PROTEINS: Structure, Function, and Genetics 40:525–537 (2000)

Scoring Docked Conformations Generated by Rigid-Body Protein-Protein Docking

Carlos J. Camacho, David W. Gatchell, S. Roy Kimura, and Sandor Vajda*
Department of Biomedical Engineering, Boston University, Boston, Massachusetts

ABSTRACT **Rigid-body methods, particularly Fourier correlation techniques, are very efficient for docking bound (co-crystallized) protein conformations using measures of surface complementarity as the target function. However, when docking unbound (separately crystallized) conformations, the method generally yields hundreds of false positive structures with good scores but high root mean square deviations (RMSDs). This paper describes a two-step scoring algorithm that can discriminate near-native conformations (with less than 5 Å RMSD) from other structures. The first step includes two rigid-body filters that use the desolvation free energy and the electrostatic energy to select a manageable number of conformations for further processing, but are unable to eliminate all false positives. Complete discrimination is achieved in the second step that minimizes the molecular mechanics energy of the retained structures, and re-ranks them with a combined free-energy function which includes electrostatic, solvation, and van der Waals energy terms. After minimization, the improved fit in near-native complex conformations provides the free-energy gap required for discrimination. The algorithm has been developed and tested using docking decoys, i.e., docked conformations generated by Fourier correlation techniques. The decoy sets are available on the web for testing other discrimination procedures. Proteins 2000;40:525–537.**

Key words: free-energy function; protein modeling; implicit solvation; fast-Fourier transform; energy minimization

INTRODUCTION

The goal of predictive protein docking is to obtain a model for the bound complex from the coordinates of the unbound component molecules.[1] The ability to predict whether and how two proteins are likely to interact would be an immensely useful tool in the analysis of biochemical processes. However, the problem is clearly a very difficult one if little or no information on the binding site is available, as the search must explore the very large configurational space of potential complex conformations.[2]

Current methods that address the above problem are based on the rigid body approximation, whereby the proteins are treated as solid objects.[1,3–6] Such methods can explore vast numbers of docked conformations, evaluating a simple function that describes the geometric fit or surface complementarity of each structure, possibly allowing for some overlap. The approach is very successful when docking bound (co-crystallized) protein conformations. In this case the structures ranked at the top are generally in the correct orientation, provided a sufficiently dense grid is used in the calculation.[7,8] However, the situation is very different when docking unbound (independently crystallized) conformations of the component proteins. Due to the incorrect conformations of some key side chains in the binding site, all near-native structures may have relatively poor surface complementarity, and hence the higher ranked conformations are frequently false positives, i.e., structures with good score but high root mean square deviation (RMSD). While we discuss this problem in the context of Fourier correlation docking methods which have generated considerable interest in recent years[7,8,10–18] it has also been well-documented in conjunction with other rigid-body docking techniques, including the geometric hashing system by Fisher et al.,[6] and the bit-mapping algorithm of Palma et al.[9]

Substantial efforts have been devoted to the development of methods for screening the docked conformations in order to eliminate the false positives. In the simplest case this involves re-ranking the structures using scoring functions that attempt to measure the affinity of the protein-protein association in the given conformation. Such functions include solvation potentials, empirical atom-atom or residue-residue contact energies, and continuum electrostatics.[19–24] For example, Jiang and Kim[5] measured the affinity of protein-protein interactions by counting the number of favorable and unfavorable pairwise interactions defined for six atom types, whereas Wallqvist and Covell[25] derived a free energy approximation from observations of surface burial of various atom pairs across the interface of known enzyme-inhibitor complexes.

More complex procedures have also been developed combining several functions and, in some cases, the refinement of the interacting surfaces. Jackson and Sternberg[20] optimized the positions of polar hydrogens and removed some steric overlaps before calculating the electrostatic

Grant sponsor: National Science Foundation; Grant number: DBI-9904834; Grant sponsor: Department of Energy; Grant number: DE-F602-96ER62263.

*Correspondence to: Sandor Vajda, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215. E-mail: vajda@enga.bu.edu

Received 20 January 2000; Accepted 13 April 2000

free energy by the Poisson-Boltzmann equation. This same group developed a more elaborate refinement of protein interfaces, including a mean-field representation of side-chain rotamers and a Langevin dipole solvation model.[23] More recently, they added a filter based on an empirical residue-residue potential.[24] Weng et al.[21] applied a systematic side-chain conformational search before calculating the binding free energies using an empirical free-energy potential. Palma et al.[9] used a neural network to optimize a scoring function that incorporated terms representing geometric complementarity, electrostatic interactions, desolvation free energy, and pairwise propensities of amino acid side chains.

The above methods improve the ranking such that some near-native conformations (with less than 5 Å RMSD) are generally within the top hundred or even within the top ten structures.[23,24,9] However, for most complexes the highest ranked structures are still false positives with large RMSD from the native. Thus, in spite of all efforts and the relative sophistication of some of the scoring functions, no currently available technique has been reported that reliably discriminates near-native conformations from the other structures generated by rigid-body docking techniques.[26] The failure to eliminate all false positives is clearly a major drawback to practical applications of predictive protein docking, and may be partially responsible for the fact that, in spite of the computer programs available on the web,[12,13,17] protein-protein docking is rarely used in applications. A notably successful exception is predicting the interaction domain for cytochrome c on cytochrome c oxydase,[27] which was shown to be consistent with and explanatory of experimental results in a companion paper.[28]

In this paper we show that the discrimination problem can be solved by putting together, in a novel way, computational elements that are already available in the literature, and can be easily added to the existing programs. The resulting procedure discriminates near-native docked conformations from other structures in two simple and computationally efficient steps. The first step includes two, distinct rigid-body filters that select a manageable number of conformations (one thousand or less) for further processing. While this is a realistic goal that has been achieved by a number of scoring methods reported in the literature, most current work is directed toward further improving the potentials in order to eliminate all false positives which, we argue, can not be achieved within the framework of the rigid-body approximation. In fact, improving the specificity of a scoring function also increases its sensitivity to the differences between bound and unbound protein conformations, and the quality of discrimination does not necessarily improve.

Complete discrimination is achieved in the second step of the scoring procedure, and is due to two factors. The first factor is the application of a free-energy function that combines molecular mechanics with an empirical solvation model. The second factor is the extensive refinement of the docked conformations by energy minimization preceding the free-energy evaluation. Before the minimization, the van der Waals energy is very sensitive even to very small overlaps, and hence provides little or no information on the overall geometry of a docked structure. Since the minimization removes such overlaps, the van der Waals energy becomes highly correlated with RMSD, and including it in the free-energy function substantially improves the discrimination.

As we mentioned, the above computational elements can be found in the literature, but have not been used in the context described here. For example, Shoichet and Kuntz[19] and Cherfils et al.[29,30] minimized the energy of conformations obtained by rigid-body docking, and also performed solvation calculations, but did not attempt to incorporate these terms into a single free-energy expression for improved scoring. Weng et al.[21] studied the conformations generated by Shoichet and Kuntz[19] and improved their results by minimizing the structures (including the search for some side-chain conformations), followed by free-energy evaluation. However, they failed to include the van der Waals term in the expression. A scoring function, similar to the one we will describe here, has been used by Totrov and Abagyan[31] in a quasi-global Monte Carlo minimization. In this work we substantially simplify the calculation by exploiting the fact that rigid-body methods can generate near-native conformations, and no further search is necessary. Since both rigid-body docking and local minimization are very efficient, computer time requirements are substantially reduced compared to traditional flexible-docking methods.

The discrimination procedure has been developed using two powerful analysis tools. The first is the systematic mapping of electrostatic and desolvation interactions between two proteins. This helped us to understand the complementary roles of these two contributions to the binding free energy, as well as the limitations of the rigid-body approximation. The second tool is testing the algorithm using specially generated docking decoys. Decoy studies have been extensively employed in the development of potentials for protein fold discrimination and structure prediction[32–35] but much less in the context of docking. An early example we already mentioned is the set of protease-protein inhibitor conformations generated by Shoichet and Kuntz.[19] The appeal of the decoy approach is shown by the facts that, shortly after their publication, these structures have been re-examined by two other groups who were able to improve the discrimination using empirical free-energy functions.[20,21] Using Fourier correlation methods it is now possible to generate more challenging sets of docking decoys, and we hope that the analysis of such decoys will result in improved scoring algorithms.

MATERIALS AND METHODS

Empirical Free Energy Functions

The scoring function we use is based on a free energy potential that combines a molecular mechanics energy with empirical solvation terms.[21,36–39] The potential has been developed for the calculation of the binding free energy $\Delta G = G^{rl} - G^{r} - G^{l}$, where G^{rl}, G^{r}, and G^{l} denote the free energies of the receptor-ligand complex, the free receptor, and the free ligand, respectively. In the most

TABLE I. Protein Complexes Studied

	Co-crystallized structures				Unbound structures	
PDB code	Receptor	Charge	Ligand	Charge	PDB code	PDB code
1cho	α-chymotrypsin	+3	OMTKY	0	5cha	2ovo
1ppf	H. Leuk. elastase	+11	OMTKY	0	1ppg	2ovo
2sni	Subtilisin novo	−1	Chymotrypsin inh. 2	0	1sbc	2ci2
2sni					1sup	2ci2
2kai	Kallikrein A	−17	PTI	+5		
1brs	Barnase	+2	Barstar	−5	1bao	1bta
1brs					1a2p	1a19
1mlc	Fab D44.1 (Antibody)	−3	Hen egg lysozyme	−3	1mlb	1lza
2ptc	Trypsin	+6	BPTI	+6	2ptn	6pti
2ptc					2ptn	4pti
2cgi	Chymotrypsinogen	+4	HPTI	−1	1chg	1hpt

general case the binding free energy is calculated by the expression

$$\Delta G = \Delta E_{elec} + \Delta E_{vdw} + \Delta G_{des} + \Delta E_{int} - T\Delta S_{sc} + \Delta G_{other} \quad (1)$$

where ΔE_{elec} and ΔE_{vdw} denote the changes in the electrostatic and van der Waals energy, respectively; ΔG_{des} is the desolvation free energy, ΔE_{int} is the internal energy change due to flexible deformations (including bond stretching, angle bending, torsional and improper energy terms), and ΔS_{sc} is the loss of side-chain entropy upon binding. The last term, ΔG_{other}, accounts for translational, rotational, vibrational, and cratic effects.[40,41] Since ΔG_{other} is a weak function of the size and shape of the interacting proteins[20,36,42–44] it will be considered constant.

The above free energy expression can be substantially simplified when used for docking or scoring. Since G^r and G^l are constant (i.e., they do not depend of the conformation of the complex), in an arbitrary reference state $\Delta G = G^{rl}$. The constant ΔG_{other} is also dropped. Furthermore, we assume that ΔE_{int} is small compared to the other terms in the free energy expression, and hence can be neglected.[36,42–48] This condition is trivially met in rigid-body association. Since $\Delta G = G^{rl}$, ΔE_{elec} and ΔE_{vdw} denote the electrostatic and van der Waals energies of the complex. These terms are calculated using version 19 of the Charmm potential[49] with polar hydrogens only, and with the distance-dependent dielectric of $\epsilon = 4r$.

The expression $\Delta G_{des} - T\Delta S_{sc}$ is replaced by the atomic contact energy term ΔG_{ACE}. The atomic contact energy[37] (ACE) is an atomic level extension of the quasichemical potential of Miyazawa and Jernigan.[50] In ACE the local interactions are given by $\sum_i \sum_j e_{ij}$, where the sum is taken over all atom pairs that are less than 6 Å apart.[37] The term e_{ij} denotes the atomic contact energy of interacting atoms i and j, and it is defined as the effective free-energy change when a solute-solute bond between two atoms of type i and j, respectively, is replaced by solute-solvent bonds. Although the atomic contact energies were estimated by a statistical analysis of atom pairing frequencies in high-resolution protein structures rather than in complexes, the function has been used to calculate the contribution $\Delta G_{des} - T\Delta S_{sc}$ to the binding free energy in a number of applications. For example, the binding free energies, calculated for nine protease-inhibitor complexes were typically within 10% of the experimentally measured values.[37]

With the above simplifications the free energy function given by Eq. 1 is reduced to the form

$$\Delta G = \Delta E_{elec} + \Delta G_{ACE} + \Delta E_{vdw}. \quad (2)$$

The function is often further simplified by assuming van der Waals cancellation. According to this assumption, the solute-solute and solute-solvent interfaces are equally well-packed, and hence the intermolecular van der Waals interactions in the bound state are balanced by solute-solvent interactions in the free state,[20,42–44,51–53] reducing the free energy to

$$\Delta G = \Delta E_{elec} + \Delta G_{ACE}. \quad (3)$$

We will assume van der Waals cancellation as a first-order approximation when evaluating the binding free energy of docked conformations in the rigid-body analysis. This approximation is necessary if no energy minimization is performed, because the docked conformations are not completely free of steric conflicts, resulting in wildly varying ΔE_{vdw} values. Since the correlation between ΔE_{vdw} and the RMSD is close to zero, in the rigid-body analysis the van der Waals term is not much more than some high frequency noise. However, the minor overlaps can be easily removed by the minimization, and ΔE_{vdw} becomes an important part of the free-energy function in the second step of the discrimination algorithm.

Mapping of Electrostatic and Desolvation Interactions

We have recently reported the systematic mapping of the electrostatic energy ΔE_{elec} and the desolvation free energy ΔG_{ACE} on a grid over the entire configurational space of diffusion-accessible encounter complexes.[54] In these encounter complexes the two proteins are brought together to a zero surface-to-surface distance, but extensive surface contacts are not necessarily established. Free-energy maps have been constructed for four complexes using bound protein conformations. In this paper we perform similar calculations for five complexes (1cho, 2sni, 1ppf, 1mlc, and 1brs) using the unbound structures of the component proteins, listed in Table I. The results will

provide the fundamentals for the discrimination procedure.

To describe the relative orientation of two proteins in an encounter complex, a coordinate system is placed at the receptor's center of mass such that the positive x axis is facing toward the binding site. We use two Euler angles, θ and ϕ, to describe the position of the ligand in this coordinate system, and three more, θ_l, ϕ_l, and ψ_l, to specify the orientation of the ligand around its own center of mass. The free energy components ΔE_{el} and ΔG_{ACE} are calculated on a grid in the space of these five Euler angles, while keeping the surface-to-surface distance between the two molecules at zero.

Docking Decoys

The construction of protein decoys used in the development of potentials for structure prediction usually involves the partial unfolding of the native conformation, either by high-temperature molecular dynamics or by introducing small random perturbations into nonregular fragments of the protein.[32–35] It is far from obvious how to generate decoys for the analysis of docking unbound protein conformations. We use two types of decoys, one obtained by docking unbound protein structures, and the other by fitting the unbound proteins to the complex and then introducing perturbations.

The decoys that will be referred to as Type I have been developed by Vakser,[13] and are available at the website http://reco3.musc.edu/. The decoy sets include a few near-native conformations among a large number of false positives with good surface complementarity, and hence are highly suitable for testing scoring functions. The false positives have been generated by docking the unbound component structures, shown in Table I, for the complexes 1cho, 2sni, 1brs, 2ptc, and 2cgi. The docking was carried out by the GRAMM program[13] at high resolution (grid size 1.2 Å, rotational step size 12°). For each complex, the top 95 structures were selected as decoys. Each set also includes the "native" conformation, obtained by superimposing the unbound proteins over the complex, as well as four near-native conformations, obtained by superimposing the unbound proteins over good matches of the bound structures. In order to remove substantial overlaps, we slightly minimized the native and near-native structures by performing 20 adopted basis Newton-Raphson (ABNR) steps with the Charmm potential. The van der Waals energies still varied widely after the minimization, and did not distinguish the near-native complexes from the decoys.

Type II decoys have been generated in our laboratory and are available at the website http://engpub1.bu.edu/bioinfo/MERL. The decoys have been designed to test if, after minimization, the combined free-energy potential $\Delta G = \Delta E_{elec} + \Delta G_{ACE} + \Delta E_{vdw}$ can correctly rank near-native conformations within 10 Å RMSD from the native complex. These decoys have been generated for five complexes (1cho, 1ppf, 2sni, 2kai, and 1mlc) by starting with the best fit of the unbound protein structures to the complex (generally less then 1.5 Å RMSD), and then perturbing the location and orientation of the ligand by small random translations and rotations. Prior to free energy evaluation, the structures have been minimized for 500 ABNR steps using the Charmm potential. As we will show, the RMSD's of the resulting conformations are evenly distributed between 1 Å and 8 Å, with a few structures deviating up to 10 Å.

Scoring Algorithm

The first step of the scoring procedure includes two rigid body filters (Steps 1a and 1b), based on ΔG_{ACE} and ΔE_{elec}, respectively. As we will show, the filters generally reduce the number of conformations to be retained for further processing. The second step involves the Charmm minimization of the filtered conformations, followed by an evaluation of the free energy $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$. The details of the procedure are as follows:

Step 1a

We calculate the desolvation free energy ΔG_{ACE} for all docked conformations, and retain only the conformations for which $\Delta G_{ACE} < G^{min}_{ACE} + C_{ACE}$, where ΔG^{min}_{ACE} is the lowest ΔG_{ACE} value found, and C_{ACE} is an empirical threshold. In this work $C_{ACE} = 5$ kcal/mol, but this value is likely to be adjusted in the light of future results. We consider ΔG_{ACE} rather than the sum $\Delta E_{elec} + \Delta G_{ACE}$ because the desolvation component exhibits much lower sensitivity to small changes in the atomic positions than the electrostatic energy, ΔE_{elec}. Thus, the differences between the bound and the unbound states affect ΔE_{elec} more than they affect ΔG_{ACE}, and we try to utilize the relatively error-free information, provided by ΔG_{ACE}, for the selection of potentially good conformations.

Step 1b

The free-energy maps indicate that some complexes with oppositely-charged component proteins may be primarily stabilized by strong electrostatic interactions.[54] For the docked conformations of such complexes we calculate the electrostatic energy ΔE_{elec} and select all conformations for which $\Delta E_{elec} < \Delta E^{min}_{elec} + C_{elec}$, where ΔE^{min}_{elec} is the lowest ΔE_{elec} value found, and C_{elec} is an empirical threshold. We use $C_{elec} = 10$ kcal/mol, but will adjust this value after solving more problems. The selected conformations are retained for further analysis, together with the ones selected in Step 1a.

Step 2

We minimize the Charmm energy of the retained docked conformations by applying a relatively large number of steps (500 to 1,000) of the adopted basis Newton-Raphson (ABNR) minimization algorithm using the distance-dependent dielectric function $\epsilon(r) = 4r$, and rank the minimized structures according to the free energy expression $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$. Notice that in this decoy study we minimize all 100 structures in each set, not only the ones that would be retained after the rigid body filtering.

RESULTS AND DISCUSSION

Electrostatic and Desolvation Free Energy Maps

Electrostatic and desolvation free energy maps, constructed for bound protein conformations,[54] have already

TABLE II. Ranking of Ten Best Binding Pockets of 4 Å RMSD Clusters of Encounter Complexes for Five Independently Crystallized Structures†

Ranking	θ	ϕ	ΔG_{ACE}	ΔE_{Coul}	RMSD Å	θ	ϕ	ΔG_{ACE}	ΔE_{Coul}	RMSD Å
	Degrees		Kcal/mol			Degrees		Kcal/mol		
5cha and 2ovo (Complex: 1cho)						1sbc and 2ci2 (Complex: 2sni)				
1	80	80	**−1.84**	**−2.66**	8.6	90	90	**−3.18**	**−0.11**	9.7
2	100	100	**−0.93**	**−2.36**	18.9	130	130	**−3.59**	**0.44**	30.6
3	80	60	**−1.08**	**−2.19**	23.7	80	80	**−3.00**	**−0.13**	11.1
4	90	70	**−1.33**	**−1.84**	21.7	90	80	**−3.14**	**0.04**	13.3
5	80	70	**−1.22**	**−1.93**	21.5	100	80	**−2.88**	**−0.02**	11.5
6	60	100	**0.32**	**−3.33**	19.0	50	130	**−2.73**	**−0.14**	32.9
7	70	110	**−1.44**	**−1.39**	23.5	50	110	**−1.81**	**−1.02**	29.5
8	70	50	**0.33**	**−3.13**	25.9	50	60	**−2.63**	**−0.16**	30.9
9	60	70	**−1.91**	**−0.71**	26.1	50	90	**−2.04**	**−0.60**	28.0
10	90	60	**−0.81**	**−1.77**	18.1	60	130	**−2.40**	**−0.15**	31.4
1ppg and 2ovo (Complex: 1ppf)						1mlb and 1lza (Complex: 1mlc)				
1	70	70	**−4.42**	**−0.09**	21.4	90	80	**−2.06**	**−0.14**	15.6
2	80	70	**−3.11**	**−1.24**	20.7	90	70	**−1.83**	**−0.18**	18.2
3	90	100	**−3.33**	**−0.98**	17.3	100	100	**−1.86**	**−0.15**	20.9
4	90	100	**−2.82**	**−0.98**	19.1	80	90	**−1.94**	**0.04**	15.0
5	60	80	**−3.36**	**−0.14**	23.9	80	80	**−1.78**	**0.02**	15.8
6	80	70	**−3.31**	**−0.18**	19.3	110	120	**−1.58**	**0.00**	30.8
7	90	80	**−3.12**	**−0.31**	10.9	90	80	**−1.14**	**−0.43**	15.1
8	90	70	**−2.78**	**−0.39**	17.0	110	80	**−1.36**	**−0.20**	23.1
9	110	90	**−2.77**	**−0.26**	16.4	110	60	**−1.45**	**−0.05**	29.2
10	130	130	**−1.52**	**−1.42**	29.8	90	100	**−1.40**	**−0.05**	15.4
1bao and 1bta (Complex: 1brs)										
1	100	120	5.55	**−5.77**	19.4					
2	90	110	3.05	**−5.54**	16.8					
3	80	90	4.64	**−5.37**	14.1					
4	90	100	2.31	**−5.21**	12.3					
5	90	90	4.12	**−5.00**	5.9					
6	120	130	4.11	**−4.74**	28.3					
7	130	130	1.76	**−4.35**	29.2					
8	80	110	4.19	**−4.06**	18.7					
9	90	100	4.29	**−3.94**	19.1					
10	90	120	4.31	**−3.80**	20.3					

†Ranking is according to the average value of $\Delta G_{ACE} + \Delta E_{elec}$ for all but the 1bao/1bta complex, which was ranked by ΔE_{elec}. In the standard Euler angle representation the complex structure is located at $(\theta, \phi; \theta_l, \phi_l, \psi_l) = (90°, 90°; 0°, 0°, 0°)$, where the first two angles correspond to the position of the center of the ligand, and the last three to its relative orientation. The center-to-center distance of the encounter complexes is constrained to the position of first contact between the VDW surfaces of the molecules (no overlaps). Mapping was done for a total of 355,014 encounter complexes corresponding to only half the surface of the receptor and ligand, i.e., $(\theta, \phi) = [40°, 50°, \cdots, 140°]$ and $\theta_l = [0°, 10°, \cdots, 90°]$ (and ϕ_l and ψ_l were sampled every 20°). Different binding pockets have a pairwise RMSD of 8 Å or more.

been reported. For the purposes of this paper it is important to note that these maps show two different types of behavior. If the receptor and the ligand have weak charge complementarity (i.e., the proteins have few charges, or one of them is neutral), then the desolvation free-energy map exhibits a well-defined minimum with a broad region of attraction. By contrast, in the case of oppositely-charged proteins, a region of low electrostatic energy surrounds the binding site. Desolvation provides some added adhesion, and for molecules with strong electrostatics, the best predictor of the binding site is the minimum of the desolvation free energy within the region of low electrostatic energy. The main point is that either the desolvation free energy (in the neutral case), or the electrostatic energy (in the strongly-charged case) has a well-defined minimum that provides approximate information on the relative orientation of the two molecules even before a tight complex is formed.

The results of examining desolvation and electrostatic maps, constructed for five complexes using unbound protein conformations, are summarized in Table II. For each protein pair the table shows the ten lowest energy regions, defined by clusters of encounter complexes within 8 Å of each other. Notice that for each (θ, ϕ) pair we show only the lowest ΔE_{elec} and ΔG_{ACE} values. Ranking is based on the sum $\Delta G = \Delta E_{elec} + \Delta G_{ACE}$ for all but the 1brs complex, which was ranked by ΔE_{elec}. As we mentioned, in terms of the Euler angles the native complex is located at $(\theta, \phi, \theta_l, \phi_l, \psi_l) = (90°, 90°, 0°, 0°, 0°)$, where the first two angles correspond to the position of the center of the ligand, and the last three to its relative orientation (see Materials and Methods).

As shown in Table II, for the weakly-charged complexes the desolvation free energy ΔG_{ACE} has relatively low values close to the binding site at $(\theta = 90°, \phi = 90°)$, whereas for the charged complex 1brs the electrostatic

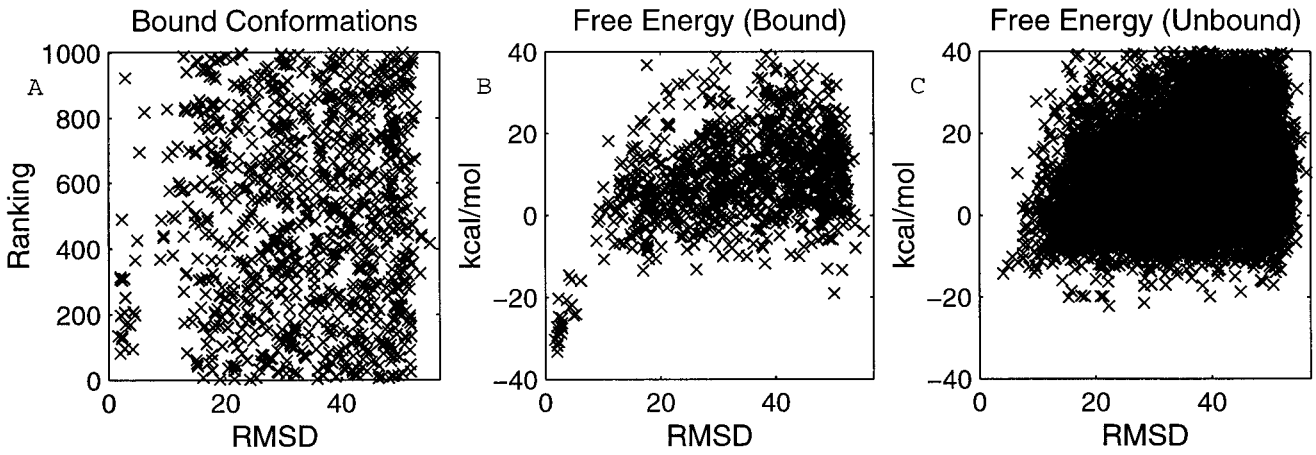

Fig. 1. Rigid-body docking of BPTI and trypsin. **A**. Surface complementarity ranking of the top 1,000 conformations obtained by docking the bound protein structures (PDB entry 2ptc). **B**. The free energy $\Delta G_{ACE} + \Delta E_{elec}$ of the docked conformations shown in part A as a function of the RMSD from the native complex. **C**. The free energy $\Delta G_{ACE} + \Delta E_{elec}$, calculated for the top 10,000 conformations obtained by docking the unbound trypsin and BPTI structures (2ptn and 6pti).

energy ΔE_{elec} has a number of minima around this point. However, unlike in the case of bound conformations, these minima are not global, and similar or even lower values are reached at other locations. For example, using the bound conformations of barnase and barstar (1brs), the electrostatic energy was shown to have a global minimum of −9.2 kcal/mol close to binding site.[54] With unbound proteins (PDB structures 1bao and 1bta), we find a local minimum, $\Delta E_{elec} = -5.0$ kcal/mol at ($\theta = 90°$, $\phi = 90°$) (see Table II). The cluster with the lowest electrostatic energy, −5.77 kcal/mol, is at 19.4 Å RMSD from the native. In the uncharged complexes 1cho and 2sni, ΔG_{ACE} has a well-defined global minimum around the binding site if the bound protein conformations are used in the calculation.[54] The use of unbound structures, 5cha/2ovo and 1sbc/2ci2, raises the ΔG_{ACE} value for these near-native encounter complexes, which then cease to be global minima. The ΔG_{ACE} map is particularly complex for the antibody-antigen complex 1mlb-1lza, where the cluster closest to the binding site at 13.45 Å RMSD ranks 40th, although the lowest free-energy region is not much further (15.66 Å). In summary, we conclude that for unbound conformations neither the desolvation free energy nor the electrostatic energy can definitely predict the binding site, and hence are unable to eliminate all false positives.

In Figure 1 we present some results for trypsin and BPTI in order to emphasize the difference between docking bound and unbound protein conformations. Figure 1A shows the surface complementarity ranking of the top 1,000 conformations obtained by docking the bound structures (PDB entry 2ptc), using a Fourier correlation docking method as implemented in the program DOT.[12] We used a grid step of 1 Å and an angular step of 12°. At this low resolution the surface complementarity ranking is not very good, even for the bound structure, and the first near-native structure ranks 90th. However, as shown in Figure 1B, the free energy function $\Delta G = \Delta G_{ACE} + \Delta E_{elec}$ provides excellent ranking of the same docked conformations, with a well-defined docking "funnel" at the binding site, and an almost 20 kcal/mol free-energy gap between near-native and other conformations.

Bound and unbound protein structures frequently differ in the conformations of some key side chains in the binding site, and this can dramatically change the result of the docking. Figure 1C shows $\Delta G_{ACE} + \Delta E_{elec}$ for the top 10,000 structures, obtained by docking unbound conformations of trypsin and BPTI (PDB entries 2ptn and 6pti, respectively). The DOT program[12] was used with the same parameters as in the docking of the bound conformations.

Figure 1C emphasizes the two new problems we are likely to encounter when docking unbound proteins. First, there are very few "hits", i.e., docked conformations with low RMSD from the native complex. In particular, for 2ptn and 6pti the lowest RMSD is 4 Å, and this structure is ranked 9,067th. Second, there is no well-defined docking "funnel" around the binding site. Thus, the free energy $\Delta G_{ACE} + \Delta E_{elec}$ does not fully discriminate between near-native and other structures, and simply re-scoring the conformations using a free-energy function can not eliminate all false positives. However, some partial affinity is retained close to the binding site, and hence ΔG_{ACE} and ΔE_{elec} still provide some structural information, particularly when used separately. The rigid-body filtering step of our discrimination algorithm will exploit this remaining affinity for reducing the number of conformations.

Rigid-Body Analysis

Figures 2 and 3 show ΔG_{ACE} and ΔE_{elec}, respectively, as functions of the RMSD for the five Type I decoy sets, each containing 100 structures. The five near-native structures are indicated by circles. As expected, the correlations between the free-energy terms and the RMSDs are not fully discriminatory, and do not enable us to eliminate all false positives. We can, however, use the relationships for filtering. As described in the Materials and Methods, we

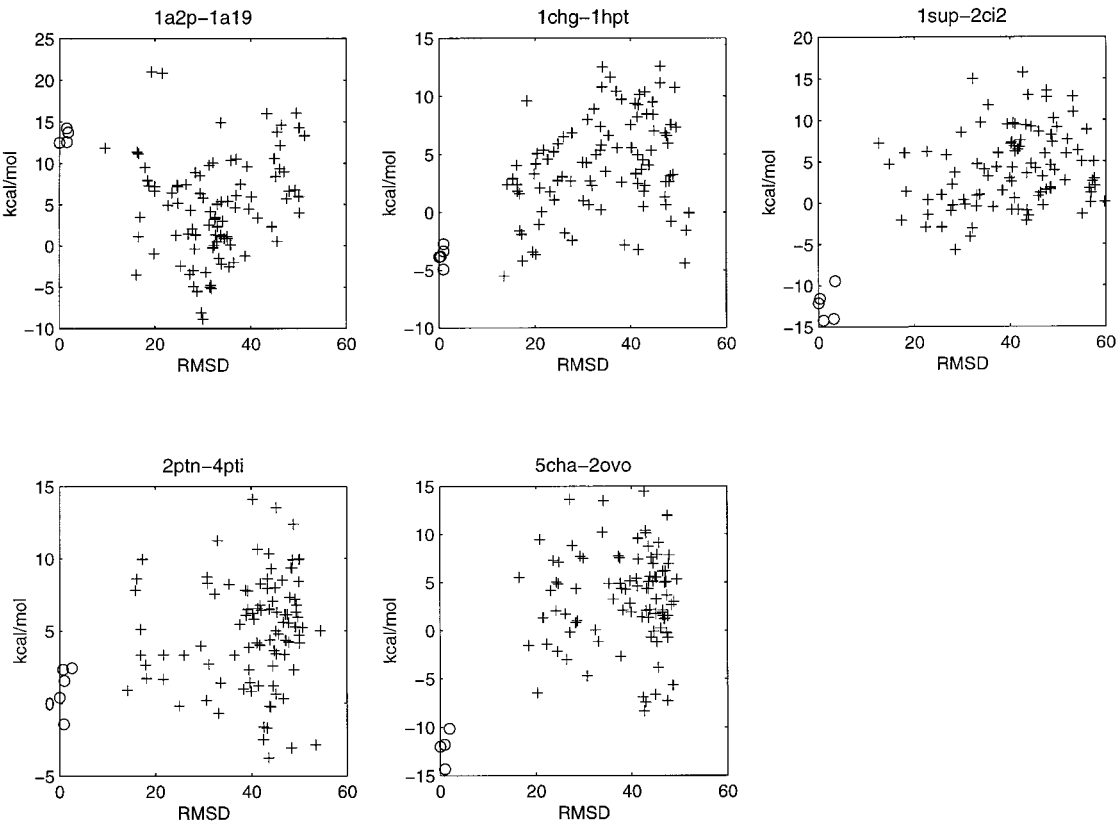

Fig. 2. Discrimination of unminimized Type I decoys by the desolvation free energy ΔG_{ACE} in the first rigid-body filter. The near-native conformations are indicated by circles.

always retain the conformations with low desolvation free energy, and for charged complexes we retain additional conformations with low electrostatics. The energy thresholds used here, $C_{ACE} = 5$ kcal/mol and $C_{elec} = 10$ kcal/mol, are empirical parameters, although we might argue that they could correspond to some minimum energy gap for binding. The larger threshold for ΔE_{elec} is due to the substantially larger uncertainty in the value of the electrostatic energy, calculated for the unbound conformations with incorrect side chains.

As shown in Figure 2, retaining all conformations within 5 kcal/mol of the lowest ΔG_{ACE} value would retain most, but not necessarily all, near-native conformations in four of the five decoy sets. The exception is the barnase/barstar complex (1a2p-1a19), which is very strongly charged (see Table I) and desolvation in the bound structure is known to be repulsive.[54] Since the native conformation of this complex is determined almost exclusively by electrostatic interactions, it is not surprising that desolvation is unable to identify the near-native folds. By contrast, ΔG_{ACE} discriminates very well the near-native structures from the decoys for the two complexes in which one of the proteins is neutral (1sup-2ci2 and 5cha-2ovo, see Table I). For the 1chg-1hpt complex, in which 1hpt is weakly charged, the near-native structures also have relatively low ΔG_{ACE} values, but the discrimination is not perfect. The same applies to 2ptn-4pti with strong but like charges.

For the two complexes with charged proteins (1a2p-1a19 and 2ptn-4pti), the filtering retains all conformations within 10 kcal/mol of the lowest ΔE_{elec} value. As shown in Figure 3, these include all near-native structures for 1a2p-1a19, the only complex among the five that is predominantly stabilized by electrostatics, and two of the five near-native conformations for 2ptn-4pti. Notice that the latter conformations would have already been selected on the basis of ΔG_{ACE}. Electrostatics does not provide any useful information for the other three complexes in which one of the proteins is neutral or weakly charged.

We emphasize that the strategy of using ΔG_{ACE} and ΔE_{elec} separately provides a much better filter than the one based on $\Delta G_{ACE} + \Delta E_{elec}$, because the sum is frequently dominated by outliers with very low electrostatic energy. These false positives are inherent to the sensitivity of the electrostatic energy to small charge overlaps. As we have shown, the same function provides very good discrimination in the docking of bound protein conformations (see Fig. 1B). Thus, the major factor limiting discrimination is the difference between bound and unbound states, rather than the inadequacy of the simple electrostatic and desolvation models. In agreement with previous results reported in the literature[23,24,9] improving the accuracy of free-energy calculations in the rigid body framework may improve discrimination, but generally does not result in the complete elimination of all false positives.

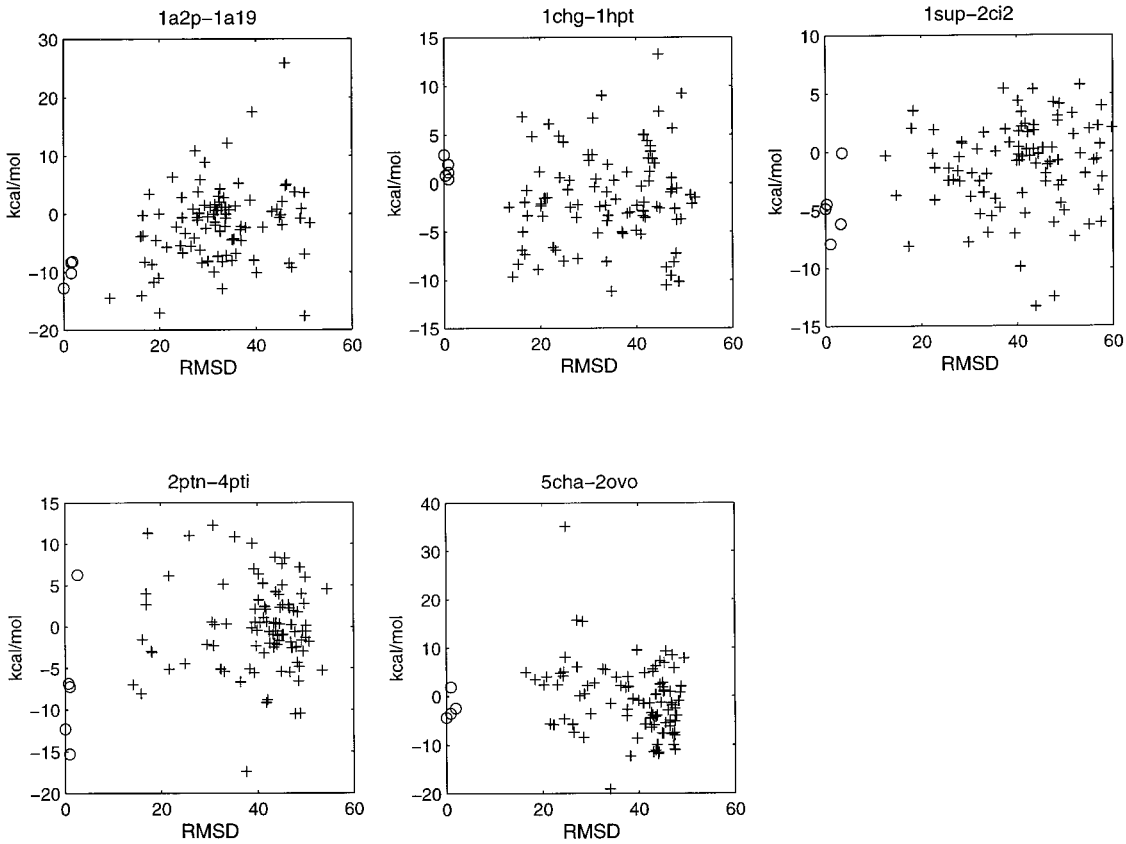

Fig. 3. Discrimination of unminimized Type I decoys by the electrostatic energy ΔE_{elec} in the second rigid-body filter. The near-native conformations are indicated by circles.

Discrimination After Energy Minimization

All structures in the five sets of Type I decoys have been minimized by applying 1000 ABNR minimization steps within the Charmm program.[49] Figure 4 shows the combined free energy expression $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$ of the minimized structures as a function of RMSD. The major result of this paper is that after extensive minimization the lowest value of ΔG is attained on a near-native structure in all five decoy sets. Furthermore, ΔE_{vdw} becomes an important factor in discriminating near-native from non-native conformations. The correlation between the RMSD and ΔE_{elec} generally improves by the minimization due to the removal of some overlaps, whereas ΔG_{ACE} remains almost invariant. Although the sum $\Delta G_{ACE} + \Delta E_{elec}$ is still unable to eliminate all false positives on its own, adding it to ΔE_{vdw} improves the discrimination. We note that the minimization may end up in local minima, and hence may be unable to reduce the van der Waals energy of some docked conformations in spite of a relatively low RMSD. For example, three of the minimized near-native structures for 1sup-2ci2 lie outside the range of the plot in Figure 3, and are thus lost. The number of such "false negatives" could have been reduced by employing a minimization method such as Monte Carlo that can cross energy barriers. However, according to our experience false negatives are rare, and therefore we perform computationally less demanding local minimizations.

Ranking Near-Native Structures

Type II decoys were used to show that the combined free energy function, $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$, not only discriminates the near-native conformations from the far-from-native decoys but also ranks conformations in the 2 to 10 Å RMSD range relatively well. We recall that the structures in the five decoy sets have already been refined by performing 500 steps of ABNR minimization using the Charmm potential.[49] Figure 5 shows the free energy ΔG of the minimized structures as a function of the RMSD, and demonstrates that for most decoy sets there is a significant correlation between the two quantities. For four of the five complexes, the free energy function ΔG reaches its minimum on docked conformations which are within 2 Å RMSD from the native, whereas for the antibody-antigen complex 1mlb-1lza the minimum occurs within 3 Å RMSD.

We use the correlation coefficients between the RMSD and various free-energy components, as well as their sums, to assess the quality of scoring. Table III shows the correlation coefficients calculated for the five sets of Type II decoys. These generally support the conclusions based on the analysis of free-energy maps and Type I decoys. In particular, the correlation coefficient between RMSD and

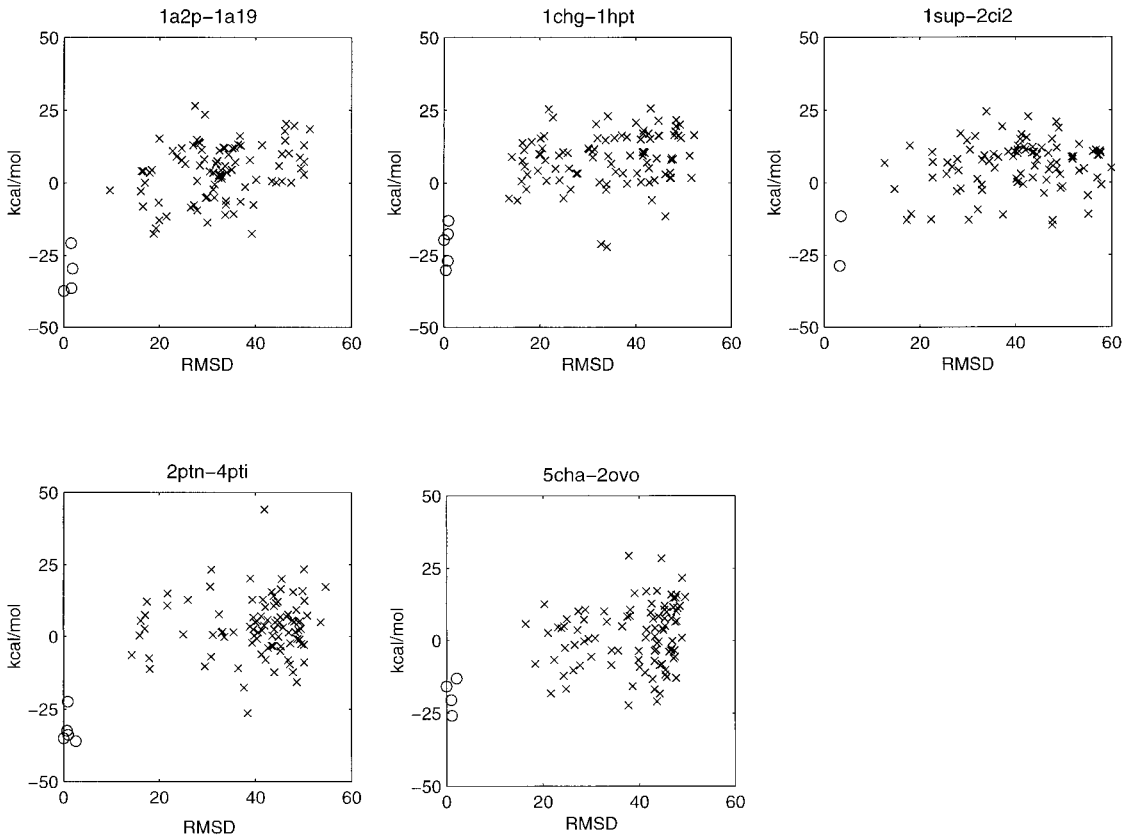

Fig. 4. Discrimination of minimized Type I decoys by the combined free-energy potential $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$. All structures have been subjected to 1,000 ABNR steps of minimization using the Charmm potential.

ΔG_{ACE} is relatively large for the three complexes in which one of the components is neutral (1cho, 2sni, and 1ppf). In contrast, the correlation is negative for the two charged complexes (1mlc and 1brs), indicating that in the presence of charges ΔG_{ACE} may be a weak discriminator. Indeed, we mentioned that the conformation of the barnase/barstar complex is determined by strong electrostatics interactions, and the correlation coefficient between RMSD and ΔE_{elec} is the highest for the barnase-barstar complex.

Ranking by the combined free-energy function $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$ is always better than ranking by any of its components, thus adding $\Delta G_{ACE} + \Delta E_{elec}$ slightly improves the discrimination provided by ΔE_{vdw} alone. The final correlation coefficient is much worse for the antibody-lysozyme complex 1mlc than for the other four systems. Adding the internal energy change, ΔE_{int} to $\Delta E_{vdw} + \Delta E_{elec} + \Delta G_{ACE}$ makes the ranking worse in all cases. This is somewhat surprising, since ΔE_{int} plays an important role in discriminating near-native protein structures from misfolded models.[55]

It is interesting that after the minimization, the correlation coefficient ρ(Electrostatics + ACE) between $\Delta E_{elec} + \Delta G_{ACE}$ and the RMSD, and the correlation coefficient ρ(Van der Waals) between ΔE_{vdw} and the RMSD, are very close to each other (Fig. 6), with a correlation coefficient of 0.89. This suggests that the quality of discrimination by these two functions is determined by a common factor, most likely by the difference between bound and unbound conformations of some key side chains.

Application to the Trypsin-BPTI Complex

We applied the discrimination algorithm to the 10,000 conformations, generated by docking unbound conformations of trypsin and BPTI (PDB entries 2ptn and 6pti), and shown in Figure 1C. Figure 7A shows the lowest 5 kcal/mol range of desolvation free energies. There are 93 structures with ΔG_{ACE} values within 5 kcal/mol of the minimum, which are thus retained for further analysis. Figure 7B shows the lowest 10 kcal/mol range of electrostatic energies. We find only eight docked structures with ΔE_{elec} values within 10 kcal/mol of the minimum, thus altogether we retain 101 conformations. Figure 7C shows the free energy $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$ for the 101 structures after 1,000 steps of ABNR minimization using the Charmm potential. The lowest value of ΔG is attained on the conformation closest to the native, at 4 Å RMSD. Although the discrimination is successful, the second lowest free energy value is at 17.6 Å, which is clearly too large. The problem is due to the fact that the set of 10,000 docked conformations includes too few "hits" (i.e., near-native conformations) to begin with. As will be further discussed, some current docking methods perform substan-

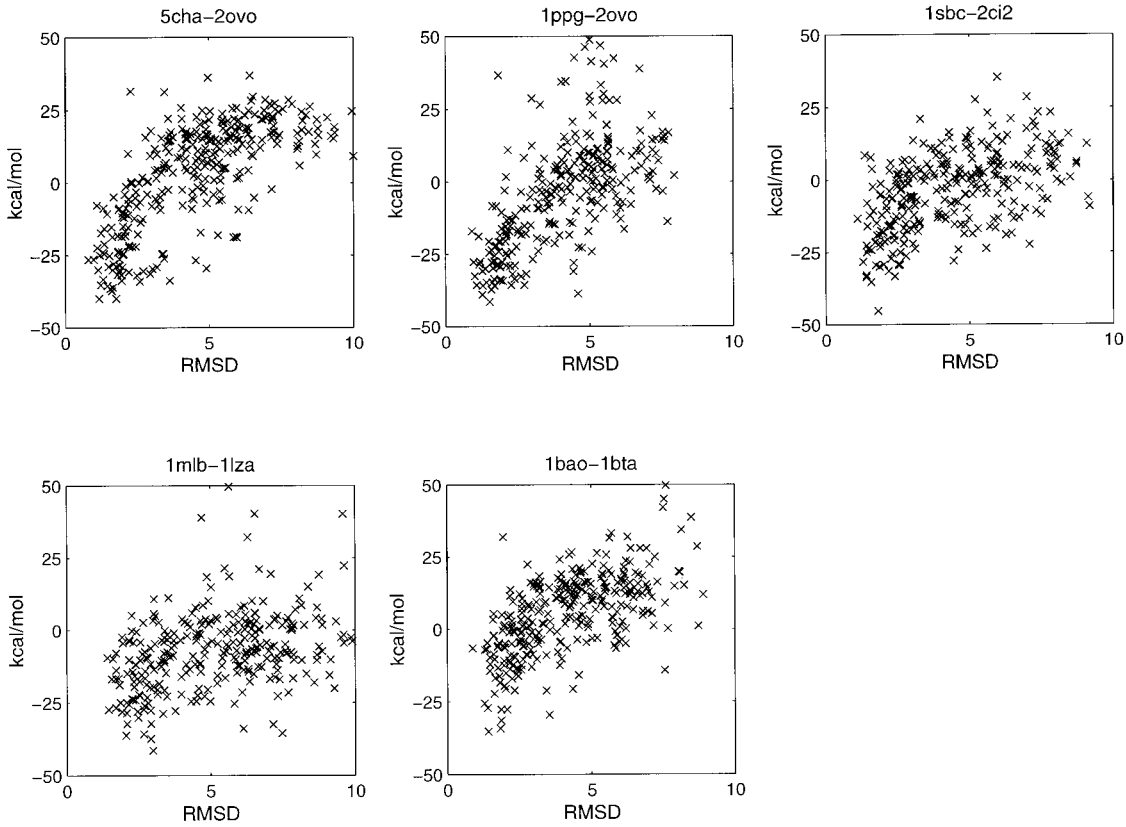

Fig. 5. Ranking of the minimized Type II decoys by the combined free-energy potential $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$. All structures have been subjected to 500 ABNR steps of minimization using the Charmm potential.

TABLE III. Correlation Coefficients Between Free-Energy Terms and RMSD

	Complex[b]				
Free energy term[a]	1cho	1ppf	2sni	1mlc	1brs
ΔG_{ACE}	0.59	0.26	0.51	−0.21	−0.02
ΔE_{elec}	0.44	0.36	0.17	0.29	0.54
$\Delta G_{ACE} + \Delta E_{elec}$	0.65	0.53	0.47	0.25	0.45
ΔE_{vdw}	0.66	0.55	0.47	0.31	0.39
$\Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$	0.69	0.60	0.52	0.33	0.54
$\Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw} + \Delta E_{int}$	0.36	0.35	0.29	0.25	0.50

[a] ΔG_{ACE}-desolvation free energy; ΔE_{elec}-electrostatic energy; ΔE_{vdw}-van der Waals energy; ΔE_{int}-internal energy.
[b] The unbound structures used in the docking are as follows: 5cha and 2ovo (1cho); 1ppg and 2ovo (1ppf); 1sbc and 2ci2 (2sni); 1mlb and 1lza (1mlc); 1bao and 1bta (1brs).

tially better,[9] and providing more hits increase the robustness of the procedure.

Why Does Minimization Improve Discrimination?

Figure 8 attempts to explain graphically the important role minimization plays in the improvement of discrimination. The continuous curve in the figure represents the free energy surface of the docked conformations formed by unbound protein structures. The open circles indicate conformations found by the Fourier correlation docking algorithm. As shown in Figure 1C, due to the incorrect conformations of key side chains the free energy does not fully discriminate the few near-native structures from the very large number of other conformations. The entire free-energy expression is dominated by the van der Waals term which, in turn, is related to surface complementarity. Therefore, the conformations found by the Fourier docking algorithm are generally close to the local minima of $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$, but most of these minima are very far from the native structure, and assuming rigid body association the potential does not provide any discrimination.

As shown in Figure 8, the minimization can substantially change the free energy surface. Indeed, starting from a relatively good docked conformation, the minimization "pushes" the ligand further into the binding site, and

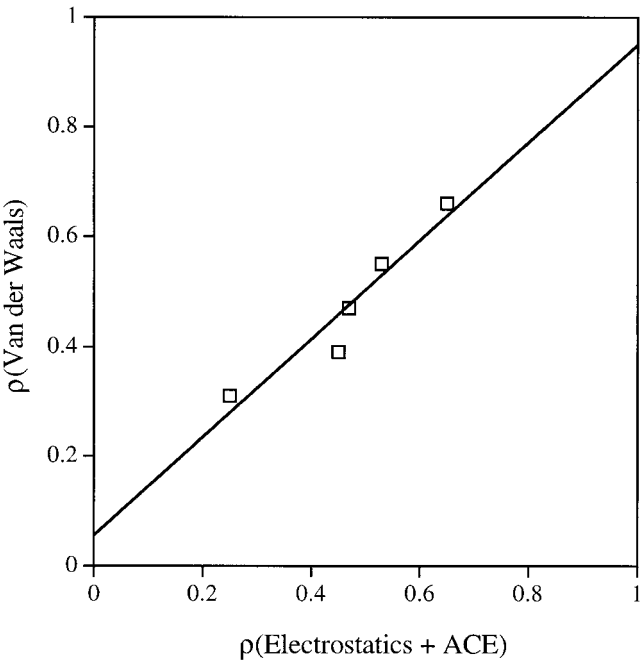

Fig. 6. The correlation coefficient between RMSD and ΔE_{vdw} as a function of the correlation coefficient between RMSD and $\Delta G_{ACE} + \Delta E_{elec}$, calculated for the five sets of Type II decoys.

yields a tight complex with large contact area, good surface complementarity, and low van der Waals energy. In contrast, starting from a docked conformation that is far from the native, the minimization is unable to create a comparable fit, and thus will not reduce the van der Waals energy as much as in the previous case. Thus, the minimization amplifies the differences between near-native and other structures, generally resulting in a free-energy gap of 5 kcal/mol or more.

The low free energy of near-native structures after the minimization is due the fact that, once minor clashes have been removed, the two molecules exhibit an overall shape complementarity. We note that the same overall shape complementarity has been exploited by Vakser in low-resolution docking methods.[13,56] This latter approach is based on the fact that low resolution representations (e.g., through the use of a coarse grid or a mean force approach) are not sensitive to local structural perturbations, including moderate conformational changes between bound and unbound states. Thus, at low resolution the rigid-body free-energy surface exhibits funnel-like behavior in the vicinity of the binding site, even when calculated for unbound conformations. Therefore, the docked ligand positions converge to this region,[56] and Fourier correlation docking can identify the general binding configuration.[13] However, it is still an open problem how these solutions can be used to build high resolution, near-native structures.[13,56]

Potential Difficulties

As shown in Figure 4, we found satisfactory discrimination of near-native structures from decoys for all five of the systems studied. However, these results are based on the analysis of a limited number of complexes, and hence it is important to assess potential problems we may face when the discrimination method is applied to other systems. In particular, most results presented here are for protease-inhibitor pairs. The atomic structure and flexibility of protein-protein recognition sites may be different in other complexes. For example, the discrimination of near-native docked structures from decoys seems to be more difficult for antibody-antigen than for protease-inhibitor complexes.[17] Indeed, the correlation coefficient between RMSD and the free energy is much lower for 1mlc than for the other four complexes (see Table III).

An obvious problem occurs if the top 10,000 or so structures from the rigid-body docking include too few "hits", i.e., near-native conformations. Notice that the subsequent minimization may be unable to reduce the van der Waals energy of some structures, and these "false negatives" will be lost. We have seen this situation in the trypsin/BPTI example, where two of the three near-native structures were lost during minimization. However, current results suggest that the number of "hits" can be increased by using improved target functions or filters within the rigid-body docking algorithms. For example, electrostatics was shown to provide moderate improvement.[12,17] Another approach is the inclusion of structure-based interaction potentials as part of the target function. For example, Palma et al.[9] substantially increased the number of "hits" within the top 1,000 structures through the use of a built-in filter, based on the pairwise propensities of amino acid side chains.

Docking Decoys

Decoy studies had a major impact on the development of potentials for protein fold discrimination and structure prediction.[32,22,34,35] The principles of decoy construction for protein structure prediction and refinement are well established.[32] Extending these principles to docking implies that good decoys (a) include near-native structures, (b) are native-like in all properties of the complex, except the particular docked conformation, (c) provide enough diversity to test various scenarios, and (d) are numerous enough for sensitive testing.

It is not completely clear how to generate docking decoys that are optimal in terms of the above properties. As described, we used two very different types of decoys that are available at http://reco3.musc.edu/ and http://engpub1.bu.edu/bioinfo/MERL, respectively. Our intention is to add further decoys to the latter webpage, and invite other groups to contribute. We strongly believe that an appropriate set of decoys will provide a powerful tool for the development of scoring functions to be used in docking.

CONCLUSIONS

We addressed the problem of discriminating near-native docked conformations from other structures generated by rigid-body docking. The discrimination algorithm has been tested on sets of decoys that included both near-native structures and others with good surface complementarity but large RMSD. Our results suggest the following conclusions:

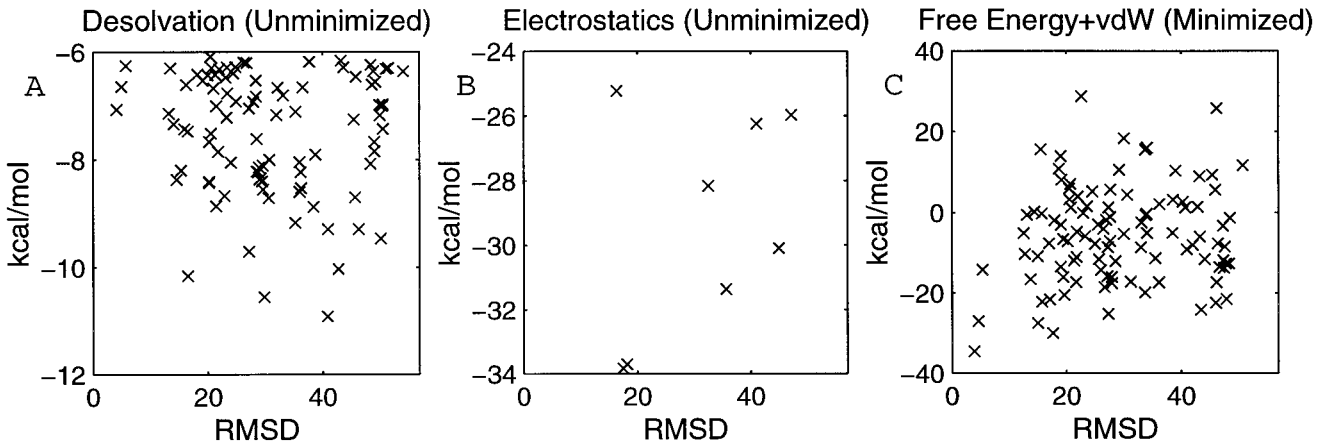

Fig. 7. An example: Docking the unbound conformations of BPTI to the unbound conformation of trypsin. **A**. 93 structures, selected from the 10,000 shown in Figure 1C, with low ΔG_{ACE} values (within 5 kcal/mol of the minimum). **B**. 8 structures, selected from the 10,000 shown in Figure 1C, with low ΔG_{elec} values (within 10 kcal/mol of the minimum). **C**. The free energy $\Delta G = \Delta G_{ACE} + \Delta E_{elec} + \Delta E_{vdw}$ of the 101 minimized structures.

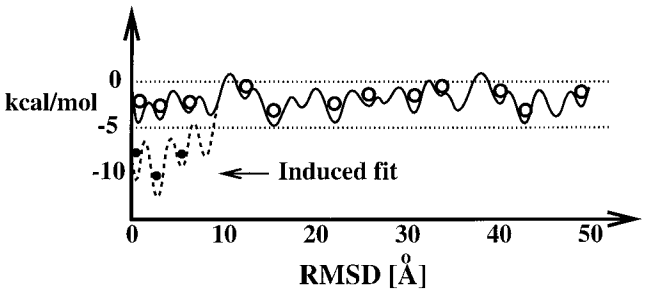

Fig. 8. Sketch of the free-energy surface of docked conformations obtained by docking unbound protein structures. The solid line shows the free energy in rigid-body docking, with open circles representing the conformations found by the Fourier correlation docking algorithm. The minimization lowers the free-energy surface in the vicinity of the native complex (dashed line), and results in lowered free energies for a number of near-native conformations, shown as filled circles.

1. The initial structure of the component proteins substantially affects the binding free-energy calculated for docked conformations. In particular, using unbound (separately crystallized) protein conformations, a number of high RMSD structures are found to be energetically comparable to near-native conformations, and complete elimination of false positives is not possible within the framework of a rigid body analysis. However, in spite of the differences in the side-chain conformations between bound and unbound states, the binding free energies calculated for near-native conformations are still among the lowest. Thus, the binding free-energy enables us to reduce the number of conformations to be retained for further analysis.
2. While the docked conformations of separately crystallized proteins are not accurate enough to provide the energy measures that could discriminate near-native structures from decoys, simple local minimizations yield "improved fits," i.e., tight complexes with a well-packed interface, and a free-energy gap between near-native complexes and other structures.
3. Decoys provide a powerful tool for the development of discriminatory potentials. Scoring functions can be tested and objectively compared. Thus, we believe that decoy studies involving docked conformations will make an important contribution to the development of discrimination methods.

ACKNOWLEDGMENTS

We thank Dr. Ilya Vakser for providing the Type I docking decoys.

REFERENCES

1. Sternberg MJE, Gabb HA, Jackson RM. Predictive docking of protein-protein and protein-DNA complexes. Curr Opin Struct Biol 1998;8:250–256.
2. Dixon JS. Evaluation of the CASP2 docking section. Proteins 1997; Suppl 1:198–204.
3. Wang H. Grid-search molecular accessible surface algorithm for solving the protein docking problem. J Comput Chem 1991;12:764–750.
4. Helmer-Citterich M, Tramontano A. PUZZLE: A new method for automated protein docking based on surface shape complementarity. J Mol Biol 1994;235:1021–1031.
5. Jiang F, Kim S.-H. Soft docking: matching of molecular surface cubes. J Mol Biol 1991;219:79–102.
6. Fischer D, Lin SL, Wolfson HL, Nussinov R. A geometry-based suite of molecular docking processes. J Mol Biol 1995;248:459–477.
7. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem A, Aflalo C, Vakser I. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA 1992;89:2195–2199.
8. Bliznyuk AA, Gready JE. Simple method for locating possible ligand binding sites on protein surfaces. J Comput Chem 1999;20:983–988.
9. Palma PN, Krippahl L, Wampler JE, Moura JG. BIGGER: A new (soft) docking algorithm for predicting protein interactions. Proteins, in press.
10. Vakser IA, Aflalo C. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. Proteins 1994;20:320–328.
11. Harrison RW, Kurinov IV, Andrews LC. The Fourier-Green's function and the rapid evaluation of molecular potentials. Protein Eng 1994;7:359–369.
12. Ten Eyck LF, Mandell J, Roberts VA, Pique ME. Surveying molecular interactions with DOT. In: Hayes A, Simmons M, editors. Proceedings of the 1995 ACM/IEEE Supercomputing Conference. New York: ACM Press, 1995.

13. Vakser I. Low-resolution docking: prediction of complexes for undetermined structures. Biopolymers 1996;39:455–464.
14. Ackermann F, Herrmann G, Kummert F, Pasch S, Sagerer G, Schomburg D. Protein docking combining symbolic descriptions of molecular surfaces and grid-based scoring functions. In: Rawlings C, Clark D, Altman R, Humter L, Lengauer TK, Wodak S, editors. Intelligent Systems for Molecular Biology. Menlo Park, CA: AAAI Press; 1995.
15. Meyerand M, Wilson P, Schomburg D. Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking. J Mol Biol 1996;264:199–210.
16. Friedman JM. Fourier-filtered van der waals contact surfaces: accurate ligand shapes from protein structures. Protein Eng 1997;10:851–863.
17. Gabb HA, Jackson RM, Sternberg MJE. Modeling protein docking using shape complementarity, electrostatics, and biochemical information. J Mol Biol 1997;272:106–120.
18. Blom NS, Sygusch J. High resolution fast quantitative docking using fourier domain correlation techniques. Proteins 1999;27:493–506.
19. Shoichet BK, Kuntz ID. Protein docking and complementarity. J Mol Biol 1991;221:327–346.
20. Jackson RM, Sternberg MJE. A continuum model for protein–protein interactions: application to the docking problem. J Mol Biol 1995;250:258–275.
21. Weng Z, Vajda S, DeLisi C. Prediction of complexes using empirical free energy functions. Protein Sci 1996;5:614–626.
22. Bacon DJ, Moult J. Docking by least–squares fitting of molecular surface patterns. J Mol Biol 1992;225:849–858.
23. Jackson RM, Gabb HA, Sternberg MJE. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. J Mol Biol 1998;276:265–285.
24. Moont G, Gabb HA, Sternberg MJE. Use of pair potentials across protein interfaces in screening predicted docked complexes. Proteins 1999;35:364–373.
25. Wallqvist A, Covell DG. Docking enzyme-inhibitor complexes using a preference–based free energy surface. Proteins 1996;35: 403–419.
26. Norel R, Petrey D, Wolfson HL, Nussinov R. Examination of shape complementarity in docking of unbound proteins. Proteins 1999;36: 307–317.
27. Roberts VA, Pique ME. Definition of the interaction domain for cytochrome c on cytochrome c oxidase. III. Prediction of the docked complex by a complete, systematic search. J Biol Chem 1999;274: 38051–38060.
28. Zhen YJ, Hoganson CW, Babcock GT, Ferguson-Miller S. Definition of the interaction domain for cytochrome c on cytochrome c oxidase. I. Biochemical, spectral, and kinetic characterization of surface mutants in subunit II of Rhodobacter sphaeroides cytochrome aa(3). J Biol Chem 1999;274:38032–38041.
29. Cherfils J, Duquerroy S, Janin J. Protein–protein recognition analyzed by docking simulation. Proteins 1991;11:271–280.
30. Cherfils J, Bizebard T, Knossow M, Janin J. Rigid body docking with mutant constraints of influenza hemagglutinin with antibody HC19. Proteins 1994;18:8–18.
31. Totrov M, Abagyan R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy. Nat Struct Biol 1994;4:259–263.
32. Park B, Levitt M. Energy functions that discriminate x-ray and near-native folds from well-constructed decoys. J Mol Biol 1996; 258:367–392.
33. Samudrala R, Moult J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol 1998;275:895–916.
34. DeBolt EE, Skolnick J. Evaluation of atomic level mean force potentials via inverse refinement of protein structures: atomic burial position and pairwise non-bonded interactions. Protein Eng 1996;8:175–186.
35. Vorobjev YN, Almagro JC, Hermans J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent. Proteins 1998;32:399–413.
36. Vajda S, Weng Z, Rosenfeld R, DeLisi C. Effect of conformational flexibility and solvation on receptor–ligand binding free energies. Biochemistry 1994;33:13977–13988.
37. Zhang C, Vasmatzis G, Cornette JL, DeLisi C. Determination of atomic desolvation energies from the structures of crystalized proteins. J Mol Biol 1996;267:707–726.
38. Gulukota K, Vajda S, DeLisi C. Peptide docking using dynamic programming. J Comput Chem 1996;17:418–428.
39. King BL, Vajda S, DeLisi C. Empirical free energy as a target function in docking and design: application to hiv-1 protease inhibitors. FEBS Lett 1996;384:87–91.
40. Brady GP, Sharp KA. Entropy in protein folding and in protein-protein interactions. Curr Opin Struct Biol 1997;2:215–221.
41. Vajda S, Sippl M, Novotny J. Empirical potentials and functions for protein folding and binding. Curr Opin Struct Biol 1997;7:222–228.
42. Novotny J, Bruccoleri RE, Saul FA. On the attribution of binding energy in the antigen–antibody complexes McPC 603, D1.3 and HyHEL-5. Biochemistry 1989;28:4735–4749.
43. Horton N, Lewis M. Calculation of the free energy of association for protein complexes. Protein Sci 1992;1:169–181.
44. Nauchitel V, Villaverde MC, Sussman F. Solvent accessibility as a predictive tool for the free energy of inhibitor binding to the HIV-1 protease. Protein Sci 1995;4:1356–1364.
45. Ajay, Murcko MA. Computational methods to predict binding free energy in ligand-receptor complexes. J Med Chem 1995;38:4953–4967.
46. Gilson MK, Given JA, Head MS. A new class of models for computing receptor-ligand binding affinities. Chem Biol 1997;4:87–92.
47. Verkhivker G, Appelt K, Freer ST, Villafranca JE. Empirical free energy calculations of ligand–protein crystallographic complexes. I. Knowledge–based ligand–protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. Protein Eng 1995;8:677–691.
48. Wallqvist A, Jernigan RL, Covell DG. A preference–based free energy parameterization of enzyme–inhibitor binding. Applications to HIV–1–protease inhibitor design. Protein Sci 1995;4:1881–1903.
49. 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.
50. Miyazava S, Jernigan R. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 1985;18:534–552.
51. Krystek S, Stouch T, Novotny J. Affinity and specificity of serine endopeptidase–protein inhibitor interactions. J Mol Biol 1993;234: 661–679.
52. Adamson AW. Physical chemistry of surfaces. New York: John Wiley & Sons; 1982. 629 p.
53. Nicholls A, Sharp KA, Honig B. Protein folding and association–insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
54. Camacho CJ, Weng Z, Vajda S, DeLisi C. Free energy landscapes of encounter complexes in protein–protein association. Biophys J 1999;76:1166–1178.
55. Janardhan A, Vajda S. Selecting near-native conformations in homology modeling: the role of molecular mechanics and solvation terms. Protein Sci 1997;7:1772–1780.
56. Vakser I, Matar OG, Lam CF. A systematic study of low-resolution recognition in protein–protein complexes. Proc Natl Acad Sci USA 1999;96:8477–8482.