PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Improving ranking of models for protein complexes with side chain modeling and atomic potentials

Shruthi Viswanath,[1,2] D. V. S. Ravikant,[3] and Ron Elber[2,4]*

[1] Department of Computer Science, University of Texas, Austin, Texas 78712

[2] Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712

[3] WalmartLabs, San Bruno, California 94066

[4] Department of Chemistry and Biochemistry, University of Texas, Austin, Texas 78712

ABSTRACT

An atomically detailed potential for docking pairs of proteins is derived using mathematical programming. A refinement algorithm that builds atomically detailed models of the complex and combines coarse grained and atomic scoring is introduced. The refinement step consists of remodeling the interface side chains of the top scoring decoys from rigid docking followed by a short energy minimization. The refined models are then re-ranked using a combination of coarse grained and atomic potentials. The docking algorithm including the refinement and re-ranking, is compared favorably to other leading docking packages like ZDOCK, Cluspro, and PATCHDOCK, on the ZLAB 3.0 Benchmark and a test set of 30 novel complexes. A detailed analysis shows that coarse grained potentials perform better than atomic potentials for realistic unbound docking (where the exact structures of the individual bound proteins are unknown), probably because atomic potentials are more sensitive to local errors. Nevertheless, the atomic potential captures a different signal from the residue potential and as a result a combination of the two scores provides a significantly better prediction than each of the approaches alone.

Proteins 2013; 81:592–606.

Key words: protein docking; mathematical programming; scoring functions; inequalities; contact potentials; energy minimization

INTRODUCTION

Detailed structural characterization of protein–protein interactions at the atomic level is essential for deducing details of protein function, protein engineering, and for structure-based drug design. Proteins carry out functions by interacting with other proteins and the details of their complex formation can be instrumental in understanding cellular processes at the molecular level. Besides, structures of protein–protein interfaces can be engineered to alter the specificity or strength of binding, and atomic details of how two proteins bind provide a basis for designing small molecules (drugs) that can inhibit the binding and impact the pathway for the cellular process that the two proteins are involved in.[1,2]

Structural characterization of protein complexes is harder to obtain experimentally than the tertiary structures of the individual proteins. Therefore, computational methods for docking protein pairs can be a useful approach when experimental structures are not available. Computational modeling of protein complexes generally consists of two steps: an initial global search in six dimensional space for an optimally docked configuration (keeping the structures of the individual proteins fixed), and a refinement step in a subspace obtained from the results for the initial search.[3] Different algorithms have been used for the global search, such as Fast-Fourier Transforms[4–7] which are by far the most popular, Monte-Carlo searches using a coarse-grained representation of proteins,[8,9] and geometric hashing.[10] The scoring functions used to guide the rigid docking search include shape matching,[5,7] statistical potentials measuring amino acid affinities,[5,7] and physical energy terms such as van der Waals and electrostatic interactions.[3]

Grant sponsor: NSF; Grant number: CCF-0833162; Grant sponsor: Welch; Grant number: F-1783
*Correspondence to: Ron Elber, 201 East 24th St. STOP C0200, Austin TX 78712-1229. E-mail: ron@ices.utexas.edu.
Received 27 August 2012; Revised 23 October 2012; Accepted 29 October 2012
Published online 23 November 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI:10.1002/prot.24214

The second step usually adjusts the rigid docking structures obtained in the first step and re-ranks them using fine-grained energy terms. RosettaDock uses an iterative Monte-Carlo search starting from rigid docking structures, first rebuilding side-chains of existing structures, and then minimizing the rigid structure of the two proteins using an elaborate energy term, the Rosetta potential.[1,9,11] Monte-Carlo approaches have also been used by others to incorporate rigid-body and side chain movements in refining docked conformations.[12] Weng and coworkers developed RDOCK,[13] a refinement algorithm which uses energy minimization and re-ranks models with a combination of electrostatics and knowledge-based potentials representing desolvation. They later developed a faster algorithm for the second step, ZRANK,[14] that is a linear combination of a knowledge-based atomic potential, ACE, with electrostatic and van der Waals terms.

Wolfson and coworkers developed the refinement algorithms, FIREDOCK,[15] which incorporates restricted side-chain flexibility and orientation adjustments and its improved version, FIBERDOCK,[16] which incorporates backbone flexibility using normal modes in addition to side-chain flexibility. GRAMM-X uses conjugate gradient minimization with a smoothed Lennard-Jones type potential and ranks the models with a scoring function that is a combination of residue-based and atom-based terms.[4] The Cluspro team developed a refinement method using Monte-Carlo runs with semi-definite programming with underestimation (SDU).[6,17] Recio and coworkers use hydrogen-bond optimization along with energy minimization of all-atom force-fields in order to refine docking poses.[18] Zhou and coworkers perform a short minimization and restricted re-sampling near existing models followed by re-ranking using DFIRE and EMPIRE energy functions.[19]

Most of the methods so far address the case in which the constituents do not undergo drastic conformational changes in the complex compared to the unbound state. Indeed, in an analysis of complexes in ZLAB benchmark 3.0, it has been found that the root mean square deviation (RMSD) of the ligand between the bound and unbound structures varies from 0.2 to 8 Å, with half of the complexes having an RMSD less than 1 Å^2.[20] This also corresponds to our analysis of 178 unbound complexes in our learning set, where we find the average TM scores (Template Modeling[32] scores) between the unbound and bound chains to be 0.8953 and 0.8875 for the receptor and ligand, respectively. In our work too, we consider cases in which no large-scale movements take place in the individual constituents.

This article describes an algorithm for the step of refinement and re-ranking. The input to this step is complexes obtained from our rigid docking package, DOCK/PIE.[5] We name the new program DOCK/PIERR, for DOCK/PIE-**R**efine**R**erank. It is based on rigid docking followed by side chain remodeling and minimization. These adjusted structures are re-ranked with a newly developed atomic potential and a previously developed residue potential.[5] We first model the side chains of interface residues from the top 1000 structures from rigid docking, using the program SCWRL.[21] In an analysis of the ZLAB Benchmark 3.0[20] set it is observed that interface residues undergo the maximum change in conformation upon docking, and most of these changes bring the interface residues from high-energy to low-energy torsion angles.[22] After modeling side-chains of the structures, we minimize them with the molecular dynamics program MOIL[23] using the OPLS force field.[24] The minimization helps to reduce interface clashes in the model. Finally, we re-rank the refined models using a combination of the new atomic and residue potentials. It has been shown that refining structures before re-ranking improves the recognition of near-native structures. Also, an optimized combination of atomic and residue potentials result in an improved discrimination as noted by others previously.[25–27] The details are described in the Methods section below.

METHODS

Learning and test sets

For optimizing the parameters of the new atomic potential, we used a learning set of 640 complexes used in previous studies,[5,28] which contains 462 bound, and 178 unbound complexes. The new methods were tested on three datasets. The first dataset comprises of 124 complexes from the ZLAB Benchmark 3.0,[20] a standard benchmark test set used by the protein–protein docking community. The second dataset comprises of 640 targets from our learning set. The third dataset is a set of 30 complexes that is independent from the learning set, and details of this dataset are available in the Results section.

Rigid docking

Given the chains of the receptor and ligand molecules, we used our previously developed docking package DOCK/PIE,[5] to generate a training set for refinement. We retain top scoring $2^{19} = 524{,}288$ fast fourier transform (FFT)-based transformations for each complex. These transformations are then clustered in rigid body space and scored using the potential, PIE,[5] which consists of a pairwise residue contact term along with van der Waals attraction and repulsion terms. Subsequently, the top scoring transformations are filtered for clashes, and clustered again using interface RMSD.

Side chain refinement

We chose the top 1000 models from DOCK/PIE rigid docking, ranked by DOCK/PIE, for each pair of proteins in the training set to refine and re-rank. The number of models must be large enough to include a near-native

model, and small enough to make the (more expensive) refinement process efficient. Our choice of the number of models was based on an examination of CAPRI targets. In our CAPRI dataset, 19 of 22 targets had near-native models (acceptable by CAPRI definition) in the top 1000, while the number dropped to 16 considering the top 500 models. Out of the remaining three targets that did not have near-native models in the top 1000, one target had a near-native model at position 3500, and the other two did not have any near-natives in the output after clustering. The choice of 1000 candidates therefore seems reasonable.

In order to reduce the number of clashes, make the rigid docking poses chemically sound, and improve the energies (scores) of the models, we first performed side-chain refinement using SCWRL4.[21] SCWRL is a side-chain prediction program that uses graph-based decomposition to identify the set of optimal rotamers for the side chains of a given model. We used a cutoff distance of 6 Å between the two proteins to identify interface residues in the top 1000 models for each target (complex). We then modeled the side chains of only those interface residues using SCWRL.

Minimization

After side chain refinement, clashes are removed by 100 steps of conjugate gradient energy minimization. The minimization was performed using the routine mini_pwl (conjugate gradient descent with Powell restart) of the molecular dynamics package MOIL[23] and the embedded OPLS-AA energy function. During the minimization, the receptor and ligand molecules are modeled as rigid bodies. Minimization both in vacuum and using implicit solvent models (GBSA) was performed. However, no noticeable difference in the results was observed between the two procedures (the structural adjustments were small: there was a change in the RMSD to the native structure by only ~0.05 Å RMSD after refinement). Therefore, we decided to use minimization in vacuum since it is more efficient. Overall, the refined structures are not more similar to the X-ray structures compared to the unrefined complexes, and distance of the refined structure in terms of RMSD from the initial structure is minute (~0.2–0.4 Å). The refinement is nevertheless useful since it allows for better ranking.

Atomic potential

We designed a distance-dependent pairwise atomic potential to re-rank the top 1000 refined structures. Using the atomic potential on the refined structures, we expect to generate more hits in the top 10 (or top 1) than the rigid docking procedure alone. The parameters for the atomic potential were learnt using mathematical programming from the top 1000 refined models (refined as described by the procedure described above) of each complex in our learning set.

The heavy atoms were collected into 32 types, as reported earlier for threading potentials.[29] We employed three distance bins: 2–3.5 Å, 3.5–5 Å, and 5–8 Å, which is the same as used in Ref. 29 to recognize approximate structures for threading/homology modeling.

Knowledge-based pairwise atomic potentials are frequently modeled by a square-well potential, that is they designate a single value, $u(i,j,d)$ for a distance range, d, and atom-type pair (i,j). For clarity, we replace the pair of interaction (i,j) by a single index α of the interaction type. If an atom of type i is found within a distance d from an atom of type j, then the value $u(\alpha,d)$ is added to the energy of the structure. The energy or score, $E(X)$ of a complex X, with a receptor A and ligand B, is a sum of all pairwise interactions between atoms i and j in A and B respectively, and is given by

$$E(X) = \sum_{\alpha,d} n(\alpha, d)\, u(\alpha, d) \qquad (1)$$

where $n(\alpha, d)$ is the number of interactions of type α (i.e., we use a single index to describe the interaction of particles i and j) at distance d. We found that a better accuracy was obtained when a continuous function is used between bins rather than steps. We use a linear interpolation between the bins, as shown in Figure 1.

The corresponding equations for the geometrical factor, $n(\alpha,d)$ are given in Eq. (2)

where r_{ab} is the distance between an atom pair (i,j) where i belongs to protein A and j belongs to protein B.

$$\begin{array}{cc}
n(\alpha,1) = 1.0 & 2\,\text{Å} \le r_{ab} < 3\,\text{Å} \\
\left.\begin{array}{l} n(\alpha,1) = 4.0 - r_{ab} \\ n(\alpha,2) = r_{ab} - 3.0 \end{array}\right\} & 3\,\text{Å} \le r_{ab} < 4\,\text{Å} \\
n(\alpha,2) = 1.0 & 4\,\text{Å} \le r_{ab} < 4.5\,\text{Å} \\
\left.\begin{array}{l} n(\alpha,2) = 4.0 - \frac{r_{ab}}{3.0} \\ n(\alpha,3) = \frac{r_{ab}}{3.0} - 3.0 \end{array}\right\} & 4.5\,\text{Å} \le r_{ab} < 6\,\text{Å} \\
n(\alpha,3) = 1.0 & 6\,\text{Å} \le r_{ab} < 8\,\text{Å}
\end{array} \qquad (2)$$

Note that the values of $n(\alpha, i)$ are fixed by the coordinates of the atoms in the structure. For every distance bin, $(i = 1,2,3)$ we identify a different multiplicative energy term $u(\alpha, i)$.

Constraints used for learning

The formulation above led to $p = 1584 = \frac{32(32+1)3}{2}$ parameters for the potential. We learn these parameters taking into account known correctly and incorrectly docked structures. The energies of correct and incorrect pairs are used to create inequalities (correctly docked pairs must have lower energies) of the type $E_{\text{incorrect}} - E_{\text{correct}} > 0$. The values of the parameters were obtained by solving the inequalities by linear programming using the LP solver, PF3.[30] More specifically the following types of inequalities are used:

(a) Inequalities comparing near-native and misdocked models

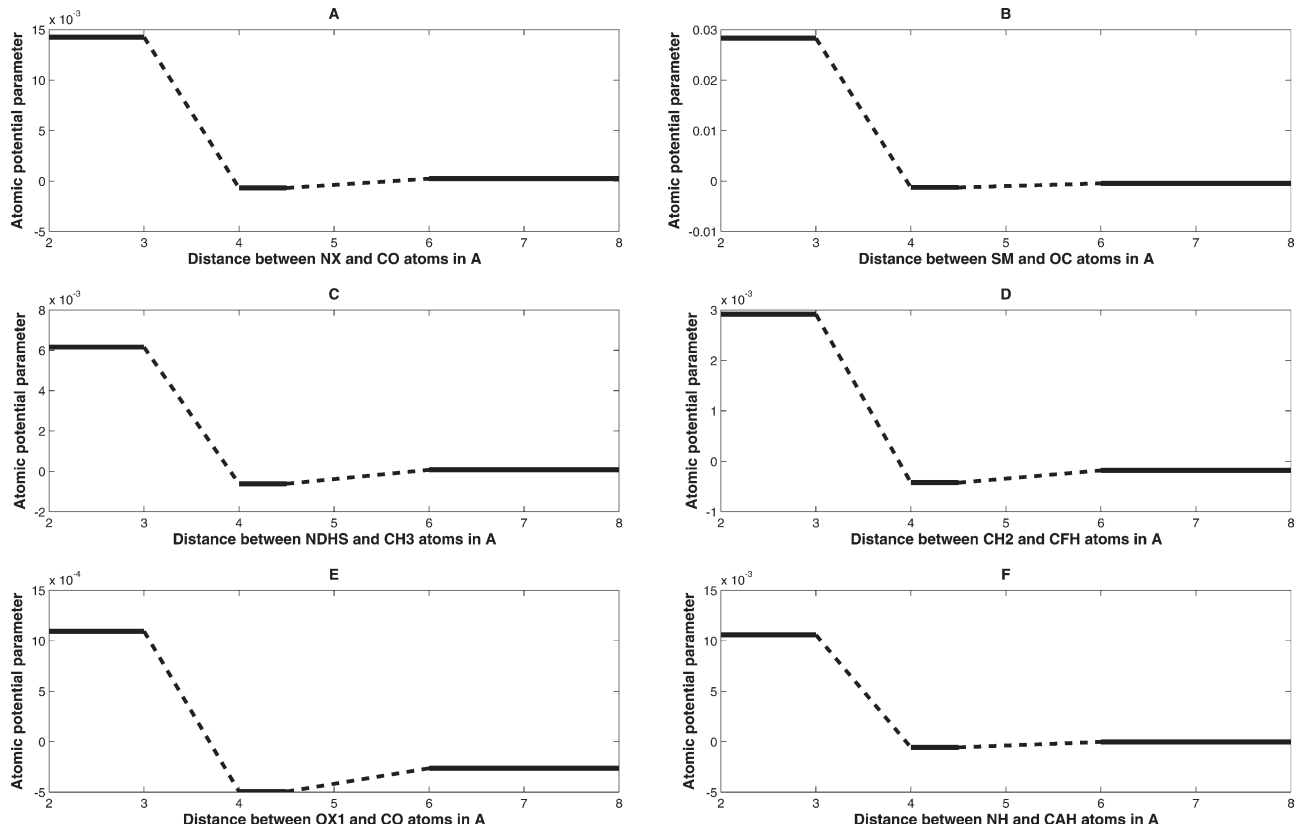

Figure 1

Examples of atomic potentials $u(\alpha,d)$ for six different pairs of atom types. **A**: NX (Lyz-NZ) and CO (carbon of backbone carbonyl). **B**: SM (MET-Sulfur) and OC (oxygen of carbonyl groups). **C**: NDHS (TRP-NE1) and CH_3 (terminal aliphatic side chain carbon). **D**: CH_2 (beta carbon) and CFH (aromatic side chain carbon). **E**: OX1 (ASP-OD1, OD2, GLU-OE1, OE2) and CO (carbonyl carbon). **F**: NH (amide nitrogen) and CAH (alpha carbon of amino acids). More details on atom types are given in Ref. 29. The three distance bins, 2–3.5 Å, 3.5–5 Å, and 5–8 Å have one single parameter value in the regions in the middle of the distance bins. The outer one-third portion of the distance bins adjacent to neighboring bins is modeled by a straight-line interpolation between the bins.

Of the top 1000 models for each complex, we define "near-native" as models having an interface RMSD less than 2.5 Å to the native PDB structure. We call conformations "misdocked" if they have an interface RMSD greater than 7 Å. We added the native structure to the set of near-natives for each target in our learning set. We then require that for each target, the atomic potential have a lower (better) energy for near-native models than for misdocked models.

$$E(\mathrm{X}_{\mathrm{misdocked}} - \mathrm{X}_{\mathrm{near\text{-}native}}) > 0 \qquad (3)$$

The above inequalities are linear in their parameters, which make them accessible to efficient calculations. We rewrite Eq. (3) in terms of Eq. (1).

$$\sum_{\alpha,d} u(\alpha,d)[n_{\mathrm{misdocked}}(\alpha,d)] - \sum_{\alpha,d} u(\alpha,d)[n_{\mathrm{near\text{-}native}}(\alpha,d)] > 0$$
$$\sum_{\alpha,d} u(\alpha,d)[n_{\mathrm{misdocked}}(\alpha,d) - n_{\mathrm{near\text{-}native}}(\alpha,d)] > 0 \qquad (4)$$

Equation (4) illustrates that the inequalities are linear in the parameters $u(\alpha,d)$. Hence, the inequalities can be solved efficiently by mathematical programming.

In Eq. (4) the solution for the parameter set is up to a multiplicative positive constant λ. Hence, if $u(\alpha,d)$ is a solution so is $\lambda u(\alpha,d)$. It is therefore convenient to put the right hand side of the equation to 1 instead of zero. This choice sets a scale for the parameter values and makes the numerical calculation easier.

Furthermore, we also allow for some errors in our solutions. It is not possible to satisfy all the inequalities because the functional form is not known exactly, and its current form is not flexible enough to solve all the constraints. On the other hand, making the functional form more complex may lead to the phenomenon of over-learning. In over-learning, the new scoring function performs better on the training set than on new targets. New targets are obviously of more interest in practical applications and we aim for comparable performance on the training set and other targets. Therefore we remain with the simpler functional form while accepting some mis-classification.

Table I
Statistics of Hits in the Learning Set

High-quality hit cutoff (hc) in Å	Number of high quality hits (*i*RMSD < hc)	Number of good hits (hc < *i*RMSD < 2.5 Å)	Number of resulting inequalities
1.0	993	5416	40,860
1.5	2341	4068	86,930
2.0	4156	2253	79,535

See text for more details.

The existence of mis-classifications is further amplified by the use of near-native structures as "correct" structures, instead of actual native complexes, and docking of unbound structures instead of bound structures of the individual chains in the complex. A near-native structure as a target and the use of unbound chains mimics better the conditions of an actual prediction. However, it also increases the noise level and introduces uncertainties to the classification. Rather than the strict constraint, we add to each inequality i a slack variable z_i. Equation (3) then becomes:

$$E(\mathrm{X}_{\mathrm{misdocked}} - \mathrm{X}_{\mathrm{near\text{-}native}}) > 1 - z_i \qquad z_i > 0 \tag{5}$$

(b) Inequalities comparing high-quality hits and good hits

We found it beneficial to add inequalities comparing near-native structures with interface RMSD less than 1.5 Å compared to the native complex (call it high-quality hits) and hits with interface RMSD between 1.5 and 2.5 Å (known as good hits). The value of the cutoff 1.5 Å was obtained based on statistics of hits in the targets of the learning set. It was chosen such that there was an even distribution of models in the high-quality and good hit categories and the number of additional inequalities was maximized, as shown in Table I. The new inequalities require that the high-quality hits will have lower energies than good hits.

$$E(\mathrm{X}_{\mathrm{good\_hit}} - \mathrm{X}_{\mathrm{hq\_hit}}) > 1 - z_i \tag{6}$$

(c) Inequalities comparing pairwise adjacent hits

We also sorted the hits (models with *i*RMSD less than 2.5 Å) by *i*RMSD, and formulated inequalities comparing energies of neighboring hits. For example, model i has lower *i*RMSD than model $i+1$. Therefore, we expect the energy of the ith model to be lower than the energy of the model ranked $i+1$ in the list. Here n_{hits} is the total number of hits for a target in the learning set.

$$E(\mathrm{X}_{\mathrm{hit}}{}^{i+1} - \mathrm{X}_{\mathrm{hit}}{}^{i}) > 1 - z_i; \quad i = 1, 2..n_{\mathrm{hits}} - 1 \tag{7}$$

Using these three types of inequalities, we had a total of 5,841,395 inequalities in our learning set. The complete set of constraints is now combined with an objective function that was minimized. The objective function is the sum of the parameters, $u(\alpha, d)$ and slack variables, z, where γ is an empirical constant that provides weight to the violations of the constraints relative to precise determination of the parameters.

$$\min \left\| \sum_{\alpha,d} u(\alpha, d) \right\|_1 + \gamma \left\| \sum_i z_i \right\|_1 \tag{8}$$

Using PF3,[30] we solved 92.8% of the inequalities. We call the atomic potential PISA [Protein Interactions Scored Atomically] henceforth. We used the value of 1.0 for γ.

For each of the complexes in the learning set we mentioned previously, we used one thousand refined models along with the native structure for the complex to generate the three kinds of inequalities discussed above. For 67 of the complexes in the learning set (58 bound and 9 unbound), one or more backbone atoms in the PDB files were missing. The missing atoms prevent us from placing side chains or minimize continuous atomic energy using MOIL.[23] We attempted to add missing backbone atoms to the complexes using Modeller.[31] However, Modeller tends to move the modeled structure away from the template. We found that the results obtained by learning based on Modeller structures were worse than the results obtained by simply using the unrefined (rigid-docking) models for complexes with missing atoms, as shown in Table II. So for these complexes, we use the unrefined models for learning and testing.

Combining atomic and residue scores for re-ranking

We observed that though the atomic potential PISA obtained above recognizes more hits in the top 100 than

Table II
Performance With and Without Modeller on 67 Targets of the Learning Set

Refinement method for 67 targets	Number of hits in top 10/top 1	Number of targets solved in top 10/top 1
Modeller for modeling missing atoms, followed by SCWRL and minimization	81/15	38/15
Using unrefined decoys in case of missing main chain atoms	86/17	40/17

A hit here is a model with an interface RMSD of 4 Å or lower to the native structure. Note that the ranking of unrefined structures is slightly better than the refined structures. However, the refined models have more sound chemical geometries.

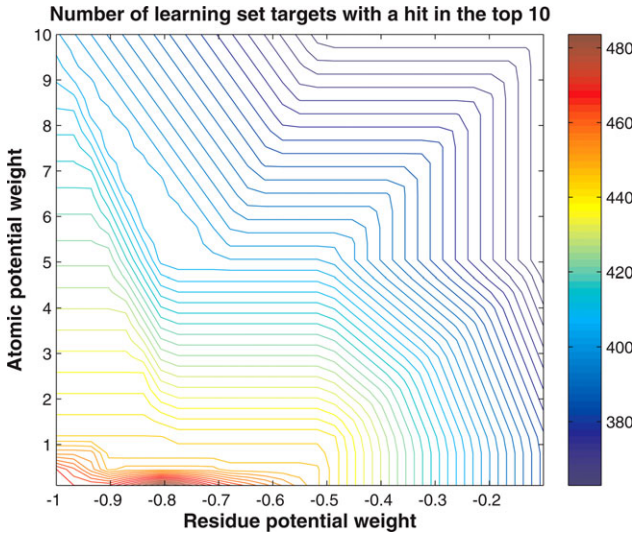

Figure 2

Contour plot showing parameter search for values of coefficients *c* and *d* in Eq. (11). The height represents the number of targets in the learning set with a hit with interface RMSD less than 4 Å in the top 10, for the combination of coefficients of the potential C3. The best values appear to be 0.1 for *c*, the atomic potential weight and −0.8 for *d*, the residue potential weight.

the previously developed residue based potential, PIE,[5] it is not sensitive enough to recognize more hits in the top 10, or top 1. The reason for the lower performance of the atomic potential at the high end of prediction may be the more significant sensitivity of atomic interactions to structural details compared to interactions at the residue level. This sensitivity of the algorithm is amplified by the use of unbound (approximate) complexes rather than just bound complexes with atomically precise interactions.

Realizing that the atomic and residue potentials encapsulate different signals (atomic potential captures shorter range interactions), we decided to combine the two, expecting the combined method to work better than the residue or atomic potentials alone. We used the following combination of potentials.

Product

For the atomic potential, PISA, the lower the energy, the better the model. Whereas for the residue score, PIE, higher the score, the better the model. Hence if we take the product of the scores of PISA and PIE for each model, the lower the score of the product, the better the model should be.

$$\mathrm{C1} = \mathrm{PISA} \times \mathrm{PIE} \tag{9}$$

Linear combination

The second combination potential consisted of a weighted linear combination of the atomic and residue potentials.

$$\mathrm{C2} = \mathrm{PISA} + a.\mathrm{PIE} \tag{10}$$

The value of coefficient *a* was set to −0.2 by learning on the learning set.

Linear combination with product

We also developed another combination potential adding the individual values of the atomic and residue potentials to their product.

$$\mathrm{C3} = c.\mathrm{PISA} + d.\mathrm{PIE} + \mathrm{PISA} \times \mathrm{PIE} \tag{11}$$

The values of the coefficients, *c* and *d*, were found to be 0.1 and −0.8, respectively, as shown in Figure 2.

Tests on other docking packages

We compared our results to the ZDOCK,[7] ZDOCK+ZRANK,[14] CLUSPRO,[17] and PATCHDOCK+FIBERDOCK[16] methods. For ZDOCK results on the ZLAB dataset, we used the latest version ZDOCK-3.0.2 with six degree Euler angle sampling and the results (RMSDs) as reported in the ZLAB website. For the other test sets, we used the downloaded packages for ZDOCK version 3.0.2 and ZRANK. For ZDOCK+ZRANK, we used the top 2000 conformations from ZDOCK as recommended.[14] We then added polar hydrogens to the models using SCWRL4[21] and re-ranked the models using ZRANK. For CLUSPRO, we used the results from the CLUSPRO server[17] which runs CLUSPRO version 2.0. We used the downloaded packages for PATCHDOCK and FIBERDOCK. FIBERDOCK is shown to be an improved refinement and re-ranking method over the same group's FIREDOCK. We used the top 500 models from PATCHDOCK as suggested[16] and refined the backbone and side chains of the models using FIBERDOCK, and re-ranked models with the FIBERDOCK energy term. The packages ZDOCK and CLUSPRO perform only rigid docking and no refinement or rescoring, and are meant to enrich the number of hits in the top 1000 or 2000 structures. For ZDOCK we therefore compared also the enrichment factor. The server for CLUSPRO reports only the top 10 models, and hence we did not compare the enrichment factor for CLUSPRO. It is interesting to note that ZDOCK in our hands scores better than ZDOCK+ZRANK.

RESULTS AND DISCUSSION

Creation of test sets

We tested DOCK/PIERR (pronounced DOCK-by-PIER), our proposed method of docking using DOCK/PIE followed by refinement and re-ranking, on three datasets. The first dataset comprises 124 complexes from the ZLAB Benchmark 3.0.[20] The second dataset com-

prises 640 targets from our learning set, described in the Methods section and used in previous work.[28]

The third dataset is a set of 30 complexes that is independent from the learning set. These are a set of complexes that were deposited in the Protein Data Bank after September 22, 2010. To construct this test set, we queried the Protein Data Bank for soluble two-chain protein–protein complexes, with no DNA, RNA, and free ligands in the structure. We discarded complexes with modified residues, and chains that were shorter than 50 residues in length. The query resulted in 126 complexes.

We then tested to see whether these complexes were similar to any of the complexes in the learning set. We used TM-Align[32] on the individual chains of the bound complex and discarded all complexes where both chains had a TM score of 0.5 or higher with the chains of a target in the learning set. Forty-five of the complexes were not similar to any in the learning set. To perform unbound–unbound docking, we searched for homologs for the individual chains of these 45 complexes using PSI-BLAST.[33] We discarded the complexes with no homologs (BLAST expectation cutoff of 10^{-3}) for either chain.

Then we constructed a homology model for each chain, using the structure of the homolog and the sequence of the chain from the bound complex. We used MODELLER[34] to build the homology model and discarded complexes where the homology models had a TM score of less than 0.8 with the chain in the bound complex. We obtained a set of 30 complexes from this procedure, with 22 having both chains unbound and 8 with one chain unbound. Table III contains the list of homologs used for the unbound docking of novel complexes.

Table III
List of Complexes in the Novel Test Set of 30 Targets Along With the Corresponding Homologs Used to Model the Unbound Receptor and Ligand for Each Complex

Target PDB ID	Receptor homolog (PDB_chain): target receptor chain	Ligand homolog (PDB_chain): target ligand chain
2xt4	2XT2_A:B	2XT2_A:A
2xty	2XTW_A:B	2XTW_A:A
3agx	3AGZ_A:A	3AGZ_A:B
3asy	1XRJ_A:A	1XRJ_A:B
3gt6	3GLA_A:A	3GLA_A:B
3lis	3LFP_A:A	3LFP_A:B
3m7f	3B7Y_A:B	1NRV_A:A
3mxj	3MXI_B:B	3MXI_B:A
3nfy	1T8P_A:B	1T8P_A:A
3oa9	3D6R_B:A	3D6R_B:B
3p2q	3KV7_A:A	3KV7_A:B
3pc6	3PC8_A:B	3PC8_A:A
3pge	3PGE_A:A	3L0W_A:B
3pra	3PRB_A:B	3PRB_A:A
3r8c	3R20_A:A	3R20_A:B
3rd6	3Q64_A:A	3Q64_A:B
3rkc	3HAG_A:B	3HAG_A:A
3t43	3LF6_A:A	3LF6_A:B
3te8	3LR5_A:B	3LR5_A:A
3u80	2UYG_A:A	2UYG_A:B
3umz	3UN0_B:A	3UN0_B:B
3vc8	3VCB_A:B	3VCB_A:A
2wfx	3HO4_B:B	2IBG_H:A
3d65	3D65_E:E	3BTM_I:I
3di3	3DI3_B:B	3DI2_C:A
3hct	1FXT_A:B	3HCT_A:A
3jrq	2IQ1_A:A	3JRQ_B:B
3l1z	3FSH_B:A	3L1Z_B:B
3m18	3M18_A:A	1I56_A:B
3nbp	1MU2_A:A	3NBP_B:B

Results on the ZLAB benchmark

In the following tables, we analyze the results by two metrics: interface RMSD and fraction of native contacts, as defined by the CAPRI assessment.[35] A hit defined in terms of interface RMSD, is a model with interface RMSD less than 4 Å, to the crystal structure of the complex, which is equivalent to an "acceptable" model in the CAPRI assessment. Similarly, a hit in terms of fraction of native contacts is a model with 10 percent or more native contacts, which is one of the criteria for an acceptable model in CAPRI.

We show in Table IV, the comparison of our docking software with ZDOCK, ZDOCK+ZRANK, CLUSPRO, and PATCHDOCK+FIBERDOCK. We compare the performance of DOCK/PIE our rigid docking package with the new DOCK/PIERR, which is DOCK/PIE with side chain remodeling, energy minimization, and re-ranking. Re-ranking is done in various ways, using the atomic potential PISA alone, or the combination potentials, C1, C2, and C3, composed of the atomic and residue potentials.

We note that DOCK/PIERR with C1 and C2 combination potentials performs better than the other DOCK/PIE versions. It is interesting to note that ZDOCK performs better than ZDOCK+ZRANK in terms of number of interface RMSD hits. DOCK/PIERR picks a smaller number of hits than ZDOCK or ZDOCK+ZRANK in the top 10. However, DOCK/PIERR and its various versions are able to solve more targets than ZDOCK. ZDOCK is able to generate a lot of reasonable models for some targets. However, for some targets it does not generate hits at all. DOCK/PIERR is more uniform in the generation of hits. Further discussion on enrichment versus ranking can be found in the next section on the novel set. CLUSPRO is one of the best methods, even though we include the results from the web server only, which does not include the more expensive refinement procedure.

Results on the novel test set

We also compared DOCK/PIE and DOCK/PIERR versions with other leading docking software on the dataset of 30 novel complexes, constructed as described earlier. This is shown in Table V. For ZRANK, the authors

Table IV
Comparison on the ZLAB 3.0 Benchmark Set of 124 Targets

	Interface RMSD		Fraction of native contacts	
Method	Number of hits in top 10/top1	Number of targets solved in top 10/top1	Number of hits in top 10/top1	Number of targets solved in top 10/top1
DOCK/PIE Rigid Docking	73/10	38/10	144/14	59/14
DOCK/PIERR Re-rank with PISA	86/17	40/17	167/28	66/28
DOCK/PIERR Re-rank with C1	107/19	50/19	190/32	72/32
DOCK/PIERR Re-rank with C2	107/19	50/19	194/30	72/30
DOCK/PIERR Re-rank with C3	102/15	46/15	175/23	63/23
CLUSPRO	63/19	50/19	172/31	69/31
ZDOCK	143/13	29/13	276/22	45/22
ZDOCK+ZRANK	96/12	23/12	208/26	50/26
PATCHDOCK+FIBERDOCK	21/2	15/2	56/4	33/4

The number of hits counts all acceptable predictions. Some of the targets can have multiple successful predictions, and all of these hits are counted in the entry "Number of hits." A target is considered solved when at least one prediction is in the top 1 or top 10 set. Only one hit per target is counted under "Number of targets solved." See text for more details.

recommend it to be used on the top 2000 models from ZDOCK. Besides applying ZRANK on the top 2000 models, we also applied ZRANK to the top 1000 models from ZDOCK, since we use the top 1000 models from our rigid docking procedure for re-ranking. We did similarly for FIBERDOCK, which is to be applied on the top 500 models from PATCHDOCK. CLUSPRO and ZDOCK have not been used so far for docking of homology models. But here we are using homology modeling to mimic a "real" docking experiment in which the bound structures are not known. The refinement to atomic structures ensures chemically reasonable conformation, but not better ranking (see also Table II).

Table VI shows the comparison of different docking methods on individual targets in the novel set. We see that targets that are hard for DOCK/PIE are generally also hard for the other docking packages. But there are some targets like 3asy, 3r8c, and 3rd6, where the only software that was able to produce a hit in the top 10 was DOCK/PIE (RR). For 3hct, only ZDOCK+ZRANK was able to produce a hit. For 3d65 and 3nbp only CLUSPRO was able to produce a hit in the top 1. Figure 3 shows some of the models produced by different methods on the novel set.

On the novel set, DOCK/PIE with the residue potential seems to perform better than DOCK/PIERR with the atomic potential. DOCK/PIE rigid docking and DOCK/PIERR with potential C3 performs better than all other docking methods. Again, performance of ZDOCK is superior to ZDOCK+ZRANK. Also, ZRANK applied to top 1000 models seems to be better than the authors' recommendations of applying it on the top 2000 models. For FIBERDOCK, the author recommendation of applying it on the top 500 models seems to work better.

We also compared the enrichment capacity of our rigid docking procedure, DOCK/PIE with that of ZDOCK on the novel set. By enrichment capacity, we mean the number of hits (with interface RMSD less than 4 Å) returned by each rigid docking program in the top 1000 models. Table VII shows that DOCK/PIE enrichment is slightly better than ZDOCK. Analogous results for CLUSPRO were not available since we used the CLUSPRO server that returns only the top 10 models for each target.

Table V
Comparison of Docking Software on the Novel Set of 30 Targets

	Interface RMSD		Fraction of native contacts	
Method	Number of hits in top 10/top 1	Number of targets solved in top 10/top 1	Number of hits in top 10/top 1	Number of targets solved in top 10/top 1
DOCK/PIE Rigid Docking	37/10	16/10	69/14	20/14
DOCK/PIERR Re-rank with PISA	33/7	12/7	50/10	17/10
DOCK/PIERR Re-rank with C1	41/7	15/7	70/11	21/11
DOCK/PIERR Re-rank with C2	43/9	17/9	75/12	21/12
DOCK/PIERR Re-rank with C3	44/10	17/10	72/12	22/12
ZDOCK	39/9	11/9	52/11	14/11
ZDOCK+ZRANK-2000	34/5	10/5	55/9	15/9
ZDOCK+ZRANK-1000	38/5	11/5	60/8	14/8
CLUSPRO	19/8	12/8	48/9	16/9
PATCHDOCK+FIBERDOCK-500	18/4	5/4	32/4	11/4
PATCHDOCK+FIBERDOCK-1000	17/3	5/3	28/3	10/3

Suffix of 1000 for example, means that the re-ranking was applied to top 1000 models from rigid docking.

Table VI
Hits in the Top 10 and Top 1 for Each Target of the Novel Dataset

	Number of irmsd hits in the top 10/top 1 for various docking software										
Targets	D/P Rigid	D/P PISA	D/P C1	D/P C2	D/P C3	ZD	ZR 2000	ZR 1000	CL	PF 500	PF 1000
2xt4	3/1	4/0	4/1	3/1	4/1	1/0	1/0	1/0	1/0	0/0	0/0
2xty	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3agx	3/1	3/1	3/1	3/1	3/1	7/1	6/1	7/1	1/1	4/1	4/1
3asy	2/0	3/1	4/0	3/0	3/0	0/0	0/0	0/0	0/0	0/0	0/0
3gt6	1/0	2/1	2/0	2/0	1/0	2/1	4/1	4/1	1/0	0/0	0/0
3lis	4/1	5/1	5/1	4/1	4/1	4/1	8/1	9/1	3/1	6/1	5/1
3m7f	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3mxj	1/0	0/0	1/1	1/0	1/1	0/0	1/1	1/1	2/1	0/0	0/0
3nfy	3/1	3/1	3/0	3/1	4/1	7/1	6/1	6/1	2/1	1/1	1/0
3oa9	1/1	4/0	3/1	2/1	2/1	7/1	2/0	3/0	1/1	0/0	0/0
3p2q	1/1	1/1	1/1	1/1	1/1	4/1	3/0	3/0	2/1	6/1	6/1
3pc6	2/1	1/0	1/0	2/1	2/1	1/1	0/0	0/0	0/0	0/0	0/0
3pge	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3pra	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3r8c	1/1	0/0	0/0	1/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0
3rd6	3/1	3/1	3/1	3/1	3/1	0/0	0/0	0/0	0/0	0/0	0/0
3rkc	2/0	1/0	3/0	3/1	3/1	1/0	1/0	1/0	1/0	1/0	0/0
3t43	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3te8	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3u80	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3umz	6/0	3/0	5/0	6/0	6/0	1/1	2/0	2/0	1/0	0/0	0/0
3vc8	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
2wfx	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3d65	2/0	0/0	1/0	2/0	2/0	0/0	0/0	0/0	3/1	0/0	0/0
3di3	2/1	0/0	2/0	3/0	3/0	4/1	0/0	0/0	0/0	0/0	1/0
3hct	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0
3jrq	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3l1z	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3m18	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3nbp	0/0	0/0	0/0	1/0	1/0	0/0	0/0	0/0	1/1	0/0	0/0

Hits correspond to models with less than 4 Å *i*RMSD. The abbreviations used are as follows: D/P Rigid: DOCK/PIE. D/P PISA: DOCK/PIE re-rank with PISA. D/P CX: DOCK/PIE re-rank with combination potential CX. ZD: ZDOCK, CL: CLUSPRO. The suffix [N] suggests that the re-ranking was applied to the top N models from rigid docking. ZR [N]: ZDOCK+ZRANK with ZRANK re-ranking applied to top N models from ZDOCK, PF [N]: PATHDOCK+FIBERDOCK with FIBERDOCK applied on top *N* models from PATCHDOCK.

Results on the learning set

We report in Table VIII, the performance of DOCK/PIE and various flavors of DOCK/PIERR on the learning set of 640 complexes.

We observe that the combination potentials generally perform better than the atomic potential, PISA alone, on all three datasets. Besides, Table VIII suggests that the atomic potential PISA seems to recover more hits in the top 10 than the residue potential in DOCK/PIE. But it is not as sensitive as the residue potential in DOCK/PIE when it comes to solving more targets in the top 10, or top 1. In other words, the atomic potential developed here is more useful for enriching the decoy set than for sensitive prediction of hits from a small set of models.

Atomic potentials may be more sensitive to noise in the learning set. One source of noise in learning is the use of unbound complexes. In unbound docking, the input receptor and ligand chains are modeled from their homologs, and are not structurally identical to the receptor and ligand, respectively, in the bound structure. In order to test this hypothesis that atomic potentials perform better on bound complexes than unbound, we compared the performance of DOCK/PIE rigid docking with the PIE potential, that includes residue based and van der Waals terms, with DOCK/PIERR re-ranking with the atomic potential. The results in Table IX show that this hypothesis is not supported, since the atomic potential PISA has a higher percentage of solved targets for the unbound complexes than for the bound complexes. The atomic potential is also better than the residue potential on the unbound complexes. Hence we still do not know what makes atomic potentials less sensitive.

Residue potentials are possibly more robust and are better able to capture enough of the overall structural features to recognize near-natives from a small set of models. Hence we use potentials that combine atomic and residue scores, hoping that they will be more robust, correlate well with RMSD, and enrich hits in the decoy set.

Approximate run times for DOCK/PIERR for different protein sizes are shown in Table X. We estimate that other software packages we compared to in the present study are faster than DOCK/PIERR by a factor of about

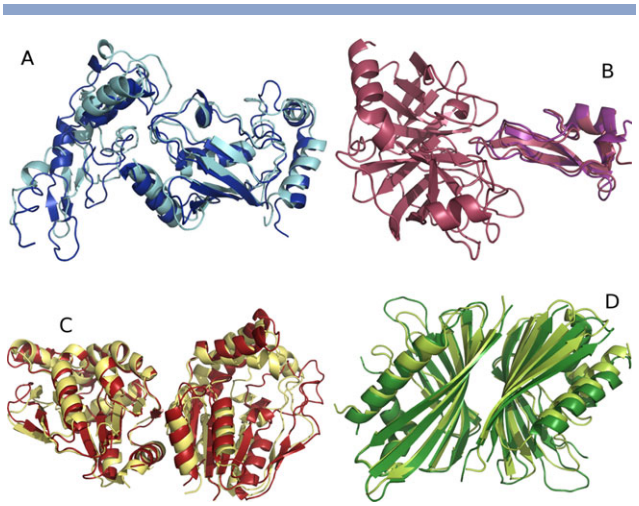

Figure 3

Structures produced by three different groups on complexes in the novel set. **A**: Native structure of 3hct (in blue) along with the best model produced for this complex, by ZDOCK+ZRANK (in cyan). Since the chains are unbound–unbound there is a slight deviation between the receptor chains in the native and model. **B**: Native structure of 3d65 (in purple) along with the best model for this enzyme–inhibitor complex by Cluspro (in raspberry). The receptor enzyme completely overlaps in this case. **C**: Native structure of 3asy (in brick red) superposed with the best model by DOCK/PIERR (in yellow). **D**: Native structure of 3rd6 (in dark green) superposed with the best model by DOCK/PIERR (in lemon yellow). DOCK/PIERR was the only method that could obtain a hit in the top 10 for the targets in (C) and (D).

10. So far we have focused our attention on getting higher accuracy and we did not focus on speed. ZDOCK is using essentially the same algorithm as DOCK/PIERR (FFT) so we are hopeful that appropriate optimization could be found. For example, DOCK/PIERR uses at present, double precision floating point number in FFT calculations while ZDOCK uses only single precision numbers.

Analysis of the new atomic potential

The complete set of parameters for the atomic potential and code for calculating energy of a structure is available on the web at http://www.ices.utexas.edu/web/dock_details.html.

On solving the linear program obtained from inequalities generated from the learning set, we can calculate for each target in the learning set, the percentage of inequalities of that target that were not satisfied by the linear programming solution. Figure 4 shows the distribution of violations among targets in the learning set. Only the 200 targets with the maximum percentage of violated inequalities are shown, as the rest of the targets have a negligible number of violations. We observe that there are a relatively small number of targets that contribute a large number of violated constraints.

Some targets can be especially hard to dock if they have a very small number of native interface contacts. We show in Figure 5, the correlation between the number of native contacts in the target and the percentage of inequalities that were violated for that target. It is observed that the targets with low number of contacts have a high percentage of violated inequalities.

To assess the extent of redundancy among the inequalities in the linear program, we calculated the cosine of the angle between any two inequality vectors (the vectors are a function of α and d and of the form $[n_{\text{misdocked}}(\alpha, d) - n_{\text{near-native}}(\alpha, d)]$) and obtained the distribution of the cosine values. We did this for three different samples of inequality vector pairs sampled at random from the inequalities in our linear program: 1500, 2500, and 3500 pairs of inequalities. Figure 6 shows the distribution of cosine values, peaked around 0.0, which shows that a significant percentage of inequalities were orthogonal to each other. This suggests that most of the constraints offer new information and are independent of each other.

Atomic and residue potentials on refined and unrefined models

Here we explore the performance of atomic potentials on rigid docking models, as opposed to refined models.

Table VII

Comparison of Number of Hits (Interface RMSD Less Than 4 Å) in the Top 1000 Models Returned by ZDOCK and DOCK/PIE Rigid Docking

Target	DOCK/PIE hits in top 1000	ZDOCK hits in top 1000
2xt4	58	26
2xty	0	0
3agx	3	24
3asy	19	9
3gt6	5	11
3lis	17	20
3m7f	0	0
3mxj	7	2
3nfy	21	31
3oa9	24	76
3p2q	3	17
3pc6	20	7
3pge	0	0
3pra	2	0
3r8c	11	6
3rd6	20	0
3rkc	33	9
3t43	7	5
3te8	0	0
3u80	0	0
3umz	61	29
3vc8	1	0
2wfx	0	0
3d65	41	35
3di3	20	17
3hct	0	2
3jrq	4	6
3l1z	6	0
3m18	11	0
3nbp	4	1
Average	13.27	11.1

Table VIII
Comparison of DOCK/PIE and DOCK/PIERR on the Learning Set of 640 Complexes

	Interface RMSD		Fraction of native contacts	
Method	Number of hits in top 10/top 1	Number of targets solved in top 10/top 1	Number of hits in top 10/top 1	Number of targets solved in top 10/top 1
DOCK/PIE Rigid Docking	1646/376	466/376	2152/400	503/400
DOCK/PIERR Re-rank with PISA	1764/334	459/334	2197/365	494/365
DOCK/PIERR Re-rank with C1	2028/410	482/410	2486/433	508/433
DOCK/PIERR Re-rank with C2	2024/411	477/411	2483/435	507/435
DOCK/PIERR Re-rank with C3	2003/413	487/413	2487/430	520/430

Table IX
Comparison of DOCK/PIE and DOCK/PIERR on 462 Bound and 178 Unbound Complexes in the Learning Set

		Interface RMSD	
Type of complexes in dataset	Method	Number of hits in top 10/top 1	Percentage of targets solved in top 10/top 1
Bound	DOCK/PIE Rigid Docking	957/278	74.1/60.4
Bound	DOCK/PIERR Re-rank with PISA	915/228	69.5/49.5
Unbound	DOCK/PIE Rigid Docking	689/98	71/55.6
Unbound	DOCK/PIERR Re-rank with PISA	849/106	78.9/60.2

If we could obtain the same performance of atomic potentials on rigid docking models as on refined models, then the extra computational expense of side chain refinement and minimization of one thousand models for each target can be avoided.

In Figure 7, we illustrate that the atomic potential PISA works best when the parameters of the potential are learnt on refined models and applied to refined models. The atomic potential has a better recognition capacity if trained and tested on refined models than when trained and tested on the rigid docking models. Interestingly, the PISA potential trained and tested on unrefined structures is much worse than residue-based rigid docking. Finally, the potential, PIE, with residue-based terms, performs worse when tested on refined structures, than when tested on unrefined structures. This is probably because it was trained on unrefined structures. The advantage of the unrefined structures is that they are more rapid to generate and consider.

Table X
Approximate Run-Times for DOCK/PIERR for Different Protein Sizes

Receptor length (number of residues)	Ligand length (number of residues)	Approximate run time in hours
105	105	1.25
202	200	1.5
418	152	4.75
272	174	5.75
554	400	9

This includes the time for rigid docking as well as refinement and re-ranking. All runs were on four nodes of a Linux cluster with eight cores each (32 cores total). Each core was an Intel Xeon X5460 processor with clock speed of 3.16 GHz. The memory size was 16 GB for each node.

Examining hit cutoffs and distance bins for the atomic potential

Tables XI–XIV show numerical experiments on different definitions of near-native and misdocked conformations and different distance bins for the ZLAB benchmark and training sets. All these different variations of the atomic potential were learnt on the training set and tested for performance on the ZLAB set. The best performing definition on the ZLAB set was chosen for the actual potential, PISA.

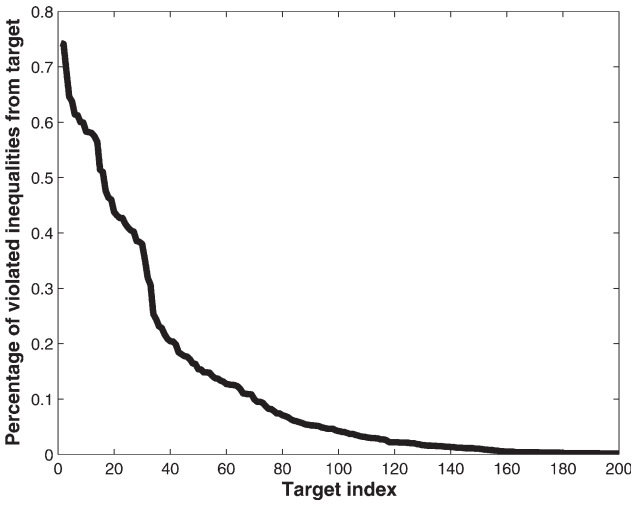

Figure 4
Percentage of violated inequalities for 200 targets of the learning set. The rest of targets are not shown as they have a negligible number of violated inequalities.

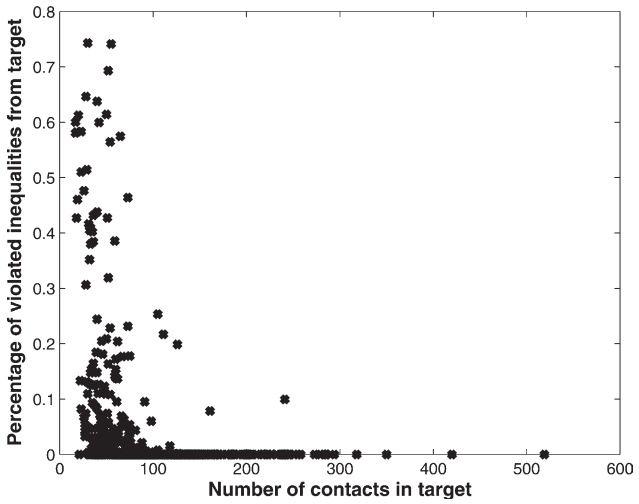

Figure 5

Percentage of violated inequalities per learning set target plotted against the number of contacts for the target.

Models were classified as near-native if they had interface RMSDs lower than the hit cutoff and were classified as misdocked models if they had an RMSD higher than the misdocked model cutoff. Near-native cutoffs of 4 Å and 2.5 Å were tried and misdocked model cutoff of 5.0, 6.0, and 7.0 A were tried and we chose near-native cutoff of 2.5 Å and misdocked model cutoff of 7 Å.

We also experimented with different distance bins for the pairwise atomic potential, PISA. We show that the distance bin 2–3.5; 3.5–5; 5–8 Å works best. Incidentally, these are also the distance bins used for detecting approximate structures for threading.[29]

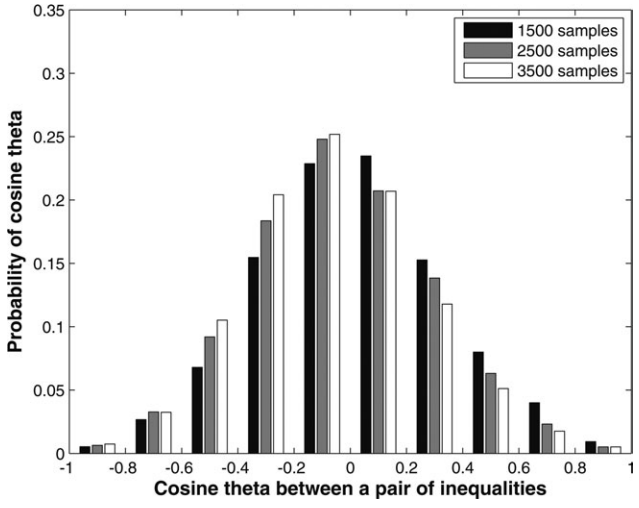

Figure 6

Probability distribution of cosine values for 1500, 2500, and 3500 pairs of inequalities sampled at random from the set of linear inequalities.

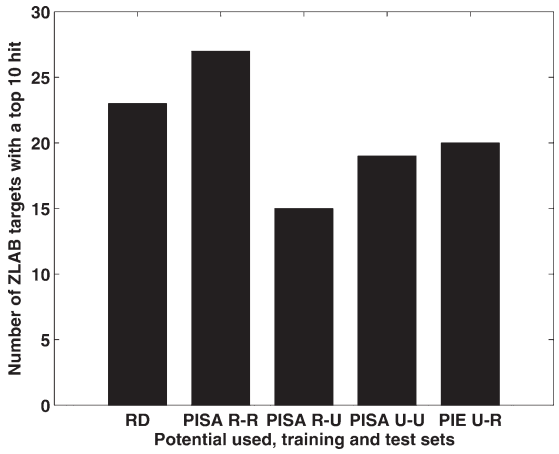

Figure 7

Comparison of performance of atomic potential PISA on refined and unrefined models. The histogram shows for each of the docking methods, the number of ZLAB targets in our test set that had a hit with interface RMSD less than 2.5 Å, in the top 10 models. Abbreviations are explained as follows: RD: rigid docking, U: unrefined, R: unrefined. PISA-R-U for example, means that the re-ranking potential was PISA, which was learnt on refined learning set models and tested on unrefined ZLAB structures.

Performance in CAPRI evaluations

Early versions of DOCK/PIERR and DOCK/PIE methods that are different from the program reported here have been tested by participation in the CAPRI evaluations. In the recent Round 26, there were two targets, T53 and T54. T53 was a protein–protein complex between artificial alpha repeat proteins REP4 and REP2. We produced one medium hit and four acceptable hits in the top 10 in the scoring round for T53, but did not produce an acceptable structure prediction of our own. Ours was one of nine groups and eight groups, out of 13, to get at least one medium hit and acceptable hit respectively, in the scoring round for T53. However other

Table XI

Comparison of Different Cutoffs for Near-Native or Misdocked Complexes on the ZLAB Benchmark of 124 Complexes

Cutoff for near-natives and misdocked models		Interface RMSD	
Near-native cutoff in A	Misdocked model cutoff in A	Number of hits in top 10/top1	Number of targets solved in top 10/top1
2.5	7.0	86/17	40/17
4.0	7.0	81/15	39/15
2.5	6.0	86/16	39/16
2.5	5.0	82/15	38/15

Table XII
Comparison of Different Cutoffs for Near-Native or Misdocked Complexes on the Learning Set of 640 Complexes

Cutoff for near-natives and misdocked models		Interface RMSD	
Near-native cutoff in A	Misdocked model cutoff in A	Number of hits in top 10/top1	Number of targets solved in top 10/top1
2.5	7.0	1764/334	459/334
4.0	7.0	1424/255	305/255
2.5	6.0	1510/280	414/280
2.5	5.0	1627/292	393/292

Table XIV
Comparison of Different Distance Bins on the Learning Set of 640 Complexes

	Interface RMSD	
Distance bins used	Number of hits in top 10/top1	Number of targets solved in top 10/top1
2–3.5; 3.5–5; 5–8 Å	1764/334	459/334
3.5–5; 5–6.5; 6.5–8 Å	1310/278	367/278
2–4.5; 4.5–6; 6–8 Å	1468/295	389/295
2–3; 3–4; 4–5; 5–8 Å	1548/272	361/272
2–3.5; 3.5–5; 5–8; 8–10 Å	1630/310	404/310

groups fared better in the prediction round for T53, with 17 out of 42 groups getting an acceptable model, nine out of 42 groups getting a medium model, and one group getting a high accuracy model. Whereas T54 was a complex between engineered neocarzinostatin and another alpha repeat protein, REP16. We did not produce any hits in the prediction or scoring round here. This target was also found to be a hard target by other groups, since no group was able to get an acceptable or better model in the scoring round for T54, and only four out of 41 groups got an acceptable model in the prediction round.

In Round 24 (the intermediate Round 25 was cancelled), there were two targets, T50 and T51. T50 was a complex of the influenza hemagglutinin bound to a designed protein. For T50, we produced one acceptable and two medium models in the server prediction, two acceptable and one medium models in the human prediction, and two acceptable models in the scoring round. Fourteen and nine other groups, out of 40, got medium and acceptable models in the prediction round for T50. Ten and six other groups out of 19 groups got an acceptable or medium model, respectively, in the scoring round for T50. T51 was a multi-domain protein for which we did not get any correct models in any of the rounds. Only 3 of 35 prediction groups and 5 of 13 scorer groups got an acceptable model for this target.

In Round 23, the targets were T47, T48, and T49. T47 was a protein–protein complex that mainly involved prediction of the interface water molecules and we did not participate in it. T48 and T49 were the same protein–protein complex starting from two different unbound structures. They involved the binding of a hetero-hexamer hydroxylase to ferredoxin. For T48, we had 1 and 2 acceptable predictions for the server and human prediction rounds, respectively. Fifteen out of 32 groups in all submitted acceptable or better models for T48. For T49, we had one acceptable prediction for the server prediction round. Fifteen out of 33 groups had an acceptable or better model for this target. The scoring for both these targets was combined, and we did not produce any acceptable model in the scoring round. Eight out of 13 scorers had produced acceptable or better models in the scoring round for T48/T49.

Round 22 consisted of only one target, T46, an unbound–unbound protein–protein complex. We obtained one acceptable prediction for the scoring round for this target and no acceptable predictions for the prediction rounds. This was a hard target as only two out of 40 groups produced acceptable models for this target. Eight out of 16 scorers produced an acceptable model in the scoring round.

The CAPRI performance is summarized in Table XV.

Table XIII
Comparison of Different Distance Bins on the ZLAB Benchmark of 124 Complexes

	Interface RMSD	
Distance bins used	Number of hits in top 10/top1	Number of targets solved in top 10/top1
2–3.5; 3.5–5; 5–8 Å	86/17	40/17
3.5–5; 5–6.5; 6.5–8 Å	86/13	39/13
2–4.5; 4.5–6; 6–8 Å	85/13	38/13
2–3; 3–4; 4–5; 5–8 Å	83/16	36/16
2–3.5; 3.5–5; 5–8; 8–10 Å	84/18	38/18

CONCLUSION

We have introduced an improvement to docking algorithms by introducing a new atomic potential and refinement and ranking algorithms. We show by extensive tests on three datasets of complexes that our methods outperform slightly, other state-of-the-art docking packages. We also observe that coarse grained potentials are more robust to noisy structures produced by unbound docking. Nevertheless, we show that atomic and residue potentials capture different signals, and hence their combination works better than either of them individually. However, the success rate of docking software even after refinement and improved re-ranking functions is still between 30 and 50%. Since rigid docking is an inexpensive approach to the problem, it makes sense to try rigid docking first. Furthermore, examining our training set for unbound complexes we found that the average TM score between

Table XV
Performance of DOCK/PIE in CAPRI Assessments

Target	Number of predictor groups with hits	Number of scorer groups with hits	DOCK/PIE server: number of top 10 hits	DOCK/PIE human: number of top 10 hits	DOCK/PIE scorer: number of top 10 hits
38	0	0	—	0	—
39	2	—	—	0	—
40	30	10	—	1 (57)	8 (6)
41	26	12	—	0	1 (40)
42	13	—	—	0	—
46	2	8	0	0	1 (1)
48	15	—	1 (13)	2 (33)	—
49	15	8	1 (24)	0	0
50	18	12	3 (4)	3 (8)	2 (15)
51	3	5	0	0	0
53	20	11	0	0	5 (7)
54	4	0	0	0	0

Only targets we participated in are shown. Hits refer to models that were of acceptable quality or better. "—" denotes no participation for that target, while 0 indicates no acceptable models were found for that target. The numbers in parenthesis indicate the rank of our best model.

the templates and the native conformations of the chains in the complex was high and about 0.89. Such high similarity supports the use of rigid docking. A promising area seems to be the development of scoring functions that can recognize near-native models from a small set of models. One could envision designing multi-body potentials, orientation-based potentials and potentials that are based on hydrogen bonding interactions to capture more structural features that can potentially lead to more accurate scoring functions and improve the success of computational docking procedures.

ACKNOWLEDGMENTS

The authors wish to thank David Hall for helping them use the Cluspro server to analyze their datasets.

REFERENCES

1. Gray JJ. High-resolution protein-protein docking. Curr Opin Struc Biol 2006;16:183–193.
2. Janin J, Bahadur RP, Chakrabarti P. Protein–protein interaction and quaternary structure. Q Rev Biophys 2008;41:133–180.
3. Smith GR, Sternberg MJE. Prediction of protein-protein interactions by docking methods. Curr Opin Struc Biol 2002;12:28–35.
4. Tovchigrechko A, Vakser IA. Development and testing of an automated approach to protein docking. Proteins 2005;60:296–301.
5. Ravikant DVS, Elber R. Energy design for protein–protein interactions. J Chem Phys 2011;135:065102.
6. Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, Li KY, Zheng JF, Vakili P, Paschalidis IC, Vajda S. Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SOU, and stability analysis in CAPRI rounds 13–19. Proteins 2010;78:3124–3130.
7. Chen R, Li L, Weng ZP. ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003;52:80–87.
8. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 2003;331:281–299.
9. Wang C, Bradley P, Baker D. Protein–protein docking with backbone flexibility. J Mol Biol 2007;373:503–519.
10. Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. Lect Notes Comput Sci 2002;2452:185–200.
11. Wang C, Schueler-Furman O, Baker D. Improved side-chain modeling for protein–protein docking. Protein Sci 2005;14:1328–1339.
12. Lorenzen S, Zhang Y. Monte Carlo refinement of rigid-body protein docking structures with backbone displacement and side-chain optimization. Protein Sci 2007;16:2716–2725.
13. Li L, Chen R, Weng ZP. RDOCK: refinement of rigid-body protein docking predictions. Proteins 2003;53:693–707.
14. Pierce B, Weng ZP. ZRANK: reranking protein docking predictions with an optimized energy function. Proteins 2007;67:1078–1086.
15. Andrusier N, Nussinov R, Wolfson HJ. FireDock: fast interaction refinement in molecular docking. Proteins 2007;69:139–159.
16. Mashiach E, Nussinov R, Wolfson HJ. Fiber Dock: flexible induced-fit backbone refinement in molecular docking. Proteins 2010;78:1503–1519.
17. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res 2004;32:W96–W99.
18. Masone D, de Vaca IC, Pons C, Recio JF, Guallar V. H-bond network optimization in protein-protein complexes: are all-atom force field scores enough? Proteins 2012;80:818–824.
19. Liang SD, Wang GC, Zhou YQ. Refining near-native protein-protein docking decoys by local resampling and energy minimization. Proteins 2009;76:309–316.
20. Hwang H, Pierce B, Mintseris J, Janin J, Weng ZP. Protein–protein docking benchmark version 3.0. Proteins 2008;73:705–709.
21. Krivov GG, Shapovalov MV, Dunbrack RL. Improved prediction of protein side-chain conformations with SCWRL4. Proteins 2009;77:778–795.
22. Guharoy M, Janin J, Robert CH. Side-chain rotamer transitions at protein–protein interfaces. Proteins 2010;78:3219–3225.
23. Simmerling CL, Elber R. Computer determination of peptide conformations in water—different roads to structure. Proc Natl Acad Sci USA 1995;92:3190–3193.
24. Jorgensen WL, Tiradorives J. The Opls potential functions for proteins—energy minimizations for crystals of cyclic-peptides and crambin. J Am Chem Soc 1988;110:1657–1666.
25. Murphy J, Gatchell DW, Prasad JC, Vajda S. Combination of scoring functions improves discrimination in protein–protein docking. Proteins 2003;53:840–854.
26. Pierce B, Weng ZP. A combination of rescoring and refinement significantly improves protein docking performance. Proteins 2008;72:270–279.

27. Vreven T, Hwang H, Weng ZP. Integrating atom-based and residue-based scoring functions for protein–protein docking. Protein Sci 2011;20:1576–1586.
28. Ravikant DVS, Elber R. PIE-efficient filters and coarse grained potentials for unbound protein–protein docking. Proteins 2010;78:400–419.
29. Qiu J, Elber R. Atomically detailed potentials to recognize native and approximate protein structures. Proteins 2005;61:44–55.
30. Wagner M, Meller J, Elber R. Large-scale linear programming techniques for the design of protein folding potentials. Math Program 2004;101:301–318.
31. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
32. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 2005;33:2302–2309.
33. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
34. Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, Pieper U, Stuart AC, Marti-Renom MA, Madhusudhan MS, Yerkovich B, Sali A. Tools for comparative protein structure modeling and analysis. Nucleic Acids Res 2003;31:3375–3380.
35. Mendez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins 2003;52:51–67.