Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy

Richard A. Friesner,*,† Jay L. Banks,‡ Robert B. Murphy,‡ Thomas A. Halgren,‡ Jasna J. Klicic,‡,|| Daniel T. Mainz,‡ Matthew P. Repasky,‡ Eric H. Knoll,† Mee Shelley,§ Jason K. Perry,§ David E. Shaw,# Perry Francis,§ and Peter S. Shenkin‡

Department of Chemistry, Columbia University, New York, New York 10036, Schrödinger, L.L.C., 120 W. 45th Street, New York, New York 10036, Schrödinger, L.L.C., 1500 SW First Avenue, Portland, Oregon 97201, and D. E. Shaw Research and Development, 120 W. 45th Street, New York, New York 10036

Received December 24, 2003

Unlike other methods for docking ligands to the rigid 3D structure of a known protein receptor, Glide approximates a complete systematic search of the conformational, orientational, and positional space of the docked ligand. In this search, an initial rough positioning and scoring phase that dramatically narrows the search space is followed by torsionally flexible energy optimization on an OPLS-AA nonbonded potential grid for a few hundred surviving candidate poses. The very best candidates are further refined via a Monte Carlo sampling of pose conformation; in some cases, this is crucial to obtaining an accurate docked pose. Selection of the best docked pose uses a model energy function that combines empirical and force-field-based terms. Docking accuracy is assessed by redocking ligands from 282 cocrystallized PDB complexes starting from conformationally optimized ligand geometries that bear no memory of the correctly docked pose. Errors in geometry for the top-ranked pose are less than 1 Å in nearly half of the cases and are greater than 2 Å in only about one-third of them. Comparisons to published data on rms deviations show that Glide is nearly twice as accurate as GOLD and more than twice as accurate as FlexX for ligands having up to 20 rotatable bonds. Glide is also found to be more accurate than the recently described Surflex method.

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

1. Introduction

The number of drug-discovery projects that have a high-resolution crystal structure of the receptor available has increased in recent years and is expected to continue to rise because of the human genome project and high-throughput crystallography efforts. A common computational strategy in such a case is to dock molecules from a physical or virtual database into the receptor and to use a suitable scoring function to evaluate the binding affinity. A number of docking programs are employed extensively in the pharmaceutical and biotechnology industries,[1−10] of which the most widely used appear to be GOLD,[1] FlexX,[2] and DOCK.[3] Over the past several years, considerable success has been reported for these programs in virtual screening applications.[11−13] However, none as of now can be viewed as offering a robust and accurate solution to the docking problem, even in the context of a rigid protein receptor.

In this paper, we describe a new docking methodology that has been implemented in the FirstDiscovery software package Glide[14] (grid-based ligand docking with energetics). Glide has been designed to perform as close to an exhaustive search of the positional, orientational, and conformational space available to the ligand as is feasible while retaining sufficient computational speed to screen large libraries. This has been accomplished via the use of a series of hierarchical filters, as described below. The current performance characteristics of Glide are as follows:

(i) Docking times average less than 1 min for data sets having 0−10 rotatable bonds on an AMD Athelon MP 1800+ processor running Linux.

(ii) Robustness in binding mode prediction is qualitatively superior to what is reported in the current literature for docking methods in widespread use. For example, a comparison with results obtained by the developers of GOLD yields an average rmsd of 1.46 Å for Glide compared with 2.56 Å for GOLD for the 72 noncovalently bound cocrystallized ligands of the GOLD test set[15] that have 10 or fewer rotatable bonds. The comparison to FlexX is even more favorable. Comparisons for ligands having up to 20 rotatable bonds yield similar results.

(iii) Binding affinity predictions, compared with experimental data for cocrystallized complexes, are reasonable (2.3 kcal/mol rmsd), though clearly subject to improvement.

(iv) Results for library screening, reported in the following paper,[16] are very encouraging. Furthermore, database enrichment factors obtained using Glide 2.5 are significantly higher than those obtained using previous versions of Glide.

The paper is organized as follows. Section 2 summarizes the computational methodology used by Glide, while section 3 describes Glide's approach to scoring relative ligand binding affinities. The fourth section

* To whom correspondence should be addressed. Phone: 212-854-7606. Fax: 212-854-7454. E-mail: rich@chem.columbia.edu.
† Columbia University.
‡ Schrödinger, L.L.C., NY.
|| Present address: Serono International, S.A, CH-1211 Geneva 20, Switzerland.
§ Schrödinger, L.L.C., OR.
# Schrödinger, L.L.C., NY and D. E. Shaw Research and Development.

10.1021/jm0306430 CCC: $27.50 © 2004 American Chemical Society
Published on Web 02/27/2004

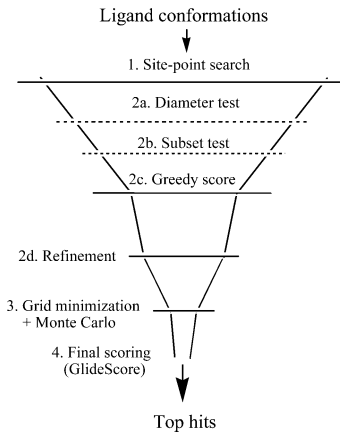


Figure 1. Glide docking "funnel", showing the Glide docking hierarchy.

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

then presents rmsd values obtained for redocking cocrystallized ligands and compares Glide to GOLD, FlexX, and Surflex, methods we believe to be representative of the current state of the art in high-throughput docking. The fifth section summarizes the results and discusses future directions. Finally, section 6 provides details of the docking methodology and of the optimization of the scoring function; this section also describes the procedure we recommend for protein preparation, which in many cases can substantially affect the quality of the results obtained in docking calculations.

2. Overview of Docking Methodology

Glide uses a series of hierarchical filters to search for possible locations of the ligand in the active-site region of the receptor (Figure 1). The shape and properties of the receptor are represented on a grid by different sets of fields that provide progressively more accurate scoring of the ligand pose. (By "pose" we mean a complete specification of the ligand: position and orientation relative to the receptor, core conformation, and rotamer-group conformations.) These fields are generated as preprocessing steps in the calculation and hence need to be computed only once for each receptor.

The next step produces a set of initial ligand conformations. These conformations are selected from an exhaustive enumeration of the minima in the ligand torsion-angle space and are represented in a compact combinatorial form. Given these ligand conformations, initial screens are performed over the entire phase space available to the ligand to locate promising ligand poses. This prescreening drastically reduces the region of phase space over which computationally expensive energy and gradient evaluations will later be performed while at the same time avoiding the use of stochastic methods; such methods can miss key phase-space regions a certain fraction of the time, thus precluding development of a truly robust algorithm. To our knowledge, Glide is unique in its reliance on the techniques of exhaustive systematic search, though approximations and truncations are required to achieve acceptable computational speed.

Starting from the poses selected by the initial screening, the ligand is minimized in the field of the receptor using a standard molecular mechanics energy function (in this case, that of the OPLS-AA force field[17]) in conjunction with a distance-dependent dielectric model. Finally, the three to six lowest-energy poses obtained in this fashion are subjected to a Monte Carlo procedure that examines nearby torsional minima. This procedure is needed in some cases to properly orient peripheral groups and occasionally alters internal torsion angles.

We and others have found that a conventional molecular mechanics energy function is a reasonable model for predicting binding modes, even in the absence of solvent. However, it is not inadequate for ranking disparate ligands, for example, ligands with different net charge. Therefore, we have implemented a modified and expanded version of the ChemScore[18] scoring function, GlideScore, for use in predicting binding affinity and rank-ordering ligands in database screens. However, we use a combination of GlideScore, the ligand−receptor molecular mechanics interaction energy, and the ligand strain energy to select the correctly docked pose. We find that this composite scoring function, which we call Emodel, is much better at selecting the correct pose than is either the molecular mechanics energy or GlideScore alone.

A final and very important issue is that the scoring function, particularly the molecular mechanics component, needs to be modified in order to accommodate the fact that the protein structure used for docking will not in general be optimized to fit a particular ligand. We have found that the most severe problem when docking a library of ligands into a single rigid receptor structure is the inability of some actives to fit into the protein cavity because the cavity is too small. Therefore, we typically scale down the van der Waals radii of selected (e.g., nonpolar) protein and/or ligand atoms to create additional space in the binding pocket. Our database enrichment studies show that this approach is effective in that context.[16] Better enrichment can often be obtained by tuning the scaling parameters for a given receptor,[16] but the default values are suitable for routine use.

3. Scoring Function

The starting point for Glide scoring is the empirically based ChemScore function of Eldridge et al.,[18] which can be written as

$$\Delta G_{\text{bind}} = C_0 + C_{\text{lipo}}\sum f(r_{\text{lr}}) + C_{\text{hbond}}\sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{metal}}\sum f(r_{\text{lm}}) + C_{\text{rotb}}H_{\text{rotb}} \quad (1)$$

The summation in the second term extends over all ligand-atom/receptor-atom pairs defined by ChemScore as lipophilic, while that in the third term extends over all ligand−receptor hydrogen-bonding interactions. In eq 1, *f*, *g*, and *h* are functions that give a full score (1.00) for distances or angles that lie within nominal limits and a partial score (1.00−0.00) for distances or angles that lie outside those limits but inside larger threshold values. For example, $g(\Delta r)$ is 1.00 if the H···X hydrogen-bond distance is within 0.25 Å of a nominal value of 1.85 Å but tails off to zero in a linear fashion if the distance lies between 2.10 and 2.50 Å. Similarly, $h(\Delta\alpha)$ is 1.00 if the Z−H···X angle is within 30° of 180° and decreases to zero between 150° and 120°.

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

Glide 2.5 employs two forms of GlideScore: (i) GlideScore 2.5 SP, used by Standard-Precision Glide; (ii) GlideScore 2.5 XP, used by Extra-Precision Glide. These functions use similar terms but are formulated with different objectives in mind. Specifically, GlideScore 2.5 SP is a "softer", more forgiving function that is adept at identifying ligands that have a reasonable propensity to bind, even in cases in which the Glide pose has significant imperfections. This version seeks to minimize false negatives and is appropriate for many database screening applications. In contrast, GlideScore 2.5 XP is a harder function that exacts severe penalties for poses that violate established physical chemistry principles such as that charged and strongly polar groups be adequately exposed to solvent. This version of GlideScore is more adept at minimizing false positives and can be especially useful in lead optimization or other studies in which only a limited number of compounds will be considered experimentally and each computationally identified compound needs to be as high in quality as possible. In what follows, we discuss the development and parametrization of Glide 2.5 SP; XP docking and scoring[19] will be described in a subsequent paper.

GlideScore 2.5 modifies and extends the ChemScore function as follows:

$$\begin{aligned}\Delta G_{\text{bind}} = {} & C_{\text{lipo-lipo}} \sum f(r_{\text{lr}}) + \\ & C_{\text{hbond-neut-neut}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{hbond-neut-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{hbond-charged-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{max-metal-ion}} \sum f(r_{\text{lm}}) + C_{\text{rotb}} H_{\text{rotb}} + \\ & C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} + \\ & C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms} \quad (2)\end{aligned}$$

The lipophilic−lipophilic term is defined as in ChemScore. The hydrogen-bonding term also uses the ChemScore form but is separated into differently weighted components that depend on whether the donor and acceptor are both neutral, one is neutral and the other is charged, or both are charged. In the optimized scoring function, the first of these contributions is found to be the most stabilizing and the last, the charged-charged term, is the least important. The metal−ligand interaction term (the fifth term in eq 2) uses the same functional form as is employed in ChemScore but varies in three principal ways. First, this term considers only interactions with anionic acceptor atoms (such as either of the two oxygens of a carboxylate group). This modification allows Glide to recognize the evident strong preference for coordination of anionic ligand functionality to metal centers in metalloproteases.[20,21] In addition, Glide 2.5 counts just the single best interaction when two or more metal ligations are found. We set the coefficient to −2.0 kcal/mol, a value we believe to be reasonable, though the parameter refinement would have preferred an even more strongly negative value. Third, we assess the net charge on the metal ion in the unligated apo protein (generally straightforward via examination of the directly coordinated protein side chains). If the net charge is positive, the preference for an anionic ligand is incorporated into the scoring function. On the other hand, if the ion is net neutral (as it is, for example, in the case of the zinc metalloprotein farnesyl protein transferase, which accepts neutral ligands such as substituted imidazoles[22]), the preference is suppressed. The seventh term, from Schrödinger's active site mapping facility, rewards instances in which a polar but non-hydrogen-bonding atom (as classified by ChemScore) is found in a hydrophobic region.

Table 1. OPLS-AA Interaction Energies with Full and Modified Charge Distributions for Ionic Centers and Groups ($\epsilon = 2r$)

system	full charges	reduced charges
charged−charged		
$Zn^{2+}\cdots MeOPO_3^{2-}$	−111.1	−16.8
$Zn^{2+}\cdots (MeO)_2PO_2^-$	−70.8	−14.5
$Zn^{2+}\cdots Me_2PO_2^-$	−68.0	−13.4
$Zn^{2+}\cdots$acetate	−80.8	−14.4
$Mg^{2+}\cdots$acetate	−90.0	−16.2
$Mn^{2+}\cdots$acetate	−71.7	−12.6
$Zn^{2+}\cdots CH_3S^-$	−48.7	−11.5
$NH_4^+\cdots$acetate	−24.8	−7.1
Me-guanidinium$\cdots$acetate	−31.6	−11.8
benzamidinium$\cdots$acetate	−27.3	−9.2
$His^+\cdots$acetate	−22.3	−7.4
Me-guanidinium$\cdots (MeO)_2PO_2^-$	−27.7	−12.0
charged−polar		
$Zn^{2+}\cdots H_2O$	−23.8	−10.3
$NH_4^+\cdots H_2O$	−7.2	−5.6
$H_2O\cdots$acetate	−8.6	−4.9
Me-guanidinium$\cdots H_2O$	−8.1	−6.4
benzamidinium$\cdots H_2O$	−6.9	−4.8
$His^+\cdots H_2O$	−7.2	−6.5
$H_2O\cdots Me_2PO_2^-$	−7.2	−4.4
$H_2O\cdots CH_3S^-$	−4.0	−2.4
polar−polar		
$H_2O\cdots H_2O$	−3.7	−3.7

The second major component is the incorporation of contributions from the Coulomb and vdW interaction energies between the ligand and the receptor. To make the gas-phase Coulomb interaction energy a better predictor of binding (and a better contributor to a composite scoring function), we reduce, by ∼50%, the net ionic charge on formally charged groups such as carboxylates and guanidiniums; we also reduce the vdW interaction energies for the atoms directly involved.[23] Table 1 illustrates the effect of these changes for some prototype systems. The wide disparities in the original interaction energies are greatly reduced, though charge−charge interactions are still favored to some extent. The Coulomb−vdW energies used in GlideScore 2.5 (but not those used in Emodel) employ these reductions in net ionic charge except in the case of anionic ligand−metal interactions, for which Glide uses the full interaction energy.

The third major component is the introduction of a solvation model. Like other scoring functions of this type, previous versions of GlideScore did not properly take into account the severe restrictions on possible ligand poses that arise from the requirement that charged and polar groups of both the ligand and protein be adequately solvated. Charged groups, in particular, require very careful assessment of their access to solvent. In addition, water molecules may be trapped in hydrophobic pockets by the ligand, also an unfavorable situation.

To include solvation effects, Glide 2.5 docks explicit waters into the binding site for each energetically

competitive ligand pose and employs empirical scoring terms that measure the exposure of various groups to the explicit waters. This "water-scoring" technology has been made efficient by the use of grid-based algorithms. Using explicit waters, as opposed to a continuum solvation model, has significant advantages. In the highly constrained environment of a protein active site containing a bound ligand, the location and environment of individual water molecules become important. Current continuum solvation models have difficulty capturing these details, but our explicit-water approach has allowed us to develop consistently reliable descriptors for rejecting a high fraction of the false positives that appear in any empirical docking calculation. Our analysis also produced trial values for the coefficients of the various penalty terms. For the most part, these coefficients were used without modification in GlideScore 2.5 XP.[19] For Glide 2.5 SP, however, the need to make the program relatively fast limits the amount of sampling that can be done during docking and hence limits the accuracy of the docked poses. As a result, optimization of the solvation penalties led to smaller coefficients that do not too heavily penalize misdocked actives. Even these smaller values, however, provide substantial enhancement in enrichment factors for many database screens, thrombin being the most prominent example (in this case primarily by rejecting false positives that bury charged groups in a hydrophobic region of the thrombin active site).

4. Docking Accuracy

This section characterizes Glide's performance in reproducing the geometries of cocrystallized ligands taken from an extensive set of 282 publically available PDB[24] complexes. This set includes most of the members of the well-known GOLD and FlexX test sets, approximately 50 PDB complexes used in evaluations of Glide by prospective customers and approximately 50 more complexes whose experimental binding affinities have been used to develop one or more of the empirical scoring functions described in the literature (e.g., ChemScore[18]). We used the latter complexes, and others included in the GOLD and FlexX test sets, to calibrate the GlideScore function. Our coverage of the GOLD and FlexX sets is not quite complete because Glide does not deal with covalently attached ligands (seven cases: 1aec, 1ase, 1blh, 1tpp, 1lmp, 3gch, and 4est) and cannot handle ligands having more than 35 rotatable bonds (one case: 2er6). In addition, we excluded one complex (6rsa) because it contains an atomic species (vanadium) for which the OPLS-AA force field used by Glide has no parameters.

All results were obtained with release 2.5021 of the FirstDiscovery suite[14] on an AMD Athelon MP 1800+ processor running Linux. All structures were prepared using the protein-preparation procedure described in section 6 or an earlier version of that procedure. For these calculations, the vdW radii of nonpolar protein atoms were not scaled, but the radii of nonpolar ligand atoms (taken to be atoms having a partial charge of less than 0.15 e^- in magnitude) were scaled down by a factor of 0.8; the same default scaling is also employed in the database-enrichment studies presented in the accompanying paper.[16]

Table 2. Average rms Deviations for Flexible Docking on 282 PDB Complexes[a]

no. of rotatable bonds	no. of cases	av rms top-ranked pose (Å)	av CPU time (min)
0−3	51	1.01	0.2
4−6	92	1.64	0.6
7−10	48	1.79	1.7
0−8	164	1.48	0.5
0−10	191	1.51	0.8
0−20	263	1.89	2.4

[a] Times are AMD Athelon MP 1800+ CPU minutes.

To examine the dependence of the results on the starting geometries, we docked five sets of ligand geometries. One set consists of the (restraint-optimized) native ligands obtained from the protein-preparation procedure. These ligands typically have an rmsd for nonhydrogen atoms of 0.3 Å or less from the original PDB coordinates. The second set uses MMFF94s-optimized[25] versions of the native ligands. Except for 1cps, 1d8f, 1hbv, 1pro, and 2cht,[26] however, the primary set used for assessing docking accuracy and for comparison to GOLD and FlexX consists of MMFF94s-optimized geometries obtained via a short MacroModel conformational search, starting from MMFF94s-optimized versions of the (restraint-optimized) native-ligand geometries. In each case, the optimizations used a $4r$ distance-dependent dielectric model. The fourth set consists of ligand geometries obtained using Corina,[27] while the fifth set contains geometries obtained by optimizing the Corina structures with MMFF94s.

We emphasize that the initial geometry of the ligand never explicitly enters as a docked conformation. However, the conformations sampled by Glide's conformation generator depend on the input bond lengths and bond angles because these variables are not optimized. Furthermore, the fact that the potential-energy landscape has multiple minima and that finite sampling is done by Glide means that different solutions can be obtained from different starting points, even when they are very close to one another. There is also a dependence on input ring conformation, though Glide by default generates and docks alternative conformations of saturated or partly saturated five- and six-membered rings when it deems them to be energetically accessible. We generated the conformationally optimized ligand geometries cited above to make sure that no "memory" of the cocrystallized pose influenced the docking results.

All results are for flexible dockings carried out using Glide's internal conformation generator. With the exception of terminal CH_3, NH_3^+, and NH_2 groups, all rotatable bonds are treated as optimization variables. In addition, as noted above, alternative ring conformations are considered for saturated or partly saturated five- and six-membered rings, and inversion at nitrogen is performed for asymmetric trigonal nitrogen centers in compounds such as sulfonamides. The reported rms deviations in coordinate positions are based on heavy (non-hydrogen) atoms, as is also done by GOLD[1] and FlexX,[28] and are computed relative to the coordinates of the restraint-optimized native-ligand structures obtained from the protein-preparation procedure.

Table 2 summarizes the rms errors obtained as a function of the flexibility of the ligand. As expected, rms errors and CPU times increase with ligand flexibility.

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

Table 3. Comparison of rms Deviations (Å) for Flexible Docking by Glide and GOLD

	≤10 rotatable bonds (72 cases)		≤20 rotatable bonds (86 cases)		all ligands (93 cases)	
method	av rmsd	max rmsd	av rmsd	max rmsd	av rmsd	max rms
Glide	1.46	8.5	1.65	8.5	1.85	13.7
GOLD	2.56	14.0	2.92	14.0	3.06	14.0

Table 4. Comparison of rms Deviations (Å) for Flexible Docking by Glide and FlexX

	≤10 rotable bonds (133 cases)		≤20 rotatable bonds (175 cases)		all ligands (189 cases)	
method	av rmsd	max rmsd	av rmsd	max rmsd	av rmsd	max rms
Glide	1.38	8.5	1.70	9.2	1.95	13.7
FlexX	2.99	12.6	3.48	13.4	3.72	15.5

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

Both, however, are quite modest for sets of ligands having 0−8 or 0−10 rotatable bonds such as are often employed in database screens carried out to find new leads. In general, the docking performance of Glide is very reasonable over a wide range of rotatable bonds and chemical functionality. Detailed results are given in Table S1 (Supporting Information); these results show that Glide reproduces the experimentally measured binding affinities for 128 cocrystallized ligands with an rms deviation of 2.3 kcal/mol and produces rms deviations from the cocrystallized ligand position that are less than 1 Å for nearly half of the 282 cases and are greater than 2 Å in only about one-third of the cases.

Comparison to GOLD and FlexX. Detailed docking accuracy results for GOLD and FlexX are posted on the GOLD[15] and FlexX[29] web sites. These data have enabled us to make the head-to-head comparisons shown in the tables below.

Tables 3 and 4 compare rms deviations (Å) given by Glide and GOLD and by Glide and FlexX for common sets of noncovalently bound ligands having up to 10 and up to 20 rotatable bonds as well as for all ligands Glide can handle (i.e., up to 35 rotatable bonds). In some cases, the PDB structure available when the GOLD or FlexX work was done is no longer accessible but a later structure is available. In such cases, we have used the later submission. For example, Glide uses 4aah whereas GOLD and FlexX use 3aah. The Glide calculations use the conformationally optimized versions of the native ligands. The comparison for ligands having 10 or fewer rotatable bonds seems to us the most relevant to database screening applications, which usually seek to find leads that are relatively inflexible. On average, Glide gives rms deviations that are less than 60% of those given by GOLD and less than half those given by FlexX. The comparison for ligands having up to 20 rotatable bonds is also favorable.

Table 5 presents the detailed results on which the summaries in Tables 3 and 4 are based. This listing shows that Glide gives a better result nearly twice as often as GOLD and more than 4 times as often as FlexX. While far from perfect, we believe that this performance represents a qualitative improvement in docking accuracy. We should note, however, that these comparisons may not be completely fair to GOLD and FlexX. The ligands were prepared in a comparable manner for all three methods, i.e., by subjecting the native ligand structures to a force-field-based optimization procedure before docking,[1,28] though Glide also used conformational search to ensure that the starting ligand structures had no memory of the cocrystallized pose. A difference arises in the preparation of the protein sites, however, in that both the GOLD and FlexX calculations retained the original PDB coordinates for non-hydrogen atoms. It is possible that these methods might give more accurate dockings if they used our protein and ligand preparations, in which steric clashes have been annealed away. However, while the GOLD study cites four cases of noncovalently bound ligands in which the crystallographic ligand geometry appears to be incorrect (1apt, 1tdb, 1hef, and 1ive),[1] it does not ascribe any of the problematic dockings to steric clashes between the ligand and the protein; for FlexX, only one such instance (1srj) is cited.[28]

One reason for suspecting that GOLD and FlexX may be less sensitive to the details of the protein preparation is that neither uses the hard 12-6 Lennard-Jones vdW potential employed by Glide. GOLD does employ an 8-4 potential,[1] but this potential is much more forgiving of nonbonded incursions. For example, it penalizes a contact at 70% of the sum of the vdW radii by only 0.9−1.8 kcal/mol, whereas Glide penalizes such a contact by 5.5−11 kcal/mol (assuming a well depth of −0.1 to −0.2 kcal/mol). FlexX does not use a molecular mechanics expression for nonbonded repulsion. Indeed, it may use no repulsive function at all, though it does reject dockings for which the "overlap volume" exceeds 2.5 Å^3 for any particular pair of ligand and protein atoms or 1.0 Å^3 averaged over all such interactions.[28] In view of these differences, it is not clear that the results for GOLD and FlexX would be materially improved by using our preparations. However, only explicit comparisons that use identical protein preparations can resolve this question.

Comparison to Surflex. Docking accuracy is also better than that recently reported by Jain for Surflex.[30] In particular, for a common set of 78 cocrystallized PDB complexes (of 81 considered by Jain), Glide gives an average rmsd from the cocrystallized ligand pose of 1.35 Å whereas Surflex gives an average rmsd of 1.82 Å. Moreover, Glide places 47 of the 78 ligands wthin 1 Å rmsd as against 38 for Surflex and makes errors of more than 2 Å in 14 cases as against 19 for Surflex.

Cross-Docking for Thymidine Kinase. Table 6 shows rms deviations to the cocrystallized pose for docking of thymidine kinase actives to the 1kim site by GOLD, FlexX, and DOCK,[12] by Surflex,[30] and by Glide. All methods have trouble docking the acv, gcv, and pcv ligands. This is expected because acv, gcv, and pcv are purine-based ligands that do not fit properly into the pyrimidine-based 1kim site. The reason is that the Gln 125 side chain undergoes a 180° rotation on going from a pyrimidine site to a purine site and the geometry that is correct for the parent site has an acceptor−acceptor and/or a donor−donor clash in the alternative site. For the seven pyrimidine-based ligands, Glide does very well except for hmtt, which does not fit quite properly into the 1kim site when the nonpolar ligand vdW radii are scaled by 0.8 (the default). Surflex and GOLD also give

Table 5. The rms Deviations (Å) for Glide, GOLD, and FlexX for Members of the GOLD and FlexX Test Sets

complex	Glide	GOLD	FlexX	complex	Glide	GOLD	FlexX	complex	Glide	GOLD	FlexX	complex	Glide	GOLD	FlexX
121p	1.57	n/a	1.29	1aaq	1.30	12.85	1.75	1phf	1.14	n/a	4.23	1phg	4.32	1.35	4.74
1abe	0.17	0.86	1.16	1abf	0.20	n/a	1.27	1poc	5.09	1.27	9.25	1ppc	7.92	n/a	3.05
1acj	0.28	4.00	0.49	1acm	0.29	0.81	1.39	1pph	4.31	n/a	4.91	1ppi	6.24	n/a	6.91
1aco	1.02	0.86	0.96	1aha	0.11	0.51	0.56	1ppk	0.45	n/a	1.54	1ppl	2.82	n/a	5.62
1ake	3.35	n/a	1.18	1apt	0.58	1.62	1.89	1ppm	0.62	n/a	8.27	1pso	13.10	n/a	1.61
1atl	0.94	n/a	2.06	1avd	0.52	n/a	1.22	1rbp	0.96	n/a	1.13	1rds	3.75	4.78	4.89
1azm	1.87	2.52	2.37	1baf	0.76	6.12	8.27	1rne	10.08	2.00	12.24	1rnt	0.72	n/a	1.90
1bbp	4.96	n/a	3.75	1bma	9.31	n/a	13.41	1rob	1.85	3.75	7.70	1slt	0.51	0.78	1.63
1byb	10.49	n/a	1.62	1cbs	1.96	n/a	1.68	1snc	1.91	n/a	7.48	1srj	0.58	0.42	2.36
1cbx	0.36	0.54	1.35	1cde	1.29	n/a	7.45	1stp	0.59	0.69	0.65	1tdb	1.46	10.48	10.10
1cdg	3.98	n/a	4.87	1cil	3.82	n/a	3.85	1thy	2.31	n/a	2.67	1tka	2.28	1.88	1.17
1com	3.64	n/a	1.62	1coy	0.28	0.86	1.06	1tlp	1.86	n/a	2.85	1tmn	2.80	1.68	0.86
1cps	3.00	0.84	0.99	1ctr	3.56	n/a	2.82	1tng	0.19	n/a	1.93	1tnh	0.33	n/a	0.56
1dbb	0.41	1.17	0.81	1dbj	0.20	0.72	1.22	1tni	2.18	n/a	2.71	1tnj	0.35	n/a	0.89
1dbk	0.47	n/a	0.76	1dbm	1.97	n/a	2.08	1tnk	0.87	n/a	1.41	1tnl	0.23	n/a	0.71
1did	3.82	3.72	4.22	1die	0.79	1.03	4.71	1tph	0.20	n/a	1.50	1tpp	1.12	0.43	1.11
1dr1	1.47	1.41	5.64	1dwb	0.25	n/a	0.54	1trk	1.64	n/a	1.57	1tyl	1.06	n/a	2.34
1dwc	0.87	n/a	1.19	1dwd	1.32	1.71	1.66	1ukz	0.37	n/a	0.94	1ulb	0.28	0.32	3.37
1eap	2.32	3.00	3.72	1eed	5.90	12.43	9.78	1wap	0.12	n/a	0.57	1xid	4.30	0.92	2.01
1ela	1.60	n/a	9.71	1elb	4.40	n/a	7.17	1xie	3.86	0.69	1.94	2ack	0.97	4.99	2.21
1elc	8.22	n/a	4.74	1eld	0.67	n/a	6.98	2ada	0.53	0.40	0.67	2ak3	0.71	5.08	0.91
1ele	2.52	n/a	10.73	1epb	1.78	2.08	2.77	2cgr	0.38	0.99	3.53	2cht	0.42	0.59	4.58
1eta	2.92	11.21	8.46	1etr	1.48	4.23	7.24	2cmd	0.65	n/a	3.75	2cpp	0.17	n/a	2.94
1fen	0.66	n/a	1.39	1fkg	1.25	1.81	7.59	2ctc	1.61	0.32	1.97	2dbl	0.69	1.31	1.49
1fki	1.92	0.71	0.59	1frp	0.27	n/a	1.89	2gbp	0.15	n/a	0.92	2lgs	7.55	n/a	4.63
1ghb	1.89	1.45	1.33	1glp	0.34	n/a	0.47	2mcp	1.30	4.37	2.07	2phh	0.38	0.72	0.43
1glq	0.29	1.35	6.43	1hdc	0.58	10.49	11.74	2pk4	0.86	1.34	1.66	2plv	1.88	13.92	7.85
1hdy	1.74	0.94	n/a	1hef	5.30	1.87	15.32	2r04	0.80	n/a	12.55	2r07	0.48	8.23	11.63
1hfc	2.24	n/a	2.51	1hgg	2.10	n/a	10.05	2sim	0.92	0.92	1.99	2tmn	0.58	n/a	5.16
1hgh	0.28	n/a	4.14	1hgi	0.28	n/a	0.97	2xis	0.85	n/a	1.54	2yhx	3.84	1.19	2.25
1hgj	0.18	n/a	3.98	1hri	1.59	14.01	10.23	2ypi	0.31	n/a	1.22	3cla	8.51	5.45	6.42
1hsl	1.31	0.97	0.59	1hti	4.40	n/a	1.54	3cpa	2.40	1.58	2.53	3hvt	0.77	1.12	10.26
1hvr	1.50	n/a	3.35	1hyt	0.28	1.10	1.62	3mth	5.48	10.12	1.59	3ptb	0.27	0.96	0.55
1icn	2.34	8.63	10.52	1ida	11.88	12.12	11.95	3tpi	0.49	0.80	1.07	4aah	0.30	0.42	5.93
1igj	1.30	9.42	7.17	1imb	0.89	n/a	4.71	4cts	0.19	1.57	1.53	4dfr	1.12	1.44	1.40
1ivb	4.97	n/a	1.29	1ivc	1.94	n/a	2.21	4fab	4.50	5.69	4.95	4fbp	0.56	n/a	1.78
1ivd	0.72	n/a	5.42	1ive	2.61	2.16	5.34	4fxn	0.44	n/a	1.04	4hmg	0.78	n/a	5.74
1ivf	0.53	n/a	6.97	1lah	0.13	n/a	0.28	4phv	0.38	1.11	1.12	4tim	1.32	n/a	4.09
1lcp	1.98	n/a	1.65	1ldm	0.30	1.00	0.74	4tln	2.24	n/a	3.68	4tmn	1.87	n/a	8.35
1lic	4.87	10.78	5.07	1lmo	0.93	n/a	4.49	4ts1	0.85	n/a	1.41	5abp	0.21	n/a	1.17
1lna	0.95	n/a	5.40	1lst	0.14	0.87	0.71	5cpp	0.59	n/a	1.49	5cts	0.28	n/a	11.61
1mbi	1.68	n/a	0.47	1mcr	4.33	6.23	10.04	5p2p	1.82	1.55	1.00	5tim	0.58	n/a	1.99
1mdr	0.52	0.36	0.88	1mld	0.32	n/a	1.45	5tmn	2.43	n/a	4.38	6abp	0.40	1.08	1.12
1mmq	0.92	n/a	0.52	1mrg	0.30	n/a	0.81	6cpa	4.58	n/a	6.61	6rnt	2.22	1.20	4.79
1mrk	1.20	1.01	3.55	1mup	4.37	3.96	3.82	6tim	1.73	n/a	1.60	6tmn	2.66	n/a	5.10
1nco	6.99	n/a	5.85	1nis	0.97	4.29	1.41	7cpa	4.14	n/a	9.11	7tim	0.14	0.78	1.49
1nsc	1.21	n/a	2.12	1pbd	0.21	0.57	0.33	8atc	0.37	n/a	0.62	8gch	0.30	0.86	8.91
1pha	0.69	1.24	n/a	1phd	1.22	0.85	0.65	9hvp	2.68	n/a	15.54				

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

Table 6. Accuracy in Cross-Docking of Thymidine-Kinase Inhibitors to the 1kim Site[a]

	rmsd (Å) of best-scoring pose[b]				
ligand	Glide	DOCK	FlexX	GOLD	Surflex
dT	0.45	0.82	0.78	0.72	0.74
ahiu	0.54	1.16	0.88	0.63	0.87
mct	0.79	7.56	1.11	1.19	0.87
dhbt	0.68	2.02	3.65	0.93	0.96
idu	0.35	9.33	1.03	0.77	1.05
hmtt	2.83	9.62	13.30	2.33	1.78
hpt	1.58	1.02	4.18	0.49	1.90
acv	4.22	3.08	2.71	2.74	3.51
gcv	3.19	3.01	6.07	3.11	3.54
pcv	3.53	4.10	5.96	3.01	3.84

[a] The last three ligands are purines that are not expected to fit properly into a pyrimidine-based site such as 1kim (see text). [b] Data for DOCK, FlexX, and GOLD are taken from Rognan and co-workers;[12] data for Surflex are taken from Jain.[30]

quite reasonable results in this case, but FlexX and DOCK fare noticeably more poorly.

Influence of Input Ligand Geometry on Docking Accuracy. Table 7 examines the effect on the final docked structures of using different starting ligand geometries. The native ligands perform best, followed by the conformationally optimized ligands. The MMFF94s-optimized geometries used for the native ligands and the Corina structures do slightly less well but are docked more accurately than are the raw Corina structures. The last comparison shows that it helps to preminimize Corina-derived ligands with a force field such as MMFF94s.

Table 7. Effect of Input Ligand Geometry on Docking Accuracy for Glide Using the GOLD Test Set

ligand set	av rms (Å)
native ligands	1.49
MMFF94s-optimized native ligands	2.00
MMFF94s-conformational search	1.85
Corina	2.35
MMFF94s-optimized Corina	2.02

MMFF94s preminimization can be performed fairly easily on a large database with the premin script provided with Glide. These minimizations employ a 4*r* distance-dependent dielectric model and use MacroMod-

el's efficient truncated Newton minimizer. Processing times are a few tenths of a second per ligand on an AMD Athelon MP 1800+ processor running Linux. The procedure keeps track of instances in which the structure recognition (e.g., atom typing) or minimization fails and automatically submits a series of MacroModel jobs that ultimately collect the good (minimized) and bad (untreatable) structures. The user can examine and fix the bad structures or can discard them.

5. Discussion and Conclusions

This paper has described the new computational algorithms for docking and scoring we have developed for Glide and has evaluated the performance of these algorithms in predicting binding modes over a wide range of cocrystallized structures. While significant errors in binding mode prediction are found in some test cases, robustness has clearly been qualitatively improved compared to widely used alternative packages such as GOLD and FlexX. Docking accuracy is also better than that reported for the recently introduced Surflex method.[30]

Further improvements can be made to Glide with regard to both accuracy and computational efficiency. The algorithms for initial screening and energy minimization are not yet fully optimized, and a 2- to 3-fold reduction in computational effort per ligand may be attainable. The ability to impose constraints on the ligand position (e.g., by requiring that a suitable ligand atom be hydrogen-bonded to a particular protein residue or be coordinated to a metal atom in the protein) is included in Glide 2.5. This approach allows the user to guarantee satisfaction of targeted protein interactions and speeds up the calculations by reducing the size of the phase space that needs to be examined. We expect to continue to improve the scoring functions used in all three phases of the calculation (initial screening, energy minimization, binding affinity prediction). Most importantly, we have developed new sampling and scoring algorithms for Extra-Precision Glide that are proving to be highly efficient at finding correctly docked poses and at rejecting false positives in database screens;[19] these efforts will be described in a subsequent paper.

6. Methods

Conformation Generation. As a first step in its docking protocol, Glide carries out an exhaustive conformational search, augmented by a heuristic screen that rapidly eliminates conformations deemed not to be suitable for binding to a receptor, such as conformations that have long-range internal hydrogen bonds. This procedure eliminates high-energy conformers by evaluating the torsional energy of the various minima using a truncated version of the OPLS-AA molecular mechanics potential function and by imposing a cutoff of the allowed value of the total conformational energy compared to the lowest-energy state. The parameters of the heuristic screening function were optimized via extensive testing on cocrystallized PDB complexes, as described below.

Each ligand is divided into a "core" region and some number of "rotamer groups" (Figure 2), each of which is attached to the core by a rotatable bond but does not itself contain additional rotatable bonds. That is, the core is what remains when each terminus of the ligand is severed at the "last" rotatable bond, as is indicated in the figure (the directly attached atom of each rotomer group is also considered to be part of the core). Carbon and nitrogen end groups terminated with hydrogen ($-CH_3$, $-NH_2$, $-NH_3^+$) are not considered rotatable because their conformational variation is of little interest. Each core region is represented by a set of core conformations, the number of which depends on the numbers of rotatable bonds, conformationally labile five- and six-membered rings, and asymmetric pyramidal trigonal nitrogen centers in the core. As Table 8 shows, this set typically contains fewer than 500 core conformations, even for quite large and flexible ligands, and far fewer for more rigid ligands. Every rotamer state for each rotamer group is enumerated, and the core plus all possible rotamer-group conformations is docked as a single object in Glide. Because each core typically has many rotamer-group combinations, the effective number of conformations being docked can easily number in the thousands or tens of thousands for molecules having several rotatable bonds.

A key issue is how closely one of the conformations matches the correct cocrystallized conformation. An exact match is not needed because the ligand subsequently undergoes flexible torsional optimization. We have found, however, that for best results one of the starting conformations needs to be within about 1.5 Å rmsd of the correct cocrystallized conformation. We have tested our algorithm by applying it to 796 cocrystallized ligands taken from the PDB and by locating the core

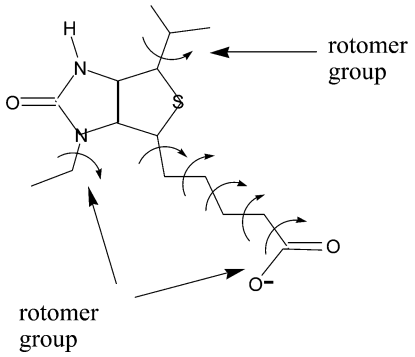


Figure 2. Definition of core and rotamer groups. The four central torsions are part of the core. Note that methyl groups are not considered rotatable.

Table 8. Maximum Number of Conformers Allowed vs Number of Core Degrees of Freedom

no. of core degrees of freedom[a]	max no. of conformers allowed	actual no. of core conformers kept for individual ligands
3	120	4, 13, 17, 18
4	120	4, 6, 18, 19, 24, 34
5	120	9, 106, 106
6	150	24, 28, 150
7	214	214, 214, ...
8	278	278, 278, ...
9	342	342, 342, ...
10	406	406, 406, ...
12	534	534, 534, ...
14	662	662, 662, ...

[a] Number of rotatable bonds in the core plus number of conformationally labile five- and six-membered rings and number of asymmetric trigonal nitrogen centers in compounds such as sulfonamides.

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

Table 9. The rms Deviations between Best-Generated and Cocrystallized Ligands, Expressed as a Percentage of Ligands Falling into Specified rms Distance Ranges

		percent of ligands having an rms deviation in the listed range, Å					
no. of rotatable bonds	no. of ligands	0.00−0.49	0.50−0.99	1.00−1.49	1.50−1.99	2.00−2.99	>3.00
1−10	495	33	37	23	15	2	0
11−15	159	3	29	45	17	6	0
16−20	87	0	10	38	31	20	1
>20	55	0	9	25	29	27	9

plus rotamer-group conformation that best matches the cocrystallized pose. The results, summarized in Table 9, show that only 7% of ligands having 1−10 rotatable bonds have rms deviations between the best core/rotamer-group conformation and the cocrystallized pose of 1.5 Å or greater. For more flexible ligands, the errors are understandably larger. Nevertheless, 77% of ligands having 11−15 rotatable bonds and 48% of ligands having 16−20 rotatable bonds have rms deviations of less than 1.5 Å. Furthermore, 89% of ligands with 11−20 rotatable bonds fall within 2 Å.

Initial Screening of Ligand Poses. For each core conformation (or for rigid docking, each ligand), Glide performs an exhaustive search of possible positions and orientations over the active site of the protein. The search begins with the selection of "site points" on an equally spaced 2 Å grid that permeates the active-site region (step 1 in Figure 1). To make this selection, precomputed distances from the site point to the receptor surface, evaluated at a series of prespecified directions and binned in 1 Å ranges, are compared to binned distances from the ligand center to the ligand surface. For flexible docking, the ligand center is defined as the midpoint of the two most widely separated atoms in the core region (which includes the directly attached atom of each rotamer group); for rigid docking, the two most widely separated atoms in the entire ligand are used. The line through these atoms is called the "ligand diameter." Glide positions the ligand center at the site point if there is a good enough match of the histograms of binned distances but skips over the site point if there is not. The ligand center can be placed at any site point on or within a box (by default, 12 Å on a side) that contains the candidate site points. The generous size of this "ligand center" box ensures that the placement of the docked ligand is not overly constrained.

The second stage examines the placement of atoms that lie within a specified distance of the ligand-diameter axis for a prespecified selection of possible orientations of the ligand diameter (step 2a). If there are too many steric clashes with the receptor, the orientation is skipped. Next (step 2b), the ligand is rotated about the ligand diameter and the subset consisting of the atoms capable of making hydrogen bonds or ligand−metal interactions with the receptor is scored ("subset test"). If this score is good enough, all interactions with the receptor are scored (step 2c).

The scoring in these three tests is carried out using a discretized version of ChemScore[18] in which precomputed scores for the ChemScore atom types are assigned to 1 Å^3 boxes. Much as for ChemScore itself, this algorithm recognizes favorable hydrophobic, hydrogen-bonding, and metal-ligation interactions and also penalizes steric clashes. This stage is called "greedy scoring" because the actual score for each atom depends not only on its position relative to the receptor but also on the best possible score it could get by moving ±1 Å in *X*, *Y*, and/or *Z*. This is done to mute the sting of the large 2 Å jumps in the site-point/ligand-center positions. The final step in stage 2 is to rescore the top greedy-scoring poses (typically ∼5000 in number) via a "refinement" procedure (step 2d) in which the ligand as a whole is allowed to move rigidly by ±1 Å in the Cartesian directions.

Energy Minimization Using a Molecular Mechanics Scoring Function. Only a small number of the best refined poses (typically 400) are minimized on precomputed OPLS-AA van der Waals and electrostatic grids for the receptor. The energy and gradient calculations are performed using standard three-dimensional interpolation methods. The Coulomb and van der Waals fields of the protein are stored at the vertexes of a grid, and the interaction of each ligand atom with these fields is evaluated using trilinear interpolation formulas for a cube. Methods of this type are in common use and have been described extensively in the literature, so we do not discuss the mathematical details here.

To ensure sufficient accuracy in regions in which the ligand and protein come into contact, Glide uses a multigrid strategy. The Coulomb/van der Waals grid is initially built using large boxes, typically 3.2 Å on a side, and is then refined hierarchically into boxes of 1.6, 0.8, or 0.4 Å depending on the distance of the box to the van der Waals surface of the protein; the smaller this distance, the higher the resolution of the box used. The resulting mesh tiles the docking volume with cubes of various size. Once the identity of the cube in which a ligand atom is contained is identified, the appropriately scaled interpolation formula can be used. While some extra bookkeeping is required, the additional computational cost as compared to a uniform mesh is negligible, whereas the reduction in memory can be more than 2 orders of magnitude.

The energy minimization typically begins on a set of Coulomb and vdW grids that have been "smoothed" to reduce the large energy and gradient terms that result from too-close interatomic contacts; it finishes on the full-scale OPLS-AA nonbonded energy surface ("annealing"). This minimization consists only of rigid-body translations and rotations when external conformations are docked rigidly. For flexible docking, however, the minimization also includes torsional motion about the core and end-group rotatable bonds. Finally, the three to six lowest-energy poses obtained in this fashion are subjected to the Monte Carlo procedure cited in section 2.

The minimized poses are then rescored using the scoring function described in section 3. As previously noted, the choice of the best-docked structure is made using a model energy score ("Emodel") that combines the energy-grid score, the binding affinity predicted by

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

GlideScore, and (for flexible docking) the internal strain energy for the model potential used to direct the conformational-search algorithm.

Optimization of the Scoring Function. The parameter optimization for GlideScore 2.5 used a simulated-annealing algorithm that has proven to be very efficient at producing large changes to the input values for the parameters when large changes are warranted. In addition to the terms described in section 3, the fitting process also considered, but ultimately rejected, a number of other prospective terms, including nearly all of those employed in ScreenScore or in its FlexX or PLP predecessors.[13]

The largest component of the objective function in the optimization process was the ranking of active compounds in a large and diverse suite of database screening tests, as measured by enrichment factors computed as described in the following paper.[16] The second component was the fit of the predicted binding affinities for a set of 125 PDB cocrystallized complexes to experimentally measured values; the rmsd achieved in this case is 2.2 kcal/mol, a reasonable result but one that can clearly be improved.

In the parametrization of GlideScore 1.8 and 2.0, we used decoy ligands assembled from cocrystallized PDB complexes and from a small portion of the Comprehensive Medicinal Chemsitry database[31] in the database screens. However, it became apparent through more extensive testing that these ligands, which average only 290 in molecular weight, are too small to make enough favorable interactions with the protein site to compete fairly with the known actives for our screens, which average about 410 in molecular weight when the very large HIV protease ligands are excluded. This disparity in molecular weight distorted both the apparent enrichment factors in the database screens and the parameters we obtained from the optimization process.

A second set of database ligands used in the present work was provided recently by a pharmaceutical colleague. This set consists of known drugs and other compounds identified in drug-discovery projects. Because our colleage intentionally chose the compounds to be "relatively small" (their average molecular weight is 337), they, too, are smaller than typical drugs and investigational compounds, which judging from Oprea's survey of the MDDR database average about 410 or 420 in molecular weight.[32] A third set of average molecular weight 350 (the "pc-350" set) consists of 1000 representative compounds drawn from a ~1 million compound database of purchasable compounds recently assembled by Schrödinger. Finally, two sets of 1000 "druglike" compounds of average molecular weight 360 and 400, the "dl-360" and "dl-400" sets, were also drawn from the million-compound database. For the pc and dl datasets, neutral database compounds were first modified by FirstDiscovery's ionizer utility to protonate or deprotonate ionizable functional groups (subject to limits of ±2 on the net charge and to a total of no more than four charged groups) to yield ionic states likely to be present in measurable concentration between pH 5 and 9. (This is to allow for shifts in pK_a induced by the protein site.) The dl sets were selected to mimic the property distribution of the drug/lead set by using a precursor to the FirstDiscovery ligparse facility. The FirstDiscovery premin utility was then employed to minimize the pc and dl ligands with MMFF94s,[25] using a "4*r*" distance-dependent dielectric. All compounds considered had 100 or fewer atoms and 20 or fewer rotatable bonds.

Table 10. Properties of 1000-Compound Ligand Databases Used in Database Screens[a]

property (av)	cmc/ pdb[b]	drug/ lead[c]	pc-350[d]	dl-360[e]	dl-400[e]
molecular weight	290	337	350	360	400
atoms	38.30	42.83	42.79	45.74	50.75
non-hydrogen atoms	22.44	26.37	25.63	28.13	31.28
higher-row atoms	0.52	0.47	0.82	0.50	0.56
% hydrophobic carbons	58.52	59.12	63.67	58.94	59.12
rings	1.53	2.09	2.33	2.23	2.51
heteroaramatic rings	0.24	0.84	0.40	0.90	0.99
rotatable bonds	5.51	5.80	5.95	6.19	6.92
amide hydrogens	0.39	0.27	0.66	0.29	0.33
neutral donors	1.61	1.46	1.02	1.56	1.70
charged donors	0.88	0.96	0.30	1.02	1.14
neutral acceptors	1.87	1.49	2.09	1.59	1.78
charged acceptors	0.63	0.42	0.11	0.45	0.49
divalent oxygens	0.55	0.68	0.78	0.73	0.82
neutral amines	0.05	0.01	0.02	0.11	0.13
acidic hydrogens	0.02	0.01	0.00	0.04	0.04

[a] Each ligand database contains 1000 compounds. [b] Taken from a subset of the Comprehensive Medicinal Chemistry database or from cocrystallized PDB complexes. [c] Compounds of relatively low molecular weight from the Derwent World Drug Index provided by a pharmaceutical collaborator. [d] Extracted from a 1-million-compound database of purchasable compounds assembled by Schrödinger in such a way as to preserve the distribution of listed properties. [e] Extracted from the million-compound database in such a way as to preserve proportionately scaled values of properties of the "drug/lead" set; however, the distribution of the last two listed properties was not controlled for.

The properties of these ligand databases are shown in Table 10. We believe that the dl-400 set is representative of ligands one would expect to find in the compound collection of a pharmaceutical or biotechnology company.

We weighted the dl-400 set the most heavily in the parametrization of GlideScore 2.5, but to broaden the parametrization, we used the others as well and also included screens run with nonstandard values for the protein and ligand scaling factors. The large number of screens employed (94 in all, embracing 16 receptor sites) should guard against overfitting and help to make database screening with Glide as tolerant as possible to variations in the ligand sets and the vdW scale factors.

The lipophilic-contact term in Glide 2.5 SP scoring (eq 2) contributes −4.85 kcal/mol to an average score of −9.33 kcal/mol for active compounds included in the database screens, and a second contact term, the vdW interaction energy, contributes −2.39 kcal/mol on average. Thus, these two terms account for nearly 80% of the total score. The hydrogen-bonding terms are next largest in importance, their average contributions for the actives being −1.12 kcal/mol, representing roughly two hydrogen bonds. One major change in the scoring concerns the way in which hydrogen bonding is evaluated. While GlideScore 2.0 differentiated hydrogen bonds on the basis of charge, our investigation of database screening results and PDB cocrystallized structures led us to a new understanding of how hydrogen bonds can be further differentiated. We believe the contribution of hydrogen bonds to binding

affinity depends substantially on the details of the hydrogen bonding in certain specific ways. On the basis of these insights, we programmed terms beyond those shown in eq 2 into Glide and optimized them against our entire range of database screens. While this is still an ongoing research area, the current parametrization leads to a substantial improvement compared to the treatment in Glide 2.0. One similarity, however, is that neutral−neutral hydrogen bonds make the largest contribution, followed by charged−neutral and then charged−charged hydrogen bonds.

Protein Preparation. Our philosophy in the application of rigid docking methods to virtual screening is that, if possible, information based on existing cocrystallized structures for the receptor of interest should be exploited. Obviously, there will be situations in which no experimentally determined structure exists, e.g., when dealing with genomic targets. In this case, a variety of strategies are possible, for example, the use of homology modeling based on cocrystallized structures for related proteins. We focus here, however, on the case in which an experimental structure is available.

Our procedure normally starts with a protein and a cocrystallized ligand. It finishes with a partially optimized protein−ligand complex to which hydrogens have been added, subject to adjustment of protonation states for ionizable residues, modification of tautomeric forms for histidine residues, and repositioning of reorientable hydrogens (e.g., side chain hydroxyl hydrogens).

The first step is to prepare the cocrystallized ligand by making sure that multiple bonds are defined correctly and that hydrogens are properly added. Normally, proteins are provided without attached hydrogens. When hydrogens are present, we usually delete all except those in peptide bonds. In such a case, it is important to note where the authors of the structure have assigned nonstandard protonation states so that these can be reimposed later if this appears warranted. Cofactors, which are included as part of the protein, need to have multiple bonds and formal charges assigned properly so that hydrogens will be added correctly in a later step.

The second step uses the pprep script that is provided with Glide. This procedure neutralizes residues that do not participate in salt bridges and that are more than a specified distance from the nearest ligand atom.[33] By default, it choses the value between 10 and 20 Å for this distance that minimizes the total charge of the receptor plus ligand complex. The script also sets the tautomeric state for His residues, which are assumed to be neutral, by considering potential metal-ligation and hydrogen-bonding interactions.

The third step is to postprocess the pprep'd receptor. This is necessary because the judgments made by the preparation procedure will not always be correct. For example, for an aspartyl protease such as HIV, both active-site aspartates will be close to one or more atoms of a properly docked ligand and hence will be represented as negatively charged. One of these aspartates, however, is typically taken to be neutral in modeling studies. Similarly, Glu 143 in thermolysin typically interacts with an acceptor oxygen of the bound ligand. This residue may need to be protonated, as may His 231, which forms a salt bridge with Asp 226 and typically places its second ring nitrogen atom within about 3 Å of a second zinc-bound oxygen of a carboxylate or phosphonate ligand. Other special circumstances can also arise. In addition, if the protein had some or all hydrogens attached, the original and prepared versions of the protein need to be compared to decide how to resolve any discrepancies.

Step four adds structural waters if any are to be kept. Nearly all waters have been removed in this work. Exceptions are 1add, 1adf, 1ebg (three waters), 1lna (two waters), and 1mdr, where a water molecule is tightly bound to a Mg^{2+} ion. By comparison, GOLD retains waters for 2ctc, 1mdr, and 1nis,[1] and FlexX does so for 1aaq, 1lna, 1xie, and 4phv.[28] Omitting structural waters can be useful because it may allow ligands to be found that are capable of displacing them. For example, we removed the water under the flaps in HIV protease to allow ligands such as the DuPont-Merck cyclic urea, which displaces this water, to dock. The resultant overly generous active site might have encouraged "false positives" in the docking. However, excellent rank orders were obtained for known HIV ligands in the database screen.[16]

The fifth step adds hydrogens to the protein, to any cofactors, and to any added structural waters, and the final step carries out a series of restrained minimizations on the protein−ligand complex. The first stage reorients repositionable side chain hydroxyls in Ser, Thr, and Tyr residues and side chain sulfhydryls of Cys. This is accomplished by tightly tethering non-hydrogen atoms (force constant = 10 kcal mol^{-1} $Å^{-2}$) and minimizing the hydrogens with torsion interactions turned off. In effect, the hydrogens are allowed to fly freely in the electrostatic wind. Subsequent steps restore the torsion potential and use progressively weaker restraints on the non-hydrogen atoms (hydrogen atoms are always free); the force constants employed are 3, 1, 0.3, and 0.1 kcal mol^{-1} $Å^{-2}$. The procedure stops when the cumulative rms deviation from the initial coordinates for non-hydrogen atoms exceeds a target value the user specifies, the default for which is 0.3 Å. The last structure having an rms deviation smaller than the target value is then selected. We perform these minimizations with either the Impact[34] or Macromodel[35] protein molecular modeling codes. When MacroModel is used, MMFF94s[25] can be employed instead of OPLS-AA.

Our experience is that this or an equivalent preparation procedure is important for attaining accurate docking with Glide. It is essential that physically untenable steric clashes often found in crystallographically determined protein sites be annealed away so that the native ligand (and others) can yield favorable vdW interaction energies for properly docked structures. It is also important that protonation states and hydrogen-bonding patterns be correct.

Acknowledgment. This work was supported in part by grants to R.A.F. from the NIH (Grants P41 RR06892 and GM 52018).

Supporting Information Available: Table S1, listing detailed results for docking ligands from 282 cocrystallized complexes into the protein sites with Glide 2.5. This ma-

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430

terial is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Jones, G.; Wilett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a generic algorithm and an empirical binding free energy function. *J. Mol. Biol.* **1997**, *267*, 727−748.

(2) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A. A fast flexible docking method using an incremental construction algorithm. *Chem. Biol.* **1996**, *261*, 470−489.

(3) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18*, 1175−1189. Ewing, T. J. A.; Makino, S.; Skillman, G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411−428.

(4) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM—a new method for protein modeling and design. Application to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488−506. Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375−382.

(5) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: fast, fully automated docking of ligands to protein binding sites. *Chem. Biol.* **1996**, *3*, 449−462.

(6) McMartin, C.; Bohacek, R. QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333−344.

(7) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. Automated docking using a Lamarkian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(8) Murray, C. W.; Baxter, C. A. Frenkel, A. D. The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin, and neuraminidase. *J. Comput.-Aided. Mol. Des.* **1999**, *13*, 547−562.

(9) McGann, M.; Almond, H.; Nicholls, A.; Grant, J. A.; Brown, F. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76−90.

(10) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. J. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289−307.

(11) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method of obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.

(12) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.

(13) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.

(14) Schrödinger, L.L.C., New York. The Glide 2.5 calculations used FirstDiscovery, version 2.5021, which was released in June 2003.

(15) The results for the GOLD test set are available at http://www.ccdc.cam.ac.uk/prods/gold/rms_tab.html.

(16) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(17) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(18) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.

(19) Murphy, R. B.; Friesner, R. A.; Halgren, T. A. Unpublished research.

(20) Babine, R. E.; Bender, S. L. Molecular recognition of protein−ligand complexes: Applications to drug design. *Chem. Rev.* **1997**, *97*, 1359−1472.

(21) Whittaker, M.; Floyd, C. D.; Brown, P.; Gearing, A. J. H. Design and therapeutic application of matrix metalloproteinase inhibitors. *Chem. Rev.* **1999**, *99*, 2735−2776.

(22) Bell, I. M.; Gallicchio, S. N.; Abrams, M.; Beese, L. S.; Beshore, D. C.; Bhimnathwala, H.; Bogusky, M. J.; Buser, C. A.; Culberson, J. C.; Davide, J.; Ellis-Hutchings, M.; Fernandes, C.; Gibbs, J. B.; Graham, S. L.; Hamilton, K. A.; Hartman, G. D.; Heimbrook, D. C.; Homnick, C. F.; Huber, H. E.; Huff, J. R.; Kassahun, K.; et al. 3-Aminopyrrolidinone farnesyltransferase inhibitors: design of macrocyclic compounds with improved pharmacokinetics and excellent cell potency. *J. Med. Chem.* **2002**, *45*, 2388−2409.

(23) The reduction in the vdW radii is needed to roughly preserve hydrogen-bonding distances, which reflect a balance between attractive electrostatic and repulsive vdW forces.

(24) Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ. http://www.rcsb.org.

(25) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720−729.

(26) For 1cps, we used the native ligand structure because the ligand contains functionality that cannot be treated by MMFF94s. The remaining four cases have larger rings that Glide's conformation generator cannot handle and for which the conformational search yielded a different ring conformation. For 1hbv and 1pro, the MMFF94s-optimized native ligand geometry was used instead, while for 1d8f and 2hct, conformational search was employed but the ring conformation was not sampled.

(27) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automated generation of 3D atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.

(28) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FlexX incremental construction algorithm for protein−ligand docking. *Proteins* **1999**, *37*, 228−241.

(29) The results for the FlexX test set are available at http://cartan.gmd.de/flexx/html/flexx-eval.html.

(30) Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular-similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499−511.

(31) CMC database is available from MDL Information Systems, Inc., San Leanrdo, CA.

(32) Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **1999**, *14*, 251−264. See Figure 2 in this reference.

(33) This distance is defined as the distance from a heteroatom that bears or shares a positive or negative formal charge in the ionized protein residue to the nearest ligand atom.

(34) Impact is the computational engine of Schrödinger's FirstDiscovery suite. Impact was developed in the laboratories of Prof. Ronald Levy (Rutgers University). It and the FirstDiscovery suite are available from Schrödinger, L.L.C., New York.

(35) MacroModel (formerly known as BatchMin) is a general-purpose molecular mechanics program available from Schrödinger, L.L.C., New York. MacroModel was developed in the laboratories of Prof. Clark Still (Columbia University).

JM0306430

Downloaded by UNIV OF TEXAS AUSTIN on September 26, 2009 | http://pubs.acs.org
Publication Date (Web): February 27, 2004 | doi: 10.1021/jm0306430