

Incorporating intermolecular distance into protein–protein docking

Hongxing Lei and Yong Duan¹

UC Davis Genome Center and Bioinformatics Program, Department of Applied Science, University of California, One Shields Avenue, Davis, CA 95616, USA

¹To whom correspondence should be addressed.
E-mail: duan@ucdavis.edu

In this work, intermolecular distance was integrated into the docking of protein–protein complexes. To develop an efficient docking procedure, 22 enzyme–inhibitor targets and 15 antibody–antigen targets were taken from a benchmark set. A three-step approach was adopted, which included global sampling by FTDOCK, filtering by intermolecular distance and ranking by a composite scoring function. For the enzyme–inhibitor targets, the composite scoring function consists of geometry and energy terms. In the set composed of the ~100 highest ranked candidates for each target, correct complexes were identified for all of the 22 enzyme–inhibitor targets. This docking strategy also succeeded on the four test targets, of which three are CAPRI targets with the same receptor but different binding modes. Interestingly, all three binding modes were correctly predicted. For the antibody–antigen targets, CDR and physical energy were also used in the filtering process and informatics terms were added to the scoring function. The composite score had successful prediction for 13 of the 15 antibody–antigen targets.

Keywords: antibody–antigen/enzyme inhibitor/
intermolecular distance/protein–protein docking

Introduction

Protein–protein interactions (PPI) are of great importance to biological systems (Jones and Thornton, 1996). However, the number of known complex structures is limited owing to the technical challenges involved in obtaining high-resolution experimental structures. Guided by surface complementarity (Katchalski-Katzir *et al.*, 1992), electrostatics (Mandell *et al.*, 2001) and desolvation energy (Zhang *et al.*, 1997; Chen and Weng, 2002), protein–protein docking (Camacho and Vajda, 2002; Smith and Sternberg, 2002) procedures have succeeded in predicting some complex structures that are very close to those determined experimentally. This is highlighted by the encouraging results in the recent CAPRI (critical assessment of predicted interactions) experiments (Ben-Zeev *et al.*, 2003; Camacho and Gatchell, 2003; Fernandez-Recio *et al.*, 2003; Law *et al.*, 2003; Mendez *et al.*, 2003; Smith and Sternberg, 2003). Docking procedures usually consist of the following steps (Halperin *et al.*, 2002): (a) generating a candidate pool by rigid-body global search, (b) enrichment of candidate pool by filtering with biochemical information, (c) selecting a small portion of candidates based upon energy ranking, then (d) clustering the selected candidates and (e) refining the

highest populated clusters by local search with certain side chain and main chain flexibility (Camacho and Vajda, 2001).

One of the greatest challenges in protein–protein docking is to reduce the number of candidates from typically hundreds of thousands to hundreds while still retaining a reasonable number of near-native complex structures. Part of the challenge can be attributed to the flexibility of ligands and receptors, especially at the interface (Betts and Sternberg, 1999), and the rough docking energy landscape. This problem is further aggravated by the marginal stability of protein–protein complexes (typically near 10 kcal/mol) (Gray *et al.*, 2003). Although protein flexibility can be modeled by Gaussian network (Bahar *et al.*, 1997) and other methods (Jacobs *et al.*, 2001), it is usually infeasible to introduce flexibility at the early stage of docking. Therefore, reduced side chain representations (Heifetz and Eisenstein, 2003; Li *et al.*, 2003b; Zacharias, 2003) and soft potentials (Fernandez-Recio *et al.*, 2002) have been applied to smooth the energy landscape in rigid-body docking. Along this line, alternative approaches such as the use of multiple copies of receptors or ligands combined with more efficient search algorithms could also be contemplated.

With the aim of predicting protein–protein interactions, comprehensive studies have been conducted to characterize protein–protein interfaces (Nooren and Thornton, 2003a; Ofra and Rost, 2003). Some of the features of the protein–protein interface identified in these studies include the number of surface patches, contact area, polarity, flatness and residue conservation (Nooren and Thornton, 2003b). Interfaces smaller than 1000 Å² usually consist of only a single patch with a core and a rim (Chakrabarti and Janin, 2002). The composition of core residues is distinctive from rim and non-interacting surfaces. Interfaces are less planar for enzyme–inhibitor complexes than other types of interfaces. Hydrophobic residues are more abundant for larger interfaces and polar residues are more abundant for smaller interfaces. ‘Hot spots’ (Ma *et al.*, 2001) are interface residues whose mutations have a significant effect on the binding free energies. Most ‘hot spots’ comprise tryptophan, tyrosine or arginine residues, whereas serine, threonine and leucine residues are least likely to be ‘hot spots’ (Bogan and Thorn, 1998). Interface residues tend to be more structurally conserved than non-interface residues (Ma *et al.*, 2003). For the enzyme–inhibitor interface, the conservation is significantly higher at the enzyme interface than the inhibitor interface (Bradford and Westhead, 2003). Residue pairing also has distinct preferences (Glaser *et al.*, 2001), favoring hydrophobic–hydrophobic and polar–polar pairing.

Universal scoring functions have been dominant in the field of protein–protein docking. However, different classes of complexes may have different binding mechanisms. It was noted that protease inhibitors interact predominantly through main chain to main chain interactions whereas sidechain interactions play a dominant role in antibody–antigen complexes (Jackson, 1999). Hence, our strategy is to develop different procedures

based on the specific features for a class of complexes (Li *et al.*, 2003a). Among the structurally determined complexes, enzyme–inhibitor and antibody–antigen are the two dominant classes. For antibody–antigen complexes, complementarity-determining regions (CDR) of antibodies are conserved both sequentially and structurally. This feature has been widely used to limit the search space for antibody–antigen complexes (Chothia and Lesk, 1987) and was shown to enhance greatly the predictive power of docking (Schneidman-Duhovny *et al.*, 2003). In addition to the CDR restriction, antibody interfaces are also abundant in ‘hot spot’ residues tryptophan and tyrosine. As for the enzyme–inhibitor class, most complexes share a common geometric feature at the interface, which is a good packing comprised of a convex inhibitor surface and a concave enzyme surface (activity pocket). Although this is a well-known geometric feature, it has not been well utilized in guiding the docking of enzyme–inhibitor complexes.

Here we explore the applicability of intermolecular distance on the evaluation of packing and subsequently the ranking of docking candidates. We adopted a three-step approach: (a) global search by rigid-body docking using FTDOCK; (b) filtering by intermolecular distance; additional filters CDR and physical energy were applied to antibody–antigen docking; (c) ranking by a composite scoring function including geometry, energy and informatics terms. The distance between the residues at the ligand interface and the center of the receptor is defined as the intermolecular distance. The geometry terms include the intermolecular distance and the interface size as approximated by the number of residues at the interface. There are three energy terms. The physical energy includes the soft electrostatics and van der Waals potentials. The statistical potential at residue level is represented by residue-pair (RP) score (Moont *et al.*, 1999), the default scoring function of the docking suite 3D-DOCK (Gabb *et al.*, 1997). The third energy term is the statistical potential at atom level, i.e. atom contact energy (ACE). The informatics terms include the CDR score and the score for the enrichment of Trp and Tyr. The informatics terms only apply to the antibody–antigen complexes. A subset of ~ 100 highest ranked candidates for each target formed the prediction set and the performance of the scoring function was evaluated based on the quality of the prediction set.

Methods

The dataset of protein complexes used in this study was compiled by Chen *et al.* (2003b). All structure coordinates of the complexes and the individual receptors and ligands were retrieved directly from the protein data bank (PDB). All of the 22 complexes in the enzyme–inhibitor class and the 19 complexes in the antibody–antigen class were included in this study. Sixteen of the enzyme–inhibitor complexes are unbound–unbound targets, where the structures of both the receptor and ligand have been solved separately. The other six are unbound–bound targets because the ligand structures are not available in free form. Among the antibody–antigen complexes, five are unbound–unbound targets and the other 14 are unbound–bound targets. Fifteen of the antibody–antigen complexes have antigen interacting with the two variable domains of antibody, which is typical for antibody–antigen interaction. The other four were complexes of enzyme and single-domain antibody; they were reassigned as the test

targets of the enzyme–inhibitor class in this study because of the similar interface features.

FTDOCK and RP score are part of the 3D-DOCK suite (Gabb *et al.*, 1997) which was obtained from the Sternberg group (<http://www.bmm.icnet.uk/docking>). The RP score is based on statistical analysis of residue pairing preferences of a non-redundant protein complex database (Glaser *et al.*, 2001). To avoid unnecessary sampling, the non-CDR domains of the antibodies were removed. The coordinates of receptors and ligands were preprocessed to remove hydrogen atoms and alternative atom coordinates by the preprocess-pdb.perl utility program in 3D-DOCK. The setup parameters for the FTDOCK runs included a 0.7 Å grid unit, 1.3 Å surface thickness and 9° rotation angle step size. Receptors (enzyme or antibody) were set as the static moiety and ligands (inhibitor or antigen) were set as the mobile moiety. The only exception was the target 1DFJ, for which the static moiety was the inhibitor because of the much larger size of the inhibitor (456 residues) compared with the enzyme (124 residues). There were three independent runs for each target; the starting coordinates for the second and third runs were generated by the randomspin utility program. For the enzyme–inhibitor targets, 10 000 candidates were saved in each run, resulting in a total of 30 000 candidates for each target. The number of saved candidates was doubled for the antibody–antigen targets to account for their larger system sizes. Each of the three runs was analyzed separately throughout this work. Following the completion of each FTDOCK run, the residue-pair potential for each candidate was calculated using the ‘rpscore’ program and normalized to 0–100 for each run.

The quality of the docking candidates was evaluated by the ligand root mean square deviation (r.m.s.d.) against the native complex. We followed the standard procedure where the receptors were aligned first and then the r.m.s.d. of the ligand C α atoms was calculated (Mendez *et al.*, 2003). A candidate with ligand r.m.s.d. <10 Å is defined as a ‘hit’. The ‘build’ utility program was used to regenerate candidate structures for r.m.s.d. evaluation and other calculations. If any atom of a residue is within 5 Å of any atom on the other protein, this residue is defined as an interface residue. This is the standard cutoff used in CAPRI. The intermolecular distances include the average and minimum distances (ADIS and MDIS) between the ligand interface and the receptor center (defined as center of geometry). These two distances were normalized to 0–100 for each run.

The charges and radii for calculating the interaction energy between the receptor and ligand were taken from a recently revised AMBER force field (Duan *et al.*, 2003). Only electrostatics and van der Waals terms were included in the interaction energy. To reduce the noise caused by atom overlap in rigid-body docking, smooth energy functions were adopted. The electrostatic energy terms were truncated when atoms were closer than their van der Waals distances. Furthermore, only the attractive London dispersion energy terms were considered and the repulsive terms were ignored. These terms were combined as interaction energy and normalized to 0–100 for each run. The parameters for calculating atomic contact energy were taken from Zhang *et al.* (1997). The contact radius was set to 6.0 Å according to the original work. The contact energy between the receptor and the ligand was calculated for each candidate. The ACE energy was also normalized to 0–100 for each run.

For enzyme–inhibitor complexes, the only filter was ADIS with a cutoff of 50. Following the filtering, a composite scoring function was applied to rank the remaining candidates. The scoring function S_{ENZ} consists of two components: energy terms and geometry terms.

$$S_{ener} = S_{phys} + S_{ACE} + S_{RP}$$

$$S_{geo} = S_{ADIS} + S_{MDIS} + S_{NR}$$

where the energy function (S_{ener}) is composed of physical interaction energy (S_{phys}), atomic contact energy (S_{ACE}) and residue-pair potential (S_{RP}). The sign of the original RP score was reversed to be consistent with scoring scheme which prefers lower score. To emphasize the physical energy over the statistical energies, the weight of the physical energy was set as 2. The geometry score (S_{geo}) is composed of two intermolecular distance scores (S_{ADIS} and S_{MDIS}) and a score reflecting the size of the receptor interface (S_{NR}). The score S_{MDIS} was added to favor those candidates with inhibitor deeply buried in the enzyme pocket. The number of residues at the receptor interface was scaled up three times to make the variation range close to that of the other terms. The sign of the last term was reversed to be consistent with the whole scheme. Candidates with lower scores were ranked higher in this scoring scheme.

For antibody–antigen complexes, the filtering consists of two steps. First, each residue of the antibody was classified into CDR residue and non-CDR residue. For each candidate, the percentage of CDR residues at the antibody interface was calculated and the candidate is retained if this percentage is above 60%. The first filter is to ensure that most of the antibody–antigen interaction involves the CDR residues. In the second filtering step, the physical interaction energy filter was also applied in addition to the distance filter and the cutoff was also set to 50. After filtering, a composite scoring function was applied to rank the remaining candidates. The scoring function S_{ANT} consists of three components: energy terms, geometry terms and informatics terms.

$$S_{ener} = S_{phys} + S_{ACE} + S_{RP}$$

$$S_{geo} = S_{ADIS} + S_{NR} + S_{NL}$$

$$S_{info} = S_{CDR} + S_{hot}$$

There were a few changes compared with the scoring function for enzyme–inhibitor docking. S_{MDIS} was removed because antibodies do not have reaction pockets like enzymes do and a score reflecting the size of the ligand interface (S_{NL}) was added. The informatics terms (S_{info}) are specific for the antibody–antigen class, which favor the candidates with a higher percentage of CDR residues (S_{CDR}) and more ‘hot spot’ residues Tyr and Trp (S_{hot}) at the antibody interface. The signs for the informatics terms were reversed to be consistent with the whole scoring scheme. The number of ‘hot spot’ residues was scaled up 10-fold to achieve a similar variation range as for other terms. The percentage of CDR residues was scaled up 4-fold because this is the most prominent feature for antibody–antigen interaction.

The prediction set of each target was composed of 33 highest ranked candidates from each of the three runs, making a total of 99 candidates in the prediction set. When the component score or individual terms were examined, the ranking was based on the specific score studied rather than the composite score.

The performance was evaluated mainly by the number of ‘hits’ in the prediction set. In addition, the lowest r.m.s.d. of the candidates in the prediction set and the highest rank of the ‘hits’ were also presented for evaluation.

Results

In this section, we first provide the results in details for the enzyme–inhibitor docking, which include the evaluation of the global search, filtering and ranking processes. Then the results from a set of four test targets, including three CAPRI targets, are presented. Finally, we present a summary of the results for the antibody–antigen docking.

Generating candidate pool by FTDOCK

Among the 22 enzyme–inhibitor targets studied here, 16 started from unbound-unbound structures and the other six started from unbound-bound structures (see Methods). There were three rigid-body docking runs for each target, starting from different orientations of both the receptor and the ligand. In each of the three runs, 10 000 candidates were saved. The quality of each candidate was evaluated by the ligand r.m.s.d. (C^α) of the candidate using the native complex as reference (see Methods). The number of hits is summarized in Table I (group definition will be described later in this section). Here we define a hit as a candidate with a ligand r.m.s.d. below 10 Å. The number of hits among the 30 000 candidates of each target ranged from 13 to 534 with a median

Table I. Summary of FTDOCK and filtering for enzyme–inhibitor targets

Complex	FTDOCK ^a		Filtering ^b			
	Hits ^c	R.m.s.d. ^d	Hits	R.m.s.d.	Candidates	P_{hits} ^e
Group 1						
1PPE	534	1.07	531	1.07	10827	99.4
1TGS	179	2.44	177	2.44	13454	98.9
1CHO	122	2.01	121	2.01	13675	99.2
1BRC	46	2.13	46	2.13	6324	100
1CGI	85	2.18	85	2.18	13468	100
1AVW	25	4.72	25	4.72	13245	100
2SIC	17	3.34	16	3.34	9771	94.1
4HTC	30	2.46	12	2.73	3953	40.0
2KAI	72	4.83	72	4.83	8030	100
2PTC	49	4.03	49	4.03	10248	100
1TAB	128	1.18	119	1.18	8510	93.0
Group 2						
2SNI	44	4.32	43	4.32	13679	97.7
1ACB	27	2.83	27	2.83	11351	100
1CSE	46	2.33	44	2.33	13082	95.7
2TEC	140	1.46	139	1.46	9955	99.3
Group 3						
1UGH	68	2.87	68	2.87	18865	100
1UDI	36	2.01	33	2.01	13459	91.7
1STF	35	1.17	34	1.17	13713	97.1
Group 4						
1BRS	63	5.44	63	5.44	20391	100
1DFJ	13	6.14	13	6.14	9758	100
1MAH	26	1.25	26	1.25	23344	100
1FSS	30	3.58	30	3.58	20094	100

^a10 000 candidates were saved in each of the three independent runs.

^bCandidate selection was based on intermolecular distance ADIS (see Methods).

^cNumber of complexes with ligand r.m.s.d. (C^α only) lower than 10 Å.

^dThe lowest ligand r.m.s.d. in the set.

^eThe percentage of hits retained after filtering.

value of 46. Most (17) of the 22 targets had fewer than 100 hits. The median hits-to-candidates ratio was 1:652, making it very challenging to pick up hits with limited selections.

The number of hits is directly correlated with the system size, especially the ligand size. The complex 1PPE, whose ligand is the smallest (only 29 residues), had the most hits (534). The complex with the second smallest ligand (1TAB, 36-residue ligand) had 128 hits. In contrast, the three complexes with the largest ligands (1DFJ, 2SIC and 1AVW) had the fewest hits (13, 17 and 25). The two complexes with the largest receptors (1FSS and 1MAH) also had few hits (26 and 30, respectively). This appears to be a sampling issue. Because the ligands were rotated at the fixed intervals, smaller ligands were sampled with smaller relative surface movement (hence finer grain sampling). Another factor is that a small angular deviation can lead to a large r.m.s.d. for complexes with a large receptor or ligand because of the way in which the ligand r.m.s.d. is evaluated. From this point of view, interface r.m.s.d. could be a better standard without bias to system size. However, interface r.m.s.d. is sensitive to side chain orientation, which is beyond the scope of this work.

Here we define the ‘best hit’ of a target as the candidate with the lowest r.m.s.d. The lowest r.m.s.d. among the 22 targets

ranged from 1.07 to 6.14 Å with a median value of 2.45 Å. The best hits of the two complexes with the smallest ligands (1PPE and 1TAB) are of very high quality (1.07 and 1.18 Å). Out of 22 targets, 20 complexes have the lowest r.m.s.d. <5.0 Å and only two complexes have the lowest r.m.s.d. >5 Å (6.14 Å for 1DFJ and 5.44 Å for 1BRS). The low quality of the best hit for 1DFJ can be attributed to the size (large ligand and receptor). Interestingly, 1BRS is a receptor that is half the size of most complexes. However, its best hit is only 5.44 Å and is considered a low-quality hit. For the six unbound/bound targets, the quality of the best hits was higher, with r.m.s.d. ranging from 1.07 to 2.46 Å. This suggests that accounting for the conformational changes will likely be helpful, but it is beyond the scope of the current study, which is based on rigid-body docking approaches and is meant to examine the issue of scoring function.

We further examined each complex structure and found that there is a great variety among these enzyme–inhibitor complexes. The classification based on interface characteristics is shown in Table I. The representative structures are shown in Figure 1. There are four major groups among these 22 complexes. The first group has 11 members with the interface comprised of a pseudo-loop of the ligand and a concave surface

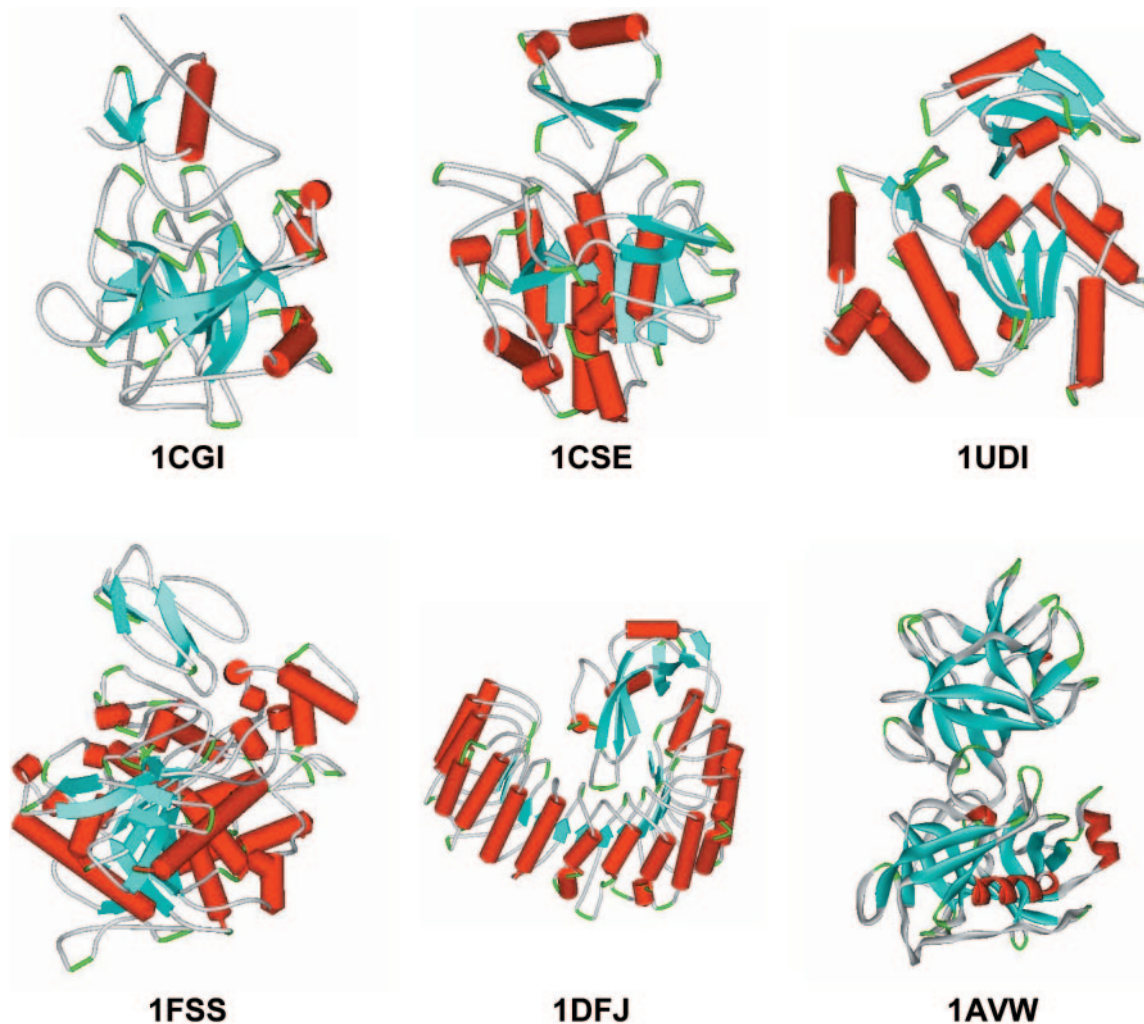


Fig. 1. Representative structures of enzyme–inhibitor complexes. Group 1 is represented by 1CGI, group 2 by 1CSE and group 3 by 1UDI. Two of the members of group 4 (1FSS and 1DFJ) are shown. Another member of group 1 (1AVW) is also shown. This figure was created by DS ViewerPro.

of the receptor. A pseudo-loop is defined here as a loop with one free end, compared with the classic loop in which both ends are attached to internal secondary structures. A representative structure of the first group is 1CGI, shown in Figure 1. Three members of the first group (1AVW, 2SIC and 4HTC) do not strictly conform to the definition of the group, but they all have at least one free end involved in the binding. The second group has four members with the interface comprised of an expanded loop of the ligand and a concave surface of the receptor. A representative structure of the second group is 1CSE (Figure 1). The third group has three members with the interface comprised of a β -strand of the ligand and a cleft surface of the receptor. A representative structure of the third group is 1UDI. The remaining four complexes do not belong to these three groups. The interface of 1BRS consists of an α -helix of the ligand and a β -sheet of the receptor. 1DFJ has a large circular inhibitor formed by internal repeats. 1FSS and 1MAH both have large globular receptors over 500 residues. We put these four complexes in the fourth group. Overall, for the fourth group, the number of hits is less and the quality of the best hits is lower.

Filtering by intermolecular distances

A simple measurement that captures the concave–convex features of enzyme–inhibitor complexes is ADIS, the average distance between residues at the ligand interface and the geometric center of the receptor. ADIS was normalized to a score between 0 and 100 in each FTDOCK run. In the filtering process, ADIS was the only filter applied to the enzyme–inhibitor targets. To avoid over-fitting to this data set, we chose to use a rather coarse cutoff. Candidates with normalized ADIS score >50 were filtered out. A summary of the filtering is shown in Table I. The total number of the remaining candidates ranged from 3953 to 23 344 with a median of 13 163. The percentage of the hits retained was $>90\%$ for all targets except for 4HTC, where a long tail of the inhibitor is involved in the interaction. For half of the 22 targets, there was no loss of hits after filtering. This shows that filtering reduced the candidate space to 43.9% (median) of the original set while retaining nearly all of the hits. Additionally, the lowest r.m.s.d. stayed the same for 21 of the 22 targets and there was a slight increase of the lowest r.m.s.d. from 2.46 to 2.73 Å for the target 4HTC.

Ranking by a composite scoring function

The composite scoring function S_{ENZ} included energy terms (S_{ener}) and geometry terms (S_{geo}). The energy terms included physical energy, atomic contact energy and residue-pair potential. The geometry terms included ADIS, MDIS and the size of the receptor interface. To evaluate the ranking ability of S_{ENZ} , 33 highest ranked candidates were selected from each of the three runs, forming the prediction set of 99 candidates. The performance by this scoring function is shown in Table II. Within the prediction set, 2–88 hits were retained for the 22 targets. The median number of hits was 11.5. According to the lowest r.m.s.d., the quality of the best hits in the prediction set remains the same as or similar to the best hits in the filtered set for most cases. The only target with significant increase of r.m.s.d. was 1ACB (the lowest r.m.s.d. increased from 2.83 to 4.56 Å). The best ranks of the hits were between 1 and 26 for the 22 targets and half of the best ranks were 1.

The contributions to the ranking ability of the composite score S_{ENZ} from the two components S_{ener} and S_{geo} are shown

Table II. Comparison of predictions by different scores for enzyme–inhibitor targets^a

Complex	S_{geo}			S_{ener}				S_{ENZ}			
	S_{ADIS}	S_{MDIS}	S_{NR}	S_{geo}	S_{phys}	S_{ACE}	S_{RP}	S_{ener}	Hits	R.m.s.d.	Rank
Group 1											
1PPE	49	21	66	56	0	11	70	51	83	1.07	1
1TGS	28	18	17	38	0	1	37	9	37	2.44	1
1CHO	0	39	9	18	8	0	31	37	44	2.01	1
1BRC	30	23	12	31	4	0	28	27	30	2.13	1
1CGI	7	9	4	8	0	4	14	2	11	2.48	10
1AVW	10	5	0	9	0	2	12	8	11	5.51	1
2SIC	8	1	0	10	0	7	5	0	3	3.34	26
4HTC	0	0	11	3	0	0	0	0	7	2.73	6
2KAI	20	6	5	16	8	2	0	6	21	4.83	1
2PTC	22	5	0	18	0	0	0	2	14	4.03	1
1TAB	16	11	17	29	0	0	0	0	14	1.18	4
Group 2											
2SNI	2	5	1	1	0	2	1	2	5	4.32	5
1ACB	4	7	2	2	0	9	0	0	2	4.56	8
1CSE	10	12	1	9	0	4	0	0	11	2.33	2
2TEC	39	58	66	62	32	62	0	52	88	1.46	1
Group 3											
1UGH	0	0	6	0	0	4	1	14	12	2.87	1
1UDI	0	0	7	0	15	0	0	23	18	2.01	2
1STF	0	7	19	20	0	20	0	8	19	1.17	1
Group 4											
1BRS	0	0	0	0	4	0	3	7	5	6.74	4
1DFJ	2	4	4	5	12	0	1	9	11	6.14	1
1MAH	0	0	0	1	10	2	3	12	8	1.25	3
1FSS	0	0	0	0	2	0	3	15	6	3.58	12

^aSee Table I and Methods.

in Figure 2. The data are also presented in Table II. In its prediction set, the geometry score retained hits for 18 targets, including all the targets in groups 1 and 2. The energy score selected hits for 17 targets, including all the targets in groups 3 and 4. The members in the first two groups are examples of the ‘classic’ enzymes with large concave receptor interface (activity pocket) and loop-like ligand interface. The geometry score seems to capture this feature very well and it is sufficient to make correct predictions by geometry score alone. For the ‘non-classical’ enzymes in groups 3 and 4, the geometry feature is no longer typical, but the energy score is sufficient to make correct predictions. This may suggest different binding mechanisms for the first and last two groups of enzyme–inhibitor complexes.

We further investigated the contributions from each individual term. The results are presented in Table II. Overall, the individual geometry terms performed better than the energy terms. For the geometric terms, the numbers of correctly identified targets are 14, 16 and 16 for S_{ADIS} , S_{MDIS} , S_{NR} , respectively. On the other hand, the numbers of correctly predicted targets are 9, 13 and 13 for energetic terms S_{phys} , S_{ACE} , S_{RP} , respectively. All of the geometric terms have similar performances on groups 1, 2 and 4. The major discrepancy was on the third group, where S_{NR} was the only term that performed well. As for the energy terms, S_{phys} performed well on group 4, S_{ACE} performed well on group 2 and S_{RP} performed well on group 4 and a major part of group 1.

RP score is the default and only score in 3D-DOCK. When compared with other individual terms used in this study, its performance was less desirable than any of the geometric terms. Since it missed nine out of the 22 targets, the ranking

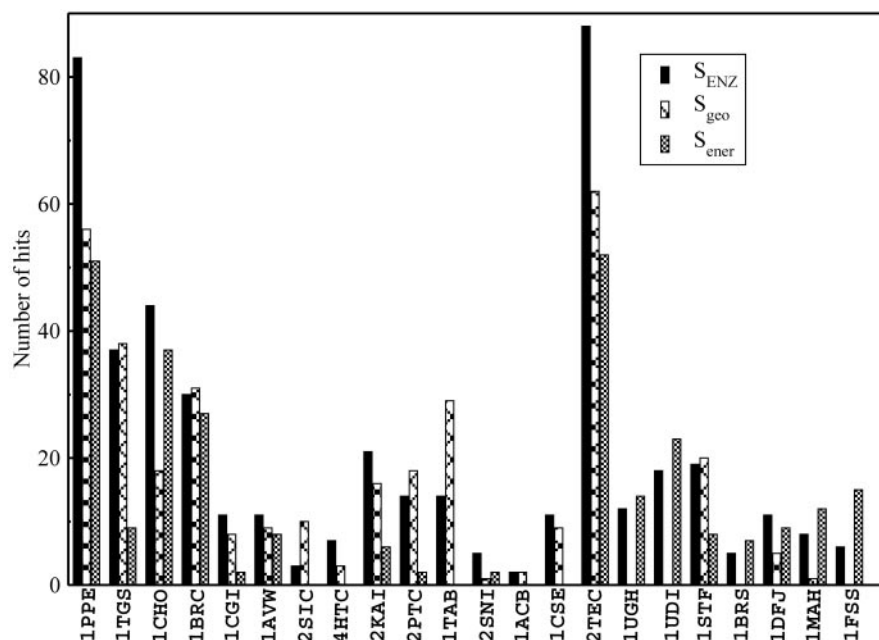


Fig. 2. Comparison of performance by geometry score, energy score and the composite score for enzyme–inhibitor docking. Targets are arranged according to the group affiliation.

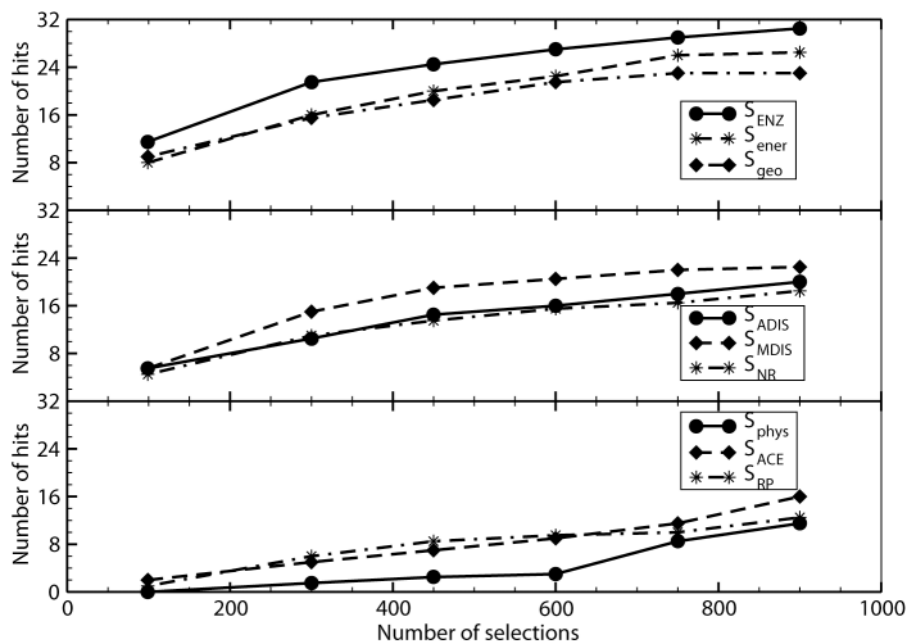


Fig. 3. Comparison of the enrichment of hits for enzyme–inhibitor docking. The median number of hits is chosen to present the enrichment. The performance of the composite scoring function S_{ENZ} was compared with that of the two component scores (S_{geo} and S_{ener}) and the individual terms (S_{ADIS} , S_{MDIS} , S_{NR} , S_{phys} , S_{ACE} and S_{RP}).

ability of RP score itself appears to be unsatisfactory for enzyme–inhibitor complexes. However, it works exceptionally well for six targets in group 1, where at least 10 hits were retained in the prediction set. The biased performance by RP score could be caused by the possible bias inherent in the training database from which the residue-pair potential was derived. It is also possible that the preferential pair interactions are more conserved in the first group.

A study of the enrichment of hits was also conducted. Among all the candidates in the filtered set, we chose 33,

100, 150, 200, 250 and 300 highest ranked candidates from each of the three runs based on the composite score, component scores or the individual scores and the number of hits was examined within the subsets of 99, 300, 450, 600, 750 and 900 selected candidates. The median number of hits within different selection sets is shown in Figure 3. For the composite score S_{ENZ} , the number of hits increased from 11.5 to 30.5 when the selection size increased from 99 to 900. The increase rate was the highest in the first segment when the selection size changed from 99 to 300, the number of hits almost doubled.

It is interesting that while the number of selected hits was significantly lower by the individual energy terms (lower panel in Figure 3) compared with the geometry terms (middle panel in Figure 3), the component score S_{ener} performed slightly better than S_{geo} when the selection size was >300 (upper panel in Figure 3).

Additional tests

In the antibody–antigen class of the same benchmark set, there are four targets with single domain antibody–formed complex with enzyme antigen, 1KXQ, 1KXT, 1KXV and 1MEL. We reassigned those targets to the enzyme–inhibitor class and conducted the test following the same procedure as described above. A major challenge is to predict the multiple binding modes of the same enzyme pancreatic α -amylase with different antibody VHH domains in the three complexes 1KXQ, 1KXT and 1KXV. As shown in Figure 4, the VHH domains bind to the major pocket, a minor groove and a flat surface of the pancreatic α -amylase in the three complexes. These are three of the seven targets in the first and second round of CAPRI experiments. The only correctly predicted binding mode was 1KXQ by any participating groups. In our test, FTDOCK generated 39, 12 and 26 hits in the candidate pool for 1KXQ, 1KXT and 1KXV, respectively. In the prediction set, the number of hits was 18, 1 and 2, respectively. The best rank of hits was 2, 25 and 22. The lowest r.m.s.d. was 1.57, 1.87 and 6.83 Å. Although the scoring function is apparently biased to the binding mode of classical enzyme–inhibitor interaction as in 1KXQ, it is encouraging that the scoring function can correctly identify all three binding modes with simple rigid-body docking. For the other target 1MEL, 31 hits were generated in the candidate pool by FTDOCK. There were eight hits within the prediction set. The lowest r.m.s.d. was 4.76 Å and the best rank was 2.

Docking the antibody–antigen targets

Fifteen antibody–antigen targets were included in this study (Table III). The antigens were lysozyme in the first seven targets. In the FTDOCK step, 60 000 candidates were saved for each target. The increased size of candidate pool compared with the enzyme–inhibitor docking was due to the significantly lower quality of candidates. The number of hits ranged from 9 to 98 with a median of 27, which was much fewer than the hits generated in the enzyme–inhibitor docking with only half of the candidates. The hits-to-candidates ratio was 1:2222 in

the candidate pool. This made it much more challenging than the enzyme–inhibitor docking. The lowest r.m.s.d. ranged from 0.82 to 6.10 Å with a median of 2.7 Å, comparable to the lowest r.m.s.d. in enzyme–inhibitor docking.

Table III. Summary of FTDOCK and filtering for antibody–antigen targets^a

Complex	FTDOCK		Filtering by CDR		Filtering by ADIS and S_{phys}			
	Hits	R.m.s.d.	Hits	Candidates	Hits	R.m.s.d.	Candidates	P_{hits}
1BVK	64	4.90	64	11873	61	4.90	7781	95.3
1JHL	98	2.70	95	15077	88	2.70	9019	89.8
1MLC	56	4.29	56	12600	53	4.29	7250	94.6
1DQJ	50	5.35	50	12279	7	5.56	3300	14.0
1BQL	33	2.25	33	11146	21	2.25	2874	63.6
1FBI	28	5.62	27	12562	25	5.62	4942	89.3
2JEL	25	2.19	24	8746	21	2.19	2680	84.0
1NMB	27	3.13	27	13661	23	3.13	6982	85.2
1NCA	23	0.82	23	12617	21	0.82	4207	91.3
1WEJ	82	1.69	81	12707	62	1.69	5185	75.6
1IAI	17	6.10	17	2652	17	6.10	1176	100
1AHW	25	2.02	25	14358	25	2.02	4974	100
1QFU	27	2.13	25	12359	25	2.13	3169	92.6
1EO8	9	2.90	9	9865	7	2.90	2475	77.8
2VIR	16	2.89	16	13806	8	2.89	2538	50.0

^aSee Table I and Methods.

Table IV. Comparison of predictions by different scores for antibody–antigen targets^a

Complex	S_{geo}	S_{ener}	S_{info}	S_{ANT}		
				Hits	r.m.s.d.	Rank
1BVK	0	1	1	4	7.49	4
1JHL	0	2	1	4	7.90	5
1MLC	4	0	0	3	7.73	6
1DQJ	0	0	0	0	12.66	>33
1BQL	2	0	0	2	3.59	6
1FBI	0	0	0	1	8.28	11
2JEL	1	0	2	9	5.18	4
1NMB	0	0	1	1	7.60	2
1NCA	14	0	2	12	1.65	1
1WEJ	6	7	4	6	6.71	1
1IAI	0	2	2	3	6.10	7
1AHW	5	17	1	8	4.28	2
1QFU	9	20	0	2	3.10	1
1EO8	0	5	0	0	17.19	>33
2VIR	0	0	0	3	3.28	17

^aSee Table I and Methods.

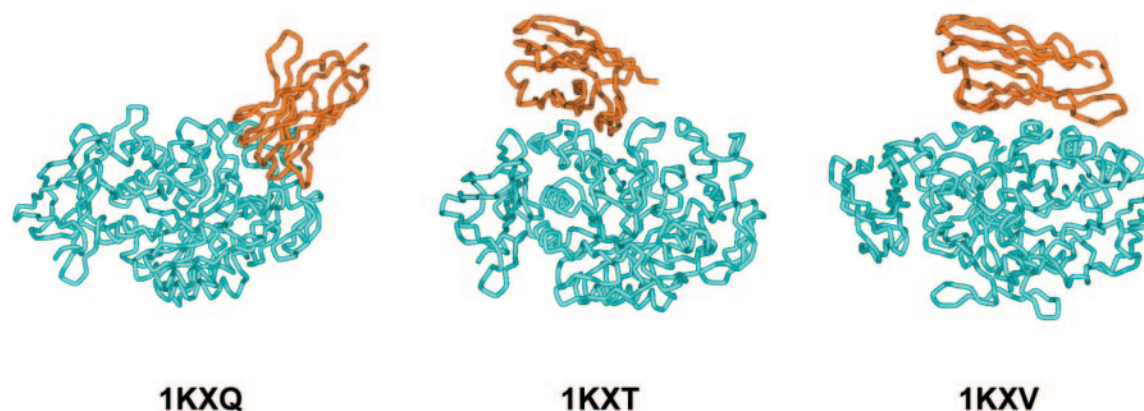


Fig. 4. The three binding modes of pancreatic α -amylase when complexed with three different VHH domains. In complex 1KXQ, the binding is in the major pocket. In complex 1KXT, the binding is in a minor groove. In complex 1KXV, the binding is on a flat surface.

The filtering process consisted of two steps. In the first step, a restriction was imposed on the antibody, which required a minimum of 60% of interface residues belonging to CDR regions. After the CDR filtering, the number of hits was almost the same for all the targets, while the size of the candidate pool shrank to 12 562 (median), 20.9% of the original size. In the second step, the intermolecular distance ADIS and the physical energy S_{phys} were both applied as filters with the same cutoff value of 50. After the filtering, the size of the candidate pool further shrank to 4207 (median), 7.0% of the original size, while 89.3% (median) of the hits were retained. The only target with over 50% of reduction of hits was 1DQJ. The lowest r.m.s.d. remained the same for all the targets except for 1DQJ, where there was a slight increase from 5.35 to 5.56 Å.

The ranking performance of the composite scoring function S_{ANT} and its component scores S_{geo} , S_{ener} and S_{info} is presented in Table IV. Although the component scores only had correct prediction on 7–8 targets, the composite score correctly predicted 13 of the 15 targets, among which three of the highest ranks are 1. It is intriguing that most of the lowest r.m.s.d. candidates were missed by S_{ANT} .

Discussion

Comparison with previous work

Chen *et al.* developed ZDOCK and tested it on the same enzyme–inhibitor set (Chen *et al.*, 2003a) examined in the current study. Several scoring schemes were tested in the work of Chen *et al.* Among them, the best performance was by a combination of pairwise shape complementarity score, desolvation energy and electrostatics. For four of the enzyme–inhibitor complexes, 2PTC, 1CSE, 2KAI and 2SNI, the first hit was ranked 193 or lower. In the worst case, only one hit was in the top 2000 for 2SNI and it was ranked 1262. The performance on the antibody–antigen targets was much less desirable. Only four targets (1AHW, 1BQL, 1NCA and 1MEL) had hits ranked among the top 100. Gray *et al.* also conducted a comprehensive test on the same complex set (Gray *et al.*, 2003). The energy terms used in their study included van der Waals, electrostatics, solvation, hydrogen bonding, residue pair probability and rotamer probability terms. In the global docking test, no hits were found among the top 25 clusters for three enzyme–inhibitor targets, 1DFJ, 1BRC and 2PTC. For the antibody–antigen targets, no hits were found among the top 25 clusters for six targets, 1DQJ, 1WEJ, 1EO8, 1IAI, 1NCA and 2VIR. In summary, there are 3–4 incorrect predictions for the enzyme–inhibitor targets and ≥ 6 incorrect predictions for the antibody–antigen targets. As a comparison, all the enzyme–inhibitor targets were correctly predicted and only two antibody–antigen targets were incorrectly predicted in the current work.

It should be noted that comparison with other studies is usually indirect. Different procedures and criteria are used by different groups. For example, the rotation step size was 6° in the ZDOCK global search and 5° in Gray *et al.*'s work (Gray *et al.*, 2003). The rotation step size was 9° in this work. When compared with 12° rotation (default in FTDOCK), candidate quality was improved at a 9° rotation step (data not shown). Unfortunately, rotation steps smaller than 9° are not allowed in FTDOCK without major modification of the program code. As a result, the candidate quality is expected to

be slightly lower in this study than that in ZDOCK and in the work of Gray *et al.* Additionally, interface r.m.s.d. was used as a criterion for hits in ZDOCK. Both the ligand r.m.s.d. and interface r.m.s.d. are used in CAPRI evaluation. Since interface r.m.s.d. is sensitive to side chain orientation on the interface and geometric features are not very sensitive to side chain orientation, we decided to use ligand C^α r.m.s.d. as criteria for hits. The highest populated clusters instead of individual candidates were used for ranking in Gray *et al.*'s work. Assuming that each cluster contains 20 members, the top 25 clusters would correspond to 500 individual candidates in the prediction set. Since a significant number of hits was found among the top 99 selections for most complexes in this work, we decided that it is not necessary to do clustering.

Future development

A considerable challenge is to account for the conformational changes during docking. Indeed, some complexes undergo large conformational changes upon binding. They are categorized into the ‘difficult’ class in the benchmark set compiled by Chen *et al.* (2003b). This class provides the greatest challenge for docking. Although this is beyond the scope of the present study, we tried to apply the scoring scheme to assess the limitation. One example is complex 1BTH, which is an enzyme–inhibitor complex in the ‘difficult’ class. Although its main structural features resemble those complexes in group one, the β -loops at the receptor interface changed from open to close state upon binding. No hits were found among the top 25 clusters in Gray *et al.*'s work (Gray *et al.*, 2003); the lowest ligand r.m.s.d. in the top 10 clusters was 16.96 Å. In our test, FTDOCK generated 21 hits in the candidate pool. The best hit had a ligand r.m.s.d. of 7.08 Å. Among the 99 selections in the prediction set, there were 12 candidates with r.m.s.d. <15 Å; the lowest r.m.s.d. was 11.84 Å. Although it did not survive the 10 Å cutoff, it is a reasonable starting point for further refinement.

Ultimately, protein–protein interactions are governed by free energy (Fernandez-Recio *et al.*, 2004). This is why most of the scoring functions for protein–protein docking employ atomic-level interaction energy terms. However, there are a few pitfalls in energy-based approaches. First, energy at the atomic level is sensitive to small perturbations. Hence incorporation of flexibility becomes necessary in the early stage of docking. This makes the global search computationally intensive. Second, entropy is an important part of free energy. Although certain solvation terms have entropy partially included in the parameterization process, it is extremely difficult to evaluate the complete entropic contribution to binding. Third, the accuracy of physical force fields is still uncertain despite decades of effort. Small systematic errors can lead to large errors exceeding binding energy when thousands of atoms are involved. These pitfalls of atomic interaction energy make it necessary to utilize information at coarse-grain level. The information does not have to be universal to all protein–protein complexes, provided that it is a general feature for certain groups of complexes. For example, restriction to CDR regions has been a common protocol for antibody–antigen docking. Interface residue conservation (Bradford and Westhead, 2003) has also been suggested as a filter for docking candidates. Intermolecular distance has been shown in this work to be of utility in selecting good candidates.

Further improvement will require introducing flexibility. This will comprise the next stage of the docking procedure.

Conclusions

We investigated the application of a new packing factor, namely intermolecular distance, to the docking of protein–protein complexes. A new selection strategy was developed on 22 enzyme–inhibitor targets and 15 antibody–antigen targets from a protein–protein docking benchmark set. A pool of docking candidates was generated by FTDOCK. This was followed by filtering with the intermolecular distance and ranking by the composite scoring function. This strategy succeeded on all of the 22 enzyme–inhibitor targets and 13 of the 15 antibody–antigen targets. Most interestingly, this selection strategy succeeded on four test targets, within which three complexes have the same receptor yet different binding modes.

Although the three-step procedure was adopted in the docking of both enzyme–inhibitor and antibody–antigen complexes, there are variations in each of the three steps. In the global sampling step by FTDOCK, 30 000 candidates were saved for the enzyme–inhibitor targets, whereas 60 000 candidates were saved for the antibody–antigen targets. The increased size for the latter case was due to the much lower quality of the candidate pool. In the filtering step, only the ADIS filter was applied for the enzyme–inhibitor targets, whereas two more filters (CDR and physical energy) were added for the antibody–antigen targets. CDR information has been routinely integrated in antibody–antigen docking, either in the sampling step or in the filtering step. The physical energy filter was added because of the good correlation between this energy and candidate quality. In the final ranking step, the composite score S_{ENZ} consists of geometry terms and energy terms, while the composite score S_{ANT} includes additional informatics terms. All these variations were due to the significantly greater difficulty of antibody–antigen docking, as been shown in many previous docking studies.

Intermolecular distance was applied in two of the three steps in this docking study. In the filtering step, it was able to reduce greatly the number of candidates without significant loss of hits. In the ranking step, the addition of the intermolecular distance was shown to be helpful in giving the hits higher ranks. The contribution of intermolecular distance to the filtering and ranking was manifested in this work for the docking of both the enzyme–inhibitor complexes and the antibody–antigen complexes. More work is under way to evaluate its application to other protein–protein complexes.

Acknowledgements

Thanks are due to Drs Michael Knaggs, Shibusish Chowdhury and Matthew Lee for helpful discussions and critical reading of the manuscript. This work was supported by research grants from NIH (GM64458 and GM67168 to Y.D.).

References

- Bahar,I., Atilgan,A.R. and Erman,B. (1997) *Fold. Des.*, **2**, 173–181.
 Ben-Zeev,E., Berchanski,A., Heifetz,A., Shapira,B. and Eisenstein,M. (2003) *Proteins*, **52**, 41–46.
 Betts,M.J. and Sternberg,M.J. (1999) *Protein Eng.*, **12**, 271–283.
 Bogan,A.A. and Thorn,K.S. (1998) *J. Mol. Biol.*, **280**, 1–9.
 Bradford,J.R. and Westhead,D.R. (2003) *Protein Sci.*, **12**, 2099–2103.
 Camacho,C.J. and Gatchell,D.W. (2003) *Proteins*, **52**, 92–97.
 Camacho,C.J. and Vajda,S. (2001) *Proc. Natl Acad. Sci. USA*, **98**, 10636–10641.
 Camacho,C.J. and Vajda,S. (2002) *Curr. Opin. Struct. Biol.*, **12**, 36–40.
 Chakrabarti,P. and Janin,J. (2002) *Proteins*, **47**, 334–343.
 Chen,R. and Weng,Z. (2002) *Proteins*, **47**, 281–294.

- Chen,R., Li,L. and Weng,Z. (2003a) *Proteins*, **52**, 80–87.
 Chen,R., Mintseris,J., Janin,J. and Weng,Z. (2003b) *Proteins*, **52**, 88–91.
 Chothia,C. and Lesk,A.M. (1987) *J. Mol. Biol.*, **196**, 901–917.
 Duan,Y. *et al.* (2003) *J. Comput. Chem.*, **24**, 1999–2012.
 Fernandez-Recio,J., Totrov,M. and Abagyan,R. (2002) *Protein Sci.*, **11**, 280–291.
 Fernandez-Recio,J., Totrov,M. and Abagyan,R. (2003) *Proteins*, **52**, 113–117.
 Fernandez-Recio,J., Totrov,M. and Abagyan,R. (2004) *J. Mol. Biol.*, **335**, 843–865.
 Gabb,H.A., Jackson,R.M., Sternberg and M.J. (1997) *J. Mol. Biol.*, **272**, 106–120.
 Glaser,F., Steinberg,D.M., Vakser,I.A. and Ben-Tal,N. (2001) *Proteins*, **43**, 89–102.
 Gray,J.J., Moughon,S., Wang,C., Schueler-Furman,O., Kuhlman,B., Rohl,C.A. and Baker,D. (2003) *J. Mol. Biol.*, **331**, 281–299.
 Halperin,I., Ma,B., Wolfson,H. and Nussinov,R. (2002) *Proteins*, **47**, 409–443.
 Heifetz,A. and Eisenstein,M. (2003) *Protein Eng.*, **16**, 179–185.
 Jackson,R.M. (1999) *Protein Sci.*, **8**, 603–613.
 Jacobs,D.J., Rader,A.J., Kuhn,L.A. and Thorpe,M.F. (2001) *Proteins*, **44**, 150–165.
 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.
 Law,D.S., Ten Eyck,L.F., Katzenelson,O., Tsigelny,I., Roberts,V.A., Pique,M.E. and Mitchell,J.C. (2003) *Proteins*, **52**, 33–40.
 Li,C.H., Ma,X.H., Chen,W.Z. and Wang,C.X. (2003a) *Protein Eng.*, **16**, 265–269.
 Li,C.H., Ma,X.H., Chen,W.Z. and Wang,C.X. (2003b) *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)*, **35**, 35–40.
 Ma,B., Wolfson,H.J. and Nussinov,R. (2001) *Curr. Opin. Struct. Biol.*, **11**, 364–369.
 Ma,B., Elkayam,T., Wolfson,H. and Nussinov,R. (2003) *Proc. Natl Acad. Sci. USA*, **100**, 5772–5777.
 Mandell,J.G., Roberts,V.A., Pique,M.E., Kotlovoyi,V., Mitchell,J.C., Nelson,E., Tsigelny,I. and Ten Eyck,L.F. (2001) *Protein Eng.*, **14**, 105–113.
 Mendez,R., Leplae,R., De Maria,L. and Wodak,S.J. (2003) *Proteins*, **52**, 51–67.
 Moont,G., Gabb,H.A. and Sternberg,M.J. (1999) *Proteins*, **35**, 364–373.
 Nooren,I.M. and Thornton,J.M. (2003a) *EMBO J.*, **22**, 3486–3492.
 Nooren,I.M. and Thornton,J.M. (2003b) *J. Mol. Biol.*, **325**, 991–1018.
 Ofran,Y. and Rost,B. (2003) *J. Mol. Biol.*, **325**, 377–387.
 Schneidman-Duhovny,D. *et al.* (2003) *Proteins*, **52**, 107–112.
 Smith,G.R. and Sternberg,M.J. (2002) *Curr. Opin. Struct. Biol.*, **12**, 28–35.
 Smith,G.R. and Sternberg,M.J. (2003) *Proteins*, **52**, 74–79.
 Zacharias,M. (2003) *Protein Sci.*, **12**, 1271–1282.
 Zhang,C., Vasmatzis,G., Cornette,J.L. and DeLisi,C. (1997) *J. Mol. Biol.*, **267**, 707–726.

Received August 15, 2004; revised October 22, 2004; accepted December 28, 2004

Edited by Bruce Tidor