

RESIDUE CONSERVATION INFORMATION FOR GENERATING NEAR-NATIVE STRUCTURES IN PROTEIN-PROTEIN DOCKING

Yuhua Duan¹, Boojala V. B. Reddy² and Yiannis N. Kaznessis^{1,2*}

¹Department of Chemical Engineering and Materials Science, and
²Digital Technology Center, University of Minnesota, Minneapolis, MN 55455

*Corresponding author

Protein-protein docking algorithms typically generate large numbers of possible complex structures with only a few of them resembling the native structure. Recently (Duan, et al, Protein Science, (2005) 14, 316-218), it was observed that the surface density of conserved residue positions is high at the interface regions of interacting protein surfaces, except for antibody-antigen complexes, where a lower number of conserved positions than average is observed at the interface regions. Using this observation we identified putative interacting regions on the surface of interacting partners and significantly improved docking results by assigning top ranks to near-native complex structures. In this paper, we combine the residue conservation information with a widely-used shape complementarity algorithm to generate candidate complex structures with a higher percentage of near-native structures (hits). What is new in this work is that the conservation information is used early in the generation stage and not only in the ranking stage of the docking algorithm. This results in a significantly larger number of generated hits and an improved predictive ability in identifying the native structure of protein-protein complexes.

We report on results from 21 well-characterized protein complexes. We compute conservation indices (CI) of residue positions on the surfaces of interacting proteins using available homologous sequences from UNIPROT and calculate the solvent accessible surface area (SASA). We combine this information with shape-complementarity scores to generate candidate protein-protein complex structures. When compared with pure shape-complementarity algorithms, performed by FTDock, our method results in significantly more hits, with the improvement being over 100% in many instances. We demonstrate that residue conservation information is useful not only in refinement and scoring of docking solutions, but also helpful in enrichment of near-native-structures during the generation of candidate geometries of complex structures.

Keywords: protein complexes; docking; molecular recognition; protein-protein interaction; sequence conservation; protein evolution.

1. Introduction

Predicting the structure of protein-protein complexes using computational methods has progressed substantially¹⁻⁷. From a practical point of view, finding an appropriate protein docking solution is generally addressed in two stages: (1) A search over the entire binding and conformation spaces in order to generate candidate complex geometries, and (2) Development of a suitable scoring function to distinguish near native structures from false solutions generated in the first stage. To address the first stage, numerous docking algorithms have been developed based on shape complementarity search algorithms,⁸ such as PUZZLE,⁹ DOCK,¹⁰ FTDock,¹¹ DOT,¹² and ZDOCK,¹³ with new techniques constantly introduced into docking procedures, such as expansion of the molecular surface and electric field in spherical harmonics,¹⁴ and genetic algorithms.^{15,16}

Our recent analysis of well-resolved protein complexes indicated that the surface density of highly conserved residues is appreciably higher in protein-protein interface positions compared to the other positions of protein surfaces.¹⁷ On the other hand, for antibody-antigen complexes, a very low number of conserved positions were observed in the interface regions.

In reference 18 we described our docking analysis and ranking of docked complex structures using conservation information for 59 benchmark complexes¹⁹ Specifically, we initially used FTDock^{11,20} to generate 10,000 docked models for each of the complexes. At a second stage, we used conserved residue position information as a filter to reduce the number of docked structures. In addition to filtering, we also used conservation

information to rank the remaining docked structures, demonstrating that conservation can be used to increase the accuracy of near-native complex structure predictions.

Obviously, if more near native structures (hits) are generated in the first stage, the docking solution at stage two will potentially have improved selectivity. Therefore, in this paper, we combine residue conservation information with shape-complementarity criteria in the first stage of docked structures generation and investigate whether more hits are generated compared to the traditional shape-complementarity algorithms.^{11,20} We then follow the ranking procedure described in ref. 18 to get the final rank for a set of 21 complexes.

2. Materials and Methods

2.1 Conservation of Residue Positions

The calculation of conservation of residue positions is described in detail in references 17 and 18. Here we only give a brief summary. The sequences of two proteins of the complex are used to obtain their homologous sequences from the annotated, non-redundant protein sequence database UNIPROT using the FASTA3 (<http://www.ebi.ac.uk/fasta3/>) sequence similarity search tool. Homologous sequences with less than 30% gaps in the sequence and greater than 35% sequence identity to the parent sequence were used for analysis. Based on amino acid substitution matrix $M(a,b)$ values introduced by Gonnet and coworkers,²¹ a similarity score S_{ii} for sequence i is calculated by summing the identical substitution values (diagonal values from $M(a,b)$). Similarly, score S_{jj} is calculated for sequence j . A similarity score S_{ij} between the sequences i and j is calculated using substitution matrix values of corresponding aligned residues between the two sequences. Using the amino acid substitution matrix of ref. 22, conservation index (CI_l) is a weighted sum of all pairwise similarities between all residues present at the position l . The CI_l value is calculated using the following equation in a given alignment and takes a value in the range [0,1].

$$CI_l = \frac{\sum_i^N \sum_{j>i}^N ED(s_i) \times ED(s_j) \times Mut(s_i(l), s_j(l))}{\sum_i^N \sum_{j>i}^N ED(s_i) \times ED(s_j)} \quad (1)$$

where N is the number of homologous sequences in the alignment; $s_i(l)$ and $s_j(l)$ are the amino acids at the alignment position l of sequences s_i and s_j respectively; $ED(s_i)$ and $ED(s_j)$ are the average evolutionary distances of $s(i)$ and $s(j)$ from the remaining homologues and defined as:

$$ED_{ij} = \left[\left(\left(1 - \frac{S_{ij}}{S_{ii}} \right) + \left(1 - \frac{S_{ij}}{S_{jj}} \right) \right) / 2 \right] \times 100 \quad (2)$$

$Mut(a,b)$ measures the similarity among the amino acids a and b as derived from amino acid substitution matrix $M(a,b)$ and defined as:

$$Mut(a,b) = \frac{M(a,b) - M(a,b)_{low}}{M(a,b)_{max} - M(a,b)_{low}} \quad (3)$$

where a, b are the pairs of amino acids at a given alignment position l . $M(a,b)_{low}$ is the lowest value in the substitution matrix (-5 in the Gonnet matrix in ref. 21) and $M(a,b)_{max}$ is the maximum value among all the possible substitution pairs in that position. Thus $Mut(a,b)$ takes values in the range [0,1].

Using the software package PSA,^{23,24} the solvent accessible surface area (SASA) of amino acids is calculated for the two structures and used to identify surface residues and

buried residues. We then identify the top 8% (defined as group 1) and 17% (defined as group 2) of highly conserved residues, which have solvent accessibility greater than 25% of their total surface area. As an example, in Table 1 we list the highly conserved surface residues of the molecular structure of the bovine α -chymotrypsin – eglin-c complex (Protein Data Bank code: 1ACB) and specifically of the E and I chains.

Table 1. The top 17% highly conserved positions for 1ACB. Solvent accessible area and the conservation index (CI in eq. 1) are given in parenthesis respectively. The residues shown in bold fall in the top 8% list.

1ACB:E	18N(72.7,0.81)	19G(31.5,0.98)	21E(52.4,0.45)	48N(35.2,0.51)
	49E(43.0,0.45)	57H(34.7,0.99)	61T(51.4,0.41)	87K(61.7,0.43)
	88V(27.4,0.46)	95N(47.9,0.48)	107K(28.1,0.55)	109S(53.3,0.43)
	110T(63.4,0.46)	111A(46.6,0.59)	113S(65.9,0.41)	114F(35.8,0.54)
	115S(57.5,0.51)	116Q(61.8,0.44)	123L(26.9,0.92)	134T(33.7,0.40)
	144T(31.9,0.40)	153D(53.5,0.42)	164S(34.6,0.46)	166T(70.8,0.58)
	169K(52.9,0.46)	170K(72.7,0.49)	173G(34.0,0.43)	178D(84.1,0.46)
	186S(49.6,0.50)	187G(49.1,0.76)	192M(54.4,0.54)	204N(101.0,0.45)
	205G(69.1,0.61)	216G(42.9,0.88)	218S(82.0,0.49)	219T(55.1,0.43)
	236N(57.1,0.47)	239Q(67.7,0.47)	240Q(67.7,0.47)	
1ACB:I	11P(55.9,1.00)	12E(55.3,1.00)	14V(38.1,0.89)	15G(63.4,1.00)
	28H(71.6,0.80)	40G(106.3,0.85)	44T(69.4,0.99)	46D(40.3,0.95)
	48R(55.4,0.82)	50N(53.6,0.80)		

For each docked model, we also calculate the difference between the values of SASA for each residue of the complex and the free protein molecules. If this difference is greater than 10%, we consider this residue an interacting interface residue. If this position is conserved we count on the number to get the number of conserved positions.

For each model complex, we add all conservation indices (CI) for each conserved position to get two values for group 1 and group 2 respectively. Then we take the average of these two values and use this average value as one of descriptors in our global ranking scheme. Specifically, when we apply the conserved residue information to identify possible model structures, we assign higher ranks to the models with high numbers of conserved position for non-antibody complexes and to the models with low numbers of conserved positions for antibody-antigen complexes.

2.2 Combination of Shape-Complementarity with Conservation Residue Position

The calculation of shape complementarity between any two proteins A and B initially projects the two molecules onto a 3-D grid of N^3 points, represented by discrete functions:

$$A(l,m,n) \text{ or } B(l,m,n) = \begin{cases} 1 & \text{inside the molecule} \\ 0 & \text{outside the molecule} \end{cases} \quad (4)$$

Then the surface and the interior of each molecule are distinguished by empirical parameters ρ and δ respectively:

$$A(l,m,n) = \begin{cases} 1 & \text{on the surface of the molecule} \\ \rho & \text{inside the molecule} \\ 0 & \text{outside the molecule} \end{cases}$$

$$B(l,m,n) = \begin{cases} 1 & \text{on the surface of the molecule} \\ \delta & \text{inside the molecule} \\ 0 & \text{outside the molecule} \end{cases} \quad (5)$$

The correlation function (score) is calculated as:

$$C(a,b,r) = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N A(l,m,n) * B(l+a,m+b,n+r) \quad (6)$$

where

$$C(a,b,r) \begin{cases} = 0 & \text{no contact between 2 molecules} \\ > 0 & \text{contact between the surfaces} \\ < 0 & \text{penetration forbidden} \end{cases}$$

with (a,b,r) the shift vector of molecule B around molecule A. We used $\rho=1$, $\delta=-15$ for the empirically chosen parameters to calculate the correlation function $C(a,b,r)$. Using a discrete fast-Fourier transform (FFT) the computation is order $N^3 \ln(N^3)$ instead of order of N^6 of the direct calculation using eq. (6). Using this scoring function, we ranked all of the possible generated complex structures.

In order to employ position-specific conservation information in the generation stage we follow the following procedure. For each shift vector (a,b,r), according to equation (6), we only retain the 10 structures with the highest correlation scores. This resulted in about one million possible models. Then we apply residue conservation information as a filter to retain a subset of models. In order to compare with the shape-complementarity based algorithm FTDock, the number of models in the retained subset is fixed around 10,000, which is the number of models generated by FTDock.

2.3 Global Normalized Ranking

We developed a normalized ranking scheme using multiple descriptors/criteria including the conservation index. For each individual descriptor, the rank was obtained by finding the maximum (V_{\max}) and minimum (V_{\min}) of their values and using the following equation:

$$NORM_RANK_i = 1 + ANINT\left(\frac{V_{\max} - V_i}{\frac{V_{\max} - V_{\min}}{N}}\right) \quad (7)$$

where V_i is the property value of candidate complex i , and N is the total number of complexes after filtering. There may be some gaps if the difference between complexes is large, and several complexes can have the same rank number if their values are very close to one another. For antibody-antigen complexes, we use $(V_i - V_{\min})$ to substitute $(V_{\max} - V_i)$ term in equation (7).

The global score is obtained by a weighted average of all normalized ranks:

$$GLOBAL_Score = \frac{1.0}{100 * M} \sum_i^M \sigma_i * NORM_RANK_i \quad (8)$$

where M is the number of rank methods (descriptors), σ_i is the weight for descriptor i . Here we only take $M=4$ which are the desolvation energy, conservation indices, pair-potential and shape-complementarity score.

3. Results

3.1 Hits of Shape-Complementarity performed by FTDock

In order to compare the conservation-based algorithm with traditional shape-complementarity, we performed FTDock for the 21 chosen complexes and obtained 10,000 docked models and their ranks according to the correlation function of shape complementarity and pair potential. For these 10,000 docked models, we calculated the root mean square deviation (RMSD) of $C\alpha$ atoms of each model structure from the native

complex structure. We then defined “hits” as the number of models having RMSD less than 4.5 Å from the native structure (shown in Table 2). Also shown in Table 2 are the lowest RMSD (LRMSD) complex obtained with FTDock and its corresponding shape-complementarity rank and pair potential rank.

Table 2. The results of FTDock and our algorithms for 21 complexes. The number of complexes with RMSD less than 4.5Å (Hits), shape-complementarity ranks (SC_rank), pair potential ranks (PP_rank) and the lowest RMSD (LRMSD) model in the 10,000 possible docked models for each complex are shown. After combining residue conservation information with shape-complementarity, the number of complexes (NC), the number of new hits, the improvement factor by combining residue conservation information with shape-complementarity (I_fact) over FTDock, the sorting global score calculated from equation (8) for the LRMSD (G_rank), the number of hits within the first 100 ranks (E_hits), and the rank for the first hit (1st_rank) are also presented.

Complex	FTDock (10,000 models)				Present Conservation-based Algorithm						
	hits	LRMSD,Å	SC_rank	PP_rank	NC	Hits	I_fact	LRMSD,Å	G_rank	E_hit	1 st _rank
1ACB	29	0.31	7497	2654	11404	68	2.06	0.16	83	21	1
1ATN	11	0.40	2960	249	8655	31	3.26	0.40	11	24	2
1AVZ	25	2.64	7580	2611	8487	46	2.17	2.33	6830	0	3024
1BRC	21	1.29	1022	12	18169	37	0.97	1.29	5	9	5
1CGI	89	1.58	3143	78	10332	104	1.13	1.58	31	50	1
1CHO	60	1.25	631	83	8511	183	3.58	1.18	11	26	1
1CSE	56	0.92	7556	9248	8467	247	5.21	0.48	23	7	2
1EFU	0	5.71	5786	550	5034	1	∞	3.48	41	1	41
1FIN	0	5.94	9597	7502	13221	2	∞	4.25	3568	0	1149
1JHL	19	0.74	9474	482	7496	87	6.11	0.74	776	1	81
1KXQ	9	0.46	2717	344	25404	35	1.53	0.65	1	4	1
1PPE	660	0.56	966	5	13853	1508	1.65	1.16	49	77	1
1SPB	40	0.95	183	163	12737	52	1.02	0.77	3	6	1
1STF	15	0.63	60	2822	10628	61	3.83	0.44	11	24	1
1TAB	80	0.72	1359	3723	12758	261	2.56	0.72	211	2	46
1TGS	25	1.64	8598	293	16855	53	1.26	1.54	5	18	1
2KAI	76	1.51	1070	1032	14832	112	0.99	1.90	1017	10	18
2PTC	59	1.42	277	1067	14560	81	0.94	1.46	568	9	5
2SIC	16	1.86	162	4	13404	20	0.93	1.28	3	19	1
2TEC	54	0.45	305	9612	12754	334	4.85	0.40	13	34	1
2VIR	5	0.80	3403	3484	6703	8	2.39	0.61	153	4	1

^a indicates that FTDock did not generate any near-native structures

3.2 Hits from Combination of Conservation Residue Information with Shape-complementarity

As described in the Method sections, we combined residue conservation information with shape-complementarity. We applied our algorithms to the 21 complexes which are taken from a non-redundant protein complexes benchmark first used by Chen and co-worker,¹⁹ and kept the total generated complex structures to be around 10,000 possible models as in FTDock. This set of possible models is not the same as the set generated from FTDock as described in above section. We also listed the number of hits, number of possible models and the LRMSD in Table 2 for comparison.

In order to clarify the improvements of conservation-based algorithms over pure shape-complementarity algorithms, we define the improvement factor (I_fact) as $I_fact = (\text{hits/models})_{\text{conserve}} / (\text{hits/models})_{\text{FTDock}}$, where hits/models is the ratio of the number of structures with RMSD < 4.5 Å from the native structure over the number of complex models generated.

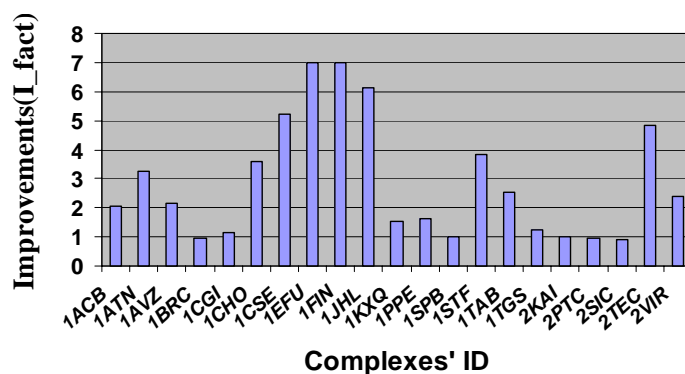


Figure 1. The improvement after employing residue conservation information. $I_fact > 1$ indicates a positive improvement.

The results are detailed in Table 2 and depicted Figure 1. It is observed that there are 17 out of 21 complexes with I_fact greater than 1.0. Most of them (12) are greater than 2.0, with the improvement over 100%. For a few complexes (1CSE, 1JHL, 2TEC) applying the filter resulted in more than 400% improvement. For 1EFU and 1FIN, no hits were obtained with FTDock.

There are only 4 out of 21 complexes (1BRC, 2KAI, 2PTC, 2SIC) with $0.9 < I_fact < 1.0$ with FTDOCK and the conservation-based algorithm generating similar numbers of hits per docking model.

3.3 Global-ranking

As described in the Methods section, we derived a global ranking function by re-normalizing the rank of each employed descriptor (eq. 7), and used weights 5, 1, 1, and 2 for the descriptors of desolvation energy, conservation indices, shape-complementarity, and pair-potential energy respectively in a new global ranking function (eq. 8). Using this function we obtained a new global rank for each model complex. Six representative example plots of the global rank versus the RMSD are shown in Figure 2.

The rank of the LRMSD structure for each complex is also listed in Table 2 (G_rank). From Figure 2 and the value of G_rank , we can see that in most of the model complexes the near native complexes have lower ranks. Comparing the G_rank with the PP_rank in Table 2, there are only 2 cases (1AVZ, 1JHL) for which the G_rank is higher than the pair-potential rank. For 19 complexes the new global ranking fairs better than the pair-potential rank.

In Table 2, we also present the number of hits (E_hits) within the first 100 ranks. For only two complexes (1AVZ, 1FIN), did the application of the global rank result in no hits in the top 100 ranked structures. For 19 complexes, E_hits values are relatively high.

Figure 3 shows model structures of the best predictions superimposed on the native structures for some of the selected targets with rank less than 10.

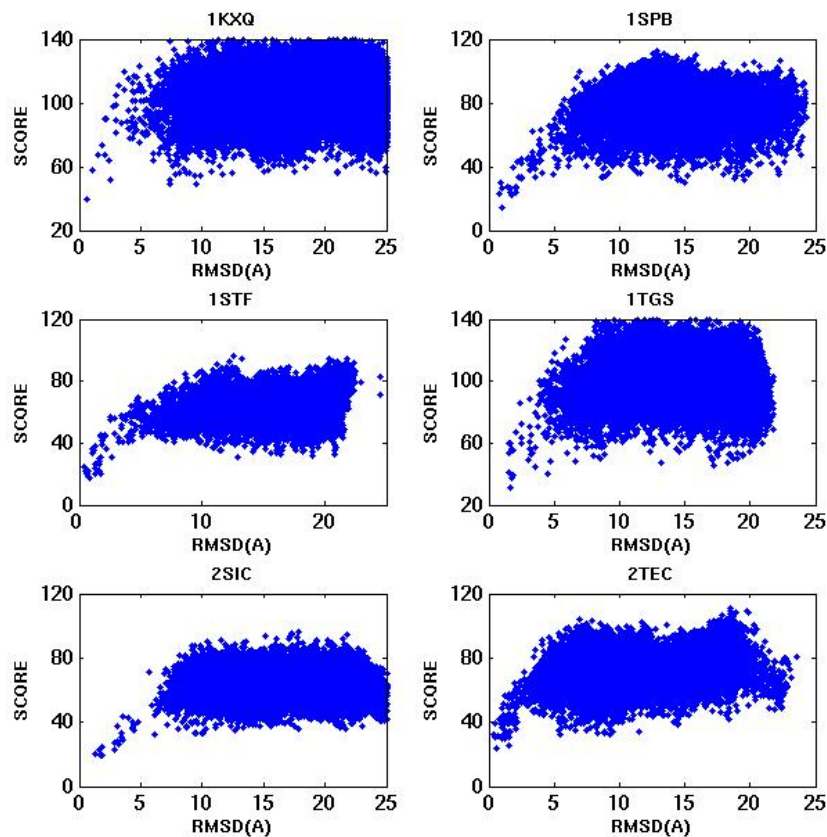


Figure 2. RMSD versus rank score (eq. 8) for six protein-protein complexes. Many of the near-native structures with low RMSD have top ranks.

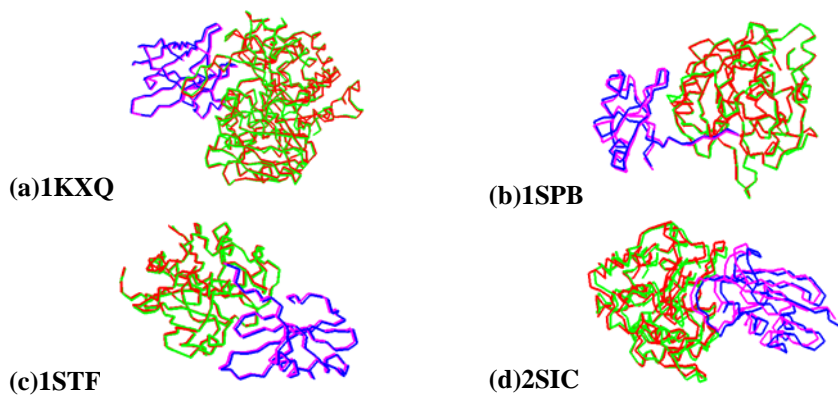


Figure 3. Representative example structures from the global predictions with the conservation-based algorithm. Red and blue indicate the experimental co-crystal. Green and purple indicate the best predicted complex structure determined by eq. (8).

4. Discussion and Concluding Remarks

There have been several reports investigating the role of conservation of interfacial residues in naturally occurring protein complexes, using evolutionary tracing of conserved residues in homologous sequences and structures.²⁵⁻³² Our analysis of well-

resolved protein complexes indicated that the density of highly conserved residues is considerably higher in protein-protein interface positions compared to the other positions of the protein surfaces,¹⁷ and we successfully used residue conservation information to filter and rank the docking solutions of protein complexes.

The generation of docking solutions is typically based on shape-complementarity arguments, retaining the protein-protein complex structures with relatively high complementarity scores. Of those only few are actually near the native protein-protein structure. This seriously hampers the identification of the actual protein-protein structure even by the most accurate filtering and ranking methods.

In this paper we demonstrated that conservation information can be employed in the generation stage of docking algorithms so that the number of near-native structures is increased. After preparing the conservation information in the form of a conservation index of each position and using a solvent accessible surface area measure, we can identify the conserved residue positions on the surface. The improvement is significant: 17 out of 21 protein-protein complexes have more hits in the docking solutions compared with the shape-complementarity algorithm FTDock. Furthermore, although for 1EFU and 1FIN pure shape-complementarity arguments fail to generate any hits, the described combination algorithms do generate 1 and 2 hits respectively.

In conclusion, residue conservation information is not only useful in refining docking solutions, but also helpful in generating more hits. Although we have used the publicly available FTDock here, generally our approach can be easily adapted to any other docking algorithms to generate more possible near-native structures.

5. Acknowledgements

This work was supported in part the National Science Foundation (BES-0425882 and EEC-0234112). This work is also supported by the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory, contract number DAAD10-01-2-0014. The content does not necessarily reflect the position or the policy of the government and no official endorsement should be inferred. The National Computational Science Alliance (TG-MCA04N033) and the Minnesota Supercomputing Institute provided access to computing facilities.

6. References

1. R. Méndez, R. Leplae, M.F. Lensink, and S.J. Wodak, *Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures*, Proteins, Structure, Function, and Bioinformatics, 2005, 60(2): p. 150-169.
2. J. Janin, *Protein-protein recognition*. Prog Biophys Mol Biol, 1995. 64(2-3): p. 145-66.
3. B.K. Shoichet and I.D. Kuntz, *Predicting the structure of protein complexes: a step in the right direction*. Chem Biol, 1996. 3(3): p. 151-6.
4. M.J. Sternberg, H.A. Gabb, and R.M. Jackson, *Predictive docking of protein-protein and protein-DNA complexes*. Curr Opin Struct Biol, 1998. 8(2): p. 250-6.
5. I. Halperin et al., *Principles of docking: An overview of search algorithms and a guide to scoring functions*. Proteins, 2002. 47(4): p. 409-43.
6. G.R. Smith and M.J. Sternberg, *Prediction of protein-protein interactions by docking methods*. Curr Opin Struct Biol, 2002. 12(1): p. 28-35.
7. C.J. Camacho and S. Vajda, *Protein-protein association kinetics and protein docking*. Curr Opin Struct Biol, 2002. 12(1): p. 36-40.
8. E. Katchalski-Katzir et al., *Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques*. Proc Natl Acad Sci U S A, 1992. 89(6): p. 2195-9.

9. M. Helmer-Citterich and A. Tramontano, *PUZZLE: a new method for automated protein docking based on surface shape complementarity*. J Mol Biol, 1994. **235**(3): p. 1021-31.
10. T.J. Ewing et al., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases*. J Comput Aided Mol Des, 2001. **15**(5): p. 411-28.
11. H.A. Gabb, R.M. Jackson, and M.J. Sternberg, *Modelling protein docking using shape complementarity, electrostatics and biochemical information*. J Mol Biol, 1997. **272**(1): p. 106-20.
12. J.G. Mandell et al., *Protein docking using continuum electrostatics and geometric fit*. Protein Eng, 2001. **14**(2): p. 105-13.
13. R. Chen, L. Li, and Z. Weng, *ZDOCK: an initial-stage protein-docking algorithm*. Proteins, 2003. **52**(1): p. 80-7.
14. D.W. Ritchie and G.J. Kemp, *Protein docking using spherical polar Fourier correlations*. Proteins, 2000. **39**(2): p. 178-94.
15. E.J. Gardiner, P. Willett, and P.J. Artymiuk, *GAPDOCK: a Genetic Algorithm Approach to Protein Docking in CAPRI round 1*. Proteins, 2003. **52**(1): p. 10-4.
16. J.M. Yang and C.C. Chen, *GEMDOCK: a generic evolutionary method for molecular docking*. Proteins, 2004. **55**(2): p. 288-304.
17. B.V.B. Reddy and Y. Kaznessis, *A quantitative analysis of interfacial amino acid conservation in protein-protein hetero complex structures*. J. Bioinformatics and Comput. Biol., 2005, in press.
18. Y. Duan, B.V. Reddy, and Y.N. Kaznessis, *Physicochemical and residue conservation calculations to improve the ranking of protein-protein docking solutions*. Protein Sci, 2005. **14**(2): p. 316-28.
19. R. Chen, et al., *A protein-protein docking benchmark*. Proteins, 2003. **52**(1): p. 88-91.
20. G. Moont, H.A. Gabb, and M.J. Sternberg, *Use of pair potentials across protein interfaces in screening predicted docked complexes*. Proteins, 1999. **35**(3): p. 364-73.
21. G.H. Gonnet, M.A. Cohen, and S.A. Benner, *Exhaustive Matching of the Entire Protein Sequence Database*. Science, 1992. **256**: p. 1443-1445.
22. W.S. Valdar and J.M. Thornton, *Protein-protein interfaces: analysis of amino acid conservation in homodimers*. Proteins, 2001. **42**(1): p. 108-24.
23. T.J. Richmond and F.M. Richards, *Packing of alpha-helices: geometrical constraints and contact areas*. J Mol Biol, 1978. **119**(4): p. 537-55.
24. A. Sali and T.L. Blundell, *Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming*. J Mol Biol, 1990. **212**(2): p. 403-28.
25. C. Dominguez, R. Boelens, and A.M. Bonvin, *HADDOCK: a protein-protein docking approach based on biochemical or biophysical information*. J Am Chem Soc, 2003. **125**(7): p. 1731-7.
26. K.E. Gottschalk, H. Neuvirth, and G. Schreiber, *A novel method for scoring of docked protein complexes using predicted protein-protein binding sites*. Protein Eng Des Sel, 2004. **17**(2): p. 183-9.
27. O. Lichtarge and M.E. Sowa, *Evolutionary predictions of binding surfaces and interactions*. Curr Opin Struct Biol, 2002. **12**(1): p. 21-7.
28. F. Glaser et al., *ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information*. Bioinformatics, 2003. **19**(1): p. 163-4.
29. C. Yan, D. Dobbs, and V. Honavar, *A two-stage classifier for identification of protein-protein interface residues*. Bioinformatics, 2004. **20**(Suppl. 1): p. i371-8.
30. I. Mihalek and O. Lichtarge, *A family of evolution-entropy hybrid methods for ranking protein residues by importance*. J Mol Biol, 2004. **336**(5): p. 1265-82.
31. O. Lichtarge et al., *Accurate and scalable identification of functional sites by evolutionary tracing*. J Struct Funct Genomics, 2003. **4**(2-3): p. 159-66.
32. E. Ben-Zeev and M. Eisenstein, *Weighted geometric docking: incorporating external information in the rotation-translation scan*. Proteins, 2003. **52**(1): p. 24-7.