# PCCP

## Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

# Journal Name

## COMMUNICATION

# What are preferred water/aromatic interactions in proteins and crystal structures of small molecules?

Goran V. Janjić,[a] Saša N. Malkov,[b] Miodrag V. Živković[b] and Snežana D. Zarić*[c,d]

**The distribution of water molecules around aromatic ring in the protein structures and crystal structures of small molecules show quite small number of the strongest OH/π interactions, larger number of parallel interactions, and the largest number of the weakest CH/O interactions.**

The understanding of the nature of the interaction between aromatic rings and water molecule is of great importance since water is in contact with aromatic molecules in many molecular systems, from materials to biological molecules.[1-5] Some of the systems where interactions of aromatic rings with water play an important role are aquaporins (water-transporting proteins),[1] nanotubes,[2] and nanoporous materials.[3] Experimental observation showed approximately one OH/π interaction between liquid water and each dissolved benzene molecule, indicating importance of water/aromatic interactions at hydrated biological interfaces.[6]

A substantial number of theoretical investigations have focused on characterizing the interaction between benzene and water as a prototype for aromatic–polar interactions.[5-9] The OH/π[7] and CH/O[8] interactions between aromatic ring and water molecule are well known interactions. The energies of these interactions are $\Delta E_{CCSD(T)(limit)}$=-3.19 kcal/mol and $\Delta E_{CCSD(T)(limit)}$= -1.41 kcal/mol, respectively.[9a] By analyzing data in crystal structures from the Cambridge Structural Database (CSD) parallel alignment interaction between water molecule and aromatic ring was recognized.[9] In parallel alignment interactions either the whole water molecule or one of its O–H bonds lies parallel to the aromatic ring. High level *ab initio* calculations reveal that the most stable conformation of parallel alignment interactions,[9b] with one O-H bond parallel to the plane of benzene ring at large horizontal displacements, out of the aromatic ring and out of the C–H bond region, can be significantly strong; $\Delta E_{CCSD(T)(limit)}$= -2.45 kcal/mol. Lone-pair/π interactions, involving water oxygen and aromatic

residues, is the least studied one, mostly by analyzing crystal structures of proteins and nucleic acids.[10,11] The interactions between water and benzene molecule were found to be attractive if the inter-planar angle was in range 0–45°,[10] due to interactions of water OH bonds with benzene π-system.[9a] The true lone-pair/π interaction, correspond to orientation with inter-planar angle of 90°, has repulsive nature (+0.55 kcal/mol, calculated at MP2/cc-pVQZ level.[8b]

Studies on interactions between water molecule and aromatic rings in proteins[12-14] have shown that water molecules within 7.0 Å of the phenylalanine ring have a tendency to be around the edges of the aromatic ring, while the interactions with the faces of aromatic rings do not occur frequently[13] and that the interactions of water molecule with neighboring polar atoms are more favorable.[14]

Here we present results on the interactions of water molecule with phenyl ring from phenylalanine (Phe) in protein structures from the Protein Data Bank (PDB) and also on the interactions of water molecule with phenyl ring in the crystal structures from the Cambridge Structural Database (CSD). High level *ab initio* calculations were used to calculate energies of some water/benzene interactions. To the best of our knowledge, this is the first comparative study on the distribution of water molecules around phenyl ring in the proteins and in the crystal structures of small molecules.

The statistical study is based on the structures archived in the PDB and the CSD. In the structures of the proteins positions of the water hydrogen atoms are not determined, and it is not possible to determine orientations of water molecules. Positions of the hydrogen atoms are determined in crystal structures from the CSD, hence, it is possible to determine orientations of water molecules. Comparison of results from the PDB and from the CSD structures enables elucidation of the water molecule orientations in proteins.

Structures of proteins from the PDB, using PDBselect[15] list of non-redundant protein chains, with the threshold 25%, were screened for intermolecular contacts of non-coordinated water molecules and phenyl rings of phenylalanine. The search yielded 5462 contacts in which water oxygen was found within the area corresponding to the ellipsoid around aromatic ring with the vertical axis of 4.0 Å and horizontal axis of 5.5 Å. The ellipsoid was chosen since the significant number of the water molecules around phenyl ring form ellipsoid (Fig. S1, ESI†) and because the contacts outside the ellipsoid mainly correspond to the interactions of water molecule with the protein backbone (Fig. S2, ESI†). The average number of water molecules obtained in each such ellipsoid is 1.3 (4152 Phe residues interact with 5462 water molecules). The contour density plot of the normal distance ($R_O$, Fig. 1) for different offset values of water oxygen ($r_O$, Fig. 1) is shown in Fig. 2.
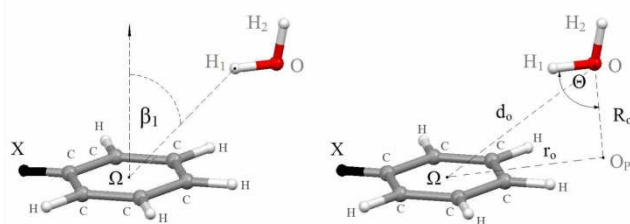


**Fig. 1** Geometrical parameters defining water–phenyl interactions: $\beta_1$ is angle between the $\Omega$··$H_1$ and the normal to the ring plane. $d_O$ is the distance between the water oxygen (O) and the aromatic center ($\Omega$). $r_O$ is displacement of Op (projection of the water oxygen to the average plane of phenyl ring) from $\Omega$. $R_O$ is the normal distance from the ring plane to the water oxygen. The parameter $\Theta$ is Op-O-$H_1$ angle ($H_1$ is H atom of the water molecule, closer to the center of benzene ring).
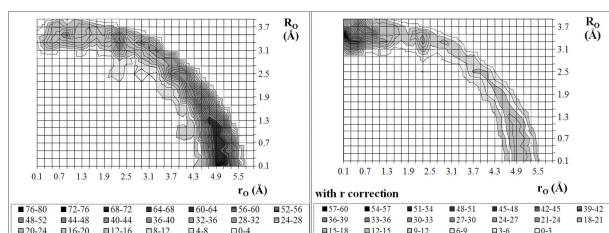


**Fig. 2** Uncorrected and corrected[16] density plot of the normal distance ($R_O$) versus the offset values of water oxygen ($r_O$) for protein structures in which water oxygen was found in the observed ellipsoid.

The density reaches a maximum in the phenyl ring plane with offset values of 4.8-5.2 Å. The data show that the water molecule has a tendency to be positioned outside of the ring, close to mean plane of the ring. In this area water molecule and phenyl ring form the CH/O[8] and the parallel alignment interactions.[9] The area corrected[16] (ESI†) density plot (Fig. 2) shows that water oxygen prefers to be at or near the centre of phenyl ring. In this area water molecule forms the OH/$\pi$ interaction, the strongest interaction between water and aromatic ring.[7] However, the corrected density plot also show high preference for water to be in the plane or close to the plane of the ring at offsets of 4.8-5.2 Å, although the CH/O and parallel interactions that are in this area are not very strong. This is in accord with previous finding that water molecules

have a tendency to be around the edges of the aromatic ring.[13] The computed solvent accessible surface area (SASA) values (Fig. S4, ESI†) show that phenylalanine residues generally tend to be away from the protein surface; SASA values are less than 10 Å². Combining the results of the crystallographic analysis with SASA values, it has been shown that the water molecules mostly interact with the phenyl group with slightly higher surface area (the maximum of distribution appears in the range of from 20 to 40 Å²), and that there are no significant correlations between water molecules positions and the surface area (Fig. S4 and S5, ESI†).

Water-aromatic interactions for tyrosine residue show similar distribution of water molecules in ellipsoid around $C_6$-aromatic ring (Fig. S7, ESI†). The average number of water molecules obtained in this ellipsoid is somewhat larger (1.4 water molecules per Tyr residue) than in case of phenylalanine residues, probable as a consequence of hydrogen bonding capability of O-H group of tyrosine residue. In the case of interactions with $C_6$-aromatic ring of tryptophan (Fig. S7, ESI†), in which $C_6$-aromatic ring is fused with the pyrrole ring, there is a more pronounced tendency towards orientation with water in the plane of aromatic ring. The average number of water molecules obtained in the ellipsoid is somewhat smaller (1.2 water molecules per Trp residue) than in case of phenylalanine.

Crystal structures in the CSD[17] were screened for intermolecular contacts of non-coordinated water molecules and phenyl rings. The CSD search program ConQuest1.14[18] was used to retrieve structures satisfying the criteria described in the ESI. Similar to water-phenyl contacts in the protein structures, the frequently populated area of water-phenyl contacts in crystal structures from the CSD corresponds to the ellipsoid with the vertical axis of 4.0 Å and horizontal axis of 5.5 Å (Fig. S8, ESI†), similar to water-phenyl contacts in proteins (Fig. S1, ESI†). In the ellipsoid we found 5223 contacts.

The density plot and area corrected[16] plot of the contacts from the CSD (Fig. 3) show very similar distribution as contacts in proteins (Fig. 2). As was mentioned above, differently than in protein structures from the PDB, in the crystal structures from the CSD position of the hydrogen atoms are available. It enables to find the orientations of the water molecules and number of OH/$\pi$, parallel, CH/O, and lone-pair/$\pi$ water/aromatic interactions. We found the largest number of CH/O and the smallest number of OH/$\pi$ interactions, what could be also anticipated from the distributions in Figs 2 and 3. The number of OH/$\pi$ interactions ($H_1$··$\Omega$ < 3.5 Å, O-$H_1$··$\Omega$ > 110° and $\beta_1$< 30°) is 141 (2.7%), the number of parallel alignment interactions (80° ≤ $\Theta$ ≤ 100°) is 837 (16.0%), while number of CH/O interactions (H··O < 3.2 Å, C-H··O > 110°, and H··O distance is shorter than H··$H_1$ distance) is 2333 (44.6%). The least frequent interactions are lone-pair/$\pi$ interactions (O··$\Omega$ distance is shorter than 3.5 Å, and O··$\Omega$ distance is shorter than $H_1$··$\Omega$ distance).[10] There are only 11 interactions; in accordance with the calculations showing that interaction is repulsive.[8b] Statistical analysis of crystal structures, determined by neutron diffraction method, is also

performed. The analysis shows a similar ratio of interactions between water molecule and aromatic group; there are 17 CH/O, 11 parallel alignment and 7 OH/$\pi$ interactions. There are no lone-pair/$\pi$ interactions.

Besides OH/$\pi$, CH/O, parallel alignment and lone-pair/$\pi$ interactions, there is large number of contacts which geometrically do not correspond to any of these interactions. This group contains 1912 contacts (36.7%). The density plots of $R_O$ versus the $r_O$ for four groups of contacts are shown in Fig. 4.
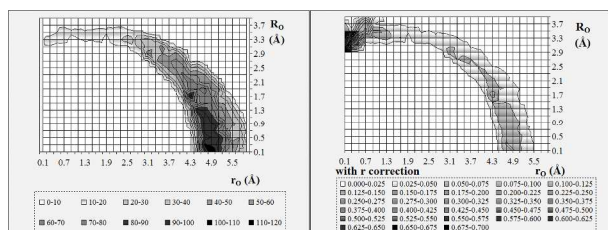


**Fig. 3** Uncorrected and corrected[16] density plot of the normal distance ($R_O$) versus the offset values of water oxygen ($r_O$) for CSD structures in which water oxygen was found in the observed ellipsoid.

Very small number of contacts (2.7%) are OH/$\pi$ interactions, in spite that these interactions have the strongest interactions energy (-3.19 kcal/mol)[9a] compared to the energies of parallel (-2.45 kcal/mol)[9b] and CH/O (-1.41 kcal/mol) interactions.[9a] Possible explanation can be that OH/$\pi$ interactions are statistically less probable because of small area around the centre of the ring, compared to area with larger offsets, where CH/O and parallel interactions occur. However, the area corrected diagrams, that take account the statistical correction, also show quite high preference for position of water molecules at large horizontal displacements (Fig. 2 and 3). It was shown that water/benzene interactions at large offsets are preferred in crystal structures[4,9] since it enables simultaneous interactions with the surrounding species in the crystal.[9] In OH/$\pi$ interactions one H of water and one face of the aromatic ring, that could form relatively strong interactions with the surrounding, are blocked. On the other hand, in CH/O interactions only O of water can be blocked, however, we find that even this O atom can form simultaneous interactions with environment. In parallel interactions water H and O atoms as well as aromatic ring are not directly blocked, however, the side of O-H bond that is close to the plane of the aromatic ring is blocked, preventing hydrogen bonds from that direction. In ESI, there are examples of crystal structures that illustrate possibilities for simultaneous interactions (Fig. S10-S12, ESI†).

We mentioned above that besides the OH/$\pi$, CH/O, parallel alignment and lone-pair/$\pi$ interactions, there are the other 1912 contacts (36.7%). Density plot in Fig. 4 shows that in other contacts the water oxygen may be placed above the ring ($r_O$ values are smaller than 2.5 Å), but it is mostly found outside of the ring, 1671 contacts. These 1671 contacts partially satisfy the first two criteria for CH/O interactions (H···O<3.2Å and C-H···O> 110°), however, in these contacts distances between phenyl hydrogen and water hydrogen are shorter than H···O distances (Fig. 1).
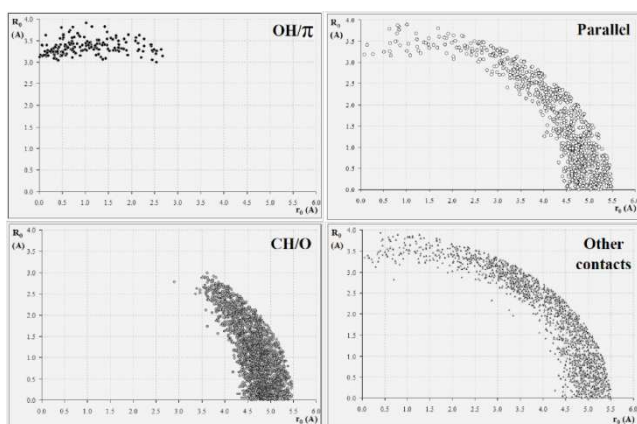


**Fig. 4** Density plots of the normal distance ($R_O$) versus the offset values of water oxygen ($r_O$) for different types of water/phenyl interactions in crystal structures from the CSD.

In order to evaluate if water molecules involved in the other contacts (Fig. 4) have attractive interactions with aromatic ring we performed *ab initio* calculations on benzene-water model systems. Model systems were made from the crystal structures by substituting all groups on aromatic ring with hydrogen atoms. The most frequent geometries of other contacts at offsets close to 0.0, 1.0, 2.0, 3.0, 4.0 and 5.0 Å were chosen as model systems (Fig. S9, ESI†).The details of the calculations are shown in ESI, (Table S1, ESI†). The results show that all interactions are attractive, however, calculated energies are quite weak, all of them are in the range of -0.01 to -0.48 kcal/mol, with the exception of one interaction with water close to the center of the ring (offset $r_o$ is 0.06 Å) and with interaction energy of -1.20 kcal/mol.

Since number of strongest OH/$\pi$ interactions between water and phenyl ring is very small, the results indicate that water molecule tends to form a weaker interactions with phenyl group. The possible explanation is that in weak interactions there are more possibilities for the water molecule to form simultaneous strong hydrogen bonds,[9] as was mentioned above. Namely, in case of OH/$\pi$ contacts, interacting OH group of water is sterically hindered to form additional bonds with neighboring fragments (Fig. S10, ESI†). On the other hand in CH/O interactions water molecule can form simultaneous interactions that additionally stabilize supramolecular structure. In crystal structure EFIPIT[19] (Figure 5), all atoms of water molecule simultaneously form strong classical hydrogen bonds or weak interactions with the atoms from the surrounding groups.

Because of the similar distribution of water/aromatic contacts in proteins and crystal structures from the CSD we can assume that the interactions in proteins are similar to the interactions in crystal structures.
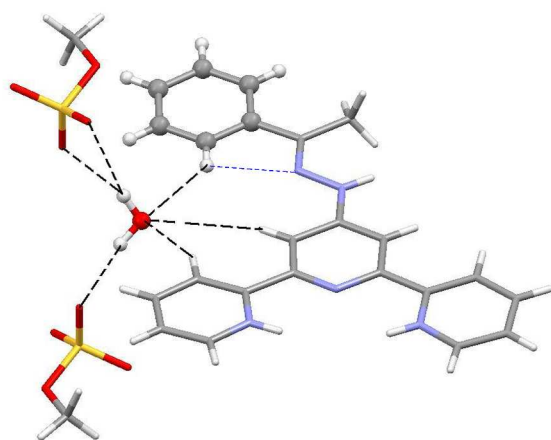
**Fig. 5** The fragment of crystal structure EFIPIT (2,2'-(4-((1-phenylethylidene) hydrazinyl)pyridine-2,6-diyl)dipyridinium nitrate methyl sulfate monohydrate) selected as an example for CH/O interaction between water molecule and phenyl group (H¨O= 2.51 Å and C-H¨O= 135.4°). The water molecule also builds additional intermolecular contacts represented by dotted lines: three classical hydrogen bonds with methyl sulfate anions (H¨O= 2.11Å and O-H¨O= 162.9°; H¨O= 2.11Å and O-H¨O= 151.0°; H¨O= 2.94Å and O-H¨O= 155.1°) and two additional CH/O interactions (H¨O= 2.35Å and C-H¨O= 151.7°; H¨O= 3.27Å and C-H¨O= 162.7°). The aryl group also builds additional intramolecular CH/N interaction, represented by blue dotted lines: H¨N= 2.43Å and C-H¨N = 99.7°).

In summary, study of interactions between water molecule and phenyl ring in proteins and crystal structures from the CSD reveal that water molecule form relatively small number of OH/π interactions (around 2%) in spite that these interactions are the strongest. On the other hand, water molecule form larger number of parallel and the largest number of CH/O interactions, in spite that these interactions are weaker. These results elucidate interactions of water with phenyl ring in protein structures and can be significant for recognizing water/aromatic interactions in proteins.

## Notes and references

[a] Institute of Chemistry, Technology and Metallurgy, University of Belgrade, Njegoševa 12, P. O. Box 473, Belgrade, Serbia.
[b] Department of Mathematics, University of Belgrade, Studentski trg 16, 11000 Belgrade, Serbia.
[c] Department of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia, Fax: (+381) 11-2184-330, E-mail: szaric@chem.bg.ac.rs
[d] Department of Chemistry, Texas A & M University at Qatar P.O. Box 23874, Doha, Qatar
†Electronic Supplementary Information (ESI) available: [Protein Data Bank search, calculations and geometry optimization]. See OI: 10.1039/b000000x/

1   (a) P. Agre, Angew. *Chem., Int. Ed.,* 2004, **43**, 4278; (b) D. D. Kozono, M. Yasui, L. S. King and P. J. Agre, *Clin.Invest.*, 2002, **109**, 1395; (c) M. Kumar, M. Grzelakowski, J. Zilles, M. Clark and W. Meier, *Proc. Natl. Acad.Sci. USA,* 2007, **104**, 20719.
2   (a) G. Hummer, J. C. Rasaiah and J. P. Noworyta, *Nature*, 2001, **414**, 188; (b) D. Takaiwa, I. Hatano, K. Koga and H. Tanaka, *Proc. Natl. Acad. Sci.USA*, 2008, **105**, 39; (c) T. A. Pascala, W. A. Goddarda and Y. Junga, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 11794; (d) E. Paineau, P. -A. Albouy, S. Rouzière, A. Orecchini, S. Rols and P. Launois, *NanoLett.*, 2013, **13**, 1751; (e) G. Pérez-Hernández and B. Schmidt, *Phys. Chem. Chem. Phys.*, 2013, **15**, 4995.
3   (a) R. Natarajan, J. P. H. Charmant, A. G. Orpen and A. P. Davis, *Angew. Chem., Int. Ed.*, 2010, **49**, 5125; (b) R. Prohens, M. Font-Bardia and R. Barbas, *CrystEngComm*, 2013, **15**, 845.
4   (a) T. J. Mooibroek and P. Gamez, *CrystEngComm*, 2012, **14**, 8462; (b) E. Voloshina, D. Usvyat, M. Schutz, Y. Dedkov and B. Paulus, *Phys. Chem. Chem. Phys.*, 2011, 13, 12041; (c) A. Simon, M. Rapacioli, J. Mascetti and F. Spiegelman, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6771; (d) L. D'Urso, C. Satriano, G. Forte, G. Compagnini and O. Puglisi, *Phys. Chem. Chem. Phys.*, 2012, **14**, 14605.
5   (a) J. F. Malone, C. M. Murray, M. H. Charlton, R. Docherty and A. J. Lavery, *J. Chem. Soc., Faraday Trans.*, 1997, **93**, 3429; (b) T. Steiner and G. Koellner, *J. Mol. Biol.*, 2001, **305**, 535; (c) S. D. Zarić, D. Popović and E. W. Knapp, *Chem.-Eur. J.*, 2000, 3935; (d) S. D. Zarić, *Eur. J. Inorg. Chem.*, 2003, 2197; (e) M. K. Milčić, Z. D. Tomić and S. D. Zarić, *Inorg. Chim. Acta*, 2004, **357**, 4327; (f) S. J. Kim, H. I. Seo and B. H. Boo, *Mol. Phys.*, 2009, **107**, 1261; (g) M. Prakash, K. Gopal Samy and V. Subramanian, *J. Phys. Chem. A*, 2009, **113**, 13845; (h) M. Besnard, Y. Danten and T. Tassaing, *J. Chem. Phys.*, 2000, **113**, 3741; (i) M. P. S. Mateus, N. Galamba and B. J. C. Cabral, *J. Chem. Phys.*, 2012, **136**, 014507; (j) M. Albertí, N. Faginas Lago and F. Pirani, *Chem. Phys.*, 2012, **399**, 232; (k) S. Suzuki, P. G. Green, R. E. Bumgarner, S. Dasgupta, W. A. Goddard and G. A. Blake, *Science*, 1992, **257**, 942.
6   K. P. Gierszal, J. G. Davis, M. D. Hands, D. S. Wilcox, L. V. Slipchenko and D. Ben-Amotz, *J. Phys. Chem. Lett.*, 2011, **2**, 2930.
7   (a) D. Z. Vojislavljević, G. V. Janjić, D. B. Ninković, A. Kapor and S. D. Zarić, *CrystEngComm*, 2013, **15**, 2099; (b) S. Tsuzuki, *Struc. Bond,* 2005, **115**, 149; (c) S. Li, V. R. Cooper, T. Thonhauser, A. Puzder and D. C. Langreth, *J. Phys. Chem. A*, 2008, **112**, 9031; (d) L. V. Slipchenko and M. S. Gordon, *J. Phys. Chem. A*, 2009, **113**, 2092; (e) L. M. Salonen, M. Ellermann and F. Diederich, *Angew. Chem., Int. Ed.*, 2011, **50**, 4808.
8   (a) D. Ž. Veljković, G. V. Janjić and S. D. Zarić, *CrystEngComm.* 2011, **13**, 5005; (b) D. P. Malenov, G. V. Janjić, D. Ž. Veljković and S. D. Zarić, *Comp. Theor. Chem.*, 2013, **1018**, 59; (c) W. Gao, J. Jiao, H. Feng, X. Xuan and L.Chen, *J. Mol. Model.*, 2013, **19**, 1273; (d) M. V Sigalov, E. P. Doronina and V. F. Sidorkin, *J. Phys. Chem. A*, 2012, **116**, 7718.
9   (a) B. D. Ostojić, G. V. Janjić and S. D. Zarić, Chem. Commun. 2008, **28**, 6546; (b) G. V. Janjić, D. Ž. Veljković and S. D. Zarić, *Cryst. Growth Des.*, 2011, **11**, 2680.
10  A. Jain, V. Ramanathan and R. Sankararamakrishnan, *Protein Sci.*, 2009, **18**, 595.
11  M. Egli and S. Sarkhel, *Acc. Chem. Res.*, 2007, **40**, 197.
12  T. M Raschke, *Curr. Opin. Struc. Biol.*, 2006, **16**, 152.
13  K. Flanagan, J. Walshaw, S. L. Price and J. M. Goodfellow, *Protein Engineering*, 1995, **8**, 109.
14  J. Walshaw and J. M. Goodfellow, J. Mol. Biol., 1993, **231**, 392.
15  S. Griep and U. Hobohm, *PDBselect 1992–2009 and PDBfilterselect. Nucleic Acids Res*, 2010, **38**, D318.
16  Z. Ciunik and G. R. Desiraju, *Chem. Commun.*, 2001, 703.
17  F. H. Allen, *Acta Crystallogr. Sect. B*, 2002, **58**, 380.
18  I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. Mc Cabe, J. Pearson and R. Taylor, *ActaCrystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 389.
19  J. E. Beves, E. C. Constable, C. E. Housecroft, M. Neuburger, S. Schaffner and J. A. Zampese, *Eur. J. Org. Chem.*, 2008, 3569.

# What are preferred water/aromatic interactions in proteins and crystal structures of small molecules?

Goran V. Janjić, Saša N. Malkov, Miodrag V. Živković and Snežana D. Zarić*[c,d]
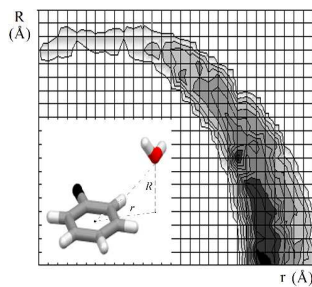
The distribution of water molecules around aromatic ring in proteins and crystal structures shows the largest number of the weakest interactions.