Kinase Inhibitors and the Case for CH···O Hydrogen Bonds in Protein–Ligand Binding

Albert C. Pierce,[*] **Kathryn L. Sandretto, and Guy W. Bemis**
Vertex Pharmaceuticals, Cambridge, Massachusetts

ABSTRACT **Although the hydrogen bond is known to be an important mediator of intermolecular interactions, there has yet to be an analysis of the role of CH···O hydrogen bonds in protein–ligand complexes. In this work, we present evidence for such nonstandard hydrogen bonds from a survey of aromatic ligands in 184 kinase crystal structures and 358 high-resolution structures from the Protein Data Bank. CH groups adjacent to the positively charged nitrogen of nicotinamide exhibit geometric preferences strongly suggestive of hydrogen bonding interactions, as do heterocyclic CH groups in kinase ligands, while other aromatic CH groups do not exhibit these characteristics. Ab initio calculations reveal a considerable range of CH···O hydrogen bonding potentials among different aromatic ring systems, with nicotinamide and heterocycles preferred in kinase inhibitors showing particularly favorable interactions. These results provide compelling evidence for the existence of CH···O hydrogen bonds in protein–ligand interactions, as well as information on the relative strength of various aromatic CH donors. Such knowledge will be of considerable value in protein modeling, ligand design, and structure–activity analysis. Proteins 2002;49:567–576.** © 2002 Wiley-Liss, Inc.

INTRODUCTION

The importance of the hydrogen bond in molecular biology is so well established that one would be hard pressed to find a single publication of structural biomolecular research that did not cite the role of hydrogen bonding in its findings. Conversely, the CH···O hydrogen bond, now widely accepted as a significant cohesive interaction,[1] is in general overlooked in these same studies of biomolecular phenomena. There was a time when the existence of CH···O interactions was controversial[2] and the lack of attention from biologic researchers may have been justified. However, nearly two decades have past since Taylor and Kennard's work[3] began laying the controversy to rest. Over the years, the hydrogen bonding character of these interactions has been verified theoretically,[4,5] spectroscopically,[6,7] and crystallographically.[3,8] In addition, a considerable body of crystal engineering research[6,9] has proven the utility of our understanding of these interactions. Never-

theless, the biologic implications of these interactions have received limited attention.[10] One early study examined phenylalanine–oxygen interactions in protein structures, revealing a weakly polar interaction between the edge of the aromatic ring and oxygen atoms.[11] Two subsequent reports considered the hydrogen bonding capacity of imidazole CH groups. In the first, a ubiquitous histidine $C^{\epsilon 2}$-H···O=C interaction was described in serine esterase structures,[12] while the second involved ab initio calculations of imidazole CH···O interactions and their force field parameterization.[13] Other studies have found convincing evidence for C^{α}—H···O and CH···π hydrogen bonding as prevalent cohesive interactions in proteins.[14,15]

While these results represent a valuable start in the exploration of the role of CH···O interactions in proteins, much remains to be done, in particular with respect to protein–ligand interactions. A significant hurdle to further studies, however, is the fact that X-ray crystal structures of proteins are in general of lower resolution than small-molecule structures and that hydrogen atom positions can only be inferred from heavy-atom positions. As a result, large numbers of observations are necessary to compensate for the inaccuracies in any single structure. A further complication is the fact that strong CH···O hydrogen bond donors seem to be relatively rare. In the survey of all of the carbon atom types in proteins, only the backbone C^{α} was found to be a frequent donor of hydrogen bonds.[14]

Fortunately, our experience with ligand-bound protein kinase structures suggested that these complexes would provide an ideal data set for the study of CH···O hydrogen bonds in protein–ligand interactions. Between the publicly available structures and structures determined as a part of our own kinase drug discovery program, there are almost 200 liganded kinase structures available for this study. And, close examination of these structures reveals what appears to be an extraordinary number of short CH···O contacts. Figure 1 shows a small subset of these contacts, from the crystal structures of CDK2 with inhibitor H717 (1G5S),[16] c-AMP-dependent protein kinase with H7 (1YDR),[17] and abl with an STI-571 analog (1FPU).[18] While there are a handful of short aliphatic CH···O contacts, as can be seen in Figure 1, the vast majority of these short contacts involve aromatic carbons. Consider-

---

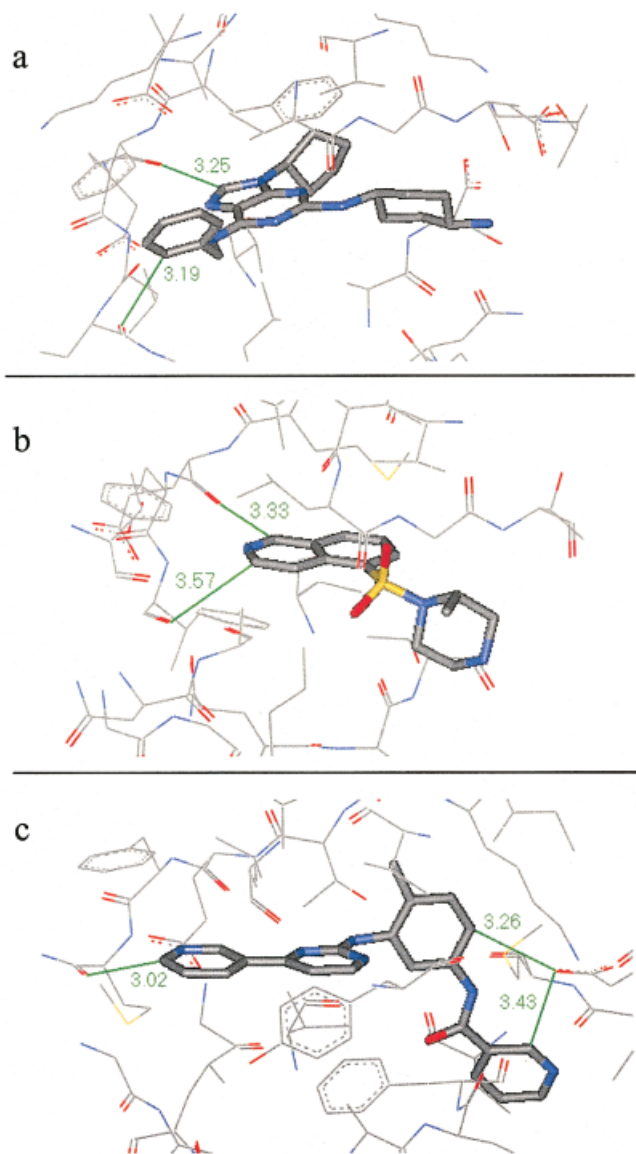*Correspondence to: Albert C. Pierce, Vertex Pharmaceuticals, 130 Waverly St., Cambridge, MA 02139. E-mail: al_pierce@vrtx.com

Fig. 1.   Short CH· · ·O contacts in crystal structures of (**a**) CDK2 with inhibitor H717, (**b**) cAMP-dependent protein kinase with inhibitor H7, and (**c**) an STI-571 analog in Abl kinase.

ing this fact, and the results of earlier studies that established the significance of aromatic CH· · ·O contacts in protein–protein interactions,[11,12] this work will focus solely on the hydrogen bonding potential of ligand aromatic carbons.

As a supplement to the survey of crystal structure CH· · ·O contacts, ab initio calculations have been performed to determine the energetic and geometric properties of aromatic CH· · ·O hydrogen bonds. These calculations serve primarily as a highly accurate standard with which to compare the geometric results of the crystal structure studies. However, they also allow the comparison of the relative interaction energies of different heterocyclic CH donors and the extension of this study to heterocycles not present in the available crystal structures. As a result, this work should not only establish the

hydrogen bonding capability of aromatic CH donors but also determine their relative hydrogen bonding strengths, enabling their use in force-field parameterization, structure–activity relationship (SAR) studies, and ligand design efforts.

METHODS
Crystal Structures

Three sets of crystal structures have been surveyed for CH· · ·O contacts. The first consists of all protein kinase structures available in the Protein Data Bank (PDB) combined with a subset of protein kinase structures solved by Vertex crystallographers. This set includes a total of 184 liganded structures, including 19 different kinases. While there are a small number of duplicates, such as ATP bound to a phosphorylated and unphosphorylated kinase, the vast majority of these complexes are unique. The resolution of these structures ranges between 1.5 and 3.0 Å, with an average resolution of approximately 2.3 Å. The second and third sets consist of all high- (<1.75 Å) and medium- (1.75–2.4 Å) resolution structures containing aromatic ligands available from the PDB in August 2001. The high-resolution structures, those with resolutions better than 1.75 Å, were gathered to serve as a comparison between the kinase results and results on a general set of protein–ligand complexes. The medium-resolution structures (1.76–2.4 Å) were examined to ensure that any features unique to the kinase data set are the result of the kinase structures and their ligands rather than an effect of their lower accuracy relative to the high-resolution set. The medium-resolution set consisted of 1837 structures and a total of 2927 ligands capable of donating an aromatic CH hydrogen. The high-resolution set consisted of 576 such ligands among 358 crystal structures. Because the results from the medium-resolution data set are qualitatively similar to the results from the high-resolution data set (data not shown), only high-resolution results are reported.

One reviewer suggested that the B-factors from these structures be examined, as some residues in the ATP site of protein kinases are known to exhibit unusually high B-factors. Overall, average B-factors for these data sets parallel their resolution at 19.4 for high resolution, 27.6 for medium resolution, and 31.6 for kinases. Certain residues in kinase active sites, such as those in the glycine-rich loop, do exhibit higher than average B-factors in 90% of CH· · ·O hydrogen-bonded structures. However, residues in the strand of β-sheet involved in most CH· · ·O hydrogen bonds have lower than average B-factors in 79% of structures. It therefore seems unlikely that high B-factors have detrimentally affected our analysis. We also note that all of the trends seen below in the full kinase data set are observable in the set of publicly available kinase structures, although the plots are considerably noisier due to the smaller quantity of data.

The automated analysis of PDB-formatted ligands is a difficult and error-prone task, but the restriction to aromatic rings simplifies the process somewhat. A ligand was determined to be any set of HETATM records with shared residue name and number, consisting of at least one
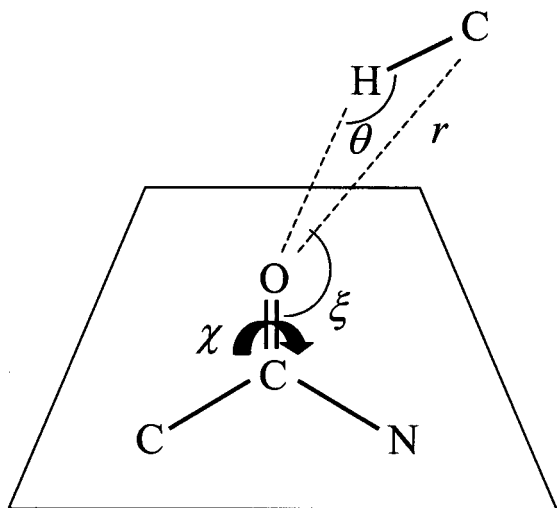
Fig. 2.   Geometric parameters used in hydrogen bond analysis. The elevation angle, ϕ, is equal to $\sin^{-1}(\sin\xi \sin\chi)$, where ξ is the C=O· · ·H angle and χ is the C—C=O· · ·H torsion.

carbon atom and five or more heavy atoms. Only the first representative of the ligand was considered in cases where multiple copies/orientations of a ligand were recorded in a PDB file. Atom connectivity was determined in a straight-forward fashion based on atom positions and standard bond lengths. Bond order, however, could not be accurately determined in this manner. Therefore, any flat, five- or six-membered ring was considered to be aromatic, with hydrogens attached at standard bond lengths (1.08 Å) and standard angles based on the appropriate ring system. Obviously, this definition will include some nonaromatic systems held flat by amide bonds and/or individual double bonds, such as cytosine or staurosporine. However, only 4 of the 476 interactions that met our criteria for short CH· · ·O contacts in the high-resolution structures in-volved nonaromatic systems. None of the short CH· · ·O contacts in the kinase systems involved nonaromatic rings.

Certain requirements were imposed on the CH· · ·O contacts for their inclusion in this analysis. First, only carbonyl and carboxylate oxygens have been considered. One of the key features of hydrogen bonds is the tendency to place the donated proton in proximity to the lone pair(s) of the acceptor atom, and the position of the lone pairs cannot be determined for hydroxyl oxygens. The paucity of hydroxyl oxygens in proteins minimizes the consequences of this omission. Second, because hydrogen bonds are directional in nature,[19] only those contacts with a C—H—O angle greater than 90° were included. Finally, only interac-tions with a protein–ligand C· · ·O distance less than 7 Å were included in the distance distribution, while a cutoff of 3.6 Å was employed to limit the angular analyses to potential hydrogen bonds. Although hydrogen bonding is attributed to C· · ·O interactions as far as 4 Å apart,[9] we used a shorter cutoff in an attempt to eliminate weaker interactions that might be difficult to characterize with our lower-resolution data sets.

The three geometric distributions considered in the CH· · ·O analyses are depicted in Figure 2. They are the C· · ·O distance (r) distribution, the C—H—O angle (θ) distribution, and the elevation angle (ϕ) distribution. The peak in the distance distribution should occur at the preferred C· · ·O distance and the height of the peak should be proportional to the strength of the CH· · ·O attraction. A higher peak at a shorter separation (r) would be indicative of possible hydrogen bonding character in an interaction. More importantly, hydrogen bonds are known to prefer a linear donor-H-acceptor angle,[19] so a peak in the θ distribution near 180° would also suggest the exis-tence of hydrogen bonds. Finally, the distribution of the elevation angle, ϕ, measures the extent to which the donated hydrogen is elevated out of the plane of the acceptor lone pairs. In an ideal hydrogen bond, the hydro-gen should be close to the lone pairs of the acceptor, suggesting a peak in the ϕ distribution at 0°.[19] All of these distributions must also be corrected for volume factors. The volume of a shell of thickness dr is proportional to $r^3 dr$, so the population of the shell between r_1 and r_2 is normalized by $1/((r_1 + r_2)/2)^3$. The corresponding volume for θ is proportional to $\sin\theta d\theta$, while for ϕ the volume is proportional to $\cos\phi d\phi$. These distributions were corrected for volume in a manner analogous to that of the distance distribution. Each distribution was then normalized by dividing the volume-corrected population of each bin by the sum of the volume-corrected populations over all bins. Finally, to determine a "standard" for each distance distri-bution to which peak heights could be compared, the average of the volume-corrected populations over the bins between 6 and 7 Å was calculated and used as the baseline.

Ab Initio Calculations

These calculations were carried out to determine the relative hydrogen bonding energies and geometries of various aromatic ring CH donors with the oxygen of a water molecule. Water was chosen as an acceptor primar-ily due to its lack of additional functionality. With water as an acceptor, nearly the full interaction energy will be the result of the CH· · ·O interaction. Also, the nature of the hydrogen bond acceptor is known to be of significantly less consequence in CH· · ·O interactions than the identity of the donor.[9,20] The ab initio energies of the methane and pyrimidine 2-CH and 4-CH interacting with water are −0.87, −1.77, and −2.71, closely following the values for the same interactions with formamide of −0.64, −1.46, and −2.56, respectively. In an additional effort to isolate the CH· · ·O interaction and accurately reproduce the hydrogen bonding potential of these complexes, the C—H—O angle was constrained to 180° in the optimiza-tion of the intermolecular geometries. Otherwise, all de-grees of freedom were optimized.

The Gaussian 98 software package[21] was employed in all calculations, using the HF/6-31G** level of theory, uncorrected for basis set superposition error (BSSE). While higher levels of theory are feasible, they are computation-ally expensive and, for the purposes of this study, unneces-sary. For large-basis set calculations including correlation and BSSE correction, the effects of electron correlation and larger basis set cancel out the correction for BSSE, at least to some extent. More importantly, for calculations on a set

of highly analogous systems such as these, a significant fraction of the error in the absolute interaction energy is likely to cancel out in the calculation of the relative energies. To verify this assertion, BSSE-corrected, MP2/cc-pVTZ single-point calculations were carried out at the 6-31G**-optimized geometries for six of the aromatic ring systems complexed with water. BSSE was corrected with the standard counterpoise method of Boys and Bernardi.[22] The average absolute difference between the energy at the two levels of theory for each complex is 0.66 kcal/mol. However, in every case, the HF/6-31G** interaction energy is simply overestimated by 0.44–0.88 kcal/mol. When the pairwise relative energies are compared, the average absolute difference is only 0.24 kcal/mol, with "errors" in ΔE ranging from −0.44 to 0.09 kcal/mol. As a final comparison, the interaction energy of the water dimer was calculated with BSSE-uncorrected HF/6-31G** and BSSE-corrected MP2/cc-pVTZ, yielding values of −5.51 kcal/mol and −4.34 kcal/mol, respectively. The most sophisticated experimental[23] and theoretical[24] methods suggest that the correct value for this quantity is −4.9 or −5.0 kcal/mol. It is therefore difficult to justify the additional computational expense of the BSSE-corrected MP2/cc-pVTZ calculations, and significantly higher levels of theory would be prohibitively expensive for a large number of systems of this size.
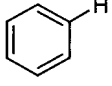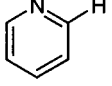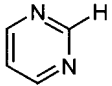
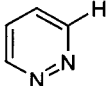## RESULTS AND DISCUSSION
Ab Initio Results
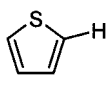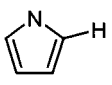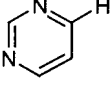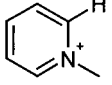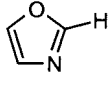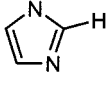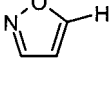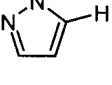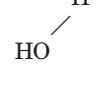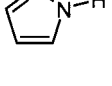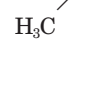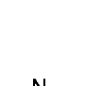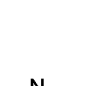
The energetic results of the ab initio calculations are shown in Table I. For the aromatic CH···O interactions, the energies range between −1.47 for benzene and −3.48 for the 5-CH of isoxazole. The comparable interaction energy is −0.87 kcal/mol for methane and −5.51 kcal/mol for the water dimer. Thus, the aromatic CH hydrogen bond reaches strengths as high as 400% of the largely van der Waals methane–water interaction and 63% of the hydrogen-bonded water dimer energy. Also, as demonstrated in Figure 3, the optimized conformations of the aromatic–water systems are in general shifted to shorter H···O distances as the interaction becomes more favorable. The H···O distance for isoxazole is 2.33 Å, relative to 2.78 Å for methane, 2.64 Å for benzene, and 2.04 Å for water. These results are in qualitative agreement with other MP2/6-31++G($2d$,$2p$) optimizations and MP2/aug-cc-pVDZ single-point calculations of the 2- and 4- CH groups of imidazole with water. The interaction energies and H···O distances were found to be −3.21 kcal/mol and 2.33 Å for 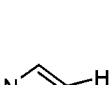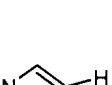the 2-CH group and −2.25 kcal/mol and 2.45 Å for the 4-CH group in this earlier study.[13] It is also interesting to note that for the 5-membered heterocycles the strength of the CH donor seems to be determined more by the position of the CH in the particular heterocycle rather than the identity of the adjacent heteroatom. For example, the difference in energy for the donors of thiophene and pyrrole is 0.02 kcal/mol, while this difference for thiophene and the 5-CH of thiazole is 0.47. This fact seems to hold for all of the heterocycles tested with the exception of isoxazole and pyrazole, which do not seem to have analogous donor strengths.

Protein–Ligand Structures

As expected from the ab initio results, examination of the protein–ligand complexes showed that different aromatic CH groups behaved differently with respect to their interactions with protein oxygen atoms. Ultimately, it was determined that separating the aromatic CH groups into three types gave the best description of their behavior. The three types were (1) those CH groups adjacent to at least one heteroatom (henceforth referred to as "heteroaromatic" CH groups), (2) those between two carbon atoms ("aromatic" CH groups), and (3) those adjacent to a positively charged nitrogen ("pyridinium" CH groups, as in nicotinamide). Figure 4 shows a plot of the relative populations of C···O contacts as a function of distance for the different types of potential CH hydrogen bond donors. Aliphatic carbons from the high-resolution PDB structures are included in this plot to allow a comparison of the more polar aromatic interactions to nonpolar van der Waals interactions.

The peak in the aliphatic plot occurs at a C···O separation of 3.8 Å (the 3.6- to 3.8-Å bin) and has a height of 1.4 times baseline. This corresponds well to the ab initio C···O separation distance of 3.87 Å, although perhaps the aliphatic peak drops off too slowly at shorter distances. By comparison, the peak for aromatic CH groups peaks at 3.6 Å, with a peak height of 2.2 times baseline. Again, this seems in reasonable agreement with the ab initio benzene C···O separation of 3.72 Å and interaction energy of −1.47 kcal/mol, compared to the methane–water interaction energy of −0.87 kcal/mol. The distance distribution of heteroaromatic CH groups has a peak shifted to even shorter distances (3.4 Å), with a peak 3.4 times its baseline value. The C···O separations for heteroaromatic CH groups based on ab initio calculations ranges from 3.40 (isoxazole 5-CH) to 3.66 (imidazole 4-CH), with intermolecular association energies ranging from −1.77 to −3.48 kcal/mol. The structural and ab initio results are again in qualitative agreement. The final distance distribution from the high-resolution PDB data set is for pyridinium CH groups. The peak in this distribution occurs at 3.2 Å, with a height of 4.9 times baseline, while the ab initio results for pyridinium suggest an ideal C···O separation of 3.19 Å with a water association energy of −11.2 kcal/mol. Interestingly, while the aromatic kinase ligands follow the same trend as the ligands in the PDB structures, the peaks in the kinase distributions are higher relative to their baseline and are shifted to even shorter C···O distances. The aromatic and heteroaromatic peaks both occur at 3.4 Å, with heights of 2.9 and 3.9 times baseline, respectively. For each category of CH donor, the percentage of interactions with a C···O distance less than 3.6 Å is PDB aliphatic 15%, PDB aromatic 24%, PDB heteroaromatic 31%, kinase aromatic 30%, kinase heteroaromatic 44%, and pyridinium 52%. Thus far, the structural and ab initio results are in qualitative agreement and seem to suggest enhanced cohesive interactions between at least some subset of aromatic CH groups and oxygen atoms. However, the θ and φ angle values are also important characteristics of hydrogen bonds, and these must be considered before any conclusions are drawn.

TABLE I. Interaction Energies—H$_2$O

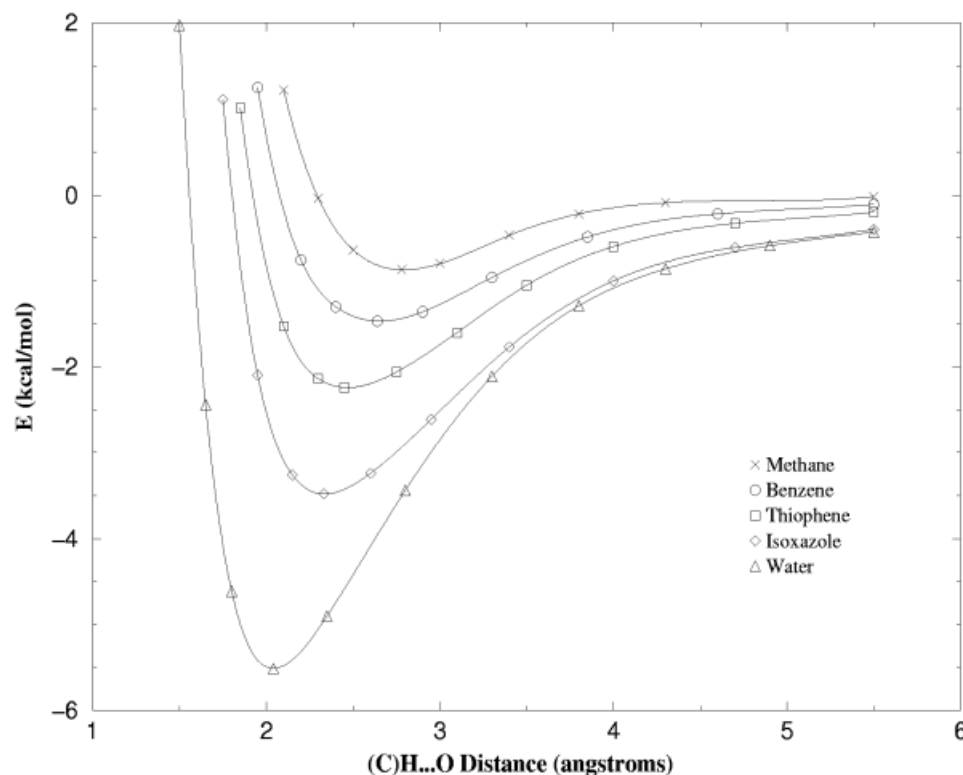|	HF/6-31G**	MP2/cc-pVTZ	$r_{x.o}$		HF/6-31G**	$r_{x.o}$
	−1.47	−1.03	3.72		−2.23	3.64
	−1.77	−1.15	3.60		−2.90	3.55
	−2.24	−1.79	3.53		−2.22	3.62
	−2.71	−1.88	3.55		−11.16	3.19
	−3.06	−2.32	3.42		−3.21	3.51
	−3.48	−2.60	3.40		−2.91	3.48
	−5.51	−4.34	2.99		−6.37	3.03
	−0.87		3.87			
	−2.41		3.57		−2.16	3.66
	−3.04		3.48		−2.08	3.64
	−2.71		3.45		−2.71	3.53

Distance Dependence of CH...O Interactions

Fig. 3. Plot of 6-31G** interaction energies for hydrogen bonds to water. The heterocyclic donors are the 2-CH of thiophene and the 2-CH of isoxazole.

Consideration of these angle distributions is in particular important for the pyridinium CH groups, as some component of the attraction evident in this CH· · ·O distance distribution will be due simply to charge–charge interactions between the positively charged aromatic system and negatively charged carboxylate oxygens in the protein.

The number of short CH· · ·O contacts considered in the angular analyses were: 45 for PDB pyridinium, 162 for PDB heteroaromatic, 269 for PDB aromatic, 109 for kinase heteroaromatic, and 348 for kinase aromatic. The distributions of the C-H-O angle, θ, are relatively flat for both PDB and kinase aromatic CH groups (Fig. 5). This would seem to argue that either these weakly polar interactions do not have an angular dependence or that these interactions are sacrificed in the majority of protein–ligand interactions to optimize hydrophobic contacts or traditional hydrogen bonds. The peak in θ for heteroaromatic CH groups in kinase structures occurs in the range of 150–180°, while the PDB pyridinium θ distribution is peaked strongly in the 165–180° range. This angular dependence is strongly suggestive of the presence of significant aromatic CH· · ·O hydrogen bonding, which should be optimal at 180°. The distribution of θ in the heteroaromatic CH groups of the PDB structures is somewhat anomalous, with a peak between 120 and 150°. This peak appears to be a result of the relatively weak heteroaromatic donors of adenine and their binding site interactions, as discussed below.

The final geometric factor that has been examined is the elevation angle distribution, ϕ, which should reflect the tendency of the donated hydrogen to lie in the plane of the lone pairs of the hydrogen bond acceptor. This angle should be zero for an ideal hydrogen bond, while nonpolar interactions should favor an angle of closer to 90°, allowing an approach unhindered by the presence of lone pair electrons. In the case of the pyridinium and kinase heteroaromatic CH groups, the elevation angle distributions show clear peaks in the 0–15° range and few interactions with ϕ greater than 45° (Fig. 6), further suggesting their tendency to be involved in CH· · ·O hydrogen bonds. This same distribution for PDB heteroaromatic CH groups is more ambiguous. There is a peak at 0° but the larger peak is at 90°, suggesting that while some heteroaromatic CH groups in the PDB structures may form hydrogen bonds the majority of close CH· · ·O contacts do not. As might have been expected from the θ distributions, the elevation angle distributions for aromatic CH groups in the PDB and kinase structures give no evidence of hydrogen bonding interactions. Both distributions have their peak at the 75–90° bin and their nadir at 0–15°.

It seems that there is a range of hydrogen bonding capacity among aromatic CH groups. In the case of aromatic CH groups not adjacent to a heteroatom, this analysis suggests that CH· · ·O hydrogen bonding is not prevalent, despite earlier work suggesting a weakly polar

C(H)..O Distance Distribution
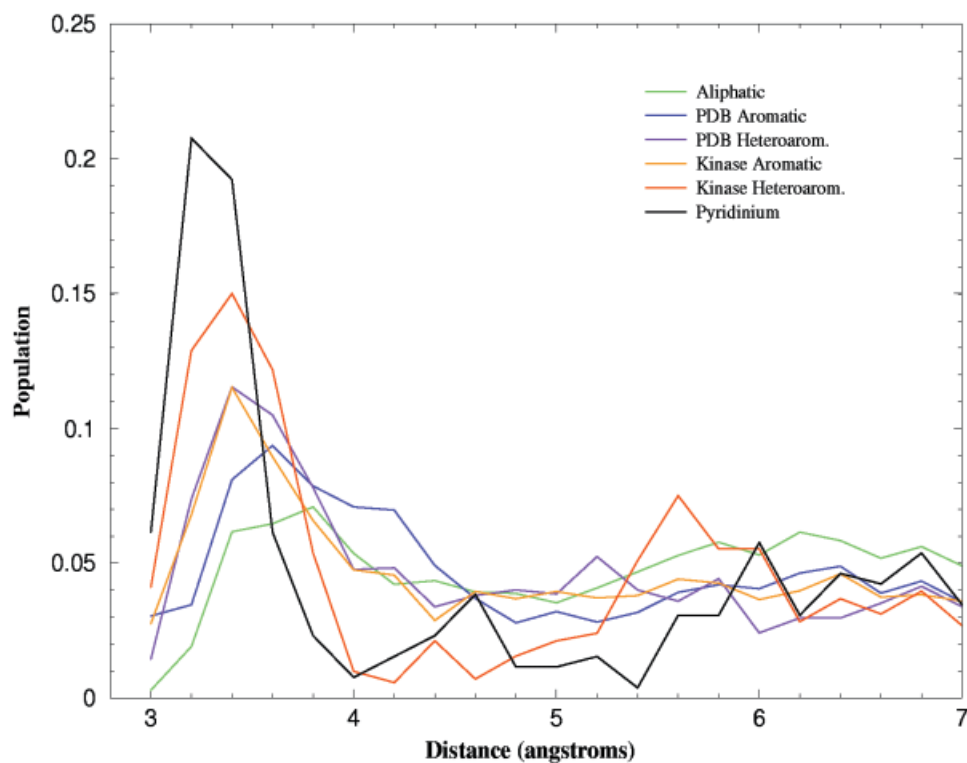
Fig. 4.   Volume-corrected distance distribution of C(H)···O contacts in protein–ligand crystal structures.

C–H–O Angle Distribution
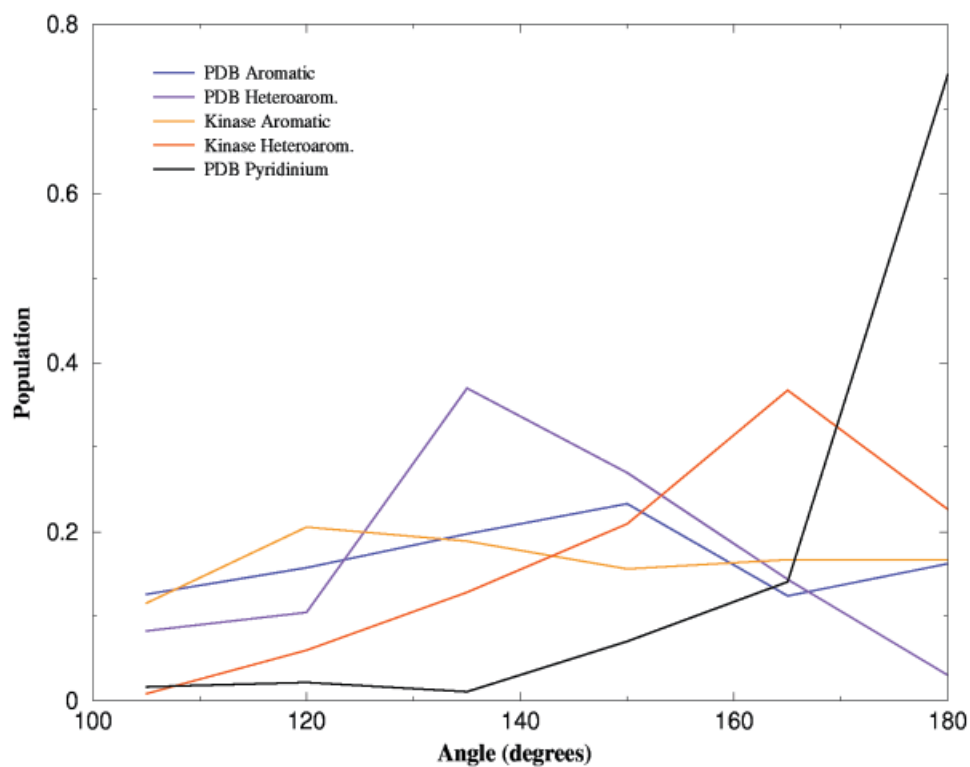
Fig. 5.   Volume-corrected C—H—O angle distribution for short ($<$ 3.6 Å) CH···O contacts.
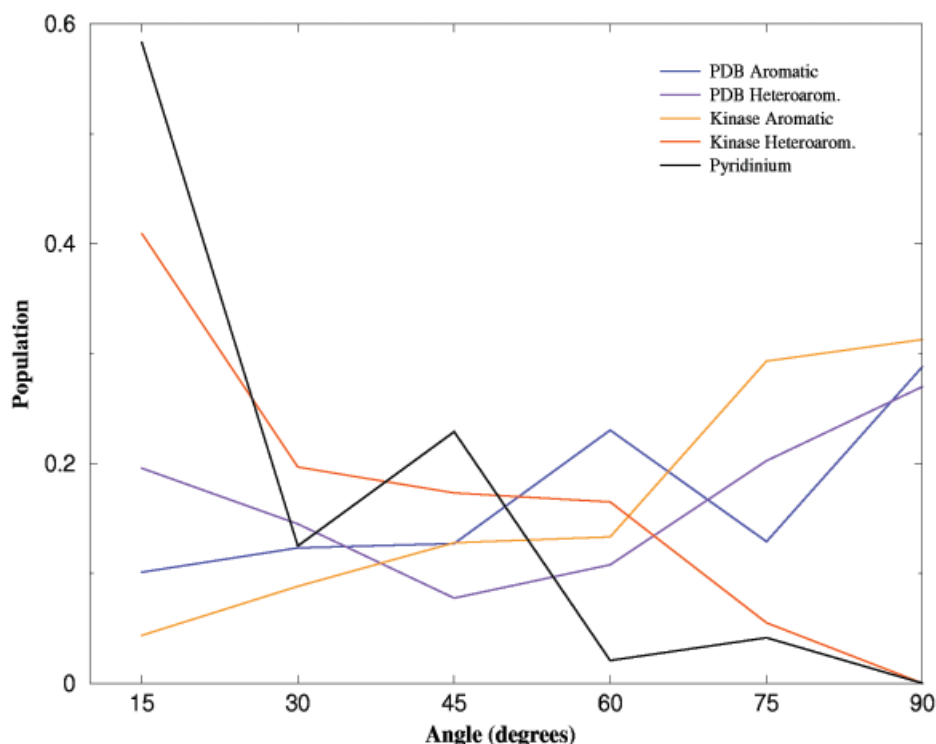
Elevation Angle Distribution

Fig. 6.   Volume-corrected elevation angle distribution for short ($<$ 3.6 Å) CH$\cdots$O contacts.

phenylalanine–oxygen interaction in proteins.[11] While it is likely that these interactions play a role in protein–ligand binding, our analysis of this data set provides little evidence for their occurrence. On the other hand, the CH groups adjacent to the nitrogen of pyridinium rings show a strong preference for short CH$\cdots$O contacts, with the hydrogen atom in the plane of the oxygen lone pairs and a linear C—H—O geometry, clearly establishing this group as a hydrogen bond donor. The heteroaromatic CH groups of kinase ligands also exhibit all geometric characteristics of hydrogen bonds. While the peaks are less sharp than in the pyridinium case, the volume-adjusted populations with θ and ϕ within 30° of their hydrogen bond ideal are 59 and 60%, respectively, relative to the 33% that would be expected from a random distribution. This represents a clear preference for hydrogen bonding interactions.

What seems most peculiar about this analysis is the fact that the heteroaromatic CH groups from the PDB structures do not show this same tendency to adopt a hydrogen bonding interaction with oxygen acceptors. The peak in the heteroaromatic distance distribution is smaller for the PDB data set than for the kinase data set, the θ distribution peaks at 120–150°, and the 0–30° peak in the ϕ distribution is smaller than the peak at 150–180°. This suggests that there should be some fundamental difference between the structures of the kinase and PDB data sets.

One possibility is that the protein structures are responsible for the difference in hydrogen bonding propensities.

The kinase data set consists entirely of protein kinases with ligands bound in the ATP binding site, while the PDB data set consists of a variety of proteins limited only by the requirement for aromatic ligands. It may be that these ATP binding sites are in particular well suited to form CH$\cdots$O hydrogen bonds with heteroaromatic ligands. However, inspection of these structures reveals that this is not the case. Of the PDB heteroaromatic CH groups involved in short CH$\cdots$O contacts, 91% involve adenine in an adenine binding pocket, with the C2 hydrogen forming a short contact with a backbone carbonyl of an amino acid in a strand of β-sheet. This is exactly the sort of interaction typical of ATP and inhibitors bound at the kinase active site. These nonkinase adenine binding pockets should therefore be equally conducive to hydrogen bond formation, so it seems likely that the different donor strengths of the PDB and kinase ligands must be responsible for the difference in hydrogen bonding propensity. Examination of the ab initio data for the 2-CH of the pyrimidine substructure of adenine reveals that this is in fact the weakest of all heteroaromatic CH groups studied. Although this is the predominant donor in the PDB data set, the short CH$\cdots$O contacts in the kinase ligands involve pyrimidine 2-CH donors only 20% of the time. The remaining 80% of these contacts involve other CH donors, including the 2-CH of pyridine, the 2-CH and 4-CH of thiazole, the 3-CH and 5-CH of pyrazole, and the 4-CH of pyrimidine. According to the ab initio results depicted in Table I, the hydrogen bonds formed by these CH groups are all

shorter and stronger than those formed by the 2-CH of pyrimidine. It seems likely that this difference in heterocycle populations, optimized in kinase inhibitors for high-affinity binding, is largely responsible for the difference in CH···O hydrogen bonding between kinase structures and protein structures from the PDB. While one might expect the 2-CH group of the imidazole of adenine ligands to be involved in hydrogen bonding, it seems that the hydrogen bonding functionality of the adjacent ribose and phosphate groups overwhelms the potential for these CH···O hydrogen bonds in most ligands.

A final piece of anecdotal evidence is included to address a question not easily answered with the above methods, "What is the penalty in binding affinity for replacing a traditional protein–ligand hydrogen bond with an aromatic CH···O hydrogen bond?" In a series of pyrazole compounds targeting a Ser/Thr kinase involved in mitosis, three direct thiazole analogs were synthesized. X-ray structures reveal that the 4-CH of the thiazole replaces the NH of the pyrazole in donating a hydrogen bond to a carbonyl oxygen of the hinge backbone. Among the three pyrazole–thiazole pairs, the K_i values are approximately equal in one case, while in the other two cases the thiazole compounds are 2- and 10-fold less potent. These differences seem remarkably small considering that the ab initio water interaction energies for the thiazole CH and pyrazole NH are −2.41 and −6.37 kcal/mol, respectively. However, these are gas-phase energies and the stronger NH donor of the pyrazole must pay a larger desolvation penalty to leave the aqueous environment to form its hydrogen bonds with the protein. So, perhaps these two effects counterbalance one another to a large extent, resulting in similar binding affinities for standard hydrogen bonds and their CH···O analogs. If traditional hydrogen bonds and CH···O hydrogen bonds are indeed roughly interchangeable, the impact on ligand design should be tremendous. NH to CH donor swaps would allow the design of novel inhibitors with similar binding affinity but potentially improved non-binding-related properties such as cell permeability or metabolic stability.

CONCLUSIONS

This survey of biomolecular crystal structures provides compelling evidence for the existence of aromatic CH hydrogen bonds in protein–ligand interactions. The evidence is most convincing for the cases of CH groups adjacent to positively charged nitrogen atoms and CH groups adjacent to heteroatoms in kinase ligands. Ab initio calculations reveal that positively charged heterocycles form a strong hydrogen bond with appropriate acceptors. These calculations also reveal that heterocycles prevalent among kinase ligands show enhanced hydrogen bond donor strengths relative to the adenine donors of nonkinase structures. The kinase ligands have been optimized for high-affinity binding, resulting in a preponderance of high-strength CH···O hydrogen bond donors. While it seems that these optimized CH···O hydrogen bonds were a serendipitous discovery in kinase inhibitors, as knowledge of these interactions becomes more widespread, it is likely that they could be exploited in the development of ligands for a wide range of proteins.

The ab initio calculations performed as part of this study reveal a considerable range of CH···O hydrogen bond strengths among aromatic CH groups. Knowledge of these relative strengths should be useful in the design of novel ligands but also in many other aspects of the ligand design process. SARs may be better understood with this knowledge. More accurate scoring functions could be developed for the screening of virtual libraries for potential lead compounds. It might even be worthwhile to incorporate CH···O hydrogen bonds into force fields for more accurate molecular simulations and crystal structure refinements. A deeper and more widespread understanding of these interactions, expedited by work reported here and elsewhere, will allow the development and assessment of these possibilities.

REFERENCES

1. Desiraju GR, Steinter T. The weak hydrogen bond in structural chemistry and biology. Oxford,UK: Oxford University Press; 1999.
2. Donohue J. In: Rich A, Davidson N, editors. Structural chemistry and molecular biology. San Francisco: W.H. Freeman & Co.; 1968. p 459–463.
3. Taylor R, Kennard O. Crystallographic evidence for the existence of C-H...O, C-H...N, and C-H...Cl hydrogen bonds. J Am Chem Soc 1982;104:5063–5070.
4. Kollman P, McKelvey J, Johansson A, Rothenberg S. Theoretical studies of hydrogen bonded dimers. Complexes involving HF, H_2O, NH_3, HCl, H_2S, PH_3, HCN, HCP, CH_2NH, H_2S, H_2CS, H_2CO, CH_4, CF_3H, C_2H_2, C_2H_4, C_6H_6, F^-, and H_3O^+. J Am Chem Soc 1975;97:955–965.
5. Gu Y, Kar T, Scheiner S. Fundamental properties of the CH...O interaction: Is it a true hydrogen bond? J Am Chem Soc 1999;121:9411–9422.
6. Desiraju GR. The C-H...O hydrogen bond in crystals: What is it? Acc Chem Res 1991;24:290–296.
7. Scheiner S, Gu Y, Kar T. Evaluation of the H-bonding properties of CH...O interactions based upon NMR spectra. J Mol Struct Theochem 2000;500:441–452.
8. Steiner T, Sanger W. Role of C-H...O hydrogen bonds in the coordination of water molecules. Analysis of neutron diffraction data. J Am Chem Soc 1993;115:4540–4547.
9. Desiraju GR. The C-H...O hydrogen bond: Sructural implications and supramolecular design. Acc Chem Res 1996;29:441–449.
10. Weiss MS, Brandl M, Suhnel J, Pal D, Hilgenfeld R. More hydrogen bonds for the (structural) biologist. Trends Biochem Sci 2001;26:521–523.
11. Thomas KA, Smith GM, Thomas TB, Feldmann RJ. Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. Proc Natl Acad Sci USA 1982;79:4843–4847.
12. Derewenda ZS, Derewenda U, Kobos P. (His)C-H...O hydrogen bond in the active sites of serine hydrolases. J Mol Biol 1994;241:83–93.
13. Ornstein RL, Zheng Y-J. Ab initio quantum mechanics analysis of imidazole C-H...O water hydrogen bonding and a molecular mechanics forcefield correction. J Biomol Struct Dynamics 1997;14(6):657–665.
14. Derewenda ZS, Lee L, Derewenda U. The occurrence of C-H...O hydrogen bonds in proteins. J Mol Biol 1995;252:248–262.
15. Brandl M. C-H...π interactions in proteins. J Mol Biol 2001;307:357–377.

16. Dreyer MK, Borcherding DR, Dumont JA, Peet NP, Tsay JT, Wright PS, Bitonti AJ, Shen J, Kim S-H. Crystal structure of human cyclin-dependent kinase 2 in complex with the adenine-derived inhibitor H717. J Med Chem 2000;44:524–530.

17. Engh RA, Girod A, Kinzel V, Huber R, Bossemeyer D. Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitors H7, H8, H89. Structural implications for selectivity. J Biol Chem 1996;271:26157–26164.

18. Schindler T, Bornmann W, Pellicena W, Miller WT, Clarkson B, Kuriyan J. Structural mechanism for STI-571 inhibition of Abelson tyrosine kinase. Science 2000;289(5486):1938–1942.

19. Legon AC, Millen DJ. Directional character, strength, and nature of the hydrogen bond in gas-phase dimers. Acc Chem Res 1987;20: 39–45.

20. Kim K, Friesner RA. Hydrogen bonding between amino acid backbone and side chain analogues: A high-level ab initio study. J Am Chem Soc 1997;119:12952–12961.

21. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC et al. Gaussian 98. Pittsburgh, PA: Gaussian, Inc.; 1998.

22. Boys SF, Bernardi F. The calculations of small molecular interaction by the difference of separate total energies. Some procedures with reduced error. Mol Phys 1970;19:553–566.

23. Fellers RS, Leforestier C, Braly LB, Brown MG, Saykally RJ. Spectroscopic determination of the water pair potential. Science 1999;284:945–948.

24. Tschumper GS, Quack M, Matthew L. Leininger BCH, Valeev E, Schaefer HF. Anchoring the water dimer potential surface with explicitly correlated computations and focal point analyses. J Chem Phys 2002;116:690–701.