ClusPro*: an automated docking and discrimination method for the prediction of protein complexes*

Stephen R. Comeau[1], David W. Gatchell[2], Sandor Vajda[1,2] and Carlos J. Camacho[1,2,]*

[1]*Bioinformatics Graduate Program and* [2]*Department of Biomedical Engineering, Boston University, 44 Cummington St, Boston, MA 02215, USA*

Pls. chk. the title given in PDF & email are different.

ABSTRACT

Motivation: Predicting protein interaction is one of the most challenging problems in functional genomics. Given two proteins known to interact, current docking methods evaluate billions of docked conformations by simple scoring functions, and in addition to near-native structures yield many false positives, i.e. structures with good surface complementarity but far from the native.

Results: We have developed an algorithm for filtering docked conformations rapidly with good surface complementarity, and ranking them based on their clustering properties. The free energy filters select complexes with lowest desolvation and electrostatic energies. Clustering is then used to smooth the local minima and to select the ones with the broadest energy wells—a property associated with the free energy at the binding site. The robustness of the method was tested on sets of 2000 docked conformations generated for 48 pairs of interacting proteins. In 31 of these cases, the top 10 predictions include at least one near-native complex, with an average RMSD of 5 Å from the native structure. The docking and discrimination method also provides good results for a number of complexes that were used as targets in the Critical Assessment of PRedictions of Interactions experiment.

Availability: The fully automated docking and discrimination server *ClusPro* can be found at http://structure.bu.edu

Contact: ccamacho@bu.edu

1 INTRODUCTION

The challenge for predictive protein docking is to start with the co-ordinates of the unbound component molecules and to obtain a model for the bound complex (Camacho and Vajda, 2002; Halperin *et al*., 2002; Smith and Sternberg, 2002). With the development of the Fourier correlation approach (Katchalski-Katzir *et al*., 1992; Vakser, 1996; Ritchie and Kemp, 2000), it became computationally feasible, for the first

*To whom correspondence should be addressed.

time, to generate and evaluate billions of possible docked conformations by simple scoring functions. Starting from unbound (separately crystallized) proteins, these methods generally yields both near-native structures and many false positives that have good surface complementarity (and electrostatics if included in the potential), but are far from the native complex. The main reason for this is the intrinsic uncertainty of the protein structures to be docked, e.g. the positions of solvent-exposed side chains (Kimura *et al*., 2001).

During the last couple of years, substantial progress has been made in developing methods that re-rank the docked conformations and attempt to select the ones close to the native, usually using a potential that accounts for the chemical activity between the molecules, and possibly refining the interacting surfaces (Weng *et al*., 1996; Gabb *et al*., 1997; Jackson *et al*., 1998; Moont *et al*., 1999; Camacho *et al*., 2000a; Norel *et al*., 2001). These procedures improve the discrimination, and 'hits', i.e. conformations with <10 Å RMSD, which sometimes can be found within the top 10 structures. For many other complexes, however, hundreds or even thousands of structures need to be retained in order to find the first hit.

In this paper, we describe an automated rigid-body docking and discrimination algorithm that rapidly filters docked conformations, and ranks them based on their clustering properties. The method has been implemented as a web server named *ClusPro*. Filtering involves the use of empirical free energy evaluation methods that select the conformations with the lowest desolvation and electrostatic energies (Camacho *et al*., 2000a). The clustering method, first implemented by Camacho and Gatchell for the 2001 Critical Assessment of PRedictions of Interactions (CAPRI) experiment (Janin *et al*., 2003), was motivated by the observation that the free energy landscapes of partially solvated receptor–ligand complexes (Camacho *et al*., 1999) showed the binding site free energy attractor as that with the greatest breadth of all the local minima. Hence, the expectation that a free energy filtered set of uniformly sampled docked conformations should in

1

fact cluster around the binding site. The method is also reminiscent of Shortle *et al.* (1998) protein structure prediction work, where the native conformation was found to be that with the highest number of structural neighbors. While we applied the method, followed by flexible docking, to a handful of proteins in the CAPRI experiment (Camacho and Gatchell, 2003), neither the method nor the clustering results have been published. Results are given both for a benchmark set of 48 proteins, and for the CAPRI targets. As will be shown, the approach predicts near-native complexes for a wide variety of proteins, including enzyme–inhibitor, antibody–antigen and signal transduction complexes, and generally provides better discrimination of near-native structures than energy-based rankings of the conformations.

2 SYSTEMS AND METHODS

2.1 Rigid body docking

Docked conformations have been generated using the docking program DOT (Mandell *et al.*, 2001; TenEyck *et al.*, 1995), based on the Fast-Fourier Transform (FFT) correlation approach. We have used version 1.0 alpha of DOT with a 10° Euler angle increment, and default values of 1 Å grid-step and 4 Å surface layer. Although the program allows the inclusion of electrostatics in the target function, only shape complementarity was used to sample approximately 10^{10} putative conformations, of which the top scoring 20 000 were retained for filtering by desolvation and electrostatics (see below). The discrimination algorithm was also tested on an independently developed benchmark set of 2000 docked conformations for 48 receptor–ligand protein pairs developed by Chen and Weng (2003), and available at http://zlab.bu.edu/~rong/dock/benchmark0.0.shtml

2.2 Filtering by empirical potentials

The rationale supporting an initial screening with surface complementarity is based on the observation that proteins generally bury relatively large surface areas upon complex formation (Chakravarti and Janin, 2002). However, the free energy of association is often dominated by desolvation and/or electrostatics contributions (Camacho *et al.*, 1999, 2000b). Complexes composed of oppositely charged proteins tend to form in free energy regions of low electrostatic potential, whereas complexes with weak charge complementarity form in free energy regions of low desolvation free energy (Camacho *et al.*, 1999). For example, systems like barnase and barstar are in the first category, protease inhibitors tend to fall in the second and antibody–antigen complexes fall somewhere in between. Therefore, we use electrostatic and desolvation potentials independently, in order to capture complexes whose binding mechanism is governed by any combination of the two. The desolvation free energy is calculated using the atomic contact potential

(ACP) (Zhang *et al.*, 1997). The electrostatic interactions are obtained by a simple Coulombic potential with the distance dependent dielectric of $4r$. We retain 500 structures with the lowest values of the desolvation free energy, and 1500 structures with the lowest values of the electrostatic energy (Camacho *et al.*, 2000a). The reason for retaining three times more electrostatic than desolvation candidates is that electrostatics is highly sensitive to small perturbations in the co-ordinates and, hence, yields many more outliers than the slowly varying ACP. The sum of desolvation and electrostatics will not provide additional information because of the rather noisy behavior of our electrostatic potential.

2.3 Clustering on the basis of pairwise RMSD

As shown by the free energy landscapes of partially solvated receptor–ligand complexes (Camacho *et al.*, 1999), the native-binding site is expected to exhibit a free energy attractor with the greatest breadth of all the local minima. Indeed, both thermodynamic and kinetic (Camacho *et al.*, 1999, 2000b) analyses suggest that the attractor is most relevant within distance separations of around a nanometer, or 10 Å. Based on these observations, we have developed a hierarchical clustering method to select and rank the docked complexes that have the largest number of neighbors within a certain fixed cluster radius ≤ 10 Å C_α RMSD.

Candidate docked conformations have one fixed molecule, typically the receptor, and one 'moving' molecule (the ligand). Moreover, we are mostly interested in the contact residues at the interface. Hence, throughout this paper RMSDs refer to rms deviations between *ligand residues* that are within 10 Å of any atom of the fixed receptor. We note that including the fixed receptor in the RMSD calculation (e.g. Chen and Weng, 2003) would decrease the RMSD by at least a factor 2, since the RMSD between receptors is 0 Å. The measure employed here has also the obvious advantage that it is not affected by parts of the molecule far from the interface and, hence, is frequently used to evaluate the results of docking programs, e.g. in the CAPRI experiment (Mendez *et al.*, 2003).

Since we cluster binding site RMSDs, for each docked conformation we need to compute the residues of the ligand within 10 Å of its receptor (typically around 28 residues), and the RMSD of these residues with all 2000 ligands. Thus, clustering 2000 docked conformations involves computing a 2000×2000 matrix of pairwise RMSD values. Based on the number of structures that a ligand has within a (default) cluster radius of 9 Å RMSD, we select the largest cluster and rank its cluster center Nr. 1. Then, the members of this cluster are removed from the matrix, and we select the next largest cluster and rank its center Nr. 2, and so on. After clustering, the ranked complexes are subjected to a straightforward (300 step and fixed backbone) van der Waals minimization using CHARMM (Brooks *et al.*, 1983) to remove potential side chain clashes.

2

3   THE ClusPro SERVER

The algorithm has been implemented as a fully automated protein docking and discrimination server *ClusPro*. The current version includes two FFT-based docking programs, DOT (Mandell *et al.*, 2001) and ZDOCK (Chen *et al.*, 2003), as its front-end. DOT runs retained 20 000 docked conformations, while ZDOCK runs retained 2000 structures. When submitting a job, *ClusPro* requires the user either to upload his/her own PDB (Berman *et al.*, 2000) files, or to input the PDB codes for automatic download from the PDB. By default, the server replies the top 10 (up to a maximum of 30) best predictions.

In order to evaluate arbitrary sets of docked conformations, the server can also upload candidate docked conformations in a flat-file format that consists of at least seven columns: one parameter (not used), three Cartesian co-ordinates of the relative position of the geometric center of the ligand (in Å) and the three Euler angles describing the relative orientation of the ligand with respect to the receptor (in radians). The rotational matrix follows the $ZX'Z''$ standard, listing the rotations in the order of $X'ZZ''$, the same used by the program DOT. Input formats to the server also include output files from the docking programs ZDOCK and GRAMM (Vakser, 1996). At least 2000 candidate complexes are required in any of these formats. The advantage of allowing the user to perform the docking independently of the server is that biochemical or other constraints can be included, if available, for the specific proteins. This option also makes the server a convenient platform for comparing the performance of different docking methods, and we plan to add other input formats and docking programs as they become available to us.

The server allows for customizing some of the parameters: (1) the clustering radius; (2) the relative number of desolvation and electrostatic best hits used by the free energy filtering and (3) the number of predictions the user would like to generate. Concerning the default values of these parameters, we have found that (1) a cluster radius of 9 Å produces the best results for proteins with around 200 residues. However, a smaller cluster radius may be more appropriate for smaller peptides. Similarly, large proteins of 700 or so residues often yield sparse sampling of hits that might produce better clusters with a larger cluster radius. (2) The free energy filtering can also be done using only either desolvation or electrostatic screening, by default retaining 500 and 1500 conformations, respectively. As shown below, this option could improve the ranking if the user knows the dominant contribution to the binding free energy. (3) The total number of predictions returned by the server can be set between 1 and 30 (default is 10).

The server runs using 16 processors on an IBM pSeries 690, each running at 1.3 GHz and sharing 32 GB of memory. The running time for each pair of average size proteins is between 3 and 4 h. The most time-consuming operation is the pairwise RMSD comparison, which currently uses only two processors. This can be optimized, and our goal is to reduce the overall running times to about 1 h per complex. Shortcomings of the server include size limitations of 11 999 atoms for the receptor and 4700 atoms for the ligand.

4   RESULTS AND DISCUSSION

4.1   Application to a benchmark set of complexes

We have tested the discrimination step of the method on the sets of 2000 docked conformations, developed by Chen and Weng (2003) for 48 protein pairs. Our results are summarized in Table 1. To assess the ranking procedure, we indicate the number of candidate complexes with a ligand RMSD under 5 Å and between 5 and 10 Å from the ligand found in the native complex structure (after the bound receptor has been overlapped with the fixed receptor from the set of docked conformations). The target function of ZDOCK accounts for shape complementarity, desolvation and electrostatics. Therefore, retaining the 2000 best scoring conformations is essentially a filtering step by empirical potentials, and we can proceed directly to clustering. Columns 5 and 6 show the best RMSD predictions after clustering and ranking using *ClusPro*.

To emphasize the role of the different components of the free energy, in columns 7 and 8 of Table 1 we show the best RMSD and rank using only the 500 conformations with the lowest desolvation (ACP) energy. Columns 9 and 10 show the same properties when clustering only the 1500 structures with the most favorable electrostatic energies. Based on these results we divide the complexes into four groups: (1) 23 complexes for which the clustering of the top 500 ACP produces the best prediction; (2) 10 complexes for which clustering of the top 1500 electrostatic complexes results in the best prediction; (3) nine complexes for which *only* the 2000 candidate complexes produce a good prediction and (4) six cases where the number of structures with <10 Å was not enough to make a prediction.

We consider the discrimination successful if a certain number of the top clusters include at least one conformation with <10 Å RMSD from the native (bold numbers in Table 1). Clustering all 2000 complexes for the benchmark set of 48 protein pairs resulted in 13, 31 and 39 successful predictions within the top 1, 10 and 30 ranked structures, respectively. Not including the six cases with poor sampling in Table 1, the success rate is 31, 74 and 93%, respectively. Independently clustering the top 500 desolvation and 1500 electrostatic structures yield overall success rates of 71 and 60% for the top 30 predictions, respectively. Given that the ACP empirical potential is not significantly affected by small overlaps and incorrect side-chain rotamers, it is not surprising that receptor–ligand pairs for which desolvation is important rank better by clustering the top 500 ACP complexes. The latter is particularly true for many enzyme–inhibitor complexes (indicated in italics in Table 1), which account for almost 50% of the proteins in the benchmark set. On the other hand, the noisier electrostatic field has a crucial contribution in only 22% of the cases (Table 1). However, the role of electrostatics

Table 1. RMSD and ranking of the best cluster for the benchmark set of 2000 docked conformations from Weng's lab (http://zlab.bu.edu/~rong/dock/benchmark0.0.shtml)

PDB codes		Number of hits		Clustered conformations					
Receptor	Ligand	0–5 Å	5–10 Å	All 2000		500 best ACP		1500 best electrostatics	
				RMSD	Rank	RMSD	Rank	RMSD	Rank
Complexes for which desolvation (ACP) alone produced best results									
1SUP	*3SSI*	31	90	**1.31**	**2**	1.31	1	2.01	2
2PTN	*1PPE(I)*	295	387	**3.95**	**1**	1.90	1	11.0	9
2PTN	*6PTI*	31	80	**2.17**	**3**	1.99	2	4.10	2
1SUP	*1SPB(P)*	99	51	**4.02**	**1**	2.06	1	4.02	1
1UDH	*1UDI(I)*	34	10	**2.70**	**10**	2.09	3	12.8	3
2PTN	*1BA7(A)*	48	33	**3.59**	**1**	2.10	2	3.59	2
5CHA	*1CSE(I)*	81	80	**2.95**	**4**	2.20	2	2.95	2
2PTN	*1HPT*	113	129	**2.74**	**1**	2.58	1	11.3	13
1BQL(LH)	1DKJ	64	37	**8.13**	**4**	3.07	2	8.12	17
1CHG	*1HPT*	91	189	**3.21**	**1**	3.21	1	3.47	1
2PKA(XY)	*6PTI*	11	36	**3.58**	**25**	3.58	7	9.71	4
1SCD	*1ACB(I)*	26	21	**3.75**	**8**	3.75	2	3.12	5
1MAA(B)	*1FSC*	27	14	**3.00**	**10**	3.96	4	3.00	7
1MEL(B)	1LZA	69	5	**4.75**	**2**	4.20	1	19.5	41
2HNT(LCEF)	*4HTC(I)*	47	19	**4.86**	**5**	5.19	2	13.5	7
1AKZ	*1UGI(A)*	22	23	**7.67**	**5**	5.27	3	12.4	15
1DQQ(LH)	3LZT	0	23	**6.50**	**29**	6.81	11	14.1	32
2BNH	*7RSA*	66	135	**7.42**	**1**	7.32	1	12.3	17
1ATN(A)	3DNI	21	14	**7.53**	**2**	7.53	1	19.4	23
1BVL(LH)	3LZT	0	55	**7.68**	**10**	7.68	2	7.68	2
1MLB(AB)	1LZA	1	88	**9.17**	**5**	9.17	1	8.22	11
1SUP	*2CI2(I)*	0	36	**9.40**	**9**	10.5	2	9.40	7
1QFU(LH)	2VIU(A)	14	2	**2.63**	**39**	2.97	10	1.66	27
Complexes for which electrostatics ($\varepsilon = 4r$) alone produced best results									
5CHA(A)	*2OVO*	90	73	**1.61**	**1**	6.11	7	1.61	1
2BTF(A)	1PNE	36	1	**1.11**	**7**	31.2	11	1.85	7
1THM	*2TEC(I)*	170	48	**3.97**	**1**	9.79	11	2.51	1
2JEL(LH)	1POH	38	69	**4.35**	**6**	16.6	5	2.74	6
2ACE(E)	*1FSC*	14	18	**4.38**	**26**	7.65	5	2.88	16
1PPN	*1STF(I)*	73	16	**4.60**	**1**	4.60	1	3.23	1
1A2P(B)	*1A19(A)*	20	52	**8.81**	**5**	9.53	2	4.09	5
1BRA	*1AAP(A)*	37	182	**6.91**	**1**	7.97	3	6.91	1
1FBI(LH)	1HHL	5	39	**8.45**	**19**	13.2	8	8.37	19
1CHN	1A0O(B)	5	13	**4.02**	**34**	21.0	7	3.93	26
Complexes for which only desolvation and electrostatic decoys produced good results									
1FGN(LH)	1BOY	75	60	**1.72**	**1**	6.79	8	20.2	4
1QBL(LH)	1HRC	14	51	**1.97**	**28**	18.8	3	11.4	5
1NCA(LH)	7NN9	62	12	**3.58**	**1**	16.2	15	14.3	30
1WER	5P21	6	95	**3.71**	**1**	11.5	7	10.6	9
2PTN	*1TAB(I)*	40	13	**4.99**	**12**	21.5	4	16.7	2
1AIF(LH)	1IAI(LH)	2	27	**6.86**	**15**	12.7	17	13.8	16
1NMB(LH)	7NN9	14	8	**7.86**	**11**	26.0	15	14.8	32
1JHL(LH)	1GHL(A)	4	55	**9.79**	**10**	14.7	6	5.91	29
1CCA	1YCC	0	11	**9.92**	**43**	15.7	3	16.2	6
Insufficient sampling near the binding site									
2BBK(LH)	1AAN	0	9	11.0	13	16.9	7	11.9	2
2VIR(AB)	2VIU(A)	3	0	15.2	45	15.3	6	15.7	23
1IGC(LH)	1IGD	3	0	15.6	34	16.4	8	15.7	28
1EO8(LH)	2VIU(A)	3	0	19.5	19	22.1	6	19.1	14
1AVV	1SHF(A)	0	0	19.7	41	23.5	8	19.9	28
1GLA(G)	1F3G	0	0	20.1	45	25.3	7	23.6	34

Enzyme–inhibitor complexes are in *italics*.

4

could in general be more relevant in filtering, since Table 1 over-represents proteases and other enzyme inhibitors. Apart from the complexes with insufficient sampling in Table 1, clustering all 2000 structures ranks at least one near-native conformation within the top 30 structures in all but three cases. It is interesting that for two of these, 1QFU/2VIU and 1CHN/1A0O, restricting consideration to either desolvation or electrostatics improves the results. For nine cases (21%), only the combination of ACP and electrostatics produce good results. In summary, clustering the combined set of docked conformations with low desolvation and/or low electrostatic energy is by far the most consistent predictor of good complex structures. However, our results also suggest that knowing the functional category of the complex could be helpful for optimizing the energy filtering of the data set.

While ZDOCK often produces an impressive number of hits, the performance of any docking method can change from one complex to another. For example, we have applied DOT, together with the default filtering and clustering algorithm, to two of the challenging cases in Table 1 that did not have hits below 10 Å. For the complex of 1GLA(G) and 1F3G the RMSD of the second best cluster is 5.17 Å, and we had 9.79 Å RMSD in the fifth ranked cluster for the complex of 1AVV and 1SHF(A). These results suggest that the performance of the discrimination method can be further improved, possibly concatenating sets of docked conformations generated by several docking programs.

4.2 Application to the CAPRI targets

Table 2 shows the predictions of the automated method for all nine target complexes from the CAPRI experiment (Janin *et al.*, 2003). For each target, we have generated 20 000 conformations using DOT, and applied the default filtering and clustering algorithm. In Rounds 1 and 2 of CAPRI (i.e. for targets 1–7), the same algorithm was used by Camacho and Gatchell (2003) to obtain a set of 25 clusters of potential complex structures that were then further refined and successfully re-ranked using SmoothDock, a flexible docking method (Camacho and Vajda, 2001). Although, the final CAPRI predictions were sometimes constrained based on available biochemical and/or structural information, no constraints were used when generating the results in Table 2. The CAPRI experiment tested the limits of our methodology. Indeed, Target 1 (1JB1/1SPH) has a significant structural change upon binding; both 1QHD and 2VIU in Targets 2 and 3, respectively, have more than 1000 residues each, forcing the sampling to be quite sparse; the free energy filtering of Target 6 (1PIF/1KXV) is relatively poor since its binding affinity is 230 nM; and Target 7 (1BEC/1B1Z) has poor affinity and surface complementarity. Nevertheless, the performance of *ClusPro* compares well with the manual predictions submitted to CAPRI (Mendez *et al.*, 2003). A significant achievement of the server is that, besides some minor changes in the ranking and RMSD, the successful predictions are reproduced

Table 2. Best RMSD and ranking predictions for the nine targets in Rounds 1, 2 and 3 of CAPRI

CAPRI		PDB codes		Best cluster	
Round	Target	Receptor	Ligand	Rank	RMSD
1	1	1JB1(tri)	1SPH(A)	5	12.6
1	2	1QHD	MCV Fab*	5	19.0
1	3	2VIU	Fab HC63*	21	**3.61**
2	4	1PIF	1KXV(C)	10	10.5
2	5	1PIF	1KXT(B)	21	**1.95**
2	6	1PIF	1KXV(C)	2	**3.68**
2	7	1BEC	1B1Z(A)	9	20.1
3	8	Nidogen-G3	1KLO	3	**6.5**
3	9	1H99	1H99	10	24.7

*Represents the Fab molecules given in CAPRI.

regardless of the original orientation of the PDBs submitted to the server. This strongly suggests that our multi-step approach is capturing some relevant markers of the general mechanism of protein recognition. This was further demonstrated in Round 3 of CAPRI, where *ClusPro* participated as an automated server, and produced one of the best predictions (the best in terms of binding site RMSD) among all participating groups for Target 8 (see http://capri.ebi.ac.uk/round3).

5 CONCLUSION

We describe the excellent performance of an automated docking and discrimination method. The algorithm filters the docked conformations by selecting the ones with favorable desolvation and electrostatics properties, clusters the retained structures using a hierarchical pairwise RMSD algorithm, and selects the centers of the most populated clusters as predictions of the unknown complex. Applied to a benchmark set of 2000 conformations, the algorithm predicts at least one experimentally relevant complex structure within the top 30 predictions, and, in about 30% of the cases, the best prediction is ranked first. For all 39 near-native predictions, the average ligand RMSD is only 5 Å from the native structure.

The method has been implemented in the framework of the server *ClusPro*. To our knowledge, *ClusPro* is the first fully integrated server that includes both docking and discrimination steps for predicting the structure of protein–protein complexes. The server can be used to discriminate a set of potential complex structures from several docking algorithms, or it can generate its own structures using DOT or ZDOCK. The performance of the server has also been successfully tested in a blind experiment in Round 3 of CAPRI, where it obtained one of the best predictions. The success rate can be further improved by using additional stages of discrimination. In particular, we have developed an algorithm, named *SmoothDock*, which adds side-chain flexibility to optimize and discriminate local clusters. For most applications, manual

discrimination among a reduced number of possibilities using biochemical constraints and bioinformatic analysis could ultimately eliminate the few remaining false positives. Hence, we expect that this new technology will be useful for the structural biology and biochemistry research communities.

ACKNOWLEDGEMENTS

REFERENCES

Pls provide first 10 author names

Berman,H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.

Camacho,C.J. and Gatchell,D. (2003) Successful discrimination of protein interactions. *Proteins*, **52**, 92–97.

Camacho,C.J., Gatchell,D.W., Kimura,S.R. and Vajda,S. (2000a) Scoring docked conformations generated by rigid-body protein–protein docking. *Proteins*, **40**, 525–537.

Camacho,C.J., Kimura,S.R., DeLisi,C. and Vajda,S. (2000b) Kinetics of desolvation-mediated protein–protein binding. *Biophys. J.*, **78**, 1094–1105.

Camacho,C.J. and Vajda,S. (2001) Protein docking along smooth association pathways. *Proc. Natl Acad. Sci. USA*, **98**, 10636–10641.

Camacho,C.J. and Vajda,S. (2002) Protein–protein association kinetics and protein docking. *Curr. Opin. Struc. Biol.*, **12**, 36–40.

Camacho,C.J., Weng,Z.P., Vajda,S. and DeLisi,C. (1999) Free energy landscapes of encounter complexes in protein–protein association. *Biophys. J.*, **76**, 1166–1178.

Chakravarti,P. and Janin,J. (2002) Disecting protein–protein recognition sites. *Proteins*, **47**, 334–343.

Pls. update.

Chen,R., Li,L. and Weng,Z. (2003) ZDOCK: an initial-stage protein docking algorithm. *Proteins*, (in press).

Chen,R. and Weng,Z. (2003) A novel shape complementarity scoring function for protein–protein docking. *Proteins*, **51**, 397–408.

Gabb,H.A., Jackson,R.M. and Sternberg,M.J.E. (1997) Modelling protein docking using shape complementarity, electrostatics, and biochemical information. *J. Mol. Biol.*, **272**, 106–120.

Halperin,I., Ma,B., Wolfson,H. and Nussinov,R. (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins*, **47**, 409–443.

Jackson,R.M., Gabb,H.A. and Sternberg,M.J.E. (1998) Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.*, **276**, 265–285.

Janin,J., Henrick,K., Moult,J., Ten Eyck,L., Sternberg,M.J., Vajda,S., Vakser,I. and Wodak,S.J. (2003) CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins*, **52**, 2–9.

Katchalski-Katzir,E., Shariv,I., Eisenstein,M., Friesem,A., Aflalo,C. and Vakser,I.A. (1992) Molecular surface recognition—determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl Acad. Sci. USA*, **89**, 2195–2199.

Kimura,S.R., Brower,R.C., Vajda,S. and Camacho,C.J. (2001) Dynamical view of the positions of key side chains in protein–protein recognition. *Biophys. J.*, **80**, 635–642.

Mandell,J.G., Roberts,V.A., Pique,M.E., Kotlovyi,V., Mitchell,J.C., Nelson,E., Tsigelny,I. and Ten Eyck,L.F. (2001) Protein docking using continuum electrostatics and geometric fit. *Protein Eng.*, **14**, 105–113.

Mendez,R., Leplae,R., De Maria,L. and Wodak,S.J. (2003) Assessment of blind predictions of protein–protein interactions: current status of docking methods. *Proteins*, **52**, 51–67.

Moont,G., Gabb,H.A. and Sternberg,M.J.E. (1999) Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins*, **35**, 364–373.

Norel,R., Sheinerman,F., Petrey,D. and Honig,B. (2001) Electrostatic contributions to protein–protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci.*, **10**, 47–61.

Ritchie,D.W. and Kemp,G.J.L. (2000) Protein docking using spherical polar Fourier correlations. *Proteins*, **39**, 178–194.

Shortle,D., Simons,K.T. and Baker,D. (1998) Clustering of low-energy conformations near the native structures of small proteins. *Proc. Natl Acad. Sci. USA*, **95**, 11158–11162.

Smith,G.R. and Sternberg,M.J.E. (2002) Prediction of protein–protein interactions by docking methods. *Curr. Opin. Struct. Biol.*, **12**, 28–35.

TenEyck,L.F., Mandell,J., Roberts,V.A. and Pique,M.E. (1995) Surveying molecular interactions with DOT. In Hayes,A. and Simmons,M. (eds), *Proceedings of the 1995 ACM/IEEE Supercomputing Conference*. ACM Press, New York.

Vakser,I.A. (1996) Low-resolution docking: prediction of complexes for underdetermined structures. *Biopolymers*, **39**, 455–464.

Weng,Z., Vajda,S. and DeLisi,C. (1996) Prediction of protein complexes using empirical free energy functions. *Protein Sci.*, **5**, 614–626.

Zhang,C., Cornette,J.L. and DeLisi,C. (1997) Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.*, **267**, 707–726.