



QM methods in structure based design: Utility in probing protein–ligand interactions

M. Paul Gleeson^{a,*}, Supa Hannongbua^{a,b}, Duangkamol Gleeson^c

^a Department of Chemistry, Faculty of Science, Kasetsart University, 50 Phaholyothin Rd, Chatuchak, Bangkok 10900, Thailand

^b Center of Nanotechnology KU, Kasetsart University, 50 Phaholyothin Rd, Chatuchak, Bangkok 10900, Thailand

^c Department of Chemistry, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

ARTICLE INFO

Article history:

Received 8 June 2010

Received in revised form

27 September 2010

Accepted 30 September 2010

Available online 8 October 2010

Keywords:

Hydrogen bond strength

Kinase inhibitors

QM

QM/MM

Structure based design

ABSTRACT

Small changes in ligand structure can lead to large unexpected changes in activity yet it is often not possible to rationalize these effects using empirical modeling techniques, suggesting more effective methods are required. In this study we investigate the use of high level QM methods to study the interactions found within protein–ligand complexes as improved understanding of these could help in the design of new, more active molecules.

We study aspects of ligand binding in a set of protein ligand complexes containing ligand efficient, fragment-like inhibitors as these structures are often challenging to determine experimentally. To assess the reliability of our theoretical models we compare the MP2/6-31+G** QM results to the original X-ray coordinates and to QM/MM B3LYP/6-31G**/UFF results which we have previously reported. We also contrast these results with data obtained from an analysis of the distribution of comparable interactions found in (a) high resolution kinase complexes (≤ 1.8 Å) from the PDB and (b) more generic, small molecule crystal structures from the CSD.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Structure based design (SBD) involves the structurally driven, iterative alteration of a molecule to increase its affinity for a given receptor by modulating the fine balance between attractive and repulsive interactions present in the protein–ligand complex. It is known that even small changes in ligand structure can often have very dramatic effects on activity confirming subtle interactions are often at play [1,2]. An ability to more effectively understand and predict these subtleties is one of the next challenges in SBD which will allow now routinely determined 3D crystallographic coordinates to be used in a more quantitative fashion to guide the structural modification of a lead series [3,4]. As a result of this requirement, an increasing number of research articles are being reported in the literature that focus on improving our fundamental understanding of the interactions between molecules including pi–pi stacking [5–8], halogen bond interactions [9,10] and the conformational preferences of functional groups [11,12] amongst others. These studies have realized valuable insights through both the mining of experimental structural databases [3,4,11–14] and quantum mechanical (QM) calculations [5–11,15,16].

Kinases represent an ideal target class to investigate the subtleties of protein–ligand binding since their implication in a wide variety of medical conditions has resulted in a huge amount of high quality structural information being amassed on a wide variety of sub-families and inhibitors [17]. The majority of kinase inhibitors bind to the ATP binding site (Type 1) by forming between 1 and 3 hydrogen bonds (H-bonds) with the 3 amino acids that constitute the hinge region (Fig. 1). These inhibitors are ATP mimics, consisting of either a single or multiple aromatic rings which inhibit the protein through a combination of polar interactions with the so called hinge and non-specific binding due to their relatively high lipophilicity. The interactions formed at the hinge are exemplified by the crystal structure of CDK2 (1WCC, Fig. 1). The inhibitor makes a single H-bond interaction to the central hinge acceptor, and as a result of the conformation of the hinge backbone, makes rather short CH mediated interaction with one of the hinge donors [18,19]. Given the ligand has only 350 μ M affinity for CDK2 and the structure is of resolution 2.2 Å it might be appropriate to assess the binding interactions with those of comparable structures in the PDB, CSD or to QM calculations. This might also be pertinent in light of research that show the ligand density associated with relatively weak binding inhibitors can be less distinct, even for protein structures of good overall resolution [20], which can complicate their application to SBD [21]. Indeed classical molecular mechanical (MM) [22,23] and QM based methods [24,25] have been specifically developed to help improve the active site placement of ligands and co-factors.

* Corresponding author. Tel.: +66 86 5242120; fax: +66 2 5793955.

E-mail address: paul.gleeson@ku.ac.th (M.P. Gleeson).

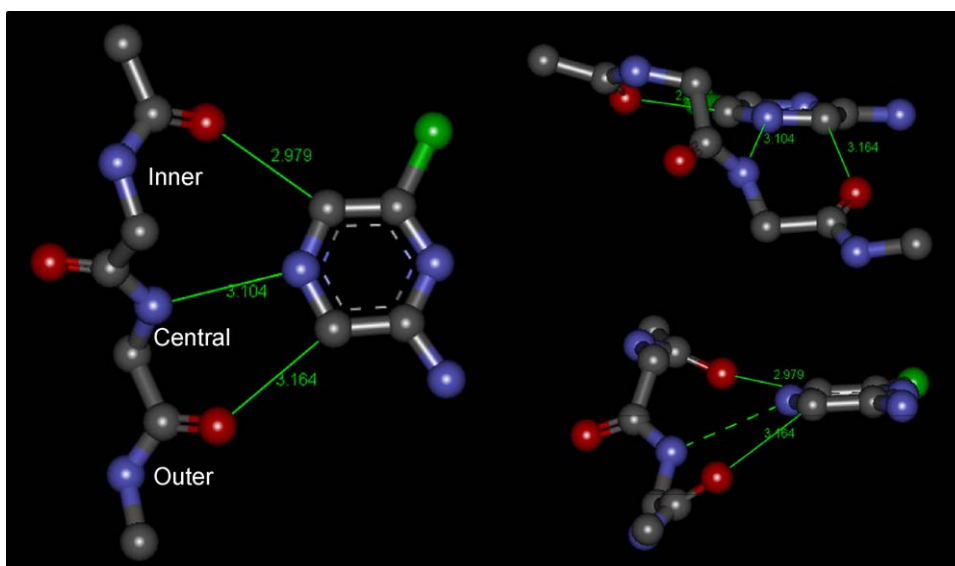


Fig. 1. Interactions formed between a small molecules inhibitor of CDK2 and the hinge backbone (PDB ID 1WCC, 2.2 Å). Clockwise from top right, the images show the binding pocket from a position (1) behind the hinge, (2) looking in from solvent, and (3) looking down from the N-lobe. Note the apparently ideal H-bond distances, angles and dihedrals formed between the inner hinge C=O and H-C of the ligand.

MM techniques are widely used in the pharmaceutical industry as they are both fast and the results are easily interpretable due to the relatively simple terms used in the derivation of their forcefields. They have proved very effective in a wide variety of applications yet it is becoming increasingly clear that these methods may lack the required accuracy to describe the subtleties associated with protein–ligand complexes due to limitations in empirical forcefield themselves [8,9,26–29]. QM calculations are a more accurate alternative to study chemical phenomenon in proteins including aspects of protein–ligand interactions [15,30–35] under investigation here, as well as their better known role in assessing chemical reactivity [36]. QM methods are typically employed on a small representation of the active site, consisting of the key residues for computational efficiency, which means the effects of potentially important longer range interactions will be lacking. This has prompted interest in the so-called QM/MM methods [37–39] which overcome some of the deficiencies associated with both QM and MM methods separately. In this type of calculation, the residues involved in the ligand binding event are treated using QM while the remainder of the protein is modeled using MM. In this way, the critical interactions between the ligand and a receptor are treated QM, the long range electrostatics can be included explicitly and medium range VDW terms accounted for classically.

QM calculations offer certain advantages over QM/MM methods including their ease of setup and the ability to use much higher levels of theory. Thus, the neglect of protein environment is balanced by the fact that we can more accurately model the local interaction between the protein and ligand.

In this study, we assess the utility of QM methods to probe aspects of protein ligand interactions using the experimental X-ray derived atomic coordinates of eight different kinase protein–ligand complexes. High level MP2/6-31+G** gas-phase models are employed to assess the ability of the method to describe the structural properties observed in the active sites. We specifically focus on small kinase inhibitors which interact solely with the 3 amino acid residues of the hinge region so that we can use the highest possible levels of QM theory.

In such cases a small gasphase model might be considered a more reasonable surrogate for the overall system since the protein environment is expected to play a much more marginal role in binding due to the nature of the ligands (at least compared to

popular QM methods to assess H-bond strength based on simple probes [15]). Even so, the lack of the protein environment provides us with a useful opportunity to assess how ideal the ligands in question interact with the key polar moieties in their respective active sites. By neglecting the nonpolar residues in our QM calculations (i.e. residues defining the ATP non-polar cavity), we expect that an inhibitor which optimally interacts with the hinge region will show a gasphase binding mode that will differ only slightly from that in the protein. This situation would suggest that the association between the inhibitor and the polar moiety is optimal (i.e. not unfavorably hindered due to the steric constraint imposed by the protein environment). Furthermore, an inhibitor whose binding mode alters substantially on gasphase optimization must make sub-optimal interactions to the polar active site feature (s) due to steric constraints arising from the protein. In the latter case, an obvious way of increasing inhibitor binding would be to maximize the interaction of the inhibitor with the polar active site features by minimizing the unfavorable steric clash that prevents the interaction from being optimal.

We assess the utility of the gasphase QM models of the 8 kinase complexes (Table 1) by comparing the predicted H-bond interaction characteristics (distances, angles and ligand RMSDs) to the original experimental results and those derived using QM/MM based models previously reported by us elsewhere [26,27] (for 4 out of 8 structures). We also contrast the results to comparable H-bond interactions values obtained from a search of the PDB (kinase type-1 inhibitors with a resolution <1.8 Å) and interactions derived from high resolution small molecule structures from the Cambridge Structural Database (CSD) [40].

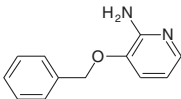
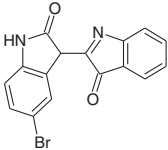

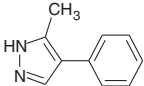
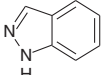
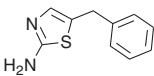
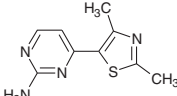
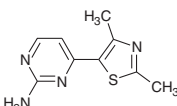
2. Computational procedures

Crystal structures of the 8 protein-kinases listed in Table 1 were downloaded from the RCSB protein databank [41]. These structures were obtained by searching the PDB for small molecule inhibitors that mediate their interaction with the hinge only (i.e. no water molecules). They are of resolutions typically employed in SBD applications.

QM models were created by extracting only the key protein residues that interact with the ligand, and the ligand itself. The QM representation used in this study is exemplified in Fig. 1. It has been

Table 1

PDB structures used in the QM analyses. Reported are the PDB ID, small molecule inhibitor, resolution, kinase target and a description of the H-bonds mediated with the hinge. Outer (O), central (C) and inner (I) H-bonds correspond to those defined in Fig. 1. CH refers to a short interaction distance between a carbonyl group of the hinge and a CH hydrogen atom of the inhibitor.

PDB ID	Inhibitor	Resolution (Å)	Target	H bonds
1W7H		2.2	P38	O(CH), C, I
2BHE ^a		1.9	CDK2	O, C, I
2UVX		2.0	PKA-B	O(CH), C, I
2UW3		2.2	PKA-B	C, I
2VTA		2.0	CDK2	O, C, I(CH)
3DND		2.3	CDK2	O(CH), C, I
2C5O ^a		2.1	CDK2(active)	O, C, I(CH)
1PXJ		2.3	CDK2(inactive)	O(CH), C, I

^a No density available to assess the X-ray atomic coordinates.

used as the core region in QM/MM calculations by both us [26,27] and others [10] to elucidate aspects of non-bonded interactions in kinase-inhibitor complexes. This suggests it is a sufficiently large representation of the key active site interactions for use in these gasphase calculations. Geometry optimization were performed in Gaussian 03 [42] at the MP2/6-31+G** level of theory to accurately account for the non-bonded interactions observed within the model systems. The C_α atoms of the truncated amino acids were frozen during geometry optimization.

QM/MM calculations were also used to compare and contrast the QM results to those of the original X-ray structure. These calculations not only employ the same QM representation but also include the effect of protein environment using an MM potential. The QM/MM results have been obtained from two distinctly different studies reported elsewhere: (a) based on a QM/MM optimization of the ligand in its native protein conformation [27] and (b) based on a QM/MM optimization of the ligand in a non-native conformation of the same protein, where the docked conformation was derived from the 3D alignment of both conformations [26]. Briefly, all QM/MM were performed using the ONIOM methodology developed by Morokuma and co-workers [43,44] and have been reported elsewhere in detail [26,27]. The QM regions were described using the B3LYP/6-31G** method and the MM using the Universal Force Field (UFF) [45]. Hydrogen link atoms were used

to satisfy atoms at the QM and MM interface. Atoms in the QM region were optimized using the electrical embedding scheme. The hydrogen link atoms and MM atoms remained fixed during optimization. Note, all atoms that directly interact with the inhibitors are treated both QM and flexibly. As noted by Bizzanti et al., in such studies it is advisable to keep the protein rigid with the exception of the residues that are known to be flexible [3]. This is because repulsive interactions, either real or due to sub-optimal refinement, can be unrealistically dissipated into the wider protein by very small changes in the overall protein conformation following MM optimization. This could arise due to the limitations of MM force-fields [8,9,24–26] or the coupling between QM and MM regions in QM/MM calculations [46].

We have also extracted information on the H-bond interactions formed between type-1 kinase inhibitors from the PDB with X-ray resolutions ≤1.8 Å to compare with the QM, QM/MM and X-ray structural models. We have also extracted protein non-specific information for comparable H-bonds formed between high resolution small molecules crystal structures in the ISOSTAR database [47]. For all models we determine the distance between HBA and HBD, the corresponding heavy atoms, and the bond angle. As X-ray structures extracted from the PDB lack hydrogens, these were added using the AMBER [48]/GAFF [49] forcefield prior to determining the HBA–HBD distance. RMSDs (in Å) of the QM opti-

Table 2
Characteristics of kinase-inhibitor H-bonds interactions with the hinge region for (a) 57 high resolution kinase-inhibitor PDB complexes, (b) 8 fragment-like kinase-inhibitor PDB complexes and (c) the equivalent QM MP2/6-31+G** optimized coordinates of the 8 PDB complexes. Also reported are the characteristics of comparable interactions found in small molecule crystal structures contained within the ISOSTAR database. In all cases the mean distances (Å), angles (degrees) and number of cases (N) are reported, along with the corresponding standard deviations in parenthesis.

	NH...N(ar)	C=O...H-N(ar)	C=O...H-C(ar)		
A. PDB ≤ 1.8 Å	Central H-Bond	outer H-Bond	inner H-Bond	outer H-Bond	inner H-Bond
X–HY distance	2.09 (0.18)	1.97 (0.14)	1.92 (0.14)	2.60 (0.31)	2.19 (0.18)
X...Y distance	3.06 (0.17)	2.86 (0.12)	2.87 (0.12)	3.29 (0.29)	3.19 (0.14)
Angle	123.4 (3.5)	102.9 (8.7)	133.5 (5.1)	91.7 (14.2)	143.8 (3.9)
N	57	30	25	27	32
B. Isostar	Generic HN...N(ar)		Generic C=O...H-N(ar)		Generic C=O...H-C(ar)
X–HY distance	1.95 (0.16)		1.92 (0.12)		2.40 (0.16)
X...Y distance	2.91 (0.32)		2.87 (0.08)		3.15 (0.08)
Angle	117.6 (12.1)		130.3 (27.7)		118.5 (16.8)
N	737		3991		1841
C. PDB (N=8)	Central H-bond	Outer H-bond	Inner H-bond	Outer H-bond	Inner H-bond
X–HY distance	1.99 (0.35)	1.97 (0.11)	1.93 (0.10)	3.04 (0.67)	2.46 (0.71)
X...Y distance	3.04 (0.35)	2.86 (0.24)	2.88 (0.07)	3.69 (0.48)	3.23 (0.37)
Angle	118.9 (6.4)	117.6 (22.8)	135.8 (5.8)	86.2 (6.7)	146.8 (89.0)
N	8	2	6	5	2
D. MP2/6-31+G** (N=8)	Central H-bond	Outer H-bond	Inner H-bond	Outer H-bond	Inner H-bond
X–HY distance	1.89 (0.04)	2.00 (0.21)	1.90 (0.07)	2.66 (0.62)	2.47 (0.07)
X...Y distance	2.91 (0.04)	2.96 (0.11)	2.91 (0.06)	3.58 (0.55)	3.25 (0.11)
Angle	121.1 (4.5)	117.0 (23.5)	134.2 (4.3)	77.7 (8.0)	150.2 (3.9)
N	2	8	6	5	2

mized coordinates were determined by optimally realigning the molecules to the original X-ray coordinates using ROCS [50]. The RMSDs were calculated using non-hydrogen atoms only.

3. Results and discussion

The theoretical results from the QM and QM/MM models for the complexes in Table 1 can be compared to their original experimental coordinates and to searches of comparable interactions from the PDB and CSD (Table 2). To understand the utility of QM calculations in SBD we must first assess their ability to describe the interactions found within the set of test protein–ligand complexes. For example, in the 2.2 Å CDK2 structure shown in Fig. 1, the ligand makes a single traditional H-bond interaction with the hinge, formed between its aromatic nitrogen acceptor and the hinge central nitrogen donor (3.1 Å). The bond angle of 111° deviates from the idealized value of ~120° that would be expected. It is also apparent that a relatively strong interaction is made between the hinge carbonyl acceptors and aromatic CH atoms ($C_{ar}H$) of the pyrazine ring. The heavy atom distances correspond to 3.0 Å and 3.2 Å for the inner and outer hinge interactions, respectively. These correspond to $C=O...HC_{ar}$ H-bond distances of 2.0 Å and 2.4 Å, respectively when hydrogen atoms are added according to the AMBER forcefield. Surprisingly, the former CH interaction distance is essentially equivalent to a conventional $C=O...HN$ H-bond formed with the central hinge donor [18]. This suggests the $C_{ar}H$ mediated interaction is either abnormally strong interaction or corresponds to a rather subtle error introduced in the PDB refinement process.

Any structural deviation could of course be due to issues with either the QM representation (i.e. level of theory, inadequacies in the gasphase model) or the X-ray derived atomic solution (i.e. the X-ray coordinates are in reality an empirically derived theoretical model fitted to the experimental electron density, solved using classical methods) [20,21,51]. To put these results in better context, and to better understand the characteristics of kinase inhibitors more generally, we have also performed 2 separate database analyses. An analysis of high resolution X-ray structures (≤ 1.8 Å) of

kinase type-1 inhibitors reported in the PDB will give us information about the frequency of interactions made with the 3 possible H-bonds of the hinge (are they equally probable?) and their relative strength. Additionally, an analysis of comparable interactions found in small molecules crystal structures from the CSD will tell us how optimal the kinase based interactions are compared to those formed in a more ideal, less sterically constricted situations.

3.1. Crystallographic analysis of high resolution kinase-inhibitor complexes

We first discuss the results obtained from the PDB and CSD database searching exercises. In Table 2 and Fig. 2, we report the distance between heavy atoms involved in the interaction, the angle associated with the interaction and the actual H-bond distance which was determined by adding hydrogen atoms to the experimental coordinates. Also reported is the number of times that the interaction is classified as being made.

Fifty-seven high resolution type-1 kinase inhibitor–protein structures were extracted from the PDB and the characteristics of their hinge–ligand interactions are summarised in Table 2a. The inhibitors contain at least 1 aromatic heterocyclic ring which is involved in the formation of a strong H-bond to the hinge H-bond donor (central donor). The average distance observed between the corresponding heavy atoms is 3.06 Å and the distribution displays a low standard deviation of just 0.17 Å. The bond angle is only slightly larger (by ~3°) than the “ideal” value of 120°. However, the corresponding average H-bond distance is 2.1 Å, which is nearly 0.15 Å longer than the average distance observed from comparable interactions in the CSD (N = 737) (Table 2b). We also find that the heavy atom distance is 0.15 Å shorter suggesting it is not an artefact of AMBER (H atoms are resolved in the CSD). Furthermore, the bond angles display comparable values to those of the 57 kinase complexes, being ~3° from the ideal value. This suggests the interaction formed between the inhibitors and the central H-bond donor may be sub-optimal due to steric factors arising from the close proximity of the outer and inner hinge acceptor features. Interestingly, this

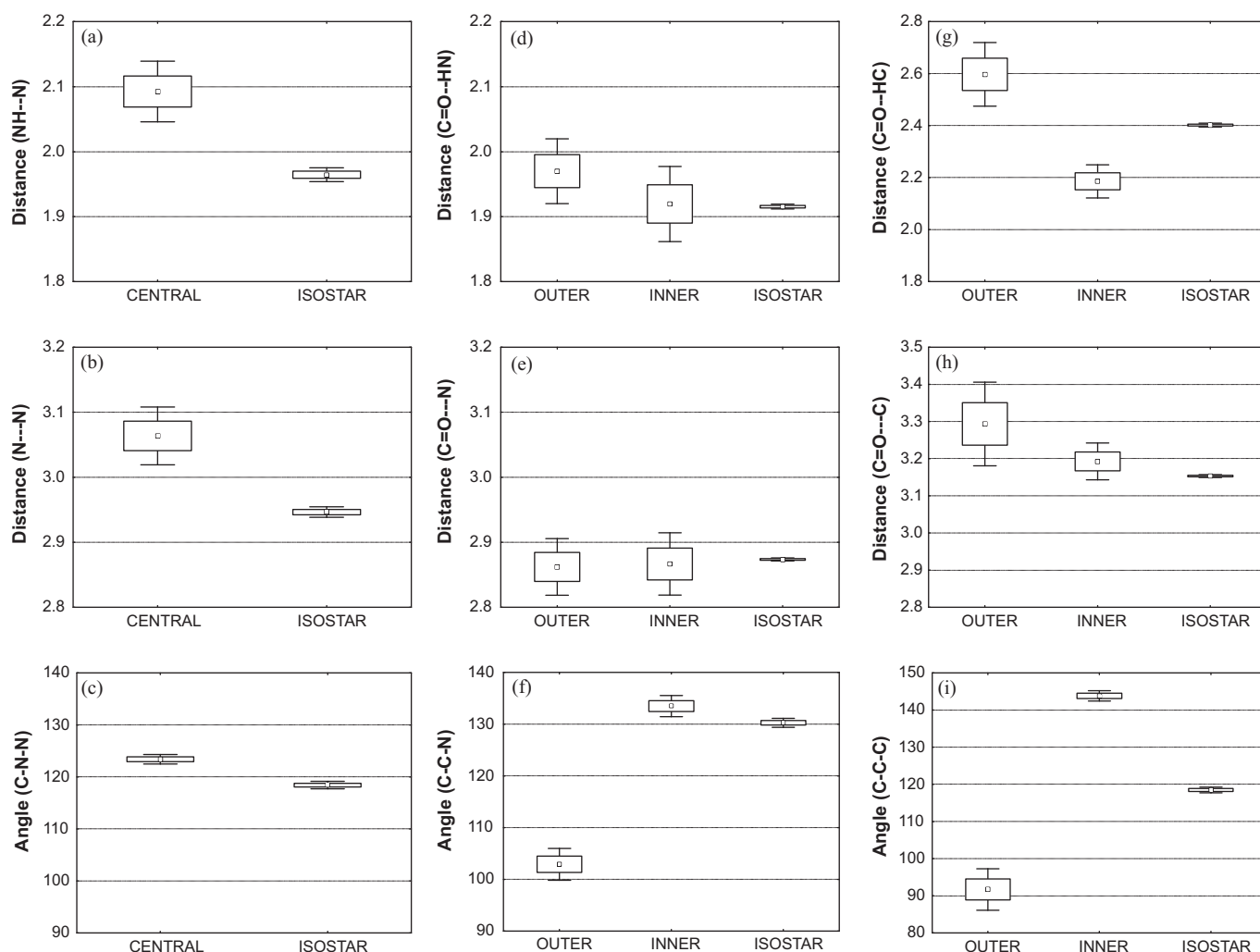


Fig. 2. The characteristics of the central and outer/inner H-bond interactions for 57 kinase X-ray structures with resolutions <1.8 Å. Also shown are the comparable ISOSTAR derived distances. All involve nitrogen containing aromatic heterocycles that interact with an amide donor. Squares denote the 95% confidence in the mean and bars the 99% confidence. Panels A–C correspond to (N–H...N) interactions, D–F correspond to (C=O...HN) interactions and G–I non-classical (C=O...HC) interactions. All distances are given in Å.

analysis suggests that the central H-bond interaction to inhibitors is sub-optimal when compared to a comparable interaction found in the CSD due to the requirement that additional interactions be made with the adjacent hinge acceptors. This can be appreciated visually from Fig. 2 where the differences in the mean values, and the corresponding errors in the mean are given.

Traditional N donor mediated H-bonds are formed to the outer carbonyl acceptor by 30 of the 57 PDB complexes while 25 are formed by the inner acceptor. It is unclear whether this means the outer H-bond distance is preferred given the magnitude of the different populations and the relatively small sample size. The average heavy atom distance in both cases is almost identical at 2.9 Å with bond angles of $\sim 103^\circ$ and 134° . Analysis of the comparable interactions formed by generic small molecules reported in the CSD ($N=3991$) reveals that the mean heavy atom distance obtained from the high resolution PDB set are almost identical (2.87 Å) (Fig. 2). They also suggest that the inner H-bond appears more optimal from an angle perspective. Additionally, a similar preference is noted when looking at the H-bond distances of 1.97 Å and 1.92 Å to the outer and inner donors, respectively. An unpaired *t*-test reveals that the differences in mean distances and angles are significant at the $>90\%$ and $>99\%$ confidence levels respectively ($>95\%$ being considered statistically significant). It is difficult there-

fore to conclude whether one hinge acceptor interaction is truly preferred over the other given that the outer interaction is more frequently made in the set of 57 PDB structures, but when made, the inner interaction appears to be stronger.

In the majority of cases, kinase inhibitors do not make all 3 H-bonds to the hinge that are possible. In the case of 1WCC, for example (Fig. 1), the inhibitor orients two carbon bonded hydrogen (CH) atoms towards the hinge acceptors. CH atoms are of course not strong hydrogen bond donors so this interaction will be viewed as sterically repulsive in SBD terms. From an analysis of the 57 high resolution PDB structures we can see that 32 (56%) of the kinase form CH interactions with the inner hinge donor while 27 form them with the outer hinge donor (47%). Interestingly, while the heavy atom distance between donor and acceptor is comparable (3.2–3.3 Å), because of the angle difference, the interaction distance to the inner H-bond is dramatically shorter (0.4 Å) with a mean value of 2.2 Å. The difference in the mean values is significant above the 99% confidence level suggesting it is a real effect. Analysis of similar H-bond interactions found in small molecule crystal structures reported in the CSD ($N=1841$) reveals that these values are intermediate to the interactions to the outer and inner acceptors found in the PDB set. While these interactions are unquestionably weak [18,19], there may be a benefit in optimizing such weak

interactions, particularly when viewed alongside halogen mediated interactions which are now considered important from a SBD perspective [9,10]. Additionally, this observation raises concerns for the application of MM methods to study such subtle interactions as it is known, for example, that relatively weak, non-standard interactions are not well described using standard MM methods [9].

3.2. QM analysis of fragment-like kinase complexes

The affinity of a ligand for a given receptor depends on a subtle balance of attractive and repulsive forces that define the extent of partitioning between itself and solution. For a good discussion, see Ref. [3]. The free energy of binding to a receptor can essentially be altered by (a) maximising the strength of polar interactions (b) reducing the conformational energy penalty associated with ligand binding to the receptor to a level approaching its energy in solution, or (c) increasing ligand lipophilicity (i.e. the hydrophobic effect), in particular at positions that result in the expulsion of high energy, labile water molecules from the active site.

QM calculations can be employed to analyse certain aspects of this process. For example, QM model calculations could be used to assess how ideal an X-ray binding mode is through the use of a model consisting of the ligand and the key polar residues it interacts with in the active site. Optimization of the QM model in the gasphase will lead to either a large change in the ligand binding conformation and an increase in interaction with the active site model, or no significant change at all. Both results provide us with useful information about the binding mode. Should gasphase optimization lead to a dramatic change in the ligand binding mode we might conclude that steric aspects associated with the protein environment prevent the ligand from forming optimal interactions with the key active site residues. Alternatively, where optimization leads to a negligible change in ligand binding, we might conclude that the interactions present are essentially optimal. This information is important since it allows us to direct synthetic efforts more precisely on finding activity gains by (a) maximizing the key non-bonded interactions in the active site, or alternatively (b) focusing elsewhere in the molecule by searching for additional, more distant polar interactions, or hydrophobic pockets to support targeted lipophilicity increases.

Prediction of the *relative* strength of interactions between simple models of a ligand and its corresponding receptor can be useful to help rationalize structure activity relationships (SAR). This relatively simple approximation has led to the development of in-silico tools using small probe molecules to estimate H-bond strength which can prove very useful in lead optimization efforts [30–35]. However, simplifications involved in these models might mask SAR in cases where the generic probes are not fully representative of the receptor, or the effect of the local active site conformation is important. Kinase ATP site inhibitors are a case in point since up to 3 H-bonds can be formed between the ligand and the important hinge region. Thus an assessment of H-bond strength would ideally be assessed using all the residues involved in the binding process, using the precise hinge conformation from the protein itself. Such target specific models could then be used to assess the strength of interaction resulting from different substituents on the core template or even modifications to the core template.

We now report the results obtained on the target specific QM model calculations of 8 kinase protein–ligand complexes. Complexes were obtained by searching the PDB for kinases containing small molecule inhibitors <350 da, with reasonable resolutions (~ 2.0 Å), where the interactions with the receptor were mediated only through H-bonding with the hinge itself (Table 1). The PDB structures have been reported as part of fragment based screening efforts and are particularly applicable to QM analysis due to their relatively small size. In addition, given their activity is pri-

marily governed by the H-bond interactions with the hinge region, these relatively small QM calculations should be particularly representative. We assess this by comparing results from the QM and QM/MM protein based models. For these structures we also look at the atomic solution derived from the electron density in greater detail, putting particular emphasis on the accuracy on the ligand placement with respect to the ligand electron density as well as conformational aspects of the ligand itself.

Analysis of the average structural parameters for the 8 ligands listed in Table 2c shows that the N-based H-bond interactions are comparable to those from the high resolution PDB set, suggesting they are quite representative set overall. However in the analysis of the often neglected CH based interactions formed with the outer and inner hinge acceptors, we can see the mean distances are ~ 0.4 Å longer than the high resolution PDB set suggesting either (a) the interactions are much more repulsive for these fragment-like

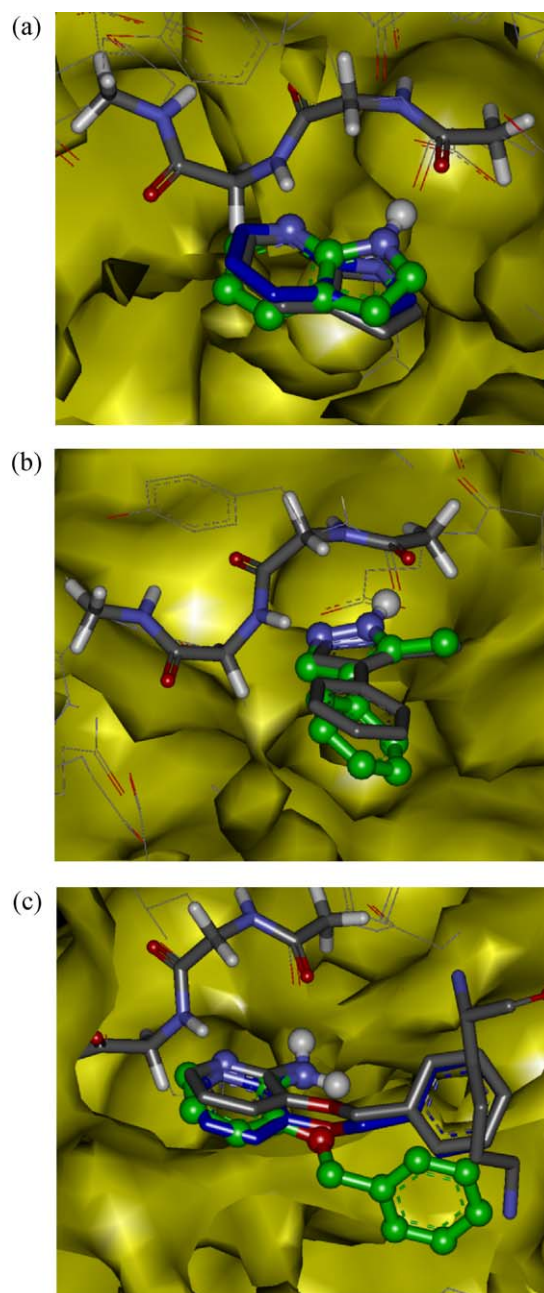


Fig. 3. Superimposition of the QM (green) and QM/MM (blue) and X-ray (grey) models for (a) 2UVX, (b) 2UW3 and (c) 1W7H.

Table 3

H-bond characteristics of 8 kinase-inhibitor complexes derived from X-ray, MP2/6-31+G** QM and B3LYP/6-31G**//UFF QM/MM models. Parameters reported are the same as those in Table 2. Distances are given in Å and angles in degrees and RMSDs in Å. QM/MM results are taken from Refs. [26,27].

ID	Model	RMSD	Outer H-bond				Central H-bond			Inner H-bond			
			X	C=O...H-X	C=O...X	C-C-X	NH...N	N...N	C-N-N	X	C=O...H-X	C=O...X	C-C-X
1W7H	XRAY	–	C	2.59	3.31	87.4	2.00	3.00	127.1	N	2.00	2.80	133.6
	MP2/6-31+G**	1.6		2.42	3.37	78.3	1.86	2.88	124.6		1.96	2.97	127.9
	B3LYP/6-31G**//UFF ^a	0.2		2.69	3.56	84.6	2.04	3.05	126.2		1.81	2.81	131.7
2BHE	XRAY	0.0	N	2.10	3.06	106.4	1.45	2.50	110.6	N	1.95	2.98	141.1
	MP2/6-31+G**	0.2		2.23	3.06	110.5	1.81	2.83	113.3		1.84	2.86	136.3
2UVX	XRAY	–	C	2.57	3.48	80.8	1.92	2.98	122.6	N	1.88	2.88	129.0
	MP2/6-31+G**	0.5		2.39	3.45	73.9	1.87	2.88	127.4		1.84	2.85	133.3
	B3LYP/6-31G**//UFF ^a	0.2		2.53	3.50	80.1	1.97	1.99	122.7		1.80	2.79	127.9
2UW3	XRAY	–	C	3.57	4.10	77.8	1.86	2.91	117.7	N	1.81	2.82	143.9
	MP2/6-31+G**	0.7		3.77	4.55	66.2	1.87	2.89	120.3		1.82	2.84	139.5
	B3LYP/6-31G**//UFF ^b	0.4		3.30	3.97	76.9	1.97	2.97	122.3		2.01	2.98	145.8
2VTA	XRAY	–	N	1.89	2.92	143.9	1.96	3.02	113.5	C	2.96	3.49	153.1
	MP2/6-31+G**	0.3		1.83	2.85	143.1	1.95	2.96	123.2		2.52	3.32	152.9
3DND	XRAY	–	C	3.94	4.29	92.6	2.66	3.70	127.5	N	2.09	2.93	136.2
	MP2/6-31+G**	2.5		2.80	3.28	84.3	1.88	2.91	123.8		1.94	2.95	131.0
2C5O	XRAY	–	N	1.93	2.59	102.6	1.84	2.88	118.2	C	1.96	2.96	140.4
	MP2/6-31+G**	0.9		1.95	2.97	97.4	1.91	2.93	117.5		2.42	3.17	147.4
	B3LYP/6-31G**//UFF ^b	1.1		1.76	2.70	124.4	1.88	2.90	124.4		2.20	3.25	152.4
1PXJ	XRAY	–	C	2.52	3.25	92.2	2.23	3.30	113.7	N	1.85	2.86	130.9
	MP2/6-31+G**	0.2		2.33	3.26	85.8	1.93	2.96	118.9		1.97	2.96	137.2

^a Obtained using the native protein conformation.

^b Obtained using a non-native protein conformation.

inhibitors or (b) the structures themselves are sub-optimal due to their only moderate resolutions.

Looking in more detail at the individual complexes we can see that the interaction characteristics displayed by 1W7H, 2UVX, 2UW3 and 2VTA lie within the expected values obtained from an analysis of the high resolution PDB structures. However, for the other four complexes, 2BHE, 3DND 1PXJ and 2C5O, analysis of their intermolecular interactions reveals uncharacteristically long or short interactions made with the hinge which in some case appear not to be supported by the electron density. We discuss these issues further based on a consideration of the original electron density as well as computationally quite rigorous MP2/6-31+G** QM calculations (Tables 2c and 3).

3.3. Validity of the QM and QM/MM Models

To assess the utility of the QM based models to probe the binding interactions we first focus on those QM models where the structures were in good agreement with the original X-ray coordinates, as well as the interaction characteristics from high resolution kinase structures from the PDB, and CSD experimental H-bonds characteristics. This will allow us to more objectively assess the performance of the QM based models. Additionally, for PDBs 2UVX and 1W7H we have previously reported QM/MM results for the native protein conformation [27] which can be compared to the QM results. In the case of 2UW3 and 2C5O we have also reported the QM/MM results obtained in a non-native protein conformation, where the binding mode was determined by aligning the protein C_α atoms [26] with subsequent optimization. The former QM/MM models provide a useful assessment of how subtle changes in the protein conformation (obtained for a different inhibitor) affect the inhibitor binding mode.

In Fig. 3, the QM, QM/MM and X-ray conformations of 2UVX, 2UW3 and 1W7H have been superimposed using their common QM atoms. The relatively small, inflexible molecules found in 2UVX and 2UW3 are both well modeled using both the QM and QM/MM approach as can be seen from a qualitative assessment of Fig. 3. Moreover, from Table 3, we can see that the QM models display RMSDs to the original X-ray of 0.5 Å and 0.7 Å, respectively, which are marginally larger than those obtained with the QM/MM models (0.2 Å and 0.4 Å, respectively). To put

these values into perspective, it is expected that the intrinsic error in the coordinates of a structure at 2 Å resolution is ~0.2. The relatively large RMSD of 0.7 Å observed for 2UW3 is a result of the subtle movement of its phenyl substituent rather than that of the pyrazole which interacts with the hinge (Fig. 3b). In fact, the interactions between the ligand and hinge are in very good agreement with the QM model results for both PDBs (Fig. 4).

In the case of 1W7H, a molecule with 2 rotatable bonds, the gas-phase representation is not accurate since it lacks the important steric effects of an active site lysine side-chain (Fig. 3c). Nonetheless, these results are still interesting from an SBD perspective since they suggest that this ligand is bound in a relatively high energy conformation and when given sufficient freedom to relax (in the gas-phase), the benzyl substituent conformation significantly alters, improving the interaction with the hinge in the process (Fig. 4). This suggests optimization of the benzyl portion of this chemotype may lead to greater dividends, potentially reducing unfavorable steric interactions in the active site, while simultaneously allowing a more optimal interaction of its pyrazole portion with the hinge. In the case of 2VTA, no QM/MM results were previously reported so we cannot make the comparison, however the predicted H-bond distances are in good agreement with the X-ray structure and the RMSD is very low at 0.3 Å. The only sizeable difference occurs for the C=O...HC interactions distance which is 2.96 Å in the crystal structure but 2.52 Å in the QM model. For comparison, the mean value from the high resolution PDB set is 2.46 Å (Table 2).

Analysis of the theoretical results for 2BHE, 3DND, 1PXJ and 2C5O shows that they are not in as good agreement compared to the other four complexes studied here. Nevertheless, the QM results do reproduce the same overall binding mode, it is only the strength of the interactions between the ligand and the protein that differ. Analysis of the intermolecular interactions found in the X-ray structures reveals significant deviations from the expected parameters derived from the PDB dataset analysis and ISOSTAR. Furthermore, the atomic solutions to the electron density are in some cases not necessarily consistent with the reported density as can be seen in Fig. 5. In each case, 1 or more interaction between the ligand and the hinge shows uncharacteristically short, or long interactions (Table 3), which are highlighted. The average resolution of the four structures is 2.1 Å with none greater than 2.3 Å.

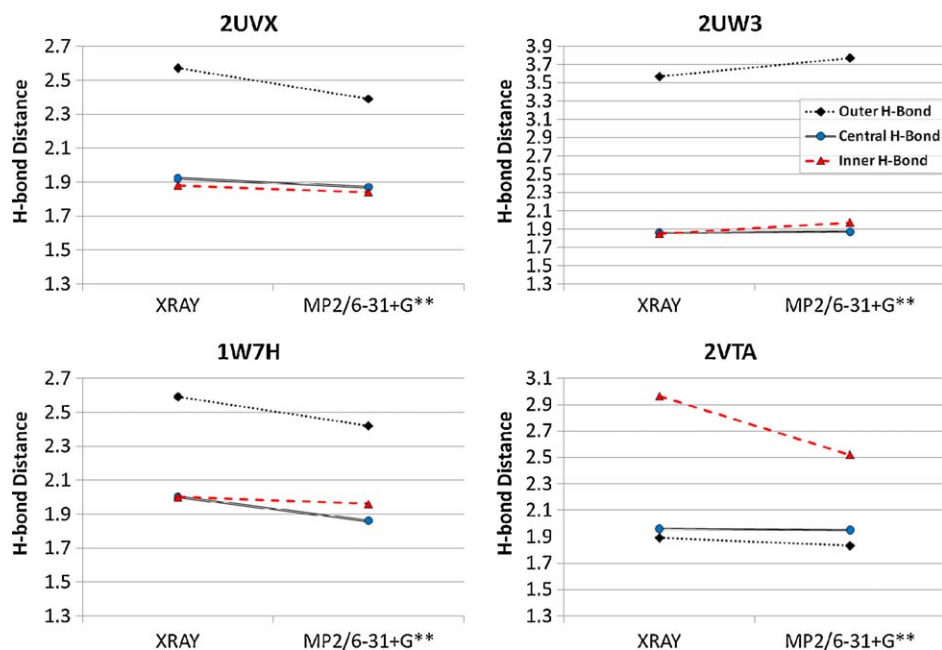


Fig. 4. A comparison of the X-ray and QM model H-bond distances for cases where good agreement is observed. Distances >2.1 Å correspond to CH mediated interactions. Circles with solid lines denote the central H-bond, triangles with dashed line denote the inner H-bond and diamonds with a dotted line denote the outer H-bond.

2BHE: The 2BHE structure is the best resolved here (1.9 Å) yet displays the clearest flaw of all PDB structures under consideration. The outer and inner $\text{C=O} \cdots \text{HN}$ interaction distances lie within those expected from the 57 high resolution kinase structures how-

ever the central $\text{N} \cdots \text{N}$ mediate H-bond is a major outlier. The N-N distance is found to be 2.5 Å, or just 1.5 Å after hydrogen atoms are added according to the AMBER forcefield. This is nearly 3 standard deviations (SD) shorter than the comparable values from the high

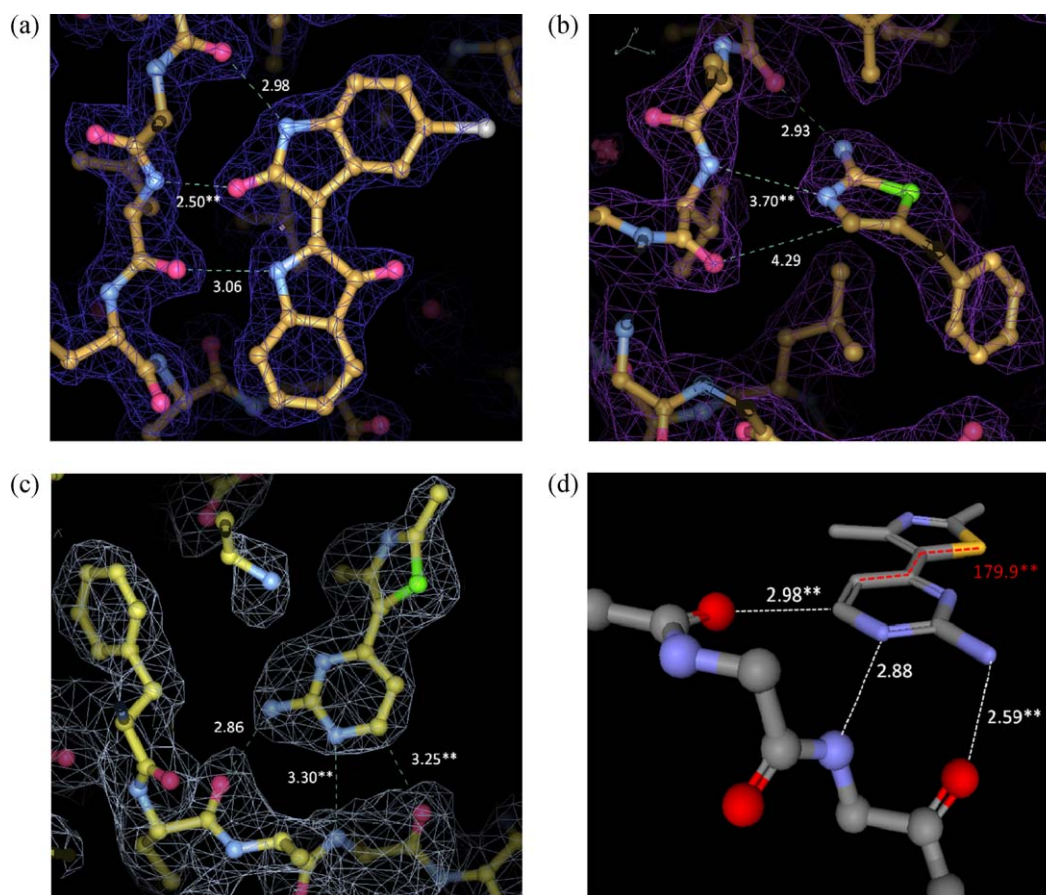


Fig. 5. Experimental atomic coordinates for 4 of the PDBs under consideration here. Clockwise from top left are 2BHE, 3DND, 1PXJ and 2C5O. Uncharacteristically short or long H-bond distances and an unnatural dihedral angle are denoted by **. Density contoured using Coot [52] at $\sim 1.5\sigma$ level. No density has been reported for 2C5O.

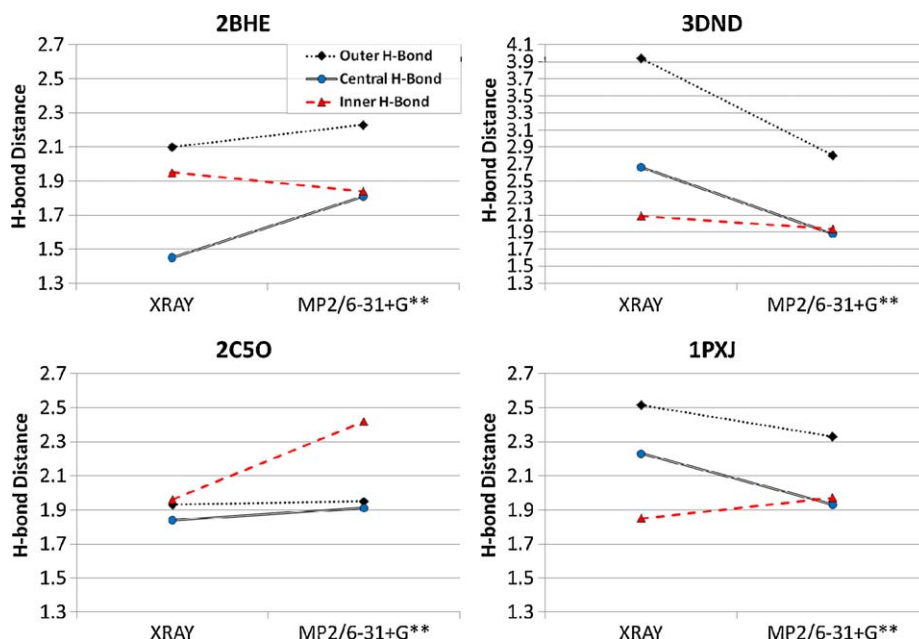


Fig. 6. A comparison of the X-ray and QM model H-bond distances for cases where poor agreement is observed Clockwise from top left are 2BHE, 3DND, 1PXJ and 2C5O. The outer H-bond for 1PXJ and 3DND and the inner H-bond distance for 2C5O correspond to CH mediated interactions. Distances are given in Å. Legend the same as that in Fig. 4.

resolution kinase PDB set (3.1 Å) suggesting it is an error arising from the protein refinement step. Analysis of the electron density in Fig. 5a certainly does not rule this out. The density around the Br atom towards the rear of the pocket is relatively poor compared to other coordinates, and it may be that attempts to better incorporate this less tightly, more flexible (relatively), part of the molecule into the observed density led to the unrealistically short H-bond. Note, while this short distance does not affect the interpretation from a visual, qualitative perspective, it will have profound implications for theoretically derived energetics if computed on this geometry.

The MP2/6-31+G** optimized model displays a central H-bond distance within expected parameters, while the H-bonds made to the outer and inner hinge are comparable to those observed in the original structure. Moreover, the RMSD associated with the QM optimized ligand is just 0.2 Å which suggests that the initial fitting of the ligand to the electron density may have introduced this subtle error.

3DND: The 2.3 Å 2BHE structure is the least resolved of all structures studied here, albeit only marginally so. Thus the difference in resolution is unlikely to explain the unnaturally long H-bond distance made by the inhibitor to the central hinge donor. The N–N distance in the X-ray structure is 3.7 Å which is almost 0.6 Å longer than the mean value displayed by the high resolution kinase PDB structures that all make an equivalent interaction. Given this is the key interaction made by almost all type-1 inhibitors it is surprising to find this distance over 3 SDs away from the mean. Analysis of the electron density in Fig. 5b suggests the ligand position is sub-optimal as it is clear the density surrounding the ligand towards the solvent exposed surface is less well defined. It is also clear that density surrounding the nitrogen donor of the inhibitor clearly supports a much shorter N–N distance. This result highlights the problems of relying on empirical solutions to what is often poor quality ligand density. In such cases ligands can be forced into unrealistic conformation that fit the density better (as in this case) but which are clearly chemically rather suspect.

The 3DND structure is relatively difficult to simulate using QM methods as the presence of a number of amino acid sidechains and water molecules towards the mouth of the pocket help to constrain the conformation of the benzyl ring. In the gasphase calculations

this constraint is lost leading to the rotation and translation of the molecule such that it can form an additional interaction between the outer hinge donor and its 2-position phenyl hydrogen atom (2.3 Å). While this deviation is clearly an artifact of the simplistic QM model, the results are interesting in that they highlight that the interaction between the hinge and the ligand in the protein are sub-optimal (or wrong) given that the H-bond distance decreases by 0.8 Å on optimization. While this structure is incongruent with the electron density, it should also be noted that so too is the original ligand solution reported in the PDB (Fig. 5). In a case such as this a QM/MM model would be more appropriate to assess the validity of the experimental coordinates as the effects of the amino acid sidechains and water molecules towards the mouth of the pocket could be directly taken into account. Nonetheless, the QM calculation still provides useful information from a SBD perspective as it highlights that the interaction of the inhibitor is sub-optimal with the hinge.

1PXJ: The 2.3 Å 1PXJ structure differs from 2C5O in that the CDK2 protein is in an active conformation, leading to a new binding mode to the hinge. The N–N distance associated with the central H-bond is 3.3 Å, corresponding to a relatively weak H-bond of 2.2 Å length on addition of hydrogen atoms. The N–N distance is close to 2 standard deviations from the average value of 2.9 Å derived from 57 high resolution kinase structures suggesting the structure may also have been placed in a somewhat suboptimal position. Furthermore, the outer hinge C=O...HC seems somewhat short, displaying a comparable heavy atom distance to that of the central hinge NH...N based interaction (3.3 Å). Analysis of the electron density is equivocal as the interaction towards the outer hinge is less well defined (Fig. 5c). This suggests this part of the molecule is more flexible meaning the atomic solution of the ligand derived from the electron density will be less reliable.

The relatively long central H-bond distance (2.23 Å) in the 1PXJ PDB structure decreases to 1.93 Å in the QM model, closer to the experimental average expected from the high resolution kinase structures (1.97 Å). The C=O...HC distance also decreases by 0.2 Å in the QM model which appears relatively short at 2.33 Å. The overall RMSD is just 0.2 Å which is within experimental error for a PDB of this overall resolution. The results from the QM analysis, database

searches of comparable high resolution kinase complexes, and a re-investigation of the original electron density suggests the relatively long interaction distances may be an artifact of the protein refinement process.

2C5O: No electron density has been reported for 2.1 Å 2C5O structure so a structural evaluation can only be performed on the atomic solution of the density only (Fig. 5c). Analysis of these coordinates reveals apparent flaws both in terms of the interaction between ligand and hinge and the ligand conformation. Firstly, it is expected that the bi-cyclic ligand molecule will not lie in the same plane to minimise unfavourable VDW repulsion (as is observed in 1PXJ). However, in this case the two rings lie in the same horizontal plane suggesting the initial ligand parameters in the refinement process were suboptimal (~5 kcal/mol energy cost at B3LYP/6-31G* level). This error will have a knock on effect on other aspects of the ligand placement and may explain the non-ideal O–C intermolecular distance observed at the inner hinge (2.98 Å). The resulting H-bond distance is 1.96 Å, almost 0.6 Å shorter than that expected from the 57 high resolution kinase structures (i.e. 2SDs). Furthermore, the outer hinge O–N distance is comparatively short (2.6 Å), 0.3 Å shorter than the ideal distance for a typical kinase C=O...HN bond.

2C5O has been explored by us from both a QM and QM/MM perspective, however both models show rather large ligand RMSDs (0.9 vs 1.1 Å, respectively). It is apparent from the X-ray coordinates that the ligand positioning is rather poor given that the addition of hydrogen atoms to the structure leads to a very short inner C=O...HC interaction distance of 2.0 Å. Furthermore, the two aromatic rings of the ligand lie in the same plane which is in contrast to the same molecule in the active CDK2 structure of 1PXJ. In this case it is difficult to comment on the reliability of the QM or QM/MM approach due to the relatively poor quality of the experimental data. Nonetheless, it is interesting to note that the QM and QM/MM derived distances are more akin to those expected from an analysis of high resolution kinase PDBs (Table 2 and Fig. 6).

In summary, we find that for 4 out of the 8 inhibitors studied here, the ligand binding conformation in the gasphase QM model is essentially identical to that observed in the experimental X-ray crystal structure. In these cases it appears that the binding mode is optimal for the given polar interaction. This is perhaps unsurprising given that the source of the structures has been fragment based screening efforts which focus on ligand efficient hits [53,54]. In the remaining 4 cases the inhibitor binding conformation differs significantly from the X-ray structure to form increased interaction with the hinge region. This suggests steric factors associated with the protein environment force these 4 inhibitors into a less optimal binding conformation compared to gasphase/solution.

4. Conclusions

In this investigation, we have assessed the utility of QM and QM/MM calculations to probe the nature of the ligand interactions found in kinase protein–ligand. QM methods are more accurate than MM based alternatives due to the fact they consider the underlying physics of a given molecular system more precisely [15,30–35]. However, they are rarely used for this purpose due to their perceived high computational demand and the relatively few cases studies reported in the literature. This type of information could certainly prove useful in the optimization of ligand affinity, and efficiency [53,54].

Information from model QM calculations could be used to specifically target molecular changes for a particular ligand to either improve the interactions made with the key polar moieties in an active site, or should they be shown to be essentially optimal, focus on alterations elsewhere to probe for additional polar

interactions, or hydrophobic pockets to allow a targeted increase in lipophilicity. In addition, QM models could also be used to assess the core templates associated with large kinase inhibitors. Calculations could be employed to study the interaction of the N-phenylquinazolin-4-amine core of the EGFR inhibitor Tykerb with the 3 hinge residues. Furthermore, alterations to the structure could be assessed in-silico to see whether simple structural changes could be made to either improve the interaction with the hinge directly, or alter the angle associated with the anilino substituent to improve its trajectory into the relatively tight opening of the backpocket.

Finally, X-ray crystal structures play an absolutely critical role in SBD and the impact of computational chemistry methods would be severely restricted without the information they provide. While this work is not representative due to the focus on small numbers of structures from a single target class, from a more general consideration of the literature [21,20,24–27,51] it does appear that the atomic coordinates derived from X-ray analyses, while capable of demonstrating the overall ligand binding modes, should not necessarily be taken as a “perfect” reference given the intrinsic errors that arise from protein mobility and choices made during structural refinement [21]. This is particularly important for the validation of ligand docking methods since the X-ray coordinates are often used in quantitative way to assess the performance of software [51]. Comparison of the observed interactions to results from comparable interactions from the CSD or RCSB can also prove useful, as can results from QM and QM/MM optimizations to determine the relative strength of interactions found in a particular protein–ligand complex.

Acknowledgments

Special thanks are due to Ben Tehan for his help during the study. This work was supported by the Thailand Research Fund grants RTA5080005 (S.H.) and RMU5180032 (D.G.). Support for research facilities at KU provided by the National Center of Excellence in Petroleum, Petrochemical Technology and Advanced Materials is gratefully acknowledged. Finally, M.P.G. would like to thank the Faculty of Science, Kasetsart University (K.U.) for support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2010.09.012.

References

- [1] M. Von itzstein, W.Y. Wu, G.B. Kok, M.S. Pegg, J.C. Dyason, B. Jin, T.V. Phan, M.L. Smythe, H.F. White, S.W. Oliver, P.M. Colman, J.N. Varghese, D.M. Ryan, J.M. Woods, R.C. Bethell, V.J. Hotham, J.M. Cameron, C.R. Penn, Rational design of potent sialidase-based inhibitors of influenza-virus replication, *Nature* 363 (1993) 418–423.
- [2] R.M. Angell, T.D. Angella, P. Bamborough, M.J. Bamford, C. Chung, S.G. Cockerill, S.S. Flack, K.L. Jones, D.I. Laine, T. Longstaff, S. Ludbrook, R. Pearson, K.J. Smith, P.A. Smee, D.O. Somers, A.L. Walker, Biphenyl amide p38 kinase inhibitors 4: DFG-in and DFG-out binding modes, *Bioorg. Med. Chem. Lett.* 18 (2008) 4428–4432.
- [3] C. Bissantz, B. Kuhn, M. Stahl, A medicinal chemist's guide to molecular interactions, *J. Med. Chem.* 53 (2010) 5061–5084.
- [4] K.A. Brameld, B. Kuhn, D.C. Reuter, M. Stahl, Small molecule conformational preferences derived from crystal structure data. A medicinal chemistry focused analysis, *J. Chem. Inf. Model.* 48 (2008) 1–24.
- [5] B. Grimme, Do special noncovalent p–p stacking interactions really exist? *Angew. Chem. Int. Ed.* 47 (2008) 3430–3434.
- [6] Y. Zhao, H.T. Ng, E. Hanson, Benchmark data for noncovalent interactions in HCOOH...benzene complexes and their use for validation of density functionals, *J. Chem. Theory Comput.* 5 (2009) 2726–2733.
- [7] L. Estevez, N. Otero, R.A. Mosquera, Computational study on the stacking interaction in catechol complexes, *J. Phys. Chem. A* 113 (2009) 11051–11058.
- [8] R.S. Paton, J.M. Goodman, Hydrogen bonding and π -stacking: how reliable are force fields? A critical evaluation of force field descriptions of nonbonded interactions, *J. Chem. Inf. Model.* 49 (2009) 944–955.

- [9] P. Zhou, J. Zou, F. Tian, Z. Shang, Fluorine bonding – how does it work in protein–ligand interactions? *J. Chem. Inf. Model.* 49 (2009) 2344–2355.
- [10] Y. Lu, T. Shi, Y. Wang, H. Yang, X. Yan, X. Luo, H. Jiang, W. Zhu, Halogen bonding – a novel interaction for rational drug design? *J. Med. Chem.* 52 (2009) 2854–2862.
- [11] Y.N. Imai, Y. Inoue, Y. Yamamoto, Propensities of polar and aromatic amino acids in noncanonical interactions: nonbonded contacts analysis of protein–ligand complexes in crystal structures, *J. Med. Chem.* 50 (2007) 1189–1196.
- [12] M.-H. Hao, O. Haq, I. Muegge, Torsion angle preferences of small-molecule ligands bound to proteins, *J. Chem. Inf. Model.* 47 (2007) 2242–2252.
- [13] G.M. Spitzer, B. Wellenzohn, P. Markt, J. Kirchmair, T. Langer, K.R. Liedl, Hydrogen-bonding patterns of minor groove–binder–DNA complexes reveal criteria for discovery of new scaffolds, *J. Chem. Inf. Model.* 49 (2009) 1063–1069.
- [14] R. Wu, T.B. McMahon, Investigation of cation– π interactions in biological systems, *J. Am. Chem. Soc.* 130 (2008) 12554–12555.
- [15] G.J. Chang, J.Z.H. Zhang, NMR scalar coupling constant reveals that intraprotein hydrogen bonds are dynamically stabilized by electronic polarization, *J. Phys. Chem. B* 113 (2009) 13898–13900.
- [16] M. Nocker, S. Handschuh, C. Tautermann, K.R. Liedl, Theoretical prediction of hydrogen bond strength for use in molecular modeling, *J. Chem. Inf. Model.* 49 (2009) 2067–2076.
- [17] J.J.-L. Liao, Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors, *J. Med. Chem.* 50 (2007) 1–16.
- [18] R. Vargas, J. Garza, D.A. Dixon, B.P. Hay, How strong is the $\text{C}\alpha\text{--H}\cdots\text{O}=\text{C}$ hydrogen bond? *J. Am. Chem. Soc.* 122 (2000) 4750–4755.
- [19] A.C. Pierce, K.L. Sandretto, G.W. Bemis, Kinase inhibitors and the case for $\text{CH}_3\cdots\text{O}$ hydrogen bonds in protein–ligand binding, *Proteins* 49 (2002) 567–576.
- [20] G.J. Kleywegt, K. Henrick, E.J. Dodson, D.M.F. van Aalten, Pound-wise but penny-foolish: how well do micromolecules fare in macromolecular refinement? *Structure* 11 (2003) 1051–1059.
- [21] A.M. Davis, S.A. St-Gallay, G.J. Kleywegt, Limitations and lessons in the use of X-ray structural information in drug design, *Drug Discovery Today* 13 (2008) 831–841.
- [22] S. Wlodek, A.G. Skillman, A. Nicholls, Automated ligand placement and refinement with a combined force field and shape potential, *Acta Cryst. D* 62 (2006) 741–749.
- [23] N.W. Moriarty, R.W. Grosse-Kunstleve, P.D. Adams, Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation, *Acta Cryst. D* 65 (2009) 1074–1080.
- [24] N. Yu, H.P. Yennawar, K.M. Merz Jr., Refinement of protein crystal structures using energy restraints derived from linear-scaling quantum mechanics, *Acta Crystallogr. D Biol. Crystallogr.* (61) (2005) 322–332.
- [25] X. Li, S.A. Hayik, K.M. Merz Jr., QM/MM X-ray refinement of zinc metalloenzymes, *J. Inorg. Biochem.* (104) (2010) 512–522.
- [26] M.P. Gleeson, D. Gleeson, QM/MM as a tool in fragment based drug discovery, *J. Chem. Inf. Model.* 49 (2009) 1437–1448.
- [27] M.P. Gleeson, D. Gleeson, Assessing the utility of hybrid quantum mechanical/molecular mechanical methods in structure-based drug design, *J. Chem. Inf. Model.* 49 (2009) 670–677.
- [28] G.L. Warren, C.A. Andrews, A. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, A critical assessment of docking programs and scoring functions, *J. Med. Chem.* 49 (2006) 5912–5931.
- [29] I.J. Enyedy, W.J. Egan, Can we use docking and scoring for hit-to-lead optimization? *J. Comput. -Aided Mol. Des.* 22 (2008) 161–168.
- [30] J.A. Platts, Theoretical prediction of hydrogen bond basicity, *Phys. Chem. Chem. Phys.* 2 (2000) 3115–3120.
- [31] J. Reynisson, E. McDonald, Tuning of hydrogen bond strength using substituents on phenol and aniline: a possible ligand design strategy, *J. Comput. -Aided Mol. Des.* 18 (2004) 421–431.
- [32] M.-H. Hao, Theoretical calculation of hydrogen-bonding strength for drug molecules, *J. Chem. Theory Comput.* 2 (2006) 863–872.
- [33] S. Raub, C.M. Marian, Quantum chemical investigation of hydrogen-bond strengths and partition into donor and acceptor contributions, *J. Comput. Chem.* (28) (2007) 1503–1515.
- [34] J. Schwobel, R.-U. Ebert, R. Kuhne, G. Schuurmann, Prediction of the intrinsic hydrogen bond acceptor strength of organic compounds by local molecular parameters, *J. Chem. Inf. Model.* 49 (2009) 956–962.
- [35] P.W. Kenny, Hydrogen bonding, electrostatic potential, and molecular design, *J. Chem. Inf. Model.* 49 (2009) 1234–1244.
- [36] P.E.M. Siegbahn, F. Himo, Recent developments of the quantum chemical cluster approach for modeling enzyme reactions, *J. Biol. Inorg. Chem.* 14 (14) (2009) 643–651.
- [37] A. Warshel, M. Levitt, Theoretical studies of enzymatic reactions: dielectric electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, *J. Mol. Biol.* 103 (1976) 227–249.
- [38] U.C. Singh, P.A. Kollman, A combined ab initio quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular-systems – applications to the $\text{CH}_3\text{Cl}+\text{Cl}^-$ exchange-reaction and gas-phase protonation of polyethers, *J. Comput. Chem.* 7 (1986) 718–730.
- [39] K. Raha, K.M. Merz 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.* 48 (2005) 4558–4575.
- [40] F.H. Allen, The Cambridge structural database: a quarter of a million crystal structures and rising, *Acta Cryst. B* 58 (2002) 380–388.
- [41] RCSB Protein databank. <http://www.rcsb.org/>.
- [42] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.
- [43] S. Dapprich, I. Komaromi, K.S. Byun, K. Morokuma, M.J. Frisch, A new ONIOM implementation in Gaussian 98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives, *J. Mol. Struct.* 1 (1999) 461–462.
- [44] T. Vreven, K.S. Byun, I. Komaromi, S. Dapprich, J.A. Montgomery, K. Morokuma, M.J. Frisch, Combining quantum mechanics methods with molecular mechanics methods in ONIOM, *J. Chem. Theory Comput.* 2 (2006) 815–826.
- [45] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard III, W.M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *J. Am. Chem. Soc.* 114 (1992) 10024–10035.
- [46] K. Senthilkumar, J.I. Mujika, K.E. Ranaghan, F.R. Manby, A.J. Mulholland, J.N. Harvey, Analysis of polarization in QM/MM modelling of biologically relevant hydrogen, *J. R. Soc. Interface* 5 (2008) S207–S216.
- [47] I.J. Bruno, J.C. Cole, J.P.M. Lommerse, R.S. Rowland, R. Taylor, M.L. Verdonk, A library of information about non-bonded interactions, *J. Comput. -Aided Mol. Des.* 11 (1997) 525–537.
- [48] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.* 117 (1995) 5179–5197.
- [49] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comp. Chem.* 25 (2004) 1157–1174.
- [50] ROCS 2.3.1: OEChem version 1.5.1 OpenEye Scientific Software Inc., Santa Fe NM, USA. www.eyesopen.com.
- [51] D. Yusuf, A.M. Davis, G.J. Kleywegt, S. Schmitt, An alternative method for the evaluation of docking performance: RSR vs RMSD, *J. Chem. Inf. Model.* 48 (2008) 1411–1422.
- [52] Coot 0.6.1, <http://www.biop.ox.ac.uk/coot/>.
- [53] A.L. Hopkins, C.R. Groom, A. Alex, Ligand efficiency: a useful metric for lead selection, *Drug Discovery Today* 9 (2004) 430–431.
- [54] C. Abad-Zapatero, J.T. Metz, Ligand efficiency indices as guideposts for drug discovery, *Drug Discovery Today* 10 (2005) 464–469.