

CARd-3D: Carbon Distribution in 3D Structure Program for Globular Proteins

Rajasekaran Ekambaram^{1*}, Akila Kannaiyan², Vijayasathy Marimuthu³, Vinobha Chinnaiah Swaminathan⁴, Senthil Renganathan⁴ & Ananda Gopu Perumal⁵

Received October 28, 2013; Revised January 27, 2014; Accepted January 30, 2014; Published March 19, 2014

¹Department of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Karunya Nagar, Coimbatore - 641114, Tamil Nadu, India; ²Department of Bioinformatics, Bishop Heber College, Trichy-620017, Tamil Nadu, India; ³Molecular Biophysics Unit, Indian Institute of Science, Bangalore - 560012, Karnataka, India; ⁴Department of Bioinformatics, Marudhu Pandiyar College, Vallam, Thanjavur - 613403, Tamil Nadu, India, ⁵Bioinformatics Research Centre, School of Computer Engineering, Nanyang Technological University, Block NS4-04-33, Singapore - 639798 Singapore; Rajasekaran Ekambaram - Email: ersekaran@gmail.com; *Corresponding author

Abstract:

Spatial arrangement of carbon in protein structure is analyzed here. Particularly, the carbon fractions around individual atoms are compared. It is hoped that it follows the principle of 31.45% carbon around individual atoms. The results reveal that globular protein's atoms follow this principle. A comparative study on monomer versus dimer reveal that carbon is better distributed in dimeric form than in its monomeric form. Similar study on solid versus liquid structures reveals that the liquid (NMR) structure has better carbon distribution over the corresponding solid (X-Ray) structure. The carbon fraction distributions in fiber and toxin protein are compared. Fiber proteins follow the principle of carbon fraction distribution. At the same time it has another broad spectrum of carbon distribution than in globular proteins. The toxin protein follows an abnormal carbon fraction distribution. The carbon fraction distribution plays an important role in deciding the structure and shape of proteins. It is hoped to help in understanding the protein folding and function.

Keywords: crystal structure, carbon distribution, protein structure, hydrophobic scale, carbon fraction, CARd3D.

Background:

On viewing the protein folding differently, a unifying spatial arrangement in protein is proposed despite there is different structure and function [1]. The nature of forces involved for the molecular conformations of proteins is reviewed some time back by Kauzmann [2]. Hydrophobic forces play an important role in structure creation, stability and to carry out specific function. Carbon involved towards these hydrophobic forces. Carbon distribution in different sequences makes the proteins to have different function. It is reported that proteins try to maintain specific fraction of carbon content along its sequence in over all [3] or in portion [4]. Carbon content and distribution is related to arrangement of amino acid along the sequence. Each amino acid in a sequence is accountable to its carbon

distribution. So the importances of carbon atoms in 20 amino acids are to be described carefully rather than simple classification as non polar versus polar/charged. Carbon substance can become magnetic when in contact with a magnetic material. Scientists say that the mechanism relies on the transfer of spin carried by electrons of the magnetic substrate to the carbon compound. A theoretical study to capture this phenomenon accepts the fact that the carbon has magnetic property [5]. It is widely believed that graphite and other forms of carbon can have ferromagnetic properties, but the effects are so weak. After measuring the magnetic properties of a meteorite sample [6], it is argued that carbon has magnetic effect. Coey and coworkers were able measure this tiny magnetic moment by using magnetic force

microscope and applying to nanotube. It is 0.1 Bohr magnetons per carbon in room temperature. It is 2.2 Bohr magnetons for iron. This magnetic property of carbon opens up a new avenue in biology of carbon compounds. The nature and extent of carbon distribution in protein structure is investigated here. The crystal and solution structure of globular proteins are taken and analyzed with a question, how the fraction of carbon varies around individual atoms in the structure?

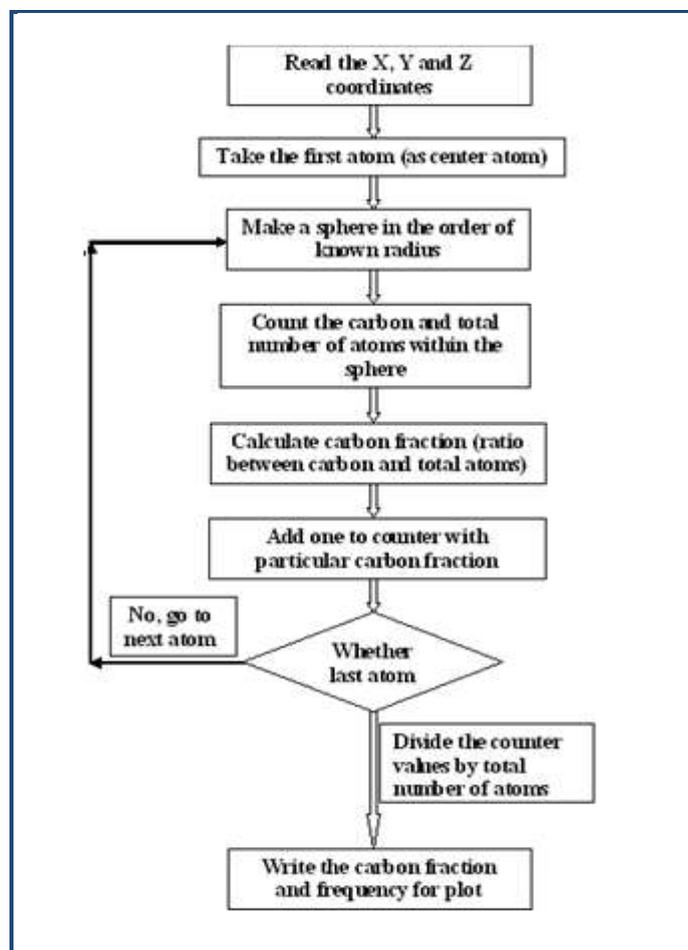


Figure 1: Flow diagram showing computation of carbon fraction distribution in structure of protein.

Methodology:

Dataset

Table 1 (see supplementary material) gives the list of PDB structure identified for comparison of carbon distribution in monomeric and dimeric proteins and differentiates solid structure with liquid structures. The PDB coordinates of proteins are taken from Protein Data Bank. The super oxide dismutase (SOD) (Uniprot ID: P00441) gives the requirement of different structure like monomer, dimer in liquid and solid form. To understand the carbon fraction distribution in fiber proteins, coronin (2AQ5) was identified. A toxin protein (1XTC) is also identified for carbon comparison study. Hydrogen atoms in X-ray structures are added using online kinemage.biochem.duke.edu (reduce version 3.16) software.

Method:

The **Flow chart (Figure 1)** describes how carbon fractions are calculated. In a simple term, (1) the crystal structure (PDB

coordinates) of a given protein is taken into array, (2) the first atom is selected as center atom, (3) a spheres around this center atom is drawn with specified radius, (4) number of carbon and total atoms are counted in the sphere, (5) the ratio between carbon and total atom is taken as carbon fraction around the center atom, (6) the center atom is accounted based on the carbon fraction, (7) the procedure is repeated until the last atom is reached and finally (8) the number of atoms in the groups are divided by total number of atoms to (9) plot number of sphere versus carbon fraction. The maximum number of spheres at a given carbon fraction is taken as probable one. A Perl program (called CARd3D) has been developed to carry out these accounting. The carbon fractions around individual atoms in different radii are captured for comparison and plotted X-Y plot for discussion.

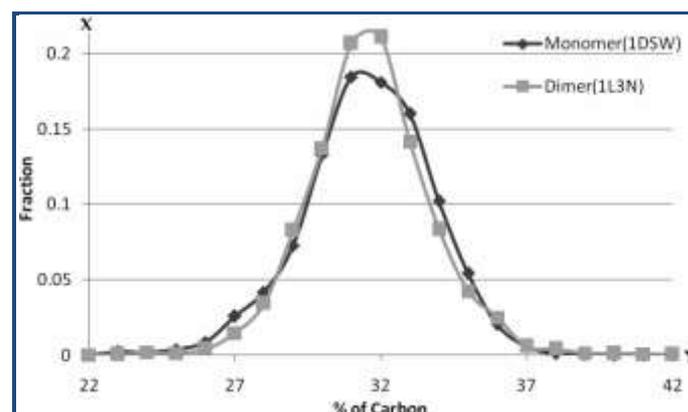


Figure 2: Carbon distribution profile of super oxide dismutase in monomeric and dimeric form. Note that the dimer shows better distribution over monomer one.

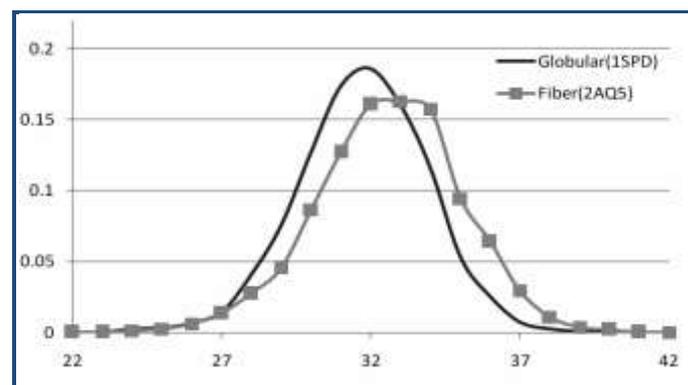


Figure 3: Carbon distribution profile of globular protein (super oxide dismutase) compared with fiber protein (Coronin). The fiber protein does not follow the normal distribution curve.

Results & Discussion:

Carbon Fraction Distribution Profile

Carbon Fraction distribution (CFD) for super oxide dismutase in monomeric and dimeric form is calculated for 15Å diameter and given in **Figure 2**. It is argued that globular proteins are trying to maintain carbon fraction around 0.3145 [4]. Here, both monomer and dimeric form are in agreement with expected term. But the dimeric form is narrow and maximum at expected value. So the carbon distribution profile stabilises the dimer over the monomer. There are instances where a sphere containing 22 and 42 percent of carbon with less

frequency. The distribution pattern is expected a normal one with both side are in equal numbers. The monomer does not show up with correct distribution while the dimeric form is in normal distribution. The existence of this protein is proportional to the normal distribution. Abnormal distributions can be disorders and diseases. Formation of protein dimer can be better understood based on this CFD formula. A sharp narrow carbon distribution profile was observed at 25Å while it was discontinuous broad pattern of distribution at 4Å. The optimum value of distribution found at 15Å which can be used as standard of measure. In fact this spherical diameter, 15Å roughly covers up the protein patterns of length 15 amino acid long which is equal to 225 atoms. Diameter 35Å and above show a same CFD for all atoms. There is no variation in carbon fraction.

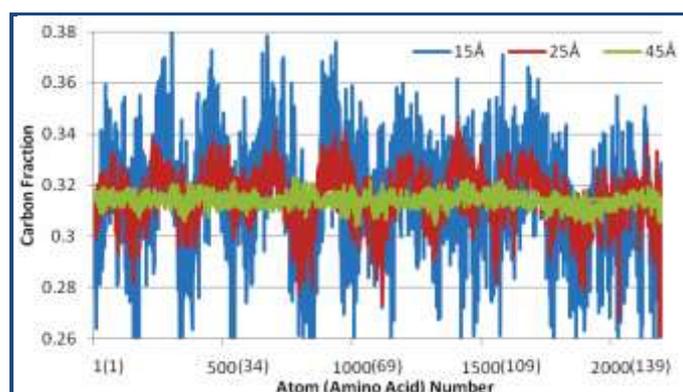


Figure 4: Atomic profile of super oxide dismutase (1DSW) in spherical diameters of 15, 25 and 45Å are shown. The carbon fraction fluctuates around 0.3145 which is measure of hydrophobic scale. In 25Å (red) the fluctuation is visible clearly. The fluctuation in hydrophobic value occurs at every 15 amino acids gap. The presence of disulphide bond between 57 and 146 amino acids makes an improper carbon distribution (from 60 to 150) in 15Å.

The carbon distribution in muscle (fiber-2AQ5) protein was analysed for existence of such pattern. The pattern was compared with globular protein (SOD-1SPD) as given in **Figure 3**. The fiber protein is in no way following the carbon distribution pattern. It is not important for its functional role. The distribution curve is flat on top. There is no narrow distribution curve but broad curve. The fibrous protein does not maintain carbon distribution profile for its survival like globular protein-(for example SOD). But both proteins maintain a definite carbon fraction distribution along the sequence. The CFD is different for fiber proteins than the globular proteins. That is why the fiber proteins are static while globular proteins are dynamic. The solution structure of dimeric form of SOD (1L3N) determined from NMR is compared (result not shown here) with X-ray structure (1SPD). The solution structure allows better CFD than the solid structure. Of course the medium influences the structure. Solid structure is slightly distorted over the solution structure.

CFD in fiber(2AQ5) and toxin (1XTC) proteins reveal that the fiber protein does not have the tendency of having maximum at 31.45% of carbon in its structure. It lacks hydrophilic elements. More over the variation in carbon fraction is found to be from 27% to 36% while it is 28% and 34% for globular

protein. The maximum fraction was greater than in globular protein. In atomic profile plot as in **Figure 4**, there was no specific gapped carbon rich or carbon less stretches. But two broad spectrum of maximum carbon fraction with length approximately 200 aminoacids were observed. These are all may be essential for fiber proteins. A clear difference in CFD and atomic profile are observed in globular and fiber proteins. On the other hand the toxin protein shows up an unusual distribution. This is in no way in tune with globular or fiber protein. The carbon fraction varies arbitrarily in toxin protein. In CFD calculation, accounting carbon in sp³/sp² hybridization and carbon in planar and flexible side chains will improve the quality of results.

CFD around individual atoms, particularly around O, N and C at lower diameters ranging from 4 to 8Å will certainly help the result. An inclusion of water molecules of the crystal structure will improve the distribution result. Protein active sites are having higher carbon content. This can be better visualized from CFD analysis. That is identification and modification of active site becomes easier from CFD analysis. The binding sites, structural domains and patterns can also be identified from this analysis. Disordered regions in diseased proteins can be identified and suggest a possible mutation for existence. It is observed that for every 2.25 residues there is increase or decrease in carbon fraction. The selection of amino acid for mutation is in such way that it follows the profile of carbon distribution to maintain 31.45% carbon. A wide range of diameters are workedout for CFD analysis for different proteins. The CFD curve is somewhat broad at 5Å and after 6Å a normal narrow distribution appears. At diameter 70Å, only two points at 31 (99.95%) and 32 (.05%) appear for SOD. The spherical diameter 15Å sounds ideal for analysis purpose. But at 25Å a smooth transition between hydrophobic and hydrophilic with a periodicity length of 15 amino acids are observed normally. Calculations involving diameter between 15 and 25Å can be used for CFD calculation. Neglecting H atom in the plot and averaging of carbon fraction values of individual atoms of given amino acids gives better observation.

Atoms in a given spherical diameter are analysed. Results reveal that there are maximum of 14 continuous amino acid's atoms are included in the vicinity of central amino acids. At higher diameter a distance amino acids are also observed nearby. The continuous stretch of amino acids varies in C-terminal and N-terminal side. It is also noticed that any difference in CFD for given diameter is counter balanced at higher diameter. So the ultimate structure is determined by not only the local atoms but also by distance amino acids. The averaged CFD for individual amino acids represent better over the atomic level representation. Further excluding the hydrogen atom from averaging, improves the results to an appropriate CFD plot. As the codons XAX and XTX (X=A, T, G or C) are playing an important role in producing proteins with adequate CFD. Producing mRNAs with this principle of CFD is important task for solving genetics diseases and production [7]. Similar studies with number of carbon rather than fraction can yield different results and observation.

The atomic profile (AP) that is carbon fraction around individual atoms of monomer SOD (1DSW) is shown in **Figure**

4. The figure is a plot of carbon fraction for individual atoms within a given spherical diameter. Though the analyses were carried out for range of diameter from 3 to 50Å, the result is shown for 15Å (blue), 25Å (red) and 45Å (green). A wide spectrum of variation in carbon fraction is observed with 15Å (see blue lines). On the other side a narrow variation among atoms are observed with 45Å (green). At 25Å a moderate variation in carbon fraction among the atoms (red) is observed. Notice that a carbon rich and carbon less stretches with specific gap is observed in this 25Å diameter. This novel set of spatial arrangement of carbon fraction variation observed with an interval of 15 amino acids is remarkable. That is a fluctuation between hydrophobic and hydrophilic elements with 15 amino acid gap is observed. This is clearly visible in diameter 25Å (red). It could be the factor deciding the protein folding, stability and existence. The globular proteins trying to maintain 31.45% carbon is evident from this factor. Alteration in carbon fraction with every 8th residues is observed. Side chain atoms (as center atom) contribute more in the carbon fraction fluctuation than the main chain atoms. A 7 amino acids long similar stretches with different amino acid sequence can be retrieved for hydrophobic and hydrophilic stretches. The presence of disulphide bond between 57 and 146 amino acids makes an improper carbon distribution (at stretch from 60 to 150) in lower diameter that is at 15Å. The globular proteins trying to maintain 31.45% carbon is evident from these observations. This trend was found weak in lower diameters (less than 15Å). It may not be the factor for existence at smaller local structure. The repeats in 7 amino acids interval are also confirmed with CARd program at sequence level. At 90Å the plot becomes flat at .3145 with no variation in AP at all.

Similar AP plot for fiber and toxin proteins are compared. The fibrous protein shows a similar trend of hydrophobic and hydrophilic patterns as in globular proteins, but the local hydrophobic/hydrophilic patterns together form another broad sense of hydrophobic peaks. In fact this is better visible with total number of carbon rather than the fraction. The periodicity of this broad stretches are at an interval of 45 amino acids. In the AP plot another very broad peaks are noticed with an interval of 280 amino acids. The toxin protein shows no such a distribution pattern. The periodicity of hydrophobic and hydrophilic patterns with 7 amino acids gap is not observed. The distribution of carbon content is important. The uneven distribution regions can be mutated for stabilization. One has to do a mutation that takes these issues in different levels of distribution, particularly at spherical diameter up to 25Å. The carbon distribution in protein structure can be modified for better stability and existence of the protein. Comparison of native and mutated form of SOD monomer suggests that the alteration in carbon fraction distribution is not at specific site rather observed in many amino acids. This gives an idea that the mutation changes the course of activity by slowing down the dynamics or aggression of the protein. On the other side the productivity is hampered. If the hydrophobic mutation occurs at binding site, the protein encourages for strong action. The hydrophilic mutation at active site does encourage water molecule to interact and neglecting the active interaction. Alteration in carbon distribution due to mutation can damage the conserved domains and motif and change the total function of the

biomolecules. Any unusual distribution can lead to disorder and can cause disease. An accumulation of carbon can manipulate the magnetic effect in fold form. What happens to this distribution analysis in terms of other elements like O, N, S and H atoms? The analyses were carried out. The results need not be presented here. The residues in unfavorable regions of Ramachandran plot can be defined by carbon distribution plot as the folding itself is due to hydrophobic elements. One can think that the beta sheet formation itself is due to hydrophobic elements. A strong H-bond donor/acceptor water molecule can penetrate into protein and fulfill required H-bonds. The H-bond only favors the beta sheet formation.

Every 2.25 amino acids alter the course of hydrophobic formulism. Diameter 15Å includes approximately 225 atoms which covers a small amphipathic pattern which is 15 amino acids length. Diameter 22.5Å includes approximately 700 atoms covering another larger 45 amino acid long amphipathic domains. Diameter 25Å contains 800 atoms in it. The inner most core of the domain contain hydrophilic atoms or residues. Eventhough the presence of sulphur atom is minimal, it affect the carbon distribution profile significantly. Particularly the cys-cys disulphide bond makes drastic changes in the carbon distribution pattern. For example the 12Å calculation on SOD shows no carbon domains between 49 and 69 while other stretches contain. So the cys57 alters the local structure. Because of this violation of carbon distribution, avoiding cys will yield better proteins. One can apply this concept of avoiding Cys for other proteins as well. General observation is, more the hydrophilic adjacent to hydrophobic pocket better the domain is or more stable. In all these calculations, the H centered statistics are not taken into account as it is significant. The carbon density around O and N is different than C atom, particularly at near surroundings (< 12Å). The O is surrounded by more carbon than that of N followed by C. The O water attraction is quenched by this way. The electrical conductance also damped. The carbon fraction distribution plot around O, N&C using 5Å diameter can be maneuvered for identification and modification of unusual amino acids, active site, hydrophilic region, hydrophobic pockets etc. Plots between 4Å and 10Å diameter will be helpful for this purpose. Analysing carbon around oxygen atoms in ASP&GLU (carboxyl, aspartate and glutamate) and ASN&GLN_SER&THR (amide, aspartamine and glutamine and alcohol, serine and threonine oxygens), reveals that greater carbon around charged carboxyl oxygen atoms. That too happens in diameter between 4 and 8Å. Any change in C fraction for individual atoms can leads to instability. In fact, one can arrange to get rule for C fraction like Chargaff's rule for protein stability analysis that can be used for comparison of two protein states and mutational effect. CARd-3D program can be better utilised for identification and modification of hydrophobic/hydrophilic/amphipathic stretches in globular proteins.

Conclusion:

Analysis leading to carbon distribution in 3D structure of proteins reveals that carbon is better distributed in dimeric form than in its monomeric form. Liquid (NMR) structure has better carbon distribution than the solid (X-Ray) structure. Fiber proteins follow the principle of carbon fraction distribution, but broad spectrum of carbon distribution than in

globular proteins. The toxin protein follows an abnormal carbon fraction distribution. These analyses reveal the role of carbon distribution deciding the structure and shape of proteins. It is hoped to help in understanding the protein structure and activity. Any deviation from this scale of measure is less probable and leads to instability. Sometime the core of the proteins contains greater carbon content which may not be required from stability point of view. An even distribution with correct carbon fraction distribution will stabilize a protein with required hydrophobicity and electrical conductivity. This needs to be validated experimentally.

References:

- [1] Mittal A & Jayaram B, *J Biomol Struc Dyn*. 2011 **28**: 443 [PMID: 21142216]
- [2] Walter Kauzmann, *Biophys J*. 1964 **4**: 43
- [3] Rajasekaran E & Vijayasathy M, *Bioinformation* 2011 **5**: 455 [PMID: 21423892]
- [4] Rajasekaran E, *Bioinformation* 2012 **8**: 508 [PMID: 22829720]
- [5] Bhattacharya S *et al. J Phys: Condens Matter*. 2011 **23**: 31600
- [6] Coey JM *et al. Nature* 2002 **420**: 156
- [7] Balamurugan P *et al. Bioinformatics* 2014 **30**: 597

Edited by P Kanguane

Citation: Rajasekaran *et al.* Bioinformation 10(3): 138-143 (2014)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: Different forms of proteins taken for study

PDB ID	Form	Method	No. of AA	Protein
1DSW	Monomeric	NMR	153	SOD (globular)
1L3N	Dimeric	NMR	306	SOD (globular)
1SPD	Dimeric	X-Ray	306	SOD (globular)
2AQ5	Monomeric	X-Ray	395	Coronin (fibrous)
1XTC	Heptameric	X-Ray	752	Cholera Toxin