Quantum Mechanics/Molecular Mechanics Strategies for Docking Pose Refinement: Distinguishing between Binders and Decoys in Cytochrome *c* Peroxidase

Steven K. Burger,*,† David C. Thompson,‡ and Paul W. Ayers†

Department of Chemistry, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, and Chemistry Systems and High Performance Computing, Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road, Ridgefield Connecticut 06877, United States

Received August 26, 2010

We investigate the effect of systematically applying molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) to docked poses in an attempt to improve the correspondence between theoretical prediction and experimental observation. The proposed scheme involves running a short time scale MD simulation on a docked ligand pose (and any known structurally important crystal structure waters in the active site), followed by QM/MM minimization. Both of these steps are relatively fast for moderately sized ligands; longer time scale MD involving the protein is not found to improve the results. The final binding energy is given in terms of the QM/MM total energy, a van der Waals correction, and a term to account for desolvation effects. This methodology is first tested with a trypsin inhibitor, for which we establish the importance of running MD before reoptimizing with QM/MM. The method is then applied to cytochrome *c* peroxidase using a set of binders and decoys. In this example, the proposed methodology affords much better discrimination between binders and decoys than the traditional docking approach used. For both systems presented, application of this protocol results in a significantly better energetic ranking and a smaller root mean squared deviation from known crystallographic ligand poses. This work highlights the importance of including polarization effects through QM/MM and of sampling with MD to refine a set of initial docked poses.

INTRODUCTION

Computational methods play an important role in the process of drug discovery.[1] Starting with the structure of a protein, these methods can help filter the number of possible compounds that need to be tested from large compound libraries. With molecular docking, compounds are algorithmically positioned in the unoccupied active site of a target protein. There are two components to a canonical docking algorithm: search and score. In the searching stage, the conformational space of the ligand is explored (either in the presence or in the absence of the protein active site). In the subsequent scoring phase, conformers (of the same ligand) or poses (of different ligands) are assessed in the context of the protein active site against a scoring function which, it is hoped, correlates well with experimental activity and allows for accurate in silico ranking of ligands. There are a wide range of academic and commercial docking programs available (for example, MOE Dock,[2] DOCK,[3,4] FlexX,[5,6] Surflex,[7] AutoDock,[8] GOLD,[9] and GLIDE[10]), and a full assessment of search strategies and scoring functions, along with recent developments in the literature, can be found in refs 11−14. There are also a number of detailed comparisons of these programs which investigate their performance as tools in both lead optimization and lead generation.[15−17]

Scoring functions used in docking programs typically come in three variants: (1) energy-based, (2) empirical, or (3) knowledge-based. Energy-based functions approximate protein−ligand interactions through inclusion of terms known to be important for protein−ligand binding, empirical functions are composed of terms believed to be important, and knowledge-based functions are derived from a statistical analysis of atom−atom interactions (frequency counts are converted into a free energy function using the Boltzmann distribution). No one scoring function has been shown to be more effective than any other, although energy-based scoring schemes are more physical and offer a systematic approach for inclusion of terms that are often neglected, such as polarization. Terms can also be included which capture solvation or entropic effects.[18−21]

Within the energy-based framework, molecular mechanics (MM) force fields are commonly used, as they are fast to evaluate and the terms involved can be decomposed easily. Parameters for these terms (charges, well depths, etc.) are included from force fields such as CHARMM[22] or Amber.[23] A better estimate of the electronic interactions can be obtained by using a quantum mechanics/molecular mechanics (QM/MM) method that refits the ligand partial charges in a field of static MM charges that represent the protein. This allows polarization effects from the protein to be included.[24−28] For the QM part, there are many density functional theory (DFT) methods that provide approximate but qualitatively reliable results without imposing crippling computational costs. We use the popular B3LYP functional.[29,30] DFT has a number of well-known shortcomings,[31] perhaps most serious being its inability to account for dispersion interactions.[32−34] Much theoretical work has been done to address

* Corresponding author e-mail: burgers@mcmaster.ca.
† McMaster University.
‡ Boehringer Ingelheim Pharmaceuticals, Inc.

10.1021/ci100329z
Published on Web 12/06/2010

this shortcoming, and there are novel functionals that include dispersion terms directly[35,36] by determining the C6 term from the density[37−39] or including empirical corrections.[40]

The QM/MM approach has been shown to give good results in a number of studies. In ref 41 it is shown that docked poses can be ranked more effectively for hepatitis C virus NS5B polymerase inhibitors using QM/MM. In ref 42 it is shown that the docking program GLIDE gives better results if the ligand partial charges are refit with QM/MM. An improvement is also shown in the root mean square deviation (RMSD) of the redocked poses with respect to the crystallographically resolved ligands. Other studies have verified that QM/MM is useful in predicting the binding free energy.[43−45] While QM/MM approaches may provide a good description of the energetics of a static system, it is generally too expensive to do any sort of dynamics study.

To capture the free energy difference between ligands, methods such as free energy perturbation[46] or thermodynamic integration[47,48] can be quite accurate, predicting binding energies to within 1 kcal/mol. However, the amount of sampling required by methods such as molecular dynamics (MD) often makes these methods too expensive to be used regularly even when a purely MM approach is used. While obtaining accurate binding energies may be too expensive, MD can be invaluable for incorporating protein flexibility and for generating refined docking poses; for a review of MD applications and refinement methods for docking see ref 49.

In the present work we demonstrate that MD and QM/MM can be used in a stepwise manner to improve docked poses, giving better binding energies and reduced RMSD's from experimentally resolved ligand structures. The protocol is established and validated for a single trypsin inhibitor and then applied to a set of cytochrome *c* peroxidase (CCP)[50] binders and decoys.

Serine proteases are of critical importance in many biological processes, not least of which is blood coagulation. Trypsin[51] is a particularly well studied serine protease that hydrolyzes peptides with the aid of the catalytic triad His-Asp-Ser. We used a crystal structure (PDB ID 1TNG) which has the inhibitor (aminomethyl)cyclohexane (AMH) in the active site. The inhibitor has a measured K_d of 1.17 mM,[52] and the binding mode has a tight salt bridge between the ammonium ion of the ligand and the negatively charged aspartic acid 189, shown in Figure 1.

CCP is a heme-containing enzyme that catalyzes the oxidation of ferricytochrome *c* in yeast mitochondria. We have used the form of CCP containing the mutation Trp191 → Gly (CCPW191G), which has a small, completely buried, active site with an overall negative charge. This particular pocket, shown in Figure 2, has been studied extensively, and there is a wealth of biological information available in the literature, including 43 known ligands and 17 "decoys" that do not bind (see Table 1 for the names and K_d (mM) values). Structurally, the binders tend to be small heterocyclic molecules with a single positive charge. Crystallographic information is known for 35 of the 43 binders, and a number of computational studies have been performed on this system. See, for example, the work of Brooks,[53,54] Rosenfeld and co-workers,[55] and Deng and Verlinde.[56]

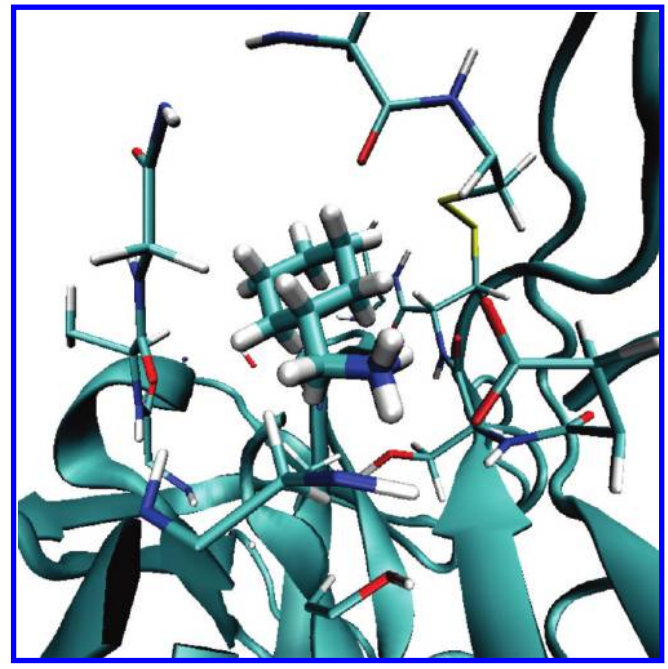

Figure 1. Trypsin active site with (aminomethyl)cyclohexane (PDB accession code 1TNG).

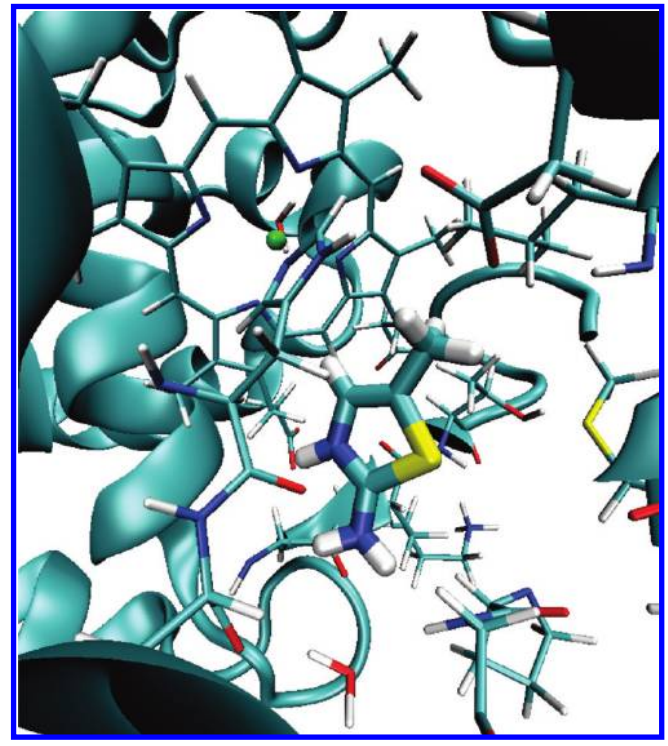

Figure 2. Active site of CCP with 2-amino-5-methylthiazole (PDB accession code 1AEH) after 40 ps of MD. The full heme group is shown with a crystal water bound to the iron. An additional crystal water that may be important for binding is shown beside the methyl group of the cognate ligand.

This paper is organized as follows: First, the general QM/MM theoretic framework is outlined, with an expression for the binding free energy. We then describe the methodology used to dock the ligands and the sequence of MD followed by QM/MM to improve upon these results. Last, we give two case studies with trypsin and CCP, and we demonstrate how the results are improved with the methodology proposed here.

Table 1. Names and Dissociation Constants (K_d) of the CCP Decoys and Binders Taken from Ref 73

	PDB code	K_d (mM)		PDB code	K_d (mM)
Binders					
2-aminopyridine	1AEO	0.05	aniline	1AEE	0.03
2-amino-5-methylthiazole	1AEN	0.01	phenol	2AS3	4.1
2,5-diaminopyridine hydrochloride	2AQD	N/A	(3-thienylmethyl)amine	2EUQ	0.05
2,6-diaminopyridine	2ANZ	0.06	4-aminopyrimidine		N/A
2-amino-5-methylpyridine	2EUP	0.04	3-fluorocatechol	2AS4	7.7
2,4-diaminopyridine	2EUN	0.05	4,5-diaminopyrimidine		N/A
2-amino-4-methylpyridine	2EUT	N/A	4-methylimidazole		0.06
3,4-diaminopyridine		N/A	cyclopentylamine	2AS6	N/A
indoline	1AEK	0.16	2,3,4-trimethylthiazole	1AC4	1.5
3-pyridylcarbinol		N/A	1-methylimidazole	1AET	0.07
4-aminopyridine	1AEG	0.04	3,4,5-trimethylthiazole	1AC8	0.2
2-amino-4-methylthiazole	1AEH	0.23	3-methylthiazole	1AEB	0.3
1,2-dimethylimidazole	1CMP	0.03	3,4-dimethylthiazole	1AED	0.05
2-iminopiperidine hydrochloride	2AS2	N/A	imidazole	1AES	0.07
3-thiophenecarboxamide hydrochloride	2AS1	0.02	benzylamine	2EUS	N/A
pyridine		0.14	1-methyl-1,6-dihydropyridin-3-amine	2EUO	N/A
2-ethylimidazole	1AEQ	0.73	imidazo[1,2-*a*]pyridine	1AEM	0.09
1*H*-imidazol-2-ylmethanol hydrochloride	2EUU	N/A	benzimidazole	1RYC	0.15
1-vinylimidazole	1AEJ	0.15	quinoline		0.42
2-ethyl-4-methylimidazole		N/A	3-aminopyridine	1AEF	0.05
2-methylimidazole	1AEU	0.07	2-aminothiazole	1AEV	0.04
4-pyridylcarbinol	2EUR	N/A			

Decoys			
3,5-difluoroaniline	3-chlorophenol	imidazole-2-thiol	dimethylammonium
toluene	3-aminopyrazole	pyrazole	methylammonium
4-aminoresorcinol hydrochloride	indazole	2-cyanoaniline	
3-amino-4-cyanopyrazole	4-amino-5-imidazolecarboxamide	isoazide	
3-cyanopyridine	pyrimidine-2-thiol	tetrazole	

THEORY

The basic theory of QM/MM calculations can be found in refs 25 and 57−59. In this work the QM/MM energy is calculated on a potential energy surface (PES), rather than the free energy surface. The total energy expression can be decomposed into QM, MM, and QM/MM terms. The QM term, $E(\mathbf{r}_{QM})$, is the QM energy of the ligand in the presence of the point charges that represent the MM part, but with the electrostatic interaction between the two removed. The MM term, $E(\mathbf{r}_{MM})$, includes all the nonbonded (electrostatic and van der Waals (vdW)) and bonded (bonds, angles, and dihedrals) force field terms. The cross terms between the QM and MM systems can written as

$$E_{QM/MM}(\mathbf{r}_{QM}, \mathbf{r}_{MM}) = E^{elec}_{QM/MM}(\mathbf{r}_{QM}, \mathbf{r}_{MM}) + E^{vdW}_{QM/MM}(\mathbf{r}_{QM}, \mathbf{r}_{MM}) + E^{cov}_{QM/MM}(\mathbf{r}_{QM}, \mathbf{r}_{MM}) \quad (1)$$

where $\mathbf{r}_{QM}$ are the coordinates of the QM subsystem, $\mathbf{r}_{MM}$ are the coordinates of the MM subsystem, and the energy terms are described below.

The electrostatic interaction between the QM and MM systems is $E^{elec}(\mathbf{r}_{QM},\mathbf{r}_{MM})$ and is most appropriately written as

$$E^{elec}_{QM/MM} = \sum_{i \in MM} \int \rho(r) q_i/(r\, dr) \quad (2)$$

where $\rho(r)$ is the density of the QM system and r is the distance between a point in the QM system and the MM charge q_i. This may be approximated with a more tractable form:

$$E'^{elec}_{QM/MM} = \sum_{j \in QM} \sum_{i \in MM} Q_j q_i / r_{ij} \quad (3)$$

where Q_j are the charges obtained by electrostatic potential (ESP) fitting of the QM system.

The term $E^{vdW}_{QM/MM}(\mathbf{r}_{QM},\mathbf{r}_{MM})$ includes the vdW interactions between the MM and QM systems and is given as

$$E^{vdW}_{QM/MM} = \sum_{j \in QM} \sum_{i \in MM} \left[-\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \right] \quad (4)$$

where A_{ij} and B_{ij} are taken from the CHARMM22[22] force field for proteins. For the ligands there are no such terms, so we developed a protocol to match ligand atoms, on the basis of their bonding, to parameters from CHARMM22 for amino acids.

$E^{cov}_{QM/MM}$ is the energy of the bonded covalent interaction between the QM and MM systems. This term is commonly calculated with either link atoms or pseudobonds.[60] Both methods are known to work well, although pseudobonds allow for a better description since they can be fine-tuned to the molecular system of interest. This term allows residues in the active site that directly interact with the ligand to be included. We chose not to include parts of the enzyme as this would significantly increase the size of the QM system, making the computational cost much higher. Instead we assumed that an MM description was sufficient for the enzyme. In this case $E^{cov}_{QM/MM} = 0$ since there are no covalent bonds linking the MM and QM systems.

To obtain E_{QM}, an ab initio calculation is performed in the presence of the point charges of the enzyme. To obtain the energy of the polarized QM system alone, E'_{QM}, the energy of these charges should be removed. To do this, we write the energy as

$$E'_{QM} = (E_{QM} + E^{elec}_{QM/MM}) - E'^{elec}_{QM/MM} \quad (5)$$

The term $E_{QM} + E^{elec}_{QM/MM}$ can be obtained with an effective Hamilton:

$$\hat{H}_{eff} = \hat{H} + \sum_{i \in MM} \sum_{j=1}^{n} q_i e / r_j \quad (6)$$

where $\hat{H}$ is the usual n-electron operator.

While eq 3 gives the correct electrostatic energy for a point charge model, it will overestimate the true value corresponding to the charge density $\rho(r)$. This is due to the rearrangement of the solvent and surrounding atoms to screen the overall charge. To account for this, a dielectric screening term can be introduced, so that we have

$$E'^{elec}_{QM/MM} = \sum_{j \in QM} \sum_{i \in MM} \frac{Q_j q_i}{\varepsilon(r) r_{ij}} \quad (7)$$

To make the screening distance dependent, we set $\varepsilon(r) = r_{ij}$. In ref 41 $\varepsilon(r) = 4r_{ij}$ is used; however, in our experience, we find that this damps out the interactions too much. Collecting terms, we arrive at an expression for the binding energy:

$$\Delta E_{bind} = E^{vdW}_{QM/MM} + E'^{elec}_{QM/MM} + \Delta E'_{QM} - 0.5 \Delta G_{solv} \quad (8)$$

where $\Delta E'_{QM} = E'_{QM} - E_{QM,gas}$ is the polarization energy from the electrostatic environment. The final term, $-0.5\Delta G_{solv}$, is included to account for desolvation effects. This represents the free energy of solvation for the ligand being transferred from a vacuum to bulk water.

While working on the PES is the simplest approach, it has been shown that this is often insufficient for fully describing the most important interactions.[41] Instead, a better strategy is to work on the mean field PES,[61−63] where the differences in energy are free energies. However, this requires extensive sampling and would be impractical for rapid refinement of docked poses. Thus, rather than working on the mean field surface, in the present study we perform a short run of MD sampling to equilibrate the system, followed by minimization on the QM/MM PES. We demonstrate that this dramatically improves the energetic and geometric components of the docking results.

METHODS

System Preparation. The procedures for preparing both trypsin and CCP were identical unless otherwise noted. The crystal structures for trypsin (PDB ID 1TNG) and CCP (PDB ID 1AEH) were independently loaded into the Molecular Operating Environment (MOE) software (version 2003.02).[2] The solvent was removed, except for a structurally important water (HOH308) in the CCP system.[64] The Protonate3D function was used to apply hydrogens.[65] In the CCP example this was done both with and without water 308. The binding site and cognate ligand were visually inspected, and the automated MOE SiteFinder algorithm was applied to generate a number of clusters of site points, with the one corresponding to the cognate ligand selected for use in the docking experiments.

Initial Docking. The 43 binders and 17 decoys were downloaded from the Shoichet group Web site[66] as SMILES strings and converted into three-dimensional structures within MOE; visual inspection and manual correction (if needed) of these 60 molecules ensured the correct SMILES perception. For trypsin, the ligand AMH had hydrogens added during the Protonate3D process followed by excision of the ligand from the binding site. AM1BCC charges were assigned, and finally the molecule was geometry optimized.

The prepared ligands were docked using the docking algorithm MOE Dock with default parameters. For trypsin the top 20 poses were used, while for CCP only the 5 top-scoring poses from each ligand were used (both in the presence and in the absence of the structurally important water molecule HOH308). In the CCP example all molecules docked except dimethylammonium, which did not satisfy the minimum number of features required by MOE Dock for the initial placement; this decoy was removed from the subsequent analysis.

MD Equilibration. To equilibrate the ligands before running QM/MM minimization, a short run of MD was done on just the ligand with the heavy atoms of the enzyme fixed. The ligand was set as the QM part with the rest of the enzyme treated with MM. Specifically for the trypsin inhibitor we used 1 ps of QM/MM MD with HF/STO-3G to speed up the MD. The forces were computed using Gaussian03 and were fed into the MD package of the program Sigma, version 2.2.[67] No parameters had to be assigned to the ligand.

For CPP an in-house program was used to assign atom types to the docked ligand poses by comparison to the CHARMM22 parameter file; missing bonded terms were assigned by analogy to existing terms. To obtain a better initial charge model for the ligands, we minimized structures in vacuo with Gaussian03[68] using the B3LYP functional and the 6-31G* basis set. Atomic charges for the resulting structures were determined by fitting the electrostatic potential using the Merz−Singh−Kollman scheme.[69]

Before QM/MM minimization was performed, the ligands (with their new force field terms), along with all nonheavy atoms in the protein, were MM minimized. Following this, we performed 40 ps of MD with Sigma using the last structure for the QM/MM minimization. The MD was run in three different ways: (1) only on the ligand in the presence of the protein, (2) on the ligand and the protein with the backbone of the protein restrained, and (3) on the ligand and the protein with the backbone and the β-carbons restrained. For the atomic restraints a 40 kcal/mol force constant was used.

QM/MM. The QM subsystem consisted of the ligand atoms only and was treated using B3LYP/6-31G*. The protein was represented with MM using the CHARMM22 force field. The QM/MM module from Sigma, version 2.2, was used for all calculations. This is a software package that interfaces the MM part of Sigma with Gaussian03. The entire protein was included as a nonperiodic calculation with no cutoff imposed. All QM optimizations were performed with Cartesian coordinates, for a maximum of 20 iterations. The point charges from the MM system were read in as a field, and the vdW terms were added to the force and energy by Sigma from the assigned parameters.

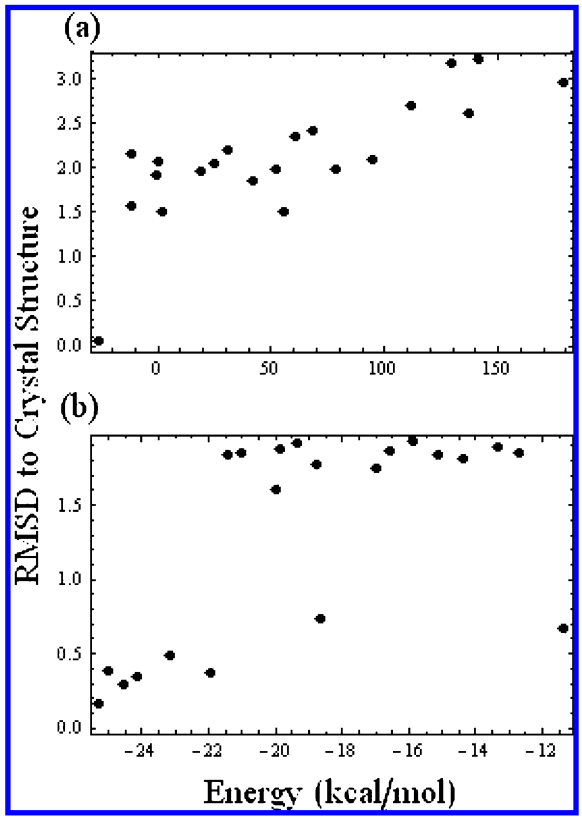

Figure 3. A comparison of the binding energy using eq 8 versus the RMSD from the crystallographic pose of the inhibitor in the PDB file 1TNG. The scatter plot shows the results of (a) just doing QM/MM minimization and (b) first running 20 ps of MD before the minimization.

Free Energy of Solvation. All ligands were optimized using B3LYP/6-31G*, and the optimized structures were used to calculate the free energy of solvation within the polarizable continuum model (PCM) using water as a solvent. Radii for the cavity were assigned with the united atom for Hartree−Fock (UAHF) model.[70]

RESULTS

Trypsin Inhibitor. The active site for trypsin is shown in Figure 1 with the inhibitor (aminomethyl)cyclohexane. Among the top 20 docked poses that MOE produced, 7 were within 2.0 Å of the somewhat arbitrary, albeit widely used, cutoff used to assess docking accuracy. None of the structures were within 1 Å. Given the crude docking scoring function used, it is not too surprising that there is a lack of correlation between the score and the RMSD of docked poses.

The 20 poses were then QM/MM minimized. Some of the docked poses could be identified as "correct" using their defined QM/MM internal energy. However, using just the internal energy, we found this approach less successful than previous authors.[41] The native structure had the lowest QM/MM internal energy, but the structures from the docking program did not change significantly during the minimization so that the best structures were still no more than 1.5 Å away from the crystal structure (Figure 3a). In ref 41 only 13 of the 20 initial poses generated with FlexX had an RMSD < 2 Å, while the remaining poses had a grossly inaccurate RMSD (>6 Å). Given what appears to be a heavily skewed bimodal distribution, refinement through QM/MM MD offers clearer discrimination between correct and incorrect poses. In the present example, the majority of the poses were within 2.5 Å and discrimination through QM/MM was more challenging.

After QM/MM MD was run on the inhibitor, followed by QM/MM minimization, the poses showed improvement (Figure 3b). The top 5 poses (selected by ΔE_{bind}) are all within 0.6 Å of the native structure. MD was also run on the native structure, which moved only slightly (∼0.2 Å) away from its initial pose. A reason for the improvement, following this refinement process, may be because it is too easy for structures to get stuck in local minima on the QM/MM potential energy surface. A very short MD run appears to be sufficient to move the structures into lower energy conformations, resulting in a lower RMSD. The QM/MM binding energy provides a good measure of poses that are close to the crystal structure. However, there is a large jump in energy between the closely docked structures with an RMSD of 0.5 Å and the structures that are further away.

In addition to highlighting the benefits of running MD, this example also demonstrates the utility of the QM/MM internal energy as a successful predictor of bioactivity for this system. This was not clear from the work in ref 41 since the gas-phase energy was as successful an indicator of good poses as the QM/MM binding energy. With (aminomethyl)-cyclohexane the system is rigid enough that there is almost no difference in internal energies between the poses, so we can discount this effect.

Cytochrome *c* Peroxidase. Docking experiments on this system are complicated by a structurally important water molecule in the active site. The average RMSD distribution of the docked poses across the binders for which crystallographic data are known is shown in Figure 4. The same improvements in RMSD observed for the trypsin system are seen for the CCP ligands. With the new methodology the average RMSD for the set of ligands was reduced from 2.43 to 1.91. There were only six structures that ended up with poses worse than the initial docked pose.

Refinement of the CCP docking poses was performed in the same manner as for the trypsin example except that MM MD was used instead of QM/MM MD. While it is possible that the polarization effects of the enzyme may be important in the MD trajectory, one can account for some of this by fitting ESP charges every 100 or 1000 steps of the MD trajectory. However, having run MD on the trypsin system for 20 ps using MM MD with the initial QM/MM point charges fit from the MOE docked poses, we observed that the results were similar to the full QM/MM MD results. From this test we felt confident not including the polarization effects during the MD run and instead used fixed force field parameters.

For the top five docked poses, MM MD was done and the final structure was QM/MM minimized. The results giving the lowest binding energy for each ligand, as determined by eq 8, are summarized in Figure 5. For this plot, the decoys appear on the right of the figure with binders on the left. The refinement method works well to separate the two groups, with the exception of a few compounds such as 4-aminoresorcinol hydrochloride, which seems to fit very well in the active site and is falsely predicted to be a binder. This particular compound is predicted to have a high free energy of solvation, and this may be the predominant reason

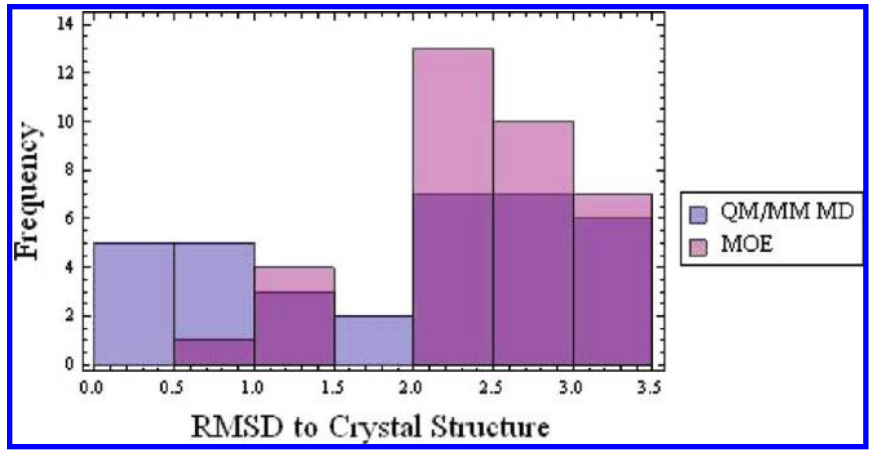

Figure 4. Histogram comparison of the RMSD from the crystal structure pose for both MOE and the hybrid MD and QM/MM methodology for the CCP ligands. Bins of size 0.5 Å were used.

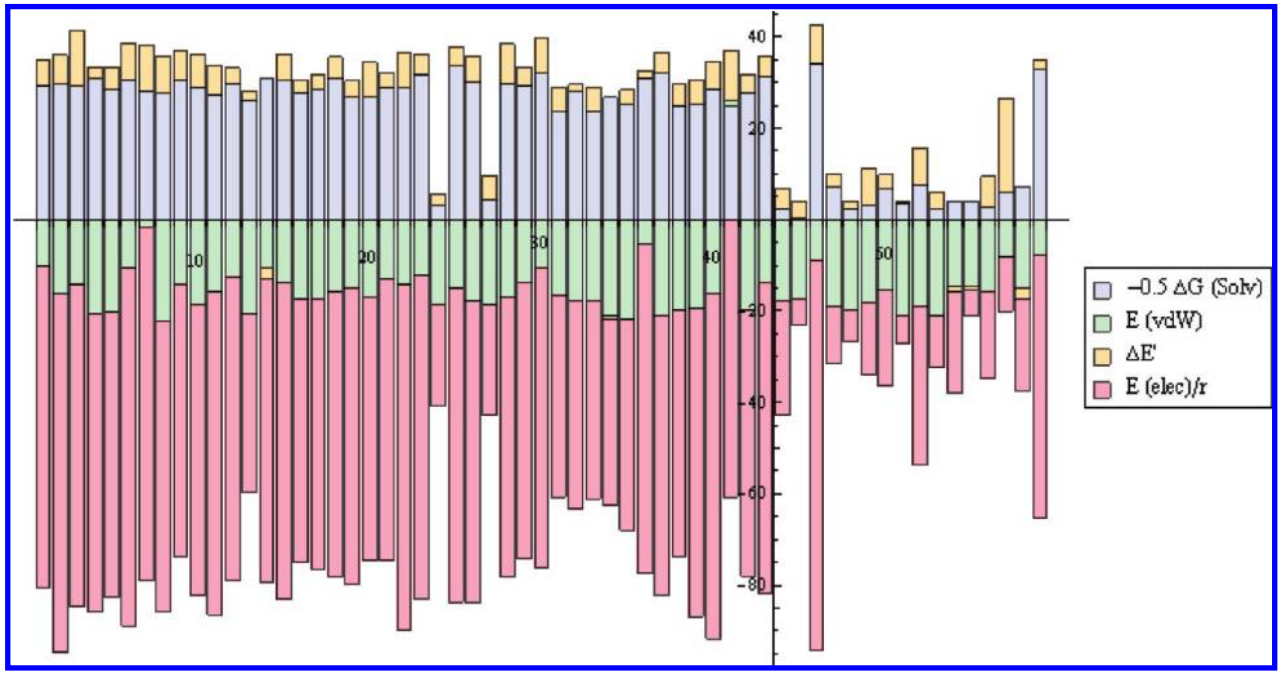

Figure 5. Bar graph illustrating the energetic contributions to the binding energy. All bars left of the axis are binders; those to the right are decoys. The order of the ligands corresponds to Table 1. The total binding energy is given by eq 8, where the terms on the right side of the equation are shown in the legend. The bottom of the bars is the total energy.

it does not bind. While this effect is partially accounted for in eq 8, it is not enough to remove some of the decoys.

While the absolute value of ΔE_{bind} from eq 8 is not directly comparable to the binding free energy of the ligands, it should act as a scoring function that is correlated with $-\log(K_d)$. To verify that this is true, linear regression was performed with $-\log(K_d)$ and ΔE_{bind}. The correlation was $R^2 = 0.50$ with a root-mean-square error of 0.63 kcal/mol. This is reasonable for a static, essentially gas phase model. If more terms were included in the fit, a better correlation could be obtained. However, the aim was to create a general model that did not involve fitting.

An investigation of the impact of allowing MD to be run on different parts of the system appears in the receiver operating characteristic (ROC)[71,72] plot in Figure 6. For this particular enzyme it is clear that including protein motions does not have much effect on the relative binding energy difference between compounds. Other less rigid enzymes or enzymes with active sites that are more solvent exposed might yield better results if flexibility in the protein was better accounted for. However, for this example, the benefits of including side chain motions does not warrant the associated additional computational effort.

In Figure 7, an ROC analysis of the rank-ordered QM/MM MD or simple docking scores (both in the presence and in the absence of HOH308) is presented. For the QM/MM MD protocol, the binders and decoys are very well separated; it appears that the results are insensitive to the presence or absence of the crystal water, which is clearly more of a problem for the naïve docking score.[56] This may be due to the fact that the water can be moved to an appropriate spot in the active site following dynamics whereas with the docking algorithm the water is fixed in its original crystallographic orientation. Overall it appears that the QM/MM MD approach is much less sensitive to crystal waters.

CONCLUSIONS

Using trypsin and CCP as example systems, we have demonstrated that performing MD followed by QM/MM

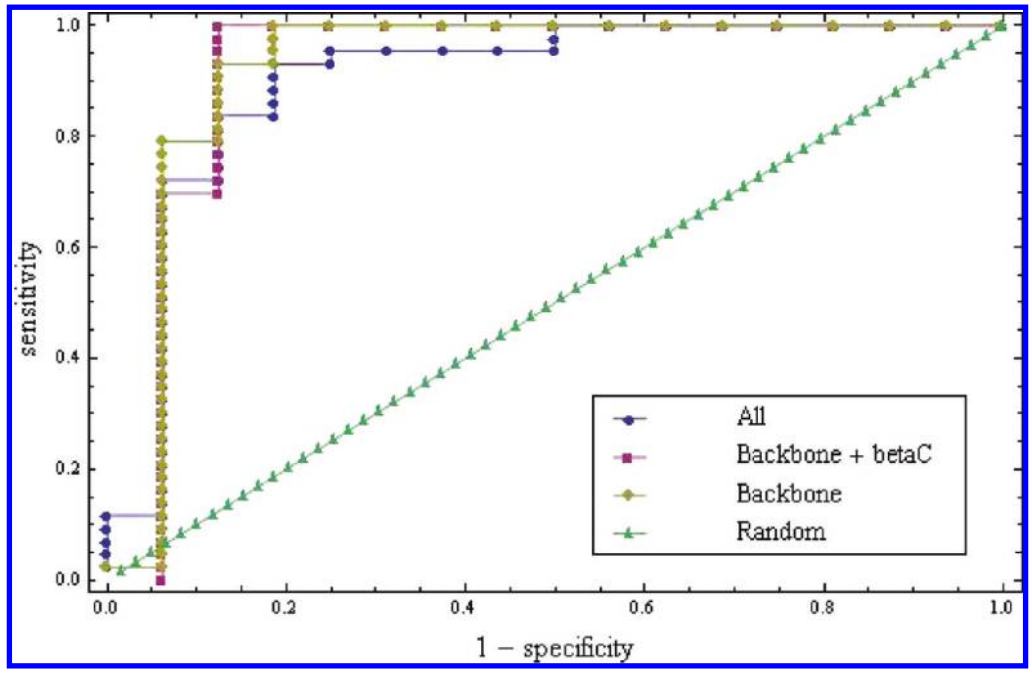

Figure 6. ROC plot for the 59 molecule set of CCP binders and decoys. Binding energy calculations were performed after 40 ps of MD with different restraints on the enzyme: (1) "All" means only on the ligand in the presence of the protein. (2) "Backbone + betaC" means the ligand and the protein with the backbone of the protein restrained. (3) "Backbone" means the ligand and the protein with the backbone and the β-carbons restrained. For the atomic restraints a 40 kcal/mol force constant was used.

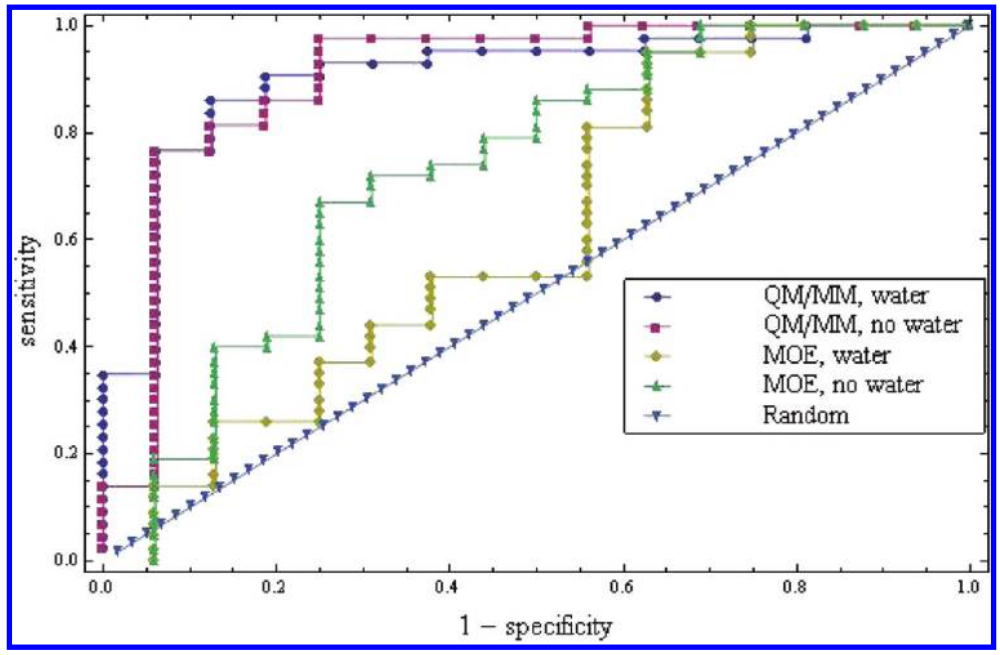

Figure 7. ROC plot of the binding energy for the 59 molecule set of CCP binders and decoys: performance of the MOE Dock scoring function compared against the QM/MM binding energy (eq 8) both with and without crystallographic water (HOH308) present.

minimization dramatically improves the rank ordering of docked poses generated by a docking program, in terms of both the RMSD from the crystal structure and correlation to the experimental binding free energy. Using this protocol, we have successfully distinguished between binders and decoys for CCP on the basis of their QM/MM energies of interaction with an extra term added to include desolvation effects. Unlike the results from the docking program used, the methodology appears largely insensitive to crystal waters in the active site and was also insensitive to the degree to which the protein was restrained during the dynamics simulation. The results illustrate that, to distinguish between binders and decoys for docking, it may not be enough to use an MM-based description. Instead, QM/MM energies, augmented with a vdW correction and a desolvation term, may be needed to correctly describe the binding energy. The general applicability of this protocol should be investigated against a broader range of protein−ligand systems, but these initial results appear encouraging.

ACKNOWLEDGMENT

This work was supported by the Department of Chemistry, McMaster University, the Hamilton Foundation, NSERC, and the Canada Research Chairs.

REFERENCES AND NOTES

(1) Jorgensen, W. L. The many roles of computation in drug discovery. *Nature* **2004**, *303*, 1813–1818.
(2) *Molecular Operating Environment (MOE)*, version 2003.02; Chemical Computing Group Inc.: Quebec, Canada, 2003.
(3) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18* (9), 1175–1189.
(4) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: Combining techniques to model RNA−small molecule complexes. *RNA* **2009**, *15* (6), 1219–1230.
(5) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
(6) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: Placing discrete water molecules during protein−ligand docking predictions. *Proteins: Struct., Funct., Genet.* **1999**, *34* (1), 17–28.
(7) Jain, A. N. Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput.-Aided Mol. Des.* **2007**, *21* (5), 281–306.
(8) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28* (6), 1145–1152.
(9) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.
(10) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2007**, *47*, 1739–1749.
(11) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 178–184.
(12) Davis, I. W.; Baker, D. ROSETTALIGAND docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009**, *385* (2), 381–392.
(13) Chen, Y.; Shoichet, B. K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5* (5), 358–364.
(14) Robeits, B. C.; Mancera, R. L. Ligand−protein docking with water molecules. *J. Chem. Inf. Model.* **2008**, *48* (2), 397–408.
(15) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
(16) Cummings, M. D.; Des Jarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* **2005**, *48*, 962–976.
(17) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y. B.; Humblet, C. Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2009**, *49* (6), 1455–1474.
(18) Lee, M. R.; Sun, Y. X. Improving docking accuracy through molecular mechanics generalized Born optimization and scoring. *J. Chem. Theory Comput.* **2007**, *3* (3), 1106–1119.
(19) 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.
(20) 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* (16), 4805–4808.
(21) Thompson, D. C.; Humblet, C.; Joseph-McCarthy, D. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.* **2008**, *45* (5), 1081–1091.
(22) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(24) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.
(25) Zhang, Y. K.; Liu, H. Y.; Yang, W. T. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.* **2000**, *112* (8), 3483–3492.
(26) Zhang, Y. K.; Lee, T. S.; Yang, W. T. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* **1999**, *110*, 46–54.
(27) Liu, H.; Lu, Z.; Cisneros, G. A.; Yang, W. Parallel iterative reaction path optimization in quantum mechanical/molecular mechanical modeling of enzyme reactions. *J. Chem. Phys.* **2004**, *121*, 697–706.
(28) Senn, H. M.; Thiel, W. QM/MM methods for biological systems. *Top. Curr. Chem.* **2007**, *268*, 173–290.
(29) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle−Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
(30) Becke, A. D., III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(31) Schreiner, P. R. Relative energy computations with approximate density functional theory−A caveat. *Angew. Chem., Int. Ed.* **2007**, *46*, 4217–4219.
(32) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. Application of 25 density functionals to dispersion-bound homomolecular dimers. *Chem. Phys. Lett.* **2004**, *394* (4), 334–338.
(33) Sherrill, D. C. Frontiers in electronic structure theory. *J. Chem. Phys.* **2010**, *132*, 110902.
(34) Alonso, J. A.; Mananes, A. Long-range van der Waals interactions in density functional theory. *Theor. Chem. Acc.* **2007**, *117* (4), 467–472.
(35) Puzder, A.; Dion, M.; Langreth, D. C. Binding energies in benzene dimers: Nonlocal density functional calculations. *J. Chem. Phys.* **2006**, *124*, 164105.
(36) von Lilienfeld, A. O.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Optimization of effective atom centered potentials for London dispersion forces in density functional theory. *Phys. Rev. Lett.* **2007**, *93* (15), 153004.
(37) Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123* (15), 154101.
(38) Angyan, J. On the exchange-hole model of London dispersion forces. *J. Chem. Phys.* **2007**, *127* (2), 024108.
(39) Ayers, P. W. A perspective on the link between the exchange(-correlation) hole and dispersion forces. *J. Math. Chem.* **2009**, *46* (1), 86–96.
(40) Wu, Q.; Yang, W. T. Empirical correction to density functional theory for van der Waals interactions. *J. Chem. Phys.* **2002**, *116*, 515–524.
(41) Parks, J. M.; Kondru, R. K.; Hu, H.; Beratan, D. N.; Yang, W. T. Hepatitis C virus NS5B polymerase: QM/MM calculations show the important role of the internal energy in ligand binding. *J. Phys. Chem. B* **2008**, *112* (10), 3168–3176.
(42) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of accurate charges in molecular docking: Mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26*, 915–931.
(43) Zhou, T.; Huang, D. Z.; Caflisch, A. Quantum mechanical methods for drug design. *Curr. Top. Med. Chem.* **2010**, *10* (1), 33–45.
(44) Zhou, T.; Huang, D.; Caflisch, A. Is quantum mechanics necessary for predicting binding free energy? *J. Med. Chem.* **2008**, *51* (14), 4280–4288.
(45) Nayeem, A.; Krystek, S.; Stouch, T. An assessment of protein-ligand binding site polarizability. *Biopolymers* **2003**, *70* (2), 201–211.
(46) Reddy, R. M.; Erion, M. D. *Free Energy Calculations in Rational Drug Design*; Springer: New York, 2001; pp 1−406.
(47) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic: New York, 2002; pp 1−664.
(48) den Otter, W. K. Thermodynamic integration of the free energy along a reaction coordinate in Cartesian coordinates. *J. Chem. Phys.* **2000**, *112* (17), 7283–7292.
(49) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26* (5), 531–568.
(50) Goodin, D. B.; Mcree, D. E. The Asp-His-Fe triad of cytochrome-c peroxidase controls the reduction potential, electronic-structure, and coupling of the tryptophan free-radical to the heme. *Biochem. J.* **1993**, *32* (13), 3313–3324.
(51) Green, N. M.; Work, E. Pancreatic trypsin inhibitor. 2. Reaction with trypsin. *Biochem. J.* **1953**, *54* (2), 347–352.
(52) Kurinov, I.; Harrison, R. W. Prediction of novel serine protease inhibitors. *Nat. Struct. Biol.* **1994**, *1*, 735–743.
(53) Banba, S.; Guo, Z. Y.; Brooks, C. L. Efficient sampling of ligand orientations and conformations in free energy calculations using the λ-dynamics method. *J. Phys. Chem.* **2000**, *104*, 6903–6910.
(54) Banba, S.; Brooks, C. L. Free energy screening of small ligands binding

to an artificial protein cavity. *J. Chem. Phys.* **2000**, *113*, 3423–3433.
(55) Rosenfeld, R. J.; Goodsell, D. S.; Musah, R. A.; Morris, G. M.; Goodin, D. B.; Olson, A. J. Automated docking of ligands to an artificial active site: Augmenting crystallographic analysis with computer modeling. *J. Comput.-Aided Mol. Des.* **2003**, *17* (8), 525–536.
(56) Deng, W.; Verlinde, C. Evaluation of different virtual screening programs for docking in a charged binding pocket. *J. Chem. Inf. Model.* **2008**, *48* (10), 2010–2020.
(57) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes*; Wiley: New York, 1991; pp 1−256.
(58) Gao, J. Methods and applications of combined quantum mechanical and molecular mechanical potentials. *Reviews in Computational Chemistry*; VCH: New York, 1995; Vol. 7, pp 119−185.
(59) Hu, H.; Yang, W. T. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.* **2007**, *59*, 573–601.
(60) Zhang, Y. K. Improved pseudobonds for combined *ab initio* quantum mechanical/molecular mechanical methods. *J. Chem. Phys.* **2005**, *112*, 024114.
(61) Hu, H.; Lu, Z. Y.; Yang, W. T. QM/MM minimum free-energy path: Methodology and application to triosephosphate isomerase. *J. Chem. Theory Comput.* **2007**, *3* (2), 390–406.
(62) Hu, H.; Yang, W. T. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.
(63) Hu, H.; Yang, W. T. Development and application of ab initio QM/MM methods for mechanistic simulation of reactions in solution and in enzymes. *THEOCHEM* **2009**, *898* (1−3), 17–30.
(64) Brenk, R.; Vetter, S. W.; Boyce, S. E.; Goodin, D. B.; Shoichet, B. K. Probing molecular docking in a charged model binding site. *J. Mol. Biol.* **2006**, *357* (5), 1449–1470.
(65) Labute, P. Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins: Struct., Funct., Bioinf.* **2009**, *75* (1), 187–205.
(66) Shoichet Laboratory Home Page. http://shoichetlab.compbio.ucsf.edu/ (accessed March, 2009).
(67) Mann, G.; Yun, R. H.; Nyland, L.; Prins, J.; Board, J.; Hermans, J. The sigma MD program and a generic interface applicable to multi-functional programs with complex, hierarchical command structure. In *Computational Methods for Macromolecules: Challenges and Applications*; Schlick, T., Gan, H. H., Eds.; Springer-Verlag: New York, 2002; Vol. 24, pp 129−145.
(68) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2003.
(69) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.
(70) Barone, V.; Cossi, M.; Tomasi, J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.* **1997**, *107* (8), 3210–3221.
(71) Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27* (8), 861–874.
(72) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48* (7), 2534–2547.
(73) Musah, R. A.; Jensen, G. M.; Bunte, S. W.; Rosenfeld, R. J.; Goodin, D. B. Artificial protein cavities as specific ligand-binding templates: Characterization of an engineered heterocyclic cation-binding site that preserves the evolved specificity of the parent protein. *J. Mol. Biol.* **2002**, *315* (4), 845–857.

CI100329Z