Atomic Details of Globular Protein

C. S. Vinobha[1] and E. Rajasekaran*[2]

[1]Department of Bioinformatics, School of Bioengineering, Faculty of Engineering and Technology, SRM University, Kattankulathur - 603 203, Kancheepuram Dt., Tamil Nadu, India
[2]Department of Bioinformatics, School of Biotechnology and Medical Sciences, Karunya University, Karunya Nagar, Coimbatore – 641 114, Tamil Nadu, India.
*Corresponding Author E-mail: sekaran@karunya.edu

Abstract

In the last few decades there has been considerable amount of work gone into protein structure, function and organization. Particularly there is lot of computational studies leading to understand protein structures. However the ultimate truth of real hideous information is not yet understood. In order to understand this buried information in the protein structures, the molecules were visualised on actual basis i.e. at atom level. The basic units of proteins are Carbon (C), Sulphur (S), Nitrogen (N), Oxygen (O) and Hydrogen (H). The distribution of these atoms along the protein sequences were studied. That is a question is asked; what is the probable number of C, S, N, O and H atoms in a given length (number of amino acids)? The results showed that given any number of lengths, the probable number of C, S, N, O and H atoms were 4.90, 0.05, 1.33, 1.50 and 7.77 per unit length respectively.

Keywords: atomic details; probable atoms; protein atoms; protein; atom distribution.

Introduction

In the last few decades there has been considerable amount of work gone into protein structure, function and organization [1-4]. Particularly there is lot of computational studies leading to understand protein structures [5-11]. However the ultimate truth of real hidden information is not yet understood. In order to understand this buried information in the protein structures, this work was carried out. The protein sequences of different organisms such as *Apis mellifera, Arabidopsis thaliana, Bos taurus, Caenorhabditis elegans, Canis familiaris, Danio rerio, Drosophila melanogaster, Gallus gallus, Homo sapiens, Mus musculus, Pan troglodytes, Plasmodium falciparum, Rattus novergicus, Saccharomyces cerevisiae, Schizosaccharomyces pombe* and *Strongylocentrotus purpuratus* were retrieved from NCBI databank in the

Fasta format. Using our own home made software, which was done using "C" language, further analysis was done. There is lot of information inside the protein, so as to retrieve the hidden information, the analysis was done at the basic level i.e. at the atomic level.

Procedure

The entire protein sequences of the above said organisms were downloaded from the NCBI database. Then the sequences were subjected to probability analyses using the homemade software SEQUANA written in 'C' language. The software calculated the protein sequences at atom level i.e., based on the probable number of Carbon (C), Sulphur (S), Nitrogen (N), Oxygen (O) and Hydrogen (H) atoms in a given length (number of amino acids) was studied. This was achieved by counting total number of these atoms present in the given stretch and grouped. The group has a highest number of same number of atom is taken as the probable one. The length differed (from 3 to 5) for each calculation and it yielded different outcomes and the stretch lengths studied were up to 45 amino acids.

Results and Discussion

The probable number of C, S, N, O and H atoms in a given length of amino acids was studied here, i.e., the total number of atoms present in the given stretch of amino acid was grouped. The stretch lengths studied were up to 45 amino acids as illustrated in Figure 1, Figure 2 and Figure 3. Given any length the number of C, S, N, O and H atoms were 4.90, 0.05, 1.33, 1.50 and 7.77 times the number of residues. As the S having no atoms and the H atoms were 8 times the number of residue, the figures of both the atoms are not required.
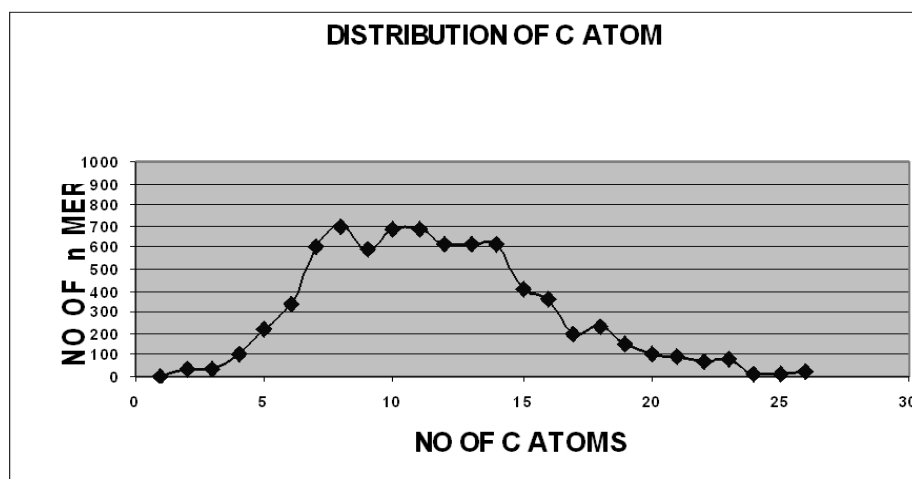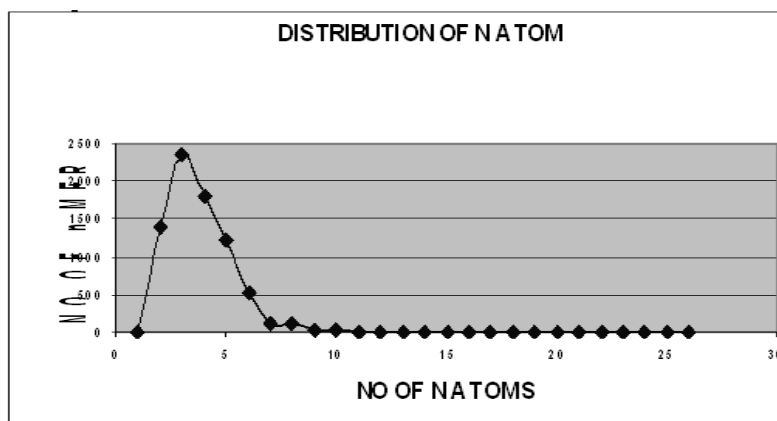

Figure 1: Distribution of Carbon atoms.

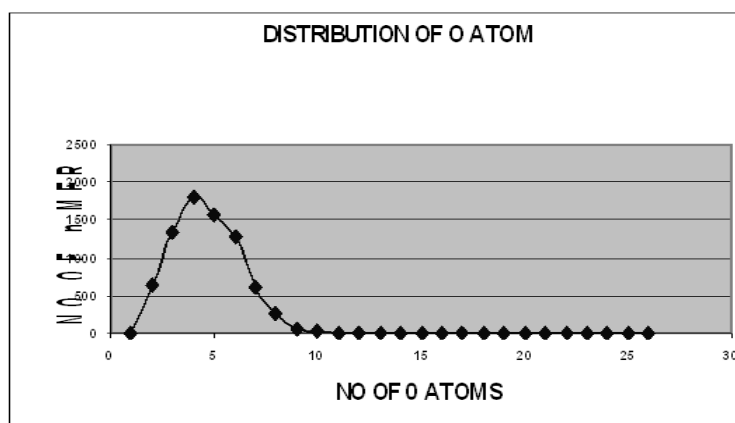**Figure 2**: Distribution of Nitrogen atoms.

Figure 3: Distribution of Oxygen atoms.

Conclusion

Given any number of lengths, the probable number of C, S, N, O and H atoms were 4.90, 0.05, 1.33, 1.50 and 7.77 per unit length respectively. The protein was highly stable at that number of atoms per residue.

References

[1] Narang P, Bhushan K, Bose S and Jayaram B 2006 Protein structure evaluation using an all-atom energy based empirical scoring function; *J. Biomol. Str. Dyn.* **23** 385-406.

[2] Venkatachalam C M 1968 Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units; *Biopolymers* **6** 1425-1436.

[3]   Ramakrishnan C and Soman K V 1982 Identification of secondary structures in globular proteins – A new algorithm; *Int. J. Peptide Protein Res.* **20** 218-237.

[4]   Sowdhamini R, Srinivasan N, Ramakrishnan C and Balram P 1992 Orthogonal ββ motifs in proteins; *J. Mol. Bio.* **229** 759-777.

[5]   Thukral L, Shenoy S.R., Bhushan K and Jayaram B 2007 ProRegIn : A regularity index for the selection of native-like tertiary structures of proteins; *J. Biosci.* **32 (1)** 71-81.

[6]   Blundell T L, Carney D, *et al* 1988 18th Sir Hans Krebs Lecture. Knowledge-based protein modeling and design; *Eur. J. Biochem.* **172** 513-520.

[7]   Bruccoleri R E and Karplus M 1987 Prediction of the folding of short polypeptide segments by uniform conformational sampling; *Biopolymers* **26** 137-168.

[8]   Bruccoleri R E and Karplus M 1990 Conformational sampling using high temperature molecular dynamics; *Biopolymers* **29** 1847-1862.

[9]   Chou P Y and Fasman G D 1974 Conformational parameters for amino acids in helical, beta-sheet and random coil regions calculated from proteins; *Biochemistry* **13** 211-222.

[10]   Donate L E, Rufino S D, Canard L H J and Blundell T L 1996 Conformational analysis and clustering of short and medium size loops connection regular secondary structures. A database for modeling and prediction; *Protein Sci.***5** 2600-2616.

[11]   Narang P, Bhushan K, Bose S and Jayaram B 2005 A computational pathway for bracketing native-like structures for small alpha helical globular proteins; *Phys. Chem. Chem. Phys.* **7** 2364-2375.