

Supplementary Information

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

Geometric analysis of data from the PDB

The crystal structures of proteins were screened for intermolecular contacts between non-coordinated water molecule and phenyl group of phenylalanine residue. The initial PDB search yielded 7 899 contacts in which water oxygen was found to be within 6.0 Å distance from the center of any phenyl ring (d distance at Fig. 1). The frequently populated area corresponds to the ellipsoid with the vertical axis of 4.0 Å and horizontal axis of 5.5 Å (Fig. S1). In this area, there are 5462 contacts (69 % of contacts with $d \leq 6.0$ Å).

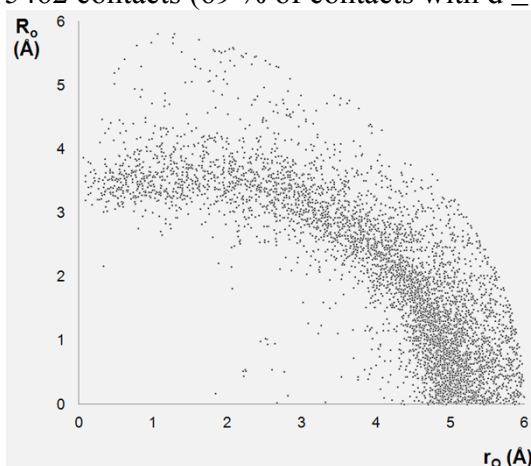


Fig. S1 The plot of the normal distances (R_o) versus offset values of water oxygen (r_o) for protein structures in which water oxygen was found within 6.0 Å distance from the center of any phenyl ring.

In the study, only the contacts in this area will be analyzed, because the contacts inside the ellipsoid correspond to interactions with the phenyl group (blue dots at Fig. S2), while the contacts outside the ellipsoid mainly correspond to the interactions of water molecule with the protein backbone (red dots at Fig. S2).

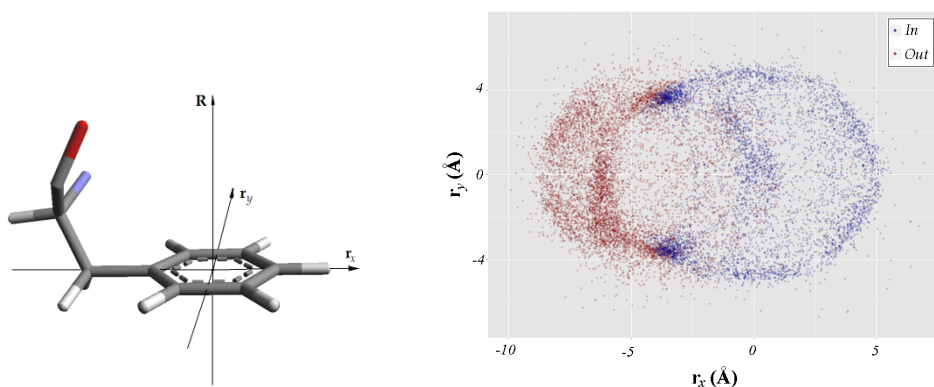


Fig. S2 Model system and the plot of r_y offset values versus r_x offset values of water oxygen for contacts inside (blue dots) and outside the ellipsoid (red dots), in which water oxygen was found to be within 10.0 Å distance from the center of any phenyl ring.

Area correction

The area correction assumes that the frequency distribution of interactions must be corrected for the fact that the acceptor, being of a multiatom type, has a finite area. The number of interactions with offset r is proportional to $2\pi r$. If N is the number of interactions in the range r and $r+\Delta r$, the area correction assumes the use of N/r in histograms instead of N .^{S1}

The non-corrected and corrected distributions of offset r_O values for water/phenyl contacts inside the ellipsoid, in proteins, are shown in Fig. S3. The non-corrected distribution of offset r_O values (Fig. S3) shows clear preference for water oxygen position with offset values of 4.5 to 5.0 Å. In this area water molecule and phenyl ring form the CH/O and the parallel alignment interaction. The area preference can be considered after area correction, that is very important for the multiatom type of acceptors.^{S1} The corrected diagram (Fig. S3) indicates that water oxygen have slightly pronounced preference to be at or near the centre of phenyl ring, in the area where water molecule forms the strongest (OH/ π) interaction with aromatic ring. However, the corrected diagram also shows high preference for water positions with r_O offset values from 4.5 to 5.0 Å, although the CH/O and parallel interactions that are in this area are not very strong.

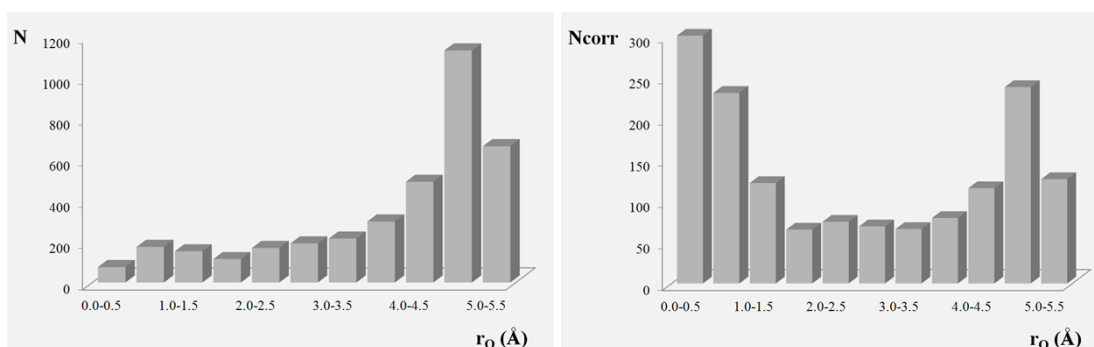


Fig. S3 Distributions of offset r_O : non-corrected (left) and area-corrected (right) diagram for water/phenyl contacts inside the ellipsoid.

The solvent accessible surface area (SASA)

For the same protein set we measured solvent accessible surface area (SASA),^{S2} using Visual Molecular Dynamics (VMD) software.^{S3} The surface of interest was the surface area of an amino acid residue, which is in the same time a part of the surface area of the protein. In terms of VMD, that is the protein surface area restricted to specific individual amino acids (i.e. belongs to the amino acid). To compute these SASA values we used VMD command "*measure sasa*" with probe radius set to 1.4 Å. We computed SASA values for all phenylalanine residues in the protein set.

The distribution of the SASA values for all phenylalanine residues in the protein set (45022 residues) shows that phenylalanine generally tends to be away from the protein surface. Most of them (around 43% of Phe residues) have the surface area less than 10 Å² (red columns at Fig. S4). The distribution of the SASA values for the Phe residues, which interact with water molecule, is different (blue columns at Fig. S4). It is relatively hard for water to reach to the aromatic ring with small open surface area. Therefore, 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).

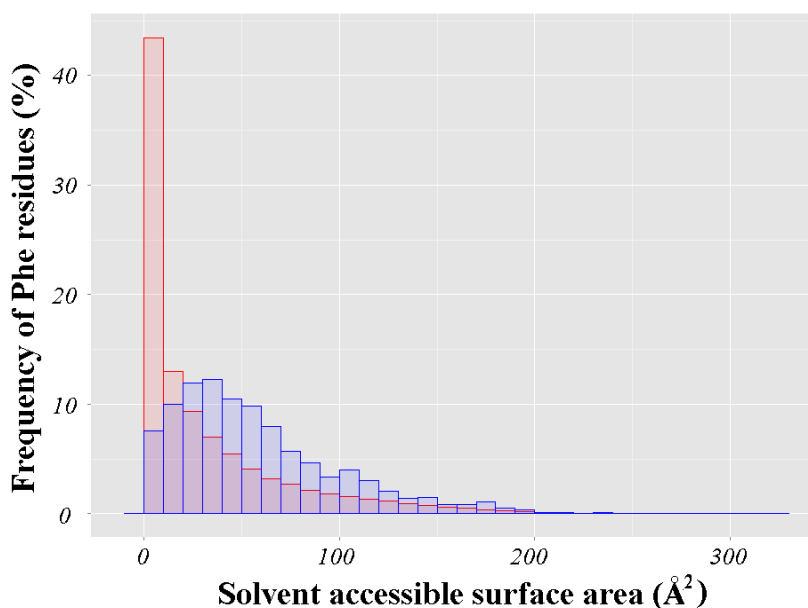


Fig. S4 The plot of the frequency versus the solvent accessible surface area for all phenylalanine residues (red columns) and phenylalanine residues interacting with water molecule (blue columns).

It is more important that there are no significant correlations between water molecules positions and the surface area. The distributions of surface areas for phenylalanine residues, shown as a function of different values of r_o offsets value are highly similar (Fig. S5).

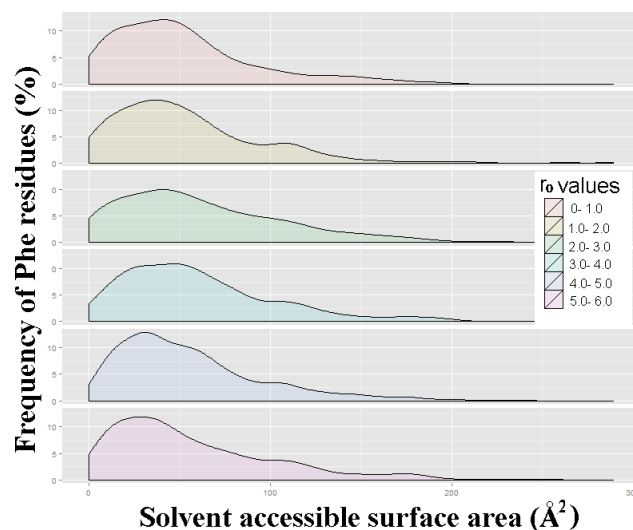


Fig. S5 The plots of the frequency versus the solvent accessible surface area for phenylalanine residues, as function of different offset r_0 values of interacting water molecule.

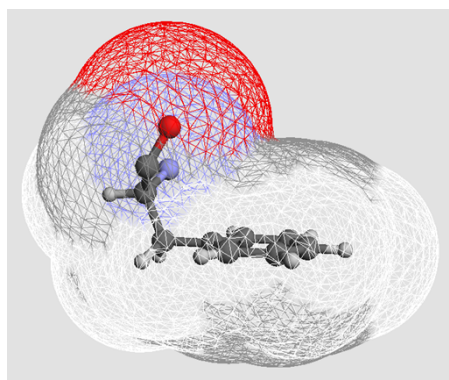


Fig. S6 The solvent accessible surface area of phenylalanine amino acid (Color Surface by-Atomic Number (white for hydrogen, grey for carbon and red for oxygen)). SASA values of phenylalanine amino acid is around 320 \AA^2 .

PDB analysis of data for Trp and Tyr residue

The distribution of water molecules around C_6 -aromatic ring of tyrosine (Fig. S7) is similar as in case of phenylalanine residues (Fig. 2). Namely, the substitution of H atom from aromatic ring by OH group does not drastically change the distribution of water molecules around C_6 -aromatic ring in the ellipsoid. However, the average number of water molecules obtained in this ellipsoid is larger (1.4 water molecules per Tyr residue) than in case of phenylalanine residues. This is a consequence of simultaneous interactions of water molecules with C_6 -aromatic and OH group from tyrosine residue or of hydrogen bonding with O-H group of tyrosine residue.

In the case of interactions with side chain of tryptophan, in which C_6 -aromatic ring is fused with the pyrrole ring, the distribution of water molecules around C_6 -aromatic ring (Fig. S7) is different than in case of phenylalanine residues (Fig. 2). There is a

more pronounced tendency towards orientation with planar arrangement. The average number of water molecules obtained in the ellipsoid around C₆-ring of tryptophan residue is smaller (1.2 water molecules per Trp residue) than in case of phenylalanine residues.

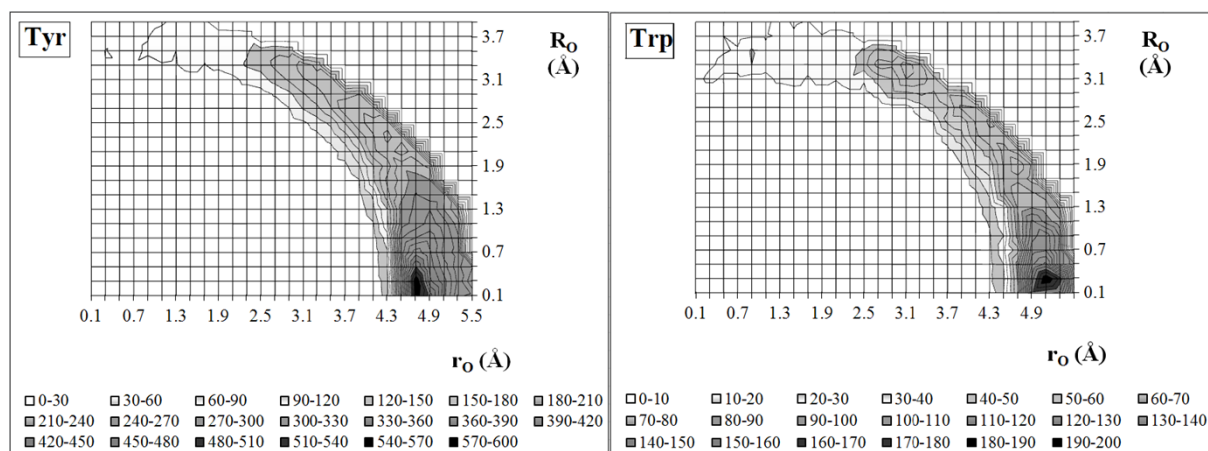


Fig. S7 Uncorrected density plot of the normal distance (R_O) versus the offset values of water oxygen (r_o) for water/aryl contacts inside the ellipsoid of C₆-aromatic ring from tyrosine residue (left figure) or from tryptophan residue (right figure).

Geometric analysis of data from the CSD

The CSD search program ConQuest1.13 was used to retrieve structures satisfying the following criteria: (a) the crystallographic R factor is $\leq 10\%$; (b) error-free coordinates according to the criteria used in the CSD; (c) H-atom positions were normalized using the CSD default bond lengths; (d) no polymer structure; (e) no powder structures; (f) not disordered structures; (g) 3D coordinates determined.

The crystal structures, involving non-coordinated water molecules and phenyl groups, were screened for intermolecular contacts. The initial CSD search yielded 8562 cases in which water oxygen was found to be within 6.0 Å distance from the center of any phenyl ring (Fig. 1). For these contacts, the plot of the normal distance (R_O) for different offset values of water oxygen (r_o) is shown in Fig. S8. The scatter plot indicates some geometrical preferences for water-phenyl interactions. The more frequently populated area corresponds to the ellipsoid with the vertical axis of 4.0 Å and horizontal axis of 5.5 Å. Those parameters describe the area where the three common types of interactions between water molecule and aromatic ring occur (O-H/ π , C-H/O, parallel alignment and lone-pair/ π interactions). We found 5223 contacts in this area (61.0 % of contacts). In the study, only the contacts in this area will be analyzed.

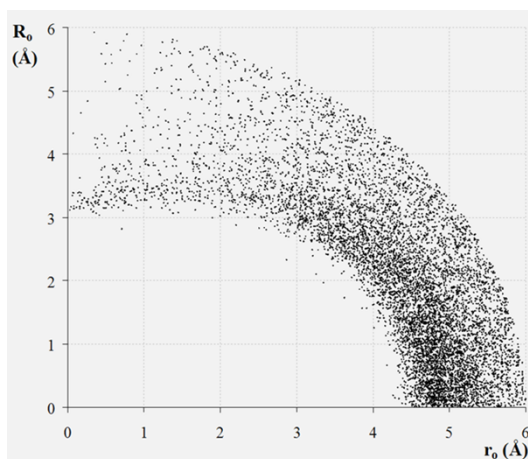


Fig. S8 The plot of the normal distances (R_o) versus the offset values of water oxygen (r_o) for structures in which water oxygen was found to be within 6.0 Å distance from the center of any phenyl ring.

Ab initio calculations

In order to evaluate whether the water molecules involved in other contacts (Fig. S9), have attractive interactions with aromatic ring, *ab initio* calculations on benzene-water model systems were performed. The calculations were performed at MP2/cc-pVQZ level, that gives results that are in excellent agreement with the very accurate CCSD(T) data for water-benzene interactions,^{S4} using the Gaussian09 program,^{S5} The basis set superposition error (BSSE) was calculated using the standard Boys–Bernardi counterpoise procedure.^{S6} The results of calculations are shown in Table S1.

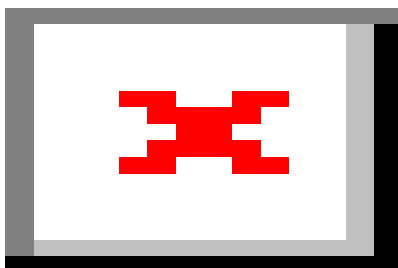


Fig. S9. The most frequent geometries of water-benzene contacts, marked as *Other contacts*, at different offset distances, that used for *ab initio* calculations.

Table S1. The geometric parameters (in Å) and calculated interaction energies (in kcal/mol) for model systems of other contacts.

	REFCODE	r_o	R_o	r_1	R_1	ΔE^*
(a)	LACDAU10	0.06	3.35	0.99	3.08	-1.20
(b)	GOHQID01	1.04	3.44	1.93	3.02	-0.07
(c)	EFOYOO	2.02	3.27	2.15	2.30	-0.01
(d)	YONNEV	3.05	2.88	3.26	2.07	-0.48
(e)	TEVSAO	4.02	2.22	4.33	1.49	-0.03
(f)	GOFHAK	4.98	0.01	4.61	0.82	-0.29

* MP2/cc-pVQZ level

Individual examples of crystal structures

It is evident that packing can also influence the frequency of interactions in crystal structures. In the following are collected a cross-section of crystal structures in which water-phenyl contacts are observed. The examples given here should provide an explanation for the differences in the frequency of investigated types of interactions. In order to visualize the relative orientation between water molecule and phenyl group a several view points are chosen in pictorial illustrations. The interactions of water molecule with surrounding molecules (groups) are also presented. For water molecule and phenyl group, ball and stick style of atoms was chosen in order best to visualize them mutual orientation. Other fragments are shown by stick style, while intermolecular interactions are shown by dotted line.

As already stated above, CH/O interactions are the most common type of contact between water molecules and phenyl groups. The earlier analysis of the CH/O interactions in the crystal structures from the CSD showed that aromatic C–H donors do not have strong preference for linear contacts. Also, the acceptor oxygen atom has tendency for simultaneous interactions. In accordance with these conclusions, the selection of crystal structure with typical geometry was carried out. The typical geometry is reflected in the example shown in Fig. 5. In crystal structure EFIPIT (Fig. 5), beside of CH/O interaction with phenyl group, the water oxygen atom simultaneously builds two additional CH/O interactions with other aromatic C-H donor groups. In addition, the water hydrogen atoms form three classical hydrogen bonds with methyl sulfate anions. These data suggest that weak CH/O interactions coexist with a large number of other interactions.

Despite being the significantly stronger, OH/ π interactions rarely occur in crystal structures. The visual analysis of these structures, provide an explanation for these unexpected results. For this purpose, the fragment of crystal structure TAQCIY^{S7} is shown at Fig. S10. In this structure, beside of OH/ π interaction, the water molecule builds two classical hydrogen bonds and one CH/O interaction with two neighboring 3-((4-amino-2-methylpyrimidin-5-yl)methyl)-5-(2-hydroxyethyl)-4-methylthiazoliumcations.

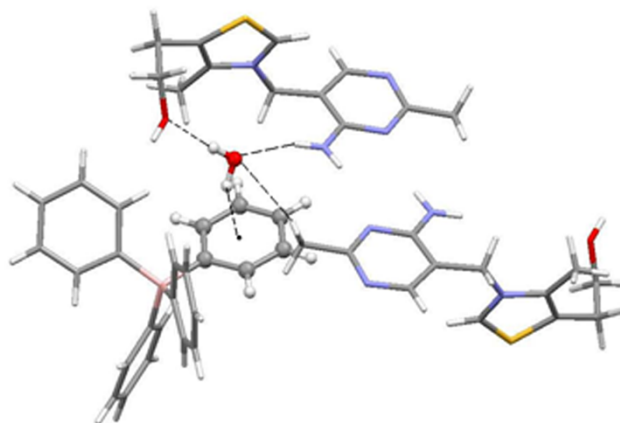


Fig. S10 The fragment of crystal structure TAQCIY (3-((4-amino-2-methylpyrimidin-5-yl)methyl)-5-(2-hydroxyethyl)-4-methylthiazolium tetraphenylborate monohydrate) selected as an example for OH/ π interaction between water molecule and phenyl group ($H_1 \cdots O = 2.62 \text{ \AA}$, $O-H_1 \cdots O = 134.4^\circ$ and $\beta_1 = 11.1^\circ$). The water molecule also builds additional intermolecular contacts: two classical hydrogen bonds ($H \cdots O = 1.93 \text{ \AA}$ and $O-H \cdots O = 174.7^\circ$; $H \cdots O = 2.02 \text{ \AA}$ and $N-H \cdots O = 159.8^\circ$) and one additional CH/O interaction ($H \cdots O = 2.74 \text{ \AA}$ and $C-H \cdots O = 122.4^\circ$) with 3-((4-amino-2-methylpyrimidin-5-yl)methyl)-5-(2-hydroxyethyl)-4-methylthiazolium cations.

A comparison of crystal packing in structures, shown in Fig. 5 and Fig. S10, it is clear that water molecule, when form CH/O interaction with phenyl group, can simultaneously build a larger number of interactions with the environment.

Fig. S11 and S12 show examples of contacts from group marked as *Other contacts*. To determine whether these contacts have attractive characters or they are the consequence of packing in the crystal structures, the benzene/water interaction energies were calculated, at MP2/cc-pVQZ level, using the Gaussian09 program. The interaction energies were calculated only for benzene/water orientations, which correspond to water/phenyl orientations in fragment of crystal structures, shown in Fig. S11 and S11. As can be seen in Fig. 4, the *Other contacts* are rare above the ring. The typical orientation of water molecule, placed above the ring, is shown in Fig. S11. The calculated value of energy for this orientation (Fig. S11c) is -1.20 kcal/mol . Obviously, this is a very weak interaction. However, this orientation allows water molecule to form a large number of additional interactions (Fig. S11). In crystal structure LACDAU10^{S8} (Fig. S11a and S11b), the position of water molecule is stabilized by three classical hydrogen bonds and two CH/O interactions.

The water molecule is mostly found outside of the ring (89.2% of *Other contacts*). Most of these contacts have the orientation where water molecule is located near to the middle plane of phenyl ring at long offset distance. Such orientation occurs in crystal structure MUNVAT,^{S9} whose fragments are shown in Fig. S12. The calculated value of energy for this orientation (Fig. S12c) is -0.84 kcal/mol . As in the previous case, it is a very weak interaction and the position of water molecule also is stabilized by large number of additional interactions. The water molecule is involved in building up additional intermolecular contacts with surrounding groups (three classical hydrogen bonds and two CH/O interactions).

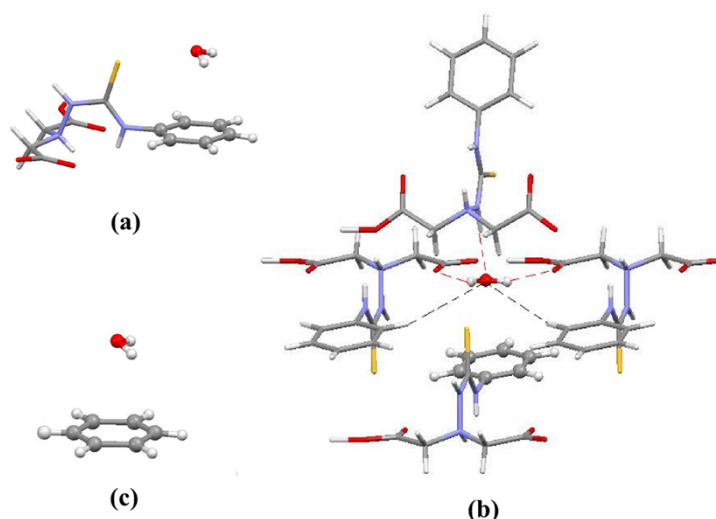


Fig. S11 (a) The fragment of crystal structure **LACDAU10** (*4-phenylthiosemicarbazide-diacetic acid monohydrate*) selected as an example of contacts from group marked as Other contacts, where water oxygen atom is positioned above the centre of phenyl ring ($r_O = 0.06\text{\AA}$ and $R_O = 3.35\text{\AA}$). (b) The water molecule also builds additional intermolecular contacts: three classical hydrogen bonds ($H\cdots O = 1.86\text{\AA}$ and $O-H\cdots O = 161.0^\circ$; $H\cdots O = 1.86\text{\AA}$ and $O-H\cdots O = 161.0^\circ$ and $O-H\cdots O = 151.0^\circ$; $H\cdots O = 2.01\text{\AA}$ and $N-H\cdots O = 171.1^\circ$) and two additional CH/O interactions ($H\cdots O = 3.01\text{\AA}$ and $C-H\cdots O = 134.1^\circ$; $H\cdots O = 3.01\text{\AA}$ and $C-H\cdots O = 134.1^\circ$) with neighbouring fragments. (c) Model system used for *ab initio* calculation, with benzene/water orientation which correspond to water/phenyl orientation in Fig. S11a.

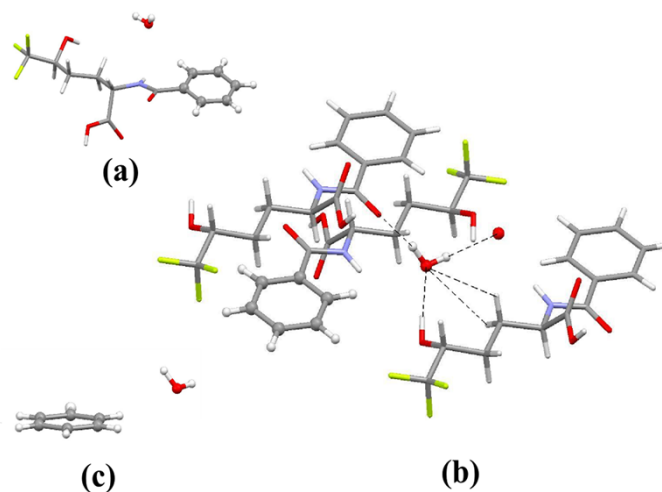


Fig. S12 The fragment of crystal structure **MUNVAT** (*N-benzoyl-6,6,6-trifluoro-5-hydroxynorleucine monohydrate*) selected as an example of contacts from group marked as Other contacts, where water oxygen atom is located near to the middle plane of phenyl ring at long offset distance ($r_O = 4.89\text{\AA}$ and $R_O = 0.08\text{\AA}$). (b) The water molecule also builds additional intermolecular contacts: three classical hydrogen bonds ($H\cdots O = 1.86\text{\AA}$ and $O-H\cdots O = 161.0^\circ$; $H\cdots O = 1.86\text{\AA}$ and $O-H\cdots O = 161.0^\circ$ and $O-H\cdots O = 151.0^\circ$; $H\cdots O = 2.01\text{\AA}$ and $N-H\cdots O = 171.1^\circ$) and two additional CH/O interactions ($H\cdots O = 3.01\text{\AA}$ and $C-H\cdots O = 134.1^\circ$; $H\cdots O = 3.01\text{\AA}$ and $C-H\cdots O = 134.1^\circ$) with neighboring fragments. (c) Model system used for *ab initio* calculation, with benzene/water orientation which correspond to water/phenyl orientation in Fig. S12a.

References

- S1. Z. Ciunik and G. R. Desiraju, *Chem. Commun.*, 2001, 703.
- S2. M. L. Connolly, *Science*, 1983, **221**, 709.
- S3. W. Humphrey, A. Dalke and K. Schulten, *J. Molec. Graphics*, 1996, **14**, 33.
- S4. S. Tsuzuki, *Struc Bond*, 2005, **115**, 149.
- S5. M. J. Frisch, et al. Gaussian 03, revision C.02; *Gaussian, Inc.: Wallingford, CT*, 2004.
- S6. S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.
- S7. N.-H. Hu, H.-Q. Jia, J.-W. Xu and K. Aoki, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 2005, **61**, o457.
- S8. V. K. Kravtsov, Y. A. Simonov, N. V. Gerbeleu, O. A. Bologa, V. I. Lozan and T. I. Malinovski, *J. Struct. Chem.*, 1994, **35**, 154.
- S9. N. A. Tolmacheva, I. I. Gerus, V. G. Dolovanyuk, I. S. Kondratov and G. Haufe, *Eur. J. Org. Chem.*, 2009, 5012.