Gene 422 (2008) 14–21

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Using protein binding site prediction to improve protein docking

Bingding Huang, Michael Schroeder *

Bioinformatics Group, Biotechnological Center, Technical University Dresden, Germany

ARTICLE INFO

Article history:
Available online 14 June 2008

Received by A. Bernardi

ABSTRACT

Predicting protein interaction interfaces and protein complexes are two important related problems. For interface prediction, there are a number of tools, such as PPI–Pred, PPISP, PINUP, Promate, and SPPIDER, which predict enzyme–inhibitor interfaces with success rates of 23% to 55% and other interfaces with 10% to 28% on a benchmark dataset of 62 complexes. Here, we develop, metaPPI, a meta server for interface prediction. It significantly improves prediction success rates to 70% for enzyme–inhibitor and 44% for other interfaces. As shown with Promate, predicted interfaces can be used to improve protein docking. Here, we follow this idea using the meta server instead of individual predictions. We confirm that filtering with predicted interfaces significantly improves candidate generation in rigid-body docking based on shape complementarity. Finally, we show that the initial ranking of candidate solutions in rigid-body docking can be further improved for the class of enzyme–inhibitor complexes by a geometrical scoring which rewards deep pockets. A web server of metaPPI is available at scoppi.tu-dresden.de/metappi. The source code of our docking algorithm BDOCK is also available at www.biotec.tu-dresden.de/~bhuang/bdock.

1. Introduction

1.1. Protein docking

Most cellular processes require 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 (Eisenstein and Katchalski-Katzir, 2004). The rapid accumulation of data on protein–protein interactions in sequence and/or structure 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 proceed in two stages. First, a set of candidate conformations is generated by searching over the entire binding and conformational spaces. A fast method using the fast Fourier transform (FFT) for this approach has been proposed by Katchalski-Katzir et al. (1992) and further developed by Gabb et al. (1997), Sternberg et al. (1998), Ritchie and Kemp (2000), Chen and Weng (2003) and Eisenstein and Katchalski-Katzir (2004). A good review about this method can be found in Eisenstein and Katchalski-Katzir (2004). In this stage, the candidate conformations are ranked by their shape complementarities. The near-native conformations should be contained in this set. In a second stage, 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 complementarity measures, desolvation energy, electrostatics, van der Waals interaction energy (Camacho et al., 2000; Gray et al., 2003; Mandell et al., 2001; Li et al., 2003b; Recio et al., 2002; Berchmanski et al., 2002) and some empirical potential functions such as residue–residue pair potential (Moont et al., 1999). A number of algorithms and many different scoring functions have been developed in the last ten years as recently reviewed by Halperin et al. (2002), Smith and Sternberg (2002) and Eisenstein and Katchalski-Katzir (2004). One of the major challenges in docking is the definition of a scoring function that differentiates between near-native complex structures and non-native ones. One approach aims to predict interaction interfaces.

Knowledge of the interfaces of known protein–protein complexes have revealed that enzyme–inhibitor, antibody–antigen, other complexes present important differences in the amino acid composition, residue interface propensity, hydrophobicity and electrostatics (Decanniere et al., 2001; Glaser et al., 2001; Huang and Schroder, 2005). Jackson (1999) 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 interfaces. This suggests that different filtering criteria should be applied to different type of complexes. Li et al. (2003a) applied a type-dependent filtering technique docking algorithm and retained

Abbreviations: CAPRI, a critical assessment of predicted interactions; EI, enzyme–inhibitor; FFT, fast Fourier transform; RMSD, root mean square deviation; SVM, support vector machine; ToF, tightness of fit.

* Corresponding author. Biotec, TU Dresden, Tatzberg 47-51, 01307 Dresden, Germany. Tel.: +49 351 46340060; fax: +49 351 46340061.
E-mail addresses: bhuang@biotec.tu-dresden.de (B. Huang), ms@biotec.tu-dresden.de (M. Schroeder).

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.06.014

Table 1
The different properties used in different prediction methods

Method	Sequence		Geometric		Physical–chemical		
	Conservation	Propensity	Planarity	Protrusion	Electrostatic	Hydrophobicity	Solvent accessibility
PPI–PRED	✓	✓	✓	✓	✓	✓	
PINUP	✓						✓
PPISP	✓	✓					✓
Promate	✓	✓	✓	✓			
SPPIDER							✓

more native-like structures and increased the probability of predicting complex structures.

1.2. Interfaces prediction

There have been many efforts to predict protein–protein interaction binding sites based on the analysis of the protein surface properties (Lichtarge et al., 1996; Jones and Thornton, 1997a,b; Neuvirth et al., 2004; Bradford and Westhead, 2005; Zhou and Shan, 2001; Espadaler et al., 2005; Aytuna et al., 2005). Jones and Thornton (1997a,b) analysed the surface patches using six parameters: solvation potential, residue interface propensity, hydrophobicity, planarity, protrusion and solvation accessible surface area. The six parameters were then combined into a global score that gave the probability of a surface patch forming protein–protein interaction. Bradford and Westhead (2005) 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 solvent accessible surface area. Using this method (PPI–PRED), 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. Neuvirth et al. (2004) developed a method called ProMate to distinguish interface region based on 13 properties. They got a success rate of 65% on a test data set of 57 proteins. Zhou et al. developed an algorithm called PPISP, which trained a neural network with sequence profiles and solvent accessibility of spatially neighbouring surface residues to predict interface residues (Zhou and Shan, 2001; Chen and Zhou, 2005). Liang et al. (2006) present an empirical scoring function, which is a linear combination of energy score, interface propensity and residue conservation score for prediction of protein binding sites (PINUP). SPPIDER is a neural-network prediction method that includes predicted relative solvent accessibility as input (Porollo and Meller, 2007). Table 1 summarizes the different properties used in these prediction methods.

Although many binding site prediction methods have been developed, only a few approaches integrate it into docking. Gottschalk et al. (2004) predicted protein–protein binding site first using their own prediction program ProMate (Neuvirth et al., 2004), of which the success rate was about 70%, and then they used these predicted binding site 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. Moreover, Vries et al. (2006) used both residue conservation and the interface residues predicted by ProMate in their data-driven docking programme HADDOCK (Dominguez et al., 2003). These approaches demonstrated interface prediction can improve the performance of current protein docking algorithms.

In this work we use predicted interfaces to improve protein–protein docking. First, we obtain the prediction results from the five state-of-the-art predictors: PPI–PRED (Bradford and Westhead, 2005), PPISP (Zhou and Shan, 2001; Chen and Zhou, 2005), PINUP (Liang et al., 2006), Promate (Neuvirth et al., 2004) and SPPIDER (Porollo and Meller, 2007). Then we propose a meta prediction approach to improve the prediction accuracy. In our docking algorithm, we propose a novel shape complementarity scoring function for the

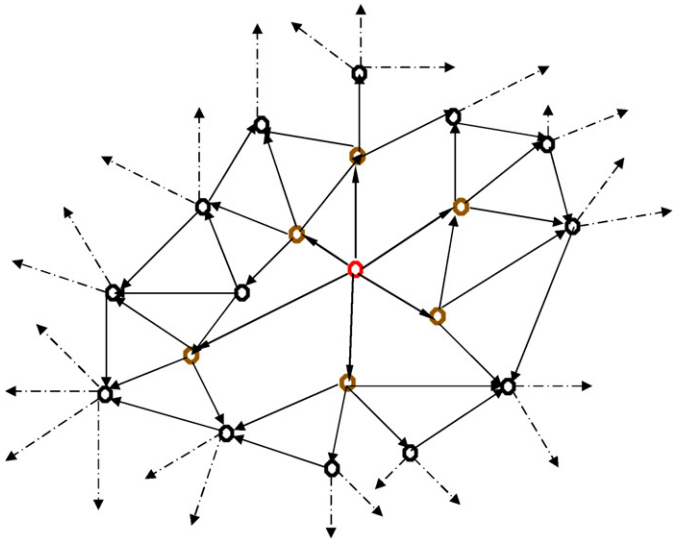

Fig. 1. Illustration of continuous surface patch generation procedure. First, the children (bottle green) of the centre vertex (red) are added into the patch. Then the children (black) of its children are added to the patch. This procedure is iterated for a certain number of times. The patches generated by this method are continuous.

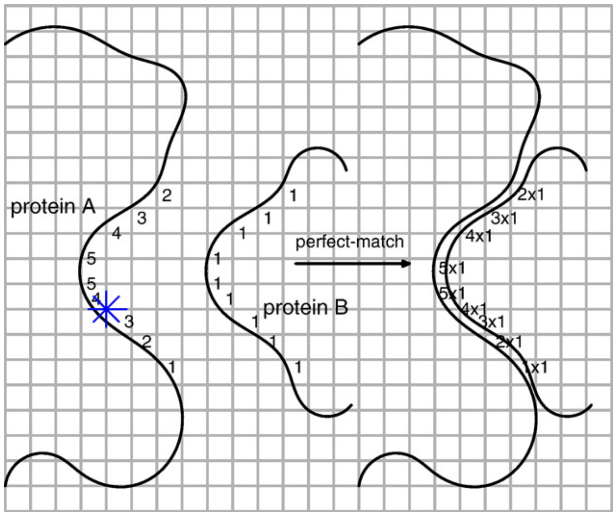

Fig. 2. 2D schematic illustration of using the degree of buriedness of surface residues in FFT docking. The dark curves represent part of the surface of protein A and B. The digits near the surface of protein A are the degrees of buriedness D_{br}, which are calculated by scanning four directions (shown in blue lines, 7 directions in 3D) to identify the number of protein–solvent–protein events. The shape complementarity score is 29 rather than 9 after shifting protein B to protein A, shown in the right side.

Table 2
The success rates of different prediction approaches on the data set (%)

Type (#)	PPI–PRED	PINUP	PPISP	Promate	SPPIDER	MetaPPI
Enzyme–inhibitor (20)	45	52	55	36	23	70
Other complex (42)	28	15	25	13	10	44
All test cases (62)	33	27	34	21	14	52

MetaPPI which integrates individual predictions improves success rate significantly. The best result for enzyme–inhibitor complexes achieved with PPISP (55%) rises to 70% with metaPPI. For other complexes, the best result of 28% (PPI–PRED) rises to 44% with metaPPI. Interestingly, the best individual predictor, PPISP, uses only a limited number of features.

initial docking stage, which takes the degree of buriedness of surface residues into account as different weights. In the filter stage, the predicted interface residues are used to calculate the tightness of fit as a scoring function to filter the docked solutions. We evaluate this approach on the unbound structures from a data set of 62 protein–protein complexes derived from Weng's docking benchmark 2.0 (Mintseris et al., 2005) and CAPRI targets (Mendez et al., 2005).

2. Materials and methods

2.1. Meta interface prediction

The unbound protein structures are submitted to the five prediction servers and prediction results are retained. We assign a confidence score to every surface residue if it is predicted as an interface residue by one prediction method. Thus, each surface residue gets a confidence score of 0 (none of the prediction methods predict) to 5 (in all prediction methods it is an interface residue). Therefore, the whole surface can be divided into different patches consisting of the residues with the same confident scores. The patch which has the highest confidence score is selected as the initial predicted binding sites. If it is one continuous patch, then the whole patch is considered as potential predicted binding site. Otherwise, the patches are clustered into different continuous sub-patches. Then the biggest sub-patch is considered as potential predicted binding site.

We propose a novel approach to generate continuous patches using a mesh representation of the protein surface. The solvent accessible surface of the protein is calculated using the Connolly algorithm (Connolly, 1983) and is represented as a mesh. Then, the vertex nearest to the mass centre of the potential predicted binding site is selected as starting point. Starting from this point, first we add its children (neighbouring) vertices to the patch. Then the children of its children are added to the patch. The procedure is iterated for a certain number of times. In the last step, the PDB atoms that are associated to the vertices in this patch are the final predicted binding site. This procedure is illustrated in Fig. 1.

Mostly, the prediction is assessed by two parameters. One is the accuracy of predicted interface, which is the fraction of correctly predicted residues in the total number of predicted interface residues; the other is the coverage of the actual interface by the predicted interface, which is the fraction of correctly predicted interface residues in the total number of native interface residues. A prediction is successful if more than half of the predicted residues are native interface residues (accuracy, 50%), as defined in Neuvirth et al. (2004) and Liang et al. (2006). A residue is a surface residue if its relative solvent accessible area is greater than 10%, calculated by NACCESS (Hubbard, 1996). The complex structures are used for defining native interface residues. A surface residue is considered as native interface residue if its accessible surface area decreased by more than 1 Å^2 upon complexation.

Before we explain how the tightness of fit is computed for predicted interfaces, we introduce FFT docking and our modification to it.

2.2. FFT docking method

In the FFT docking (fast Fourier transform) (Katchalski-Katzir et al., 1992; Gabb et al., 1997), the protein structure is projected into a three-dimensional grid of N complex numbers. The shape of the protein is depicted in the real part as described below:

$$\text{protein A}: \quad a_{p,q,r} = \begin{cases} 1 & \text{surface cell} \\ p & \text{interior cell} \\ 0 & \text{elsewhere} \end{cases}$$

$$\text{protein B}: \quad b_{p,q,r} = \begin{cases} q & \text{interior cell} \\ 0 & \text{elsewhere} \end{cases}$$

Thus, for protein A, the surface grid points have the value 1, interior grid points have the value p (usually −15), and grid points outside protein are 0. For protein B, the grid points on the surface and in the interior are set to value q, which is usually 1. These two grids can then be superimposed and the mobile grid (protein B) is translated by shifts α, β and γ.

The value of $a_{l,m,n} \times b_{l-\alpha,m-\beta,n-\lambda}$ gives the extent of shape complementarity of two proteins at grid point l,m,n. A value of 1 for the product grid indicates that the cell of protein B is superimposed on the surface of A which indicates favourable shape complementarity. A value of − 15 indicates a steric clash with the cell from B superimposed on the core of A. The value of −15 is chosen to penalise but not totally prevent steric clashes. A value of zero means that there is no overlap between the two proteins. Thus, the total shape complementarity $c_{\alpha,\beta,\gamma}$ for the two superimposed grids is calculated as:

$$c_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} a_{l,m,n} \times b_{l-\alpha,m-\beta,n-\gamma} \tag{1}$$

The value of c is a convolution and its calculation has a complexity of $O(N^6)$, which is reduced by the discrete fast Fourier transform (FFT) to $O(N^3 \log N^3)$, where N is the size of the grid.

We modify this basic approach to take the degree of buriedness of surface residues into account. For protein A, the grid points of the surface layer are scored with the degree of buriedness (D_{br}):

$$a_{p,q,r} = \begin{cases} 1 + D_{\text{br}} & \text{surface cell} \\ \text{p} & \text{interior cell} \\ 0 & \text{elsewhere} \end{cases}$$

The degree of buriedness is calculated by scanning the x, y, z directions and four cubic diagonals to identify the number of protein–solvent–protein events (PSP) (Huang and Schroeder, 2006). The grid points near deep buried surface residues will have protein–solvent–

Table 3
The average accuracy and coverage of different prediction approaches (%)

Type	PPI–PRED		PINUP		PPISP		Promate		SPPIDER		MetaPPI	
	Accu	Covr	Accu	Covr	Accu	Covr	Accu	Covr	Accu	Covr	Accu	Covr
Enzyme–inhibitor	41	42	46	43	46	39	40	19	38	45	60	35
Other complexes	26	29	24	22	27	24	15	8	26	41	38	21
All test cases	31	33	31	29	33	29	24	12	29	42	45	25

MetaPPI improves the best average accuracy from 46% to 60% for enzyme–inhibitor complexes. For other complexes, the best average accuracy rises from 27% to 38% with metaPPI. However, metaPPI does not improve the average coverage.

Table 4
The prediction results of different prediction approaches on the data set

PDB	Type[a]	Receptor[b]							Ligand[b]						
		Inf_n[c]	C_{pp}/P_{pp}[d]	C_{pi}/P_{pi}	C_{ps}/P_{ps}	C_{pr}/P_{pr}[e]	C_{sp}/P_{sp}[f]	C_{mt}/P_{mt}	Inf_n	C_{pp}/P_{pp}	C_{pi}/P_{pi}	C_{ps}/P_{ps}	C_{pr}/P_{pr}	C_{sp}/P_{sp}	C_{mt}/P_{mt}
1ACB	EI	25	15/27	10/23	14/20	11/25	14/45	14/16	17	1/12	11/20	7/20	2/8	17/49	10/15
1AY7	EI	17	2/18	8/16	6/22	4/10	4/15	3/11	15	0/10	11/20	9/11	5/9	10/21	8/11
1BVN	EI	33	24/47	17/36	9/15	2/15	13/54	10/17	24	6/7	16/23	14/21	7/8	23/50	10/11
1CGI	EI	29	19/33	11/23	16/20	12/25	15/46	16/16	19	7/11	13/17	13/22	4/6	12/12	7/7
1D6R	EI	21	17/36	4/15	1/20	8/23	4/23	6/11	15	9/15	2/19	7/30	3/6	15/52	8/18
1DFJ	EI	34	10/20	10/14	5/14	4/13	0/0	9/14	42	14/43	21/32	14/15	3/5	1/1	13/16
1E6E	EI	23	3/75	9/39	2/21	0/0	8/43	0/21	28	8/9	13/17	17/20	8/11	19/23	14/15
1EAW	EI	25	18/23	8/21	14/20	7/24	3/29	12/12	21	13/15	19/25	16/20	5/6	15/18	9/9
1EWY	EI	19	11/38	2/31	3/9	0/0	8/45	1/11	19	5/13	7/19	13/21	4/10	9/18	8/14
1F34	EI	32	12/39	7/31	1/27	1/14	13/53	4/13	33	13/20	11/15	14/20	4/4	22/39	8/11
1MAH	EI	29	23/77	12/44	13/20	0/13	8/30	1/6	20	3/16	6/15	2/20	0/7	16/31	3/19
1PPE	EI	21	16/25	8/17	10/20	10/23	2/9	9/14	16	8/8	16/20	12/16	3/3	16/29	12/13
1TMQ	EI	30	20/36	16/32	12/20	3/11	12/28	12/16	25	0/12	17/22	17/27	6/12	20/55	14/15
1UDI	EI	27	21/27	13/24	10/20	1/5	7/15	7/11	23	8/16	13/18	16/20	5/9	21/59	13/14
2PCC	EI	14	11/58	9/28	0/14	0/8	0/5	7/13	17	9/15	3/12	4/22	2/11	0/5	6/14
2SIC	EI	21	14/29	8/16	18/20	7/12	8/30	8/11	14	0/25	14/19	0/21	9/11	11/44	0/17
2SNI	EI	18	11/31	7/13	16/22	6/20	6/23	6/10	14	9/14	12/21	13/20	6/7	14/40	8/14
7CEI	EI	17	8/14	8/16	9/13	2/2	2/6	6/8	16	6/14	0/22	4/11	0/5	0/0	6/15
T16	EI	23	0/48	16/24	7/13	3/28	12/25	5/9	31	21/34	7/20	4/9	2/6	1/1	9/9
T17	EI	20	0/23	0/21	0/18	0/26	2/28	0/9	20	9/27	10/15	4/20	1/15	7/16	13/14
T18	EI	24	12/26	8/22	3/20	4/18	6/17	12/14	29	1/47	0/15	0/24	0/26	2/42	0/8
1AK4	Ot	16	14/24	6/22	8/20	1/17	1/5	5/10	12	0/31	12/17	0/18	0/5	12/90	0/11
1ATN	Ot	16	0/58	0/20	0/22	0/4	16/191	0/12	16	0/31	0/17	2/6	0/10	0/3	0/10
1B6C	Ot	18	10/15	9/20	7/23	4/11	7/16	10/12	24	6/67	6/24	3/20	0/0	13/40	6/11
1BUH	Ot	17	0/30	0/23	0/20	5/10	5/43	0/13	18	12/18	7/15	6/22	5/8	17/60	7/12
1E96	Ot	20	3/15	8/17	7/20	0/2	15/46	5/13	14	0/28	8/20	1/5	0/0	9/23	5/12
1FQ1	Ot	14	0/24	4/19	1/11	0/2	3/27	0/11	13	1/74	4/19	0/21	0/10	7/59	0/14
1FQJ	Ot	28	21/37	13/25	8/15	0/2	13/36	12/17	21	14/27	12/39	7/13	2/12	4/15	10/11
1GCQ	Ot	11	0/29	9/13	1/21	2/22	6/18	7/10	15	9/13	3/19	8/11	0/7	12/33	6/10
1GHQ	Ot	8	0/38	0/22	0/19	0/0	1/16	0/15	9	2/21	0/17	5/20	1/6	4/38	0/10
1GRN	Ot	20	5/25	3/15	6/22	1/5	10/55	3/9	22	0/20	5/20	11/20	0/0	13/33	1/6
1H1V	Ot	46	14/58	6/18	3/14	2/4	38/188	6/13	41	0/99	0/31	0/13	0/0	0/20	0/5
1HE1	Ot	26	14/23	1/27	9/20	0/0	8/10	8/12	28	12/18	10/16	14/20	0/2	23/49	13/17
1HE8	Ot	21	0/210	0/48	0/20	0/2	7/67	0/11	18	7/30	5/17	9/22	0/0	14/27	3/7
1I2M	Ot	28	11/33	7/15	5/20	0/0	10/47	6/14	39	23/60	10/20	0/7	0/3	0/11	0/16
1IBR	Ot	50	17/28	7/17	15/20	0/0	22/52	11/15	56	16/68	2/36	2/27	3/10	11/65	1/14
1KAC	Ot	19	0/25	0/15	0/20	0/7	2/32	0/10	21	10/14	8/16	0/9	4/8	1/8	6/10
1KTZ	Ot	8	2/31	4/18	7/22	2/12	8/70	5/11	10	3/19	8/19	8/20	2/11	8/23	8/12
1KXP	Ot	43	26/46	0/20	7/21	2/4	40/189	13/16	46	27/63	0/31	0/16	2/5	1/30	0/14
1KXQ	Ot	36	27/66	1/27	11/16	0/15	22/51	11/12	28	2/10	6/19	7/20	0/3	25/85	5/12
1M10	Ot	29	14/21	6/24	3/7	0/4	2/5	7/12	33	0/28	0/31	0/20	0/9	8/19	0/11
1QA9	Ot	17	0/19	0/22	0/25	0/3	0/9	0/12	19	0/28	0/17	0/14	0/2	0/6	1/15
1SBB	Ot	17	0/34	1/20	0/18	0/3	0/28	0/10	16	0/31	0/22	3/4	0/4	1/2	0/12
1WQ1	Ot	31	17/35	0/18	2/5	0/0	11/27	10/15	34	14/22	14/20	14/20	0/0	18/21	9/9
2BTF	Ot	25	19/48	0/19	8/21	1/4	22/188	8/15	25	12/17	12/31	3/22	1/2	11/34	0/13
T04	Ot	28	1/42	0/30	0/11	0/11	1/32	0/14	22	15/21	6/21	7/22	3/12	21/90	9/14
T05	Ot	22	2/42	2/30	0/11	0/11	0/32	0/14	23	7/18	7/19	2/26	3/12	22/99	9/10
T06	Ot	34	25/42	14/30	9/11	1/11	18/32	10/14	26	1/22	6/20	5/20	1/12	23/90	4/14
T07	Ot	13	0/24	5/22	0/2	0/14	0/3	4/12	18	0/40	1/18	0/13	0/9	0/18	1/13
T08	Ot	30	0/42	0/32	0/20	0/0	1/21	0/17	26	3/38	4/20	4/16	2/17	5/19	3/10
T09	Ot	40	8/31	6/19	15/20	8/23	24/64	8/10	40	5/31	6/18	15/20	10/23	25/63	5/11
T11	Ot	22	11/24	6/19	13/22	7/14	5/6	13/15	18	13/23	1/20	15/24	0/8	15/40	9/11
T12	Ot	24	10/22	6/20	14/22	7/14	6/7	15/16	15	2/9	6/21	10/20	2/6	15/43	5/9
T14	Ot	43	3/35	2/31	3/21	0/22	5/17	1/10	38	1/51	7/18	7/29	0/6	29/84	5/10
T15	Ot	9	1/20	2/26	3/18	2/11	0/1	0/14	9	1/11	6/21	3/20	2/9	7/37	3/19
T20	Ot	45	10/35	12/21	18/33	0/0	21/52	17/18	36	23/99	17/19	10/20	0/0	14/75	13/18
T21	Ot	15	1/33	4/21	0/22	2/12	0/24	0/10	11	0/22	0/19	0/21	4/12	8/60	0/13
T22	Ot	15	0/20	0/18	6/29	0/14	9/27	1/11	12	9/15	5/27	9/20	1/7	10/24	5/6
T23	Ot	53	33/50	12/21	0/15	5/31	1/6	9/14	53	33/50	12/21	0/15	7/31	1/6	9/14
T24	Ot	15	11/22	8/15	12/20	4/11	11/24	10/15	13	8/16	0/17	1/21	0/10	0/2	0/12
T25	Ot	22	18/22	10/15	15/20	8/11	18/24	15/15	22	20/36	4/16	0/22	0/13	1/6	7/12
T26	Ot	33	0/49	9/31	12/20	0/0	4/13	9/12	24	2/11	4/19	4/11	0/11	1/2	3/14
T27	Ot	9	2/30	0/20	0/23	0/2	0/8	0/10	9	0/32	3/18	8/20	0/7	3/17	2/10

[a] Type of complexes, EI: enzyme–inhibitor complexes, Ot: other complexes.
[b] Receptor: the bigger protein in complex; Ligand: the smaller one.
[c] Inf_n: the number of residues on real interface in complex.
[d] C_{pp}: the number of residues predicted correctly to be on interface from PPI_PRED; P_{pp}: the number of total residues predicted to be on interface from PPI_PRED; pi: PINUP; ps: PPISP; pr: Promate; sp: SPPIDER; mt: metaPPI.
[e] Promate fails in some proteins, i.e., $P_{pr}=0$. It is specifically designed for enzyme–inhibitor complexes.
[f] For some proteins, SPPIDER (version 2) predicts very big interface which is unrealistic.

protein events ranging from 3 (deep buried) to 7 (very deep buried). The values of degree of buriedness is the number of protein–solvent–protein event minus 2, ranging from 1 to 5.

$$D_{br} = \begin{cases} N_{PSP} - 2 & \text{If } N_{PSP} \geq 3 \\ 0 & \text{else} \end{cases}$$

Fig. 2 illustrates how the degree of buriedness is calculated and how it improves the shape complementarity score between two protein surfaces.

This approach is integrated into a system called BDOCK, implemented using the BALL library (Kohlbacher and Lenhof, 2000). To see how the degree of buried influences the docking results, we also evaluate the BDOCK implementation that does not add D_{br} to the surface layer grid points. We call it BDOCK^{nb} (BDOCK without buriedness).

2.3. Tightness of fit

In the filtering step of docking, a good scoring function is needed to filter out a large number of false positives. Here we use the tightness of fit (ToF) proposed by Gottschalk et al. (2004) and make a small modification, which takes both sides of predicted interface residues into account. ToF is calculated as follows:

$$\text{ToF} = \frac{d_{\text{inter}}}{d^{\text{r}}_{\text{all}} + d^{\text{l}}_{\text{all}}} \tag{2}$$

where

$$d_{\text{inter}} = \frac{1}{n \times m} \sum_{i=1}^{n} \sum_{j=1}^{m} D_{\text{inter};ij} \tag{3}$$

$$d^{\text{r}}_{\text{all}} = \frac{1}{n \times q} \sum_{i=1}^{n} \sum_{j=1}^{q} D_{\text{all};ij} \tag{4}$$

$$d^{\text{l}}_{\text{all}} = \frac{1}{m \times p} \sum_{i=1}^{m} \sum_{j=1}^{p} D_{\text{all};ij} \tag{5}$$

$D_{\text{inter};ij}$ is the distance of the C_α atom of predicted interface residue i of the receptor to the C_α atom of predicted interface residue j of the ligand. $D_{\text{all};ij}$ is the distance of C_α atom of predicted interface residue i of the receptor to the C_α atom of surface residues j of the ligand. For the receptor, there are n predicted interface residues and p surface residues (m and q for the ligand, respectively).

For each solution generated in the initial stage, we compute the Z-score for the tightness of fit based on predicted interface residues:

$$Z\text{-score}_i = \frac{X_i - \bar{X}}{\sigma} \tag{6}$$

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

It is obvious 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 filtering, the improvement factor IF is calculated according to

$$\text{IF} = \frac{\text{hits}_{\text{af}}/N_{\text{af}}}{\text{hits}_{\text{bf}}/N_{\text{bf}}} \tag{7}$$

where N_{bf}, N_{af} are the number of the complex structures; hits_{bf} and hits_{af} are the number of near-native structures (L_RMSD ≤ 10 Å) in the pool, before and after filtering.

2.4. Test dataset

We use 41 complexes from the Weng docking benchmark data set (Mintseris et al., 2005), where there is only one chain in each component since the five prediction server can only predict interface for proteins with one chain. We also include 21 recent single-chain CAPRI targets (Janin et al., 2003; Mendez et al., 2003). Totally we have 62 complexes, which consist of 20 enzyme–inhibitor complexes and 42 non-enzyme–inhibitor (other types of complex). The unbound structures are used for interface prediction and docking (some CAPRI targets are bound or homolog modelling structures). The complex structures (bound) are used for validation of the prediction and docking results. To compare these docking solutions to the native complex structure, we calculate the RMSD of the ligand (the smaller protein) in the predicted versus native complexes after the receptors (the bigger protein) is superimposed. This ligand RMSD, denoted L_RMSD, is computed on C_α atoms. We define a solution as near-native structure (hit) if the L_RMSD is below 10 Å, which corresponds to the "acceptable prediction" defined in CAPRI experiments (Mendez et al., 2005). Then we compare the number of hits within the top 10000 solutions for BDOCK^{nb} and BDOCK. The RMSD value (bRMS) of the best hit (with lowest L_RMSD) are also compared.

3. Results and discussion

3.1. Prediction results

The prediction success rates for each method are summarised in Table 2. Our meta approach achieves 70% success rate for the enzyme–inhibitor complexes, rising from the best result of 55% which are achieved with PPISP. For other complexes, the best success rate 28% (PPI–PRED) rises to 44% with metaPPI. Totally, our approach achieves 52% success rate for all the test cases. Overall, our approach significantly improves the success rate than individual prediction method. We also show the average accuracy and average coverage of each prediction method in Table 3. Our approach has higher accuracy (66% for EI, 43% for non EI and 51% for all) than the other methods. However, our method does not have higher coverage than the other methods. Since the goal of the prediction in this work is to help docking, we care more about the accuracy than coverage. The detailed prediction results for each method for each complex are also shown in Table 4.

Fig. 3 shows the overlapping prediction results of the five methods for all the test cases. All the methods predict 5 interfaces successfully (accuracy ≥ 50%). Four of them are successful for 4 interfaces (11 for three, 23 for two and 25 for one of them, respectively). For 56 interfaces, none of these five methods gives correct prediction.

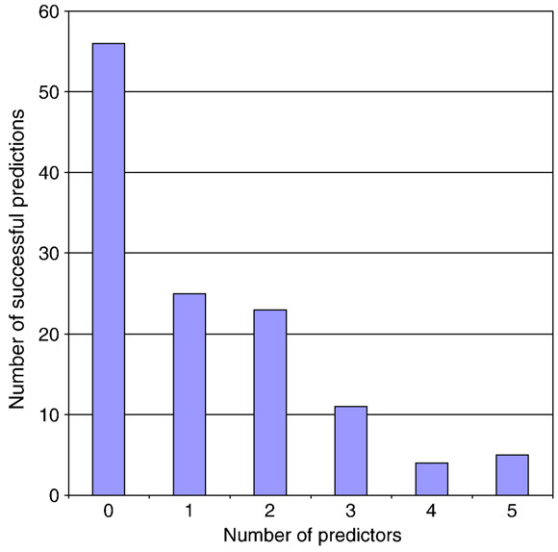

Fig. 3. The overlapping performance of the five prediction methods. All the methods predict 5 interfaces successfully (accuracy ≥50%). Four of them are successful for 4 interfaces (11 for three, 23 for two and 25 for one of them, respectively). For 56 interfaces, none of the five methods gives correct predictions.

Table 5
The docking results by ToF for BDOCK[nb] and BDOCK[a]

Complex	BDOCK[nb]					BDOCK					Prediction accuracy(%)[f]	
	$hits_{bf}$[b]	$hits_{af}$/Naf[c]	$bRMS_{bf}$[d]	$bRMS_{af}$[d]	IF[e]	$hits_{bf}$	$hits_{af}/N_{af}$	$bRMS_{bf}$	$bRMS_{af}$	IF	Receptor	Ligand
EI												
1ACB	13	13/1682	5.1	5.1	5.9	80	80/1721	2.5	2.5	5.8	87	66
1AY7	12	10/1820	2.9	2.9	4.6	7	7/1648	3.0	3.0	6.1	27	72
1CGI	67	60/1636	4.5	4.5	5.5	132	127/1631	3.1	3.1	5.9	100	100
1D6R	3	1/2038	5.0	8.3	1.6	166	0/1585	2.9	13.8	–	54	44
1DFJ	5	5/1621	4.4	4.4	6.2	51	51/1444	3.7	3.7	6.9	64	81
1E6E	38	36/1498	4.5	4.5	6.3	85	71/1840	3.9	3.9	4.5	0	93
1EAW	44	40/1731	5.3	5.3	5.3	15	14/1859	7.1	7.1	5.0	100	100
1EWY	126	58/1793	2.3	2.3	2.6	518	402/2137	4.0	4.0	3.6	9	57
1F34	8	8/1761	6.6	6.6	5.7	1	1/2069	4.5	4.5	4.8	30	72
1MAH	7	6/1710	4.1	4.1	5.0	76	73/1543	2.4	2.4	6.2	16	15
1PPE	164	123/1987	1.5	1.5	3.8	601	576/1868	0.6	0.6	5.1	64	92
1TMQ	14	14/1864	3.4	3.4	5.4	0	0/1982	34.3	34.3	–	75	93
1UDI	45	24/1821	4.7	6.5	2.9	120	45/1799	3.9	5.5	2.1	63	92
2PCC	28	26/1683	7.3	7.3	5.5	5	5/966	8.6	8.6	10.4	53	42
2SIC	25	0/1693	3.5	16.4	–	329	0/1977	1.2	19.0	–	72	0
2SNI	17	13/1795	8.7	8.7	4.3	4	4/1801	9.2	9.2	5.6	60	57
7CEI	25	25/1654	4.7	4.7	6.0	2	2/1456	8.3	8.3	6.9	75	40
T16	6	6/1645	7.0	7.0	6.1	45	0/1973	2.2	12.8	–	55	100
T17	0	0/1642	10.6	23.0	–	13	0/1570	7.4	32.3	–	0	92
T18	1	0/1533	8.2	41.6	–	16	0/2118	4.0	47.2	–	85	0
Other												
1AK4	0	0/1636	12.1	12.3	−1	0	0/1690	12.3	12.3	–	50	0
ATN	0	0/1706	12.3	41.4	–	0	0/2017	25.1	25.1	–	0	0
1B6C	26	26/1819	3.3	3.3	5.5	0	0/1699	13.2	13.2	–	83	54
1BUH	7	0/1594	6.4	12.5	–	0	0/1405	15.9	24.8	–	0	58
1E96	11	9/1714	4.4	4.4	4.8	9	9/1524	6.1	6.1	6.6	38	41
1FQ1	1	0/1520	7.4	17.5	–	8	0/1725	7.3	19.3	–	0	0
1FQJ	4	4/1741	5.7	5.7	5.7	0	0/1847	10.3	10.3	–	70	90
1GCQ	0	0/1753	10.6	10.6	–	2	2/1888	9.0	9.0	5.3	70	60
1GHQ	0	0/1570	10.2	12.1	–	0	0/1629	10.0	12.8	–	0	0
1GRN	16	16/1653	3.2	3.2	6.0	10	10/1749	6.5	6.5	5.7	33	16
1H1V	0	0/1713	13.1	28.0	–	0	0/1688	14.6	24.3	46	0	
1HE1	9	9/1562	3.6	3.6	6.4	1	1/1634	6.5	6.5	6.1	66	76
1HE8	5	0/1503	6.9	21.6	–	0	0/1881	13.7	70.7	–	0	42
1I2M	10	0/1735	7.1	17.3	–	58	14/1789	4.4	6.6	1.3	42	0
1IBR	7	0/1818	8.0	19.8	–	9	0/1640	8.1	20.4	–	73	7
1KAC	8	0/1589	6.2	21.9	–	0	0/1719	15.1	29.7	–	0	60
1KTZ	1	1/1964	4.6	4.6	5.1	0	0/1936	17.6	17.6	–	45	66
1KXP	7	0/1649	1.8	35.8	–	60	0/1379	1.8	26.7	–	81	0
1KXQ	33	30/1784	0.7	0.7	5.1	27	5/1775	1.3	8.5	1.0	91	41
1M10	1	1/1776	9.5	9.5	5.6	0	0/1478	10.5	10.5	–	58	0
1QA9	2	0/1836	7.1	42.7	–	0	0/1655	23.4	38.9	–	0	6
1SBB	2	0/1691	6.2	12.6	–	0	0/1584	20.4	22.5	–	0	0
1WQ1	2	2/1736	7.5	7.5	5.8	0	0/1581	11.1	11.1	–	66	100
2BTF	16	0/1618	5.5	13.0	–	0	0/1860	15.0	15.0	–	53	0
T04	17	7/1737	2.6	7.9	2.4	31	26/1636	6.1	6.1	5.1	0	64
T05	15	4/1799	0.9	7.9	1.5	0	0/1803	11.1	11.1	–	0	90
T06	35	1/1704	1.2	8.8	0.2	1	1/1665	9.5	9.5	6.0	71	28
T07	4	4/1788	2.5	2.5	5.6	0	0/1855	27.6	27.6	–	33	7
T08	0	0/1735	10.9	16.8	–	0	0/2319	10.2	32.8	–	0	30
T09	0	0/1534	14.7	14.7	–	0	0/1641	14.7	14.7	–	80	45
T11	19	19/1694	3.3	3.3	5.9	2	2/1629	6.8	6.8	6.1	86	81
T12	64	7/1723	1.5	7.6	0.6	1	0/2102	5.9	12.4	–	93	55
T14	4	4/1858	2.9	2.9	5.4	2	0/2308	4.5	25.8	–	10	50
T15	75	7/1855	4.3	6.0	0.5	0	0/2217	11.6	11.6	–	0	15
T20	0	0/1811	25.7	25.7	–	0	0/1863	25.3	25.3	–	94	72
T21	13	0/1825	4.3	21.1	–	1	0/1980	8.5	25.6	–	0	0
T22	11	3/1679	3.9	3.9	1.6	0	0/1724	13.5	14.8	–	9	83
T23	1	1/1571	9.7	9.7	6.4	0	0/1901	26.6	28.5	–	64	64
T24	0	0/1720	16.6	16.7	–	0	0/1709	17.0	17.0	–	66	0
T25	18	18/1752	1.9	1.9	5.7	0	0/1768	11.8	11.8	–	100	58
T26	62	0/1808	2.6	17.8	–	0	0/1678	10.1	18.0	–	75	21
T27	0	0/1717	38.7	38.7	–	0	0/1711	35.0	35.0	–	0	20

[a] Only the top 10,000 solutions are kept for filtered by ToF (threshold: ≤−1.0).
[b] $hits_{bf}$: the number of hits in the top 10,000 solutions before filtering.
[c] N_{af} is the number of docking results remaining in the pool after filtering. $hits_{af}$ is the number of hits in this pool.
[d] $bRMS_{bf}$ and $bRMS_{af}$ are the L_RMSD value (Å) of the best hit before/after filtering.
[e] Improvement factor (IF) after filtering, calculated according to Eq. (7).
[f] The accuracy of interface prediction by metaPPI; Receptor: the bigger protein; Ligand: the smaller one.

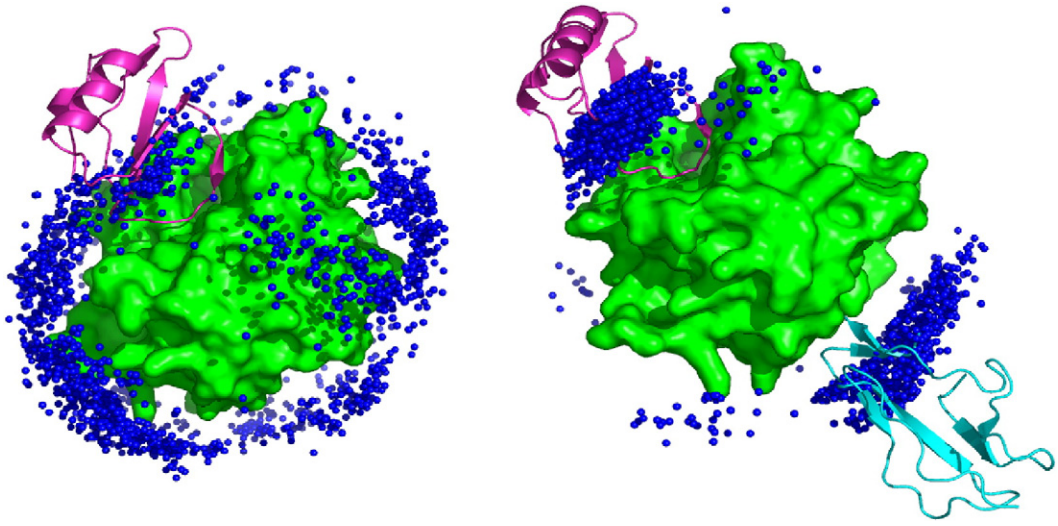

Fig. 4. Comparison of the performance of BDOCK^{nb} (left) and BDOCK (right) on chymotrypsin–inhibitor complex (pdbcode:1acb). The mass center of the inhibitor of the top 2000 docked solutions are shown in blue spheres. Green: enzyme, magenta: native inhibitor. Scoring surface layer grids equally, BDOCK^{nb} sample the solutions over the whole surface. In BDOCK the solutions are mainly clustered around two pocket sites which correspond to two binding sites. The top left ligand in magenta is the Eglin C inhibitor in 1acb and the bottom right ligand in cyan is the complement C1S protease domain in 1elv. The catalytic domain of 1elv is superimposed with the chymotrypsin domain of 1acb (RMSD: 1.8 Å). Both of them are in the same SCOP (Murzin et al., 1995) family of eukaryotic proteases.

3.2. Docking results

In the FFT-based docking algorithms, the shape complementarity is the first criterion to discriminate the non-native and near-native complex structures. The goal of the initial stage of docking is to generate as many near-native complex structures (hits) as possible. The docking results of BDOCK and BDOCK^{nb} are shown in Table 5 for all 62 complexes. For each approach, the number of hits, the RMSD value of the best hit (with the lowest L_RMSD) based on shape complementarity are listed.

For 20 enzyme–inhibitor complexes, for 13 cases BDOCK produces significantly more hits and better bRMSD[1] (L_RMSD of the best hit) values than BDOCK^{nb}. These results show that adding the degree of buriedness to the surface layer grids significantly improves the docking performance. This is the case since the binding sites of enzyme–inhibitor complexes usually involve a very big cleft. The "lock-and-key" principle plays an important role in this type of complex.

For 42 non-enzyme–inhibitor complexes, both of the approaches produce less hits than those on enzyme–inhibitor complexes. This is because the interface of non enzyme–inhibitor complexes do usually not involve a deep pocket. Therefore, comparing to BDOCK^{nb}, BDOCK does not produce more hits, as it does on enzyme–inhibitor complexes. In contrast, BDOCK^{nb} performs better than BDOCK in most of cases.

Fig. 4 shows the spatial distributions of the top 2000 docked solutions of BDOCK and BDOCK^{nb} on the complex of 1ACB. For good visualisation, we only take the top 2000 solutions instead of 10000. In BDOCK^{nb}, the surface layer grids being scored equally, the solutions are distributed around the whole receptor surface. However the solutions of BDOCK mainly accumulate around two big pocket sites, by adding more scores to deeply buried surface residues. These clustered docked solutions suggest two different binding sites of two partners, one is the Eglin C inhibitor (chain I, 1ACB) and another is the Complement C1S protease domain in 1ELV. This observation demonstrates the potential ability of our approach to identify multiple binding sites of proteins.

[1] Note, that RMSD after filtering can be worse, if the best solution has been filtered out.

3.3. Filtering results by metaPPI prediction

Table 5 also shows the filtering results of docking solutions by the tightness of fit (ToF) for BDOCK^{nb} and BDOCK, using the interface prediction results from metaPPI. For the complexes for which both sides of interfaces are predicted successfully (accuracy ≥ 50%) and hits are found in the solution pool, filtering with ToF with a threshold ≤ −1.0 rejects a large part of the false positives without rejecting correct solutions. For 20 enzyme–inhibitor complexes, metaPPI predicts both sides of interface correctly for 9 cases, for which the ToF works very well both for BDOCK^{nb} and BDOCK. For 10 cases, metaPPI only predicts one side interface correctly and fails in the other side (accuracy < 50%). Among these cases, ToF still works for 7 and 5 cases, for BDOCK^{nb} and BDOCK respectively. However, for 42 non enzyme–inhibitor complexes, the docking and prediction results are worse. In most of the cases, BDOCK^{nb} generates more hits than BDOCK but there are some difficult cases for which no hits are found in both approaches. MetaPPI predicts both side interfaces successfully for only 10 cases and one side successfully for 16 cases. For the rest of the cases, metaPPI fails in both sides of the interfaces and hence ToF can not be applied. Moreover, for the complexes for which the predictions are correct but no hits are found in the top 10,000 solutions, ToF does not work either.

3.4. Summary

To summarise, for interface prediction, our meta method improves the prediction success rates of individual prediction approaches. The tightness of fit scoring function based on these correctly predicted interface residues effectively discriminates between near-native complex structures and non-native ones. This approach is implemented in BDOCK^{nb} and is applicable to all types of complexes. Adding further background for special classes of complexes, such as enzyme–inhibitor complexes, these results can further be improved. Our implementation of BDOCK does so for enzyme–inhibitor complexes by integrating a gemometric scoring for pockets into the FFT scoring.

4. Conclusion

Rigid-body docking aims to predict complexes from unbound component structures. In the first stage, a pool of candidates is generated

based on shape complementarity and in the second stage, candidates are filtered out. Using predicted interfaces for filtering has been shown a promising strategy by Neuvirth et al. (2004). Here we build on this work and make three contributions.

First, we develop a meta server, MetaPPI, for interface prediction, which integrates five servers. We develop a novel algorithm which generates the best continuous patch out of the given predictions. MetaPPI significantly improves success rates by 15% on the best individual predictions indicating that all of the integrated tools have their own strengths and that the sum of predictions is more than its parts.

Second, following Neuvirth et al. (2004), we integrate the predicted interfaces into the second stage of docking, the filtering of candidates, in our BDOCK system. We confirm the suitability of this approach on a comprehensive benchmark of 62 complexes including enzyme–inhibitor and other complexes.

Third, in our BDOCK system, we show that a further improvement for the class of enzyme–inhibitor complexes, which are characterized by deep pockets, can be achieved by modifying the first stage of docking incorporating into the shape complementarity scores factors rewarding pockets.

Finally, the developed code and systems are freely available for academic use. BDOCK is implemented using the BALL library (Kohlbacher and Lenhof, 2000) in C++, taking an object-oriented and generic programming approach, which is very easy to extend and adapt new docking algorithms. There is an option to switch BDOCK to BDOCK^{nb}, which is better for docking non enzyme–inhibitor complexes. Generally docking a complex of moderate size using BDOCK takes about 1 to 2 h at 10° rotational angle step on current Linux workstation. The source code of BDOCK is freely available for academic users from www.biotec.tu-dresden.de/~bhuang/bdock. The metaPPI web server is available at scoppi.biotec.tu-dresden.de/metappi.

Acknowledgements

We are very grateful to H. Zhu and A. Hildebrandt, for their help about the BALL library and Frank Dressel for comments. We thanks the authors of the five predictors for making their prediction public. We also appreciate computational support from High Performance Comuter Center of Dresden Technical University. Funding by EFRE project CODI, FoldUnfold, and Sealife is kindly acknowledged. Thanks to the ZIH for providing access to the supercomputer.

References

Aytuna, A.S., Gursoy, A., Keskin, O., 2005. Prediction of protein–protein interactions by combining structure and sequence conservation in protein interfaces. Bioinformatics 21 (12), 2850–2855.
Berchmanski, A., Katchalski Katzir, E., Eisenstein, M., 2002. Electrostatics in protein–protein docking. Protein Sci. 11, 571–587.
Bradford, J., Westhead, D., 2005. Improved prediction of protein–protein binding sites using a support vector machines approach. Bioinformatics 21 (8), 1487–1494.
Camacho, C., Gatchell, D., Kimura, S., Vajda, S., 2000. Scoring docked conformations generated by rigid-body protein–protein docking. Proteins 40 (3), 525–537.
Chen, H., Zhou, H.X., 2005. Prediction of interface residues in protein–protein complexes by a consensus neural network method: test against NMR data. Proteins 61 (1), 21–35.
Chen, R., Weng, Z., 2003. A novel shape complementarity scoring function for protein–protein docking. Proteins 51, 397–408.
Connolly, M., 1983. Analytical molecular surface calculation. J. Appl. Cryst. 16, 548–558.
Decanniere, K., Transue, T., Desmyter, A., Maes, D., Muyldermans, S., Wyns, L., 2001. Degenerate interfaces in antigen–antibody complexes. J. Mol. Biol. 313, 473–478.
Dominguez, C., Boelens, R., Bonvin, A., 2003. HADDOCK: a protein–protein docking approach based on biochemical and/or biophysical information. J. Am. Chem. Soc. 125, 1731–1737.
Eisenstein, M., Katchalski-Katzir, E., 2004. On proteins, grids, correlations, and docking. C. R., Biol. 327, 409–420.
Espadaler, J., Romero-Isart, O., Jackson, R.M., Oliva, B., 2005. Prediction of protein–protein interactions using distant conservation of sequence patterns and structure relationships. Bioinformatics 21 (16), 3360–3368.
Gabb, H., Jackson, R., Sternberg, M., 1997. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J. Mol. Biol. 272 (1), 106–120.
Glaser, F., Steinberg, D., Vakser, I., Ben-Tal, N., 2001. Residue frequencies and pairing preferences at protein–protein interfaces. Proteins 43, 82–102.
Gottschalk, K., Neuvirth, H., Schreiber, G., 2004. A novel method for scoring of docked protein complexes using predicted protein–protein binding sites. Protein Eng. 17, 183–189.
Gray, J., et al., 2003. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J. Mol. Biol. 331, 281–299.
Halperin, I., Ma, B., Wolfson, H., Nussinov, R., 2002. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47, 409–443.
Huang, B., Schroder, M., 2005. Using residue propensities and tightness of fit to improve rigid-body protein–protein docking. In: Rarey, M., Torda, A., Kurtz, S., Willhoeft, U. (Eds.), Proceedings of German Bioinformatics Conference, pp. 159–173.
Huang, B., Schroeder, M., 2006. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. BMC Struct. Biol. 6, 19.
Hubbard, S., 1996. Naccess http://wolf.bms.umist.ac.uk/naccess.
Jackson, R., 1999. Comparison of protein–protein interactions in serine protease–inhibitor and antibody–antigen complexes: implications for the protein docking problem. Protein Sci. 8, 603–613.
Janin, J., et al., 2003. Capri: a critical assessment of predicted interactions. Proteins 52, 2–9.
Jones, S., Thornton, J., 1997a. Analysis of protein–protein interaction sites using surface patches. J. Mol. Biol. 272, 121–132.
Jones, S., Thornton, J., 1997b. Prediction of protein–protein interaction sites using patches analysis. J. Mol. Biol. 272, 133–143.
Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A., Aflalo, C., Vakser, I., 1992. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. PNAS 89, 2195–3199.
Kohlbacher, O., Lenhof, H., 2000. BALL — rapid software prototyping in computational molecular biology. Bioinformatics 16 (9), 815–824.
Liang, S., Zhang, C., Liu, S., Zhou, Y., 2006. Protein binding site prediction using an empirical scoring function. Nucl. Acids Res. 34 (13), 3698–3707.
Lichtarge, O., Bourne, H.R., Cohen, F.E., 1996. An evolutionary trace method defines binding surfaces common to protein families. J. Mol. Biol. 257 (2), 342–358.
Li, C., Ma, X., Chen, W., Wan, C., 2003a. A protein–protein docking algorithm dependent on the type of complexes. Protein Eng. 16, 265–269.
Li, L., Chen, R., Zhiping, W., 2003b. RDOCK: refinement of rigid-body protein docking predictions. Proteins 53 (3), 693–707.
Mandell, J., et al., 2001. Protein docking using continuum electrostatics and geometric fit. Protein Eng. 14, 103–113.
Mendez, R., R., L., Maria, L., Wodak, S., 2003. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins 52, 51–67.
Mendez, R., Leplae, R., Lensink, M.F., Wodak, S.J., 2005. Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. Proteins 60 (2), 150–169.
Mintseris, J., et al., 2005. Protein–protein docking benchmark 2.0: an update. Proteins 60 (2), 214–216.
Moont, G., Gabb, H., Sternberg, M., 1999. Use of pair potentials across protein interfaces in screening predicted docked complexes. Proteins 35 (3), 364–373.
Murzin, A., Brenner, S., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540.
Neuvirth, H., Raz, R., Schreiber, G., 2004. Promate: A structure based prediction program to identify the location of protein–protein binding sites. J. Mol. Biol. 338, 181–199.
Porollo, A., Meller, J., 2007. Prediction-based finger-prints of protein–protein interactions. Proteins 66 (3), 630–645.
Recio, J., Totrov, M., Abagyan, R., 2002. Soft protein–protein docking in internal coordinates. Protein Sci. 11, 280–291.
Ritchie, D., Kemp, J., 2000. Protein docking using spherical polar Fourier correlations. Proteins 39 (2), 178–194.
Smith, G., Sternberg, M., 2002. Prediction of protein–protein interactions by docking methods. Curr. Opin. Struct. Biol. 12, 28–35.
Sternberg, M., Gabb, H., Jackson, R., 1998. Predictive docking of protein–protein and protein–DNA complexes. Curr. Opin. Struct Biol. 8 (2), 265–269.
Vries, S.D., van Dirj, A., Bonvin, A., 2006. WHISCY: What information does surface conservation yield? Application to data-driven docking. Proteins 63 (3), 479–489.
Zhou, H., Shan, Y., 2001. Prediction of protein interaction sites from sequence profile and residue neighbor list. Proteins 44 (3), 336–343.