SKATE: Decoupling Systematic Sampling From Scoring

JIANWEN A. FENG, GARLAND R. MARSHALL
Department of Biochemistry and Molecular Biophysics, Center for Computational Biology, Washington University School of Medicine, 660 S. Euclid Ave, St. Louis, Missouri 63110

Received 25 March 2009; Revised 16 February 2010; Accepted 21 February 2010
DOI 10.1002/jcc.21545
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: SKATE is a docking prototype that decouples systematic sampling from scoring. This novel approach removes any interdependence between sampling and scoring functions to achieve better sampling and, thus, improves docking accuracy. SKATE systematically samples a ligand's conformational, rotational and translational degrees of freedom, as constrained by a receptor pocket, to find sterically allowed poses. Efficient systematic sampling is achieved by pruning the combinatorial tree using aggregate assembly, discriminant analysis, adaptive sampling, radial sampling, and clustering. Because systematic sampling is decoupled from scoring, the poses generated by SKATE can be ranked by any published, or in-house, scoring function. To test the performance of SKATE, ligands from the Asetex/CDCC set, the Surflex set, and the Vertex set, a total of 266 complexes, were redocked to their respective receptors. The results show that SKATE was able to sample poses within 2 Å RMSD of the native structure for 98, 95, and 98% of the cases in the Astex/CDCC, Surflex, and Vertex sets, respectively. Cross-docking accuracy of SKATE was also assessed by docking 10 ligands to thymidine kinase and 73 ligands to cyclin-dependent kinase.

 J Comput Chem 00: 000–000, 2010

Key words: consensus scoring drug discovery; discriminant analysis; systematis search; virtual screening

AQ2

Introduction

Small molecule docking programs are used extensively in the pharmaceutical industry and increasingly in academia for the discovery of novel lead compounds. A number of docking programs are available either as commercial software or from academic labs.[1-13] Molecular docking programs have three main components: a representation of the system, a sampling algorithm and a scoring function.[14] A docking program must be able to sample near-native poses to rank them as top-scoring poses. A pose defines the relative orientation and conformation of a ligand when bound to a receptor. Velec et al. noted that highly accurate ligand poses (below 1 Å root-mean-square deviation) are a prerequisite to improving scoring functions.[15]

Evolutionary algorithms and other stochastic search methods are common types of sampling algorithm. They rely on scoring functions to guide their stochastic steps, so the search and scoring processes are necessarily coupled. Scoring functions need to evaluate anywhere from thousands to millions of poses in a docking experiment. To speed up the calculations, the energy functions are simplified so that they can be evaluated quickly. The tradeoff is a less accurate energy function that at best approximates the binding energy of a pose. If a coarse energy function scores a near-native pose poorly, the pose will be discarded. This problem of false negatives is often the root cause of poor docking performance. The interdependence of sampling and scoring in current docking programs makes it difficult to determine whether a sampling error or a scoring error caused poor performance in a docking experiment. SKATE is a novel docking program that decouples systematic sampling from imperfect scoring. It employs a rigorous search method to systematically sample conformational, orientation and rotational degrees of freedom of a ligand to find optimal docking poses. Any naive brute-force approach, literally rotating each bond, results in combinatorial explosion and becomes computationally intractable. In SKATE, efficient systematic sampling was achieved by pruning the combinatorial tree using aggregate assembly, discriminant analysis, adaptive sampling, radial sampling and clustering. The resulting sterically allowed poses of a ligand bound to a receptor were then ranked independently using three scoring functions. The docking performance of SKATE was evaluated by three large test sets in terms of self-docking, and two test sets in terms of cross-docking.

Correspondence to: G. R. Marshall; e-mail: garland@biochem.wustl.edu

Contract/grant sponsor: NIH; contract/grant number: RO1 GM068460

Contract/grant sponsor: Computational Biology Training; contract/grant number: GM 008802

Contract/grant sponsor: Division of Biology and Biomedical Science of Washington University, Kauffman Foundation

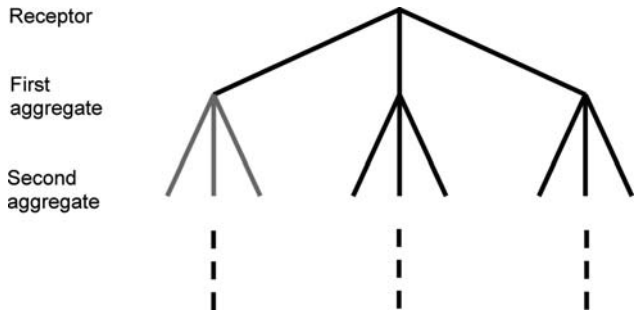


Figure 1. The tree structure of systematic search of conformational space for a ligand hydrogen-bonded to a receptor. Vertices of the tree represent ligand aggregates; edges represent discrete torsion values of a ligand's rotatable bonds and a ligand-receptor hydrogen-bond. Red edges represent "pruning" of the search tree by eliminating branches of the tree where the addition of an aggregate is sterically prohibited for any torsion value. Sterically allowed conformations are represented by the tree leaves that are connected by black edges. The first aggregate is hydrogen-bonded to the receptor and the bonding geometries are determined from a set of geometric parameters. At each branch point, a new aggregate may be added to the existing partial conformation if it is sterically allowed (black lines). Each black line represents a torsion value of a rotatable bond where an aggregate is added to the existing partial molecule. The assembly of a sterically allowed conformation continues until aggregates along every branch have been systematically evaluated. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Overview of Docking sMethodology and Data Sets

Sampling

Hydrogen-bonding interactions are essential in drug specificity and high affinity binding. SKATE takes advantage of this natural phenomenon by forming all possible hydrogen-bond pairings between the ligand and the receptor pocket to anchor systematic search. Once a sterically allowed hydrogen bond is formed between a receptor atom and a ligand atom, SKATE systematically samples the ligand's torsional degrees of freedom. The simplest systematic approach to find all sterically allowed conformations of a flexible ligand is to iteratively rotate each rotatable bond. Assuming a ligand molecule of N atoms with T rotatable bonds, and a receptor pocket of M atoms, if each rotatable bond of the ligand is explored at angular increments of A degrees, there are 360/A values to be examined for each T resulting in $(360/A)^T$ possible conformations to be examined for steric conflict. The 3D coordinates that determine the geometry of a conformation can be generated by applying appropriate transformation matrices to different subsets of atoms. These conformers must be checked for van der Waals (VDW) overlap to eliminate sterically impossible conformations. To a first approximation, there are N(N-1)/2 pair-wise distance calculations that must be performed for each conformation. Then M x N pair-wise distance calculations must be performed between atoms in each conformation and those in the receptor pocket. These distances are checked against the allowed sum of VDW radii for the two atoms involved. The number of VDW comparison V for a single hydrogen-bond formed between the receptor and the ligand is given by $V = \left(\frac{360}{A}\right)^T \times \left(\frac{N(N-1)}{2} + M \times N\right)$. The rate-limiting step in this brute force approach is the sheer number of VDW comparisons that must be performed to find sterically allowed poses. For example, sampling at torsional increments of 10 degrees for a ligand with 6 rotatable bonds and 50 atoms and a receptor pocket of 1000 atoms will result in 1.4 e13 VDW calculations. Assuming there is a combination of 50 possible hydrogen-bonds that can be formed, it would take 22 years to complete this calculation on a modern, single CPU computer that is capable of processing 1 million VDW comparisons per second.

Such a brute force approach to systematic search is obviously both inefficient and unnecessary. SKATE implements a number of strategies that truncate the combinatorial explosion. Sterically allowed poses of a ligand as constrained by a receptor pocket are systematically sampled by a step-wise build up of aggregates (Fig. 1). An aggregate is defined as a set of atoms whose relative positions are invariant to rotational degrees of freedom.[16] A ligand is divided into individual aggregates around internal rotatable bonds (Fig. 2). An aggregate capable of hydrogen-bonding is transformed by rigid-body translation and rotation to form an energetically favorable hydrogen-bond with the receptor. The geometries of the newly formed hydrogen-bond are determined by a set of hydrogen-bonding geometric parameters. A second aggregate that shares a common rotatable bond with the first aggregate is spliced onto the partial molecule by applying the appropriate transformations. The range of sterically allowed torsions for this rotatable bond is analytically determined by discriminant analysis.[16] Discrete values in the range of allowed torsions are sampled by rotating the second aggregate around the rotatable bond that joins the first and second aggregates. The step-wise assembly of sterically allowed conformations of the ligand within the receptor pocket continues until all aggregates have been added. As shown in Figure 1, the possible conformations of a flexible ligand hydrogen-bonded to a receptor can be represented by a search tree. The tree is anchored by a receptor atom that forms a hydrogen bond with the ligand. SKATE systematically finds sterically allowed ligand poses (tree leaves) by performing a depth-first search of the tree. Systematic search is performed for each possible pairing of hydrogen-bonding atoms between the ligand and receptor. A more detailed explanation of systematic search and discriminant analysis is provided in the Methods section.

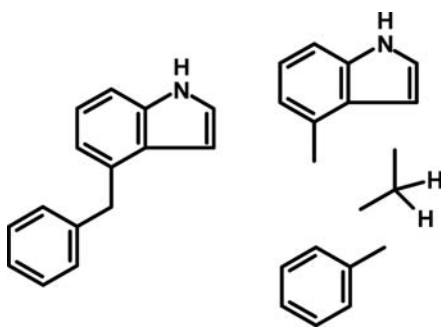


Figure 2. A simple molecule (left) is divided into its aggregates (right) by partition at its rotatable bonds.

Scoring

SKATE decouples systematic sampling from scoring. A unique feature of SKATE is that any scoring function may be used to rank the ensemble of poses generated by SKATE. SKATE itself does not use a scoring function per se to determine if a pose is low energy, but relies on a calibrated set of VDW radii and the assumption of a hydrogen bond with a set of defining parameters. It uses discriminant analysis and incremental build-up of the ligand to find the set of sterically allowed poses. Those poses are clustered with a heavy atom root-mean-square deviation (RMSD) cutoff of 0.5 Å. In this work, we used energy functions in FRED,[17] Rosetta[6] and X-Score[18] to rank or score the poses generated by SKATE. These scoring functions are made available by their respective authors at no charge to academic groups.

FRED or Fast Rigid Exhaustive Docking is a commercial docking program developed by OpenEye that can also be used to score poses generated by other programs.[17] We used FRED's default consensus scoring function that is an equal-weighted sum of ranks by chemgauss3, PLP, and oechemscore. Chemgauss3 uses smooth Gaussian functions to represent the shape and chemistry of molecules.[17] PLP, or Piecewise Linear Potential, is a minimal scoring function that includes a steric term and a hydrogen-bonding term, but no electrostatic term.[19] Oechemscore is an OpenEye variant of chemscore, an empirical scoring function.[20] We also examined how FRED scoring is affected by a fast, rigid-body local optimization of SKATE-generated poses before scoring.

Rosetta's energy function was originally trained for protein structure prediction and was extended to score protein-ligand interactions.[15] The energy function consists of a weighted sum of force-field-based and knowledge-based terms calculated from the receptor and ligand coordinates. Hydrogen atoms are explicitly treated. The terms include VDW interactions, an implicit solvent model, an explicit orientation-dependent hydrogen-bonding potential, and an electrostatics model. For this work, we used Rosetta's energy function, referred to as Rosetta-Score, to rank poses generated by SKATE.

X-Score is an empirical scoring function that treats hydrophobic effect by using three different functions and averaging the results.[18] Each of the three functions includes a VDW interaction term, a hydrogen-bonding term, a hydrophobic effect term, a torsional-entropy penalty, and a regression constant. X-score was trained to reproduce the known binding affinity of 200 protein-ligand complexes.

Data Sets

SKATE was tested on five data sets in assessing its self-docking and cross-docking performance. Results from four of the data sets can be compared directly to results from published docking programs.

Astex/CCDC Diverse Set

Hartshorn et al. prepared a set of 85 high-quality and diverse protein-ligand complexes and made them publicly available as a validation set for testing docking performance.[21] Protein targets were selected based on their relevance to drug discovery or agrochemical research. Consequently, only complexes with drug-like ligands were allowed in this set. To ensure complex diversity, no receptor was represented more than once. Furthermore, the ligands contained distinct molecular recognition types. A special focus was placed on selecting very high-quality experimental structures for which the experimental binding mode of the ligands was easily assessed. Protein structures were prepared by removing solvents and small ions. Exceptions were made for water molecules that coordinate a metal ion and for small ions that mimic a cofactor. His, Asn and Gln side-chain placements in the crystal structure that were not consistent with hydrogen-bonding patterns were rotated if such rotations would significantly improve hydrogen-bonding. This is reasonable because crystallographers usually cannot orient His, Asn, and Gln side chains with absolute certainty based on electron density alone. This data set was downloaded from the Cambridge Crystallographic Data Centre (http://www.ccdc.cam.ac.uk).

Surflex Set

To compile a test set for Surflex, Jain filtered 134 protein-ligand complexes in the GOLD data set by removing complexes that (i) contained ligands with more than 15 rotatable bonds, (ii) were covalently attached to the protein, and (iii) contained obvious errors in structure.[1,22] The resulting 81 complexes were made available on http://jainlab.ucsf.edu. The protein files in the original GOLD set were prepared by removing water molecules and by adding hydrogen atoms while taking protonation states into account. Exceptions were made to keep water molecules and metal atoms that coordinated ligand binding.[1]

Vertex Set

Perola et al. prepared a test set of 150 protein-ligand complexes to compare the performances of Glide, GOLD and ICM.[23] These complexes were selected for their relevance to modern drug discovery programs. Ligands were selected for (i) their drug-like properties; (ii) molecular weights between 200 and 600 Daltons; (iii) having between 1 and 12 rotatable bonds; and (iv) structural diversity. The ligands in the Vertex set were prepared by extracting them from their respective PDB files and assigning bond orders and correct protonation states by visual inspection. Protein structures were prepared by removing subunits, ions, solvent and other small molecules not involved in binding. Metal ions and tightly bound water molecules in the ligand binding site were preserved.[23] Hydrogen atoms were added to the protein. The structures of ligand, protein, and cofactor were minimized as a complex for 1,000 steps using Macromodel and the OPLS-AA force field. All heavy atoms were constrained to their original positions during minimization. The structures with optimized hydrogen positions were saved. Of the 150 complexes, 100 are PDB entries and 50 are corporate structures. The files of the 100 PDB complexes are available on the Jain Lab website (http://jainlab.ucsf.edu).[10] Seven complexes in the Vertex set are also included in either the Astex/CDCC set or the Surflex set (Table 1). T1

Thymidine Kinase Set

Bissantz et al. tested the virtual screening capability of docking programs by using the crystal structure of HSV-1 thymidine ki-

Table 1. Results for Skate on the Astex/CDCC, Surflex and Vertex complexes.[a]

Vertex Set				Astex/CDCC set				Surflex set			
		RMSD				RMSD				RMSD	
PDB code	nrot	Best pose	Top rank	PDB code	nrot	Best pose	Top rank	PDB code	nrot	Best pose	Top rank
13gs	3	0.49	0.83	1g9v	5	1.49	1.51	1abe	0	0.45	0.45
1a42	8	0.77	1.41	1gkc	8	1.27	0.95	1acj	1	0.46	0.68
1a4k	5	0.66	1.81	1gm8	4	1.59	2.24	1ack	2	0.53	3.83
1a8t	8	1.00	7.99	1gpk	1	0.27	0.30	1acm	6	0.70	0.65
1afq	9	0.86	8.23	1hnn	1	0.67	0.98	1aco	2	0.27	0.35
1aoe	3	0.48	0.86	1hp0	2	0.45	0.36	1aha	0	0.26	0.18
1atl[b]	8	0.91	1.26	1hq2	1	0.43	0.26	1atl	9	1.35	2.86
1azm	2	0.51	1.28	1hvy	8	1.77	1.64	1baf	4	0.64	0.92
1bnw	5	0.64	5.48	1hwi	9	0.61	1.11	1bbp	9	0.75	0.76
1bqo	6	0.30	0.48	1hww	1	0.22	0.14	1bma	9	1.65	2.47
1br6	3	0.39	1.14	1ia1	2	0.26	0.36	1cbs	0	0.20	0.36
1cet	7	0.93	7.45	1ig3	4	0.44	1.20	1cbx	5	0.29	0.43
1cim	3	0.28	1.09	1j3j	2	0.18	0.30	1com	4	0.46	0.79
1d3p	12	1.14	1.19	1jd0	1	0.72	3.36	1coy	1	0.32	0.51
1d4p	3	0.24	0.60	1jje	7	0.58	7.97	1dbb	1	0.25	0.51
1d6v	7	0.92	2.17	1jla	7	0.70	0.77	1dbj	1	0.32	0.54
1dib	7	0.80	2.88	1k3u	6	0.27	0.29	1dr1	3	0.28	1.48
1dlr	4	0.42	0.64	1ke5	1	0.34	0.29	1dwd	8	1.16	2.97
1efy	3	0.42	1.76	1kzk	9	0.65	0.89	1eap	10	0.82	0.81
1ela	8	0.44	0.68	1l2s	2	0.31	0.51	1epb	0	0.91	0.74
1etr[b]	8	0.46	0.60	1l7f	8	0.33	0.44	1etr	8	0.92	0.93
1ett	6	0.51	0.98	1lpz	6	0.71	1.00	1fen[d]	0	–	–
1eve	6	1.38	1.01	1lrh	2	1.32	1.42	1fkg	9	0.78	1.60
1exa	4	0.25	0.32	1m2z	3	0.19	0.60	1fki	0	0.30	0.34
1ezq	10	0.39	0.71	1meh	7	1.12	1.07	1frp	7	0.26	0.92
1f0r	4	0.40	0.75	1mmv	8	0.81	0.58	1glq	12	1.62	9.14
1f0t	5	0.73	2.57	1mzc	7	1.27	2.26	1hdc	6	1.41	1.61
1f4e	2	0.41	1.09	1n1m	3	0.82	0.57	1hdy	0	0.90	0.74
1f4f	8	0.67	2.23	1n2j	4	0.67	0.47	1hri	9	2.87	10.18
1f4g	11	1.23	1.49	1n2v	3	0.45	1.08	1hsl	4	0.36	0.42
1fcx	4	0.28	0.32	1n46	5	0.43	0.66	1hyt	5	0.65	0.78
1fcz	3	0.26	0.39	1nav	5	0.39	0.73	1lah	6	0.36	0.36
1fjs	8	1.16	2.01	1of1	2	0.32	0.32	1lcp	4	0.56	1.13
1fkg[b]	9	0.64	1.33	1of6	4	0.32	0.64	1ldm	0	0.18	0.39
1fm6	6	0.68	0.74	1opk	2	0.35	0.56	1lic	15	5.05	5.07
1fm9	11	0.48	2.31	1oq5	3	0.37	5.00	1lna	10	0.64	0.74
1frb	5	0.24	0.23	1owe	2	1.01	1.84	1lpm	8	0.89	6.82
1g4o	5	1.04	3.59	1oyt	4	0.40	0.62	1lst	7	0.29	0.21
1gwx	10	1.61	2.19	1p2y	1	1.67	4.87	1mdr	3	0.20	0.47
1h1p	3	0.43	0.43	1p62	3	0.16	0.40	1mrg	0	0.29	0.59
1h1s	4	0.46	0.66	1pmn	6	2.51	6.70	1mrk	3	0.36	0.99
1h9u	3	0.26	0.39	1q1g	3	0.36	0.69	1nco	9	0.91	0.68
1hdq	5	0.98	1.03	1q41	1	0.34	0.54	1phg	3	0.87	4.39
1hfc	10	0.60	0.52	1q4g	3	0.27	0.64	1rds	8	1.03	1.74
1hpv	12	0.88	0.89	1r1h	10	0.43	0.53	1rob	5	0.94	1.41
1htf	13	0.92	2.10	1r55	8	0.98	0.86	1snc	6	0.54	0.80
1i7z	5	0.39	0.48	1r58	9	0.77	0.91	1srj	2	0.40	0.40
1i8z	6	4.82	4.80	1r9o	3	0.52	0.76	1stp	5	0.40	0.79
1if7	7	0.88	5.13	1s19	5	0.38	0.62	1tka	8	1.21	1.46
1iy7	5	0.30	0.64	1s3v	5	0.38	0.73	1tmn	13	0.75	1.48
1jsv	1	0.46	0.39	1sg0	3	0.32	0.49	1tng	2	0.10	0.69
1k1j	8	0.45	1.61	1sj0	6	0.54	0.66	1tni	5	0.57	1.92
1k22	9	0.42	0.49	1sq5	6	0.81	1.68	1tnl	2	0.19	0.40

(*continued*)

Table 1. (*Continued*).

Vertex Set				Astex/CDCC set				Surflex set			
		RMSD				RMSD				RMSD	
PDB code	nrot	Best pose	Top rank	PDB code	nrot	Best pose	Top rank	PDB code	nrot	Best pose	Top rank
1k7e	4	0.18	1.00	1sqn	2	0.22	0.27	1trk	9	0.40	0.58
1k7f	5	0.66	1.20	1t40	6	0.65	0.69	1ukz	4	0.20	0.24
1kv1	1	0.24	0.81	1t46	4	0.43	0.38	1ulb	0	0.20	0.46
1kv2	6	0.53	0.55	1t9b	3	0.61	0.47	1wap	4	0.17	0.32
1l2s[c]	1	0.17	0.42	1tow	4	0.56	4.81	2ada	3	0.15	0.19
1l8g	3	0.33	2.13	1tt1	4	0.29	0.73	2ak3	4	0.42	0.55
1lqd	5	0.56	1.00	1tz8	5	0.58	2.29	2cgr	5	1.66	1.80
1m48	7	0.56	0.96	1u1c	6	0.60	1.12	2cht	3	0.67	1.36
1mmb	13	0.82	6.79	1u4d	1	0.28	0.91	2cmd	6	0.57	0.45
1mnc	10	0.80	1.55	1uml	9	0.54	0.62	2ctc	4	0.24	0.41
1mq5	3	0.26	0.41	1unl	6	0.55	0.95	2dbl	6	1.62	1.35
1mq6	4	0.27	0.32	1uou	2	0.73	0.73	2gbp	2	0.19	0.26
1nhu	8	0.44	0.46	1v0p	6	0.50	0.45	2lgs	5	1.13	1.74
1nhv	8	1.15	7.46	1v48	6	0.45	0.35	2phh	2	0.25	0.41
1o86	12	7.97	8.63	1v4s	3	0.29	0.28	2r07	8	1.64	8.96
1ohr	11	0.39	0.46	1vcj	8	0.59	0.84	2sim	5	0.40	1.21
1ppc	8	1.08	1.96	1w1p	0	0.24	0.28	3aah	3	0.79	0.39
1pph	6	0.76	2.03	1w2g	2	0.39	0.56	3cpa	6	0.88	0.92
1qbu	10	0.72	0.77	1x8x	4	0.44	0.78	3hvt	1	2.33	1.62
1qhi	4	0.25	0.40	1xm6	5	0.53	1.23	3ptb	1	0.23	0.42
1ql9	3	0.37	0.37	1xoq	5	0.95	4.03	3tpi	7	0.26	0.26
1qpe	2	0.42	0.55	1xoz	1	0.26	0.28	4cts	2	0.27	0.44
1r09	3	0.75	0.60	1y6b	6	0.39	0.34	4dfr	8	0.84	1.19
1syn	7	0.59	2.48	1ygc	10	2.55	3.85	6abp	1	0.34	0.39
1thl	11	0.82	0.95	1yqy	4	0.18	0.51	6rnt	4	0.38	7.01
1uvs	8	1.30	1.49	1yv3	2	0.30	0.31	6rsa	2	0.47	0.76
1uvt	5	0.49	0.45	1yvf	4	0.58	0.60	7tim	3	0.47	1.13
1ydr	1	0.36	0.36	1ywr	5	0.54	0.45	8gch	8	1.56	2.20
1yds	4	0.44	0.37	1z95	5	0.31	0.34				
1ydt	7	0.87	3.46	2bm2	7	0.45	1.48				
2cgr[b]	5	0.60	0.78	2br1	6	1.12	1.58				
2csn	4	1.54	3.01	2bsm	6	0.50	0.74				
2pcp	2	0.24	0.30								
2qwi	5	0.24	1.01								
3cpa[b]	7	0.45	1.07								
3erk	3	0.25	0.60								
3ert	8	0.48	1.03								
3std	5	0.26	0.26								
3tmn	6	0.47	0.71								
4dfr[b]	8	0.98	1.42								
4std	4	0.26	0.35								
5std	4	0.52	0.32								
5tln	9	0.80	1.06								
7dfr	8	0.77	1.44								
7est	6	0.43	0.82								
830c	7	0.29	0.52								
966c	7	0.30	0.63								

[a]Top rank poses are scored by FRED-Opt-Score.
[b]Complex is also part of the Surflex set.
[c]Complex is also part of the Astex/CDCC set.
[d]The ligand in complex 1fen does not have any atom that is capable of hydrogen bonding.

J_ID: ZQY Customer A_ID: 09-0130.R3 Cadmus Art: JCCT21545 Date: 27-MARCH-10 Stage: I Page: 6

nase (TK) (PDB ID: 1KIM), 10 known ligands, and 990 randomly chosen decoys.[24] In this work, the 10 known ligands were docked to the 1KIM structure to test SKATE's performance in cross-docking. The structures were prepared as described in Bissantz et al. No optimization of ligand or receptor coordinates was performed.

Cyclin-Dependent Kinase 2 Set

Seventy-three known ligands that have been cocrystallized with cyclin-dependent kinase 2 (CDK2) were docked to a single high resolution CDK2 structure (PDB ID: 2B54, 1.85 Å).[25] These ligands occupy the ATP-binding site of CDK2. To prepare the receptor, water molecules and cocrystallized ligands were removed from the 2B54 structure, and hydrogen atoms were added to the receptor using Sybyl 8.1.[26] The ligand structures were extracted from their respective complexes, and were assigned correct bond orders and protonation states by visual inspection. To create reference coordinates of the 73 known ligands, their respective cocrystallized receptors were aligned to the 2B54 structure and the ligands were extracted and saved as mol2 files.

Docking Setup

The files in the Astex/CDCC, Surflex and Vertex sets were downloaded from their respective websites and used as obtained. Hydrogen atoms were already added to protein and ligand structures by the test sets' respective authors. No further optimization of protein or ligand geometries were performed since it could lead to biased results.[27] For the Vertex set, its author did minimize the ligand and receptor hydrogen atoms while constraining the heavy atoms to their original locations.[23] Ligand and protein files in PDB or MOL formats were converted to the mol2 format and assigned Tripos atom types. The coordinates of these ligand files were used as references when calculating RMSD values.

In this study, the experimentally determined ligand was used to define the binding pocket for the purpose of docking. A list of potential hydrogen-bond donors and acceptors were created by inspecting the receptor atoms that are within 5 Å of the cocrystallized ligand. The coordinates from this list of atoms served as anchor points in systematic search. SKATE attempted to dock all possible pairings of ligand hydrogen-bond donors with protein hydrogen-bond acceptors, and vice versa. A vast majority of these attempted pairings resulted in immediate search termination because they were not sterically allowed. The resulting sterically allowed poses generated by SKATE were written to a file in the mol2 format to be ranked by scoring functions.

Results and Discussion

Sampling Accuracy

To rank a near native pose as the top-scoring pose, a docking program must be able to sample such poses. The interdependence of sampling and scoring in current docking programs makes it difficult to determine whether it is a sampling error or a scoring error that caused a program to fail on a given test case. SKATE approaches the docking problem by decoupling systematic sampling from scoring. It anchors a search by pairing a ligand hydrogen-bond donor to a receptor hydrogen-bond acceptor and vice versa. For each hydrogen-bond formed, SKATE systematically samples a ligand's torsional degrees of freedom to find poses that sterically fit within a receptor pocket. To delete conformational memory of the experimentally determined ligand structures, their torsions were reset to 180 degrees before docking. Bond angles, bond lengths and ring conformations were not modified. Receptor atoms within 5 Å of the cocrystallized ligand defined the binding pocket. Figure 3 shows the cumulative proportion of best poses, as measured by RMSD to the experimental structure (reference), that were generated by SKATE for the complexes in the Astec/CDCC, Surflex and Vertex self-docking test sets. A pose is considered best if its heavy atom RMSD to the reference structure is the lowest. Table 1 lists the RMSD values of the best poses and top-scoring poses for each complex in the three test sets. For an RMSD threshold of 2 Å, sampling accuracy rates are 98, 95, and 98% for the Astex/CDCC, Surflex and Vertex sets, respectively. For an RMSD threshold of 1 Å, the respective sampling accuracy rates are 86, 80, and 88% for the three self-docking sets. Highly accurate ligand poses that approximate the native pose below 1 Å RMSD are a prerequisite to improving solutions to the scoring problem.[15] For all but a few test cases in the three test sets, SKATE was able to sample poses that were within 2 Å. The highly accurate sampling of SKATE can be attributed to the systematic sampling algorithm. It is essential for a docking program to sample near-native poses of the complex to give scoring functions the opportunity to rank them as top-scoring poses.

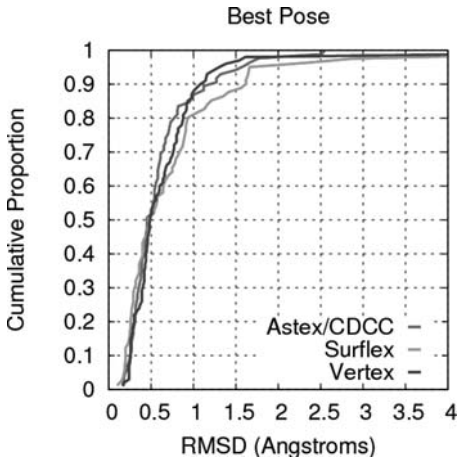


Figure 3. Cumulative proportion of best RMSD poses for the Astex/CDCC (red), Suflex (green), and Vertex (blue) sets. There are 85 complexes in the Astex/CDCC set, 81 complexes in the Surflex set and 100 complexes in the Vertex set. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Systematic sampling in SKATE never repeatedly samples the same point in conformational space. In practice, two conformations can be clustered when their only difference is a 10 degree torsional variation in a terminal rotatable bond. To speed up sampling, SKATE implements heuristics to further reduce conformational space. As shown in Figure 1, the possible conforma-

ID: jaishankarn Date: 27/3/10 Time: 20:40 Path: N:/3b2/JCCT/Vol00000/100045/APPFile/C2JCCT100045

tions of a ligand that is hydrogen-bonded to a receptor can be represented by a search tree. The edges, nodes, and leaves of the tree represent torsion values of rotatable bonds, aggregates and sterically allowed poses, respectively. SKATE traverses this tree using a depth-first search approach. Upon reaching a leaf of the tree by traversing down a branch from the root, a sterically allowed conformation is found. SKATE determines if it is necessary to travel down a branch of the tree by checking if the partial ligand constructed thus far is similar to a ligand pose that was already found from visiting previous tree branches. If the RMSD between atoms in a partial ligand and the corresponding atoms in a previously generated pose is less than 0.3 Å, then SKATE terminates the search of the current branch. If the search were to be continued, the resulting poses would be very similar to the previously generated pose and would be discarded during clustering.

Discriminant analysis determines the range of torsions that are sterically allowed for a rotatable bond. The allowed range of torsions is discretized and converted into a list of torsions to be sampled. Not all conformers assembled from this list of values will be low in energy. Ligands in the receptor-bound state rarely adopt strained conformations where the torsions of rotatable bonds deviate significantly from the energy minima of the +gauche, -gauche, and anti rotations. SKATE truncates conformational space by limiting allowed torsions to be within 30 degrees of +gauche, -gauche and anti torsions for rotatable bonds that (i) are not terminated by oxygen or sulfur atoms, and (ii) contain atoms that are bonded to fewer than four heavy atoms. For terminal aggregates of a ligand, only the three torsions that are nearest to +gauche, -gauche and antivalues are sampled for rotatable bonds that meet the above criteria (i) and (ii). SKATE uses a combination of 180 geometric parameters to predict potential hydrogen-bonding interactions between a ligand acceptor and receptor donor, and vice versa. The parameters that represent the most common geometries are tried first. SKATE skips the remaining parameters if a pose is found. These heuristics that reduce the search space to speed up performance are optional and can be enabled or disabled by the user.

Analysis of Failed Sampling Cases

SKATE was able to sample a pose that is within 2 Å RMSD of the reference structure for 98, 95, and 98% of the test cases in the Astex/CDCC, Surflex, and Vertex data sets, respectively. Two of the ligands in the 85 complexes Astex/CDCC set barely missed the 2 Å RMSD threshold; their RMSD values were 2.51 and 2.55. SKATE was unable to sample a pose that was within 2 Å RMSD of the native structure for test cases 1O86 and 1I8Z in the Vertex set. In 1O86, lisinopril, an antihypertension drug, is bound to the human angiotensin converting enzyme (ACE). There are 12 rotatable bonds in lisinopril. The ACE active site consists of a zinc-coordinated narrow groove flanked by two large hydrophobic pockets. Poses found by SKATE occupied either one of the pockets exclusively but were not able to bridge the two. To correctly dock lisinopril to ACE, a docking program must sample a pose where the carboxyl group of lisinopril correctly coordinates the zinc atom and still fits sterically into a very narrow channel. For test case 1I8Z, the ligand also coordinates a zinc atom, but SKATE failed to generate a pose that captured this interaction. For the Surflex set, SKATE failed to find near-native poses for 1FEN, 1HRI, 1LIC, and 3HVT. The 1FEN ligand does not have any hydrogen-bonding atoms and SKATE could not anchor its search since it could not form a hydrogen bond between the ligand and the receptor. For 1HRI, the ligand does not form a hydrogen bond with the receptor. SKATE sampled a pose (RMSD = 2.87 Å) where the ligand did form a hydrogen bond with the receptor but its orientation was inverted. Similarly, the 3HVT ligand does not form a hydrogen bond with its receptor and the best pose RMSD value was 2.33 Å. The 1LIC ligand is a simple alkyl chain molecule that has 15 rotatable bonds; it is a poor candidate for testing docking programs because it does not represent drug- or lead-like compounds and should not have been included in the Surflex test set. For the Astex/CDCC and the Vertex sets, SKATE sampled near-native poses ($\leq$2.0 Å) for 98% of the test cases. Ligands in the Astex/CDCC were selected for an unambiguous fit to the experimental electron density. Protons in the Vertex set were optimized to alleviate poor steric contacts. The likelihood of intermolecular penetration of VDW surfaces in these two test sets is lower because of high structural resolution in one case and proton optimization in another.

PDB structures are static models that best fit the available electron density data. Errors in lower resolution structures may result in poor modeling of small molecule ligands. This could lead to poor intermolecular steric contacts and even incorrect fitting of the electron density.[28] It is important to keep this in mind when assessing a docking program's ability to reproduce experimentally determined ligand poses.

Scoring Accuracy

SKATE focuses on the systematic sampling of sterically allowed poses of a ligand where its search space is constrained by a binding pocket. It does not provide a scoring function to rank order the generated poses *per se*, but takes advantage of the many published scoring functions' ability to rerank docked poses. In this article, we presented data from using X-Score, Rosetta, and FRED energy functions to rank SKATE-generated poses. X-Score is an empirical scoring function that estimates the hydrophobic effect by using three different functions and averaging the results.[18] Rosetta's energy function was originally trained for protein-structure prediction and was extended to score protein-ligand interactions.[15] In this article, Rosetta's energy function will be referred to as Rosetta-Score. FRED[17] itself is a docking program, but could also be used to rank previously generated poses with a consensus scoring function that consists of chemgauss3, PLP, and oechemscore. It will be referred to as FRED-Score.

We also evaluated whether rigid-body, local optimization of SKATE-generated poses would improve overall docking performance. SKATE allows some VDW penetration by scaling atomic VDW radii within the systematic sampling algorithm (see methods section for details). FRED's consensus scoring function could rank a near native pose poorly due to poor contacts. Before scoring, poses were optimized by performing fast, small-scale, rigid-body translations (0.75 Å) or rotations (0.5 Å),

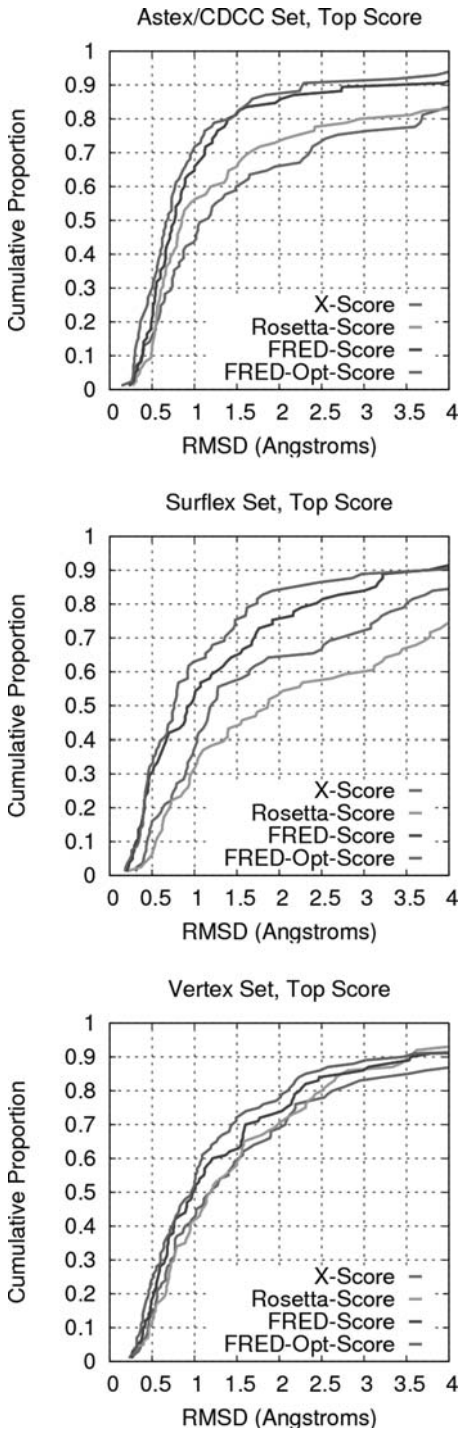


Figure 4. Cumulative proportion of top scoring RMSD for 85 complexes in the Astex/CDCC set (A), 81 complexes in the Surflex set (B) and 100 complexes in the Vertex set (C). Poses were generated by SKATE and were ranked by X-Score(red), Rosetta-Score (green), FRED-Score (blue), and FRED-Opt-Score (magenta). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

a total of 72 systematic transformations, using FRED.[17] The optimized pose was selected by using the PLP scoring function. The receptor atoms were fixed throughout the optimization process. We emphasize that we only used the rigid-body, local optimization feature of FRED, not its full-fledged docking capabilities. The process of optimizing and scoring with FRED will be referred to as FRED-Opt-Score.

The results of using X-score, Rosetta-Score, FRED-Score, and FRED-Opt-Score to rank SKATE-generated poses for the Astex/CDCC test set are shown in Figure 4A. For an RMSD threshold of 2.0 Å, the success rates were 87%, 85%, 73%, and 66% for FRED-Opt-Score, FRED-Score, Rosetta-Score and X-Score, respectively. SKATE coupled with FRED-Opt-Score ranking performed particularly well in identifying poses that were less than 1 Å RMSD as the best pose for the Astex/CDCC set. Its accuracy rate was 72%. This is very encouraging because only 86% of the test cases had a pose that was less than 1 Å RMSD. Taking that into account, the scoring accuracy rate is 84% for ranking a pose that is within 1 Å RMSD from the native structure. This could partly be attributed to the high quality of the x-ray structures comprising the Astex/CDCC set. F4

Similar to the results in the Astex/CDCC set, FRED-Opt-Score performed best in identifying poses that were within 2 Å RMSD as the best pose for the Surflex set. FRED-Opt-Score's accuracy rate was 84% (Fig. 4B). FRED-Score, X-Score and Rosetta-Score's accuracy rates were 75%, 64%, and 52%, respectively.

The results for the Vertex test set are shown in Figure 4C. For an RMSD threshold of 2.0 Å, the success rates were 77%, 73, 70, and 69% for FRED-Opt-Score, FRED-Score, Rosetta-Score and X-Score, respectively. For an RMSD threshold of 1 Å, the scoring accuracy of FRED-Opt-Score was 53% and that of FRED-Score was 50%. The performance of FRED-Score and that of FRED-Opt-Score were comparable in identifying a pose that is within 1 Å RMSD as the best pose.

Existing literature that evaluates docking program performances usually focuses on overall docking results such as the fraction of correctly predicted protein-bound conformations.[23,29] However, this kind of comparison is not conducive to pinpointing the cause of poor performance, i.e. whether poor performance is attributable to poor sampling, inaccurate scoring or both, thereby making it difficult to isolate and fix problem areas. In this work, the same set of high quality SKATE-generated poses was ranked by FRED-Score, X-Score, and Rosetta-Score. Because the sampling and the scoring were separated, it allowed for a fair comparison of the scoring function performances.[30] Although comparing scoring performance is not the main purpose of this work, it is still valuable to discuss the results. In all three self-docking test sets, FRED-Score was the most accurate scoring function (Fig. 4). FRED-Score summed the individual ranks by chemgauss3, PLP, and oechemscore to produce a consensus rank. This "rank-by-rank" strategy was also employed by Wang et al. in a study evaluating consensus scoring functions.[30] They showed that combining results from three complementary scoring functions improved the recognition of near-native poses ($\leq$ 2.0 Å) as best poses. Coincidently or not, FRED-Score and one of the best consensus functions in Wang et al. both included the PLP scoring function. Rosetta-Score is an extension of

ID: jaishankarn Date: 27/3/10 Time: 20:40 Path: N:/3b2/JCCT/Vol00000/100045/APPFile/C2JCCT100045

C O L O R

Figure 5. The RMSD between the 1JJE native pose (blue) and top-scoring pose (gray) of the ligand was 7.97 Å. The two ring systems of the top scoring pose were oriented in opposite directions.

Rosetta's energy function which was designed for *in silico* protein structure prediction. It may not have been optimally parameterized to score protein-ligand interactions. X-score was successful in ranking a pose that is within 2 Å of the experimental conformation in the range of 64–69% for the three test sets. This is consistent with a 66% success rate observed by Wang et al. in evaluating X-score on a 100 complexes test set.[30]

Generally, rigid-body local optimization of SKATE-generated poses improved FRED scoring. At RMSD thresholds between 1 and 2 Å, optimization followed by FRED scoring improved accuracy by up to nine percentage points. Poses with RMSD values under 2 Å are often considered near-native but some may contain poor contacts that cause a scoring function to rank them poorly. A quick rigid-body local optimization or minimization of those poses alleviated those poor contacts and resulted in better scores. PLP was the scoring function used in the rigid-body optimization of SKATE poses. Despite its simplicity, PLP has been shown to be one of the top performing scoring functions and is incorporated in multiple docking programs.[12,19,30] Results from using oechemscore or chemgauss3 as the scoring function for optimization were similar to those from using PLP.

Examples of Scoring Errors

The best poses for the 1JJE and 1OQ5 complexes in the Astex/CDCC were 0.52 Å and 0.37 Å, respectively. However, the RMSD of the top-scoring pose, ranked by FRED-Opt-Score, for 1JJE was 7.97 Å and that for 1OQ5 was 5.00 Å. Upon closer inspection of the 1JJE poses, we found the shapes of the top scoring pose and the native pose were essentially superimposable. The middle parts of the two poses overlap very well but the two ring systems on the ligand were placed in opposite orientations in the top-scoring pose (Fig. 5). Due to the symmetric nature of this ligand, this was a challenging case for scoring, because a small difference in 3D docked shape may be "flipped" to yield a large apparent RMSD. FRED-Score, X-Score and Rosetta-Score also failed to rank a near-native pose as the top-scoring pose. 1OQ5 is another example where the shapes of the top-scoring pose overlapped well with the native pose (Fig. 6). A phenyl group was swapped with a trichloromethyl group in the top-scoring pose. FRED-Score, X-Score and Rosetta-Score also failed to rank a near-native pose as the top-scoring pose.

F5

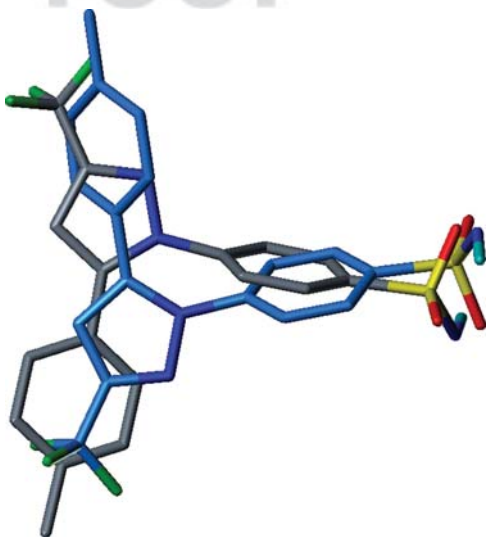
F6

Comparison with Other Docking Programs

Directly comparable docking results for the Astex/CDCC set are available for GOLD.[1] The success rate for GOLD[21] was 81% (Fig. 7A); which was 6% lower than SKATE. GOLD and SKATE used the crystallographically determined ligand bond angles and bond lengths. GOLD does allow ring flipping while SKATE does not sample ring flexibility. GOLD defined its binding site as all protein atoms within 6 Å of a nonhydrogen ligand atom. SKATE defined its binding site as all protein atoms within 5 Å of hydrogen and nonhydrogen ligand atoms. The two binding sites are nearly identical because the hydrogen-heavy atom bond-length is 1 Å.

F7

Seventy-seven of 81 complexes in the Surflex set were docked by the authors of Glide.[11] Glide's success rate for this subset was 82% for an RMSD threshold of 2.0 Å. The same subset was also docked by the authors of MolDock[12] with a resulting success rate of 87%. For comparison, the success rate of Surflex[22] was 77% and that of SKATE/FRED-Opt-Score was 73% for the entire set of 81 complexes (Fig. 7B). The energy minimized ligands, with torsions set to 180, were the input structures for SKATE. SKATE's systematic sampling was anchored by pairing H-bond donors and acceptors between the ligand and receptor H-bond atoms (defined in docking setup), and the ligand was allowed to sample regions outside of this 5 Å radius. It is hard to directly compare the results of Glide, MolDock, Surflex, and SKATE for several reasons. First, Glide and MolDock's success rates are based on 77 complexes, a subset of the 81 complexes in Surflex. Both Suflex and SKATE's success rates are based on the entire 81 complexes in the Surflex set. Second, MolDock basically trained its scoring function on this set of 77 complexes as pointed out by Hawkins et al.[28] Third, Glide calculated RMSD using optimized ligand coordinates instead of experimentally determined coordinates. Glide also used the optimized ligand and protein coordinates in its docking setup. The fact that the same energy function, OPLS/AA, was used in both complex optimization and pose scoring means Glide biased its methods by guaranteeing that the initial coordinates were at a local energy minimum per the OPLS/AA scoring function.[10,11,13,28]



C O L O R

Figure 6. The RMSD between the 1OQ5 native pose (blue) and top-scoring pose (gray) was 5.00 Å. A phenyl group was swapped with a trichloromethyl group in the top-scoring pose.

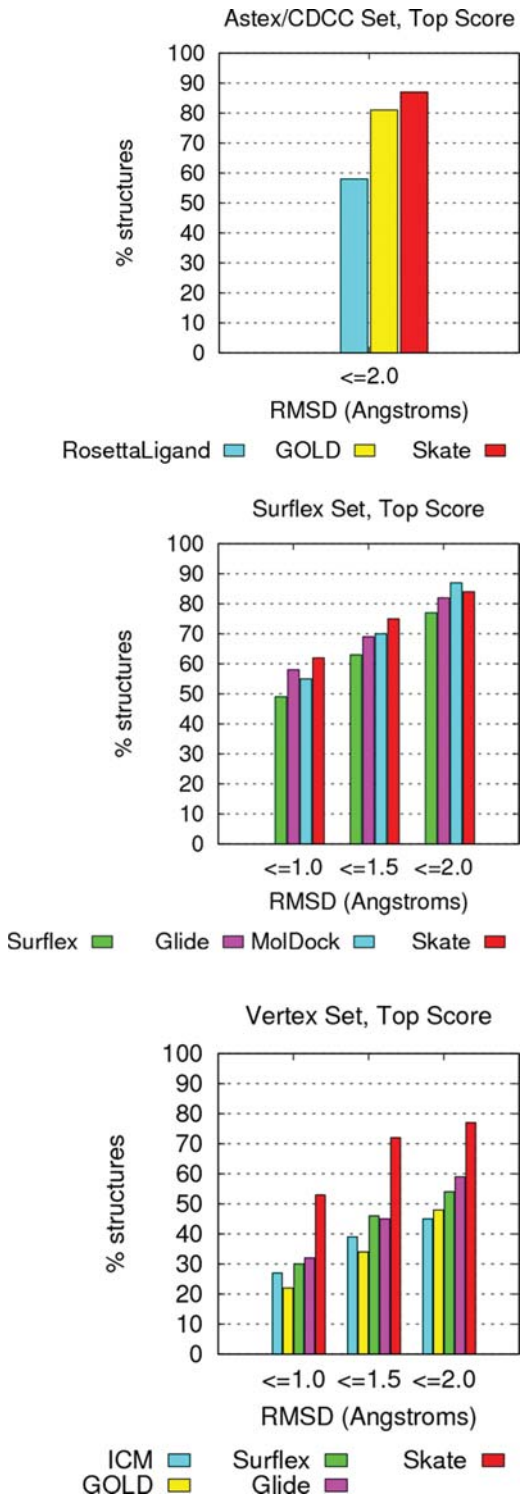


Figure 7. Distribution of the RMSD values between the top-ranked docking poses and the corresponding crystal structures in the Astex/CDCC set (A), Surflex set (B), and Vertex set (C). RMSD values were calculated on the coordinates of the heavy atoms of the ligands. X-axis: RMSD cutoffs; Y-axis: percentage of top-ranked docking poses within a given RMSD cutoff from the crystallographic pose.

Perola et al.[23] prepared a test set of 150 protein-ligand complexes to compare the performances of Glide, GOLD and ICM. Of the 100 publicly available PDB structures, Glide correctly identified a docked pose that was within 2.0 Å RMSD of the experimental structure in 59% of the cases, versus 48% by GOLD (Fig. 7C). The success rate of ICM with this subset of 100 PDB structures was not available, but its success rate with the entire 150 complexes was 45%. Jain docked the same 100 PDB complexes using Surflex and its success rate was 54%[10]. SKATE's systematic sampling coupled with FRED-Opt-Score ranking was successful in identifying a pose that was within 2.0 Å RMSD of the native structure as the best pose for 56% of the cases. This result was obtained by using six initial ligand conformations, one generated by Corina[31] and five generated by Omega.[17] The five lowest energy Omega conformers were added to increase ring diversity. SKATE does not perform ring sampling, if the initial ligand conformation contains experimentally determined ring conformation, the SKATE result for this set improved to 71% (Fig. 7C, SKATE*). SKATE's systematic sampling was anchored by H-bond anchor points and the ligand was allowed to sample regions outside of this 5 Å radius. The docking site volume in SKATE is similar to that used by other docking programs.

Comparing results from different docking programs are not always straightforward.[27,28] Results depend on, by varying degrees, receptor preparation, initial structure of the ligand, docking site volume, and quality and composition of test sets. Generous sharing of protein and ligand files by test set authors has made it easier to do fair comparisons. In this work, whenever possible, we aimed for minimally-biased comparisons by using conditions (described above) that were parallel to those used in other docking programs. Receptor files were downloaded from each test set's respective websites and were used without modification (other than file format conversion).

Bias, whether intentional or unintentional, is invariably introduced in results reported by authors of docking programs. One of the best available methods to compare docking program performance is blind docking assessments like the community wide Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) experiment (http://sampl.eyesopen.com).

Cross Docking

The thymidine kinase data set from the comparative article of Bissantz et al.[24] and the cyclin-dependent kinase 2 data set from Yang et al.[25] were used to test the cross-docking performance of SKATE. The TK set was originally used to quantitatively compare the performance of GOLD, DOCK, and FlexX. Data on this set are also available for Glide and Surflex. The TK structure used for docking was the deoxythymidine-bound structure (PDB code 1KIM). Table 2 summarizes the top-scoring RMSD values generated by the different docking programs for 10 thymidine kinase ligands. The ligand and receptor structures were prepared as described in Bissantz et al. FRED-Opt-Score was used to rank the poses generated by SKATE. Five of the 10 ligands were docked to the 1KIM structure with an RMSD of less than 1.8 Å. Another five failed to dock and their RMSD values were between 3 and 4 Å. Of the five failed cases,

Table 2. Accuracy in Cross Docking of Thymidine Kinase Inhibitors to the 1KIM site.[a]

	RMSD (Å) of top-scoring pose					
Ligand	SKATE[b]	Glide	DOCK	FlexX	GOLD	Surflex
dT	0.21	0.45	0.82	0.78	0.72	0.74
ahiu	0.61	0.54	1.16	0.88	1.63	0.87
mct	0.52	0.79	7.56	1.11	1.19	0.87
dhbt	1.74	0.68	2.02	3.65	0.93	0.96
idu	0.22	0.35	9.33	1.03	0.77	1.05
hmtt	3.24	2.83	9.62	13.30	2.33	1.78
hpt	4.03	1.58	1.02	4.18	0.49	1.90
acv	3.57	4.22	3.08	2.71	2.74	3.51
gcv	3.33	3.19	3.01	6.07	3.11	3.54
pcv	3.80	3.53	4.10	5.96	3.01	3.84

[a]Data for DOCK, FlexX, and GOLD are taken from Bissantz et al.;[24] data for Surflex are taken from Jain[22]; data for Glide are taken from Friesner et al.[11]
[b]FRED-Opt-Score was used to rank the poses generated by SKATE.

SKATE generated poses that were less than 2.0 Å RMSD for four ligands. However, FRED-Opt-Score failed to rank them as top-scoring poses. The RMSD values of the best pose for ligands hpt, hmtt, gcv, pcv, and acv were 0.35 Å, 1.19 Å, 2.11 Å, 1.65 Å, and 1.37 Å, respectively. As pointed out by Friesner et al.[11], ligands acv, gcv and pcv are purine-based ligands and do not fit properly into the pyrimidine-based ligand site. All six docking programs did not sufficiently sample receptor flexibility and therefore all failed to dock these three ligands. The cross-docking results by SKATE are comparable to Glide, GOLD, and Surflex.

The CDK2 test set consists of 73 complexes and the ligands were docked to a single CDK2 protein structure (PDB ID 2B54). The resolution of 2B54 is 1.85 Å and is cocrystallized with 6-(3,4-dihydroxybenzyl)-3-ethyl-1-(2,4,6-trichlorophenyl)-1H-prrazolo[3,4-d]pyrimidin-4(5H)-one. The 2B54 structure was selected to be the model receptor because it is the best-resolution structure with no missing residues or side-chain atoms. Two sets of VDW scaling parameters were tested in docking the 73 ligands to 2B54. The default VDW scaling value for intermolecular interactions is 0.9. A second set of parameters allows even more VDW penetration by using a scaling value of 0.8. Reducing VDW radii is a technique docking programs can employ to mimic receptor side-chain flexibility. Admittedly, this generates poor mimicry of receptor flexibility, yet is nevertheless useful until more advanced features are added to SKATE. To evaluate sampling and scoring accuracy, we used heavy atom RMSD from the native structure. To transform the reference coordinates into the same global coordinates, 72 of the 73 complexes were structurally aligned to 2B54 using pymol[32] and ligands were extracted and saved in the Tripos mol2 format. The sampling results from using the two different VDW scaling parameters are shown in Figure 8A. More permissive VDW parameters allow for more VDW penetration; hence more receptor flexibility result in improved sampling. SKATE was able to sample a pose that was within 2 Å RMSD of the native structure for 81% of the ligands (Fig. 8A dotted curve). However, this level of VDW scaling was not accommodated in FRED-Opt-Score. A low RMSD pose will score poorly if there are severe VDW penetrations. The percentage of top-scoring poses as a function of RMSD is shown in Figure 8B. At an RMSD cutoff of 2.0 Å RMSD, the success rate was 38%. Scaling atomic VDW radii by a factor of 0.8 improved sampling but a similar improvement was not achieved in scoring. The percentage of top-scoring ligand poses plotted as a function of RMSD threshold was similar for the two sets of VDW scaling parameters (Fig. 8B). In terms of overall docking accuracy, there was no significant advantage to using a VDW scaling value of 0.8. Thus, modifications to SKATE to systematically include receptor flexibility using the same approach are under consideration.

Pitfalls in Complex Preparation

The Vertex data set was prepared by performing a constrained minimization of the complexes using MacroModel and the

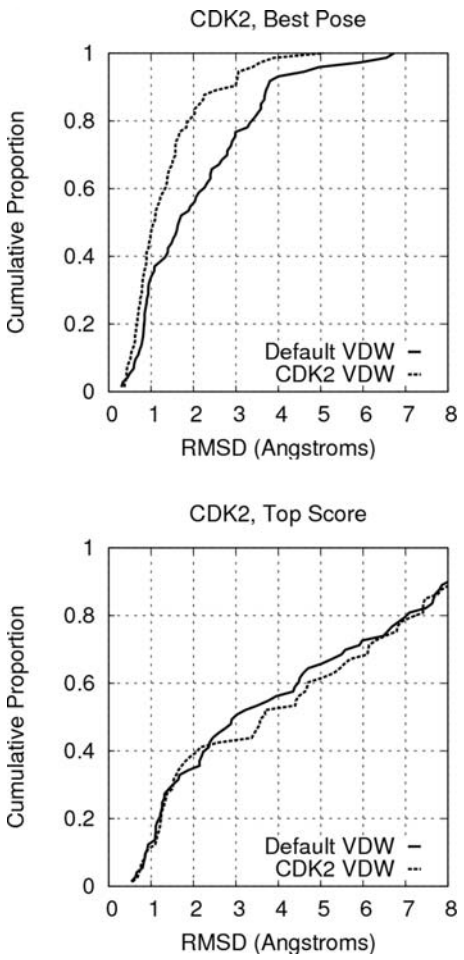


Figure 8. (A) Cumulative proportion of best RMSD poses for the CDK2 cross-docking set using two sets of VDW scaling parameters. (B) Cumulative proportion of the RMSD between the top-ranked poses and the native structure. FRED-Opt-Score was used to rank the poses. The default intermolecular VDW scaling value was 0.9. The CDK2 VDW scaling value was 0.8.

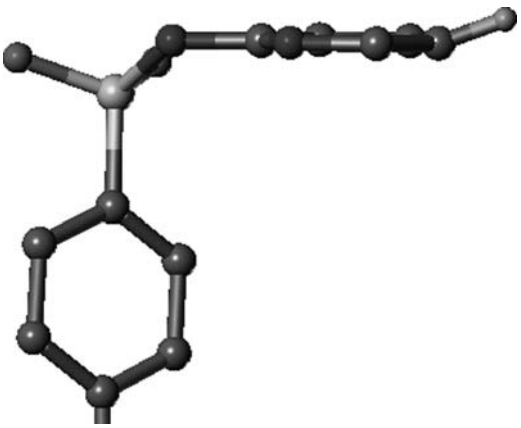


Figure 9. An aromatic proton in test case 13GS of the Vertex set was bent out of plane by an optimization step in complex preparation. Heavy atoms in a complex were fixed while protons were allowed to optimize. This proton (cyan) was bent out of plane by 25o to relieve steric overlap with a proton on residue Pro202 of the receptor. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

OPLS/AA force field.[33,34] Heavy atoms were constrained to their original position while hydrogen atoms were allowed to optimize. While this alleviated poor contacts between software-added hydrogen atoms, it could lead to artifacts where a hydrogen atom can bend out of plane to relieve steric interactions. F9 Shown in Figure 9 is an example where an aromatic hydrogen was bent 25 degrees out of plane during the minimization step. Hawkins et al. pointed out additional pitfalls in complex preparation and X-ray structure quality.[28] However, not optimizing software-added hydrogen atoms also has its problems. A proton on the ligand penetrated the VDW surface of a proton atom on a lysine side-chain in complex 1YGC of the Astex/CDCC set. In this case, the poor contacts could have been alleviated by a quick minimization.

Computational Time

Using SKATE, the average docking time per ligand-protein hydrogen-bond pair was less than 5 min for ligands with six or fewer rotatable bonds, and 10 min for ligands with 8 rotatable bonds. Total docking time was proportional to the number of possible hydrogen bonds that a ligand can form with the receptor. For the Astex/CDCC set, the median docking time was 42 min and the average docking time was 98 min on a single CPU (Pentium 4, 2.4 GHz) computer running Linux. SKATE allows simple parallelization by submitting each possible hydrogen-bond pairing to a different CPU in a computing cluster.

SKATE has not been optimized as it is still a prototype under development to evaluate the separation of docking from scoring. Limited optimization, however, should allow significant reduction in computational time. Further speed improvement in SKATE can be made by implementing look-ahead technologies to further prune the combinatorial search tree.[16,35] Knowledge about distance constraints between pharmacophore points can also be used to prune the search tree. Additional heuristics can be applied to reduce the number of discrete torsions sampled. Speed improvement will make SKATE more amenable to virtual screening applications of large compound libraries for which it is not appropriate in its current version.

Conclusions

We implemented a novel docking concept in SKATE that decouples systematic sampling from scoring to improve overall docking accuracy. SKATE's systematic sampling coupled with FRED's optimization and scoring was comparable to commercially available program across three large data sets. Systematic sampling in SKATE was robust as tested by three large self-docking test sets and two cross-docking test sets. The high-quality poses generated by SKATE could be used to train scoring functions to distinguish between near-native and poorly docked poses.

The problem of false negatives is often the root cause of poor performance in docking programs. If a docking program never samples near-native poses, then there is zero chance that a scoring function can rank them as top-scoring poses. Unfortunately, modern docking programs' sampling methods are dependent on scoring functions that, at best, approximate experimental binding energies. The interdependence of sampling and scoring makes it difficult to determine whether a sampling error or a scoring error caused significant problems in a docking experiment. SKATE breaks this dependence by systematically sampling all sterically allowed poses of a ligand as constrained by a receptor pocket. This work shows that improved sampling contributed to docking accuracy.

An executable version of SKATE and the five test data sets are available for download from http://www.ccb.wustl.edu/software.

Methods

An aggregate is defined as a set of atoms whose relative positions are invariant to rotational degrees of freedom.[16] Atoms in an aggregate could be directly bonded, have a 1–3 relationship defined by a bond angle, be part of a ring system, or have bonds between them conjugated by resonance. Table 1 lists the number of rotatable bonds, sampled by SKATE, for the ligands in the three self-docking test sets. Figure 2 illustrates how a simple molecule was divided into three aggregates. There are $T+1$ aggregates and T torsional degrees of freedom in a flexible molecule. Sterically allowed conformations of a ligand are generated by assembling its aggregates. Since the distance between two atoms within an aggregate is constant, it is not necessary to check for VDW clashes between atoms within the same aggregate.

In SKATE, sterically allowed poses of a ligand are constructed in a stepwise fashion by reassembling the aggregates comprising the ligand. Starting with an initial aggregate that contains an atom that forms a hydrogen bond with a receptor atom, a second aggregate is added via the rotatable bond that joins the two aggregates (Fig. 1). Some torsion values around this shared rotatable bond will lead to VDW overlaps between atoms in aggregate two and atoms in aggregate one, as well as atoms in the receptor. It is extremely inefficient to assemble two

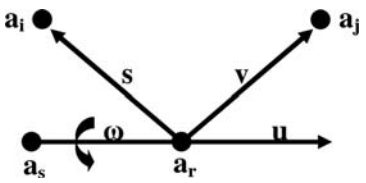


Figure 10. The variable distance between a fixed atom ai and a rotatable atom aj is a function of a single torsional variable ω. Atoms ai, as and ar are rigid with respect to each other and they belong to the sterically allowed conformation of a partially docked ligand. Atoms as and ar forms the rotatable bond and determine the rotatational axis. u is a unit vector along the axis of rotation. The torsional variable of ω is being evaluated by discriminant analysis to determine the range of torsions where atom aj does not clash with any atoms in the partial conformation.

aggregates for a given torsion only to find out that it is a sterically impossible conformation. Discriminant analysis solves this problem by analytically calculating the range of sterically allowed torsions within which two aggregates can be assembled without steric overlap. The result is that only allowed torsions are sampled. In theory, systematic sampling should find all sterically allowed poses of a ligand. In practice, SKATE discretizes the continuous conformational space and then uses adaptive torsion sampling and radial sampling to ensure sufficient sampling.[35]

Discriminant analysis was first applied to systematically search the conformational hyperspace available to a flexible molecule to define three-dimensional quantitative structure-activity relationships (3D-QSAR) and biological receptor mapping.[36] In the construction of a molecule from stepwise addition of aggregates, there are two sets of atoms to consider. First are those in the sterically allowed partial molecule (set A) previously constructed. Second are those in the next aggregate (set B) to be added to the existing partial molecule. Atoms in set B must be checked against those in set A to find torsions that are sterically allowed. Distance constraint equations are used analytically to determine the possible torsion ranges such that a new aggregate can be added without steric overlap between atoms in the new aggregate (set B) and the partial conformation (set A). These equations, derived elsewhere,[36] describe the variable distance between any two atoms as a function of a single torsion angle (ω). The square of the interatomic distance between a_j and a_i in Figure 10 is given by:

F10

$$d_{ij}^2(\omega) = d_1 + d_2\cos(\omega) + d_3\sin(\omega) \quad (1)$$

where coefficients d1, d2, and d3 are defined as follows:

$$d_1 = |\vec{s}|^2 + |\vec{v}|^2 - 2(\vec{s} \bullet \vec{v}_1)$$

$$d_2 = -2(\vec{s} \bullet \vec{v}_2)$$

$$d_3 = -2(\vec{s} \bullet \vec{v}_3)$$

$v1$, $v2$, and $v3$ are the three orthogonal components of the vector v in Figure 10 where

$$\vec{v} = a_j - a_r$$

$$\vec{v}_3 = \vec{u} \times \vec{v}$$

$$\vec{v}_2 = \vec{u} \times \vec{v}_3$$

$$\vec{v}_1 = \vec{v} - \vec{v}_2$$

Equation (1) can be rewritten as

$$g_{ij}^2(\omega) = \frac{ax^2 + bx + c}{1 + x^2}$$

where

$$a = d_1 - d_2$$

$$b = 2d_3$$

$$c = d_1 + d_2$$

$$x = \tan\left(\frac{\omega}{2}\right)$$

Let c_{ij} be the sum of the VDW radii for atoms i and j, then differential distance function $\delta_{ij}^2(\omega) = d_{ij}^2(\omega) - c_{ij}^2$ is evaluated to determine whether or not the two atoms are in contact. The differential distance function can be converted to a quadratic form

$$\delta_{ij}^2(\omega) = \frac{(a - c_{ij}^2)x^2 + bx + (c - c_{ij}^2))}{1 + x^2}$$

$$D = b^2 - 4(a - c_{ij}^2)(c - c_{ij}^2)$$

$$x = \frac{-b \pm \sqrt{D}}{2a}$$

$$\omega = 2\tan^{-1}(x)$$

The resulting discriminant D can be used to determine if there is a real or imaginary solution to $\delta^2{}_{ij}(\omega)$. If $D > 0$, then $\delta^2{}_{ij}(\omega)$ has real roots and the upper and lower bound values of the torsional range (ω) can be calculated from the above equa-

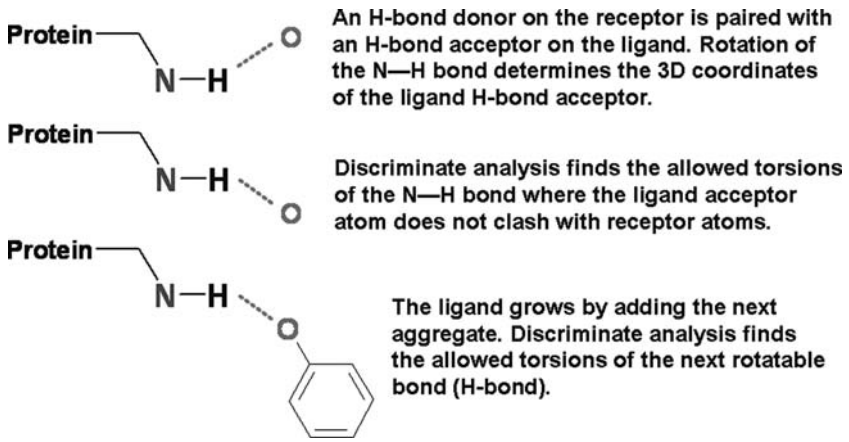


Figure 11. Docking of a ligand to a receptor by pairing H-bonding partners. Rotatable bonds in the ligand are searched systematically to find allowed torsions that generate a bound pose for further evaluation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

tions. If $D \leq 0$, $\delta^2_{ij}(\omega)$ has complex or real double degenerate roots. For $c - c_{ij}^2 \geq 0$, $\delta^2_{ij}(\omega)$ is positive for all values of ω implying that atom i and atom j never come in contact for any torsional value of ω. For $c - c_{ij}^2 \geq 0$, $\delta^2_{ij}(\omega)$ is negative for all values of ω and there is no sterically allowed way to add the new aggregate. In this case, the new partial conformation will be discarded and this search branch of the tree truncated.

The distance constraint equations minimize the number of pair-wise intramolecular and intermolecular distances that must be evaluated in a systematic search. They prune the search tree by analytically determining torsion ranges that result in sterically allowed partial or complete conformers. The intersection of allowed torsion ranges for every atom pair spanning a rotatable bond results in discontinuous slices of torsion ranges in which a new aggregate is added during the step-wise construction process. The torsional ranges are discretized by adaptive sampling and radial sampling to ensure sufficient sampling.[35] Adaptive sampling, as opposed to uniform sampling, ensures that SKATE does not over-sample or under-sample a torsional range. Radial sampling determines the increment in degrees between two sampled torsions. In SKATE, a rotation of an aggregate around its rotatable bond displaces an atom in the aggregate by a maximum of 0.25 Å.

SKATE pairs an H-bond donor of a receptor with an H-bond acceptor of a ligand, and vice versa, to anchor systematic search. In SKATE, three parameters are used to define a hydrogen bond, the distance between the hydrogen atom and the acceptor atom; the angle formed by the acceptor, hydrogen, and donating atoms; the angle formed by the acceptor base, acceptor, and hydrogen atoms. Figure 11 illustrates how SKATE initializes its H-bond pairing and systematic search process. A receptor H-bond donor is paired with a ligand H-bond acceptor. Rotation of the N—H bond on the receptor determines the 3D coordinate of the ligand acceptor atom. Using discriminant analysis, SKATE quickly determines the allowed torsions of the N—H bond such that the ligand acceptor atom does not clash with receptor atoms. The next bond to be rotated is the H-bond between the receptor and the ligand. This determines the allowed torsions of the H-bond such that ligand atoms in the first aggregate do not clash with the receptor atoms. The remaining aggregates are then systematically searched by recursion. Sterically allowed poses for a given ligand-receptor hydrogen-bond represent leaves of a tree graph where nodes represent aggregates and edges represent discrete torsions of rotatable bonds (Fig. 1). SKATE travels this tree using a depth first search approach as illustrated by the following pseudo code.

```
Systematic Search Pseudo Code
MAIN( )
  DOCK (receptor, ligand)
  SEARCH (receptor, ligand, torsions, agg_idx)
SEARCH (receptor, ligand, torsions, agg_idx)
  UPDATE (ligand, agg_idx)
  VALIDATE(receptor, ligand, torsions, agg_idx)
  for each allowed torsion of aggregate agg_idx
    ROTATE (ligand, torsions)
    if last aggregate
      RECORD (ligand)
    else
      SEARCH (receptor, ligand, torsions, agg_idx+1)
    end if
  end for
```

The DOCK procedure transforms the coordinates of a ligand H-bond partner such that it forms a hydrogen bond with a receptor partner. The resulting H-bond geometry is determined by a set of geometric H-bond parameters.

The UPDATE procedure transforms the atoms in aggregate agg_indx to be in the same local coordinates as the previously searched aggregates and partially assembled molecule.

The VALIDATE procedure performs discriminant analysis to find allowed torsions of the rotatable bond that connects aggregate agg_indx with the previously searched aggregates of the ligand. A list of allowed torsions is stored in the torsions data structure.

The ROTATE procedure simply rotates an aggregate to an allowed torsion that was calculated by the VALIDATE procedure.

Due to inherent errors in X-ray structure determination, there are often VDW clashes between ligand and receptor atoms in crystal structures. We employed a VDW scaling factor to reduce the VDW radii of protein and ligand atoms to ensure the reproduction of experimental structures.[37] A general scaling factor of 0.95 is applied to ligand intramolecular interactions. A 1,4 scaling factor of 0.87 is applied to ligand atoms in 1–4 relationships. Intermolecular interactions are scaled by a factor of 0.9 and hydrogen-bond interactions are scaled by a factor of 0.6.

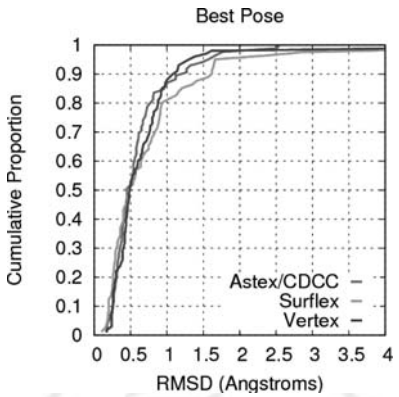
Acknowledgments

The authors thank Dr. Chris Ho for providing C libraries to read mol2 files and to build data structures. They thank Robert Yang for sharing the CDK2 data set, Dr. Ajay Jain for sharing the Surflex set and Vertex set, and Dr. Marcel Verdonk for sharing the Astex/CDCC set.

References

1. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. J Mol Biol 1997, 267, 727.
2. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. J Mol Biol 1996, 261, 470.
3. Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. J Computer-Aided Mol Des 2001, 15, 411.
4. Abagyan, R.; Totrov, M.; Kuznetsov, D. J Comput Chem 1994, 15, 488.
5. Mcmartin, C.; Bohacek, R. S. J Computer-Aided Mol Des 1997, 11, 333.
6. Meiler, J.; Baker, D. Proteins 2006, 65, 538.
7. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. J Comput Chem 1998, 19, 1639.
8. Mcgann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Biopolymers 2003, 68, 76.
9. Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. J Mol Graphics Model 2003, 21, 289.
10. Jain, A. N. J Computer-Aided Mol Des 2007, 21, 281.
11. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. J Med Chem 2004, 47, 1739.
12. Thomsen, R.; Christensen, M. H. J Med Chem 2006, 49, 3315.
13. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P. J Mol Graphics Model 2007, 26, 198.
14. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Proteins 2002, 47, 409.
15. Velec, H. F.; Gohlke, H.; Klebe, G. J Med Chem 2005, 48, 6296.
16. Dammkoehler, R. A.; Karasek, S. F.; Shands, E. F.; Marshall, G. R. J Computer-Aided Mol Des 1989, 3, 3.
17. OpenEye. Santa Fe, NM, 2008. AQ4
18. Wang, R.; Lai, L.; Wang, S. J Computer-Aided Mol Des 2002, 16, 11.
19. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Chem Biol 1995, 2, 317.
20. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. J Computer-Aided Mol Des 1997, 11, 425.
21. Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T.; Mortenson, P. N.; Murray, C. W. J Med Chem 2007, 50, 726.
22. Jain, A. N. J Med Chem 2003, 46, 499.
23. Perola, E.; Walters, W. P.; Charifson, P. S. Proteins 2004, 56, 235.
24. Bissantz, C.; Folkers, G.; Rognan, D. J Med Chem 2000, 43, 4759.
25. Yang, R.; Parnot, C.; Marshall, G. R. (in press). AQ5
26. Tripos, Saint Louis, MO, 2008.
27. Jain, A. N. J Computer-Aided Mol Des 2008, 22, 201.
28. Hawkins, P. C.; Warren, G. L.; Skillman, A. G.; Nicholls, A. J Computer-Aided Mol Des 2008, 22, 179.
29. Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; Lalonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. J Med Chem 2006, 49, 5912.
30. Wang, R.; Lu, Y.; Wang, S. J Med Chem 2003, 46, 2287.
31. Networks, M. Erlangen, Germany, 2006.
32. Delano, W. L. Delano Scientific LLC: Palo Alto, CA, 2009. AQ6
33. Schrodinger: New York, 2006.
34. Jorgensen, W. L.; Tirado-Rives, J. Abstr Pap Am Chem Soc 1998, 216, U696.
35. Beusen, D. D.; Shands, E. F. B.; Karasek, S. F.; Marshall, G. R.; Dammkoehler, R. A. Theochem-J Mol Struct 1996, 370, 157.
36. Motoc, I.; Dammkoehler, R. A.; Marshall, G. R. Mathematics and Computational Concepts in Chemistry; Trinajstic, N., Ed.; Ellis Horwood: Chichester, 1986; pp. 222–251.
37. Iijima, H.; Dunbar, J. B.; Marshall, G. R. Proteins 1987, 2, 330.

SGML and CITI Use Only
DO NOT PRINT



AQ1: Kindly check whether the short title is OK as given.

AQ2: Kindly check whether ALL section heads and subheads in the manuscript have been set as intended.

AQ3: Please confirm whether the color figures should be reproduced in color or black and white in the print version. If the color figures must be reproduced in color in the print version, please fill the color charge form immediately and return to Production Editor. Or else, the color figures for your article will appear in color in the online version only.

AQ4: Is this a book-type reference, if so, kindly provide the author names and book title for references 17, 26, and 31–33.

AQ5: Kindly update ref. 5.

AQ6: Please note that the references have been renumbered as the originally numbered ref. 32 seemed to be an incorrect repetition of the ref. 17, hence the subsequent refs. are made sequential per journal style.