



## C-H...O INTERACTIONS STABILIZE THE STRUCTURE OF THE THERAPEUTIC PROTEINS: A COMPUTATIONAL STUDY

K. RAMANATHAN, V. SHANTHI AND RAO SETHUMADHAVAN\*

School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India.

Email: rsethumadhavan@vit.ac.in (R. Sethumadhavan)

Received: 22 Dec 2010, Revised and Accepted: 20 Jan 2011

### ABSTRACT

Stability of therapeutic proteins contributes significantly towards the discovery of new drug molecules. We have investigated the role of C-H...O interactions on the structural stability of therapeutic proteins. In our study we have analyzed the interactions in a set of 48 therapeutic proteins. A total of 1099 interactions were observed with an average of 23 C-H...O interactions per therapeutic protein. The analysis of the position of interacting residues in therapeutic proteins showed that C-H...O interactions are mainly formed by long range contacts. Phe has the highest occurrence among the donor residues while Glu has the highest occurrence among the acceptor residues. The C-H...O interaction residues involved in these proteins were found highly conserved with more than 50% of the interacting residues either as a donor or as an acceptor having conservation score of  $\geq 6$ . The results obtained in the present study should be useful in understanding the contribution of C-H...O interactions towards the structural stability and specificity of therapeutic proteins.

**Keywords:** C-H...O Interactions, Therapeutic proteins, Interaction range, Conservation score.

### INTRODUCTION

The importance of therapeutic proteins has grown rapidly since the emergence of the biotechnology industry more than 30 years ago. There are approximately 140 therapeutic proteins approved in the United States and Europe and an additional 500 in clinical trials<sup>1</sup>, with an even large number in preclinical development. In recent years, the number of recombinant proteins used for therapeutic applications has increased dramatically. This increasing trend has driven the development of a variety of improvements in protein expression and stability analysis. Therapeutic proteins had several advantages over small molecules (e.g. higher affinity/specificity to target and lower toxicity profiles) and antibodies (e.g. room temperature storage and better tissue penetration because of their smaller size). However, the widespread use of therapeutic proteins is hampered by their rapid elimination from the circulation. Despite being better than small chemical drugs, the overall stability of therapeutic proteins is still very low. Because of low stability and selectivity, a high dosage of peptides is essential in order to elicit the proper therapeutic effects which is often not economical and is beyond the clinically acceptable level. Therefore stability is a major factor to be considered for choosing the potential therapeutic proteins<sup>2</sup>.

Although the three dimensional structure of a bio macromolecules determined by its covalent structure, *i.e.* its amino acid sequence, the forces responsible for the folding and stabilization of the structure are mainly non-covalent in nature. Covalent interactions play a key role in chemistry while non-covalent interactions play a decisive role in bio disciplines. These non-covalent interactions include hydrogen bonds (H-bonds), electrostatic interactions and the so-called hydrophobic effect. Many excellent and theoretical perspectives of hydrogen bond in protein have appeared in the literature<sup>3,4,5</sup>. In their standard incarnation hydrogen bonds result from the approach of a proton donor molecule toward an acceptor, forming a bridge of the sort A-H...B. The donor A atom is thought to be very electronegative, e.g., O or N, as is the acceptor atom B, which must also contain at least one lone pair of electrons by which the bridge is formed. Another type of hydrogen bond that has undergone a resurgence of attention is that characterized by a CH donor, in place of the more common OH or NH groups. When first proposed many years ago<sup>6,7</sup>, there was some resistance due to the low electro negativity of C which was presumed to make it a weak proton donor. However, Support was later added to this idea on the

basis of IR data<sup>8,9,10</sup> and the geometry of molecular complexes in the gas phase<sup>11,12,13</sup> and in crystalline environment<sup>14,15</sup>. This type of interaction called C-H...O interactions. This C-H...O hydrogen bond systems share numerous features with the more traditional hydrogen bonds, such as geometric preference, NMR chemical shifts, and electron density patterns. It is only now gaining wide acceptance as a genuine hydrogen bond<sup>16,17</sup>. The importance of this C-H...O interactions is not limited to small molecules but has been extended to biological system. Possibilities of C-H...O interactions have also been investigated in nucleic acid and carbohydrate structures<sup>18</sup> and are involved in protein-nucleic acid interactions<sup>19</sup>. These interactions are also observed at protein-protein interfaces<sup>20,21</sup> and could be relevant in the ligand-binding sites<sup>22</sup>, where the ligands are observed to use their stronger hydrogen bond capabilities for use with the protein residues, leaving the weaker C-H...O interactions to bind water<sup>23</sup>. Moreover, these interactions do not necessarily disappear when the solid loses its regular structure, as C-H...O interactions can persist into the liquid phase<sup>24,25</sup>. Such a type of continuous upshot prompted us to study the C-H...O interactions in the structural stability of therapeutic proteins.

Recently, we published our results on the cation- $\pi$  and C-H... $\pi$  interactions in the structural stability of therapeutic proteins<sup>26,27</sup>. However, there has been no analysis of C-H...O interactions in the structural stability of therapeutic proteins. It is the objective of the present paper to carry out just this sort of systematic analysis of the nonconventional C-H...O interactions in which an therapeutic proteins structures were stabilized. It is noteworthy to mention here all the 48 therapeutic proteins in our data set showed significant number of C-H...O interactions and hence we emphasize that this investigation is very significant in the sense that, C-H...O interactions in therapeutic proteins do play a major role in structural stability. We strongly hope that the ingenuity and success of the discovery efforts discussed above bode well for the future prospects of finding new therapeutics which could result into massive reductions in therapeutics development time.

### MATERIALS AND METHODS

#### Data set

We have considered a set of 49 therapeutic proteins from the Protein Data Bank (PDB)<sup>28</sup> for our investigation. The details of which are given in Table 1. According to the structural classification of proteins, 42% of the proteins come under alpha group, 29%

comes under beta 11% comes under alpha and beta and remaining 18% comes under small proteins in the therapeutic protein data set.

#### C-H...O Interactions

C-H...O interactions were identified using the program available for this purpose called HBAT<sup>29</sup>. The C-H...O interactions considered here were between all possible donor C-H groups in the therapeutic proteins structures (C<sub>α</sub>-H, C<sub>β</sub>-H and C<sub>γ</sub>-H) and oxygen containing proton acceptor molecule. The oxygen atoms in proteins are of the hydroxyl, carbonyl and carboxyl type. In terms of their electro negativity, this increases in the order O-H < C=O < C-O<sup>-</sup>. The distances

from C and H to the main chain carbonyl O is  $\leq 3.8 \text{ \AA}$  and  $\leq 3.3 \text{ \AA}$  and the angles CH...O and H...O=C is  $\geq 120^\circ$  and  $\geq 90^\circ$  respectively.

The C-H...O interaction types are represented by a two-letter code in which the first letter indicates the donor atom and the second the acceptor: M and S represent the main-chain and side-chain atom respectively. We classified the C-H...O interactions into four types namely, main-chain to main-chain C-H...O interactions (MM-C-H...O), main-chain to side-chain C-H...O interactions (MS-C-H...O), side-chain to main-chain C-H...O interactions (SM-C-H...O) and side-chain to side-chain C-H...O interactions (SS-C-H...O)<sup>29</sup>. The position and geometry is adapted from earlier work of Babu<sup>30</sup>.

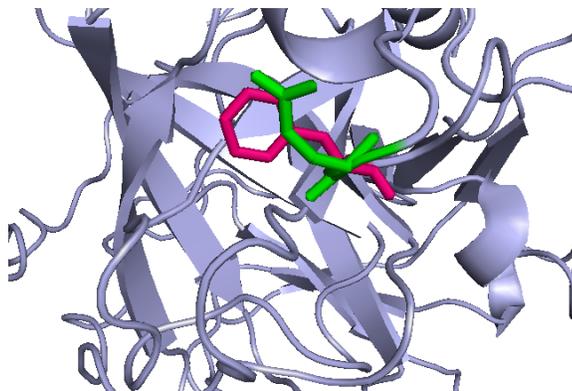


Fig. 1: PyMOL view of C-H...O interaction between Phe (715) (Hot pink) and Glu (714) (Green) in therapeutic proteins 1BML.

#### Sequential distance

The definition of short, medium and long-range interactions in a protein is based on (i) the amino acid residues and (ii) their respective locations in the sequence. For each residue, the sequential distance of surrounding residues (within a sphere of 8 Å radius) was analyzed<sup>31,32,33</sup>. The residues that are within a distance

of two residues are considered to contribute to short-range interactions, whereas, those within a distance of  $\pm 3$  or  $\pm 4$  residues contribute to medium-range and those more than four residues apart contribute to long-range interactions. This classification enables us to evaluate the contribution of short range, medium-range and long-range contacts in the formation of C-H...O interactions<sup>34</sup>.

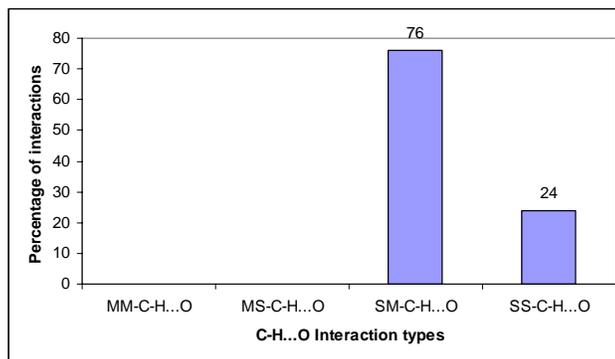


Fig. 2: CH...O interaction types and contribution in therapeutic proteins

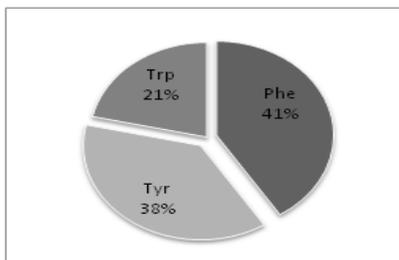


Fig. 3: Donor amino acids contribution to the stability

Table 1: The PDB codes and total number of interactions in therapeutic proteins data set.

S. No	PDB Code	TNI	S. No	PDB code	TNI	S. No	PDB code	TNI	S.No	PDB code	TNI
1	1BML	37	13	1I4Z	33	25	1QQR	14	37	1Z92	10
2	1BMP	15	14	1M4C	18	26	1QRZ	40	38	2B5I	13
3	1C4P	14	15	1M47	11	27	1QVN	11	39	2ERJ	20
4	1CA9	20	16	1M48	15	28	1R46	60	40	2GMF	8
5	1CD9	8	17	1M49	14	29	1R47	67	41	2GOO	14
6	1CN4	37	18	1N8Y	53	30	1REU	13	42	2H62	15
7	1DDJ	41	19	1N8Z	41	31	1REW	12	43	2H64	13
8	1EER	11	20	1NBP	13	32	1RHG	6	44	2IWG	23
9	1ES7	11	21	1PGG	78	33	1RJX	23	45	2OSL	31
10	1GNC	6	22	1PGR	6	34	1TPG	8	46	1YYG	60
11	1HTZ	27	23	1PW6	11	35	1WAQ	14	47	3BMP	16
12	1I4D	28	24	1PY2	12	36	1YY8	35	48	3INK	14

TNI: Total Number of Interactions

### Conservation Score

We computed the conservation score of C-H...O interacting amino acid residues in each therapeutic protein structure using the ConSurf program<sup>35</sup>. This program computes the conservation based on the comparison of the sequence of a PDB chain with the proteins deposited in Swiss-Prot<sup>36</sup> and finds the ones that are homologous to the PDB sequence. The number of PSI-BLAST iterations and the *E*-value cutoff used in all similarity searches were 1 and 0.001, respectively. All the sequences that are evolutionarily related with each one of the proteins in the data set were used in the subsequent multiple alignments. Based on these protein sequence alignments the residues are classified into nine categories from highly variable to highly conserved. Residues with a score of 1 are considered highly variable and residues with a score of 9 are highly conserved.

## RESULTS

### C-H...O Interactions

The C-H...O interactions were calculated using the program HBAT<sup>29</sup> and total number of interactions in each therapeutic proteins listed in Table 1. The selected group of therapeutic proteins show a total of 1099 interactions, there was an average of 23 C-H...O interactions per therapeutic protein. The PyMOL view of C-H...O interaction in therapeutic proteins, PDB Code 1BML were shown in Figure 1. Figure 2 shows the various types of C-H...O interactions in therapeutic proteins data set. It shows that, SM-C-H...O interactions are the predominant type of interaction in the therapeutic protein data set. The preferences of amino acid residues that are involved in C-H...O interactions were analyzed and the results are presented in Fig. 3 and Fig. 4. We observed that Phe has the highest occurrence,

around 40.90%, followed by Tyr, around 37.62% whereas minimum occurrence, only 21.47%, was observed for Trp. In acceptor residues, Glu has highest occurrence with 13.11% whereas Trp has minimum occurrence with 0.48%.

### Distribution of interatomic distances

Another important issue is related to interatomic distances. The currently accepted van der Waals' radii for an oxygen atom (1.5 Å) and a carbon bonded hydrogen (1.2 Å) were taken into account, statistical analysis showed that an average, the H...O hydrogen bond distance is 0.3 Å shorter than the sum of the atomic van der Waals' radii. The H...O and C...O interatomic distances studied in the therapeutic proteins data set shown in Fig. 5 and 6 Briefly, these parameters were: the distance C...O, representing the donor-acceptor distance; the distance H...O, or the hydrogen bond distance; the crystallographic refinement programs treat close C...O distances as repulsive, and as a consequence, the distribution analyzed in our study may be effectively biased. The linearity of CH...O contacts is a function of the C...O distances: the more linear the hydrogen bond, the longer the donor acceptor distance, even though the bond energy may change only slightly. Finally, even though 67 percentage of the H...O distance should ideally be less than 2.8 Å, Coulombic attraction is a long range force inversely proportional to the interatomic distance; therefore, longer distances may still represent significant cohesive interactions. Nonetheless, All cohesive interatomic interactions, whatever their specific nature, contribute to the overall stability of any macromolecule. However, it is interesting to note that, the maximum interactions observed in the range of 2.4-2.8Å is identical to the earlier reports given by Taylor and Kennard<sup>37</sup>.

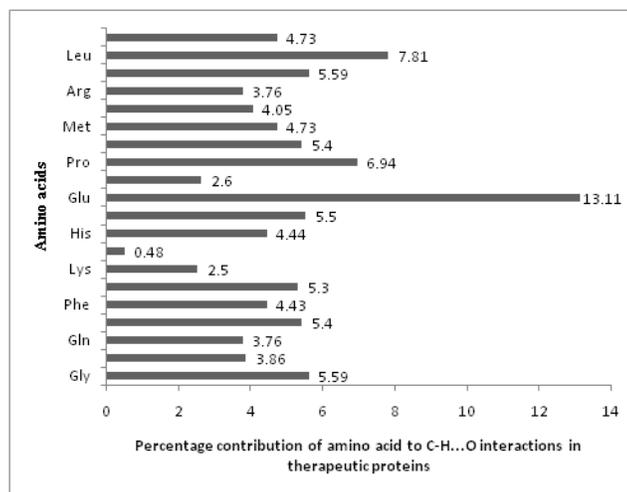


Fig. 4: Acceptor amino acids contribution to the stability

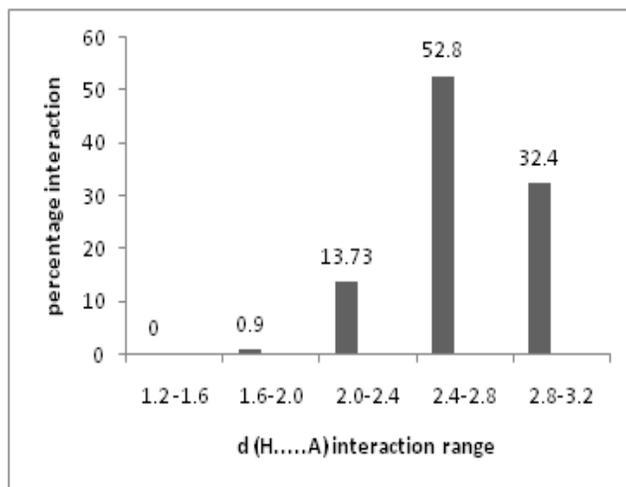


Fig. 5: Observed distribution of C-H...O interactions as a function of interatomic (H...O) distance in the therapeutic proteins data set

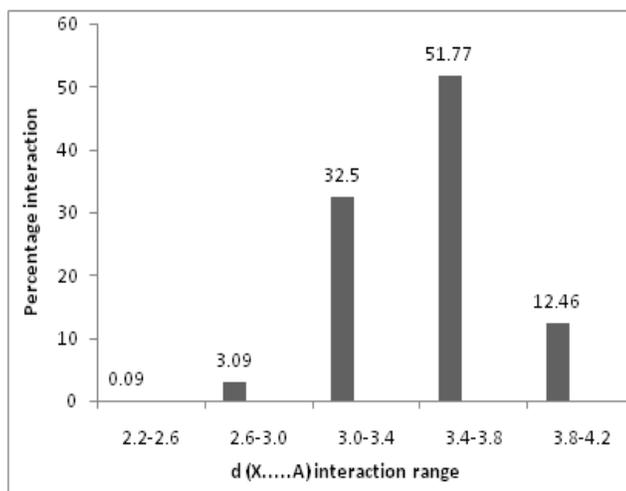


Fig. 6: Observed distribution of C-H...O interactions as a function of interatomic (C...O) distance in the therapeutic proteins data set

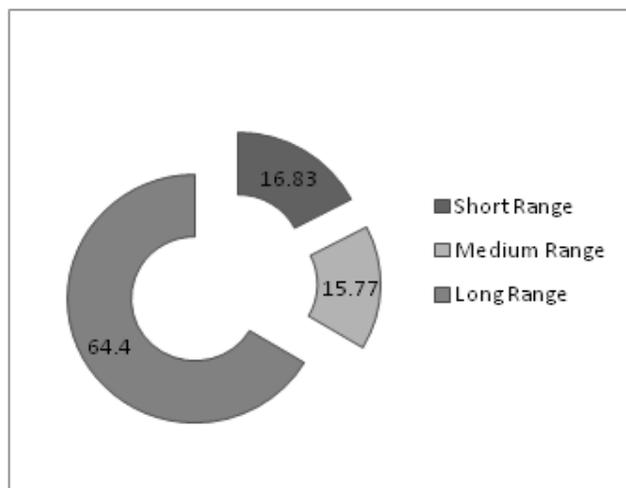


Fig. 7: Sequential distance analysis

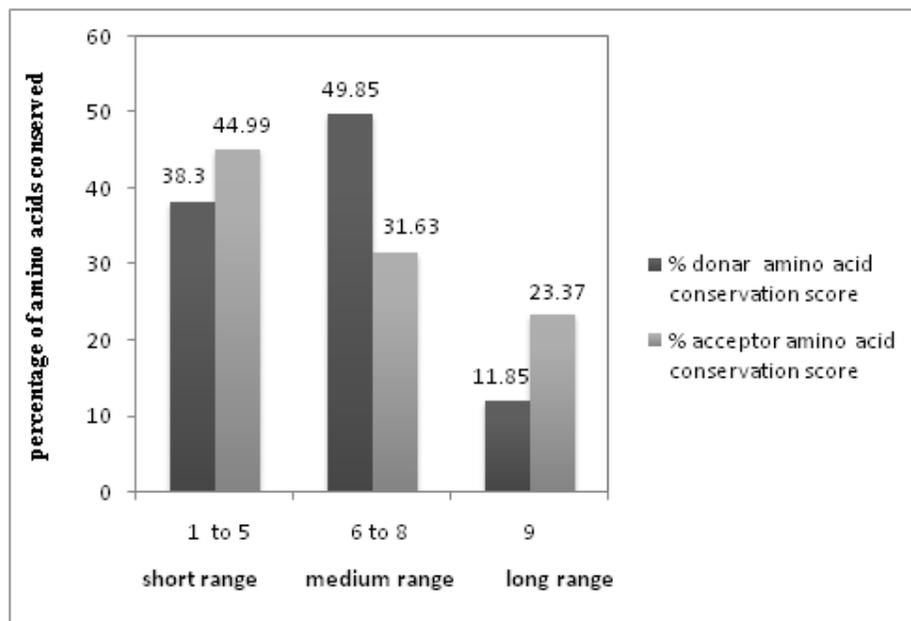


Fig. 8: Conservation scores of donor and acceptor amino acids involved in C-H...O interaction

#### Interaction range

We classified the C-H...O interactions into three categories named short, medium and long range interactions based on the position of donor and acceptor residues. It is clear from Fig. 7 that long range interactions are the predominant in the set of therapeutic proteins. Almost 64.4% of interactions fall in the group of long range contacts followed by 16.83% and 15.77% interactions were from short and medium range interactions respectively.

#### Computation of conservation score

We used the ConSurf program to compute the conservation score of amino acid residues involved in C-H...O interactions in therapeutic proteins and the results are shown in Fig. 8. 12% of the amino acid residues that contributed donor atoms in C-H...O interactions had the highest conservation score of 9, while 50% of the amino acid residues had a conservation score, in the range of 6–8. Thus, 62% of the donor amino acid residues had a higher conservation score. In the case of amino acid residues that contributed acceptor atoms in C-H...O interactions, 23% of the acceptor amino acid residues had the highest conservation score of 9, while 32% of the amino acid residues had a conservation score, in the range of 6–8. Thus, 55% of the acceptor amino acid residues had a higher conservation score. From these observations, we were able to infer that most of the amino acid residues involved in C-H...O interactions might be conserved in therapeutic proteins.

#### DISCUSSION

All cohesive interatomic interactions, whatever their specific nature, contribute to the overall stability of any macromolecule. Hydrogen bonds, however, are known to play a key part in many other phenomena, including enzymatic catalysis. Here we have investigated the role of C-H...O interactions in the structural stability of therapeutic proteins. The selected group of therapeutic proteins show a total of 1099 interactions, there was an average of 23 C-H...O interactions per therapeutic protein. The geometry of these interactions demonstrates that the C-H...O interactions are preferred to be long range contacts. More than 55% of the interacting residues either as a donor or as an acceptor had a high conservation score of 6–9. This high conservation of amino acid residues may in some cases be linked to their involvement in C-H...O interactions and to the stability or the function of the protein<sup>38</sup>. In terms of energetic contribution, theoretical *ab initio* calculations<sup>39,40,41,42</sup> have clearly

revealed that the energy of these C-H...O interactions is less than the energy of a conventional hydrogen bond (O/N-H...O). Even though the C-H...O bond is not only comparable in strength to a traditional hydrogen bond but cumulatively can make a quantitatively greater energetic contribution to folding and stability. The frequency of occurrence and extent of conservation obtained in the present study unequivocally shows that the C-H...O interactions cannot and must not be neglected. Hence, without ambiguity, we can confirm that C-H...O interactions play an important role in the structural stability of therapeutic proteins.

#### ACKNOWLEDGMENTS

The authors thank the management of Vellore Institute of Technology, for providing the facilities to carry out this work.

#### REFERENCES

- Walsh G. Biopharmaceutical benchmarks. *Nat Biotechnol* 2003; 21:865-870.
- Tu Z, Hao J, Kharidia R, Meng XG, Liang JF. Improved stability and selectivity of lytic peptides through self-assembly. *Biochem Biophys Res Commun* 2007; 361(3):712-717.
- Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* 1984; 44:97–179.
- Jeffrey GA, Saenger W. Hydrogen bonding in biological structures. Springer-Verlag, Berlin, 1991.
- Scheiner S. Hydrogen bonding: A theoretical perspective. Oxford University Press, New York, 1997.
- Glasstone S. The structure of some molecular complexes in the liquid phase. *Trans Faraday Soc* 1937; 33:200–214.
- Dippy JFJ. The dissociation constants of monocarboxylic acids; their measurement and their significance in theoretical organic chemistry. *Chem Rev* 1939; 25:151–211.
- Allerhand A, Schleyer PVR. A survey of C-H groups as proton donors in hydrogen bonding. *J Am Chem Soc* 1963; 85:1715–1723.
- DeLaat AM, Ault BS. Infrared matrix isolation study of hydrogen bonds involving carbonhydrogen bonds: alkynes with oxygen bases. *J Am Chem Soc* 1987; 109:4232–4236.
- Kariuki BM, Harris KDM, Philp D, Robinson JMA. A triphenylphosphine oxide-water aggregate facilitates an exceptionally short C-H...O hydrogen bond. *J Am Chem Soc* 1997; 110:12679–12680.

11. Astrup EE, Aomar AM. The Molecular Structure of 2, 2-dimethoxypropane, CH<sub>3</sub>-O- C(CH<sub>3</sub>)<sub>2</sub>- O-CH<sub>3</sub>, in the Gas Phase. *Acta Chem Scand* 1975; A29:794-798.
12. Peterson KI, Klemperer W. Waterhydrocarbon interactions: Rotational spectroscopy and structure of the water-acetylene complex. *J Chem Phys* 1984; 81:3842-3845.
13. Fujii A, Fujimaki E, Ebata T, Mikamiet N. A new type of intramolecular hydrogen bonding: hydroxyl methyl interactions in the o-cresol cation. *J Am Chem Soc* 1998; 120: 13256-13257.
14. Sutor DJ. The C-H...O hydrogen bonds in crystals. *Nature* 1962; 195:68-69
15. Kaufmann R, Knochel A, Kopf J, Oehler J, Rudolph G. Anion Activation, IV. Crystal and Molecular Structure of the Malononitrile Adduct of 18-Crown-6 (1, 4, 7, 10, 13, 16-Hexaoxacyclooctadecane). *Chemische Berichte* 1977; 110:2249-2253.
16. Wahl MC, Sundaralingam M. C-H...O hydrogen bonding in biology. *Trends Biochem Sci* 1997; 22:97-102.
17. Desiraju GR, Steiner T. The weak hydrogen bond in structural chemistry and biology. Oxford University Press, New York, 1999. p. 29-121.
18. Steiner T, Saenger W. Geometry of CH...O hydrogen bonds in carbohydrate crystal structures: Analysis of neutron diffraction data. *J Am Chem Soc* 1992; 114:10146 10154.
19. Mandel-Gutfreund Y, Margalit H, Jernigan RL, Zhurkin VB. A role for C-H...O interactions in protein-DNA recognition. *J Mol Biol* 1998; 277:1129-1140.
20. Jiang L, Lai L. C-H...O hydrogen bonds at protein-protein interfaces. *J Biol Chem* 2002; 277:37732-37740.
21. Saha RP, Bhattacharyya R, Chakrabarti P. Interaction geometry involving planar groups in protein-protein interfaces. *Proteins* 2007; 67:84-97.
22. Pierce AC, Sandretto KL, Bemis GW. Kinase inhibitors and the case for C-H...O hydrogen bonds in protein-ligand binding. *Proteins Struct Funct Genet* 2002; 49:567-576.
23. Sarkhel S, Desiraju GR. N-H...O, O-H...O and C-H...O hydrogen bonds in protein-ligand complexes: strong and weak interactions in molecular recognition. *Proteins Struct Funct Bioinform* 2004; 54:247-259.
24. Jedlovsky P, Turi L. Role of the C-H...O hydrogen bonds in liquids: A Monte Carlo simulation study of liquid formic acid using a newly developed pair-potential. *J Phys Chem* 1997; 101:5429-5436.
25. Mizuno K, Ochi T, Shindo Y. Hydrophobic hydration of acetone probed by nuclear magnetic resonance and infrared: Evidence for the interaction C-H...OH<sub>2</sub>. *J Chem Phys* 1998; 109:9502.
26. Shanthi V, Ramanathan K, Rao Sethumadhavan. Role of the Cation- $\pi$  Interaction in Therapeutic Proteins: A Comparative Study with Conventional Stabilizing Forces. *Journal of Computer Science & Systems Biology* 2009; 2:051-068
27. Ramanathan K, Shanthi V, Rao Sethumadhavan. Significance of CH/ $\pi$  Interactions on the Stability of Therapeutic Proteins. *The Open Structural Biology Journal* 2009; 3:1-10
28. Berman HM, Westbrook JZ, Feng G, Gillilandm TN, Bhat H. The Protein Data Bank. *Nucleic Acids Res* 2000; 28:235-242.
29. Tiwari A, Panigrahi SK. HBAT: A complete package for analyzing strong and weak hydrogen bonds in macromolecular crystal structures. *In Silico Biology* 2007; 7:0057.
30. Babu MM. NCI: A server to identify noncanonical interactions in protein structures. *Nucleic Acid Res* 2003; 31:3345-3348.
31. Gromiha MM, Selvaraj S. Influence of medium and long range interactions in different structural classes of globular proteins. *J Biol Phys* 1997; 23:151-162.
32. Gromiha MM, Selvaraj S. Inter-residue interactions in protein folding and stability. *Prog Biophys Mol Biol* 2004; 86:235-277.
33. Gromiha MM, Santhosh C, Shandar A. Structural Analysis of Cation- $\pi$  Interactions in DNA Binding Proteins. *Int J Biol Macromol* 2004; 34:203-211.
34. Selvaraj S, Gromiha MM. Role of hydrophobic clusters and long-range contact networks in the folding of ( $\alpha$ / $\beta$ ) 8 barrel proteins. *Biophys J* 2003; 84:1919-1925.
35. Glaser F, Pupko T, Paz I, Bell RE, Bechor D. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 2003; 19:163-164.
36. Boeckman B, Bairoch A, Apweiler R, Blatter MC, Estreicher A. et al. The SWISS-PROT Protein knowledge base and its supplement TrEMBL in 2003. *Nucleic Acids Res* 2003; 31: 365-370.
37. Taylor R, Kennard O. Crystallographic evidence for the existence of C-H...O, C-H...N and CH...Cl hydrogen bonds. *J Am Chem Soc* 1982; 104:5063-5070.
38. Brandl M, Weiss MS, Jabs A, Suhnel J, Hilgenfeld R. C-H... $\pi$  interactions in proteins, *Journal of Molecular Biology* 2001; 307:357-377.
39. Levitt M, Perutz MF. Aromatic rings act as hydrogen bond acceptors. *J Mol Biol* 1988; 201:751-754.
40. Scheiner S, Kar T, Gu Y. Strength of the CH...O hydrogen bond of amino acid residues. *J Biol Chem* 2001; 276:9832-9837.
41. Vargas R, Garza J, Dixon DA, Hay BP. How strong is the C-H...O=C hydrogen bond? *J Am Chem Soc* 2000; 122:4750-4755.
42. Duan G, Smith VH, Weaver DF. An *ab initio* and data mining study on aromatic amide interactions. *Chem Phys Lett* 1999; 310:323-332.