PROTEINS: Structure, Function, and Bioinformatics 65:538–548 (2006)

ROSETTALIGAND: Protein–Small Molecule Docking with Full Side-Chain Flexibility

Jens Meiler[1]* and David Baker[2]*
[1]*Vanderbilt University, Department of Chemistry, Center for Structural Biology, Nashville, Tennessee*
[2]*University of Washington, Department of Biochemistry, Seattle, Washington*

ABSTRACT **Protein–small molecule docking algorithms provide a means to model the structure of protein–small molecule complexes in structural detail and play an important role in drug development. In recent years the necessity of simulating protein side-chain flexibility for an accurate prediction of the protein–small molecule interfaces has become apparent, and an increasing number of docking algorithms probe different approaches to include protein flexibility. Here we describe a new method for docking small molecules into protein binding sites employing a Monte Carlo minimization procedure in which the rigid body position and orientation of the small molecule and the protein side-chain conformations are optimized simultaneously. The energy function comprises van der Waals (VDW) interactions, an implicit solvation model, an explicit orientation hydrogen bonding potential, and an electrostatics model. In an evaluation of the scoring function the computed energy correlated with experimental small molecule binding energy with a correlation coefficient of 0.63 across a diverse set of 229 protein– small molecule complexes. The docking method produced lowest energy models with a root mean square deviation (RMSD) smaller than 2 Å in 71 out of 100 protein–small molecule crystal structure complexes (self-docking). In cross-docking calculations in which *both* protein side-chain and small molecule internal degrees of freedom were varied the lowest energy predictions had RMSDs less than 2 Å in 14 of 20 test cases. Proteins 2006;65:538–548.** **© 2006 Wiley-Liss, Inc.**

Key words: docking; protein–ligand docking; binding energy; Monte Carlo minimization; ROSETTA

INTRODUCTION

Protein–small molecule (referred to as "ligand" in what follows) interactions play central roles in numerous basic processes in life, such as enzyme catalysis, activation by naturally occurring ligands, and inhibition by human-designed drugs. Thus our capability of modeling such interactions at atomic resolution is crucial to enhance our understanding of biochemistry.

A large number of docking programs have been developed in the last 20 years based on a variety of search algorithms.[1,2] The use of such programs in conjunction with one or more scoring functions to evaluate and rank potential ligands from chemical collections is a standard step in virtual drug screening. While several successful applications of this methodology have been described in recent publications,[3,4] frequently protein flexibility is neglected.[5] While this approach is suitable for rapid virtual screening, inclusion of protein flexibility is needed if the protein–ligand interface is to be modeled in atomic detail. Side-chain conformational changes frequently occur upon ligand binding. Hence side-chain coordinates taken from a complex with a different ligand, an unbound structure, or a homology model can be inaccurate.

Docking programs seek to identify the lowest free energy pose of the ligand in the protein binding site. In screening the goal is to identify the ligand with the highest binding affinity.[5] Currently, docking and screening appear to be best carried out with different methods: DOCK,[6] AUTODOCK,[7,8] FLEXX,[9] and GOLD[10] are widely used docking programs.[4,5,7,11,12]

A wide variety of empirical scoring functions have been used for virtual screening, the best of which include X-SCORE,[13] DRUGSCORE,[14] CHEMSCORE,[15,16] and PLP.[17] The different aims in docking and screening justify usage of different scoring methods. However, because both searches are driven by the same biophysics, a method which mimics nature should perform well in both high-resolution docking and ranking in screening.

In recent years, several docking algorithms have been reported that include protein flexibility. SLIDE[18] captures small side-chain motions without rotamer changes and inclusion of side-chain flexibility in ICM[19–21] was

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: Human Frontier Science Program (HFSP); Grant sponsor: HHMI; Grant sponsor: Protein Design Project (DARPA).

*Correspondence to: Jens Meiler, Vanderbilt University, Department of Chemistry, Center for Structural Biology, Nashville, TN 37235-8725. E-mail: jens@jens-meiler.de or David Baker, University of Washington, Department of Biochemistry, J Wing, Health Sciences Building, Box 357350 Seattle, WA 98195.
E-mail: dabaker@u.washington.edu

Received 3 November 2005; Revised 18 March 2006; Accepted 4 May 2006

Published online 13 September 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21086

shown to improve docking for protein kinases.[22] FDS,[23] SKELGEN,[24] and GLIDE[25] are most similar to our approach in using rotamer libraries to represent side-chain flexibility. Finally, QXP[26] allows for small protein structural changes during an energy minimization and FLEXE[27] docks small molecules into ensembles of protein structures that represent protein flexibility.

ROSETTADOCK is a protein–protein docking algorithm that starts with a low-resolution search followed by a high-resolution refinement stage in which side-chain and rigid body degrees of freedom are optimized simultaneously using a Monte Carlo minimization protocol.[28] The performance of ROSETTADOCK was recently improved by introducing a gradient-based minimization step during the cycling between rotamers to allow efficient sampling of off rotamer conformations.[29] While the algorithm was already reasonably successful in earlier CAPRI experiments,[30] with the improved treatment of side-chain flexibility predictions of unprecedented accuracy were made in the recent CAPRI experiment.[31]

In this work, we extend the ROSETTADOCK approach to protein–ligand docking. Full side-chain flexibility is achieved by extending the repacking methodology introduced in ROSETTADESIGN[32–35] to repack protein–ligand interfaces. We report the performance of the method in extensive self-docking and cross-docking benchmarks. In the self-docking benchmark, ligand conformation and protein backbone remain unaltered in order to evaluate the ability of our method to simultaneously optimize protein side-chain degrees of freedom and ligand orientation. In the second cross-docking benchmark, however, ligand and backbone flexibility are included to test the capabilities of the method thoroughly. Ligand flexibility is represented by a conformational ensemble excluding the native bound conformation. Backbone flexibility was taken into account by using multiple crystal structures of the proteins. We also evaluate the ability of our energy function to reproduce the experimental binding energies of a large set of protein–ligand complexes obtained from the ligand–protein database (LPDB).[36]

MATERIALS AND METHODS

The algorithms are written in C++ and implemented in the ROSETTA package as ROSETTALIGAND mode, which can be combined with ROSETTADOCK and ROSETTADESIGN.

Atom Types

To allow modeling of most organic molecules, including protein cofactors and drugs, as well as frequently occurring metal ions in proteins the atoms and ions F, Cl, Br, I, P, Zn^{2+}, Fe^{2+}, Fe^{3+}, Mg^{2+}, Ca^{2+}, Na^{+}, K^{+} were introduced. The parameters for modeling VDW interactions were taken from the CHARMM27 force field for the halogen atoms and P and from the MM3 force field for the metal ions. The volumes and free energies as necessary to model solvation according to Lazaridis–Karplus[37] were estimated from the atom/ion radii and similarities to the classical atom types that occur in proteins (Table S1).

Docking Monte Carlo Minimization Protocol

Focus was not put on the actual identification of potential binding sites for small molecules because a large number of algorithms are available for this purpose.[5,11] Rather, the exact prediction of the protein conformation when binding the ligand was the objective of this work. Here we believe lie the shortcomings of many of the currently used docking tools for ligands and here the sampling of protein degrees of freedom can add most to the field.

The Monte Carlo minimization protocol (cf. Fig. 1) starts from a random starting position and orientation of the ligand in the binding site of the protein. The ligand center of mass was placed randomly in a cube of $(10\ \text{Å})^3$, allowing complete reorientation of the ligand. Each Monte Carlo minimization cycle consists of the following three steps: (1) The position of the ligand is perturbed by random translations of mean 0.1 Å in each direction and by random rotations of mean 0.05° around

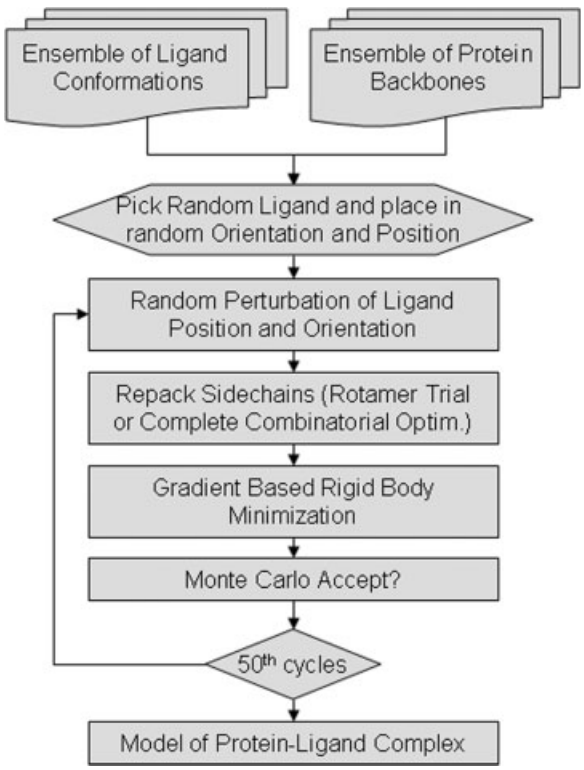

Fig. 1. Flow chart diagram of the high-resolution docking protocol. First the ligand is placed in a random position and orientation into the binding site of interest, requiring only that the backbone of the protein and the ligand non-hydrogen atoms do not clash. Fifty cycles of the Monte Carlo minimization protocol including small perturbations of the ligand pose, side-chain repacking, and gradient minimization as described in the methods section are carried out. This protocol is repeated *N* times. *N* is between 1000 and 5000 depending on the size of the ligand, its flexibility (and therefore the size of the conformational ligand ensemble), and the binding site.

each axis; (2) side-chain conformations are repacked using either rotamer trials or a full combinatorial search as described below; (3) the rigid body orientation and side-chain χ angles of the ligand are optimized using the gradient based Davidson–Fletcher–Powell algorithm. The move (steps 1–3) is accepted based on the difference of staring and final energy according to the standard Metropolis criterion probability $P = \min\{1,\exp[-(E_{\text{final}} - E_{\text{start}}/kT]\}$ with kT set to 2 kcal according to the Rosetta energy function. The move is always accepted if the energy decreases, if the energy increases the acceptance probability decreases; because minimization is carried out at each step each move consists of a transition between local minima on the free energy landscape. Each docking trajectory consists of 50 of these Monte Carlo minimization cycles. No simulated annealing was carried out. Between 1000 and 5000 trajectories are computed for each docking experiment.

Rotamer Trials and Side-Chain Repacking

A backbone-dependent rotamer library (http://dunbrack.fccc.edu/bbdep)[38,39] supplemented with additional rotamers for the side-chain dihedral angles χ_1 and χ_2 was used in all calculations. The side-chain optimization algorithms were previously described[28] except that the total energy was modified to include all intraprotein interactions and protein–ligand interactions. After most steps side-chain conformations were optimized by successively substituting each rotamer at each position and using quasi-Newton minimization to refine their torsion angles (rotamer trials with minimization); a full, combinatorial rotamer optimization was performed only once every eight cycles.

Force Field

The force field describing the interactions between ligand and protein is comprised of (1) a standard 12–6 Lennard Jones potential to model attractive interactions ($E < 0$) with van der Waals radii and well depths from the CHARMM27 parameter set; (2) a repulsive term that connects in amplitude and slope with the 12–6 potential at $E = 0$ and then ramps linearly until the two atoms are 0 Å apart (this is less repulsive than a 12–6 potential and compensates to some extent for the use of a fixed backbone and rotamer set); (3) a solvation term similar to the Lazaridis–Karplus implicit solvation model[37] for proteins; (4) an explicit hydrogen bonding potential[40]; (5) a Coulomb model with a distance-dependent dielectric constant[41] using partial charges from the CHARMM27 force field.[41–43] The partial charges inside the ligand were linearly scaled to reproduce the total charge of the small molecule. This rather simple model is pairwise additive which allows rapid computation. The utilized force field parameters are summarized in Table I. A generalized Born model was also tested, however results did not profit from its introduction while the computation time increased significantly. Inside the protein the electrostatics is represented using a pair potential (6) derived from the protein data bank (PDB) statistics.[44] Backbone-dependent internal free energies (7) of the rotamers are estimated from PDB statistics.[33,45]

TABLE I. Comparison of the Weights as Determined for Ligand Binding Sites in Proteins with the Standard Weights Used for Proteins Only[52]

	Ligand	Standard
LJ-atractive	0.80	0.80
LJ-repulsive	0.60	0.73
Solvation	0.50	0.52
Hydrogen bonding	1.20	1.39
Pair energy	0.50	0.27
Rotamer probability	0.32	0.32
Phi psi probability	0.32	0.41
LJ-atractive (ligand)	0.80	N/A
LJ-repulsive (ligand)	0.60	N/A
Solvation (ligand)	0.50	N/A
Hydrogen bonding (ligand)	1.20	N/A
Electrostatics (ligand)	0.25	N/A

Weights

A database of 100 native protein–ligand complexes was compiled for testing the method. The weights for linearly combining the energy terms to build the composite energy function were initially taken from the protein force field and were fitted by maximizing the correlation between the composite energy and the square root of the RMSD in Å for sets of 1000 randomly generated docking poses as well as 50 native-close poses for each of the 100 complexes. The weights do not change significantly when different subsets of 20 complexes are used in this fitting procedure. Since the resulting weights are very close to the original protein weights, they are robust to changes in the composition of the training dataset, and were trained on random poses rather than docked models. We assume them to be sufficiently independent from the dataset to be trusted as independent. The protein–protein and protein–ligand force fields are identical except for the replacement of the intra-protein pair potential (6) with the Coulomb energy (5) in the protein–ligand interface. The weights are encouragingly similar to the standard ROSETTA protein forces, suggesting that the underlying physical chemistry is modeled reasonably well.

Ligand Flexibility

Ligands were represented as a set of discrete conformations. To generate these conformations, first all torsional degrees of freedom in the ligand were identified. For each of these torsion angles short list of likely conformations was compiled from atom type and hybridization state of the linked atoms. For a dihedral angle between two sp^3 hybridized atoms (e.g., $-CH_2-CH<$), three states — 180°, 60°, and −60° — were considered; for a torsion angle between two sp^2 hybridized atoms

(e.g., —CH=CH—) two states — 180° and 0° — were considered; for all other combinations 12 states (180°, 150°, . . . , −150°) were considered. To build a ligand conformation, each torsion angle was put in one of the considered states. Conformations with internal clashes of ligand atoms were not considered. The conformation of closed ring systems was not altered. No internal ligand energy was evaluated and no energy minimization was applied.

An ensemble of ligands was built using the following protocol: (1) a random non-clashing conformation was generated and accepted as first member of the ensemble; (2) 10 new random conformations were generated and their RMSD to all accepted members of the ensemble was evaluated; (3) the conformation with the largest RMSD was added to the accepted set of conformations. Steps (2) and (3) were repeated until 10 conformations were accepted into the ensemble. In step (2) only conformations with a RMSD larger 1 Å to all accepted conformations were considered. If no such conformation could be built, the algorithm was stopped and the conformational ensemble for this particular ligand was left incomplete.

This procedure ensures that the ensembles spun a maximal range of the conformational space. The minimal RMSD in the conformational ensemble to the crystal structure conformation was 0.43 Å on average (Table S2). Hence this procedure samples the ligand conformational space sufficiently dense for these examples. Larger ligands will require additional sampling. The sampling of ligand conformations by choosing local low energy torsion angles resembles the sampling of protein side-chain conformations by a rotamer search.

The crystal structures of protein–ligand complexes were obtained from the PDB (http://www.rcsb.org/pdb/).[46,47] Additional protein chains that do not interact with the ligand, the binding site water molecules, and additional ligands in alternative binding sites, were removed prior to calculation. Hydrogen atoms were added using standard bond lengths and bond angles. sp^3-Nitrogen atoms were generally assumed to be protonated and positively charged; carboxyl groups were assumed deprotonated and negatively charged. Metal ions, sulfate, and phosphate ions were assumed to carry their net charge. The assigned bond states and charges for all ligands were checked by visual inspection.

The calculations were performed on a 32-node Linux personal computer (PC) cluster each with two Intel Pentium 4 processors running at 1.9 GHz and 512 GB memory. Depending on the size of the protein–ligand complex, the computation time was between 5 and 10 min per model on one processor. Building 500 models each for 10 ligand conformations on the cluster took approximately 4 h.

RESULTS AND DISCUSSION

The major focus of this work is to adapt ROSETTADOCK to simultaneously optimize side-chain and rigid body degrees of freedom in protein–ligand interfaces. The computer program ROSETTA hosts the fundamental protein structure prediction, docking, and design algorithms for proteins as used throughout this work. The program was expanded and modified to make it capable of handling small organic molecules. This comprised the inclusion of all relevant atom types for modeling organic molecules as well as metals, the modification of the docking protocol to cope with ligands, and the adaptation of the force field.

Docking

The docking protocol is illustrated in Figure 1 and was derived from the ROSETTADOCK protocol for protein–protein docking.[28] In the first experiment the ability of our approach to model protein conformational changes rather than ligand flexibility was evaluated. The ligand was kept rigid in its bound conformation and a single rigid protein backbone was used. All amino acid side-chains in the ligand binding site as well as in the second shell were allowed to alter their conformation. The discrete set of conformations allowed for these side-chains was taken from Dunbrack's updated rotamer library (http://dunbrack.fccc.edu/bbdep).[38]

For a database of 100 native protein–ligand complexes, 5000 models were generated as discussed in the methods section. In 71 of 100 cases the lowest energy model had an RMSD smaller than 2 Å, indicating that the correct ligand pose was not only sampled but also detected based on its low energy (Table II, Fig. S1, and Fig. S2). For 18 additional cases, at least one of the 10 lowest energy models had an RMSD smaller than 2 Å. For 11 protein–ligand complexes no low energy pose had an RMSD smaller 2 Å. In most of these cases the native complex is not recognized as a particularly low energy pose even after minimization.

The success rate of the self-docking experiments, 71% to 80%, is slightly below the best rates reported for other methods discriminating near-native from non-native models with ranges from 80% to 90%.[5] However, one has to keep in mind that the protein side-chain structural space is sampled in this test together with the ligand pose, but the ligand conformational space is not. Hence it is difficult to compare the actual sizes of the search spaces. While a RMSD < 2 Å is counted as success in our experiment as well as in the literature,[5] in our case the RMSD includes binding site side-chains as well as hydrogen atoms and is therefore more sensitive to structural changes of the protein and ligand. In some cases, reported success rates refer to scoring of existing ensembles of models enriched with low RMSD models, which makes recognition much easier.[5] GLIDE, GOLD, and ICM were recently compared on a different, more drug-like benchmark set with success rates of 61%, 48%, and 45%, respectively.[11]

Cross-Docking

This experiment is designed to mimic the situation in drug discovery where a crystal structure of the protein

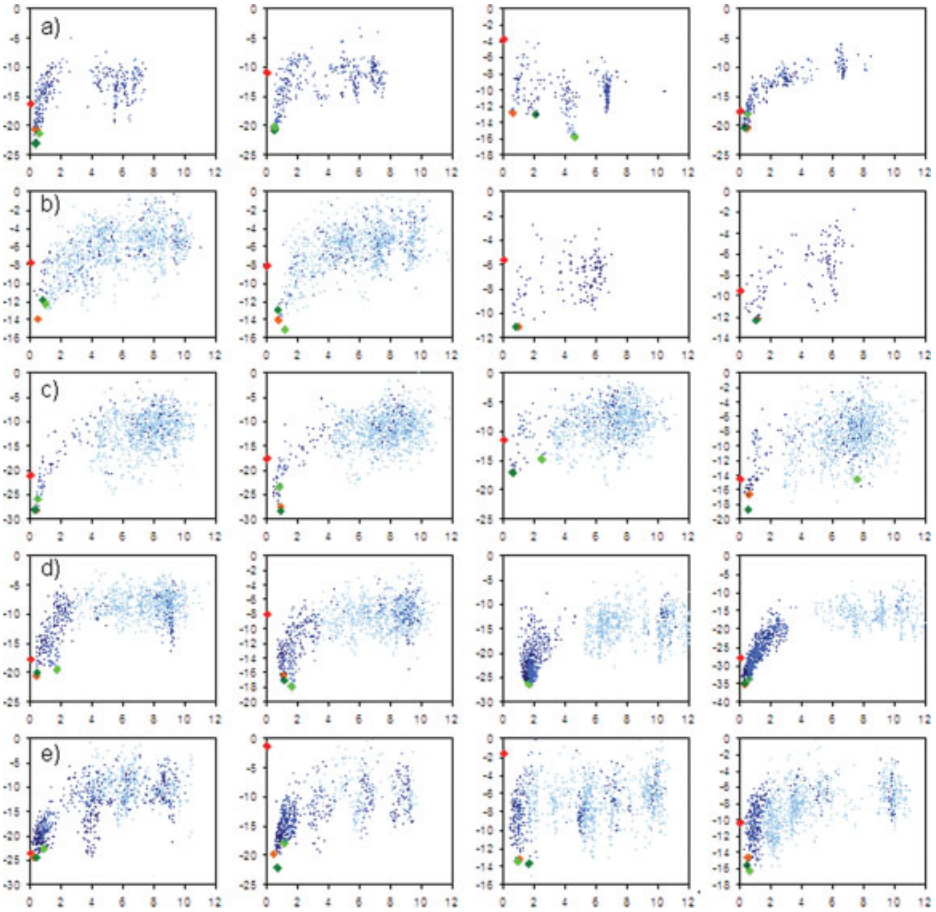

Fig. 2. Results of cross-docking benchmark of pairs of complexes A/B: (**a**) 1aq1/1dm2, (**b**) 1dbj/2dbl, (**c**) 1dwc/1dwd, (**d**) 1fm9/2prg, (**e**) 1p8d/1pq6, (**f**) 1p8d/1pqc, (**g**) 1ppc/1pph, (**h**) 1pq6/1pqc, (**i**) 2ctc/7cpa, (**j**) 4tim/6tim: The diagrams show from the left to the right ligand A docked in protein A, ligand A docked in protein B, ligand B docked in protein A, and ligand B docked in protein B. The ROSETTALIGAND binding energy is shown on the *y*-axis and the RMSD in Ånstroms on the *x*-axis. In all diagrams the energy of the crystal structure (red diamond), the minimized native structure (orange diamond), the lowest energy model obtained utilizing the crystal structure conformation of the ligand (dark green diamond), and the lowest energy model obtained including ligand flexibility (light green diamond) are shown. All other models obtained from the crystal structure conformation of the ligand are shown as dark blue points and models obtained including ligand flexibility are marked with light blue points.

with a ligand bound in the binding site is frequently known. In the course of developing inhibitors the binding of different ligands to the same protein needs to be evaluated accounting for potential changes in the protein side-chain or backbone structure.

Ten proteins were selected that were crystallized with two different ligands and each ligand was docked into the backbone of the original crystal structure (self-docking) and the structure determined with the other ligand (cross-docking). Ligand flexibility was introduced by generating diverse conformational ensembles with up to 10 ligand conformations by systematically altering torsion angles (see Methods section). Two hundred fifty models were obtained for each of the ligand conformations in both protein structures. For 14 of the 20 cross-docking cases and 16 out of 20 self-docking cases, the lowest energy model had an RMSD smaller than 2 Å. The average RMSD of the successful cross-docking models (1.3 Å) was somewhat higher than that of the self-docking experiment (0.8 Å). When comparing the results for the rigid ligand with flexible ligand docking, the success rate decreases from 35 to 30 out of 40 examples. The average RMSD of the lowest energy models in the successful runs increases from 0.7 Å to 1.0 Å. Figure 2 and Table II summarize these results.

The modeling of side-chain conformations is of particular interest for our docking algorithm. In Figure 3 we analyze side-chain conformations in four representative lowest scoring models from the cross-docking benchmark in detail. Frequently, side-chain conformations change little in structure of the same protein crystallized with two different ligands. In the immunoglobulin test case [Fig. 3(a)], the ligand binding interface is formed by several aromatic amino acids whose confirmations change little upon

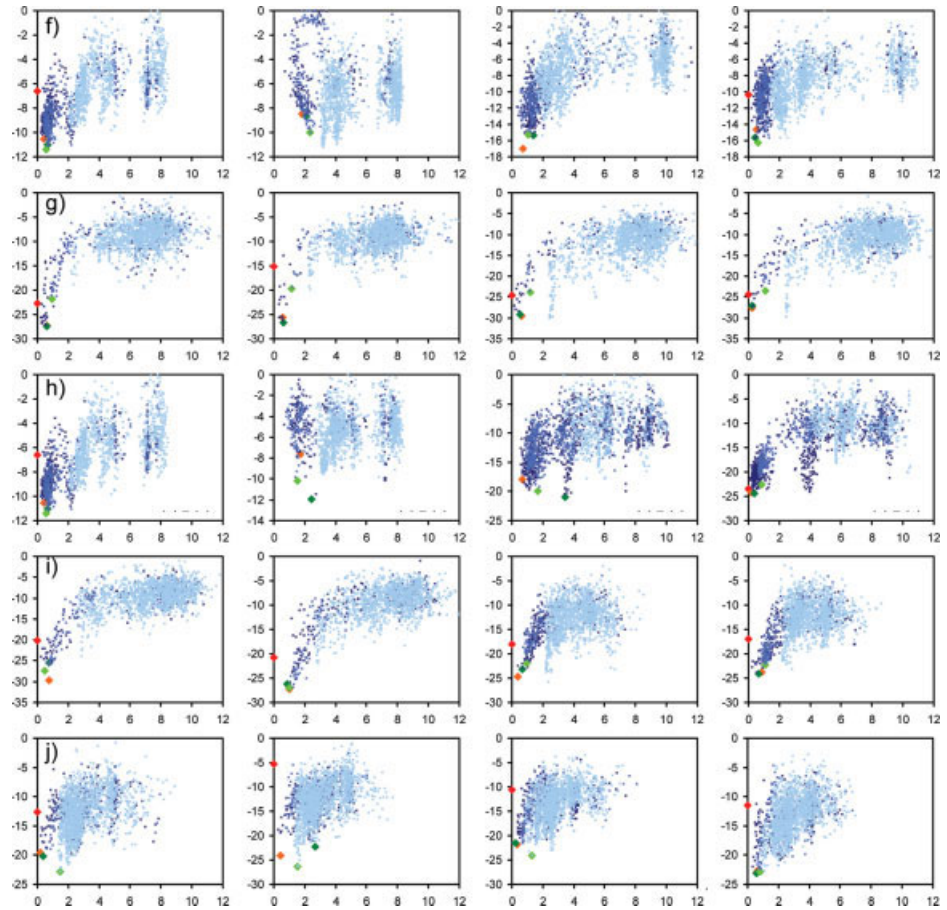

Figure 2. (Continued.)

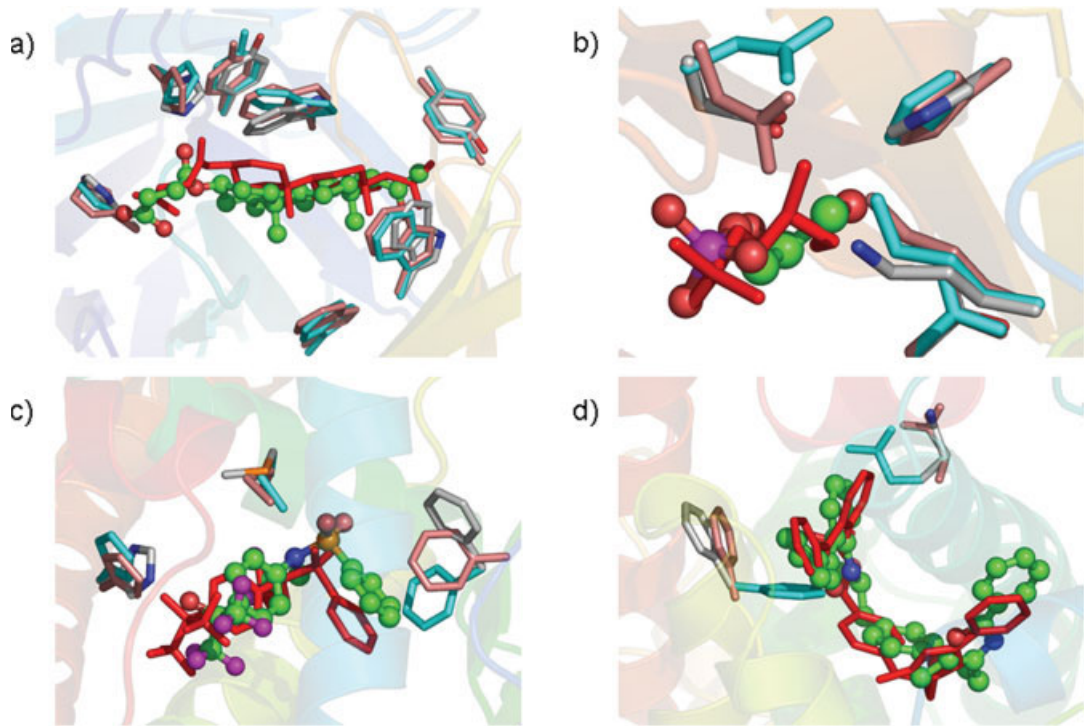

Figure 3.

TABLE II. Benchmark Results for Docking Rigid Ligand Conformations to Their Protein Crystal Structures Including Sampling of Protein Side-Chain Conformations (Self-Docking)

ID	PDB	RMSD Å[a]	Rank #/Å[b]	ID	PDB	RMSD Å[a]	Rank #/Å[b]	ID	PDB	RMSD Å[a]	Rank #/Å[b]
1	1a07	10.6	99[c]/1.8	**35**	**1epo**	**0.6**	**1/0.6**	**69**	**1qbt**	**0.4**	**1/0.4**
2	*1a1b*	*2.5*	*2/1.4*	**36**	**1ets**	**0.2**	**1/0.2**	**70**	**1qbu**	**0.5**	**1/0.5**
3	**1a1e**	**0.7**	**1/0.7**	**37**	**1ett**	**0.6**	**1/0.6**	**71**	**1ql7**	**0.5**	**1/0.5**
4	**1a28**	**0.3**	**1/0.3**	*38*	*1f0r*	*2.2*	*2/2.0*	**72**	**1qpe**	**0.4**	**1/0.4**
5	**1a6w**	**1.1**	**1/1.1**	**39**	**1f0s**	**0.6**	**1/0.6**	**73**	**1rnt**	**1.4**	**1/1.4**
6	*1a9u*	*7.4*	*3/0.7*	*40*	*1fax*	*10.0*	*2/1.2*	**74**	**1sln**	**0.4**	**1/0.4**
7	**1abf1**	**1.2**	**1/1.2**	**41**	**1fbl**	**0.6**	**1/0.6**	**75**	**1srg**	**0.7**	**1/0.7**
8	*1abf2*	*4.1*	*3/1.8*	**42**	**1hiv**	**0.3**	**1/0.3**	76	1srh	8.4	28/1.7
9	**1apu**	**0.5**	**1/0.5**	**43**	**1hos**	**0.6**	**1/0.6**	77	1tlp	10.2	5/0.8
10	**1b6n**	**0.7**	**1/0.7**	**44**	**1hpv**	**1.1**	**1/1.1**	**78**	**1tmn**	**0.2**	**1/0.2**
11	**1b9v**	**0.8**	**1/0.8**	45	1hsb	3.3	13/0.5	**79**	**1tnh**	**0.5**	**1/0.5**
12	**1bl7**	**0.8**	**1/0.8**	**46**	**1htf**	**0.4**	**1/0.4**	**80**	**1uvs**	**0.6**	**1/0.6**
13	**1byg**	**0.8**	**1/0.8**	**47**	**1htf**	**0.4**	**1/0.4**	**81**	**1uvt**	**0.3**	**1/0.3**
14	**1c2t**	**0.6**	**1/0.6**	**48**	**1hvr**	**0.2**	**1/0.2**	*82*	*1vgc*	*12.1*	*2/1.4*
15	**1c5x**	**0.6**	**1/0.6**	**49**	**1hyt**	**0.6**	**1/0.6**	*83*	*1ydr*	*10.5*	*8/0.6*
16	*1c83*	*4.5*	*2/1.3*	*50*	*1icn*	*12.6*	*2/0.5*	*84*	*1yds*	*11.5*	*6/0.6*
17	*1cin*	*6.7*	*3/0.4*	51	1ivc	4.1	42/1.9	**85**	**1ydt**	**1.2**	**1/1.2**
18	**1ckp**	**0.5**	**1/0.5**	**52**	**1ivd**	**1.9**	**1/1.9**	**86**	**2aad**	**1.1**	**1/1.1**
19	**1cps**	**1.4**	**1/1.4**	**53**	**1ivq**	**0.6**	**1/0.6**	87	2fox	4.4	91/2.9
20	*1cqp*	*2.5*	*5/0.5*	**54**	**1jap**	**0.5**	**1/0.5**	*88*	*2ifb*	*3.6*	*2/0.7*
21	**1ctt**	**0.6**	**1/0.6**	**55**	**1lic**	**0.6**	**1/0.6**	**89**	**2mip1**	**0.5**	**1/0.5**
22	**1d0l**	**0.4**	**1/0.4**	**56**	**1lyb**	**0.3**	**1/0.3**	**90**	**2mip2**	**0.5**	**1/0.5**
23	**1d4p**	**0.4**	**1/0.4**	**57**	**1mmb**	**0.6**	**1/0.6**	**91**	**2qwk**	**0.4**	**1/0.4**
24	*1dd7*	*2.9*	*7/0.7*	**58**	**1mnc**	**0.6**	**1/0.6**	**92**	**2tmn**	**0.8**	**1/0.8**
25	*1dg5*	*5.3*	*2/0.9*	**59**	**1mts**	**1.3**	**1/1.3**	**93**	**2ypi**	**0.8**	**1/0.8**
26	**1dhf**	**0.3**	**1/0.3**	**60**	**1mtw**	**1.3**	**1/1.3**	**94**	**3cpa**	**1.5**	**1/1.5**
27	**1dmp**	**0.3**	**1/0.3**	61	1mup	4.6	67/1.6	*95*	*3nos*	*6.2*	*2/1.1*
28	1dy9	3.0	23/1.7	**62**	**1ngp**	**0.4**	**1/0.4**	**96**	**4er2**	**0.4**	**1/0.4**
29	**1ejn**	**0.3**	**1/0.3**	*63*	*1nsd*	*6.9*	*10/1.5*	**97**	**4lbd**	**0.5**	**1/0.5**
30	**1ela**	**0.5**	**1/0.5**	**64**	**1ppc**	**0.5**	**1/0.5**	**98**	**4tpi**	**0.3**	**1/0.3**
31	1elb	5.1	99[c]/**2.8**	**65**	**1pph**	**0.4**	**1/0.4**	**99**	**5er1**	**1.0**	**1/1.0**
32	1elc	12.1	99[c]/**2.0**	**66**	**1ppl**	**0.4**	**1/0.4**	**100**	**6cpa**	**0.6**	**1/0.6**
33	1eld	7.3	25/1.3	**67**	**1pso**	**0.3**	**1/0.3**	**71**	**Model < 2 Å scores best**		
34	1ele	2.6	99[c]/**1.4**	**68**	**1ptv**	**0.3**	**1/0.3**	*89*	*Model < 2 Å scores top 10*		

[a]RMSD of top scoring decoy in Ångstroms measured over all ligand and all side-chain atoms in the binding site of the protein. RMSDs smaller 2 Å are displayed in **bold** letters.
[b]Rank of first decoy with RMSD smaller 2 Å and its RMSD in Ångstroms. Rank 1 is displayed in **bold** letters. Ranks 2 to 10 are displayed in *italic* letters.
[c]Ranks larger than 99 are displayed as 99 in this table.

exchanging the ligands. The algorithm builds all side-chains in conformations very close to both crystal structures in the cross-docking calculations. In two crystal structures of triosephosphate isomerase [TIM, Fig. 3(b)] replacing 2-phosphoglycerate with glycerol-3-phosphate results in a conformational change of one glutamate. This change is modeled correctly in the cross-docking experiment. The binding pocket of the liver X receptor 1pqc/1p8d [Fig. 3(c)] is particularly flexible, can accommodate a variety of different ligands, and is therefore challenging for cross-docking. Some of the loop regions shift by more than 1 Å in between the two crystal structures. As a result, side-chain conformations of amino acids in secondary structure elements are frequently predicted with higher accuracy than in loop regions. For the histidine in Figure 3(c) a small shift is modeled to allow the formation of a hydrogen bond while the ability of the algorithm to capture the conformational change of the phenylalanine is limited because of the significant shift in the backbone. Similar results are found for the human nuclear receptor structures 1fm9/2prq [Fig. 3(d)], in which the conformational changes of a glutamate and a phenylalanine are predicted correctly.

Fig. 3. Side-chain conformational changes in cross-docking. Lowest energy cross-docking results for ligand 2dbl in protein 1dbj (a), ligand 6tim in protein 4tim (b), ligand 1pqc in protein 1pq6 (c), and ligand 1fm9 in protein 2prq (d) are shown (protein backbone rainbow coloring scheme, amino acid side-chains carbon atoms white, and ligand carbon atoms green). Selected amino acid side-chain conformations in the binding site of the protein crystal structures are shown in light blue (conformation in protein crystal structure obtained used for cross-docking the ligand) and light red (conformation in crystal structure obtained with this ligand). While in case (a) all side-chains in both crystal structures superimpose almost perfectly and therefore no changes need to be modeled, conformational changes are observed for some side-chains in (b), (c), and (d) and modeled at different levels of accuracy.

TABLE III. Benchmark Results for Docking Flexible Ligand Conformations to Alternative Protein Crystal Structures Including Sampling of Side-Chain Conformations (Cross-Docking)

	Native[a] RMSD of Minimized Starting Structure in Å		Rigid Ligand + Flexible Side-Chains[b]						Flexible Ligand + Flexible Side-Chains[c]					
			RMSD of Best Scoring Model in Å			Rank of First Model with RMSD < 2 Å			RMSD of Best Scoring Model in Å			Rank of First Model with RMSD < 2 Å		
	1aq1	1dm2		1aq1	1dm2		1aq1	1dm2		1aq1	1dm2		1aq1	1dm2
1aq1	**0.42**	**0.57**	1aq1	**0.28**	*2.04*	1aq1	**1**	*3*	1aq1	**0.49**	4.59	1aq1	**1**	27
1dm2	**0.44**	**0.35**	1dm2	**0.48**	**0.34**	1dm2	**1**	**1**	1dm2	**0.49**	**0.57**	1dm2	**1**	**1**
	1dbj	2dbl		1dbj	2dbl		1dbj	2dbl		1dbj	2dbl		1dbj	2dbl
1dbj	**1.12**	**0.94**	1dbj	**0.99**	**0.78**	1dbj	**1**	**1**	1dbj	**0.99**	**0.78**	1dbj	**1**	**1**
2dbl	**0.72**	**0.47**	2dbl	**0.72**	**0.83**	2dbl	**1**	**1**	2dbl	**1.14**	**1.37**	2dbl	**1**	**1**
	1dwc	1dwd		1dwc	1dwd		1dwc	1dwd		1dwc	1dwd		1dwc	1dwd
1dwc	**0.55**	**0.56**	1dwc	**0.66**	**0.59**	1dwc	**1**	**1**	1dwc	6.28	6.67	1dwc	58	244
1dwd	**0.84**	**0.36**	1dwd	**0.88**	**0.30**	1dwd	**1**	**1**	1dwd	**0.79**	6.66	1dwd	**1**	*2*
	1fm9	2prg		1fm9	2prg		1fm9	2prg		1fm9	2prg		1fm9	2prg
1fm9	**0.30**	**1.52**	1fm9	**0.27**	**1.56**	1fm9	**1**	**1**	1fm9	**0.55**	**1.62**	1fm9	**1**	**1**
2prg	**1.03**	**0.38**	2prg	**1.10**	**0.43**	2prg	**1**	**1**	2prg	**1.60**	**1.71**	2prg	**1**	**1**
	1p8d	1pq6		1p8d	1pq6		1p8d	1pq6		1p8d	1pq6		1p8d	1pq6
1p8d	**0.54**	**1.00**	1p8d	**0.45**	**1.64**	1p8d	**1**	**1**	1p8d	**0.64**	**1.75**	1p8d	**1**	**1**
1pq6	**0.44**	**0.18**	1pq6	**0.67**	**0.39**	1pq6	**1**	**1**	1pq6	**1.11**	**0.87**	1pq6	**1**	**1**
	1p8d	1pqc		1p8d	1pqc		1p8d	1pqc		1p8d	1pqc		1p8d	1pqc
1p8d	**0.54**	**0.70**	1p8d	**0.45**	**1.39**	1p8d	**1**	**1**	1p8d	**0.64**	**1.98**	1p8d	**1**	**1**
1pqc	**1.82**	**0.44**	1pqc	*2.11*	**0.66**	1pqc	*7*	**1**	1pqc	3.19	**0.56**	1pqc	162	**1**
	1ppc	1pph		1ppc	1pph		1ppc	1pph		1ppc	1pph		1ppc	1pph
1ppc	**0.26**	**0.61**	1ppc	**0.25**	**0.54**	1ppc	**1**	**1**	1ppc	2.44	2.45	1ppc	15	27
1pph	**0.55**	**0.62**	1pph	**0.61**	**0.60**	1pph	**1**	**1**	1pph	2.33	**0.97**	1pph	*2*	**1**
	1pq6	1pqc		1pq6	1pqc		1pq6	1pqc		1pq6	1pqc		1pq6	1pqc
1pq6	**0.18**	**0.65**	1pq6	**0.39**	*3.43*	1pq6	**1**	*7*	1pq6	**0.87**	**1.66**	1pq6	**1**	**1**
1pqc	**1.72**	**0.44**	1pqc	2.44	**0.66**	1pqc	12	**1**	1pqc	**1.51**	**0.56**	1pqc	**1**	**1**
	2ctc	7cpa		2ctc	7cpa		2ctc	7cpa		2ctc	7cpa		2ctc	7cpa
2ctc	**0.84**	**0.37**	2ctc	**0.68**	**0.69**	2ctc	**1**	**1**	2ctc	*2.80*	2.34	2ctc	*2*	16
7cpa	**1.00**	**0.76**	7cpa	**0.86**	**0.76**	7cpa	**1**	**1**	7cpa	**1.02**	**0.48**	7cpa	**1**	**1**
	4tim	6tim		4tim	6tim		4tim	6tim		4tim	6tim		4tim	6tim
4tim	**0.61**	**0.32**	4tim	**0.57**	**0.26**	4tim	**1**	**1**	4tim	**0.81**	**1.29**	4tim	**1**	**1**
6tim	**0.42**	**0.20**	6tim	*2.67*	**0.39**	6tim	*6*	**1**	6tim	**1.54**	**1.48**	6tim	**1**	**1**

[a]RMSD of ligand minimized in original and alternative protein crystal structure in Ångstroms. RMSDs smaller 2 Å are displayed in **bold** letters.
[b]RMSD of rigid ligand docked in original and alternative protein crystal structure in Ångstroms and rank of first decoy with RMSD smaller 2 Å. Rank 1 is displayed in **bold** letters. Ranks 2 to 10 are displayed in *italic* letters.
[c]RMSD of flexible ligand docked in original and alternative protein crystal structure in Ångstroms and rank of first decoy with RMSD smaller 2 Å. Rank 1 is displayed in **bold** letters. Ranks 2 to 10 are displayed in *italic* letters.

Because side-chain flexibility is modeled we observe only a slight reduction in performance between self-docking (80% success rate) and cross-docking (70% success rate) benchmark. Ferrara and colleagues[5] report cross-docking experiments on two systems, HIV-1 proteases and trypsin, and report small and large deteriorations, respectively. Cross-docking experiments on protein kinases have been reported for ICM, which also utilizes Monte Carlo minimization.[22] The success rate is ~65% when using the 2 Å RMSD criteria for ligand atoms only. It is clear that the difficulty of cross-docking calculations depends critically on the amount of backbone conformational change and the variety of ligands the binding site can accommodate. For rigid protein binding sites, one cross-docking benchmark is reported for FLEXX.[48] However, here a ligand is docked into a list of 2 to 10 other X-ray structures and only the best result obtained is reported, which simplifies the test somewhat. Using the 2 Å criterion a success rate of 76% is achieved. More extensive comparison to previous methods is not possible because no cross-docking benchmark covering a comparable variety of systems has to our knowledge been published. However, in a virtual screening experiment with 10 proteins, McGovern and Shoichet report a decreasing enrichment of known ligands when going from the holo (ligand-bound) protein to the apo (ligand-free) protein to a homology model.[49]

Comparison of structures of protein–protein and protein–ligand complexes with the unbound structures show that about 80% of the side-chain rotamer confor-

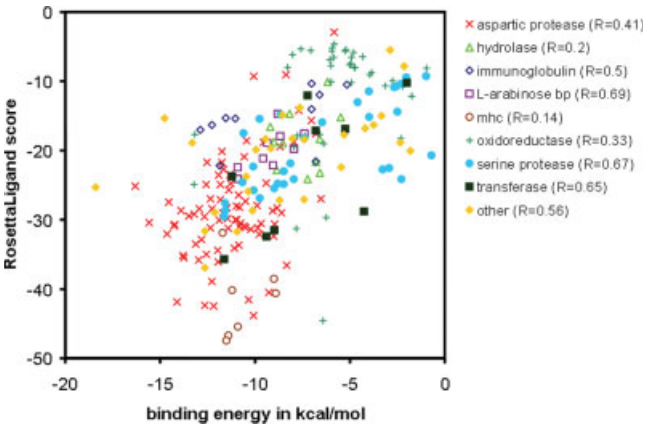

Fig. 4. Correlation between experimental (*x*-axis) and predicted (*y*-axis) binding energy for a set of 229 diverse protein–ligand complexes taken from the LPDB.[36] The overall correlation coefficient is $R = 0.63$. While particularly good correlations are obtained for aspartic proteases and serine proteases, the correlations are worse for hydrolases, mhc's, and oxidoreductases. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

mations are retained upon binding.[18,29] In our cross-docking benchmark, 23% of all side-chains in the first and second shell of the binding site change to a different rotamer (Table SII). Hence in real world docking applications it is advantageous to include input side-chain conformations in the rotamer library for the search.[29] Although this is an option in ROSETTA, it was not used here to keep the benchmarks as strict and unbiased as possible. The Monte Carlo rotamer search improved the docking results in particular when side-chain conformational changes occur. Rotamer search and minimization performed significantly better than straight minimization from the crystal structure coordinates, which cannot traverse side-chain torsional barriers (data not shown).

Binding Free Energy Prediction

Two hundred twenty-nine protein–ligand complexes taken from the LPDB (http://lpdb.scripps.edu/)[36] were scored using our energy function. To avoid uncorrelated noise due to minor clashes the Lennard Jones repulsive term was neglected in the total energy computed. These energies are plotted in Figure 4 versus the experimental binding free energies. The overall correlation coefficient was found to be $R = 0.63$, the standard deviation is SD = 2.9 kcal/mol. This agreement did not further improve upon minimizing of the initial structures in our force field. However, after minimization the Lennard Jones repulsive energy can be included without reduction of the correlation coefficient. As already seen for protein–protein interfaces,[34] the computed ROSETTA energies are larger than the experimental binding free energies by a factor of 2.7, at least in part because of the neglect of entropy decrease associated with binding.

Only CHEMSCORE[15,16] achieves a similar correlation coefficient of $R = 0.65$ for this set of protein–ligand complexes. The correlation coefficients of all other reported energy functions were significantly lower.[5] This is particularly remarkable because our energy function is in contrast to CHEMSCORE not specialized to predict binding energies but for use in docking and design calculations.

Correlation and prediction quality vary largely with the type of protein: L-arabinose binding proteins (SD = 0.90 kcal/mol; $R = 0.69$, $N = 9$), hydrolases (SD = 1.82 kcal/mol; $R = 0.20$; $N = 11$), mhc's (SD = 2.18 kcal/mol; $R = 0.18$; $N = 7$), immunoglobines (SD = 2.53 kcal/mol; $R = 0.50$; $N = 10$), aspartic proteases (SD = 2.64 kcal/mol; $R = 0.41$; $N = 82$), serine proteases (SD = 2.68 kcal/mol; $R = 0.67$; $N = 31$), transferases (SD = 2.79 kcal/mol; $R = 0.65$; $N = 9$), oxidoreductases (SD = 3.33 kcal/mol; $R = 0.33$; $N = 39$), and other (SD = 3.46 kcal/mol; $R = 0.56$; $N = 28$). These large differences and the finding $R_{\text{serine proteases}} > R_{\text{other}} > R_{\text{aspartic proteases}} > R_{\text{oxidoreductases}}$ are in agreement with data reported for most other energy functions.[5]

CONCLUSIONS

We present a novel approach for modeling protein–ligand interfaces that allows the parallel optimization of protein side-chain conformations and ligand translational and rotational degrees of freedom. We find an energy function comprised of LJ-attractive and repulsive interactions, an implicit solvent model, an explicit orientation dependent hydrogen bonding potential, and electrostatics successful in distinguishing low energy from alternative native conformations in more than 70% of all docking experiments computed.

Such successful docking runs are frequently accompanied with the formation of a distinct binding funnel (see Figure 3). The kinetics of binding can be computed from the dimensions of the aperture to the binding funnels using the solution of the diffusion equation for asymmetric rigid bodies with orientational constraints.[50] The energy function fails to distinguish native from non-native conformations only for small molecules with only a few atoms and hence a very limited number of interactions.

The Monte Carlo minimization procedure used to sample ligand rotational and translational degrees of freedom as well as protein side-chain conformational space is found to be efficient for sampling all but one case of the benchmark comprised of 140 self-docking and cross-docking experiments. Ligand flexibility and protein backbone degrees of freedom are currently considered by performing multiple runs from a set of alternate conformations.

The algorithm compares well in prediction accuracy with existing methods in self-docking experiments. This is with the addition of the degrees of freedom on the protein side which is a harder test than most extensive benchmarks recorded so far.The binding energies computed with the energy function correlate with the experimental values with a correlation coefficient of $R = 0.63$. This is comparable to the currently best energy functions used for this problem. Our approach has the advantage that the same function is used for docking and for scoring.

Future improvements to the method will include modeling ligand flexibility using gradient minimization inside the ROSETTA docking procedure. Larger changes in ligand conformation will be modeled by Monte Carlo sampling of torsion angles of the ligand in a manner similar to sidechain rotamers. The protein structure prediction and loop modeling capabilities of ROSETTA[51] will be used to model loop flexibility during docking.

ACKNOWLEDGMENTS

The authors thank Charles L. Brooks for stimulating discussions, and Jeff Gray, Chu Wang, and Ora Furman for developing the ROSETTADOCK methods. Supplementary material is available at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

REFERENCES

1. Taylor RD, Jewsbury PJ, Essex JW. A review of protein-small molecule docking methods. J Comput Aided Mol Des 2002;16:151–166.
2. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 2002;47:409–443.
3. Shoichet BK, McGovern SL, Wei B, Irwin JJ. Lead discovery using molecular docking. Curr Opin Chem Biol 2002;6:439–446.
4. Schneider G, Bohm HJ. Virtual screening and fast automated docking methods. Drug Discov Today 2002;7:64–70.
5. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL 3rd. Assessing scoring functions for protein-ligand interactions. J Med Chem 2004;47:3032–3047.
6. Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 2001;15:411–428.
7. Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit 1996;9:1–5.
8. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 2002;46:34–40.
9. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470–489.
10. Willett P, Glen RC, Leach AR, Taylor R, Jones G. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
11. Perola E, Walters WP, Charifson PS. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 2004;56:235–249.
12. Buzko OV, Bishop AC, Shokat KM. Modified AutoDock for accurate docking of protein kinase inhibitors. J Comput Aided Mol Des 2002;16:113–127.
13. Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des 2002;16:11–26.
14. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 2000;295:337–356.
15. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 1997;11:425–445.
16. Murray CW, Auton TR, Eldridge MD. Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model. J Comput Aided Mol Des 1998;12:503–519.
17. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol 1995;2:317–324.
18. Zavodszky MI, Kuhn LA. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. Protein Sci 2005;14:1104–1114.
19. Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins 1997; Suppl 1:215–220.
20. Totrov M, Abagyan R. Protein-ligand docking as an energy optimization problem. Drug Receptor Thermodynamics 2001:603–624.
21. Abagyan R, Totrov M. High-throughput docking for lead generation. Curr Opin Chem Biol 2001;5:375–382.
22. Cavasotto CN, Abagyan RA. Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol 2004;337:209–225.
23. Taylor RD, Jewsbury PJ, Essex JW. FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. J Comput Chem 2003;24:1637–1656.
24. Alberts IL, Todorov NP, Dean PM. Receptor flexibility in de novo ligand design and docking. J Med Chem 2005;48:6585–6596.
25. Schroedinger. Glide 2.5. New York: Schroedinger; 2003.
26. McMartin C, Bohacek RS. QXP: powerful, rapid computer algorithms for structure-based drug design. J Comput Aided Mol Des 1997;11:333–344.
27. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395.
28. 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.
29. Wang C, Schueler-Furman O, Baker D. Improved side-chain modeling for protein-protein docking. Protein Sci 2005;14:1328–1339.
30. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura KM, Morozov AV, Baker D. Protein-protein docking predictions for the CAPRI experiment. Proteins 2003;52:118–122.
31. Schueler-Furman O, Wang C, Baker D. Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. Proteins 2005;60:187–194.
32. Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D. Accurate computer-based design of a new backbone conformation in the second turn of protein L. J Mol Biol 2002;315:471–477.
33. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic level accuracy. Science 2003;302:1364–1368.
34. Kortemme T, Baker D. Computational design of protein-protein interactions. Curr Opin Chem Biol 2004;8:91–97.
35. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein-protein interaction specificity. Nat Struct Mol Biol 2004;11:371–379.
36. Roche O, Kiyama R, Brooks CL 3rd. Ligand-protein database: linking protein-ligand complex structures to binding data. J Med Chem 2001;44:3592–3598.
37. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
38. Dunbrack RL Jr, Karplus M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J Mol Biol 1993;230:543–574.
39. Bower MJ, Cohen FE, Dunbrack RL Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. J Mol Biol 1997;267:1268–1282.
40. Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. J Mol Biol 2003;326:1239–1259.
41. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 1983;4:187–217.
42. MacKerell AD Jr, Brooks BR, Brooks CL, Nilsson L, Roux B, Won Y, Karplus M.CHARMM: The energy function and its parameterization with an overview of the program. In: The Encyclopedia of Computational Chemistry. Vol. 1. Chichester: Wiley; 1998. p 271–277.
43. MacKerell AD Jr, Banavali N, Foloppe N. Development and current status of the CHARMM force field for nucleic acids. Biopolymers 2000;56:257–265.

44. Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 1999;34:82–95.
45. Dunbrack RL, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci 1997;6:1661–1681.
46. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr., Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
47. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2002;58:899–907.
48. Kramer B, Rarey M, Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins 1999;37:228–241.
49. McGovern SL, Shoichet BK. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. J Med Chem 2003;46:2895–2907.
50. Schlosshauer M, Baker D. Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. Protein Sci 2004;13:1660–1669.
51. Rohl CA, Strauss CE, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. Proteins 2004;55:656–677.
52. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci U S A 2000;97:10383–10388.