doi:10.1016/j.jmb.2008.11.010 J. Mol. Biol. (2009) 385, 381–392

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

RosettaLigand Docking with Full Ligand and Receptor Flexibility

Ian W. Davis[1] and David Baker[1,2]*

[1]*Department of Biochemistry, University of Washington, Seattle, WA 98195-7350, USA*

[2]*Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195-7350, USA*

Received 18 August 2008; received in revised form 7 November 2008; accepted 10 November 2008
Available online 18 November 2008

Edited by D. Case

Computational docking of small-molecule ligands into protein receptors is an important tool for modern drug discovery. Although conformational adjustments are frequently observed between the free and ligand-bound states, the conformational flexibility of the protein is typically ignored in protein–small molecule docking programs. We previously described the program RosettaLigand, which leverages the Rosetta energy function and side-chain repacking algorithm to account for flexibility of all side chains in the binding site. Here we present extensions to RosettaLigand that incorporate full ligand flexibility as well as receptor backbone flexibility. Including receptor backbone flexibility is found to produce more correct docked complexes and to lower the average RMSD of the best-scoring docked poses relative to the rigid-backbone results. On a challenging set of retrospective and prospective cross-docking tests, we find that the top-scoring ligand pose is correctly positioned within 2 Å RMSD for 64% (54/85) of cases overall.

Keywords: small-molecule docking; protein flexibility; backbone flexibility; Rosetta

Introduction

Many biochemical processes depend on the specific interactions between protein receptors and their small-molecule ligands—metabolites, cofactors, hormones, and drugs. Knowledge of the atomic interactions involved is useful both for understanding natural processes and for engineering new therapeutics, but experimentally determining structures of complexes can be difficult, time-consuming, and expensive. Thus, a variety of programs have been developed to computationally "dock" receptors with ligands at varying levels of detail.

Docking programs often ignore conformational changes in the protein receptor, although these are frequently observed upon binding. Better modeling of receptor flexibility is one of the major challenges for the field.[1,2] Of methods that do account for receptor flexibility, most require significant prior knowledge of the system. For instance, 21 different crystal structures were used to generate composite energy grids for docking into HIV-1 protease with AutoDock.[3] Similarly, FlexE can combine parts from multiple structures during docking.[4] On the other hand, FLIPDock allows many kinds of flexibility to be modeled *de novo*, but each degree of freedom must be selected manually.[5] Likewise, the Glide/Prime method needs only one starting structure, but flexibility is initially modeled simply by truncating three or fewer selected side chains to alanine.[6] Such methods are least useful in situations where docking could be most informative: where binding involves interactions and conformations that have not been observed previously and thus could not have been easily foreseen.

Many approaches to docking with receptor flexibility have been recently reviewed by Cozzini and colleagues,[7] with a focus on either coarse-grained or atomistic molecular dynamics calculations to generate ensembles of receptor structures. Separate docking calculations can be carried out on each member of the ensemble, or the receptor–ligand energy landscapes can be combined into an ensemble average. Neither approach is perfect: in separate docking, the individual structures constitute only a tiny fraction of all low-energy receptor conformations (even if the algorithm recombines the structures on the fly), while averaging may lead to ligand poses that are not compatible with any real receptor state.[7] Both approaches also implicitly assume the

**Corresponding author.* Department of Biochemistry, University of Washington, Seattle, WA 98195-7350, USA. E-mail address: dabaker@u.washington.edu.

Abbreviations used: JNK3, c-Jun N-terminal kinase 3; MCM, Monte Carlo minimization.

Table 1. Impact of initial ligand placement and receptor backbone flexibility on the Meiler and Baker cross-docking test set

	Simple protocol				Rigid-backbone protocol				Flexible-backbone protocol				AutoDock 4.0			
	RMS of #1 (Å)		Rank<2 Å		RMS of #1 (Å)		Rank<2 Å		RMS of #1 (Å)		Rank<2 Å		RMS of #1 (Å)		Rank<2 Å	
Receptor	Self	Cross	Self	Cross	Self	Cross	Self	Cross	Self	Cross	Self	Cross	Self	Cross	Self	Cross
1aq1	7.48	5.40	2	2	1.39	0.80	1	1	1.39	0.42	1	1	0.43	3.80	1	—
1dm2	0.97	6.28	1	—	0.82	6.22	1	—	1.14	1.36	1	1	0.52	0.67	1	1
1dbj	0.55	1.59	1	1	0.38	0.94	1	1	0.46	1.22	1	1	0.90	1.22	1	1
2dbl	1.25	5.77	1	3	1.24	0.58	1	1	1.35	0.85	1	1	1.26	0.96	1	1
1dwc	5.59	3.85	—	—	7.05	6.88	2	—	1.15	8.23	1	—	5.92	2.64	—	4
1dwd	7.78	7.54	—	—	1.40	1.48	1	1	4.47	1.34	15	1	4.01	5.76	19	—
1fm9	3.18	2.33	—	2	5.90	8.35	19	18	5.72	8.39	11	2	3.09	4.20	—	14
2prg	1.20	8.80	1	—	8.28	10.62	17	—	8.30	4.07	5	34	2.77	4.57	6	—
1p8d	1.21	3.76	1	78	1.49	2.06	1	158	1.39	2.24	1	6	2.11	4.82	22	2
1pq6	2.22	4.34	3	7	2.05	1.20	11	1	2.08	1.27	3	1	1.45	2.91	1	48
1p8d	1.21	1.47	1	1	1.49	1.78	1	1	1.39	2.29	1	19	2.11	4.42	22	7
1pqc	0.70	1.43	1	1	0.49	0.98	1	1	0.53	0.98	1	1	1.62	4.57	1	—
1pq6	2.22	3.49	3	12	2.05	3.46	11	7	2.08	3.61	3	25	1.45	3.65	1	45
1pqc	0.70	8.10	1	93	0.49	3.70	1	181	0.53	2.20	1	128	1.62	2.73	1	—
1ppc	7.61	6.37	—	—	3.16	4.39	4	2	1.62	1.16	1	1	3.69	2.64	4	—
1pph	4.37	8.94	—	—	0.90	3.04	1	8	1.00	5.63	1	2	0.75	3.93	1	8
2ctc	0.90	6.31	1	—	0.70	8.49	1	—	0.73	3.64	1	—	5.07	4.44	28	—
7cpa	4.63	0.52	—	1	4.30	4.51	—	3	4.62	4.60	—	4	6.25	5.04	—	36
4tim	5.27	1.25	2	1	1.07	1.35	1	1	0.95	1.34	1	1	1.39	1.37	1	1
6tim	1.22	2.12	1	5	1.16	2.13	1	7	1.25	0.70	1	1	1.19	2.58	1	—
Average	3.01	4.48			2.29	3.65			2.11	2.78			2.38	3.35		

Ten pairs of proteins were docked against their own co-crystalized ligand ("self") or the ligand from the other structure ("cross"). "RMS of #1" denotes the RMSD of ligand heavy atom positions *versus* the crystal structure, for the best-scoring docked pose. "Rank <2 Å" gives the rank of the first pose to get within 2 Å RMSD of the crystal structure; some pairs did not have a pose under 2 Å within the top 5% of structures produced. See the text for description of the protocols.

ligand-bound conformation is sufficiently low energy in the absence of the ligand to be included in the ensemble. Nonetheless, these approaches have improved performance in some cases.[3,8]

Here we describe a new approach to protein–ligand docking that explicitly models full side-chain, backbone, and ligand flexibility with ligand and receptor degrees of freedom explored simultaneously. The new method is an extension of the ROSETTALIGAND algorithm, which uses Monte Carlo sampling and the Rosetta full-atom energy function. We analyze the performance of the new method on an extensive set of retrospective and prospective cross-docking scenarios and find it competitive with and possibly superior to the best existing methods.

Results

As described in Materials and Methods, we have made important enhancements to the ROSETTALIGAND docking program, notably the inclusion of full ligand flexibility (conformers plus torsional minimization), better initial ligand placement, and receptor backbone flexibility. Here we demonstrate, with both retrospective and prospective tests, that these features improve the accuracy of ROSETTALIGAND docking predictions. The retrospective assessment used the Meiler and Baker set: 10 pairs of co-crystallized receptor–ligand structures from our previous work,[9] which allow for self- and cross-docking experiments within each pairing. Many of them are quite challenging, with large, flexible ligands and multiple side chains changing rotamer. This set was used repeatedly to test the effectiveness of various algorithms. The prospective assessment was conducted as part of the community-wide Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL-1) blind-docking experiment (see Materials and Methods)†. The challenge was to dock nearly 100 ligands into representative structures of their cognate receptors [c-Jun N-terminal kinase 3 (JNK3) kinase and urokinase-type plasminogen activator]. We made truly blind predictions during SAMPL, but the initial coarse-grained ligand placement and backbone flexibility features were not implemented at that time. We

† http://sampl.eyesopen.com/

reevaluated our final method with these features on this set a few months later, but we did not use it for testing in between, and there was no human intervention in the second docking run. Thus, we believe this provides a fair and nearly blind assessment of the current version of RosettaLigand.

Meiler and Baker cross-docking

Table 1 shows the impact of initial ligand placement and receptor backbone flexibility on the Meiler and Baker set. The "simple protocol" is similar to the original RosettaLigand in that the ligand is placed randomly in a sphere centered on the binding site and the backbone is held fixed throughout. The "rigid-backbone protocol" adds the coarse screen for initial ligand placement, which only accepts as starting points those poses with good ligand–receptor shape complementarity. Finally, the "flexible-backbone protocol" adds the restrained minimization of the receptor backbone at the end of every docking trajectory.

Better initial ligand placement means more search effort is focused in plausible regions of the energy

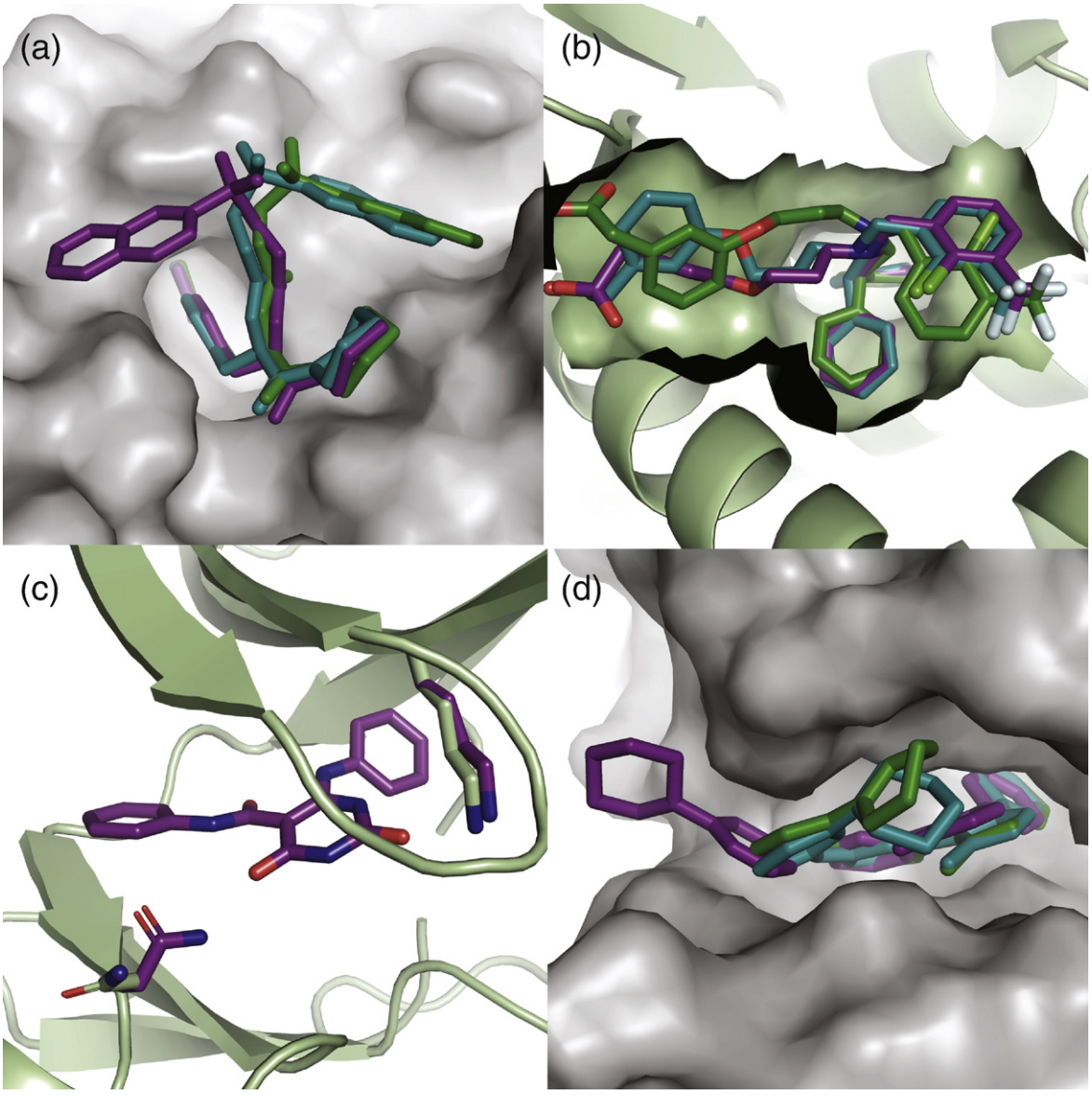

Fig. 1. Challenges in searching ligand conformations. (a) Docking the 1ppc ligand into the 1pph receptor. The actual binding mode (green) requires fitting the ligand into three separate pockets, successful in the cyan docked pose. Without the initial shape complementarity search, typically only two of the three pockets are filled (purple). (b) Generating the bioactive conformation of the flexible 1pq6 ligand is difficult. The nearest conformers generated by Omega (cyan, purple) fill roughly the same space as the native (green) but do not make the same chemical interactions. (c) Including side-chain flexibility is important for capturing the native binding mode of ligand 56 for JNK3 kinase (purple). Only two hydrogen bonds hold it in place, and in the provided SAMPL crystal structure Asn152 (green, at lower left) is pointed out into solvent. (d) The narrow binding cleft of JNK3 kinase requires precise ligand placement, particularly for large, flexible ones like compound 8. Random attempts at placement generally leave one end of the ligand projecting into solvent (purple), but the initial shape complementarity search guides docked poses (cyan) into the native binding mode (green).

landscape. As a result, the program is more likely to sample closer to the true binding mode. Thus, more cases have the lowest-energy structure within 2 Å of the native structure, and the average RMSD to the native structure decreases. Also, more cases have some low-energy predictions with RMSD <2 Å, even if the single lowest-energy prediction is not. Not surprisingly, ROSETTALIGAND performs better

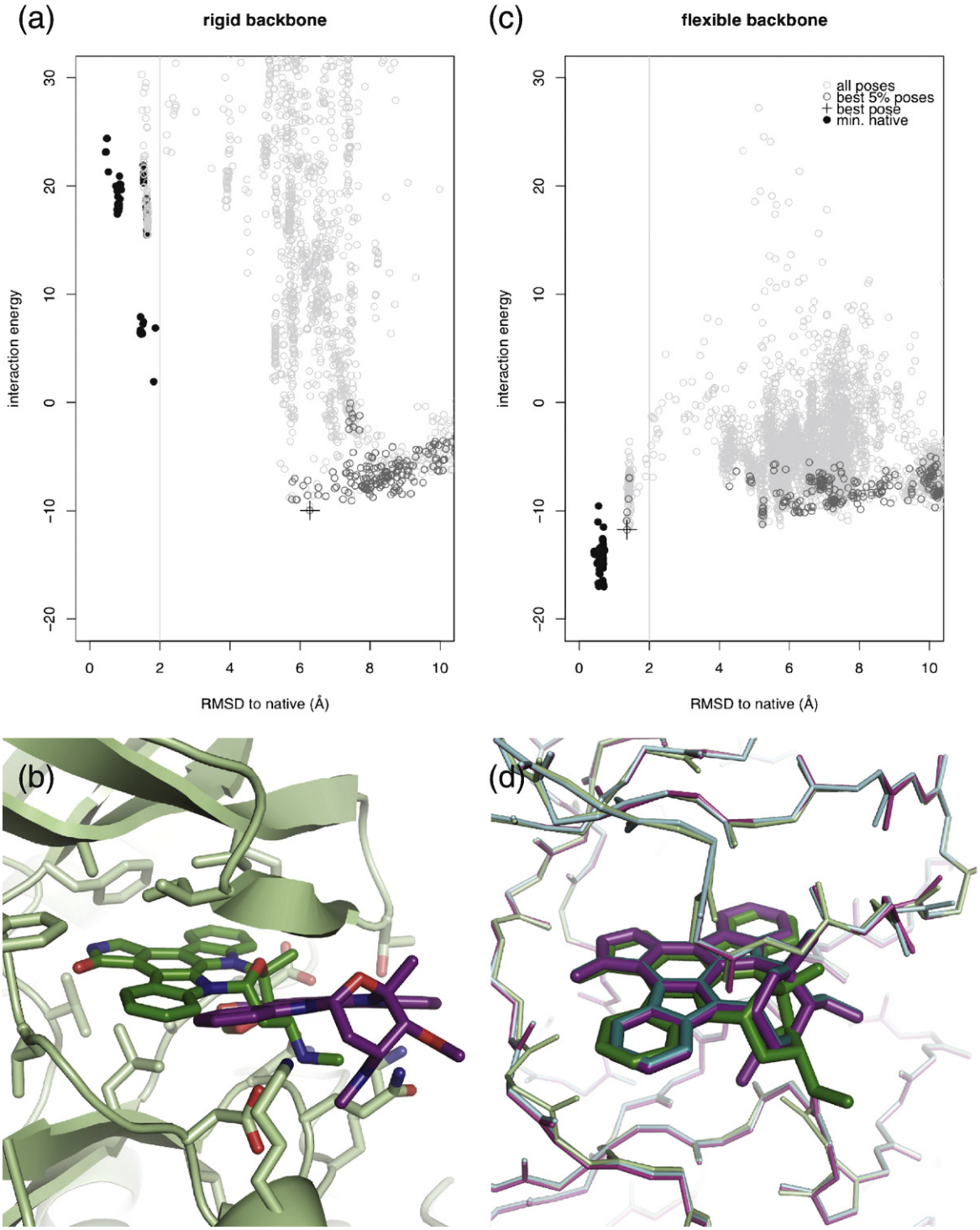

Fig. 2. Impact of backbone flexibility on 1dm2/1aq1 (CDK2/staurosporine). (a) The rigid-backbone algorithm (left) does sample poses (light circles) near the native (black dots), but clashes with the receptor backbone make them significantly less favorable than the lowest-energy poses (dark circles). (b) The predicted pose (black cross/purple sticks) is docked in the mouth of the binding pocket, far from the native binding mode (green sticks). (c) By allowing a small amount of backbone flexibility for the receptor (right), near-native poses are identified as the lowest in energy. (d) Two low-energy predictions (cyan, purple) show that relatively small backbone changes occur relative to the crystal structure (green), yet the consequences for scoring are tremendous.

on the self-docking cases than the cross-docking ones, but the improvement afforded by better ligand placement is evident for both.

Figure 1a shows the 1pph/1ppc (trypsin) cross-docking case, which benefits from the initial shape complementarity search. This ligand has three arms that fit tightly in distinct pockets. Totally random placement is unlikely to find all three interactions, so even the best predictions generally have one arm projecting into solvent (purple). The initial rough screen for shape matches, however, narrows the search space and leads to docking conformations (cyan) that match the native (green). Another example is shown in Fig. 1d (JNK3 kinase), where a long, flexible ligand must fold into a Z shape. Again, the highly specific sterics of this pocket produce a narrow funnel in the energy landscape, one that is difficult to discover (purple) without focusing the search on plausible poses (cyan).

In contrast to ligand placement, which makes the search process more efficient, introducing receptor backbone flexibility actually alters the receptor–ligand energy landscape. Again, compared to the rigid-backbone protocol, there are more cases in which docked models are within 2 Å of the native complex and there are more cases in which the lowest-energy model is within 2 Å. The average RMSD also drops again, although by a smaller amount.

Figure 2 shows a particularly dramatic example of the benefits of backbone minimization. This receptor structure was determined with a smaller ligand bound, so the binding site had closed in slightly. The docked ligand, being rigid and somewhat larger, clashed with the receptor even after side-chain and ligand torsions were minimized. Note that even the minimized native pose has positive binding energies (steric clashes), and although some docking trajectories sampled near the native pose, their energies kept them out of the 5% lowest-energy structures. Instead, the ligand preferred to stick in the mouth of the pocket. On the other hand, allowing the backbone to relax even a small amount allowed a near-native pose to be recognized as lower in energy.

While backbone and side-chain flexibility were helpful in correctly docking the ligands, the magnitude of changes in the receptors was small, as suggested by Fig. 2. Within 6 Å of the ligand, the average RMSDs between the lowest-energy models and their starting structures were 0.45 Å over backbone atoms and 0.88 Å over all atoms (0.7 and 1.7 Å maximum, respectively). The RMSDs between the pairs of starting structures span the same ranges. On average, backbone flexibility did not improve receptor RMSD to the co-crystal complex, possibly because the changes that enabled correct ligand docking were balanced by spurious changes elsewhere. Regardless, the rugged energy landscape created by hydrogen bonding and repulsive van der Waals terms means that these small conformational changes can have a large impact on docking success.

Table 2. Impact of initial ligand placement and receptor backbone flexibility on JNK3 kinase targets from the SAMPL-1 blind docking challenge

		SAMPL-1 predictions		Flexible backbone	
Compound	Torsions	RMS of best	Rank<2 Å	RMS of best	Rank<2 Å
jnk.pp.1-1	6	0.84	1	1.17	1
jnk.pp.1-3	4	1.40	1	0.51	1
jnk.pp.1-7	5	1.07	1	1.00	1
jnk.pp.1-8	5	3.40	22	0.73	1
jnk.pp.1-9	4	2.23	9	1.92	1
jnk.pp.1-10	5	1.10	NA*	0.96	1
jnk.pp.1-11	5	0.91	1	1.07	2
jnk.pp.1-12	5	0.74	1	0.71	1
jnk.pp.1-14	3	1.28	1	1.22	1
jnk.pp.1-15	4	4.00	6	0.36	1
jnk.pp.1-18	7	0.94	2	1.19	2
jnk.pp.1-19	7	0.64	21*	0.67	1
jnk.pp.1-22	4	1.55	1	0.47	1
jnk.pp.1-24	3	5.56	NA	2.70	23
jnk.pp.1-25	5	0.69	4	0.67	1
jnk.pp.1-26	4	0.61	1	0.71	1
jnk.pp.1-27	5	0.90	1	0.50	1
jnk.pp.1-28	3	0.79	1	3.41	31
jnk.pp.1-32	4	1.39	1	0.45	1
jnk.pp.1-33	6	3.44	NA	2.78	NA
jnk.pp.1-34	3	0.89	1	0.71	1
jnk.pp.1-35	6	0.87	1	0.65	1
jnk.pp.1-37	7	2.25	4	1.27	1
jnk.pp.1-38	4	2.48	6	0.83	1
jnk.pp.1-39	4	0.47	NA*	0.68	1
jnk.pp.1-40	5	0.55	NA*	0.64	1
jnk.pp.1-41	5	1.52	NA*	1.13	1
jnk.pp.1-42	6	0.95	1	0.85	1
jnk.pp.1-46	8	4.89	NA	5.34	NA
jnk.pp.1-48	4	5.81	5	1.21	2
jnk.pp.1-49	3	1.15	1	0.96	1
jnk.pp.1-51	1	1.39	1	1.33	1
jnk.pp.1-52	5	0.64	1	6.89	NA
jnk.pp.1-55	6	1.92	1	1.01	1
jnk.pp.1-56	4	9.47	6	0.73	2
jnk.pp.1-59	5	4.18	3	0.66	1
jnk.pp.1-60	4	10.91	20	0.98	7
jnk.pp.1-62	5	10.61	NA	5.28	NA
#1<2 Å		0.66		0.84	
#1<3 Å		0.74		0.89	
Avg. RMS		2.49		1.43	

"RMS of best" denotes the RMSD of ligand heavy-atom positions *versus* the crystal structure for the best of the three highest-scoring docked poses. "Rank <2 Å" gives the rank of the first pose to be within 2 Å RMSD of the crystal structure; some pairs did not have a pose under 2 Å within the top 5% of structures produced. Compounds marked with asterisks were eventually positioned manually in the binding site for the actual SAMPL-1 predictions; all docking was fully automatic with the flexible-backbone protocol. Decoy ligands were not included in the table, giving rise to discontinuous compound numbering. See the text for description of the protocols.

SAMPL-1 cross-docking

Tables 2 and 3 show the corresponding comparison of the simple docking algorithm with the flexible-backbone algorithm for the two receptors used in the SAMPL-1 blind docking experiment. Because the initial ligand placement and flexible-backbone portions of ROSETTALIGAND were developed later, our actual submitted predictions for SAMPL were made with the simpler algorithm; the

Table 3. Impact of initial ligand placement and receptor backbone flexibility on urokinase targets from the SAMPL-1 blind docking challenge

		SAMPL-1 predictions		Flexible backbone	
Compound	Torsions	RMS of best	Rank<2 Å	RMS of best	Rank<2 Å
uk.pp.1-1	0	3.27	3	2.19	16
uk.pp.1-2	5	1.01	1	1.19	1
uk.pp.1-3	6	0.76	1	1.04	1
uk.pp.1-5	4	3.78	NA	2.27	81
uk.pp.1-6	3	5.59	3	5.85	10
uk.pp.1-7	3	1.46	1	0.80	1
uk.pp.1-8	3	2.34	5	2.17	13
uk.pp.1-9	8	3.28	34	7.39	NA
uk.pp.1-10	3	1.16	1	1.95	3
uk.pp.1-11	1	0.72	1	0.54	1
uk.pp.1-13	5	1.48	1	2.03	NA
uk.pp.1-14	1	0.52	1	0.64	1
uk.pp.1-15	4	2.38	2	2.10	11
uk.pp.1-16	5	2.94	3	2.56	15
uk.pp.1-18	2	2.09	2	1.40	1
uk.pp.1-19	4	2.17	2	2.21	NA
uk.pp.1-20	6	3.83	5	3.50	50
uk.pp.1-21	3	8.27	37	7.93	49
uk.pp.1-22	8	3.75	NA	4.33	NA
uk.pp.1-23	1	0.58	1	0.43	1
uk.pp.1-24	3	2.41	14	3.09	10
uk.pp.1-25	2	1.31	1	0.67	1
uk.pp.1-29	6	1.45	1	1.70	3
uk.pp.1-30	3	0.62	1	0.59	1
uk.pp.1-31	4	1.17	1	3.04	5
uk.pp.1-32	3	4.11	NA	2.22	49
uk.pp.1-34	4	2.12	2	1.98	2
#1<2 Å		0.44		0.44	
#1<3 Å		0.70		0.74	
Avg. RMS		2.39		2.44	

See Table 2 for details.

flexible-backbone algorithm was evaluated after the fact.

Adding in the initial shape-complementarity search and receptor backbone flexibility was quite beneficial for the JNK3 kinase ligands, but had virtually no effect on the urokinase ligands. For JNK3, more predictions are within 2 Å of native and the average RMSD is lower. Also, the five compounds that had to be positioned manually with the original algorithm now find their native binding modes without any intervention. For the most part, the JNK3 binding site is rigid. Two representative receptor structures were provided, and although several nonpolar side chains change rotamer, the overall change in the shape and electrostatics of the binding site are minimal. However, Fig. 1c shows an example of a binding mode that would not have been discovered with a rigid receptor model. This compound does not make the backbone hydrogen bonds that most of the others do, and instead is held in place by side-chain hydrogen bonds to Lys93 and Asn152. Asn152 is not positioned correctly to make this interaction in either of the representative structures, but Rosetta's repacking algorithm discovers the correct conformation and, thus, the correct binding mode. Other successful cases of mandatory side-chain repacking were previously described;[9] the new ROSETTALIGAND continues to treat them successfully.

For urokinase, on the other hand, some compounds fare worse and some fare better, but on average the performance is unchanged. Figure 3a and b show a representative docking success and failure, respectively, for this system. The success accurately captures all the key interactions, but in the failed case the compound is buried too deeply in the receptor. In one low-energy case it occupies a pocket that should hold a buried water molecule (water molecules are not modeled in our protocol), and in another it displaces a tyrosine side chain. Since this receptor does not change conformation appreciably upon binding, allowing receptor flexibility introduces opportunities for this kind of mistake.

Comparison to other methods

To aid comparison, we have docked the Meiler and Baker set with AutoDock 4.0.[10] AutoDock is among the most used docking programs[2] and is available under an open-source license‡. Parameters were as recommended by the authors' online tutorial: 60×60×60 grid points spaced 0.375 Å apart, 25 million energy evaluations per run, and 50 runs of the Lamarckian genetic algorithm. Other parameters were set to their default values in the AutoDockTools graphical interface, and the receptor remained in its crystallized conformation. Docking required 5–22 processor hours per compound, compared to 40–80 for ROSETTALIGAND. Average RMSDs of the top-ranked models (Table 1) are similar to the ROSETTALIGAND rigid-backbone protocol, although ROSETTALIGAND generated *some* models within 2 Å RMSD for significantly more cases. The ROSETTALIGAND *flexible-backbone* protocol performs better than AutoDock by both criteria. (These findings should be interpreted cautiously, as we are not experts in using AutoDock; better performance is likely possible.)

At the request of a reviewer, we have also tested the full ROSETTALIGAND flexible-backbone protocol against the Astex Diverse Set[11] to facilitate comparison with other docking tools. The Astex Diverse Set contains only self-docking problems; as such, the receptor flexibility allowed by ROSETTALIGAND becomes a disadvantage. Nonetheless, the top-scoring pose was positioned within 2 Å RMSD in 58% of cases, with additional near misses: one of the top three poses was within 2 Å in 73% of cases, and the top pose was within 3 Å in 71% of cases (see Table 4). By comparison, GOLD docks 75% of these compounds within 2 Å RMSD.[11] When the experimentally determined ligand geometry is used, GOLD improves to 80–99% correct, and ROSETTALIGAND improves to 99% correct, suggesting that ligand conformer generation is the limiting factor. Thus, it appears that ROSETTALIGAND maintains reasonable performance on self-docking tests while offering

‡ http://autodock.scripps.edu

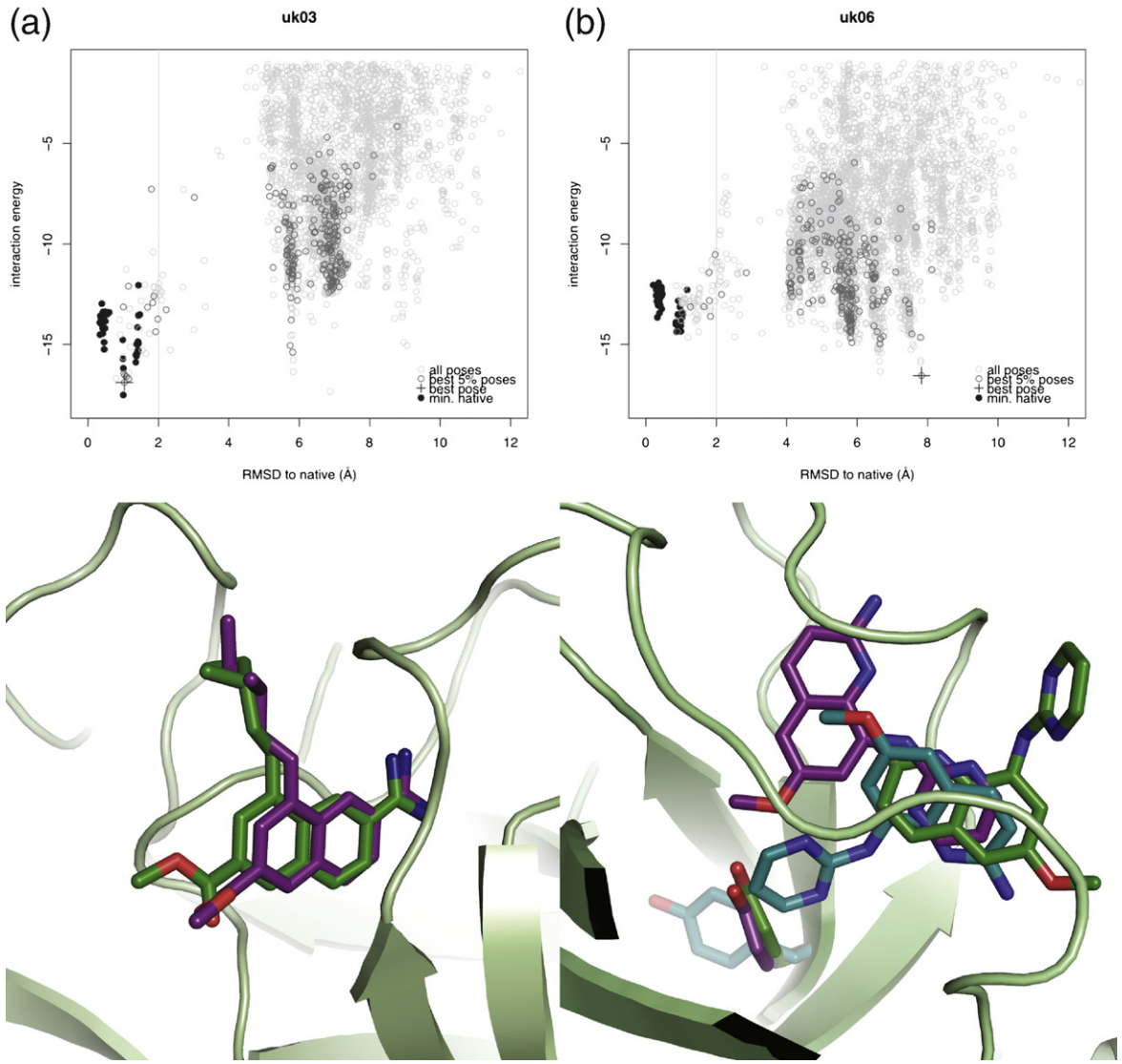

Fig. 3. Examples of docking against urokinase. (a) Compound 3 demonstrates a typical successful docking, with the lowest-energy poses (dark circles) similar in RMSD and energy to the minimized native structure (black dots). (b) Compound 6 illustrates a failed docking, in which the flexibility allowed by ROSETTALIGAND permits the ligand to penetrate too deeply into the receptor (cyan, purple) compared to the true binding mode (green).

additional benefits for more difficult cross-docking scenarios.

Discussion

Overall, we find that the top-scoring ligand pose is correctly positioned within 2 Å RMSD for 64% (54/85) of cross-docking scenarios. To put this in context, Warren *et al.* studied 19 docking protocols on a diverse set of cross-docking scenarios and found that when considering only the top-scoring pose, even the best program for each receptor rarely placed more than ~60% of compounds within 2 Å RMSD, and on average, each program did significantly worse.[12] While it is difficult to meaningfully compare different docking studies, ROSETTALIGAND does show promise for accurate docking in real-world conditions, particularly for receptors with a moderate amount of structural plasticity.

The current version of ROSETTALIGAND captures much of the possible conformational variation in small molecules, but those with more than 8–10 rotatable bonds still present a significant challenge. Figure 1b shows docking results for the 1pq6 ligand, with 14 rotatable bonds. Although the docked poses occupy the correct pocket in approximately the correct orientation, the details of the interaction are mostly wrong. In part, this is a limitation of a pre-generated conformer library: even with explicit local optimization of ligand degrees of freedom, a few hundred or thousand conformers are not adequate to explore such a large conformation space. (The nearest conformer has an RMSD of 1.7 Å to the native, making a docking within 2 Å nearly impossible.) Thus, future work will consider generating new conformers on the fly,

Table 4. Results for the Astex Diverse Set of self-docking cases

	Omega only		Omega + xtal geometry	
PDB code	RMS of #1	Rank < 2 Å	RMS of #1	Rank < 2 Å
1g9v	7.23	25	0.30	1
1gkc	1.32	1	0.24	1
1gm8	3.03	68	3.08	184
1gpk	3.68	2	0.22	1
1hnn	1.12	1	0.72	1
1hp0	1.65	1	0.80	1
1hq2	1.02	1	0.23	1
1hvy	2.46	NA	0.59	1
1hwi	0.72	1	0.31	1
1hww	3.84	2	0.58	1
1ia1	0.47	1	0.26	1
1ig3	1.48	1	0.53	1
1j3j	0.34	1	0.40	1
1jd0	5.49	40	1.12	1
1jje	8.03	31	0.28	1
1jla	0.84	1	0.35	1
1k3u	2.75	2	0.31	1
1ke5	0.46	1	0.25	1
1kzk	0.28	1	0.17	1
1l2s	1.04	1	1.03	1
1l7f	1.55	1	0.23	1
1lpz	3.65	NA	0.34	1
1lrh	2.14	3	0.38	1
1m2z	0.40	1	0.33	1
1meh	1.66	1	0.64	1
1mmv	1.70	1	0.40	1
1mzc	8.90	NA	0.26	1
1n1m	2.36	2	0.78	1
1n2j	4.29	44	0.37	1
1n2v	1.82	1	0.33	1
1n46	0.60	1	0.23	1
1nav	0.65	1	0.22	1
1of1	0.61	1	0.22	1
1of6	0.57	1	0.24	1
1opk	1.83	1	0.17	1
1oq5	1.45	1	0.47	1
1owe	2.36	7	0.23	1
1oyt	0.73	1	0.33	1
1p2y	1.01	1	0.58	1
1p62	2.83	2	0.32	1
1pmn	0.86	1	0.56	1
1q1g	6.38	69	0.35	1
1q41	1.92	1	0.33	1
1q4g	0.92	1	0.37	1
1r1h	1.74	1	0.30	1
1r55	1.60	1	0.15	1
1r58	3.13	3	0.68	1
1r9o	7.93	2	0.59	1
1s19	0.90	1	0.30	1
1s3v	0.53	1	0.19	1
1sg0	0.52	1	0.30	1
1sj0	4.95	NA	0.35	1
1sq5	5.71	4	0.56	1
1sqn	1.36	1	0.12	1
1t40	5.88	NA	0.23	1
1t46	0.42	1	0.21	1
1t9b	1.66	1	0.62	1
1tow	3.44	6	0.68	1
1tt1	0.28	1	0.12	1
1tz8	2.38	2	0.36	1
1u1c	1.03	1	0.14	1
1u4d	0.78	1	0.55	1
1uml	7.67	NA	0.36	1
1unl	0.71	1	0.29	1
1uou	1.44	1	0.28	1
1v0p	1.26	1	0.22	1
1v48	2.35	5	0.35	1
1v4s	1.00	1	0.38	1
1vcj	4.22	35	0.71	1
1w1p	3.03	18	0.56	1
1w2g	2.28	3	0.33	1
1x8x	3.56	6	0.18	1
1xm6	2.26	3	0.26	1
1xoq	8.08	30	0.63	1
1xoz	6.99	2	0.27	1
1y6b	5.17	NA	0.24	1
1ygc	3.08	11	0.50	1
1yqy	1.02	1	0.25	1
1yv3	0.38	1	0.28	1
1yvf	2.82	2	0.36	1
1ywr	0.59	1	0.55	1
1z95	7.87	5	0.28	1
2bm2	1.36	1	0.34	1
2br1	1.03	1	0.40	1
2bsm	0.61	1	0.19	1
#1 < 2 Å	0.58		0.99	
#(1–3) < 2 Å	0.73		0.99	
#1 < 3 Å	0.71		0.99	
Avg. RMS	2.49		0.41	

The full ROSETTALIGAND protocol including backbone flexibility was used. Dockings were performed either with the ligand conformers generated by Omega ("Omega only") or with those conformers plus the experimentally observed conformation of the co-crystallized ligand ("Omega + xtal geometry"). "RMS of #1" denotes the RMSD of ligand heavy-atom positions *versus* the crystal structure for the best-scoring docked pose. "Rank <2 Å" gives the rank of the first pose to be within 2 Å RMSD of the crystal structure; some pairs did not have a pose under 2 Å within the top 5% of structures produced.

as well as incremental ligand-building strategies (K. Kaufmann and J. Meiler, unpublished results).

The core ROSETTA algorithms and energy function have been successfully applied in various problems where an entire protein is flexible.[13,14] However, we found that unconstrained minimization of the receptor backbone impaired discrimination between native and nonnative poses; other authors report similar findings.[15] On the other hand, we found that constrained minimization produced better results than simply softening the repulsive potential (data not shown). For example, using ROSETTA's soft-repulsive potential with a rigid backbone leads to the same nonnative pose for the 1dm2/1aq1 case as the hard-repulsive potential with a rigid backbone does (Fig. 2); performance for other cases is also similar to the hard-repulsive potential. Similar observations have previously been made for high-resolution structure prediction,[13] protein design,[14] and protein–protein docking.[16] Although successful docking into highly flexible receptors (such as HIV-1 protease) or homology models will require more flexibility than constrained minimization can offer, it appears to be a useful technique for receptors with more modest backbone changes.

As expected, incorporating so much conformational flexibility into ROSETTALIGAND requires considerable computational expense, typically 40–80 processor hours for the cases presented here. Obviously, the method is not suited to large-scale virtual screening applications, but its requirements

are similar to those of other programs that allow substantial protein flexibility. For example, Glide/Prime requires 30 min per compound on a cluster of 20 processors,[6] while ROSETTALIGAND requires 2–4 h on the same cluster. Furthermore, Glide/Prime cannot use more than 20 processors, while ROSETTALIGAND calculations scale linearly with the number of processors available. There are also ways to make the ROSETTALIGAND calculations faster. Reducing the translational search from 5 Å to 3 Å would allow the same density of sampling in fivefold less time, albeit over a smaller volume. Ligands with few rotatable bonds present a smaller search space and so require several times less than 5000 trajectories to sample the correct binding mode; the larger number was used for all cases here simply for uniformity. Finally, foreknowledge of likely binding modes or specific interactions can be input as atom–atom distance or angle restraints; the reduction in search time depends on the specificity of the restraints.

As others have noted, retrospective assessments are no substitute for making truly prospective predictions.[1] However, we have endeavored to avoid common pitfalls and to make the retrospective results presented here fairly represent a real-world docking scenario: neither the ligand's bound conformation nor its co-crystallized receptor's structure (excepting the 17 self-docking cases) was used; the receptors were preminimized without the ligands present to avoid bias; and binding site definitions were made generously large to allow exploration of alternative binding modes. Also, all side chains and backbone near the binding site were allowed to change conformation, rather than a few residues carefully selected from multiple structures. Finally, no system-specific knowledge was applied (e.g., pharmacophore constraints), although such knowledge is often available in real drug-discovery exercises.

Despite these efforts, we believe that experiments like SAMPL provide a still more stringent test and fill a critical need within the field. The full results of the SAMPL-1 experiment have not yet been released, so we are unable to make a detailed comparison between the performance of ROSETTALIGAND and other docking software; in any case, performance on just two receptors is not enough to make statistically valid comparisons. Nevertheless, it provides a promising indication of ROSETTALIGAND's potential. There

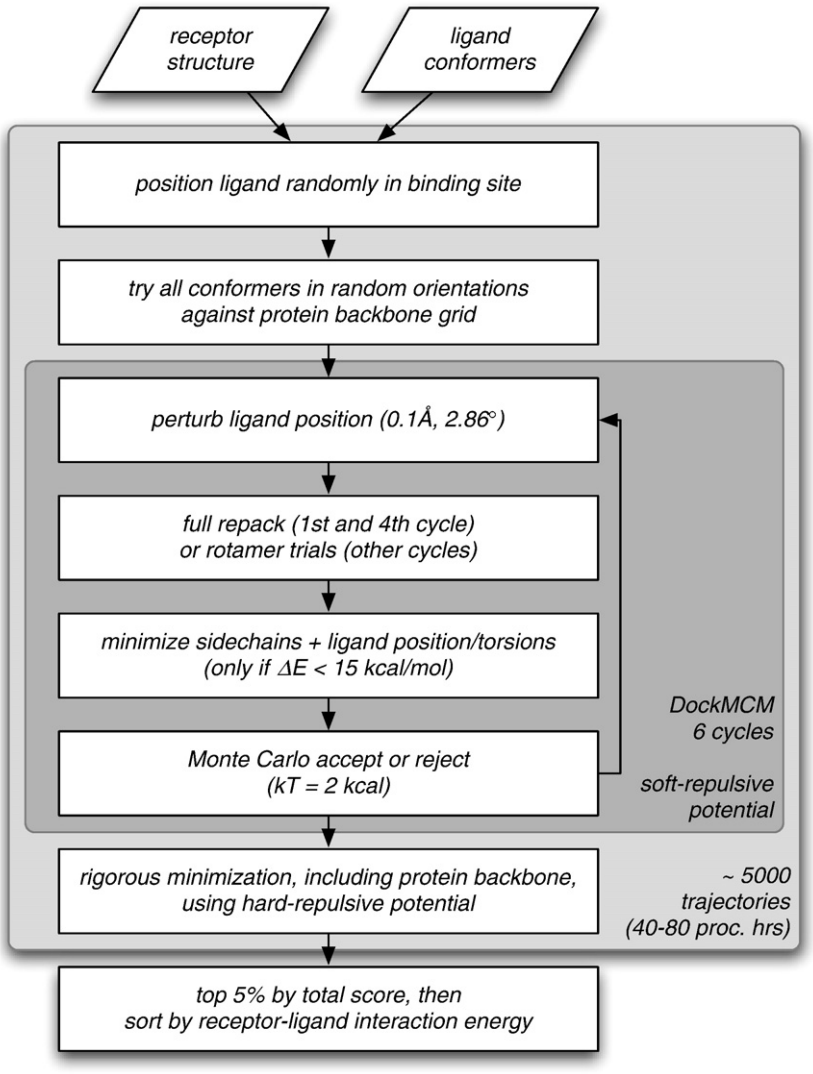

Fig. 4. Flowchart of the ROSETTALIGAND docking algorithm. See Materials and Methods for details. Each trajectory takes approximately 30–60 s, with 2–3 s spent on initial placement, about 10 s spent on the final minimization, and the remainder spent on the DockMCM search.

is, however, still much work to be done to improve the method, particularly in ranking the affinity and predicting the binding energy of diverse ligands.

Materials and Methods

Three stages of the docking algorithm

Our docking algorithm is based on the RosettaLigand method described by Meiler and Baker,[9] but with enhancements that make it faster and more accurate. It also samples internal degrees of freedom in the ligand and protein backbone, in addition to the rigid-body and protein side-chain degrees of freedom sampled in the original RosettaLigand. The protocol is comprised of three stages, which progress from coarse-grained sampling and scoring to fine optimization and a detailed all-atom energy function (Fig. 4). In the first, coarse-grained stage, the ligand is placed randomly within the binding site: a selected "neighbor atom" (usually one of the centermost atoms) is constrained to lie within a 10 Å diameter sphere. Each conformer is superimposed on the neighbor atom, rotated as a rigid body into 1000 random orientations, and scored for shape compatibility with the receptor. We define a shape complementarity score S as follows: If there are N total ligand atoms, R ligand atoms within 2.25 Å of the receptor backbone or C^{β}s (repulsive clashes), and A ligand atoms between 2.25 and 4.75 Å of any receptor atom (attractive contacts), then $S = R - \min(A/N, 0.85)$. More negative scores are better. Only heavy atoms are considered, and receptor side chains are ignored for the repulsive term R because they may move significantly during docking. The attractive contact count A is capped at 85% of its maximum, with the expectation that in the native complex up to 15% of ligand atoms will project into solvent or through the unbound positions of receptor side chains. Typically, dozens of ligand poses are found with "perfect" scores—no clashes and ≥85% good contacts. The best-scoring poses are filtered by RMSD to eliminate near-duplicates, with a threshold of $0.65\sqrt{N}$Å, where N is the number of nonhydrogen ligand atoms (although a wide range of values produced acceptable results). Finally, one of the remaining poses is selected at random to continue this trajectory.

The second, high-resolution stage employs the Monte Carlo minimization (MCM) scheme from the original RosettaLigand: the ligand position and orientation are randomly perturbed by a small amount (normal distributions with σ 0.1 Å and 3°); receptor side chains are repacked using a rotamer library; the ligand position, orientation, and torsions and protein side-chain torsions are simultaneously optimized using quasi-Newton minimization; and the end result is accepted or rejected based on the Metropolis criterion ($kT = 2$). Scoring uses the full-atom Rosetta energy function with softened van der Waals repulsion and the weights described previously,[9] except that hydrogen bonding was downweighted from 2.0 to 1.3 based on our experience designing enzymes.[17,18] This MCM process is repeated for 6 cycles rather than the 50 cycles used originally because it was found most trajectories converge very early. The side-chain rotamers are searched simultaneously in cycles 1 and 4 ("full repack") and one at a time in the remaining cycles ("rotamer trials"). Both rotamer algorithms are stochastic and are described elsewhere.[19,20] Briefly, the full repack makes ~10^6 random rotamer substitutions at random positions and accepts or rejects each on the Metropolis criterion, as kT is lowered from 100 to 0.3. Rotamer trials chooses the single best rotamer at a position in the context of the current state of the rest of the system, with the positions visited once each in random order. The ligand is treated as a single residue, and its input conformers serve as rotamers here. The minimization step, finely sampled rotamer library, and soft-repulsive energy function allow access to off-rotamer conformations.

The third and final stage is a more stringent gradient-based minimization of the ligand position, orientation, and torsions and the receptor torsions for both side chains and backbone. Scoring uses the same Rosetta energy function, but with a hard-repulsive van der Waals potential (see below); this creates a more rugged energy landscape that is less suited to the search stage (above), but is better at discriminating native from nonnative binding modes. Consistent with the experiences of other groups[15] and our own work on protein–protein docking,[21,22] we find that the experimental *apo* structure contains information not fully captured by our energy function. As a result, we place harmonic restraints on the receptor C^{α} atoms, and we follow Wang *et al.*[22] in not penalizing the native side chain conformations (energy equal to the statistically best side chain conformation at that position).

As shown in the Results section, both the initial shape complementarity screen and the final minimization of the receptor backbone make important contributions toward accurate docking. In simplified protocols where the shape complementarity step is omitted, the ligand is placed in a random conformation, orientation, and location with its centroid within 5 Å of the native centroid. In simplified protocols where the backbone is held rigid, the final minimization includes all non-backbone degrees of freedom.

The coarse-grained first step of the protocol may appear wasteful in that only one pose is carried forward when a few tens of poses score equally well, and because similar poses may be explored by multiple trajectories. However, the first step takes only 3–10% of the time for each trajectory, and by not requiring multiple processes to communicate, we can easily achieve linear scaling with the number of available processors. This enables us to use thousands of processors simultaneously, if desired, including distributed computing schemes such as Rosetta@home§. Furthermore, we find empirically that repeated discovery of a low-energy pose in multiple trajectories provides partial evidence that that pose is in fact the correct native structure, both for RosettaLigand and for other Rosetta protocols.

All-atom Rosetta scoring function

The Rosetta scoring function used in the second and third stages consists of a weighted sum of physical and knowledge-based terms calculated from the three-dimensional coordinates of the receptor and ligand, including all hydrogen atoms. The function and its uses have recently been reviewed in Ref. 23 and a fuller description appears in Ref. 14; considerations specific to ligand docking are discussed in Ref. 9. In brief, the Rosetta energy E with weights w is given by $E = w\cdot(E_{vdw}, E_{sol}, E_{hb}, E_{sor}, E_{elec}, E_{restr})$. E_{vdw} is the Lennard–Jones 6–12 van der Waals energy; however, the function switches to linear at 60% or 91% of the ideal contact distance (hard- and soft-repulsive forms, respectively) to prevent extreme

§ http://boinc.bakerlab.org

energies from large clashes. E_{sol} is the electrostatic and hydrophobic desolvation energies calculated by Lazardis and Karplus' EEF1 implicit solvation model.[24] E_{hb} is an orientation-dependent hydrogen-bonding term derived from statistics of structures in the Protein Data Bank (PDB) and quantum mechanics calculations.[25,26] E_{tor} is protein side-chain and backbone torsional potentials derived from the PDB.[27] E_{elec} is a simple Coulombic electrostatics term with linearly distance-dependent dielectric, which is evaluated only between the receptor and ligand.[9] E_{restr} includes harmonic restraints to prevent chain breaks and limit backbone movement during the flexible-backbone minimization.

Analysis of docking results

For most of the results described here, we performed 5000 such docking trajectories for each receptor–ligand pair, requiring 40–80 processor-hours on our Linux cluster. (Integrated treatment of ligand conformers and fewer Monte Carlo cycles make this up to 3 orders of magnitude faster than the equivalent computations with the original ROSETTALIGAND.) The top 5% of structures were selected by total ROSETTA energy, and those were ranked by the receptor–ligand interaction energy; this eliminates minor noise from distant parts of the structure. Reported RMSDs account for any chemical symmetry within the ligand.

Algorithm implementation and parameters

The current version of ROSETTALIGAND is based on a new version of the ROSETTA core, rewritten from the ground up in object-oriented C++ by our laboratory and others. The entire receptor–ligand system is represented as a single tree of atoms ("atom tree" or "fold tree"), in which edges represent chemical bonds or rigid-body transformations.[23] In the calculations in this study, the tree was constructed for docking with loop modeling.[21] Nonflexible portions of the receptor backbone were held fixed relative to one another, and together defined a coordinate frame for rigid-body placement of the ligand. Each flexible segment of the receptor backbone was randomly divided into two pieces, which were attached to their respective flanking rigid segments. As in other ROSETTA protocols, a pseudo-energy term was included to maintain proper covalent geometry between the two pieces of the divided segments during minimization.

To make large receptors tractable, a generous region around the ligand was allowed to be flexible (typically 20–40 residues of backbone and 40–80 side chains), while the rest of the protein was held rigid. Side chains were included if any atom could reach within 6 Å of a ligand atom (considering all rotamers). Backbone torsions were varied if the C^{β} was within 7 Å of a ligand atom or if it was within 3 residues (sequence-wise) of such a residue. Docking results were not sensitive to the exact cutoffs used. To avoid any scoring artifacts, the entire receptor was minimized (side chains only) before beginning.[21] The ligand was not present during this prepacking step to avoid biasing the results.

The unified internal representation of protein residues and small molecules allows us to simultaneously optimize receptor and ligand. Taking advantage of this, we treat ligand conformations as "rotamers" that are sampled at the same time that protein side chains are repacked. This allows us to dock hundreds of ligand conformers in about the same time as a single conformer. Currently, ligand conformers are generated externally: we used the program Omega (v. 2.2.1, OpenEye) with its default settings and restrained ligand torsions with a harmonic potential while minimizing. [Future work will consider more elaborate conformational sampling and scoring schemes for ligands (K. Kaufmann and J. Meiler, unpublished results).] The crystal conformation was not included in the conformer library, as this led to artificially high success rates. Ligand atoms were represented by the most similar ROSETTA atom type (here determined automatically), and partial charges were assigned by OpenEye's AM1-BCC implementation. The methods for conformer generation, atom typing, and partial charges can easily be replaced by other procedures.

Availability

The ROSETTALIGAND source code is available through Rosetta Commons and the University of Washington Tech Transfer office at no cost to academic researchers.

SAMPL-1 blind docking experiment

The first annual SAMPL† was organized by Nichols and colleagues at OpenEye Inc. in early 2008. Participation was open to all and explicit invitations were issued to many academic and commercial groups; participants could also choose to submit predictions but remain anonymous in any published results. The experiment consisted of blind predictions on not publicly available data sets donated by private companies. There were three stages: virtual screening, docking, and affinity prediction; we only submitted docking predictions. The provided inputs for docking consisted of one structure of urokinase-type plasminogen activator and two structures of JNK3 kinase, both without ligands, and single 3-D structures of 34 putative urokinase ligands and 62 putative JNK3 ligands. Some of the provided compounds were decoys that did not actually bind these receptors. The possible presence of decoys was revealed in advance, but not their identity or number; compound numbers in Tables 2 and 3 are discontinuous due to decoys. For our blind predictions, the three best-scoring poses for each ligand were examined manually for consistency with observed binding patterns, and a few JNK3 ligands were manually placed near the assumed correct starting position. For consistency in comparison, the lowest RMSD of the top three poses is reported for the flexible-backbone algorithm, which was run after the correct answers were revealed, although in this case none of the ligands needed to be positioned manually. Most of the results have not yet been released to either the participants or the general public, although OpenEye has expressed the intention to publish a summary of the results at some point in the future.

Acknowledgements

We thank Andrew Leaver-Fay for design and implementation of core functionality in the new ROSETTA code base ("Mini"), Kristian Kaufmann and Jens Meiler for many helpful conversations, and OpenEye Inc. for providing software. I.W.D. gratefully acknowledges support from the University of Washington Genome Training Grant.

References

1. Leach, A. R., Shoichet, B. K. & Peishoff, C. E. (2006). Prediction of protein–ligand interactions. Docking and scoring: successes and gaps. *J. Med. Chem.* **49**, 5851–5855.
2. Sousa, S. F., Fernandes, P. A. & Ramos, M. J. (2006). Protein–ligand docking: current status and future challenges. *Proteins*, **65**, 15–26.
3. Osterberg, F., Morris, G. M., Sanner, M. F., Olson, A. J. & Goodsell, D. S. (2002). Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins*, **46**, 34–40.
4. Claussen, H., Buning, C., Rarey, M. & Lengauer, T. (2001). FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **308**, 377–395.
5. Zhao, Y. & Sanner, M. F. (2007). FLIPDock: docking flexible ligands into flexible receptors. *Proteins*, **68**, 726–737.
6. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. (2006). Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **49**, 534–553.
7. Cozzini, P., Kellogg, G. E., Spyrakis, F., Abraham, D. J., Costantino, G., Emerson, A. *et al.* (2008). Target flexibility: an emerging consideration in drug discovery and design. *J. Med. Chem.* **51**, 6237–6255.
8. Totrov, M. & Abagyan, R. (2008). Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **18**, 178–184.
9. Meiler, J. & Baker, D. (2006). ROSETTALIGAND: protein–small molecule docking with full side-chain flexibility. *Proteins*, **65**, 538–584.
10. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662.
11. Hartshorn, M. J., Verdonk, M. L., Chessari, G., Brewerton, S. C., Mooij, W. T., Mortenson, P. N. & Murray, C. W. (2007). Diverse, high-quality test set for the validation of protein–ligand docking performance. *J. Med. Chem.* **50**, 726–741.
12. Warren, G. L., Andrews, C. W., Capelli, A. M., Clarke, B., LaLonde, J., Lambert, M. H. *et al.* (2006). A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **49**, *5912–5931.*
13. Bradley, P., Misura, K. M. & Baker, D. (2005). Toward high-resolution de novo structure prediction for small proteins. *Science*, **309**, 1868–1871.
14. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.
15. Graves, A. P., Shivakumar, D. M., Boyce, S. E., Jacobson, M. P., Case, D. A. & Shoichet, B. K. (2008). Rescoring docking hit lists for model cavity sites: predictions and experimental testing. *J. Mol. Biol.* **377**, 914–934.
16. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A. & Baker, D. (2003). Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331**, 281–299.
17. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A. *et al.* (2008). De novo computational design of retro-aldol enzymes. *Science*, **319**, 1387–1391.
18. Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J. *et al.* (2008). Kemp elimination catalysts by computational enzyme design. *Nature*, **453**, 190–195.
19. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA*, **97**, 10383–10388.
20. Leaver-Fay, A., Kuhlman, B. & Snoeyink, J. (2005). *Workshop on Algorithms in Bioinformatics, Mallorca, Spain.*
21. Wang, C., Bradley, P. & Baker, D. (2007). Protein–protein docking with backbone flexibility. *J. Mol. Biol.* **373**, 503–519.
22. Wang, C., Schueler-Furman, O. & Baker, D. (2005). Improved side-chain modeling for protein–protein docking. *Protein Sci.* **14**, 1328–1339.
23. Das, R. & Baker, D. (2008). Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382.
24. Lazaridis, T. & Karplus, M. (2000). Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* **10**, 139–145.
25. Kortemme, T., Morozov, A. V. & Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. *J. Mol. Biol.* **326**, 1239–1259.
26. Morozov, A. V., Kortemme, T., Tsemekhman, K. & Baker, D. (2004). Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc. Natl Acad. Sci. USA*, **101**, 6946–6951.
27. Dunbrack, R. L. & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, *1661–6981.*