Lone pair $\cdots \pi$ interactions between water oxygens and aromatic residues: Quantum chemical studies based on high-resolution protein structures and model compounds

Alok Jain,¹ Venkatnarayan Ramanathan,²
and Ramasubbu Sankararamakrishnan^{1*}

¹Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India

²Department of Chemistry, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India

Received 7 November 2008; Revised 19 December 2008; Accepted 29 December 2008

DOI: 10.1002/pro.67

Published online 11 January 2009 proteinscience.org

Abstract: The π electron cloud of aromatic centers is known to be involved in several noncovalent interactions such as C—H $\cdots\pi$, O—H $\cdots\pi$, and $\pi\cdots\pi$ interactions in biomolecules. Lone-pair (lp) $\cdots \pi$ interactions have gained attention recently and their role in biomolecular structures is being recognized. In this article, we have carried out systematic analysis of high-resolution protein structures and identified more than 400 examples in which water oxygen atoms are in close contact (distance < 3.5 Å) with the aromatic centers of aromatic residues. Three different methods were used to build hydrogen atoms and we used a consensus approach to find out potential candidates for lp $\cdots\pi$ interactions between water oxygen and aromatic residues. Quantum mechanical calculations at MP2/6-311++G(d,p) level on model systems based on protein structures indicate that majority of the identified examples have energetically favorable interactions. The influence of water hydrogen atoms was investigated by sampling water orientations as a function of two parameters: distance from the aromatic center and the angle between the aromatic plane and the plane formed by the three water atoms. Intermolecular potential surfaces were constructed using six model compounds representing the four aromatic amino acids and 510 different water orientations for each model compound. Ab initio molecular orbital calculations at MP2/6-311++G(d,p) level show that the interaction energy is favorable even when hydrogen atoms are farthest from the aromatic plane while water oxygen is pointing toward the aromatic center. The strength of such interaction depends upon the distance of water hydrogen atoms from the aromatic substituents. Our calculations clearly show that the lp $\cdots\pi$ interactions due to the close approach of water oxygen and aromatic center are influenced by the positions of water hydrogen atoms and the aromatic substituents.

Keywords: noncovalent interaction; dispersion energy; B-factors; water–aromatic interactions; ab initio calculation; crystal structure analysis

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Ministry of Human Resources and Development (MHRD).

*Correspondence to: Ramasubbu Sankararamakrishnan, Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India. E-mail: rsankar@iitk.ac.in

Introduction

Water is ubiquitous in high-resolution crystal structures of biomolecules. In protein structures, they are usually seen in different sites of the molecule and are found on surfaces, in crevices, at binding interfaces and at times buried in the interior.¹ These water molecules are often integral part of the structures, and have

been observed to play a critical role in protein folding, structure, activity, dynamics, and protein–ligand interaction.^{2–11} These properties are, in general, driven by water's hydrogen bonding capacity with main chain carbonyl, amide and other electronegative atoms of side chain. In addition to hydrogen-bonding interactions, water oxygens (O_w) may also participate in $O_w-H\cdots\pi$ interaction with the π electron clouds present in aromatic rings of phenylalanine, tryptophan, histidine, and tyrosine residues.¹² Short contacts between O_w and aromatic residues in very high-resolution protein structures were analyzed by Steiner and validated the existence of aromatic hydrogen bonding in protein structures.¹³ Nevertheless it is difficult to assign the exact position of hydrogen atoms in water molecules to infer the true nature of interactions. Besides hydrogen bonding, it has been reported that the two lone-pair (lp) electrons of O_w are found to be involved in a novel $lp\cdots\pi$ interaction.^{14–18} This novel interaction along with anion $\cdots\pi$ interaction is gaining attention in the recent past.^{12,15,17,19–29} In $O_w-H\cdots\pi$ interactions, LUMO of the water and HOMO of the aromatic ring are involved and in stark contrast, in $lp\cdots\pi$ interactions, HOMO of the water and LUMO of the aromatic ring seem to interact.³⁰ Recently, Egli and Sarkhel have proposed water–aromatic ring contact where the lone pairs of O_w interact with purine systems in RNA after solving the crystal structure of a ribosomal frame-shifting RNA pseudoknot from a beet western yellow virus.¹⁵ Subsequently, the authors revisited the problem to validate this interaction by surveying the structures of DNA and Cambridge Structure Database (CSD) coupled with quantum mechanical calculations.¹⁷ Gallivan and Dougherty¹⁴ have shown the existence of similar interaction between lone pairs of O_w and electron deficient aromatic ring by performing various ab initio and DFT calculations. Recent theoretical kinetics studies coupled with quantum mechanical calculation have shown that $lp\cdots\pi$ interaction involving oxygen lone pairs plays an important role in protein-coupled electron transfer reaction along with $\pi\cdots\pi$ interaction.³¹ More recently, Gung *et al.* reported a quantitative study of the interaction between oxygen lone pairs and electron rich aromatic ring by determining the free energy of interaction through low temperature 1H NMR studies and quantum mechanical (QM) calculations.²⁸ In our earlier report, we have investigated the proximity of carbonyl oxygen and aromatic residues in protein structures and we showed that $lp\cdots\pi$ interactions cannot be completely ruled out between these groups.²⁵ The close contacts between backbone $C=O$ and aromatic centers occurred predominantly in helical structures and to some extent in β -strands. The favorable and stable nature of this interaction was investigated using QM and molecular dynamics (MD) studies and it was concluded that such interactions could be stabilizing the secondary structures in proteins.

In this article, we have extended the study further to identify and investigate the water–aromatic interaction between water oxygen atom and the π -electron cloud of the aromatic ring. We have carried out extensive analysis of high-resolution protein structures (resolution ≤ 1.8 Å) with a motivation to establish the existence of lone-pair $\cdots\pi$ interaction. Our results are supported with adequate quantum mechanical calculation on different model systems. We have clearly established that water hydrogen atoms have the capability to influence the strength of the lone-pair $\cdots\pi$ interactions even when they are positioned away from the aromatic center.

Results

The initial database search yielded 427 cases in which water oxygen was found to be within 3.5 Å distance from one of the aromatic centers [Fig. 1(A)]. After building hydrogen atoms, GROMACS-OPLS, GROMACS-ffg43a1, and InsightII methods resulted, respectively in 138, 139, and 112 examples that can be considered for potential $lp\cdots\pi$ interactions. These cases were obtained after several stages of rigorous filtering criteria as described in the “Materials and Methods” section. For further analysis and characterization, we wanted to consider only those examples that are found in all the three methods. This criterion led us to only 21 examples in which the lone-pair electrons of water oxygen atoms could possibly interact with the π electrons of aromatic rings (Table I). Analysis of the distance ‘ d ’ reveals that water oxygen atom is too close to the aromatic centers in two structures 1H2R (2.76 Å) and 1QNF (2.47 Å). Such an arrangement is likely to result in van der Waals clashes between water oxygen and the aromatic atoms (see below). As a result of this observation, when we characterized the geometry of the two interacting groups, the data for these two structures were not included for the analysis. The parameter d in the remaining 19 cases varied from 3.13 to 3.49 Å. The angle (θ) involving O_w atom, aromatic center (AC), and one of the aromatic carbons (CA) was calculated for each of the 19 structures [Fig. 2(A)]. The distribution shows that this angle exhibits a preference to be in the range of 80–110°. Although our choice of parameter ‘ r ’ will certainly have an influence on the distribution of θ , it must be noted that a similar preference was observed for anion $\cdots\pi$ interactions in a number of small chemical compounds.²³ In our earlier studies on protein structures, the same parameter showed an identical trend in interactions between the $C=O$ functional groups and aromatic centers.²⁵ We have speculated the possible involvement of lone-pair electrons from the $C=O$ groups in interacting with the π electrons of the aromatic residues.

The atomic displacement parameters (B values) determined by high-resolution X-ray crystallographic studies represent smearing of atomic electron densities around their equilibrium positions due to thermal

Table I. Examples of $lp \cdots \pi$ Interactions Involving Water Oxygen Atoms and Aromatic Residues of High-Resolution Protein Structures

PDB ID ^a	Aromatic ^b	Water ^c	d (Å) ^d	r (Å) ^d	E^e	E^f
1A8I	W215	HOH42	3.24	0.99	-0.15	-1.83
1A8I	F749	HOH588	3.48	0.72	-0.63	-0.16
1AKO	H38	HOH83	3.34	0.11	-0.15	-1.40
1BD0/A	Y239	HOH68	3.40	0.25	-0.93	-0.67
1BFD	F31	HOH535	3.28	0.51	-0.63	-0.36
1C24/A	H236	HOH7	3.49	0.99	-1.61	-5.45
1CNZ/B	W205	HOH505	3.24	0.62	-1.71	-2.85
1GOF	W290	HOH193	3.34	0.5	-0.61	-0.19
1GOF	F523	HOH229	3.29	0.57	-0.91	0.01
1H2R/S	F236	HOH3232	2.76	0.72	2.89	2.25
1HXN	W354	HOH132	3.31	0.23	-0.75	-1.55
1NFN	W34	HOH201	3.33	0.77	-1.76	-1.00
1QH8/A	W292	HOH527	3.13	0.15	-1.35	-1.64
1QNF	W286	HOH583	2.47	0.72	6.57	8.54
1QUS/A	F260	HOH520	3.34	0.5	-1.12	-0.68
1TC1/B	Y33	HOH79	3.30	0.59	-2.12	-0.47
1TOA/A	H253	HOH576	3.38	0.87	-0.11	-0.42
2LIS/A	Y5	HOH2246	3.27	0.39	-1.67	-1.41
2LIS/A	Y117	HOH2079	3.41	0.1	-1.15	-0.33
2MYR	F469	HOH217	3.24	0.2	-0.25	-1.06
7ATJ/A	F68	HOH1040	3.45	0.74	-0.56	-0.89

^a Four-letter unique Protein Data Bank ID/Chain ID.

^b Aromatic residue in one letter code followed by the residue number.

^c Water number as given in the respective PDB structures.

^d For the definition of geometric parameters, see Figure 1(A).

^e BSSE corrected MP2/6-311++G(d,p) energy in kcal/mol; hydrogen atoms were constructed using InsightII.

^f BSSE corrected MP2/6-311++G(d,p) energy in kcal/mol; hydrogen atoms were built by GROMACS-OPLS.

motion and positional disorder.³² In other words, it gives an idea of degree of flexibility of any atom in a protein crystal structure. If the observed contact between water oxygen and the aromatic center is due to a favorable interaction, then it is likely to result in a small B -value for water oxygens. First, we extracted B -values of water oxygen and side-chain atoms of aromatic residues from all the PDB files and normalized. Average B -value for all water oxygens (excluding those that are involved in interactions with aromatic centers) were computed in each PDB file and compared with that of the specific water molecules in contact with the aromatic center. We also compared the B -values of aromatic residues involved in possible $lp \cdots \pi$ interactions with the average B -values of all aromatic residues from the protein structures under consideration. We have found in most of the cases that the normalized B -values of interacting water molecules are less compared to average B -values of all water molecules present in the same PDB files and the results are shown in Figure 2(B). Similarly, the average B -value of the interacting aromatic residues is slightly less than that of all aromatic residues. The difference between the average B values of interacting water molecules and all the water molecules was found to be extremely statistically significant in an unpaired t test (the two-tailed P value < 0.0001 ; <http://www.graphpad.com/quickcalcs/ttest1.cfm>). Thus, the lower B -values of interacting water molecules strongly indicate that they could be involved in possible stabilizing interactions.

This prompted us to characterize the nature of such contacts between water molecules and aromatic residues. Ab initio quantum chemical calculations were carried out on these systems as described in the “Materials and Methods” section. As mentioned earlier,

Figure 1. (A) Geometrical parameters defining water oxygen - aromatic interactions. d is the distance between the water oxygen (O_w) and the aromatic center (AC). r gives a measure of displacement of O_w from AC. θ is the angle between O_w, AC and one of the aromatic carbons (CA). (B) The parameter δ is the angle rotated about an axis that is perpendicular to the sixfold/fivefold rotational axis of the aromatic ring. At $\delta = 0^\circ$, this axis along with H—O—H bisector axis and the sixfold/fivefold rotational axis of aromatic ring will be perpendicular to each other. At $\delta = 90^\circ$, the bisector will be collinear to the sixfold/fivefold rotational axis.

Figure 2. (A) Distribution of angle ' θ ' for the 19 examples extracted from high-resolution protein structures in which water oxygen atom is likely to interact with the π -electron cloud of the aromatic ring. (B) B-factor analysis of aromatic residues and water molecules. Dark grey bar indicates the average B-factors of those aromatic residues or water molecules involved in $\text{lp}\cdots\pi$ interactions. Light grey bar represents the average B-factor values of all aromatic residues or all water molecules from the high-resolution protein structures.

for each of the 21 examples, three different methods were used to generate hydrogen atoms. However, we could not consider the coordinates generated using GROMACS-ffG43a1^{33,34} because it employs united atom approach (hydrogens bonded to carbon atoms

are not constructed). Hence, only the systems generated using the other two methods (InsightII and GROMACS-OPLS) were used for calculating the point energy using ab initio quantum chemical calculations at MP2/6-311++G(d,p) level of theory. The

Figure 3. Examples of $\text{lp}\cdots\pi$ interactions in high-resolution protein structures. The four-letter PDB code and the chain ID of each structure are shown on the upper right corner of each example. Each aromatic residue is identified with its one letter amino acid code and residue number. The participating water molecule is displayed with the number as given in the respective PDB file. The distance between the oxygen atom and the aromatic center is shown. The BSSE corrected energy along with the interplanar angle (δ) are also displayed. Among the InsightII-generated coordinates, (A) the least and (B) the most energetically favorable aromatic–water interactions are shown. (C) Water oxygen atom interacts strongly with the aromatic center of protonated imidazole. In this case, hydrogen coordinates were generated using GROMACS-OPLS. (D) An example in which the distance between the two water hydrogen atoms from the aromatic center differs by 0.88 Å (4.23 vs. 3.35 Å) is shown and the hydrogen coordinates in this case were generated using InsightII. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

coordinates generated by both the methods show similar results. Our QM calculations show that 18 of the 21 cases examined have negative interaction energies in both methods indicating the favorable nature of the contacts between water oxygen and aromatic center (Table I). In one example (1GOF: F523-HOH229), while InsightII coordinates showed a clearly favorable interaction (-0.91 kcal/mol), GROMACS-OPLS energy is close to 0.0 kcal/mol. Some of the examples showing favorable interaction energies are shown in Figure 3. In the case of InsightII-generated coordinates, the interaction energy values varied from -0.11 to -2.12 kcal/mol after BSSE correction (Table I) and the magnitude of nearly half of them is close to or exceeds 1.0 kcal/mol. Examples of most and least favorable interactions are shown in Figures 3(A,B), respectively. The range of interaction energies among the favorable cases in GROMACS-OPLS systems varied from -0.16 to -5.45 kcal/mol and more than half of them have magnitude equal to or greater than 1.0 kcal/mol.

In two cases (1H2R and 1QNF) in which $d < 3.0$ Å the observed contacts result in a high positive energy indicating that they are not favorable. These water molecules are also involved in stronger interactions with side-chain of a lysine residue (1H2R) or other water molecules (1QNF). These attractive interactions are likely to offset the unfavorable contacts between the water molecules and the aromatic centers. It is also interesting to note that the B factors of these two water oxygen atoms are higher compared to the respective average B-factor values calculated for all the water molecules observed in the corresponding protein structures.

When interaction energies of both Insight-II and GROMACS-OPLS generated coordinates were considered, the most favorable energy was found to occur in 1C24 between a histidine residue and a water molecule and the hydrogen atom coordinates in this case were generated using GROMACS-OPLS [Fig. 3(C)]. The energy -5.45 kcal/mol is equivalent to or more favorable³⁵ than a conventional hydrogen bond. This is due to the fact that a protonated histidine was generated when GROMACS-OPLS was used to construct hydrogen atoms (Figure S3; See Supporting Info.). It has been reported earlier that when protonation occurs, the imidazole ring attracts water oxygen atom more than its hydrogen atoms.¹²

Although $O_w \cdots$ aromatic contacts are in general favorable, the energy difference between the least and most favorable interactions is more than 2.0 kcal/mol (even if we do not take the water-protonated imidazole interaction into account and excluding the two cases in which the distance $d < 3.0$ Å; Table I). The interaction energies determined using the coordinates generated by two different methods are not identical and in a few cases, the energy difference is more than 1.0 kcal/mol (Table I). The variation in the interaction energies could have come from the differences in the geometry of

Figure 4. Snapshots during the sampling of water orientations with respect to the aromatic plane. The interplanar angle δ , is changed in steps of 10° . δ is the angle between the aromatic plane and the plane formed by the three atoms of the water molecule [Fig. 1(B)]. In this rotation, both water hydrogen atoms are equally displaced away from the aromatic ring. Four orientations representing four different δ values are shown for a fixed d value of 3.4 Å. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

water–aromatic compounds. The optimal geometry for maximum favorable interaction depends on several parameters. This includes the distance of water oxygen from the aromatic center, the amount of displacement of water oxygen from the aromatic center and the proximity of water hydrogens to the aromatic atoms (defined by the interplanar angle between the plane formed by the aromatic ring and the plane formed by the three water atoms). The interaction energy also depends upon the nature of substituents in the aromatic ring. To answer some of the questions, we have performed quantum chemical calculations varying some of the geometrical parameters on a set of model compounds.

Interaction energies were calculated on model systems by varying two parameters the interplanar angle (δ) and the distance ' d ' between water oxygen and aromatic center (see Materials and Methods). The selected model systems mimic the aromatic moiety of four aromatic amino acids. The generated structures for each model compound differed in their interplanar angles and/or the distances d . Example of varying interplanar angle δ for a fixed distance d is shown for benzene–water system for four different δ values (see Fig. 4). The intermolecular potential surfaces thus generated for 510 different water orientations for each model system are shown in Figure 5. The results of all model compounds are very similar except that of protonated imidazole ring. Hence, the nature of interactions in water-protonated imidazole ring will be discussed later. In all other cases, the most favorable interaction energy occurs when d is between 3.2 and 3.4 Å and the interplanar angle is

Figure 5. Contour diagrams showing the intermolecular potential surface for the systems: (A) Benzene–water, (B) Imidazole–water, (C) Indole 6-membered ring–water, (D) Indole 5-membered ring–water, (E) Phenol–water, and (F) Protonated imidazole–water. Intermolecular potential surfaces are color coded; the colors red, orange, yellow, cyan, dark blue, dark brown indicate increasing level of favorable nature of oxygen···aromatic interactions with red and dark brown at the two extremes. Grey regions indicate unfavorable interactions.

o°. Such orientation places the two water hydrogen atoms closer to the atoms in the aromatic ring and the interaction energy in this orientation varies from -1.44 to -1.7 kcal/mol (Table S1; Supp. Info.). Hence, the favorable nature of interaction could be considered partly due to the electrostatic interaction between the electron-deficient water hydrogen atom and the π -electron cloud of aromatic ring. This is similar to the O–H··· π interaction, although in this case, the hydrogen atoms are not directly pointing toward the aromatic center. In general, the δ value of 45° – 60° is not a preferred orientation when water oxygen points toward the aromatic ring and in the case of benzene, this region is highly unfavorable for such interactions. The interplanar angle 90° is an ideal ori-

entation for the lone-pair··· π interaction between water oxygen and the aromatic ring. In this orientation, water hydrogens will be farthest from any of the aromatic atoms. However at 90° , our calculations show that the interaction energies are still favorable in benzene, imidazole, phenol, and indole (6-membered ring) systems. At 90° , the most favorable energies vary between -0.17 and -0.54 kcal/mol in these systems (Table S1; Supp. Info.) and the optimal distance in this orientation is in the range of 3.2 – 3.4 Å. In the case of indole 5-membered ring, the 90° orientation is much more favorable (-0.83 kcal/mol) with the optimum distance of 3.2 Å. Among the five systems studied, the lp··· π interaction is the strongest when the pyrrole ring of indole is involved.

In all the five systems, the orientation that gives rise to the most favorable interaction energy is 0° . Our calculations clearly show that the strength and nature of water–aromatic interactions are determined by the positions of water hydrogen atoms with respect to the aromatic ring. This is evident when the distance d is fixed and only the parameter δ is varied. In the high-resolution protein structures, although the water oxygens are close to the aromatic ring, the strength of interactions will be known only when the positions of the water hydrogen atoms are unambiguously determined. The intermolecular potential surface reveals that the interaction is attractive even when the water hydrogens are located away from the aromatic ring at $\delta = 90^\circ$. It must be noted that the more favorable nature of interactions observed at 0° is not clearly due to the conventional $\text{O}—\text{H}\cdots\pi$ interactions as described in the literature.^{24,36,37}

In the case of protonated imidazole ring, a dramatically different picture is observed. Not only the most favorable interaction occurs at 90° with the interaction energy of -9.25 kcal/mol but also it is several fold more stable than that observed for other model compounds for any orientation. The optimal distance of 2.9 Å is also 0.3 – 0.5 Å closer to the aromatic ring compared to the other aromatic compounds. In general, both the parameters d and δ show wide regions in which water oxygen having favorable interactions with the imidazole ring. Previous reports have suggested that the aromatic moiety in the protonated form of imidazole ring is electron deficient compared to the neutral form of imidazole ring.¹² This explains the reason why the histidine observed in the PDB structure (PDB ID: 1C24), when protonated, resulted in a highly favorable interaction with water molecule [Fig. 3(C)].

Discussion

High-resolution crystal structures from PDB were searched for possible lone-pair $\cdots\pi$ interactions involving water oxygen atoms and aromatic rings of four aromatic residues. In 427 examples, water oxygen atoms were within 3.5 Å from the centre of aromatic rings. As the positions of hydrogen atoms have to be clearly defined to characterize the nature of such contacts, we used three different methods to construct hydrogen atoms. A series of stringent steps were then followed to discard those cases which are either ambiguous (occupancy, hydrogen bonds with aromatic atoms) or the contacts are due to the result of $\text{O}—\text{H}\cdots\pi$ interactions. Each method, after building the hydrogen atoms and applying stringent filtering criteria, identified more than 100 cases out of 427 examples as likely candidates for lp $\cdots\pi$ interactions. However, when we adopted a consensus approach, we could finally find only 21 examples common to all three methods that could be investigated for lone-pair $\cdots\pi$ interactions. Analysis of the geometrical parameters indicates that

the water–aromatic compounds have similar features observed in anion $\cdots\pi$ or $\text{C}=\text{O}\cdots\pi$ interactions.^{23,25} The smaller B -values for these water molecules provided a hint that they could be involved in stabilizing interactions. Quantum mechanical calculations were carried out on two sets of coordinates that differed only in hydrogen atom positions; in one case, hydrogens were generated using InsightII and in the second case, they were built with GROMACS-OPLS. In both sets, in at least 18 of 21 cases, QM calculations showed that the contacts are indeed the result of favorable interactions. Closer examination of the interaction energy values revealed that the results are not identical in both the sets. For example, in the case of 1TCI, the interaction energy is -2.12 kcal/mol from the InsightII generated coordinates, whereas the OPLS-generated system gave rise to only -0.47 kcal/mol. The energy difference of 1.65 kcal/mol in this case could be clearly attributed to the different hydrogen atom positions generated by InsightII and OPLS methods. The other example is 1C24 in which the energy difference of 3.84 kcal/mol is due to protonated and neutral imidazole rings. The methods used to construct hydrogen atom positions in InsightII and GROMACS-OPLS are different and hence the interaction energies heavily depend on the generated hydrogen atom positions.

To further explore this point, we considered the simpler prototype systems: 1:1 dimers of water and aromatic moieties that mimic the aromatic amino acids. Quantum chemical calculations were carried out by varying two geometrical parameters, distance d and interplanar angle δ . By varying d , we investigated the influence of π -electron cloud as a function of distance. The δ value determines the influence of hydrogen atom positions. The interplanar angle was varied in such a way that both hydrogens were equally displaced from the aromatic centre. Our results clearly show that when water hydrogens are closer to the aromatic atoms at $\delta = 0^\circ$, the interaction energy is the most favorable for all model systems with the exception of the protonated imidazole ring. At $\delta = 90^\circ$, the hydrogen atoms are farthest from the aromatic center and oxygen lone-pair electrons will be directly pointing towards the aromatic ring. In such cases, the interaction energy is still favorable in these systems but the magnitude of the energy is less than 1.0 kcal/mol. The only exception again is protonated imidazole. In the case of protonated imidazole, the interaction energy at $\delta = 90^\circ$ is close to -9.25 kcal/mol making it one of the strongest noncovalent interactions. In summary, when water interacts with aromatic molecules with its oxygen pointing toward the aromatic center, then the strength of the interaction is determined by the positions of the hydrogen atoms and the substituents of the aromatic ring.

The complete sampling of water–aromatic interactions involves several parameters with many degrees of freedom. In the present study, the parameter δ was varied by rotating the axis that is perpendicular both to

the bisector of H—O—H angle (at $\delta = 0^\circ$) and the C6/C5 rotational axis of aromatic ring. In this operation, both the water hydrogens are equally displaced away from the aromatic center. In reality, this need not be the case [Fig. 3(D)] and in such situation one water hydrogen will be closer to the aromatic ring than the other. Thus, one can also rotate the water molecule about an axis that could make one of the water hydrogens closer to the aromatic atoms. The parameter r which determines the displacement of water oxygen from the aromatic center is also likely to influence the interaction and could be varied. As the number of possible ways a water molecule can be oriented with respect to its interacting aromatic ring is enormous, exhaustive sampling of water orientations will be computationally prohibitive. In this study, we have carried out limited sampling to demonstrate the influence of hydrogen atom orientations on the strength of water–aromatic interactions that cannot be described as O—H $\cdots\pi$ interactions. In all the aforementioned cases, the role of lone-pair electrons in oxygen atom could be partial ($\delta = 0^\circ$) or it could completely dominate ($\delta = 90^\circ$).

In systems like indole and imidazole, even at $\delta = 90^\circ$, the orientation of water hydrogens with respect to the aromatic substituents (the nitrogen in the 5-member ring of indole or the two imidazole nitrogens) could be significant. To investigate this point, we carried out quantum chemical calculations on indole–water system by placing the water molecule above the five-member pyrrole ring and rotating it around the C5 axis in steps of 20° with $d = 3.2 \text{ \AA}$ which is the optimal distance found for $\delta = 90^\circ$. The same level of theory was used in these molecular orbital calculations and they showed that the interaction energy varied from -0.78 to -1.27 kcal/mol (Figure S1, see Supp. Info.). The minimum distance between water hydrogens and the indole nitrogen was calculated and plotted as a function of interaction energy (Figure S2, see Supp. Info.). The regression coefficient for this plot is 0.99 indicating a strong correlation between water hydrogen–indole nitrogen distance and the interaction energy. When the water hydrogen is closer to the indole nitrogen, the interaction between water and the pyrrole ring becomes weaker. Even at $\delta = 90^\circ$, when the water hydrogen atoms are farthest from the aromatic ring, the positions of hydrogen atoms with respect to the aromatic substituents could still influence the strength of interactions.

Most of the previous studies on model compounds to investigate the lp $\cdots\pi$ interactions have focused on aromatic molecules with electron withdrawing groups.^{14,24} Studies of lp $\cdots\pi$ interactions and their importance in biomolecular structures are few and are reported recently. The abilities of aromatic groups of the amino acids Phe, Tyr, Trp, and His to participate in various interactions with water molecules were compared by Scheiner *et al.*¹² They found that the protonated imidazole group attracts the oxygen atom of the

water molecule, a conclusion supported by the present study. The reported interaction energy is -8.1 kcal/mol , and this is much more favorable than the hydrogen bonds formed between water oxygen with the aromatic ring atoms or the π electron. Egli and Sarkhel¹⁷ have identified lp $\cdots\pi$ interactions in some nucleic acid and protein structures. Their subsequent ab initio calculations showed that uracil and protonated cytosine are most likely to participate in lp $\cdots\pi$ interactions in nucleic acid structures. Their study concluded that this interaction can significantly contribute to the stability of the structure, especially when the nucleobase is positively polarized due to chemical environment or when it involves protonated cytosine. Recently, we have analyzed high-resolution protein structures and found several examples in which carbonyl oxygens approach the aromatic centers within $3.5 \text{ \AA}$.²⁵ Our ab initio calculations showed that such an approach is energetically favorable and the resulting interactions could involve lone-pair electrons of oxygen atoms.

The nature of lp $\cdots\pi$ interactions in general is weak and attractive. Recently, Gung *et al.*²⁸ used low-temperature NMR spectroscopy, X-ray analysis and quantum chemical calculations on a series of triptycene-based model compounds and have quantified the lp $\cdots\pi$ interactions. On the basis of the molecular orbital calculations at the Hatree-Fock (HF) level and at the MP2/aug-cc-pVTZ level of theory, they concluded that the attractive nature of lp $\cdots\pi$ interaction observed in electron-rich aromatic rings must be due to dispersion forces. The magnitude of lp $\cdots\pi$ interactions for the triptycene compounds with electron-donating groups is similar to what we have observed in the present study for aromatic amino acids at $\delta = 90^\circ$. Further characterization in our study using molecular orbital calculations indicate that the model compounds with the exception of protonated imidazole do not have energy minimum at the HF level and they show repulsive interaction (2.20 – 2.83 kcal/mol). The difference between calculations at the HF level and MP2 level is correlation energy, which is mainly made up of the dispersion energy. Hence in this case also, the attractive nature of lp $\cdots\pi$ interactions observed between the aromatic amino acids and water oxygen atoms must have come from the dispersion forces. Other studies on anion $\cdots\pi$ interactions have reached similar conclusions.^{38,39} In the present study, we have also demonstrated the importance of water hydrogen atoms and their positioning with respect to the aromatic substituents.

Materials and Methods

Database survey

Five hundred high-resolution protein structures (resolution $\leq 1.8 \text{ \AA}$) (Ref. 40; <http://kinemage.biochem.duke.edu/databases/top500.php>) were searched to find out close contacts between O_w and the aromatic

centers (AC) of four aromatic residues. All O_w lying within a distance (d) of 3.5 Å from five possible aromatic centers (histidine, tyrosine, phenylalanine and both rings of tryptophans) were retrieved [Fig. 1(A)]. If the occupancy value of any of the atoms is not one, then those cases were discarded. In the case of histidine and tryptophan, further filtering was achieved by excluding the examples that participate in hydrogen bond through their imidazole or indole nitrogen donors with O_w as acceptor. The program HBPLUS⁴¹ was used for this purpose to identify potential hydrogen bonds based on conventional H-bond geometric criteria.⁴² Water hydrogen atoms are not observed in the crystal structures and the positions of these atoms are important to characterize the nature of water–aromatic interactions. Several methods are available for predicting hydrogen positions and they have been compared.⁴³ Here, we used three methods to build hydrogen atoms to understand the water–aromatic interactions. Homology module of InsightII (Accelrys, San Diego, CA) and *pdb2gmx* tool in GROMACS (version 3.2.1)⁴⁴ with two different force-fields ffG43a1^{33,34} and OPLS-AA⁴⁵ were used to construct hydrogen atom positions. The coordinates generated by these methods will be referred respectively by the names of these methods. If any of the three methods gives rise to O_w –H $\cdots\pi$ interactions (when the distance between water hydrogen and aromatic center is less than that of O_w and AC, then this contact could be considered as O_w –H $\cdots\pi$ interaction), then those cases were not considered further. If all three methods result in positions of water hydrogens away from aromatic center (thus keeping O_w closer to aromatic center), then those examples were analyzed for the possible lone-pair $\cdots\pi$ interactions.

In the candidate examples chosen to study the lone-pair $\cdots\pi$ interactions, O_w atoms were projected on to the aromatic plane and its distance (r) from aromatic center was calculated [Fig. 1(A)]. Only those interactions, in which $r \leq 1.0$ Å, were ultimately considered. This final elimination ensures that the oxygen atom of water molecule is proximal to the center of the aromatic plane and precludes its interaction with the periphery of the ring. As a favorable interaction is likely to exhibit reduced B values (atomic displacement parameters) of respective groups [See Results and Fig. 2(B)], we have analyzed the normalized B values of both water molecules and the aromatic groups. Characterization of interactions using QM calculations was carried out to investigate the nature of contacts between O_w and aromatic centers.

Quantum mechanical calculations

For each candidate example selected to characterize the possible lone-pair $\cdots\pi$ interaction, point energy was calculated at MP2/6-311++G(d,p) level of theory using Gaussian 03 suite of program.⁴⁶ The input geometry consisted of O_w and the corresponding aromatic amino

acid after the C_β carbon with the heavy atom coordinates taken from the PDB data. All hydrogen atom coordinates were built from either InsightII or GROMACS-OPLS.

Furthermore, ab initio quantum chemical calculations were also carried out to investigate the favorable nature of this interaction in simpler prototype systems: 1:1 dimers of water and aromatic substrates like phenol, indole, benzene and imidazole mimicking tyrosine, tryptophan, phenylalanine and histidine, respectively.¹² In the case of imidazole, both protonated and neutral forms were considered. The orientation of water molecule and its hydrogen atoms with respect to the aromatic plane are likely to influence the nature of water–aromatic interactions. To explore this point, exhaustive sampling of water orientations with respect to the aromatic ring must be considered. In this study, for each model compound, sampling of water molecule was done by rotating the water molecule about an axis that is perpendicular to the sixfold/fivefold symmetry axis of the aromatic ring [Fig. 1(B)]. This angle is denoted by the symbol ' δ ' and is defined in such a way that when $\delta = 0^\circ$, the axis by which water is rotated, the bisector of the H–O–H angle and the rotational axis of the aromatic ring are all perpendicular to each other. This interplanar angle between the aromatic plane and the plane formed by the three water atoms was varied from 0 to 90° with 10° increments. At $\delta = 90^\circ$, the water bisector axis will coincide with the sixfold/fivefold rotational axis of the aromatic ring. In addition to the interplanar angle, the distance d between O_w atom and the aromatic center was also varied from 2.0 to 7.0 Å in steps of 0.1 Å. Each of these monomers were individually optimized at the MP2/6-311++G(d,p) level of theory. Single point energy was calculated at the same level of theory for each of the generated complex and a total of 510 water orientations were considered for each model compound. The interaction energy for all the systems studied is subjected to basis set superposition error (BSSE) correction by using the Boys-Bernardi counterpoise method.⁴⁷

Conclusions

Nonbonded interactions are important for a protein's structure, stability and function. Lone-pair $\cdots\pi$ interactions is the least studied one in biomolecular structures and in this study, we have attempted to identify and characterize lp $\cdots\pi$ interactions involving water oxygen and aromatic residues. In a database of high-resolution protein structures, we have found a large number of examples in which water oxygen atom is in close contact with the aromatic center of aromatic residues. It will be possible to describe the nature of interactions due to such close contacts only if the positions of the water hydrogen atoms are known. In this study, using three different methods to construct hydrogen atom positions and by consensus approach,

we have identified nearly 20 examples that can be described as lp $\cdots\pi$ interactions involving water oxygen atoms and aromatic residues. Quantum chemical calculations on compounds based on protein structures indicate that the lone-pair electrons of water oxygen atom can interact favorably with the π -electron cloud of the aromatic residues. However, the strength of this interaction depends upon the distance of water hydrogen atoms with respect to the different substituents of aromatic residues. This is evident from the interaction energy values obtained for systems in which the only difference between the two systems is due to the hydrogen atom positions constructed using two different methods. Intermolecular potential surface of model compounds obtained from the limited sampling of water orientations confirmed the role of hydrogen atoms in influencing the oxygen-aromatic interactions.

Acknowledgments

AJ and VR thank CSIR, India and IIT-Kanpur for a research fellowship. Part of the calculations in this study was performed in the Computer Center, IIT-Kanpur. R.S. is a Joy Gill Chair Professor in the Department of Biological Sciences and Bioengineering, IIT-Kanpur. The authors thank Mr. Brajesh Mishra and Mr. Tuhin Kumar Pal for helpful discussions. Mr. Mainpal Rana is acknowledged for his help in the preparation of one of the figures.

References

- Li Z, Lazaridis T (2007) Water at biomolecular binding interfaces. *Phys Chem Chem Phys* 9:573–581.
- Steinbach PJ, Brooks BR (1993) Protein hydration elucidated by molecular dynamics simulation. *Proc Natl Acad Sci USA* 90:9135–9139.
- Loris R, Langhorst U, Vos SD, Decanniere K, Bouckaert J, Maes D, Transue TR, Steyaert J (1999) Conserved water molecules in a large family of microbial ribonucleases. *Proteins: Struct Funct Genet* 36:117–134.
- Scheidig AJ, Burmester C, Goody RS (1999) The pre-hydrolysis state of p21_{ras} in complex with GTP: new insights into the role of water molecules in the GTP hydrolysis reaction of ras-like proteins. *Structure* 7:1311–1324.
- Tarek M, Tobias DJ (1999) Environmental dependence of the dynamics of protein hydration water. *J Am Chem Soc* 121:9740–9741.
- Ben-Naim A (2002) Molecular recognition—viewed through the eyes of the solvent. *Biophys Chem* 101–102:309–319.
- Cheung MS, Garcia AE, Onuchic JN (2002) Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc Natl Acad Sci USA* 99:685–690.
- Dunn RV, Daniel RM (2004) The use of gas-phase substrates to study enzyme catalysis at low hydration. *Philos Trans R Soc Lond B* 359:1309–1320.
- Halle B (2004) Protein hydration dynamics in solution: a critical survey. *Philos Trans R Soc Lond B* 359:1207–1224.

- Smolin N, Oleinikova A, Brovchenko I, Geiger A, Winter R (2005) Properties of spanning water networks at protein surfaces. *J Phys Chem B* 109:10995–11005.
- Chaplin M (2006) Do we underestimate the importance of water in cell biology? *Nature Rev Mol Cell Biol* 7:861–866.
- Scheiner S, Kar T, Pattanayak J (2002) Comparison of various types of hydrogen bonds involving aromatic amino acids. *J Am Chem Soc* 124:13257–13264.
- Steiner T (2002) Hydrogen bonds from water molecules to aromatic acceptors in very high-resolution protein crystal structures. *Biophys Chem* 95:195–201.
- Gallivan JP, Dougherty DA (1999) Can lone pairs bind to a π system? The water hexafluorobenzene interaction. *Org Lett* 1:103–106.
- Sarkhel S, Rich A, Egli M (2003) Water-nucleobase “stacking”: H- π and lone pair- π interactions in the atomic resolution crystal structure of an RNA pseudoknot. *J Am Chem Soc* 125:8998–8999.
- Stollar EJ, Gelpí JL, Velankar S, Golovin A, Orozco M, Luisi BF (2004) Unconventional interactions between water and heterocyclic nitrogens in protein structures. *Proteins: Struct Funct Bioinf* 57:1–8.
- Egli M, Sarkhel S (2007) Lone pair-aromatic interactions: To stabilize or not to stabilize. *Acc Chem Res* 40:197–205.
- Mooibroek TJ, Gamez P, Reedijk J (2008) Lone pair- π interactions: a new supramolecular bond? *Cryst Eng Comm* 10:1501–1515.
- Schneider H-J, Werner F, Blatter T (1993) Attractive interactions between negative charges and polarizable aryl parts of host-guest systems. *J Phys Org Chem* 6:590–594.
- Alkorta I, Rozas I, Elguero J (1997) An attractive interaction between the π -cloud of C₆F₆ and electron-donor atoms. *J Org Chem* 62:4687–4691.
- Danten Y, Tassaing T, Besnard M (1999) On the nature of the water-hexafluorobenzene interaction. *J Phys Chem A* 103:3530–3534.
- Mascal M, Armstrong A, Bartberger MD (2002) Anion-aromatic bonding: a case for anion recognition by π -acidic rings. *J Am Chem Soc* 124:6274–6276.
- Quiñonero D, Garau C, Rotger C, Frontera A, Ballester P, Costa A, Deyà PM (2002) Anion- π interactions: do they exist? *Angew Chem Int Ed* 41:3389–3392.
- Raimondi M, Calderoni G, Famulari A, Raimondi L, Cozzi F (2003) The benzene/water/hexafluorobenzene complex: a computational study. *J Phys Chem A* 107:772–774.
- Jain A, Purohit CS, Verma S, Sankaramakrishnan R (2007) Close contacts between carbonyl oxygen atoms and aromatic centers in protein structures: $\pi\cdots\pi$ or lone-pair- π interactions? *J Phys Chem B* 111:8680–8683.
- Quinonero D, Frontera A, Escudero D, Ballester P, Costa A, Deyà PM (2007) A theoretical study of anion- π interactions in seven-membered rings. *Chemphyschem* 8:1182–1187.
- Ringer AL, Senenko A, Sherrill CD (2007) Models of S/ π interactions in protein structures: comparison of the H₂S benzene complex with PDB data. *Protein Sci* 16:2216–2223.
- Gung BW, Zou Y, Xu Z, Amicangelo JC, Irwin DG, Ma S, Zhou HC (2008) Quantitative study of interactions between oxygen lone pair and aromatic rings: substituent effect and the importance of closeness of contact. *J Org Chem* 73:689–693.
- Imai YN, Inoue Y, Nakanishi I, Kitaura K (2008) Cl- π interactions in protein-ligand complexes. *Protein Sci* 17:1129–1137.
- Reyes A, Fomina L, Rumsh L, Fomine S (2005) Are water-aromatic complexes always stabilized due to π -H

- interactions? LMP2 study. *Int J Quantum Chem* 104: 335–341.
31. DiLabio GA, Johnson ER (2007) Lone pair- π and $\pi\cdot\pi$ interactions play an important role in proton-coupled electron transfer reactions. *J Am Chem Soc* 129: 6199–6203.
 32. Parthasarathy S, Murthy MRN (2000) Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Eng* 13:9–13.
 33. van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Kruger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag AG an der ETH. Zurich, Switzerland. pp 1–1042.
 34. Daura X, Mark AE, van Gunsteren WF (1998) Parametrization of aliphatic CH_n united atoms of GROMOS96 force field. *J Comput Chem* 19:535–547.
 35. Weiss MS, Brandl M, Sühnel J, Pal D, Hilgenfeld R (2001) More hydrogen bonds for the (structural) biologist. *Trends Biochem Sci* 26:521–523.
 36. Feller D (1999) Strength of the benzene-water hydrogen bond. *J Phys Chem A* 103:7558–7561.
 37. Mishra BK, Sathyamurthy N (2007) Van der waals complexes of small molecules with benzenoid rings: influence of multipole moments on their mutual orientation. *J Phys Chem A* 111:2139–2147.
 38. Garau C, Frontera A, Quinonero D, Ballester P, Costa A, Deya PM (2004) Cation- π versus anion- π interactions: energetic, charge transfer, and aromatic aspects. *J Phys Chem A* 108:9423–9427.
 39. Kim D, Tarakeshwar P, Kim KS (2004) Theoretical investigations of anion- π interactions: the role of anions and the nature of π systems. *J Phys Chem A* 108: 1250–1258.
 40. Lovell SC, Davis IW, Arendall III WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by $\text{C}\alpha$ geometry: ϕ , ψ and C_β deviation. *Proteins: Struct Funct Genet* 50:437–450.
 41. McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol* 238:777–793.
 42. Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* 44:97–179.
 43. Forrest LR, Honig B (2005) An assessment of the accuracy of methods for predicting hydrogen positions in protein structures. *Proteins: Struct Funct Bioinf* 61: 296–309.
 44. Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 7:306–317.
 45. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236.
 46. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JR, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian 03, revision B.05; Gaussian, Inc., Pittsburgh PA.
 47. Boys SF, Bernardi F (1970) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol Phys* 19:553–566.