

RDOCK: Refinement of Rigid-body Protein Docking Predictions

Li Li,¹ Rong Chen,¹ and Zhiping Weng^{1,2*}

¹Bioinformatics Program, Boston University, Boston, Massachusetts

²Department of Biomedical Engineering, Boston University, Boston, Massachusetts

ABSTRACT We present a simple and effective algorithm RDOCK for refining unbound predictions generated by a rigid-body docking algorithm ZDOCK, which has been developed earlier by our group. The main component of RDOCK is a three-stage energy minimization scheme, followed by the evaluation of electrostatic and desolvation energies. Ionic side chains are kept neutral in the first two stages of minimization, and reverted to their full charge states in the last stage of brief minimization. Without side chain conformational search or filtering/clustering of resulting structures, RDOCK represents the simplest approach toward refining unbound docking predictions. Despite its simplicity, RDOCK makes substantial improvement upon the top predictions by ZDOCK with all three scoring functions and the improvement is observed across all three categories of test cases in a large benchmark of 49 non-redundant unbound test cases. RDOCK makes the most powerful combination with ZDOCK2.1, which uses pairwise shape complementarity as the scoring function. Collectively, they rank a near-native structure as the number-one prediction for 18 test cases (37% of the benchmark), and within the top 4 predictions for 24 test cases (49% of the benchmark). To various degrees, funnel-like energy landscapes are observed for these 24 test cases. To the best of our knowledge, this is the first report of binding funnels starting from global searches for a broad range of test cases. These results are particularly exciting, given that we have not used any biological information that is specific to individual test cases and the whole process is entirely automated. Among three categories of test cases, the best results are seen for enzyme/inhibitor, with a near-native structure ranked as the number-one prediction for 48% test cases, and within the top 10 predictions for 78% test cases. RDOCK is freely available to academic users at <http://zlab.bu.edu/~rong/dock>. *Proteins* 2003;53:693–707.

© 2003 Wiley-Liss, Inc.

Key words: RDOCK; protein docking; docking refinement; energy minimization; neutralized ionic side chains; binding free energy

INTRODUCTION

Protein-protein docking entails the computational prediction of the 3-dimensional (3D) structure of a protein complex starting from the 3D structures of the receptor and the ligand. This problem has fascinated biophysicists and computational biologists for over two decades^{1,2} because of its biological importance: a complex structure necessarily reveals all structural details of the molecular interaction. It continues to receive considerable attention in the post-genomic era. Structural genomics aims to determine the 3D structure of every protein in the cell, and high throughput techniques such as yeast-two-hybrid and protein microarray will eventually identify all protein-protein interactions. Docking algorithms can be invaluable in the detailed investigation of these interactions, as well as in the design of novel pharmaceuticals.

Despite a large body of work, the protein-protein docking problem is far from being solved. When individually determined (or unbound) structures of the receptor and the ligand are used, near-native orientations may be scored poorly by a docking algorithm, since proteins undergo conformational changes upon complex formation, and most docking algorithms cannot capture these changes effectively. A divide-and-conquer strategy is widely accepted in the docking field,^{3,4} with initial-stage algorithms focused on retaining near-native structures (also called hits) in a reasonably short list of predictions^{5–14} and refinement algorithms aimed at ranking a hit at the top of the list.^{8,15–20}

The need of refinement algorithms for unbound docking became evident in the early 1990s when initial-stage rigid-body algorithms were powerful enough to explore the entire 6D translational and rotational space. In each case, high-scoring hits were found, among many false positives with similar or better scores.^{21–25} Shoichet and Kuntz performed a systematic investigation of refining unbound docking predictions.²² They tested three protease/inhibitor complexes, restraining the search to the known binding

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Grant sponsor: NSF; Grant numbers: DBI-0078194, DBI-0133834, DBI-0116574.

*Correspondence to: Zhiping Weng Boston University, 44 Cummington Street, Boston, MA 02215. E-mail: zhiping@bu.edu

Received 2 January 2003; Accepted 28 February 2003

sites of the proteases. They reported that most scoring functions, including surface area burial, solvation free energy, packing and mechanism-based filtering, could not distinguish hits from false positives, although total interaction energy and electrostatic interaction energy of the complex were somewhat better than other functions. Two groups responded to this challenge: (1) Jackson and Sternberg¹⁵ first optimized hydrogen-bonding networks and checked for unrealistically short distances between non-bonded atoms in the structures, and then reevaluated an energy function that included a continuum electrostatic free energy obtained from solving the Poisson-Boltzmann equation, a surface-area based hydrophobic free energy and side chain conformational entropy loss. They reported reliable discrimination between hits and false positives in all three test cases, but only when molecular surface was used as the descriptor of the hydrophobic effect, in contrast to solvent accessible surface. (2) Weng et al. performed energy minimization using CHARMM,²⁶ and then computed an effective binding free energy function containing the Coulombic electrostatic energy with a distance-dependent dielectric constant, a desolvation free energy based on the Atomic Solvation Parameter (ASP) model²⁷ and surface area based side chain entropy loss.²⁸ This effective binding free energy reliably differentiated hits from false positives. Furthermore, systematic search of interface side chain conformations using CONGEN²⁹ prior to energy minimization notably improved the discrimination.¹⁶

Sternberg and colleagues subsequently developed their own initial-stage docking algorithm FTDOCK.⁵ They performed a global search on six unbound enzyme/inhibitor and four antibody/antigen test cases. They then subjected the top predictions from the global search to filtering, which was similar to using known receptor binding site information. These filtered structures were later supplied to a refinement algorithm MultiDOCK developed by the same team.¹⁷ MultiDOCK included a two-step structure refinement procedure (side chain rotamer search and energy minimization), followed by the evaluation of an energy scheme that included vacuum electrostatics, the van der Waals (vdW) energy and a sophisticated desolvation energy based on a soft sphere Langevin dipole model. They also applied MultiDOCK to the previous Shoichet and Kuntz dataset and showed good discrimination. The results on the FTDOCK dataset were somewhat mixed: a hit was ranked in the top four predictions for four out of five enzyme/inhibitor test cases, but discrimination was seen only for two out of four antibody/antigen test cases. The same team consequently developed an empirical residue-residue pair potential as an intermediate step between FTDOCK and MultiDOCK. The combination of the pair potential and MultiDOCK was applied to the above FTDOCK dataset. It placed a hit within the top five predictions for all enzyme/inhibitor test cases, and within the top 40 predictions for all antibody/antigen test cases.¹⁸

Some researchers do not make the refinement step an independent method. Norel et al.³⁰ applied a computer vision-based docking algorithm to 19 unbound test cases

without any information of the binding site. The algorithm only used shape complementarity as the scoring function. The rank for the best hit ranged from 1 to 619, and was within the top 10 for five test cases. The authors attempted to refine the resulting predictions with buried surface area, unsatisfied buried charges, and hydrogen bonds, but observed no improvement.³⁰ In another method by Palma et al.,⁸ a Boolean-operation module first performed a global search, and the resulting top 1,000 structures were subjected to refinement by a function that incorporated geometric complementarity, electrostatics, desolvation, and a residue-residue pair potential. The relative contribution of each of these terms was determined by a neural network trained on roughly half of the test cases. The refinement step significantly improved the results. Excluding homodimers, there were 14 unbound test cases, among which eight were in the training set. For these test cases, the rank for the best hit ranged from 1 to 43. For the remaining six test cases, no hit was found within the top 1,000 predictions for three test cases, and the ranks for the best hits were 6, 13, and 50 for the remaining test cases.⁸ Totrov and Abagyan³¹ developed a two-step docking procedure: rigid-body docking followed by ligand side chain optimization using their Internal Coordinate Mechanics strategy. The method was applied to 24 unbound test cases, and the search was restricted to the known binding site of the receptor. The scoring functions for both stages of the procedure were quite elaborate, including smoothed vdW (with separate terms for hydrogen and heavy atoms), the Coulombic electrostatics with a distance dependent dielectric constant, a hydrogen bonding potential, the ASP-based solvation energy, and a hydrophobicity potential. The scoring function for the refinement stage also included the internal energy of the ligand. The coefficients for these energy terms were calibrated on the test set. Fernandez-Recio and coworkers¹⁴ reported significant improvement by the refinement, in terms of both side chain conformation and the rank of the best hit. A hit was ranked as the best prediction for seven test cases, and within the top 10 predictions for another seven test cases.

Camacho and Vajda²⁰ are entirely focused on the refinement stage of protein docking, and perform testing on decoy sets. In one approach, a two-step filtering procedure was first used to retain 100 structures for each test case, 500–1,000 steps of energy minimization were then performed on all structures, followed by the evaluation of a free energy expression that was the sum of the Coulombic electrostatics, the Atomic Contact Energy (ACE)-based desolvation free energy,³² and vdW. Successful discrimination was reported on two sets of decoys; both were mixtures of hits and false positives generated by a combination of superposition of bound docking, unbound docking, and random perturbation of superposed complex structures.¹⁹ Later on, the same group developed another method that sampled six rotational and translational degrees of freedom with vdW constraints.²⁰ Structures were minimized between consecutive rigid-body search steps. The method was tested on eight unbound test cases, and could consistently obtain a hit with ~ 2 Å ligand Root

Mean Square Deviation (RMSD) from refining <10 Å RMSD starting structures. The starting structures were hand picked from the outputs of an initial stage algorithm DOT.¹¹ Both methods seem promising, and it would be extremely interesting to test them on a complete set of global search predictions generated by an initial stage algorithm.

Our lab has invested substantial effort on developing a fully automated rigid-body docking algorithm ZDOCK, which focuses on the initial stage of docking unbound protein structures.^{13,33,34} Due to the diverse nature of protein interactions, the goal of ZDOCK is to retain at least one hit in the first 2,000 predictions for as many test cases as possible. To facilitate our own as well as others' docking efforts, we have developed a large benchmark of non-redundant test cases.³⁵ All of our ZDOCK developments have been extensively tested on this benchmark. ZDOCK performed competitively at the first CAPRI challenge, a community-wide blind test of docking algorithms.³⁶ Subsequent to CAPRI, we have been developing refinement methods to improve upon ZDOCK predictions.

Since ZDOCK is a rigid-body docking algorithm, it "tolerates" conformational changes by allowing small clashes between the receptor and the ligand structures, provided that the scores of false positives do not benefit significantly from this treatment. All variants of ZDOCK scoring functions have been designed with this consideration in mind. Thus, the most direct approach to refining ZDOCK predictions is to first relieve these structures from clashes and then re-rank them with more detailed scoring functions. Here we report such a simple and highly effective refinement method RDOCK. We perform energy minimization on the top 2,000 ZDOCK predictions per test case, with special treatment for ionic side chains. All structures that still have vdW clashes after minimization are discarded. We then compute electrostatics and desolvation energies for the remaining structures and rank them according to the sum of these two energy terms.

RDOCK improves ZDOCK predictions generated using all three scoring functions that have been recently developed in our laboratory.³⁴ It is particularly effective on refining the predictions generated with the pair-wise shape complementarity (PSC) scoring function.³³ At a 6° rotational sampling density, PSC retains at least one hit in the top 2,000 predictions for 42 out of 49 test cases in the benchmark.³³ Among these 42 test cases, RDOCK ranks a hit as the number one prediction for 18 test cases, among the top 10 predictions for 28 test cases and among the top 100 predictions for 38 test cases. This is far superior to the best performance of ZDOCK, which is achieved with a scoring function that combines PSC with electrostatics and desolvation: 7, 18, and 27 test cases with a hit ranked in the top 1, 10, and 100 predictions, respectively. Furthermore, for 10 test cases, RDOCK reports hits for all top three predictions, and thus identifies the binding mode unambiguously. We observe well-formed binding funnels for these test cases. To the best of our knowledge, this is the first demonstration of binding funnels for a large number of test cases starting from global searches. RDOCK

performs the best for enzyme/inhibitor test cases, and a hit is ranked as the top 1 and within the top 10 for 48% and 78% test cases, respectively.

It is difficult to make an exact comparison of RDOCK with previous methods, since most previous work used receptor binding site information, most of them were tested primarily on enzyme/inhibitor, and the definitions of hits could also be slightly different. Nevertheless, considering that our benchmark is the largest and most diverse test set to date and our method is rapid to compute and entirely automated, we believe that our results demonstrate an improvement over previous work on docking refinement.

METHODS

Energy Minimization of ZDOCK Structures

We start with the top 2,000 predictions for each test case, generated using ZDOCK at a 6° rotational sampling density with each of the following three scoring functions: pairwise shape complementarity (PSC), the combination of PSC with desolvation (PSC+DE; abbreviated PD), the combination of PSC with desolvation and electrostatics (PSC+DE+ELEC; abbreviated PDE). We previously assigned different versions of ZDOCK to reflect different scoring functions: ZDOCK2.1 for PSC, ZDOCK2.2 for PD, and ZDOCK2.3 for PDE.³⁴ In order to test RDOCK, we used version 0.0 of the benchmark developed in our lab,³⁵ which contains 23 enzyme/inhibitor, 16 antibody/antigen, and 10 other types of test cases. ZDOCK with PSC, PD, or PDE retains at least one hit in the top 2,000 predictions for 42, 40, and 44 test cases, respectively. In this report, we restrict ourselves to these test cases. Hits are predictions with RMSD below 2.5 Å after superposition. Superposition and RMSD calculation only involve the C_α atoms of interface residues, which are receptor (or ligand) residues with at least one atom within 10 Å of any atoms of the ligand (or receptor). PSC, PD, and PDE have been compared extensively in a previous study,³⁴ and the above data sets are available from our website (<http://zlab.bu.edu/~rong/dock/>).

Each ZDOCK prediction, which is a protein complex structure, is subjected to 130 steps of Adopted Basis Newton-Raphson (ABNR) energy minimization using the molecular mechanics software CHARMM²⁶ in three stages:

1. *Removing clashes.* We first perform 50 steps of minimization of vdW plus internal (bond and angle) energies. All atoms are free to move. We have discovered that this scheme is extremely effective in removing small clashes. Most structures have favorable vdW energies after minimization.
2. *Optimizing polar interactions.* We then minimize the total energy (the sum of Coulombic electrostatics, vdW, and internal energies) for 60 steps. Non-hydrogen atoms are harmonically restrained to their starting positions with a mass-weighted force constant of 20Kcal/(mol*Å) during the minimization. We keep all ionic side chains (Arg, Lys, Glu, and Asp) in their neutral states during this stage of minimization. A distance depen-

dent dielectric constant $4r$ (r being the distance) is used throughout this study. “Turning off” charges prevents them from dominating the total energy, and thus allows polar interactions to be optimized first.

3. *Optimizing charge interactions.* We minimize the total energy for another 20 steps without any constraint. All charged residues are in their charged states during this stage of minimization. The goal here is to optimize charge interactions. Since the minimization is done in vacuum (although the distance dependent dielectric constant partially accounts for desolvation), too many steps of minimization would inevitably lead to overestimation of electrostatics, and therefore should be avoided.

We use the CHARMM 19 potential with extended atom types and polar hydrogens throughout the study. It is important to divide the energy minimization into three stages with ionic side chains set to neutral in the first two stages. Nonetheless, the exact number of steps in each stage does not make a large impact on the performance of RDOCK.

Re-ranking of Minimized Structures

We assume that if a complex structure still has clashes after the energy minimization, it cannot be a hit. Therefore, we discard all structures with vdW energies higher than 10 kcal/mol after minimization. The electrostatics and desolvation energies of the remaining structures are computed. The electrostatics energy (ΔE_{elec}) is computed using CHARMM, with a dielectric constant $4r$ and all ionic side chains in their full charge states. We use the Atomic Contact Energy (ACE)³² to estimate desolvation. ACE is defined as the free energy difference between breaking two protein-atom/water contacts and forming a protein-atom/protein-atom contact and a water-water contact. The total desolvation score of a protein complex (ΔG_{ACE}) is simply the sum of the ACE scores of all receptor-ligand atom pairs within a distance cutoff of 6 Å.

The scoring function of RDOCK is the sum of desolvation and electrostatics:

$$\Delta G_{binding} = \Delta G_{ACE} + \beta \times \Delta E_{elec} \quad (1)$$

where β is a scaling factor, set to 0.9 in this study. For ZDOCK, we combined electrostatics with desolvation and the coefficient for electrostatics β was set to 1/7.³⁴ In RDOCK, since the electrostatics is rather accurate for structures free of vdW clashes, the optimal value of β is very close to 1. In fact, $\beta = 1$ leads to only slightly inferior results.

Performance Evaluation

The performance of RDOCK and ZDOCK is evaluated using success rate, as defined previously.³³ Given the number of predictions being evaluated for each test case (N_p), the success rate is the percentage of test cases in the benchmark for which at least one hit has been found. Since RDOCK is a refinement algorithm, we are only interested in the success rates at $N_p \leq 100$, ideally at $N_p = 1$. The minimization procedure in RDOCK only makes small local

changes to a structure, almost never large enough to turn a non-hit into a hit. Thus, RDOCK cannot improve the test cases for which ZDOCK hasn't retained any hits in the top 2,000 predictions. However, in order to be consistent across the publications from our laboratory, we have included all such test cases in success rate calculations.

RESULTS

RDOCK's Overall Performance on the Benchmark

The number of hits, the rank of the first hit, and the RMSD of the first hit for each complex according to ZDOCK(PSC)+RDOCK, which stands for ZDOCK with PSC as the scoring function followed by RDOCK refinement, are listed in Table I. In Table I, we have also reproduced ZDOCK results with PSC and PDE as the scoring functions from our previous work to facilitate comparison.^{33,34} The reason that the “Hits” column for ZDOCK(PSC)+RDOCK does not match up with the “Hits” column for ZDOCK(PSC) is that the vdW interaction energies for some hits are still higher than 10 kcal/mol after minimization and these hits are discarded. In most cases, all hits survive the vdW cutoff. For five enzyme-inhibitor test cases (4HTC being the most extreme, and 1TAB, 1ACB, and 1CHO being the others), the majority of hits are lost, but the ranks of the remaining hits are still very high. The percentage of hits out of all predictions remains largely unchanged after the vdW filtering. Nonetheless, false positives with unfavorable vdW energies tend to have favorable electrostatics and desolvation, and, thus, filtering these out improves RDOCK performance. Success rates are improved by ~4% throughout $N_p \leq 100$ (data not shown).

It is apparent from Table I that RDOCK substantially improves the best rank of hits. Out of the 22 enzyme/inhibitor test cases for which PSC retains at least one hit in the top 2,000 predictions, RDOCK improves 14 test cases, doesn't affect 5 test cases (all of which have a hit ranked as the number one prediction), and worsens 3 test cases slightly (with ranks increased by 4, 5, and 10, respectively). When ZDOCK(PSC)+RDOCK is compared with ZDOCK(PDE), the former achieves a better rank for the first hit for 15 test cases, worse for 4 test cases, and they tie for 4 test cases (for all of these, a hit is ranked as the number-one prediction). Out of the 14 antibody/antigen test cases for which PSC retains at least one hit in the top 2,000 predictions, RDOCK improves the rank of the first hit for 11 test cases, and worsens the rank for the remaining three test cases. When ZDOCK(PSC)+RDOCK is compared with ZDOCK(PDE), the former achieves better rank for the first hit in 11 test cases. Among the 6 test cases in the “others” category for which ZDOCK(PSC) retains at least one hit in the top 2,000 predictions, RDOCK improves the ranks for 4 of them, worsens the rank for only one of them, and does not alter the rank for the last one (with a hit ranked as the number-one prediction). If we compare ZDOCK(PSC)+RDOCK with ZDOCK(PDE), the former achieves better rank for 4 test cases, worse for 1, and the two methods tie for 1 test case (with a hit ranked as the number-one prediction).

TABLE I. Docking Performance with 6° Rotational Sampling Interval

Complex ^a	ZDOCK(PSC)+RDOCK			ZDOCK(PSC)			ZDOCK(PDE)		
	Hits ^b	Rank ^c	RMSD ^d	Hits ^b	Rank ^c	RMSD ^d	Hits ^b	Rank ^c	RMSD ^d
1CGI	52	8	2.24	54	4	2.41	77	4	2.41
1CHO	37	1	1.28	66	1	1.67	99	3	1.57
2PTC	2	2	1.12	2	1,655	2.11	48	193	1.83
1TGS	101	8	2.03	107	3	1.83	109	3	2.22
2SNI	0	—	—	0	—	—	1	1,262	2.22
2SIC	23	1	1.17	24	241	1.68	52	11	2.37
1CSE	3	1	1.17	3	1,537	1.17	29	198	2.20
2KAI	3	141	2.36	3	1,399	2.47	16	388	1.61
1BRC	15	3	2.41	16	173	1.58	54	24	2.32
1ACB	20	1	1.86	38	25	1.33	93	18	1.33
1BRS	32	13	1.23	34	61	2.23	21	65	2.13
1JTG	60	13	1.54	69	3	1.52	82	1	1.52
1MAH	6	1	0.91	6	849	1.49	28	24	1.29
1UGH	4	1	2.08	4	305	2.37	20	8	2.25
1DFJ	10	1	2.48	15	37	2.48	51	1	2.48
1FSS	5	42	1.52	5	731	1.52	15	50	1.52
1AVW	24	2	2.00	28	45	2.07	52	3	2.07
1PPE ^e	270	1	0.69	272	1	0.59	393	1	0.90
1TAB ^e	21	10	0.76	47	65	1.21	50	79	1.21
1UDI ^e	15	3	1.06	16	31	1.19	35	5	1.19
1STF ^e	42	1	1.04	42	1	0.88	83	1	0.88
2TEC ^e	77	1	0.83	77	1	0.39	185	1	0.76
4HTC ^e	2	1	1.46	54	1	1.79	57	3	2.46
1MLC	3	2	1.65	3	1,106	2.10	17	128	1.65
1WEJ	4	4	0.91	4	1,396	1.07	22	183	1.04
1AHW	28	1	1.61	28	26	1.57	67	7	1.82
1DQJ	1	952	2.45	1	1,341	2.45	0	—	—
1BVK	2	1314	1.64	2	974	1.89	2	821	2.34
1FBI ^e	1	53	2.15	2	1,786	2.15	5	642	2.03
2JEL ^e	57	301	1.70	62	112	1.82	35	233	1.46
1BQL ^e	15	1	1.18	16	172	1.18	70	13	1.07
1JHL ^e	15	41	0.88	15	404	1.46	12	333	1.37
1NCA ^e	50	8	0.83	55	2	0.85	67	1	1.06
1NMB ^e	6	1	1.11	6	693	1.10	9	135	0.98
1MEL ^e	51	1	1.37	52	12	1.19	71	3	1.19
2VIR ^e	3	80	1.19	3	476	1.03	3	1,101	1.03
1EO8 ^e	0	—	—	0	—	—	2	1,497	0.96
1QFU ^e	10	29	0.95	10	407	1.17	18	388	1.14
1IAI ^e	0	—	—	0	—	—	3	997	1.70
2PCC	0	—	—	0	—	—	0	—	—
1WQ1	25	16	1.91	26	5	1.37	54	15	1.31
1AVZ	0	—	—	0	—	—	0	—	—
1MDA	0	—	—	0	—	—	0	—	—
1IGC ^e	6	21	1.18	6	22	1.20	3	153	1.20
1ATN ^e	1	1	0.80	1	360	0.80	24	7	0.80
1GLA ^e	0	—	—	0	—	—	0	—	—
1SPB ^e	68	1	0.70	75	1	0.61	112	1	0.61
2BTF ^e	13	1	0.95	13	32	0.68	35	2	0.95
1A0O ^e	2	11	2.46	2	833	2.46	4	284	2.45

^a4-letter Protein Data Bank (PDB) code for the crystal complex of a test case.

^bOnly the first 2,000 predictions are evaluated. Hits are defined as docked structures with interface C_a RMSD ≤ 2.5 Å from the crystal complex.

^cRank of the best ranked hit. “—” indicates that no hit was found in the top 2,000 predictions.

^dInterface C_a RMSD for the best ranked hit. “—” indicates that no hit was found in the top 2,000 predictions.

^eUnbound/bound test cases.

Figure 1(a) compares the success rates of three ZDOCK scoring functions on the entire benchmark (PSC, PD, and PDE), with and without RDOCK refinement. In general, RDOCK improves the predictions of all of these scoring

functions (solid curves being higher than dashed curves over the whole range of N_p). However, the improvement is not uniform. As indicated in our previous publication,³⁴ PD outperforms PSC, and PDE outperforms both PD and

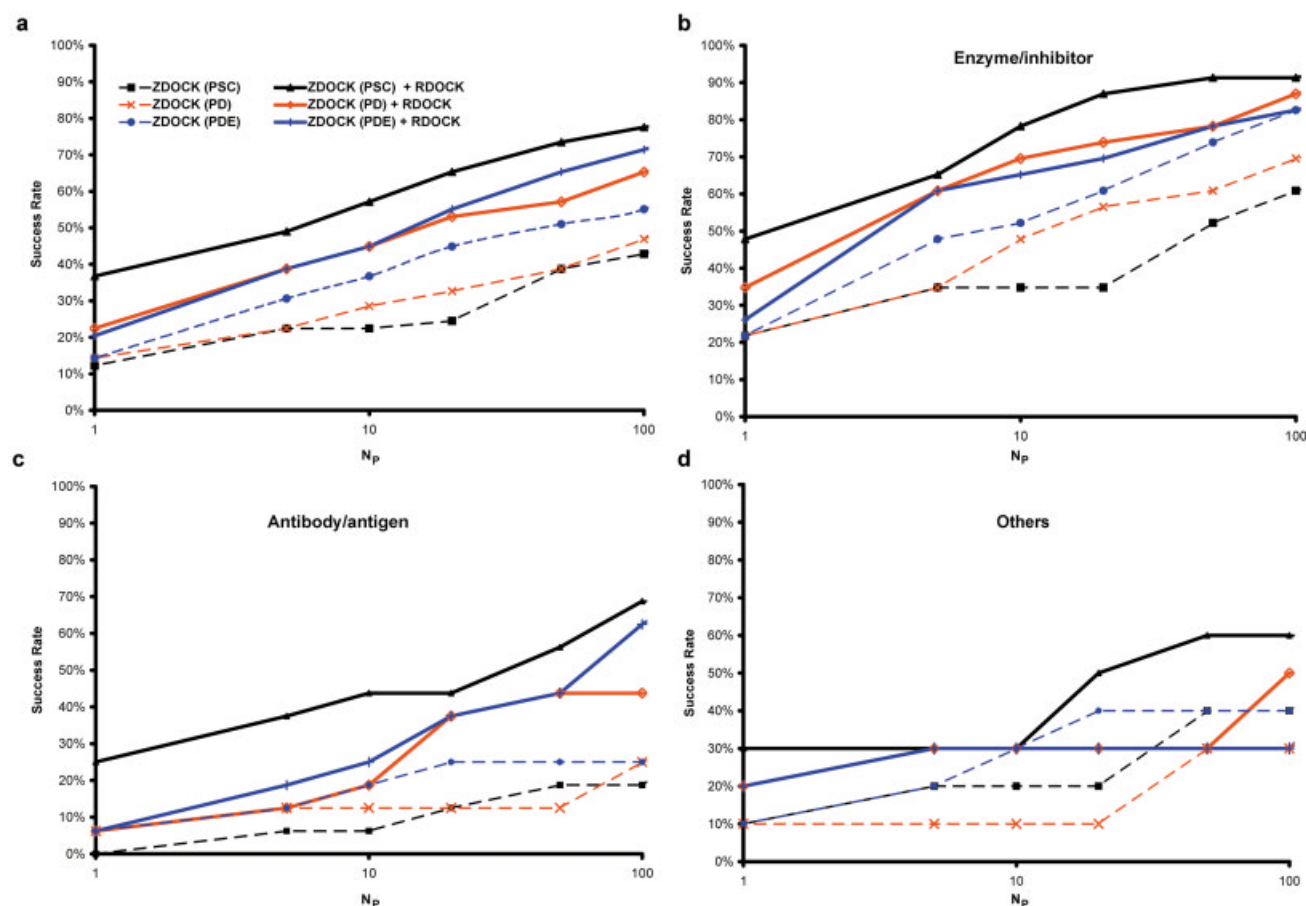


Fig. 1. Success rates of ZDOCK with (solid lines) or without (dashed lines) RDOCK refinement. We compare three ZDOCK scoring functions: PSC (black), PD (red), PDE (blue). We consider the performance on (a) the entire benchmark, as well as in the (b) Enzyme/inhibitor, (c) Antibody/antigen, and (d) Others categories. The success rate is defined in Methods. It reflects the average ability of a method in ranking the first hit. Hits are predictions with interface C_α RMSD < 2.5 Å (see Methods for more details).

PSC. However, the success rates of RDOCK on the predictions made by these three scoring functions do not remain in this order. ZDOCK(PDE)+RDOCK is in general better than ZDOCK(PD)+RDOCK. Interestingly, ZDOCK(PSC)+RDOCK outperforms both ZDOCK(PD)+RDOCK and ZDOCK(PDE)+RDOCK over the entire range of $N_p \leq 100$, with the improvement being most pronounced at $N_p = 1$.

The discrepancy in RDOCK's improvement can be better understood if we analyze the success rates according to individual categories [Fig. 1(b–d)]. For enzyme/inhibitor [Fig. 1(b)], RDOCK improves ZDOCK(PSC) so much that even though ZDOCK(PSC) is much worse than ZDOCK(PDE), ZDOCK(PSC)+RDOCK is much better than both ZDOCK(PD) and ZDOCK(PDE). The success rate for ZDOCK(PSC)+RDOCK is 78% at $N_p = 10$ and 87% at $N_p = 20$. Such accuracy is particularly impressive considering that no biological information is used and the whole process is entirely automatic. Interestingly, ZDOCK(PD)+RDOCK outperforms ZDOCK(PDE)+RDOCK at most N_p values, even though ZDOCK(PDE) outperforms ZDOCK(PD) by a great deal.

For antibody/antigen test cases [Fig. 1(c)], RDOCK also improves upon ZDOCK with all three scoring functions.

ZDOCK(PDE) outperforms ZDOCK(PD), and accordingly ZDOCK(PDE)+RDOCK outperforms ZDOCK(PD)+RDOCK. Similar to enzyme/inhibitor test cases, RDOCK improves upon ZDOCK(PSC) to the greatest extent, so that ZDOCK(PSC)+RDOCK outperforms both ZDOCK(PDE)+RDOCK and ZDOCK(PD)+RDOCK. The improvement ($\sim 19\%$) is especially evident for $N_p \leq 10$. The success rate of ZDOCK(PSC)+RDOCK is 44% at $N_p = 10$; although not as high as that for enzyme/inhibitor, it is much higher than what has been reported in the past for this category of test cases.

For four out of a total of ten test cases in the “others” category (2PCC, 1AVZ, 1MDA, and 1GLA), ZDOCK with any of the three scoring functions did not retain any hits within the top 2,000 predictions, due to a number of reasons including backbone conformational changes, low binding affinities, and ZDOCK's tendency of identifying surface pockets.³⁴ As mentioned before, RDOCK is not able to improve upon these test cases. Thus, the success rate of RDOCK is at most 60% for this category. Figure 1(d) indicates that RDOCK improves upon ZDOCK(PDE) and ZDOCK(PD) only at small N_p values ($N_p < 10$ for PDE and $N_p < 50$ for PD). Yet, RDOCK improves upon

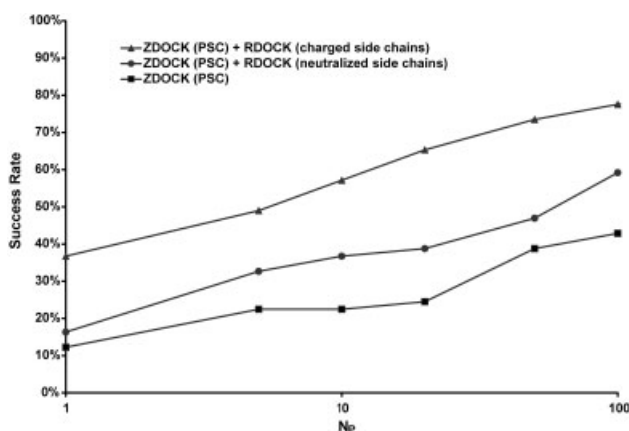


Fig. 2. Comparison of the success rates: ZDOCK(PSC) (squares), ZDOCK(PSC) followed by the first two stages of RDOCK energy minimization with all ionic side chains kept neutral (circles), and ZDOCK(PSC) followed by all three stages of RDOCK energy minimization, with ionic side chains reverted to their full charge states in the last stage of minimization (triangles).

ZDOCK(PSC) throughout the entire N_p range, and ZDOCK(PSC)+RDOCK is clearly the best performer. Except for the four aforementioned test cases for which ZDOCK does not retain any hits, RDOCK improves the first rank for the remaining 6 test cases to 21 or better, and thus achieves its maximal success rate of 60% at $N_p = 21$.

Impact of Neutral Side Chains During Minimization

Our strategy of using all neutralized side chains for most minimization steps and performing a brief minimization with full charges at the end is crucial to RDOCK's performance. If we keep all charged residues in their full charge states throughout the minimization, RDOCK provides no improvement over ZDOCK with any scoring function (data not shown). Figure 2 indicates the impact of our special treatment of ionic side chains. At the end of the second stage of minimization (first 50 steps without electrostatics and the next 60 steps with electrostatics but with neutralized side chains; see Methods), RDOCK achieves an improvement of 5–10% in success rate over ZDOCK(PSC). The last 20 steps of minimization with full charges further improve the results drastically, with a 20–25% increase in success rate at all N_p values.

Impact of Rotational Sampling Density

Previously, we carefully investigated the impact of rotational sampling interval (Δ) in ZDOCK.^{33,34} We discovered that ZDOCK achieved higher success rates at larger Δ values (such as 15°) for all scoring functions; however, for some test cases we observed under-sampling.³⁴ In addition to being more consistent, smaller Δ (such as 6°) led to many more hits. The advantage of a finer sampling becomes evident in RDOCK. Figure 3 compares the two Δ values, with or without RDOCK. Even though in general ZDOCK(PSC) achieves better success rates for $\Delta = 15^\circ$, ZDOCK(PSC)+RDOCK performs considerably better for

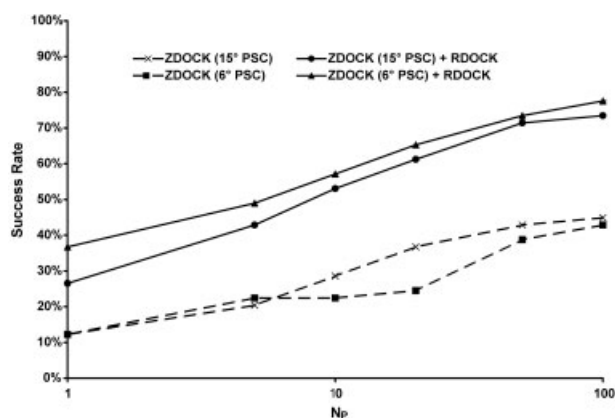


Fig. 3. The success rates of ZDOCK(PSC) with (solid lines) and without (dashed lines) RDOCK refinement are compared between two rotational sampling densities: 6° (triangles and squares) and 15° (circles and crosses).

$\Delta = 6^\circ$, at all $N_p \leq 100$, with the largest improvement of 10% seen at $N_p = 1$.

To better understand this phenomenon, we compare the best-ranked hits at $\Delta = 6^\circ$ before and after RDOCK. For merely 8 test cases, the best-ranked hit by PSC remains the best according to RDOCK. In all other test cases, RDOCK improves the rank of the first hit by promoting another hit that was somewhat down the list. Therefore, finer rotational sampling in an initial stage algorithm is clearly beneficial for the refinement, since more hits are generated, and there is a better chance that the energy of at least one of these hits will be substantially minimized after refinement.

RMSD of the Best Hit

In Table I, we provide the RMSDs of the best-ranked hits according to three methods. Since no currently available scoring function can resolve RMSD differences below 0.2 \AA , we use this as the cutoff for deeming two predictions different. If we compare ZDOCK(PSC)+RDOCK with ZDOCK(PSC), the former picks a lower RMSD structure as the best ranked hit for 12 test cases and the latter for 5 test cases. If we compare ZDOCK(PSC)+RDOCK with ZDOCK(PDE), the former picks a lower RMSD hit for 14 test cases and the latter for 4 test cases. Both differences are significant with a P value of < 0.08 according to the sign test.³⁷ Thus, we conclude that RDOCK can correctly rank structures even at the RMSD range of below 2.5 \AA .

Score-RMSD Relationship for Individual Test Cases

Since RDOCK is a refinement algorithm, we are naturally interested in knowing if there is a correlation between energies computed using Equation (1) and the RMSDs of predicted structures from the crystal complex. We divide the 42 test cases into five groups according to RDOCK's performance, and discuss them in order.

The first group contains 10 test cases, for which ZDOCK(PSC)+RDOCK reports hits for all top three predictions. Thus, the binding modes of these test cases have been unambiguously determined by the combination of

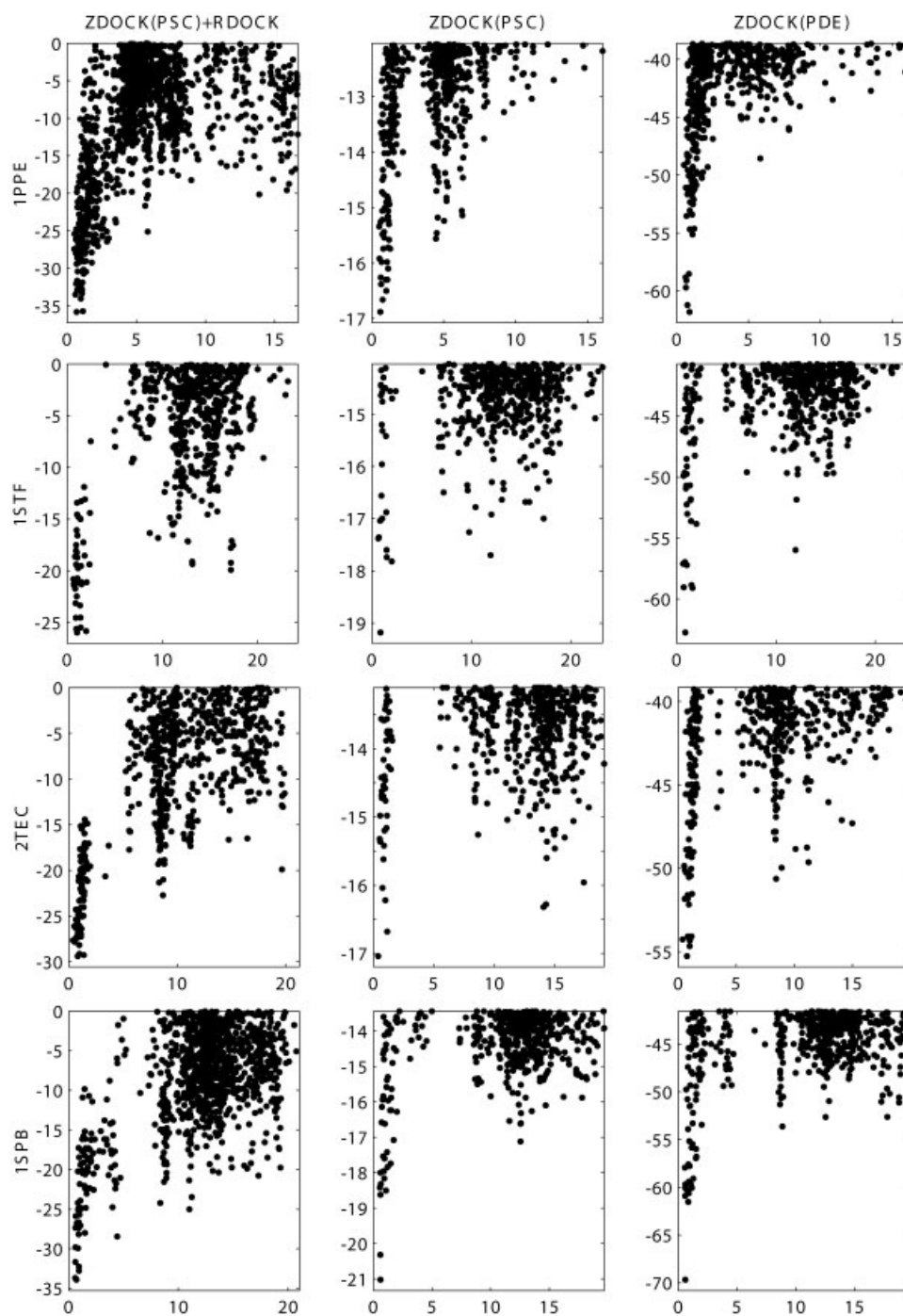


Fig. 4. Binding energy (ordinate) vs. RMSD (abscissa) graphs are compared among ZDOCK(PSC)+RDOCK (left), ZDOCK(PSC) (middle) and ZDOCK(PDE) (right) for the first group of test cases (10 in total). For these test cases, ZDOCK(PSC)+RDOCK reports hits for all top three predictions. Only predictions with favorable binding energy (lower than 0) are shown for ZDOCK(PSC)+RDOCK, and only the first 500 predictions are shown for ZDOCK(PSC) and ZDOCK(PDE), similarly for Figures 5 and 6.

ZDOCK and RDOCK. We plot energy vs. RMSD for these test cases in Figure 4. The ZDOCK(PSC)+RDOCK graphs exhibit the classic funnel-like energy landscape: a cluster of low RMSD structures (hits) dominates the lowest energy region, and as structures move toward higher RMSD, their energies increase. For 1PPE, 1STF, 2TEC, and 1SPB, all

three methods produce perfect energy-RMSD graphs. Even for these test cases, the improvement of RDOCK upon ZDOCK(PSC) is evident, and ZDOCK(PSC)+RDOCK typically outperforms ZDOCK(PDE). For 1CHO, ZDOCK(PSC) also achieves good discrimination. So does ZDOCK(PDE) for 1DFJ. For the remaining four test cases (2SIC, 1ACB,

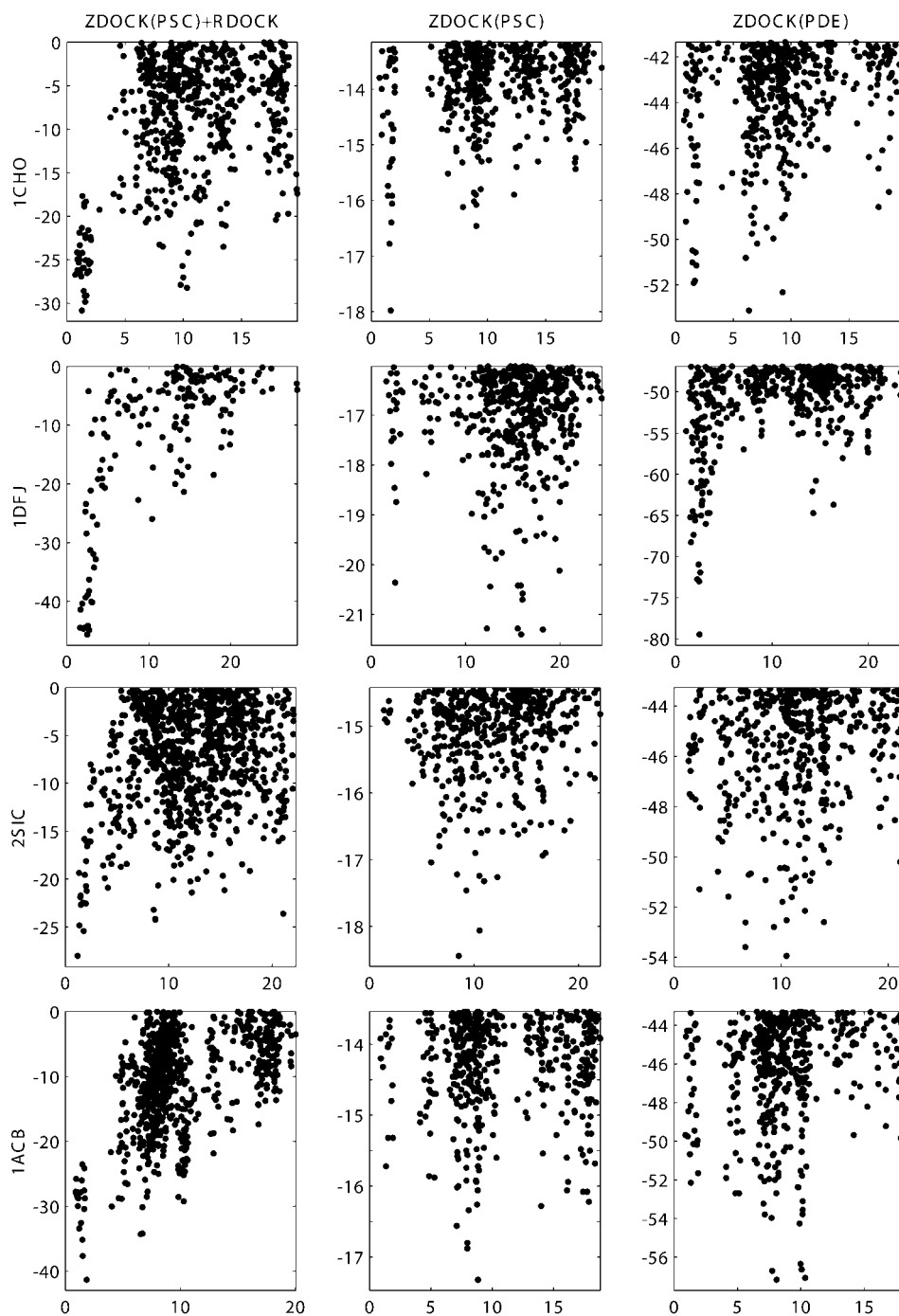


Figure 4. (Continued.)

1BQL, and 2BTF), only ZDOCK(PSC)+RDOCK exhibits perfect energy-RMSD relationships, representing a drastic improvement over both ZDOCK(PSC) and ZDOCK(PDE).

There are 8 test cases in the second group (1CSE, 1MAH, 1UGH, 4HTC, 1AHW, 1NMB, 1MEL, and 1ATN), and for all of them ZDOCK(PSC)+RDOCK ranks a hit as the number-one prediction. However, a false positive is ranked within the top three predictions. For 4HTC, ZDOCK(PSC) retains 54 hits within the top 2,000. Only 2 of the hits survive the vdW cutoff of RDOCK. They are

ranked as the top two predictions by RDOCK. For 1ATN, ZDOCK(PSC) only retains one hit within the top 2,000, ranked at 360. After RDOCK refinement, this hit is ranked number one. For 1CSE, ZDOCK(PSC) only retains 3 hits within the top 2,000, with the highest rank of 1,537. RDOCK improves all three hits to the top 10. The situation is very similar for 1MAH, 1UGH, 1NMB, and 1AHW: RDOCK improves the ranks of all hits retained by ZDOCK(PSC) to within the top 100. For 1MEL, ZDOCK(PSC) provides 52 hits in the top 2,000, and

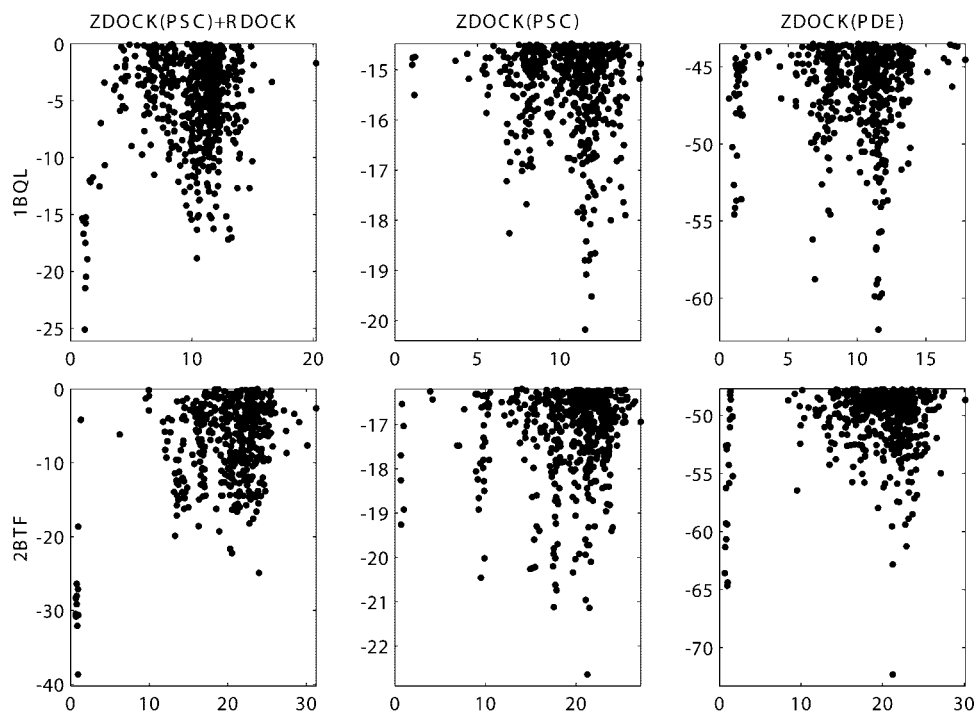


Figure 4. (Continued.)

RDOCK increases the number of hits within the top 100 from 6 to 41. The energy-RMSD graphs for these test cases are included in Figure 5, and ZDOCK(PSC)+RDOCK demonstrates marked improvement over both ZDOCK(PSC) and ZDOCK(PDE) for all cases.

The third group includes 11 test cases: 2PTC, 1BRC, 1BRS, 1AVW, 1TAB, 1UDI, 1MLC, 1WEJ, 1FBI, 1QFU, and 1A00. There is no doubt about the improvement of ZDOCK(PSC)+RDOCK over ZDOCK(PSC) and ZDOCK(PDE), both in terms of the number of hits in the top 100 and the rank of the first hit. For six of these test cases, ZDOCK(PSC)+RDOCK ranks a hit among the top 4 predictions. We show their energy-RMSD graphs in Figure 6. Most of the ZDOCK(PSC)+RDOCK graphs show good discrimination, except for a few scattered false positives. Notably, the graph for 1AVW is perfect except for one false positive at RMSD of 11Å. For 2PTC, 1MLC, and 1WEJ, the ranks of the best hits according to PSC are very poor (1655, 1106, and 1396). They are improved to 2, 2, and 4 by RDOCK, which are also far superior to the ranks according to ZDOCK(PDE): 193, 128, and 183, respectively.

The fourth group has five test cases (1FSS, 1JHL, 2VIR, 1DQJ, and 1IGC). The improvement made by RDOCK over ZDOCK(PSC) is evident for these test cases. Nevertheless, the final results of ZDOCK(PSC)+RDOCK are less impressive than those for the first three groups of test cases. For example, ZDOCK(PSC) provides 15 hits in the top 2,000, with the best one ranked at 404. RDOCK brings one hit to the top 100, with the rank of 41. This is compared with the rank of 303 according to ZDOCK(PDE).

The last group includes the remaining eight test cases. RDOCK makes little improvement or worsens the results somewhat. For none of these test cases, RDOCK worsens

the result significantly. The most severe case is 1NCA. The rank of the best hit is 1, 2, and 8 for ZDOCK(PSC), ZDOCK(PDE), and ZDOCK(PSC)+RDOCK respectively. Moreover, all top three predictions are hits according to ZDOCK(PDE).

DISCUSSION

We present a simple and effective refinement algorithm, RDOCK, for protein-protein docking. The main component of RDOCK is a three-stage energy minimization scheme, followed by the evaluation of electrostatic and desolvation energies. We do not perform any explicit side chain or backbone searches, or refinement of the relative orientation between the receptor and the ligand. We do not cluster resulting structures either. Thus, RDOCK represents the simplest approach toward refining protein-protein docking predictions. Side chain conformational search has been shown by several groups to significantly improve protein docking.^{14,16,17,38} We expect that it will have a similar impact on RDOCK. Despite its simplicity, RDOCK makes substantial improvement upon the top predictions made by an initial stage rigid-body docking algorithm ZDOCK³⁴ with all three scoring functions: PSC, PD, and PDE. The improvement is observed across all three categories of test cases in a large benchmark.³⁵

RDOCK makes the most powerful combination with ZDOCK2.1, which uses PSC as the scoring function. Collectively, they rank a hit as the number-one prediction for 18 test cases (37% of the benchmark), and within the top 4 predictions for 24 test cases (49% of the benchmark). To various degrees, funnel-like energy landscapes are observed for these 24 test cases (Figs. 4–6). To the best of our knowledge, this is the first report of binding funnels

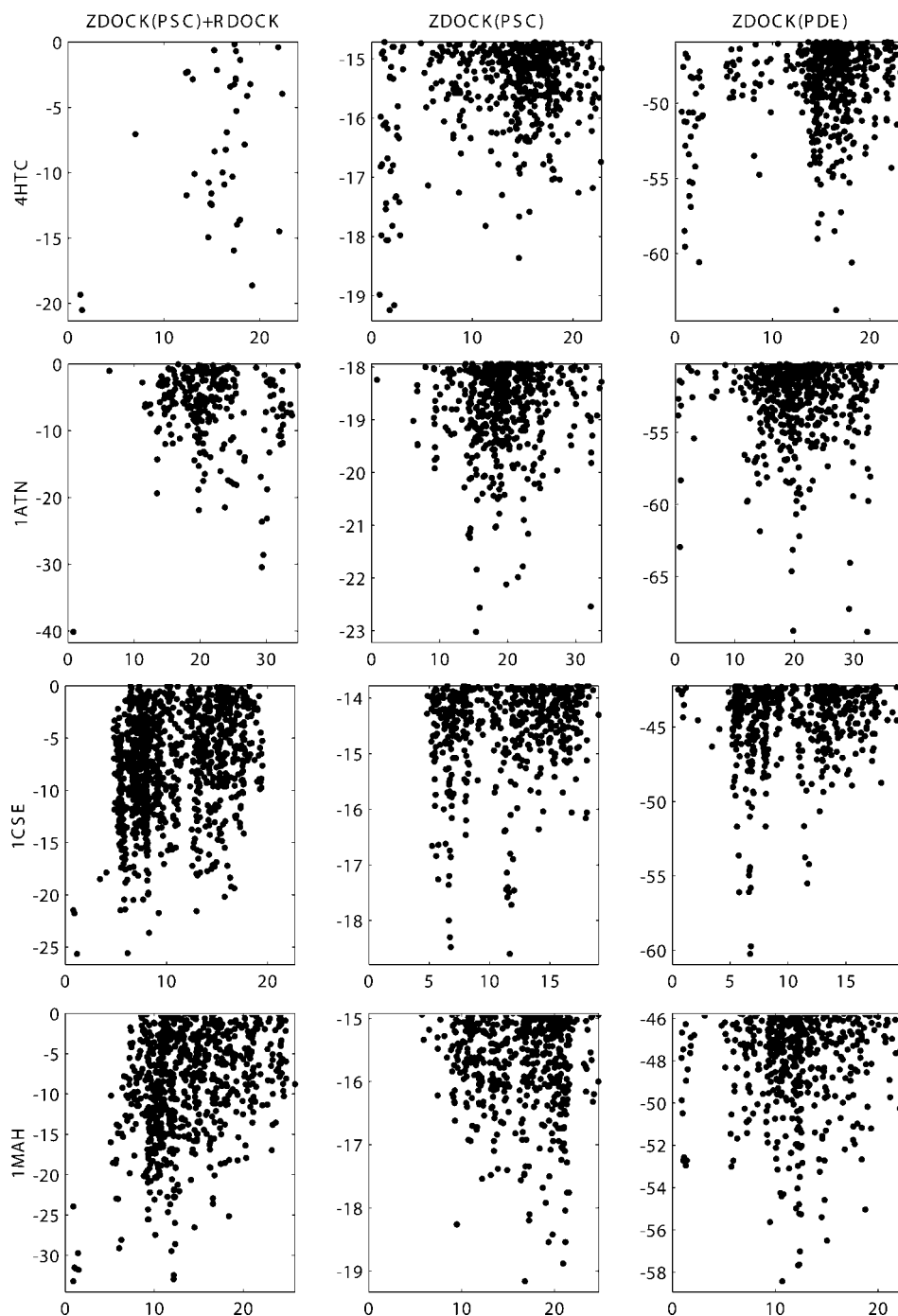


Fig. 5. Binding energy (ordinate) vs. RMSD (abscissa) graphs are compared among ZDOCK(PSC)+RDOCK (left), ZDOCK(PSC) (middle), and ZDOCK(PDE) (right) for the second group of test cases (8 in total). ZDOCK(PSC)+RDOCK ranks a hit as the number-one prediction for these test cases.

starting from global searches for a broad range of test cases. These results are particularly exciting, given that (1) we have not used any biological information that is specific to individual test cases and the whole process is entirely automated, and (2) most previous unbound docking algorithms have been tested on substantially fewer than 24 unbound cases.^{5,7,8,11,15–20,22,30}

Although the ideal number of predictions per test case is 1, it is possible to use biological information to process 20 predictions generated by an automated method to produce 1–5 predictions for experimental testing. At such a prediction level, ZDOCK(PSC)+RDOCK succeeds for 32 test cases (65% of the benchmark). Note that RDOCK can only operate on 42 test cases for which ZDOCK(PSC) produces

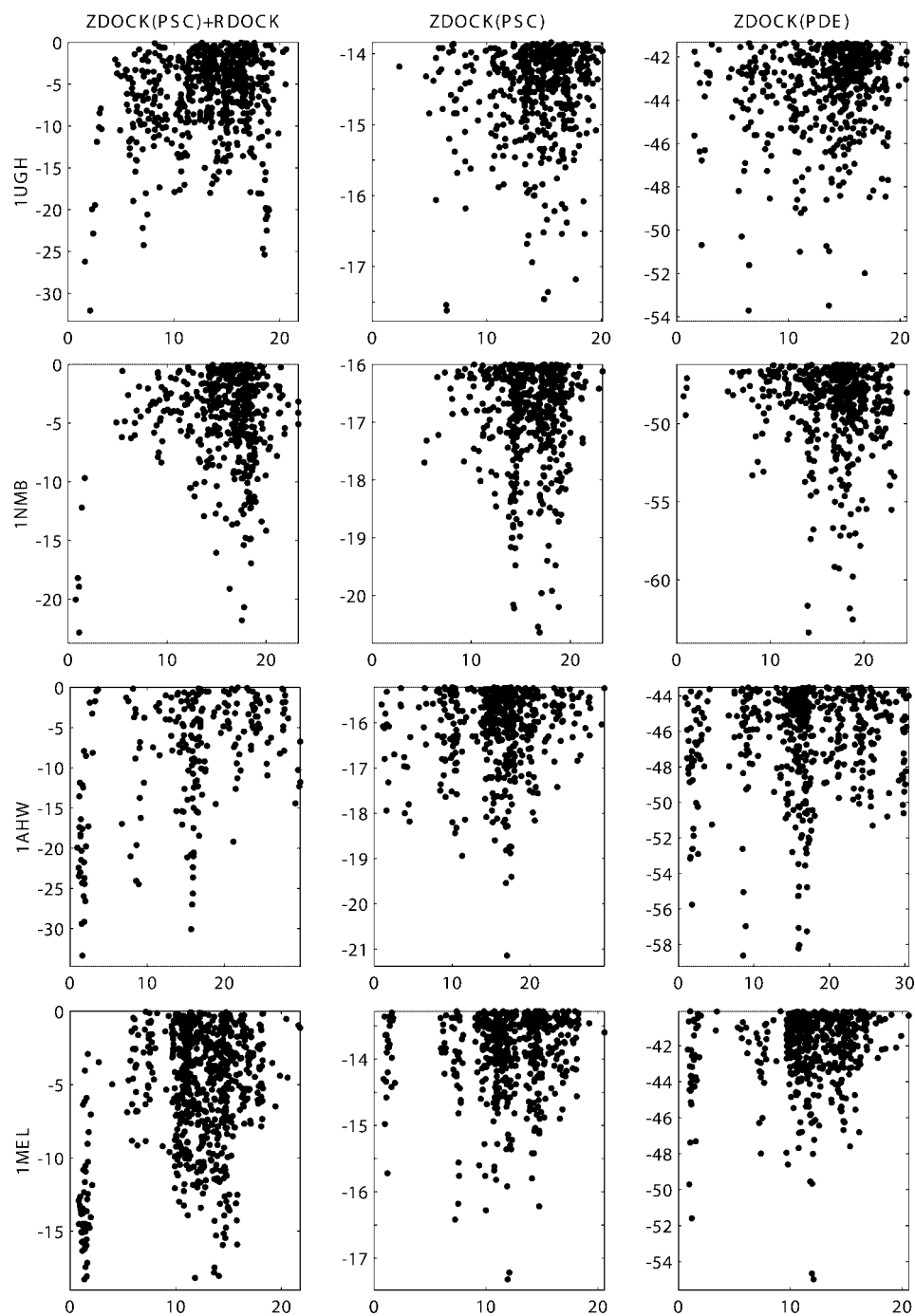


Figure 5. (Continued.)

at least one hit, thus its maximal success rate is 86% for this benchmark. Among three categories of test cases, best results are seen for enzyme/inhibitor, with a success rate of 48% at $N_P = 1$ and 78% at $N_P = 10$. Even though ZDOCK has the most difficulty with test cases in the “others” category, the improvement made by RDOCK on the available test cases is rather impressive: a hit is ranked as the number-one prediction for 3 test cases, and ranked at 11, 16, and 21 for the remaining 3 test cases (Table I).

It is difficult to make direct comparison with previous refinement methods. Most previous methods assumed known binding site information, and/or used ≤ 10 unbound test cases.^{5,7,9,11,14-18,22,38} Others focused on discriminating against low RMSD false positives.^{19,20} Also, the definitions of hits vary slightly among methods. We compared ZDOCK 1.3 with the genetic algorithm by Gardiner et al.¹⁰ and our method was slightly better.¹³ RDOCK is significantly better than ZDOCK 2.3, which was shown to be better than ZDOCK1.3.³⁴ Fernandez-Recio¹⁴ et al. at-

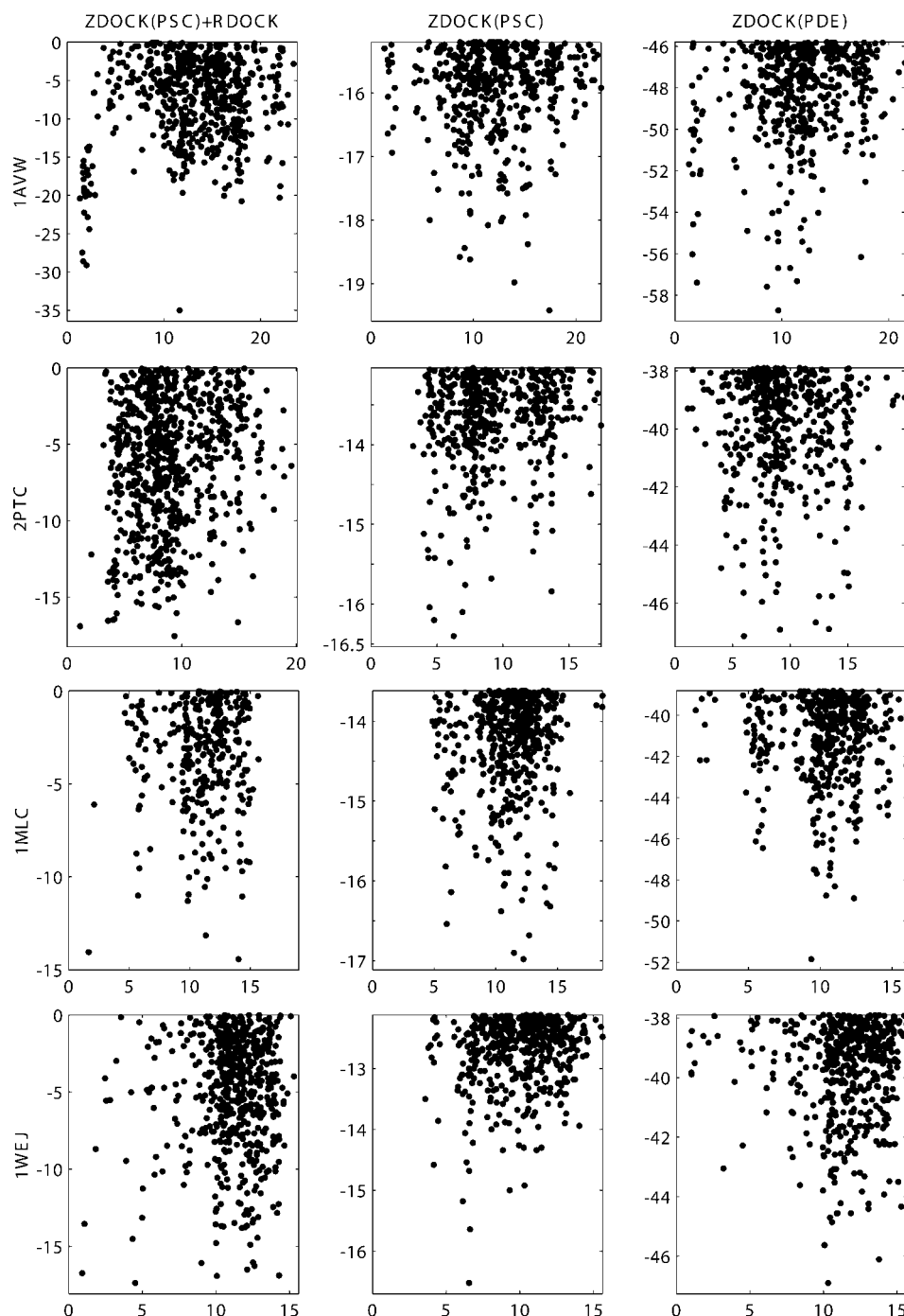


Fig. 6. Binding energy (ordinate) vs. RMSD (abscissa) graphs are compared among ZDOCK(PSC)+RDOCK (left), ZDOCK(PSC) (middle), and ZDOCK(PDE) (right) for the third group of test cases (6 in total). ZDOCK(PSC)+RDOCK ranks a hit among the top 4 predictions for these test cases.

tempted a comparison with four earlier methods,^{5,8,12,17} and showed that their method was the best. However, a direct comparison between our results and those of Fernandez-Recio et al.¹⁴ is not straightforward, since they used known binding site information for the receptors, and the vast majority of their test cases were enzyme/inhibitor. If we limit ourselves to this category, our global search results can be favorably compared with theirs from semi-

global searches: we rank a hit in the top 1 for 48% test cases and in the top 10 for 78% test cases, and the corresponding percentages for their method are 35 and 65%, respectively.¹⁴

There are two aspects that a refinement algorithm can improve docking performance: (1) it can refine structures and improve the binding energies for hits by computing them more accurately (hopefully to a greater extent than

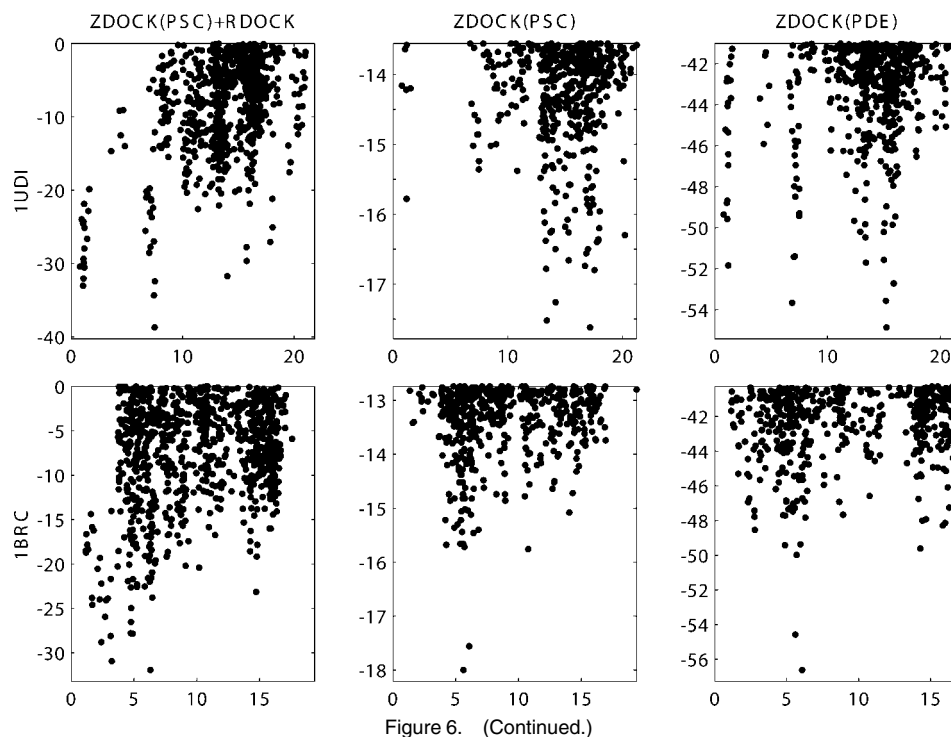


Figure 6. (Continued.)

for false positives), and (2) it can discriminate against false positives by using energy components that are not present at the initial stage algorithm. The first point is clearly supported by our results since we observe a consistent improvement by RDOCK over all three sets of ZDOCK predictions for most test cases. The second point deserves more discussion. It is intriguing that even though ZDOCK(PDE) substantially outperforms ZDOCK(PSC),³⁴ ZDOCK(PSC)+RDOCK substantially outperforms ZDOCK(PDE)+RDOCK (Fig. 1). We believe this is due to the partial overlap between PDE³⁴ and the scoring function of RDOCK (Equation 1). In addition to the PSC component, PDE also contains Coulombic electrostatics and ACE-based desolvation, although they are computed somewhat differently from RDOCK. In order to save computing time, these two terms in ZDOCK are grid-based and they are formulated to fit the Fast Fourier Transform algorithm. Furthermore, the desolvation term is not pairwise.³⁴ In other words, they are only approximations of the terms in Equation 1. This explains the improvement of RDOCK over ZDOCK(PDE) (Fig. 1). Nevertheless, the fact that ZDOCK(PDE) significantly outperforms ZDOCK(PSC) indicates that the formulation of PDE is sufficiently accurate. Therefore, we expect ZDOCK(PSC) and ZDOCK(PDE) to generate different populations of false positives: the former would generate false positives with the best shape complementarity, and the latter with the highest composite scores containing shape complementarity, desolvation, and electrostatics energies. RDOCK is particularly effective in discriminating against PSC false positives, since most of them do not have favorable desolvation or electrostatics. In comparison, it is less effective against PDE false positives. This finding raises an important issue for the development of refinement methods: a

successful refinement algorithm should try to cover energy terms that are not contained in the initial stage method. We believe that the development of such orthogonal energy functions should be a promising future direction.

A key component of RDOCK is the way we treat ionic side chains: they are kept neutral throughout the first two stages of energy minimization, and only reverted to full charge for the last stage of brief minimization (20 steps). Moreover, a distance dependent dielectric constant of $4r$ was used throughout. Although Coulombic electrostatics with a distance-dependent dielectric constant is commonly used in docking algorithms, none of the above-cited work neutralizes ionic side chains. In another work by Gatchell et al.³⁹ on the topic of protein folding, a refinement method was developed to discriminate near-native protein structures from misfolded models in several decoy sets, and neutralized side chains were shown to be essential for the discrimination. Interestingly, a similar procedure was not used in two other reports on docking refinement.^{19,20} Lazaridis and Karplus,^{40,41} developers of the CHARMM force field, also reported that fully charged side chains could not distinguish near-native protein structures from misfolded ones. Note that the CHARMM force field has been developed with molecular dynamics as the primary application. Without explicit solvent molecules, oppositely charged side chains become very close to each other after energy minimization. This corresponds to extremely favorable electrostatics energies in vacuum, which outweigh all other energy terms. Lazaridis and Karplus recommended using neutralized ionic side chains in combination with a distance dependent dielectric constant.^{40,41} This is supported by our results. However, we are the first group to show that reverting to full charges after minimization

significantly improves results (Fig. 2). We conclude that the full charge state is, in fact, more accurate than the neutral state for ionic side chains. However, they must not be used during energy minimization, since this could lead to inaccurate structures and thus inaccurate energies.

Another factor that contributes to RDOCK's good performance is its favorable starting point. ZDOCK retains many hits within the top 2,000 structures, which are supplied to RDOCK. Camacho et al.¹⁹ commented that they had difficulty in obtaining sufficient hits in the top 10,000 structures from rigid-body docking algorithms to test their refinement algorithm. This is clearly not a concern with ZDOCK. We have also observed that having more hits (the 6° vs. 15° rotational sampling in ZDOCK) benefits RDOCK (Fig. 3).

In summary, we have presented a simple refinement algorithm RDOCK for unbound protein docking. It significantly improves over our initial-stage docking algorithm ZDOCK, and is highly competitive when compared with previous docking algorithms. The next obvious steps for improving RDOCK are incorporating side chain conformational search and developing orthogonal energy components.

ACKNOWLEDGMENTS

We thank Kevin Wiehe for critically reading the manuscript. We are grateful to the Scientific Computing Facilities at Boston University and the Advanced Biomedical Computing Center at NCI, NIH, for computing support.

REFERENCES

- Levinthal C, Wodak SJ, Kahn P, Dadivanian AK. Hemoglobin interaction in sickle cell fibers. I: theoretical approaches to the molecular contacts. *Proc Natl Acad Sci USA* 1975;72:1330–1334.
- Wodak SJ, Janin J. Computer analysis of protein-protein interaction. *J Mol Biol* 1978;124:323–342.
- Sternberg MJ, Gabb HA, Jackson RM, Moont G. Protein-protein docking. Generation and filtering of complexes. *Methods Mol Biol* 2000;143:399–415.
- Smith GR, Sternberg MJ. Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol* 2002;12:28–35.
- Gabb HA, Jackson RM, Sternberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Proteins* 1997;S1:226–230.
- Vakser IA. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins* 1997; (Suppl 1):226–30.
- Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. *Proteins* 2000;39:178–194.
- Palma PN, Krippahl L, Wampler JE, Moura JJ. BiGGER: A new (soft) docking algorithm for predicting protein interactions. *Proteins* 2000;39:372–384.
- Taylor JS, Burnett RM. DARWIN: a program for docking flexible molecules. *Proteins* 2000;41:173–191.
- Gardiner EJ, Willett P, Artymiuk PJ. Protein docking using a genetic algorithm. *Proteins* 2001;44:44–56.
- Mandell JG, Roberts VA, Pique ME, Kotlovoy V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF. Protein docking using continuum electrostatics and geometric fit. *Protein Eng* 2001;14:105–113.
- Norel R, Sheinerman F, Petrey D, Honig B. Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci* 2001;10:2147–2161.
- Chen R, Weng Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins* 2002;47:281–294.
- Fernandez-Recio J, Totrov M, Abagyan R. Soft protein-protein docking in internal coordinates. *Protein Sci* 2002;11:280–291.
- Jackson RM, Sternberg MJE. A continuum model for protein-protein interactions: Application to the docking problem. *J Mol Biol* 1995;250:258–275.
- Weng Z, Vajda S, Delisi C. Prediction of protein complexes using empirical free energy functions. *Protein Sci* 1996;5:614–626.
- 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.
- Moont G, Gabb HA, Sternberg MJ. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins* 1999;35:364–373.
- Camacho CJ, Gatchell DW, Kimura SR, Vajda S. Scoring docked conformations generated by rigid-body protein-protein docking. *Proteins* 2000;40:525–537.
- Camacho CJ, Vajda S. Protein docking along smooth association pathways. *Proc Natl Acad Sci USA* 2001;98:10636–10641.
- Cherfils J, Duquerroy S, Janin J. Protein-protein recognition analyzed by docking simulation. *Proteins* 1991;11:271–280.
- Shoichet BK, Kuntz ID. Protein docking and complementarity. *J Mol Biol* 1991;221:327–346.
- 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.
- Walls PH, Sternberg MJE. New algorithm to model protein-protein recognition based on surface complementarity. applications to antibody-antigen docking. *J Mol Biol* 1992;228:277–297.
- Cherfils J, Janin J. Protein docking algorithms: simulating molecular recognition. *Curr Opin Struct Biol* 1993;3:265–269.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983;4: 187–217.
- Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. *Nature (London)* 1986;319:199–203.
- Pickett SD, Sternberg MJE. Empirical scale of side-chain conformational entropy in protein folding. *J Mol Biol* 1993;231:825–839.
- Bruccoleri RE, Haber E, Novotny J. Structure of antibody hyper-variable loops reproduced by a conformational search algorithm [published erratum appears in *Nature* 1988 Nov 17;336(6196): 266]. *Nature* 1988;335:564–568.
- Norel R, Petrey D, Wolfson HJ, Nussinov R. Examination of shape complementarity in docking of unbound proteins. *Proteins* 1999;36: 307–317.
- Totrov M, Abagyan R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy. *Nat Struct Biol* 1994;1:259–263.
- Zhang C, Vasmataz G, Cornette JL, DeLisi C. Determination of atomic desolvation energies from the structures of crystallized proteins. *Proteins* 1997;51:397–408.
- Chen R, Weng Z. A novel shape complementarity scoring function for protein-protein docking. *Proteins* Forthcoming.
- Chen R, Li L, Weng Z. ZDOCK: An initial-stage protein docking algorithm. *Proteins* Forthcoming.
- Chen R, Mintseris J, Janin J, Weng Z. A protein-protein docking benchmark. *Proteins* Forthcoming.
- Chen R, Tong W, Mintseris J, Li L, Weng Z. ZDOCK Predictions for the CAPRI Challenge. *Proteins* Forthcoming.
- Pearson WR. Comparison of methods for searching protein sequence databases. *Protein Sci* 1995;4:1145–1160.
- Lorber DM, Udo MK, Shoichet BK. Protein-protein docking with multiple residue conformations and residue substitutions. *Protein Sci* 2002;11:1393–1408.
- Gatchell DW, Dennis S, Vajda S. Discrimination of near-native protein structures from misfolded models by empirical free energy functions. *Proteins* 2000;41:518–534.
- Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins* 1999;35:133–152.
- Lazaridis T, Karplus M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J Mol Biol* 1999;288:477–487.