Lone Pair−Aromatic Interactions: To Stabilize or Not to Stabilize

MARTIN EGLI* AND SANJAY SARKHEL
Department of Biochemistry, Vanderbilt University, School of Medicine, Nashville, Tennessee 37232

Received July 31, 2006

ABSTRACT
The ability of aromatic rings to act as acceptors in hydrogen bonds has been demonstrated extensively both by experimental and by theoretical means. Countless examples of D−H···π (H···π, D = O, N, C) interactions have been found in the three-dimensional structures of proteins. Much less is known with regard to the occurrence of other possible noncovalent interactions with aromatics in macromolecular structures, those with a geometry that points oxygen lone pairs into the face of a π system. There has been a growing interest in such lp···π interactions in recent years, but the binding energies have mostly been studied using small-molecule model systems. We have conducted a survey of lp···π interactions in crystal structures of DNA, RNA, and proteins and used ab initio calculations to estimate their energies. Our results demonstrate that such interactions are more common in nucleic acids and that significant binding energies only result when the aromatic system is positively polarized, for example, due to protonation of a nucleobase.

I. Introduction

Noncovalent interactions form the backbone of supramolecular chemistry and include hydrogen bonds (H-bonds), stacking, electrostatic, hydrophobic, and charge-transfer interactions as well as metal ion coordination.[1] Among these interactions, H-bonds play a central role in the structure, function, and dynamics of chemical and biological systems.[2] A "conventional" H-bond may be represented as D−H···A whereby D (donor) and A (acceptor) are both electronegative atoms (usually N and O). However, H-bonding is not restricted to N and O, but may also involve less electronegative donor or acceptor functionalities. These "non-conventional" H-bonds have attracted enormous interest in chemistry and structural biology,[3] and interactions such as C−H···O,[4] C−H···π,[5] and N−H···π[6] (H···π) are ubiquitous in the structures of macromolecules (Figure 1A).

Whereas H···π interactions are expected simply from electrostatic arguments, a stabilizing effect of the interaction between a lone pair of electrons and the face of the π system (lp···π interaction) appears counterintuitive. Ab initio calculations (counterpoise-corrected, cc, MP2(full)/6-31G**) revealed that for the water−hexafluorobenzene system the magnitude of the lp···π interaction (−2.1 kcal/mol) is comparable to that of the H···π interaction between water and benzene (−1.8 kcal/mol) (ref 7 and cited lit.). A more recent report using the cc-LMP2/6-31+G* level of theory compared the H···π and lp···π interactions for the water−benzene complex; the energies are −2.7 and −0.6 kcal/mol, respectively.[8] Clearly, the higher stability of the lp···π interaction between water and hexafluorobenzene as compared to the water−benzene system is due to the presence of electron-withdrawing fluorine atoms. In the case of water−aromatic systems, the H···π arrangement is stabilized mainly by the interaction between the LUMO of water and the aromatic HOMO. On the other hand, the lp···π arrangement is stabilized by the interaction between the HOMO of water and the aromatic LUMO. The switch-point between these two arrangements occurs at molecular electronegativities somewhere between 0.125 and 0.162 atomic units (au).[8] Besides these theoretical data, the relative orientations of carbonyl groups and arene rings in crystal structures provide evidence for potentially stabilizing lp···π interactions.[9]

We had previously referred to the close contact between a lone pair of 2′-deoxycytidine O4′ and the adjacent guanine base at CpG steps in left-handed Z-DNA as an $n \rightarrow \pi^*$ interaction.[10] In the present Account, we will use the term lp···π interaction to describe noncovalent contacts between oxygen lone pair(s) and aromatics independent of the covalent chemical environment of the oxygen atom. Thus, the lone pair can be contributed by water,[7,11] ether,[10] or carbonyl[12] moieties, or even anions.[13,14] The aromatic moiety may exhibit a range of polarities or carry a positive charge. Ab initio calculations have provided evidence that lp···π interactions can afford a significant degree of stability when the aromatic moiety is protonated.[15] Thus, the interaction energy between water and protonated imidazole (Im^+) obtained at the MP2/6-31+G** level of theory was −8.1 kcal/mol.

Theoretical studies regarding lp···π interactions published to date[7,8,12−16] have been limited to simple model systems and have largely ignored the existence of such interactions in macromolecules (i.e., ref 10). In RNA U-turns,[17] a backbone phosphate group is positioned

Martin Egli (born March 23, 1961, in Schaffhausen, Switzerland) obtained his diploma in Chemistry from the ETH-Zürich and his Dr. sc. nat. in Chemical Crystallography from the same institution. He was a postdoctoral fellow in structural biology at MIT and held academic appointments at ETH and Northwestern University before being appointed as Professor of Biochemistry at Vanderbilt University. His research interests include the 3D-structures of DNA, RNA, and their analogues, stereoelectronic effects in macromolecules, the specificity and inhibition of Ser/Thr kinases, the structure and function of the cyanobacterial KaiABC circadian clock, and trans-lesion DNA polymerase−DNA interactions.

Sanjay Sarkhel (born March 11, 1971, in Guwahati, India) obtained his B.Sc. and M.Sc. in Chemistry from the Banaras Hindu University (B.H.U) and his Ph.D. in Structural and Medicinal Chemistry from the same institution for work carried out at the Central Drug Research Institute, Lucknow. He worked on anticancer drug design and the role of noncovalent interactions in protein−ligand complexes at the School of Chemistry, Hyderabad, before joining the Egli lab in 2002 as a Research Associate. His research interests include the 3D-structures of biological macromolecules and understanding structure−activity relationships, molecular dynamics, the role of water in macromolecular structure, the design of β-hairpins, and the interplay of weak and strong noncovalent interactions in molecular recognition and applications in structure-based drug design.

* To whom correspondence should be addressed. E-mail: martin.egli@vanderbilt.edu.

10.1021/ar068174u CCC: $37.00
Published on Web 12/15/2006

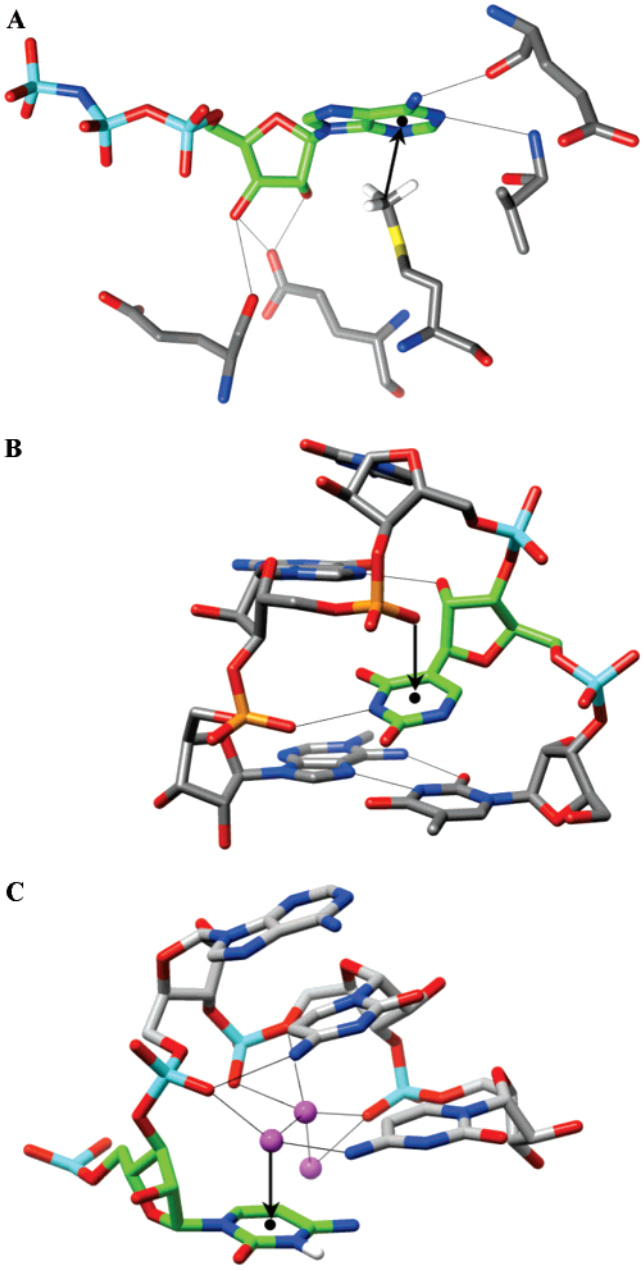

FIGURE 1. H···π versus lp···π interactions (indicated by an arrow). (A) H···π interaction between the ϵ-methyl group of Met146 and AMPPnP at the active site of death-associated protein kinase (PDB ID 1ig1). The Cϵ···centroid (solid dot) distance is 2.86 Å. (B) lp(anion)···π interaction in the Ψ-turn of $tRNA^{Phe}$ (PDB ID 1ehz). The distances between phosphate oxygens of me^1A58 and G57 (phosphorus atoms highlighted in orange) and N3 and ring centroid of Ψ55 are 2.77 and 2.92 Å, respectively. (C) lp(water)···π interaction in the C-turn of the RNA pseudoknot from Potato Leaf Roll Virus (PDB ID 2a43). Note that C at the beginning of the turn is protonated. The distance between O_W and ring centroid is 2.92 Å.

above a nucleobase, and a lone pair from the phosphate oxygen is directed into the aromatic system (Figure 1B). The observation of a water molecule sitting on top of a functionally important, protonated cytosine in the crystal structure of a luteoviral RNA pseudoknot provides unequivocal evidence for the existence of an lp···π interaction (Figure 1C).[11] An inspection of atomic-resolution crystal structures of proteins found that in quite a few

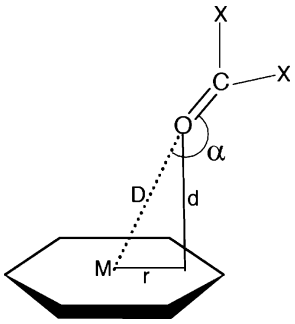

FIGURE 2. Various parameters used for characterizing $X_2C{=}O{\cdots}\pi$ interactions retrieved from the CSD. The six-membered ring represents an aromatic system. X denotes any atom, and M is the ring centroid. The CSD search allowed for any atom in all positions of the aromatic system. ω is the dihedral angle between the planes defined by $X_2C{=}O$ and the aromatic system.

cases the contacts between water molecules and the π faces of aromatic amino acids were seemingly inconsistent with an H···π interaction.[18] However, the author did not explicitly imply the existence of water−amino acid lp···π interactions in these structures.

In this Account, we will discuss the experimental evidence for the occurrence of lp···π interactions in the crystal structures of macromolecules. To assess whether the lp···π interactions seen in the structures actually contribute to stability or are merely tolerated short contacts, we have carried out ab initio calculations (equilibrium geometry or point-energy) at the MP2/6-31G* level of theory for simplified model systems in selected cases. A geometric analysis of the interactions between carbonyl groups and aromatics in small-molecule crystal structures deposited in the Cambridge Structural Database (CSD) provides further support for the existence of lp···π interactions. The combined experimental and theoretical data allow conclusions with regard to the chemical environment conducive to lp···π interactions as well as their geometry and interaction energies.

II. Carbonyl···π Interactions: CSD Analyses

We searched the Cambridge Structural Database (version 5.26, Nov. 2004) for C=O···π contacts using the following distance and angle constraints: The distance D between the carbonyl oxygen atom and the ring-centroid M had to be between 2.8 and 3.8 Å, and the dihedral angle ω between the planes defined by the $O{=}CX_2$ moiety and the aromatic ring had to be 90° or smaller (Figure 2). Any atoms were allowed in the aromatic ring; however, structures that contained H-bonds D−H···A (D, A = N/O) were excluded to ascertain that the weaker C=O···π contacts did not merely exist due to the presence of their stronger counterparts. Hits with 2.8 Å $\leq D \leq$ 3.8 Å and $\alpha \leq 180°$ (Figure 2) were grouped into three regions: $0° \leq \omega \leq 24°$ (region 1; 1191 hits), $25° \leq \omega \leq 64°$ (region 2; 329), and $65° \leq \omega \leq 90°$ (region 3; 240) (Figure 3A−C). Region 1 represents cases where the carbonyl group is stacked on the ring plane. In region 2, the carbonyl group takes an angular approach towards the ring, and in region 3, the carbonyl oxygen heads directly into the ring plane.

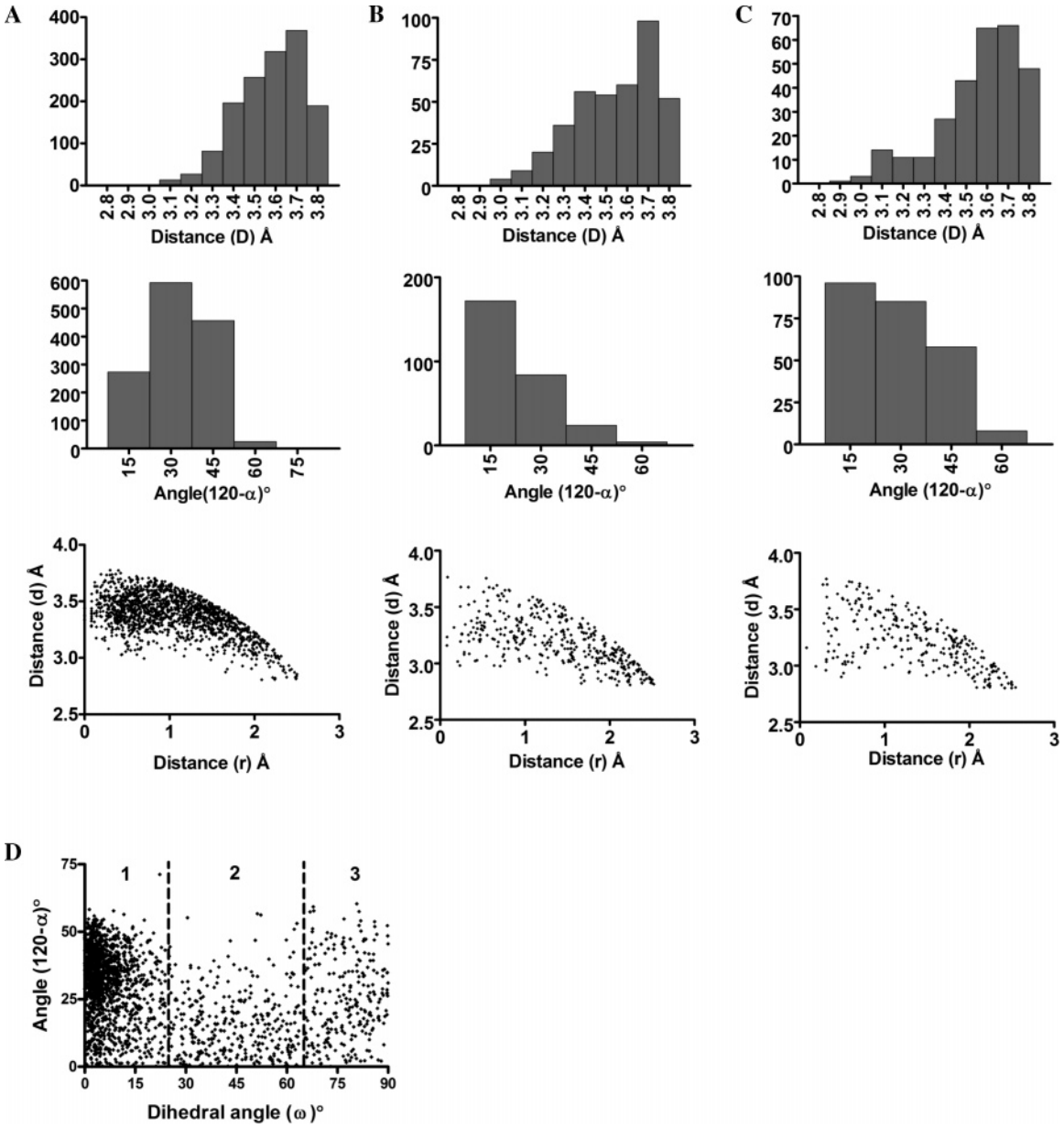

FIGURE 3. Histograms and scatter plots for $X_2C{=}O\cdots\pi$ interactions retrieved from the CSD. (A) Region 1, (B) region 2, and (C) region 3. For explanations, see main text. (D) Scatter plot of dihedral angle ω versus angle ($120 - \alpha$) for regions 1, 2, and 3.

An orientation of the carbonyl group more or less parallel to the ring plane (region 1) renders the oxygen lone pairs available for H-bond donors. However, because our search criteria excluded structures featuring conventional H-bonds, this stacked orientation of the carbonyl group is probably a consequence of the interaction of the π-orbital of the C=O bond with the π-face of the aromatic ring. Such a $\pi-\pi$ interaction between the (carboxylate) carbonyl group and the phenyl ring was thought to be responsible for weakening of H-bonds, thermal dehydration behavior, and for preventing formation of new Cu−O bonds in the crystal lattice of $[Cu_2(sgly)_2(H_2O)]-1H_2O$.[19] The mean distance D and angular distribution (deviation of α from 120°; Figure 2) are 3.58 Å and 30.6°, respectively (Figure 3A). By comparison, the mean distance D and angular distribution in region 2 are 3.54 Å and 16.8°, respectively (Figure 3B). The correlation for the scatter plot of d versus r for such an orientation of the carbonyl group (Pearson $r = -0.646$) is marginally higher as compared to the data for region 1 (Pearson $r = -0.556$). This suggests that as the carbonyl oxygen gets closer to the ring plane, it tends to be pushed away from the center (Figure 3D). The significant angular preference of such an approach is noteworthy; the mean deviation of 16.8° from the ideal geometry of 120° is almost one-half of that observed in region 1. For region 3, with the carbonyl group taking a more or less head-on approach towards the ring centroid, the mean distance D is 3.56 Å with a mean angular distribution of 25.0° about the ideal geometry (Figure 3C). The d versus r scatter plot for region 3 shows a correlation of −0.622 (Pearson r); although the correlations for regions 2 and 3 are significant, they do not exceed that for region 1 by much.

The relatively low correlations imply that the contacts do not have strict directional preferences, as one would expect for a weak interaction. In a recent survey of the environment of amide groups in protein−ligand complexes, the authors noted that such contacts are actually avoided because the electron-rich C═O oxygen and the aryl ring π electrons should repel each other.[20] The smaller number of hits for this type of contact (region 3) as compared to the number of hits in regions 1 and 2 appears to support the authors' conclusion, and it is also consistent with the results of possibly the earliest analysis of the oxygen environment of phenylalanine aromatic rings in protein structures.[21] Interestingly, another report relying on crystallographic data provides direct evidence of energetically and structurally significant interactions between C═O groups and arene rings, where the carbonyl group points toward the centroid of the aromatic ring with O···centroid distances between 2.8 and 3.2 Å.[9] At first sight, the results of the two studies appear to contradict each other. However, it may be possible to reconcile them if we consider that the molecular electronegativities of aromatic rings can vary considerably in the individual structures. Accordingly, aromatic systems with a molecular electronegativity of <0.15 au tend to favor an H···π contact, whereas the lp···π interaction is preferred for molecular electronegativities >0.15 au.[8]

In light of these observations, it is worth discussing another class of interactions that at first appears counterintuitive but has now been well documented, the anion···π interaction. Based on crystallographic evidence, anion···π interactions have been shown in Ag(I)−aromatic complexes,[14] between *s*-triazine and chloride,[22] and between *s*-tetrazine and the AsF_6^- ion.[23] In addition, crystallographic and computational studies have established that isocyanuric acids are suitable binding units for anions.[13] Recent reports regarding interactions in solution have highlighted halide recognition through aromatic receptors based on anion···π interactions[24,25] and anion binding at the peripherial nitrogen of a C_6F_5-substituted N-fused porphyrin.[26,27] The stability of anion···π interactions is derived from electrostatic and ion-induced polarization forces. The former is represented by the permanent and positive quadrupole moment of the aromatic system,[28] whereas the latter depends on the molecular polarizability of the system.[29]

III. H···π versus lp···π Interactions in H_2O−Aromatic Amino Acid and H_2O−Nucleobase Systems

Using ab initio calculations, Scheiner et al. compared the relative energies of various types of H-bonds involving aromatic amino acids.[15] Model systems for the side chains of Phe, Tyr, Trp, and His with water H-bonding either in the plane of the aromatic moiety or pointing into the face of the π system with lone pair or hydrogen were treated at the cc-MP2/6-31+G** level of theory. For imidazole (Im; His model), the H···π and lp···π interaction energies were −3.1 (Im) and −8.1 kcal/mol (Im^+), respectively (Figure 4A, REF ΔE). Provided that the His side chain is protonated, the lp···π interaction is similar in stability to that of a standard (water) O−H···N (Im) H-bond (−6.0 kcal/mol, Figure 4A). We used the published data for the Im/Im^+···water system to calibrate our ab initio calculations. Equilibrium geometry calculations in Spartan (Spartan '04 for MacIntosh, Wavefunction Inc., http://www.wavefun-.com)[30] performed at various levels of theory indicated reasonable correspondence between the published interaction energies and those from MP2/6-31G*-type calculations (Figure 4A). The aim here was not to use the highest level of theory available but one that has been demonstrated to be competent and handle the water−aromatic systems reasonably well (i.e., refs 7 and 15).

One noteworthy insight from the theoretical analysis concerns the different relative energies afforded by H···π and lp···π interactions between water and uracil (U) or cytosine (C; Figure 4B and C, respectively). Thus, the interaction energy of the lp···π water−U complex is comparable to that of the H···π water−C and water−Im complexes. U appears to be more positively polarized than either of the other aromatic systems. However, like His, C can be protonated (Figure 1C)[11] and the energy of the lp···π interaction between water and C^+ matches that of a strong H-bond (Figure 4C). Another difference that is apparent from the calculations concerns the distance between water oxygen and the centroid of the aromatic ring. This distance is typically somewhere between 2.7 and 3.1 Å for lp···π interactions and between 3.2 and 3.5 Å for H···π interactions. Consistent with the higher stabilization of the lp···π interaction for the water−U as compared to the water−C complex, the distance between O_W and the ring plane is considerably shorter in the former (2.85 vs 3.13 Å, respectively).

IV. Lone Pair···π Interactions in Nucleic Acids

A. Observations in Crystal Structures of Oligonucleotides. In the canonical right-handed double helical conformations, the nucleobases are stacked in the core and are typically unavailable for H···π and lp···π interactions. However, structural motifs such as bulged nucleosides or hairpin loops render the bases available for interactions other than stacking. The Z-DNA duplex differs from the A- and B-form duplexes not only because it is left-handed, but it also exhibits an unusual "stacking" of the cytidine 2′-deoxyribose on G. This leads to the aforementioned lp···π interaction between O4′ (C) and the guanidinium moiety of G (Figure 5A).[10] Close contacts between sugars and nucleobases are not uncommon in the structures of nucleic acids. Except for the 2′-OH group in RNA, the sugar−phosphate backbones of oligonucleotides are devoid of H-bond donors. Packing interactions between duplexes can be mediated by metal ions, H-bonding between bases, end-to-end base stacking, and close contacts between sugars and bases, among others. As expected purely from electrostatics, the close spacing of sugars and bases typically involves H···π interactions

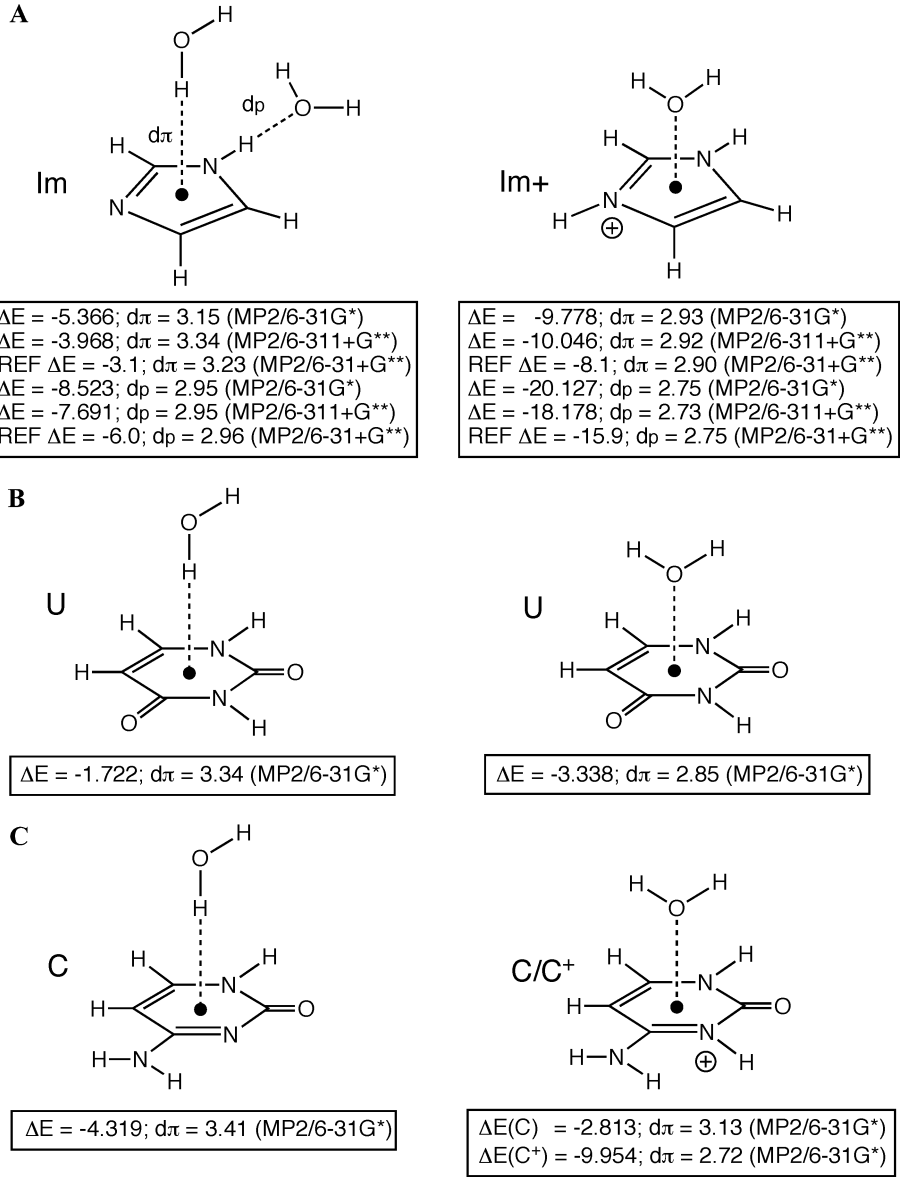

FIGURE 4. Equilibrium energies in kcal/mol and distances in Å for H···π and lp···π interactions between water and (A) imidazole, (B) uracil, and (C) cytosine calculated at various levels of theory. REF values refer to published data.[15]

(i.e., ref 31). However, lp(O4′)···A interactions with distances similar to the those for the lp(O4′)···π interactions in Z-DNA duplexes (2.78−2.96 Å)[10] were also observed in the crystal structure of cyclic r(ApAp) (see Figure 5A in ref 32). In the crystal structure of a DNA dodecamer complexed with propamidine, the minor groove binder displaces the spine of water molecules in the central A-tract and establishes H-bonds with its amidinium groups to the DNA bases and deoxyribose sugars.[33] The O4′ atoms lie as close as 3.2 Å from the centroids of the positively polarized aromatic moieties of the drug molecule and may favorably affect the stability of the complex through lp···π interactions.

RNA molecules exhibit intricate tertiary structural motifs, and the so-called uridine turn (U-turn) is among the most common building blocks in RNA.[17] It was first identified in the pseudouridine (Figure 1B) and anticodon loops of tRNA^{Phe} [34] and is also present in the structure of the hammerhead ribozyme.[35−37] At U-turns, the RNA backbone makes a sharp turn between the first and the second nucleotide of the UNR sequence (N = any base, R = purine; Ψ55−C56−G57 in Figure 1B). The turn is stabilized by two H-bonds, between the 2′-OH of Ψ55 and N7 of G57 and between the N3 of Ψ55 and the phosphate of me^{1}A58, as well as stacking between the bases of the second and third residues (Figure 1B). Another conserved feature of the U-turn is the close contact between the phosphate group of the third residue and the face of the nucleobase of the first (G57 and Ψ55, respectively, in the pseudouridine turn of tRNA^{Phe}; Figure 1B) that can be classified as an lp···π interaction. Model systems for the theoretical analyses that contain the phosphate−base lp···π interaction and/or the phosphate−base "pair" are depicted in Figures 5B and 6B, respectively. Subsequently, it was found that the ubiquitous RNA GNRA-tetraloops also make a U-turn.[17] In the structure of the hammerhead ribozyme published by Scott et al., the sequence of the tetraloop is GUAA[37] and the phosphate of the first A stacks

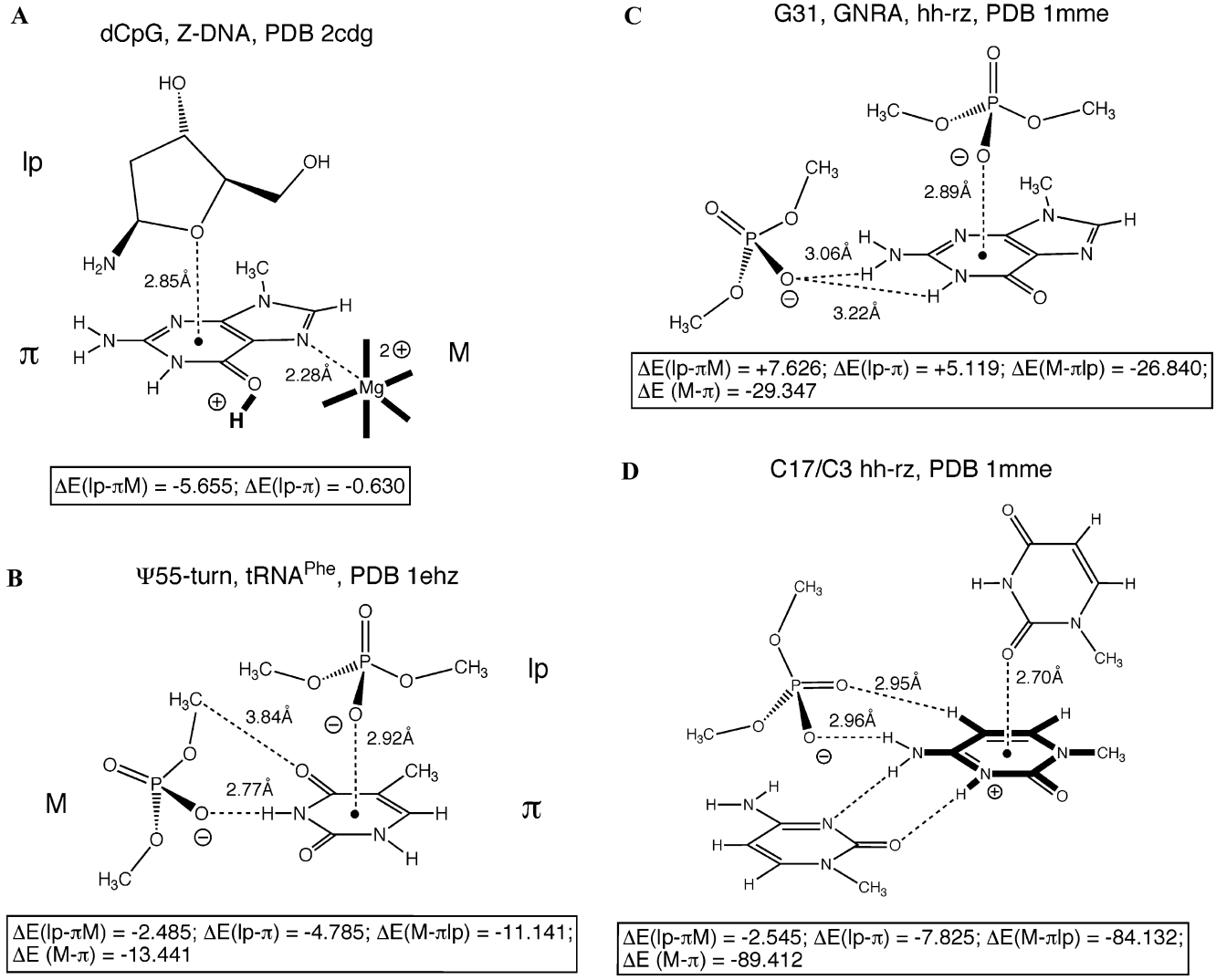

FIGURE 5. Point-energies of lp···π interactions in kcal/mol calculated at the DFT/6-31G* level of theory using crystallographic coordinates. The four systems analyzed are (A) the 2′-deoxyribose sugar of C stacked on the 3′-adjacent G in Z-DNA,[10] (B) the Ψ-turn in tRNA^{Phe},[41] (C) the GUAA-turn in the structure of the all-RNA hammerhead ribozyme,[37] and (D) the protonated C adjacent to the cleavage site in the same hammerhead ribozyme. π refers to the aromatic system (nucleobase) and lp and M (modulator) to the moieties interacting with the nucleobase face-on and in-plane, respectively. In the model shown in panel A, both protonation of G and a Mg^{2+} coordinated to N7 were assessed as modulators, and the energy ΔE(lp−πM) is for the complex with G^+. In the model shown in panel D, only C17 (thick lines) was considered in the calculation, but C3 is included to show the formation of the putative C^+:C base pair with two H-bonds. The phosphodiester moiety in DNA and RNA was modeled as a dimethyl phosphate anion.

on G. A model system for this lp···π interaction is depicted in Figure 5C.

In the initial crystal structures of hammerheads,[36,37] residue C17 in the substrate strand whose 2′-OH attacks the 3′-adjacent phosphate, subsequently resulting in strand cleavage, pairs with C3 from the ribozyme strand near the U-turn of the latter. Although the pairing clearly involves two H-bonds judging from the N3(C17)−O2(C3) and N4(C17)−N3(C3) distances in the structures, both author teams assigned only a single H-bond to the C17:C3 pair (N4−N3; see, i.e., Figure 5b in ref 36). Thus, they may have overlooked that C17 is protonated. Protonation is consistent with the base-pairing mode, the presence of a phosphate at H-bonding distance from N4 and C5 of C17, an lp···π interaction involving O2 of a neighboring U and the face of C17 (see Figure 4a of ref 37), and the pH conditions used in the crystallization experiments (6–6.5; Figure 5D). The pairing mode of C is a reliable indicator of its protonation state and initially led us to conclude that a functionally important C in the structure of a frameshifting RNA pseudoknot is protonated and engages in an lp···π interaction (Figure 1C; ref 11 and cited lit.).

B. Results of Ab Initio Calculations. To assess the stabilities of the lp···π interactions described above, we calculated point-energies at the DFT/6-31G* level for model systems (Figure 5) based on the crystallographic coordinates retrieved from the Protein Data Bank (http://www.rcsb.org).[38] The stacking of a deoxyribose on the guanine base as observed in Z-DNA is practically neutral in terms of the interaction energy (Figure 5A, ΔE(lp−π)). However, a cooperative effect by a modulator such as Mg^{2+} coordinated to N7 of G results in a sizeable stabilization (ca. −2 kcal/mol). Not surprisingly, a hypothetical

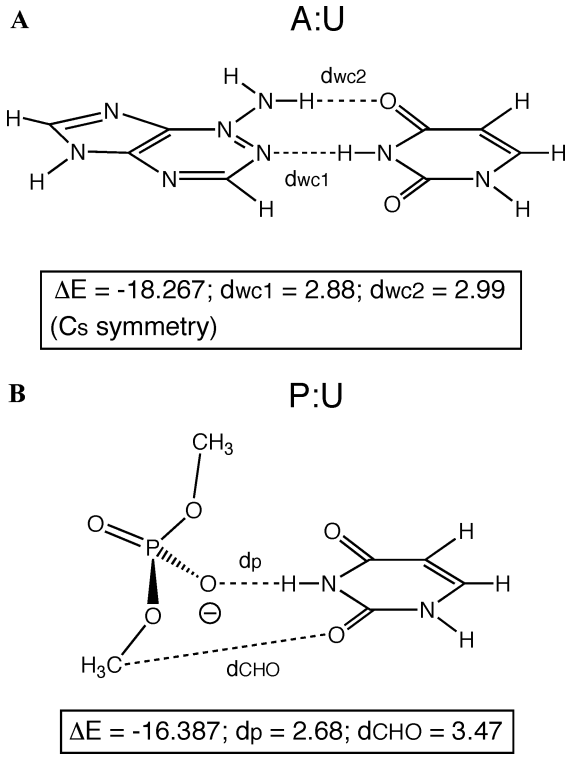

FIGURE 6. Comparison between the equilibrium interaction energies in kcal/mol (distances in Å) obtained at the MP2/6-31G* level for (A) a Watson–Crick A:U base pair and (B) an in-plane uracil–phosphate interaction (P:U). The P:U "pair" is a component of the RNA U-turn motif (see Figures 1B, 5B).

protonation of G at O6 leads to even higher stability of the lp···π interaction (>−5.6 kcal/mol; Figure 5A). Similarly, the lp···π interaction between phosphate and pseudouridine in the U-turn is stabilizing when considered alone (Figure 5B). However, the in-plane H-bond between a phosphate group and Ψ diminishes the strength of the lp···π interaction. This is not surprising, as one would expect a H-bond between the aromatic moiety and a strong acceptor such as a phosphate to exert a negative cooperative effect on the lp···π interaction. In some ways, this scenario constitutes the opposite of the coordination of a metal cation to guanine in Z-DNA, which leads to a positive polarization of the nucleobase and thus enhances the lp···π interaction. Cooperative effects of conventional and unconventional hydrogen bonds involving imidazole have recently been analyzed in detail with theoretical means.[39]

Ab initio equilibrium geometry calculations (MP2/6-31G*) indicate that the "pairing" between a phosphate group and U virtually matches the base pairing energy of A and U (Figure 6). The close correspondence of the binding energies in the two pairs is noteworthy. In light of this observation, it is probably reasonable to view the stacking of the phosphate on uracil (or Ψ) in the RNA U-turn as an interaction that is tolerated rather than significantly stabilizing. Clearly, as compared to the contributions to stability of the turn due to base stacking and H-bonds, the phosphate–base sandwich is of secondary importance. This is consistent with the results of the point-energy calculations for the lp···π interaction

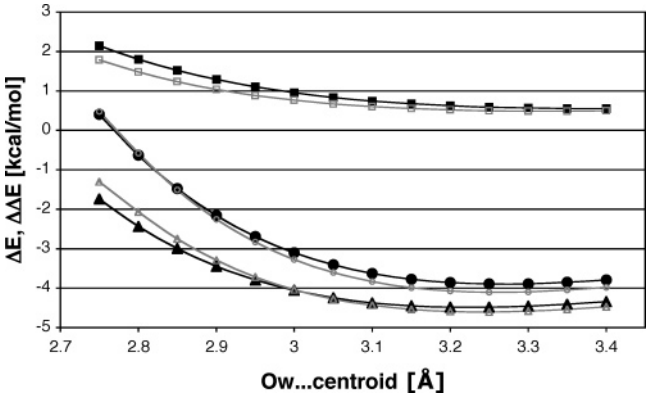

FIGURE 7. Dependence of the point-energies ΔE obtained from ab initio calculations at the MP2/6-31G* level on the distance between oxygen (water) and the ring centroid of Tyr or Trp (five-membered ring) for H···π (Tyr ●, Trp ○)- and lp···π (Tyr ■, Trp □)-type interactions. Only the side chains of Tyr (*p*-cresol, 4-methylphenol) and Trp were considered in the calculations. The water molecule approaches along the normal to the ring plane with either an O–H bond or the oxygen lone pairs pointing into the face of the π system. The lower curves (Tyr ▲, Trp △) show values for $\Delta\Delta E = \Delta E$(H···$\pi$) − ΔE(lp···π).

between phosphate and G in the model system for the GUAA tetraloop, which actually indicate a destabilizing effect (Figure 5C). On the contrary, the lp···π interaction between O2(U) and C^+ near the scissile bond in the structure of the hammerhead ribozyme makes a favorable contribution to stability, although the in-plane H-bond to a phosphate group again leads to a reduced net stability of the interaction between carbonyl oxygen and base (Figure 5D). While the numerical data certainly need to be treated with the necessary caution, it is probably justified to conclude that lp···π interactions in nucleic acids can make significant contributions to stability when the nucleobase is positively polarized due to the particular chemical environment or when it involves protonated cytosine. By analogy, the presence of an lp···π interaction in the structures of DNA and RNA may serve as a reporter of an unusual polarization or protonation state of a nucleobase.

V. Lone Pair···π Interactions in Proteins

Steiner inspected 75 protein crystal structures at resolutions <1.1 Å with the goal to establish O_W–H···π interactions between water and aromatic residues.[18] The positions of water hydrogen atoms cannot be reliably determined in X-ray crystal structures of proteins even at high resolution. Therefore, the local environment of water was screened in the hope of confirming the existence of an H···π interaction based on the distribution of acceptor and donor atoms around the water molecule. Of the 18 water molecules found in close contact with Phe, Tyr, or Trp, five were deemed to be "likely" engaged in an H···π interaction, and two were deemed "unlikely" cases. For the majority of the remaining cases (9), it was concluded that it was impossible to determine whether the close water–aromatic contact corresponded to an H···π inter

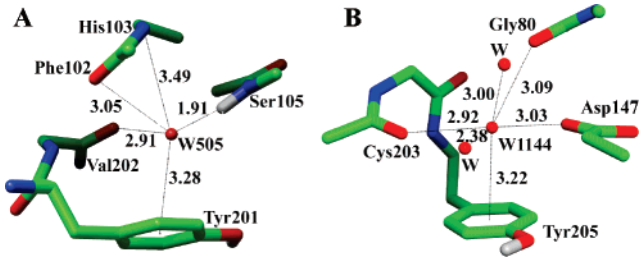

FIGURE 8. Possible lp···π interactions between water and Tyr in high-resolution crystal structures of proteins. (A) PDB ID 1qj4; W505 is in the vicinity of three H-bond acceptor groups (Val 202, Phe 102, and His 103) and one donor functionality (Ser 105). (B) PDB ID 1fsg; W1144 is in the vicinity of three H-bond acceptor groups (Cys 203, Gly 80, and Asp 147) and two water molecules (that may be involved in a cooperative network of H-bonds). Residues are labeled, and distances are in Ångstrom.

action; only in two cases the possibility of an H···π interaction was rejected.[18]

We revisited the water coordination in each case and also looked into the second and third coordination shells to see if these would shed light on the orientation of the water molecule relative to the aromatic moiety. Indeed, the H-bonding environment of water alone cannot establish whether the water engages in an H···π or an lp···π interaction. However, the distance between O_W and ring centroid is ≥3.1 Å in all cases, and none of the contacts involves His. In all instances of experimentally observed lp···π interactions discussed before, the distance between oxygen and ring centroid is clearly below 3 Å. This distance range is confirmed by the results of ab initio calculations (Figure 4).[7,15] Therefore, the distances of O_W···π interactions observed in the protein structures argue against an orientation of the water with either one or both lone pairs directed into the face of Phe, Tyr, or Trp. Point-energies calculated for H···π and lp···π interactions between water and either Tyr or Trp as a function of the distance between water oxygen and ring centroid essentially support this conclusion (Figure 7). Neither Tyr nor Trp exhibit a polarization that is conducive to formation of an lp···π interaction, contrary to a protonated His that favors interaction with a lone pair (Figure 4A).

What about the two cases for which Steiner had deemed the formation of an lp···π interaction impossible? Figure 8 depicts the two water···Tyr contacts, with the water molecule surrounded by two or more conventional H-bond acceptors in both cases. The most obvious scenario is that water donates in H-bonds; therefore, it appears likely that the water directs a lone pair into the ring. However, in both cases it is possible to orient the water molecule in a way that both satisfies the surrounding acceptors (i.e., by sharing a water hydrogen atom) and allows formation of an H···π interaction (data not shown). As pointed out above, the O_W···centroid distances are considerably longer than expected for an lp···π interaction. On the other hand, not every short contact observed in a crystal lattice needs to be stabilizing. Accordingly, in both examples shown in Figure 8, it is possible that water forms strong H-bonds to the different acceptors even if that resulted in a destabilizing lp···π interaction with Tyr. Irrespective of such considerations, we can infer that the formation of an lp···π interaction between amino acids and water in the structures of proteins occurs only very rarely.

VI. Conclusions

We have reviewed known and potential lp···π interactions in the crystal structures of small molecules, oligonucleotides, and proteins. Ab initio calculations provide evidence that among the aromatic moieties encountered in biomacromolecules, U (and Ψ), C^+, and His^+ are the most likely to engage in an lp···π interaction. When the interaction involves a positively charged aromatic system, the resulting binding energy can match that of a strong hydrogen bond. Although there are examples of oxygen lone pairs interacting with the π face of U (or Ψ; carbonyl, phosphate anion) and C^+ (water), the experimental observation of an lp···His^+ contact remains elusive. The different distances between oxygen and the aromatic plane for lp···π (≪3 Å) and H···π interactions (≫3 Å) render the distinction between them quite straightforward. The decision whether the interaction is of the lp···π or the H···π type can be challenging when the oxygen atom approaching the aromatic system belongs to a water molecule. This is because the location of hydrogen atoms is not reliably known in crystal structures of proteins even at very high resolution. The renewed interest in neutron macromolecular crystallography should prove helpful in this respect and may lead to the identification of lp(water)···His^+ interactions. At the very least, the reader should appreciate that not every close contact between an oxygen atom and the π face of an aromatic moiety represents an H···π interaction. Clearly, stabilizing lp···π interactions are going to be more common with nucleic acids than with proteins because nucleobases can be positively polarized (unlike Phe, Tyr, and Trp) and the backbone of the former is devoid of potential H-bond donors except for the RNA 2′-hydroxyl group. Unfortunately, a systematic search of the structures of nucleic acids and nucleic acid–protein complexes stored in the Protein Data Bank with the goal to retrieve additional examples of lp···π interactions is currently impossible as search tools such as Relibase[40] are only designed to handle protein–protein and protein–water interactions.

Financial support from the National Institutes of Health is gratefully acknowledged (grant R01 GM55237 to M.E.). This project was carried out during a sabbatical stay by M.E. in the Department of Chemistry, Oxford University, Oxford, UK, and M.E. would like to thank Professor Graham Richards for helpful discussions and financial assistance.

References

(1) Lehn, J.-M., Atwood, J. L., Davies, J. E. D., MacNicol, D. D., Vögtle, F., Eds. *Comprehensive Supramolecular Chemistry*; Pergamon: Oxford, UK, 1996.
(2) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, 1991.
(3) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond in Structural Chemistry and Biololgy*; Oxford University Press: Oxford, U.K., 1999.

(4) Pierce, A. C.; Sandretto, K. L.; Bemis, G. W. Kinase inhibitors and the case for C-H···O hydrogen bonds in protein-ligand binding. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 567−576.
(5) Nishio, M.; Hirota, M.; Umezawa, Y. *The CH/π Interaction. Evidence, Nature, and Consequences*; Wiley−VCH: New York, 1998.
(6) Steiner, T.; Koellner, G. Hydrogen bonds with π-acceptors in proteins: frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* **2001**, *305*, 535−557.
(7) Gallivan, J. P.; Dougherty, D. A. Can lone pairs bind to a π system? The water···hexafluorobenzene interaction. *Org. Lett.* **1999**, *1*, 103−105.
(8) Reyes, A.; Fomina, L.; Rumsh, L.; Fomine, S. Are water-aromatic complexes always stabilized due to π-H interactions? LMP2 study. *Int. J. Quantum Chem.* **2005**, *104*, 335−341.
(9) Ganis, P.; Valle, G.; Pandolfo, L.; Bertani, R.; Visentin, F. Further crystallographic evidence of NH···π (system) and CO···π (system) interactions: the structures of bis(diarylhydrazonecarbonyl)-methylene derivatives [{ArPh═NNH−C(O)}$_2$CH$_2$] (Ar = Ph, 2-C$_5$H$_4$N, 2-C$_4$H$_3$S). *Biopolymers* **1999**, *49*, 541−549.
(10) Egli, M.; Gessner, R. V. Stereoelectronic effects of deoxyribose O4′ on DNA conformation. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 180−184.
(11) Sarkhel, S.; Rich, A.; Egli, M. Water-nucleobase "stacking": H−π and lone pair−π interactions in the atomic resolution crystal structure of an RNA pseudoknot. *J. Am. Chem. Soc.* **2003**, *125*, 8998−8999.
(12) Li, Y.; Snyder, L. B.; Langley, D. R. Electrostatic interaction of π-acidic amides with hydrogen-bond acceptors. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3261−3266.
(13) Frontera, A.; Saczewski, F.; Gdaniec, M.; Dziemidowicz-Borys, E.; Kurland, A.; Deyà, P. M.; Quiñonero, D.; Garau, C. Anion−π interactions in cyanuric acids: a combined crystallographic and computational study. *Chem.-Eur. J.* **2005**, *11*, 6560−6567.
(14) Schottel, B. L.; Chifotides, H. T.; Shatruk, M.; Chouai, A.; Pérez, L. M.; Bacsa, J.; Dunbar, K. R. Anion−π interactions as controlling elements in self-assembly reactions of Ag(I) complexes with π-acidic aromatic rings. *J. Am. Chem. Soc.* **2006**, *128*, 5895−5912.
(15) Scheiner, S.; Kar, T.; Pattanayek, J. Comparison of various types of hydrogen bonds involving aromatic amino acids. *J. Am. Chem. Soc.* **2002**, *124*, 13257−13264.
(16) Tarakeshwar, P.; Kim, K. S.; Brutschy, B. Fluorobenzene···water and difluorobenzene···water systems: an *ab initio* investigation. *J. Chem. Phys.* **1999**, *110*, 8501−8512.
(17) Jucker, F. M.; Pardi, A. GNRA tetraloops make a U-turn. *RNA* **1995**, *1*, 219−222.
(18) Steiner, T. Hydrogen bonds from water molecules to aromatic acceptors in very high-resolution protein crystal structures. *Biophys. Chem.* **2002**, *95*, 195−201.
(19) Yang, X.; Wu, D.; Ranford, J. D.; Vittal, J. J. Influence of the C═O···π interaction on the thermal dehydration behavior of [Cu$_2$(sgly)$_2$(H$_2$O)]·1H$_2$O. *Cryst. Growth Des.* **2005**, *5*, 41−43.
(20) Cotesta, S.; Stahl, M. The environment of amide groups in protein-ligand complexes: H-bonds and beyond. *J. Mol. Model.* **2006**, *12*, 436−444.
(21) Thomas, K. A.; Smith, G. M.; Thomas, T. B.; Feldmann, R. J. Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4843−4847.
(22) Demeshko, S.; Dechert, S.; Meyer, F. Anion−π interactions in a carousel copper(II)-triazine complex. *J. Am. Chem. Soc.* **2004**, *126*, 4508−4509.
(23) Schottel, B. L.; Bacsa, J.; Dunbar, K. R. Anion dependence of Ag-(I) reactions with 3,6-bis(2-pyridyl)-1,2,4,5-tetrazine (bptz): isolation of the molecular propeller compound [Ag$_2$(bptz)$_3$][AsF$_6$]$_2$. *Chem. Commun.* **2005**, 46−47.
(24) Rosokha, Y. S.; Lindeman, S. V.; Rosokha, S. V.; Kochi, J. K. Halide recognition through diagnostic anion-π interactions: molecular complexes of Cl-, Br-, and I- with olefinic and aromatic receptors. *Angew. Chem., Int. Ed.* **2004**, *43*, 4650−4652.
(25) de Hoog, P.; Gamez, P.; Mutikainen, I.; Turpeinen, U.; Reedijk, J. An aromatic anion receptor: anion-π interactions do exist. *Angew. Chem., Int. Ed.* **2004**, *43*, 5815−5817.
(26) Maeda, H.; Furuta, H. *N*-confused porphyrins as new scaffolds for supramolecular architecture. *J. Porphyrins Phthalocyanines* **2004**, *8*, 67−75.
(27) Maeda, H.; Osuka, A.; Furuta, H. Anion binding properties of *N*-confused porphyrins at the peripheral nitrogen. *J. Inclusion Phenom. Macrocycl. Chem.* **2004**, *49*, 33−36.
(28) Quiñonero, D.; Garau, C.; Rotger, C.; Frontera, A.; Ballester, P.; Costa, A.; Deyà, P. M. Anion−π interactions: do they exist? *Angew. Chem., Int. Ed.* **2002**, *41*, 3389−3392.
(29) Garau, C.; Frontera, A.; Quiñonero, D.; Ballester, P.; Costa, A.; Deyà, P. M. A topological analysis of the electron density in anion-π interactions. *ChemPhysChem* **2003**, *4*, 1344−1348.
(30) Hehre, W. J. *A guide to molecular mechanics and quantum chemical calculations*; Wavefunction: Irvine, CA, 2003.
(31) Li, F.; Sarkhel, S.; Wilds, C. J.; Wawrzak, Z.; Prakash, T. P.; Manoharan, M.; Egli, M. 2′-Fluoroarabino- and arabinonucleic acid show different conformations, resulting in deviating RNA affinities and processing of their heteroduplexes with RNA by RNase H. *Biochemistry* **2006**, *45*, 4141−4152.
(32) Frederick, C. A.; Coll, M.; van der Marel, G. A.; van Boom, J. H.; Wang, A. H.-J. Molecular structure of cyclic deoxyadenylc acid at atomic resolution. *Biochemistry* **1988**, *27*, 8350−8361.
(33) Nunn, C. M.; Neidle, S. Sequence-dependent drug binding to the minor groove of DNA: crystal structure of the DNA dodecamer d(CGCAAATTTGCG)$_2$ complexed with propamidine. *J. Med. Chem.* **1995**, *38*, 2317−2325.
(34) Quigley, G. J.; Rich, A. Structural domains of transfer RNA molecules. *Science* **1976**, *194*, 796−806.
(35) Wedekind, J. E.; Kay, D. B. Crystallographic structures of the hammerhead ribozyme: relationships to ribozyme folding and catalysis. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 475−502.
(36) Pley, H. W.; Flaherty, K. M.; Kay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **1994**, *372*, 68−74.
(37) Scott, W. G.; Finch, J. T.; Klug, A. The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **1995**, *81*, 991−1002.
(38) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(39) Kar, T.; Scheiner, S. Cooperativity of conventional and unconventional hydrogen bonds involving imidazole. *Int. J. Quantum Chem.* **2006**, *106*, 843−851.
(40) Bergner, A.; Günther, J.; Hendlich, M.; Klebe, G.; Verdonk, M. Use of Relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects. *Biopolymers (Nucleic Acid Sci.)* **2002**, *61*, 99−110.
(41) Shi, H.; Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **2000**, *6*, 1091−1105.

AR068174U