JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Protein−Protein Docking Using EMAP in CHARMM and Support Vector Machine: Application to Ab/Ag Complexes

Jon D. Wright,[†,‡] Karen Sargsyan,[†] Xiongwu Wu,[§] Bernard R. Brooks,[§] and Carmay Lim[†,∥,*]

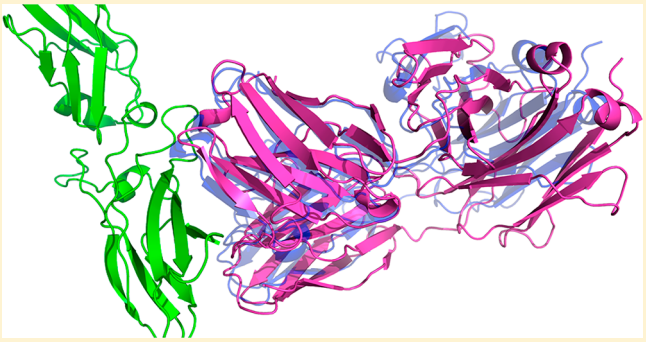
[†]Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan

[‡]Genomics Research Institute, Academia Sinica, Taipei 115, Taiwan

[§]Laboratory of Computational Biology, NHLBI, National Institutes of Health, Bethesda, Maryland, United States

[∥]Department of Chemistry, National Tsinghua University, Hsinchu 300, Taiwan

ABSTRACT: In this work, we have (i) evaluated the ability of the EMAP method implemented in the CHARMM program to generate the correct conformation of Ab/Ag complex structures and (ii) developed a support vector machine (SVM) classifier to detect native conformations among the thousands of refined Ab/Ag configurations using the individual components of the binding free energy based on a thermodynamic cycle as input features in training the SVM. Tests on 24 Ab/Ag complexes from the protein−protein docking benchmark version 3.0 showed that based on CAPRI evaluation criteria, EMAP could generate medium-quality native conformations in each case. Furthermore, the SVM classifier could rank medium/high-quality native conformations mostly in the top six among the thousands of refined Ab/Ag configurations. Thus, Ab-Ag docking can be performed using different levels of protein representations, from grid-based (EMAP) to polar hydrogen (united-atom) to all-atom representation within the same program. The scripts used and the trained SVM are available at the www.charmm.org forum script repository.



1. INTRODUCTION

Protein−protein interactions, defined as specific physical contacts between protein pairs that occur by selective molecular docking in a particular biological context,[1] can generally be divided into two types:[2,3] (i) *obligate* interactions of protomers that cannot exist on their own *in vivo* (e.g., multisubunit enzymes) and (ii) *nonobligate* interactions of protomers that can exist independently (e.g., antibody (Ab)−antigen (Ag) complexes). They are critical to many biological functions such as signal transduction and immune response and are therapeutic drug targets.[4,5] Hence, a detailed understanding of the mechanisms of protein association is of wide interest and of importance for drug design. Knowledge of the 3-dimensional (3D) structure of the protein−protein complex is prerequisite for understanding how proteins associate. However, experimental determination of these protein−protein complex structures by X-ray, NMR, and cyroelectron microscopy is time-consuming and is limited by the size of the complex. Thus, *in silico* protein−protein docking approaches, which can predict the complex structure from the coordinates of the unbound component proteins, complement experimentally determined protein complex structures. In this work, we describe the EMAP method[6,7] implemented in the CHARMM program[8] for producing Ab/Ag configurations, and the use of a support vector machine (SVM)[9] to distinguish native, low-energy Ab/Ag conformations from other low-energy ones.

Since there are many reviews on flexible protein−protein docking methods[10−27] and the performance of docking predictions,[25,28−35] we summarize below the two key problems common to all protein−protein docking methods: The first is the "conformational search problem" to generate the correct conformation of the two proteins together, and the second is the "target identification problem" to recognize a native conformation among the millions of configurations generated. In practice, the first problem is often the simpler of the two problems to solve, as the increase in computer power has increased the number of different configurations that can be produced in a reasonable amount of time. This almost guarantees that one of the millions of configurations generated is close to the native conformation if the individual proteins do not undergo extensive conformational change upon binding. To solve the problem of recognizing a native or near-native conformation from the millions of non-native ones that are generated, many different scoring functions have been developed over the past 20 years. The methods available to solve these two problems are summarized below.

Conformational Search Problem (Problem 1). Most methods generate the initial docked configurations in a two-step process: The first step is a rigid-body global/local search for the relative position and orientation of the two proteins and

Received: February 11, 2013
Published: August 16, 2013

the second step is refinement of a subset of the configurations generated to account for backbone and side chain flexibility. If the conformational change upon binding is not significant and no *a priori* information on the complex is available, fast Fourier transform (FFT)-based algorithms[10,12,17,36−44] and geometric hashing[45] can be used to globally search the entire rotational and translational space of the proteins to identify approximate binding modes. FFT-based approaches describe the protein by grid points, scan the entire solution space by systematically rotating and translating one protein grid about the other, and evaluate the docked configurations on a grid using correlation-type scoring functions.[46] In geometric hashing,[47] the protein surface is described by critical points and the point representations of the two proteins are then compared to detect conformations with optimal geometric complementarity.[48] Rather than screening all possible rigid-body docked configurations, Monte Carlo (MC) simulated annealing,[49] and genetic algorithms[50] explore only the potentially useful regions of the conformational space. Although these methods can treat more degrees of freedom, the search is confined to the starting structure neighborhood and may miss the correct conformation without any *a priori* binding mode information. If reliable information on interface residues is available, then restraint-based docking can be used to guide the search toward the right region of the conformational space in which the restraints are satisfied.[51]

Rigid-body docking using FFT and geometric hashing generally yield a few thousand putative complex structures, which often include steric overlap. These structures have to be refined to account for backbone and side chain conformational changes that occur upon protein association. One way to refine the structures is to employ energy minimization using all-atom force fields to remove steric overlap and optimize the van der Waals (vdW) and electrostatic interactions. Other methods used to refine structures include molecular dynamic simulations,[51−53] MC and related algorithms,[49,54−57] genetic algorithms,[50] multicopy approaches,[58] self-consistent mean-field approaches,[59,60] principal component analysis,[61,62] and normal modes analysis.[63] However, structure refinement can still fail to produce a native docking pose if the conformational changes upon binding are large.

Target Identification Problem (Problem 2). In determining the 'correct' solution from the few thousand putative docked orientations, several scoring functions have been developed. The scoring functions commonly employ criteria such as terms related to shape complementarity, chemical properties, solvation terms, and evolutionary conservation to rank the orientations. They can generally be grouped into three classes: forcefield-based, empirical, and knowledge-based scoring functions. Shape complementarity is harder to detect in protein−protein interactions especially in Ab/Ag with flat interfaces compared to receptor−drug complexes with concave/convex interfaces.[34,64] Scoring functions specific for three classes of complexes; viz., (I) enzyme−inhibitor/substrate complexes, (II) Ab/Ag complexes, and (III) other protein−protein complexes not belonging to class I or II have been obtained using various physical, chemical, and biological parameters to train separate SVMs for each class.[65,66] Each of these class-specific functions has been shown to outperform several nonclass-specific scoring functions for the respective type of protein−protein complexes.[65,66]

This work provides a calibration of the EMAP docking module implemented in the CHARMM program to generate near-native conformations of Ab/Ag complexes and describes the training and testing of a SVM classifier to recognize native Ab/Ag conformations. EMAP allows protein structures to be represented using map objects instead of atoms. Each map object represents a distribution of an atomic property such as vdW energies, electric charges, or electrostatic potentials on grid points.[6] Separate maps can be docked together and the map interaction scored and minimized using grid-threading MC minimization to determine good and bad map interaction orientations. The key advantage of EMAP relative to other protein−protein docking methods is that it allows the use of CHARMM force fields, implicit solvation methods such as Poisson−Boltzmann or generalized Born, and ready access to the wide range of computational tools in CHARMM. However, no calibration of the EMAP method for protein−protein docking had been performed. Hence, we have evaluated the ability of EMAP to generate Ab/Ag complex structures (problem 1). We have also used individual components of the binding free energy and the solvent-accessible surface area (SASA) change upon binding to train and test a SVM to spot native conformations for Ab/Ag complexes (problem 2). The free energy components are based on a thermodynamic cycle (see Scheme 1 in Methods) and include the gas-phase vdW and electrostatic interaction energies as well as the electrostatic and nonelectrostatic solvation *free* energy changes upon binding. Thus, some of the entropic effects upon binding of an Ab to an Ag are taken into account; however, vibrational and conformational side chain entropy changes have been neglected (see Discussion).

2. METHODS

2.1. Benchmark Data Set. The docking methodology was calibrated using the Ab/Ag structures from the protein−protein docking benchmark version 3.0,[67] except 1e4k. This complex was excluded because the Ag is an Fc receptor bound by the Fc rather than the Fv region of the Ab. For 12 of the 24 Ab/Ag complexes, the free structures of both Ab and Ag had been solved, but for the remaining complexes, only the free Ag structures had been solved, so the free Ab structure was taken from the bound complex in the docking benchmark. The free structures were docked, and their docked conformations were compared to the respective solved complex structures. For each system, the Ab was assigned as the ligand and the Ag as the receptor. All preparation, EMAP docking, and postdocking refinement calculations were performed using version 35 of the CHARMM program.[8]

2.2. Preparing the Protein Structures for Docking. First, water molecules in the free protein structure were removed. Since hydrogen atoms are generally absent in the crystal structures, polar hydrogen atoms were added using the CHARMM united atom force field.[68,69] Histidine residues were protonated/deprotonated using the Reduce program,[70] which attempted to optimize the number of hydrogen bonds and reduce steric clashes in the local geometry. All Asp/Glu residues were assumed to be deprotonated, while Lys and Arg residues were assumed to be protonated. After adding polar hydrogen atoms, steepest descent minimization using a distance-dependent dielectric was performed for 100 steps with constraints on the heavy atoms to restrain them to their positions in the X-ray structure.

2.3. Rigid Body Protein−Protein Docking Using EMAP. Since the Ab CDR loops are involved in binding the Ag,[71] the docking search was limited to the vicinity of the

CDRs: only atoms within 8 Å of the Ab CDR loop atoms were considered for the Ab ligand map objects, but all atoms were included for the Ag receptor map objects. The map object grid was set to 2 Å. To speed up the search, two parallel dockings were carried out for each system: The entire receptor search space was partitioned into two independent search areas, one for each side of the receptor, allowing two processors to be used for the search simultaneously. The receptor search grid was set to a minimum of 6 grid points per side. If the map object created by the receptor was >72 Å on any one side, the number of starting positions was increased to allow for a minimum of one search grid starting point every 12 Å on each side. Thus, the number of receptor search points varies from 152 to 728. For each receptor search grid point, six starting rotations were assigned to the ligand; only the grid search points on the outer ligand protein surface defined by atoms within 8 Å of the CDR loops were used. These restrictions yielded 11 552 and 55 328 initial Ab/Ag orientations for the smallest and largest systems, respectively. Each search involved 50 cycles of grid-threading MC minimization using a distance-dependent dielectric for 20 steps, which allowed a maximum of 1.0 Å translation and 10° rotation for the ligand, generating between 11 552 000 and 55 328 000 sampled orientations per complex.

2.4. Refining the Rigid Body Configurations Using MC Minimization. The top EMAP-scoring 2000 Ab/Ag configurations generated for each of the two independent searches were refined by a postdocking MC and minimization protocol using the standard CHARMM energy function and a united atom force field. Each of the 4000 configurations for a given Ab/Ag complex was minimized using adopted-basis Newton−Raphson for 200 steps excluding electrostatic interactions. Subsequently, the electrostatic interactions were included using a distance-dependent dielectric and the minimized configuration was subjected to 4000 steps of Metropolis MC[72] at 300 K allowing two types of moves: (i) the ligand could move via rigid body translation up to 0.3 Å or (ii) torsion angles defined by side-chain atoms of close contact site residues (whose ligand and receptor heavy-atoms are separated by ≤5 Å) could be rotated by ≤30°. During the 4000 attempted MC moves, the ligand translation was attempted for 1/3 of the moves, while the torsion rotation was attempted for 2/3 of the moves. The resulting configuration was reminimized using adopted-basis Newton−Raphson for 300 steps including electrostatic interactions.

2.5. Measuring the Fit between the Docked and X-ray Structures. To evaluate the quality of the docked configurations, we used CAPRI criteria, which employ three metrics, f_{nat}, L_rms, and I_rms.[32,73] f_{nat} is the fraction of the native residue−residue contacts (where an Ab residue was considered to be in contact with an Ag residue if any of their atoms were within 5 Å) in the X-ray structure that is retained in the docked configuration. L_rms measures the overall geometric fit between the docked and X-ray structures by the root-mean-square deviation (RMSD) of the backbone (N, C, C^{α}, and O) atoms of the "ligand" (the smaller of the two proteins) in the docked structure from those in the respective X-ray structure after superposition of the backbone atoms of the larger proteins in the two structures using PyMOL. I_rms measures how the docked structure fits the X-ray structure in the interface region by the RMSD of the backbone atoms of the interface residues (defined as those with one or more heavy atoms within 10 Å of a heavy atom belonging to the other protein) in the docked structure from those in the respective X-ray structure after superposition of the backbone atoms of these residues in the two structures using PyMOL. We also calculated I_C^{α}-rms, which is the same as I_rms except that the superposition was performed only for the C^{α} atoms.[42]

2.6. Defining Native Conformations. According to the CAPRI criteria, a docked configuration is deemed to be highly accurate if it retained half or more of the native residue−residue contacts in the X-ray structure ($f_{nat} \geq 0.5$) and the overall geometric or interface fit between the docked and X-ray structures is ≤1 Å (L_rms or I_rms ≤1 Å). It is deemed to be of medium accuracy if f_{nat} is ≥0.5 but L_rms or I_rms is >1 Å, but if f_{nat} is <0.5 but ≥0.3, then L_rms should be ≤5 Å or I_rms ≤2 Å. A docked configuration is considered to be "acceptable" if f_{nat} is ≥0.3, L_rms >5 Å, and I_rms >2 Å, or if f_{nat} is <0.3 but ≥0.1, then L_rms should be ≤10 Å or I_rms ≤4 Å. Docked configurations ranked as high or medium accuracy were defined as natives, while the remaining docked configurations were regarded as non-natives. These results were compared with those obtained by ZDOCK[42] where a hit was defined as a docked structure with I_C^{α}-rms ≤ 2.5 Å.[74]

2.7. Training and Test Data Sets for SVM. The SVM was trained on configurations that were nonidentical to the 4000 postdocking refined configurations (test set). To obtain an independent training set of native configurations, each native configuration from the test set for a given Ab/Ag complex was subjected to 2000 MC steps at 600 K by rotating any torsion angle for the close contact site residues by ≤30°. The final accepted configuration was used as the starting point for another 2000 MC steps. This was repeated 14 times giving an enrichment of ≤15 times more native configurations, denoted as eNAT. To obtain an independent training set of non-native configurations, 600 non-native configurations from the test set were randomly selected for each complex. Each of these 'misses' was subjected to the same MC protocol used to enrich the native structures except that only one set of 2000 MC steps was performed per structure. This enrichment procedure yielded 600 new non-native structures per complex, denoted as e-nonNAT. Thus, the training data set for the SVM consists of all enriched eNAT and e-nonNAT configurations, whereas the test data set consists of the "original" refined configurations (4000 for each Ab/Ag complex).

2.8. Input Features for SVM Training. The 4000 Ab/Ag configurations refined with MC followed by energy minimization were scored by a SVM, which predicted whether a docked configuration was native or not.[65,66] The input features for training a SVM were the components of the free energy of binding an Ab to an Ag ($\Delta G^{protein}$) based on the following thermodynamic cycle:[75]

$$\begin{array}{ccccc} & & & \xrightarrow{\Delta G_{gas}} & \\ \text{Ab (gas)} & + & \text{Ag (gas)} & \rightarrow & \text{Ab/Ag (gas)} \\ \Delta G_{solv}\text{(Ab)} \downarrow & & \downarrow \Delta G_{solv}\text{(Ag)} & & \downarrow \Delta G_{solv}\text{(Ab/Ag)} \\ \text{Ab (protein)} & + & \text{Ag (protein)} & \rightarrow & \text{Ab/Ag (protein)} \\ & & & \Delta G_{protein} & \end{array}$$

(scheme 1)

The free energy change upon binding of the Ab and Ag in the gas phase, ΔG_{gas}, can be approximated by the gas-phase energy change, ΔE_{gas}, which in turn can be estimated as a sum of the vdW (ΔE_{gas}^{vdW}) and electrostatic (ΔE_{gas}^{ele}) interaction energies; that is, $\Delta G_{gas} \sim \Delta E_{gas}^{vdW} + \Delta E_{gas}^{ele}$. The solvation free energy, ΔG_{solv}, can be approximated as a sum of the electrostatic (ΔG_{solv}^{elec}) and nonelectrostatic ($\Delta G_{solv}^{nonelec}$)

Table 1. EMAP Docking Results for the Ab/Ag Complexes Using the Respective Unbound−Unbound or Bound-Unbound Structures

Ab/Ag[a]	Ab	Ag	I_C$^{\alpha}$-rms[b] (Å)	EMAP combined[c]	EMAP separated[d]	acceptable combined[e]	acceptable separated[f]
			Using Unbound−Unbound Structures				
1ahw	1fgn	1tfh	0.69	1243	352	89	139
1bgx*	1ay1	1cmw	1.48	875	580	86	77
1bvk	1bvl	3lzt	1.24	252	216	88	77
1dqj	1dqq	3lzt	0.75	871	552	102	84
1e6j	1e6o	1a43	1.05	372	69	95	96
1jps	1jpt	1tfh	0.51	351	153	73	73
1mlc	1mlb	3lzt	0.60	492	256	39	40
1vfb	1vfa	8lyz	1.02	1101	818	79	69
1wej	1qbl	1hrc	0.31	75	32	163	157
2fd6	2fat	1ywh	1.07	469	46	73	73
2i25	2i24	3lzt	1.21	705	538	123	111
2vis	**1gig**	**2viu**	**0.80**	***34 926***	***18 950***	***0***	***0***
			Using Bound-Unbound Structures				
1bj1	1bj1	2vpf	0.50	3499	1613	8	8
1fsk	1fsk	1bv1	0.45	136	51	140	140
1iqd	1iqd	1d7p	0.48	1	1	75	75
1i9r	1i9r	1aly	1.30	1497	731	28	27
1k4c	1k4c	1jvm	0.53	1287	784	62	60
1kxq	1kxq	1ppi	0.72	37	37	154	149
1nca	1nca	7nn9	0.24	153	57	48	48
1nsn	1nsn	1kdc	0.35	91	18	156	156
1qfw	1qfw	1hrp	1.31	1996	1022	34	34
2hmi**	2hmi	1s6p	2.26	3335	1685	6	6
2jel	2jel	1poh	0.17	121	85	192	192
2qfw	2qfw	1hrp	0.73	2161	1172	39	39

[a]* and ** mean the test cases are assigned as medium (1.5 Å < I_C$^{\alpha}$-rms ≤ 2.2 Å) and difficult (I_C$^{\alpha}$-rms > 2.2 Å), based on the structural difference (I_C$^{\alpha}$-rms) between the bound and the unbound forms of the binding partners.[67] The results in the 1st 12 rows were derived from free structures of both Ab and Ag, while the remaining results were derived from free structure of Ag only [b]I_C$^{\alpha}$-rms of interface residues after finding the best superposition of the experimental bound and unbound interfaces. [c]Lowest EMAP rank of a native conformation among the combined top 4000 rigid-body configurations from both sets of docking calculations [d]Lowest EMAP rank of a native conformation among the top 2000 + 2000 rigid-body configurations from each of the two parallel sets of docking calculations. [e]Number of "acceptable" or better configurations generated by EMAP among the combined top 4000 rigid-body configurations from both sets of docking calculations. [f]Number of "acceptable" or better configurations generated by EMAP among the top 2000 + 2000 rigid-body configurations from each of the two parallel sets of docking calculations.

terms; that is, $\Delta G_{solv} \sim \Delta G_{solv}{}^{elec} + \Delta G_{solv}{}^{nonelec}$. Thus, the change in the solvation free energy upon binding, $\Delta\Delta G_{solv}$, can be estimated as a sum of $\Delta\Delta G_{solv}{}^{elec} = \Delta G_{solv}{}^{elec}(Ab/Ag) - [\Delta G_{solv}{}^{elec}(Ab) + \Delta G_{solv}{}^{elec}(Ag)]$ and $\Delta\Delta G_{solv}{}^{nonelec} = \Delta G_{solv}{}^{nonelec}(Ab/Ag) - [\Delta G_{solv}{}^{nonelec}(Ab) + \Delta G_{solv}{}^{nonelec}(Ag)]$. Hence,

$$\Delta G_{protein} \sim \Delta E_{gas}{}^{vdW} + \Delta E_{gas}{}^{ele} + \Delta\Delta G_{solv}{}^{elec} + \Delta\Delta G_{solv}{}^{nonelec} \quad (1)$$

The individual components of $\Delta G^{protein}$ were used as input features, namely,

1. the gas-phase electrostatic ($\Delta E_{gas}{}^{ele,nonCDR}$) interaction energies with the Ag receptor from Ab non-CDR atoms within 4.2 Å of any Ag atom in the complex structure,
2. the gas-phase electrostatic ($\Delta E_{gas}{}^{ele,CDR}$) and vdW ($\Delta E_{gas}{}^{vdW,CDR}$) contributions involving only the Ab CDR loops with the Ag, which were separated into three ranges depending on the distance R_{ij} between an Ab CDR loop atom i and an Ag atom j in the complex structure: short-ranged if R_{ij} was <5.4 Å, medium-ranged if 5.4 ≤ R_{ij} < 8.0 Å, long-ranged if 8.0 ≤ R_{ij} < 20 Å,
3. solvent-accessible surface area (SASA) change of the Ag receptor and Ab ligand, which was computed with a probe radius of 1.4 Å using CHARMM, and
4. electrostatic ($\Delta\Delta G_{solv}{}^{elec}$) and nonelectrostatic ($\Delta\Delta G_{solv}{}^{nonelec}$) solvation free energy changes upon binding computed using the FACTS[76] method in CHARMM.

The electrostatic solvation free energy of the protein was estimated using the Generalized Born formula:

$$\Delta G_{solv}^{elec} = -\frac{1}{2}\left(\frac{1}{\varepsilon_{protein}} - \frac{1}{\varepsilon_{water}}\right) \times \sum_{ij=1}^{N} \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/\kappa R_i R_j)}} \quad (2)$$

where the summation is over the N atoms in the system, $\varepsilon_{protein}$ is the dielectric constant of the protein, ε_{water} is the dielectric constant of water, q_i is the partial charge of atom i in electronic units, r_{ij} is the interatomic distance between atoms i and j in Å, R_i is the effective Born radius in Å, and κ is a constant. In computing $\Delta G_{solv}{}^{elec}$, $\varepsilon_{protein} = 2$, $\varepsilon_{water} = 78.5$, $\kappa = 4$, and R_i was derived from the electrostatic solvation free energy of atom i, which was computed using an empirical equation whose

dx.doi.org/10.1021/ct400508s | J. Chem. Theory Comput. 2013, 9, 4186−4194

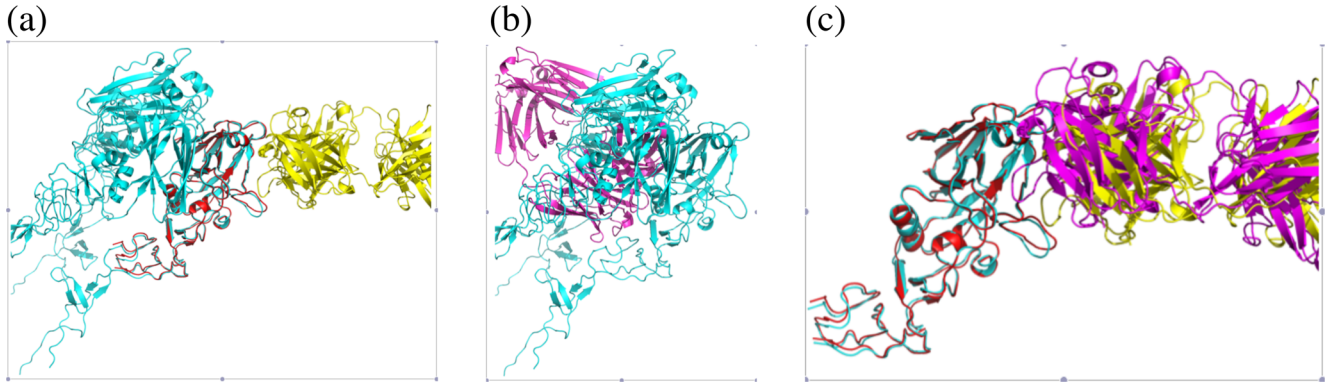


Figure 1. (a) X-ray structure (2vis) of influenza virus hemagglutinin Ag (red) bound to a neutralizing Ab (yellow); the bound Ag, which is a monomer consisting of 267 residues, is superimposed on the X-ray structure of the free Ag (2viu, turquoise), which is a trimer consisting of 320 residues in each chain. (b) The best EMAP-docked structure from two parallel docking runs of the free Ab (1gig, magenta) and the free Ag trimer (2viu, turquoise). (c) The best-refined structure from two parallel docking runs of the free Ab (1gig, magenta) and the free Ag 267-residue monomer (turquoise) superimposed on the X-ray structure of the Ag (red)/Ab (yellow) complex (2vis).

parameters were obtained by fitting to atomic solvation free energies from numerical solution of the Poisson equation.[76] The nonelectrostatic solvation free energy ($\Delta G_{solv}^{nonelec}$) was estimated as a sum of the atomic SASA values, S_i:

$$\Delta G_{solv}^{nonelec} = \gamma \sum_{i=1}^{N} S_i \quad (3)$$

where γ is an empirical surface tension parameter. The SASA of atom i (S_i) was computed from an empirical equation whose parameters were derived by fitting to SASA values calculated by an exact analytical method.[77]

2.9. Training and Testing of SVM. Using the above features, the possibility of a docked configuration to be a native conformation was ranked via the ε-regression approach with a radial basis function (Gaussian) kernel using LIBSVM.[78] The ε-regression approach employed the ε-insensitive loss function that assigns no penalty if the predicted value differed from the real value by less than ε. The parameter ε was searched in the proximity of its default value, 0.1. The Gaussian kernel has two parameters, C and γ, which were searched using a grid in the interval (2^{-8}, 2^{8}). The best C, γ, and ε parameters were obtained using 10-fold cross validation. An optimal value of C = 4, γ = 2, ε = 0.125 yielded the best cross-validation accuracy for the training set. To prevent overfitting, only models where the number of configurations in the training data set divided by the number of support vectors is greater than 2.5 were accepted. These C, γ, and ε values and the classifier based on the training data set were used to predict native conformations among the 4000 postdocking refined configurations.

2.10. Computation Time. Using two cores from either an Intel i7 or AMD PhenomII CPU, rigid-body docking of an Ab to an Ag using EMAP to produce a complex such as 1ahw with 630 residues took ~4 h. Refining the top EMAP-scoring 2000 Ab/Ag configurations from each docking run using MC minimization takes around 17 h. The trained SVM classifier takes around 20 s to rank the 4000 refined configurations.

3. RESULTS

3.1. EMAP Performance. In the following, we will discuss the results in relation to the two key challenges that need to be overcome for a successful protein−protein prediction (see Introduction). The first challenge is the initial generation of a native conformation for the bound complex. To determine if EMAP docking followed by grid-threading Monte Carlo minimization could produce native conformations for the 24 Ab/Ag complexes, we computed the CAPRI quality of the lowest energy configurations obtained after grid-threading Monte Carlo minimization from each search point of an Ab/Ag complex (see Methods). The resulting values indicate that the EMAP method could produce native conformations of medium accuracy (see section 2.6) for all 24 Ab/Ag complexes. Thus, EMAP could solve the first problem of protein−protein docking in that the set of EMAP-docked structures contained native conformations.

For each Ab/Ag complex, two parallel sets of docking calculations were carried out (see Methods), and the docked configurations were ranked according to their EMAP scores. To determine an "optimal" subset of EMAP-docked structures for postdocking refinement, the lowest rank of a native configuration for each Ab/Ag complex was determined for two subsets of EMAP-docked structures: subset I contains the "combined" EMAP-scored configurations from both sets of docking calculations, while subset II contains the "separate" EMAP-scored configurations from the parallel sets of docking calculations. The results in Table 1 show that for each Ab/Ag complex, the lowest rank of a native configuration from subset II is less than or equal to that from subset I. Furthermore, for 23/24 Ab/Ag complexes with EMAP-docked structures, EMAP could rank one or more native conformations within the top 2000 from each docking calculation, except 2vis. Thus, the top-scoring 2000 Ab/Ag configurations from each set of docking calculations rather than the top-scoring 4000 Ab/Ag configurations from both sets were selected for postdocking refinement. Note, however, that the total numbers of "acceptable" or better conformations from the combined top 4000 or the top 2000 + 2000 rigid-body configurations are comparable.

A plausible reason why EMAP failed to yield a native conformation among the configurations selected for refinement for 2vis is because the *free* Ag (2viu, turquoise), influenza virus hemagglutinin, is a trimer consisting of 320 residues each, whereas the *bound* Ag (2vis, red) complexed with a neutralizing Ab (2vis, yellow) is a monomer consisting of 267 residues (Figure 1a). Using the trimeric structure of the free Ag (2viu), EMAP attempted to dock the Ab CDRs in the cavity formed by the trimer (Figure 1b), which was scored more favorably than

the solved complex. Since the Ag in the bound 2vis complex consists of 267 residues, the rigid-body docking was repeated using a 267-residue truncated single chain from the unbound structure. This produced four medium quality conformations in the top 2000 + 2000 configurations and the best-refined configuration is shown in Figure 1c.

3.2. Refinement Performance. To evaluate whether the refinement improved the quality of the EMAP docked configurations, the top-scoring 2000 Ab/Ag configurations from each set of docking calculations were evaluated using the CAPRI metrics (f_{nat}, L_rms, and I_rms) before and after the MC minimization refinement stage. Comparison of the f_{nat}, L_rms, and I_rms for the lowest I_rms native configurations from the top 2000 + 2000 EMAP configurations before and after refinement in Table 2 show that the MC minimization process could refine nine of the 23 Ab/Ag complexes to produce at least one of high accuracy, whereas EMAP did not generate any configurations of high accuracy. Furthermore, the MC minimization produced a conformation whose interface was generally closer to the observed Ab/Ag interface: the f_{nat} and/or I_rms of the best-refined configuration were smaller than that of the best EMAP-docked configuration except in the case of 2i25. For the "difficult" test system, 2hmi, with an I_C^{α}-rms of 2.26 Å, the Ab binds to half of the site but in a different direction from the 2hmi structure, hence the L_rmsd is very large although f_{nat} is 0.53. Although most of the f_{nat} values in Table 2 decrease after refinement, they are still ~0.5 or greater. If the f_{nat}, L_rms, and I_rms for the largest f_{nat} rather than the lowest I_rms values from the top 2000 + 2000 EMAP configurations before and after refinement were compared, then the average drop in the f_{nat} values was only ~0.05.

Table 2. CAPRI Assessment of the Lowest I_rms Native Configurations from Top 2000 + 2000 EMAP Configurations before and after Refinement

	EMAP configurations[a]			refined configurations[b]		
PDB ID	f_{nat}	L_rms (Å)	I_rms (Å)	f_{nat}	L_rms (Å)	I_rms (Å)
Using Unbound–Unbound Structures						
1ahw	0.75	4.19	1.88	0.67	3.29	1.12
1bgx	0.68	3.42	2.44	0.61	3.23	2.38
1bvk	0.83	3.83	1.99	0.85	2.92	1.59
1dqj	0.75	3.11	1.82	0.48	3.12	1.75
1e6j	0.86	4.95	2.04	0.49	2.05	1.22
1jps	0.80	6.95	2.04	0.69	6.37	1.29
1mlc	0.93	5.85	1.42	**0.68**	**2.64**	**0.96**
1vfb	0.72	4.36	2.16	0.77	2.90	1.53
1wej	0.98	3.83	1.61	**0.91**	**1.37**	**0.65**
2fd6	0.79	6.63	2.25	0.57	4.90	1.76
2i25	0.61	7.48	1.84	0.50	9.40	2.05
Using Bound–Unbound Structures						
1bj1	0.64	8.13	1.95	0.67	7.77	1.34
1fsk	0.89	3.36	1.39	**0.83**	**1.15**	**0.56**
1iqd	0.80	3.50	1.55	0.85	2.20	1.04
1i9r	0.83	5.12	2.06	0.66	4.27	1.56
1k4c	0.79	3.53	1.59	**0.74**	**1.55**	**0.75**
1kxq	0.80	3.32	1.46	**0.86**	**0.43**	**0.69**
1nca	0.95	3.17	1.55	**0.76**	**1.12**	**0.39**
1nsn	0.86	3.42	1.68	**0.76**	**1.09**	**0.50**
1qfw	0.85	4.09	1.81	**0.75**	**2.84**	**0.83**
2hmi	0.50	50.07	5.56	0.53	59.82	5.63
2jel	0.75	3.48	1.59	**0.73**	**1.73**	**0.50**
2qfw	0.92	5.11	2.29	0.73	2.54	1.32

[a]The f_{nat}, L_rms, and I_rms for the lowest I_rms native configurations from the top 2000 + 2000 EMAP configurations; there are no high quality conformations at this stage. [b]The f_{nat}, L_rms, and I_rms for the lowest I_rms native configurations from the top 2000 + 2000 EMAP configurations after refinement using Monte Carlo and minimization. Bold values indicate highly accurate conformations with $f_{nat} \geq 0.5$ and L_rms $\leq$ 1 Å or I_rms $\leq$ 1 Å.

3.3. SVM Performance. The second challenge is to spot the 'correct' solution from the 4000 refined structures containing at least one native conformation for the 23 Ab/Ag complexes. To determine if the SVM-based scoring function could identify native conformations among the 4000 refined configurations for each of the 23 Ab/Ag complexes, the SVM scores of docked configurations were used to rank the configurations such that the highest score corresponds to rank 1. The results in Table 3 show that the SVM classifier derived from the training data set could rank a native conformation among the top six for all the 23 Ab/Ag complexes except 1mlc, where the lowest rank of a native conformation was 36. Notably, it ranked a native conformation as the top one for 5 of the 12 unbound–unbound test cases and half of the 12 bound–unbound test cases.

3.4. Comparison with Previous Works. Weng and co-workers[74] had employed the program ZDOCK 3.0 to dock 11 of the 24 Ab/Ag complexes and used ZRANK[79] to rerank all 54 000 rigid-body configurations without any refinement. ZRANK uses a scoring function that is a linear weighted sum of vdW attractive and repulsive energies, electrostatics short and long-range attractive and repulsive energies computed with a $1/r$ distance-dependent dielectric and desolvation energy.[79] For each Ab/Ag complex, the top-scoring 20 rigid-body configurations from ZDOCK and ZRANK were refined using RosettaDock,[54] which generated 300 conformations per rigid-body model; these 300 refined structures were rescored by ZRANK and the top-scoring conformation was selected. The resulting 20 refined conformations for each Ab/Ag complex were reranked using ZRANK and the lowest rank of a native conformation (defined as a docked structure with I_C^{α}-rms $\leq$ 2.5 Å[74]) was reported and shown in parentheses in Table 3. Comparison of the SVM and ZRANK ranks shows that the SVM classifier generally performs better than ZRANK, considering that the definition of a native conformation used by Weng and co-workers[74] is less stringent than that used herein, where a native conformation is defined as a docked structure of at least medium accuracy with I-rms $\leq$ 2 Å as well as $f_{nat} \geq 0.3$ but <0.5, or $f_{nat} \geq 0.5$ and I-rms is >1 Å. The ZRANK values are larger than the SVM ranks for seven of the 11 Ab/Ag complexes, equal for one Ab/Ag complex (1 fsk), but are smaller than the SVM ranks for three Ab/Ag complexes (1mlc, 2vis, and 1nca). Furthermore, the percentage number of complexes for which a native conformation was ranked as the top one or within the top six when scored according to ZRANK (18% or 55%) is less than that when scored by the SVM (48% or 96%).

DISCUSSION

This work has shown that (i) EMAP could produce native conformations in the top few thousand configurations, which can subsequently be refined and (ii) the SVM could rank native conformations mostly among the top six. Notably, EMAP could produce a configuration quite close to the native one with I_C^{α}-rms $\leq$ 3.5 Å for the most difficult case 2hmi, which

Table 3. SVM Docking Results for the Ab/Ag Complexes Using the Respective Unbound Structures

Ab/Ag[a]	Ab	Ag	lowest rank[a]	f_{nat}[b]	L_rms[b] (Å)	I_rms[b] (Å)	I_C$^{\alpha}$-rms[b] (Å)	CAPRI quality[d]
Using Unbound−Unbound Structures								
1ahw	1fgn	1tfh	2	0.60	6.83	1.52	1.53	**
1bgx	1ay1	1cmw	1	0.53	6.35	3.09	3.09	**
1bvk	1bvl	3lzt	6	0.58	2.27	1.65	1.65	**
1dqj	1dqq	3lzt	4	0.46	4.83	2.33	2.31	**
1e6j	1e6o	1a43	2 (8)	0.59	3.58	1.35	1.35	**
1jps	1jpt	1tfh	2 (5)	0.69	6.37	1.29	1.27	**
1mlc	1mlb	3lzt	36 (9)	0.66	4.79	1.29	1.29	**
1vfb	1vfa	8lyz	1	0.68	3.03	1.74	1.78	**
1wej	1qbl	1hrc	1 (4)	0.56	11.05	2.23	2.21	**
2fd6	2fat	1ywh	1	0.64	13.61	3.78	3.79	**
2i25	2i24	3lzt	1	0.50	9.57	2.07	2.08	**
2vis	1gig	2viu	−[c](15)				−[c]	
Using Bound−Unbound Structures								
1bj1	1bj1	2vpf	4 (16)	0.66	7.81	1.46	1.42	**
1fsk	1fsk	1bv1	1 (1)	0.52	7.76	2.67	2.69	**
1iqd	1iqd	1d7p	1 (18)	0.71	10.84	2.79	2.77	**
1i9r	1i9r	1aly	6	0.52	16.16	3.37	3.36	**
1k4c	**1k4c**	**1jvm**	**1**	**0.74**	**1.55**	**0.75**	**0.76**	***
1kxq	1kxq	1ppi	3 (6)	0.80	1.77	1.15	1.12	**
1nca	**1nca**	**7nn9**	**3 (1)**	**0.89**	**1.42**	**0.66**	**0.66**	***
1nsn	1nsn	1kdc	1	0.63	5.16	2.83	2.80	**
1qfw	1qfw	1hrp	1	0.72	7.88	2.32	2.32	**
2hmi	2hmi	1s6p	1	0.53	59.82	5.63	5.57	**
2jel	2jel	1poh	3	0.46	3.35	1.72	1.70	**
2qfw	2qfw	1hrp	2 (5)	0.50	30.18	5.09	5.10	**

[a]Lowest SVM rank of a native conformation among the top 2000 + 2000 refined structures; numbers in parentheses are taken from Weng and co-workers[74] who used ZDOCK, ZRANK, and RosettaDock to rank 11 of the 24 Ab/Ag complexes. [b]Values correspond to the lowest SVM-ranked native conformation among the top 2000 + 2000 refined structures. The conformations are of medium accuracy with (i) f_{nat} is ≥0.5 and L_rms or I_rms > 1 Å, or (ii) f_{nat} is <0.5 but ≥0.3 and L_rms ≤ 5 Å or I_rms ≤2 Å, except those appearing in bold, which are of high accuracy with f_{nat} is ≥0.5 and I_rms ≤ 1 Å. [c]A dash means no native conformations were found among the top 2000 + 2000 EMAP configurations (see Table 1). [d]*** indicates a high quality model according to the CAPRI metrics, while ** indicates a medium quality one.

exhibits the largest I_C$^{\alpha}$-rms between the bound and the unbound proteins among the 24 test systems. The success of the SVM scoring function implies that the vibrational and conformational entropy changes upon binding of an Ab to an Ag, which were not used as input features to train the SVM, do not generally play a key role in discriminating between native and non-native conformations of Ab/Ag complexes.

The key novelty of our approach lies in using the components of the free energy change upon binding of an Ab to an Ag according to Scheme 1 to train a SVM. In contrast, a previously trained SVM for Ab/Ag complexes had employed a different set of input features, which include specialized scoring functions, knowledge-based pairwise potentials at the atom and residue level, gap volume, measures of tightness of fit, evolutionary conservation, residue-interface propensities, and temperature factors.[65] Hence, the SVM-scoring function herein differs from previous scoring functions (see Introduction). Notably, it differs from the ZRANK[79] scoring function in two ways: First, it includes gas-phase electrostatic energies using a constant dielectric equal to one, whereas ZRANK computes electrostatic energies using a $1/r$ distance-dependent dielectric. Second, it includes the solvation free energy change, but ZRANK estimates the desolvation energy by pairwise atomic contact energy. The binding free energy components in Scheme 1 (which could include vibrational and conformational entropy changes) and the procedure used to train and calibrate a SVM can be generalized to other classes of protein complexes.

The integration of EMAP into the CHARMM molecular mechanics package has several advantages for protein-macromolecule docking. EMAP can handle large proteins by reducing specific parts to MAP objects for initial rigid-body docking, but atomic coordinates can easily be regenerated from the MAP positions for rescoring the best configurations using atom−atom interactions, including interactions from outside the predicted interface region. The ability to include the noninterface regions is important in docking large proteins because the docked region under consideration is usually kept small for efficiency, but when the entire protein is generated using the predicted interface configuration, the noninterface regions may have large unfavorable steric interactions, showing the configuration to be a false positive. Since molecular dynamics, MC, and continuum dielectric methods are incorporated in CHARMM, it is possible to freeze the predicted noninterface areas of two docked proteins and to perform local MD or MC on the interface residues to repack the interface without resorting to additional external programs and force fields. Further analysis can be extended to free-energy decompositions to derive the relative contributions of the residues toward binding using an all-atom force field.

AUTHOR INFORMATION

Corresponding Author

*Fax: 886-2-2788-7641. E-mail: carmay@gate.sinica.edu.tw.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by research grants from the National Science Council and Academia Sinica, Taiwan. Software development efforts by X.W. and B.R.B. were supported by the National Institutes of Health (NIH) intramural research program (Z01 HL001027-30).

■ REFERENCES

(1) Rivas, J. D. L.; Fontanillo, C. *PLoS Comput. Biol.* **2009**, *6*, e1000807.
(2) Nooren, I. M. A.; Thornton, J. M. *EMBO J.* **2003**, *22*, 3486−3492.
(3) Chen, Y. C.; Lim, C. *Nucleic Acids Res.* **2008**, *36*, 7078−7087.
(4) Arkin, M. R.; Wells, J. A. *Nat. Rev. Drug Discov.* **2004**, *3*, 301−317.
(5) González-Ruiz, D.; Gohlke, H. *Curr. Med. Chem.* **2006**, *13*, 2607−2625.
(6) Wu, X.-W.; Milne, J. L. S.; Borgnia, M. J.; Rostapshov, A. V.; Subramaniam, S.; Brooks, B. R. *J. Struct. Biol.* **2003**, *141*, 63−69.
(7) Wu, X.-W.; Brooks, B. R.Modeling of macromolecular assemblies with map objects. *Int. Conf. Bioinf. Comput. Biol., BIOCOMP 2007*, Las Vegas, NV, 2007.
(8) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(9) Cortes, C.; Vapnik, V. *Mach. Learn.* **1995**, *20*, 273−293.
(10) Cherfils, J.; Janin, J. *Current Opinion Struct. Biol.* **1993**, *3*, 265−269.
(11) Lorber, D. M.; Shoichet, B. K. *Protein Sci.* **1998**, *7*, 938−950.
(12) Sternberg, M. J. E.; Gabb, H. A.; Jackson, R. M. *Curr. Opin. Struct. Biol.* **1998**, *8*, 250−256.
(13) Elcock, A. H.; Sept, D.; J.A., M. *J. Phys. Chem. B* **2001**, *105*, 1504−1518.
(14) Camacho, C. J.; Vajda, S. *Curr. Opin. Struct. Biol.* **2002**, *12*, 36−40.
(15) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Struct. Funct. Bioinf.* **2002**, *47*, 409−443.
(16) Smith, G. R.; Sternberg, M. J. *Curr. Opin. Struct. Biol.* **2002**, *12*, 28−35.
(17) Eisenstein, M.; Katchalski-Katzir, E. *C. R. Biol.* **2004**, *327*, 409−420.
(18) Deremble, C.; Lavery, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 171−175.
(19) van Dijk, A. D.; Boelens, R.; Bonvin, A. M. *FEBS J.* **2005**, *272*, 293−312.
(20) Bonvin, A. M. *Curr. Opin. Struct. Biol.* **2006**, *16*, 194−200.
(21) Gray, J. J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 183−193.
(22) Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H. J. *Proteins: Struct. Funct. Bioinf.* **2008**, *73*, 271−289.
(23) Ritchie, D. W. *Curr. Prot. Peptide Sci.* **2008**, *9*, 1−15.
(24) Vajda, S.; Kozakov, D. *Curr. Opin. Struct. Biol.* **2009**, *19*, 164−170.
(25) Janin, J. *Mol. Biosyst.* **2010**, *6*, 2351−2362.
(26) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. *J. Comput. Chem.* **2010**, *31*, 317−342.
(27) Zacharias, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 180−186.
(28) Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J. E.; Vajda, S.; Vakser, I.; Wodak, S. J. *Proteins: Struct. Funct. Genet.* **2003**, *52*, 2−9.
(29) Mendez, R.; Leplae, R.; De Maria, L.; Wodak, S. J. *Proteins* **2003**, *52*, 51−67.
(30) Vajda, S.; Camacho, C. J. *Trends Biotechnol.* **2004**, *22*, 110−116.
(31) Wodak, S. J.; Mendez, R. *Curr. Opin. Struct. Biol.* **2004**, *14*, 242−249.
(32) Mendez, R.; Leplae, R.; Lensink, M. F.; Wodak, S. J. *Proteins* **2005**, *60*, 150−169.
(33) Janin, J. *Protein Sci.* **2005**, *14*, 278−283.
(34) Vajda, S. *Proteins* **2005**, *60*, 176−180.
(35) Pons, C.; Grosdidier, S.; Solernou, A.; Perez-Cano, L.; Fernandez-Recio, J. *Proteins* **2010**, *78*, 95−108.
(36) Vakser, I. A.; Aflalo, C. *Proteins: Struct., Funct. Genet.* **1994**, *20*, 320−329.
(37) Vakser, I. A. *Protein Eng.* **1995**, *8*, 371−377.
(38) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. E. *J. Mol. Biol.* **1997**, *272*, 106−120.
(39) Ritchie, D. W.; Kemp, G. J. *Proteins* **2000**, *39*, 178−194.
(40) Mandell, J. G.; Roberts, V. A.; Pique, M. E.; Kotlovyi, V.; Mitchell, J. C.; Nelson, E.; Tsigelny, I.; Ten Eyck, L. F. *Protein Eng.* **2001**, *14*, 105−113.
(41) Del Carpio-Munoz, C. A.; Ichiishi, E.; Yoshimori, A.; Yoshikawa, T. *Proteins: Struct., Funct. Genet.* **2002**, *48*, 696−732.
(42) Chen, R.; Li, L.; Weng, Z. *Proteins: Struct., Funct. Genet.* **2003**, *52*, 80−87.
(43) Carter, P.; Lesk, V. I.; Islam, S. A.; Sternberg, M. J. E. *Proteins: Struct., Funct. Bioinf.* **2005**, *60*, 281−288.
(44) Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. *Proteins: Struct., Funct. Bioinf.* **2006**, *65*, 392−406.
(45) Fischer, D.; Liang-Lin, S.; Wolfson, H. L.; Nussinov, R. *J. Mol. Biol.* **1995**, *248*, 459−477.
(46) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195−2199.
(47) Norel, R.; Fischer, D.; Wolfson, H. J.; Nussinov, R. *Prot. Eng.* **1994**, *7*, 39−46.
(48) Schneidman-Duhovny, D.; Inbar, Y.; Polak, V.; Shatsky, M.; Halperin, I.; Benyamini, H.; Barzilai, A.; Dror, O.; Haspe, l. N.; Nussinov, R.; Wolfson, H. J. *Proteins* **2003**, *52*, 107−112.
(49) Abagyan, R.; Totrov, M. *J. Mol. Biol.* **1994**, *235*, 983−1002.
(50) Zhao, Y.; Sanner, M. F. *Proteins* **2007**, *68*, 726−737.
(51) Dominguez, C.; Boelens, R.; Bonvin, A. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.
(52) Camacho, C. J.; Gatchell, D. W. *Proteins: Struct., Funct. Genet.* **2003**, *52*, 92−97.
(53) Rajamani, D.; Thiel, S.; Vajda, S.; Camacho, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11287−11292.
(54) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. *J. Mol. Biol.* **2003**, *331*, 281−299.
(55) Schueler-Furman, O.; Wang, C.; Baker, D. *Proteins* **2005**, *60*, 187−194.
(56) Fernandez-Recio, J.; Abagyan, R.; Totrov, M. *Proteins: Struct., Funct. Bioinf.* **2007**, *52*, 113−117.
(57) Wang, C.; Bradley, P.; Baker, D. *J. Mol. Biol.* **2007**, *373*, 503−519.
(58) Zacharias, M. *Protein Sci.* **2003**, *12*, 1271−1282.
(59) Koehl, P.; Delarue, M. *J. Mol. Biol.* **1994**, *239*, 249−275.
(60) Bastard, K.; Thureau, A.; Lavery, R.; Prevost, C. *J. Comput. Chem.* **2003**, *24*, 1910−1920.
(61) Zacharias, M. *Proteins: Struct., Funct. Bioinf.* **2004**, *54*, 759−767.
(62) Cavasotto, C. N.; Kovacs, J. V. A., R. *J. Am. Chem. Soc.* **2005**, *127*, 9632−9640.
(63) May, A.; Zacharias, M. *Proteins* **2008**, *70*, 794−809.
(64) Ponomarenko, J. V.; Bourne, P. E. *BMC Struct. Biol.* **2007**, *7*, 64−82.
(65) Martin, O.; Schomburg, D. *Proteins* **2008**, *70*, 1367−1378.
(66) London, N.; Schueler-Furman, O. *Structure* **2008**, *16*, 269−279.
(67) Hwang, H.; Pierce, B.; Mintseris, J.; Janin, J.; Weng, Z. *Proteins: Struct., Funct. Bioinf.* **2008**, *73*, 705−709.
(68) Nilsson, L.; Karplus, M. *J. Comput. Chem.* **1986**, *7*, 591−616.
(69) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(70) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Rchardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735−1747.
(71) Janeway, C. A.; Travers, P.; Walport, M. *Immunobiology: The Immune System in Health and Disease*, 4th ed.; Elsevier Science Ltd.: London, 1999.

(72) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(73) Dessailly, B. H.; Lensink, M. F.; Wodak, S. J. *BMC Bioinf.* **2007**, *8*, 141.
(74) Pierce, B.; Weng, Z. *Proteins: Struct., Funct. Bioinf.* **2008**, *72*, 270−279.
(75) Noskov, S.; Lim, C. *Biophys. J.* **2001**, *81*, 737−750.
(76) Haberthür, U.; Caflisch, A. *J. Comput. Chem.* **2008**, *29*, 701−715.
(77) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.
(78) Chang, C.-C.; Lin, C.-J. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27:21−27:27.
(79) Pierce, B.; Weng, Z. *Proteins: Struct., Funct. Bioinf.* **2007**, *67*, 1078−1086.