doi:10.1016/S0022-2836(03)00670-3 J. Mol. Biol. (2003) 331, 281–299

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations

Jeffrey J. Gray, Stewart Moughon, Chu Wang, Ora Schueler-Furman Brian Kuhlman, Carol A. Rohl and David Baker*

Howard Hughes Medical Institute and Department of Biochemistry, University of Washington, J-567 Health Sciences, Box 357350, Seattle WA 98195, USA

**Corresponding author*

Protein–protein docking algorithms provide a means to elucidate structural details for presently unknown complexes. Here, we present and evaluate a new method to predict protein–protein complexes from the coordinates of the unbound monomer components. The method employs a low-resolution, rigid-body, Monte Carlo search followed by simultaneous optimization of backbone displacement and side-chain conformations using Monte Carlo minimization. Up to 10^5 independent simulations are carried out, and the resulting "decoys" are ranked using an energy function dominated by van der Waals interactions, an implicit solvation model, and an orientation-dependent hydrogen bonding potential. Top-ranking decoys are clustered to select the final predictions. Small-perturbation studies reveal the formation of binding funnels in 42 of 54 cases using coordinates derived from the bound complexes and in 32 of 54 cases using independently determined coordinates of one or both monomers. Experimental binding affinities correlate with the calculated score function and explain the predictive success or failure of many targets. Global searches using one or both unbound components predict at least 25% of the native residue–residue contacts in 28 of the 32 cases where binding funnels exist. The results suggest that the method may soon be useful for generating models of biologically important complexes from the structures of the isolated components, but they also highlight the challenges that must be met to achieve consistent and accurate prediction of protein–protein interactions.

Keywords: protein–protein docking; protein binding; biomolecular modeling; biomolecular free energy functions; conformational change

Introduction

The protein docking problem, that is, the task of assembling two separate protein components into their biologically relevant complex structure, is important for several reasons. First, it is of extreme relevance to cellular biology, where function is accomplished by proteins interacting with themselves and with other molecular components. Second, the protein docking problem presents a fundamental test of our understanding of the energetics of macromolecular interactions, as the native complex structure is almost certainly at a global free energy minimum. Finally, an important post-genomic goal is the characterization of the structures of protein–protein complexes, and computational tools offer an inexpensive means to carry out large-scale studies.

Protein–protein docking has been studied for some time now, and there are several excellent review articles available.[1–3] Many early and current docking strategies involve grid-based search algorithms.[4–11] These algorithms are quite successful at joining the components of a separated complex because of the excellent shape complementarity at the interface. However, proteins and protein interfaces are flexible, and the conformations of the bound partners often differ from

Present address: J. J. Gray, Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA.

Abbreviations used: rmsd, root-mean-squared distance.

E-mail address of the corresponding author: dabaker@u.washington.edu

those of the isolated components. If the unbound monomer components are used, it is no longer trivial to match the shapes together. Strategies to address this include softening the interface or coarsening the grid to allow more uncertainty in the matching process.[8,12,13] Chemical and physical information can be incorporated by including this information while matching the surfaces.[6,10,11,14] Finally, some algorithms explicitly include side-chain flexibility, although in most cases in only one of the protein partners.[15–17] Accurate and consistent prediction of correct complex structures from unbound components remains elusive, and few algorithms have been tested on large sets of targets.[8,14,17]

Recent large-scale studies have examined up to 27 targets. Fernández-Recio *et al*.[17] tested their method on unbound components using soft docking with side-chain refinement, localizing the search space to one-half of the receptor. Their algorithm admirably found correct solutions in the top 20 models in 17 of 24 cases, including seven of 11 protease-inhibitor cases for which the top-ranked solution was correct. Palma *et al*.[8] similarly found correct solutions of rank 20 or less in 14 of 25 cases (bound, semi-bound and unbound) using a soft docking algorithm designed to capture side-chain flexibility. Chen & Weng[14] used target functions that are tolerant of conformational change to study 27 systems; they predicted 12 structures within the top 20 ranked decoys, and three systems for which the correct solution was top-ranked. While these results are encouraging, current search algorithms are not sufficient to efficiently explore conformational space, and free energy functions are unable to consistently recognize correct complexes. There are still unsolved problems in the field of protein–protein docking, and insight could come from new approaches.

Like protein docking, protein folding requires a vast search and an accurate free energy or scoring function. Recently, progress has been noted in single-protein, *ab initio* structure prediction algorithms.[18] In particular, the Rosetta program developed at the University of Washington is now able to construct crude (~5 Å) models of many short (less than 150 amino acid residues) protein sequences.[19] This progress has emerged through the application of the following core philosophies and techniques. Physical forces are modeled whenever possible, but they are parameterized using data from high-resolution protein structures (rather than small molecules, as in traditional molecular mechanics approaches). The sampling problem is attacked with supercomputing clusters to create very large numbers of decoys. Residue-scale potential functions and backbone fragments are used to enable faster computation and to average interactions over long length and time-scales during the initial search of conformational space.[20,21] Physically based all-atom potentials are then used for refinement of decoys and accurate discrimination.[22] Finally, algorithm convergence, as measured by solution degeneracy after decoy clustering, is used as a final criterion in decoy selection.[19]

In this work, we adapt and expand the Rosetta techniques and philosophies for the protein–protein docking problem. Our algorithm includes a fast search using low-resolution potentials followed by an atomic-scale refinement step incorporating simultaneous optimization of side-chain positions and rigid-body displacement. The process mimics the steps involved in a diffusional encounter between two macromolecules, although the treatment is certainly not a rigorous physical simulation. Scoring functions include both physical and physically inspired statistical potentials derived from structures in the Protein Data Bank (PDB).[23] We employ small-perturbation studies to examine the quality of the scoring function. We use backbones of bound complexes to tune the algorithm, but we extend also to unbound complexes. Finally, we perform global searches using only the unbound component structures, emulating the situation of a blind search application.

A preliminary version of our protocol performed respectably in the opening rounds of the CAPRI challenge.[24–26] Since then, the protocol has been developed further, and here we present a complete description of the algorithm plus rigorous tests of its performance. Our goals are to develop a search strategy appropriate for global searches on unknown structural targets, and to assemble and optimize an energy function capable of differentiating decoys close to the native structure from those far away.

Results

Docking algorithm

Figure 1 shows a flowchart of the method. Creation of a decoy begins with a random orientation of each partner and a translation of one partner along the line of protein centers to create glancing contact between the proteins. In the first stage of the algorithm, we employ a rigid-body Monte Carlo search, translating and rotating one partner around the surface of the other through 500 Monte Carlo move attempts. Step sizes are adjusted continually to maintain a 50% move acceptance rate, with initial Gaussian perturbation sizes of mean value 0.7 Å (along all three axes) and 5° (spin around the axis of protein centers and tilt off this axis in a randomly-chosen direction). The low-resolution, residue-scale interaction potentials[20,21] are based on a Bayesian expansion of the probability of the correctness of each decoy. These potentials include residue environment and residue–residue interaction terms derived from a database of interfaces, a score to reward contacting residues, a penalty for overlapping residues and, for antibody targets, an alignment score to direct the interface toward the complementarity-determining

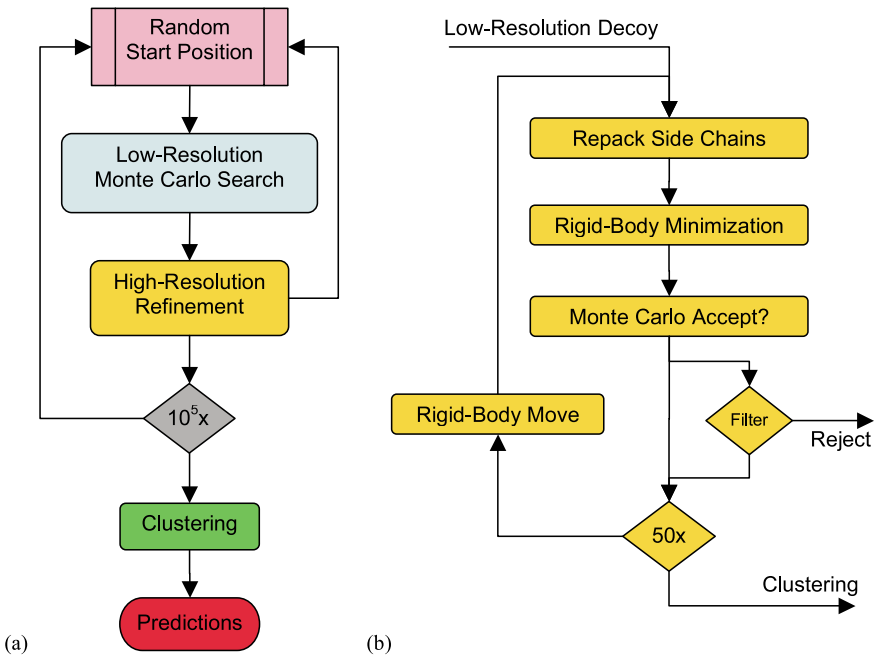

Figure 1. Docking protocol. (a) Process flowchart; (b) detail of the refinement stage.

loops. All scores at this stage are based on a reduced representation of the amino acid residues based on side-chain centroid positions.[20] These scores are described in detail in Materials and Methods.

After the low-resolution search, explicit side-chains are added to the protein backbones using a backbone-dependent rotamer packing algorithm.[27] Rotamer choices are created from nine χ_1 angles (three major rotamer angles and each of those angles plus and minus the standard deviation of that angle), three χ_2 angles (major rotamer angles), plus major rotamer options for χ_3 and χ_4 angles for appropriate residue types.[28] Polar hydrogen atoms are placed for use in hydrogen bond formation. The optimal combination of rotamers is found using a simulated-annealing Monte Carlo search.

Once the proteins have side-chains, the rigid-body displacement is optimized. The gradient of the scoring function determines the starting direction (in the rigid-body translation/rotation space) for a sequence of line minimizations. The Davidon–Fletcher–Powell quasi-Newton minimization technique[29] finds the local minimum of the energy function to within a given tolerance, here a loose 1.0 scoring unit.

To simultaneously optimize the side-chain conformations and the rigid-body position, the side-chain packing and minimization operations are repeated 50 times (Figure 2). Before each cycle, the position of one protein is perturbed by random translations of mean 0.1 Å in each direction of Cartesian space and by random rotations of mean 0.05° around each Cartesian axis. After each move, packing, and minimization, a score is calculated. The new position is kept or rejected according to the standard Metropolis acceptance criterion.[30] After the final cycle, the lowest-scoring conformation is minimized once more to a fine tolerance of 0.02 scoring unit. The repetition of rigid-body and side-chain conformational moves is new to this work; the predictions for the first CAPRI experiments were completed with a preliminary version of the protocol.[26]

Several measures are taken to ensure computational efficiency during these repeated cycles. First, the side-chain packing algorithm usually varies the conformation of only one residue at a time while keeping the other side-chains fixed; a

Figure 2. The voyage over the free energy surface during one refinement cycle. The steps are: (1) a random perturbation (rigid-body translation and rotation) moves the structure on the potential surface; (2) a packing step optimizes the side-chain positions, thus changing the energy surface; (3) an explicit minimization finds the nearest local minimum accessible *via* a rigid-body translation and rotation. Start and finish positions are compared by the Metropolis criterion, and the cycle is repeated 50 times.

full, combinatorial rotamer optimization is performed only once every eight cycles. Second, a filter is employed periodically to detect inferior decoys and reject them without further refinement: decoys are discarded if the alignment profile determines that the interface contains two or more forbidden residues (Materials and Methods), or if the score is not within 500, 10, and 5 scoring units of a reference score before the first minimization cycle, after the first minimization cycle, and after five minimization cycles, respectively. The reference score is the maximum score of the top 1% of 5000 decoys created in a calibration run.

In the packing and minimization refinement steps and for decoy discrimination, the full-atom scoring function includes van der Waals interactions with the repulsive part of the potential partially replaced with a linear term to avoid singularities; solvation using a pair-wise Gaussian solvent-exclusion model;[31] hydrogen bonding energies using an orientation-dependent function derived from high-resolution protein structures;[32] a rotamer probability term;[27,28] residue–residue pair interactions derived statistically from a database of protein structures;[27] a simple electrostatic term; and a surface area and atomic solvation term (for decoy discrimination only, due to the expense of calculation).[33] While the weights of most of the terms in the scoring function are of the same order of magnitude, the dominant contributions to discrimination are the van der Waals (packing) interactions, followed by solvation. Full scoring functions and their respective weights and contributions are detailed in Materials and Methods.

The search procedure is repeated to create approximately 10^5 decoys per target. Each final decoy is rescored, reducing the weight on the repulsive van der Waals energy term and including the surface area-based solvation term. The 200 best-scoring decoys are then clustered on the basis of pair-wise root-mean-squared distance (rmsd) using a hierarchical clustering algorithm.[34] Structures within a 2.5 Å clustering threshold are designated as a set, and the lowest-scoring decoy within the set represents that position. The clusters with the most members are selected as the final predictions, ranked according to the cluster sizes. The cluster size, or the degeneracy of the docked position, may be related to the entropy of the bound complex.

In addition to the global docking searches, we performed perturbation studies to explore the nature of the docking energy funnel. These studies were performed on both bound complexes with native side-chain conformations removed and on unbound monomer components superimposed on the native, bound complex structures. Random starting positions are created by translating one protein partner by Gaussian random distances of 3 Å standard deviation along the line of protein centers and 8 Å standard deviation in the two perpendicular directions, and by spinning the mobile partner by a Gaussian random angle of 8° standard deviation around the axis of centers and by tilting a Gaussian random angle of 8° standard deviation off the axis in a random direction. In this way, a set of 1000 random starting positions create a diffuse cloud that covers a reasonable area (~20 Å radius rmsd) with moderate density around the native ligand position. In the perturbation runs, the filtering and clustering steps are omitted.

Calculations were performed on computer clusters of approximately 50-processor Linux workstations with clock speeds near 1 GHz. Sets of 1000 perturbation decoys were created in one-half to two hours of cluster time per target. Complete processing of global searches (~10^5 decoys) required between one-half and two days of cluster time for each of the 54 targets, for a total computing time exceeding eight processor-years.

Perturbation studies

The docking problem requires two components: first, a search strategy must be able to create good decoys, and second, a scoring function must be able to differentiate structures close to and far from the native structure. To assess the accuracy of the scoring function somewhat independently of the search problem, we generated sets of structures enriched for near-native conformations by seeding the protocol with perturbed native structures, where the reference structure is either the bound complex itself or one or both unbound component(s) optimally superimposed upon on the bound complex coordinates. In both cases, the native crystal structure side-chain conformations were removed and regenerated by the program.

Bound backbones

Figures 3–5 show the results from various perturbation studies on the major classes of interfaces in Chen *et al.*'s benchmark set.[35] For each target, the all-atom score of 1000 decoys is plotted as a function of the rmsd from the native structure (calculated over the C^{α} atoms of the smaller docking partner in the fixed coordinate frame of the larger partner). If the search algorithm and scoring functions are successful, we expect that the lowest-scoring decoys will have low rmsds, creating a score "funnel." Indeed, in many cases funnels are apparent, and the widths of the binding funnels range from 2 Å to 6 Å rmsd. To quantify the presence of the funnel, we examine the five lowest-scoring decoys for each target. If three or more of those decoys are within 5 Å rmsd of the native structure, we say that the target exhibits a score funnel (Table 1). Of the 54 benchmark targets, 42 have funnels, including 21 of 22 enzyme–inhibitor pairs and ten of 16 antibody–antigen pairs (Table 2). Note that this is not the typical, trivial bound–bound docking problem: we have removed all side-chain information before starting the protocol. Therefore, although the perturbation studies

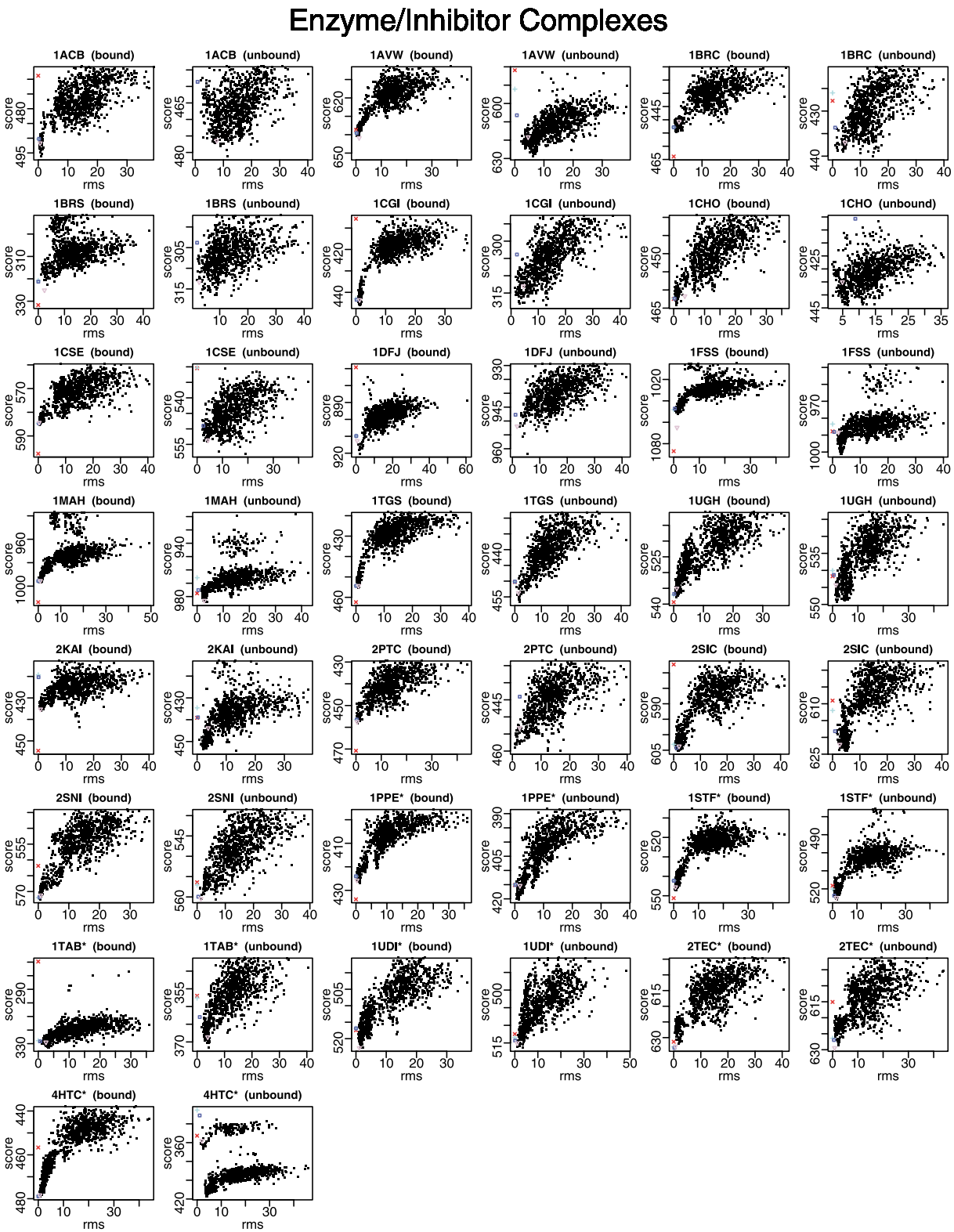

Figure 3. Perturbation studies on enzyme/inhibitor complexes. Plots show rmsd *versus* score for 1000 decoys created from random starting positions near the native structure. Bound indicates that backbone coordinates were taken from the bound complex; Unbound indicates that backbone coordinates were taken from one or both unbound monomer components. (×) Native structure, (+) native backbones with side-chains replaced using the algorithm, (□) that position after rigid-body minimization, (▽) after 50 cycles of packing and minimization, i.e. equivalent to the high-resolution refinement algorithm as applied to the docking decoys. Here and in subsequent Figures and Tables, asterisks (∗) indicate targets for which only one unbound monomer structure is known.

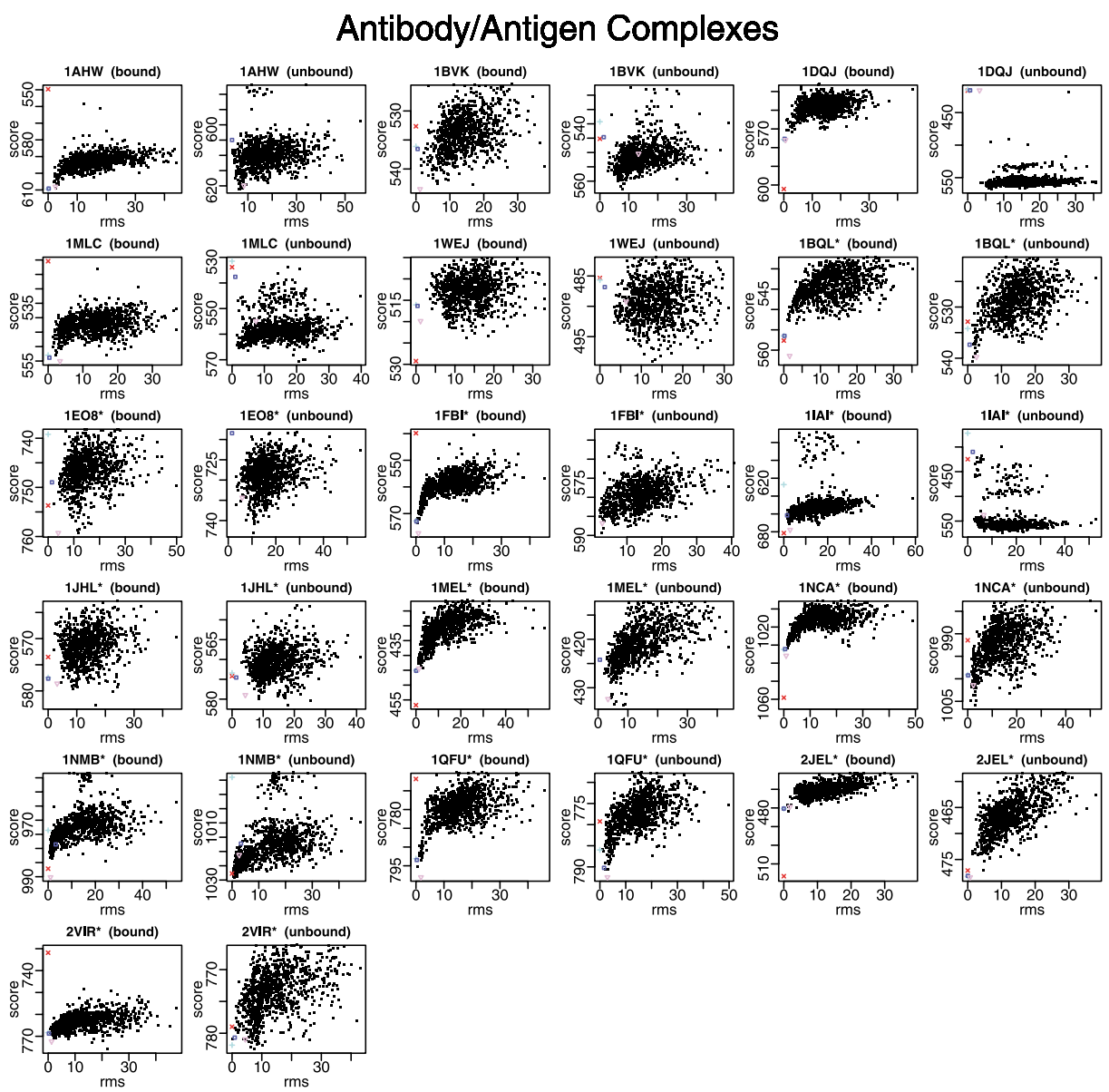

Figure 4. Perturbation studies on antibody/antigen complexes. See the legend to Figure 3.

were designed to isolate the score function problem, they also test the ability of the search algorithm to find the correct rigid-body displacement and side-chain conformations. In most cases, the program is able to re-create near-correct conformations. However, in some cases the search can be limiting: when we attempted perturbation studies on the antibody/antigen targets in which the antibody side-chains were fixed in the native positions (as given in the recent CAPRI experiment[24]), the number of funnels rose from ten to 14 of 16 cases. Therefore, finding correct side-chain packing is an issue in a few of the cases lacking funnels.

The native complex must have lower free energy than the many alternative decoy complexes, and this provides a test of the quality of the score function. In most cases, particularly those where funnels are present, the native structure scores as well as the best decoys or better (Figures 3–5). Scores are shown for three stages of refinement of the native structure: after removing the side-chains and replacing them with an optimal combination of rotamers, after performing a single rigid-body minimization on the repacked structure, and after performing a complete set of 50 refinement cycles in a treatment identical with that used for decoys. Typically, the minimization step improves the score by a small amount (0–2 score points), and the repeated refinement improves the score significantly more (1–10 + points). The difference between the native structure score and that with repacked side-chains, however, can be either positive or negative (and often of substantial absolute value, typically 20 points). Some unusually high-scoring native structures (e.g.

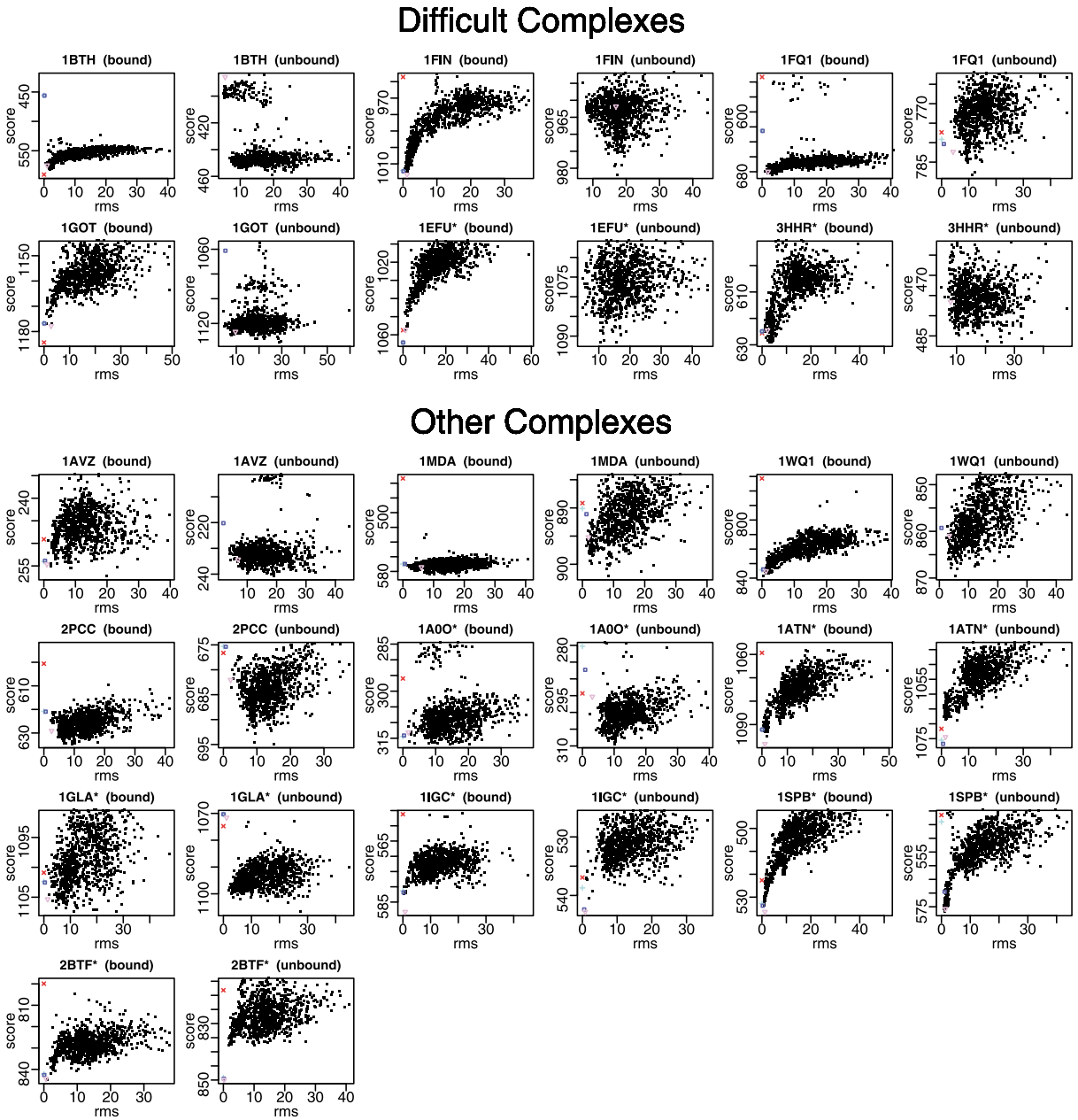

Figure 5. Perturbation studies on difficult and other complexes. See the legend to Figure 3.

1AHW or 1QFU), contain side-chains with non-standard χ angles that receive poor rotamer probability scores and result in artificially poor rankings. More often, the native structures exhibit exceptionally good packing, which is difficult to achieve in model structures, and even the optimal rotamer combination is not close to the native score (e.g. 1NCA). Finally, it should be noted that the refined native structure represents one of the best scores of which the program's refinement protocol is capable of generating, and its score relative to that of the other (false positive) decoys indicates whether correct solutions are possible to detect at all.

Figure 6 shows experimentally-measured binding affinities for targets for which they are available and correlates them with the difference in the score between the lowest-rmsd decoy from the bound perturbation studies and a reference score calculated when the two partners are in isolation. Targets that do and do not create funnels are noted. Funnels are observed consistently for targets with binding affinities K_a above approximately $10^7\,M^{-1}$ or, equivalently when the score difference exceeds 19 units. Furthermore, failures seem to be explained by their low binding affinities, below the level of uncertainty in the scoring functions. It is intriguing that the complexes for which clear binding funnels are not observed are the very low-affinity complexes: first, this suggests that the scoring function has some relation to the actual free energy function, and second, it suggests

Table 1. Summary of the docking of bound and unbound components from perturbation studies and global searches

	Perturbation studies					Global searches							
	Bound	Unbound											
PDB	*N5Å*	*N10Å*	*N25C*	Best contacts	Best rmsd	*R5Å*	*R10Å*	*R75C*	*R50C*	*R25C*	*R01C*	Best contacts	Best rmsd
A. *Enzyme/inhibitor complexes*													
1ACB	5 ·	2	1	0.462	9.08	>25	14	>25	14	14	5	0.051	11.84
1AVW†	5 ·	5 ·	5 ·	0.723	6.11	>25	2	>25	2	2	2	0.702	5.76
1BRC†	5 ·	1	2	0.938	3.76	>25	>25	>25	>25	>25	9	0.156	15.96
1BRS	4 ·	4 ·	4 ·	0.667	3.89	>25	2	>25	23	2	1	0.389	8.67
1CGI	5 ·	4 ·	4 ·	0.377	3.15	2	2	>25	>25	2	1	0.415	3.27
1CHO	5 ·	3 ·	3 ·	0.525	5.61	>25	2	>25	>25	2	2	0.425	6.88
1CSE	5 ·	2	0	0.233	8.65	>25	7	>25	>25	24	1	0.186	8.87
1DFJ	5 ·	4 ·	4 ·	0.581	3.92	>25	>25	>25	>25	>25	2	0.023	22.09
1FSS	5 ·	5 ·	5 ·	0.682	2.98	5	5	>25	5	5	1	0.636	3.19
1MAH†	5 ·	5 ·	5 ·	0.744	2.36	1	1	>25	1	1	1	0.641	2.40
1TGS†	5 ·	5 ·	5 ·	0.688	1.39	2	2	>25	2	2	1	0.562	3.31
1UGH	5 ·	5 ·	4 ·	0.676	1.69	1	1	>25	1	1	1	0.595	2.15
2KAI†	0	4 ·	4 ·	0.881	2.04	1	1	1	1	1	1	0.833	2.40
2PTC	5 ·	2	2	0.872	0.98	>25	>25	>25	>25	>25	5	0.179	10.35
2SIC	5 ·	5 ·	5 ·	0.841	2.41	2	2	2	2	2	2	0.818	2.98
2SNI	5 ·	4 ·	4 ·	0.881	1.61	7	5	7	5	5	1	0.857	1.36
1PPE*†	5 ·	5 ·	5 ·	0.881	0.61	2	1	>22	2	1	1	0.738	0.59
1STF*†	5 ·	5 ·	5 ·	0.897	1.31	3	3	3	3	3	3	0.795	0.56
1TAB*†	5 ·	5 ·	5 ·	0.727	3.67	1	1	>25	1	1	1	0.636	3.35
1UDI*†	5 ·	5 ·	5 ·	0.694	1.16	8	2	>25	2	2	2	0.500	3.75
2TEC*†	5 ·	5 ·	5 ·	0.810	2.25	3	3	>25	3	3	1	0.690	2.09
4HTC*†	5 ·	5 ·	5 ·	0.591	3.81	1	1	>25	1	1	1	0.591	3.84
B. *Antibody/antigen complexes*													
1AHW	5 ·	5 ·	5 ·	0.533	6.17	>25	3	>25	>25	3	1	0.489	7.24
1BVK	0	5 ·	0	0.241	6.25	>25	1	>25	6	3	1	0.517	6.15
1DQJ	4 ·	2	2	0.347	5.86	>25	>25	>25	>25	>25	20	0.000	17.75
1MLC	4 ·	0	0	0.182	18.89	6	6	>25	6	6	1	0.515	2.52
1WEJ	0	0	2	0.250	10.65	>25	>25	>25	>25	>25	2	0.031	17.90
1BQL*	4 ·	5 ·	5 ·	0.794	1.40	1	1	>25	1	1	1	0.676	2.09
1EO8*	0	1	4 ·	0.406	7.38	>25	>25	>25	>25	>25	1	0.062	56.51
1FBI*	5 ·	3 ·	3 ·	0.545	2.71	18	18	>25	18	18	2	0.030	22.25
1IAI*†	2	0	1	0.306	14.11	>25	>25	>25	>25	>25	>25	0.000	33.96
1JHL*	0	1	0	0.217	8.56	>25	15	>25	>25	15	1	0.174	14.91
1MEL*†	5 ·	5 ·	5 ·	0.528	5.38	>25	1	>25	8	1	1	0.556	5.80
1NCA*	3 ·	5 ·	5 ·	0.750	1.24	>25	>25	4	4	4	1	0.750	31.47
1NMB*†	4 ·	5 ·	5 ·	0.818	0.65	13	13	>25	>25	>25	2	0.030	72.48
1QFU*	5 ·	5 ·	5 ·	0.636	2.00	4	4	>25	4	4	4	0.515	3.20
2JEL*	3 ·	5 ·	4 ·	0.727	4.71	6	1	10	3	1	1	0.848	1.45
2VIR*†	0	4 ·	1	0.483	6.12	>25	>25	>25	>25	>25	7	0.172	41.47
C. *Other complexes*													
1AVZ	1	0	0	0.080	10.14	>25	>25	>25	>25	>25	19	0.000	20.03
1MDA	0	3 ·	0	0.083	7.86	>25	>25	>25	>25	>25	19	0.000	53.71
1WQ1	5 ·	3 ·	4 ·	0.511	3.99	>25	23	>25	23	23	13	0.000	44.03
2PCC	0	3 ·	1	0.444	7.28	>>25	>25	>25	>25	>25	8	0.056	19.07
1A0O*	0	1	4 ·	0.478	8.22	>25	10	>25	>25	1	1	0.391	5.08
1ATN*	5 ·	5 ·	5 ·	0.875	0.96	1	1	1	1	1	1	0.800	2.30
1GLA*	0	1	1	0.737	5.95	>25	>25	>25	>25	>25	9	0.158	25.66
1IGC*	5 ·	2	2	0.875	1.92	>25	>25	>25	>25	>25	3	0.062	24.19
1SPB*	5 ·	5 ·	5 ·	0.733	0.91	4	4	>25	4	4	4	0.717	1.32
2BTF*	4 ·	4 ·	4 ·	0.762	1.90	>25	>25	>25	>25	>25	4	0.119	27.52
D. *Difficult test cases*													
1BTH†	5 ·	0	1	0.254	18.23	>25	>25	>25	>25	>25	3	0.017	16.96
1FIN	5 ·	0	0	0.125	16.56	>25	>25	>25	>25	>25	4	0.014	26.84
1FQ1	5 ·	2	2	0.387	9.61	>25	>25	>25	>25	>25	16	0.000	24.64
1GOT	5 ·	0	0	0.062	19.27	22	22	>25	>25	>25	1	0.094	55.84
1EFU*	5 ·	0	0	0.087	10.64	16	16	>25	>25	>25	12	0.000	29.51
3HHR*	5 ·	0	0	0.041	10.75	>25	>25	>25	>25	>25	>25	0.000	33.90
Totals	42/54	34/54	32/54	43/54	45/54	20/54	28/54	7/54	23/54	28/54	46/54	28/54	28/54
†Subset	13/16	13/16	12/16	16/16	14/16	10/16	11/16	2/16	11/16	11/16	15/16	11/16	11/16

For the perturbation studies on bound backbones, *N5*Å is the number of the top five decoys (by all-atom score) that have rmsd <5 Å. For the perturbation studies on unbound components, *N10*Å is the number of the top five decoys that have rmsd <10 Å, *N25C* is the number of the top five decoys that correctly predict more than 25% of native residue–residue contacts, *best contacts* is the best fraction of native residue–residue contacts in the top five decoys, and *best rmsd* is the best rmsd in the top five decoys. For global unbound searches: *R5*Å and *R10*Å are the ranks of the first clusters with rmsd <5 Å or <10 Å, respectively; *R75C*, *R50C*,

Table 2. Correct predictions by interface type

	All			Unbound–unbound			Semibound		
	BB-p	XU-p	XU-g	BB-p	UU-p	UU-g	BB-p	BU-p	BU-g
Enzyme/inhibitor	21/22	18/22	17/22	15/16	12/16	11/16	6/6	6/6	6/6
Antibody/antigen	10/16	9/16	8/16	3/5	1/5	3/5	7/11	8/11	5/11
Other	5/10	5/10	3/10	1/4	1/4	0/4	4/6	4/6	3/6
Difficult	6/6	0/6	0/6	4/4	0/4	0/4	2/2	0/2	0/2
Total	42/54	32/54	28/54	23/29	14/29	14/29	19/25	18/25	14/25

BB-p denotes bound perturbation studies; UU-p, BU-p and XU-p denote perturbation studies on unbound, semibound, and all targets; and UU-g, BU-g, and XU-g denote global searches on unbound, semibound, and all targets. XU in columns two and three indicates that unbound coordinates were used for both partners for the 29 targets (columns 4–6) for which they were available, and for one of the two partners for the 24 targets (columns 7–9) where only one uncomplexed structure was available. Successful perturbation runs have three of five top-scoring decoys with rmsd <5 Å (BB) or three of five top-scoring decoys with greater than 25% contacts (UU or BU). Global search success means that a decoy of rank 10 or less contains at least 25% of correct native residue–residue contacts.

that the method could ultimately be useful in determining whether two proteins of known structure interact with high affinity.

Unbound backbones

Using coordinates from independently determined X-ray structures, score funnels still exist for many of the targets (Figures 3–5), although not as frequently. The "unbound" problem is considerably more difficult (and more realistic) than the bound component problem, as there may be significant changes in backbone conformation upon binding and hence the unbound components may have much less of a lock-and-key fit than the bound components.

Unbound perturbation funnels are quantified by counting structures that either have rmsd less than 10 Å or, alternatively, a native residue–residue contact fraction above 25% (a residue–residue contact is identified when any of the inter-residue atom–atom distances is less than 4 Å, excluding hydrogen atoms). Table 1 shows the number of decoys meeting these criteria among the five top-scoring decoys from the unbound perturbation studies. In addition, the Table lists the best rmsd and best fraction of native residue–residue contacts found in the five top-scoring decoys. Of the 54 targets, 32 display funnels as measured by the 25% contact fraction measure. Since the perturbation studies are designed to sample the configuration space close to the native structure, this measure represents an upper limit to the fraction of successes expected from the protocol in a global search.

Scores for the unbound "native" structures are based on the conformation created by superimposing the unbound monomer components onto the bound complex.[35] Sometimes, this superposition does not create a good match at the interface (even after repacking the side-chain conformations); however, once refined, the structure can provide a much better scoring target. For example, the unbound "native" antigen/antibody pair 1AHW moves 8 Å upon refinement, and the score drops over 50 points to the level of the bottom of the potential well. For this reason, a 10 Å limit is more realistic to detect the presence of a score funnel.

Global searches using unbound components

The results for global searches on the 54 targets from fully randomized starting positions are detailed in Table 1, using rmsd or the fraction of correct native residue–residue contacts. Ranks of the first predicted structure (after clustering, ranked by cluster size) that meet the given criteria reveal the performance of the algorithm. The native residue–residue contact fraction measures whether specific interactions occur and is a fairly strict test: of the 32 cases for which binding funnels exist, only seven predict at least 75% of the native contacts correctly in a model of rank 10 or less. However, 23 of 32 cases predict at least 50% of the contacts, and 28 cases predict more than 25%. Finally, even when a high level of detail is not predicted, our algorithm quite often (in 46 cases) produces a structure of rank 10 or less that has enough of the interface patches matched to make at least one correct contact. The scoring functions are sensitive enough to detect the general recognition region. We can measure the results by rmsd, although this is not the most precise measure, since it relies on the imperfect superposition of backbones. A total of 20 targets can be

R25C and *R01C* are the ranks of the first clusters with more than 75%, 50%, 25% and 1% of native residue–residue contacts; *best contacts* is the best fraction of native residue–residue contacts in the top ten clusters; and *best rmsd* is the best rmsd in the top ten clusters. Totals count either the number of targets that have top five measures (*N5*Å, *N10*Å, *N25C*) greater than or equal to 3 (indicated by ·), rank measures (*R5*Å, *R10*Å, *R75C*, *R50C*, *R25C*, *R01C*) less than or equal to 10, *best contacts* greater than or equal to 0.25, or *best rmsd* less than or equal to 10 Å. The independent subset totals on the bottom line include the targets, indicated with a dagger (†), which were not used in determining the score weightings.

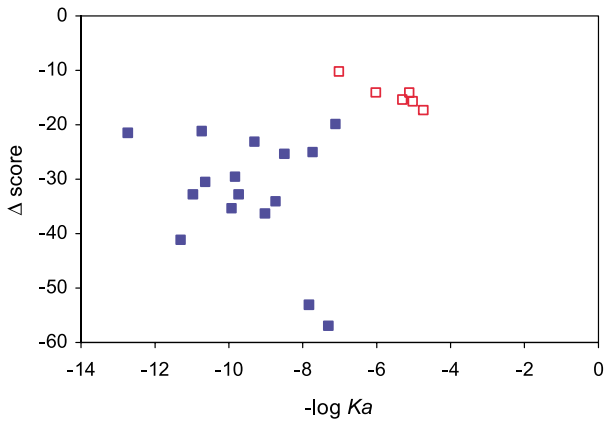

Figure 6. Negative logarithm of the experimental binding equilibrium constant, K_a, *versus* the difference in score between the lowest-rmsd docked complexes from bound perturbation studies and the separate monomer components for targets for which the binding affinities are known.[45–60] Filled and open symbols represent targets for which binding funnels were and were not observed, respectively.

predicted within 5 Å with at least one of the top ten models, and 28 can be predicted within 10 Å.

Results for enzyme–inhibitor pairs are particularly successful: of the 22 enzyme–inhibitor pairs, 14 structures can be predicted in global unbound searches with at least 50% of native residue–residue contacts, and 17 can be predicted with at least 25% native contacts; this is all but one of the structures for which binding funnels are present in the unbound perturbation studies.

Results for antibody–antigen pairs are encouraging. Of the 16 targets, ten can be docked from bound backbones, and nine can be docked from unbound backbones in the perturbation studies. In global searches, eight of the 16 targets can be predicted successfully at 25% native residue–residue contacts or more, with models of rank 10 or less. Two of the five unbound–unbound targets exhibit a funnel, and three can be found in a global search to 25% native residue–residue contacts in ten models.

The targets in the "other" category are challenging, particularly on the global search. Although five of the ten targets exhibit funnels in both the bound and unbound cases, only three targets can be predicted successfully in a global search. The "difficult" targets are so designated because they exhibit a significant backbone conformational change upon docking. As might be expected when not modeling any backbone movement, the program is unable to re-dock any of these five targets in the unbound perturbation studies, even though all exhibit funnels in the bound backbone perturbation studies.

Table 2 separates the unbound–unbound targets (for which both docking partners have monomer coordinates available) from the semibound targets (for which one docking partner's backbone was taken from the co-crystal coordinates, designated with an asterisk in the Figures and Tables). The overall success rates do not differ by much: 14 of 29 global searches correctly predicted 25% of the native residue–residue contacts in the unbound–unbound cases, compared with 14 of 25 semibound cases. In the various categories, the differences were more marked. All of the four failures in the enzyme–inhibitor category were of the unbound–unbound type. The antibody/antigen category actually had a higher success fraction for the unbound–unbound cases (three of five *versus* five of 11), although the number of unbound–unbound targets is small. None of the unbound–unbound targets was predicted correctly in the difficult or other category, reflecting the challenge created by the larger backbone movement in these categories.

Figure 7 shows pictures of the best predictions superimposed with the native structures for selected targets in each category. The enzyme–inhibitor predictions are excellent: in 1AVZ, 1CGI, and 2SIC, the native structure and the prediction overlay almost perfectly. Although the 1BRS fit is less exact, it matches 33% of the native residue–residue contacts while the inhibitor is somewhat twisted within the active site of the receptor. The program has recognized the general shape and chemical complementarity but has not fit the interface exactly. Both unbound–unbound and semibound antibody/antigen targets match closely, as well as two semibound targets from the "other" category. This Figure displays the encouraging performance of the algorithm on a wide variety of interface types.

Discussion

The algorithm here imitates the physical process of protein docking. Camacho hypothesizes that docking occurs in two stages.[36] During a "recognition" stage, the two proteins diffuse near each other until the interface patches come sufficiently close to begin the "binding" stage, when a high-affinity interaction is formed by modification of the side-chain and backbone conformations. The low-resolution search stage simulates the recognition process, capturing short individual encounters between the proteins diffusing in a solution with a Monte Carlo search of translations and rotations. The high-resolution refinement simulates the binding stage, as side-chain conformations are rearranged while the protein simultaneously can adjust its rigid-body position. This framework for modeling protein interactions should provide a basis for incorporating further realistic motions of the protein, such as local backbone flexibility and even global rearrangement, if the computational requirements can be made more tractable.

The key elements to the algorithm's success are rapid searching during the first search stage enabled by low-resolution protein representations and scoring functions, the simultaneous optimization of side-chain configurations and rigid-body

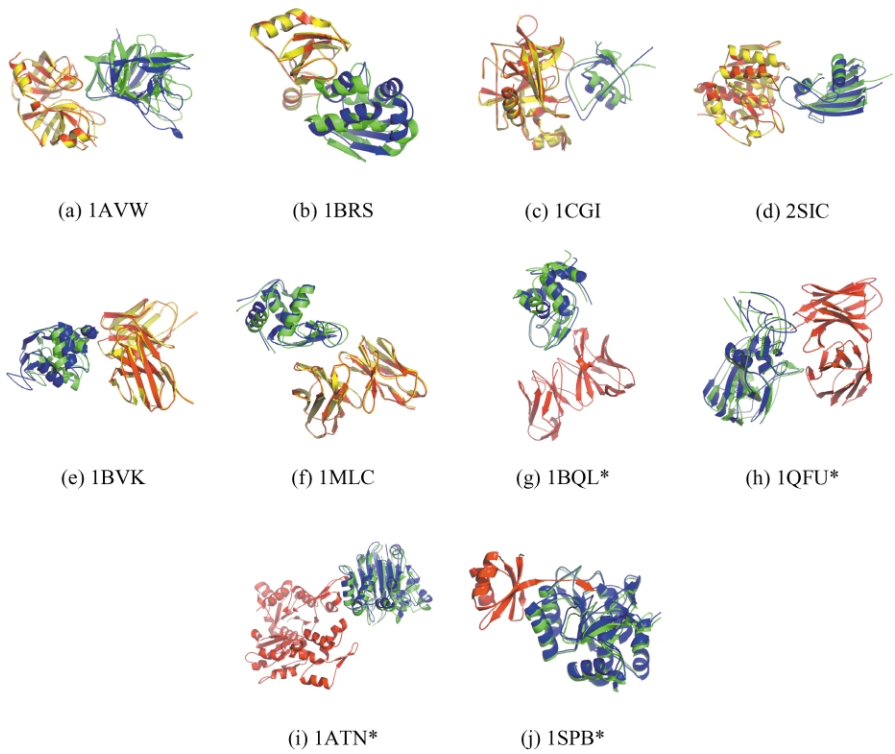

Figure 7. Selected structures from global prediction runs. Red and blue indicate the experimental co-crystal. Yellow and green indicate the best prediction of rank less than 10 (determined by global rmsd). (a)–(d) Enzyme–inhibitor systems; (e)–(h) antibody–antigen systems; (i)–(j) other systems.

displacement in the refinement, the free energy function, the large numbers of decoys generated, and the decoy clustering. This approach differs significantly from most other docking algorithms.

The existence of the funnels in the bound cases and the correlation between experimental binding affinities and simulation binding scores (Figure 6) indicates that the scoring function is beginning to approach a true free energy function. The score is dominated by packing interactions, solvation, rotamer probability, and hydrogen bonding, and includes contributions from electrostatics (Table 4, Materials and Methods). Each of these contributions is important for effective discrimination of good and poor decoys. However, the repulsive van der Waals interactions are particularly important. The steep function employed here (Materials and Methods) increases the decoy enrichment significantly better than previously used modified van der Waals potentials with low penalties for overlap.[22,27]

The search strategy and the scoring function are intimately linked. With a strict van der Waals function, it is difficult to search conformation space due to the large energetic barriers between many free energy minima. The refinement algorithm resolves the search problem and is tailored for the strict repulsive van der Waals potential. First, by using a large number of rotamers in the packing algorithm, realistic side-chain conformations that avoid steric clashes can be found for correct interface positions, while incorrect structures are likely to contain atomic overlaps as well as voids at the interface. Second, the Monte Carlo minimization cycles, with simultaneous optimization of the side-chain conformations and the rigid-body displacement, efficiently explore a local area of conformation space to find the lowest of the nearby minima in a very noisy energy landscape.

The results presented here outline challenges for future exploration. Success ratios near 50% while considering up to ten models leave much uncertainty for biologists or engineers interested in using the predicted structures. Failures may occur for a variety of reasons. The most likely reason is familiar to all who model biomolecules: insufficiently accurate free energy functions. The score function could yet be missing a contribution from a physical component that is essential for a particular target. This might include explicit water molecules that make specific hydrogen bonds not described accurately by the continuum solvent model, or interactions with ions or small ligands such as phosphate or heme groups. Several of the score function components have some redundancy (such as solvation and electrostatics), and certainly electrostatics is treated in a rudimentary fashion. In addition, the scoring may not be tuned optimally for targets with particular interface types. Finally, Figure 6 shows that our algorithm fails consistently for weak binding pairs. Any of the previously mentioned issues would be exacerbated in the case of a weak binding interaction, or it may be that native-like configurations simply do not

exhibit enough of a favorable score relative to the noise in the current scoring function. In any case, energy function issues can be best isolated by examining the bound backbone perturbation studies, where other uncertainties are controlled. Further individual examination of targets that do not exhibit funnels in the bound backbone studies (2KAI, 1BVK, 1WEJ, etc.) may lead to the identification of the most important shortcomings.

Similarly, the search algorithm may be responsible for cases that fail. For some antibody targets (the unbound 1WEJ and 1DQJ), the search is clearly insufficient because the low-resolution stage does not create many decoys close to the native position, likely prohibiting the high-resolution refinement from finding the binding funnel. In some cases, the rotameric representation may be insufficient; a finer representation may be necessary to achieve correct packing. In particular, atoms at the ends of long rigid side-chains such as the aromatics move significantly when represented by rotamers. These search problems complicate the evaluation of a free energy function; if good positions are not sampled, it is impossible to tune a scoring function properly.

Other researchers have noted difficulty in predicting cases with significant backbone motion.[15] While the algorithm is successful in predicting many targets using the unbound backbones, it includes no means to adjust the backbone conformation. It is completely unable to predict the six difficult targets and there were several others (1ACB, 1BRC, 1CSE, etc.) that exhibited funnels for the bound backbones but not for the unbound backbones. These targets would make good cases for further studies with a generalized algorithm capable of relaxing the backbone coordinates.

The current protocol produces predictions containing between 35% and 80% of native residue–residue contacts. On one hand, it is encouraging that the free energy function can discriminate such positions from false-positive structures. The model structures could represent pre-docked conformations before the full binding stage is complete. Since our algorithm relies on clustering, it could be detecting positions with high degeneracy (large configurational entropy). The interfaces could exhibit multiple conformations *in vivo*, and the models represent some of the non-crystallographic conformations of the docking (or pre-docking) interaction. On the other hand, it is likely that better modeling of the binding phase (with a more accurate all-atom potential function and a deeper search strategy) will lead to higher-quality predictions with more correct contacts.

Although it is hard to compare studies that use different measures of the quality of docking algorithms, our method's success rate near 50% for unbound–unbound and bound–unbound targets (using the 25% or 50% native residue–residue contacts criterion) is similar to that found in other leading studies. The method of Fernández-Recio *et al*. predicted correct solutions in the top 20 models in 17 of 24 cases, and Palma *et al*. found correct solutions of rank 20 or less in 14 of 25 cases. Our method finds 25 of 54 in the top 20 with 50% of contacts, and 31 of 54 with 25% contacts.

Conclusion

We have created a new docking algorithm that mimics the physical process of docking, in that it contains a low-resolution recognition stage and a high-resolution binding stage. The high-resolution refinement simultaneously optimizes the rigid-body displacement and the side-chain conformations. To our knowledge, this study is the largest scale benchmark test to date. Perturbation studies reveal docking funnels in 42 of 54 cases using bound backbones and 32 of 54 cases using one or both unbound backbones. In global searches, one of the top ten models predicts correctly more than 25% of the native residue–residue contacts in 28 of 54 cases. Still, despite progress, this work highlights the challenges associated with modeling biomolecular systems. The search space is large and highly corrugated, and the free energy function used is becoming more accurate but is still insufficient for some systems. Protein docking continues to be a significant scientific challenge to structural biologists and the biomolecular modeling community.

Materials and Methods

Benchmark set

Targets are selected from the benchmark set described by Chen *et al.*† (and see Table 3).[35] This set includes enzymes and inhibitors, antibodies and antigens, several "difficult" cases which have significant backbone conformational change in the binding interfaces, and some others. Targets in the benchmark set were selected for the availability of structures of both the bound complex and the unbound components. Since the number of such targets is small, the set is supplemented with structures for which one of the two partners has a known unbound structure.

Low-resolution scoring functions

In the low-resolution representation, each residue is represented by the four backbone atoms (N, C^{α}, C and O) and one pseudo-atom, the "centroid," to represent the side-chain. The location of the centroid, determined using known structures from the PDB, is the average location of the side-chain atoms in residues of the same identity and with ϕ and ψ angles in the same $10^{\circ} \times 10^{\circ}$ bin.[20]

Here, we detail the development of a low-resolution (residue scale) potential function. In searching for a correctly docked structure, one would like to maximize the probability of the complex structure being correct given

† http://zlab.bu.edu/~rong/dock/benchmark.shtml

Table 3. Targets in the benchmark set[35]

Complex	Receptor	Ligand	Description	ΔG (kcal/mol)	K_a (1/M)	Ref.
A. *Enzyme/inhibitor complexes*						
1ACB(E:I)	5CHA(A)	1CSE(I)	α-Chymotrypsin/Eglin C	−13.0	5.0×10^9	44
1AVW(A:B)	2PTN	1BA7(A)	Trypsin/Soybean trypsin inhibitor	−12.3	*1.1×10^9*	45
1BRC(E:I)	1BRA	1AAP(A)	Trypsin/APPI			
1BRS(A:D)	1A2P(B)	1A19(A)	Barnase/Barstar	−17.2	*4.5×10^{12}*	45
1CGI(E:I)	1CHG	1HPT	α-Chymotrypsinogen/Human pancreatic secretory trypsin inhibitor			
1CHO(E:I)	5CHA(A)	2OVO	α-Chymotrypsin/Ovomucoid third domain			
1CSE(E:I)	1SCD	1ACB(I)	Subtilisin Carlsberg/Eglin C	−13.1	6.6×10^9	44
1DFJ(I:E)	2BNH	7RSA	Ribonuclease inhibitor/Ribonuclease A			
1FSS(A:B)	2ACE(E)	1FSC	Acetylcholinesterase/Fasciculin II	*−14.9*	9.1×10^{10}	46
1MAH(A:F)	1MAA(B)	1FSC	Mouse acetylcholinesterase/inhibitor			
1TGS(Z:I)	2PTN	1HPT	Trypsinogen/Pancreatic secretory trypsin inhibitor			
1UGH(E:I)	1AKZ	1UGI(A)	Human uracil-DNA glycosylase/inhibitor			
2KAI(AB:I)	2PKA(XY)	6PTI	Kallikrein A/Trypsin inhibitor			
2PTC(E:I)	2PTN	6PTI	β-Trypsin/Pancreatic trypsin inhibitor	−14.2	4.0×10^{10}	47
2SIC(E:I)	1SUP	3SSI	Subtilisin BPN/Subtilisin inhibitor			
2SNI(E:I)	1SUP	2CI2(I)	Subtilisin Novo/Chymotrypsin inhibitor 2			
1PPE*(E:I)	2PTN	1PPE(I)	Trypsin/CMT-1			
1STF*(E:I)	1PPN	1STF(I)	Papain/Stefin B	*−13.5*	8.3×10^9	48
1TAB*(E:I)	2PTN	1TAB(I)	Trypsin/BBI			
1UDI*(E:I)	1UDH	1UDI(I)	Virus uracil-DNA glycosylase/inhibitor			
2TEC*(E:I)	1THM	2TEC(I)	Thermitase/Eglin C			
4HTC*(LH:I)	2HNT(LCEF)	4HTC(I)	α-Thrombin/Hirudin	*−15.4*	2.0×10^{11}	49
B. *Antibody/antigen complexes*						
1AHW(DE:F)	1FGN(LH)	1BOY	Antibody Fab 5G9/Tissue factor	*−11.5*	2.9×10^8	50
1BVK(DE:F)	1BVL(LH)	3LZT	Antibody Hulys11 Fv/Lysozyme			
1DQJ(AB:C)	1DQQ(LH)	3LZT	Hyhel-63 Fab/Lysozyme			
1MLC(AB:E)	1MLB(AB)	1LZA	IgG1 D44.1 Fab fragment/Lysozyme	−9.7	*1.4×10^7*	45
1WEJ(LH:F)	1QBL(LH)	1HRC	IgG1 E8 Fab fragment/Cytochrome *c*	−9.5	*9.7×10^6*	45
1BQL*(LH:Y)	1BQL(LH)	1DKJ	Hyhel-5 Fab/Lysozyme	−14.5	*4.6×10^{10}*	45
1EO8*(LH:A)	1EO8(LH)	2VIU(A)	Bh151 Fab/Hemagglutinin			
1FBI*(LH:X)	1FBI(LH)	1HHL	IgG1 Fab fragment/Lysozyme			
1IAI*(MI:LH)	1AIF(LH)	1IAI(LH)	IgG1 Idiotypic Fab/Igg2A anti-idiotypic Fab			
1JHL*(LH:A)	1JHL(LH)	1GHL(A)	IgG1 Fv Fragment/Lysozyme			
1MEL*(B:M)	1MEL(B)	1LZA	Vh single-domain antibody/Lysozyme	*−10.5*	5.0×10^7	51
1NCA*(LH:N)	1NCA(LH)	7NN9	Fab NC41/Neuraminidase			
1NMB*(LH:N)	1NMB(LH)	7NN9	Fab NC10/Neuraminidase			
1QFU*(LH:A)	1QFU(LH)	2VIU(A)	Igg1-κ Fab/Hemagglutinin			
2JEL*(LH:P)	2JEL(LH)	1POH	Jel42 Fab fragment/A06 Phosphotransferase			
2VIR*(AB:C)	2VIR(AB)	2VIU(A)	Igg1-λ Fab/Hemagglutinin			
C. *Other complexes*						
1AVZ(B:C)	1AVV	1SHF(A)	HIV-1 NEF/FYN tyrosin kinase SH3 domain	*−6.4*	*$<5 \times 10^4$*	52
1MDA(LH:A)	2BBK(LH)	1AAN	Methylamine dehydrogenase/Amicyanin	*−7.3*	2.2×10^5	53
1WQ1(G:R)	1WER	5P21	RAS activating domain/RAS			
2PCC(A:B)	1CCA	1YCC	Cytochrome *c* peroxidase/Iso-1-cytochrome *c*	−7.0	*1.4×0^5*	45
1A0O*(A:B)	1CHN	1A0O(B)	Che Y/Che A	−8.1	*9.1×10^5*	45
1ATN*(A:D)	1ATN(A)	3DNI	Actin/Deoxyribonuclease I	*−11.8*	5.0×10^8	54
1GLA*(G:F)	1GLA(G)	1F3G	Glycerol kinase/GSF III	*−6.7*	9.1×10^4	55
1IGC*(LH:A)	1IGC(LH)	1IGD	IgG1 Fab fragment/Protein G	*−12.6*	2.0×10^9	56
1SPB*(S:P)	1SUP	1SPB(P)	Subtilisin/Subtilisin prosegment			
2BTF*(A:P)	2BTF(A)	1PNE	β-Actin/Profilin			
D. *Difficult test cases*						
1BTH(LH:P)	2HNT(LCEF)	6PTI	Thrombin mutant/Pancreatic trypsin inhibitor			
1FIN(A:B)	1HCL	1VIN	CDK2 cyclin-dependant kinase 2/Cyclin	*−9.9*	2.1×10^7	57
1FQ1(B:A)	1B39(A)	1FPZ(F)	CDK2/KAP			
1GOT(A:BG)	1TAG	1TBG(AE)	Transducin Gt-α, Gi-α chimera/Gt-β-γ			
1EFU*(A:B)	1D8T(A)	1EFU(B)	*E. coli* EFtu/Efts	*−10.7*	7.0×10^7	58
3HHR*(B:A)	3HHR(B)	1HGU	Human growth hormone/Receptor	−11.7	6.3×10^7	59

The Table shows targets in the benchmark set (http://zlab.bu.edu/~rong/dock/benchmark.shtml), as well as experimentally measured binding free energies and equilibrium constants. ΔG and K_a values in standard type are those obtained from the literature; values in italics are calculated assuming $T = 298.15$ K.

the sequences and structures of the monomers, P(Cstr|seq,Mstr). We follow the Bayesian decomposition:

$$P(\text{Cstr}|\text{seq},\text{Mstr}) = \frac{P(\text{seq}|\text{Cstr},\text{Mstr})P(\text{Cstr}|\text{Mstr})}{P(\text{seq}|\text{Mstr})} \quad (1)$$

P(seq|Cstr,Mstr) is the probability of a given complex structure having a particular amino acid sequence; this term isolates the chemical component of the score and can be expanded further. P(Cstr|Mstr) is the probability of the complex structure given the monomer structures, and it contains the characteristics of well-docked complexes, e.g. shape complementarity. P(seq|Mstr), the probability of the sequences given the structures of the monomers, is constant and can be neglected in the analysis. In practice, the negative natural logarithm of the probability is used as a score for the Monte Carlo search. The score is decomposed into residue environment and pair terms (S_{env}, S_{pair}) which arise from P(seq|Cstr,Mstr) and contact, bump, and alignment terms (S_{contact}, S_{bump}, S_{align}) which arise from P(Cstr|Mstr):

$$\begin{aligned} -\ln P(\text{Cstr}|\text{seq},\text{Mstr}) &= S \\ &= S_{\text{env}} + S_{\text{pair}} + S_{\text{bump}} + 2S_{\text{contact}} + S_{\text{align}} \end{aligned} \quad (2)$$

Residue environment and residue pair potentials

The chemical or sequence-specific terms arise from a series expansion of P(seq|Cstr,Mstr):[20]

$$\begin{aligned} &P(\text{seq}|\text{Cstr},\text{Mstr}) \\ &\cong \prod_{aa_i \in \text{seq}_1,\text{seq}_2} P(aa_i|E_i) \\ &\times \prod_{aa_j \in \text{seq}_1, aa_k \in \text{seq}_2} \frac{P(aa_j, aa_k|\text{int}, E_j, E_k)}{P(aa_j|\text{int}, E_j, E_k)P(aa_k|\text{int}, E_j, E_k)} \end{aligned} \quad (3)$$

The first product is a residue environment term, and the second is a residue pair term. The environment term comprises amino acid propensities that capture the solvation effects that drive the burial of hydrophobic residues. For docking, we defined four environments using two criteria: a residue in a protein complex can be either buried or exposed to solvent, and either at the interface or not at the interface of the complex. A buried residue has at least 16 neighbors (defined as the residues with their centroid pseudo-atom within 10 Å of the centroid of the original residue). An interface residue is defined as a residue whose centroid lies within 6 Å of any residue centroid of the other docking partner. The cutoff of 6 Å was chosen because it best matches an atomic-resolution criterion of defining interface residue pairs as those with at least one atom within 4 Å of an atom in the other docking partner. The advantage of the centroid-based criterion is that it can be calculated using the low-resolution protein representation (i.e. without explicit positions of the side-chain atoms), thus allowing for a fast first stage of the algorithm.

The second-order term in equation (3) captures the specific chemical interactions between amino acid groups across the interface. The numerator, $P(aa_j, aa_k|\text{int}, E_j, E_k)$, is the frequency of interactions between residues of type j and k in a calibration set of interfaces, and the denominator normalizes by the total frequencies of residues of type j or k in the set of interface residue pairs. Again, a residue is designated at the interface if its centroid is within 6 Å of any centroid in the other protein partner. By defining potentials through the environment/pair decomposition, the dominant solvation contributions due to burial do not overwhelm the more subtle, environment-dependent pair interactions. (Other studies prepare pair potentials without this decomposition.[37])

Environment and pair probabilities were calculated on a large, non-redundant set of interfaces described by Glaser *et al.*[38] Since antibody–antigen interfaces have a chemical composition different from other transient biological interfaces, environment statistics for this category of interfaces were calculated on a separate subset of antibody–antigen interfaces only (see below).

Contact and bump scores

The structural part of a decoy's probability can be approximated by rewarding contacts at the interface and penalizing overlaps. The contact and bump scores roughly capture the attractive and repulsive part of the van der Waals interactions between proteins. Since attractions between proteins maximize the number of residues at the interface, a contact score is calculated as:

$$S_{\text{contact}} = \begin{cases} 12 & \text{if } N_{\text{int}} = 0 \\ 9.5 & \text{if } N_{\text{int}} = 2 \\ 10 - 0.5N_{\text{int}} & \text{if } 2 < N_{\text{int}} < 20 \\ -10 & \text{if } N_{\text{int}} \geq 20 \end{cases} \quad (4)$$

where N_{int} is the number of residues at the interface, defined by the 6 Å centroid–centroid distance criterion. The score is linear with interface residues except that extra penalties discourage the complete separation of the docking partners and interfaces larger than 20 residues are not favored additionally.

The bump score detects steric clashes between residues. A minimal approach distance $d_{\alpha\beta}$ for each type of atom–atom pair (where the types of atom, α or β, include the four backbone atoms plus each of the 20 types of centroid pseudo-atom) is determined using statistics from high-resolution crystal structures (1.6 Å or better) from the PDB (updated tables, unpublished).[21] A penalty is imposed for each atom–atom distance less than the minimal approach distance, so that:

$$S_{\text{bump}} = \sum_{ij} (d_{\alpha\beta}^2 - r_{ij}^2)^2 / d_{\alpha\beta}^2 \quad (5)$$

where the sum is only over residues i and j in different docking partners for which $r_{ij} < d_{\alpha\beta}$.

Alignment scores and other constraints

Biological information may be used to assess the probability of a structure being correct. For example, an alignment and conservation analysis[39,40] might indicate that a certain patch of residues is conserved and likely to be involved in binding. For the current study, we used an alignment only to guide the docking of antibody–antigen pairs. Antibody complementarity-determining regions (CDRs) are known to contain the residues that typically bind to an antigen. We created an alignment profile using a set of 55 antibody–antigen pairs and identified interface residues using the 6 Å centroid–centroid distance criterion. The profile was simplified into three categories: sites that never contacted the antigen (False); sites that contacted an antigen in one or two structures in the set (Neutral); and sites that

contacted the antigen in three or more complexes (True). The alignment score S_{align} is:

$$S_{\text{align}} = \sum_i \delta_i^{\text{int}} p_i \quad (6)$$

where the sum is over all residues i, δ_i^{int} is 1 if residue i is an interface residue using the 6 Å centroid criterion and 0 otherwise, and p_i is 1.5 for False residue sites, 0.5 for Neutral sites, and 0 for True sites. The profile used in this work has been reported.[26]

Similarly, constraint scores can incorporate other biological information. For example, if a certain residue is phosphorylated in the interaction, it can be constrained to the interface using a function that, beyond a minimum constraint distance, penalizes according to the distance between that residue and the other partner. Our algorithm used constraints during the CAPRI challenge;[26] however, no such constraint is used in the current study.

High-resolution scoring functions and refinement details

In the high-resolution search, all heavy atoms and polar hydrogen atoms are represented explicitly. A complete all-atom scoring function should capture van der Waals interactions, solvation, hydrogen bonding, and electrostatics in addition to local internal energies such as torsion angle strains. No perfect free energy model exists, so we have attempted to assemble one containing each of these contributions in a manner that can best discriminate docking decoys in a short amount of computation time per structure. While our representation is far from optimal (including some partial redundancy), it is designed to capture the major interactions as tested by the ability of the scoring function to discriminate between good and poor decoys.

Weight fitting

To determine which scoring functions to include, bound-backbone perturbation decoys were used to fit scoring weights. Weights for the scoring function were determined using the statistical calculation language R,[34] following a method modified from that used by Tsai *et al*.[22] Decoys created in small perturbation runs were designated as "good" if they were in the top 5% by rmsd and below the mean of the repulsive van der Waals score. A logistic regression was used to determine weights that could maximally separate the good decoys from the others. The regression included data from 68 targets including 38 of the 54 targets in the benchmark set (using a set of decoys distinct from that used in testing the algorithm performance). A dramatic improvement in the performance of the scoring functions resulted from the use of the repulsive van der Waals score to partition the good decoys for fitting.

The significance of each term in the scoring function is attested by its weight as well as the z-value of the assignment of the weight, which shows which terms aid in discriminating. Indeed, repulsive van der Waals interactions dominate the score (i.e. they can most easily be used to eliminate poor models), followed by attractive van der Waals, solvation, rotamer probability, and hydrogen bonding.

The full-atom score is a linear combination of an attractive van der Waals score (S_{atr}), a repulsive van der Waals score (S_{rep}), an implicit solvation score (S_{sol}), a surface area-based solvation term (S_{sasa}), a hydrogen bonding score (S_{hb}), a rotamer probability term (S_{dun}), a residue–residue pair probability term (S_{pair}), and simple electrostatic terms divided into short-range and long-range attractive and repulsive components ($S_{\text{elec}}^{\text{sr-rep}}, S_{\text{elec}}^{\text{sr-atr}}, S_{\text{elec}}^{\text{lr-rep}}, S_{\text{elec}}^{\text{lr-atr}}$):

$$\begin{aligned} S = {} & w_{\text{atr}} S_{\text{atr}} + w_{\text{rep}} S_{\text{rep}} + w_{\text{sol}} S_{\text{sol}} + w_{\text{sasa}} S_{\text{sasa}} + w_{\text{hb}} S_{\text{hb}} \\ & + w_{\text{dun}} S_{\text{dun}} + w_{\text{pair}} S_{\text{pair}} + w_{\text{elec}}^{\text{sr-rep}} S_{\text{elec}}^{\text{sr-rep}} + w_{\text{elec}}^{\text{sr-atr}} S_{\text{elec}}^{\text{sr-atr}} \\ & + w_{\text{elec}}^{\text{lr-rep}} S_{\text{elec}}^{\text{lr-rep}} + w_{\text{elec}}^{\text{lr-atr}} S_{\text{elec}}^{\text{lr-atr}} \end{aligned} \quad (7)$$

Values of the weights (w) used during packing, minimization, and discrimination stages of the algorithm are noted in Table 4.

Finally, although many of the test targets were used also for weight calibration, over-fitting errors are unlikely, since a small number of parameters were fit with a very large set of data. A tally of results from an independent subset of 16 targets in Table 1 supports this assertion: the success ratios are similar to those of the benchmark set as a whole. An 81% success rate on bound-perturbation studies in the independent set compares with a rate of 78% in the full benchmark set and, similarly in global searches, a 69% success rate on the prediction of 25% or more of native residue–residue contacts compares with a rate of 52% in the full benchmark set.

van der Waals interactions

The van der Waals interactions are split into attractive and repulsive components and represented with a modified Lennard-Jones 6–12 potential. To avoid the singularity as the atom–atom distance $r_{ij} \to 0$, the short-range repulsive score is extrapolated linearly below $r_{ij} = 0.6\sigma_{ij}$, where σ_{ij} is the sum of the atomic radii of atoms i

Table 4. Weights used in high-resolution scoring

Score	Weight (P)	Weight (M)	Weight (D)	z-Value
Repulsive van der Waals	0.80	0.338	0.08	73.0
Attractive van der Waals	0.80	0.338	0.338	45.0
Surface area solvation	–	–	0.344	28.5
Gaussian solvent-exclusion	0.80	0.279	0.279	27.2
Rotamer probability	0.79	0.069	0.069	19.6
Hydrogen bonding	2.1	0.441	0.441	
SC/SC + SC/BB				14.9
BB/BB				6.8
Residue pair probability	0.66	0.164	0.164	6.9
Simple electrostatics				
Short-range repulsive	–	0.025	0.025	3.2
Short-range attractive	–	0.025	0.025	8.3
Long-range repulsive	–	0.098	0.098	15.1
Long-range attractive	–	0.0020	0.0020	0.4

(P) indicates weights used during side-chain packing, (M) indicates weights used during rigid-body minimization, and (D) indicates weights used for discrimination. The z-value indicates the importance of each term in the selection of good decoys in a calibration set. SC and BB indicate side-chain and backbone, respectively.

and j. Thus:

$$s_{ij}^{\mathrm{atr}} = \varepsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - 2\frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right) \qquad \text{for } 0.89\sigma_{ij} < r_{ij} < 8\ \text{Å},$$

$$s_{ij}^{\mathrm{rep}} = \begin{cases} \varepsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - 2\frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right) & \text{for } 0.6\sigma_{ij} < r_{ij} < 0.89\sigma_{ij}, \\ \varepsilon_{ij}(A + (0.6\sigma_{ij} - r_{ij})B) & \text{for } r_{ij} < 0.6\sigma_{ij}, \end{cases}$$

$$\text{with } A = \frac{\sigma_{ij}^{12}}{(0.6\sigma_{ij})^{12}} - 2\frac{\sigma_{ij}^{6}}{(0.6\sigma_{ij})^{6}}$$

$$\text{and } B = -12\frac{\sigma_{ij}^{12}}{(0.6\sigma_{ij})^{13}} + 12\frac{\sigma_{ij}^{6}}{(0.6\sigma_{ij})^{7}} \qquad (8)$$

The energy well depths ε_{ij} and the atomic radii sums σ_{ij} are taken from the CHARMM19 set.[41] The repulsive potential rises more steeply than the linearizations used in previous work;[22,27] the stringency provides significantly better decoy discrimination. The total scores S_{atr} and S_{rep} (as well as the other pair-wise all-atom scores) are sums over all atom–atom pairs ij, including intermolecular pairs as well as intra-molecular pairs whose distances depend on more than one set of backbone ϕ and ψ angles. That is, atoms in adjacent residues are excluded to avoid redundancy with the backbone-dependent rotamer probability term, which certainly comprises contributions from local steric interactions.

Solvation

Solvation is scored using two different measures. First, an implicit solvation score is computed according to the pair-wise Gaussian solvent-exclusion model developed by Lazaridis & Karplus.[31] Since the sequence identities are constant, atom reference energies are omitted.

The second solvation score acts to minimize hydrophobic surface area. A surface area calculation to determine the solvent-accessible surface area is performed rapidly with Boolean logic and look-up tables,[42] and the surface is weighted by atomic solvation parameters.[33,43] Due to the expense of the calculation, this score is used only for the final decoy discrimination.

Orientation-dependent hydrogen bonding

The hydrogen bonding scores are represented with a function derived from high-resolution protein structures in the PDB and detailed by Kortemme & Baker (see the Supplementary Material).[44] In this work, however, hydrogen bonds are not weighted according to their degree of burial.

Internal residue energies

The effects of local internal and packing stresses are approximated by backbone-dependent rotamer probabilities computed by Dunbrack & Cohen.[28] Probabilities of non-canonical angles are calculated assuming a Gaussian distribution about the means using the tabulated variances.[28] The score is a sum of the negative logarithms of the individual probabilities P_i for each residue I:

$$S_{\mathrm{dun}} = \sum_i -\log P_i(\mathrm{rot}|\phi_i, \varphi_i) \qquad (9)$$

While this score is responsible for the low ranking of some native structures (1FIN, 1WQ1, etc.), it is essential for the discrimination of decoys.

Residue pair potentials and simple electrostatics

Electrostatic preferences are modeled by both residue pair potentials and a simple distance-dependent dielectric model. Residue pair potentials are derived from pair distribution functions with the residue–residue distance measured between relevant "action centers" of the side-chains. The formulation is given in detail by Kuhlman & Baker (see the Supplementary Material).[27]

Electrostatic contributions are added using a Coulomb model:

$$s_{ij}^{\mathrm{elec}} = 332 q_i q_j / \varepsilon \hat{r}_{ij} = 332 q_i q_j / \hat{r}_{ij}^2 \qquad (10)$$

where $\hat{r} = \min(r_{ij}, 3\ \text{Å})$ to avoid singularities as $r_{ij} \to 0$, and the dielectric is assumed to increase linearly with distance ($\varepsilon \sim \hat{r}_{ij}$). Charges are modeled only on the most polar side-chain atoms: a charge of $-1/2$ on each of the δ-oxygen atoms in Asp and the ε-oxygen atoms in Glu, a charge of $+1/2$ on the η-nitrogen atoms in Arg, and a full positive charge on the ζ-nitrogen atom in Lys. Interactions are separated into attractive and repulsive categories as well as short-range (<5 Å) and long-range categories. In this way, the weight-fitting routine can account for dielectric distance-dependence and discount the attractive interactions that may be captured by the hydrogen bonding function.

Side-chain packing

Side-chains are placed using a simulated annealing algorithm that searches through backbone-dependent rotamers, as described by Kuhlman & Baker.[27] In addition to the major χ_1 angles used as described, angle sets were expanded to include the major χ_1 angles plus and minus one standard deviation of those angles.[28] In addition, major rotamer options for χ_3 and χ_4 angles were included.

To save computation time, the side-chains of each protein monomer are added and optimized before docking ("pre-packing"). Then, during the refinement stage of the algorithm, only residues with side-chain centroid positions within 8 Å of a side-chain centroid of the other protein partner are included in the packing routine. Disulfides must be treated with care in packing algorithms. During the pre-packing of the monomer components, disulfides are allowed to form, and extra rotamer conformations are allowed for cysteine residues (W.R. Schief & D.B., unpublished results). During the docking run, disulfide bonds are considered part of the backbone and are frozen, while unpaired cysteine residues are not allowed to form disulfide bonds.

Assessment criteria

The root-mean-squared distance (rmsd) between two structures is calculated over the C^{α} atoms of the smaller docking partner in the fixed coordinate frame of the larger partner. In practice, the larger partner's backbone

position is fixed during the search routine, so that the calculation becomes:

$$\text{rmsd} = \sqrt{\sum_i |x_i - y_i|^2} \quad (11)$$

where x_i and y_i are the positions of the two structures' C^α atoms in each residue i, and the sum is over the residues in the smaller binding partner.

The contact fraction is determined by first identifying all residue–residue contacts, defined as those pairs for which at least one inter-residue heavy atom–heavy atom distance is less than 4 Å. The native residue–residue contact fraction is then calculated as the fraction of the contacts identified in the native structure that are present also in the predicted structure. The complementary criterion, the fraction of predicted structure contacts that are present in the native structure, produces values in the vicinity of the reported native contact fractions (data not shown).

Acknowledgements

The authors thank the many scientists who have participated in the development of the suite of computational tools used in the Baker laboratory for computations on the structure of proteins. In particular, Kira Misura & William Wedemeyer refined the Lennard-Jones model, Jerry Tsai implemented the surface area solvation calculation and assisted with the logistic regression of scoring weights, and William Schief developed methods to treat disulfide bonds. Discussions with Tanja Kortemme on the nature of protein–protein interfaces were helpful and stimulating. Keith Laidig of Formix, Inc. built and maintained reliable, state-of-the-art computing resources. J.J.G. thanks NIH/NHGRI for support through award K01-HG02316. O.S.-F. carries Damon-Runyon Fellowship DRG-1704-02 from the Damon Runyon Cancer Research Foundation. Additional funding was provided by NIH.

References

1. Smith, G. R. & Sternberg, M. J. (2002). Prediction of protein–protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **12**, 28–35.
2. Camacho, C. J. & Vajda, S. (2002). Protein–protein association kinetics and protein docking. *Curr. Opin. Struct. Biol.* **12**, 36–40.
3. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. (2002). Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins: Struct. Funct. Genet.* **47**, 409–443.
4. Shoichet, B. K. & Kuntz, I. D. (1991). Protein docking and complementarity. *J. Mol. Biol.* **221**, 327–346.
5. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A., Aflalo, C. & Vakser, I. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl Acad. Sci. USA*, **89**, 2195–2199.
6. Gabb, H. A., Jackson, R. M. & Sternberg, M. J. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **272**, 106–120.
7. Camacho, C. J., Gatchell, D. W., Kimura, S. R. & Vajda, S. (2000). Scoring docked conformations generated by rigid-body protein–protein docking. *Proteins: Struct. Funct. Genet.* **40**, 525–537.
8. Palma, P. N., Krippahl, L., Wampler, J. E. & Moura, J. J. (2000). BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins: Struct. Funct. Genet.* **39**, 372–384.
9. Mandell, J. G., Roberts, V. A., Pique, M. E., Kotlovyi, V., Mitchell, J. C., Nelson, E. *et al.* (2001). Protein docking using continuum electrostatics and geometric fit. *Protein Eng.* **14**, 105–113.
10. Norel, R., Sheinerman, F., Petrey, D. & Honig, B. (2001). Electrostatic contributions to protein–protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci.* **10**, 2147–2161.
11. Heifetz, A., Katchalski-Katzir, E. & Eisenstein, M. (2002). Electrostatics in protein–protein docking. *Protein Sci.* **11**, 571–587.
12. Vakser, I. A., Matar, O. G. & Lam, C. F. (1999). A systematic study of low-resolution recognition in protein–protein complexes. *Proc. Natl Acad. Sci. USA*, **96**, 8477–8482.
13. Jiang, F., Lin, W. & Rao, Z. (2002). SOFTDOCK: understanding of molecular recognition through a systematic docking study. *Protein Eng.* **15**, 257–263.
14. Chen, R. & Weng, Z. (2002). Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins: Struct. Funct. Genet.* **47**, 281–294.
15. Jackson, R. M., Gabb, H. A. & Sternberg, M. J. (1998). Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.* **276**, 265–285.
16. Lorber, D. M., Udo, M. K. & Shoichet, B. K. (2002). Protein–protein docking with multiple residue conformations and residue substitutions. *Protein Sci.* **11**, 1393–1408.
17. Fernandez-Recio, J., Totrov, M. & Abagyan, R. (2002). Soft protein–protein docking in internal coordinates. *Protein Sci.* **11**, 280–291.
18. Moult, J., Fidelis, K., Zemla, A. & Hubbard, T. (2001). Critical assessment of methods of protein structure prediction (CASP): round IV. *Proteins: Struct. Funct. Genet.* **45**, 2–7.
19. Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C. E. & Baker, D. (2001). Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins: Struct. Funct. Genet.* **45**, 119–126.
20. Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225.
21. Simons, K. T., Ruczinski, I., Kooperberg, C., Fox, B. A., Bystroff, C. & Baker, D. (1999). Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins: Struct. Funct. Genet.* **34**, 82–95.
22. Tsai, J., Bonneau, R., Morozov, A.V., Kuhlman, B., Rohl, C.A., & Baker, D. (2003). An improved protein decoy set for testing energy functions for protein structure prediction. *Proteins: Struct. Funct. Genet.*
23. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne,

P. E. (2000). The Protein Data Bank. *Nucl. Acids. Res.* **28**, 235–242.
24. Janin, J., Henrick, K., Moult, J., Ten Eyck, L., Sternberg, M. J. E., Vajda, S., Vakser, I. A. & Wodak, S. J. (2003). CAPRI: a critical assessment of predicted interactions. *Proteins: Struct. Funct. Genet.* **52**, 2–9.
25. Méndez, R., Laplae, R., de Maria, L. & Wodak, S. J. (2003). Assessment of blind predictions of protein–protein interactions: current status of docking methods. *Proteins: Struct. Funct. Genet.* **52**, 51–67.
26. Gray, J. J., Moughon, S. E., Kortemme, T., Schueler-Furman, O., Misura, K. M. S., Morozov, A. V. & Baker, D. (2003). Protein–protein docking predictions for the CAPRI experiment. *Proteins: Struct. Funct. Genet.* **52**, 118–122.
27. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA*, **97**, 10383–10388.
28. Dunbrack, R. L., Jr & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661–1681.
29. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd edit., Cambridge University Press, Cambridge.
30. Allen, M. P. & Tildesley, D. J. (1989). *Computer Simulation of Liquids*, Oxford University Press, New York.
31. Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins: Struct. Funct. Genet.* **35**, 133–152.
32. Kortemme, T., Morozov, A. V. & Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. *J. Mol. Biol.* **326**, 1239–1259.
33. Koehl, P. & Delarue, M. (1994). Polar and nonpolar atomic environments in the protein core: implications for folding and binding. *Proteins: Struct. Funct. Genet.* **20**, 264–278.
34. Ihaka, R. & Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314.
35. Chen, R., Mintseris, J., Janin, J. & Weng, Z. (2003). A protein–protein docking benchmark. *Proteins: Struct. Funct. Genet.* **52**, 88–91.
36. Camacho, C. J. & Vajda, S. (2001). Protein docking along smooth association pathways. *Proc. Natl Acad. Sci. USA*, **98**, 10636–10641.
37. Moont, G., Gabb, H. A. & Sternberg, M. J. (1999). Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins: Struct. Funct. Genet.* **35**, 364–373.
38. Glaser, F., Steinberg, D. M., Vakser, I. A. & Ben-Tal, N. (2001). Residue frequencies and pairing preferences at protein–protein interfaces. *Proteins: Struct. Funct. Genet.* **43**, 89–102.
39. Goh, C. S. & Cohen, F. E. (2002). Co-evolutionary analysis reveals insights into protein–protein interactions. *J. Mol. Biol.* **324**, 177–192.
40. Elcock, A. H. & McCammon, J. A. (2001). Identification of protein oligomerization states by analysis of interface conservation. *Proc. Natl Acad. Sci. USA*, **98**, 2990–2994.
41. Neria, E., Fischer, S. & Karplus, M. (1996). Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **105**, 1902–1921.
42. Legrand, S. M. & Merz, K. M. (1993). Rapid approximation to molecular-surface area *via* the use of boolean logic and look-up tables. *J. Comput. Chem.* **14**, 349–352.
43. Eisenberg, D. & McLachlan, A. D. (1986). Solvation energy in protein folding and binding. *Nature*, **319**, 199–203.
44. Kortemme, T. & Baker, D. (2002). A simple physical model for binding energy hot spots in protein–protein complexes. *Proc. Natl Acad. Sci. USA*, **99**, 14116–14121.
45. Ascenzi, P., Amiconi, G., Menegatti, E., Guarneri, M., Bolognesi, M. & Schnebli, H. (1988). Binding of the recombinant proteinase inhibitor eglin c from leech *Hirudo medicinalis* to human leukocyte elastase, bovine alpha-chymotrypsin and subtilisin Carlsberg: thermodynamic study. *J. Enzyme Inhib.* **2**, 167–172.
46. Stites, W. E. (1997). Protein–protein interactions: interface structure, binding thermodynamics, and mutational analysis. *Chem. Rev.* **97**, 1233–1250.
47. Eastman, J., Wilson, E. J., Cerveñansky, C. & Rosenberry, T. L. (1995). Fasciculin 2 binds to the peripheral site on acetylcholinesterase and inhibits substrate hydrolysis by slowing a step involving proton transfer during enzyme acylation. *J. Biol. Chem.* **270**, 19694–19701.
48. Menegatti, E., Guarneri, M., Bolognesi, M., Ascenzi, P. & Amiconi, G. (1987). Binding of the bovine pancreatic secretory trypsin inhibitor (Kazal) to bovine serine (pro)enzymes. *J. Mol. Biol.* **198**, 129–132.
49. Turk, B., Krizaj, I., Kralj, B., Dolenc, I., Popovic, T., Bieth, J. G. & Turk, V. (1993). Bovine stefin C, a new member of the stefin family. *J. Biol. Chem.* **268**, 7323–7329.
50. Ayala, Y. M., Vindigni, A., Nayal, M., Spolar, R. S., Record, M. T., Jr & Di Cera, E. (1995). Thermodynamic investigation of hirudin binding to the slow and fast forms of thrombin: evidence for folding transitions in the inhibitor and protease coupled to binding. *J. Mol. Biol.* **253**, 787–798.
51. Huang, M., Syed, R., Stura, E. A., Stone, M. J., Stefanko, R. S., Ruf, W. *et al.* (1998). The mechanism of an inhibitory antibody on TF-initiated blood coagulation revealed by the crystal structures of human tissue factor, Fab 5G9 and TF.G9 complex. *J. Mol. Biol.* **275**, 873–894.
52. Ward, E. S., Gussow, D., Griffiths, A. D., Jones, P. T. & Winter, G. (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature*, **341**, 544–546.
53. Lee, C. H., Leung, B., Lemmon, M. A., Zheng, J., Cowburn, D., Kuriyan, J. & Saksela, K. (1995). A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein. *EMBO J.* **14**, 5006–5015.
54. Davidson, V. L., Graichen, M. E. & Jones, L. H. (1993). Binding constants for a physiologic electron-transfer protein complex between methylamine dehydrogenase and amicyanin. Effects of ionic strength and bound copper on binding. *Biochim. Biophys. Acta*, **1144**, 39–45.
55. Mannherz, H. G., Goody, R. S., Konrad, M. & Nowak, E. (1980). The interaction of bovine pancreatic deoxyribonuclease I and skeletal muscle actin. *Eur. J. Biochem.* **104**, 367–379.
56. Pettigrew, D. W., Meadow, N. D., Roseman, S. & Remington, S. J. (1998). Cation-promoted association of *Escherichia coli* phosphocarrier protein IIAGlc with regulatory target protein glycerol kinase: substitutions of a Zinc(II) ligand and implications for inducer exclusion. *Biochemistry*, **37**, 4875–4883.

57. Akerstrom, B. & Bjorck, L. (1986). A physicochemical study of protein G, a molecule with unique immunoglobulin G-binding properties. *J. Biol. Chem.* **261**, 10240–10247.
58. Heitz, F., Morris, M. C., Fesquet, D., Cavadore, J. C., Doree, M. & Divita, G. (1997). Interactions of cyclins with cyclin-dependent kinases: a common interactive mechanism. *Biochemistry*, **36**, 4995–5003.
59. Romero, G., Chau, V. & Biltonen, R. L. (1985). Kinetics and thermodynamics of the interaction of elongation factor Tu with elongation factor Ts, guanine nucleotides, and aminoacyl-tRNA. *J. Biol. Chem.* **260**, 6167–6174.
60. Uchida, H., Banba, S., Wada, M., Matsumoto, K., Ikeda, M., Naito, N., Tanaka, E. & Honjo, M. (1999). Analysis of binding properties between 20 kDa human growth hormone (hGH) and hGH receptor (hGHR): the binding affinity for hGHR extracellular domain and mode of receptor dimerization. *J. Mol. Endocrinol.* **23**, 347–353.

Edited by F. E. Cohen

(Received 18 February 2003; received in revised form 6 May 2003; accepted 16 May 2003)

Note added in proof:
The complete protocol, as described in this paper, was used on target 8 in the third round of CAPRI (January 2003). With some human intervention to prepare the monomers and select models, the program correctly predicted 53% of the native residue-residue contacts in a model with 4.6 rmsd to the native structure. This model was one of the two best predictions for this target in the challenge.