

## Assessing the Role of Polarization in Docking

Christopher J. R. Illingworth,<sup>†</sup> Garrett M. Morris,<sup>‡</sup> Kevin E. B. Parkes,<sup>§</sup> Christopher R. Snell,<sup>§</sup> and Christopher A. Reynolds<sup>\*,†</sup>

Department of Biological Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, The Scripps Research Institute, Department of Molecular Biology, MB-5 10550 North Torrey Pines Road La Jolla, California 92037-1000, and Medivir UK Ltd., Chesterford Research Park, Little Chesterford, Essex CB10 1XL, United Kingdom

Received: October 19, 2007; Revised Manuscript Received: August 11, 2008

We describe a strategy for including ligand and protein polarization in docking that is based on the conversion of induced dipoles to induced charges. Induced charges have a distinct advantage in that they are readily implemented into a number of different computer programs, including many docking programs and hybrid QM/MM programs; induced charges are also more readily interpreted. In this study, the ligand was treated quantum mechanically to avoid parametrization issues and was polarized by the target protein, which was treated as a set of point charges. The induced dipole at a given target atom, due to polarization by the ligand and neighboring residues, was reformulated as induced charges at the given atom and its bonded neighbors, and these were allowed to repolarize the ligand in an iterative manner. The final set of polarized charges was evaluated in docking using AutoDock 4.0 on 12 protein–ligand systems against the default empirical Gasteiger charges, and against nonpolarized and partially polarized potential-derived charges. One advantage of AutoDock is that the best rmsd structure can be identified not only from the lowest energy pose but also from the largest cluster of poses. Inclusion of polarization does not always lead to the lowest energy pose having a lower rmsd, because docking is designed by necessity to be rapid rather than accurate. However, whenever an improvement in methodology, corresponding to a more thorough treatment of polarization, resulted in an increased cluster size, then there was also a corresponding decrease in the rmsd. The options for implementing polarization within a purely classical docking framework are discussed. 2008112

### Introduction

A variety of virtual screening protocols are used in drug discovery for both hit generation and lead optimization and consequently there is much effort to improve both the efficiency and the accuracy of these methods. Here we present initial quantum mechanical/molecular mechanics (QM/MM) studies on the inclusion of polarization in both the ligand and the target. Current virtual screening approaches range from pharmacophore-based methods,<sup>1</sup> through empirical or knowledge-based methods<sup>2–7</sup> to physics-based methods.<sup>8–15</sup> Each approach has its merits. For instance, pharmacophore-based methods are not susceptible to the same kinds of errors as physics-based ones, whereas empirical methods can be parametrized to reproduce known binding constants—provided that they are used on systems similar to those used in the parametrization. Various combinations of these approaches are also used. Typically, the physics-based methods employ molecular mechanics to evaluate the ligand–target interaction energy and the errors in such force-field-based methods are well understood from their use in molecular dynamics and Monte Carlo simulations.<sup>16</sup> Explicitly modeled molecular mechanics polarization is frequently absent from most such simulations, from many hybrid QM/MM (quantum mechanical/molecular mechanics) studies and from most docking experiments.<sup>16,17</sup>

Past approaches to modeling MM polarization include those based around a fluctuating charge model,<sup>18–22</sup> those based on a Drude Oscillator model<sup>23–27</sup> and those based around induced dipoles.<sup>28–42</sup> The latter approach is the most widespread, and the SIBFA implementation<sup>43</sup> is one of a small number of programs with a long track record in terms of application to ligand–receptor interactions. Other recent specific approaches to polarization are reviewed elsewhere.<sup>44</sup> Our method as presented here is distinct from these, as it is based upon modifications to the MM charges of a QM/MM system, induced through the influence of the QM region. Compared to purely MM approaches, QM treatment is convenient in that it avoids the need for parametrization of new ligands. In contrast to the induced dipole and Drude Oscillator models, it holds the advantage of being expressed purely in modifications to already existent point charges, making for easier integration with packages, such as AutoDock, that use point charges to model the electrostatics of a protein system. Previous work on the modeling of polarization through induced charges in QM/MM models has shown significant improvement in calculated energy values.<sup>45</sup>

Our approach to docking employs AutoDock and is based on an extension to the QM/MM approach of Friesner et al.<sup>46</sup> It includes not only the quantum mechanical polarization of the ligand as in Friesner et al.'s approach but also the polarization of the target macromolecule. The implementation described here involves (a) knowledge of the structure of the complex and (b) iteration of quantum mechanical calculations and is not presented as a practical tool in docking but rather to catalyze developments that eventually lead to routine docking to polar-

\* Corresponding author. E-mail: reync@essex.ac.uk.

<sup>†</sup> The University of Essex.

<sup>‡</sup> The Scripps Research Institute.

<sup>§</sup> Medivir UK.

ized enzymes and receptors, particularly as the methodology can be implemented in a purely classical framework.<sup>47,48</sup>

## Methods

The 12 ligand–protein systems studied were selected from a larger list of complexes that in other studies had proven difficult for AutoDock. They were randomly chosen subject to the restriction that the number of rotatable bonds in the ligand was relatively low (fewer than eight torsions) so that evaluation of the results was not complicated by the possibility of limited sampling. Six distinct docking experiments were carried out for each of the 12 systems tested, each experiment using a different set of partial charges calculated as described below.

In the first experiment, the ligand was docked using AutoDock with Gasteiger<sup>49</sup> (i.e., default) charges for both the ligand and the receptor. In a second experiment, Gasteiger charges were used for the ligand, and potential-derived charges on the receptor were taken from the AMBER force field.<sup>50,51</sup> In the remaining experiments, quantum mechanical calculations were carried out at the B3LYP/6-31G\* level of theory<sup>52,53</sup> using GAUSSIAN03.<sup>54</sup> For the third experiment, a QM calculation was carried out on the ligand in vacuo, to obtain a distributed multipole analysis (DMA). The mulfit program<sup>55–57</sup> was used to assign potential-derived charges to the ligand atoms from the DMA, and AMBER charges were again used for the receptor. For the fourth experiment, a QM/MM calculation was carried out on the ligand in its crystallographic position, with the QM region defined as the ligand, and the MM region defined as all residues of the protein with at least one atom within 5.5 Å of the ligand, the MM region being represented as point charges. Though it is true that the energetic effects of polarization will extend beyond this cutoff,<sup>58</sup> the effect of polarization on the magnitude of individual charges is generally short-ranged.<sup>45</sup> Therefore, for the purposes of assessing the effect of incorporating polarization into an MM docking model, a 5.5 Å residue-based cutoff is justified. Within the QM/MM calculations, the interaction between the atomic charges in the MM region and the quantum mechanical wave function is included as a perturbation in the core Hamiltonian. Essentially, the point charges are treated as additional nuclear centers that do not have basis functions, as described elsewhere.<sup>45,59–61</sup> Mulfit was again used to assign potential-derived charges to the ligand based on the converged wave function, which had been polarized by the AMBER charges used for the MM region. These QM/MM-derived charges were used for the ligand, and unmodified AMBER charges were used for the receptor.

In the fifth experiment, the MM region of the system, defined above, was polarized by the QM wave function, using the induced charge method.<sup>45</sup> In addition to the polarizing effect of the wave function, the MM region was allowed to mutually polarize itself. The MM region was divided into residues, and the polarizing effect of the point charges of other residues in the MM region was added to the atoms of each residue. Mulfit charges were used for the ligand (as in the fourth experiment). The new polarized MM charges were used for the atoms in residues within 5.5 Å of the ligand, whereas for other parts of the protein, standard AMBER charges were used.

In the sixth and final experiment, the new MM charges of the fifth run were allowed to repolarize the QM region, and the induced charge algorithm was iterated four times, allowing for convergence between the QM wave function and the MM charges. This generated new polarized QM and MM charges for the ligand and the protein, which were used for a final docking experiment.

The above strategy was implemented into the latest version of AutoDock 4.0 (beta test),<sup>15,62</sup> partly because of the ease of implementation and partly because the clustering in AutoDock<sup>63,64</sup> provides an energy-independent way of assessing docked poses. By default, AutoDock uses Gasteiger charges<sup>49</sup> to compute electrostatics. Regardless of the origin of the charges (Gasteiger charges, QM derived charges or QM/MM derived charges), the force field used in the docking was entirely classical. The use of potential-derived charges for the protein<sup>50</sup> and the ligand<sup>55,57,65</sup> raises the magnitude of the charges and so the relative scaling by AutoDock of the electrostatic and other terms was adjusted to keep these in balance and hence retain the association between the computed charges and experimental  $pK_i$ 's (see Supporting Information). One hundred Lamarckian GA (Genetic algorithm) runs were calculated for each docking, with the results being clustered with a tolerance of 0.5 Å rmsd. The number of energy evaluations was largely set in line with the number of rotatable bonds. The protonation state of the ligand and the protein residues in the binding site were generally determined by visual inspection. Where there was ambiguity about the location of hydrogens, MolProbity,<sup>66</sup> an online tool for the validation and correction of protein structures, was used to optimize the hydrogen bonding network. (As a result, all amino acids had the usual protonation state, except that His 57 in 1TTP (tryptophan synthase) and His 110 in 2ACS (aldose reductase) were protonated.) A large grid box was selected that generally covered about a quarter of the protein surface and provided scope for ligands to dock in regions other than the native binding site. Where the protein structure contained additional molecules that were different to the ligand being docked, these were retained for the docking, and similarly, nonidentical additional protein chains were also retained. Metal ions were retained, with the exception of mercury, and all water molecules were removed from the system. In all docking runs, the receptor was kept rigid.

Each docking experiment produced 100 different ligand positions, each with an associated AutoDock 4 energy value,  $\epsilon_i$ , in kcal mol<sup>-1</sup>, and a crystallographic rmsd value,  $r_i$ , in Å, where  $i$  denotes the docked conformation in question. To compare docking experiments, the Boltzmann distribution was applied to give a graphical representation of the effectiveness of the docking experiment. For each ligand position,  $i$ , in the docking run, where  $R$  is the Universal gas constant, and  $T$  is the absolute temperature, here 298 K, the relative population  $p_i$  was calculated as

$$p_i = \frac{e^{-1000\epsilon_i/RT}}{\sum_{j=1}^{100} e^{-1000\epsilon_j/RT}} \quad (1)$$

These relative population values can be used to give an estimate for the proportion of docked ligands that fall within any given rmsd cutoff. For example, the proportion of ligands that fall within 2 Å of the crystallographically observed position is given by summing the values  $p_i$  for all  $i$  such that  $r_i \leq 2$  Å. By steadily increasing the rmsd cutoff, the proportion of ligands falling within any rmsd value was plotted for a range of rmsd values.

## Results

The root-mean-square deviations (RMSDs) between the docked and crystallographic positions of the ligand for the lowest energy poses are given in Table 1. For cytochrome *c* peroxidase and methylimidazole (PDB code 1AET), the rmsd upon redocking the ligand with the default Gasteiger charge was 3.0 Å

**TABLE 1: RMSD Values in Å of the Lowest Energy Poses for Each Docking Run<sup>a</sup>**

system	protein	no. of ligand torsions	no. energy evals ('000s)	rmsd value of the lowest energy docking (Å)					
				original	AMBER charges	QM ligand	QM pol	MM pol	MM converged
1AET	cytochrome <i>c</i> peroxidase	0	250	3.0	3.0	3.2	0.4	0.3	0.3
1AVN	carbonic anhydrase	3	1000	13.4	6.5	6.9	6.4	6.6	6.4
1DOG	glucoamylase	5	250	0.9	0.9	0.9	0.9	0.9	0.9
1DWB	thrombin	2	250	13.4	13.4	13.5	13.3	13.3	13.4
1ETT	bovine thrombin	7	2500	0.6	0.5	4.0	3.7	3.7	4.1
1LGR	glutamine synthetase	7	2500	13.3	2.6	12.8	12.5	12.8	12.3
1PPH	bovine trypsin	7	2500	0.7	0.7	0.6	0.6	0.9	0.7
1TNI	trypsin	5	1000	2.1	1.9	2.1	2.1	2.3	2.2
1TTP	bovine trypsin	6	2500	1.3	1.3	11.0	0.7	11.1	11.0
2ACS	human aldose reductase	6	2500	16.6	6.5	6.3	6.4	1.6	2.0
4TS1	mutated tyrosyl-T/RNA synthetase	5	2500	2.6	4.8	6.3	6.3	6.3	14.7
9AAT	mitochondrial aspartate amino-transferase	6	2500	7.0	2.7	2.7	0.9	0.9	0.9

<sup>a</sup> Column 4 gives RMSD scores for the run using default charges, and columns 5–9 give scores for runs adding AMBER receptor charges, in vacuo QM ligand charges, QMMM ligand charges, polarized MM receptor charges, and converged ligand and receptor charges, respectively. The ligand names are given in Table 2.

**TABLE 2: RMSD Values in Å from the Crystallographic Structure for the Lowest Energy Pose in the Largest Cluster of Dockings Produced by AutoDock for Each Run<sup>a</sup>**

system	ligand	smallest crystallographic rmsd value in Å in largest cluster and (size of cluster)					
		original	AMBER charges	QM ligand	QM pol	MM pol	converged
<b>1AET</b>	protonated benzimidazole	3.0 (61)	2.9 (51)	0.2 (77)	0.2 (91)	0.2 (85)	0.2 (89)
<i>1AVN</i>	histamine	14.5 (18)	6.8 (9)	6.5 (6)	6.7 (5)	6.4 (7)	6.3 (5)
1DOG	1-deoxyojirimycin	0.8 (54)	0.6 (53)	0.8 (48)	0.7 (41)	0.6 (58)	0.7 (56)
<b>1DWB</b>	benzamidine	9.8 (14)	9.8 (21)	13.2 (30)	0.5 (39)	0.4 (48)	0.4 (59)
<i>1ETT</i>	NPAP inhibitor	0.5 (64)	0.4 (51)	3.9 (18)	3.9 (25)	3.9 (14)	3.9 (14)
<i>1LGR</i>	AMP	1.0 (60)	1.0 (18)	12.6 (8)	12.0 (4)	12.6 (5)	12.2 (4)
1PPH	3TPAT	0.4 (39)	0.4 (46)	0.4 (32)	0.4 (23)	0.5 (29)	0.4 (34)
1TNI	4-phenylbutylamine	2.3 (15)	1.0 (22)	1.3 (19)	1.0 (15)	2.1 (21)	2.4 (10)
<i>1TTP</i>	<i>p</i> -amidinophenylpyruvate	0.6 (31)	0.5 (45)	1.6 (23)	0.6 (17)	1.6 (18)	1.3 (10)
2ACS	citrate	6.4 (11)	6.4 (6)	6.1 (28)	6.1 (23)	1.7 (11)	1.9 (11)
4TS1	tyrosine	0.7 (18)	4.7 (12)	6.2 (16)	6.2 (17)	0.7 (13)	0.8 (23)
<b>9AAT</b>	phosphonomethylphenylalanine	10.0 (5)	0.8 (32)	0.8 (23)	0.7 (31)	0.7 (24)	0.8 (23)

<sup>a</sup> The system name is marked in bold type where the size of the largest cluster increases as the charges are improved or in italic type where the size of the largest cluster decreases as the charges are improved.

(column 4), a result which can be described as only just of intermediate quality.<sup>46</sup> Using potential-derived charges for the protein (column 5, rmsd 3.0 Å) or for the ligand and the protein (column 6, rmsd = 3.2 Å) essentially does not change this. Allowing the potential-derived charges for the ligand to be polarized by the protein (during the QM/MM calculation) reduced the rmsd to 0.4 Å (column 7) and including polarization of the protein reduces this further to 0.3 Å (column 8), regardless of whether the induced charges are iterated (column 9) or not. The results for the docking of histamine to carbonic anhydrase (PDB code 1AVN) are improved numerically but not qualitatively, as the result remains incorrect for all methods, probably because of the removal of water from the binding site. Omission of water molecules was also identified as a likely cause of poor results for the dockings of benzamidine to thrombin (PDB code 1DWB), although in this case a cluster of well-docked poses (at higher energy) was identified with the addition of MM polarization. The results for docking 1-deoxyojirimycin to glucoamylase II-(471) (PDB code 1DOG) and for 4-phenylbutylamine to trypsin (PDB code 1TNI) are good and adequate respectively and this does not change even when the charge determination method is improved.

In the dockings of adenosine monophosphate to glutamine synthetase (PDB code 1LGR), poor results were caused by the ligand being incorrectly docked into an alternative binding site around 13 Å from the site in the crystal structure. Another

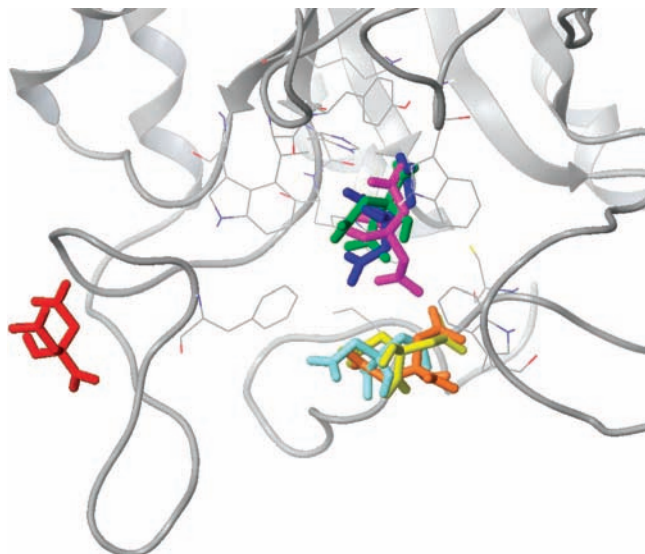
crystallographic study of this protein shows this alternative binding site occupied by phosphothrycin.<sup>67</sup> 1LGR represents a monomer unit of glutamine synthetase, where ligand binding occurs in a cleft between two protein chains, and it is possible that including the second protein chain would have given better results.

For 2ACS there is a dramatic improvement in the result, but only when protein polarization is included. Figure 1 shows the 16.3 Å rmsd lowest energy pose in pdb code 2ACS, generated using Gasteiger charges (red), the ~6.4 Å rmsd lowest energy poses (orange, yellow and light blue) generated using intermediate methods and the ~2.0 Å rmsd lowest energy pose that results after MM polarization (dark blue, purple). Such a large improvement in rmsd from 16.3 to 2.0 Å is a little surprising as the polarization of the atomic charges is nowhere near as large. However, the small changes that do occur may be sufficient to guide the ligand docking process toward a correct pose rather than an incorrect one. In this sense, RMSDs of 16.3 Å and 6.4 Å are equivalent as they represent equally incorrect poses, as illustrated in Figure 1.

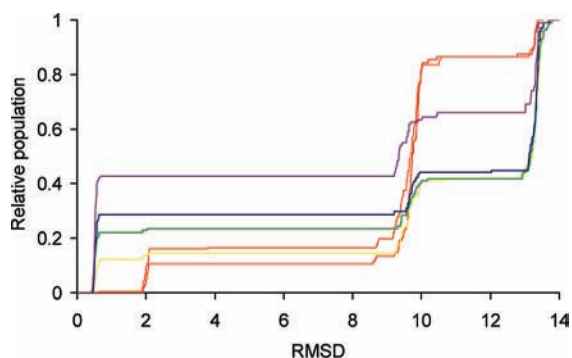
Overall, however, some systems gave improved results as the methodology improved whereas for others the reverse is true.

A possible way to analyze the results more thoroughly is to consider the full Boltzmann-weighted set of 100 poses to see whether including polarization increases the proportion of poses





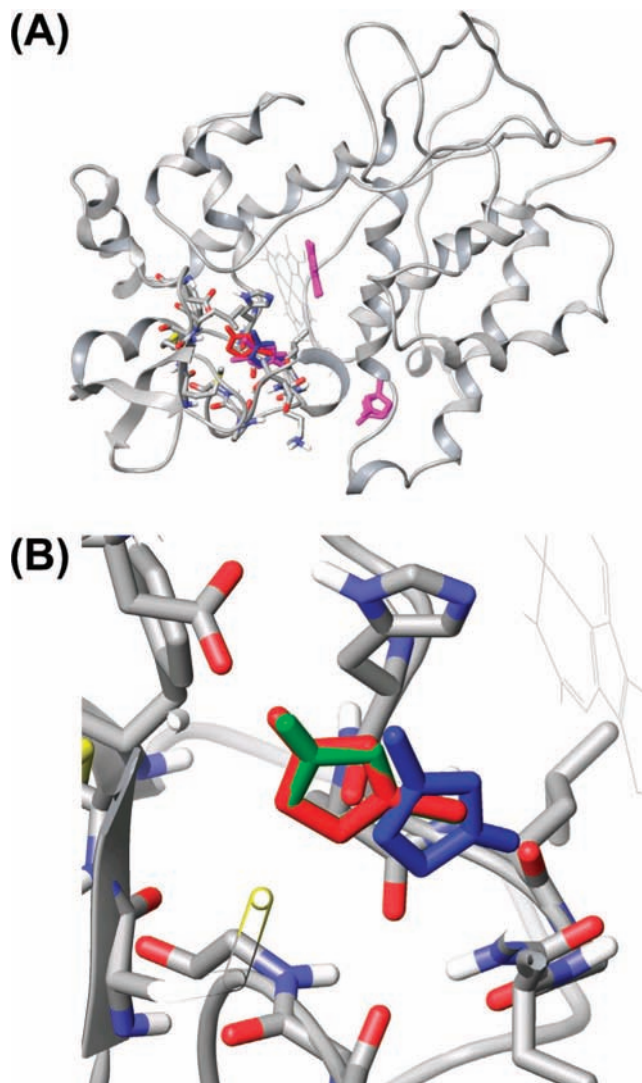
**Figure 1.** Lowest energy docked poses for the PDB code 2ACS, using default Gasteiger charges (red), AMBER protein charges and Gasteiger ligand charges (orange), AMBER protein charges and in vacuo potential-derived ligand charges (yellow), AMBER protein charges and QM/MM potential-derived ligand charges (light blue), MM polarized protein charges and QM ligand charges after one iteration of the MM polarization algorithm (dark blue), and converged MM polarized charges (purple), compared to the crystallographic ligand structure (green). The residues within the 5.5 Å of the MM region are shown in a stick fashion, and the remainder of the protein is shown as ribbons.



**Figure 2.** Boltzmann-weighted docking results for the PDB code 1DWB. Following the introduction of AMBER charges for the receptor, the relative population of the dockings with rmsd under 2 Å steadily increases as the electrostatics are improved. The improvement follows the color of the rainbow: original Gasteiger charges (red), AMBER charges (orange), QM ligand charges (yellow), polarized ligand charges (green), polarized enzyme charges (blue), converged polarized enzyme charges (violet).

with a low rmsd, and this is certainly the case for PDB code 1DWB, where, as shown in Figure 2, the relative population of systems with an rmsd of 2.0 Å or less increases from 8% with the Gasteiger charges, dips through 3% with the AMBER charges, rises to 14% with a QM ligand, to 23% with a QM polarized ligand, to 29% with the introduction of MM polarization in the receptor, and then to 43% when the polarization algorithm is converged. Other PDB codes showed mixed results, with changes in rmsd commensurate with the results tabulated.

Given the well-documented deficiencies in docking programs in general, such as limited sampling, lack of explicit solvation, rigidity of target, etc., arising from the need for computational speed, there is no guarantee that an improvement in one aspect of the methodology will increase the association between a lower energy and a lower rmsd. In this respect, the clustering of docked



**Figure 3.** (a) Lowest energy binding modes in each of the clusters obtained after clustering 100 independent dockings of 1-methylimidazole to cytochrome *c* peroxidase (PDB code 1AET), in which the default Gasteiger charges were used. The binding mode from the largest cluster is blue and has a crystallographic rmsd of 3.0 Å, whereas the structure from the cluster with lowest crystallographic rmsd (0.2 Å) is red. Other clusters, some of which appear superimposed, are purple. (b) Close-up of the binding site showing the lowest energy docked conformations in each of the largest clusters from two docking experiments of 1-methylimidazole to cytochrome *c* peroxidase (PDB code 1AET). The experiment using the default Gasteiger charges is shown in blue whereas that using the converged MM polarized charges is shown in red; the crystallographic binding mode of 1-methylimidazole is also shown (in green). When the polarized charges are introduced, the binding mode with the lower rmsd is identified more often (in 89 out of 100 dockings).

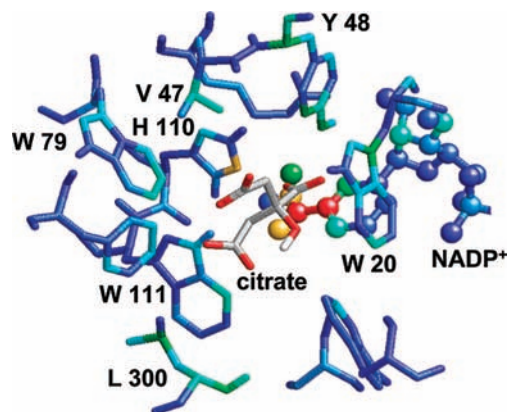
poses as implemented in AutoDock is useful, as an increase in the size of the largest cluster tends to be associated with an improvement in rmsd.<sup>63,64</sup> It is difficult to see why this is true but the corollary is easier to illustrate in that if the ligand–protein complementarity is poorly represented, then the ligand is likely to dock to multiple locations besides the correct one, as there is no reason the poses should be centered on the native binding site. Thus, Figure 3a shows the location of the lowest energy poses in each of the 8 clusters in the docking of 1-imidazole to cytochrome *c* peroxidase (1AET) with Gasteiger charges. The lowest energy representative of the largest cluster, which is blue, has an rmsd of 3.0 Å. A smaller cluster, in green, has a

representative with rmsd of 0.2 Å. Figure 3b shows the lowest energy pose in each of the largest clusters produced by the docking runs on 1AET with both the original and the converged charges compared to the crystal structure ligand, and it can be seen that, as a result of the modification to the charges, the population of the cluster with rmsd 3.0 Å has decreased, such that the cluster with rmsd 0.2 Å is now the largest cluster.

Thus, Table 2 shows the crystallographic rmsd of the lowest energy pose in the largest cluster, and here the general conclusion is that whenever an improvement in the physics results in an increased cluster size then there is also a decrease in the rmsd. For example, for the docking of benzamidine to thrombin (PDB code 1DWB), the original rmsd was 9.8 Å, with a cluster size of 14 but when the 1DWB enzyme is polarized in response to the ligand as described in the Methods above (fifth docking experiment), the rmsd drops to 0.4 Å and the cluster size increases to 59. Similar improvements are seen for PDB codes 1AET (cytochrome *c* peroxidase/1-methylimidazole) and 9AAT (aspartate aminotransferase/4'-deoxy-4'-aminopyridoxal-5'-phosphate), whereas PDB codes 1DOG (glucoamylase II-(471)/1-deoxynojirimycin), 1TNI (trypsin/4-phenylbutylamine) and 1PPH (trypsin/*m*-amidinophenyl-3-alanine) maintain consistently good RMSDs with reasonable clustering. For 1ETT ( $\epsilon$ -thrombin/*p*-amidinophenyl-3-alanine) and 1LGR (glutamine synthetase/adenosine monophosphate) the initial results are good (the rmsd is 0.5 and 1.0 Å with cluster sizes of 64 and 60, respectively). The indication that the rmsd does not improve on inclusion of polarization as expected, comes from the smaller cluster sizes of 14 and 4, respectively (final column). The only ambiguous result is for the docking of citric acid to aldose reductase (PDB code 2ACS); the original rmsd was 6.4 Å, with a cluster size of 11 but when the 2ACS enzyme is polarized in response to the ligand as described in the Methods above (fifth docking experiment), the rmsd drops to 2.0 Å and the cluster size returns to 11 (having gone through a maximum of 28 for intermediate calculations).

## Discussion

Polarization is the key term that is generally omitted from molecular mechanics studies such as molecular dynamics simulations, Monte Carlo simulations and also docking.<sup>16,44</sup> Though molecular dynamics simulations of enzymes usually include a flexible enzyme, this is not usually true in docking studies. Rather, flexibility, if it is included at all, is usually handled via the torsional movement of polar hydrogen atoms or soft repulsion potentials,<sup>2,8,68–70</sup> by inclusion of a small number of flexible side chains<sup>8,71,72</sup> or by analysis of a small number of snapshots on a molecular dynamics trajectory.<sup>73</sup> These approaches all reduce the potential for steric clashes between the ligand and the enzyme and so to some extent they also allow for the reduction of electrostatic noncomplementarities. Elsewhere we have shown that electrostatic noncomplementarities can exist among well-docked structures and that they can be alleviated by use of induced charges.<sup>45,74</sup> This factor contributes to the improvement in docking seen for a number of the difficult cases presented in Tables 1 and 2. Thus, Figure 4 shows the extent of polarization (through color coding) in the aldose reductase binding site within the enzyme (PDB code 2ACS). The biggest changes in charge are in the vicinity of the amide group of NADP+ (nicotinamide adenosine diphosphate), indicated by the red and green color, then on Trp 20, Tyr 48, His 110 and Leu 300 indicated by the green color, and with smaller changes on Val 47, Trp 79 and Trp 111, indicated by the cyan color; such information could be useful in drug design,



**Figure 4.** Induced charges on key residues in the binding site of the citrate ligand within the 2ACS structure of aldose reductase, color-coded according to their magnitude (red and orange, high; green and blue, low). The residues with changes in atomic charges are labeled. The QM citrate ligand (with three carboxylate groups and one hydroxyl group) is colored by atom type; the NADP+ (ball and stick, truncated) is part of the MM system.

as it would indicate regions where polarizing groups on the ligand could elicit a complementary response in the target protein. Polarization does not improve every case studied here and this is to be expected because other effects that have been ignored here (as is often the case in state-of-the-art docking<sup>75</sup>) may be equally important. The neglect of hydration may have led to the poor results achieved for at least two of the systems tested, as mentioned above. For the purposes of this study, protein flexibility was also neglected, and it may be that allowing for mechanical, as well as electrostatic, flexibility in the receptor, would lead to improved results. Moreover, addition of an extra term could be detrimental to the balance of the force field. Nevertheless, the scheme described here presents a staged improvement from empirical Gasteiger charges, through potential-derived charges to polarized charges. By exploiting the clustering within AutoDock, we see for these ligands that whenever the clustering improves in concert with an improvement in methodology, then the crystallographic rmsd improves. Here we treated the ligand quantum mechanically to avoid parameterization issues, but elsewhere we have shown that induced charges can be implemented within a purely classical framework.<sup>47,48</sup> Here we also focused only on the redocking problem, but elsewhere Friesner et al. have developed an iterative survival of the fittest approach<sup>46</sup> that could be used within an entirely classical approach and so knowledge of the crystal structure is not essential for development of this method. Moreover, the computational cost of induced charges is low compared to the cost of induced dipoles and because iteration and derivatives would not necessarily be required within docking programs, the induced charge method could be implemented with relatively little additional computational cost,<sup>48</sup> offering an approach to alleviating electrostatic noncomplementarity that is more cost-effective than implementation of protein flexibility. In some respects, although induced dipoles are more difficult to implement in existing docking codes, they are more accurate. However, elsewhere we have shown that compared with quantum mechanical benchmarks at the same basis set level, induced dipoles only become superior to induced charges at the level of very large basis sets.<sup>76</sup> In other words, the induced charge approach is more in keeping with the accuracy achieved from quantum mechanical methods unless very large basis sets with diffuse functions are used. From this perspective, therefore, induced charges offer the appropriate level of convenience and



accuracy for including polarization in docking, and in many other biomolecular studies.

## Conclusions

We have described a method for including polarization in docking through the use of induced charges that has the advantage of being readily implemented into a number of different programs, including both hybrid QM/MM programs and docking programs. A hierarchy of improved treatments of electrostatics ranging from Gasteiger charges through potential-derived charges to polarization of the ligand (via hybrid QM/MM calculations) or polarization of both the ligand and the enzyme were compared by redocking 12 difficult protein–ligand complexes and evaluating the results based on the rmsd of the lowest energy structure and the rmsd of the lowest energy structure in the largest cluster. The most striking result is that there is always a decrease in the crystallographic rmsd whenever an improvement in methodology results in an increase in cluster size. This particular clustering feature of AutoDock is extremely useful because the state-of-the-art in virtual screening is such that absolute energies alone are not necessarily sufficiently accurate to indicate the correct docked structure. Here the docking involved costly QM/MM calculations and knowledge of the experimental result but extensions of the method to a classical framework where the result is not known a priori are eminently possible.

**Acknowledgment.** We acknowledge support from the BBSRC, Medivir (UK) (C.J.R.I.) and the NIH (G.M.M.).

**Supporting Information Available:** Ligand charges and coordinates and further descriptions of AutoDock charges. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- Hindle, S. A.; Rarey, M.; Buning, C.; Lengauer, T. Flexible docking under pharmacophore type constraints. *J. Comput. Aided Mol. Des.* **2002**, *16* (2), 129–149.
- Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* **2003**, *21* (4), 289–307.
- Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* **2001**, *15* (5), 411–428.
- Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261* (3), 470–489.
- Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308* (2), 377–395.
- DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Docking flexible ligands to macromolecular receptors by molecular shape. *J. Med. Chem.* **1986**, *29* (11), 2149–2153.
- Bacon, D. J.; Moul, J. Docking by least-squares fitting of molecular surface patterns. *J. Mol. Biol.* **1992**, *225* (3), 849–858.
- 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.
- Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196.
- 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.* **2004**, *47* (7), 1739–1749.
- Laurie, A. T.; Jackson, R. M. Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinformatics* **2005**, *21* (9), 1908–1916.
- Jackson, R. M. Q-fit: a probabilistic method for docking molecular fragments by sampling low energy conformational space. *J. Comput. Aided Mol. Des.* **2002**, *16* (1), 43–57.
- Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMM-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24* (13), 1549–1562.
- Trosset, J. Y.; Scheraga, H. A. Reaching the global minimum in docking simulations: a Monte Carlo energy minimization approach using Bezier splines. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (14), 8011–8015.
- Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19* (14), 1639–1662.
- Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (19), 6665–6670.
- Jorgensen, W. L. Special issue on polarization. *J. Chem. Theory Comput.* **2007**, *3* (6), 1877.
- Field, M. J. Hybrid quantum mechanical molecular mechanical fluctuating charge models for condensed phase simulations. *Mol. Phys.* **1997**, *91* (5), 835–845.
- Rick, S. W.; Berne, B. J. Dynamical fluctuating charge force fields: The aqueous solvation of amides. *J. Am. Chem. Soc.* **1996**, *118* (3), 672–679.
- Rappe, A. K.; Goddard, W. A. Charge Equilibration for Molecular-Dynamics Simulations. *J. Phys. Chem.* **1991**, *95* (8), 3358–3363.
- Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical Fluctuating Charge Force-Fields - Application to Liquid Water. *J. Chem. Phys.* **1994**, *101* (7), 6141–6156.
- Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L., III. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004**, *25* (12), 1504–1514.
- Cao, J. S.; Berne, B. J. Theory and Simulation of Polar and Nonpolar Polarizable Fluids. *J. Chem. Phys.* **1993**, *99* (9), 6998–7011.
- Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119* (10), 5185–5197.
- Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119* (6), 3025–3039.
- Whitfield, T. W.; Varma, S.; Harder, E.; Lamoureux, G.; Rempe, S. B.; Roux, B. Theoretical study of aqueous solvation of K<sup>+</sup> comparing ab initio, polarizable, and fixed-charge models. *J. Chem. Theory Comput.* **2007**, *3* (6), 2068–2082.
- Anisimov, V. M.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D. Polarizable empirical force field for the primary and secondary alcohol series based on the classical drude model. *J. Chem. Theory Comput.* **2007**, *3* (6), 1927–1946.
- Applequist, J.; Carl, J. R.; Fung, K. K. Atom Dipole Interaction Model for Molecular Polarizability - Application to Polyatomic-Molecules and Determination of Atom Polarizabilities. *J. Am. Chem. Soc.* **1972**, *94* (9), 2952.
- Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103* (2), 227–249.
- Lybrand, T. P.; Kollman, P. A. Water Water and Water Ion Potential Functions Including Terms for Many-Body Effects. *J. Chem. Phys.* **1985**, *83* (6), 2923–2933.
- Caldwell, J.; Dang, L. X.; Kollman, P. A. Implementation of Nonadditive Intermolecular Potentials by Use of Molecular-Dynamics - Development of A Water Water Potential and Water Ion Cluster Interactions. *J. Am. Chem. Soc.* **1990**, *112* (25), 9144–9147.
- Voisin, C.; Cartier, A. Determination of Distributed Polarizabilities to be Used for Peptide Modeling. *J. Mol. Struct. (THEOCHEM)* **1993**, *105*, 35–45.
- Gresh, N. Energetics of Zn<sup>2+</sup> Binding to A Series of Biologically Relevant Ligands - A Molecular Mechanics Investigation Grounded on Ab-Initio Self-Consistent-Field Supermolecular Computations. *J. Comput. Chem.* **1995**, *16* (7), 856–882.
- Meng, E. C.; Caldwell, J. W.; Kollman, P. A. Investigating the anomalous solvation free energies of amines with a polarizable potential. *J. Phys. Chem.* **1996**, *100* (6), 2367–2371.
- Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. *J. Comput. Chem.* **2002**, *23* (16), 1515–1531.
- Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. *J. Phys. Chem. A* **2004**, *108* (4), 621–627.

- (37) Ren, P.; Ponder, J. W. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.* **2002**, *23* (16), 1497–1506.
- (38) Borodin, O.; Smith, G. D. Development of quantum chemistry-based force fields for poly(ethylene oxide) with many-body polarization interactions. *J. Phys. Chem. B* **2003**, *107* (28), 6801–6812.
- (39) Borodin, O.; Smith, G. D. Development of many-body polarizable force fields for Li-battery components: 1. Ether, alkane, and carbonate-based solvents. *J. Phys. Chem. B* **2006**, *110* (12), 6279–6292.
- (40) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P. Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J. Chem. Theory Comput.* **2007**, *3* (6), 1960–1986.
- (41) Gresh, N. Development, validation, and applications of anisotropic polarizable molecular mechanics to study ligand and drug-receptor interactions. *Curr. Pharm. Des.* **2006**, *12* (17), 2121–2158.
- (42) Friesner, R. A. Modeling polarization in proteins and protein-ligand complexes: Methods and preliminary results. *Adv. Prot. Chem.* **2006**, *79*, +
- (43) Gresh, N.; Claverie, P.; Pullman, A. Theoretical-Studies of Molecular-Conformation - Derivation of An Additive Procedure for the Computation of Intramolecular Interaction Energies - Comparison with Abinitio Scf Computations. *Theor. Chim. Acta* **1984**, *66* (1), 1–20.
- (44) Jorgensen, W. L. Special issue on polarization, edited by W. L. Jorgensen. *J. Chem. Theory Comput.* **2007** *3* (6),
- (45) Illingworth, C. J.; Gooding, S. R.; Winn, P. J.; Jones, G. A.; Ferenczy, G. G.; Reynolds, C. A. Classical polarization in hybrid QM/MM methods. *J. Phys. Chem. A* **2006**, *110* (20), 6487–6497.
- (46) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26* (9), 915–931.
- (47) Winn, P. J.; Ferenczy, G. G., III. Polarization through modified atomic charges. *J. Comput. Chem.* **1999**, *20* (7), 704–712.
- (48) Ferenczy, G. G.; Reynolds, C. A. Modelling polarization through induced atomic charges. *J. Phys. Chem. A* **2001**, *105*, 11470–11479.
- (49) Gasteiger, J.; Marsili, M. New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.* **1978**, *3181*–3184.
- (50) Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.
- (51) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999–2012.
- (52) Becke, A. D. Density-Functional Thermochemistry. 3. the Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (53) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. 12. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular-Orbital Studies of Organic-Molecules. *J. Chem. Phys.* **1972**, *56* (5), 2257.
- (54) 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.; and Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (55) Ferenczy, G. G. Charges Derived from Distributed Multipole Series. *J. Comput. Chem.* **1991**, *12* (8), 913–917.
- (56) Ferenczy, G. G.; Winn, P. J.; Reynolds, C. A.; Richter, G. Effective distributed multipoles for the quantitative description of electrostatics and polarisation in intermolecular interactions. *Abstr. Pap. Am. Chem. Soc.* **1997**, *214*, 38.
- (57) Winn, P. J.; Ferenczy, G. G.; Reynolds, C. A. Toward improved force fields 0.1. Multipole-derived atomic charges. *J. Phys. Chem. A* **1997**, *101* (30), 5437–5445.
- (58) Illingworth, C. J. R.; Parkes, K. E.; Snell, C. R.; Marti, S.; Moliner, V.; Reynolds, C. A. The effect of MM polarization on the QM/MM transition state stabilization: application to chorismate mutase. *Mol. Phys.* **2008**, *106*, 1511–1515.
- (59) Reynolds, C. A.; Richards, W. G.; Goodford, P. J. Prediction of selective bioreductive anti-tumor, anti-folate activity using a modified abinitio method for calculating enzyme-inhibitor interaction energies. *J. Chem. Soc., Perkin Trans. II* **1988**, (4), 551–576.
- (60) Reynolds, C. A.; Richards, W. G.; Goodford, P. J. Introducing selectivity into dihydrofolate-reductase inhibitors. *Anti-Cancer Drug Des.* **1987**, *1* (4), 291–295.
- (61) Reynolds, C. A.; Ferenczy, G. G.; Richards, W. G. Methods for determining the reliability of semiempirical electrostatic potentials and potential derived charges. *J. Mol. Struct. (THEOCHEM)* **1992**, *88*, 249–269.
- (62) 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.
- (63) 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.
- (64) Ruvinsky, A. M.; Kozintsev, A. V. New and fast statistical-thermodynamic method for computation of protein-ligand binding entropy substantially improves docking accuracy. *J. Comput. Chem.* **2005**, *26* (11), 1089–1095.
- (65) Ferenczy, G. G.; Winn, P. J.; Reynolds, C. A. Toward improved force fields. 2. Effective distributed multipoles. *J. Phys. Chem. A* **1997**, *101* (30), 5446–5455.
- (66) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., III; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W375–W383.
- (67) Gill, H. S.; Eisenberg, D. The crystal structure of phosphothricin in the active site of glutamine synthetase illuminates the mechanism of enzymatic inhibition. *Biochemistry* **2001**, *40* (7), 1903–1912.
- (68) Jiang, F.; Kim, S. H. “Soft docking”: matching of molecular surface cubes. *J. Mol. Biol.* **1991**, *219* (1), 79–102.
- (69) Gschwend, D. A.; Good, A. C.; Kuntz, I. D. Molecular docking towards drug discovery. *J. Mol. Recognit.* **1996**, *9* (2), 175–186.
- (70) Schnecke, V.; Swanson, C. A.; Getzoff, E. D.; Tainer, J. A.; Kuhn, L. A. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins* **1998**, *33* (1), 74–87.
- (71) Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **1994**, *235* (1), 345–356.
- (72) Schaffer, L.; Verkhivker, G. M. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. *Proteins* **1998**, *33* (2), 295–310.
- (73) Meagher, K. L.; Carlson, H. A. Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case. *J. Am. Chem. Soc.* **2004**, *126* (41), 13276–13281.
- (74) Gooding, S. R.; Winn, P. J.; Maurer, R. I.; Ferenczy, G. G.; Miller, J. R.; Harris, J. E.; Griffiths, D. V.; Reynolds, C. A. Fully polarizable QM/MM calculations: An application to the nonbonded iodine-oxygen interaction in dimethyl-2-iodobenzoylphosphonate. *J. Comput. Chem.* **2000**, *21* (6), 478–482.
- (75) Gouldson, P. R.; Kidley, N. J.; Bywater, R. P.; Psaroudakis, G.; Brooks, H. D.; Diaz, C.; Shire, D.; Reynolds, C. A. Toward the active conformations of rhodopsin and the beta2-adrenergic receptor. *Proteins* **2004**, *56* (1), 67–84.
- (76) Illingworth, C. J. R.; Parkes, K. E. B.; Snell, C. R.; Ferenczy, G. G.; Reynolds, C. A. Toward a consistent treatment of polarization in model QM/MM calculations. *J. Phys. Chem. A* **2008**, *112*, 12151–12156.