Article

C–H/O interactions of aromatic CH donors within proteins: a crystallographic study

Jovan Lj. Dragelj, Ivana M Stankovic, Dragana M. Bozinovski, Tim Meyer, Dusan Z Veljkovic, Vesna B Medakovi#, Ernst-Walter Knapp, and Snezana D. Zaric

Cryst. Growth Des., **Just Accepted Manuscript** • DOI: 10.1021/acs.cgd.5b01543 • Publication Date (Web): 04 Mar 2016

C–H/O interactions of aromatic CH donors within proteins: a crystallographic study

Jovan Lj. Dragelj,[a] Ivana M. Stanković,[b] Dragana M. Božinovski,[c] Tim Meyer,[a] Dušan Ž. Veljković,[d] Vesna B. Medaković,[d] Ernst-Walter Knapp,[a] and Snežana D. Zarić[d,e]*

[a] Fachbereich Biologie, Chemie, Pharmazie/Institute of Chemistry and Biochemistry, Freie Universität Berlin, Fabeckstrasse 36A, 14195 Berlin, Germany.

[b] Institute of Chemistry, Technology and Metallurgy, University of Belgrade, Njegoševa 12, 11000 Belgrade, Serbia.

[c] Helmholtz Centre for Environmental Research – UFZ, Department of Proteomics, Permoserstr. 15, Leipzig, Germany.

[d] Department of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia.

[e] Department of Chemistry, Texas A&M University at Qatar, P. O. Box 23874, Doha, Qatar.

* Corresponding author; E-mail: szaric@chem.bg.ac.rs

KEYWORDS: Aromatic amino acids; C–H/O interactions; Hydrogen bond; PDB.

ABSTRACT

C–H/O interactions of aromatic C–H donors within proteins have been studied by analysing the data in the Protein Data Bank (PDB). The C–H/O interactions were studied between aromatic donors; phenylalanine, tyrosine and tryptophan and the acceptors; alcohol, backbone amide and side-chain amide groups. The analysis of the C–H–O angle indicates that protein C–H donors do not show the preference for linear contacts. Although there is no tendency for linear C–H/O interactions, there are only around 3% of bifurcated C–H/O interactions. Furthermore, the analyses of the C–H/O interactions indicate an influence of simultaneous classical hydrogen bonds, especially for the tyrosine systems. The calculated electrostatic potential maps for model systems can explain the results of the crystallographic analysis. These results can be important for recognizing the C–H/O interaction of aromatic rings in the crystal structures of proteic systems.

INTRODUCTION

C–H/O interactions represent a wide group of weak hydrogen bonds that have important roles in many biological macromolecule structures like in the stability of proteins, in different types of interactions e.g. protein-protein, protein-ligand and protein nucleic acid interactions as well as in crystal structures and in enzymatic activity.[1-3] A few studies have presented that weak C–H/O interactions exist between parallel and anti-parallel beta sheets in proteins.[4-6] It was established that the role of C–H/O interactions is primarily to stabilize protein structures where they contribute up to 25% among the total number of hydrogen bonds detected in proteins.[5-13]

The C–H/O interactions are widely studied using spectroscopic methods,[14-17] theoretical calculations,[9,11,14,18-23] and structure analysis of data in the Cambridge Structural Database (CSD)[24,25] and the Protein Data Bank (PDB).[10,12,13,26]

Although C–H/O interactions are considered to belong to the group of weak, linear interactions, their energies can vary from very weak, -0.3 kcal/mol, to quite strong, $E < -4$ kcal/mol.[8] In the case of aromatic C–H donors, the interaction energy depends on the aromatic ring substituents and on the acceptor.[20,27,28]

Previous studies of C–H/O interactions, involving non-aromatic molecules, showed the tendency for linear geometry. However, our recent studies of C–H/O interactions of aromatic C–

H groups did not show the tendency for linear geometry.[29-31] The analysis of the crystal structure data from the Cambridge Structural Database indicated that aromatic C–H donors do not show a preference for linear contacts and that the preference depends on the type of atom or group in *ortho*-position to the interacting C–H group. The acceptor oxygen atom has the possibility to form simultaneous C–H/O interactions with atoms or groups in *ortho*-position to the interacting C–H group. Also, our results of quantum chemical calculations showed that linear C–H/O interactions of aromatic C–H groups are not energetically favored.[29-31] Namely, the calculated interaction energy for the linear C–H/O interaction between water and benzene is -1.28 kcal/mol, while the interaction energy for the bifurcated C–H/O interaction is -1.38 kcal/mol calculated at MP2/cc-pVTZ level with BSSE correction.[30] The interaction energies for the bifurcated C–H/O interactions between water and pyridine are also stronger than for the linear ones.[29] The result on pyridine also showed that the simultaneous hydrogen bond with pyridine N atom strengthens the C–H/O interaction by about 20%. The calculations on water-benzene-water system showed that two water/benzene C–H/O interactions weaken each other.[32]

The goal of our study was to elucidate the geometry of protein aromatic C–H/O interactions. We hypothesized that linear aromatic C–H/O interactions in proteins are not very frequent, since our previous data showed that bifurcated interactions are more stable than linear ones.[29-31]

Here we present the results based on the analysis of data in the Protein Data Bank (PDB) and on quantum chemical calculations of electrostatic potentials. We studied C–H/O interactions of three aromatic amino acids; phenylalanine, tyrosine and tryptophan, with several acceptors.

COMPUTATIONAL METHODS

PDB Search

The crystallographic analysis is based on the protein crystal structures archived in the Protein Data Bank (PDB).[33] To study C–H/O interactions in proteins, contacts involving three different amino acids, phenylalanine, tyrosine and tryptophan (Figure 1a) and different types of acceptors were screened. As acceptors, following groups were considered: double bounded oxygen atom from polypeptide backbone chain (backbone amide group), oxygen atom from side chain groups of amino acids serine, threonine and tyrosine (alcohol group) as well as the double bonded

oxygen atoms from side chain groups of amino acids asparagine and glutamine (side-chain amide group). The geometric criteria for C–H/O interactions were the same as in the study of C–H/O interactions with a benzene molecule.[30]

We used the data from the PDB,[33] employing the following criteria to assemble the set:

(1) no theoretical model structures and no NMR structures were accepted,

(2) only crystal structures with a resolution of 2.5 Å or better and R_{free}-$R_{work} \leq 0.05$ were accepted,

(3) only crystal structures with sequence identity < 30% were used to assure the nonredundancy,

(4) structures not chosen are those in which the ligands are ions, solvent molecules, or disordered species,

(5) if the distance between donor and any non-protein residue is less than 5 Å, the C–H/O interaction of this donor is not considered,

(6) a residue is not considered if it does not contain all heavy atoms resolved crystallographically,

(7) PDB structures with more than 3000 residues were not considered because large structures have repeating motifs and can interfere with statistics,

(8) a residue is not considered if it represents a physically impossible overlap ≥ 0.4 Å or clash, with nonbonded atoms,[34]

(9) a residue is not considered if its side chain rotamer is disfavored,[34]

The validation test for steric clashes (8) and rotamers (9) have been calculated with the MolProbity validation tool,[35] under Phenix.[36]

If not already present, hydrogen atoms in polar groups were added using the program HBPLUS.[37] The ring hydrogens were added by extrapolation of the ring centre – C atom direction with the C–H distances (1.0800 Å) taken from the CHARMM force field.[38] Scripts for the search were written in Python (http://www.python.org/) and for PDB file parsing MDAnalysis,[39] python library was used (http://mdanalysis.googlecode.com/svn/trunk/doc/html/index.html#).

The geometric parameters used for the PDB search and description of C–H/O interactions are shown in Figure 1. A contact was considered C–H/O interaction if the distance between a hydrogen atom from the C–H group of an amino acid and an oxygen atom from acceptor group (d) was less than 2.9 Å and angle $\alpha \geq 110°$ (Figure 1b).

To avoid the parallel alignment interactions[40-42] and short H···H contacts between aromatic ring and an –OH group, an additional criterion was used; the distance d should be shorter than the distances between interacting hydrogen and the hydrogen atom of the –OH group.

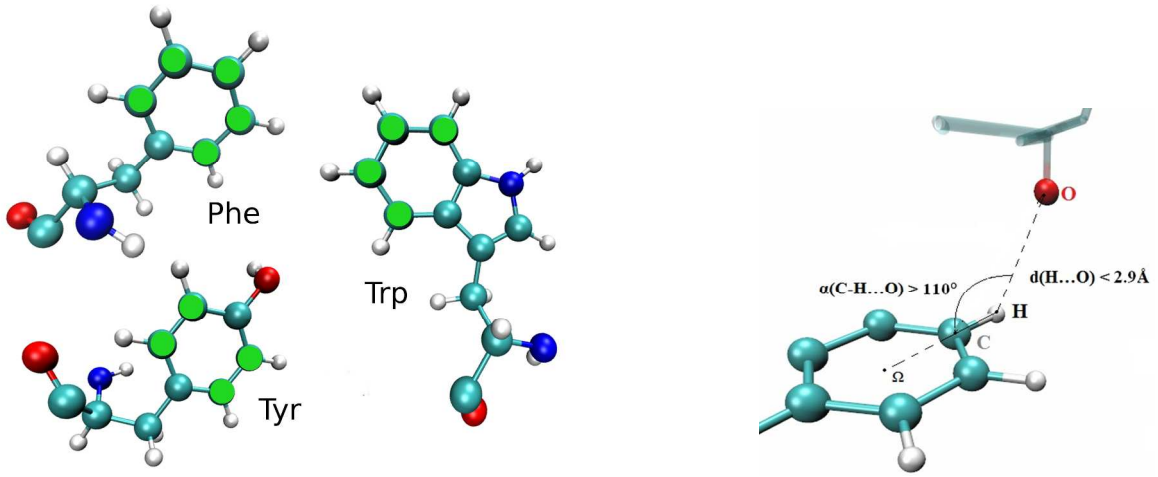
The influence of simultaneous classical hydrogen bonds was also studied as data analyses obtained by PDB search revealed that the acceptors, alcohol, backbone amide and side-chain amide groups, in some of the structures can simultaneously form classical hydrogen bonds. The package HBPLUS was used to identify whether acceptors form classical hydrogen bonds simultaneously with C–H/O interactions. The criterion for the classical hydrogen bond was angle greater than 110° and donor atom – acceptor atom distance not greater than 4Å.

The preference for linear geometry can be observed by the distributions of angle α (Figure 1b). To obtain more reliable data, cone correction should be used.[43] Another parameter was used to determine the position of acceptor relative to donor, the angle φ. This is the angle in the plane of the aromatic ring formed by the acceptor`s oxygen atom projection to the plane (O_p), the centre of the aromatic ring (Ω) and the reference C atom (Figure 2). The reference C atom is the one bonded to the protein main chain. The angle ranges were up to φ = 180° for phenylalanine and tyrosine, and up to φ = 360° for tryptophan because of its absence of symetry. These regions have angle φ values in the range of 0 - 30° for phenylalanine, 0 - 30° and 150 - 180° for tyrosine, and 0 - 30° and 270 - 360° for tryptophan.

The histograms for the angles α and φ have bins of 10°. For the parameter d, the range is 2.0 Å < d < 2.9 Å with the bin of 0.1 Å. The frequency of the interactions was expressed in percentage in relation to the total number of interactions in certain donor-acceptor system. For the cone correction of the angle α, in the factor 1/sin α, the α value was taken to be the α value from the beginning of a bin.[43]

Calculations of electrostatic potentials

The structures of donors and acceptors were optimized at MP2/cc-PVTZ level. Calculations were performed using Gaussian 09 series of programs.[44] From wavefunction files, electrostatic potential maps were calculated and visualized using the Wavefunction Analysis Program (WFA-SAS).[45,46]



a) **b)**

Figure 1. (a) C–H donor residues with carbon atoms involved in C–H/O interactions indicated in green. (b) The geometric parameters and atom labelling used for the search of the PDB and analysis of C–H/O interactions between amino acid residues (Phe, Tyr, Trp) and different oxygen acceptors. The distance between amino acid (Phe, Tyr, Trp) C–H group and the acceptor oxygen is d. Angle α is the C–H–O angle.

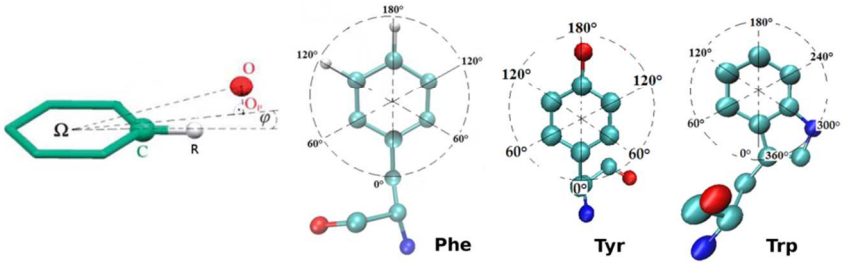


Figure 2. Angle φ defines the acceptor position relative to donor. The angle values are represented for the three donors: Phe, Tyr and Trp. The reference position for φ = 0° is the carbon atom bonded to the protein main chain.

RESULTS AND DISCUSSION

Analyses of the data from protein crystal structures

By searching the PDB using the criteria described in the Methodology section, large number of C–H/O interactions for three aromatic amino acids, phenylalanine, tyrosine and tryptophan, with different types of acceptors has been found (5168 for Phe, 4123 for Tyr, and 1282 for Trp). The amino acid phenylalanine was shown before to be a good donor among the amino acids.[47]

The obtained data revealed that the acceptors, amide group from polypeptide backbone chain, alcohol group from side chain groups of amino acids serine, threonine and tyrosine and amide group from side chain groups of amino acids asparagine and glutamine, in some of the structures can simultaneously form classical hydrogen bonds. It was shown that the hydrogen bond can influence C–H/O interactions of pyridine.[29] Therefore the C–H/O interactions with and without the presence of the classical hydrogen bonds were separately analysed. The number of the C–H/O interactions for three aromatic amino acids with and without hydrogen bonds are given in Table 1. The overall percentage of C–H/O contacts in presence of other, stronger hydrogen bonds, is 66%. Upon comparison, the number of bifurcated C–H/O interactions for each system is small (Table 1).

In our previous studies we found a substantial number of bifurcated interactions between benzene and water (33% of contacts),[30] showing their importance in aromatic systems. In this study using the same geometric criteria for the bifurcated C–H/O interactions, we found significantly smaller number of bifurcated interactions; 2.05% for C–H/O interactions with hydrogen bonds and 4.65% for C–H/O interactions without hydrogen bonds, Table 1. This might be a consequence of the fact that acceptor and donor groups are arranged closer in proteins than in small molecule crystals, so an acceptor could form simultaneous C–H/O interactions with different donors rather than with the same one as in bifurcation. The number of C–H/O interactions in the presence of classical hydrogen bonds is around twice as high as in the absence, except for the side-chain amid group acceptor.

Table 1. Number of C–H/O interactions of donors with different acceptor molecules in crystal structures from the PDB. The numbers of total and bifurcated C–H/O interactions for each system with and without hydrogen bonds are presented.

	C–H/O interactions with simultaneous classical hydrogen bonds							C–H/O interactions without simultaneous classical hydrogen bonds						
	O-H (Tyr, Ser, Thr)		$O=CNH_2$ (Asn, Gln)		O=CNH (backbone)			O-H (Tyr, Ser, Thr)		$O=CNH_2$ (Asn, Gln)		O=CNH (backbone)		
Phe	Total	589	Total	108	Total	2718	Phe	Total	72	Total	148	Total	1533	
	Bif	31	Bif	2	Bif	47		Bif	9	Bif	3	Bif	88	
Tyr	Total	394	Total	120	Total	2162	Tyr	Total	44	Total	125	Total	1278	
	Bif	13	Bif	1	Bif	25		Bif	2	Bif	0	Bif	51	
Trp	Total	127	Total	35	Total	705	Trp	Total	15	Total	37	Total	363	
	Bif	9	Bif	1	Bif	14		Bif	2	Bif	1	Bif	12	

The distributions of the distance d (Figure S1) are in the range of 2.5 - 2.9 Å, similarly to the small aromatic molecules.[29,30] There are no significant differences in C–H/O bond lengths for 9 different systems with and without classical hydrogen bond.

The geometries and preference for linear contacts in C–H/O interactions were studied separately for three aromatic amino acids (Phe, Tyr and Trp) with all chosen acceptors (alcohol, side-chain amide and backbone amide groups). The preference for linear geometry can be analysed by the distributions of angle α. To obtain more reliable data, cone correction was done. As the probability of finding interactions decreases with increasing angle α for purely geometrical reasons, the angular distribution must be cone corrected by a factor of 1/sin α to properly reflect angular preferences. The corrected total distributions of angle α, with and without classical hydrogen bonds are shown in Figure 3. The noncorrected distributions are shown in Figure S2.

For all the three investigated aromatic amino acids, the corrected distributions of the angle α show very small number of the interactions with angle α close to 180°, hence there is no any preference toward linear arrangements. The most of the interactions have angle α in the range of 135° - 165°.

The donors exhibit different angle maxima: phenylalanine around the angles of 150° and 160°, tyrosine 130° - 170° and tryptophan around 130° and 150° - 160°.

Tyrosine shows more pronounced tendency of forming interactions around $\alpha = 170°$ with the side-chain amide group than with alcohol or backbone amide group, while phenylalanine exhibits the opposite behaviour. Unlike phenylalanine, tyrosine contains the –OH group substituent which can influence the C–H/O interaction with the side-chain amide acceptor. This could be a consequence of the simultaneous hydrogen bonding between the –OH substituent of tyrosine and a strong acceptor, the nitrogen atom in amide group.

The deviation in the distributions of the angle α in the systems tryptophan/side-chain amide and tryptophan/alcohol could be a consequence of the small number of interactions found in the PDB, 72 and 142 respectively.

Our recent study of benzene C–H/O interactions with different acceptors,[30] and the one with the pyridine–water system do not indicate the preference for linear C–H/O contacts.[29] Similar results are obtained in this work. Bearing in mind the Steiner's previous study on C–H/O interactions of non-aromatic donors, the difference between two types of the donors, non-aromatic and aromatic, shows a clear tendency towards linearity and non-linearity, respectively.[24]

The total distributions of the angle φ, with and without hydrogen bonds, for C–H/O interactions of aromatic amino acids (Phe, Tyr and Trp) with different acceptors are shown in Figure 4.

As mentioned above, the systems tryptophan/side-chain amide and tryptophan/alcohol contain small number of interactions, 72 and 142 respectively; hence the distributions for these systems are not particularly reliable.

The donors show different angle maxima: phenylalanine around the angles of 50° - 60° and smaller number of interactions around 130° and 170°, tyrosine around 50° - 70°, 120° and 150°, and tryptophan in the vicinity of the angles 50° and 60° and smaller peaks in the range of 180° - 270°. The results show that there is no tendency for pronounced peaks around $\varphi = 60°$, 120°, 180°, 240°, i.e. in the regions of the donor CH bonds, with exception of the system tryptophan/backbone amide with a peak around $\varphi = 60°$.

The influence of the presence of a classical hydrogen bond on the C–H/O frequency and geometry has been studied by analysing the φ and α angle distribution, with and without classical

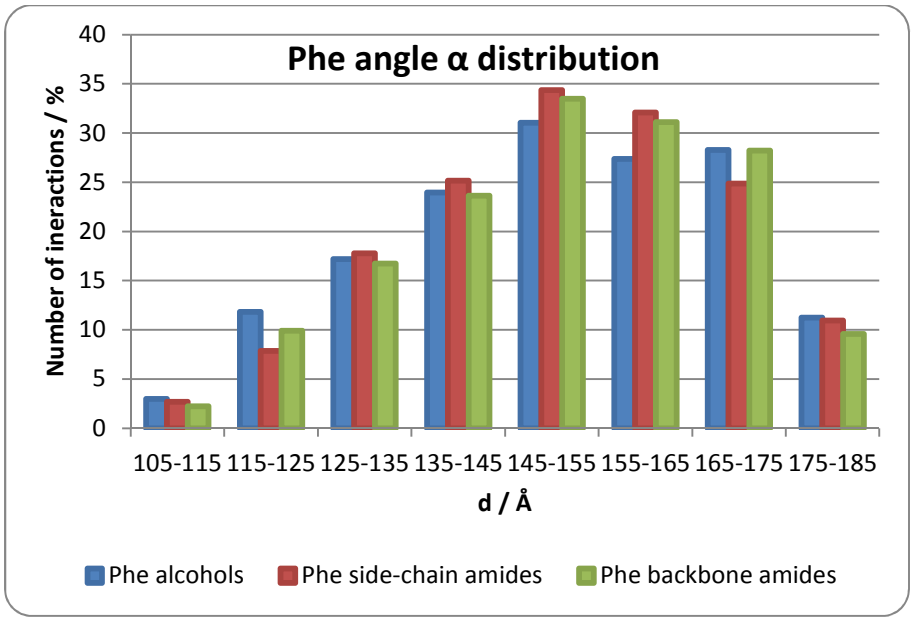
hydrogen bonds separately. The diagrams for the angle φ distributions are given in Figures 5-7. The diagrams for the angle α distributions are given in Figures S3-S5.

The distributions for systems tryptophan/alcohol and tryptophan/side-chain amide are not presented because of small number of interactions making statistics unreliable.

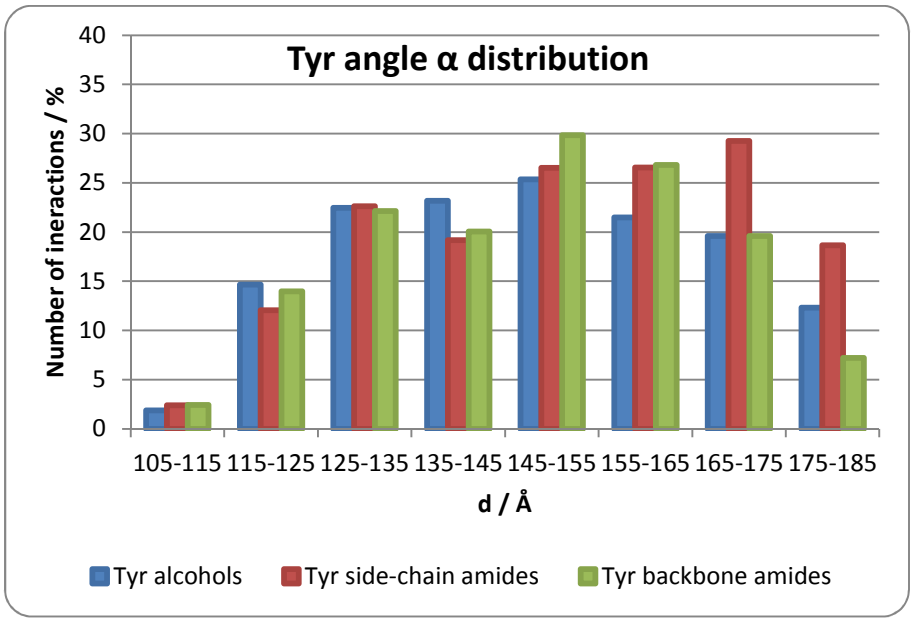
The distributions of the angle α indicate that the geometries of C–H/O interactions are more linear in the absence of the classical hydrogen bonds (Figures S3-S5). This is observed in all the systems with phenylalanine and in the systems tyrosine/backbone amide and tyrosine/side-chain amide. It indicates that the classical hydrogen bonds influence geometry of simultaneous C–H/O interactions, what can be anticipated based on strength of the interactions; namely, classical hydrogen bonds are stronger than C–H/O interactions.[24,25]

The most striking impact of the simultaneous classical hydrogen bonds of the acceptors is visible on the angle φ in the systems with tyrosine (Figure 6). In the system tyrosine/alcohol (Figure 6a), the presence of the classical hydrogen bonds favours C–H/O interactions at φ = 60° and 150° and decreases their probability around φ = 70°, 100° and 130°. In the system tyrosine/side-chain amide (Figure 6b) the presence of classical hydrogen bonds favours C–H/O interactions at φ = 150° and reduces the tendency for C–H/O interactions around φ = 70° and 120°. In case of tyrosine/backbone amide system (Figure 6c) the presence of classical hydrogen bonds favours C–H/O interactions at φ = 50° and 150°. The results show that presence of hydrogen bonds involving –OH substituent influence C–H/O frequency and geometry in peak at φ = 150°.

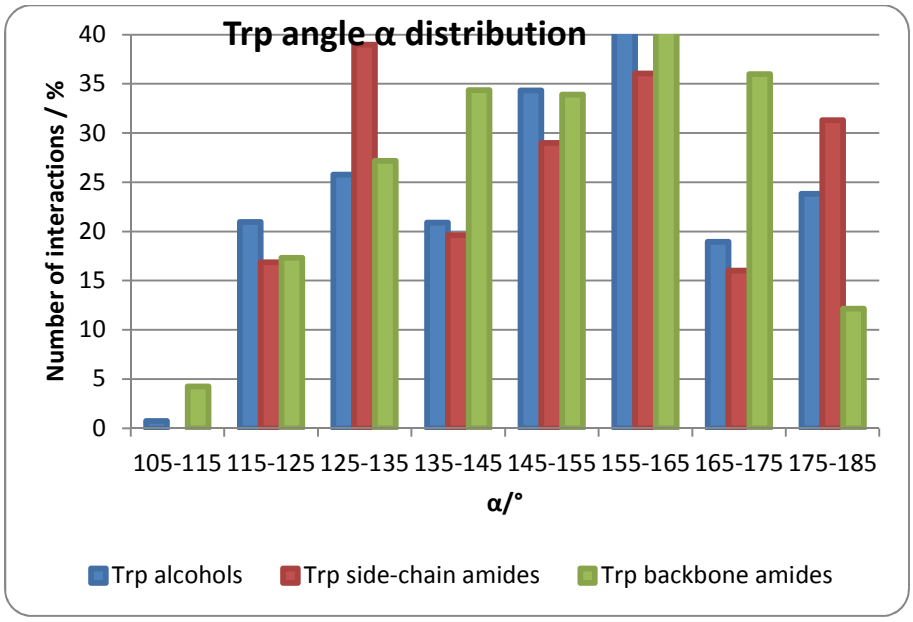
The presence of the hydrogen bonds causes a multifold increase in the frequencies of interactions around φ = 150° for all acceptors in the case of tyrosine (Figure 4). This tendency is higher for side-chain amide group than for alcohol or backbone amide group acceptor because the side-chain amide group possesses a stronger hydrogen bond acceptor, the nitrogen atom, as shown below in electrostatic potentials. In tyrosine the –OH substituent is located at φ = 180°, in *ortho*-position to the C–H bond at φ = 120°. An acceptor could form simultaneously interactions with this CH bond and a classical hydrogen bond with the –OH substituent, which explains the peaks at φ = 150°. The detailed inspection showed that in the peak at φ = 150° all simultaneous classical hydrogen bonds are formed with –OH of tyrosine. To represent this, one PDB structure containing simultaneous hydrogen bond with –OH group was shown in Figure 8.


Phe angle α distribution
Number of ineractions / %
40
35
30
25
20
15
10
5
0
105-115 115-125 125-135 135-145 145-155 155-165 165-175 175-185
d / Å
Phe alcohols
Phe side-chain amides
Phe backbone amides

a)

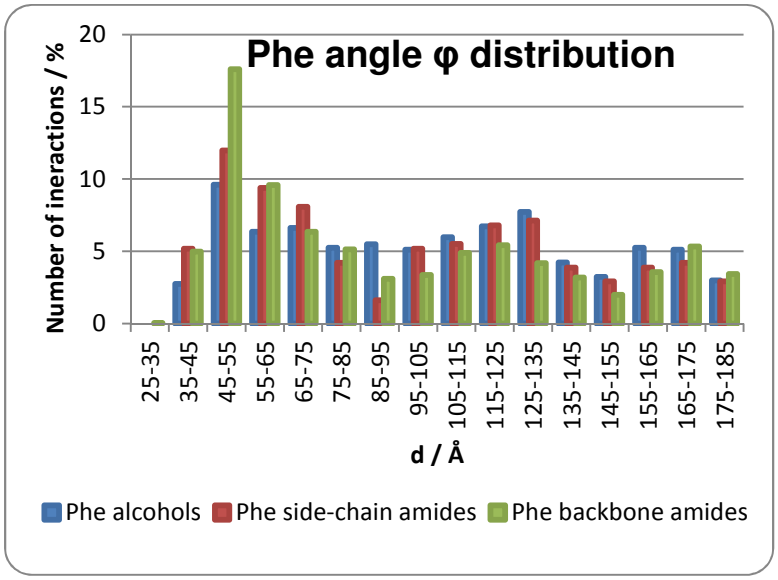

Tyr angle α distribution
Number of ineractions / %
40
35
30
25
20
15
10
5
0
105-115 115-125 125-135 135-145 145-155 155-165 165-175 175-185
d / Å
Tyr alcohols
Tyr side-chain amides
Tyr backbone amides

b)

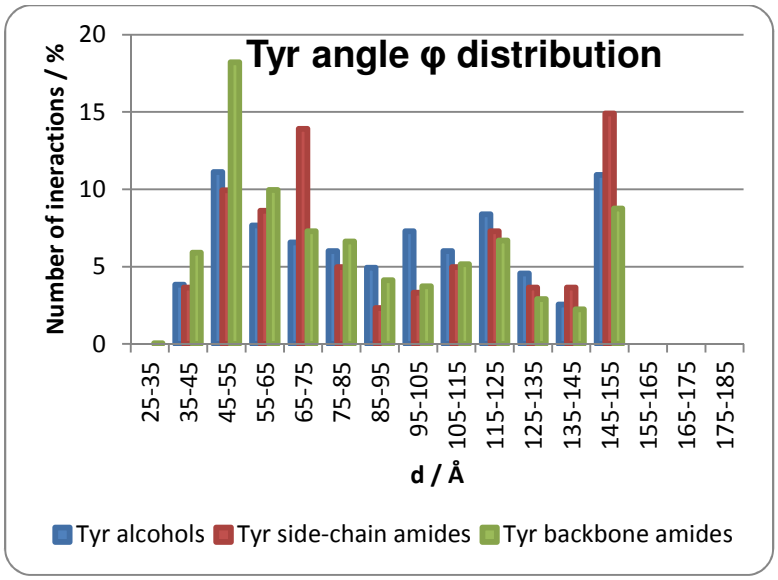


c)

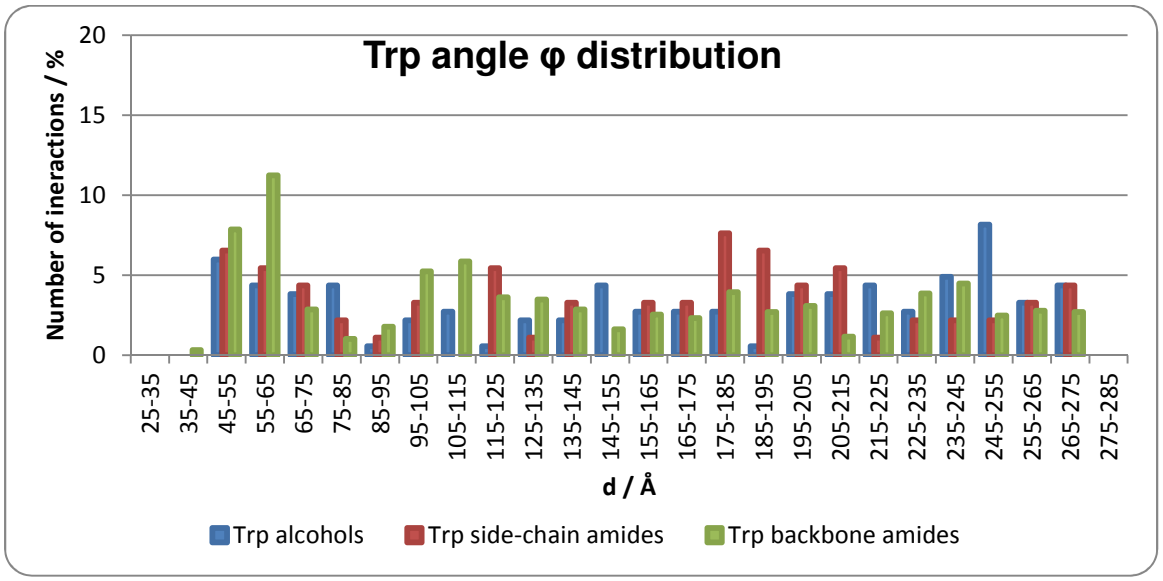
Figure 3. The corrected distributions of angle α for the three donors: (a) phenylalanine, (b) tyrosine and (c) tryptophan. The acceptors are represented in different colors: blue for the alcohol group, red for the side-chain amide group and green for the backbone amide group.



a)

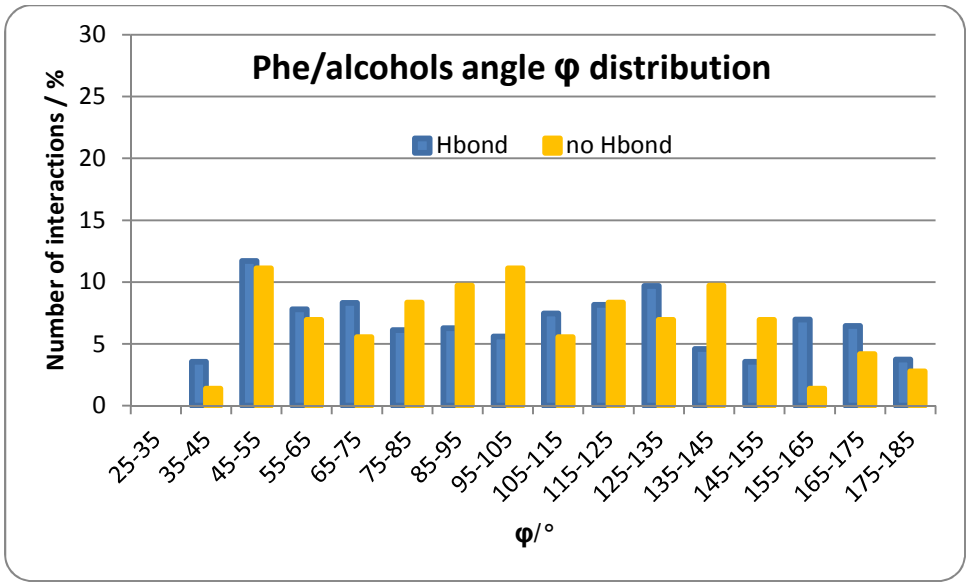


b)

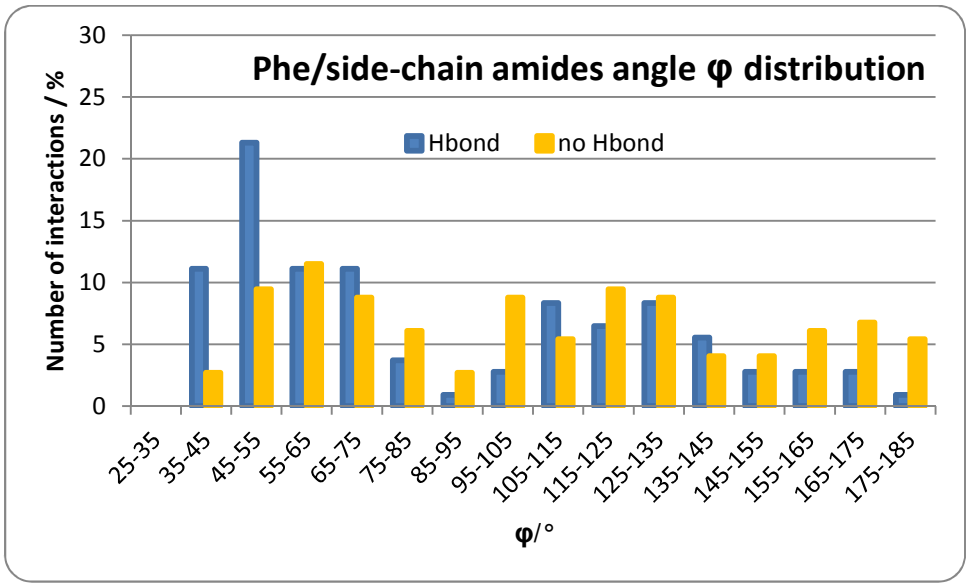


c)

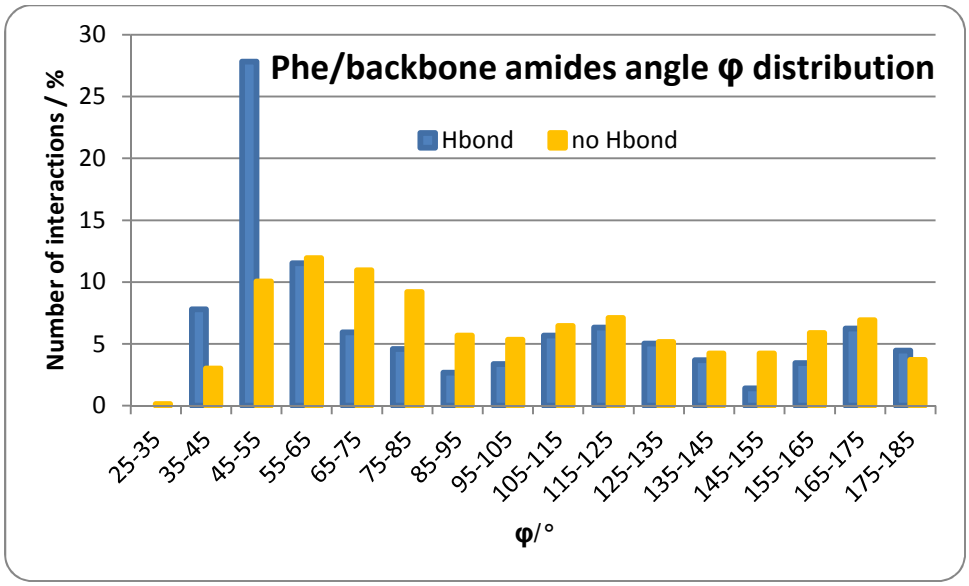
Figure 4. Distributions of the angle φ for the three donors: (a) phenylalanine, (b) tyrosine and (c) tryptophan. The acceptors are represented in different colours: blue for the alcohol group, red for the side-chain amide group and green for the backbone amide group.


Phe/alcohols angle φ distribution
Hbond
no Hbond
Number of interactions / %
30
25
20
15
10
5
0
25-35
35-45
45-55
55-65
65-75
75-85
85-95
95-105
105-115
115-125
125-135
135-145
145-155
155-165
165-175
175-185
φ/°

a)

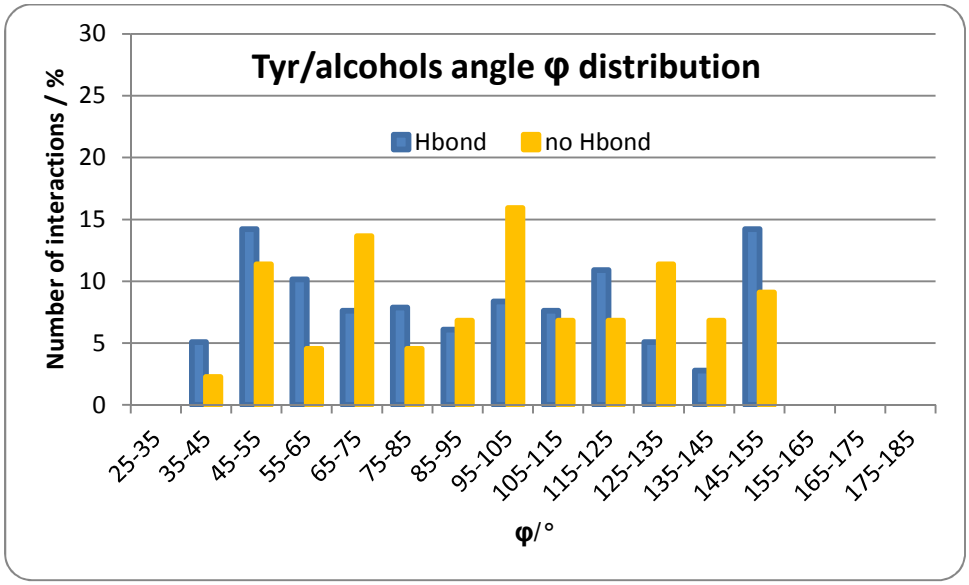

Phe/side-chain amides angle φ distribution
Hbond
no Hbond
Number of interactions / %
30
25
20
15
10
5
0
25-35
35-45
45-55
55-65
65-75
75-85
85-95
95-105
105-115
115-125
125-135
135-145
145-155
155-165
165-175
175-185
φ/°

b)

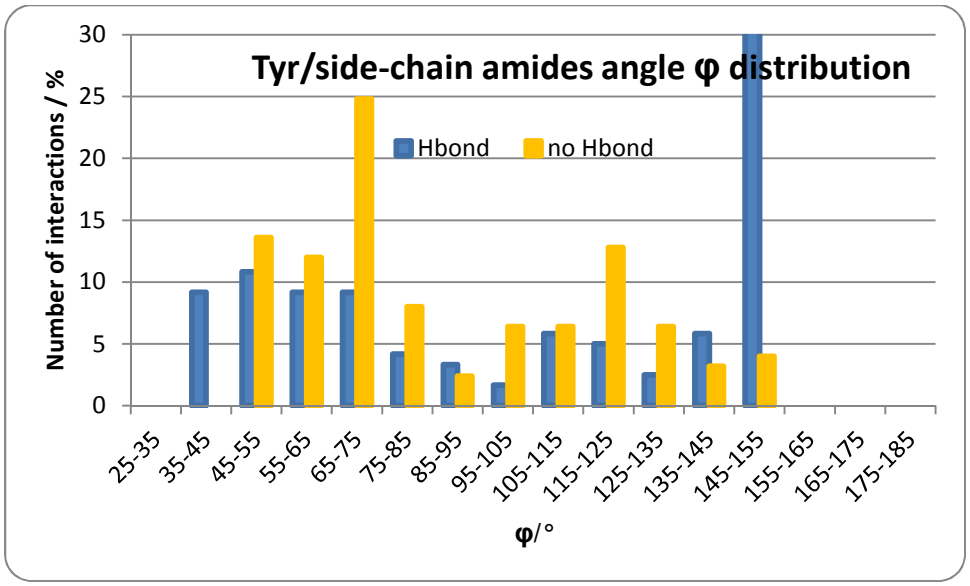


c)

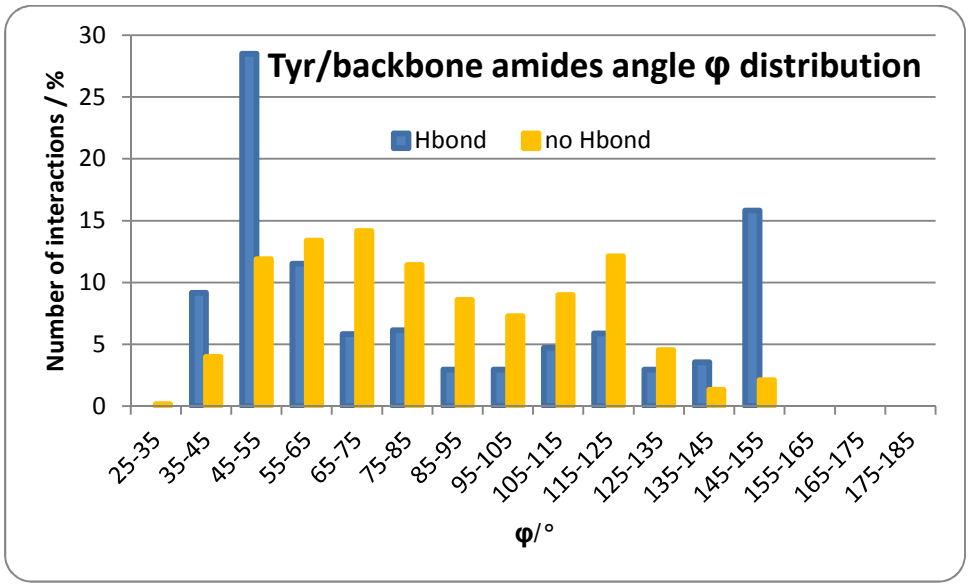
Figure 5. The distributions of angle φ for C–H/O interactions with (blue) and without (yellow) simultaneous classical hydrogen bonds for phenylalanine with different acceptors: (a) alcohol group, (b) side-chain amide group and (c) backbone amide group.



a)



b)



c)

Figure 6. The distributions of angle φ for C–H/O interactions with (blue) and without (yellow) simultaneous classical hydrogen bonds for tyrosine with different acceptors: (a) alcohol group, (b) side-chain amide group and (c) backbone amide group.

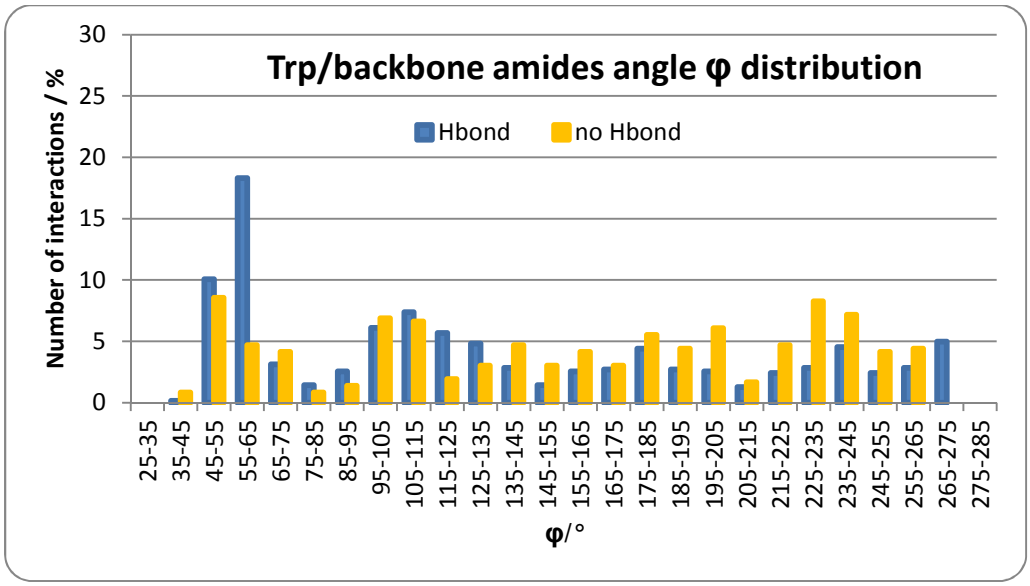


Figure 7. The distributions of angle φ for C–H/O interactions with (blue) and without (yellow) simultaneous classical hydrogen bonds for tryptophan with backbone amide group. The systems tryptophan/alcohol and tryptophan/side-chain amide have small number of interactions making statistics unreliable.

The C–H/O interactions involving backbone amide group as acceptor have a larger tendency towards angle φ values of 50° and 60° in comparison to the other acceptors (Figure 4). The analysis of backbone amide group interactions with and without simultaneous classical hydrogen bonds (Figures 5c, 6c, 7) showed that the pronounced peak for backbone amide groups is the consequence of the presence of simultaneous classical hydrogen bonds. Namely, in this region hydrogen bonds could be formed with a donor backbone, i.e. with the –NH or –C=O groups.

Figure 9 presents the example of a C–H/O interaction where simultaneous hydrogen bond is formed with –NH group of the donor, while Figure S6 presents the example of simultaneous hydrogen bond with –NH group of the neighboring residue to the donor.

It has been analysed whether backbone amide groups have a higher tendency to form hydrogen bonds with donor backbone, and backbone atoms of its neighbouring residues, in comparison to the side-chain amides and alcohols (Figure S6). For every peak in the region around the angle φ values of 50° and 60° (Figures 5, 6, 7, blue graphs), a number of hydrogen bonds with hydrogen donors from backbone has been calculated. For each donor, the most relevant region has been chosen, i.e. a region of highest differences between the acceptors. The Table S1 shows fraction of simultaneous classical hydrogen bonds with backbone. The results show the highest tendency

of the backbone amide acceptor to form simultaneous hydrogen bonds with donor atoms from backbone (90.3%), while tendencies are lower for side-chain amides (81.1%) and alcohols (34.8%). These results can be elucidated by the electrostatic potentials of donors and acceptors that are presented in the next section.

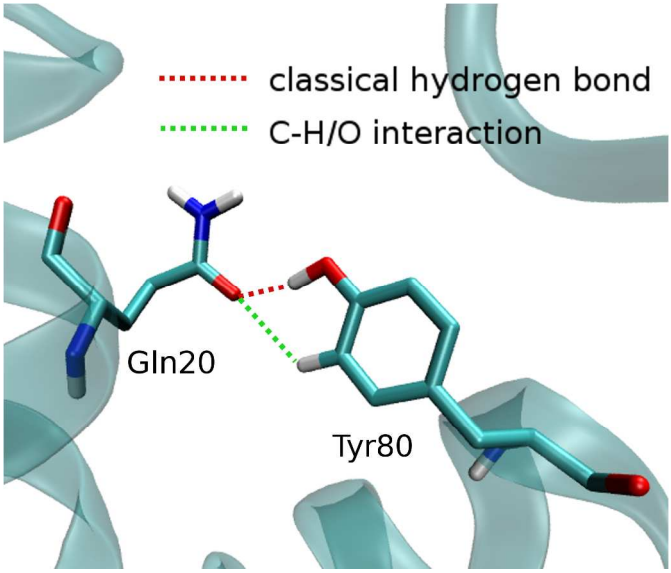


Figure 8. The structure of the bacterocin transport accessory protein from streptococcus pneumoniae, PDB ID 1ZMA,[48] showing simultaneous hydrogen bond and C–H/O interaction of tyrosine residue with the side-chain amide acceptor. The acceptor oxygen atom (in the amino acid Gln20) is positioned at the angle φ = 149.93°.

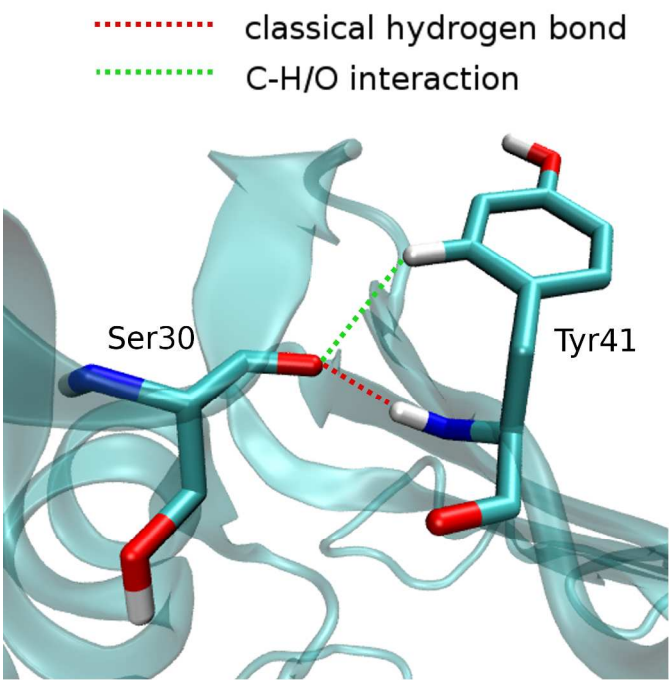


Figure 9. A representative structure presenting C-H/O interaction, in the region around φ = 50°, with simultaneous classical hydrogen bond, between tyrosine and the backbone amide acceptor. This is the structure of the beta-lactamase inhibitory protein-like protein (PDB ID 3GMX).[49] The acceptor oxygen atom (in the amino acid Ser30) is positioned at the angle φ = 50.00°.

Quantum chemical calculations of electrostatic potential maps

To understand the results on geometries of C-H/O interactions in protein structures, electrostatic potential maps were calculated. The calculations were done on model systems for acceptors and for amino acid molecules as donors (Figures 10 and 11). Ethanol, acetamide and N-methylacetamide were chosen as model systems for the hydrogen atom acceptors; alcohol, side-chain amide, and backbone amide groups, respectively. The model donor groups were phenylalanine, tyrosine and tryptophan. The areas of strong positive potential are coloured in red, the areas of strong negative potential are coloured in blue. The points in which electrostatic potential exhibits maximal or minimal values (maxima or minima) are displayed as black and blue dots on the surface of electrostatic potential maps.

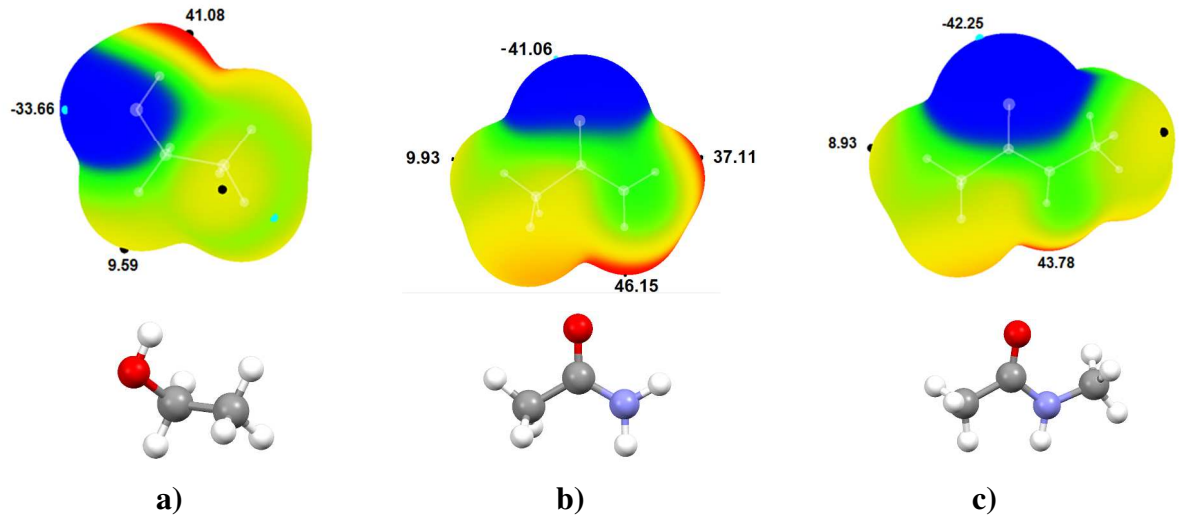


Figure 10. Computed MP2/cc-PVTZ electrostatic potentials on the 0.001 au surface of: (a) ethanol, (b) acetamide and (c) N-methylacetamide. Colour ranges, in kcal/mol, are: red, greater than 22.40; yellow, from 0.00 to 22.40; green, from -15.00 to 0.00; blue, more negative than -15.00.

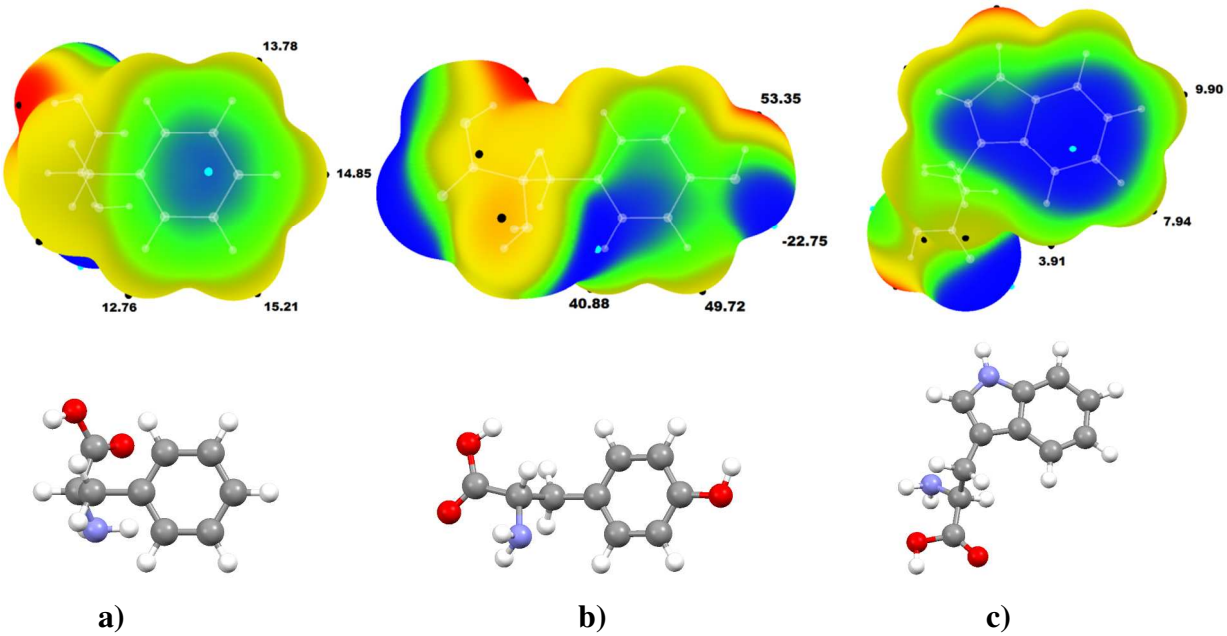


Figure 11. Computed MP2/cc-PVTZ electrostatic potentials on the 0.001 au surface of: (a) phenylalanine, (b) tyrosine and (c) tryptophan. Colour ranges, in kcal/mol, are: red, greater than 35.96; yellow, from 0.00 to 35.96; green, from -12.05 to 0.00; blue, more negative than -12.05.

As can be anticipated the most positive areas are around hydrogen atoms attached to electronegative atom like oxygen and nitrogen (red areas in Figures 10a, 10b, 11b and 11c), while most negative areas are around oxygen atoms of acceptors and donors and above aromatic rings of donors (blue areas in Figures 10 and 11).

The Figure 11 represents the electrostatic maps for the C–H/O donors, phenylalanine, tyrosine and tryptophan. The hydrogen atoms on aromatic rings in all donor model systems have positive electrostatic potential. The values of electrostatic potential at maxima are very similar for all hydrogen atoms on aromatic rings what explains their similar ability to be involved in C–H/O interactions.

Among the acceptors, amide groups probably form stronger hydrogen bonds because of their more negative potential at the oxygen atom in comparison to the alcohol group (Figure 10). The critical electrostatic points for side-chain amide, backbone amide and alcohol oxygen atoms are -41.06, -42.25 and -33.66 kcal/mol, respectively. The alcohol group has a small area of negative potential in comparison to the other acceptors. These properties of acceptors are closely related to their tendency for bifurcated interactions and to the number of the C-H/O interactions in presence of simultaneous hydrogen bonds (Figures 5-7).

In the case of alcohol group as hydrogen acceptor, the presence of the simultaneous classical hydrogen bonds influences significantly C-H/O interaction only in the system Tyr/alcohol; the number of C–H/O interactions at $\varphi = 60°$ is increased (Figure 6a). The presence of the classical hydrogen bonds does not influence C-H/O interaction significantly in the system Phe/alcohol (Figure 5a), while for the system Trp/alcohol, small number of interactions makes statistics unreliable. The fact that presence of classical hydrogen bond does not have such a strong influence on interactions of alcohol group is a consequence of small area of negative potential (Figure 10) that does not favour bifurcated interactions, as already mentioned.

In the case of side-chain amide group as hydrogen acceptor, the number of C–H/O interactions in presence of simultaneous, classical hydrogen bonds is increased at $\varphi = 50°$ only in the system Phe/side-chain amide (Figure 5b). In this region there is possibility for simultaneous hydrogen bonds with the backbone atoms of the phenylalanine donor. In Tyr/side-chain amide system, the number of C–H/O interactions in presence of simultaneous hydrogen bonds is increased at $\varphi = 150°$, while at the position $\varphi = 50°$ the influence cannot be observed (Figure 6b). Increased number of interactions at $\varphi = 150°$ can be a consequence of strong attraction of very positive

potential of Tyr in the area around φ = 150° (Figure 11b), and very negative potential of side-chain amide (Figure 10b). For the system Trp/side-chain amide, small number of interactions makes statistics unreliable.

When backbone amide group is hydrogen acceptor, the presence of the simultaneous classical hydrogen bonds favours C–H/O interactions at φ = 50° or 60° in all three systems Phe/backbone amide, Tyr/backbone amide and Trp/backbone amide (Figures 5c, 6c, 7). In Tyr/backbone amide system, the presence of the classical hydrogen bonds also favours interactions at φ = 150° (Figure 6c). The higher propensity of backbone amide in comparison to the side-chain amide group towards forming bifurcated bonds, one C-H/O interaction and the other classical hydrogen bond, might be caused by the presence of the positive potential (37.11 kcal/mol) near the oxygen in the model representing side-chain amide acceptor (Figure 10b).

CONCLUSIONS

The analysis of the distribution of the C–H–O angle in the crystal structures from the Protein Data Bank (PDB) indicates no preference for linear C–H/O interactions between aromatic donors and acceptors (alcohol, side-chain amide and backbone amide groups) in protein structures.

Although there is no tendency for linear C–H/O interactions, there is no significant number of bifurcated C–H/O interactions, while in previous work it was shown that benzene in crystal structures forms significant number of bifurcated interactions.[30] This might be a consequence of difference in conformation flexibility and availability of donors and acceptors in proteins and in crystals.

The analyses also indicate an influence of simultaneous classical hydrogen bonds. The influence is particularly observed in case of tyrosine. Namely, the –OH substituent of tyrosine plays an important role by simultaneously forming a hydrogen bond with the C–H/O interaction in *ortho*-position to the –OH substituent.

There are significant differences among the acceptors. The most remarkable is the increase of C–H/O frequency involving backbone and side-chain amide groups in the presence of the simultaneous hydrogen bonds as a consequence of the electrostatic potential near oxygen atom favouring simultaneous interactions.

These investigations could help in future C–H/O interactions studies in proteins or other proteic systems. Since our results show that linear C–H/O interactions do not have high

frequency in the protein structures, in the future studies one should consider non-linear C–H/O interactions as important contacts.

Supporting Information. Distance d(H···O) distributions (Figure S1). Noncorrected angle α distributions (Figure S2). C-H/O angle α distributions with and without classical hydrogen bond (Figures S3-S5). C-H/O interactions with simultaneous classical hydrogen bonds of the acceptors (Figure S6, Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

***Corresponding author:**

Dr Snežana D. Zarić
Department of Chemistry, University of Belgrade
Studentski trg 12-16
11000 Belgrade, Serbia
E-mail: szaric@chem.bg.ac.rs

Acknowledgements

This work was supported by the Serbian Ministry of Education, Science and Technological Development (grant 172065). We thank Prof. Peter Politzer and Dr. Jane S. Murray from the University of New Orleans for support.

REFERENCES

(1) Steiner, T.; Saenger, W. *Acta Crystallogr. Sect. B-Struct. Sci.* **1994**, *50*, 348–357.

(2) Mandel-Gutfreund, Y.; Margalit, H.; Jernigan, H.; Zhurkin, V. B. *J. Mol. Biol.* **1998**, *277*, 1129–1140.

(3) Brik, A.; Alexandratos, J.; Lin, Y.–C.; Elder, J. H.; Olson, A. J.; Wlodawer, A.; Goodsell, D. S.; Wong, C.–H. *ChemBioChem*, **2005**, *6*, 1167–1169.

(4) Jiang, L.; Lai, L. *J. Biol. Chem.* **2002**, *277*, 37732–37740.

(5) Derewenda, Z. S.; Lee, L.; Derewenda, U. *J. Mol. Biol.* **1995**, *252*, 248–262.

(6) Fabiola, G. F.; Krishnaswamy, S.; Nagarajan, V.; Pattabhi, V. *Acta Crystallogr.,Sect. D: Biol. Crystallogr.* **1997**, *53*, 316–320.

(7) Weiss, M. S. *Trends Biochem. Sci.* **2001**, *26*, 521–523.

(8) Castellano, R. K. *Curr. Org. Chem.* **2004**, *8*, 845–865.

(9) Scheiner, S.; Kar, T. *J. Phys. Chem. B*, **2005**, *109*, 3681–3689.

(10) Chakrabarti, P.; Bhattacharyya, R. *Prog. Biophys. Mol. Biol.* **2007**, *95*, 83–137.

(11) Scheiner, S. *J. Phys. Chem. B*, **2006**, *110*, 18670–18679.

(12) Park, H.; Yoon, J.; Seok, C. *J. Phys. Chem. B*, **2008**, *112(3)*, 1041–1048.

(13) Manikandan, K.; Ramakumar, S. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 768–781.

(14) Vaz, P. D.; Nolasco, M. M.; Gil, F. P. S. C.; Ribeiro-Claro, P. J. A.; Tomkinson, J. *Chem. Eur. J.* **2010**, *16*, 9010–9017.

(15) Sigalov, M. V.; Doronina, E. P.; Sidorkin, V. F. *J. Phys. Chem. A*, **2012**, *116*, 7718–7725.

(16) Baures, P. W.; Beatty, A. M.; Dhanasekaran, M.; Helfrich, B. A.; Pérez-Segarra, W.; Desper, J. *J. Am. Chem. Soc.* **2002**, *124*, 11315–11323.

(17) Scheiner, S.; Kar, T. *J. Phys. Chem. A*, **2008**, *112*, 11854–11860.

(18) Compaan, K.; Vergenz, R.; Schleyer, P. V. R.; Arreguin, I. *Int. J. Quantum Chem.* **2008**, *108*, 2914–2923.

(19) Joseph, J.; Jemmis, E. D. *J. Am. Chem. Soc.* **2007**, *129*, 4620–4632.

(20) Scheiner, S.; Kar, T.; Pattanayak, J. *J. Am. Chem. Soc.* **2002**, *124*, 13257–13264.

(21) Gu, Y.; Kar, T.; Scheiner, S. *J. Am. Chem. Soc.* **1999**, *121*, 9411–9422.

(22) Yang, H.; Wong, M. W. *J. Am. Chem. Soc.* **2013**, *135*, 5808–5818.

(23) Dumitru, F.; Legrand, Y. M.; Barboiu, M.; van der Lee, A. *Acta Crystallogr. Sect. B-Struct. Sci.* **2013**, *69*, 43–54.

(24) Steiner, T. *Chem. Commun.* **1997**, 727–734.

(25) Steiner, T.; Desiraju, G. R. *Chem. Commun.* **1998**, 891–892.

(26) Nanda, V.; Schmiedekamp, A. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 489–497.

(27) Niebel, C.; Lokshin, V.; Sigalov, M.; Krief, P.; Khodorkovsky, V. *Eur. J. Org. Chem.* **2008**, 3689–3699.

(28) Venkatesan, V.; Fujii, A.; Mikami, N. *Chem. Phys. Lett.* **2005**, *409*, 57–62.

(29) Dragelj, J. Lj.; Janjić, G. V.; Veljković, D. Ž.; Zarić, S. D. *CrystEngComm*, **2013**, *15*, 10481–10489.

(30) Veljković, D. Ž.; Janjić, G. V.; Zarić, S. D. *CrystEngComm*, **2011**, *13*, 5005–5010.

(31) Veljković, D. Ž.; Medaković, V. B.; Andrić, J. M.; Zarić, S. D. *CrystEngComm*, **2014**, *16*, 10089–10096.

(32) Malenov, D. P.; Janjić, G. V.; Veljković, D. Ž.; Zarić, S. D. *Comput. Theor. Chem.* **2013**, *1018*, 59–65.

(33) Meyer, T.; Knapp, E. W. *Proteins: Struct., Funct., Bioinf.* **2014**, *82(5)*, 744–751.

(34) Read, R. J.; Adams, P. D.; Arendall, W. B. 3rd; Brunger, A. T.; Emsley, P.; Joosten, R. P.; Kleywegt, G. J.; Krissinel, E. B.; Lütteke, T.; Otwinowski, Z.; Perrakis, A.; Richardson, J. S.; Sheffler, W. H.; Smith, J. L.; Tickle, I. J.; Vriend, G.; Zwart, P. H. *Structur*e, **2011**, *19*, 1395–1412.

(35) Chen, V. B.; Arendall, W. B. 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S., Richardson, D. C. *Acta Crystallogr D Biol Crystallogr.* **2010**, *66*, 12–21.

(36) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr D Biol Crystallogr.* **2010**, *66*, 213–221.

(37) McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *238*, 777–793.

(38) MacKerell Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wirkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B*, **1998**, *102*, 3586–3616.

(39) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

(40) Ostojić, B. D.; Janjić, G. V.; Zarić, S. D. *Chem. Commun.* **2008**, 6546–6548.

(41) Janjić, G. V.; Veljković, D. Ž.; Zarić, S. D. *Cryst. Growth Des.* **2011**, *11*, 2680–2683.

(42) Janjić, G. V.; Malkov, S. N.; Živković, M. V.; Zarić, S. D. *Phys. Chem. Chem. Phys.* **2014**, *16*, 23549–23553.

(43) Kroon, J.; Kanters, J. A. *Nature*, **1974**, *248*, 667–669.

(44) Gaussian 09, Revision **D.01**, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.;

Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

(45) Bulat, F. A.; Toro–Labbe, A. WFA: A suite of programs to analyse wavefunctions, *unpublished*.

(46) Bulat, F. A.; Toro–Labbé, A.; Brinck, T.; Murray, J. S.; Politzer, P. *J. Mol. Model.* **2010**, *16*, 1679–1691.

(47) Ramanahan, K.; Shanthi, V.; Sethumadhavan, R. *Int. J. Pharm. Pharm. Sci.* **2011**, *3*, 324–329.

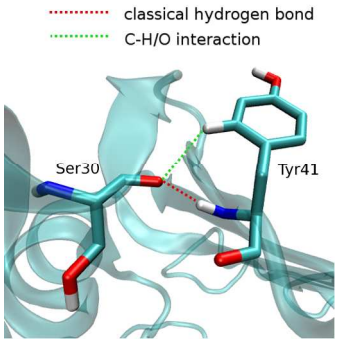
(48) Kim, Y.; Hatzos, C.; Abdullah, J.; Collart, F.; Joachimiak, A. The Crystal Structure of the Bacterocin Transport Accessory Protein from Streptococcus pneumoniae, *unpublished*.

(49) Gretes, M.; Lim, D.C.; de Castro, L.; Jensen, S. E.; Kang, S. G.; Lee, K. J.; Strynadka, N. C. *J. Mol. Biol.* **2009**, 389, 289–305.

For Table of Contents Use Only

C–H/O interactions of aromatic CH donors within proteins: a crystallographic study

Jovan Lj. Dragelj, Ivana M. Stanković, Dragana M. Božinovski, Tim Meyer, Dušan Ž. Veljković, Vesna B. Medaković, Ernst-Walter Knapp and Snežana D. Zarić[*]



Analysis of the distribution of the C–H–O angle in the crystal structures from the Protein Data Bank (PDB) did not show preference for linear C–H/O interactions between aromatic donors and different acceptors in protein structures.