

# QM/MM Calculations in Drug Discovery: A Useful Method for Studying Binding Phenomena?

M. Paul Gleeson<sup>\*,†</sup> and Duangkamol Gleeson<sup>‡</sup>

Computational & Structural Chemistry, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, United Kingdom, and Department of Chemistry, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

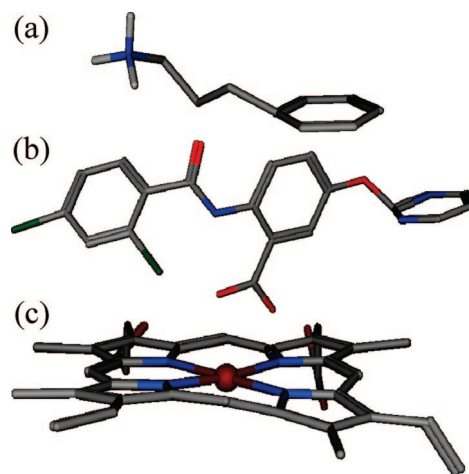
Received November 13, 2008

Herein we investigate whether QM/MM could prove useful as a tool to study the often subtle binding phenomena found within pharmaceutical drug discovery programs. The goal of this investigation is to determine whether it is possible to employ high level QM/MM calculations to answer specific questions around a binding event in a cycle time that is aligned with medicinal chemistry synthesis. To this end QM/MM calculations have been performed on four protein kinase-ligand complexes using five different levels of theory, using standard hardware, in an effort to assess their utility. We conclude that the accuracy and turnaround time of such calculations mean they could prove valuable to (1) probe the subtle nature of the interactions within protein active sites, (2) facilitate the interpretation of poorly resolved electron density, and (3) study the impact of substituent changes on the binding conformation or in the assessment of alternate scaffolds. In practice, the successful application of such methods will be limited by the size of the system under investigation, the level of theory used, and whether there is a need for conformational sampling.

## 1. INTRODUCTION

Structure-based drug design (SBDD) is an important component of the drug discovery process which sees the optimization of a lead molecule in a rational, structurally enabled manner, to realize a more potent, developable molecule in the fewest number of chemistry iterations.<sup>1–4</sup> Determination of high resolution protein–ligand complexes, via X-ray crystallography or NMR, means computational chemistry methods can be used to virtually design small numbers of molecules to probe the existing interactions found in the protein–ligand complex and also explore the active site for any additional hydrophilic or hydrophobic interactions to increase affinity.<sup>5–8</sup>

The resolved atomic coordinates from an X-ray crystallography experiment, the most common source of data, represents an interpretation of the electron density of a static crystal structure, which itself is an average of a dynamic system.<sup>9,10</sup> Although the iterative refinement of the electron density involves a cross-validation procedure coupled with predetermined libraries of bond distances, angles, and dihedrals compiled from existing solved complexes,<sup>11</sup> the results are unlikely to be error-free, especially with low resolution diffraction data.<sup>2,7,12–14</sup> For example 3-phenylpropyl-amine bound to trypsin (1TNK) and an agonist of PPAR-Gamma (1WM0) both display pyrimidal SP<sup>2</sup> carbon atoms, while the resolved haem group of P450-2C9 (1R9O) shows significant distortion (Figure 1). An additional limitation of X-ray crystallography is that to test a hypothesis, a molecule must be designed, synthesized, crystallized, and



**Figure 1.** An illustration of errors found in high resolution X-ray coordinates of (a) 3-phenylpropylamine bound to trypsin (1TNK, 1.8 Å), (b) a PPAR-Gamma agonist (1WM0), and (c) the haem group found in P450-2C9 (1R9O, 2.0 Å) (bottom).

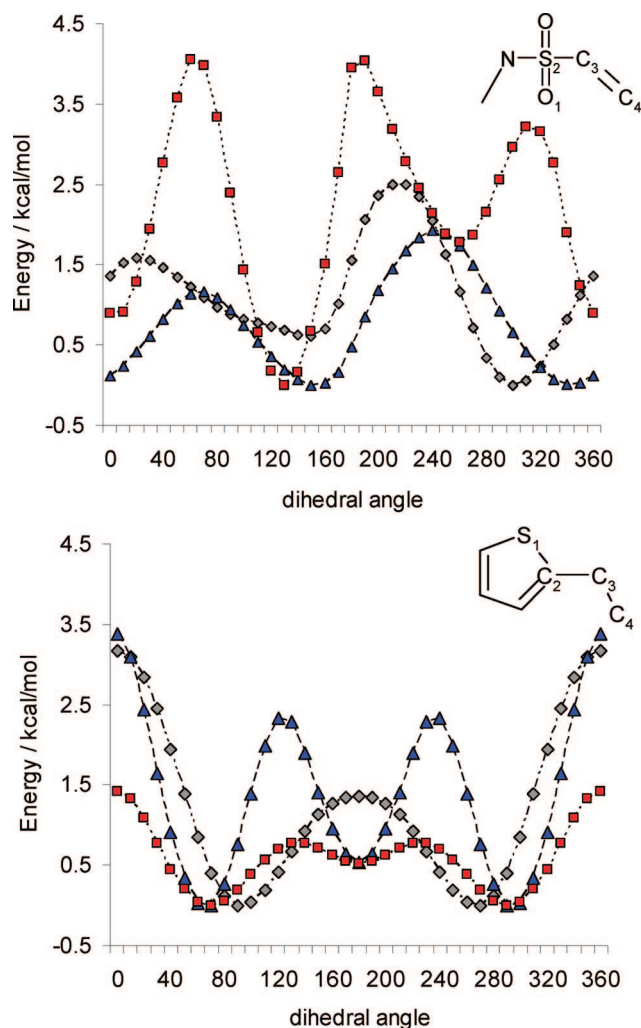
diffracted and finally the structure solved before it can be used by a program team. This process can sometimes take months to realize a result if at all in difficult cases, meaning the cycle time can be incompatible with those of medicinal chemistry.

Computational chemistry in contrast allows the testing of a hypothesis in minutes to hours using conventional, high throughput empirical methods such as GOLD,<sup>15</sup> GLIDE,<sup>16</sup> or COMFA.<sup>17</sup> However reservations have been raised in the literature recently regarding the accuracy of such methods, particularly in the fields of docking and scoring<sup>18–21</sup> and QSAR.<sup>22,23</sup> These failures can be attributed to a number of factors including inaccurate or incomplete force-field parameters, such as for metals,<sup>24</sup> the neglect of protein

\* Corresponding author phone: +44 (0)1438 768682; fax: +44 (0) 1438 763352; e-mail: paul.x.gleeson@gsk.com.

<sup>†</sup> GlaxoSmithKline Medicines Research Centre.

<sup>‡</sup> King Mongkut's Institute of Technology Ladkrabang.



**Figure 2.** Plot of dihedral energy scans for a sulfonamide fragment at the B3LYP/6-31G\* (red), OPLS (blue), and MMFF (gray).<sup>29</sup> The corresponding results for thiophene were also obtained using the same methodology as described in ref 29. For the thiophene fragment, the OPLS method shows the same energy profile as the DFT result but with markedly higher barriers, while MMFF displays a subtly different profile but comparable overall barriers. For the sulfonamide fragment both MM methods show considerably different profiles and energy barriers when compared to the DFT results, indicating the subtle errors associated with MM based methods.

flexibility or ensemble averaging,<sup>9,10</sup> or due to deficiencies with the empirical docking score. This is essentially a global QSAR model that will suffer if the validation procedure is not rigorous<sup>25,26</sup> or because of the neglect of the domain of applicability,<sup>27</sup> a concept not considered in docking and scoring.

Quantum mechanics (QM) is a theoretically rigorous method capable of accurately reproducing experimental phenomena and is a potential solution in the iterative SBDD process.<sup>7,28</sup> QM methods offer significant advantages over empirical molecular mechanics (MM) methods, for example, being capable of accounting for charge delocalization/polarization and the accurate treatment of metal atoms in protein systems.<sup>24</sup> SAR investigations by Senger et al.<sup>29</sup> for example highlight the issues with empirical force field methods through the calculation of dihedral energy barriers for fragments using B3LYP/6-31G\*, MMFF,<sup>30</sup> and OPLS<sup>31</sup> methods. The results in Figure 2 show that a considerable

difference exists between the MM methods and DFT based results, which is disconcerting given the latter method has been shown to be in good agreement with experiment based on population distributions derived from the CSD.<sup>32</sup>

QM methods allows one to accurately calculate internal energies and describe the range of complex interactions observed between a protein and ligand including very strong hydrogen bonds (H-Bonds) between 1.2 and 1.5 Å, often observed within catalytically active proteins, moderately strong H-bonds (1.5–2.2 Å) such as NH-O/N, OH-O/N), weak H-bonds such as CH-O and CH-N (~2.0–3.0 Å),<sup>33</sup>  $\pi$ - $\pi$  stacking (edge to face or parallel displaced), to cation- $\pi$  interactions.<sup>34,35</sup> To model a typical sized protein–ligand complex using reasonable levels of QM theory alone (i.e. B3LYP/6-311++G\*\*) would be computationally unfeasible, but compromise solutions exist, including a range of hybrid QM/MM methods,<sup>36–40</sup> linear scaling semiempirical,<sup>41</sup> and ab initio methods.<sup>42</sup> We focus on the former method where the key active site residues and the bound ligand can be treated using more accurate QM methods (termed inner region here), and the remaining less critical residues are modeled using an MM force field (outer region). By incorporating the MM point charges of the outer region directly into the QM calculation the crucial longer range electrostatic effects can be assessed directly.

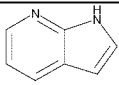
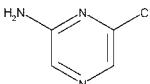
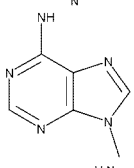
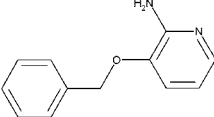
QM/MM methods<sup>36–40</sup> have been used to model reaction mechanisms of pharmaceutically relevant targets, to help our understating of P450 mediated metabolism<sup>43</sup> and in efforts to improve the docking and scoring problem.<sup>44</sup> It has been suggested that conformational sampling in conjunction with QM/MM will be necessary to predict biological parameters such as binding infinities or rates.<sup>45</sup>

In the ONIOM<sup>39,40</sup> based implementation used here the total QM/MM energy is computed as the QM energy of the inner region including the MM charges, plus the MM energy of the whole outer region, minus the MM energy of the inner region.

$$E_{QM/MM} = E_{inner}^{QM+MMcharges} + (E_{outer}^{MM} - E_{inner}^{MM}) \quad (1)$$

In this study we explore whether QM/MM can be used to supplement X-ray crystallography in SBDD programs, whereby high quality theoretical structures of analogs incorporating relatively small iterative changes could be derived from the former, to rapidly answer specific hypotheses in a cycle time more aligned with medicinal chemistry synthesis. We do this by assessing the performance of a range of computational methods to describe the interactions in 4 separate protein kinase complexes (Table 1), these representing a target class where SBDD plays a crucial role. To this end we employ a QM/MM model where the inner region consists of the ligand and the backbone of 3 amino acids of the so-called hinge,<sup>46</sup> and the remaining protein is modeled using MM. The QM region is fully optimized using 4 different levels of theory (B3LYP/6-311++G\*\*, B3LYP/6-31G\*\*, HF/3-21G\*, and PM3), and the MM region is treated using the Universal Force Field (UFF). A purely force field based solution is also derived using the Merck Molecular Force Field (MMFF) for the purpose of comparison. To test the effectiveness of each theoretical method we assess the concordance of the theoretical coordinates to those of the original crystallographic structures using the root mean

**Table 1.** List of Protein–Kinase–Inhibitor Complexes Used in This Study

Kinase	PDB ID	Structure	Resolution
PKA/B	2uvx		2.0Å
CDK2	1wcc		2.2Å
GSK3	1o9u		2.4Å
P38	1w7h		2.2Å

squared deviation (RMSD), the computed H-bond distances, and the electron density where available.

## 2. COMPUTATIONAL PROCEDURES

Crystal structures of 4 protein-kinases (2UVX, 1WCC, 1O9U, 17WH) containing small molecule ligands were downloaded from the RCSB protein databank<sup>47</sup> (Table 1). As is common in many theoretical studies,<sup>19</sup> cofactors, ions, or water molecules are removed from the protein–ligand complexes for computational efficiency, and the impact of this will be discussed later. Each protein was subsequently prepared using the protein preparation wizard in Maestro.<sup>48</sup> Hydrogen atoms were added to the system, and ionizable amino acid side chains were protonated assuming a pH of 7.4. The system then underwent restrained minimization using the IMPREF utility, only to optimize hydrogen atoms and to remove any high energy contacts or distorted bonds, angles, and dihedrals. These coordinates were used for the subsequent QM/MM calculations.

All QM/MM calculations were performed in Gaussian 03<sup>49</sup> using the ONIOM methodology developed by Morokuma and co-workers.<sup>39,40</sup> Our strategy is to use the highest levels of QM possible for a reasonable computational resource, to mirror how one might use the methodology in a program environment where rapid turnaround is required. To this end we have employed the smallest acceptable QM region consisting of the bound ligand and the backbone of the 3 amino acids of the hinge, with the point charges of the outer MM region being electrically embedded. Where bonds cross the QM and MM interface, the valences were satisfied by hydrogen link atoms. In this case the outer region is kept fixed, and the charges are assigned using the charge equalization method.<sup>50</sup> The latter method has been demonstrated to produce molecular charges that correlate well with more expensive QM calculations. The inner QM region was treated using a range of different levels of theory: B3LYP/6-311++G\*\*, B3LYP/6-31G\*\*, HF/3-21G\*, and PM3. Electrical embedding was used in all cases apart from PM3. Mechanical embedding was used as the former is not supported (Gaussian 03 enables both mechanical and electrical embedding to be used in ONIOM calculations, unlike earlier versions where only the former is supported). The van der Waals contribution to the protein–ligand complexes

**Table 2.** QM/MM and MM Results for the 4 Optimized Protein–Inhibitor Complexes<sup>a</sup>

ID	model	RMSD	HB1 (center)	HB2 (inner)
PKA/B 2UVX	XRAY (2.0 Å Res.)	-	1.98	1.93
	B3LYP/6-311++G**	0.19	1.99	1.81
	B3LYP/6-31G**	0.19	1.97	1.80
	HF/3-21G*	0.19	1.92	1.76
	PM3	0.20	1.87	1.81
	MMFF	0.26	1.82	1.65
CDK2 1WCC	XRAY (2.2 Å Res.)	-	2.11	-
	B3LYP/6-311++G**	0.31	2.08	-
	B3LYP/6-31G**	0.33	2.06	-
	HF/3-21G*	0.34	1.98	-
	PM3	0.34	1.89	-
	MMFF	0.34	2.41	-
GSK3 1O9U	XRAY (2.4 Å Res.)	-	2.24	1.82
	B3LYP/6-311++G**	0.77	2.04	1.76
	B3LYP/6-31G**	0.78	2.01	1.74
	HF/3-21G*	0.78	2.01	1.72
	PM3	0.77	1.86	1.82
	MMFF	0.74	2.11	1.76
P38 1W7H	XRAY (2.2 Å Res.)	-	2.06	2.04
	B3LYP/6-311++G**	0.26	2.07	1.82
	B3LYP/6-31G**	0.24	2.04	1.81
	HF/3-21G*	0.41	1.95	1.81
	PM3	0.30	2.35	1.90
	MMFF	0.61	1.88	1.48

<sup>a</sup> Listed are the heavy atom RMSD to the original crystallographic coordinates and the H-bond distance between the inhibitor acceptor and the central hinge donor (HB1), and between the inhibitor donor and the inner hinge acceptor (HB2). Hydrogen atoms were added to the original X-ray structure using the OPLS-AA force field as implemented in Maestro.

was treated classically using the universal force field (UFF).<sup>51</sup> While this combination has been successfully applied in the past for calculations on large representations of 3-dimensional zeolites catalysts,<sup>52–54</sup> some concerns have been raised with QM/MM methods, including the use of nonideal empirical functions (i.e., Lennard-Jones potentials) to compute interactions across the QM and MM interface and the issue of overpolarization of the QM region.<sup>55</sup> With these issues in mind, we assess the utility of the method to protein kinases by comparison of the theoretical coordinates to those from the original X-ray experiment.

All Gaussian 03 calculations were submitted as single processor jobs on a Linux cluster and took no more than 2 days to converge.

For the purpose of comparison the prepared protein–ligand complexes were also assessed using force field methods alone. In this case the whole protein was fixed, and the ligands alone were optimized using MMFF as implemented in MOE.<sup>56</sup> More conventional force field methods for simulating protein–ligand complexes were not considered due to the relatively time-consuming setup required to prepare each ligand correctly.<sup>57,58</sup>

## 3. RESULTS

The RMSDs and H-bond distances obtained from the five different calculations on each of the four protein–ligand complexes are reported in Table 2. We first discuss this output obtained in terms of the RMS deviation between the experimental and theoretical coordinates as this is the most commonly used parameter in such studies. The results in



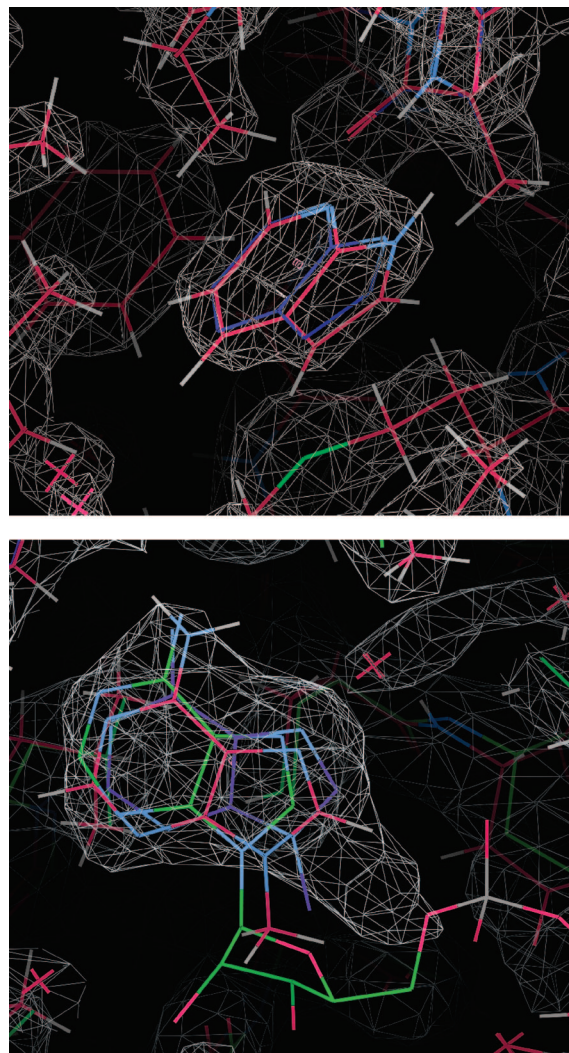
Table 2 unsurprisingly show that the lower the resolution of the X-ray structure the larger the RMSD, irrespective of the theoretical model. The lowest resolution structure, 1O9U, is solved to 2.4 Å resolution, and the average RMSD of theoretical models is  $\sim 0.8$  Å. In contrast 2UVX, the highest resolution structure at 2.0 Å, displays a mean RMSD of  $\sim 0.2$  Å. In between these, 1WCC and 1W7H are solved to a resolution of 2.2 Å, and the RMSDs of the theoretical structures are  $\sim 0.3$  Å on average. It is reported that the experimental atomic positional errors for complexes with resolutions between 1.8 and 2.0 Å will range from  $\sim 0.2$ – $0.3$  Å.<sup>12,14</sup>

The highest level B3LYP/6–311++G\*\*//UFF results do not show demonstrably different RMSDs than any of the other theoretical models. This however may be a reflection of the fact that the 1-dimensional RMSD value captures an average effect and because the X-ray coordinates themselves are not error free. In fact, the focus on this parameter in docking studies has come in for criticism recently as they themselves are a fitted atomic model derived from the electron density.<sup>59</sup> We therefore conclude that starting from the X-ray coordinates, all methods give qualitatively the same minima. However, as we shall see later, the HF and DFT methods more effectively reproduce the H-bond distances and orientations than the more approximate PM3 and MMFF based methods.

RMSD values of  $\sim 0.3$  Å could be naively interpreted as meaning the theoretical models are equally plausible representations of the electron density since they lie within the error of the X-ray coordinates. To assess this further we compared the X-ray structure and B3LYP/6–311++G\*\* optimized structure to the original electron density for the two complexes where it has been reported (Figure 3).

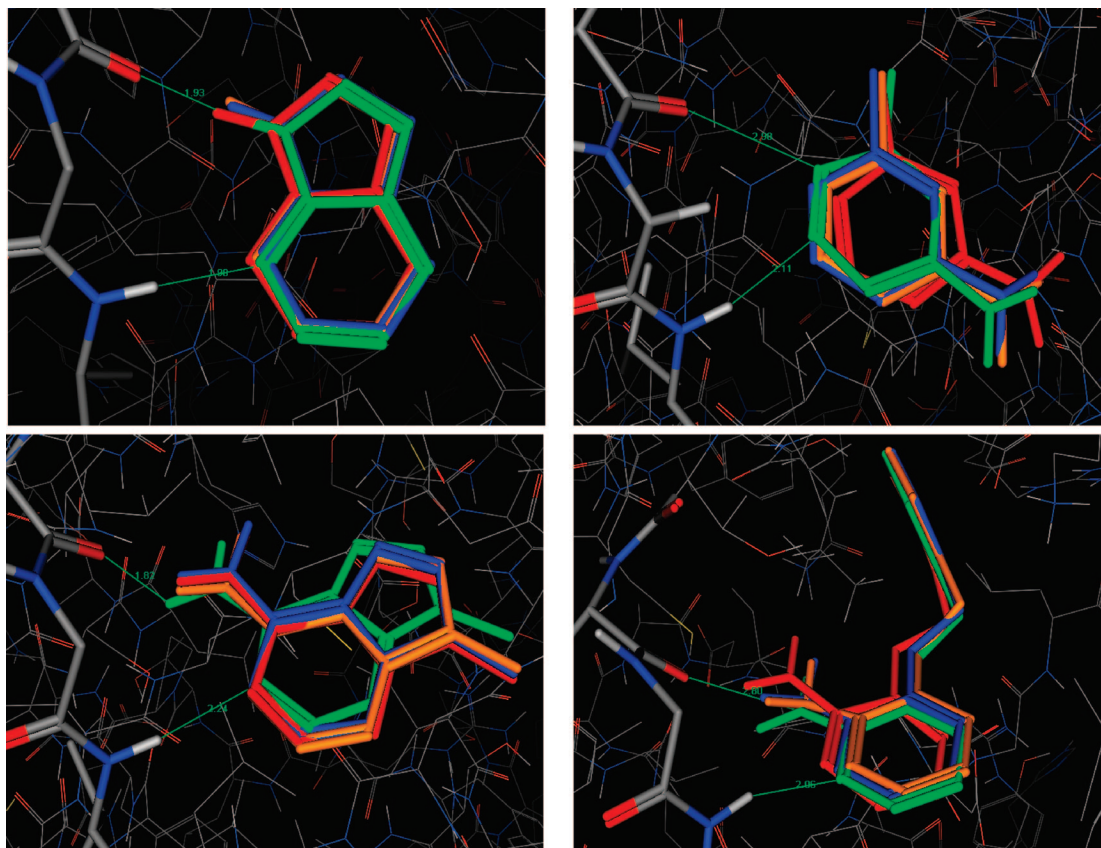
In 2UVX (2.0 Å) the solved heavy atom coordinates of the ligand lie directly within the well defined electron density, giving us confidence in the experimental result. The QM/MM results can be distinguished from the latter by the presence of the hydrogen atoms, and from Figure 3a it is apparent that the B3LYP/6–311++G\*\* optimized coordinates also fit perfectly within the electron density. Interestingly, the indole proton is not planar with the aza-indole ring, distorting to a dihedral angle of  $\sim 12^\circ$  to maximize the H-bond interaction with the inner hinge acceptor. While the gas phase optimized structures are found to be planar, the difference in energy is typically  $< 0.5$  kcal/mol meaning the improved H-bonding interaction overcomes the small distortion cost. Experimental evidence for this type of effect can be seen from both microwave studies of common substituents such as aniline<sup>60</sup> and from an analysis of small molecule structures reported in the CSD where hydrogen atom positions have been resolved. A search for aniline containing small molecule structures also reveals numerous examples of NH bonds deviating from the expected planarity where favorable H-bonds are made ( $\sim 50\%$  deviating  $> 10^\circ$ ). These subtleties would not be picked up using force field methods. More computational intensive calculation may be needed to study this issue further (i.e. MP2) due to known deficiencies of DFT methods.

For 1O9U (2.4 Å) the situation is less clear. As discussed above the theoretically optimized coordinates, including those at the highest B3LYP/6–311++G\*\* level of theory, display rather large RMSDs to the solved crystal coordinates, being



**Figure 3.** Illustration of the reported electron density of 2UVX (2.0 Å) (top) and 1O9U (2.4 Å) and how they relate to the solved X-ray coordinates (blue) and those determined using QM/MM (red). For 1O9U the higher resolution structure (1j1b, 1.8 Å) is shown for the purpose of comparison (green). The electron density (2fo-fc) is contoured at the  $1.0\sigma$  and  $1.5\sigma$  levels, respectively.

the largest observed for all the 4 complexes considered here. This is inline with the relatively low resolution indicating the result could be a function of the experimental data, the theoretical method, or a combination of both. In Figure 3b one observes a sizable shift between the X-ray coordinates of 1O9U and those derived at the B3LYP/6–311++G\*\*//UFF level of theory, which could be taken to mean the latter method has performed poorly. However, the solved X-ray coordinates do not fit into the experimental electron density particularly well, with a sizable portion of undefined density lying toward the phosphate binding pocket. Comparison of this X-ray structure to the ATP mimetic found in 1J1B reveals the latter lies much closer to the hinge and does not put its ribose ring anywhere near the ambiguous density. The 1O9U result could have arisen due to incomplete crystal lattice occupancy. This is where other unit cells might contain differing mixtures of ligand, solvent and co-factors, and in such situations efforts to obtain an atomic solution for one ligand will be detrimentally affected by the ambiguous density. From Figure 3b it is clear the QM/MM optimized 1O9U structure occupies a halfway position between the two X-ray structures. It is again interesting to note that the



**Figure 4.** The optimized coordinates of the protein–inhibitor complexes for PKA/B-2UVX, CDK2-1WCC, GSK3-1O9U, and P38-17wh displayed counterclockwise from top left. Ligands are colored as follows: B3LYP/6–311++G(d,p)//UFF (blue), PM3//UFF (orange), and UFF (red), and the original crystallographic coordinates (green). Ligands and protein hinge are denoted in stick form.

aromatic-amine substituent of the QM/MM optimized structure is nonplanar, one hydrogen being  $\sim 4^\circ$  and the other  $\sim 13^\circ$  to the plane of the ring to maximize the interaction with the hinge acceptor.

**Relative Performance of the Different QM/MM Methods.** To allow for a more detailed examination of the differences between the different theoretical structures and experimental structure, a graphical illustration of the 3D coordinates is given in Figure 4 and a plot of the corresponding H-bond distances in Figure 5. A visual inspection of the 3D coordinates derived using the B3LYP/6–311++G\*\*//UFF, PM3//UFF and MMFF methods shows that they are all qualitatively very similar. However, from Figure 5 it is clear that the more accurate computational methods generally display H-bond distances (either to the outer hinge donor, central acceptor, or inner donor) that are closer to the original X-ray coordinates. PM3 and MMFF show much greater variability than the DFT or HF methods, MMFF in particular displaying either much too short (1WCC and 17wh) or too long H-bonds (2UVX). It is reasonable to assume that this is a result of approximations in empirical and semiempirical methods to make them computationally more attractive and the lack of polarization in both methods.

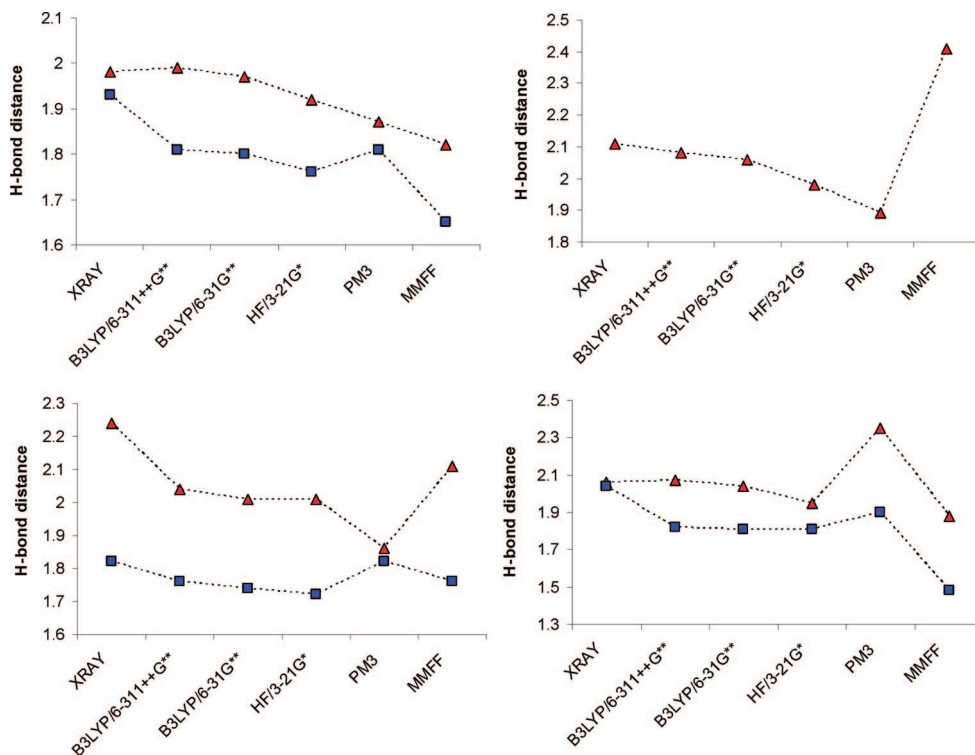
For 2UVX all theoretical methods display comparable RMSDs, but analysis of Figure 4 shows MMFF to be shifted toward the hinge, explaining the unnaturally short H-bond distances observed in Figure 5. It is found for QM/MM methods that as the underlying QM description increases in rigor so too does the correspondence with experiment, in particular performing better than the empirical MMFF model.

Nonetheless, the DFT and HF methods predict H-bonds distances of unequal length to the hinge, unlike the X-ray structure. The highest level B3LYP/6–311++G\*\*//UFF results lie within the error of the experimental data based on the RMSD, and given that the X-ray structure is itself based using an iteratively solved, parametrized method that is known to introduce errors, it is interesting to consider which more accurately reflects reality.

For 1WCC the QM/MM methods display much lower RMSDs than the MMFF conformation, and it can be seen from Figure 4 that this is due to the rotation of the ligand as a result of the perceived unfavorable CH–N interaction with the hinge. This however can be categorized as a weak H-bond interaction<sup>33</sup> which is picked up the QM based methods. We also see that as the accuracy of the QM method increases the predicted H-bond distance approaches that of the X-ray structure.

In 1O9U all QM/MM methods display rather large RMSDs, and as discussed earlier, this is likely to have arisen due to experimental limitation of the lower resolution data. Each of the theoretically derived structures displays a comparable RMSD, and the 3D coordinates are essentially equivalent as can be seen from Figure 4. This makes comparisons of the H-bond distances rather difficult. However, the higher resolution (1.8 Å) ATP mimetic found in 1J1B displays a H-bond distance to the hinge acceptor of 1.8 Å and 2.1 Å to the inner hinge donor, which are very similar to those obtained theoretically. The main difference observed between the different theoretical structures is the short H-bond to the inner hinge donor predicted using PM3.





**Figure 5.** Plot of the H-bond distances observed in the crystal structure and each of the 6 different theoretical models considered. Displayed clockwise from top left are as follows: PKA/B-2UVX, CDK2-1WCC, GSK3-1O9U, P38-17wh. The hinge donor mediated H-bond is denoted with a triangle and the acceptor mediated H-bond with a square.

Both H-bonds to the hinge are equivalent length according to PM3, in contrast to experiment and higher level calculations. This is likely to be a result of the fundamental approximations used within the semiempirical method and also the lack of electrical embedding. Interestingly this is not observed for the purely empirically MMFF method.

For the 1W7H protein–ligand complex the DFT based calculations show distinctly lower RMSDs and H-bonding pattern closer to experiment compared to the other methods used. A visual inspection of the complexes in Figure 4 shows that the MMFF optimized ligand is shifted in toward the hinge, displaying unnaturally short H-bonds (1.5 Å to the hinge acceptor). In contrast, the PM3 derived structure lies more distant from the hinge (both H-bonds >2.3 Å). As we have observed with 2UVX, the B3LYP/6–311++G\*\* result lies closer to the X-ray results overall; however, the predicted H-bonds are of different lengths. This is because the aniline-like amine hydrogens are not planar with the ring, distorting to maximize the interaction with the hinge acceptor as discussed earlier. In fact the amine nitrogen atoms and the other heavy atoms of the ring lie almost on top of each other as can be seen from Figure 4.

An additional advantage of QM/MM methods are that ligand polarization is directly accounted for, unlike conventional MM methods.<sup>24</sup> The net ESP charge on the ligand at the B3LYP/6–311++G\*\*//UFF level point to significant charge transfer within the QM region with values of 0.16 au for 2UVX, 0.12 for 1WCC, 0.13 for 1O9U, and 0.31 for 1W7H. This indicates electron density is being pulled from the ligand via the H-bond interactions onto the amino acid backbone of the hinge. This finding is in line with reports of other QM/MM simulations on other proteins.<sup>61</sup>

It is also necessary to acknowledge the importance of protein flexibility or the presence or absence of water

mediated H-bonds on these types of calculations. A larger representation of the QM region, coupled with flexible active site amino acid side chains in the MM region, could be used; however, this will be more time-consuming and will not produce the more substantial movements seen in protein kinases such as LCK (DFG loop outward rotation<sup>62</sup>) or EGFR (C  $\alpha$ -helix shift<sup>63</sup>).

For QM/MM methods to be useful in SBDD they must be relatively quick to compute. The B3LYP/6–311++G\*\*//UFF calculations performed here took no more than 2 days to complete as single processors jobs. Even though the ligands used in this study were fragmentlike in size, studies on more druglike molecules would be feasible if submitted as multiprocessor jobs or run at a more modest level of theory such as B3LYP/6–31G\* or HF/3–21G. Additionally, it could be argued that there is an even greater need for more accurate methods to describe protein–ligand complexes of larger, druglike molecules since the greater number of interactions present means the flaws of empirical or semiempirical method will have an additive detrimental effect on the theoretically derived coordinates and by extension the interaction energies or score.

#### 4. CONCLUSIONS

QM/MM calculations on four protein kinase complexes have been undertaken to assess the utility of QM/MM in SBDD environment. The results indicate that irrespective of the method, the optimized coordinates for each method are qualitatively similar in terms of the computed RMSD. However, looking at the results in a less simplistic way one finds the more rigorous B3LYP/6–311++G\*\*//UFF and B3LYP/6–31G\*\*//UFF methods perform better over all four cases, leading to optimized structures that deviate least from the

experimental coordinates in terms of the H-bond distances and orientation. The semiempirical and empirical methods displayed more variable performance. These results collectively suggest that the use of a DFT/UFF ONIOM scheme is sufficiently accurate to probe protein–ligand complexes that are of interest to drug discovery programs.

It can be argued that such accuracy is not required in the majority of SBDD applications, especially when empirical methods give results that are qualitatively similar. On the other hand the computation of accurate energetics, to allow the ranking/scoring of different docking poses for example, must logically require more accurate 3D coordinates and energy functions, given the failure of traditional empirical docking and scoring methods.

The utility of QM/MM to a program team will ultimately be determined by the nature of the problem facing the program, the size of the system needing simulation, the available computational resources, and the program timescales. For the relatively small number of programs that can currently comply with these conditions, the extra precision of the QM/MM methods may prove useful in (1) probing the nature of the interactions within the protein active site and the effect of molecule conformation,<sup>64</sup> (2) facilitating the atomic interpretation of poorly resolved electron density, and (3) studying the impact of subtle substituent changes on the binding conformation or in the assessment of alternate scaffolds.

This study is the first step in our assessment of QM/MM methods in SBDD. These results have allowed us to determine the most efficient model that can reproduce the experimentally determined structures to a high degree of accuracy. This information is now being used in a cross-docking evaluation of fragment-like leads to assess the utility of the method in a fragment based drug discovery/reduced complexity screening environment.<sup>65–67</sup>

#### ACKNOWLEDGMENT

M. P. Gleeson would like to thank Drs. Stefan Senger, Andrew Leach, Mike Hann, Paul Bamborough, Francis Atkinson, Joelle Le, Don Somers, Chun-wa Chung, Iain McLay, Darren Green, and Colin Edge for proofreading the manuscript and providing very useful input.

#### REFERENCES AND NOTES

- Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **1994**, *37*, 1035–1054.
- Structure based ligand design*; Gubernator, K.; Bohm, H. J., Eds.; Wiley-VCH: Weinheim, 1998.
- Babine, R. E.; Bender, S. L. Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design. *Chem. Rev.* **1997**, *97*, 1359–1472.
- Anderson, A. The process of structure based drug design. *Chem. Biol.* **2003**, *10*, 787–797.
- Marrone, T. J.; Briggs, J. M.; McCammon, J. A. Structure-based drug design: computational advances. *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71–90.
- Kubinyi, H. Combinatorial and computational approaches in structure-based drug design. *Curr. Opin. Drug Discovery Dev.* **1998**, *1*, 16–27.
- Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M., Jr. The role of quantum mechanics in structure-based drug design. *Drug Discovery Today* **2007**, *12*, 725–731.
- Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem., Int. Ed.* **2003**, *42*, 2718–2736.
- Teague, S. J. Implications of protein flexibility for drug discovery. *Nature* **2003**, *2*, 527–541.
- Davis, A. M.; Teague, S. J. Hydrogen Bonding, Hydrophobic Interactions, and Failure of the Rigid Receptor Hypothesis. *Angew. Chem., Int. Ed.* **1999**, *38*, 736–749.
- Engh, R. A.; Huber, R. Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1991**, *47*, 392–400.
- Kleywegt, G. J.; Jones, T. A. Good Model-building and Refinement Practice. *Methods Enzymol.* **1997**, *277*, 208–230.
- Branden, C. I.; Jones, T. A. Between objectivity and subjectivity. *Nature* **1990**, *343*, 687–689.
- Bostrom, J. Reproducing the conformations of protein-bound ligands: A critical evaluation of several popular conformational searching tools. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1137–1152.
- GOLD*; CCDC, Business & Administration, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, U.K. <http://www.ccdc.cam.ac.uk> (accessed Nov 20, 2008).
- GLIDE*; Schrodinger, Dynamostasse 13, D-68165 Mannheim, Germany. [www.schrodinger.com](http://www.schrodinger.com) (accessed Nov 20, 2008).
- COMFA*; Tripos, 1699 South Hanley Road, St. Louis, MO 63144-2319 U.S.A. [www.tripos.com](http://www.tripos.com) (accessed Nov 20, 2008).
- Tame, J. R. H. Coring functions: A view from the bench. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 99–108.
- Warren, G. L.; Andrews, C. A.; Capelli, A.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- Enyedy, I. J.; Egan, W. J. Can we use docking and scoring in lead optimization. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 161–168.
- Kim, R.; Skolnick, J. Assessment of Programs for Ligand Binding Affinity Prediction. *J. Comput. Chem.* **2008**, *29*, 1316–1331.
- Johnson, S. R. The Trouble with QSAR (or How I Learned To Stop Worrying and Embrace Fallacy). *J. Chem. Inf. Model.* **2008**, *48*, 25–26.
- Hawkins, D. M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12.
- Halgren, T. A.; Damm, W. Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242.
- Stouch, T. R.; Kenyon, J. R.; Johnson, S. R.; Chen, X. Q.; Doweyko, A.; Li, Y. J. In silico ADME/Tox: why models fail. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 83–92.
- Barril, X.; Hubbard, R. E.; Morley, S. D. Virtual Screening in Structure-Based Drug Discovery. *Mini-Rev. Med. Chem.* **2004**, *4*, 779–791.
- Weaver, S.; Gleeson, M. P. The importance of the domain of applicability in QSAR modeling. *J. Mol. Graphics Modell.* **2008**, *26*, 1315–1326.
- Peters, M. B.; Raha, K.; Merz, K. M., Jr. Quantum mechanics in structure-based drug design. *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 370–379.
- Senger, S.; Chan, C.; Convery, M. A.; Hubbard, J. A.; Shah, G. P.; Watson, N. S.; Young, R. J. Sulfonamide-related conformational effects and their importance in structure-based design. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2931–2934.
- Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *5–6*, 490–519.
- Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- Hao, M.-H.; Haq, O.; Muegge, I. Torsion angle preferences of small-molecule ligands bound to proteins. *J. Chem. Inf. Model.* **2007**, *47*, 2242–2252.
- Rozas, I. On the nature of hydrogen bonds: an overview on computational studies and a word about patterns. *Phys. Chem. Chem. Phys.* **2006**, *9*, 2782–2790.
- Grimme, B. Do Special Noncovalent p–p Stacking Interactions Really Exist. *Angew. Chem., Int. Ed.* **2008**, *47*, 3430–3434.
- Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210–1250.
- Warshel, A.; Levitt, M. Theoretical Studies of Enzymatic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.
- Singh, U. C.; Kollman, P. A. A Combined Abinitio Quantum-Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular-Systems - Applications to the CH<sub>3</sub>Cl+Cl<sup>-</sup> Exchange-Reaction and Gas-Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718–730.
- Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

- (39) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. *J. Mol. Struct.* **1999**, *1–21*, 461–462.
- (40) 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.
- (41) Dixon, S. L.; Merz, K. M., Jr. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104*, 6643–6649.
- (42) Claeysens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. A.; Schutz, M.; Thiel, S.; Thiel, W.; Werner, H. High-Accuracy Computation of Reaction Barriers in Enzymes. *Angew. Chem., Int. Ed.* **2006**, *45*, 6856–6859.
- (43) Mulholland, A. J. Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discovery Today* **2005**, *10*, 1393–1402.
- (44) Cho, A. E.; Guallar, V.; Berne, B.; Friesner, R. A. Importance of Accurate Charges in Molecular Docking: Quantum Mechanical/Molecular Mechanical (QM/MM) Approach. *J. Comput. Chem.* **2005**, *26*, 915–931.
- (45) Bowman, A. L.; Ridder, L.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. Molecular Determinants of Xenobiotic Metabolism: QM/MM Simulation of the Conversion of 1-Chloro-2,4-dinitrobenzene Catalyzed by M1–1 Glutathione S-Transferase. *Biochemistry* **2007**, *46*, 6353–6363.
- (46) Liao, J. J.-L. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.* **2007**, *50*, 1–16.
- (47) RCSB Protein databank. <http://www.rcsb.org/> (accessed Nov 20, 2008).
- (48) *Maestro*; Schrodinger, Dynamostrasse 13, D-68165 Mannheim, Germany. [www.schrodinger.com](http://www.schrodinger.com) (accessed Nov 20, 2008).
- (49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; *Gaussian 03, Revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.
- (50) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenare, J. P. The Electronegativity Equalization Method I: Parametrization and Validation for Atomic Charge Calculations. *J. Phys. Chem. A* **2002**, *106*, 7887–7894.
- (51) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (52) Namuangruk, S.; Khongpracha, P.; Pantu, P.; Limtrakul, J. Structures and Reaction Mechanisms of Propene Oxide Isomerization on H-ZSM-5: An ONIOM Study. *J. Phys. Chem. B* **2006**, *110*, 25950–25957.
- (53) Namuangruk, S.; Tantanak, D.; Limtrakul, J. Application of ONIOM calculations in the study of the effect of the zeolite framework on the adsorption of alkenes to ZSM-5. *J. Mol. Catal. A: Chem.* **2006**, *256*, 113–121.
- (54) Jungsuttiwong, S.; Limtrakul, J.; Truong, T. N. Theoretical Study of Modes of Adsorption of Water Dimer on H-ZSM-5 and H-Faujasite Zeolites. *J. Phys. Chem. B* **2005**, *109*, 13342–13351.
- (55) Senthilkumar, K.; Mujika, J. I.; Ranaghan, K. E.; Manby, F. R.; Mulholland, A. J.; Harvey, J. N. Analysis of polarization in QM/MM modelling of biologically relevant hydrogen. *J. R. Soc. Interface* **2008**, *5*, S207–S216.
- (56) *Molecular Operating Environment (MOE)*; Chemical Computing Group, 1010 Sherbrooke St. W, Suite 910 Montreal, Quebec, Canada H3A 2R7. [www.chemcomp.com](http://www.chemcomp.com) (accessed Nov 20, 2008).
- (57) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (58) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kucsera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III.; Roux, B.; Schlenkrich, M.; Smith, J. 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.
- (59) Yusuf, D.; Davis, A. M.; Kleywegt, G. J.; Schmitt, S. An Alternative Method for the Evaluation of Docking Performance: RSR vs RMSD. *J. Chem. Inf. Model* **2008**, *48*, 1411–1422.
- (60) Lister, D. G.; Tylern, J. K. Non-planarity of the Aniline Molecule. *Chem. Commun.* **1966**, *6*, 152–153.
- (61) Raha, K.; Merz, K. M., Jr. Large scale validation of a quantum mechanics scoring function: predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J. Med. Chem.* **2005**, *48*, 4558–4575.
- (62) Yun, C.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K.; Meyerson, M.; Eck, M. J. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2070–2075.
- (63) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.* **2002**, *9*, 268–272.
- (64) Brameld, K. A.; Kuhn, B.; Reuter, D. C.; Stahl, M. Small Molecule Conformational Preferences Derived from Crystal Structure Data. A Medicinal Chemistry Focused Analysis. *J. Chem. Inf. Model.* **2008**, *48*, 1–24.
- (65) Carr, R.; Jhoti, H. Structure-based screening of low-affinity compounds. *Drug Discovery Today* **2002**, *2*, 522–527.
- (66) Leach, A. R.; Hann, M. M.; Burrows, J. N.; Griffen, E. J. Fragment screening: an introduction. *Mol. Biosyst.* **2006**, *2*, 430–446.
- (67) Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856–864.

CI800419J