Scoring by Intermolecular Pairwise Propensities of Exposed Residues (SIPPER): A New Efficient Potential for Protein−Protein Docking

Carles Pons,[†,‡] David Talavera,[§] Xavier de la Cruz,[∥] Modesto Orozco,[‡,○,⊥] and Juan Fernandez-Recio*[,†]

[†]Life Sciences Department, Barcelona Supercomputing Center and [‡]Computational Bioinformatics, National Institute of Bioinformatics, Barcelona 08034, Spain

[§]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, U.K.

[∥]CSIC-IBMB, Barcelona 08028, Spain, and ICREA, Barcelona 08018, Spain

[○]Joint Research Program in Computational Biology, Institute of Research in Biomedicine and Barcelona Supercomputing Center and [⊥]Departament de Bioquímica, Universitat de Barcelona, Barcelona 08028, Spain

S *Supporting Information*

ABSTRACT: A detailed and complete structural knowledge of the interactome is one of the grand challenges in Biology, and a variety of computational docking approaches have been developed to complement experimental efforts and help in the characterization of protein−protein interactions. Among the different docking scoring methods, those based on physico-chemical considerations can give the maximum accuracy at the atomic level, but they are usually computationally demanding and necessarily noisy when implemented in rigid-body approaches. Coarser-grained knowledge-based potentials are less sensitive to details of atomic arrangements, thus providing an efficient alternative for scoring of rigid-body docking poses. In this study, we have extracted new statistical potentials from intermolecular pairs of exposed residues in known complex structures, which were then used to score protein−protein docking poses. The new method, called SIPPER (scoring by intermolecular pairwise propensities of exposed residues), combines the value of residue desolvation based on solvent-exposed area with the propensity-based contribution of intermolecular residue pairs. This new scoring function found a near-native orientation within the top 10 predictions in nearly one-third of the cases of a standard docking benchmark and proved to be also useful as a filtering step, drastically reducing the number of docking candidates needed by energy-based methods like pyDock.
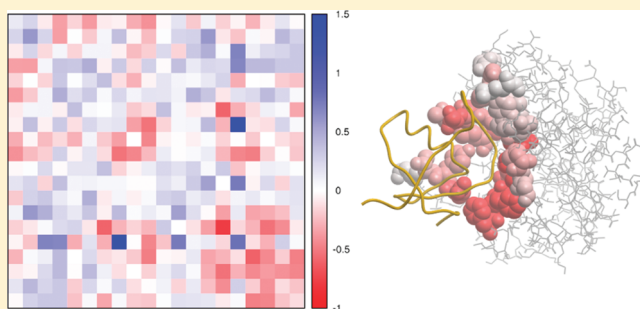
■ INTRODUCTION

The majority of protein functions are mediated by their interactions with other molecules in the cell, especially with other proteins. Thus, the knowledge of the structure and properties of protein−protein complexes is essential in order to understand how cells act and to open new applications in the drug discovery field.[1] Technical difficulties to experimentally solve the structures of protein complexes,[2] together with current high performance computational capabilities, have propelled the development of new protein−protein docking tools that aim to predict by computational means the atomic details of a protein complex starting from its individual components.

The docking problem is usually addressed by an initial sampling of the configurational space of the interacting proteins (to generate potential docking poses), followed by the scoring and/or refinement of the previously generated docking candidates (to determine which of them is the correct one). The different approaches for sampling, scoring, and refinement can take into account geometrical features (like shape complementarity), physicochemical properties, residue conservation, statistical information, and/or experimental information[3] (see the recent

CAPRI experiments at http://www.ebi.ac.uk/msd-srv/capri/capri.html for an updated overview of the field). Sampling is typically performed by using rigid-body strategies, which have significantly improved with the use of Fast Fourier Transform (FFT) algorithms.[4−8] In fact, while current computers and algorithms can achieve a quite complete sampling of the protein−protein interaction space, we do not have yet efficient methods for scoring and refinement, which can often hinder us from the identification of the near-native conformations within the docking set.

Ideally, scoring should provide an accurate description of the docking free energy, but these calculations are usually computationally expensive and prone to error when dealing with the expected inaccuracies in the interprotein contacts caused by the rigid-body docking approach. This justifies the use of statistical potentials derived from the interface pairwise propensities found in structural databases,[7,9−14] thus adapting an approach that was very successful for structural prediction in monomeric

proteins.[15] Some of the reported statistical potentials for protein—protein docking scoring have been derived from structural databases that included oligomeric complexes (mostly homodimers)[11,12] and might not accurately reflect the particularities of heterodimeric protein—protein interactions.[16] More recent approaches used atom-based statistical pair potentials derived from heterodimeric complexes, and they indeed gave better predictive success rates when included in the FFT approach in combination with other scoring terms.[7] However, current limitations of this approach are (i) the high computational cost when explicitly applied to score individual docking poses, due to the atom-level representation and the necessity of a number of scoring terms, and (ii) the higher noise of the atom-based scoring of rigid-body geometries. Consequently, publicly available propensities derived at the residue level from heterodimeric complexes that can be efficiently applied as scoring functions are still needed.

The first goal of the present work is thus to derive new residue—residue interface propensities for the optimal scoring of docking poses. In addition, we want to bridge the gap here between the statistical potentials and other physical-based energy descriptors, not only to improve the predictive results but also to obtain a better understanding of the mechanism of association. Among the different terms involved in docking, and that need to be accordingly represented in the scoring function, desolvation plays in general the most important role for protein—protein binding and has the advantage that is the one that is less dependent on atomic details. For instance, a simple ASA-based solvation description with atomic parameters (ASP) optimized for protein docking[17] can be used to find protein surface regions that correlate well with protein interaction sites, using the ODA method.[18−21] This desolvation descriptor can efficiently score docking poses, and indeed, it is used by some of the most successful methods, like ICM,[22] Q-SiteFinder interface cleft analysis,[23] pyDock,[24] HADDOCK,[25] or FRODOCK.[8]

In this work, we describe new interface potentials for docking, defined at the residue level from a large, well-characterized, and manually curated data set of heterodimeric complexes, thus overcoming current limitations of other available propensity sets. In addition, the uniqueness of our approach is that we complement these new statistical potentials with our previously successful energy-based scoring protocols, adapted here for their use at the residue level and therefore at very low computational cost.[18] The new method SIPPER efficiently integrates knowledge-based potentials and physical energy at the residue level in a successful scoring function that can be used to identify a near-native solution within the top 10 docking poses in over 28% of the cases of a standard docking benchmark.

■ MATERIALS AND METHODS

Statistical Potentials from Interface Pair Propensities. Interface residue—residue statistical potentials were computed from a previously compiled data set of 70 heterodimeric complexes, containing 31 enzyme—inhibitors, 18 antibody—antigens, and 21 "other" protein—protein complexes.[26] Residue pair propensities were determined for each of the 210 residue pair types (combinations with repetition formed by the 20 amino acids, i.e. without considering the order of the residues) from the observed relative frequency of each residue pair type at protein—protein interfaces with respect to its expected relative

frequency given the residue composition in a random distribution (eq 1).

$$P_{kl} = f_{\text{obs}}/f_{\text{exp}} \qquad (1)$$

The observed relative interface frequency f_{obs} for a given residue pair of type kl was calculated as in eq 2:

$$f_{\text{obs}} = n_{kl}^{\text{I}}/N^{\text{I}} \qquad (2)$$

where n_{kl}^{I} is the number of solvent-exposed residues of type k from one protein that have any of their non-hydrogen atoms closer than a given distance cutoff to any of the non-hydrogen atoms from a solvent-exposed residue of type l of the other protein and N^{I} is the total number of interprotein residue pairs between solvent-exposed residues of all types. Four different distance cutoffs have been used to obtain the interprotein contacts: 4.0, 4.5, 5.0, and 6.0 Å.

The expected relative interface frequency f_{exp} for a residue pair of type kl was computed as the probability of forming such pair by chance given the residue composition in a random model (eq 3):

$$f_{\text{exp}} = n_k^0 n_l^0/(N^0)^2 \qquad (3)$$

where n_k^0, n_l^0, and N^0 are the number of residues of type k, the number of residues of type l, and the total number of residues, respectively, in a random model. We used here two different random models, according to (i) the residue composition of the protein surfaces or (ii) that of the protein interfaces. In the protein surface random model, we considered the residues of the separated partners with relative ASA ≥ 5% (with respect to that of the given residue type when it is exposed). In the interface random model, we reduced that selection to the interface residues, defined as those ones for which their ASA decreased upon complex formation.

Finally, propensities were converted to free energy estimates or statistical potentials at room temperature:

$$\Delta G_{kl}^{\text{stat}} = -RT \ln(P_{kl}) \qquad (4)$$

Scoring of Docking Poses by Pairwise Statistical Potentials. We used the pairwise statistical potentials to score docking poses generated by FTDock, a very popular method that performs rigid body docking using FFT approaches.[5] FTDock was run with no electrostatics and 0.7 Å grid resolution, keeping 10 000 docking poses per case. For each docking pose D, all interprotein contacting residue pairs were computed. A contacting residue pair was defined as that between two residues that have at least one heavy atom-heavy atom contact below the distance threshold used to define the propensities. Increasing the minimal number of interatomic contacts to define a residue—residue pair did not significantly change the global scoring results, probably because of the use of rigid-body proteins, in which the interacting side-chain conformations might not be optimal. For each interprotein residue pair ij, the propensity-based statistical potential value corresponding to that residue pair type kl was linearly added to the final scoring:

$$\Delta G_D^{\text{stat}} = \sum_{ij} \Delta G_{kl}^{\text{stat}} \qquad (5)$$

Optimal Docking Areas (ODA) Calculations. ODA (optimal docking area) screens the protein surface in search of areas with low desolvation energy, as previously described.[18]

For each initial point for the ODA calculations, defined as the geometric center of each residue side-chain, patches of increasing size were created and their desolvation energy calculated, using the same atomic solvation parameters as in pyDock.[24] The lowest energy among the patches was then assigned to the residue from which they were created.

In the present study, we calculated the ODA values of the unbound subunits and we used them to score docking poses as follows. For a given docking pose D, the precomputed ODA values of all interface residues i, j (defined as those that had at least a non-hydrogen atom within a distance threshold from any non-hydrogen atom of the partner molecule), from receptor and ligand respectively, were added up to form the final ODA score, as in eq 6:

$$E_D^{ODA} = \sum_i ODA_i + \sum_j ODA_j \qquad (6)$$

SIPPER. SIPPER (Scoring by Intermolecular Pairwise Propensities of Exposed Residues) is a new fast scoring function that uses statistical information and desolvation energy to rank docking poses. For a given docking pose D, the final scoring is formed by the scoring based on the pairwise propensities and that of the ODA values (both described in previous subsections), with appropriate weighting factor w as in eq 7.

$$E_D^{SIPPER} = \Delta G_D^{stat} + w E_D^{ODA} \qquad (7)$$

The weighting factor w was optimized by minimizing the function $F(w)$ (eq 8) on a proper benchmark set (described later in this Methods section):

$$F(w) = \sum_m \ln(rank_m^w) \qquad (8)$$

where $rank_m^w$ is the best rank of a near-native solution (see below) for the docking case m, once the docking set is reevaluated with a given weight w for the ODA scoring. Weighting values for w ranging from 0.0 to 2.0 were systematically explored (with step 0.001) to determine the lowest value for $F(w)$.

pyDock. pyDock[24] evaluates the binding energy of rigid-body docking orientations based on Coulombic electrostatics with distance-dependent dielectric constant, ASA-based desolvation with atomic solvation parameters optimized for rigid-body docking, and 6–12 Lennard-Jones van der Waals potential weighted by 0.1. Both van der Waals and electrostatics use atomic parameters from the AMBER 94 force field. To account for the clashes in rigid-body docking poses, van der Waals is truncated to +1 kcal/mol (to allow some overlap between the structures) and electrostatics to ±1 kcal/mol (to avoid artificial high scores from overlap structures). Given pyDock success in scoring rigid-body docking poses, as it has been proven for several CAPRI targets,[27,28] it was used here as the "gold standard" in evaluating the performance of SIPPER.

Benchmark Set and Evaluation of Success Rates. We used the unbound structures of the protein–protein docking benchmark 3.0[29] as a test set for SIPPER. The benchmark includes enzyme–inhibitor, antibody–antigen, and "other" types of complexes, with different difficulty levels depending on their flexibility and size. In order to clearly differentiate the test set from the data set used to derive the propensities, from the 124 cases of the benchmark, 23 were discarded because both ligand and receptor had more than 70% sequence identity with the cases

used to generate the propensities matrices. Rigid-body docking poses were generated for the 101 remaining cases with the FFT-based program FTDock.[5]

We considered a docking pose as a hit (or a near-native solution) if its C-alpha ligand rmsd with respect to the experimental reference was below 10 Å, after superimposing both receptors (which would be defined as acceptable by CAPRI criteria as long as it had fraction of native contacts above 0.1). In practice, our definition is arguably stricter than that in CAPRI, since typically very few models with ligand rmsd below 10 Å are defined as incorrect based on the fraction of native contacts, while a much larger number of models with ligand rmsd above 10 Å are still acceptable by CAPRI if they have interface rmsd below 4 Å.[30] The success rate was calculated as the percentage of cases of the benchmark for which at least a hit was found within the top 10 predictions (which is a manageable number of models in a realistic situation and also the number of predictions evaluated by CAPRI).[30] Success rates were calculated only for the 81 cases for which FTDock generated at least one hit within the set of 10 000 docking poses. The remaining cases represent especially complex situations, where sampling errors preclude successful scorings.

Other Propensity Sets. For comparison purposes, we have used other available sets of propensity parameters extracted from crystallographic structures to score our docking poses. *Glaser* propensities[11] were generated with a set of 621 complexes, from which around 66% were homodimers. They defined the residue pair contacts using a distance threshold of 6 Å between $C\beta$ atoms, with a random model based on the interface residues. *Skolnick* propensities,[12] were trained on a set of 340 complexes that included 271 homodimers, using a distance cutoff of 4.5 Å between heavy-atoms to obtain the interprotein contacts. Their random model took into account all protein surface residues. The *RPScore* set was generated with 90 interfaces using a 4.5 Å cutoff between any atom from different residues to get the observed interprotein contacts.[10,31]

■ RESULTS AND DISCUSSION

New Interface Residue–Residue Propensities. We have generated 8 different sets of propensity-based statistical potentials for the 210 residue pairs (applying 4 different distance cutoffs and 2 random models, one based on the residue composition of protein surfaces and the other on that of protein–protein interfaces; see Materials and Methods). For a 5 Å distance cutoff and interface random model (*Interface 5 Å*; Figure 1A), the pairs with more favorable statistical energy were Met-Met, Phe-Phe, and Met-His, while the pairs with more unfavorable statistical energy were Cys-Pro, Cys-Cys, and Cys-Thr. For a 5 Å distance cutoff and surface random model (*Surface 5 Å*; Figure 1B), the most favored pairs were Met-Met, Met-His, and Cys-Trp, whereas the most disfavored were Pro-Cys, Lys-Ala, and Thr-Ala. See detailed values of potentials, observed occurrence of pairs, and statistical errors in the Supporting Information. The associated statistical errors are not large, and actually, only in a few cases is the error above 0.2: Lys-Lys, Pro-Cys, Met-Met, and Cys-Cys for the Interface 5 Å random model and Lys-Lys, Pro-Cys, Met-Met, and Met-Asp for the Surface 5 Å random model. These errors can be explained because some of these pairs had very low counts (Cys-Cys, Pro-Cys, Lys-Lys) and the other pairs (Met-Met, Met-Asp) occurred in only one complex. An interesting case for discussion is Arg-Arg, which has negative statistical
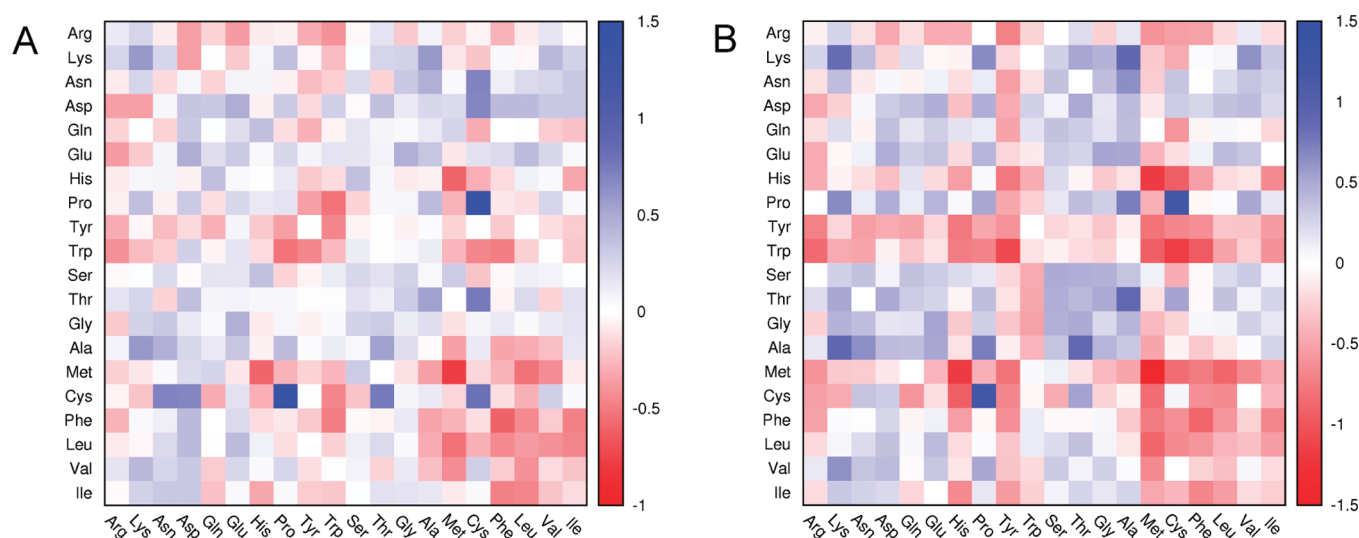
Figure 1. Propensity-based pairwise statistical potentials obtained in this work using a 5 Å distance cutoff to define the intermolecular residue−residue pairs, with two different random models based on interface composition (A) or on surface composition (B). See detailed values in the Supporting Information online.

Table 1. Correlation between Different Interface Residue−Residue Propensity Sets

	INT4	INT4.5	INT5	INT6	SUR4	SUR4.5	SUR5	SUR6
INT4	1	0.947	0.914	0.869	0.877	0.836	0.804	0.768
INT4.5	0.947	1	0.976	0.926	0.821	0.850	0.821	0.778
INT5	0.914	0.976	1	0.945	0.797	0.833	0.835	0.788
INT6	0.869	0.926	0.945	1	0.738	0.769	0.769	0.787
SUR4	0.877	0.821	0.797	0.738	1	0.976	0.962	0.946
SUR4.5	0.836	0.850	0.833	0.769	0.976	1	0.990	0.970
SUR5	0.804	0.821	0.835	0.769	0.962	0.990	1	0.979
SUR6	0.768	0.778	0.788	0.787	0.946	0.970	0.979	1
RPScore	0.681	0.677	0.659	0.615	0.765	0.771	0.760	0.745
Skolnick	0.487	0.504	0.495	0.407	0.693	0.723	0.723	0.694
Glaser	0.473	0.493	0.475	0.407	0.554	0.575	0.564	0.532

potential with the Surface 5 Å random model, while it has positive value with the Interface 5 Å random model. The occurrence of this pair is not insignificant and the propensity value seems to strongly depend on the reference model used to calculate the expected frequencies. The differences are because this pair is found less often than expected according to the residue composition of interfaces but slightly more often than expected according to the residue composition of protein surfaces. Indeed, Arg is a frequent residue in protein−protein interfaces, and its capabilities to establish aliphatic interactions and to act with either hydrophobic or polar character have already been discussed.[11]

Table 1 shows the similarities between the generated propensity sets. As can be seen, the random model (interface, INT; surface, SUR) introduces more variability in the propensity sets than the distance cutoffs used to select the interprotein residue pairs. We have also compared our propensities with other sets reported in the literature. The best agreements (Table 1) are found with RPScore (Pearson correlations around 0.7), while the largest discrepancies are found with Glaser propensities (Pearson correlations typically 0.4−0.5). In any case, our propensities generated with surface composition as random model were more similar to the other propensities reported in the literature than

those obtained with an interface-adapted random model. All of the above is just probably reflecting the different composition of the databases used to derive each propensity set (Skolnick and Glaser ones are enriched in homodimers and RPScore has a larger proportion of heterodimers, while ours is formed exclusively by heterodimers).

Scoring of Docking Poses by Propensity-Based Statistical Potentials. In order to evaluate the use of the new propensities in docking, we evaluated the scoring results on a test set formed by 101 unbound−unbound docking cases, different from those in the training set used to derive the propensities (see Materials and Methods). For each case, we generated docking sets using FTDock as described in Materials and Methods, using previously optimized parameters to maximize the number of cases with at least one near-native solution.[32] In these conditions, FTDock found at least one near-native solution for 81 cases of our benchmark (26 enzyme/inhibitor, 13 antibody/antigen, 42 other complexes). The 20 cases for which FTDock did not find any near-native solution (indicating severe sampling problems) were discarded for the rest of this study.

We scored these unbound docking sets with the statistical potentials derived for each one of the eight different propensity sets above-described (eq 5). Figure 2A shows the success rates for each propensity set, for the top 10, 100, and 1000 predictions, clearly improving the FTDock default scoring in all cases. As a general trend, scoring with the interface random model propensities was more successful than with the surface random model propensities for the top 10 and 100 predictions, with the best results obtained at a 5 Å distance cutoff (success rates were 23.5% and 45.7% for the top 10 and 100 predictions, respectively). This better enrichment in near-native orientations at lower ranks indicated that correcting the interface propensities by the residue interface composition (as compared with residue surface composition) gives a more accurate description of protein−protein binding preferences.

For comparison purposes, we also scored the docking sets with other propensity sets available in the literature, using identical procedure and the threshold distances used for their generation (Skolnick 4.5 Å; Glaser 6 Å; RPScore 4.5 Å). Scoring with our
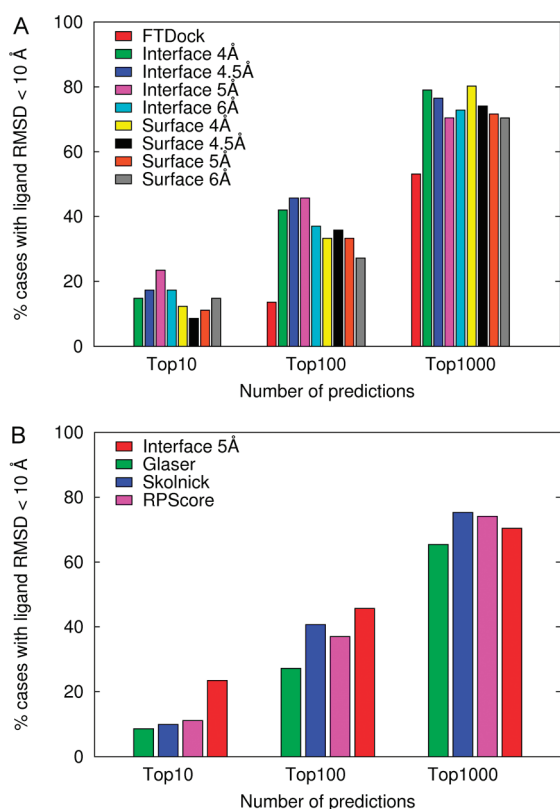
Figure 3. Success rates of SIPPER scoring function, as compared to the scoring with statistical potentials alone (Interface 5 Å), ODA values, and pyDock. The expected success rates from a random distribution have been computed by randomly shuffling the ranking values of the docking results and recalculating the success rates. This was repeated 100 times, and the averaged obtained success rates are shown.

Figure 2. (A) Success rates (i.e., percentage of cases with at least a near-native solution within the top *N* solutions) of the propensity-based statistical potentials obtained in this work on the FTDock-generated docking sets. For comparison, the success rate of FTDock original scoring function is shown. (B) Success rates of our interface random model propensities compared to those of other propensity sets available in literature (Glaser, Skolnick, and RPScore; see Materials and Methods).

Interface 5 Å propensities yielded the best success rates for the top 10 and 100 predictions (see Figure 2B). For the top 1000 predictions, the Skolnick and the RPScore propensities gave better success rates, but not as good as those from our Surface 4 Å set (see Figure 2A). The worse success rates of Glaser propensities are probably due to the fact that their propensities were derived from data sets enriched in homodimers and considering only the $C\beta$ atoms.

Scoring of Docking Poses by Interface Residue ODA Values. We scored the docking sets using the ODA values previously calculated on the unbound molecules, using an optimal 5 Å distance cutoff to select the interface residues contributing to the scoring (eq 6). Success rates reached 17.3% and 35.8% for the top 10 and top 100 predictions, respectively (see Figure 3). These predictive values were slightly worse than those by pyDock (pyDock energy includes electrostatics and van der Waals, in addition to desolvation), but a major advantage is the much lower computational cost (ODA-based scoring is defined at residue level and ASA is calculated only once, while pyDock scoring is atom-based for the two extra terms and ASA is computed for each docking pose).

SIPPER: Combining Statistical Potentials and ODA. Success rates obtained using ODA values for scoring were well below those obtained when using the Interface 5 Å propensities (Figure 3). Interestingly, both scoring functions showed low correlation ($r = 0.166$) and thus were likely to represent different
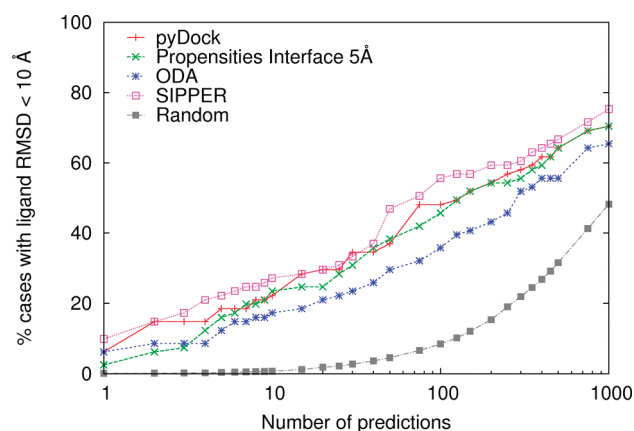
contributions to binding. Therefore, we studied the predictive value of the combination of the statistical energies and the ODA scoring, implemented in a new function called SIPPER (eq 7).

Given the different range of values of the two scoring functions, a scaling factor was defined in order to optimally combine them. The value for this weighting factor *w* was set as the one minimizing the sum of best ranks (in logarithmic units) on the 81 cases of our benchmark for which FTDock generated at least a near-native solution (eq 8). We explored values for *w* ranging from 0.0 to 2.0, with step of 0.001, obtaining the optimal weight $w = 0.019$. With this *w* value for SIPPER, for the top 1, 10, and 100 predictions the success rates were 9.9%, 27.2%, and 55.6%, respectively, a noteworthy improvement with respect to the statistical energies and ODA scorings alone, and outperforming our standard pyDock scoring function as well (see Figure 3). To discard any possibility of overfitting, we performed a leave-one-out cross-validation with our data set that resulted in the same value for *w* in all but four cases (4.9% of our benchmark). Moreover, this did not change the number of cases in which a near-native solution was found among the top 10 predictions. As described in Materials and Methods, the benchmark set was formed by cases with less than 70% sequence identity with respect to those in the training set used to derive the propensities. However, to further reduce the possibility of overfitting, we defined a new test set by applying a stricter criteria in the selection of the cases, keeping only those ones in which both receptor and ligand had less than 30% of sequence identity with any protein in the data set used to derive the propensities. For this new benchmark, composed of 31 cases (25 of which had at least a near-native solution generated by FTDock), the results for the top 10 predictions not only did not worsen but even slightly improved, with 28.0% success rate for SIPPER (as compared to 24.0% for the statistical energies with Interface 5 Å, 12.0% for RPScore, 8.0% for Skolnick and Glaser, and 28.0% for pyDock).

Analyzing the results of SIPPER by complex type, we obtained top 10 success rates of 34.6%, 38.5%, and 19.0% for enzyme−inhibitor, antibody−antigen, and "other" complex types, respectively. In Figure 4 we can see an example of an enzyme−inhibitor complex (PDB 1CGI) in which SIPPER finds the correct
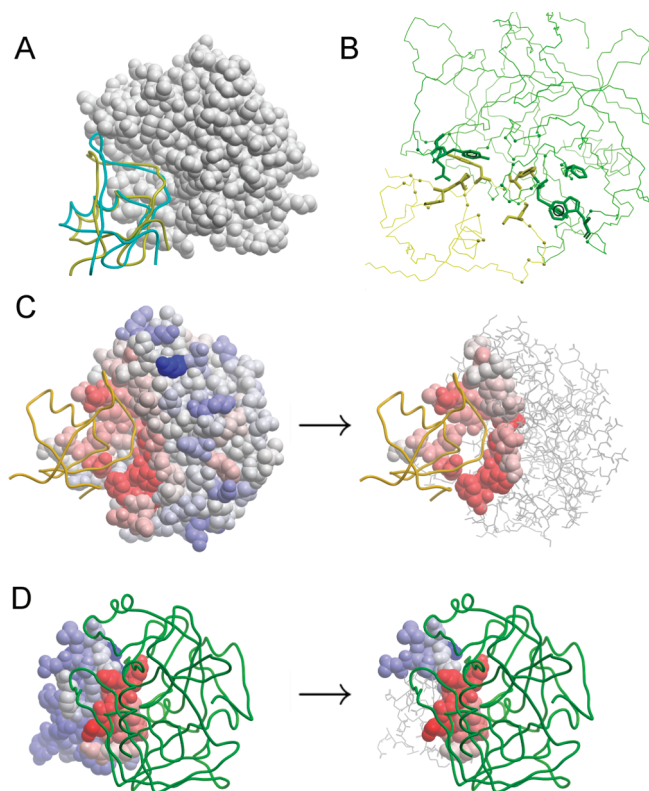
Figure 4. Application of SIPPER scoring to the benchmark case 1CGI. (A) First prediction of SIPPER in yellow, with 2.9 Å ligand rmsd with respect to the reference (in cyan), after superimposing receptors (white cpk). (B) Interface details of SIPPER prediction (receptor in green, ligand in yellow). The Cα atoms of the residues involved in the pairwise propensities score are represented as balls. The side chains of residues forming pairs with propensities < −0.30 are displayed. (C) ODA values for the receptor (red favorable desolvation; blue unfavorable desolvation), with the predicted ligand represented in yellow. The second panel shows the ODA values that are accounted for the ODA scoring of the receptor (only those values of residues being at the interface; see Materials and Methods). (D) Analysis as in C but with ODA values for the ligand.

docking orientation with rank 1. The critical pairs contributing to the scoring are shown, and the important role of the ODA values for that prediction is clearly seen.

We further analyzed some of the cases for which our approach did not find a near-native solution among the top 10 predictions. For the benchmark complexes 1K5D and 1K74, the propensity-based scoring placed a near-native solution in positions 21 and 10, respectively, but the ODA scoring performed much worse, yielding near-native rank values of 357 and 4073, respectively. For the 1K5D case, half of the receptor interface had a poor desolvation score, whereas the ligand had no clear patch with a strong desolvation signal. On the contrary, for 1K74 there was a excessively strong desolvation signal in most of the surface for both receptor and ligand, which made it impossible for the ODA scoring to discriminate the correct interface. These cases represent well two of the typical problems of ODA predictions (either too small or too large signal). These incorrect ODA scorings artificially unbalanced the good results obtained from the propensity-based scorings. We also found the opposite situation. In the cases of 1KXP and 1S1Q, the ODA scoring yielded good near-native rank values (5 and 44, respectively), but the

propensity-based scoring was rather bad (near-native rank values of 1164 and 2931, respectively). The top-ranked model for 1KXP according to the propensity-based scoring correctly predicted the interface for the ligand, but it was completely wrong for the receptor. The decoy had a high number of pairs with favorable propensities, like Leu-Met, Val-Met, and Val-Leu, which mislead our scoring. Actually, we checked that the native orientation had a higher proportion of pairs with unfavorable propensity, like Ala-Thr, Glu-Asp, Val-Lys, or Phe-Asp, which penalized the propensity-based ranking. For the case 1S1Q, the top-ranked model according to the propensity-based score had incorrect predicted interface, which was abundant on favorable Ile-Leu and Val-Leu pairs and caused its high scoring. Similar to the previous case, the interface of the native orientation was abundant on unfavorable residue pairs, like Ala-Lys, Ala-Asn, or Val-Lys pairs. They misguided our scoring and could not be overcome by the few favorable contacts present in the interface.

Docking Set Enrichment in Near-Native Docking Solutions by SIPPER. In a previous study, we found that the success of pyDock predictions strongly depended on the density of near-native solutions within the docking set.[32] Indeed, in cases with only one near-native solution among 10 000 docking poses, pyDock was unable to place such solution in the top 10 predictions. On the contrary, results clearly improved as the relative number of near-native solutions increased. Here we analyzed the use of the statistical energies (Interface 5 Å), ODA scoring and SIPPER to enrich the docking sets in near-native solutions, so that pyDock or other scoring functions could benefit from their application to the enriched sets. For this purpose, we created different subsets of docking poses selecting the top 1%, 5%, and 10% ranked predictions according to the different scoring criteria (Table 2).

The density of near-native solutions dramatically increased in all these subsets generated after filtering by all the scoring criteria (Table 2). Interestingly, filtering by ODA scoring increased the density of near-native solutions more than filtering by the statistical potentials. However, the number of cases with a near-native solution when filtering by ODA scoring was smaller than when filtering by statistical potentials, due to the high efficiency of ODA in certain complexes where binding is mostly driven by hydrophobic terms, while the statistical potentials have a more general application. Indeed, SIPPER (which combines both effects) significantly improved the average enrichment of near-native solutions with respect to the whole initial set to 5.3, 8.0, and 24.3 for the top 10%, 5%, and 1% SIPPER predictions, respectively. It also increased the number of cases with near-native solutions in the filtered sets with respect to the individual scorings. These results are comparable to enrichment factors reported by other scoring functions like ENDES, a combination of energy, conservation, and individual propensities.[33]

Filtering with SIPPER (Taking a Sip) and Scoring by pyDock. Given the good enrichment in near-native docking solutions by SIPPER, we next studied the performance of pyDock scoring on these enriched subsets. As SIPPER is around 300 times faster than pyDock on average, any reduction in the number of solutions to be scored without compromising the predictive results would be relevant. Figure 5 shows the success rates of pyDock scoring on the subsets generated from the top 1%, 5%, and 10% SIPPER predictions. Notably, pyDock success rates for top 10 and 100 predictions improved in all SIPPER subsets, as compared to the results in the complete docking sets. Remarkably, for the top 1% SIPPER subset, the success rates reached 28.4% and 55.6%, respectively (as compared to 22.2%

Table 2. Enrichment of near-Native Solutions in the Subsets Generated by Selecting the Top 1%, 5%, and 10% Docking Orientations As Sorted by Statistical Energies, ODA Scoring, and SIPPER

	statistical energies			ODA scoring			SIPPER		
	top 10%	top 5%	top 1%	top 10%	top 5%	top 1%	top 10%	top 5%	top 1%
% NN filtered in[a]	42.2	30.6	15.1	48.3	37.1	17.5	53.3	40.1	24.3
NN density (%)[b]	0.41	0.61	1.43	0.53	0.80	1.51	0.54	0.82	2.30
% cases with NN	70.4	64.2	45.7	65.4	55.6	35.8	75.3	66.7	55.6
global enrichment[c]	4.2	6.1	15.1	4.8	7.4	17.5	5.3	8.0	24.3

[a] Percentage of near-native (NN) solutions in each top n% subset with respect to the total number of near-native solutions in the whole set, averaged for the 81 cases with near-native solutions. [b] Density of near-native solutions in each subset, i.e. percentage of docking poses in each subset that are near-native solutions, averaged for the 81 cases with near-native solutions. [c] Averaged density (see above) of near-native solutions in each subset with respect to that in the whole set.
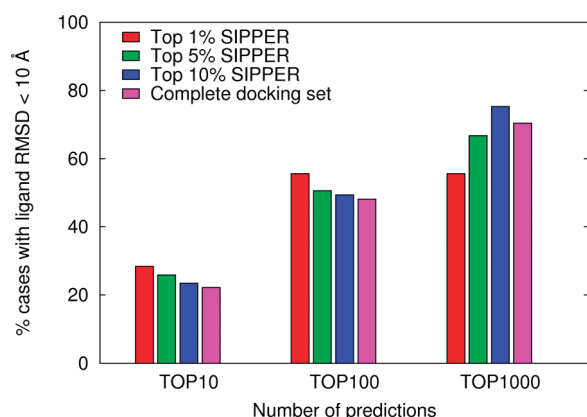
Figure 5. Success rates of pyDock scoring on subsets generated by top 1%, 5%, and 10% docking poses as sorted by SIPPER, as compared to that on the complete docking sets.

and 48.1% when running pyDock on the complete docking sets). Not only the predictive results improve but also the total computer cost is dramatically reduced (roughly 100 times).

As we described in previous subsections, SIPPER scoring (27.2% success rate for top 10) outperforms that of pyDock (22.2%) and would be the desired strategy to score large sets of rigid-body docking poses, since it is much faster. However, here we have shown that when pyDock is applied to a 1% subset selected with SIPPER, it can achieve better top 10 success rate (28.4%). This is interesting for future flexible refinement of docking poses: since any refinement strategy would be too slow to be applied to the whole docking set, SIPPER subsets could largely reduce the number of solutions for refinement and atom-based pyDock energy rescoring.

Interestingly, our propensities were statistically derived from a small, but carefully curated, data set of complex structures and give docking success rates that are comparable to those of other recently reported pairwise potentials generated by using a more sophisticated methodology based on docking decoys and mathematical programming.[34] It is noticeable how such different approaches are able to capture similar relevant determinants for protein–protein binding at residue level.

CONCLUSIONS

We have computed here a new set of intermolecular residue–residue propensities from a well-characterized and manually curated data set of heterodimeric complexes, which can be efficiently applied as statistical potentials for rigid-body docking. Our propensities were more successful in scoring docking poses than other available propensities extracted from different structural data sets. In combination with a coarse-grained desolvation score, the performance substantially improved and even outperformed energy-based functions like pyDock. Moreover, SIPPER can also be efficiently used to dramatically enrich the docking set in near-native solutions.

ASSOCIATED CONTENT

S **Supporting Information.** Interface 5 Å (SI-1) and Surface 5 Å (SI-2) statistical potentials; statistical errors for the Interface 5 Å (SI-3) and Surface 5 Å (SI-4) statistical potentials; and observed occurrences of residue pairs at a distance threshold of 5 Å (SI-5). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: juanf@bsc.es.

ACKNOWLEDGMENT

REFERENCES

(1) Arkin, M. R.; Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug. Discov.* **2004**, *3*, 301–317.

(2) Russell, R. B.; Alber, F.; Aloy, P.; Davis, F. P.; Korkin, D.; Pichaud, M.; Topf, M.; Sali, A. A structural perspective on protein-protein interactions. *Curr. Opin. Struct. Biol.* **2004**, *14*, 313–324.

(3) Ritchie, D. W. Recent progress and future directions in protein-protein docking. *Curr. Protein Pept. Sci.* **2008**, *9*, 1–15.

(4) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195–2199.

(5) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **1997**, *272*, 106–120.

376

dx.doi.org/10.1021/ci100353e |*J. Chem. Inf. Model.* 2011, 51, 370–377

(6) Ritchie, D. W.; Kemp, G. J. Protein docking using spherical polar Fourier correlations. *Proteins* **2000,** *39,* 178–194.

(7) Mintseris, J.; Pierce, B.; Wiehe, K.; Anderson, R.; Chen, R.; Weng, Z. Integrating statistical pair potentials into protein complex prediction. *Proteins* **2007,** *69,* 511–520.

(8) Garzon, J. I.; Lopéz-Blanco, J. R.; Pons, C.; Kovacs, J.; Abagyan, R.; Fernandez-Recio, J.; Chacon, P. FRODOCK: a new approach for fast rotational protein-protein docking. *Bioinformatics* **2009,** *25,* 2544–2551.

(9) Lo Conte, L.; Chothia, C.; Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **1999,** *285,* 2177–2198.

(10) Moont, G.; Gabb, H. A.; Sternberg, M. J. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins* **1999,** *35,* 364–373.

(11) Glaser, F.; Steinberg, D. M.; Vakser, I. A.; Ben-Tal, N. Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins* **2001,** *43,* 89–102.

(12) Lu, H.; Lu, L.; Skolnick, J. Development of unified statistical potentials describing protein-protein interactions. *Biophys. J.* **2003,** *84,* 1895–1901.

(13) Zhang, C.; Liu, S.; Zhou, H.; Zhou, Y. An accurate, residue-level, pair potential of mean force for folding and binding based on the distance-scaled, ideal-gas reference state. *Protein Sci.* **2004,** *13,* 400–411.

(14) Huang, S.; Zou, X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins* **2008,** *72,* 557–579.

(15) Sippl, M. J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **1990,** *213,* 859–883.

(16) Ofran, Y.; Rost, B. Analysing six types of protein-protein interfaces. *J. Mol. Biol.* **2003,** *325,* 377–387.

(17) Fernández-Recio, J.; Totrov, M.; Abagyan, R. Identification of protein-protein interaction sites from docking energy landscapes. *J. Mol. Biol.* **2004,** *335,* 843–865.

(18) Fernandez-Recio, J.; Totrov, M.; Skorodumov, C.; Abagyan, R. Optimal docking area: a new method for predicting protein-protein interaction sites. *Proteins* **2005,** *58,* 134–143.

(19) Fernández, D.; Vendrell, J.; Avilés, F. X.; Fernández-Recio, J. Structural and functional characterization of binding sites in metallo-carboxypeptidases based on Optimal Docking Area analysis. *Proteins* **2007,** *68,* 131–144.

(20) Federici, L.; Du, D.; Walas, F.; Matsumura, H.; Fernandez-Recio, J.; McKeegan, K. S.; Borges-Walmsley, M. I.; Luisi, B. F.; Walmsley, A. R. The crystal structure of the outer membrane protein VceC from the bacterial pathogen Vibrio cholerae at 1.8 A resolution. *J. Biol. Chem.* **2005,** *280,* 15307–15314.

(21) Bolanos-Garcia, V. M.; Fernandez-Recio, J.; Allende, J. E.; Blundell, T. L. Identifying interaction motifs in CK2beta--a ubiquitous kinase regulatory subunit. *Trends Biochem. Sci.* **2006,** *31,* 654–661.

(22) Fernández-Recio, J.; Totrov, M.; Abagyan, R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* **2003,** *52,* 113–117.

(23) Burgoyne, N. J.; Jackson, R. M. Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* **2006,** *22,* 1335–1342.

(24) Cheng, T. M.; Blundell, T. L.; Fernandez-Recio, J. pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins* **2007,** *68,* 503–515.

(25) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003,** *125,* 1731–1737.

(26) Chakrabarti, P.; Janin, J. Dissecting protein-protein recognition sites. *Proteins* **2002,** *47,* 334–343.

(27) Grosdidier, S.; Pons, C.; Solernou, A.; Fernández-Recio, J. Prediction and scoring of docking poses with pyDock. *Proteins* **2007,** *69,* 852–858.

(28) Pons, C.; Solernou, A.; Perez-Cano, L.; Grosdidier, S.; Fernandez-Recio, J. Optimization of pyDock for the new CAPRI challenges: Docking of homology-based models, domain-domain assembly and protein-RNA binding. *Proteins* **2010,** *78,* 3182–3188.

(29) Hwang, H.; Pierce, B.; Mintseris, J.; Janin, J.; Weng, Z. Protein-protein docking benchmark version 3.0. *Proteins* **2008,** *73,* 705–709.

(30) Méndez, R.; Leplae, R.; De Maria, L.; Wodak, S. J. Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **2003,** *52,* 51–67.

(31) Carter, P.; Lesk, V. I.; Islam, S. A.; Sternberg, M. J. Protein-protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI. *Proteins* **2005,** *60,* 281–288.

(32) Pons, C.; Grosdidier, S.; Solernou, A.; Pérez-Cano, L.; Fernández-Recio, J. Present and future challenges and limitations in protein-protein docking. *Proteins* **2010,** *78,* 95–108.

(33) Liang, S.; Meroueh, S. O.; Wang, G.; Qiu, C.; Zhou, Y. Consensus scoring for enriching near-native structures from protein-protein docking decoys. *Proteins* **2009,** *75,* 397–403.

(34) Ravikant, D. V. S.; Elber, R. PIE-efficient filters and coarse grained potentials for unbound protein-protein docking. *Proteins* **2010,** *78,* 400–419.