

A protein–protein docking algorithm dependent on the type of complexes

Chun Hua Li, Xiao Hui Ma, Wei Zu Chen and
Cun Xin Wang¹

College of Life Science and Bioengineering, Beijing University of
Technology, Beijing 100022, China

¹To whom correspondence should be addressed.
E-mail: cxwang@bjpu.edu.cn

An efficient ‘soft docking’ algorithm is described to assist the prediction of protein–protein association using three-dimensional structures of molecules. The basic tools are the ‘simplified protein’ model and the docking algorithm of Wodak and Janin. The side chain flexibility of Arg, Lys, Asp, Glu and Met residues at the protein surface is taken into account. The complex type-dependent filtering technique on the basis of the geometric matching, hydrophobicity and electrostatic complementarity is used to select candidate binding modes. Subsequently, we calculate a scoring function which includes electrostatic and desolvation energy terms. In the 44 complexes tested including enzyme–inhibitor, antibody–antigen and other complexes, native-like structures were all found, of which 30 were ranked in the top 20. Thus, our soft docking algorithm has the potential to predict protein–protein recognition.

Keywords: binding free energy/molecular flexibility/
molecular recognition/protein docking/protein–protein
interactions

Introduction

Recently, experimental and computational efforts have increasingly been devoted to the investigations of protein–protein interactions, which is very significant for understanding biochemical processes, e.g. signal transduction, cell regulation and immune response. Given the difficulties in experimentally determining the structures of protein complexes, the docking method to predict potential binding modes computationally is currently of great interest. The principles of docking and the progress that has been made during the last decade have been described (Cherfils and Janin, 1993; Lengauer and Rarey, 1996; Sotriffer *et al.*, 2000; Halperin *et al.*, 2002). Many promising algorithms, such as the fast Fourier transform (FFT)-based matching (Katchalski-Katzir *et al.*, 1992; Gabb *et al.*, 1997), Geometric Hashing (Fischer *et al.*, 1995; Norel *et al.*, 1999a) and BIGGER (Palma *et al.*, 2000), have been developed. However, because of the complexity of the problem, protein–protein docking is still largely at the theoretical stage and there is still considerable scope for the development of methodology.

In protein–protein docking, because of the large number of atoms and degrees of freedom involved, it would be impracticable to treat molecular flexibility in an explicit way. Currently, the solutions to this problem are mainly limited to the techniques that may tolerate a limited degree of molecular flexibility by using a ‘soft’ representation of the molecular surface (Jiang

and Kim, 1991; Walls and Sternberg, 1992; Sandak *et al.*, 1995; Vakser, 1995; Palma *et al.*, 2000; Ritchie and Kemp, 2000). Jiang and Kim used a cube representation of the molecular surface and volume in their docking procedure (Jiang and Kim, 1991). Ritchie and Kemp introduced a ‘soft’ model of electrostatic complementarity in the algorithm (Ritchie and Kemp, 2000). Palma *et al.* proposed a surface-implicit method in which the surface is represented by values 0 and 1 on two grids, the surface and core grids (Palma *et al.*, 2000). This digitization introduces the first level of ‘softness’ in the algorithm. In this paper, the flexible amino acid residues Arg, Lys, Asp, Glu and Met at the protein surface are softened on the basis of the ‘simplified protein’ model (Levitt, 1976). This softness treatment improves the effect of unbound docking to some degree.

A search procedure may produce millions of docked structures. How to reduce these solutions drastically by filtering to a range manageable by the scoring functions is a serious and challenging topic of current research. The docking method is generally based on the idea of complementarity between the interacting molecules. This complementarity may be geometric, electrostatic or hydrophobic, or all three. Most docking algorithms developed so far used the extent of geometric complementarity of the protein surfaces as the filtering criterion to eliminate a large number of solutions with poor surface matching. It is generally recognized, however, that a single filtering criterion is not sufficient to discriminate between the native-like and incorrect docked structures except in a very few cases (Shoichet and Kuntz, 1991). Recently, investigations on the interfaces of known protein–protein complexes (Jones and Thornton, 1996; Betts and Sternberg, 1999; Lo Conte *et al.*, 1999; Norel *et al.*, 1999b; Decanniere *et al.*, 2001) have revealed that enzyme–inhibitor, antibody–antigen and other complexes present important differences in the interface residue composition, hydrophobicity and electrostatics. Jackson compared protein–protein interactions in these different types of complexes and concluded that enzyme–inhibitor interfaces are more static and hence more easily predicted than antibody–antigen interfaces (Jackson, 1999). This suggests that different filtering criteria should be applied to different types of complexes. In this paper, we focus on the type-dependent filtering technique in which, in addition to the geometric matching, we also take hydrophobicity and electrostatic complementarity into account.

Materials and methods

The selected test set

A collection of 44 protein–protein complexes from the Protein Data Bank (PDB) was used as test sets (Table I). They were chosen from different types of complexes, including 23 enzyme–inhibitor, 11 antibody–antigen and 10 other complexes. For 24 systems, docking was performed with the unbound experimental structures of both the receptor and the

Table I. The selected 44 protein–protein complexes used to test the docking algorithm

Complex	Description	Receptor PDB	Ligand PDB	C α r.m.s.d. ^a receptor/ligand
<i>Enzyme–inhibitor</i>				
1ACB	α -Chymotrypsin/Eglin C	5cha	1cse	0.69/0.63
1BRC	Trypsin/APPI	1bra	1aap	0.40/0.43
1BRS	Barnase/Barstar	1a2p	1a19	0.49/0.44
1CHO	α -Chymotrypsin /Ovomucoid 3rd domain	5cha	2ovo	0.47/1.16
1CGI	α -Chymotrypsinogen/Human pancreatic secretory trypsin inhibitor	1chg	1hpt	1.12/1.78
1CSE	Subtilisin Carlsberg/Eglin C	1scd	1acb	0.36/0.63
1FSS	Acetylcholinesterase/Fasciculin II	2ace	1fsc	0.48/0.78
2KAI	Kallikrein A/Trypsin inhibitor	2pka	6pti	0.69/2.01
1MAH	Mouse acetylcholinesterase/inhibitor	1maa	1fsc	0.65/0.74
2PTC	β -Trypsin/Pancreatic trypsin inhibitor	2ptn	6pti	0.34/0.36
1UGH	Human uracil–DNA glycosylase/inhibitor	1akz	1ugi	0.39/0.64
2SNI	Subtilisin novo/Chymotrypsin inhibitor 2	1sup	2ci2	0.25/0.46
1TGS	Trypsinogen/Pancreatic secretory trypsin inhibitor	2ptn	1hpt	1.19/1.88
1BTH	Thrombin mutant/Pancreatic trypsin inhibitor	2hnt	6pti	0.89/0.48
2SIC	Subtilisin BPN/Subtilisin inhibitor	1sup	3ssi	0.25/0.64
1FQ1	CDK2/KAP	1b39	1fpz	3.29/0.80
1TAB ^c	Trypsin/BBi	2ptn	1tab ^b	0.34/0.00
1PPE ^c	Trypsin/CMT-1	2ptn	1ppe ^b	0.35/0.00
1UDJ ^c	Virus uracil–DNA glycosylase/inhibitor	1udh	1udi ^b	0.47/0.00
2TEC ^c	Thermitase/Eglin C	1thm	2tec ^b	0.23/0.00
1STF ^c	Papain/Stefin B	1ppn	1stf ^b	0.32/0.00
4HTC ^c	α -Thrombin/Hirudin	2hnt	4htc ^b	0.65/0.00
1EFU ^c	E.coli EFtu/Efts	1d8t	1efu ^b	2.60/0.00
<i>Antibody–antigen</i>				
1WEJ	IgG1 E8 Fab fragment/Cytochrome c	1qbl	1hrc	0.34/0.35
1AHW	Antibody Fab 5G9/Tissue factor	1fgn	1boy	0.60/0.82
1BVK	Antibody Hulys11 Fv/Lysozyme	1bvl	3lzt	0.51/1.22
1DQJ	Hyhel63 Fab/Lysozyme	1dqj	3lzt	0.50/0.84
1FBI ^c	IgG1 Fab fragment/Lysozyme	1fbi ^b	1hhl	0.00/0.68
2JEL ^c	Jel42 Fab fragment/A06 phosphotransferase	2jel ^b	1poh	0.00/0.70
1JHL ^c	IgG1 Fv fragment/Lysozyme	1jhl ^b	1ghl	0.00/0.51
2VIR ^c	IgG1-lambda Fab/Hemagglutinin	2vir ^b	2viu	0.00/0.64
1QFU ^c	IgG1-kappa Fab/Hemagglutinin	1qfu ^b	2viu	0.00/0.32
1NMB ^c	Fab NC10/Neuraminidase	1nmb ^b	7nn9	0.00/0.28
1BQL ^c	Hyhel5 Fab/Lysozyme	1bql ^b	1dkj	0.00/0.84
<i>Others</i>				
1MDA	Methylamine dehydrogenase/Amicyanin	2bbk	1aan	1.53/2.60
1WQ1	RAS activating domain/RAS	1wer	5p21	0.97/0.48
1AVZ	HIV-1 NEF/FYN tyrosine kinase SH3 domain	1avv	1shf	0.68/1.00
1FIN	CDK2 cyclin-dependent kinase 2/Cyclin	1hcl	1vin	3.50/0.41
1A00 ^c	Che A/Che Y	1chn	1a0o ^b	1.45/0.00
1IGC ^c	IgG1 Fab fragment/Protein G	1igc ^b	1igd	0.00/1.05
1ATN ^c	Actin/Deoxyribonuclease I	1atn ^b	3dni	0.00/0.41
1GLA ^c	Glycerol kinase/GSF III	1gla ^b	1f3g	0.00/0.42
1SPB ^c	Subtilisin/Subtilisin prosegment	1sup	1spb ^b	0.59/0.00
3HHR ^c	Human growth hormone/Receptor	3hhr ^b	1hgu	0.00/3.59

^aThe r.m.s.d. (units Å) of the atom C α between the docked receptor or ligand and corresponding bound structure.

^bThe proteins with the bound structures (not free form).

^cThe complexes are reconstructed from the structures of unbound/bound or bound/unbound proteins.

ligand. For the remaining 20 systems, the unbound experimental structure of only one molecule was available, and therefore the bound structure was taken for the other molecule.

Treating molecular flexibility

For docking, we used the ‘simplified protein’ model (Levitt, 1976) with one sphere per residue and radii listed in the reference (Levitt, 1976) except for Arg, Lys, Asp, Glu and Met at the protein surface. As conformational changes often affect their flexible side chains (Cherfils and Janin, 1993; Lo Conte *et al.*, 1999; Zhao *et al.*, 2001), these residues were represented with spheres centered on the C β atom with a small radius of 1.5 Å, making the molecular surface ‘softer’ to some extent at these positions than elsewhere in the protein.

Searching

We used Wodak and Janin’s docking algorithm (Wodak and Janin, 1978) implemented in the program DOCK (Cherfils *et al.*, 1991). The six parameters that defined the position and orientation of one molecule relative to the other were five Euler rotation angles (θ_1 , ϕ_1 , θ_2 , ϕ_2 and χ) and an intermolecular distance ρ . Angles θ_1 and ϕ_1 located the center of the ligand relative to the receptor; θ_2 and ϕ_2 located the center of the receptor relative to the ligand; χ was a spin angle about the center line. The five angles were systematically searched in steps of 7.5°. We explored the full range of θ_2 ($\pm 90^\circ$), ϕ_2 and χ ($\pm 180^\circ$), that is, the full surface of the ligand. For the receptor, we restrict the search range of the θ_1 and ϕ_1 to $\pm 30^\circ$

around the active site. With a 7.5° step, about 4.86×10^6 different docked structures were generated for each complex.

Filtering

In this work, the docked structures with an interface area not less than 500 \AA^2 were retained and subjected to filtering. The filtering technique based on geometric matching, hydrophobicity (Zhang *et al.*, 1997) and electrostatic complementarity was from the analysis in the Combined filtering section in Results and discussion (see below). We used interface areas to score geometric and electrostatic energy for electrostatic complementarity.

For the enzyme–inhibitor complexes, first the top 500 docked structures were sorted according to descending interface area. For the following ones, their interface area values were compared with the sorted lookup list containing those of the 500 best geometric matching solutions found so far. If its surface matching is poorer than that of the worst solution in the list, it will be discarded. Otherwise, it will be saved and inserted in the list at the same time the worst element is eliminated. In this way, 500 solutions were retained by geometric matching. Then, in the solutions left, in the same way, those of the 500 least desolvation free energy solutions were added. Finally, the total 1000 solutions retained by geometric matching and hydrophobicity were combined as the last retained solutions.

For the antibody–antigen complexes, the 500 solutions first retained by geometric matching and the 500 solutions then retained by electrostatic complementarity were combined as the final retained solutions. For the other complexes, the total of 1500 solutions retained by geometric matching, hydrophobicity and electrostatic complementarity were combined as the final retained solutions.

Finally, for every system, several binding modes with similar structure were replaced with an average conformation (Cherfils *et al.*, 1991).

The scoring functions

After clustering, the following scoring function was evaluated:

$$\text{score} = \Delta E_{elec} + \Delta G_{des}(ACE) \quad (1)$$

where ΔE_{elec} denoted the changes in the electrostatic energy. A soft-core Coulombic potential was used to calculate electrostatic energy:

$$\Delta E_{elec} = k \cdot \frac{q_i q_j}{4(r_{ij} + c)^2} \quad (2)$$

where k was a constant including the electrical permittivity of vacuum and r_{ij} was the distance between atoms i and j . The constant c was set to 1.2 \AA . The charge parameters were from the CHARMM force field (Brooks *et al.*, 1983). $\Delta G_{des}(ACE)$ was the desolvation free energy based on the atomic contact energy (ACE) (Zhang *et al.*, 1997, 1999):

$$\Delta G_{des}(ACE) = \sum_i \sum_j n_{ij} e_{ij} \quad (3)$$

where e_{ij} denoted the ACE between atoms i and j , and n_{ij} was a switch function (Zhang *et al.*, 1999) applied to e_{ij} in the range $6\text{--}10 \text{ \AA}$ in order to avoid a sharp distance cutoff. The sum was taken over all atom pairs less than 10 \AA apart.

Results and discussion

Treatment of conformational flexibility

In order to examine the effect of the molecular flexibility treatment in our molecular model, we compared the docked

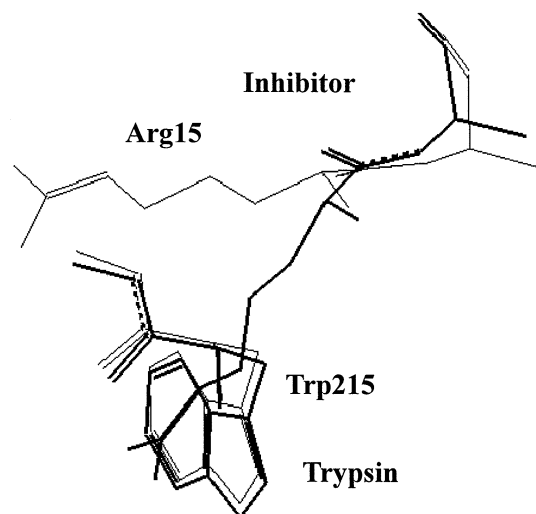


Fig. 1. Detail at the interface of the experimental structure of 1BRC and the structure docked starting from the enzyme trypsin (1bra) and inhibitor APPI (1aap) using the modified molecular model. Thin lines correspond to the experimental structure of 1BRC and thick lines show the conformations of the docked structure.

structure with our modified molecular model with the experimental structure. Figure 1 shows results obtained from the comparisons above for the complex 1BRC. The docking was performed starting from the superimposed structures (reference structure) of the enzyme trypsin (1bra) and its inhibitor APPI (1aap) upon the complex 1BRC, but far apart (200 \AA). Actually, in the association of the two molecules, an obvious conformational change occurs on the Arg15 side chain of the inhibitor APPI, which can be found by comparing the bound and unbound structures of the inhibitor APPI. From Figure 1, the docking using the modified molecular model tolerates the appropriate overlap between the Arg15 side chain of APPI and Trp215 of trypsin, whereas a major clash would be expected to appear if the docking is performed with the original molecular model. This means that our modified molecular model can reasonably allow the side chain flexibility of the surface residues.

Combined filtering

Since different types of the complexes have important differences in interface hydrophobicity and electrostatics, we attempted to apply different filtering criteria to different types of the complexes. In order to compare the filtering effect of the geometric matching, hydrophobicity and electrostatic complementarity for different types of the complexes, the numbers of native-like structures in the three lists containing those of the 1000 best interface matching, 1000 least desolvation free energy and 1000 best electrostatic complementarity solutions are listed in columns S, A and E (under Filtering) in Table II. The ratio of the number of native-like structures to that of the retained solutions is a key factor in evaluating the filtering effect. A docked structure is considered a native-like structure if the root mean square deviation (r.m.s.d.) of the backbone atoms (N, C α , C, O) from the reference structure is not greater than 4.0 \AA .

From Table II, for the enzyme–inhibitor complexes, in most cases native-like solutions selected by electrostatic complementarity are fewer than those selected by geometric matching and by hydrophobicity. Furthermore, it is clear that only if the native-like solutions selected by these two criteria are combined

Table II. Results of molecular docking calculations for 44 protein–protein complexes^a

Complex	Filtering			Scoring		Complex	Filtering			Scoring	
	S	A	E	Rank	R.m.s.d.		S	A	E	Rank	R.m.s.d.
<i>Enzyme–inhibitor</i>											
1ACB	9	143	19	16	3.64	1TGGS	39	57	9	15	2.01
1BRC	0	107	19	1	3.84	1BTH	146	64	32	15	3.98
1BRS	21	7	19	89	3.73	2SIC	9	77	6	4	2.12
1CHO	25	208	58	19	3.04	1FQ1	9	0	0	321	3.41
1CGI	44	122	36	7	3.85	1TAB ^b	68	66	20	17	3.39
1CSE	31	159	8	1	3.68	1PPE ^b	464	280	186	1	2.03
1FSS	60	51	106	1	3.89	1UDI ^b	48	7	19	49	3.50
2KAI	14	38	19	4	2.33	2TEC ^b	20	80	12	13	2.91
1MAH	70	148	17	3	1.55	1STF ^b	19	32	6	4	2.56
2PTC	4	60	9	13	1.45	4HTC ^b	35	6	0	19	2.72
1UGH	27	6	12	164	1.55	1EFU ^b	0	8	7	10	4.00
2SNI	30	66	27	9	2.12						
<i>Antibody–antigen</i>											
1WEJ	13	0	17	17	2.38	1JHL ^b	3	0	14	103	3.66
1AHW	5	0	38	15	2.32	2VIR ^b	9	4	4	221	2.92
1BVK	8	2	6	200	3.48	1QFU ^b	7	5	9	273	2.39
1DQJ	9	2	6	198	3.61	1NMB ^b	3	3	3	205	3.14
1FBI ^b	22	0	3	18	3.54	1BQL ^b	6	0	0	57	1.02
2JEL ^b	33	0	9	20	3.62						
<i>Others</i>											
1MDA	9	56	8	1	2.72	1IGC ^b	42	93	9	17	2.64
1WQ1	20	3	9	19	3.23	1ATN ^b	4	5	0	3	1.82
1AVZ	8	0	7	131	3.73	1GLA ^b	6	52	7	7	3.38
1FIN	4	22	3	9	3.78	1SPB ^b	52	9	8	12	1.09
1A00 ^b	6	8	31	176	3.68	3HHR ^b	17	3	7	350	4.00

^aThe Filtering column gives the number of native-like docked structures retained according to the different filtering criteria: S, geometric matching filter; A, hydrophobicity filter; E, electrostatic complementarity filter. Interface area, electrostatic energy (Equation 2) and atomic contact energy (Equation 3) are used to evaluate the geometric matching, electrostatic complementarity and hydrophobicity, respectively. Scoring presents the highest ranking position of a native-like docked structure and the corresponding r.m.s.d. (units Å) relative to the reference structure.

^bThe complexes are reconstructed from the structures of unbound/bound or bound/unbound proteins.

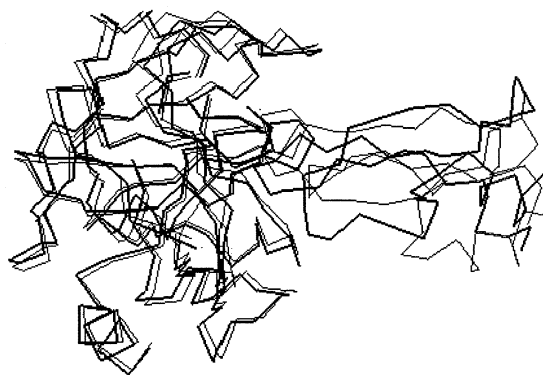


Fig. 2. Superposition of the experimental structure of protein complex 1BRC and the best ranked native-like prediction reported in Table II. Thick lines, C α trace of experimental structure; thin lines, C α trace of predicted mode.

as the retained solutions is the approach universal. For example, for 1BRC, the native-like solutions obtained by geometric matching are zero, but 107 native-like structures are selected by hydrophobicity. For 1FQ1, however, we find no native-like structures by hydrophobicity and nine native-like structures by geometric matching. Therefore, the combined filtering criterion based on geometric matching and hydrophobicity is implemented for enzyme–inhibitor complexes. For antibody–antigen complexes, the filtering effect of the hydrophobicity is bad and that of geometric matching and electrostatic complementarity seems to be relatively good. For the other complexes,

it is clear that the three filtering criteria are all important in different cases.

As we know, the principles governing protein–protein recognition have obvious differences for different types of complexes. Probably it is the biological function that determines those differences. From an evolutionary perspective, enzyme–inhibitor complexes have evolved over a long period to optimize the interfaces performing their biological functions, which makes the interfaces more like the interior of proteins. Therefore, hydrophobic interaction is prominent in the association. In contrast, the antibody–antigen recognition is a ‘happstance’ not subject to evolutionary optimization over more than a few days. The contribution of the hydrophobic interaction to antibody–antigen association is relatively poor, whereas the electrostatic interaction seems to be very important. There could be some other biological principles governing antibody–antigen recognition. For other complexes, since the biological functions are diverse, there are no evident principles of recognition observed in the analysis above. Perhaps the other complexes can be divided into homodimers and heterocomplexes based on their structures. This can be done with the increasing structures of those complexes.

Scoring putative complexes

Table II summarizes the docking results (under Scoring). The ranking position of the first native-like structure is listed for each of the 44 complexes, followed by the corresponding r.m.s.d. relative to the reference structure. There are 30 cases in which the first native-like structures are ranked within the

top 20. These cases include the complexes 1CHO, 1CGI, 1TGS, 1EFU*, 1MDA, 1FIN and 1IGC*, in which the relatively large integral conformational changes of the receptors or ligands occur during the complex formation (see the last column for C α r.m.s.d. in Table I). It should be noted, however, that the native-like structures do not always correspond to the best scoring solutions and, often, incorrect docked structures are ranked first. Perhaps the effect of evaluating native-like structures will be better if in the scoring function, H-bond and van der Waals energies are taken into account in addition to electrostatic and desolvation energies after energy minimization. Additionally, properly combined with experimental information on the complex, the method will increase the successful probability of predicting the complex structures.

Figure 2 shows a comparison between the experimental structure of the complex 1BRC and the best-ranked native-like prediction reported in Table II. Although there is a major clash between the Arg15 side chain of APPI and Trp215 of the enzyme trypsin (see Figure 1), the native-like structure is ranked first and it is clear that the binding site is satisfactorily identified.

Conclusions

It should be pointed out that the docking simulations in this paper are based on the assumption that the binding region on one of the two proteins is known. In the spherical polar coordinates used in this work, this information is given as a simple constraint in just one or two of the angular degrees of freedom. The computation time is much reduced. Ritchie and Kemp also used the same coordinates in their docking algorithm and successfully predicted the structures of some protein-protein complexes (Ritchie and Kemp, 2000). In their test, when the search ranges of two angle degrees of freedom are limited to $\pm 30^\circ$ around the active site, the first native-like structures of 11 out of 18 complexes are ranked the top 20 (Ritchie and Kemp, 2000). In this paper, the first native-like structures of 30 out of 44 tested complexes are ranked the top 20. This indicates that our algorithm captures some important factors in the protein-protein association and can provide useful help for the study of molecular recognition.

The guidance of docking by the characters on protein-protein interfaces will be important. Currently, many important features of antibody-antigen interfaces have been reported. For example, tyrosine residues represent over a quarter of the total interaction energy donated by the antibody (Jackson, 1999). Therefore, we might add this information to filtering or scoring just for antibody-antigen docking.

In summary, our soft docking algorithm has several advantages: (1) the modified molecular model can improve the simulation result for the unbound protein-protein docking; (2) the type-dependent filtering technique can retain much more native-like structures and increase the successful probability of predicting complex structures; and (3) the scoring function based on the binding free energy can effectively distinguish the correct from the incorrect structures. However, the main shortcoming of this algorithm is that only a partial binding space is searched. This is obviously a limitation for the docking simulations in which no information about the binding site is known. Work on improving our docking algorithm is in progress.

Acknowledgements

We thank Professor Janin and Dr Cherfils for providing the rigid-body protein-protein docking program. We also thank Dr Ben Zhuo Lu for helpful

discussions. This work was supported in part by the Chinese Natural Science Foundation (No.s 29992590-2, 30170230 and 10174005).

References

- Betts,M.J. and Sternberg,M.J.E. (1999) *Protein Eng.*, **12**, 271–283.
- Brooks,B.R., Brucocoleri,R.E., Olafson,B.D. and States,D.J. (1983) *J. Comput. Chem.*, **4**, 187–217.
- Cherfils,J. and Janin,J. (1993) *Curr. Opin. Struct. Biol.*, **3**, 265–269.
- Cherfils,J., Duquerroy,S. and Janin,J. (1991) *Proteins: Struct. Funct. Genet.*, **11**, 271–280.
- Decanniere,K., Transue,T.R., Desmyter,A., Maes,D., Muyldermans,S. and Wyns,L. (2001) *J. Mol. Biol.*, **313**, 473–478.
- Fischer,D., Lin,S.L., Wolfson,H.J. and Nussinov,R. (1995) *J. Mol. Biol.*, **248**, 459–477.
- Gabb,H.A., Jackson,R.M. and Sternberg,M.J. (1997) *J. Mol. Biol.*, **272**, 106–120.
- Halperin,I., Ma,B., Wolfson,H. and Nussinov,R. (2002) *Proteins: Struct. Funct. Genet.*, **47**, 409–443.
- Jackson,R.M. (1999) *Protein Sci.*, **8**, 603–613.
- Jiang,F. and Kim,S.H. (1991) *J. Mol. Biol.*, **219**, 79–102.
- Jones,S. and Thornton,J.M. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
- Katchalski-Katzir,E., Shariv,I., Eisenstein,M., Friesem,A.A., Aflalo,C. and Vakser,I.A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2195–2199.
- Lengauer,T. and Rarey,M. (1996) *Curr. Opin. Struct. Biol.*, **6**, 402–406.
- Levitt,M. (1976) *J. Mol. Biol.*, **104**, 59–107.
- Lo Conte,L., Chothia,C. and Janin,J. (1999) *J. Mol. Biol.*, **285**, 2177–2198.
- Norel,R., Retey,D., Wolfson,H.J. and Nussinov,R. (1999a) *Proteins: Struct. Funct. Genet.*, **35**, 403–419.
- Norel,R., Retey,D., Wolfson,H.J. and Nussinov,R. (1999b) *Proteins: Struct. Funct. Genet.*, **36**, 307–317.
- Palma,P.N., Krippahl,L., Wampler,J.E. and Moura,J.J.G. (2000) *Proteins: Struct. Funct. Genet.*, **39**, 372–384.
- Ritchie,D.W. and Kemp,G.J.L. (2000) *Proteins: Struct. Funct. Genet.*, **39**, 178–194.
- Sandak,B., Nussinov,R. and Wolfson,H.J. (1995) *Comput. Appl. Biosci.*, **11**, 87–99.
- Shoichet,B.K. and Kuntz,I.D. (1991) *J. Mol. Biol.*, **221**, 327–346.
- Sotriffer,C.A., Flader,W., Winger,R.H., Rode,B.M., Liedl,K.R. and Varga,J.M. (2000) *Methods*, **20**, 280–291.
- Vakser,I.A. (1995) *Protein Eng.*, **8**, 371–377.
- Walls,P.H. and Sternberg,M.J. (1992) *J. Mol. Biol.*, **228**, 277–297.
- Wodak,S.J. and Janin,J. (1978) *J. Mol. Biol.*, **124**, 323–342.
- Zhang,C., Vasmatzis,G., Cornette,J.L. and DeLisi,C. (1997) *J. Mol. Biol.*, **3**, 707–726.
- Zhang,C., Chen,J. and DeLisi,C. (1999) *Proteins: Struct. Funct. Genet.*, **34**, 255–267.
- Zhao,S., Goodsell,D.S., Olson,A.J. (2001) *Proteins: Struct. Funct. Genet.*, **43**, 271–279.

Received August 8, 2002; revised January 2, 2003; accepted February 11, 2003