BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 6(8)

Allotment of carbon is responsible for disorders in proteins

Ekambaram Rajasekaran*, Kannaiyan Akila, Marimuthu Vijayasarathy

Department of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Coimbatore - 641114, Tamil Nadu, India; Ekambaram Rajasekaran - Email: ersekaran@gmail.com; *Corresponding author

Received May 11, 2011; Accepted June 24, 2011; Published July 06, 2011

Abstract:
Sequence stretches in proteins that do not fold into a form are referred as disordered regions. Databases like Disport describe disordered regions in proteins and web servers like PrDOS and DisEMBL, facilitate the prediction of disordered regions. These studies are often based on residue level features. Here, we describe proteins with disordered regions using carbon content and distributions. The distribution pattern for proteins with disordered regions is different from those that do not show disordered regions.

Keywords: Disorder protein; carbon distribution; intrinsic disorder; Carbana; carbon allotment.

Background:
Proteins that are not fold itself in a specified form for normal functioning are defined as disorder proteins. There are many databases available for disordered proteins (e.g. DISPROT). It provides sequence and structural information about disordered proteins. What makes a protein to be disordered? Sethi *et al*., predicted disordered protein based on dipeptide composition and comparison of features of secondary structure **[1].** Based on SVM (support vector machine) method they were able to predict a protein is disordered or not. Kumar *et al.*, reported that the intrinsically disordered proteins form a structured conformation at high temperature due to changes in hydrophobicity **[2]**. Even different solvent environments can alter the hydrophobicity followed by structured conformation. It is reported that proteins prefers to have 31.45% carbon for its stability **[3]**. This fraction of carbon content can be used as a standard of carbon measurement and comparison. Here we report that the carbon distribution along the sequence is responsible for protein disorder. The order or disorders in proteins can be predicted based on carbon distribution. Carbon is a very special element because it plays an important role in hydrophobic interaction. The carbon content and distribution in disordered proteins is expectedly different from normal one because the organization of amino acids in these proteins is different.

Materials and Methodology:
Dataset:
The protein sequences of disordered proteins are collected from Disprot, a curated database (www.disprot.org) of disordered proteins. It provides information about proteins that lack fixed 3D structure in their native form. In this study only one known disordered protein that is Methyl-CpG-binding protein (Uniprot ID: DP00539, Uniprot ID: P51608) of human is taken. It has 486 amino acids with 91% disorder.

Method:
The carbon distribution along the sequence is obtained using CARBANA program available online **[3]**. The detailed methodology on how the carbon content is calculated is given in reference **[3]**. A window size of 500 atoms is chosen in all calculations. The output of the CARBANA program is plotted for better visualization of results as shown in figures.

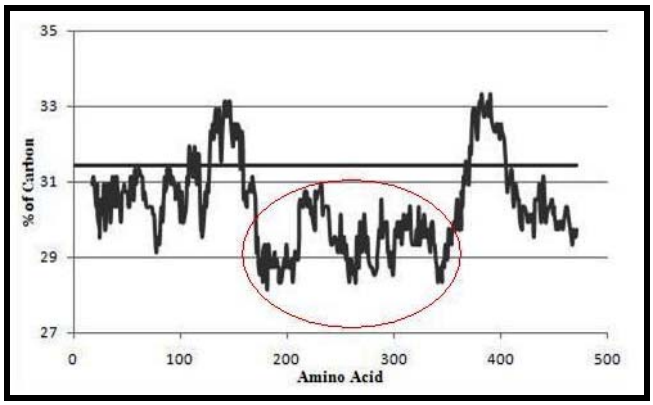

Figure 1: Carbon distribution in Methyl-CpG-binding protein. The region within the circle is disordered.

Results and discussion:
Though several disordered proteins are analysed and tested for carbon distribution, only one protein (Methyl-CpG-binding protein) is reported here. This protein is important because of its characterization in 3D-structure, alternative splicing, chromosomal rearrangement, disease mutation, DNA-binding activity, mental retardation, polymorphism and transcription regulation. It contains 486 amino acids and 91% of them are disordered. The carbon distribution of this protein is given in **Figure 1**. It is a plot of mean distribution of carbon content along the protein chain. If the protein is stable and ordered, it is expected to maintain a carbon content of 31.45% of carbon all along the sequence. Normally globular proteins maintain it. This order and

disorder is a coded property of mRNA/DNA. It is discussed earlier that higher or lower carbon content along the protein sequence might not fold properly due to compression or expansion by water. Most of the time these disorder proteins maintain higher carbon content. Any region with higher carbon content is likely to misfold and makes the protein disordered **[4]**. Lower level of carbon content over long stretches is also considered as to cause of disorder in protein. These low carbon content regions interact with water and unfold from regular structure that makes the protein disorder. In Methyl-CpG-binding protein, a lower carbon content at 155-366 amino acids is maintained (circled region in **Figure 1).** This is a clear indication of protein disorder. This is predicted by several other methods and reported in disordered protein databases. DISPROT for example predicts this protein sequence as a disordered one. So the point is carbon distribution pattern can predict if a protein is in order or not. Importantly, it can locate the portion of sequence that is responsible for disorder. One can also go for an alternative sequence that will have proper carbon distribution for highly functional proteins. It must be noted that even a disordered protein has link to other interactomes **[5]**. The first 108 and last 80 amino acids of this methyl-CpG-binding protein show up again with less amount of carbon content. Overall the hydrophobicity of this protein is less. It is likely to have more water molecules inside. The disordered proteins sometimes match with viral proteins **[6].** One can develop a dedicated tool for prediction of disordered proteins based on carbon content. The development of database of disordered proteins will be useful in developing a disease free living system tomorrow as most of the animal proteins are already in disorder. Particularly in humans, several proteins are in disorder due to evolutionary cause. This is mainly due to reduction of thymine in coding frames of nucleic acids during evolution **[7, 8]**.

Conclusion:

Prediction of disordered portion based on carbon allotment along the protein sequence is reported here. The misfold due to carbon rich regions and unfold due to carbon less regions are responsible for disorders in proteins. This new method is cable of predicting the protein's disordered regions. This can aid in development of ordered proteins for better function.

References:

[1] Sethi D *et al. Amino Acids* 2008 **35**: 599 [PMID: 18425404]
[2] Kumar N *et al*. *Proteins* 2008 **71**: 1123 [PMID: 18004752]
[3] Rajasekaran E & Vijayasarathy M. *Bioinformation* 2011 **5**: 455 [PMID: 21423892]
[4] Sneha NJ *et al. CMBB-2010 IEEE.* 2011 231-232.
[5] Menon R & Farina C. *PLoS ONE.* 2011 **6**: e18660 [PMID: 21533026]
[6] Akila K & Rajasekaran E. *International Journal of Bioinformatics Research* 2009 **1**: 3
[7] Anandagopu P *et al*. *Bioinformation* 2008 **2**: 304 [PMID: 18478084]
[8] Rajasekaran E & Anandagopu P. *J Adv Biotech.* 2010 **9**: 9

Edited by P Kangueane
Citation: Rajasekaran *et al.* Bioinformation 6(8): 291-292 (2011)