Assessment of QM/MM Scoring Functions for Molecular Docking to HIV-1 Protease

Pedro Fong,[†] Jonathan P. McNamara,[‡] Ian H. Hillier,[‡] and Richard A. Bryce*[,†]

School of Pharmacy and Pharmaceutical Sciences and School of Chemistry, University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom

Received November 26, 2008

We explore the ability of four quantum mechanical (QM)/molecular mechanical (MM) models to accurately identify the native pose of six HIV-1 protease inhibitors and compare them with the AMBER force field and ChemScore and GoldScore scoring functions. Three QM/MM scoring functions treated the ligand at the HF/6-31G*, AM1d, and PM3 levels; the fourth QM/MM function modeled the ligand and active site at the PM3-D level. For the discrimination of native from non-native poses, solvent-corrected HF/6-31G*:AMBER and AMBER functions exhibited the best overall performance. While the electrostatic component of the MM and QM/MM functions appears important for discriminating the native pose of the ligand, the polarization contribution in the QM/MM functions was relatively insensitive to a ligand's binding mode and, for one ligand, actually hindered discrimination. The inclusion of a desolvation penalty, here using a generalized Born solvent model, improved discrimination for the MM and QM/MM methods. There appeared to be no advantage to binding mode prediction by incorporating active site polarization at the PM3-D level. Finally, we found that choice of the protonation state of the aspartyl dyad in the HIV-1 protease active site influenced the ability of scoring methods to determine the native binding pose.

INTRODUCTION

Structure-based virtual screening by molecular docking is now a widely used strategy in rational drug design. The effectiveness of this approach relies on the quality of the computed binding mode and the affinity of a ligand in its receptor binding site. The prediction of the binding affinity in a molecular docking tool is estimated by a scoring function, which generally needs to be both fast and accurate. However, accurate scoring of protein−ligand interactions remains problematic for a number of reasons, including the neglect of solvation effects, oversimplistic treatments of ligand−protein interactions, and other deficiencies in the scoring function, along with inaccuracies in protein structure, ligand size dependence, and uncertainties in protonation states.[1] Moreover, there is a tendency for scoring functions to perform well only for certain types of protein−ligand interactions. These difficulties lead to the prediction of a large number of false positives and false negatives in current scoring functions.[2,3]

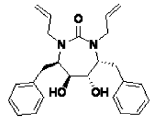
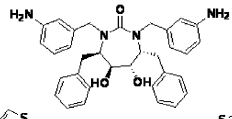
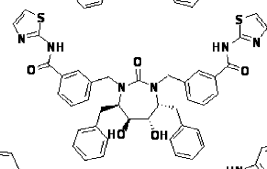
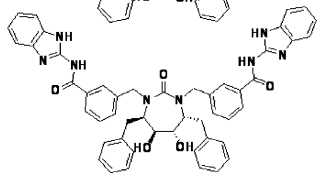
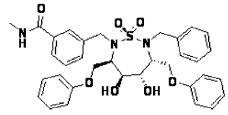
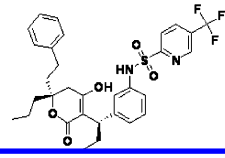
Scoring functions are typically grouped into three approaches: (i) knowledge-based potentials, such as PMF[4] and DrugScore,[5] that derive a quasi-potential of mean force from crystallographic protein−ligand data; (ii) empirical potentials, such as ChemScore,[6] derived from fitting to a set of protein−ligand binding affinities and structures; and (iii) physics-based potentials. For the latter, for example GoldScore,[7] the potentials are often related to molecular mechanical force fields, such as AMBER[8] and CHARMm.[9] A disadvantage of these approaches is the need for specific force field parameters to span the diverse range of chemical functionalities that exist within a virtual screening library. Therefore, there is growing interest in employing combined quantum mechanical/molecular mechanical (QM/MM) approaches as general physics-based scoring functions for docking[10,11] and binding affinities.[12−14] Although currently prohibitive in cost, QM/MM methods could be useful tools for screening large ligand libraries with diverse chemical structures. The fundamental concept of the QM/MM method is to partition a system into two regions: the ligand and part of the protein is represented quantum mechanically and the protein/solvent environment is modeled by an empirical MM force field. Treating the ligand and proximal amino acids at the QM level incorporates physical details such as electronic polarizability that could potentially provide improved accuracy in the scoring of docked poses and ranking between ligands. Recognition of the importance of including explicit polarizability in modeling biomolecular recognition is reflected by ever-increasing numbers of polarizable force field parametrizations.[11,15]

In this study, we explore the ability of a two-level ONIOM treatment[16] to accurately identify the native bound conformations of six HIV-1 protease inhibitors, **1**−**6** (Table 1). The QM/MM coupling is via electrostatic embedding,[16,17] such that the partial charges from the MM region enter the QM Hamiltonian. By this approach, electronic polarization of the ligand by its protein environment is explicitly considered. Previous work has found that polarization can contribute up to around one-third of the total electrostatic binding energy of HIV-1 protease inhibitors in their native pose.[18] Polarization effects are suggested to play an important role in transition state stablilization.[19] If significant orientation dependence of this term is exhibited, it could prove important for accurate discrimination of the native pose by docking algorithms. Here, we examine the dependence on orientation

* Corresponding author e-mail: Richard.Bryce@manchester.ac.uk.
[†] School of Pharmacy and Pharmaceutical Sciences.
[‡] School of Chemistry.

10.1021/ci800432s CCC: $40.75
Published on Web 03/23/2009

Table 1. Structure and Binding Affinity of HIV-1 Protease Inhibitors

No.	Name	PDB code	
1	XK216[25]	1HWR	
2	DMP450[26]	1DMP	
3	XV638[27]	1BWA	
4	SD146[27]	1BWB	
5	AHA047[28]	1G2K	
6	tipranavir[29]	1D4Y	

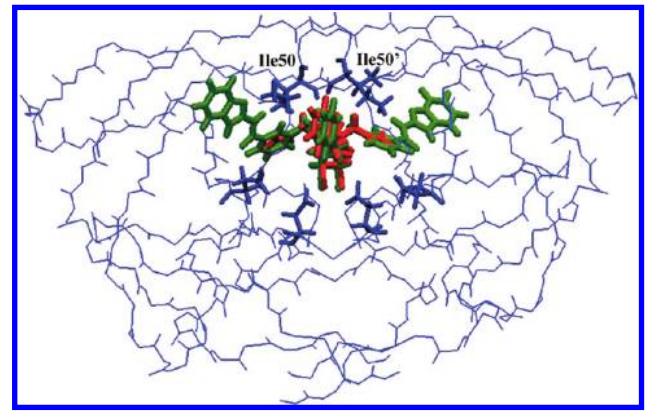


Figure 1. HIV-1 protease with overlay of bound poses of ligands **1** and **4**.

Table 2. Decomposition of Total Interaction Energy (ΔE_{inter}) at the HF/6-31G*:AMBER Level for the Native Pose of Ligands **1**−**6** and Ligand Efficiency (Δe_{inter}) and Solvent-Corrected Total Interaction Energy (ΔE^{aq}_{inter}) and Ligand Efficiency (Δe^{aq}_{inter})[a]

	inhibitor					
contribution	**1**	**2**	**3**	**4**	**5**	**6**
ΔE_{inter}	−95.5	−136.4	−156.7	−184.9	−122.1	−152.1
Δe_{inter}	−3.18	−3.41	−2.90	−2.98	−2.71	−3.62
ΔE_{vdw}	−54.1	−75.4	−87.9	−95.4	−76.8	−67.3
ΔE_{el}	−41.4	−61.0	−68.8	−89.5	−45.3	−84.8
ΔE_{perm}	−36.4	−53.8	−59.5	−75.6	−38.5	−69.7
ΔE_{pol}	−5.0	−7.2	−9.6	−13.9	−6.8	−15.1
ΔE_{stab}	−9.9	−14.4	−18.6	−28.2	−13.6	−29.7
ΔE_{dist}	5.0	7.2	9.3	14.3	6.8	14.7
$\Delta E_{pol}/\Delta E_{el}$	0.12	0.12	0.14	0.16	0.15	0.18
ΔE_{solv}	33.9	58.5	69.4	97.8	61.1	66.9
ΔE^{aq}_{inter}	−61.6	−77.9	−87.3	−87.1	−61.0	−85.3
Δe^{aq}_{inter}	−2.05	−1.95	−1.62	−1.40	−1.36	−2.03

[a] All energies in kilocalories per mole.

of protein−ligand energetics for six HIV-1 protease inhibitors at four distinct QM/MM levels of theory: for three ONIOM models, we describe the ligand at the HF/6-31G*, AM1d,[20] and PM3[21] levels of quantum mechanics, in conjunction with the AMBER force field[8] to represent the protein. We also explore a fourth ONIOM treatment where, in addition to the ligand, 31 residues of the protein active site are modeled quantum mechanically. For this, we use the dispersion-corrected PM3 method (PM3-D) for biomolecular interactions,[22] which modifies the PM3 Hamiltonian for dispersion effects via an empirical function. These QM/MM methods are compared with a force field description of the protein−ligand interactions using AMBER, and with the ChemScore[7] and GoldScore[23] scoring functions. The contribution of a solvent to the binding landscape is also considered using a generalized Born (GB) continuum model.

METHODS

Structure Preparation. Initial structures of the six HIV-1 protease inhibitors, **1**−**6**, and their geometric decoys were obtained from the Ligand Protein Data Bank (LPDB)[24] with respective PDB codes: 1HWR,[25] 1DMP,[26] 1BWA,[27] 1BWB,[27] 1G2K,[28] and 1D4Y[29] (Table 1). These structures correspond to wild-type HIV-1 protease (1HWR, 1DMP, and 1G2K), a double mutant, Val82Phe/Ile84Val (1BWA and 1BWB), and a triple mutant, Gln7Lys/Leu33Ile/Leu63Ile (1D4Y). These protein/ligand structures had been previously prepared for the LPDB by the removal of crystallographic waters,[24] addition of hydrogen atoms, and energy minimization using the CHARMm force field.[9] For each of the six ligands, geometric decoys had been generated for the LPDB using replica-based Langevin molecular dynamics with 25 ligand copies, in conjunction with a nuclear-Overhauser-effect-like restraint to distance the ligand from its native pose. The ligand was then energy-minimized to obtain its final decoy position. Over the six ligands, the decoys spanned a range of root-mean-square deviations (rmsd) in atomic positions of 0.03−15.80 Å. Between 27 and 50 decoys were obtained for each ligand.

A complicating factor in modeling of the ligand/HIV-1 protease interactions concerns the protonation states of the two catalytic aspartate groups (the Asp dyad) at the base of the active site. Potentially, there are nine combined protonation states of Asp25 and Asp25′: one dianionic, four monoanionic, and four neutral forms (Figure 1S, Supporting Information). In the absence of an inhibitor, NMR studies suggest that the aspartates are dianionic at a pH of 6.[30] However, in the presence of an inhibitor, the dyad can be either monoanionic or neutral, depending on the nature of the inhibitor and assay conditions.[31−34] DMP323, a non-peptide cyclic urea (CU)-based HIV-1 protease inhibitor, is similar to molecules **1**−**5** in this study (Table 1). NMR and X-ray evidence[31] on DMP323 argue that the aspartyl dyad is neutral with a hydrogen-bond network formed by the protonated Asp25/25′ carboxyl side chains and the diol groups of DMP323.[35] Conversely, on the basis of kinetic data[36] and following prior modeling studies of tipranavir,

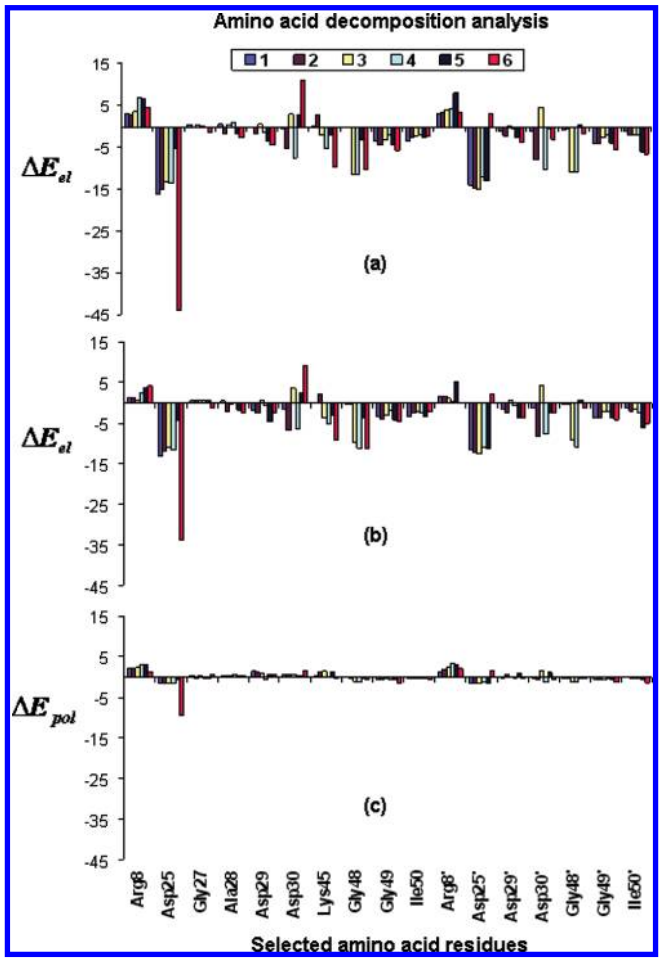


Figure 2. Electrostatic interaction energies (ΔE_{el}) of six HIV-1 protease native pose inhibitors with selected residues calculated at the (a) HF/6-31G*:AMBER and (b) AMBER levels and (c) associated polarization energy contributions (ΔE_{pol}) at the HF/6-31G*:AMBER level. Energies in kilocalories per mole. Negative energy values indicate stabilization of ligand binding.

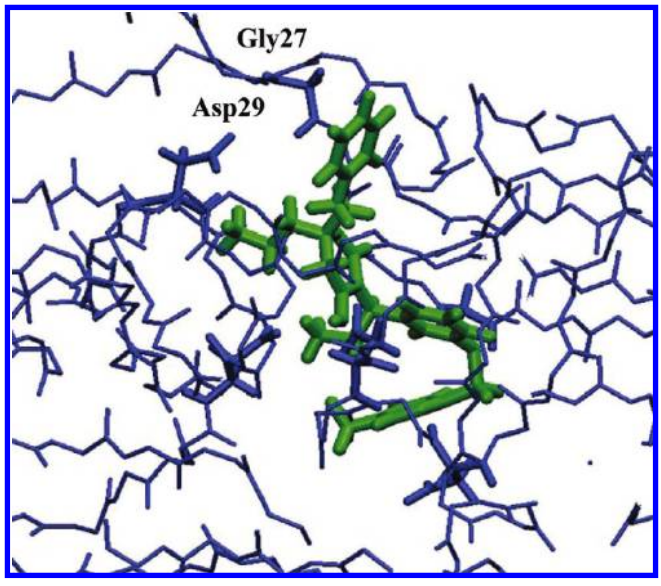


Figure 3. Pose of ligand **6** in the active site of HIV-1 protease.

6,[18,37] we adopt a monoanionic description of the Asp dyad in the presence of that ligand. Energy minimization studies using AMBER[8] with the *ff99* parameter set[38] and a generalized Born solvent model[39] were used to determine the optimal neutral and monoanionic form of the dyad in the presence of native-pose ligands **1**−**5** and **6**, respectively. Decoys were then assigned their native pose protonation configurations and similarly energy-minimized. We found a good correspondence in the ligand/dyad hydrogen-bond network derived for DMP323[31] with the network predicted here for **2**. These geometries were used subsequently for the evaluation of all scoring functions.

QM/MM Methodology. We use an electronic embedding[16,17] approach as implemented using the two-layer ONIOM[16,40,41] framework. A range of levels of theory was used. For treatment of the ligand at a QM level and protein at a MM level, the ligand was described using the ab initio HF/6-31G* and semiempirical PM3 and AM1d levels of theory; the latter method incorporates d orbitals on second row elements and is thus pertinent to sulfur-containing inhibitors **3**,[27] **5**,[28] and **6**.[29] We also explore a fourth ONIOM treatment where, in addition to the ligand, the active site is modeled quantum-mechanically. As is common practice,[42] the QM active site region was selected according to residues falling within a distance cutoff; here, we treat as QM all amino acids within 3.6 Å across all six ligands. This region encompasses residues Arg8, Asp25, Gly27, Ala28, Asp29, Asp30, Glu34, Lys45, Ile47, Gly48, Gly49, Ile50, Phe52, Pro81, Val82, Asn83, Ile84, Arg8′, Leu23′, Asp25′, Gly27′, Ala28′, Asp29′, Asp30′, Val32′, Gly48′, Gly49′, Ile50′, Pro81′, Val82′, and Ile84. It should be recognized that, due to the asymmetry of some ligands, this selection protocol leads to six residues being treated at a different level of theory than their symmetry-related counterparts. For a calculation of this size, a semiempirical QM approach was required, and we used the dispersion-corrected PM3-D method.[22] Here, dispersion effects are accounted for with an empirical function:[43]

$$E_{\text{PM3-D}} = E_{\text{PM3}} - s_6 \sum_i \sum_j \frac{C_6^{ij}}{R_{ij}^6} f_{\text{dmp}}(R_{ij}) \qquad (1)$$

The protein was modeled using AMBER[8] with the *ff99* parameter set.[38] Semiempirical QM/MM calculations were performed using the electronic structure package MUSE (Manchester University Semi-Empirical program).[44] Ab initio QM/MM calculations employed Gaussian 03.[45] QM/MM scores were computed according to $\Delta E_{\text{inter}} = E_{\text{complex}} - E_{\text{ligand}} - E_{\text{protein}}$. These were compared with GoldScore and ChemScore and a full force field approach. For the latter, ligand parameters and charges were determined with the *antechamber*[46] module of AMBER[47] based on the general atom force field[46] and AM1-BCC charge scheme.[48] The effect of a simple solvent correction was explored for the QM/MM and MM protein−ligand scoring functions using a recent implementation of the GB model[49] and Bondi radii.[39] The solvation energy corrections to ΔE_{inter} of each native and decoy protein−ligand pose (ΔE_{solv}) were calculated as the difference between the solvation energies of the complex, protein, and ligand at the AMBER level of theory. For the ChemScore function, we used the total free energy change on binding with inclusion of a clash penalty and internal torsional term.

Within our QM/MM scheme, the total ligand−protein interaction energy (ΔE_{inter}) consists of van der Waals (ΔE_{vdw}) and electrostatic (ΔE_{el}) contributions. The latter can be

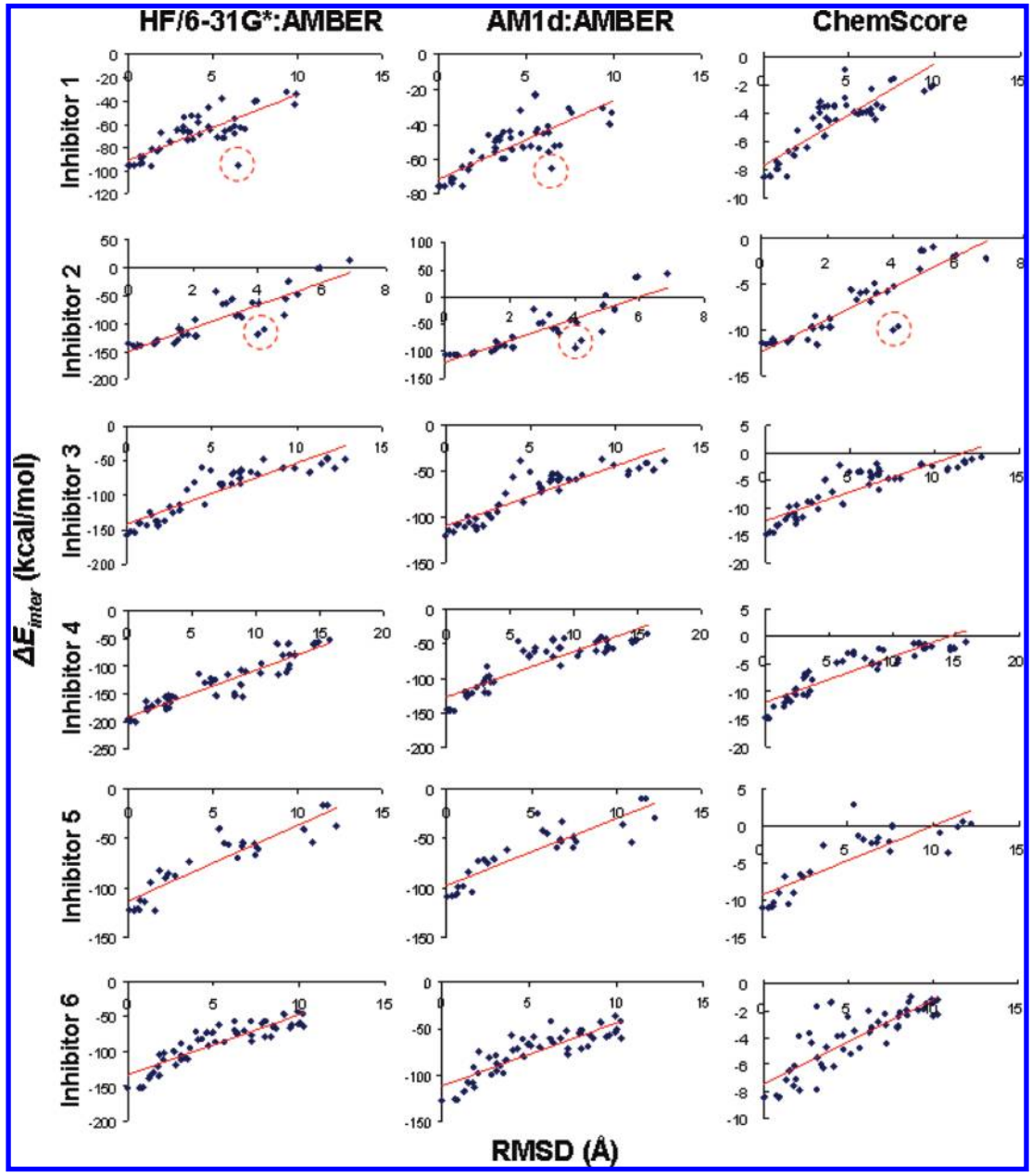


Figure 4. Interaction energies (ΔE_{inter}) of the six HIV-1 protease inhibitors' native poses and their geometric decoys. Columns from left to right: ΔE_{inter} calculated at HF/6-31G*:AMBER, AM1d:AMBER, and ChemScore levels of theory. Outlying decoy score for **1** is circled.

resolved[18,50] into a permanent electrostatic interaction energy component, ΔE_{perm} given by $\langle \Psi^0 | \hat{H}_{QM/MM} | \Psi^0 \rangle$, and a polarization energy component, ΔE_{pol}, defined as

$$\Delta E_{pol} = \langle \Psi | \hat{H}_{QM} + \hat{H}_{QM/MM} | \Psi \rangle - \langle \Psi^0 | \hat{H}_{QM} + \hat{H}_{QM/MM} | \Psi^0 \rangle \quad (2)$$

where Ψ and Ψ^0 refer the electronic wave function of the ligand in the protein environment and in the gas phase, respectively. This polarization energy can further be resolved into a distortion energy component for reorganizing the electron distribution of the ligand in the protein binding site (ΔE_{dist}) and a stabilization energy component of the ligand caused by this protein-polarized charge redistribution (ΔE_{stab}).[50] Additionally, we perform a per residue electrostatic interaction energy analysis, exploring individual amino acid contributions to the total ΔE_{el} and its polarization component (by deletion of MM point charges of selected residues).

Statistical Analysis. We use a combination of Z-score and discriminative power (DP) to evaluate the ability of the scoring functions to differentiate between well-docked and misdocked structures.[2,51] The Z-score, Z(E), for a given docked ligand pose is given as $(E - \bar{E})/\sigma$, where E is the interaction energy score of the given protein−ligand complex, $\bar{E}$ is the mean energy score over the set of docked ligand geometries, and σ is the standard deviation of the energy score distribution.

From this, the discriminative power of a given scoring function for a single protein−ligand complex (DP_i) and for N complexes (DP) is defined as

$$DP = \frac{1}{N}\sum_{i=1}^{N} DP_i = \frac{1}{N}\sum_{i=1}^{N} f_i(Z^{D}_{min} - Z^{M}_{min}) \quad (3)$$

where f_i is the fraction of the correctly docked structures with Z-scores lower than those of the lowest Z-score

Table 3. Total Discriminative Power (DP) and Individual DP_i for Scoring of HIV-1 Protease Complex *i* Using a Range of Approaches Including Total Interaction Energy (ΔE_{inter}) and Its Solvent-Corrected Value (ΔE^{aq}_{inter}) and Total Interaction Energy with Polarization Contribution Omitted ($\Delta E'_{inter}$) and Its Solvent-Corrected Value ($\Delta E'^{aq}_{inter}$)

		DP_i						
method	score	**1**	**2**	**3**	**4**	**5**	**6**	DP
HF/6-31G*:AMBER	ΔE_{inter}	−0.01	−0.40	−1.22	−1.97	−1.63	−2.04	−1.21
	$\Delta E'_{inter}$	−0.38	−0.40	−1.28	−1.90	−1.65	−2.08	−1.28
	ΔE^{aq}_{inter}	*−1.19*	*−0.53*	*−1.27*	*−2.08*	*−1.50*	*−2.44*	−1.50
	$\Delta E'^{aq}_{inter}$	*−1.51*	*−0.53*	*−1.35*	*−1.96*	*−1.49*	*−2.12*	−1.49
AM1d:AMBER	ΔE_{inter}	−0.91	−0.25	−1.25	−1.92	−1.62	−1.98	−1.32
	$\Delta E'_{inter}$	−1.31	−0.26	−1.30	−1.82	−1.58	−2.00	−1.38
	ΔE^{aq}_{inter}	*−1.29*	*−0.38*	*−1.39*	*−1.44*	*−1.18*	*−2.16*	−1.31
	$\Delta E'^{aq}_{inter}$	*−1.24*	*−0.40*	*−1.36*	*−1.25*	*−1.13*	*−1.73*	−1.19
PM3:AMBER	ΔE_{inter}	−0.57	−0.24	−1.30	−1.86	−1.65	−1.94	−1.26
	$\Delta E'_{inter}$	−1.10	−0.25	−1.34	−1.77	−1.63	−1.95	−1.34
	ΔE^{aq}_{inter}	*−1.26*	*−0.39*	*−1.37*	*−1.45*	*−1.17*	*−2.15*	−1.30
	$\Delta E'^{aq}_{inter}$	*−1.21*	*−0.41*	*−1.42*	*−1.18*	*−1.26*	*−1.81*	−1.22
PM3-D:AMBER[a]	ΔE_{inter}	−0.71	−0.26	−1.39	−1.87	−1.63	−2.03	−1.32
	ΔE^{aq}_{inter}	*−1.46*	*−0.44*	*−1.27*	*−1.37*	*−1.38*	*−2.21*	−1.36
AMBER	ΔE_{inter}	−0.17	−0.27	−1.29	−1.90	−1.69	−2.08	−1.23
	ΔE^{aq}_{inter}	*−1.41*	*−0.40*	*−1.37*	*−1.90*	*−1.73*	*−2.37*	−1.53
van der Waals	ΔE_{vdw}	−0.65	−0.20	−0.99	−1.35	−1.06	−0.93	−0.86
ChemScore	ΔG_{bind}	−1.83	−0.29	−1.24	−2.02	−1.73	−0.92	−1.34
GoldScore	score	−1.27	−0.52	−1.53	−1.43	−1.59	−1.44	−1.30
average		−1.03	−0.36	−1.31	−1.71	−1.49	−1.91	−1.30

[a] Includes active site residues in the QM region.

misdocked structure. Z^{D}_{min} and Z^{M}_{min} are the Z-scores for the lowest-energy structure from the cluster of docked and misdocked conformations, respectively. Misdocked conformations are defined as structures with a rmsd from the native of larger than 4 Å, following Vieth et al.[51] We compared definitions of correctly docked conformations as structures with a rmsd from the native of less than 1 or 2 Å.[52] However, as we find both definitions produce a similar trend, we only present the results of the 2 Å cutoff.

RESULTS AND DISCUSSION

We evaluate the performance of a range of scoring functions, including four QM/MM-based approaches, to correctly distinguish the native pose of six HIV-1 protease inhibitors (Table 1).

HIV-1 protease contains two identical 99-residue monomers which associate to form a cylindrical active site cavity with the aspartyl dyad at the base, a flap region containing Ile50/Ile50′ at the top, and large hydrophobic pockets to the sides (Figure 1). The active site of the protein is typically resolved into subsites (S) which interact with the amino acid side chains (P) of a putative substrate peptide.[53] Thus, subsites S1, S2, S3, and so forth are defined by HIV-1 protease residues which interact with substrate side chains P1, P2, P3, and so forth on the N-terminal side of the cleaved peptide bond between residues 1 and 1′. Likewise, the subsites S1′, S2′, S3′, and so forth interact with substrate side chains P1′, P2′, P3′, and so forth on the C-terminal side of the cleaved peptide bond between residues 1 and 1′. Given the homodimer's 2-fold axis of symmetry, residues of pockets S1 and S1′, S2 and S2′, and so forth are related by symmetry. The residues of these pockets are not mutually exclusive. Pockets S1/1′ and S2/2′ are quite hydrophobic in nature, although the S2/2′ is better able to accommodate polar moieties. The preference of S3/3′ is less well defined; beyond S3/3′, the complementary substrates begin to approach the limits of the protein's binding groove.[54]

Analysis of Native Binding Mode. We first consider the geometries, calculated interaction energies, and polarization contributions to binding of the native, crystallographic poses of each of ligands **1**−**6** (Table 1). Of these six inhibitors, symmetric ligands **1**−**4** and asymmetric ligand **5** share the same central CU ring structure with two pendant hydroxyl groups (Table 1). Ligands **1**−**4** possess phenyl groups in the analogous P1 and P1′ substrate positions, whereas **5** has two phenoxy groups. From **1** to **4**, the size of the P2 and P2′ groups progressively increases (Table 1). In the native pose of ligands **1**−**5**, the P1/1′ and P2/2′ substituents occupy the expected S1/1′ and S2/2′ subsites in HIV-1 protease; this is exemplified by the native poses of ligands **1** and **4** in the active site (Figure 1).

Accordingly, the total calculated protein−ligand interaction energies steadily increase with greater filling of subsites S2 and S2′. For ligands modeled at the HF/6-31G* level of theory, the calculated QM/MM ΔE_{inter} increases from −95.5 kcal/mol for **1** to −184.9 kcal/mol for **4** (Table 2). The value of ΔE_{inter} for ligand **5** (−122.1 kcal/mol) is close to that of **2** (−136.4 kcal/mol), which also contains four aromatic rings as substituents. Improved packing of larger P2/2′ groups into S2/2′ is also reflected by an increase in ΔE_{vdw} from −54.1 to −95.4 kcal/mol for **1** to **4** (Table 2). Concurrently, ΔE_{el} increases from −41.4 to −89.5 kcal/mol, with ligand **4** making eight hydrogen bonds more than **1**.[25] The corresponding polarization energy (ΔE_{pol}) and its stabilization/distortion energy contributions (which follow closely the linear response values) increase only slightly across the four ligands, approximately following the number of ligand aromatic rings (Table 2); similarly, ΔE_{pol} as a proportion of the total electrostatic interaction energy (ΔE_{el}) gradually increases from 12% for ligand **1** to 16% for **4** using HF theory (Table 2). This is close to the values of 12% and 18% found for **1** and **4** using AM1d to describe the ligand wave function (Table 1S, Supporting Information) but is less than the 15% and 20% for **1** and **4** obtained via PM3.

Molecular details of the interactions of inhibitors **1**−**6** are given by a per residue electrostatic interaction energy analysis of the active site amino acids (Figure 2a). The largest contribution to the interaction for **1**−**5** is from the two catalytic Asp residues, with ΔE_{el} values of −5.2 to −15.0 kcal/mol for Asp25 at the HF/6-31G*:AMBER level, due to hydrogen bonding to the CU ring hydroxyls. The CU carbonyl group displays interactions with the flap residues (Gly48/48′, Gly49/49′, and Ile50/50′), to an approximately equal extent for the five inhibitors. Discrimination appears to occur at residues Asp29/29′ and Asp30/30′ of the S2 and S2′ subsites; as expected, these residues interact to different extents with the differing P2/2′ groups of **1**−**5**. Interestingly, ligands **2** and **4** interact strongly with Asp30/30′ and ligands **1**, **3**, and **5** less so. For **1**, S2 is insufficiently occupied for the ligand to fully interact with Asp30/30′, whereas the N-aniline P2/2′ groups of **2** have interactions with Asp30/30′. For ligand **3**, the sulfur lone pair repels the carboxyl oxygens of Asp30/30′; this is relieved by replacing the thiazole ring of **3** with the benzimidazole ring of **4**, such that the NH group is involved in an extra hydrogen bond.[25] After inclusion of the solvent effects, all levels of theory, with the exception of the PM3-D active site model, predict

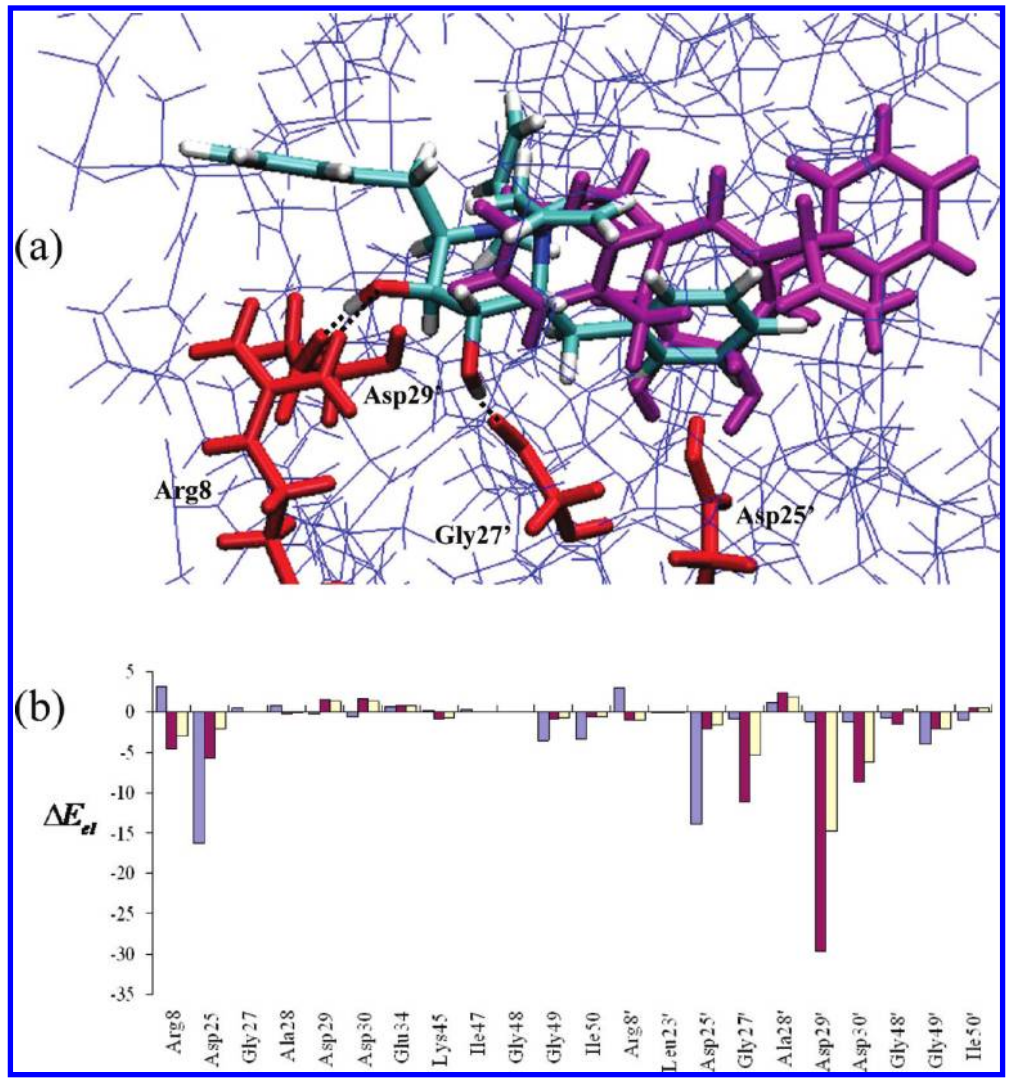


Figure 5. (a) Major active site interaction of inhibitor **1** in its native pose (purple) and as a decoy pose with rmsd = 6.50 Å. (b) Residue contributions (kcal/mol) to total electrostatic interaction ΔE_{el} of inhibitor **1** in its native (blue) and decoy poses (magenta) at the HF/6-31G*:AMBER level of theory, and for the decoy at the AM1d level (yellow).

3 and **4** to be approximately equal in interaction energy (Table 2S, Supporting Information). The origin of this correction is the penalization of additional buried lipophilic surface area of **4**. Thus, in terms of ligand efficiency, Δe_{inter} (Table 2), **4** is the same efficiency as **3**, with binding efficiencies of −2.98 and −2.90 kcal/mol, respectively. Interestingly, both are less efficient binders than **1** and **2**. We also note that the interaction of Arg8/8′ is repulsive across all ligands, and in particular for **5**. This feature arises from repulsion of the positively charged guanidine group of Arg8/8′ and the protons of the aromatic P1/1′ group (which, as a phenoxy moiety, is slightly more prominent for **5**).

Inhibitor **6** has a scaffold distinct from those of **1**−**5**, based on a central 4-hydroxy-dihydropyridone moiety (Table 1). In the native pose of **6**, the pendant aliphatic and aromatic groups occupy S1/1′, S2/2′, and S3 subsites (Figure 3). As for the other five ligands, the central ring carbonyl function displaces a bound water in the flap region, and the catalytic Asp side chains interact with the central ring hydroxyl group(s). However, as distinct from **1**−**5**, HIV-1 protease is thought to have a monoanionic dyad when ligand **6** is bound. The calculated greater strength of interaction of Asp25 with **6** is clearly shown, being of the order of −43 kcal/mol at the HF level (Figure 2a). Of this, 21% is due to ΔE_{pol} (Figure 2c). ΔE_{inter} is predicted to be only −34 kcal/mol via AMBER, which therefore may be lacking a full account of this contribution (Figure 2b). However, in general over the six ligands, AMBER reproduces the amino acid interaction energy analysis of the HF method rather well (Figure 2a and b).

The complex of ligand **6** with HIV-1 protease has been studied previously[18] at the AM1/CHARMM level,[55] using molecular dynamics and an explicit solvent. The protein was similarly modeled with monoanionic Asp25/25′. The study calculated ΔE_{el} to be −62.6 kcal/mol. We find here a similar estimate of ΔE_{el} at the AM1d:AMBER level of −58.8 kcal/mol. Both values are somewhat less than the value of −84.8 kcal/mol we obtain at the HF/6-31G*:AMBER level (Table 2). Whereas a previous study observed a polarization contribution to binding of **6** of 37%,[18] we find the $\Delta E_{pol}/\Delta E_{el}$ ratio to be 23% at AM1d and PM3 levels, and somewhat smaller (18%) at the HF level (Table 1S, Supporting Information). The difference may lie in part in the results of ensemble averaging, which could lead to contributions from configurations with suboptimal electrostatic interactions. For these structures, it is likely that ΔE_{pol} becomes a more substantial proportion of ΔE_{el}; this is the case for decoy orientations studied here, as discussed below.

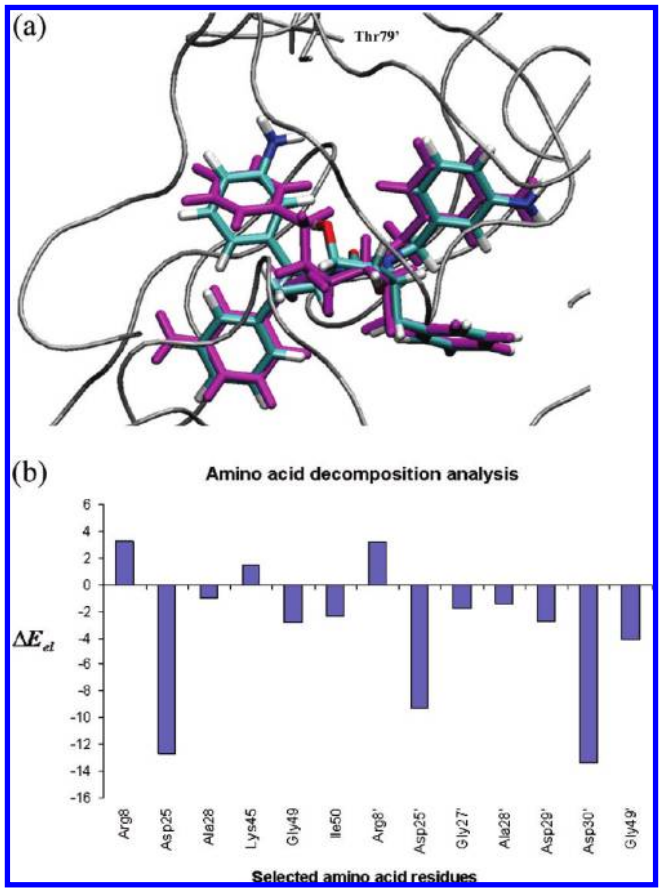


Figure 6. (a) Major active site interaction of inhibitor **2** in its native pose (purple) and as a geometric decoy with rmsd = 4.01 Å. (b) Residue contributions (kcal/mol) to total electrostatic interaction ΔE_{el} of inhibitor **2** in the decoy pose at the HF/6-31G*:AMBER level of theory.

To conclude, the polarization contribution to binding for the QM/MM potentials is non-negligible (12−18% of total electrostatic interaction). However, it varies only modestly with ligand polarizability, by 10 kcal/mol over the six ligands (Table 2). The PM3 Hamiltonian appears to overestimate the $\Delta E_{pol}/\Delta E_{el}$ ratio relative to HF/6-31G*, but AM1d is in closer agreement. Although it would be useful to assess the ability of the scoring functions to rank ligands in order of potency, valuable in a virtual screening campaign, this would not be appropriate without inclusion of, for example, entropic effects of bound water displacement.

Detection of Native Binding Mode. We now turn to consider the ability of these QM/MM, MM, ChemScore and GoldScore scoring functions to distinguish the native from the decoy pose for each HIV-1 protease inhibitor. For the QM/MM functions, we evaluate the role of polarization effects on their predictive ability.

At the simplest level of discrimination, we can ask if the native pose has the most negative interaction energy (ΔE_{inter}) among a set of decoys. All approaches in this study predict the native pose or its decoys with a rmsd of less than 2 Å to have the most favorable interaction energy (Table 3S, Supporting Information), as evident from calculated topologies of the HF/6-31G*:AMBER, AM1d:AMBER, and ChemScore binding surfaces (Figure 4). This is also observed using simply the van der Waals contribution of the QM/MM and MM functions, ΔE_{vdw} (Table 3S, Supporting Information).

Furthermore, the interaction energy surfaces also reflect the funnel shape expected for protein−ligand binding (Figure 4). The correlation coefficient for these surfaces provides a crude estimate of discrimination between low- and high-rmsd poses (Table 4S, Supporting Information). All potentials perform well across the six ligands, with an average r of 0.89 at the HF level. There is a similar correlation with the inclusion of solvent effects via the generalized Born model[56] (r of 0.85 for HF).

A more sensitive measure of discrimination is afforded by discriminative power, DP (eq 3). Here, we observe greater variation across ligands and scoring methods (Table 3). To assess the importance of electrostatic contributions to pose prediction, we apply solely the ΔE_{vdw} contribution to scoring and find it the least sensitive discriminator of all potentials (DP of −0.86). Thus, for the QM/MM and MM approaches, the electrostatic models applied in conjunction with a van der Waals term improve discrimination for each of the six ligands. Interestingly, there is little difference in predicted DP for in vacuo semiempirical QM (with a DP range from −1.26 to −1.32), ab initio QM (−1.21), and force field (−1.23) models of ligands (Table 3). It does also appear, at least according to DP, that there is not a major benefit in including the 31 active site residues quantum mechanically, in combination with the dispersion correction; the use of PM3-D for ligands and active sites leads to a DP of −1.32 (Table 3).

Considering the variation in DP with the ligand, it appears that ligands **1** and **2** are treated least satisfactorily by the QM/MM and MM methods, with low average DP values of −0.47 and −0.28, respectively, for these ligands in the gas phase. The treatment of **2** is also poor using the ChemScore and GoldScore scoring functions, with DPs of −0.29 and −0.52, respectively (Table 3). DP_1, the calculated DP of ligand **1**, via HF/6-31G*:AMBER is −0.01 and via the AMBER force field is −0.17 (Table 3); this reflects nearly no discrimination of the native pose in the gas phase. This failure for ligand **1** is notably worse than for the semiempirical QM/MM approaches and appears to mask a generally superior performance of the ab initio QM/MM and force field scoring functions in discriminating native from non-native docked conformations for the remaining five ligands.

For ligand **1**, the van der Waals term actually provides an improved DP_1 of −0.65 compared to those for HF/6-31G*:AMBER and AMBER (Table 3). The degradation of DP_1 due to the electrostatic model appears to stem from both ΔE_{pol} and ΔE_{perm} contributions. To more fully understand the reasons for this, we recalculate DP_1 using $\Delta E'_{inter}$, where ΔE_{pol} is omitted from ΔE_{inter}. This leads to an improved DP_1 at the HF level of −0.38 (Table 3). This contrasts with a DP_i using $\Delta E'_{inter}$ for **2**−**6**, which is approximately similar to the use of ΔE_{inter}. To understand this, it is useful to consider the degree of overlap of the native and misdocked distributions, as estimated by f_i in eq 3 (Table 5S, Supporting Information). It is evident that there is considerable overlap of native and non-native distributions at these levels of theory, with respective f_1 values of 0.44 and 0.63 for HF/6-31G*:AMBER and AMBER (relative to 0.81−0.94 across the semiempirical approaches).

This overlap and the poor DP for ligand **1** arise from an unusually strongly interacting, high rmsd decoy pose: this misdocked pose (marked as an outlier, Figure 4) has a rmsd of

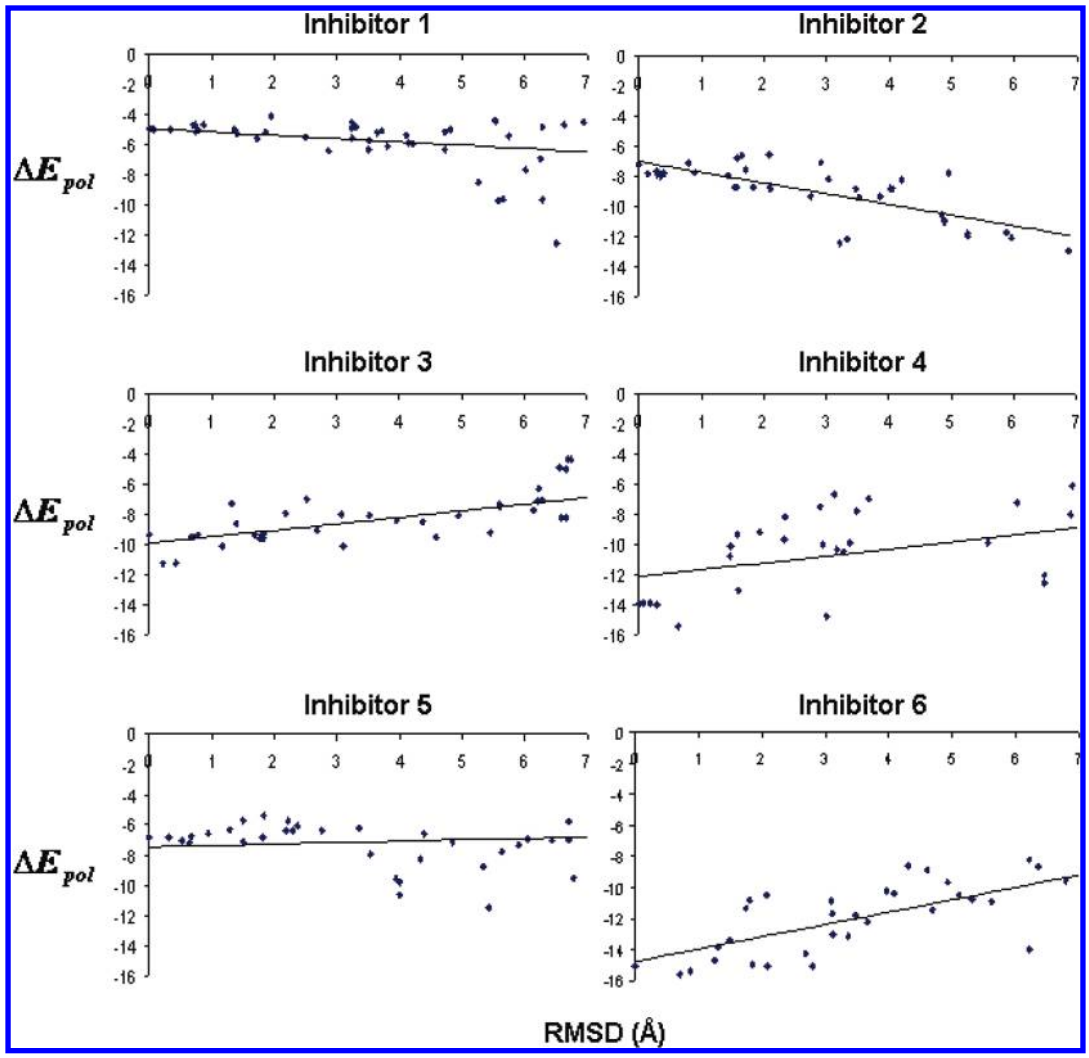


Figure 7. Polarization energy contribution (ΔE_{pol}) to binding of the six HIV-1 protease inhibitors and their geometric decoys with a rmsd less than 7 Å (kcal/mol). All calculations were performed at the HF/6-31G*:AMBER level of theory.

Table 4. Correlation Coefficient, r, between ΔE_{pol} at the HF/6-31G*:AMBER Level and rmsd of Pose, Polarization Energy Contribution to Binding Averaged over Native and All Decoy Poses, $\langle\Delta E_{pol}\rangle$, at This Level of Theory, and Difference in Total Interaction Energy Averaged over Native and All Decoy Poses, $\langle\Delta\Delta E_{inter}\rangle$, between the HF/6-31G*:AMBER and AMBER Levels of Theory[a]

	inhibitor					
	1	**2**	**3**	**4**	**5**	**6**
r	−0.35	−0.75	0.80	0.70	0.21	0.85
$\langle\Delta\Delta E_{inter}\rangle$	−6.3 (2.6)	−8.9 (3.1)	−8.1 (6.1)	−9.3 (5.2)	−5.1 (1.8)	−10.8 (3.8)
$\langle\Delta E_{pol}\rangle$	−5.8 (1.6)	−9.0 (1.8)	−7.6 (2.0)	−8.9 (3.1)	−6.6 (1.2)	−10.5 (2.9)

[a] All energies in kilocalories per mole. Standard deviations in parentheses.

6.50 Å and an interaction energy of −95.3 kcal/mol at the HF/6-31G*:AMBER level, compared to −95.5 kcal/mol for the native pose. This decoy pose forms a quite distinct but nevertheless favorable orientation in the active site of HIV-1. It forms strong hydrogen bonds from its two CU ring OH groups to Arg8 and Asp29′ side chains and the backbone carbonyl of Gly27′ (Figure 5a); the involvement of these residues is reflected in a per residue ΔE_{el} analysis, with favorable contributions for all three amino acids (Figure 5b). These interactions replace the interactions with the active site Asp25 and Asp25′, which now interact with the P1 phenyl ring. The second phenyl ring in this decoy pose presents as largely noninteracting, projecting out of the pocket.

By contrast, AM1d:AMBER and PM3:AMBER functions predict the native pose of ligand **1** rather better, and when ΔE_{pol} is omitted to give $\Delta E'_{inter}$, a reasonable DP of −1.31 and −1.10 is found for AM1d and PM3, respectively (Table 3). This appears to arise from underestimation at these levels of theory of the interaction of the decoy pose with HIV-1 protease, as indicated by the per residue ΔE_{el} analysis at AM1d relative to HF/6-31G* (Figure 5b). Thus, for Gly27′, Asp29′, and Asp30′, there is notable underestimation of residue interactions at the AM1d level. Also interesting is the 3−4 kcal/mol underestimate of the favorable interaction of the carboxylate side chain of Asp25 with the phenyl ring of **1**.

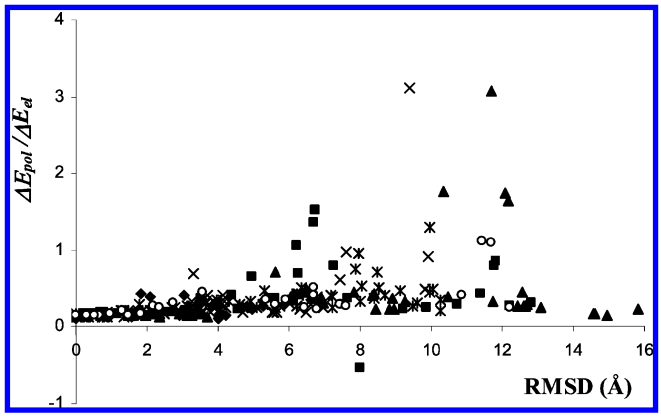


Figure 8. $\Delta E_{pol}/\Delta E_{el}$ at the HF/6-31G*:AMBER level as a function of rmsd from the native pose, for ligands **1** (cross), **2** (diamond), **3** (square), **4** (triangle), **5** (circle), and **6** (asterisk).

The addition of a GB solvent correction to the AMBER and HF/6-31G*:AMBER functions markedly reduces the binding energy of this decoy—the change in solvation free energy on binding of this decoy is 66.9 kcal/mol, as opposed to an average solvation penalty of 34.2 kcal/mol for nativelike conformers. This is in accord with burial of the decoy's hydrophobic groups next to the Asp dyad. Thus, on solvation, the DP improves at the HF/6-31G*:AMBER and AMBER levels from −0.01 and −0.17 to −1.19 and −1.41, respectively. The improved treatment of ligand **1** then leads to an overall superior performance of HF/6-31G*:AMBER and AMBER across the six ligands compared to the semiempirical QM/MM approaches. For the semiempirical QM/MM levels of theory, the inclusion of solvent effects appears to have a relatively small impact on the average DP for ligands **1**−**6**.

For the docking of ligand **2**, the performance of all methods is least satisfactory, with an average DP of −0.36, relative to an average DP of greater than −1.0 for the other ligands (Table 3). Similar results are found with or without including ΔE_{pol} in the QM/MM calculation of interaction energy. Ligand **2** is somewhat larger than **1**, and for the six ligands, the P1/1′ substituents of ligand **2** have the greatest similarity in size and shape to those of P2/2′. The highly scoring decoy here, with a rmsd of 4.01 Å, features as an outlier on the binding landscape (Figure 4). The ligand's conformation involves an alternative pucker of the CU ring, leading to occupancy of a similar volume in the active site and similar interactions (Figure 6a). The decoy CU ring pucker orients the two pendant hydroxyls to form a less ideal hydrogen-bonding network with the protonated Asp25 and Asp25′ residues: a per residue energy analysis indicates the ligand electrostatic interaction with these two residues to reduce by 2 and 6 kcal/mol, respectively (Figure 6b), relative to the native pose (Figure 2). This size similarity of P1/1′ and P2/2′ substituents for **2** is also reflected by a flat binding funnel from native up to 2 Å, and low-energy decoys at around 4 Å (Figure 4).

However, on solvation via the GB model, there is a modest improvement for **2** in DP, obtained across all levels of theory (Table 3). This observation appears linked with the interchange of the functional groups filling one of the protease's large S1 hydrophobic pockets—for the native pose, it is the phenyl group; for the highly scored 4.01 Å decoy, S1 is filled by the aniline ring (Figure 6a). Thus, burial of the more polar aniline in a hydrophobic environment does not appear adequately compensated by a weak polar interaction of the ligand aniline NH_2 group with Thr179 (a O···H distance of 2.90 Å, Figure 6a). However, even with solvent correction, this decoy pose scores highly: at the HF/6-31G*:AMBER level, the in vacuo ΔE_{inter} of −118.2 kcal/mol for this decoy pose is reduced to −55.8 kcal/mol in solution. This compares to an average ΔE_{inter} for misdocked structures of −54.3 kcal/mol (gas phase) and −1.8 kcal/mol (solution) and, for the native pose (Table 3), an ΔE_{inter} of −136.4 kcal/mol (gas-phase) and −77.9 kcal/mol (solution). The high score of this decoy, and its poor discrimination from the native binding mode, appears to be a consequence of its near-symmetric shape and degenerate interactions. Although one might be tempted to speculate that this pose could constitute a genuine alternative binding mode to HIV-1 protease, we find that the optimized native and decoy ligand conformations differ by 15.7 kcal/mol in internal energy at the HF/6-31G* level. Correspondingly, only one bound ligand pose is witnessed in the X-ray structure.[25]

Unlike ligand **1**, a similar level of discrimination is found for **2** with or without inclusion of ΔE_{pol} in the calculation of ΔE_{inter} (Table 3). Indeed, for ligands **2**−**6**, the gas-phase DP is very similar with or without ΔE_{pol}. The same is true for a solvent-corrected DP with (ΔE^{aq}_{inter}) or without ΔE_{pol} ($\Delta E'^{aq}_{inter}$, Table 3). This would suggest that there is not a large benefit here in incorporating ligand polarizability (by using a QM/MM model) and indeed could be to some extent detrimental to identification of the native pose via docking. However, although the omission of ΔE_{pol} does improve the scoring of **1**, the most substantial improvement is gained through application of a solvent model, and the benefits of the solvent model are felt equally by a force field approach (the MM ΔE_{inter}, it can be argued, equates to a calculation of ΔE_{perm}) and an HF/6-31G*:AMBER approach. We note that this finding differs somewhat from a previous work[11] where ligand polarization was incorporated. However, in that study, polarization was treated indirectly: prior to docking with an empirical potential, ligand partial point charges were first derived for the native pose in the presence of a protein receptor (from fitting to the QM/MM electrostatic potential).

Polarization Energy Contributions. Undeniably, ΔE_{pol} is a non-negligible contribution to ΔE_{el}: for poses of the six ligands with a rmsd from the native of less than 2 Å, this contribution is 16% at the HF/6-31G* level of theory, 21% at PM3, and 18% at AM1d. As for the calculated interaction energy of the native pose, PM3 appears to overestimate this contribution. Interestingly, for non-native poses of less than 4 Å rmsd, this contribution at the HF level rises from 16% to 19%. The larger contribution of ΔE_{pol} with the inclusion of higher rmsd poses arises in the main from poorer non-native electrostatic interactions and thus a smaller ΔE_{perm} contribution to binding. This may partly explain the large $\Delta E_{pol}/\Delta E_{el}$ of ∼38% for HIV-1 protease ligand interactions found in the AM1/CHARMm molecular dynamics study of Hensen et al.;[18] there, the MD ensemble will necessarily sample some suboptimal protein−ligand contacts in the course of its trajectory. We find here that $\Delta E_{pol}/\Delta E_{el}$ at the AM1d level rises from 18% for poses of rmsd < 2 Å to 26% for poses < 4 Å.

The underlying reason for the lack of improvement in DP by inclusion of ΔE_{pol}, and thus (i) the similarity of scoring

via AMBER and HF/6-31G*:AMBER and (ii) the similarity of scoring via ΔE_{perm} and ΔE_{el} at the HF/6-31G*:AMBER level, appears to lie in the relative insensitivity of ΔE_{pol} to decoy orientation. Thus, the profile of ΔE_{pol} against rmsd is fairly flat (Figure 7), and the average ΔE_{pol} contribution to ligand interaction energies ($<\Delta E_{pol}>$ in Table 4) has a relatively small standard deviation of 1−3 kcal/mol. The greatest discrimination of a native from a non-native pose is for **6** and is likely a result of favoring of the pose which achieves the closest proximity to the charged Asp25 side chain. For ligands **1** and **2** (Figure 7), the correlation coefficients are negative, with values of −0.35 and −0.75, respectively (Table 4), indicating the more favorable polarization energies for decoys due to the non-native electrostatic contacts that these decoys make. For example, the rmsd 6.50 Å decoy for ligand **1** has a contribution from ΔE_{pol} of −13 kcal/mol, arising mainly from Gly27′ and the charged Asp29′ (Figure 5b). This is rather larger than that for the native pose, which interacts with the neutral Asp25/25′ groups and has an ΔE_{pol} contribution of −5 kcal/mol (Table 2).

Corresponding to the rather flat profile of ΔE_{pol} against rmsd, the ratio of ΔE_{pol} to ΔE_{el} is quite constant as well, up to about 4 Å (Figure 8). After this fairly distinct cutoff for **2**−**6**, there is dispersion of the ΔE_{pol} to ΔE_{el} ratio. As the ΔE_{pol} contributions remain rather constant as a function of rmsd (Table 4, Figure 7), this effect arises mainly from ΔE_{perm}, as the electrostatic interaction between the gas-phase ligand wave function and the protein partial charges becomes increasingly less favorable with increasing rmsd; indeed, for the rmsd 7.99 Å decoy of **3**, the ratio becomes negative (Figure 8).

Consequently, ΔE_{el} calculated via the AMBER force field mimics the orientation dependence of HF/6-31G* without explicit ligand polarization; both ab initio and semiempirical QM descriptions of the ligand lead to energetic trends similar to those for AMBER, with a correlation coefficient of 0.97−1.00 across ligands for ΔE_{inter} at the PM3, AM1d, and HF/6-31G* levels (Table 6S, Supporting Information). In terms of absolute interaction energies, AMBER systematically underestimates the interaction energy of the ligand with the protein at the HF/6-31G*:AMBER level; this average difference over the native and decoy poses ($<\Delta\Delta E_{inter}>$ in Table 4) ranges from −5.1 kcal/mol for **5** to −10.8 kcal/mol for **6** (Table 4). Interestingly, these discrepancies correspond closely to the average value of ΔE_{pol} for each ligand (Table 4); this is somewhat expected given that the electrostatic model of the AMBER force field is fitted to the electrostatic potential of Ψ^0, and its electrostatic interaction corresponds in essence to ΔE_{perm}.

Thus, we find here that the most discriminative methods have a DP of around −1.5 over the six HIV-1 protease inhibitors. The best QM/MM scoring function was the solvent-corrected HF/6-31G*:AMBER model (Table 3). This is an improvement on the performance of the empirical potential, ChemScore, which has a DP of −1.34. Interestingly, a study of 52 aspartic protease complexes for nine different scoring functions, which used the same criteria to calculate DP, observed a modestly improved performance of ChemScore relative to this study (DP of −1.45).[2] The best performing functions overall in that study were the CHARMm force field and an AMBER-based approach,[57] with DPs of −1.66 and −1.69, respectively (QM/MM functions were not considered). Both of these functions contained a distance-dependent solvent correction. Knowledge-based potentials also featured in this study and did less well (DP typically below −1.0).

Crucially, in that work,[2] the authors chose uniformly to model the active site aspartyl dyad of HIV-1 protease in the monoanionic form. For ligand **6**, where we model the dyad as monoanionic, our study finds the best gas-phase DP at the HF/6-31G*:AMBER level of −2.04 (Table 3). However, modeled as the neutral form, the DP for **6** drops to −1.45 (Table 7S, Supporting Information). Conversely, the average DP for ligands **1**−**5** increases from −1.12 at the gas-phase HF/6-31G* level to −1.29 if the dyad is modeled as monoanionic (Table 7S, Supporting Information).

A recent study[58] examined the effect of the tautomer and ionization state on docked poses of native ligands; for the three protein−ligand systems studied, binding modes and underlying energy landscapes appeared relatively insensitive to changes of the protonation and tautomeric state. Thus, the native ligands were able to satisfy hydrogen-bond demands of the active site in various ways. Here, the qualitative features of the binding surface are similar for the monoanionic and neutral aspartyl dyads. However, a stronger interaction of ligand OH groups with a monoanionic over a neutral dyad increases the selectivity for native-like poses that satisfy this motif, resulting in a larger discriminatory power for the monoanionic dyad. This is clearly illustrated for ligands **1**−**5**, by comparison of the binding funnels for docking to monoanionic and neutral HIV-1 protease (Figure 4 and Figure 2S, Supporting Information), where the funnels of the monoanionic state are steeper, and from the respective per residue ΔE_{el} analyses of amino acids (Figure 2 and Figure 3S, Supporting Information), showing more favorable dyad−ligand interactions for the monoanionic state.

CONCLUSIONS

We examine the ability of QM/MM, MM, ChemScore and GoldScore scoring functions to discriminate native from non-native poses in six HIV-1 protease complexes. The best QM/MM scoring function was the solvent-corrected HF/6-31G*:AMBER model with a DP of around −1.5 over the six HIV-1 protease inhibitors, exceeding the performance of the ChemScore and GoldScore scoring functions. This QM/MM function, however, performed comparably to a solvent-corrected MM scoring function. Indeed, the ΔE_{pol} component of the QM/MM potentials, although comprising 12−18% of the total protein−ligand electrostatic interaction for the native pose, appears relatively insensitive to changes in rmsd from the native pose. For one ligand, ΔE_{pol} actually impeded accurate discrimination of native from non-native poses. This stabilization of certain decoy poses by electric polarization may be physically correct but distorts the smooth funnel-shaped topology of the binding surface required for efficient docking. Correspondingly, the MM approach, which omits explicit ligand polarization, performed similarly to the Hartree−Fock QM/MM model. Semiempirical QM/MM methods and underestimated interaction of native and decoy poses relative to ab initio QM/MM and exhibited lower discriminative power.

A solvent correction improved discrimination of the native poses by MM and QM/MM functions. The GB solvent

correction applied here was derived in the presence of the fixed partial point charges of the AMBER force field. It is possible that a more sophisticated treatment of the solvent for the QM/MM systems, for example, a coupled generalized Born−QM/MM treatment,[59] could improve discrimination of the ligand poses further. Finally, we observe that, for the receptor HIV-1 protease, the choice of protonation state of the aspartyl dyad affected the ability of scoring functions to discriminate the native binding pose.

ACKNOWLEDGMENT

We thank the EPSRC for funding.

Supporting Information Available: Semiempirical QM/MM calculations of ΔE_{pol}; rmsd of top-ranked poses for all methods and ligands; correlation of rmsd with ΔE_{inter}; fraction of correctly docked poses; correlation of QM/MM to MM methods; and discriminative power, interaction energy profiles, and per residue electrostatic interaction energy analysis for alternative dyad protonation states. 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. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: Indianapolis, IN, 2002; Vol. 18, pp 41−87.

(2) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing Scoring Functions for Protein-Ligand Interactions. *J. Med. Chem.* **2004**, *47*, 3032–3047.

(3) Wang, R. X.; Lu, Y. P.; Wang, S. M. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.

(4) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791–804.

(5) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.

(6) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.

(7) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 609–623.

(8) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(9) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(10) Beierlein, F.; Lanig, H.; Schurer, G.; Horn, A. H. C.; Clark, T. Quantum Mechanical/Molecular Mechanical (QM/MM) Docking: an Evaluation for Known Test Systems. *Mol. Phys.* **2003**, *101*, 2469–2480.

(11) Friesner, R. A. Modeling Polarization in Proteins and Protein-Ligand Complexes: Methods and Preliminary Results. *Adv. Protein Chem.* **2006**, *72*, 79–104.

(12) Grater, F.; Schwarzl, S. M.; Dejaegere, A.; Fischer, S.; Smith, J. C. Protein/Ligand Binding Free Energies Calculated With Quantum Mechanics/Molecular Mechanics. *J. Phys. Chem. B* **2005**, *109*, 10474–10483.

(13) Khandelwal, A.; Balaz, S. Improved Estimation of Ligand-Macromolecule Binding Affinities by Linear Response Approach Using a Combination of Multi-Mode MD Simulation and QM/MM Methods. *J. Comput.-Aided. Mol. Des.* **2007**, *21*, 131–137.

(14) Khandelwal, A.; Lukacova, V.; Comez, D.; Kroll, D. M.; Raha, S.; Balaz, S. A Combination of Docking, QM/MM Methods, and MD Simulation for Binding Affinity Estimation of Metalloprotein Ligands. *J. Med. Chem.* **2005**, *48*, 5437–5447.

(15) Patel, S.; Brooks, C. L. CHARMM Fluctuating Charge Force Field for Proteins: I Parameterization and Application to Bulk Organic Liquid Simulations. *J. Comput. Chem.* **2004**, *25*, 1–15.

(16) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods With Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.

(17) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go From Here. *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(18) Hensen, C.; Hermann, J. C.; Nam, K. H.; Ma, S. H.; Gao, J. L.; Holtje, H. D. A Combined QM/MM Approach to Protein-Ligand Interactions: Polarization Effects of the HIV-1 Protease on Selected High Affinity Inhibitors. *J. Med. Chem.* **2004**, *47*, 6673–6680.

(19) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. L. Importance of Substrate and Cofactor Polarization in the Active Site of Dihyrofolate Reductase. *J. Comput. Chem.* **2003**, *24*, 177–190.

(20) Winget, P.; Horn, A. H. C.; Selcuki, C.; Martin, B.; Clark, T. AM1* Parameters for Phosphorus, Sulfur and Chlorine. *J. Mol. Model.* **2003**, *9*, 408–414.

(21) Stewart, J. J. P. Optimisation of Parameters for Semi-Empirical Methods. 2. Applications. *J. Comput. Chem.* **1989**, *10*, 221–264.

(22) McNamara, J. P.; Hillier, I. H. Semi-Empirical Molecular Orbital Methods Including Dispersion Corrections for the Accurate Prediction of the Full Range of Intermolecular Interactions in Biomolecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362–2370.

(23) 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*, 727–748.

(24) Roche, O.; Kiyama, R.; Brooks, C. L. Ligand-Protein DataBase: Linking Protein-Ligand Complex Structures to Binding Data. *J. Med. Chem.* **2001**, *44*, 3592–3598.

(25) Ala, P. J.; DeLoskey, R. J.; Huston, E. E.; Jadhav, P. K.; Lam, P. Y. S.; Eyermann, C. J.; Hodge, C. N.; Schadt, M. C.; Lewandowski, F. A.; Weber, P. C.; Mccabe, D. D.; Duke, J. L.; Chang, C. H. Molecular Recognition of Cyclic Urea HIV-1 Protease Inhibitors. *J. Biol. Chem.* **1998**, *273*, 12325–12331.

(26) Hodge, C. N.; Aldrich, P. E.; Bacheler, L. T.; Chang, C. H.; Eyermann, C. J.; Garber, S.; Grubb, M.; Jackson, D. A.; Jadhav, P. K.; Korant, B.; Lam, P. Y. S.; Maurin, M. B.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Reid, C.; Sharpe, T. R.; Shum, L.; Winslow, D. L.; EricksonViitanen, S. Improved Cyclic Urea Inhibitors of the HIV-I Protease: Synthesis, Potency, Resistance Profile, Human Pharmacokinetics and X-Ray Crystal Structure of DMP 450. *Chem. Biol.* **1996**, *3*, 301–314.

(27) Ala, P. J.; Huston, E. E.; Klabe, R. M.; Jadhav, P. K.; Lam, P. Y. S.; Chang, C. H. Counteracting HIV-1 Protease Drug Resistance: Structural Analysis of Mutant Proteases Complexed With XV638 and SD146, Cyclic Urea Amides With Broad Specificities. *Biochemistry* **1998**, *37*, 15042–15049.

(28) Schaal, W.; Karlsson, A.; Ahlsen, G.; Lindberg, J.; Andersson, H. O.; Danielson, U. H.; Classon, B.; Unge, T.; Samuelsson, B.; Hulten, J.; Hallberg, A.; Karlen, A. Synthesis and Comparative Molecular Field Analysis (CoMFA) of Symmetric and Nonsymmetric Cyclic Sulfamide HIV-1 Protease Inhibitors. *J. Med. Chem.* **2001**, *44*, 155–169.

(29) Thaisrivongs, S.; Skulnick, H. I.; Turner, S. R.; Strohbach, J. W.; Tommasi, R. A.; Johnson, P. D.; Aristoff, P. A.; Judge, T. M.; Gammill, R. B.; Morris, J. K.; Romines, K. R.; Chrusciel, R. A.; Hinshaw, R. R.; Chong, K. T.; Tarpley, W. G.; Poppe, S. M.; Slade, D. E.; Lynn, J. C.; Horng, M. M.; Tomich, P. K.; Seest, E. P.; Dolak, L. A.; Howe, W. J.; Howard, G. M.; Schwende, F. J.; Toth, L. N.; Padbury, G. E.; Wilson, G. J.; Shiou, L. H.; Zipp, G. L.; Wilkinson, K. F.; Rush, B. D.; Ruwart, M. J.; Koeplinger, K. A.; Zhao, Z. Y.; Cole, S.; Zaya, R. M.; Kakuk, T. J.; Janakiraman, M. N.; Watenpaugh, K. D. Structure-Based Design of HIV Protease Inhibitors: Sulfonamide-Containing 5,6-Dihydro-4-Hydroxy-2-Pyrones As Non-Peptidic Inhibitors. *J. Med. Chem.* **1996**, *39*, 4349–4353.

(30) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. H. Ionization States of the Catalytic Residues in HIV-1 Protease. *Nat. Struct. Biol.* **1996**, *3*, 946–950.

(31) Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Chang, C. H.; Weber, P. C. Nmr and X-Ray Evidence That the Hiv Protease Catalytic Aspartyl Groups Are Protonated in the Complex Formed by the Protease and A Nonpeptide Cyclic Urea-Based Inhibitor. *J. Am. Chem. Soc.* **1994**, *116*, 10791–10792.

(32) Czodrowski, P.; Sotriffer, C. A.; Klebe, G. Atypical Protonation States in the Active Site of HIV-1 Protease: A Computational Study. *J. Chem. Inf. Model.* **2007**, *47*, 1590–1598.
(33) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. Ab Initio Molecular Dynamics-Based Assignment of the Protonation State of Pepstatin A/HIV-1 Protease Cleavage Site. *J. Am. Chem. Soc.* **2001**, *123*, 8730–8737.
(34) Brik, A.; Wong, C. H. HIV-1 Protease: Mechanism and Drug Discovery. *Org. Biomol. Chem.* **2003**, *1*, 5–14.
(35) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; EricksonViitanen, S. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas As Hiv Protease Inhibitors. *Science* **1994**, *263*, 380–384.
(36) Hyland, L. J.; Tomaszek, T. A.; Meek, T. D. Human Immunodeficiency Virus-1 Protease. 2. Use of Ph Rate Studies and Solvent Kinetic Isotope Effects to Elucidate Details of Chemical Mechanism. *Biochemistry* **1991**, *30*, 8454–8463.
(37) Hou, T.; McLaughlin, W. A.; Wang, W. Evaluating the Potency of HIV-1 Protease Drugs to Combat Resistance. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 1163–1174.
(38) 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.
(39) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes With a Modified Generalized Born Model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383–394.
(40) Maseras, F.; Morokuma, K. Imomm - A New Integrated Ab-Initio Plus Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition-States. *J. Comput. Chem.* **1995**, *16*, 1170–1179.
(41) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO+MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t-Bu)(3))(2)+H-2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
(42) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. A Quantum Mechanical/Molecular Mechanical Model of Inhibition of the Enzyme Phospholipase A2. *J. Chem. Soc., Perkin Trans. 2* **1991**, 18199–1832.
(43) Grimme, S. Accurate Description of Van Der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.
(44) Mohr, M.; McNamara, J. P.; Wang, H.; Rajeev, S. A.; Ge, J.; Morgado, C. A.; Hillier, I. H. The Use of Methods Involving Semi-Empirical Molecular Orbital Theory to Study the Structure and Reactivity of Transition Metal Complexes. *Faraday Discuss.* **2003**, *124*, 413–428.
(45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, 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*; Gaussian, Inc.: Wallingford, CT, 2004.
(46) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
(47) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
(48) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
(49) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model With a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.
(50) Gao, J. L.; Xia, X. F. A Priori Evaluation of Aqueous Polarization Effects Through Monte-Carlo QM-MM Simulations. *Science* **1992**, *258*, 631–635.
(51) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L. Assessing Energy Functions for Flexible Docking. *J. Comput. Chem.* **1998**, *19*, 1612–1622.
(52) Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. *J. Chem. Inf. Model.* **2006**, *46*, 401–415.
(53) Schechter, I.; Berger, A. On Size of Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.
(54) Wlodawer, A.; Erickson, J. W. Structure-Based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem.* **1993**, *62*, 543–585.
(55) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; 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.
(56) Case, D. A.; Darden, T.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R.; Luo, R.; Merz, K. M., Jr.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8.0*; University of California: San Francisco, CA, 2004.
(57) Ewing, T. J. A.; Kuntz, I. D. Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening. *J. Comput. Chem.* **1997**, *18*, 1175–1189.
(58) Todorov, N. P.; Monthoux, P. H.; Alberts, I. L. The Influence of Variations of Ligand Protonation State and Tautomerism on Protein-Ligand Recognition and Binding Energy Landscape. *J. Chem. Inf. Model.* **2006**, *46*, 1134–1142.
(59) Pellegrini, E.; Field, M. J. A Generalized-Born Solvation Model for Macromolecular Hybrid-Potential Calculations. *J. Phys. Chem. A* **2002**, *106*, 1316–1326.

CI800432S