PROTEINS: Structure, Function, and Bioinformatics 68:187–194 (2007)

Identification of Near-Native Structures by Clustering Protein Docking Conformations

Stephan Lorenzen and Yang Zhang*
Center of Bioinformatics, Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66046

ABSTRACT **Most state-of-the-art protein–protein docking algorithms use the Fast Fourier Transform (FFT) technique to sample the six-dimensional translational and rotational space. Scoring functions including shape complementarity, electrostatics, and desolvation are usually exploited in ranking the docking conformations. While these rigid-body docking methods provide good performance in bound docking, using unbound structures as input frequently leads to a high number of false positive hits. For the purpose of better selecting correct docking conformations, we structurally cluster the docking decoys generated by four widely-used FFT-based protein–protein docking methods. In all cases, the selection based on cluster size outperforms the ranking based on the inherent scoring function. If we cluster decoys from different servers together, only marginal improvement is obtained in comparison with clustering decoys from the best individual server. A collection of multiple decoy sets of comparable quality will be the key to improve the clustering result from meta-docking servers. Proteins 2007;68:187–194.**

Key words: protein–protein docking; Fast Fourier Transformation; protein structure clustering; docking decoy; docking meta-server; TM-score

INTRODUCTION

The goal of protein–protein docking is to elucidate the quaternary structure of protein complexes by reassembling the solved monomer structures. Since the possible orientations of receptor and ligand towards each other are in the order of billions (three translational and three rotational degrees of freedom), an exhaustive search of all possible conformations using traditional energy scoring methods is not feasible. The introduction of the grid-based Fast Fourier Transform (FFT) method by Katchalski-Katzir et al. in 1992 made an extensive screening of orientations possible for the first time.[1] The proposed FFT method starts with a grid-based representation of the target receptor and ligand molecules by three-dimensional discrete functions that distinguish between the surface and the interior of protein molecules. A correlation function that assesses the degree of molecular shape match can then be quickly calculated with the aid of the Fourier transformation technique.

FFT methods have been enhanced by several means. While the original approach summed up the number of overlapping surface grid points as a measure of shape complementarity, Chen and Weng introduced a pair-wise shape complementarity function which provides higher scores in regions of high curvature and favors continuous contact patches over several small contact patches.[2] In addition to shape complementarity, terms for desolvation[3] and electrostatic energy[4] have been introduced for the account of water and protein atom interactions. Common methods of postdocking processes for refining the docking conformations include conjugate gradient rigid-body energy minimization,[5,6] simultaneous backbone rigid-body and side-chain rotamer optimization,[7] and the use of computationally more expensive scoring functions to rerank the FFT decoys.[8]

For the purpose of identifying near-native docking decoys, another efficient approach is the structural clustering of the conformations generated by the docking search algorithms. The basic assumption behind the clustering approach is that the free energy landscape exhibits a broader and deeper well near the native structure than near non-native structures.[9] Therefore, a reasonable docking algorithm should have more structures in near-native regions. Clustering methods have been introduced and extensively used in tertiary protein structure prediction for picking up correct protein folds.[10,11] Vakser et al.[12,13] introduced clustering of docking decoys in low-resolution docking studies. It is pointed out that, in contrast to high-resolution docking, where a tight lock between the receptor and the ligand is needed and thus only few discrete decoys fall into one energy minimum, low-resolution representation of potentials lead to broader wells in the energy landscape and thus produce many "partially correct" hits which fall in the same energy minimum and can thus be detected by clustering. Law et al.[14] used docking decoys generated by

Grant sponsor: KU Start-up Fund; Grant number: 06194; Grant sponsor: NFGRF; Grant number: 2302003.

*Correspondence to: Yang Zhang, Center of Bioinformatics, Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66046. E-mail: yzhang@ku.edu

Received 31 August 2006; Revised 27 November 2006; Accepted 19 January 2007

Published online 30 March 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21442

DOT[15] and, after shape complementarity filtering, clustered the remaining decoys using a six-dimensional distance measure describing the rotations and translations between different ligand orientations. Camacho and Gatchell[16] used an all C_α RMSD and a C_α binding site RMSD to cluster decoys generated by DOT.[15] The algorithm using C_α binding site RMSD is now made publicly available as a server called ClusPro,[17,18] which shows improvements in selecting docking decoys over the inherent docking scoring function. The RMSD matrix in ClusPro is calculated for ligand residues which are less than 10 Å away from the receptor and the default clustering RMSD cutoff is set to 9 Å. Clustering is known to be used by many other docking algorithms after sampling the possible orientations of receptor and ligand towards each other.[19] However, the parameters used in clustering differ between the groups, and to our knowledge no systematic approach of optimizing parameters like the RMSD cutoff or the number of decoys used for clustering has been published. Kozakov et al.[20] recently suggested using a clustering threshold corresponding to the minimum following the first peak of the bimodal pairwise RMSD histogram, which usually lies between 5 and 10 Å.

In this work, we will apply the clustering algorithm to structure decoys generated by four different FFT-based protein–protein docking methods and study in detail the impact of various clustering parameters on the clustering performance. Another purpose was inspired by the idea of meta-servers in the protein tertiary structure prediction, where the consensus models collected from different servers on average perform better than that of the best individual server.[21–23] We will test whether additional gain can be obtained in the quaternary structure prediction if we combine and cluster structure decoys from different FFT docking methods. According to Camacho et al.,[24] the differences in refinement and scoring functions between docking algorithms make a rescoring of the pooled decoys of different methods less successful than rescoring the decoys from each single method. It is not known whether the clustering behavior of a pool of decoys from metaserver is influenced by differences in the energy scoring methods. However, a rough scoring function might result in only few hits around the native conformation, whereas a smoother function is more tolerant to changes in the ligand orientation and will thus result in bigger clusters.

MATERIALS AND METHODS

Rigid-Body Docking

We took unbound structures of 59 nonredundant protein complexes from Weng's benchmark library.[25] Prior to docking, ligand structures were randomly translated and rotated. ZDOCK[26] version 2.3, FTDOCK[7] version 2.0 with rescoring by RPScore,[8] GRAMM-X[5,27] version 1.0.1, and MOLFIT[28] version 2 were used to generate rigid-body docking conformations.

In all these programs, the conformational search is performed using the FFT technique of Katchalski-Katzir et al.[1] The methods differ in their scoring functions: both FTDOCK[7] and MOLFIT[28] use shape complementarity and electrostatic potential in their FFT scoring functions. FTDOCK[7] provides an additional program (RPScore[8]) to rerank docking decoys by pair-wise residue potentials. In ZDOCK,[26] the surface-overlap shape complementarity function is replaced by a pair-wise shape complementarity, and terms for average Atomic Contact Energies are added as a measure of desolvation energy. GRAMM-X[5] includes a grid-projection of a smoothed Lennard-Jones potential, combined with a postdocking procedure of rigid-body energy minimization and structure clustering.

All programs were run with default parameters. If not stated otherwise, 2000 highest-scoring decoys of each method were retained for further analysis. As a control, the same randomly translated and rotated structures were submitted to the ClusPro server for clustering.[17]

Structure Clustering

For a given set of docking decoys, an all-against-all RMSD matrix was calculated by comparing the ligand coordinates of different orientations after superposing the C_α atoms of the receptor. The first cluster center is defined as the docking decoy with the highest number of structural neighbors below a given RMSD threshold. After excluding the first cluster center and its structural neighbors, the second cluster center is identified as the decoy with the highest number of structural neighbors in the remaining decoy pool. Subsequent cluster centers can be obtained by repeating the procedure. This clustering procedure is similar to that used in SPICKER,[11] which was designed to cluster tertiary structure conformations.

Decoy Quality Evaluation

While the pairwise RMSD of complete ligands was used for clustering, the quality of docking predictions is evaluated by the RMSD to native of interface residues, where interface residues are defined as the residues of the receptor (ligand) that have at least one atom within 10 Å to the ligand (receptor) in the native structure. One purpose of using the interface RMSD rather than complex RMSD or ligand RMSD is to minimize the influence of the sizes of ligand and receptor. The same measure has been used to evaluate docking quality in the CAPRI experiment.[29] A 'hit' here is defined as a docking prediction with an interface RMSD below 2.5 Å.

By definition, the RMSD calculation averages the squares of distances of all residue pairs with the same weight.[30] A local error (e.g., a tail misorientation) may cause a big RMSD value although the global topology of the model is correct. Alternatively, we will use the TM-score,[31] which weights small distances stronger than larger distances and therefore is more sensitive to the global topology of models. Moreover, the TM-score value

TABLE I. Comparison of Four Rigid-Body Docking Algorithms on 59 Benchmark Protein Complexes

	Number of hits in top-10	Number of hits in top-2000	Number of cases with >50 hits in top-2000
ZDOCK	19	42	15
FTDOCK	5	35	0
MOLFIT	6	26	0
GRAMM-X (default)	17	42	0
GRAMM-X (filtered)	13	36	3
ClusPro on ZDOCK	22	—	—

is rescaled by the average distance of random structure pairs in the PDB library, which makes the meaning of a specific TM-score value independent on the target protein sizes.

RESULTS AND DISCUSSION

Benchmark Docking of Four FFT-Based Docking Algorithms

In Table I, we summarize the docking results of the four FFT-based docking algorithms on the 59 targets taken from Weng's benchmark set.[25] If ranking the decoys by the inherent scoring function, ZDOCK has at least one hit within the first 10 decoys in 19 cases. If the top 2000 decoys are considered, the number of cases with hits increases to 42. In 15 cases, ZDOCK has more than 50 hits in the top 2000 decoys. FTDOCK and MOLFIT have hits in the top 10 decoys in five and six cases, respectively. Interestingly, the number of cases with hits in the top 2000 decoys is comparable in all four methods. The higher number of hits in top 10 by ZDOCK indicates the power of its inherent scoring function in distinguishing good decoy from bad ones.

The number of hits in the GRAMM-X run is comparable to that of ZDOCK (17 in top-10 and 42 in top-2000). With consideration of the inherent postdocking procedure, the good performance of GRAMM-X may be partially due to the combined effect of the minimization and clustering procedures. As shown in row six of Table I, if we omit the clustering and minimization procedures and allow only one decoy for each of the FFT rotation steps, the number of hits by GRAMM-X is reduced to 13 (in top-10) and 36 (in top-2000). We will refer to the first type of GRAMM-X decoys with minimization as 'default GRAMM-X' and to the second type of GRAMM-X decoys as 'filtered GRAMM-X'.

In the last row of Table I, we show the clustering result of ZDOCK decoys by ClusPro, which improves the number of hits in top 10 from 19 to 22.

Number of Structural Neighbors

In Figure 1, we show the average number of structural neighbors, as judged by the pair-wise ligand RMSD after superimposing the receptor, for all and near-native decoys. Here, a near-native structure is defined as a structure with an interface RMSD to native below 5 Å.

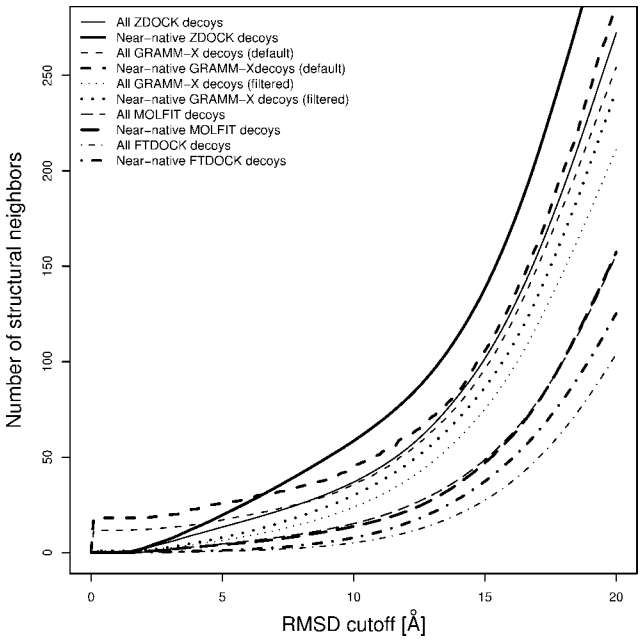

Fig. 1. Average number of structural neighbors versus the RMSD cutoff. For each program, two sets of decoys are calculated: all decoys (thin curves) and near-native decoys (thick curves).

Ideally, in a docking algorithm of reasonable scoring potential, a correct docking conformation should have more structural neighbors than an incorrect one because the native structure often shows a wider well in the free energy landscape.[9] On average, near-native decoys in ZDOCK have much more structural neighbors than non-native ones and therefore can be clearly differentiated from the latter by structure clustering. In contrast, the clustering of near-native decoys generated by MOLFIT is barely distinguishable from non-native ones. In FTDOCK, the near-native decoys cluster relatively better than the non-native decoys. However, the overall decoys are very divergent.

For the default GRAMM-X run, because of the energy minimization procedure, the decoy conformations tend to be driven to some common energy minima. Therefore, the default GRAMM-X decoys shows a large number of neighbors even with a very small RMSD cutoff. If the minimization step is omitted and only one decoy is kept per FFT rotation step, the redundant structural neighbors vanish (see Fig. 1, dotted lines). The near-native filtered GRAMM-X decoys have also more structural neighbors than non-native decoys. Thus, the decoys produced by ZDOCK and the filtered GRAMM-X decoys seem to be the most promising for the subsequent clustering.

Clustering of Individual Decoy Sets

In Figure 2, we show the clustering results of docking decoys generated by different algorithms using various

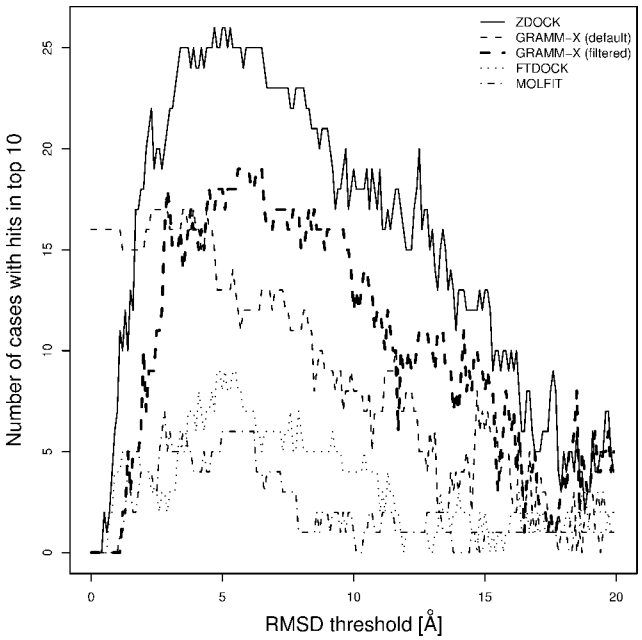

Fig. 2. Number of cases with hits in top 10 decoys when the decoys are clustered using different RMSD cutoffs.

RMSD cutoffs. The number of cases with hits show a peak around RMSD = 5 Å for all decoys (except for the default GRAMM-X). These data demonstrate that a RMSD of 5 Å may be used as a uniform cutoff for clustering of all docking decoys. The maximum at small RMSD cutoffs in the default GRAMM-X is due to the inherent energy minimization procedure which drives the decoys into some common energy minima. Because of the redundancy of the decoy structures, using different RMSD cutoffs below 5 Å just identifies the same set of tightly clustered decoys. As shown in both Figures 1 and 2, using the filtered GRAMM-X eliminates the redundancy and improves the clustering of GRAMM-X decoys. Overall, the number of hits in ZDOCK clustering is higher than all other individual methods because of its better decoy quality and the relatively higher number of structural neighbors of near-native decoys compared to non-native ones as shown in Figure 1. When using a cutoff of 5 Å, our clustering approach improves the number of ZDOCK hits from 19 to 25 cases, slightly more than the 22 cases detected by ClusPro. This difference of performance should be due to the different clustering strategies. In ClusPro, the RMSD matrix is calculated from ligand interface residues, where the RMSD is more sensitive to the ligand spin/rotation. In our method, the RMSD matrix is calculated using the complete ligands, where the RMSD is more sensitive to the receptor spin/rotation. The different sensitivity in these two cases can be easily understood by the example of spinning a ligand with the receptor fixed, where the change on the global ligand RMSD is smaller than that on the ligand interface RMSD. On the other hand, when we spin a receptor with the ligand fixed, after superposing the receptor, the change on the global ligand RMSD will be larger than that on the ligand interface RMSD.

TABLE II. The Distribution of 59 Benchmark Cases Using Different Ranking Methods

ZDOCK score	ClusPro	Current clustering	Number of cases (A/E/O/D)
			30 (12/4/7/7)
	x		3 (2/1/0/0)
	x	x	4 (1/2/1/0)
		x	3 (1/2/0/0)
x	x		1 (0/1/0/0)
x		x	4 (1/2/1/0)
x	x	x	14 (2/10/2/0)

'x' denotes the case that has at least one hit in top 10 decoys. The targets are categorized as antibody–antigen (A), enzyme-inhibitor (E), 'other' (O), and 'difficult test case' (D) according to Chen and Weng.[25]

In Table II, we list the distribution of the 59 targets when the ZDOCK decoys are ranked by different methods, i.e. ZDOCK inherent scoring function, ClusPro, and current clustering. There are 30 cases that have no hit by any ranking, although 13 of them have at least one hit in the decoy pool. Among them, 12 cases are antibody–antigen complexes, four are enzyme-inhibitor complexes, and seven are other complexes. The remaining seven cases are 'difficult test cases' in Chen and Weng's assignment.[25]

In 14 easy cases, the correct hits can be identified by all three methods. The majority of these (10) are enzyme-inhibitor complexes. In the remaining 15 cases, the correct decoys can be picked up by at least one of the methods, which show the potential power of combining different ranking methods.

Clustering of Combined Decoys From Different Programs

Meta-servers have been extensively used in protein tertiary structure prediction. It has been demonstrated that the consensus prediction derived from models generated by different modeling servers usually outperforms the best individual server.[21–23] Inspired by the meta-server idea, here we check whether a combination of decoys from different docking servers leads to further improvement of the results.

In Figure 3, we show the results of clustering decoys from two different methods. Obviously, because the FTDOCK and MOLFIT decoys are much more divergent than ZDOCK decoys, as demonstrated in Figure 1, the combination of ZDOCK with FTDOCK or MOLFIT does not change the results much in comparison to clustering ZDOCK decoys alone. However, it was somewhat unexpected that the combination of ZDOCK and GRAMM-X performed worse than clustering ZDOCK decoys only (thin short-dashed curve in Fig. 4), although both methods show a similar number of hits in top-10 and top-2000 decoys in Table I. One reason might be that the

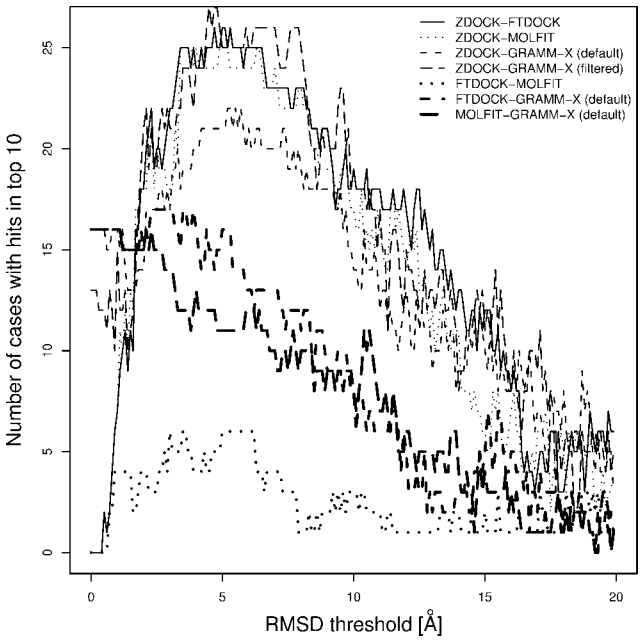

Fig. 3. Clustering results of combined decoys from two different docking programs.

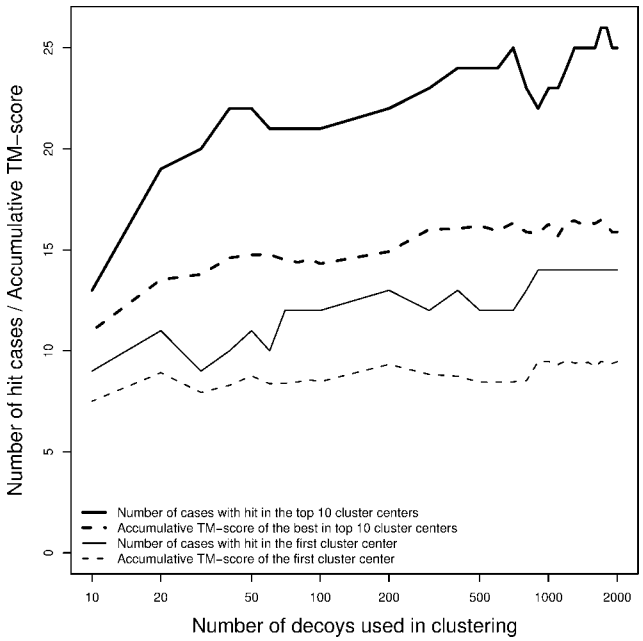

Fig. 4. Dependence of clustering results on the number of used decoys.

energy optimization procedure in the default GRAMM-X run has driven the decoys into common structure basins. Therefore, the redundant GRAMM-X decoys dominate the combined clustering and mask the cluster centers from ZDOCK. As shown in the thin long-dashed curve in Figure 3, omission of the energy optimization procedure and filtering of the redundant GRAMM-X decoys by choosing one decoy for each FFT rotation step leads to better clustering performance than the combination of ZDOCK and default GRAMM-X. However, compared with the ZDOCK clustering alone (25 cases in top 10 hits), the increase in the number of hits (27 cases in top 10 hits) is still marginal. Hit rates when combining three or four programs are not better than results from combining two programs (data not shown). The lack of obvious improvement when combining different programs might be due to the big differences in quality of docking decoys and clustering behavior between the different programs. A collection of comparable decoys from new good servers will be helpful. However, differences in the softness of the scoring functions of the programs used might cause different tolerance towards small movements of the ligand, which can influence the number of structural neighbors included in their high-scoring decoys.

How Many Decoys Should be Used in Clustering?

The choice of a subset of clustering decoys from billions of docking conformations generated in FFT scans is always a compromise between the number of hits in the pool and the noise contamination. Since the decoys are usually ranked and preselected by the inherent scoring function, a pool of fewer decoys should have on average better quality but include fewer hits. On the other hand, the inclusion of more decoys in the clustering will lead to more hits in the pool but be in the danger of including more noise. In the above clustering calculations, we have by default used 2000 decoy structures in our clustering pool. Here we examine the numbers in more detail.

In Figure 4, we plot the clustering result versus the number of used decoys. Besides the binary count of the number of hit cases, we also calculate a continuous accumulative TM-score of all targets, where TM-score is defined as[31]

$$\text{TM-score} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{1+d_i^2/d_0^2}. \quad (1)$$

Here, d_i is the $C_\alpha - C_\alpha$ distance of the ith pair of residues in the ligand between the decoy and the native structure after superimposing the receptor. $d_0 = 1.24\sqrt[3]{N-15} - 1.8$ and N is the number of residues in the ligand molecule.

For both hits in the first cluster and in top-10 clusters, there is a gradual increase when including more decoys in the clustering pool. After the total decoy number reaches around 1000, there is no obvious change on the overall performance. The number of hits in the first cluster is the same when using 1000 and 2000 decoys. While the hit number in top-10 clusters fluctuates around 1000, it is the same when using 2000 or 750 decoys. The accumulative TM-score stays stable after the number of decoys increases to 1000. These data confirm that 2000 decoys should be a safe choice for the docking decoy clustering.

TABLE III. TM-Score of Cluster Center, Cluster Centroid, and The Conformation Superimposed on the Cluster Centroid (Superpos.)

ID	Center	Centroid	Superpos.
1A0O	0.029	0.031	0.031
1ACB	0.383	0.406	0.404
1AHW	0.639	0.686	0.687
1ATN	0.885	0.909	0.909
1AVW	0.630	0.616	0.613
1AVZ	0.021	0.018	0.018
1BQL	0.640	0.721	0.719
1BRC	0.222	0.245	0.243
1BRS	0.127	0.127	0.125
1BTH	0.071	0.070	0.069
1BVK	0.038	0.038	0.038
1CGI	0.538	0.532	0.533
1CHO	0.307	0.381	0.382
1CSE	0.125	0.112	0.110
1DFJ	0.593	0.678	0.678
1DQJ	0.070	0.067	0.067
1EFU	0.022	0.023	0.023
1EO8	0.050	0.051	0.051
1FBI	0.115	0.121	0.119
1FIN	0.021	0.022	0.022
1FQ1	0.073	0.073	0.073
1FSS	0.109	0.119	0.116
1GLA	0.011	0.010	0.010
1GOT	0.102	0.100	0.100
1IAI	0.090	0.090	0.090
1IGC	0.006	0.006	0.006
1JHL	0.264	0.280	0.278
1KKL	0.016	0.017	0.017
1KXQ	0.649	0.649	0.648
1KXT	0.041	0.042	0.043
1KXV	0.015	0.014	0.014
1L0Y	0.014	0.014	0.014
1MAH	0.035	0.039	0.037
1MEL	0.757	0.787	0.790
1MLC	0.134	0.118	0.117
1NCA	0.834	0.888	0.888
1NMB	0.099	0.098	0.098
1PPE	0.365	0.358	0.365
1QFU	0.031	0.031	0.031
1SPB	0.693	0.761	0.765
1STF	0.683	0.750	0.748
1TAB	0.328	0.307	0.315
1TGS	0.550	0.685	0.685
1UDI	0.406	0.671	0.672
1UGH	0.473	0.469	0.466
1WEJ	0.081	0.081	0.079
1WQ1	0.512	0.505	0.503
2BTF	0.916	0.930	0.931
2JEL	0.090	0.098	0.097
2KAI	0.074	0.071	0.069
2MTA	0.021	0.019	0.019
2PCC	0.085	0.086	0.085
2PTC	0.228	0.244	0.243
2SIC	0.219	0.236	0.236
2SNI	0.113	0.121	0.119
2TEC	0.636	0.684	0.688
2VIR	0.047	0.045	0.045
3HHR	0.023	0.023	0.023
4HTC	0.535	0.644	0.654
Sum	*15.884*	*17.017*	*17.020*

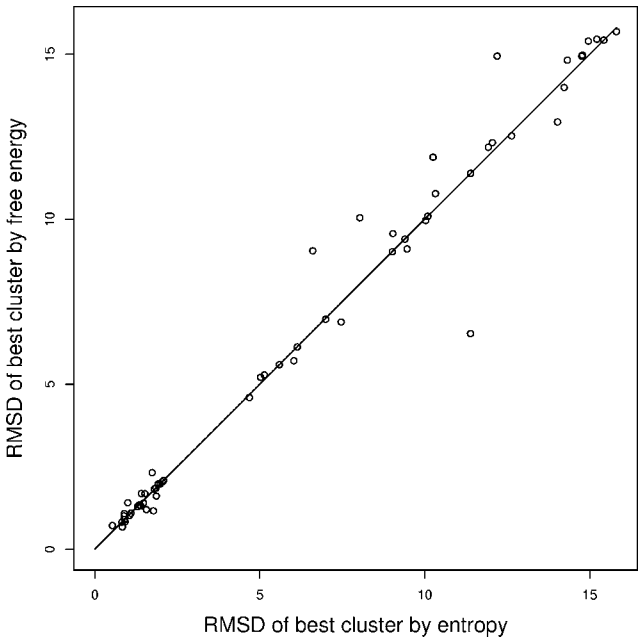

Fig. 5. Lowest RMSD within first 10 clusters with and without weighting by Boltzmann factor.

Cluster Centroid and Cluster Center

In our cluster algorithm, the cluster center (the decoy which has the most structural neighbors) was returned as the final docking prediction. Here, we also try the option of averaging the atom coordinates of all clustered ligand conformations and generating an artificial decoy called 'cluster centroid'. As shown in Table III, the cluster centroid in general has a higher TM-score than the cluster center. However, the cluster centroid is usually structurally distorted because of the averaged coordinates. The distortion will become more serious if the RMSD cutoff of the clustering is large. In the last column of Table III, we generate a new docking conformation by superimposing the ligand structure onto the cluster centroid. On average, the TM-score of the docking conformation generated by the centroid superposition is as good as the cluster centroid, which is about 7% higher than the cluster center.

Ranking of the Clusters by Free Energy

The general procedure of clustering the FFT-based docking decoys in RMSD space is actually to identify the state of the maximum entropy, because the energy or scoring function is not used to drive the FFT searching. This is different from the clustering of decoys generated by Metropolis Monte Carlo simulations, where the biggest cluster corresponds to the lowest free-energy because ideally the Monte Carlo conformation is inherently weighted by the Boltzmann factor.[11] To identify the lowest free-energy state in FFT-based docking, we can simply rerank the clusters in a way that each decoy is weighted by the Boltzmann factor $\exp(-\beta E)$, where E is

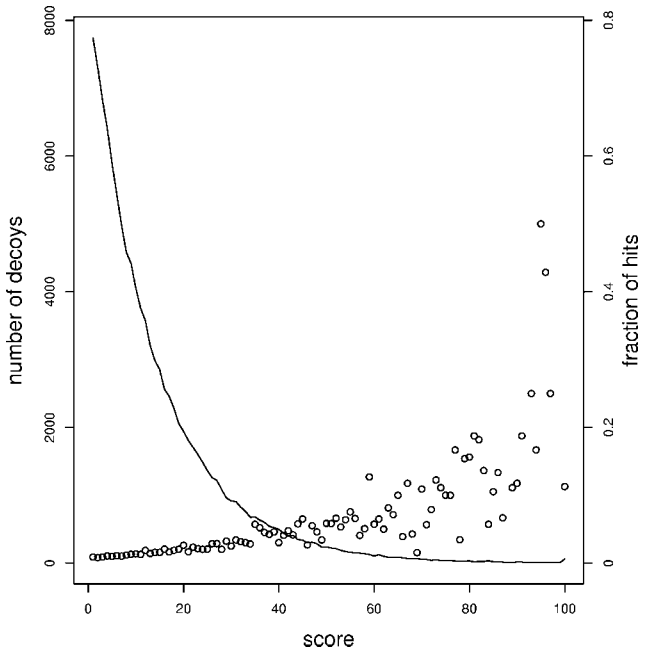

Fig. 6. Number of decoys within a bin with width equal to 1% (solid curve) and fraction of hits with interface RMSD <2.5 Å in decoys of different scores ([circo]).

the energy/scoring function of the decoys in the cluster and β is the reciprocal temperature. In Figure 5, we show the comparison of clusters of the lowest free-energy and that of the maximum entropy for ZDOCK decoys. The result indicates some difference but overall the number of hit is quite similar.

The result is understandable based on Figure 6, which shows that, for the majority of the decoys, the ZDOCK score is not sensitive to the real quality of the conformation (see left part of Fig. 6). Therefore, the Boltzmann factor in these cases is close to random and does not improve the selection of docking decoys. To improve the free-energy based clustering on the decoy selection, a scoring function with strong correlation to the decoy quality will be needed.

CONCLUSIONS

In summary, we develop different clustering algorithms for picking up near-native docking conformations generated by the four widely used FFT-based protein–protein docking methods ZDOCK,[26] GRAMM-X,[5,27] FTDOCK,[4] and MOLFIT.[28] The performance of our clustering algorithms varies considerably depending on the quality and structural distribution of the decoys. However, in all cases, the decoy ranking based on clustering is better than that by the inherent scoring function.

A strong correlation of the performance of the individual docking methods and their clustering behavior is demonstrated. On average, in ZDOCK, the top 2000 decoys are more tightly clustered than in the other three methods and near-native decoys are more densely distributed than non-native decoys. For MOLFIT and FTDOCK, the decoys are either divergently distributed or there is no difference in the number of structural neighbors between near-native and non-native decoys. These clustering behaviors are closely related to the performances of the docking methods.

We also tried to combine and cluster docking decoys generated from different methods. Compared with the clustering of ZDOCK decoys alone, only marginal improvement was obtained by combining filtered GRAMM-X and ZDOCK decoys. The result of other combinations is almost the same as that by clustering only ZDOCK decoys. Since the distribution of ZDOCK decoys is more convergent, they will dominate the clustering procedure if combined with other divergent decoys. Therefore, to obtain additional gain from the combination of meta-docking servers, different docking decoy sets with comparable quality and balanced structural distribution will be the key.

We further examine the docking quality of cluster centers and the conformations generated by superimposing the ligand onto the cluster centroids. The latter shows to be closer to the native structure than the cluster centers, with a TM-score increase by about 7%.

Finally, we compare clustering based on free-energy with clustering based on entropy. While differences occur in specific protein targets, there is not obvious difference in the overall result. A main reason is that there is no strong correlation between the energy/scoring function and the decoy quality in the majority of the pooled decoys.

ACKNOWLEDGMENTS

The calculations of this work have been performed on the KU ITTC clusters where the help from Drs. V. Frost and A. Hock are gratefully acknowledged. We also thank Drs. M. Eisenstein, M. J. Sternberg, A. Tovchigrechko, I. A. Vakser, and Z. Weng for allowing us to use MOLFIT, FTDOCK, GRAMM-X, and ZDOCK programs.

REFERENCES

1. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA 1992;89:2195–2199.
2. Chen R, Weng Z. A novel shape complementarity scoring function for protein-protein docking. Proteins 2003;51:397–408.
3. Chen R, Weng Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. Proteins 2002;47:281–294.
4. Gabb HA, Jackson RM, Sternberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 1997;272:106–120.
5. Tovchigrechko A, Vakser IA. Development and testing of an automated approach to protein docking. Proteins 2005;60:296–301.
6. Li L, Chen R, Weng Z. RDOCK: refinement of rigid-body protein docking predictions. Proteins 2003;53:693–707.
7. Jackson RM, Gabb HA, Sternberg MJ. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. J Mol Biol 1998;276:265–285.
8. Moont G, Gabb HA, Sternberg MJ. Use of pair potentials across protein interfaces in screening predicted docked complexes. Proteins 1999;35:364–373.

9. Camacho CJ, Weng Z, Vajda S, DeLisi C. Free energy landscapes of encounter complexes in protein-protein association. Biophys J 1999;76:1166–1178.
10. Shortle D, Simons KT, Baker D. Clustering of low-energy conformations near the native structures of small proteins. Proc Natl Acad Sci USA 1998;95:11158–11162.
11. Zhang Y, Skolnick J. SPICKER: a clustering approach to identify near-native protein folds. J Comput Chem 2004;25:865–871.
12. Vakser IA. Low-resolution docking: prediction of complexes for underdetermined structures. Biopolymers 1996;39:455–464.
13. Vakser IA, Matar OG, Lam CF. A systematic study of low-resolution recognition in protein–protein complexes. Proc Natl Acad Sci USA 1999;96:8477–8482.
14. Law DS, Ten Eyck LF, Katzenelson O, Tsigelny I, Roberts VA, Pique ME, Mitchell JC. Finding needles in haystacks: Reranking DOT results by using shape complementarity, cluster analysis, and biological information. Proteins 2003;52:33–40.
15. Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF. Protein docking using continuum electrostatics and geometric fit. Protein Eng 2001;14:105–113.
16. Camacho CJ, Gatchell DW. Successful discrimination of protein interactions. Proteins 2003;52:92–97.
17. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res 2004;32(Web Server issue):W96–W99.
18. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 2004;20:45–50.
19. Mendez R, Leplae R, Lensink MF, Wodak SJ. Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. Proteins 2005;60:150–169.
20. Kozakov D, Clodfelter KH, Vajda S, Camacho CJ. Optimal clustering for detecting near-native conformations in protein docking. Biophys J 2005;89:867–875.
21. Fischer D. 3D-SHOTGUN: a novel, cooperative, fold-recognition meta-predictor. Proteins 2003;51:434–441.
22. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 2003;19:1015–1018.
23. Lundstrom J, Rychlewski L, Bujnicki J, Elofsson A. Pcons: a neural-network-based consensus predictor that improves fold recognition. Protein Sci 2001;10:2354–2362.
24. Camacho CJ, Ma H, Champ PC. Scoring a diverse set of high-quality docked conformations: a metascore based on electrostatic and desolvation interactions. Proteins 2006;63:868–877.
25. Chen R, Mintseris J, Janin J, Weng Z. A protein-protein docking benchmark. Proteins 2003;52:88–91.
26. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003;52:80–87.
27. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. Nucleic Acids Res 2006;34(Web Server issue):W310–W314.
28. Heifetz A, Katchalski-Katzir E, Eisenstein M. Electrostatics in protein-protein docking. Protein Sci 2002;11:571–587.
29. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ. CAPRI: a critical assessment of predicted interactions. Proteins 2003;52:2–9.
30. Kabsch W. A discussion of the solution for the best rotation to relate two sets of vectors. Acta Crystallogr A 1978;34:827–828.
31. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins 2004;57:702–710.