Optimal Design of Protein Docking Potentials: Efficiency and Limitations

Dror Tobi and Ivet Bahar*
Department of Computational Biology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

ABSTRACT **Protein–protein docking is a challenging computational problem in functional genomics, particularly when one or both proteins undergo conformational change(s) upon binding. The major challenge is to define scoring function soft enough to tolerate these changes *and* specific enough to distinguish between near-native and "misdocked" conformations. Using a linear programming technique, we derived protein docking potentials (PDPs) that comply with this requirement. We considered a set of 63 nonredundant complexes to this aim, and generated 400,000 putative docked complexes (decoys) based on shape complementarity criterion for each complex. The PDPs were required to yield for the native (correctly docked) structure a potential energy lower than those of all the nonnative (misdocked) structures. The energy constraints applied to all complexes led to ca. 25 million inequalities, the simultaneous solution of which yielded an optimal set of PDPs that discriminated the correctly docked (up to 4.0 Å root-mean-square deviation from known complex structure) structure among the 85 top-ranking (0.02%) decoys in 59/63 examined bound–bound cases. The high performance of the potentials was further verified in jackknife tests and by ranking putative docked conformation submitted to CAPRI. In addition to their utility in identifying correctly folded complexes, the PDPs reveal biologically meaningful features that distinguish docking potentials from folding potentials. Proteins 2006;62:970–981.** © 2005 Wiley-Liss, Inc.

Key words: protein–protein interaction; binding geometry; contact energies; docking simulations; coarse-grained conformations; linear programming

INTRODUCTION

Interest in protein–protein interactions has rapidly grown in the past few years given that protein–protein interactions provide an underlying framework through which cellular activities are executed and controlled. Although methods such as X-ray crystallography, nuclear magnetic resonance spectroscopy, and cryo-electron microscopy provide valuable information on docking mechanisms and conformations, many complexes are too transient to lend themselves to experimental characterization. Computational methods have recently gained importance as possible means of efficiently providing information on protein–protein interactions.

Several algorithms have been developed to date for protein docking.[1–11] The usual approach is to exploit the geometric fit (or shape complementarity) of the complex-forming proteins, with or without considering their chemical affinities. Docking algorithms that reconstruct known complexes using the structures of the proteins in the bound form (*bound docking problem*) generally show reasonable levels of success. However, when the structures in the unbound form are used as input (*unbound docking problem*), the same algorithms may perform poorly. This is mainly attributed to the inability of the algorithms to take account of the conformational changes that (may) accompany complex formation.[12] Additional biological data are usually resorted to in order to single out the native or native-like conformation(s) among putative docked conformations.[13]

In the absence of prior data, a complete search of the space of rigid docking geometries at sufficiently fine intervals may necessitate scanning up to 10^9 different positions and orientations in search of the optimal binding,[1] which makes it difficult, if not impossible, to consider the additional degrees of freedom imparted by protein flexibility. Many groups instead use a two-stage strategy: a rigid-body docking followed by a refinement stage. The first stage screens all possible geometries and retains a reasonably small set that potentially includes the native (or native-like) conformation(s).[6] Because of computing time limitation, approximate but fast algorithms with tractable functions and parameters are adopted at this stage. The refinement stage reexamines the retained structures, to rank-order them using more sophisticated physicochemical criteria or scoring functions, and energy optimization schema.[8,14,15]

The criteria for evaluating the putative docked conformations are crucial to the success of the algorithms. Various scoring strategies have been adopted in previous work. The docking algorithms DOT,[4] ClusPro,[2] and ZDOCK[6] use

Grant sponsor: National Institutes of Health; Grant numbers: LM007994-01A1; Grant sponsor: National Science Foundation; Grant number: EIA-0225636.

*Correspondence to: Department of Computational Biology, School of Medicine, University of Pittsburgh, 3064 BST 3, 3501 Fifth Avenue, Pittsburgh, PA 15213. E-mail: bahar@ccbb.pitt.edu

Received 1 June 2005; Revised 27 September 2005; Accepted 19 October 2005

Published online 29 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20859

atom-based potentials whereas BiGGER,[1] FTDock,[16] and RosettaDock[11]combine atomic potentials with coarse-grained residue-level potentials. Atom-based potentials may, however, be too sensitive to the precise position of the interacting atoms, and may not be flexible enough to tolerate rearrangements induced upon binding, hence their limited detection (or low coverage) of native-like geometries. So far, coarse-grained residue level potentials have not shown good ability to distinguish between near-native and nonnative complex geometries. There is a need for constructing relatively soft potentials, with coarser-grained representations, while maintaining sufficient specificity.

Designing coarse-grained optimal protein docking potentials (PDPs) is a challenging task for the following reasons. The learning set of X-ray resolved *transient* protein complex structures available in the Protein Data Bank (PDB)[17] is relatively small,[12] and on average only 22 amino acids from each protein participate in the interfacial interactions.[18] Thus, not only limited structures are available, but each structure also provides limited data to derive statistically reliable information. Additionally, some types of complexes are overrepresented in the PDB. A typical example is the set of complexes containing proteases. Knowledge-based potentials for docking usually perform well, as a consequence, on protease complexes, but to a lesser degree on other complexes.[16] The limited (incomplete) and biased structural data on protein–protein complexes poses serious problems on the applicability/accuracy of the inverse Boltzmann principle to derive effective PDPs. In view of this problem, PDPs have been derived from datasets that contain multimeric protein interfaces[11,19,20] and even from the examination of protein interiors.[1,16] However, *transient* complex interfaces differ from those of multimers and protein interiors as pointed out by Lo Conte et al.[18]

In the present study, we present an effective methodology based on a linear programming (LP) technique, which utilizes as a learning dataset *transient* protein complexes and associated misdocked decoys. The approach enables us to include in our training set false positives (FPs) in addition to the true positives (TPs), thus alleviating the statistical problem of limited data. Additionally, the present PDPs take account of side-chain–side-chain, side-chain–backbone, and backbone–backbone interaction, thus providing a more realistic description of interresidue interactions. The utility of the new set of PDPs is illustrated by performing a jackknife test on the learning set, and by ranking independent test set of putative docked conformation submitted to CAPRI.

METHODS

LP Approach for Constructing PDPs

We consider the set of nonredundant complexes compiled by Lo Conte et al.[18] (Table I, first column), apart from the largest complexes (one of monomers has 400 or more residues) and those containing large hetero-groups at the interface (e.g., FKBP12 immunophilin-calcineurin). This set contains 63 representative members for various types of transient complexes such as enzyme-inhibitor and antigen-antibody. For each complex, we generated 400,000 putative docked conformations, a few of which conform to the native (or native-like) conformation(s), and the remainder are "*misdocked*." In the derivation of the PDPs, the basic requirement is to obtain for each native complex a free energy lower than that of *any* misdocked complex for the same pair of proteins. Each misdocked complex (decoy) thus defines a constraint (in the form of an inequality) and the objective is to determine a set of interresidue potentials that satisfy these constraints.

The advantages of LP approach over statistical approaches are twofold: (a) statistical approaches learn from known native states on other native states, i.e., the (inverse) Boltzmann statistics is applied to a set of known protein complexes. In the LP approach, however, we learn from a set of native states as well as large sets of nonnative states (FPs) about how a native state should, and should not, look. Therefore, we have more varied data to optimize the empirical potentials. (b) The LP approach is not sensitive to over- or underrepresentation or to sequence/structure homologies in the training set.

LP techniques have been previously used in deriving protein *folding* potentials[21–25] after the pioneering work of Maiorov and Crippen.[26] The terminology adopted here follows that of our previous studies.[21,22]

The PDP construction problem is set as follows. Let X be the coordinate vector that represents a protein conformation, $(\hat{X}_a, \hat{X}_b)$ the coordinates of the native complex (composed of monomers a and b), and $\{\hat{X}_a, X_b^i\}_{i=1}^N$ the set of N decoys generated by alternative rotations and translations of monomer b. The potential energy $U(\hat{X}_a, \hat{X}_b)$ of the native complex is required to be lower than that of any misdocked conformation, as expressed by the set of N inequalities

$$\{U(\hat{X}_a, X_b^i) - U(\hat{X}_a, \hat{X}_b) > \epsilon_i\}_{i=1}^N \quad (1)$$

where ϵ_i is a nonnegative constant, set to 0.1 in our calculations.

Protein Models

An intermediate level of complexity is adopted here for modeling the structures, in which each amino acid is represented by *three* interaction sites: the side-chain centroid (S), the amide nitrogen (N), and the carbonyl oxygen (O) on the backbone (B). This representation presents the advantage of accounting for the hydrogen-bond formation propensities of backbone atoms, as well as specificity of side-chains. Six types of interactions are involved between these three types of atoms: S-S, S-O, S-N, O-N, N-N, and O-O, the former three being residue-specific. This results in a set of 210 (S-S) + 20 (S-0 + S-N) + 1 (O-N + N-N + O-O) = 253 parameters.

Parameter Optimization

The total energy is expressed as a linear combination of a parameter set $\{p_l\}_{l=1}^L$

$$U(X_a{:}X_b, P) = \sum_{l=1}^{L} S_l(X_a{:}X_b)p_l \quad (2)$$

TABLE I. Ranking of the Best Native-Like Conformation Among the 400,000 Decoys Generated for Each Complex (Bound Docking Results)

	Results obtained from different ranking criteria						No. of decoys with RMSD ≤ 4.0 Å
	Shape		PDP		MJ		
Complex	Rank[a]	RMSD[b]	Rank[a]	RMSD[b]	Rank[a]	RMSD[b]	
Protease-inhibitor							
2PTC	2	2.23	1	3.39	1	2.23	108
1AVW*	1	1.75	1	2.62	3	2.76	29
1MCT	1	2.16	1	1.09	1	1.56	344
3TPI	3	1.31	1	2.94	24	1.36	16
1TGS	1	1.06	1	1.28	1	1.06	104
1CHO	1	2.77	1	1.85	1	2.79	528
1ACB	10	2.83	1	2.83	8	2.83	47
1CBW	6	1.19	1	3.90	5	3.56	259
1PPF	45	2.59	1	3.44	1	2.42	85
1FLE	1	1.14	1	1.54	1	1.54	65
2KAI	51	1.52	1	2.17	4	3.99	69
1HIA	1	2.72	4	3.60	1	2.61	57
3SGB	7	1.17	2	1.54	3	1.26	1,657
1MKW	7,531	2.91	36	3.11	166,269	2.91	5
1CSE*	1	1.73	1	1.81	1	3.93	39
2SIC	287	3.22	1	3.22	7	3.22	13
2SNI	1	1.77	1	2.70	1	1.33	379
1STF	1	1.64	1	2.87	1	2.71	62
4CPA	9	1.27	1	3.00	2	3.41	1,273
Large protease complexes							
1BTH	1	2.05	7	2.05	1	2.05	18
4HTC	1	0.75	1	1.09	1	0.75	17
1TBQ*	1	1.41	1	1.41	1	1.41	5
1TOC	NA		NA		NA		0
Antigen-antibody							
1VFB	21	1.54	2	1.86	430	1.86	103
1MLC	1,493	2.58	47	3.52	248	2.58	11
1JHL	580	1.05	3	2.65	14,793	2.65	53
3HFL	27	1.90	3	2.58	136	2.58	100
3HFM	1,062	3.25	1	3.25	282	2.92	36
1FBI	32	2.37	2	2.32	93	2.32	62
1MEL	1	1.05	1	1.05	1	1.05	56
2JEL	59	2.81	1	2.81	2	2.81	132
1NSN	772	3.72	15	3.37	22	3.72	26
1OSP*	1,694	3.87	1	3.87	4,288	3.87	14
1NCA	1	3.60	2	3.60	5	3.60	57
1NMB	42,843	3.28	13	3.28	2,396	3.28	5
1DVF	174	2.79	78	2.85	1,113	2.79	26
1IAI	13,949	2.78	35	2.33	7,599	2.33	4
1NFD	3,176	3.92	58	1.79	46,167	3.79	23
1KB5	82	3.50	5,121	3.91	1,971	3.01	7
Enzyme-complexes							
1BRS*	2	1.86	1	2.36	1	2.36	301
1DFJ	1	2.02	1	3.06	1	2.02	18
1DHK	1	3.49	1	3.49	1	3.11	15
1FSS	12	1.64	168	3.30	241	1.64	46
1GLA	437	2.19	85	2.86	93	2.40	32
1UDI	1	1.80	1	2.42	1	1.38	602
1YDR*	1	0.74	1	2.74	1	0.90	68
2PCC	377	3.60	1	3.60	3,210	3.02	36
G-proteins, cell cycle, signal transduction							
1TX4*	34	2.68	26	2.19	1	2.68	21
1GUA	50	2.57	3	2.24	1,435	3.53	91
1GG2	8	3.62	2	3.62	1	3.48	27
1GOT	1	3.82	2	3.82	1	3.82	4
2TRC	1	1.50	2	1.50	1	1.53	201
1AGR	169	2.23	9	2.23	11,263	2.82	38
1FIN	1	1.32	1	3.65	1	1.64	17
1A0O	550	1.20	7	1.20	134	3.62	61

TABLE I. (Continued)

Complex	Results obtained from different ranking criteria						No. of decoys with RMSD ≤ 4.0 Å
	Shape		PDP		MJ		
	Rank[a]	RMSD[b]	Rank[a]	RMSD[b]	Rank[a]	RMSD[b]	
Miscellaneous							
1FC2	21	3.22	5	2.95	11	3.22	87
1IGC	245	1.53	1	2.12	1,092	1.02	677
1AK4	15,120	3.42	2	2.50	426	3.82	17
1EFN	8	1.85	1	2.55	72	2.55	160
1ATN*	309	3.85	67	3.31	8	3.85	18
1EBP	3	1.63	5	1.21	2	1.63	181
1HWG	NA		NA		NA		0
1YCS	779	2.16	1	2.16	9,366	2.16	28

[a]Among 400,000 putative docked conformations.
[b]Between the C^{α} coordinates of the computed and experimentally observed docked substrate.
*Designates the structures that have been subjected to jackknife tests (see text).

where $\{S(X_a{:}X_b)\}_{l=1}^{L}$ designates the set of $L = 253$ basis functions. The parameters $\{p_l\}_{l=1}^{L}$ are linear, therefore, can be optimized using LP algorithms such as Simplex or Interior-Point. Consistent with many existing knowledge-based potentials,[16,19,21,27,28] we express $S_l(X_a{:}X_b)$ as a step function of the form

$$S_l(X_a{:}X_b) = \sum_{n\epsilon a, m\epsilon b} H'(|r_n - r_m|)\delta(l,nm) \qquad (3)$$

where

$$H'(|r_n - r_m|) = \begin{cases} 1 & if\ |r_n - r_m| < r^l_{cutoff} \\ 0 & otherwise \end{cases} \qquad (4)$$

is the Heaviside function that extracts all interacting pairs m and n (in the respective monomers a and b) separated by a distance $|r_n - r_m|$ shorter than a cutoff distance r^l_{cutoff} characteristic of the type (l) of interaction between the two sites; $\delta(l,nm)$ is the Kronecker delta equal to 1 if the type of interaction between sites m and n is of type l ($1 \le l \le 253$), zero otherwise. We adopted r^l_{cutoff} values of 4.0 Å for B-B, 5.6 Å for B-S, and 6.8 Å for S-S interactions. The total energy for a given decoy i reduces to the summation over the effective PDPs p_l, weighted by the numbers n^i_l of contacts of type l occurring in the examined decoy. The N inequalities for each decoy i ($1 \le i \le N$) in Equation 1 become

$$\sum_{l=1}^{253} p_l(n^i_l - \hat{n}_l) > \epsilon_i \qquad (5)$$

where $\hat{n}_l$ is the number of contacts of type l in the native complex.

The adoption of a simplified model for protein structure and energetics increases the tolerance to small conformational changes. It becomes, however, harder to distinguish between native and native-like complexes. We use a threshold root-mean-square deviation (RMSD) of 4.0 Å for defining native-like conformations (TPs). Thus, the decoys having RMSD > 4.0 Å form the set of FPs used for training the PDPs.

Generation of a Population of Putative Docked Conformations

A set of 400,000 decoys was generated for each complex from the bound subunits, using the docking algorithm and surface matching criteria proposed by Palma et al.[1] Accordingly, geometric fit was assessed from the number of overlapping *surface nodes* between the receptor (large protein) and the substrate (small protein). On average, 399,360 nonnative (FP) and 640 native-like structures (TPs) were found for a given complex, which mapped to a total of $N = 25{,}159{,}988$ inequalities (Eq. 5).

Numerical Solution of Inequalities

Each inequality divides the parameter space into two regions, one accessible (any point within this space is a valid solution), and the other excluded. A given inequality may give rise to three outcomes: it may (1) reduce the space allowed for the parameter set (most desirable), (2) have no effect on the allowed space, or (3) impose an impossible condition (reducing the allowed parameter space to zero). A small subset of inequalities fell in the third group. The problem was then to obtain an optimal set of parameters that would satisfy *most* of the inequalities. In some cases, especially with large complexes, the RMSD criterion was observed to be too severe to capture native-like conformations. A slight tilt of the substrate at the binding site would, for example, lead to a large displacement at the opposite (far) end of the substrate, and yield a relatively high RMSD although most of the native interfacial contacts were accurately captured. Penalizing such cases would severely restrict the space accessible to the PDPs. Such geometries are characterized by a relatively small number of non-zero ($n^i_l - \hat{n}_l$) terms (Eq. 5). We filtered out the cases having <37 non-zero terms, which led to a set of 25,069,820 inequalities. These were solved with the interior point program BPMPD[29] on an Opteron 246 processor with 8 GB of memory. Memory is the prime limiting factor in the computations, determining the maximal number of inequalities that can be solved simultaneously. Solving a large number of inequalities ($N > 25$ millions in the present case) requires an iterative proce-

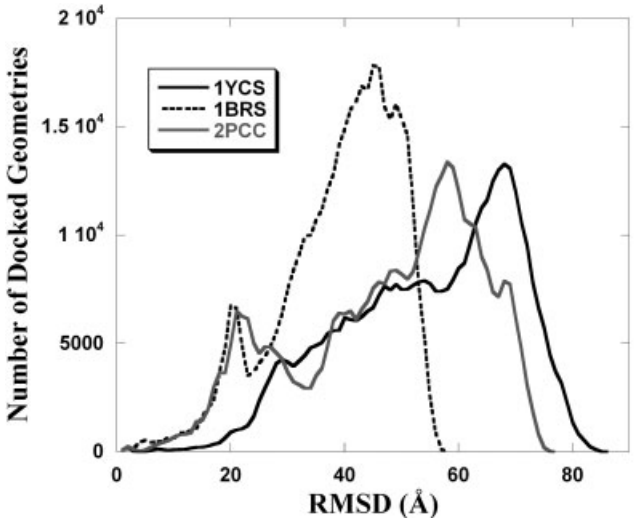

Fig. 1. Distribution of the RMSD from native (correctly docked) complex for the 400,000 putative docked structures, illustrated for three complexes: 1YCS, 1BRS, and 2PCC. The RMSD is measured between the C$^{\alpha}$ atoms of the substrate in the putative docked conformation and in the native complex, after optimal superposition of the receptor proteins of the two structures. Decoys showing an RMSD of up to 4.0 Å form the set of native-like conformations (TPs). They amount to $<$0.16% of all generated decoys, on average (over all the 63 cases). Those above the threshold 4.0 Å provide the constraints implemented in the LP algorithm.

dure, starting with a small set (about 200) of representative inequalities from each complex, and repeating the calculations with larger subsets that include the inequalities not satisfied at the preceding iteration. Usually three iterations were sufficient to obtain the PDPs that satisfy 99.999% of the inequalities.

RESULTS

PDPs

The decoys generated by using the bound forms of the complex-forming proteins show a diversity of conformations illustrated in Figure 1 for three examples. The figure displays the distribution of the RMSDs from the native structures for the 400,000 putative complexes, RMSDs referring to the differences in substrate backbone coordinates after optimal superposition of the receptors. The generated dock structure is accepted to be native-like (TP) if the corresponding RMSD is below 4Å. All decoys having RMSD above this threshold were used as FPs in our LP algorithms.

The resulting PDPs are presented in Table II. It is of interest to compare the PDPs with knowledge-based potentials commonly used for exploring protein folding or stability characteristics.[16,19,21,27,28] Figure 2(A) displays the map of the docking potentials presently obtained; the maps of two folding potentials, the Miyazawa-Jernigan potentials[27] and the TE potentials,[21] are displayed at Figure 2(B) and (C), respectively. Whereas folding and docking potentials exhibit similarities, particular pairs exhibit significant differences, which will be further discussed below.

TABLE II. PDPs

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL	N(H)	(C)O
ALA	−3.56	4.17	1.69	−1.45	−3.00	−2.52	3.56	0.54	−0.07	−0.65	−0.58	4.26	5.39	−0.17	4.50	1.23	−0.21	−3.36	−0.60	−2.05	0.60	−0.44
ARG	4.17	−1.66	−0.12	−2.00	−1.02	4.20	−1.80	−0.04	0.29	2.80	−1.80	0.90	−1.10	−3.04	1.33	−0.02	−0.55	−4.65	−3.18	−2.69	1.60	−1.37
ASN	1.69	−0.12	−0.33	−1.09	6.52	−2.25	−2.20	1.13	0.95	−0.86	2.20	1.22	−1.33	−1.83	−0.42	−0.09	−0.99	−2.81	−1.58	0.66	0.17	−0.71
ASP	−1.45	−2.00	−1.09	0.97	2.57	0.63	1.45	0.67	−0.71	0.01	−0.62	−2.45	1.21	6.05	1.11	−2.42	−0.92	−0.90	−1.03	−1.61	0.85	0.07
CYS	−3.00	−1.02	6.52	2.57	10.00	−0.63	−3.34	0.94	−4.93	−1.13	−2.02	−2.95	0.00	6.01	4.87	−1.74	5.30	2.60	−0.36	1.37	0.44	−0.72
GLN	−2.52	4.20	−2.25	0.63	−0.63	9.72	1.25	−0.76	1.65	−1.83	−2.63	1.23	1.94	8.05	0.80	−0.75	1.01	−3.60	−2.36	−0.68	1.35	−1.28
GLU	3.56	−1.80	−2.20	1.45	−3.34	1.25	1.49	0.89	−0.81	3.26	0.05	−3.01	−0.91	0.36	2.76	−1.26	−0.63	−0.53	−1.74	1.04	−0.38	0.05
GLY	0.54	−0.04	1.13	0.67	0.94	−0.76	0.89	−0.34	0.44	−1.86	0.35	−0.26	−0.90	−0.16	2.87	1.20	0.43	0.10	−1.63	−1.06	−1.28	0.49
HIS	−0.07	0.29	0.95	−0.71	−4.93	1.65	−0.81	0.44	1.04	−1.82	2.20	0.42	−4.03	−2.13	1.65	1.24	0.35	1.68	−4.25	−0.20	−0.65	−1.15
ILE	−0.65	2.80	−0.86	0.01	−1.13	−1.83	3.26	−1.86	−1.82	−3.27	−4.09	1.34	0.50	−6.05	0.25	−1.24	1.66	−5.01	−2.88	−3.78	2.55	−0.09
LEU	−0.58	−1.80	2.20	−0.62	−2.02	−2.63	0.05	0.35	2.20	−4.09	−3.43	0.80	−1.90	−2.21	−1.40	0.29	0.72	−0.39	−1.43	−1.80	−0.70	−0.09
LYS	4.26	0.90	1.22	−2.45	−2.95	1.23	−3.01	−0.26	0.42	1.34	0.80	3.24	3.64	−2.39	3.85	1.71	0.65	−3.37	−2.34	−2.01	1.02	−0.57
MET	5.39	−1.10	−1.33	1.21	0.00	1.94	−0.91	−0.90	−4.03	0.50	−1.90	3.64	10.00	−4.17	−4.02	−0.55	−2.81	−5.89	−2.28	−1.32	0.74	−1.22
PHE	−0.17	−3.04	1.83	6.05	6.01	8.05	0.36	−0.16	−2.13	−6.05	−2.21	−2.39	−4.17	−6.42	−3.96	−0.29	−1.08	−2.65	−3.51	−2.17	0.49	−0.79
PRO	4.50	1.33	−0.42	1.11	4.87	0.80	2.76	2.87	1.65	0.25	−1.40	3.85	−4.02	−3.96	−2.50	−0.29	−1.18	−4.56	−2.23	0.22	0.01	−0.69
SER	1.23	−0.02	−0.09	−2.42	−1.74	−0.75	−1.26	1.20	1.24	−1.24	0.29	1.71	−0.55	0.17	−0.29	0.45	−1.01	−0.86	−1.38	1.27	0.65	−0.94
THR	−0.21	−0.55	−0.99	−0.92	5.30	1.01	−0.63	0.43	0.35	1.66	0.72	0.65	−2.81	−1.08	−1.18	−1.01	0.22	2.43	−1.76	0.83	0.76	−1.22
TRP	−3.36	−4.65	−2.81	−0.90	2.60	−3.60	−0.53	0.10	1.68	−5.01	−0.39	−3.37	−5.89	−2.65	−4.56	−0.86	2.43	8.64	−2.83	−1.87	−0.28	−0.70
TYR	−0.60	−3.18	−1.58	−1.03	−0.36	−2.36	−1.74	−1.63	−4.25	−2.88	−1.43	−2.34	−2.28	−3.51	−2.23	−1.38	−1.76	−2.83	−2.15	−1.37	1.38	−1.50
VAL	−2.05	−2.69	0.66	−1.61	1.37	−0.68	1.04	−1.06	−0.20	−3.78	−1.80	−2.01	−1.32	−2.17	0.22	1.27	0.83	−1.87	−1.37	−0.37	−0.22	−0.27
N(H)	0.60	1.60	0.17	0.85	0.44	1.35	−0.38	−1.28	−0.65	2.55	−0.70	1.02	0.74	0.49	0.01	0.65	0.76	−0.28	1.38	−0.22	6.84	−2.14
(C)O	−0.44	−1.37	−0.71	0.07	−0.72	−1.28	0.05	0.49	−1.15	−0.09	−0.09	−0.57	−1.22	−0.79	−0.69	−0.94	−1.22	−0.70	−1.50	−0.27	−2.14	1.82

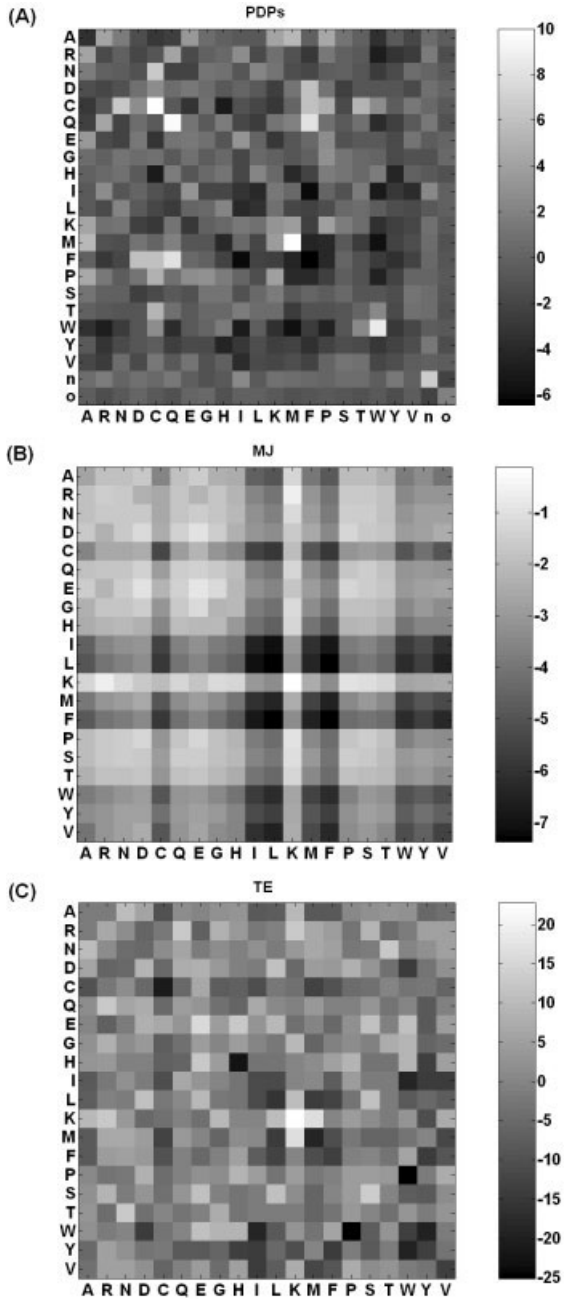

Fig. 2. Interresidue contact energies for docking and folding. **A:** The map corresponding to the presently obtained PDPs. **B and C:** The maps for the MJ[27] and the TE[21] potentials. The last two potentials were obtained from interresidue interactions in protein interiors rather than those at the interface of transient complexes. Animo acids are indicated using (upper-case) single-letter code; the amide nitrogen and the carbonyl oxygen are indicated with the respective symbols n and o in panel A.

Discriminating Ability of Docking Potentials

To evaluate the ability of the PDPs to distinguish between native-like and nonnative structures, we rank-ordered the 400,000 docked structures for each complex according to their PDPs *alone*. Results are presented in Table I, columns 4 and 5. In 32 of 63 complexes, a native-like geometry (RMSD $\leq$ 4 Å) was ranked in the first place (lowest energy) and in 46 of 63 cases it was ranked within the first five places, and finally among the top 100 places in 59 of 63 complexes. Thus, the PDPs exhibit a good ability to detect the correctly docked conformations. For two complexes, (1TOC) and (1HWG), no native-like structure could be sampled among the 400,000 putative docked structures.

The last column in Table I lists the number of native-like docked conformations generated for each complex. We note that the ranking success is correlated with the ability of the structure-generation methodology to sample native-like conformations. A rank $<$10 is obtained when the decoys set contains $>$50 native-like structures.

TABLE III. Performance of Different Criteria for Identifying Correctly Docked Proteins

Methodology for rank-ordering the decoys[a]	Number of complexes whose native structure is correctly identified in the top-ranking R decoys[b]		
	$R = 1$	$R = 5$	$R = 100$
Shape complementarity	21	25	42
MJ potentials	24	32	41
KB potentials	22	29	40
Present PDPs	32	46	59
TE potentials for folding	4	7	16

[a]400,000 decoys generated for each complex.
[b]From a total set of 63 complexes, using bound structures of the proteins in docking simulations.

Comparison With the Performance of Other Knowledge-Based Potentials

We tested the discriminating ability of three other sets of knowledge-based potentials, the Miyazawa-Jernigan (MJ)[27] potentials, the Keskin et al. (KB)[20] potentials, and the Tobi et al. (TE)[21] folding potentials.

The MJ solvent-mediated potentials were derived by a quasi-chemical approximation from a set of soluble protein subunits and represent S-S *intramolecular* interactions. The KB potentials are also solvent-mediated, but derived from the interfacial regions of a set of protein complexes and multimeric proteins, and thereby represent S-S *intermolecular* interactions. The two sets are highly correlated (correlation coefficient: 0.89), and their use in rank-ordering the decoys also yield similar results, the MJ potentials slightly outperforming the KB. The results for the MJ potentials are presented in Table I, columns 6 and 7. A native-like orientation was ranked in the first position in 24/63 cases, and in the top five position in 32/63 cases. With the KB potentials (not shown) these figures decreased to 22/63 and 29/63, respectively. The relatively high performance of the MJ and KB potentials may be attributed to their incorporation of solvent mediation.

The TE potentials were constructed for discriminating correctly *folded* structures, using the same LP technique, requiring the energy of a sequence in its native fold to be lower than that of any misfolded decoy. Decoys were generated by threading 572 sequences onto the corresponding set of structures.[21] Upon using these (TE) potentials for discriminating the corrected docked decoys, only in

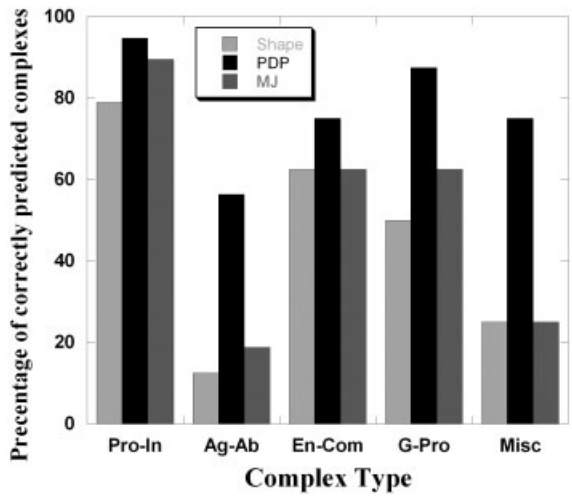

Fig. 3. Success rate of the PDPs in different families. Results are presented for protease-inhibitor complexes (Pro-In), antigen-antibody (Ag-Ab) complexes, enzyme complexes (En-Com), G-proteins, cell cycle, signal transduction (G-Pro), and miscellaneous (Misc). The success rate is based on the fraction of complexes whose native (or native-like) structure is detected among the top-ranking 10 putative docked conformations, rank-ordered according to the indicated three criteria (shape complementarity, present PDPs, and MJ potentials).

TABLE IV. Jackknife Test

Complex name	Rank[a]	RMSD[b]
1ACB	6	2.83
1BTH	13	2.05
1EBP	3	1.21
1EFN	10	8.49
1FC2	—[c]	—
1FIN	1	3.65
1FLE	1	1.54
1GUA	14	9.30
1MCT	7	1.09
1MEL	1	1.05
1TGS	1	1.28
1UDI	1	2.42
2JEL	2	2.81
2KAI	63	1.52
2PTC	1	3.39
2SNI	1	2.70
2TRC	2	2.29
3HFL	1	2.58
3SGB	3	3.38
3TPI	1	2.94

[a]Among 400,000 putative docked conformations.
[b]Between the C^{α} coordinates of the computed and experimentally observed docked substrate.
[c]No near-native complex (RMSD $\leq$ 10) was found among the top 2,000 ranked decoys.

seven cases was a native-like conformation ranked within the top five positions, and in 16 cases in the top 100. The performances of the different potentials are summarized in Table III, supporting the utility of presently derived PDPs.

How Effective Is Shape Complementarity in Discriminating Correct Docking Geometries?

To assess the utility of shape complementarity, decoys were rank-ordered based on the size of their matched surface (i.e., number of shared surface nodes). The results are presented in the columns 2 and 3 of Table III. A native-like structure was found to rank in the first position in 21/63 cases, and in the top five positions in 25/63 cases. Notably, the native structure of about 1/3 of the complexes can be effectively discriminated by shape complementarity alone. This observation supports the important role of shape, or contact topology, in defining stable structures. Yet, the consideration of residue-specific PDPs increases the number of correctly identified docked structures by 50% (Table III).

Dependence of the PDPs' Performance on the Types of Complexes

Figure 3 shows the success rate of the PDPs as a function of the families of the examined complexes. Results are presented for five types of complexes: protease-inhibitors, antigen-antibody complexes, enzyme-complexes, G-proteins cell cycle signal transduction complexes, and "miscellaneous." The bars display the percent of members in each group, for which a native-like geometry ranked in the top 10 decoys. Three rank-ordering criteria—shape complementarity, PDPs, and MJ potentials—are compared.

The PDPs successfully discriminate the native structure of >75% of each type of complex within the top-ranking 10 positions, except for the antigen-antibody complexes. Lawrence and Colman[30] suggested that the enzyme-inhibitor and antigen-antibody complexes represent two different classes of binding and that the latter as a whole exhibits poorer shape complementarity at the interface, which is also in accord with the reduced (to <60%) ability of the PDP to assign a high rank to the native-like conformations in this group. The same group shows, however, the largest enhancement (by a factor of 3–4) in the fraction of correctly discriminated docked structures, compared with those detected by MJ potentials or by shape complementarity.

Validation of the PDPs by Jackknife Tests

The performance of the PDPs was further examined by jackknife tests applied to eight complexes selected from different families (indicated by asterisks in Table I). Six of these complexes were accurately predicted by the PDPs as the first ranking decoy, and two (1TX4 and 1 ATN) were ranked in the 26th and 67th positions. New sets of potentials were obtained for each case, by solving the reduced set of inequalities in which the constraints associated with the tested proteins were *not* included. The question was to find out whether the level of success originally obtained with these complexes could be reproduced with the new PDPs. In six of eight cases, the native-like structure was ranked in the first place as before; in one case (ITX4), a native-like orientation was ranked at the 98th place. The PDPs did not perform well in 1ATN, where the rank of the first native-like conformation was 1,265th. A similar test was applied for a set of 20

TABLE V. Evaluation of CAPRI Structures Using Present PDPs

	Properties of CAPRI targets			PDP-based rank of near-native complex[c]	MJ-based rank of near-native complex[c]
		[a]	[b]		
1	HPr kinase/HPr	102	8	4	5
2	Rotavirus VP6/Fab	88	4	7	5
3	Hemagglutinin/Fab HC63	90	4	48	18
4	α-Amylase/camelide VH_1	66	1	2	40
5	α-Amylase/camelide VH_2	65	1	3	27
6	α-Amylase/camelide VH_3	65	9	4	4
7	T cell receptor/exotoxin A	70	19	1	8
8	Nidogen-G3/laminin EGF	179	12	3	6
9	LicT homodimer	162	32	1	1

[a]Number of predicted complexes submitted for each CAPRI target.
[b]Number of correctly predicted (hit) structures among those submitted.
[c]Near-native is defined by RMSD ≤ 10 Å, in accord with Comeau et al.[2]

proteins by excluding four proteins each time, solving the reduced set of inequalities and ranking the test proteins using the obtained potentials; results are presented in Table IV. In 18/20 cases, a native-like structure (RMSD ≤ 10) was ranked among the top 15 ranked orientations, in one case it was ranked in the 63rd place, and in one case no native-like complex was found among the top 2,000 ranked orientations. The results confirmed the validity of the potentials. The computing times required to score 400,000 putative docked conformations for 1ACB and 1UDI (Table IV) were 40.3 and 49.25 min, respectively, using a Pentium IV 2.8-GHz processor, which shows the computational efficiency of the PDPs.

Rescoring CAPRI Predictions: Extension to the Unbound Protein Docking Problem

We used our PDPs and the MJ potentials to score the structures submitted for targets 1–9 in the CAPRI assessment.[31] Using energetic criteria, Comeau et al.[2] (see Table II therein) were able to rank correctly a near-native structures (RMSD ≤ 10 Å) for four of nine targets, and only two of these near-native structures were ranked among the top 10 clusters. The predicted complexes submitted by the various participating groups are well-refined low-energy decoys and therefore present a challenging task to the PDPs. In the case of targets 4 and 5, no near-native (RMSD ≤ 10 Å) structure was found among the submitted predictions, and therefore a near-native structure was included by superimposing the unbound subunits onto the coordinates of the complex structure.

The results are presented in Table V. Using the PDPs, in seven of the nine targets a near-native complex is detected among the top-ranking five low-energy structures and in eight of the nine targets a near-native structure is ranked in the top 10 low-energy structures. Using the MJ potentials, however, four of the nine targets yielded a near-native complex among the top-ranking five structures, and six of nine targets led to a near-native structure among the top 10 results. Therefore, the PDPs perform better than the MJ potentials on this set of decoys. Some of the current docking algorithms share comparable scoring functions that are a combination of the ACE[32] electrostatic and van der Waals interactions. The present results suggest the PDPs that are derived using a different (LP) approach may considerably improve the prediction (or discriminatory) ability of the scoring functions/potentials if combined with the currently used scoring functions. Figure 4 shows the RMSD versus energy plots for targets 5, 6, 8, and 9.

The Degree of Tolerance of the PDPs to the Unbound Protein Conformation

Protein docking in the unbound case is a significantly more challenging task than reconstruction of a complex knowing the three-dimensional (3D) structure of its subunits in the bound form. The extent of structural change that occurs upon complex formation may vary from local changes in side-chain conformations to more global ones that involve protein backbone or domain movements.

We tested the ability of the above derived PDPs to discriminate between correctly docked and misdocked complexes on 17 complexes using the unbound structures of one or both proteins. Twelve of these test cases were "unbound–unbound" problems, in the sense that the structures of both proteins in the unbound form were used as input, and the remaining five were "bound–unbound" cases, i.e., the bound form was used for one of the proteins—receptor or substrate, whereas the unbound form was taken for the other.

Results are presented in Table VI. Shape complementarity alone (columns 2 and 3) led to very poor results for the rank of the first native-like structure (RMSD < 4.0 Å) within the list of rank-ordered putative docked structures. Column 4 in Table VI presents the rank of the first native-like structure in the set of decoys rank-ordered according to the newly developed PDPs. In 10/17 cases, a native-like structure is detected among the top-ranking 100 decoys out of the set of 400,000 generated for each case; and in one case, it was ranked in the 604th place. In 5/17 cases where binding was accompanied by a large conformational change in one of the complex subunits (RMSD between bound and unbound forms >1.72 Å for interfacial C^{α} atom, and >2.01Å for all atoms; see columns 6 and 7 in Table VI), the energy function failed to rank a native-like structure within the first 2,000 decoys. The docking algorithm failed to capture a native-like orientation within the original set of 400,000 docked orientations

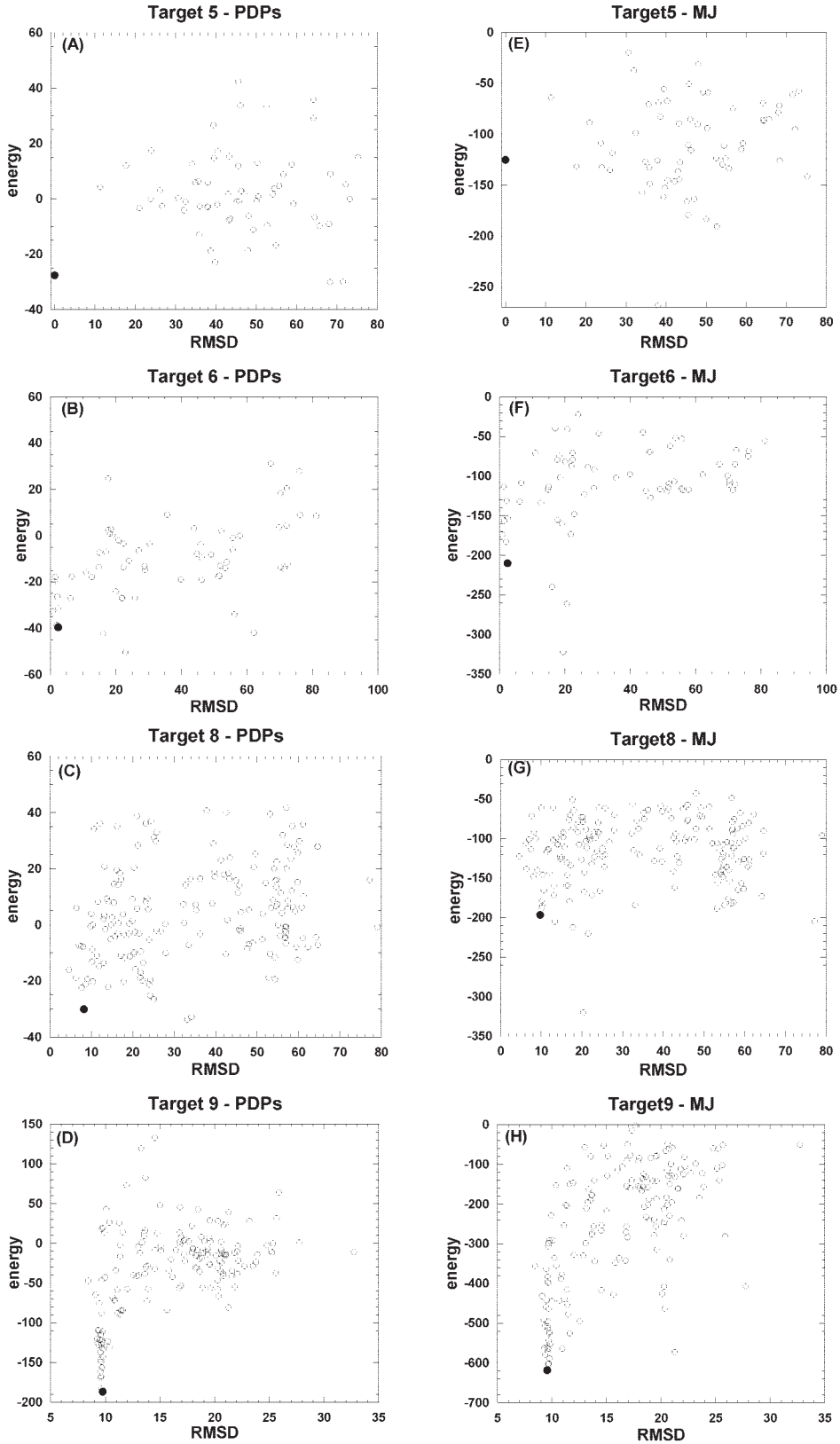

Fig. 4. RMSD versus energy (based on PDPs) plot for four CAPRI targets using the PDPs (**A–D**) and the MJ potentials (**E–H**). Results are shown for targets 5, 6, 8, and 9. The structures identified by the potentials (Table V) are shown by the filled circles. Targets 5 and 6 are antibody-antigen complexes. None of the submitted predictions for target 5 have ligand RMSD $\leq$ 10 Å from the native structure; therefore a near-native structure was included. For target 5, only two of the putative complexes have energy values below that of the native structure upon using the PDPs, compared with 26 using the MJ potentials. In as much, for targets 6 and 8, the PDPs show improved discrimination ability compared with that of the MJ potentials. For target 9, both potentials score the near-native structures better than any misdocked complex.

TABLE VI. Ranking of Docked Orientations in the Unbound Case

	Results based on shape complementarity		Results based on docking potentials		RMSD between bound and unbound forms	
Complex	Rank	RMSD	Rank	RMSD	Receptor[a]	Substrate[b]
2PTC	9,729	1.54	6	3.34	0.31; 0.47	0.23; 1.21
1AVW	25,223	1.18	26	2.83	0.49; 1.12	0.33; 1.05
1MCT	1	0.49	57	1.98	0.51; 0.99	0.0; 0.0
3TPI	13,393	0.91	5	3.68	0.38; 0.60	0.32; 0.94
1TGS	1,763	0.54	2	0.64	4.22; 4.03	0.0; 0.0
1CHO	5,185	1.43	73	3.8	0.37; 0.96	0.61; 1.39
1ACB	43,591	2.34	26,402	3.37	0.90; 0.84	2.48; 3.14
1CBW	4,032	0.86	23	1.53	0.46; 0.70	0.30; 1.08
1PPF	17,375	1.88	604	2.71	0.0; 0.0	0.83; 1.53
1FLE	4,334	2.19	22,785	4.06	0.56; 1.17	2.47; 3.53
1CSE	370	0.82	17,901	3.7	0.33; 0.53	1.98; 3.07
1VFB	74,463	1.27	2,085	4.09	0.43; 0.77	1.72; 2.01
1NCA	14,604	1.23	7	3.1	0.0; 0.0	0.35; 0.80
1BRS	812	1.54	11	3.2	0.50; 0.86	1.17; 1.48
1YDR	8	0.73	11	3.56	0.94; 1.53	0.0; 0.0
1TX4	22,723	2.8	7,667	3.63	0.49; 1.55	2.57; 4.09
1ATN	NF	—	NF	—	6.34; 6.38	0.36; 0.69

[a] C^{α} atoms RMSD; all atoms RMSD, for the receptor interfacial amino acids after overlapping the bound and unbound structures. Interfacial amino acids are defined as those whose distance is ≤6.5 Å from any amino acid of the other complex subunit. Distance was measured between side-chains' heavy atoms.
[b] C^{α} atoms RMSD; all atoms RMSD, for the substrate interfacial amino acids after overlapping the bound and unbound structures.

for one complex, actin-DNase I (1ATN). The all-atoms RMSD of the interfacial amino acids between the bound (actin-ADP) and unbound (actin-ATP) forms is 6.38 Å in this complex. These results indicate that the PDPs can satisfactorily distinguish between near-native and misdocked complexes, provided that the subunit structural changes accompanying binding do not exceed approximately 1.7 Å for C^{α} atoms.

Comparison Between Folding and Docking Potentials

The TE folding potentials, shown in our previous work to distinguish between native and nonnative *folded* structures, could not distinguish here between native-like and nonnative *docked* structures (Table III). Figure 5(A) compares the S-S contact potentials found in the two cases. The two sets of potentials do not correlate (correlation coefficient: 0.17). The same analysis also shows that the PDPs do not correlate with the MJ potentials [Fig. 5(B); correlation coefficient: 0.25] whereas the TE and the MJ potentials yield a correlation coefficient of 0.57 [Fig. 5(C)].

The differences between folding and docking potentials are consistent with the work of Lo Conte et al.[18] that shows that the amino acid composition at the interfaces of complexes is different from that in the proteins interiors or that at the interfaces of *multimers*. Cys-Cys interaction, for example, is the third most attractive interaction in the TE folding potential. However, it is the most repulsive interaction in the present PDPs. This can be rationalized by the fact that the Cys-Cys pairings tend to stabilize folded structures via formation of a disulfide bond; however, formation of a covalent bond in protein interface is undesirable because complexes are transient, and complexes association and dissociation need to be synchronized for normal cell function. Conversely, Arg-Trp interaction is one of the most attractive interactions in the PDP, whereas it is almost neutral in folding potentials. Arg-Tyr interaction is very favorable PDP, but repulsive in the folding potentials. Similarly Lys-Trp interaction is a favorable PDP, but slightly unfavorable in folding. Not surprisingly, Arg, Tyr, and Trp are distinguished by their high frequency at protein–protein interfaces' hot spots.[33] Hydrophobic interactions such as Phe-Phe, Leu-Leu, Ile-Val, and Ile-Trp remain highly attractive in both potentials, whereas Lys-Lys, Glu-Glu, and Arg-Gln are the most repulsive interactions in both potentials.

DISCUSSION AND CONCLUSION

In this work, we constructed a new set of PDPs using an LP technique at intermediate-resolution model that is relatively insensitive to the precise orientations of side-chains and still retains the ability to distinguish between native and nonnative conformations. The PDPs were shown to accurately distinguish between native-like and nonnative conformations for different types of complexes tested. The advantages of the presently proposed potentials are twofold: (a) they enable us to score putative docked structures very fast, and (b) they are relatively insensitive to the precise geometry of the complex. These advantages support the use of the PDPs in exhaustive search protocols used in rigid-body docking algorithms.

The idea of using coarse-grained PDPs for protein–protein interactions is not new. A challenge is, however, to design the *optimal* set of potentials given the limitations on the available data on *transient* complexes. Most of the previously derived statistical potentials for protein dock-

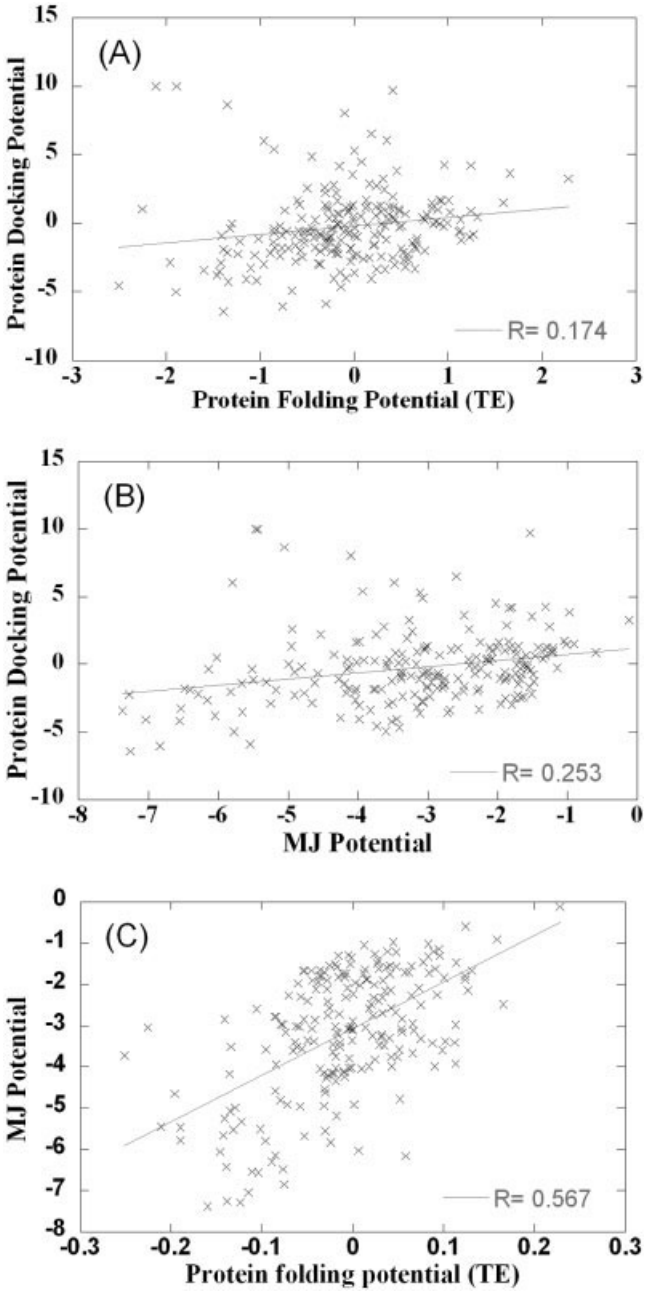

Fig. 5. Comparison of side-chain–side-chain contact potentials for protein docking and protein folding. **A and B:** Comparison of the optimally designed PDPs (ordinate) with the TE (A) and MJ (B) potentials derived from folded structures. Weak correlation is observed between the two sets. **C:** Comparison of the two sets of folding potentials. The best fitting lines and correlation coefficients are shown on the panels.

ing have used in their training combined sets intermolecular interactions in complexes and multimeric proteins[16,19,20] or even intramolecular interactions in proteins domains.[16] Moont et al.[16] recently reevaluated a subset of docked conformations generated by FTDock[3] using different potentials. Interestingly, the best results were found with the potentials extracted from *intramolecular* interactions in nonhomologous protein *domains*, as opposed to *intermolecular* interactions in nonredundant *complexes*. This is at odds with the fact that the amino acid composition at complexes' interface is distinct from that in protein interiors, and from that at oligomeric interfaces.[18] A possible explanation for the weaker performance of intermolecular potentials is the limited data exploited to derive them using the statistical approaches. Using the LP technique, we were able to alleviate this problem and derive an optimal set of PDPs strictly from a set of *transient* protein complexes, which outperform the commonly used MJ potentials (Table III). Most significantly, the PDPs were able to identify (among the top 10 rank-ordered structures) native-like complex structures.

The potentials tolerate small deviations in side-chain conformations with respect to those in the unbound form, as evidenced by the unbound–unbound docking tests (Table VI). As an attempt to increase the tolerance of the potentials to larger changes in side-chain orientations, we increased the gap ϵ_i between the energy of the native complex structure and that of the ith decoy by adopting a proportionality of the form $\epsilon_i \propto \min\{RMSD(\hat{X}_a{:}\hat{X}_b, \hat{X}_a{:}X_b^i), 10\}$, i.e., a higher ϵ_i value is adopted if the RMSD is larger, whereas the PDPs were constrained to be $<|10|$ units. This approximation led, however, to infeasible solution, and we could not even obtain an approximate solution that would satisfy the majority of the inequalities. This suggests that further tolerance to side-chain conformational changes may necessitate the adoption of functional forms more sophisticated than the simple expression (Eq. 3) presently adopted.

Comparison of the present PDPs with the folding potentials reveals some similarities in interresidue interactions in accord with their physical characteristics. However, others seem to reflect biological functional aspects. A most striking example is the Cys-Cys interactions that are very attractive in Keskin et al.[20] potential. In addition, our potential favors interaction involving Arg, Tyr, and Trp. This is consistent with the experimentally observed higher frequency of these residues at protein interfaces' hot spots.[33]

The conclusions that can be drawn from the present study, which may assist in future protein–protein docking computations are: (i) the important role of contact topology, evidenced by the ability of shape complementarity criterion alone (or scoring function based on the "number" of contacts) to accurately predict the docked structure in 1/3 of examined bound-docking cases, (ii) the possibility of improving the former success rates by 100% upon adoption of optimal PDPs, (iii) the qualitative and quantitative differences between folding and docking potentials of mean force and inadequacy of adopting folding potentials for docking simulations.

ACKNOWLEDGMENTS

Partial support from the National Institutes of Health and National Science Foundation is gratefully acknowledged.

REFERENCES

1. Palma PN, Krippahl L, Wampler JE, Moura JJ. BiGGER: a new (soft) docking algorithm for predicting protein interactions. Proteins 2000;39:372–384.
2. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an

automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 2004;20:45–50.
3. Gabb HA, Jackson RM, Sternberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 1997;272:106–120.
4. Mandell JG, Roberts VA, Pique ME, et al. Protein docking using continuum electrostatics and geometric fit. Protein Eng 2001;14: 105–113.
5. Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. Proteins 2000;39:178–194.
6. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003;52:80–87.
7. Norel R, Sheinerman F, Petrey D, Honig B. Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. Protein Sci 2001;10:2147–2161.
8. Fernandez-Recio J, Totrov M, Abagyan R. Soft protein-protein docking in internal coordinates. Protein Sci 2002;11:280–291.
9. Vakser IA, Matar OG, Lam CF. A systematic study of low-resolution recognition in protein-protein complexes. Proc Natl Acad Sci USA 1999;96:8477–8482.
10. Camacho CJ, Vajda S. Protein-protein association kinetics and protein docking. Curr Opin Struct Biol 2002;12:36–40.
11. Gray JJ, Moughon S, Wang C, et al. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 2003;331:281–299.
12. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 2002;47:409–443.
13. Ben-Zeev E, Eisenstein M. Weighted geometric docking: incorporating external information in the rotation-translation scan. Proteins 2003;52:24–27.
14. Li L, Chen R, Weng Z. RDOCK: refinement of rigid-body protein docking predictions. Proteins 2003;53:693–707.
15. Murphy J, Gatchell DW, Prasad JC, Vajda S. Combination of scoring functions improves discrimination in protein-protein docking. Proteins 2003;53:840–854.
16. Moont G, Gabb HA, Sternberg MJ. Use of pair potentials across protein interfaces in screening predicted docked complexes. Proteins 1999;35:364–373.
17. Berman HM, Battistuz T, Bhat TN, et al. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2002;58:899–907.
18. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol 1999;285:2177–2198.
19. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N. Residue frequencies and pairing preferences at protein-protein interfaces. Proteins 2001;43:89–102.
20. Keskin O, Bahar I, Badretdinov AY, Ptitsyn OB, Jernigan RL. Empirical solvent-mediated potentials hold for both intramolecular and inter-molecular inter-residue interactions. Protein Sci 1998;7:2578–2586.
21. Tobi D, Shafran G, Linial N, Elber R. On the design and analysis of protein folding potentials. Proteins 2000;40:71–85.
22. Tobi D, Elber R. Distance-dependent, pair potential for protein folding: results from linear optimization. Proteins 2000;41:40–46.
23. Meller J, Elber R. Linear programming optimization and a double statistical filter for protein threading protocols. Proteins 2001;45: 241–261.
24. Vendruscolo M, Najmanovich R, Domany E. Can a pairwise contact potential stabilize native protein folds against decoys obtained by threading? Proteins 2000;38:134–148.
25. Loose C, Klepeis JL, Floudas CA. A new pairwise folding potential based on improved decoy generation and side-chain packing. Proteins 2004;54:303–314.
26. Maiorov VN, Crippen GM. Contact potential that recognizes the correct folding of globular proteins. J Mol Biol 1992;227:876–888.
27. Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 1996;256: 623–644.
28. Betancourt MR, Thirumalai D. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. Protein Sci 1999;8:361–369.
29. Meszaros C. Fast Cholesky factorization for interior point methods of linear programming. Comp Math Appl 1996;31:49–51.
30. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. J Mol Biol 1993;234:946–950.
31. Janin J, Henrick K, Moult J, et al. CAPRI: a critical assessment of PRedicted interactions. Proteins 2003;52:2–9.
32. Zhang C, Vasmatzis G, Cornette JL, DeLisi C. Determination of atomic desolvation energies from the structures of crystallized proteins. J Mol Biol 1997;267:707–726.
33. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol 1998;280:1–9.