Detailed Analysis of Grid-Based Molecular Docking: A Case Study of CDOCKER—A CHARMm-Based MD Docking Algorithm

GUOSHENG WU,[1] DANIEL H. ROBERTSON,[1] CHARLES L. BROOKS III,[2] MICHAL VIETH[1]

[1]*Eli Lilly and Company, Lilly Research Laboratories, Lilly Corporate Center, DC 1513, Indianapolis, Indiana 46285*

[2]*Department of Molecular Biology, TPC6, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, California 92037*

Abstract: The influence of various factors on the accuracy of protein-ligand docking is examined. The factors investigated include the role of a grid representation of protein-ligand interactions, the initial ligand conformation and orientation, the sampling rate of the energy hyper-surface, and the final minimization. A representative docking method is used to study these factors, namely, CDOCKER, a molecular dynamics (MD) simulated-annealing-based algorithm. A major emphasis in these studies is to compare the relative performance and accuracy of various grid-based approximations to explicit all-atom force field calculations. In these docking studies, the protein is kept rigid while the ligands are treated as fully flexible and a final minimization step is used to refine the docked poses. A docking success rate of 74% is observed when an explicit all-atom representation of the protein (full force field) is used, while a lower accuracy of 66–76% is observed for grid-based methods. All docking experiments considered a 41-member protein-ligand validation set. A significant improvement in accuracy (76 vs. 66%) for the grid-based docking is achieved if the explicit all-atom force field is used in a final minimization step to refine the docking poses. Statistical analysis shows that even lower-accuracy grid-based energy representations can be effectively used when followed with full force field minimization. The results of these grid-based protocols are statistically indistinguishable from the detailed atomic dockings and provide up to a sixfold reduction in computation time. For the test case examined here, improving the docking accuracy did not necessarily enhance the ability to estimate binding affinities using the docked structures.

© 2003 Wiley Periodicals, Inc.   J Comput Chem 24: 1549–1562, 2003

Key words: MD docking; grid approximations; CDOCKER; CHARMm; *p* value

Introduction

In this era of computer-aided structure-based drug design, molecular docking is frequently used to predict the putative geometry of a protein-ligand complex. The success of this computational methodology can be rapidly verified by parallel crystallography efforts. In addition, docking is often used in conjunction with scoring functions to predict binding affinities of ligands in virtual screening experiments[1] and in studying structure activity relationships to prioritize synthesis of new compounds.[2,3] The force field or docking function defining the energetics of the system is the underlying foundation for the methodology, as it is this target function that the docking algorithms attempt to optimize. The major goal of a good docking function is to discriminate between the manifold of true solutions, usually defined as poses within 2.0 Å root mean square deviation (RMSD) from the X-ray geometry, and false solutions or misdocked structures.[4,5]

While many interesting reviews[1,6] have provided insights into the utility of docking in drug design, there is continuing need to critically examine the approximations and strategies used in these algorithms. One approximation that is commonly used is holding the protein structure rigid. While this assumption is justified in many instances, researchers have proposed methodology to handle the cases where it is not a reasonable constraint.[7] Nevertheless, the ability of a docking algorithm to find the best-docked ligand structure (as defined in the original X-ray ligand-protein complex) is still lacking. Even with the rigid protein constraint, the conformational space of the system can be overwhelmingly large as it

includes the six external translational-rotational degrees of freedom as well as the numerous internal degrees of freedom of the ligand. To adequately explore the conformation space, many different optimization methods and search strategies have been developed, including distance-geometry,[8] Monte Carlo (MC) simulated-annealing,[9] genetic algorithms (GAs),[10] tabu search,[11] and molecular dynamics.[12–16] Other factors that complicate the characterization of an algorithm's performance to reproduce the X-ray structure geometries involve the choice of the validation set, the protocol used for the testing, or the preparation of the ligand structure. Even the most optimistic benchmarks show that achieving high docking accuracy is a challenging task.[11,17,18] Because docking is often subsequently followed by scoring of the docked ligands, the performance of ligand ranking—which is usually an essential measure of the success of computer-aided drug design in real projects—can potentially be undermined by poor accuracy during the docking process.

Molecular dynamics (MD) is a general simulation technique that is included in many molecular modeling packages such as CHARMm[19] and AMBER.[20] Despite its popularity for the simulation of biomolecules,[21,22] it is interesting that MD is rarely used for protein-ligand docking[12–16] and has not currently been included as an available docking methodology in commercial docking packages. A primary reason behind MD's lack of popularity for docking can be attributed to the length of the computation time,[3] which makes virtual screening of large compound sets prohibitively time-consuming. The role of MD in the drug design process is likely to grow as computer power increases and as these MD methodologies are incorporated into practical docking packages. This article studies our implementation of a grid-based MD docking algorithm, CDOCKER (CHARMm-based DOCKER), which offers all the advantages of full ligand flexibility (including bonds, angles, dihedrals), the CHARMm[19] family of force fields, the flexibility of the CHARMm engine, and reasonable computation times. CDOCKER is based on previous work demonstrating the improved efficiency and accuracy of automated MD docking with a soft-core potential over MC and GA in searching a large configuration space when using a detailed atomic force field.[15]

This article extends the previous work by providing a detailed analysis of the effects of a grid-based implementation of this methodology in an attempt to provide computational efficiency while retaining much of the accuracy of the full force field method. While this study focuses on MD as a search engine, the majority of conclusions are generally applicable to other grid-based search strategies. A grid-based representation of interactions has been applied widely in many docking protocols using a tri-linear interpolation to approximate potentials for ligand atoms located between the grid points.[9,12,23] This interpolation is very effective in reducing the computational time. Various grid approximations have been investigated for their ability to reproduce all-atom energetics,[24] but a detailed analysis of the grid compared with a full force field approach has not been performed using docking accuracy as a gauge of success. It is generally believed that using the grid approximation will result in some loss of docking accuracy, but to what extent has not been studied. Some researchers use hierarchical approaches to refine the final structures with lengthy all atom simulations.[25] Can one find a more balanced approach that minimizes the grid approximations without sacrificing the compu-

tation time? Furthermore, is the tri-linear interpolation sufficient to provide similar accuracy for grid-based docking to the full force field approach? In this article, these questions are addressed with an extensive analysis of the MD docking of 41 diverse protein-ligand complexes.

Methodology

Docking Strategy

The general docking protocol and potential functions employed in CDOCKER have been described in prior articles.[4,15] In this work, we use a variation of the "short annealing" schedules described in detail previously.[3] The basic strategy involves the generation of several initial ligand orientations in the target protein's active site followed by MD-based simulated annealing, and final refinement by minimization. The annealing schedule includes multiple heating and cooling stages and is implemented as a set of scripts for the CHARMm package using its replica feature.[26]

An important advance to the CDOCKER protocols has been the introduction of the soft-core potentials. Soft-core potentials are found to be effective in exploring the conformational space of small organics and macromolecules and are being used in various applications, including docking[4,27,28] and the prediction of protein loop conformations.[29] During the docking process, the nonbonded interactions [including van der Waals (vdW) and electrostatics] are softened at different levels, but this softening is removed for the final minimization. In this study, the soft-core potentials are implemented as previously described for the DOMCOSAR protocol[3] and the details of the annealing schedule and soft-core potentials are summarized in Table 1.

Test Set of Protein-Ligand Complexes

A test set of 41 diverse protein-ligand complexes was selected from the PDB.[30] This set of 41 comprises a subset of the test suite introduced by Baxter et al.,[11] which excludes complexes where the ligand-protein interactions are mediated by metal ions. The ligands available in this test set are structurally diverse with varying degrees of flexibility, including from 0 to 22 rotatable bonds. Two-dimensional representations of these ligands together with several simple molecular descriptors are shown in Table 2. For the target proteins in this study, we used a polar hydrogen representation and the param19/toph19 parameter set.[19] The proteins were prepared for docking by adding and minimizing polar hydrogens while holding the heavy atoms of the protein fixed. The ligands were modeled using an all-atom representation with formal charges assigned by a set of rules to match protonation states of the ionizable groups at pH = 7. The partial atomic charges and atom types for the ligands were assigned by an automated protocol designed to match the Momany-Rone force field[31] as implemented in INSIGHT2000.[32] The initial ligand geometries were automatically created from the 2D representation using CORINA.[33] The influence of initial geometry on the docking results is also included in this study. We compare docking results that start from a randomly positioned ligand's X-ray geometry to those that start from the CORINA geometry. The active site center is assigned as the

Table 1. Parameters for the MD Docking Protocol.

Docking stage	Sim. time[a] (ps)	Initial T (K)	Final T (K)	Soft-core potentials[b]		
				E_{max}(vdW)	E_{max}(att)	E_{max}(rep)
1	4.5	300	700	1.5	−10.0	20.0
2	21.0	700	300	0.6	−0.4	8.0
3	10.5	500	300	3.0	−20.0	40.0
4	4.5	400	50	30.0	−200.0	400.0

[a]A time step of 1.5 fs was used at all stages.
[b]The soft-core function form is described by eq. 1[3]. E_{max}(vdW), E_{max}(att), and E_{max}(rep)[kcal $\cdot$ mol^{-1}] is for the van der Waals, electrostatic attractive, and electrostatic repulsive parts, respectively. A distance dependent dielectric constant of 3, and vdW switching functions with inner switching distance at 6 Å and outer at 8 Å were employed.

center of mass of the ligand from X-ray structure of the complex. The ligand's center of mass is restrained to this active site using a flat-bottomed harmonic restraint with a distance of 6.0 Å before the constraining force activates. Using this ligand restraint, the effective search space explored varies from approximately 4200 Å^3 for the smaller ligands to approximately 21000 Å^3 for the larger ligands.

Grid and Force Field

The grid origin was located at the center of the active site with a minimum of 21 Å or the largest ligand dimension + 5 Å as the side length. A grid spacing of 0.5 Å was used. No significant differences in docking accuracy were observed using a grid spacing ranging from 0.25 to 1.0 Å. For each defined vdW or electrostatic probe, the interactions with all protein atoms were stored at these grid points. For ligand atoms located between grid points, a trilinear interpolation was used to approximate the energies. A harmonic potential with the force constant of 300 kcal $\cdot$ mol^{-1} was applied outside the grid boundary.

A +1 point-charge probe was used to map electrostatic interactions to the grid. The grid's vdW interactions were generated in one of two ways. The first used probe radii that were equally spaced (0.65 through 2.55 Å with 0.1 Å distance) in a scheme referred to as GRID I. In the second approach they were chosen discretely based on the distribution of atomic radii in a log P database of 11 K diverse compounds[34] (scheme referred to as GRID II). In GRID II, the 11 selected probes represent 88% of the total atoms in this database and the remaining 12% of the atoms have radii within 0.05 Å to these selected probes (Fig. 1). The need to use this larger number of probes was necessitated by the complexity of CHARMm's vdW potentials, whereas a single probe could have sufficed for a generic 12-6 potential. The forces were implemented by one of two approaches: computed analytically from the interpolation formula of energy grid, in the scheme referred to as GRID I; or interpolated directly from the precomputed force grids in GRID II. Our initial expectation was that the GRID II energies and forces would minimize the degree of grid approximation and improve docking accuracy.

For both GRID I and GRID II, the energies and forces were generated with two options for the soft-core potentials in the heating-cooling stages, and with normal nonbond potentials in the

final minimization step. The soft core was approximated by the following function:
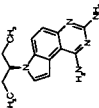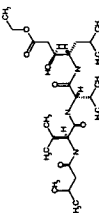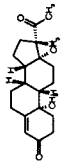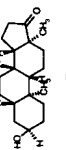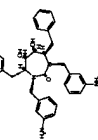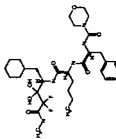
$$E_{ij}(r_{ij}) = E_{max} - a \cdot r_{ij}^b \quad \text{if } |E_{ij}^*| > \frac{|E_{max}|}{2} \qquad (1)$$
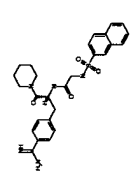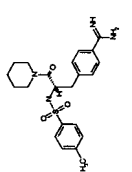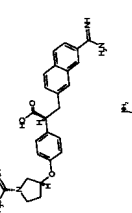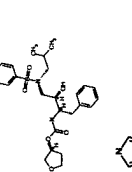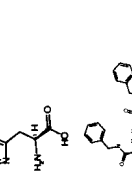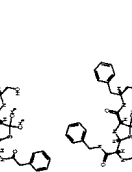
where E_{ij}^* is the energy of the regular nonbond (vdW or electrostatic) potential. The coefficients a and b were extracted from two equations that express equality of the regular and the soft potential and forces at the switching distance. Note that for the electrostatic attraction both E_{max} and a are negative.[3] For the first two stages in Table 1, the grid file is generated with E_{max}(vdW) = 0.6, E_{max}(att) = −0.4, and E_{max}(rep) = 8.0. For the last two stages, the grid file is generated with E_{max}(vdW) = 3.0, E_{max}(att) = −20, and E_{max}(rep) = 40.0 (*att* stands for electrostatic attraction, *rep* stands for electrostatic repulsion). Note that in stages 1 and 4, the soft-core parameters are different between the grid and full force field calculations, but our tests showed little difference in docking accuracy with these four grids. Figures 2A and 2B display the interaction energies and forces for the x direction of a soft-core potential [E_{max}(vdW) = 3, E_{max}(att) = −20, and E_{max}(rep) = 40][3] between a probe atom (nonbond parameters σ = 1.55 Å, ε = 1.0 kcal/mol, and q = +1.0) and protein 1abe. Despite some differences between the grid energies and forces from those of the full force field, the grid-based approximations retain similar overall curvature. Although the error in energies for both the grid methods is small, a discontinuity of the forces is present in GRID I.

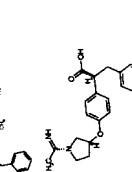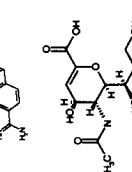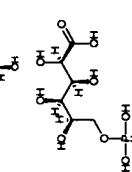### *Details of the CDOCKER Docking Protocol*

In the standard protocol, 50 replicas[26] for each ligand are generated and randomly distributed around the center of the active site. The internal coordinates for each of the replicas are kept the same as those originally generated from CORINA. The MD simulated annealing process is performed using a rigid protein and flexible ligand. The ligand-protein interactions are computed from either GRID I, GRID II, or the full force field. A final minimization step is applied to each of the ligand's docking poses. The minimization consists of 50 steps of steepest descent followed by up to 200 steps of conjugate-gradient using an energy tolerance of 0.001 kcal $\cdot$ mol^{-1}. These minimized docking poses are then clustered based
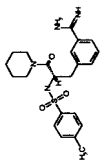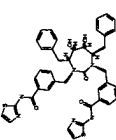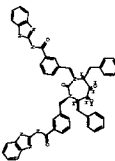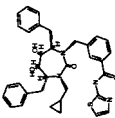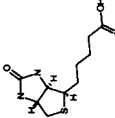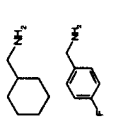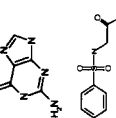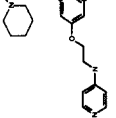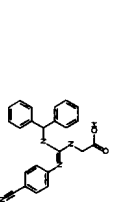
Table 2. Properties of Ligands and Detailed Docking Results.

Structure[a]	PDB[b]	pK_i Data[c]	Rot[d]	MW[e]	RMSD (C−X)[f]	Best RMSD[g]	RMSD (Full mean)[h]	RMSD (GRID I mean)[h]	RMSD (GRID II mean)[h]	Full FF[i]	GRID I[i]	GRID II[i]	Docked-X-ray energy difference[j]
	1abe	9.58	0	150.1	0.1	0.3	0.55	0.38	0.44	0.9	1	1	−6.39
	1abf	7.40	0	164.2	0.1	0.3	0.40	0.40	0.52	1	1	1	−8.12
	1aoeA	13.18	3	269.4	0.5	0.3	2.25	1.59	1.05	0.7	0.8	0.9	−0.26
	1apu	10.51	19	485.7	1.5	0.7	2.05	2.62	1.94	0.8	0.5	0.9	−0.37
	1dbb	12.28	1	314.5	0.4	0.5	0.59	0.59	1.10	1	1	0.9	−2.98
	1dbj	10.48	0	290.4	0.2	0.3	0.41	0.41	0.41	1	1	1	−9.68
	1dmp	13.04	8	536.7	3.3	0.4	0.87	1.01	0.90	1	1	1	−2.49
	1dog	5.48	1	163.2	0.6	3.7	3.46	3.36	3.37	0	0	0	−0.05
	1dwb	3.99	0	120.2	1.1	0.8	0.82	0.82	0.82	1	1	1	−2.86
	1epo	10.86	20	653.8	2.3	0.9	3.54	3.66	2.30	0.6	0.4	0.8	−9.05

1ets	11.63	11	521.6	3.1	0.6	2.43	2.18	2.83	0.7	0.7	0.6	−0.02
1ett	8.44	8	428.6	2.3	1	1.95	2.85	1.72	0.7	0.3	0.6	−9.41
1fax	10.08	8	444.5	2.6	1.1	2.99	1.85	2.55	0.8	0.9	0.8	−18.49
1hpv	12.59	13	505.6	1.5	0.8	1.59	3.69	1.68	0.9	0.6	0.9	−0.04
1hslA	9.96	3	155.2	1.5	0.3	0.81	0.91	1.05	1	1	1	−7.51
1htf	11.05	15	574.7	4.3	6.3	9.18	8.82	8.37	0	0	0	−2.35
1hvr	12.98	8	606.8	3.4	0.5	1.05	0.92	1.17	1	1	0.9	−6.70
1mtw	7.50	8	444.5	2.9	1.3	1.86	2.53	2.49	0.6	0.2	0.5	−8.07
1nsd	7.23	6	291.3	0.9	0.7	0.84	0.88	0.92	1	1	1	−0.02
1pgp	7.78	7	276.1	1.9	1.5	2.05	1.54	1.55	0.8	1	1	−5.96

(continued)

Table 2. (*Continued*)

Structure[a]	PDB[b]	pK_i Data[c]	Rot[d]	MW[e]	RMSD (C−X)[f]	Best RMSD[g]	RMSD (Full mean)[h]	RMSD (GRID I mean)[h]	RMSD (GRID II mean)[h]	Full FF[i]	GRID I[i]	GRID II[i]	Docked-X-ray energy difference[j]
	1pph	8.49	8	428.6	1.8	0.3	2.76	3.37	3.45	0.5	0.4	0.2	−0.92
	1qbr	14.42	14	758.9	2.9	0.3	1.07	1.14	1.21	0.9	0.9	0.9	−2.74
	1qbt	14.50	14	824.9	3	0.2	1.90	1.41	0.78	0.4	0.7	1	−0.04
	1qbu	13.97	11	596.8	3	0.9	5.04	4.61	6.49	0.3	0.6	0.3	−4.22
	1stp	18.29	5	244.3	1.1	0.6	0.87	0.77	0.87	1	1	1	−0.05
	1tng	4.00	1	113.2	0.6	0.3	0.33	0.33	0.34	1	1	1	−11.67
	1tnh	4.60	1	125.1	0.6	0.6	0.66	0.66	0.65	1	1	1	−0.10
	1ulb	7.23	0	151.1	0.1	0.7	1.37	0.75	0.75	0.9	1	1	−8.70
	1uvs	7.37	11	467.6	2.6	4.2	6.24	5.31	6.06	0	0	0	−7.53
	1uvt	10.43	8	383.5	2.4	2.3	3.86	4.59	4.65	0	0	0	−0.16
	2cgr	9.93	8	384.4	1.7	0.9	2.32	4.21	2.83	0.8	0.4	0.7	−0.09

Structure	PDB[b]	pK_i[c]	Rot.[d]	MW[e]	RMSD CORINA[f]	RMSD[g]	Mean RMSD FF[h]	Mean RMSD GRID I[h]	Mean RMSD GRID II[h]	Freq FF[i]	Freq GRID I[i]	Freq GRID II[i]	ΔE[j]
	2gbp	10.37	1	180.2	0.1	0.3	0.44	0.47	0.44	1	1	1	−3.54
	2ifb	7.41	14	256.4	2.2	0.7	1.02	1.03	1.14	1	1	1	−19.62
	2pph	6.39	9	559.3	3	2.1	5.53	4.61	3.15	0	0	0.3	−5.43
	2r04	8.49	10	342.4	1.3	1.1	3.87	1.68	1.67	0.6	0.8	0.7	−0.01
	2ypi	6.58	3	156.0	0.9	1	1.40	1.56	1.57	1	1	1	−17.36
	3ptb	6.47	0	120.2	1.1	0.9	1.06	1.06	1.06	1	1	1	−3.25
	3tpi	5.88	7	230.3	1.1	0.6	0.94	1.17	0.99	1	0.9	1	−0.20
	4dfrA	13.24	10	454.4	1.5	1	4.65	3.76	4.42	0.1	0.3	0.1	−0.04
	4tpi	3.89	6	216.3	0.9	0.6	1.10	1.10	1.26	1	1	0.9	−11.14
	5abp	9.06	1	180.2	0.1	0.3	0.77	1.04	0.97	1	1	1	−6.02

[a] Two-dimensional drawings of the ligand chemical structure.
[b] PDB code for the complex; chain identifier, if any, is included as the fifth character.
[c] pK_i data [$-\log(K_i[M])$] for the complexes.[11]
[d] The number of ligand rotatable bonds.
[e] Ligand molecular weight.
[f] Heavy atom RMSD in Å between the initial CORINA geometry and the ligand geometry from X-ray of complex.
[g] Docking results (smallest RMSD of the best pose over 10 independent docking runs) from full force-field.
[h] The mean RMSD of the best pose over 10 independent docking runs for the full force-field, GRID I, and GRID II.
[i] Frequencies of docked structures over 10 runs for the grid and full force-field methods.
[j] Energy difference in [kcal · mol^{-1}] between the lowest energy structure from docking with the corresponding RMSD (See footnote g) and the minimized X-ray structure.
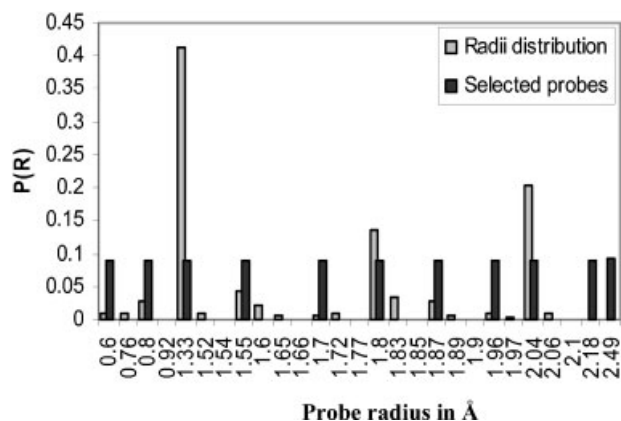
Figure 1. The selection of van der Waals probes based on radii distribution of atoms in a diverse log *P* database of 10,971 molecules from Biobyte.[34] There are 27 different radii (represented by gray bars) representing all atom types in this set. Eleven probes (black bars) were selected and they cover about 88% of all of the atoms.

on a heavy atom RMSD approach using a 1.5 Å tolerance.[35] The final ranking of the ligand's docking poses is based on the total docking energy (including the intramolecular energy for ligands and the ligand-protein interactions). A ligand-protein docking is considered a success if the RMSD between the top ranking (lowest energy) docking pose and the ligand's X-ray position is less than 2.0 Å. The docking accuracy is then computed as the percentage of successfully docked ligands from this 41-complex test set.

To investigate CDOCKER's variability, 10 independent runs were performed for each method (full force field, GRID I, and GRID II) so that a statistical analysis could be conducted. For each of these independent runs, all of the ligand starting orientations (50 replicas) were chosen using a different random seed. The observed variation of docking results from these independent runs is a measure of the variability of the algorithm and allows one to statistically differentiate the effects of the various grid approximations.

Results and Discussion

General Statistics of Docking Accuracy

Figure 3A shows the docking success rates from 10 independent runs for all three methods. The full force field runs give the highest mean success rate of 74% (30.4/41), with GRID II exhibiting a lower rate of 68% (27.9/41), and GRID I showing a slightly lower success rate of 66% (27.1/41). Both grid approaches significantly under-perform the full force field runs as indicated by *p* value of 0.0003. There is no significant difference in the docking accuracy between these two grid approaches (*p* value of 0.2868).[36]

To improve these initial grid results, a series of runs was performed where the full force field was used in the minimization steps for each of the GRID methods rather than using their own grids. Switching to the full force field during the final minimization did not sacrifice much efficiency as this minimization takes less

than 1% of the total computing time. The observed improvement in accuracy of the grid-based dockings is demonstrated in Figure 3B. The mean success rate for GRID I improved to 70% (1.8 more ligands on average), and to 76% for GRID II (3.4 more ligands on average). The *p* value[3,36] of 0.2168 indicates that the full force field results are now comparable to the GRID II results, while both are statistically better than GRID I (*p* values smaller than 0.05 indicate that the means are significantly different). These results clearly indicate that full force field minimization leads to a significant improvement in docking accuracy. Additionally, the better representation of energies and forces (Fig. 2) in GRID II protocol leads to docking success rates that are statistically indistinguishable from the results using the full force field. The results demonstrate that a good implementation of a grid-based energy representation can perform competitively with full force-field-based results. As indicated in Figure 2B, the forces for GRID II retain the overall curvature of the full force field, which appears to be sufficient to work reliably in the docking search; however, the final minimization effect is correlated with the detailed variations of the interpolated forces. Thus, we conclude that the optimal way to correct for the inadequacies of grid representation at this fine level of detail is to utilize the full force field in the final minimization step. Although an improvement in the accuracy was also observed for GRID I using the full force field in the final minimization step, it appears that this minimization could not overcome the problems related to the more substantial approximations and discontinuities of the forces represented in GRID I.

An additional measure of the quality of the grid approximation is to compare the final energies obtained from the docking using grid and full force field strategies. The mean values of docking energies in the 10 runs are calculated for the top ranking poses and shown in Figure 4A (grid potential for final minimization) and in Figure 4B (full force field for final minimization). The large differences in the final off- and on-grid energies are substantially decreased (Fig. 4B) as a result of the utilization of full force field minimization. The RMS (root-mean-square) value of the energy difference between grid and full force field for the best poses is 10.24 kcal · mol^{-1} for the on-grid GRID I method and 7.69 kcal · mol^{-1} for the on-grid GRID II method. When the full force field minimization is applied in the final step, these differences are reduced to 1.38 kcal · mol^{-1} and 1.58 kcal · mol^{-1} for GRID I and GRID II, respectively.

As the full force field final minimization has proven to be beneficial for both reducing the grid's energetic discrepancies and improving the docking accuracy, it has been adopted as the default docking strategy in CDOCKER. This hybrid strategy is consistently used for all the remaining studies presented in this article. The accuracy of the grid-based energy representation could be improved using other interpolation schemes such as B-splines[24] or logarithmic interpolation.[37] These schemes predominantly improve the repulsive hard-core interactions, which in our docking scheme are already substantially reduced by using the soft-core potential. Because the inadequacies of the GRID II force field have been largely corrected by using the full force field during the final minimization step, an enhanced interpolation scheme is less likely to significantly improve the results of our current protocol.
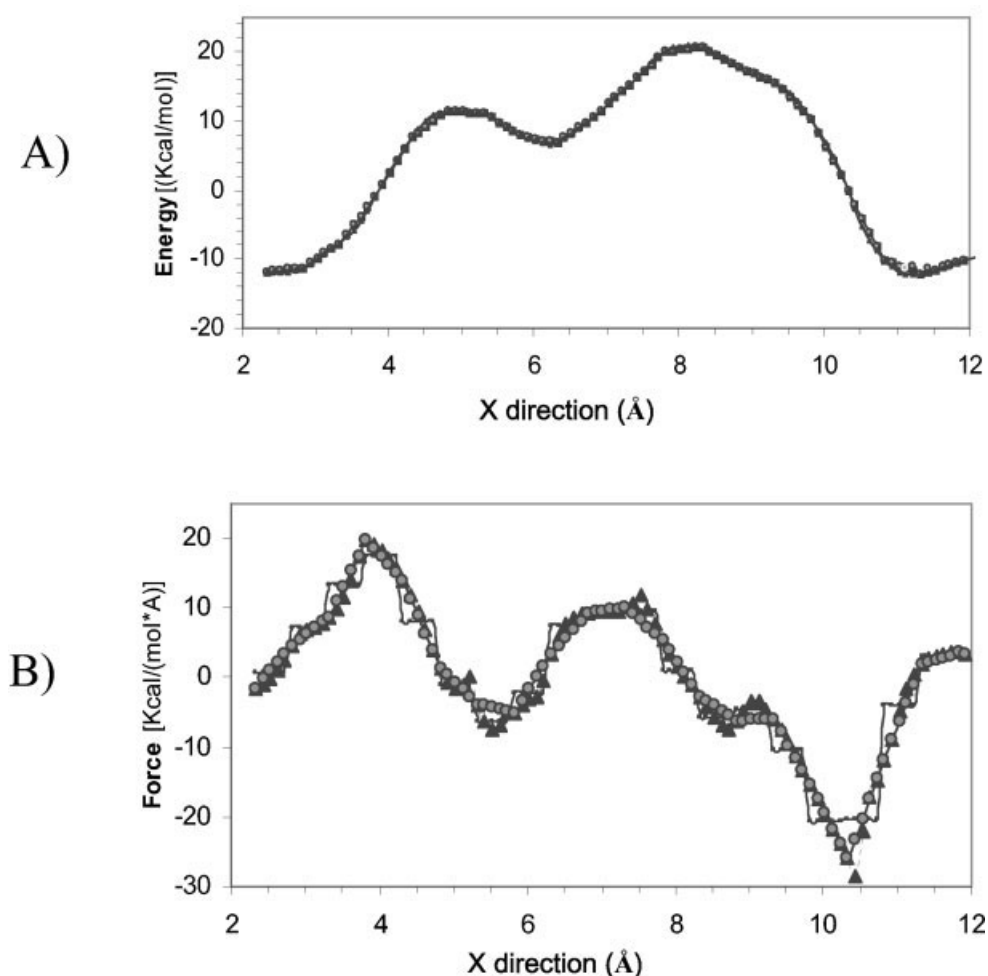
Figure 2. (A) Example of energies between a probe (nonbond parameters $\sigma = 1.55$ Å, $\varepsilon = 1.0$ kcal · mol^{-1}, and q = +1.0) and receptor (1abe) from GRID I (black cubes), GRID II (gray circles), and the full force field (black triangles). A soft-core potential (see details in article) is used and the coordinates for y, z of the probe are kept constant at the values of the active center. (B) Example of forces in the x direction between a probe and the receptor (1abe) from GRID I, GRID II, and the full force field.

Detailed Docking Results

The observed variability of results between the 10 independent docking trials for each ligand uncovers a potential sampling inadequacy within these single runs. Figure 5 demonstrates the distribution of docking frequency for each of these ligands across 10 independent runs. The overall distributions are similar for all three methods: 46% of the ligands (19/41) were docked successfully in all 10 runs, roughly 70% (29/41) were successfully docked in more than seven runs, and less than 12% (5/41) of the ligands never successfully docked. These observations are qualitatively correlated with ligand flexibility expressed as the number of rotatable bonds, and not with the size or potency of the ligands (data from Table 2). The docking success frequencies detailed on a per ligand-protein complex basis are tabulated in Table 2. In addition, this table includes the mean RMS of the top poses across 10 independent runs for all docking schemes. If the 41 ligands are divided into two groups based on rotatable bonds (8 being the

median), the first group (20 molecules with less than eight rotors) has significantly better docking success, with mean docking RMS of 1.0 Å and a mean docking success frequency of 0.91. The second group of 21 molecules has a mean docking RMS of 3.1 Å and a mean docking success frequency of 0.56. Interestingly, division of ligands into equal size groups using molecular weight or binding affinity (pK$_i$) criteria gives much less differentiation in docking accuracy. In particular, 21 ligands with $M_w < 314$ have a mean RMS of 0.74 Å, with the remaining 20 ligands showing mean RMS of 1.30 Å. Twenty-one less potent ligands with pK$_i <$ 9.580 have a mean RMSD of 1.09 Å, while 20 more potent ligands exhibit a mean RMS of 0.96 Å. However, as the ligands become even more complex there is no clear correlation between docking success frequency and number of rotors. Also, the differences between different docking schemes for these ligands (full force field, GRID I, and GRID II) can be substantial. These results demonstrate that docking algorithm accuracy depends most
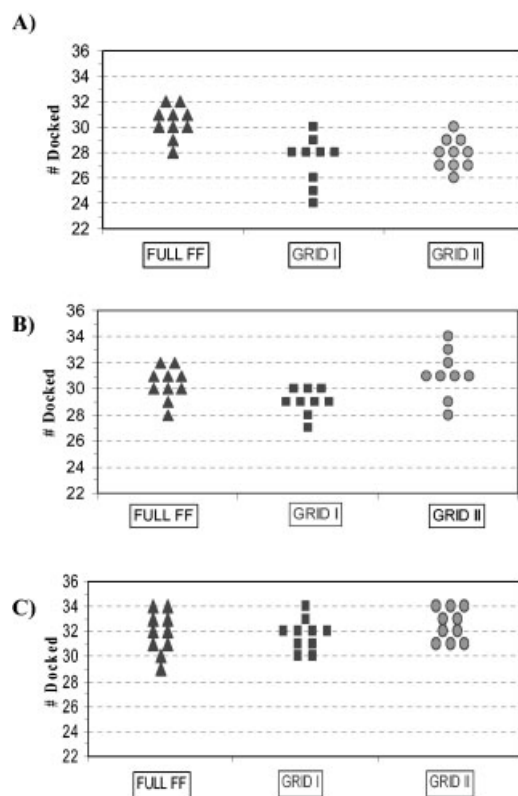
Figure 3. Comparison of docking success rates. Success rate is expressed by a number of successfully docked ligand structures in each of 10 independent docking trials. Full force field results are represented by black triangles, GRID I results by black squares, and GRID II by gray circles. Statistical analysis of results is reported for each case.

(A) Grid final minimization is performed for GRID I and GRID II. CORINA structure is used as the initial ligand geometry.

Docking potential	Mean value (%)	Standard deviation	p-Value (comparison to full FF)	p-Value (comparison to GRID II)
Full FF	30.4 (74%)	1.7	—	0.0003
GRID II	27.9 (68%)	1.2	0.0003	—
GRID I	27.1 (66%)	2.0	0.0003	0.2868

(B) Docking success rate for full forcefield final minimization for GRID I and GRID II.

Docking potential	Mean value (%)	Standard deviation	p-Value (comparison to full FF)	p-Value (comparison to GRID II)	p-Value[a] (comparison to 3A)
GRID II	31.3 (76%)	1.2	0.2168	—	0.0189
GRID I	28.9 (70%)	1.0	0.0086	0.0018	0.0001

[a]p-Value in this column refers to the comparison of GRID minimization results from 3A and full force field minimization.

(C) Success rates when ligand X-ray conformation is used to start the docking.

Docking potential	Mean value (%)	Standard deviation	p-Value (comparison to full FF)	p-Value (comparison to GRID II)	p-Value[b] (comparison to 3B)
Full FF	31.9 (78%)	1.7	—	0.3765	0.0357
GRID II	32.5 (79%)	1.3	0.3765	—	0.1055
GRID I	31.7 (77%)	1.3	0.1749	0.7647	0.0058

[b]p-Value in this column refers to the comparison of GRID minimization results from 3A and full force field minimization.
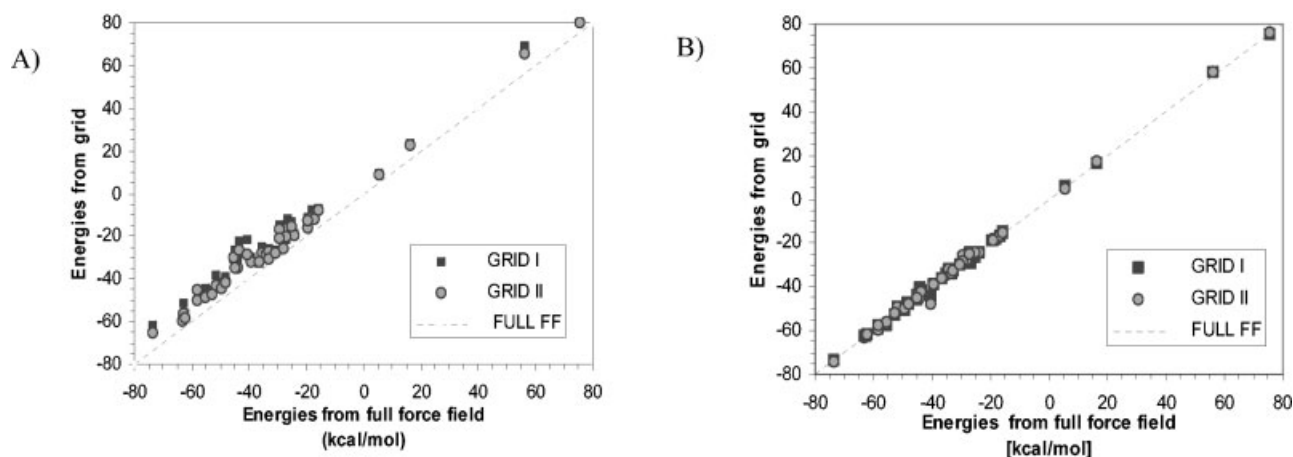
Figure 4. Comparison of grid-based docking energies averaged over 10 runs with the full force field for 41 complexes. (A) Mean grid minimized docking energies for GRID I (black squares) and GRID II (gray circles). (B) Mean full force field minimized docking energies averaged over 10 for GRID I and GRID II.

heavily on the ligand flexibility, thus, to make further improvements, better internal ligand sampling or representations of a ligand's internal force field are required.

Among five complexes that are never docked, two are thrombin ligands (1uvs, 1uvt), for which the protein flexibility and the variable water structures are essential for the protein-ligand interaction.[38] One is highly flexible HIV inhibitor GR126045 (1htf) that exhibits two binding modes, one is a highly charged and flexible APR molecule (2phh), and one small 1-deoxynoryjmicin (1dog) that binds in two neighboring sites. In these cases CDOCKER force field representation is inadequately representing the real system, as the energies of the misdocked structures were lower than those of the minimized X-ray geometries. The conclusion is supported by the last column in Table 2, which shows the energy difference between the lowest energy docked structure and the minimized X-ray structure. In all five cases docking gives lower energy structures.

For some ligands, the results from the GRID II method were better than those from the detailed full atomic force field. This may be attributed to small differences in the implementation of docking

protocols for the full force field and grid. There are four different types of soft-core potentials applied at different stages of full force field annealing, but only two types of soft-core grids were used for GRID I or GRID II (to minimize the number of grids to be stored and manipulated).

Based on these results, it can be concluded that the sampling quality in the current MD docking protocol is acceptable, and a single run (assuming 50 replicas for each complex) would on average give a 70% probability of the correct docking. There is still opportunity to improve the sampling efficiency and a more robust sampling scheme is currently under development. This variability in the docking results can be compared to the recent study by Paul and Rognan, which shows analysis of docking 100 protein-ligand complexes[39] with success rates of 39, 51, and 56% for DOCK,[8] FlexX,[40] and GOLD,[10] respectively. Despite the favorable initial conditions (X-ray geometry used), only 25% of the complexes were successfully docked with all three methods, indicating that significant improvement can be achieved by combining various methodologies. Even though the test sets, search spaces, and starting ligand conformations were different in that study, CDOCKER compares favorably with these established algorithms in terms of docking accuracy and variability of results.
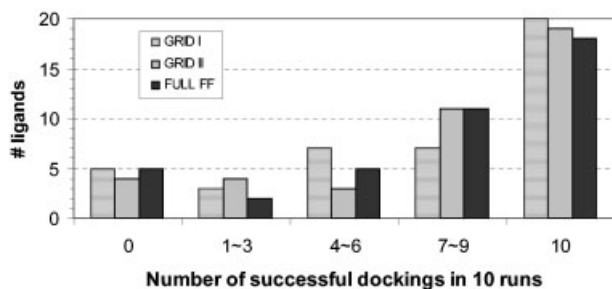
Timing

The mean computational time for docking using the full force field potential is 30 min using 50 replicas on a single node of a 733 MHz PIII CPU. The mean docking time is approximately three times faster for GRID I ($\approx$10 min), and it can be as much as six times faster for smaller or more rigid ligands (i.e., 2ypi). GRID II is on average 10% slower than GRID I due to the different method for computing the forces during dynamics. The relatively modest speed-up achieved by the grid introduction results from the substantial time requirements of internal ligand energetics that are treated explicitly with all CHARMm local energy terms,[23] and this timing cannot be improved without significantly impacting the

Figure 5. Distribution of docking success frequency in 10 independent runs. The bins on the x axis correspond to 0, 1–3, 4–6, 7–9, and all 10 successful attempts. Each successful attempt contains the best docking pose with RMSD less than 2 Å from X-ray position.
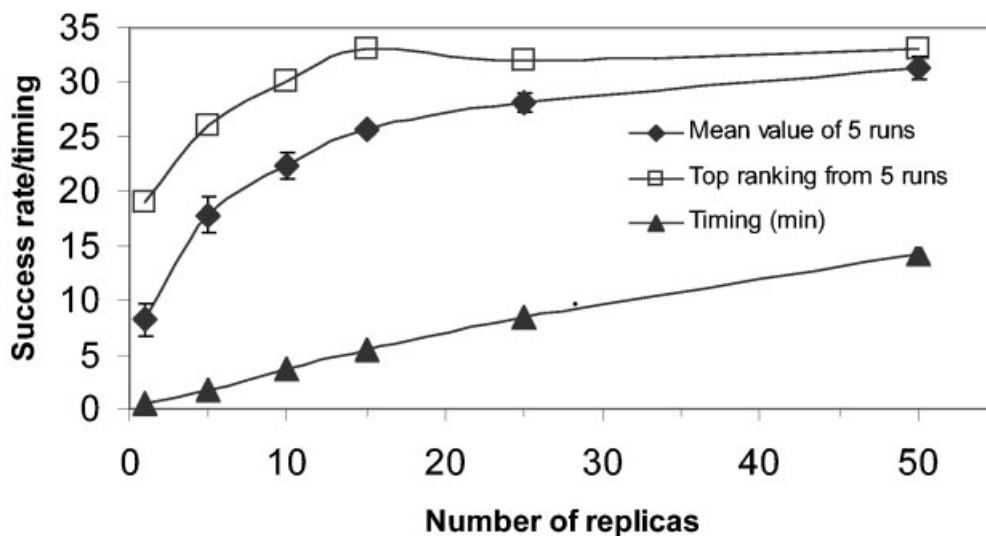
Figure 6. Mean docking success rates (number of docked complexes represented as black diamonds) and timing (in minutes represented as black triangles) for five runs with a different number of replicas for each ligand using the GRID II method with full force field minimization. Also plotted are the success rates for top ranking structures from all five runs (open squares).

accuracy by reducing the sampling rate for small or semirigid ligands as discussed below.

Docking Accuracy as a Function of Sampling

To further examine docking accuracy, the performance of the CDOCKER algorithm was studied as a function of the ligand-sampling rate. Figure 6 shows the docking success rates of the GRID II method when increasing the sampling regime by changing the number of replicas used in the docking runs. Various numbers of replicas including 1, 5, 10, 15, 25, and 50, have been investigated using five independent runs each. For example, using only 10 replicas gives a reasonable docking accuracy (success rate as $22/41 = 54\%$) with the mean docking time reduced to 4 min—a time scale that makes CDOCKER a viable application for screening modestly large numbers of compounds with the utilization of Linux clusters.

Figure 6 also shows the success rates for the energetically best pose observed in all five independent runs. The docking accuracy increases most rapidly for a very small number of replicas as five runs effectively increase the sampling rate by five times. The improvement of such complete sampling for 50 replicas (equivalent to a run of 250 replicas as we take the best docking pose of five independent runs of 50 replicas) is small (one structure), indicating that increasing the sampling level beyond 50 replicas will lead to little additional improvement in accuracy. As might be expected, the results from five runs of 10 replicas give consistent docking accuracies ($30/41 = 73\%$) to one single run of 50 replicas (76% with standard deviation of 3%).

Starting Structures

To investigate the dependence of the algorithm on initial conformation, docking experiments were performed using the internal complex X-ray geometry as a starting point instead of the CORINA geometry. All of the results discussed to this point have used the CORINA initial geometry. For many ligands, CORINA provides a good approximation of the X-ray ligand structures, while for others the difference between the CORINA structure and that from protein-ligand X-ray can be large, especially for ligands with a large number of rotors (see Table 2). On average the heavy atom RMSD between the X-ray and CORINA structures is 0.7 Å for ligands with less than eight rotatable bonds and 2.5 Å for other more flexible ligands.

Identical docking strategies were performed for all the ligands starting from the X-ray geometries (randomly orientated and translated around the active site). Figure 3C shows the success rates for GRID I, GRID II, and the full force field when the X-ray conformation is used as initial geometry. The mean successful dockings for all methods increases to 78%. This improvement is largest for GRID I, which is the method with the most room to improve. p values (Fig. 3C caption) for the docking success frequency means coming from CORINA and X-ray starting geometries indicate the improvement is significant for the full force field and GRID I schemes but not for GRID II.

Scoring Accuracy Dependence on Docking Accuracy

Up to this point factors influencing the docking accuracy have been examined. Because docking is frequently used in combination with scoring functions, the dependence of scoring these complexes based on docking accuracy was also examined. For the 41 test complexes the experimental pIC_{50} [log $(1/IC_{50})$] data were taken from the previous works.[5,11] Using leave-one-out cross validation and multiple regression, the Spearman rank correlation was computed between pIC_{50} and several variables, including PMF scoring,[41] ligand-protein vdW energies, molecular weight,
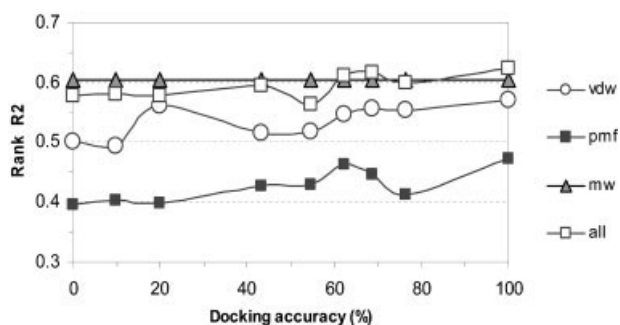
Figure 7. Spearman Rho (rank correlation) between experimental pIC$_{50}$ and van der Waals energies (vdW, open circles), PMF scores (pmf, black squares), molecular weight of ligands (mw, gray triangles), or all three sets combined (all, open squares) by leave-one-out cross validation. Electrostatic contribution is inversely correlated with binding affinities for this set. The van der Waals energies and PMF scores were calculated from the docked geometries based on GRID II. Different levels of docking accuracy correspond to docking results from 1, 5, 10, 15, 25, and 50 replicas, except that 100% data were based on minimized X-ray structures, 0 and 10% data were calculated from structures of several runs of the one replica case.

and the combination of all these three parameters. Figure 7 shows the leave-one-out rank correlation as a function of docking accuracy (note that the rank correlation increases by 15% if all data regression is used instead of leave-one-out cross validation). For this data set, the CDOCKER docking, in combination with the scoring functions examined, does not exhibit a correlation between docking and scoring accuracy. A similar conclusion was reached by Bissantz et al.[17] from several docking/scoring combinations used for virtual screening. Possible reasons may include the limitation of scoring functions or the size and characteristics of our test data set. Although a high correlation is observed when all three variables are applied, it is predominantly from the size effect—molecular weight alone gives a Rank R^2 of approximately 0.6. The lack of dependence of scoring ability on the docking accuracy is correlated with the fact that CDOCKER optimizes the energy of the system to a similar extent for all ligands in different sampling regimes. Thus, even poor sampling/docking accuracy leads to the energy terms that rank the ligands in a similar way to the good sampling/accuracy results.

Conclusions

In this article we present a thorough investigation of the performance of several CDOCKER MD docking approaches on a diverse test set of 41 protein-ligand complexes. Our analysis demonstrates that a combined strategy involving grid docking and full force field minimization is optimal for a docking protocol. The better grid representation of energy and forces in GRID II (as compared to those implemented in GRID I) leads to docking energies, structures (RMSD from X-ray), and a successful docking rate of 76% that are indistinguishable from the full force field MD simulations. The results from the individual ligand dockings de-

pend on the ligand flexibility more than on the ligand size, suggesting that further improvements may be gained by better sampling of the ligand's internal geometries and/or better representing the ligand's true force field. A detailed full force field final minimization of docked structures improves the docking accuracy by up to 10%, suggesting that the details of the energies and forces have more effect on the fine tuning of docking solutions and less influence on the docking process. Finally, for rigid or small ligands, less sampling can lead to a reduced computational time while retaining reasonable docking accuracy. When using a smaller number of replicas, the CDOCKER grid protocol can be usefully applied to obtain more accurate dockings in medium size virtual screening experiments.

References

1. Muegge, I.; Rarey, M. In Reviews in Computational Chemistry; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 2001; p 1.
2. Holloway, K. M.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M. D.; Vacca, J. P.; Dorsey, B. D.; Levin, R. B.; Wayne, J. T.; Chen, L. J.; deSolms, J. S.; Gaffin, N.; Ghosh, A. K.; Giuliani, E. A.; Graham, S. L.; Guare, J. P.; Hungate, R. W.; Lyle, T. A.; Sanders, W. M.; Tucker, T. J.; Wiggins, M.; Wiscount, C. M.; Woltersdorf, O. W.; Young, S. D.; Darke, P. L.; Zugay, J. A. J Med Chem 1995, 38, 305.
3. Vieth, M.; Cummins, D. J. J Med Chem 2000, 43, 3020.
4. Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks III, C. L. J Comput Chem 1998, 19, 1612.
5. Roche, O.; Kiyama, R.; Brooks, C. L. I. J Med Chem 2001, 44, 3592.
6. Oshiro, C. M.; Kuntz, I. D.; Knegtel, R. M. In The Encyclopedia of Computational Chemistry; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., et al., Eds.; John Wiley & Sons: Chichester, 1998; p 3056.
7. Broughton, H. B. J Mol Graphics Model 2000, 18, 247.
8. Kuntz, I., D.; Blaney, J., M.; Oatley, S., J.; Langridge, R.; Ferrin, T. E. J Mol Biol 1982, 161, 269.
9. Goodsell, D. S.; Olson, A. J Proteins 1990, 8, 195.
10. Jones, G.; Willet, P.; Glen, R. C. J Comput-Aided Mol Des 1995, 9, 532.
11. Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Proteins 1998, 33, 367.
12. Luty, B. A.; Wasserman, Z. R.; Stouten, P. F. W.; Hodge, C. N.; Zacharias, M. et al. J Comput Chem 1995, 16, 454.
13. Given, J. A.; Gilson, M. K. Proteins: Struct, Funct, Genet 1998, 33, 475.
14. Mangoni, M.; Roccatano, D.; Di Nola, A. Proteins: Struct, Funct, Genet 1999, 35, 153.
15. Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks III, C. L. J Comput Chem 1998, 19, 1623.
16. Wang, J.; Kollman, P. A.; Kuntz, I. Proteins 1999, 36, 1.
17. Bissantz, C.; Folkers, G.; Rognan, D. J Med Chem 2000, 43, 4759.

18. Halgren, T. A.; Murphy, R. B.; Banks, J.; Mainz, D.; Klicic, J.; Perry, J. K.; Friesner, R. A. Schrodinger: New York; Abstracts of Papers, 224th ACS National Meeting, Boston, MA, August 18–22, 2002; American Chemical Society: Washington, DC.

19. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; et al. J Comput Chem 1983, 4, 187.

20. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Wang, J.; et al. AMBER, 7th ed.; University of California: San Francisco, 2002.

21. Brooks III, C. L.; Karplus, M.; Petitt, B. M. Proteins: A theoretical perspective of dynamics structure and thermodynamics; John Wiley & Sons, Inc: New York, 1988; p 1.

22. Shea, J. E.; Brooks III, C. L. Ann Rev Phys Chem 2001, 52, 499.

23. Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. J Comput Chem 1992, 13, 505.

24. Oberlin Jr., D.; Scheraga, H. A. J Comput Chem 1998, 19, 71.

25. Verkhivker, G.; Bouzida, D.; Gehlhaar, D.; Rejto, P.; Arthurs, S.; et al. J Comput-Aided Mol Des 2000, 14, 731.

26. Miranker, A.; Karplus, M. Proteins 1991, 11, 29.

27. Schoichet, B. K.; Kuntz, I. D. J Mol Biol 1991, 221, 327.

28. Walls, P. H.; Sternberg, M. J. E. J Mol Biol 1992, 228, 277.

29. Tappura, K.; Lahtela-Kakkonen, M.; Teleman, O. J Comput Chem 2000, 21, 388.

30. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; et al. Nucleic Acids Res 2000, 28, 235.

31. Momany, F. A.; Rone, R. J Comput Chem 1992, 13, 888.

32. Insight, Release 2000; Accelrys Inc.: San Diego, CA, 2002.

33. Sadowski, J.; Gasteiger, J.; Klebe, G. J Chem Inf Comput Sci 1994, 34, 1000.

34. Biobyte Corporation and Pomona College Medchem Database. Daylight CIS, Inc., 2002.

35. Jain, A. K.; Dubes, R. C. Algorithms for clustering data; Prentice Hall: Englewood Cliffs, NJ, 1988.

36. Bulmer, M. G. Principles of Statistics; Dover Publications, Inc.: New York, NY, 1979.

37. Diller, D. J.; Verlinde, C. L. M. J. J Comput Chem 1999, 20, 1740.

38. Engh, R. A.; Brandstetter, H.; Sucher, G.; Eichinger, A.; Baumann, U.; et al. Structure 1996, 4, 1353.

39. Paul, N.; Rognan, D. Proteins 2002, 47, 521.

40. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. J Mol Biol 1996, 261, 470.

41. Muegge, I.; Martin, Y. C. J Med Chem 1999, 42, 791.