThe number of complexes whose RMSD is below 4.5 Å in 10000 solutions.

^bThe best RMSD from 10000 docking results.

^cZ-score. Threshold: ≥ 1.0 .

^dZ-score. Threshold: ≤ -1.5 .

^e N_{af} is the number of docking results remained in the pool after filter. $hits_{af}$ is the number of hits in this pool.

^fCalculated according to equation 6. N_{bf} is 10000.

^gThe best RMSD after filter using IP and ToF.

Table II. The docking results for antibody-antigen complexes^a

Complex	$hits_{bf}$	RMSD	Filter by IP ^b		Filter by ToF ^c		Filter by IP and ToF		
			$hits_{af}/N_{af}$	IF	$hits_{af}/N_{af}$	IF	$hits_{af}/N_{af}$	IF	RMSD
1AHW	5	2.62	5/1316	7.60	0/1770	0	0/1106	0	5.94
1BQL	8	1.58	8/1641	6.10	8/2236	4.47	8/1377	7.26	1.58
1DQJ	20	2.85	19/1135	8.37	20/1978	5.05	19/1106	8.60	2.85
1EO8	6	1.25	6/1840	5.43	6/1745	5.73	6/900	11.11	1.25
1FBI	9	2.84	6/1831	3.64	9/1932	5.17	6/1560	4.27	2.84
1IAI	4	1.85	4/1804	5.54	4/1931	5.17	4/1138	8.79	1.85
1MLC	0	4.55	0/1478	0	0/1626	0	0/1177	0	4.55
1NCA	13	1.17	13/1876	5.33	13/2307	4.33	13/1499	6.67	1.17
1QFU	8	1.29	8/1745	5.73	3/2279	1.64	3/1021	3.80	1.29
1WEJ	9	3.16	9/1253	7.99	9/1685	5.83	9/1195	8.36	3.16
2JEL	5	3.48	5/1745	5.73	5/1694	5.90	5/1055	9.47	3.48
2VIR	2	1.40	2/1695	5.90	2/1852	5.40	2/986	10.14	1.40

^asee Table I.

^bThreshold ≥ 1.0 .

^cThreshold ≤ -1.0 .

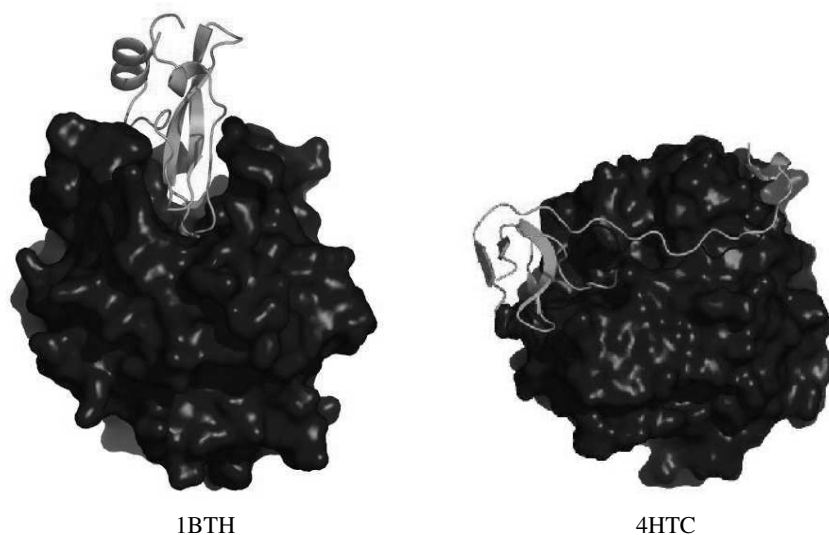


Figure 6: The complex structures of 1BTH and 4HTC. Tightness of fit fails to figure out near-native structures for them. Blue: enzyme, green: inhibitor.

Antibody-antigen complexes 12 complexes belong to this A/A, i.e. 1AHW, 1BQL, 1DQJ, 1EO8, 1FBI, 1IAI, 1MLC, 1NCA, 1QFU, 1WEJ, 2JEL, 2VIR. For these 12 complexes, the antibody consist of two domains: V set domains (antibody variable domain-like) (SCOP ID b.1.1.1) and C1 set domains (antibody constant domain-like) (SCOP ID b.1.1.2). These two domains bind together and they are both involved in interaction with antigen. Only four residues (MET, PHE, TRP and TYR) have propensity ≥ 1.5 and are regarded as predicted interface residues (see Fig 3). The docking and filter results are showed in table II. Based on shape complementarity, BDOCK generates only a few near-native structures in 10000 docking results, from 1 to 20 except for 1MLC (best RMSD 4.55). When comparing the unbound structures (Antibody D44.1 Fab fragment (1mlb) and lysozyme (1lza)) with the complex structure of 1MLC, a significant conformational change is observed on lysozyme, with Pro70 moving as much as 4 Å and dragging the main chain atoms (see Fig 7, modified from [PKWM00]). As a result, rigid-body docking method fails to generate near-native structures for this case and flexibility needs to be taken into account.

Filter by propensity (≥ 1.0), the improve factor is from 3 to 8. Filter by ToF (≤ -1.0), it is from 4 to 6. ToF fails in 1AHW because the high propensity residues have some distribution on non-interface surface although they do accumulate around the real interface in the receptor of this complex. Using both filter criteria, the IF is 4 to 11 and the number of candidates in the pool is from 900 to 1600. Comparing to the docking results of E/I complexes, less hits are found in the initial 10000 solutions for A/A complexes based on shape complementarity which confirms the conclusion by [VC04] that antibody-antigen is less predictable by current rigid-body docking methods. That is also the reason why the threshold for ToF here is ≤ -1.0 .

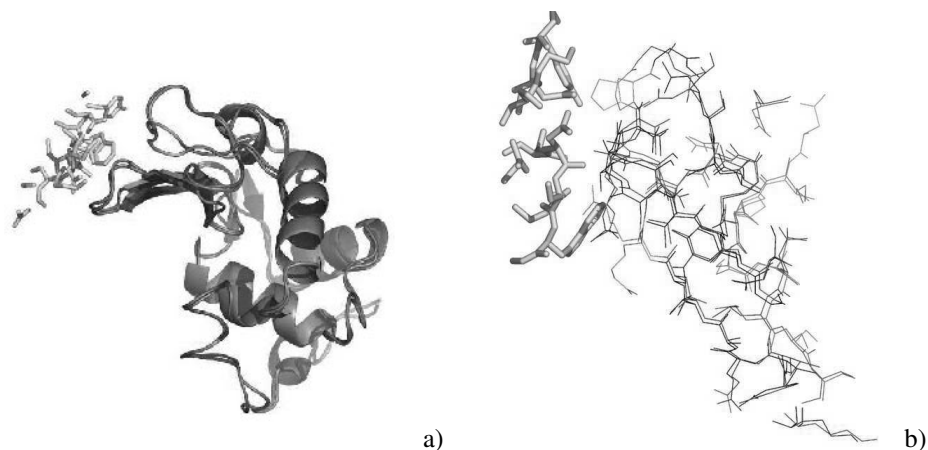


Figure 7: Detail of the interaction between monoclonal antibody D44.1 Fab fragment (1mlb) and lysozyme (1lza). The blue one corresponds to the lysozyme in the complex and the yellow one Fab fragment (only show interface), while the red one show the conformations of the non-complexed lysozyme structures, when superposed on that of the complex (1mlc). Significant conformational changes of PRO70 and ARG45 are shown in cyan and green. a). main chain level. b). side chain level, the ARG45 (green) of 1lza penetrating to Fab fragment.

4 Conclusion

Protein docking aims to predict complex structure from unbound component structures. A major challenge in this field is to extract the near-native structures from a pool of a large number of solutions by using appropriate scoring function. In this paper, we describe a novel scoring function using interface residue propensity and the tightness of fit between high propensity residues. We evaluated this scoring function on 24 structures from the Chen data set and obtained improvement factors of 4-30 for enzyme-inhibitor and 4-11 for antibody-antigen complexes. A key insight is that focusing on a few residues, which are of particular importance (in our case the high propensity residues) and considering the tightness of fit for these residues improves scoring. Since our review highlights various other approaches to predict interaction interfaces ([BW05, JT97a, JT97b]), we hypothesize that the approaches can lead to further improvements of scoring functions. We will continue this line of research using other definitions of hot spots and geometric measures to distinguish pockets from flat surface patches.

5 Acknowledgement

Funding by EFRE project CODI no. 4212/04-07 is kindly acknowledged.

References

- [AHB⁺04] A. Andreeva, D. Howorth, S.E Brenner, T. Hubbard, C. Chothia, and A.G Murzin. SCOP database in 2004: refinements integrate structure and sequence family data. *Nucl. Acid Res.*, 32:D226–D–229, 2004.
- [BKKE02] A Berchmanski, E Katchalski-Katzir, and M Eisenstein. Electrostatics in protein-protein docking. *Protein Sci*, 11:571–587, 2002.
- [BW05] James Bradford and David Westhead. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics*, 21(8):1487–1494, 2005.
- [BWF⁺00] HM Berman, J Westbrook, Z Feng, G Gilliland, TN Bhat, H Weissig, IN Shindyalov, and PE Bourne. The Protein Data Bank. *Nucleic Acids Res*, 28(1):235–42, 2000.
- [BZE03] Efrat Ben-Zeev and Miriam Eisenstein. Weighted geometric docking: Incorporating External Information in the Rotation-Translation Scan. *Proteins: Structure, Function and Genetics*, 52:24–27, 2003.
- [CGVC04] S.R. Comeau, D.W. Gatchell, S. Vajda, and C.J. Camacho. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, 20:45–50, 2004.
- [CMJW03] R Chen, J Mintseris, J Janin, and Z Weng. A protein-protein docking benchmark. *Proteins*, 52:88–91, 2003.
- [CW03] R Chen and Z Weng. A Novel Shape Complementarity Scoring Function for Protein-Protein Docking. *Proteins*, 51:397–408, 2003.
- [DBG⁺04] Panos Dafas, Dan Bolser, Jacek Gomoluch, Jong Park, and Michael Schroeder. Using convex hulls to extract interaction interfaces from known structures. *Bioinformatics*, 20(10):1486–1490, 2004.
- [DRK05] Y. Duan, V.B. Reddy, and YN. Kaznessis. Physicochemical and residue conservation calculations to improve the ranking of protein-protein docking solutions. *Protein Science*, 14:316–328, 2005.
- [DTD⁺01] K. Decanniere, T. Transue, A. Desmyter, D. Maes, S. Muyldermans, and L. Wyns. Degenerate interfaces in antigen-antibody complexes. *J. Mol. Biol*, 313:473–478, 2001.
- [EKK04] M. Eisenstein and E. Katchalski-Katzir. On proteins, grids, correlations, and docking. *C.R.Biologies*, 327:409–420, 2004.
- [GJS97] H. Gabb, R. Jackson, and M. Sternberg. Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information. *J.Mol.Biol*, 272(1):106–120, 1997.
- [GMW⁺03] JJ Gray, S Moughon, C Wang, O Schueler-Furman, B Kuhlman, CA Rohl, and D Baker. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.*, 331:281–299, 2003.
- [GNS04] K.E. Gottschalk, H Neuvirth, and G. Schreiber. A novel method for scoring of docked protein complexes using predicted protein-protein binding sites. *Protein Eng. Des.*, 17:183–189, 2004.
- [GS99] G. Moont H.A. Gabb and M.J. Sternberg. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins*, 35(3):364–373, 1999.

- [GSVBT01] F. Glaser, DM. Steinberg, IA. Vakser, and N. Ben-Tal. Residue Frequencies and Pairing Preferences at Protein-Protein Interfaces. *Proteins*, 43:82–102, 2001.
- [HMWN02] I Halperin, B Ma, H Wolfson, and R Nussinov. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins*, 47:409–443, 2002.
- [Jac99] RM Jackson. Comparison of protein-protein interactions in serine protease-inhibitor and antibody-antigen complexes: implications for the protein docking problem. *Protein Sci*, 8:603–613, 1999.
- [JT97a] Susan Jones and Janet Thornton. Analysis of protein-protein interaction sites using surface patches. *J.Mol.Biol*, 272:121–132, 1997.
- [JT97b] Susan Jones and Janet Thornton. Prediction of protein-protein interaction sites using patches analysis. *J.Mol.Biol*, 272:133–143, 1997.
- [JTA02] J.Recio, M Totrov, and R Abagyan. Soft protein-protein docking in internal coordinates. *Protein Sci*, 11:280–291, 2002.
- [KKSE⁺92] E Katchalski-Katzir, I Shariv, M Eisenstein, A A Friesem, C Aflalo, and I A Vakser. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proc Natl Acad Sci USA*, 89:2195–3199, 1992.
- [KL00] O. Kohlbacher and H.P. Lenhof. BALL - Rapid Software Prototyping in Computational Molecular Biology. *Bioinformatics*, 16(9):815–824, 2000.
- [LCZ03] L Li, R Chen, and W Zhiping. RDOCK: refinement of rigid-body protein docking predictions. *Proteins*, 53(3):693–707, 2003.
- [LMCW03] ChunHua Li, XiaoHui Ma, WeiZu Chen, and CunXin Wan. A protein-protein docking algorithm dependent on the type of complexes. *Protein Eng.*, 16:265–269, 2003.
- [NRS04] Hani Neuvirth, Ran Raz, and Gibeon Schreiber. ProMate: A structure based prediction program to indentify the location of protein-protein binding stes. *J.Mol.Biol*, 338:181–199, 2004.
- [PKWM00] P.N. Palma, L Krippahl, J.E. Wampler, and J.G. Moura. BiGGER: A new (soft) docking algorithm for predicting protein interactions. *Proteins*, 39(4):372–384, 2000.
- [PLT01] J. Park, M. Lappe, and S.A. Teichmann. Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast. *J Mol Biol*, 307(3):929–38, 2001.
- [RK00] D.W. Ritchie and J.L. Kemp. Protein docking using spherical polar Fourier correlations. *Proteins: Structure, Function, and Genetics*, 39(2):178–194, 2000.
- [SGJ98] MJ Sternberg, HA Gabb, and RM Jackson. Predictive docking of protein-protein and protein-DNA complexes. *Curr. Opin. Struct Biol.*, 8(2):265–269, 1998.
- [SS02] GR Smith and MJ Sternberg. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct Biol.*, 12:28–35, 2002.
- [TWV02] A. Tovchigrechko, C.A. Wells, and I. Vakser. Docking of protein models. *Protein Sci.*, 11:1888–1896, 2002.
- [VC04] S Vajda and CJ Camacho. Protein-protein docking: is the glass half-full or half-empty? *Trends Biotechnol.*, 22(3):110–116, 2004.
- [ZS01] HX Zhou and Y. Shan. Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins: Structure, Function, and Genetics*, 44(3):336–343, 2001.