Investigation of MM-PBSA Rescoring of Docking Poses

David C. Thompson,[†,§] Christine Humblet,[‡] and Diane Joseph-McCarthy*[,†]

Wyeth Research Chemical and Screening Sciences 200 Cambridge Park Drive, Cambridge, Massachusetts 02140, and Wyeth Research Chemical and Screening Sciences 865 Ridge Road Princeton, New Jersey 08543

Received December 20, 2007

Target-based virtual screening is increasingly used to generate leads for targets for which high quality three-dimensional (3D) structures are available. To allow large molecular databases to be screened rapidly, a tiered scoring scheme is often employed whereby a simple scoring function is used as a fast filter of the entire database and a more rigorous and time-consuming scoring function is used to rescore the top hits to produce the final list of ranked compounds. Molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) approaches are currently thought to be quite effective at incorporating implicit solvation into the estimation of ligand binding free energies. In this paper, the ability of a high-throughput MM-PBSA rescoring function to discriminate between correct and incorrect docking poses is investigated in detail. Various initial scoring functions are used to generate docked poses for a subset of the CCDC/Astex test set and to dock one set of actives/inactives from the DUD data set. The effectiveness of each of these initial scoring functions is discussed. Overall, the ability of the MM-PBSA rescoring function to (i) regenerate the set of X-ray complexes when docking the bound conformation of the ligand, (ii) regenerate the X-ray complexes when docking conformationally expanded databases for each ligand which include "conformation decoys" of the ligand, and (iii) enrich known actives in a virtual screen for the mineralocorticoid receptor in the presence of "ligand decoys" is assessed. While a pharmacophore-based molecular docking approach, PhDock, is used to carry out the docking, the results are expected to be general to use with any docking method.

1. INTRODUCTION

The virtual screening of large databases to select smaller subsets for biological testing is one of the common tasks in drug discovery. For structure-based virtual screening in order to screen large numbers of molecules, fast docking and scoring methods are routinely used.[1,2] Structure-based virtual high-throughput screening has become a companion tool proven to add value to experimental high-throughput screening (HTS).

Practically, within a hierarchical virtual screening workflow, the affinity of a small molecule for a given target receptor is first assessed with a coarse scoring function; it is the crudity of the scoring function that permits millions of compounds to be rapidly assessed.[3,4] A smaller subset of compounds that pass this initial coarse filter is then subjected to a more accurate scoring scheme, which is expected to result in an enrichment of true positives. This general approach of the application of progressively more sophisticated scoring schemes to an increasingly smaller collection of ligands with a better fit to the target has been shown to be quite efficient. Thus, the later scoring functions usually attempt to include solvation and entropic effects. Comparisons of scoring schemes have shown a number of postdocking functions to be effective.[5–9] In the present work, we explore and assess the practical applicability of an MM-PBSA rescoring approach for its effectiveness in retrieving the best hits during late-stage docking.

In general, the MM-PBSA method, as developed by Kollman and Case,[10,11] combines molecular mechanical (MM) energies, a continuum solvent Poisson–Boltzmann (PB) model for polar solvation, and a solvent-accessible surface area dependent nonpolar solvation term (SA). In the initial implementation, there was also a term to account for solute entropy, which is estimated through either quasi-harmonic or normal-mode analysis, although this term is smaller than either the MM or PBSA contributions. Full molecular dynamics (MD) simulations were performed on the ligand, the receptor, and the complex, and average quantities were then extracted from the resulting MD trajectories. Recently, Pearlman compared the usefulness of the MM-PBSA approach to other computational methods such as thermodynamic integration (TI), one window free energy grid (OWFEG), ChemScore, pairwise linear potential (PLP), and the Dock Energy Score. In a test set of P38 MAP kinase inhibitors studied, the MM-PBSA results were inferior to those from TI, OWFEG, and even the DOCK Energy Score at ranking the congeneric series of ligands in terms of binding affinity.[12] However, a number of examples of applying MM-PBSA to predict protein–ligand binding show more promise.[13–17] Also, in a recent paper it was shown that applying the MM-PBSA energy function to a single, relaxed, complex structure is a valuable postdocking filter which enriches virtual screening results.[18] All of these MM-PBSA calculations involved molecular dynamics simulations which are computationally intensive. A related approach is the linear interaction energy (LIE)[19] method, which averages interac-

* To whom correspondence should be addressed at AstraZeneca, R&D Boston, 35 Gatehouse Dr., Waltham, MA 02451. E-mail: diane.joseph-mccarthy@astrazeneca.com.

† Wyeth Research Chemical and Screening Sciences, Cambridge, MA.

‡ Wyeth Research Chemical and Screening Sciences, Princeton, NJ.

§ Present address: Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road, Ridgefield, CT 06877.

10.1021/ci700470c CCC: $40.75
Published on Web 05/09/2008

tion energies across an MD simulation to estimate the absolute free energy. Huang and Caflisch have developed a relatively high throughput linear interaction energy (LIE) workflow, LIECE, which replaces MD sampling with a simple energy minimization of the ligand and combines LIE with a PB description of electrostatics.[20]

As an alternative to a numerical solution of the Poisson–Boltzmann equation, generalized Born charges (GB) have been used, resulting in the so-called MM-GBSA method, which has been shown to be less computationally demanding and nearly as accurate.[21] Lee and co-workers[22] investigated the use of DOCK4.0 with a GB approach to solvation. In addition, a rescoring MM-GBSA approach has recently been successfully applied to rank order known inhibitors of four kinase targets; using PRIME (Schrödinger, Inc., New York, NY, 2006), each ligand is minimized by itself and in the presence of the fixed protein with the GB solvation term included.[23] Here the energy of the ligand minimized by itself and the energy of the protein alone extracted out of the complex are both subtracted from the complex energy to yield the estimate of binding free energy.

In this study, the application of a very high throughput MM-PBSA rescoring function is presented. In this implementation, the ligand is simply minimized in the fixed protein structure with an MM function.[24] The resulting bound conformation of the ligand is also taken as the conformation of the ligand alone. Docked poses are generated using PhDock[25,26] with the set of standard scoring functions described below and then rescored using this MM-PBSA method. The accuracy of the MM-PBSA method with respect to regeneration of the known bound conformation of the ligand within the target-binding site is measured. In the succeeding sections, the ability of this rapid and simplified MM-PBSA scoring function to discriminate between correct and incorrect docking poses is further investigated and assessed.

2. METHODOLOGY

2.1. Protein–Ligand Test Sets. The CCDC/Astex[27] test set, which contains a diverse set of protein–ligand complexes, is often used for the testing of molecular-docking programs and scoring schemes. The smallest and highest resolution (<2.0 Å) of the test sets as described in the original Astex paper consists of 92 complexes. In the present study, complexes with covalent interactions between the ligand and the receptor, containing a heme group, and with ligands containing fewer than 4 pharmacophoric features were removed. The ascension codes for the remaining 68 complexes are included in the Supporting Information (Table S1) along with plots of the distribution of the 68 ligands versus the number of rotatable bonds and molecular weight (Figure S1A,B), illustrating that the present data set includes higher coverage for less flexible, low-MW ligands. Also, higher chemical diversity is achieved, as indicated by a mean pairwise similarity of 0.33 calculated using MACCS fingerprints and the Tanimoto coefficient.

For the enrichment test discussed below, the actives and inactives for the mineralocorticoid receptor (MR) target were taken from the DUD (Directory of Useful Decoys) data set.[28] Huang and co-workers, in preparing the DUD data set, used a similarity and physical property analysis to ensure that the inactives (decoys) have similar physical properties (e.g., molecular weight, number of hydrogen bond donors/acceptors, number of rotatable bonds, log *P*) to the actives. As such, this data set is expected to allow a bias-free assessment of virtual screening methods. For the MR target, there are 15 actives (12 of which docked into the binding site) and 535 inactives.

2.2. Complex Preparation. For each complex studied, water molecules were removed, all hydrogens were added, and His, Asn, and Gln side chains were reviewed and flipped for optimal geometries. Finally, the hydrogen positions were minimized with MacroModel (Schrödinger, Inc., New York, NY, 2006) using the MMFF94s force field[29] and default parameters. This curated set of ligand and protein structures was saved separately as mol2 files.

2.3. Ligand Database Preparation. To generate docking poses, we used the PhDock approach,[25,26] as implemented within DOCK4.0.[30] In a PhDock database, conformers of the same or different molecules are overlaid by their largest three-dimensional (3D) pharmacophore. During the docking, each database ensemble (a 3D pharmacophore and associated conformers) is simultaneously docked into the target-binding site. MCSS2SPTS[31] was used to determine the site points that are used for the matching during the docking. Theoretical pharmacophore points, representing hot spots in the binding site, allow for preferential orientation of the ligands in productive modes. Recognized database and site point pharmacophore types include hydrogen bond acceptors, hydrogen bond donors, duals (that can act as both acceptors and donors), and ring centroids.

X-ray Conformer Database. For each ligand, a 3D pharmacophore was generated by extracting the pharmacophore features from the bound conformation of the ligand in the X-ray complex structure. The overlaid pharmacophore and X-ray conformation were then translated out of the frame of reference of the X-ray complex structure so as not to bias the docking results.

Conformationally Expanded Databases. Conformationally expanded PhDock databases were generated for each test set ligand; the translated bound conformation of each ligand was also included in the database. Starting from this bound conformation of the ligand, a maximum of 1000 conformers were generated using OMEGA2.0 (OpenEye, Santa Fe, NM, 2006) with a 25 kcal/mol internal energy cutoff evaluated with the MMFF94s force field.[29] OMEGA2.0 performs an exhaustive depth first torsion search of freely rotatable bonds (excluding hydrogen rotors) and uses root-mean-squared deviation (rmsd) to account for symmetry and to remove duplicate poses. OMEGA2.0 has been shown to be very effective at reproducing the bioactive conformer for a broad range of ligands.[32]

2.4. Docking and Scoring Protocols. To orient each of the database pharmacophores into the target-binding site, distances between pairs of pharmacophore points in the database pharmacophore are matched to distances between pairs of pharmacophoric site points within the binding site. Partial pharmacophore matches are allowed. The fit of each individual conformer is scored, and in the end the *N* top-scoring poses per molecule are saved.

For the purposes of this study, docked poses were generated with PhDock using (i) the DOCK contact (CNT) score, (ii) the DOCK CNT score with minimization (CNT/

MIN), (iii) the DOCK energy (NRG), and (iv) the DOCK NRG score with minimization (NRG/MIN). The CNT score is a summation of the number of heavy-atom contacts arising between the ligand and the receptor within 4 Å (default parameter).[33] The NRG function consists of the nonbonded terms (the Coulombic term and the Lennard-Jones 6-12 van der Waals (vdW) terms) of the Amber molecular mechanics force field calculated on a grid.[34–36] For docking with minimization, for each docked pose 25 steps of simplex rigid-body minimization of the ligand in the fixed protein structure was performed. *N* docked poses are saved and subjected to the MM-PBSA rescoring scheme described below.

2.5. MM-PBSA Rescoring. An MM-PBSA rescoring approach is employed whereby the free energy of binding, ΔG, is approximated through

$$\Delta G_{\text{Bind}} \approx \Delta H^{\text{Sol}}_{\text{Elec}} + \Delta H_{\text{vdW}} - T\Delta S_{\text{HPhob}} - T\Delta S_{\text{RotB}}$$

For each complex, the ligand is subjected to up to 1000 steps of Cartesian coordinate minimization within the fixed protein structure using the Szybki minimizer (OpenEye, Santa Fe, NM, 2006) and the MMFF94s force field[29,37] with the default convergence criterion of 0.1 kcal/(mol Å) on the gradient vector norm. The default gradient tolerance was used instead of a more strict criterion to allow for the rapid minimization of large numbers of conformers and is expected to yield reasonable results, since the protein is held rigid. Each of the terms in ΔG_{Bind} are approximated by calculating the values for the complex, the protein alone, and the ligand alone, where the structures for the protein and ligand are simply taken from the minimized complex. $\Delta H^{\text{Sol}}_{\text{Elec}}$ represents the Poisson–Boltzmann electrostatics calculated using ZAP (OpenEye, Santa Fe, NM, 2006) with Bondi radii.[38] ΔH_{vdW} is the vdW energy. The hydrophobic term is

$$T\Delta S_{\text{Hphob}} = (\text{surface area buried upon complex formation}) \times 0.006 \text{ kcal/mol } \text{Å}^2$$

where the surface area term is computed with ZAP (OpenEye, Santa Fe, NM, 2006). The coefficient 0.006 kcal/mol Å^2 is taken from refs 39 and 40 and accounts for the partitioning of solute molecules between aqueous and organic phases. A rotatable bond penalty

$$T\Delta S_{\text{RotB}} = \text{number of rotatable bonds} \times 0.7 \text{ kcal/mol}$$

is also included to account for the loss of binding energy due to the freezing of the internal degrees of freedom of the ligand.[41]

This rescoring approach is implemented in a C-shell script (Figure 1) that spawns jobs over a parallel architecture and then gathers and sorts the results into a single multi-mol2 file with an associated scoring report. The processing of each docked molecule involves first removing any degenerate poses; uniqueness in the MM-PBSA script is defined such that two conformers are considered equivalent if their rmsd is within 0.5 Å. Next, up to three rotatable hydrogens (specifically the first three in the coordinate file) per ligand are sampled. This sampling generates a maximum of 27 conformers for each of the unique poses and was intended as a balance between effective sampling and reasonable computational time. For a representative docking run, a comparison of the results after no rotatable hydrogen sampling, sampling of up to three rotatable hydrogens, and full sampling is given in the Results and Discussion.

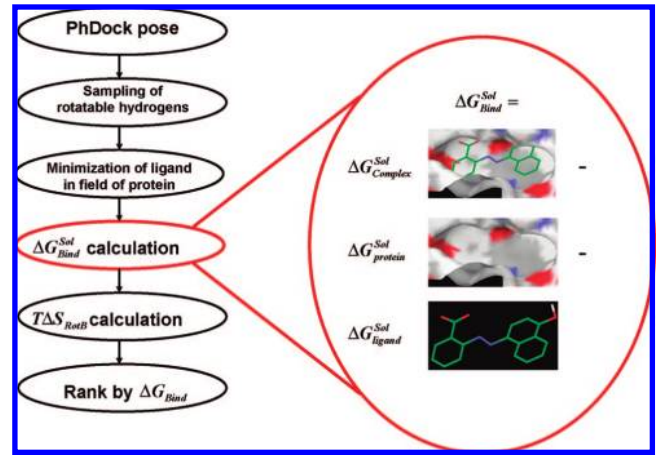

Figure 1. Schematic of the high-throughput MM-PBSA rescoring approach.

2.6. Analysis of Poses. For each saved pose, the heavy-atom rmsd from the position of the ligand in the X-ray complex structure was computed. Specifically, an OEChem utility for calculating the rmsd of a pose was employed to examine the top-scoring pose and the best rmsd pose both after docking and after rescoring. Thereby, sampling vs scoring issues can be distinguished.

3. RESULTS AND DISCUSSION

Each test set ligand was docked into its respective target binding site using PhDock and various scoring schemes, ultimately to test the ability of the MM-PBSA approach outlined above to discriminate between correct and incorrectly docked ligand poses. In Test One, the regeneration of each X-ray protein–ligand complex in the test set by redocking the ligand in its X-ray conformation is investigated as a function of scoring scheme. In Test Two, a conformationally expanded database for each ligand (seeded with the bound conformation of the ligand) is docked into its corresponding protein structure to further investigate the effect of scoring function on overall performance. Finally, an enrichment experiment is performed on the MR example taken from the DUD set[28] to begin to explore the effect of MM-PBSA rescoring on distinguishing between different ligands as well as different poses of the same ligand.

3.1. Test One: Regenerating X-ray Complexes. As a first test of the various scoring schemes employed, the X-ray bound conformation of each ligand was docked back into its corresponding protein structure. Initially, the CNT score was used for the docking and up to 10 poses were rescored with the MM-PBSA function to assess the ability of the approach to regenerate the "correct" pose (i.e., within 2 Å of the X-ray position). In Table 1, for each protein–ligand complex in the test set, the rmsd of the top-scoring pose and the best-scoring pose both after docking and after rescoring with MM-PBSA are given. Overall, 52/68 complexes (76%) have correctly docked poses after rescoring with the MM-PBSA function. For high-throughput docking of large databases, the greatest speedup with PhDock is obtained when using the CNT score and a 76% recovery of X-ray complexes after MM-PBSA rescoring is well within the current industry standard.

Of the 76% that are correctly docked, 38/68 complexes (56%) are correct both after the initial docking stage and

Table 1. Top- and Best-Scoring RMSDs for Each of the Ligands in the Test Set[a,b]

PDB complex code	CNT TOP	CNT BEST	CNT BEST POSE RANK	MM-PBSA TOP	MM-PBSA BEST	MM-PBSA BEST POSE RANK
1A28	1.11	0.51	9	0.31	0.23	3
1A4Q	0.74	0.19	6	0.28	0.21	2
1A6W	0.79	0.50	8	0.85	0.52	6
1ABE	2.51	2.32	7	2.57	1.30	25
1ABF	3.25	2.93	8	3.29	2.76	9
1AOE	0.88	0.86	2	0.36	0.36	1
1AQW	0.32	0.32	1	0.43	0.26	12
1ATL	1.17	0.52	4	0.71	0.50	4
1BMA	0.39	0.39	1	0.69	0.68	5
1BYB	1.01	0.89	2	0.56	0.46	2
1C5C	0.40	0.40	1	0.27	0.27	1
1C83	0.53	0.53	1	0.41	0.40	3
1CIL	3.70	3.19	3	2.74	1.98	3
1COY	2.98	1.46	6	0.93	0.92	4
1D3H	2.74	0.87	10	0.65	0.36	2
1EJN	0.87	0.64	2	0.30	0.30	1
1F3D	0.80	0.43	6	0.26	0.26	1
1FLR	1.27	0.23	7	0.62	0.30	4
1GLP	0.93	0.27	6	0.35	0.32	3
1GLQ	0.72	0.26	8	0.27	0.27	1
1HFC	0.82	0.35	8	4.84	0.28	12
1HPV	0.76	0.50	10	0.53	0.37	4
1HSL	0.55	0.35	6	0.21	0.21	1
1HVR	1.15	0.74	9	0.37	0.37	1
1HYT	6.03	0.57	3	0.74	0.64	3
1JAP	1.90	0.57	7	7.34	0.58	18
1KEL	0.68	0.46	4	0.94	0.89	4
1LCP	4.68	1.32	10	1.29	1.29	1
1LNA	6.39	1.47	5	1.02	1.02	1
1LST	0.43	0.15	3	0.17	0.15	5
1MLD	0.73	0.29	4	0.90	0.34	6
1MMQ	0.78	0.30	10	6.09	0.20	16
1MRG	3.00	0.83	6	0.63	0.63	1
1MRK	1.34	0.92	7	1.06	0.56	78
1MTS	3.82	2.01	9	1.86	1.84	2
1NCO	14.36	12.72	7	14.36	12.82	161
1PPC	1.60	1.60	1	1.57	1.57	1
1PPH	2.55	1.38	3	0.77	0.77	1
1QBR	0.46	0.46	1	0.37	0.30	32
1QBU	1.28	1.20	8	0.53	0.53	1
1RDS	1.08	1.08	1	0.89	0.85	6
1RNT	0.72	0.55	9	1.22	0.99	12
1ROB	6.59	5.12	4	8.43	5.30	60
1SLT	8.47	6.46	2	7.02	6.42	12
1SNC	1.17	0.76	7	0.69	0.69	1
1SRJ	7.68	0.87	3	7.66	0.69	5
1TMN	7.16	1.17	2	0.53	0.53	1
1TYL	1.87	1.58	2	1.38	0.86	6
1UKZ	3.80	2.40	2	1.02	1.02	1
1WAP	0.78	0.37	8	0.14	0.14	1
1XID	4.07	3.08	5	4.39	2.90	63
1XIE	4.23	0.67	2	5.28	0.57	264
2AK3	6.16	2.78	2	3.42	3.14	27
2CMD	3.06	0.63	5	0.25	0.25	2
2CTC	1.21	0.74	2	0.20	0.20	1
2FOX	0.37	0.20	5	0.20	0.17	18
2GBP	1.75	1.75	1	0.19	0.19	1
2H4N	3.95	3.95	1	5.33	3.59	6
2QWK	0.83	0.83	1	0.37	0.33	2
2TMN	0.73	0.61	3	0.83	0.74	2
2TSC	1.37	1.01	6	0.91	0.90	2
3CLA	5.75	2.52	6	5.60	2.33	33
3ERT	5.29	0.32	3	0.45	0.36	6
3TPI	1.05	0.64	4	0.26	0.26	1
4DFR	8.20	7.15	7	7.78	7.30	2
5ABP	3.01	0.93	9	0.20	0.20	1
6RNT	6.26	0.98	2	0.93	0.73	3
7TIM	1.14	1.14	1	0.87	0.87	3

[a] The CNT scoring function is used, with up to 10 saved poses passed on for MM-PBSA rescoring. [b] The results are color coded: (green) top CNT and MM-PBSA pose rmsd <2 Å; (light green) best CNT and top MM-PBSA pose rmsd <2 Å, top CNT pose >2 Å; (blue) best CNT pose rmsd >2 Å and top MM-PBSA pose <2 Å; (pink) best CNT pose rmsd <2 Å and top MM-PBSA pose rmsd >2 Å, top CNT pose >2 Å; (red) top CNT pose rmsd <2 Å and top MM-PBSA pose rmsd >2 Å; (yellow) best CNT pose rmsd >2 Å and top MM-PBSA pose rmsd >2 Å.

after the MM-PBSA rescoring (dark green rows in Table 1). See Figure 2 for an illustrative example. Of the remaining 14/68 complexes (21%), the MM-PBSA rescoring improves the results. Of those 14 ligands with a top-scoring CNT pose >2 Å, 12 have a "best CNT pose" (out of the 10 or less CNT poses that are passed on for rescoring) with rmsd < 2 Å from the X-ray position and are subsequently recovered by the MM-PBSA rescoring to yield a correct pose (light green rows in Table 1). For the remaining two of the 14, even the "best CNT pose" has an rmsd somewhat greater

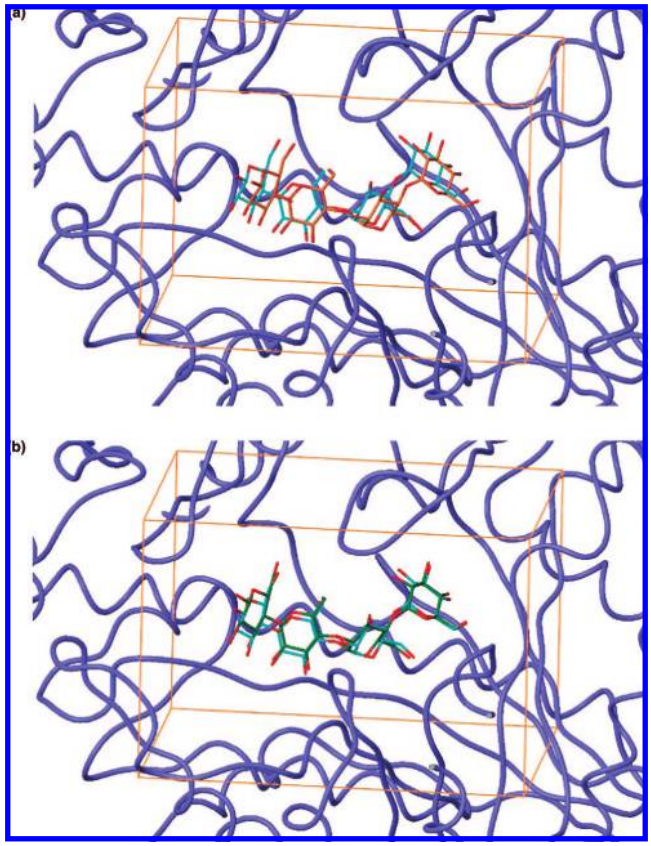

Figure 2. Superposition of the 1BYB X-ray structure (ligand carbons shown in cyan) and (a) the top-scoring CNT pose (ligand carbons in brown) and (b) the top scoring pose after MM-PBSA rescoring (ligand carbons in green). Only protein atoms were superposed. The rmsd of the top-scoring pose is reduced from 1.01 to 0.56 Å following rescoring.

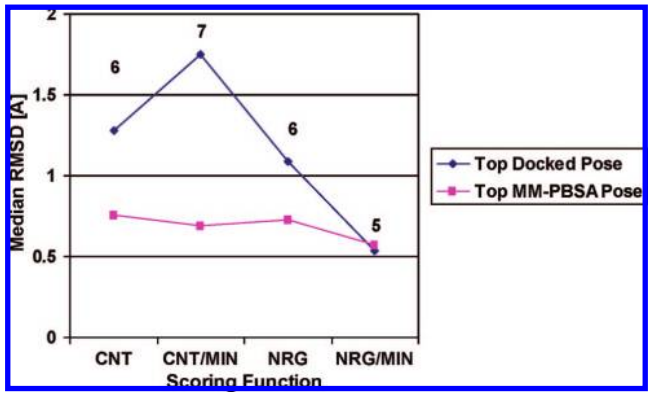

Figure 3. Median rmsd of the top-scoring poses across the test set as a function of scoring scheme for Test One. The X-ray conformer was docked, and up to 10 poses were passed on for MM-PBSA rescoring. The actual number of unique poses passed on for rescoring is shown above each data point on the plot.

than 2 Å (2.01 and 2.40 Å, respectively; corresponding to the blue rows in Table 1) and both are still recovered by the MM-PBSA rescoring. In many cases, the minimization during the MM-PBSA rescoring is able to shift the ligands closer to their true minimum energy position within the protein structure.

For 3/68 complexes (4%), however, the top CNT pose is correct but the MM-PBSA rescoring selects a ligand pose outside the 2 Å rmsd limit (red rows in Table 1). In all three cases, the top-ranked MM-PBSA pose is very poor (rmsd >4.8 Å), even though two or more correct poses for each ligand are passed on for rescoring. For these three complexes (1HFC, 1JAP, and 1MMQ) the MM-PBSA step selects a decoy pose instead of the correct one. These three complexes consist of matrix metalloproteases with bound hydroxamic acid ligands chelating a zinc ion. Similarly, 2/68 complexes (3%) with an incorrect top-scoring CNT pose have a best CNT pose <2 Å but are not recovered by the MM-PBSA rescoring step (pink rows in Table 1). 1SRJ (streptavidin in complex with naphthyl-HABA) has a fairly solvent exposed binding cavity, and for the best pose (with an rmsd <1 Å) the vdW term is less favorable than for the top pose (rmsd >7 Å). Isothermal titration calorimetry experiments on the 1SRJ complex suggest that favorable entropy contributions due both to water displacement from the binding site and retention of ligand flexibility in the bound state contribute to the binding affinity.[42] In 1XIE (D-xylose isomerase) the ligand interacts with a bound Mn^{2+} ion.

For the three matrix metalloproteases for which the MM-PBSA rescoring selects an incorrect pose (1HFC, 1JAP, and 1MMQ), rescoring with an analytical PB term included during the ligand minimization (option 4 with the Szybki minimizer) has a small effect on the final position of the ligand in the binding site (e.g., the rmsd of the top-scoring pose for the 1HFC ligand differs by only 0.3 Å from the best pose following molecular mechanics minimization) but does select a correct pose. This result suggests that the value of the final PB calculation can be quite sensitive to the precise position of the ligand in the binding pocket, especially when the ligand chelates an ion. Rescoring with a grid-based PB term included during the ligand minimization (option 1 with the Szybki minimizer) selects a correct pose in two of the three cases (1JAP and 1MMQ). Rescoring with the analytical PB term included during the ligand minimization increases the time for the calculation of the PB term by 23- to 70-fold, making it less practical for rescoring large sets of molecules. Furthermore, the overall results for the test set were not improved by including the analytical PB term during the ligand minimization.

For the test set overall, 16% (11/68) of the ligands are incorrectly docked using the CNT scoring function (best CNT pose >2.3 Å) and remain so after the MM-PBSA rescoring (yellow rows in Table 1). For 7 of the 11, the CNT score is not sufficient to generate a "largely correct" pose with rmsd <3 Å; the best CNT pose rmsd in those cases ranges from 3.1 to 12.7 Å. The worst results were for the 1NCO complex (holo-neocarzinostatin + chromophore and apo-neocarzinostatin + methyl-pentanediol); for the docking, the chromophore which is quite large (48 heavy atoms and 10 rotatable bonds) was taken as the ligand. In two of these 11 cases (1ABE and 1CIL), the MM-PBSA rescoring does generate a pose with an rmsd <2 Å but does not select it as the top pose. The L-arabinose ligand in the 1ABE complex is small and highly symmetric; furthermore both α- and β-L-arabinose are observed in the complex structure with equal occupancies while only the α-L-arabinose was docked. While the top-scoring MM-PBSA pose for 1ABE has a poor rmsd (2.57 Å), it makes most (4/5) of the key pharmacophoric interactions seen in the X-ray pose (Figure 4). In 1CIL, a complex of carbonic anhydrase II with the ETS inhibitor, the binding interaction occurs through a zinc ion and a

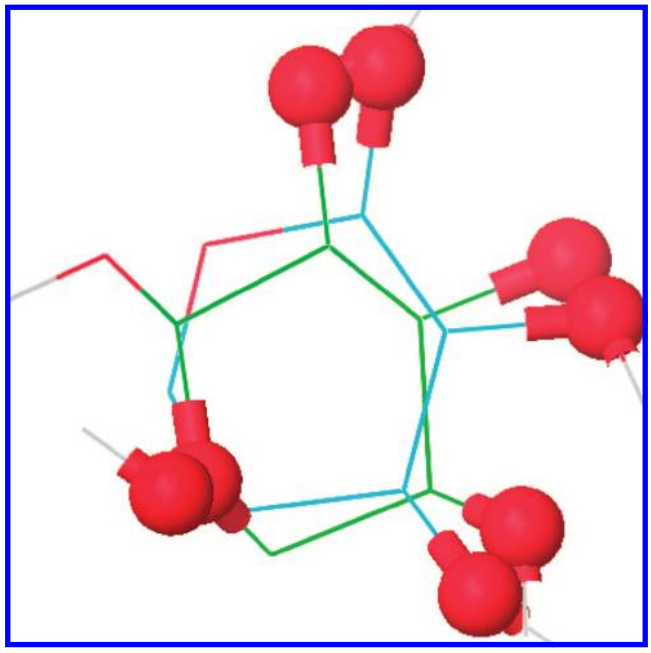

Figure 4. Superposition of the 1ABE X-ray structure and the top-scoring pose following CNT scoring and MM-PBSA rescoring. The ligand, α-L-arabinose, is colored by element. Only protein atoms (not shown) were superimposed. Although the rmsd of the top pose is large (2.57 Å), the majority of interactions (4/5) important for binding are preserved.

Table 2. Effect of Initial Scoring Function on Results following MM-PBSA Rescoring[a,b]

	CNT	CNT/MIN	NRG	NRG/MIN
correct	76	79	72	78
green	56	57	59	72
light green	18	21	13	4
blue	3	1	0	1
pink	3	9	3	6
red	4	0	1	3
yellow	16	12	24	13

[a] Test One with up to 10 docking poses saved; percentages are reported for the 68-complex set. [b] Results are color coded: (correct) top MM-PBSA pose <2 Å (green + light green + blue); (green) top CNT and MM-PBSA pose rmsd <2 Å; (light green) best CNT and top MM-PBSA pose rmsd <2 Å, top CNT pose >2 Å; (blue) best CNT pose rmsd >2 Å and top MM-PBSA pose <2 Å; (pink) best CNT pose rmsd <2 Å and top MM-PBSA pose rmsd >2 Å, top CNT pose >2 Å; (red) top CNT pose rmsd <2 Å and top MM-PBSA pose rmsd >2 Å; (yellow) best CNT pose rmsd >2 Å and top MM-PBSA pose rmsd >2 Å.

sulfonamide functional group on the ligand. Visual inspection of the MM-PBSA poses reveals a grossly correct binding mode prediction, with the correct orientation of the key sulfonamide group toward the zinc ion; however, the main core of the small molecule is sufficiently displaced to account for the rmsd.

A similar analysis was carried out using CNT/MIN, NRG, and NRG/MIN scoring during the docking followed by MM-PBSA rescoring. Prior to rescoring, 54% of the ligands were correctly docked with CNT, 57% with CNT/MIN or NRG, and 71% with NRG/MIN. After MM-PBSA rescoring, 76% of the ligands were correctly docked with CNT (as described in detail above), 79% with CNT/MIN, 72% with NRG, and 78% with NRG/MIN (see Table 2). As expected, rigid-body minimization of the ligand during the docking improves the results when using either CNT (79 vs 76%) or NRG scoring (78 vs 72%). For those complexes with no correct poses (<2 Å rmsd) passed onto rescoring after the initial docking, the chances of the MM-PBSA recovering a satisfactory pose are small (i.e., there are only a few entries which are "blue" in Table 2). The improvement in regeneration of X-ray complexes due to the MM-PBSA rescoring, however, is small (7%) when using NRG/MIN for the docking.

Table 3. Effect of Scoring Function on Regenerating Protein−Ligand X-ray Complexes[a,b]

docking	CNT	CNT/MIN	NRG	NRG/MIN
average	2.62	2.56	3.03	1.99
median	1.28	1.75	1.09	0.54
MM-PBSA rescoring	CNT	CNT/MIN	NRG	NRG/MIN
average	1.89	1.52	2.17	1.97
median	0.76	0.69	0.73	0.57

[a] Test One with up to 10 docking poses saved. [b] The average and median rmsd in Å (over the test set) for the top scoring pose.

In terms of the accuracy of the final poses selected, the median rmsd values of the top scoring poses across the entire test set are plotted in Figure 3. MM-PBSA rescoring improves the median rmsd over the test set when used subsequent to docking with CNT, CNT/MIN, and NRG scoring. With NRG/MIN scoring, the average and median rmsd over the test set is not improved (Figure 2 and Table 3). Obviously when the X-ray conformer is docked back into its respective protein structure, rigid-body minimization of the ligand should have essentially the same effect as full minimization of the ligand.

This analysis therefore indicates that (i) simple "shape" (as measured by the CNT score) is sufficient for the sampling part of a docking protocol and that (ii) the solvation part of the MM-PBSA term may not be significantly improving the results when regenerating X-ray complex structures. To further test hypothesis (ii), the CNT poses were rescored using the MM-PBSA script with the PB and SA terms turned off. Each CNT pose was subject to full minimization in the field of the fixed protein the Sybki minimizer with the MMFF94s force field as before. The final score includes only the MM terms as described in the methods. Using this MM rescoring script, 79% of the ligands in the test set are correctly docked, compared to 76% after the MM-PBSA rescoring. Furthermore, the median rmsd over the test set is 0.66 Å after MM rescoring, compared to 0.76 Å after the MM-PBSA rescoring.

In addition, in Figure 5 ΔG_{Bind} is plotted against $\Delta H^{\mathrm{Sol}}_{\mathrm{Elec}}$ for each complex in the test set using CNT scoring for the docking. A linear regression analysis reveals a correlation coefficient (R^2) value of 0.96, indicating that the loss area term and rotatable bond penalty term are not contributing significantly to the overall rescore value. Furthermore, examination of $T\Delta S_{\mathrm{Hphob}}$ and $T\Delta S_{\mathrm{RotB}}$ reveals that for any given complex these two terms are similar in magnitude, while differing in sign. Indeed, the contribution of the electrostatic and vdW terms is approximately 90% of the total binding free energy in this implementation. While the loss area term is likely to be similar for many different poses of a given ligand to a given receptor and the rotatable bond penalty is of course constant, what is somewhat surprising

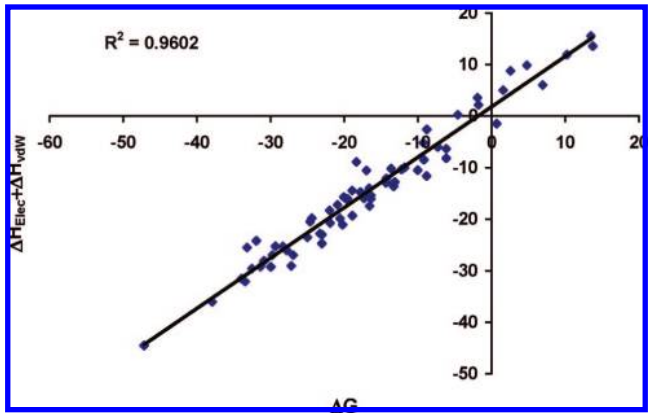

Figure 5. Plot of estimated ΔG_{Bind} versus $\Delta H_{\mathrm{Elec}}^{\mathrm{Sol}} + \Delta H_{\mathrm{vdW}}$ for each test set ligand.

Table 4. Effect of the RMSD of the Pose Passed onto MM-PBSA Rescoring[a,b]

rmsd of best pose passed on	no. of complexes	median rmsd of top rescored pose
$0 \le x < 0.5$	19	0.37
$0.5 \le x < 1.0$	24	0.64
$1.0 \le x < 1.5$	9	0.89
$1.5 \le x < 2.0$	3	1.38
$2.0 \le x < 2.5$	3	1.86
$2.5 \le x < 3.0$	3	3.42
≥ 3.0	7	7.02

[a] Test One with up to 10 docking poses saved. [b] CNT scoring function used for docking.

here is that ΔG_{Bind} is well correlated to $\Delta H_{\mathrm{Elec}}^{\mathrm{Sol}} + \Delta H_{\mathrm{vdW}}$ across the entire test set.

Finally, the effect of the quality of poses passed on for rescoring on the final accuracy (as measured by rmsd from the X-ray position) after MM-PBSA rescoring was examined (Table 4). Using the CNT score for the docking, if a pose with rmsd <2.5 Å is passed on for MM-PBSA rescoring, in general over the entire test set the rescoring improves the results and yields a median rmsd of 1.86 Å. For poses with an initial rmsd above 2.5 Å, this effect is lost. If a docked pose is too far from the X-ray position, the rescoring with minimization often forces the ligand further away from the correct binding position. This effect is most pronounced for docked complexes with an initial rmsd greater than 3 Å.

3.2. Test Two: Conformationally Expanded Ligand Database. As a more stringent test, the effect of MM-PBSA rescoring in the presence of conformational decoys has been investigated. A conformationally expanded PhDock database was created for each test set ligand, as described in the Methodology section. Since the objective of this study is to evaluate different scoring schemes and not conformer generation methods, the X-ray conformer was also included in each database.

As for Test One, initially the effect of the scoring scheme used during docking on the median rmsd for the test was investigated (Figure 6 and Table 5). With the conformational decoys added, the overall median rmsd of the top scoring poses is generally poorer than with only the X-ray conformer, as anticipated. It is not surprising that NRG without minimization would perform poorly, since the vdW term is

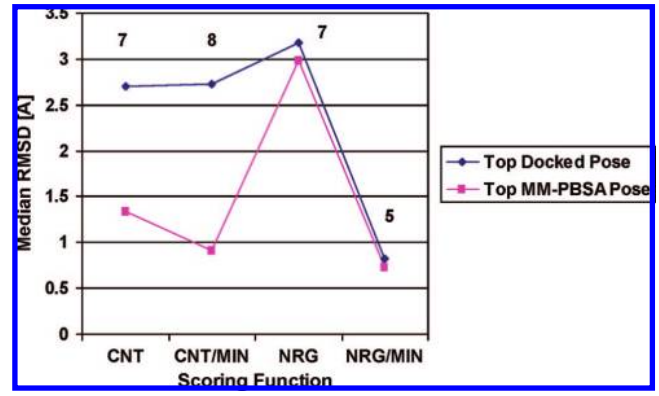

Figure 6. Median rmsd of the top-scoring poses across the test set as a function of scoring scheme for Test Two. The conformationally expanded ligand databases were docked, and up to 10 poses were passed on for MM-PBSA rescoring. The actual number of unique poses passed on for rescoring is shown above each data point on the plot.

Table 5. Effect of Scoring Function on Regenerating the X-ray Pose When Docking the Conformationally Expanded Database[a,b]

docking	CNT	CNT/MIN	NRG	NRG/MIN
average	3.35	3.25	4.04	2.23
median	2.70	2.73	3.18	0.82
MM-PBSA rescoring	**CNT**	**CNT/MIN**	**NRG**	**NRG/MIN**
average	2.44	2.01	3.88	2.35
median	1.34	0.91	2.99	0.73

[a] Test Two with up to 10 docked poses saved. [b] The average and median rmsd in Å (over the whole set) for the top-scoring pose.

very sensitive to the exact position of the ligand in the binding pocket. With NRG scoring during the docking, the median rmsd over the test set is 3.18 Å before rescoring; after MM-PBSA rescoring it is 2.99 Å. As was shown in section 3.1, poses that are too far from the correct position cannot be rescued by the MM-PBSA rescoring. For the other three scoring schemes, the MM-PBSA rescoring improves the results. After MM-PBSA rescoring for Test Two, 59% of the ligands were correctly docked with CNT, 68% with CNT/MIN, 41% with NRG, and 65% with NRG/MIN (see Table 6). For the CNT scoring followed by MM-PBSA, the average rmsd was also examined in terms of the number of rotatable bonds in the ligands; for the ligands with <7 rotatable bonds (44 complexes) the average rmsd was 1.72 Å, while for ligands with ≥7 rotatable bonds (24 complexes), it was 3.78 Å (Table 7). As expected, more flexible ligands are more difficult to dock accurately. As such, for production runs with PhDock, we are typically screen our lead-like databases containing only ligands with <7 rotatable bonds.

A timing analysis of the CNT and MM-PBSA workflow as compared to the NRG/MIN workflow reveals the following: CNT followed by MM-PBSA rescoring takes 4.42 h to process all 68 complexes; this is as compared to 1.69 h for the initial NRG/MIN docking. The majority of time (~60%) for the CNT/MM-PBSA calculation is spent performing the flexible minimization with Szybki, while the rigid-body minimization during the docking with NRG/MIN is very rapid, as only one minimization cycle of 25 steps is performed.

Table 6. Effect of Initial Scoring Function on Results following MM-PBSA Rescoring[a,b]

	CNT	CNT/MIN	NRG	NRG/MIN
correct	59	68	41	65
green	32	44	31	56
light green	22	18	9	9
blue	4	6	1	0
pink	1	6	3	1
red	3	1	6	7
yellow	37	25	50	26

[a] Test Two with up to 10 docking poses saved; percentages are reported for the 68-complex set. [b] Results are color coded: (correct) top MM-PBSA pose <2 Å (green + light green + blue); (green) top CNT and MM-PBSA pose rmsd <2 Å; (light green) best CNT and top MM-PBSA pose rmsd <2 Å, top CNT pose >2 Å; (blue) best CNT pose rmsd >2 Å and top MM-PBSA pose <2 Å; (pink) best CNT pose rmsd <2 Å and top MM-PBSA pose rmsd >2 Å, top CNT pose >2 Å; (red) top CNT pose rmsd <2 Å and top MM-PBSA pose rmsd >2 Å; (yellow) best CNT pose rmsd >2 Å and top MM-PBSA pose rmsd >2 Å.

Table 7. Effect of Rotatable Bonds on the Average RMSD of the Top- and Best-Scoring Poses[a]

no. of rotatable bonds[b]	CNT[c] top	CNT[c] best
<7	2.84	1.56
≥7	4.28	2.87
no. of rotatable bonds[b]	**MM-PBSA top**	**MM-PBSA best**
<7	1.72	1.29
≥7	3.78	2.65

[a] Test Two with up to 10 docked poses saved. [b] Number of rotatable bonds per ligand calculated within the Szybki minimizer. [c] CNT scoring function used.

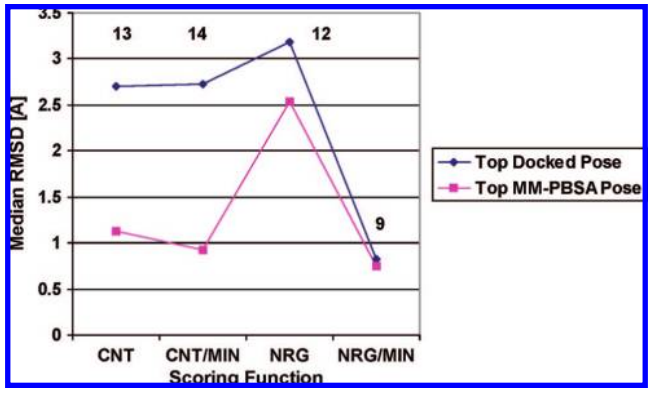

Figure 7. Median rmsd of the top-scoring poses across the test set as a function of scoring scheme for Test Two. The conformationally expanded ligand databases were docked, and up to 20 poses were passed on for MM-PBSA rescoring. The actual number of unique poses passed on for rescoring is shown above each data point on the plot.

To probe if increasing the number of saved docked poses that are passed onto the rescoring step would improve the results, Test Two was repeated, saving up to 20 docked poses instead of 10. The median rmsd for the top-scoring poses after docking and after rescoring is shown in Figure 7 and should be compared to Figure 6. Increasing the number of saved poses from 10 to 20 has little effect on the median rmsds over the test set, and indeed the general trend with respect to the scoring scheme used is preserved. With CNT scoring followed by MM-PBSA rescoring, there is a small improvement in the median rmsd on increasing the number of poses, but this is offset by an increase in computational time. In addition with CNT followed by rescoring, 59% of the ligands are correctly docked when up to 10 poses are rescored, compared to 56% when up to 20 poses are rescored.

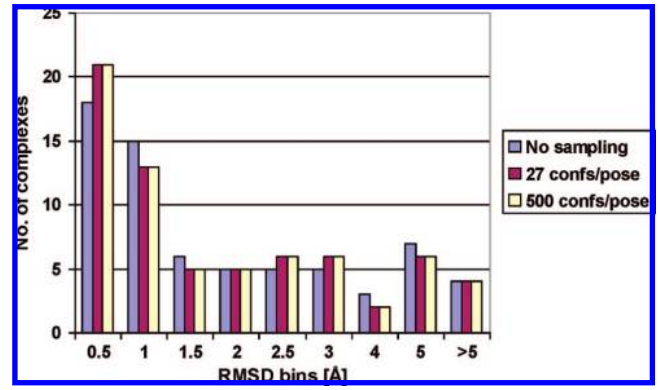

Figure 8. Histogram of the number of complexes in the test set with the top-scoring docking pose within each rmsd bin following CNT scoring and MM-PBSA rescoring for Test Two with different degrees of sampling of rotatable hydrogens during the rescore step. Blue bars represent no sampling of rotatable hydrogens, red bars represent sampling of up to 3 rotatable hydrogens per ligand, and yellow bars represent sampling of up to 56 rotatable hydrogens per ligand.

To investigate the effect of sampling the rotatable hydrogen positions on the results of the MM-PBSA rescoring, a histogram showing the number of complexes in the test set with top scoring poses within each rmsd bin plotted after no sampling of rotatable hydrogen positions, generation of up to 27 conformers per unique docked pose (the default used with the MM-PBSA script for all calculations discussed throughout this paper), and generation of up to 500 conformers per unique docked pose (Figure 8). Generating and rescoring up to 27 structures per unique docked pose increases the time for the rescore by a factor of 3 (from 0.98 h for the test set overall without sampling the rotatable hydrogen positions to 3.45 h with the limited sampling, with the mean time per complex changing from 52 to 182 s). Thus, the sampling of the rotatable hydrogen positions is a costly step and has little effect on the accuracy of the overall results.

To further examine the effect of the PBSA term on the rescoring, Test Two (with conformational expansion of the ligands) was carried out with the CNT score followed by MM rescoring (see Table 8). With or without the PBSA term included (Table 6 vs Table 8), 59% of the complexes are correctly regenerated. As with Test One, over a diverse set of targets, the PBSA term does not appear to improve the results when distinguishing between different poses of the same ligand in a given target-binding site. For generating binding modes for use in lead optimization, the MM terms with full minimization of the ligand in the field of the protein are probably sufficient. It is still possible that the PBSA term may help to rank order the binding of different ligands to a given target binding site and may therefore improve the quality of virtual screening results.[43,44] If this is true, one would expect to see a greater enrichment of known actives for a given target with the PBSA term included. Also, the PBSA term may aid in eliminating false positives in a large virtual screen.

As an initial look into this question, PhDock was used with CNT scoring followed by MM-PBSA and MM rescor-

Table 8. Effect of MM Rescoring on CNT Docking Results[a,b]

	CNT
correct	**59**
green	35
light green	21
blue	3
pink	3
red	0
yellow	38

[a] Test Two with up to 10 docking poses saved; percentages are reported for the 68-complex set. [b] Results are color coded (correct) top MM-PBSA pose <2 Å (green + light green + blue); (green) top CNT and MM-PBSA pose rmsd <2 Å; (light green) best CNT and top MM-PBSA pose rmsd <2 Å, top CNT pose >2 Å; (blue) best CNT pose rmsd >2 Å and top MM-PBSA pose <2 Å; (pink) best CNT pose rmsd <2 Å and top MM-PBSA pose rmsd >2 Å, top CNT pose >2 Å; (red) top CNT pose rmsd <2 Å and top MM-PBSA pose rmsd >2 Å; (yellow) best CNT pose rmsd >2 Å and top MM-PBSA pose rmsd >2 Å.

ing, respectively, to dock the actives and inactives for the mineralocorticoid receptor (MR) target taken from the DUD data set.[28] A receiver operating characteristic (ROC) curve is shown in Figure 9, and overall the MM-PBSA and MM curves are similar, with respective AUC measures of 0.97 and 0.98. The MM-PBSA exhibits early enrichment, and with 1% of the total database screened, 33% of the actives (4/12) have been identified by the MM-PBSA rescoring, whereas 25% (3/12) are identified with MM rescoring; with 3% screened, both curves show 67% of the actives as found, while at 10% the curves have crossed and 75% (9/12) are identified by the MM-PBSA compared to 100% (12/12) with the MM. It is still possible, therefore, that at the very top of a virtual screening hit list, which is of course the most relevant part, the PBSA term may play a small role in eliminating more false positives. Future work will involve a more systematic investigation of this possibility.

4. SUMMARY

In this paper, the ability of a high-throughput MM-PBSA rescoring function to discriminate further between correct and incorrect docking poses is investigated. First, the ability of the rescoring scheme to distinguish between different docked poses of the X-ray conformation of the bound ligand in its corresponding target binding site is assessed for a subset of the CCDC/Astex test set. While our in-house docking approach, PhDock, is used with various initial scoring schemes to generate the poses that are passed on for MM-PBSA rescoring, the results should be general to its use with any docking approach. In terms of sampling during docking, a simple shape-based scoring function (i.e., the CNT score) appears adequate. Next, the ability of the MM-PBSA rescoring to regenerate the X-ray complexes when docking conformationally expanded databases for each ligand that include "conformation decoys" of the ligand is studied. In general, for regenerating the X-ray complexes, the MM-PBSA rescoring did not improve the results over MM rescoring with full minimization of the ligand in the fixed protein structure. The solvation and surface area terms did not further improve the discrimination between different poses of the same ligand. Finally, the ability of the MM-PBSA rescoring function to enrich known actives in a virtual screen for the MR target in the presence of "ligand decoys" is assessed; the DUD data set of actives and inactives for MR is employed.[28] For virtual screening, the PBSA rescoring shows a minor improvement in terms of the number of actives in top 1% of the ranked database for this test case.

The time pressures in the modern pharmaceutical industry routinely require the rapid assessment of millions of compounds, and yet the intricacies of the interactions between a small molecule and a target receptor are subtle and often require a complex description; indeed, finding an acceptable balance between accuracy and speed of use is a challenge.

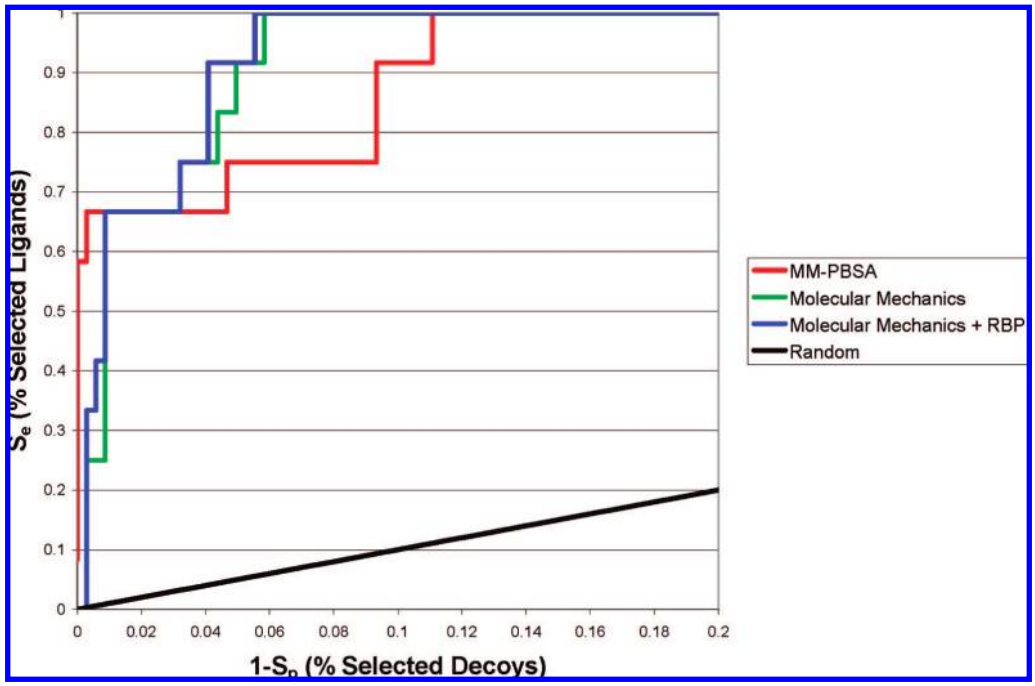

Figure 9. ROC plot for the MR target actives vs inactives taken from the DUD data set.[28] PhDock with CNT scoring was used to dock the conformationally expanded database followed by MM-PBSA rescoring (red curve), MM rescoring (green curve), and MM rescoring including the rotatable bonds penalty (blue curve), respectively. Random expectation is indicated by the black curve.

The use of a rapid scoring scheme such as NRG/MIN or CNT followed by MM-PBSA rescoring is a reasonable approach, although it seems likely that the true benefit of the rescoring comes from full minimization of the ligand in the field of the protein and possibly the inclusion of the solvent-accessible area loss term. Additional work evaluating other docking and scoring schemes is in progress.

ACKNOWLEDGMENT

We thank Jason Cross, Natasja Brooijmans, Brain Cole, Iain J. McFadyen, and Eric Feyfant for their scientific insights during the preparation of this paper. We also thank John C. Morris and Vassilios Pantazopoulos for assistance automating various scripts for parallel computing.

Supporting Information Available: Table S1, giving pdb ascension codes for the 68 complexes in the ASTEX test set used and Figures S1A and S1B, giving plots of the distribution of the ligands in the test set versus the number of rotatable bonds and molecular weight. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Bohm, H. J.; Stahl, M. The use of scoring functions in drug discovery applications. In *Reviews in Computational Chemistry*, Lipkowitz, K. B., Boyd, D. B., Eds. Wiley-VCH: Hoboken, NJ, 2002; Vol. 18, pp 41–87.
(2) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
(3) Joseph-McCarthy, D. Structure-based Lead Optimization. *Annu. Rep. Comput. Chem.* **2005**, *1*, 169–183.
(4) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discov. Today* **2002**, *7*, 1047–1055.
(5) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.
(6) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44*, 3417–3423.
(7) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
(8) Velec, H. F. G.; Gohlke, H.; Klebe, G. DrugScore^{CSD}-Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *J. Med. Chem.* **2005**, *48*, 6296–6303.
(9) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions 0.1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Design* **1997**, *11*, 425–445.
(10) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897.
(11) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
(12) Pearlman, D. A. Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase. *J. Med. Chem.* **2005**, *48*, 7796–7807.
(13) Kuhn, B.; Kollman, P. A. Binding of a Diverse Set of Ligands to Avidin and Strepavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **2000**, *43*, 3786–3791.
(14) Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. Molecular mechanics methods for predicting protein-ligand binding. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5166–5177.
(15) Huang, N.; Jacobson, M. P. Physics-based methods for studying protein-ligand interactions. *Curr. Opin. Drug Devel.* **2007**, *10*, 325–331.
(16) Sims, P. A.; Wong, C. F.; McCammon, J. A. A Computational Model of Binding Thermodynamics: The Deisgn of cyclin-dependent kinase 2 inhibitors. *J. Med. Chem.* **2003**, *46*, 3314–3325.
(17) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.
(18) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005**, *48*, 4040–4048.
(19) Aqvist, J.; Medina, C.; Samuelsson, J. E. New Method for Predicting Binding-Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385–391.
(20) Huang, D.; Caflisch, A. Efficient Evaluation of Binding Free Energy Using Continuum Electrostatics Solvation. *J. Med. Chem.* **2004**, *47*, 5791–5797.
(21) Lee, M. C.; Yang, R.; Duan, Y. Comparison between Generalized-Born and Poisson-Boltzmann methods in physics-based scoring functions for protein structure prediction. *J. Mol. Model* **2005**, *12*, 101–110.
(22) Lee, M. R.; Sun, Y. Improving Docking Accuracy through Molecular Mechanics Generalized Born Optimization and Scoring. *J. Chem. Theory Comput.* **2007**, *3*, 1106–1119.
(23) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring. *J. Med. Chem.* **2006**, *49*, 4805–4808.
(24) Rush, T., III; Manas, E. S.; Tawa, G. J.; Alvarez, J. C. Solvation-Based Scoring for High Throughput Docking. In *Virtual Screening in Drug Discovery*; Alvarez, J. C., Shoichet, B., Eds.; Marcel Dekker: New York, 2005; pp 249–278.
(25) Joseph-McCarthy, D.; McFadyen, I. J.; Zou, J.; Walker, G.; Alvarez, J. C. Pharmacophore-based molecular docking: A practical guide. In *Virtual Screening in Drug Discovery*, Alvarez, J. C., Shoichet, B., Eds.; Marcel Dekker: New York, 2005; pp 327–347.
(26) Joseph-McCarthy, D.; Thomas, B. E., IV; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins* **2003**, *51*, 172–188.
(27) Nissink, J. W. M.; Murray, C. W.; Hartshorn, M. J.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein-ligand interaction. *Proteins* **2002**, *49*, 457–471.
(28) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
(29) Halgren, T. A. MMFF VI. MMFF94s Option for Energy Minimization Studies. *J. Comput. Chem.* **1999**, *20*, 720–729.
(30) Ewing, T.; Makino, S.; Skillman, A.; Kuntz, I. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Design* **2001**, *15*, 411–428.
(31) Joseph-McCarthy, D.; Alvarez, J. C. Automated generation of MCSS-derived pharmacophoric DOCK site points for searching multiconformation databases. *Proteins* **2003**, *51*, 189–202.
(32) Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T. Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations. *J. Chem. Inf. Model.* **2006**, *46*, 1848–1861.
(33) Shoichet, B.; Bodian, D. L.; Kuntz, I. D. Molecular Docking Using Shape Descriptors. *J. Comput. Chem.* **1992**, *13*, 380–397.
(34) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2nd Generation Force-Field For the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(35) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking With Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.
(36) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Amber, a Package of Computer-Programs For Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
(37) Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17*, 616–641.
(38) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
(39) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* **1991**, *252*, 106–109.

(40) Sitkoff, D.; Ben-Tal, N.; Honig, B. Calculation of Alkane to Water Solvation Free Energies Using Continuum Solvent Models. *J. Phys. Chem.* **1996**, *100*, 2744–2752.
(41) Bohm, H. J. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein Ligand Complex of Known 3-Dimensional Structure. *J. Comput.-Aided Mol. Design* **1994**, *8*, 243–256.
(42) Weber, P. C.; Pantoliano, M. W.; Simons, D. M.; Salemme, F. R. Structure-Based Design of Synthetic Azobenzene Ligands for Streptavidin. *J. Am. Chem. Soc.* **1994**, *116*, 2717–2724.
(43) Bernacki, K.; Kalyanaraman, C.; Jacobson, M. P. Virtual ligand screening against Escherichia coli dihydrofolate reductase: Improving docking enrichment using physics-based methods. *J. Biomol. Screening* **2005**, *10*, 675–681.
(44) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-Based Scoring of Protein-Ligand Complexes: Enrichment of Known Inhibitors in Large-Scale Virtual Screening. *J. Chem. Inf. Model.* **2006**, *46*, 243–253.

CI700470C