Middle-East Journal of Scientific Research 13 (6): 764-771, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.13.6.2344

Graphical Representation of Protein Sequences by CGR: Analysis of Pentagon and Hexagon Structures

[1]K. Manikandakumar, [2]K. Gokulraj, [3]S. Muthukumaran and [4]R. Srikumar

[1]Department of Physics, Bharathidasan University College,
Aranthangi, Pudukkottai District, Tamil Nadu, India
[2]Department of Computer Science, Jamal Mohamed College,
Tiruchirappalli-620 020, Tamil Nadu, India
[3]Department of Physics, The H.H. Raja's College,
Pudukkottai - 622 001, Tamil Nadu, India
[4]Tamil Nadu State AIDS Control Society (TANSACS), Egmore,
Chennai - 600 008, Tamil Nadu, India

Abstract: Graphical illustration of DNA and Protein sequences by Chaos Game Representation (CGR) reveals the intrinsic structures of gene and protein sequences. Even though CGR of DNA and Protein sequences have been deliberated broadly, there are only a few reports of similar studies on protein sequences. We have applied a new method of CGR for protein sequences and found that there are no intrinsic fractal structures of different families of protein sequences. The families of protein sequences could be viewed only using the CGR method. We apply the CGR method of structural analysis of protein sequences by considering the twenty different amino acid residues, which are classified into five and six groups based on their characteristics namely, Pentagon and Hexagon structure. Hence, this analysis elucidated exclusively in Pentagon & Hexagon type of CGR method of protein sequences. This analysis is used for varying the family of protein sequences which are in different patterns and it will generate the structure of protein sequences without any experimental techniques. This analysis is used to show intrinsic structures and self similar structures of protein sequences. This study is used to distinguish between the families of protein sequences separately.

Key words: Chaos Game Representation · Pentagon structure · Hexagon structure · Protein sequences · Residues · Amino acids.

INTRODUCTION

The Chaos Game Representation (CGR) reveals the intrinsic structures of gene and protein sequences. Even though CGR of DNA and Protein sequences have been deliberated broadly, only a few studies are found on protein sequences. CGR analysis is used to generate the structure of protein sequences without the experimental techniques. The families of protein sequences could be viewed using only the CGR method.

To a large extent, deliberation has newly focused on analyzing the biological sequences of both Deoxyribo Nucleic Acid (DNA) and proteins using the patterns observed in their graphical representations [1-15] and mathematical descriptions. Chaos Game Representation (CGR) for gene sequences was introduced by [7, 8] and the essential structures of genome sequences of a few model organisms were obtained using CGR plots. [2] analyzed 12 sided regular polygon method. So, we have tried and applied a new method of CGR for protein sequences analysis. The amino acids are classified into five and six groups based on their characteristics. The five groups of amino acids are called pentagon type. The six groups of amino acid are called hexagon type. Hence, this analysis elucidated exclusively the Pentagon & Hexagon type of CGR method of protein sequences. This analysis has showed intrinsic structures and self similar structures of protein sequences. We describe a

Corresponding Author: K. Manikandakumar, Department of Physics, Bharathidasan University College, Aranthangi, Pudukkottai District, Tamil Nadu, India.

new way of applying CGR method to different families of protein sequences, to produce 'Pentagon and Hexagon' type structures.

CGR is an algorithm that reveals the sequence pattern in DNA in the form of graphical representations. [6] used the CGR method to explain the observed patterns by calculating the dinucleotide and trinucleotide frequencies. [3] converted CGR pattern of DNA sequences into structures to find out phylogenetic proximity. There are several other attempts made to graphically represent and numerically characterize the DNA sequences and reviewed by [9] outlines most of these methods, their merits and demerits. The major outcome of this review was that the DNA invariants can be used to find the similarity or dissimilarity among the biological sequences.

Even large number of works had been done on the CGR analysis of DNA sequences [1, 3, 4, 6, 7, 9, 10-13], number of works carried out by [2, 5] on CGR analysis of protein sequences. The CGR method has been extended [5] to visualize and analyse both the primary sequences and 3-D structures of proteins. Further, CGR plots of different families of proteins on a 12-sided regular polygon were obtained by [2] by means of representing the twenty amino acids into 12 groups based on the conservative substitutions.

In this paper, we describe a new way of applying CGR method to different families of protein sequences, to produce 'Pentagon and Hexagon' type graphical representations. It possess self-similar structure. The percentage values of amino acid residues have been computed and used for further analysis. We find that there are four major structural classes of proteins namely, All α, All β, α plus β and α by β which are used to produce the graphical representations.

MATERIALS AND METHODS

Details of Data Used: There are four major classes of Structu ral Classification of Proteins (SCOP) namely, all α, all β, α plus β and α by β. In the present study, we have chosen a few representative families of proteins from each of these four classes. The selected protein sequences are downloaded from the Protein Data Bank (PDB) website using SCOP option. Further, we have excluded the residues denoted as 'X' and '?' which represent 'any' and 'unknown' residues respectively, from the downloaded data. We eliminate the unknown residues for calculations. These sequences are used for identifying the patterns and generating the CGR of the corresponding protein sequences.

The pentagon and hexagon structural analysis have been carried out for more than 50 families of protein sequences from the first four structural classifications of proteins. But, here we mentioned only 8 families of proteins, by taking 2 families from each of the above SCOP options.

We apply the CGR method to protein sequences by considering the twenty different amino acid residues into five groups namely, Hydrophobic-strong (H_s), Hydrophobic-weak (H_w), Polar-uncharged (P_{uc}), Polar-charged (P_c) and Charged (C) amino acids. The list of amino acids selected in each of these groups is provided in Table 1.

The H_s, H_w, P_{uc}, P_c and C positions are represented by the vertices (0.0, 0.0), (0.0, 1.0), (1.0, 0.0), (1.0, 1.0) and (0.5, 2.0) respectively for pentagon structure. We apply the CGR method to protein sequences similar to the way mentioned for obtaining the pentagon structure method. Here, we use the single letter amino acid codes of the primary structures of protein sequences belonging to different families of data. We begin with the initial point (0.0, 0.0). The first amino acid residue is read in and depending on the residue, we will identify the group it belongs to. For example, if the first amino acid residue is Histidine (H) then it belongs to Polar-charged (P_c) group and hence it corresponds to the vertex (1.0, 1.0). Next, we take the second amino acid residue and if it is Proline (P), then it belongs to the Hydrophobic-weak (H_w) group with the vertex (0.0, 1.0). By repeating this procedure for all the remaining amino acid residues of the protein sequences in a protein family, we obtain the CGR of the protein sequence family. Using this exclusive way of CGR method, the protein sequences of different families produce the intrinsic structure of 'Pentagon' type. We find that all the four structural classes of protein families produce the self-similar structure. We apply a novel way of grouping the amino acid residues to obtain the CGR of protein sequences belonging to different families.

Grouping of Amino Acids in Hexagon Type Structural Analysis of Protein Sequences: The CGR method is applied to Hexagon type structural analysis of protein sequences by considering the twenty different amino acid residues into six groups namely, Hydrophobic- strong (H_s), Hydrophobic-weak (H_w), Polar-uncharged (P_{uc}), Polar-charged (P_c), Charged- positive (C_p) and Charged-negative (C_n) amino acids. The list of amino acids selected in each of these groups is provided in Table 2.

Table 1: Grouping of amino acid residues - Pentagon structural analysis

S. No.	Name of the Group	Related amino acids
1.	H-Strong (H_s)	Val (V), Leu (L), Ile (I)
2.	H-Weak (H_w)	Ala (A), Phe (F), Pro (P), Met (M)
3.	P-Uncharged (P_{uc})	Ser (S), Thr (T), Asn (N), Gln (Q), Gly (G)
4.	P-Charged (P_c)	Tyr (Y), Cys (C), His (H), Trp (W)
5.	Charged (C)	Lys (K), Arg (R) & Asp (D), Glu (E)
6.	Unknown amino acid	Unknown amino acid (?) / (X)

Table 2: Grouping of amino acid residues - Hexagon structural analysis

S. No.	Name of the Group	Related amino acids
1.	H-Strong (H_s)	Val (V), Leu (L), Ile (I)
2.	H-Weak (H_w)	Ala (A), Phe (F), Pro (P), Met (M)
3.	P-Uncharged (P_{uc})	Ser (S), Thr (T), Asn (N), Gln (Q), Gly (G)
4.	P-Charged (P_c)	Tyr (Y), Cys (C), His (H), Trp (W)
5.	C-Positive (C_p)	Lys (K), Arg (R)
6.	C-Negative (C_n)	Asp (D), Glu (E)
7.	Unknown amino acid	Unknown amino acid (?) / (X)

Table 3: CGR Calculation of Pentagon structural analysis

Sl. No.	Name of the Proteins	No. of pro.	Hs(%)	Hw(%)	Puc(%)	Pc(%)	C(%)
			(All Alpha Proteins)				
1.	DNA/RNA-binding 3-helical bundle	1746	20.04	18.26	26.98	9.83	24.89
2.	Ferritin-like	586	20.61	19.04	24.47	9.35	26.54
			(All Beta Proteins)				
3.	Nucleoplasmin-like/VP	592	20.03	19.89	33.41	9.09	17.57
4.	Trypsin-like serine proteases	1970	20.32	15.32	33.58	11.54	19.23
			(Alpha plus Beta Proteins)				
5.	Ferredoxin-like	2132	20.39	20.21	26.12	9.89	23.40
6.	TBP-like	325	20.84	20.03	27.43	8.63	23.08
			(Alpha by Beta Proteins)				
7.	Flavodoxin-like	1368	20.88	20.77	25.89	8.85	23.61
8.	Alpha/beta-Hydrolases	610	20.55	20.56	28.57	10.10	20.22

Table 4: CGR Calculation of Hexagon structural analysis

Sl. No.	Name of the Proteins	No. of pro.	Hs(%)	Hw(%)	Puc(%)	Pc(%)	Cp(%)	Cn(%)
				(All Alpha Proteins)				
1.	Cytochrome c	594	19.82	21.11	27.33	9.41	10.52	11.81
2.	Heme-dependent peroxidases	297	19.30	21.16	29.22	8.59	9.94	11.79
				(All Beta Proteins)				
3.	Carbonic anhydrase	198	19.85	16.93	28.84	10.86	11.48	12.03
4.	Supersandwich	329	19.73	18.82	28.78	10.81	9.23	12.62
				(Alpha plus Beta Proteins)				
5.	Cysteine proteinases	314	19.77	16.66	30.52	10.96	10.17	11.92
6.	Zincin-like	431	19.58	17.75	29.53	10.54	9.73	12.88
				(Alpha by Beta Proteins)				
7.	Alpha/beta-Hydrolases	610	20.55	20.56	28.57	10.10	8.88	11.34
8.	Periplasmic binding protein-like II	468	20.26	19.61	26.70	8.73	12.33	12.37

The H_s, H_w, P_{uc}, P_c, C_p and C_n positions are represented by the vertices (0.5, 0.0), (0.0, 0.5), (1.0, 0.5), (1.0, 1.0), (0.0, 1.0) and (0.5, 1.5) respectively, for Hexagon structure. We apply the CGR method to protein sequences analysis similar to the way mentioned for obtaining the pentagon method. We begin with the initial point (0.0, 0.0). The first amino acid residue is read in and depending on the residue, we will identify the group it belongs to. For example, if the first amino acid residue is Valine (V), then it belongs to Hydrophobic-strong (H_s) group and hence it corresponds to the vertex (0.5, 0.0). Next, we take the second amino acid residue and if it is Glutamic acid (E), then it belongs to the Charged-negative (C_n) group with the vertex (0.5, 1.5). By repeating this procedure for all the remaining amino acid residues of the protein sequences in a protein family, we obtain the CGR method of the protein sequence family. Using this unique way of CGR technique, the protein sequences of different families produce the structure of 'Hexagon' type. The percentage values of amino acid residues of the six groups available in the family of proteins under consideration are also computed. Table 3 and Table 4 show the CGR calculation of Pentagon and Hexagon structure of proteins.

The CGR calculation is carried by the grouping of amino acid residues which have been classified into Hydrophobic-strong (H_s) region, Hydrophobic-weak (H_w) region, Polar- uncharged (P_{uc}) region, Polar-charged (P_c) region, Charged-positive (C_p) and Charged-negative (C_n) region of amino acid residues. All values are given in %. The difference of Pentagon and Hexagon structure is charged group and it is classified into two types as Hexagon structure (i.e. charged-positive and charged-negative amino acid residues).

RESULTS AND DISCUSSION

CGR of Pentagon Structural Analysis

The All-Alpha Proteins: The DNA/RNA-binding 3-helical bundle family of proteins is having 1746 protein sequences and 280085 numbers of residues. The hydrophobic-strong group is having 20.04% of amino acid residues. The hydrophobic-weak group is having 18.26% of residues. The polar-uncharged group is having 26.98% of residues. The polar-charged group is

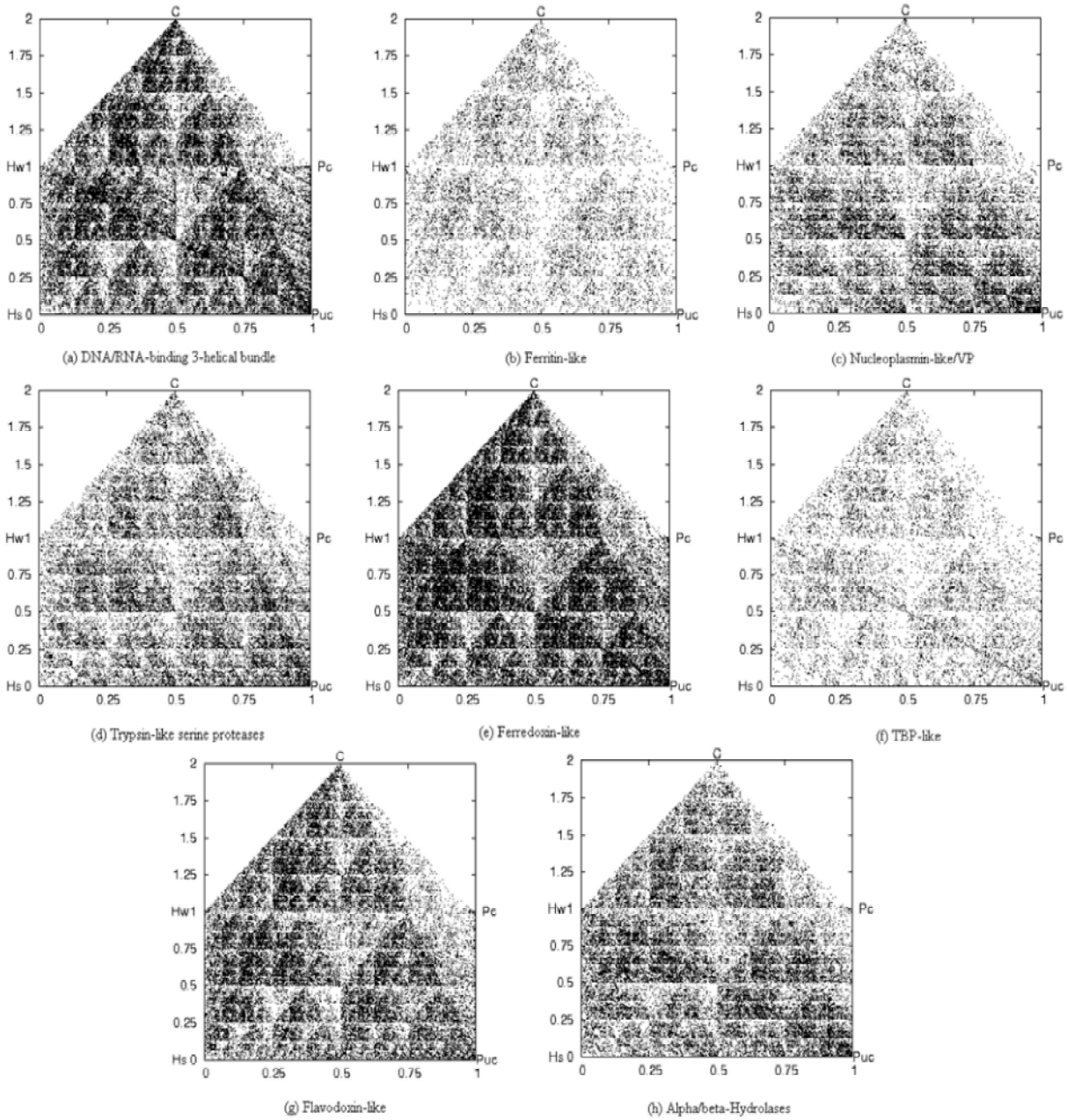

Fig. 1: CGR of pantagon structures

having 9.83% of residues. The charged group is having 24.89% of residues. The CGR plot Fig. 1. (a) Shows, the P_{uc}-H_w-C regions are forming a triangle structure. The centre of the plot is having low number of residues. The P_{uc}-P_c-C regions are having low number of residues. The clustering of residues shows in various regions. It shows some scoop regions. The scoop regions do not have more residues.

The Ferritin-like family of proteins is having 586 protein sequences and 140347 numbers of residues. The hydrophobic-strong group is having 20.61% of amino acid residues. The hydrophobic-weak group is having 19.04% of residues. The polar-uncharged group is having 24.47% of residues. The polar-charged group is having 9.35% of residues. The charged group is having 26.54% of residues. The CGR plot Fig. 1. (b) Shows that, this family of protein is having the total number of amino acid residues which are very less. So, it shows sunny. The residues are distributed in all regions. The centre of the plot is having low residues. The P_{uc}-P_c-C region is having low residues.

The All-Beta Proteins: The Nucleoplasmin-like/VP family of proteins is having 592 protein sequences and 133686 numbers of residues. The hydrophobic-strong group is having 20.03% of amino acid residues. The hydrophobic-weak group is having 19.89% of residues. The polar-uncharged group is having 33.41% of residues. The polar-charged group is having 9.09% of residues. The charged group is having 17.57% of residues. The CGR plot Fig. 1. (c) Shows, the H_w and P_{uc} regions are connected in straight line. The P_{uc}-H_w-C regions are forming triangle structure. The centre of the plot is having low number of residues. The P_{uc}-P_c-C region is having low number of residues.

The Trypsin-like serine proteases family of proteins is having 1970 protein sequences and 299111 numbers of residues. The hydrophobic-strong group is having 20.32% of amino acid residues. The hydrophobic-weak group is having 15.32% of residues. The polar-uncharged group is having 33.58% of residues. The polar-charged group is having 11.54% of residues. The charged group is having 19.23% of residues. The CGR plot Fig. 1. (d) Shows, the P_{uc}-C region connected in straight line. The C-P_{uc}-P_c region is having low number of residues. The centre of the plot is having low number of residues. The H_s-C-P_{uc} regions form a triangle.

The Alpha plus Beta Proteins: The Ferredoxin-like family of proteins is having 2132 protein sequences and 524881 numbers of residues. The hydrophobic-strong group is having 20.39% of amino acid residues. The hydrophobic-weak group is having 20.21% of residues. The polar-uncharged group is having 26.12% of residues. The polar-charged group is having 9.89% of residues. The charged group is having 23.40% of residues. The CGR plot Fig. 1. (e) Shows, this protein family is forming more triangle structures. The P_{uc}-C-P_c region is having low number of residues. The centre of the plot is having low number of residues. The P_{uc}-H_w-C and P_{uc}-H_w-H_s regions are showed in triangle structure.

The TBP-like family of proteins is having 325 protein sequences and 54545 numbers of residues. The hydrophobic-strong group is having 20.84% of amino acid residues. The hydrophobic-weak group is having 20.03% of residues. The polar-uncharged group is having 27.43% of residues. The polar-charged group is having 8.63% of residues. The charged group is having 23.08% of residues. The CGR plot Fig. 1. (f) Shows, this family of protein is having low number of residues. So, it shows sunny. The P_{uc}-H_w region is connected in straight line. The P_{uc}-H_w-H_s regions are forming a triangle structure. The P_{uc}-C-P_c region is having low number of residues.

The Alpha by Beta Proteins: The Flavodoxin-like family of proteins is having 1368 protein sequences and 380807 numbers of residues. The hydrophobic-strong group is having 20.88% of amino acid residues. The hydrophobic-weak group is having 20.77% of residues. The polar-uncharged group is having 25.89% of residues. The polar-charged group is having 8.85% of residues. The charged group is having 23.61% of residues. The CGR plot Fig. 1. (g) Shows, it forming triangle between the points H_s-P_{uc}-C regions. The P_{uc}-C-P_c region is having low number of residues. The centre of the plot shows sunny. This region does not have more residues.

The Alpha/beta-Hydrolases family of proteins is having 610 protein sequences and 233802 numbers of residues. The hydrophobic-strong group is having 20.55% of amino acid residues. The hydrophobic-weak group is having 20.56% of residues. The polar-uncharged group is having 28.57% of residues. The polar-charged group is having 10.10% of residues. The charged group is

having 20.22% of residues. The CGR plot Fig. 1. (h) Shows, the P_{uc}-H_s-H_w and P_{uc}-H_w-C regions are forming triangle structure. The P_{uc}-C-P ,region is having low number of residues. The H_w-P_c region is connected in straight line.

CGR of Hexagon Structural Analysis

The All-Alpha Proteins: The Cytochrome c family of proteins is having 594 protein sequences and 120851 numbers of residues. The hydrophobic-strong group is having 19.82% of amino acid residues. The hydrophobic-weak group is having 21.11% of residues. The polar-uncharged group is having 27.33% of residues. The polar-charged group is having 9.41% of residues. The charged-positive group is having 10.52% of residues. The charged-negative group is having 11.81% of residues. The CGR plot Fig. 2. (a) Shows, the P_{uc} region is having more number of residues. The H_w-C_n-C_p, C_p-P_c-C_n and H_s-P_{uc}-P_c regions are having less number of residues. The centre of the plot is having more number of residues.

The Heme-dependent peroxidases family of proteins is having 297 protein sequences and 107401 numbers of residues. The hydrophobic-strong group is having 19.30% of amino acid residues. The hydrophobic-weak group is having 21.16% of residues. The polar-uncharged group is having 29.22% of residues. The polar-charged group is having 8.59% of residues. The charged-positive group is having 9.94% of residues. The charged-negative group is having 11.79% of residues. The CGR plot Fig. 2. (b) Shows, the P_{uc}-H_w region forming a straight line. The P_{uc}-H_w-H_s region forms a triangle structure. The H_w-C_n-C_p, C_p-P_c-C_n and H_s-P_{uc}-P_c regions have low number of residues.

The All-Beta Proteins: The Carbonic anhydrase family of proteins is having 198 protein sequences and 51003 numbers of residues. The hydrophobic-strong group is having 19.85% of amino acid residues. The hydrophobic-weak group is having 16.93% of residues. The polar-uncharged group is having 28.84% of residues. The polar-charged group is having 10.86% of residues. The charged-positive group is having 11.48% of residues. The charged-negative group is having 12.03% of residues. The CGR plot Fig. 2. (c) Shows, this family of protein is having a total number of residues which is low. So, it shows sunny. The maximum number of residues are formed in P_{uc} region. The other regions are having low number of residues.

The Supersandwich family of proteins is having 329 protein sequences and 258206 numbers of residues. The hydrophobic-strong group is having 19.73% of amino acid residues. The hydrophobic-weak group is having 18.82% of residues. The polar-uncharged group is having 28.78% of residues. The polar-charged group is having 10.81% of residues. The charged-positive group is having 9.23% of residues. The charged-negative group is having 12.62% of residues. The CGR plot Fig. 2. (d) Shows, the P_{uc} and H_s regions are having more number of residues. The P_{uc}-H_s-P_c, H_w-C_n-C_p regions have low residues. The H_s-C_n regions are connected in straight line.

The Alpha plus Beta Proteins: The Cysteine proteinases family of proteins is having 314 protein sequences and 85748 numbers of residues. The hydrophobic-strong group is having 19.77% of amino acid residues. The hydrophobic-weak group is having 16.66% of residues. The polar-uncharged group is having 30.52% of residues. The polar-charged group is having 10.96% of residues. The charged-positive group is having 10.17% of residues. The charged-negative group is having 11.92% of residues. The CGR plot Fig. 2. (e) Shows, the P_{uc} and H_s regions having more residues and they are showed in clustering regions. The H_w-C_n-C_p, C_p-P_{uc}-C_n and H_s-P_{uc}-P_c regions are having low number of residues.

The Zincin-like family of proteins is having 431 protein sequences and 140344 numbers of residues. The hydrophobic-strong group is having 19.58% of amino acid residues. The hydrophobic-weak group is having 17.75% of residues. The polar-uncharged group is having 29.53% of residues. The polar-charged group is having 10.54% of residues. The charged-positive group is having 9.73% of residues. The charged-negative group is having 12.88% of residues. The CGR plot Fig. 2. (f) Shows, the P_{uc} and H_s regions are having more residues. The P_{uc}-H_s regions are connected in straight line. The H_w-C_n-C_p and H_s-P_{uc}-P_c regions are having low number of residues. The C_n-H_s regions are forming a straight line.

The Alpha by Beta Proteins: The Alpha/beta-Hydrolases family of proteins is having 610 protein sequences and 233802 numbers of residues. The hydrophobic-strong group is having 20.55% of amino acid residues. The hydrophobic-weak group is having 20.56% of residues. The polar-uncharged group is having 28.57% of residues. The polar-charged group is having 10.10% of residues. The charged-positive group is having 8.88% of residues.

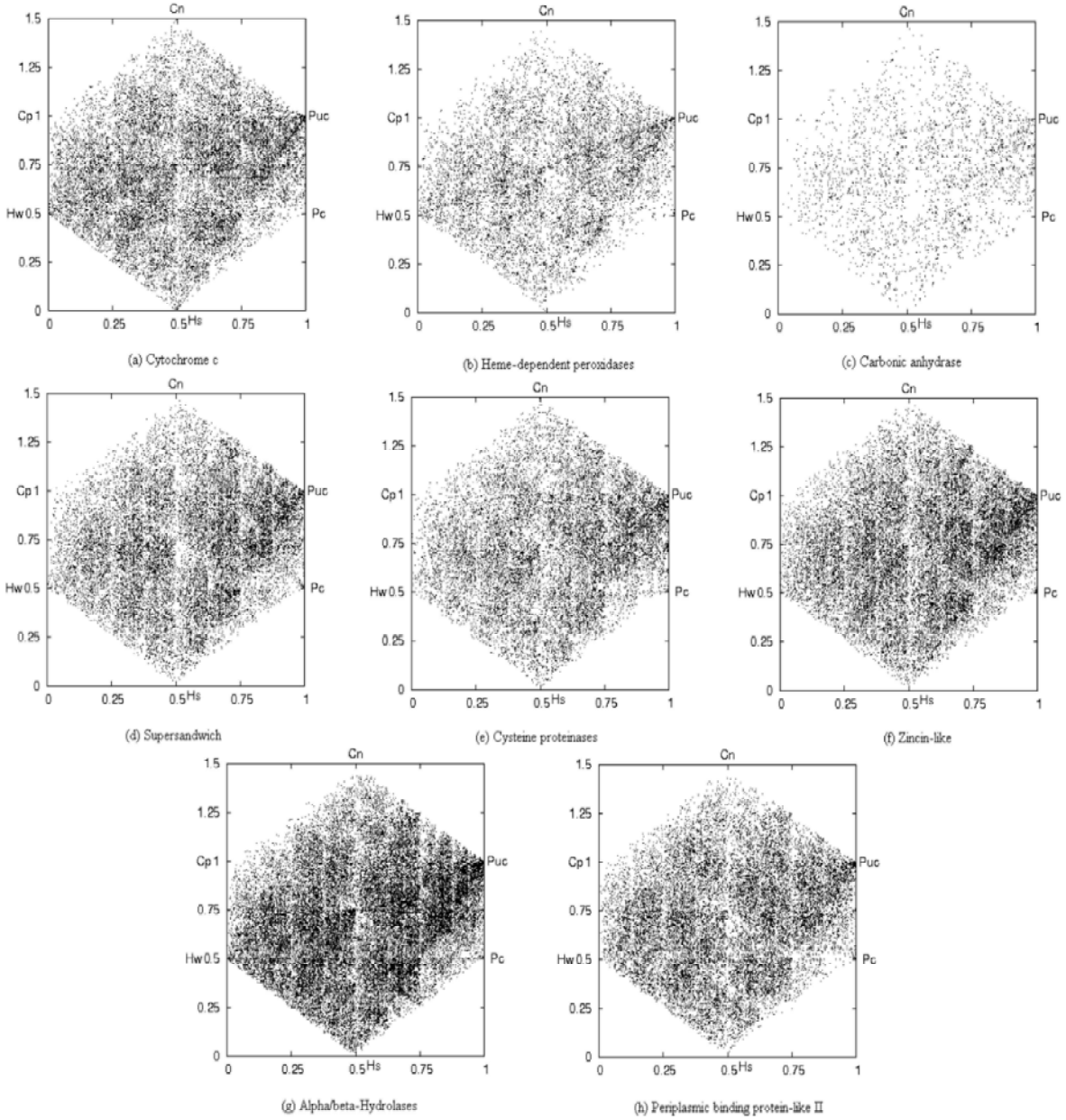

Fig. 2: CGR Hexagon structures

The charged-negative group is having 11.34% of residues. The CGR plot Fig. 2. (g) Shows, the P_{uc}-H_s regions are having more number of residues. The P_{uc}-H_s-H_w regions are forming triangle structure. The H_w-C_n-C_p, H_s-P_c-P_{uc} regions are having low number of residues. The centre of the plot has more residues.

The Periplasmic binding protein-like II family of proteins is having 468 protein sequences and 150380 numbers of residues. The hydrophobic-strong group is having 20.26% of amino acid residues. The hydrophobic-weak group is having 19.61% of residues. The polar-uncharged group is having 26.70% of residues. The polar-charged group is having 8.73% of residues. The charged-positive group is having 12.33% of residues. The charged-negative group is having 12.37% of residues. The CGR plot Fig. 2. (h) Shows, the P_{uc} and H_s regions have more residues. The P_{uc}-P_c-H_s, H_w-C_n-C_p and P_{uc}-C_p-C_n regions have low number of residues. The centre of the plot has more residues.

CONCLUSION

This analysis has been used to generate the structure of protein sequences without the experimental techniques. The experimental techniques are used to visualize the small number of residues only. But, the families of protein sequences are viewed only using this method. Hence, we view the families of protein sequences in single plot. The CGR plot of the Pentagon structure shows the P_c region is not having more residues in all plots. The H_w region is having more residues, but it is not used to form a structure. The H_s - C and P_{uc} - C regions are forming a CGR pentagon structure. The centre of the plot does not have more residues and it shows some scoop region. The Hexagon structure of CGR plot is shown in all family of protein sequences. In general, the Hexagon structure of CGR plot has the H_s-P_c-P_{uc}, H_w-C_p-C_n, C_p-P_{uc}-C_n regions which are showed separately. Because, these regions are having low number of residues. In this study, we identify the charged group of amino acid residues may be differentiated between these two structures. But, it is not related to other families of protein sequences. This analysis showed that some intrinsic structures and self-similar structures of protein sequences are found in pentagon and hexagon structures. This analysis is used for varying the family of protein sequences kept in different patterns.

REFERENCES

1. Almeida, J.S., J.A. Carrico, A. Maretzek, P.A. Noble and M. Fletcher, 2001. Analysis of genomic sequences by Chaos Game Representation. Bioinformatics, 17: 429-437.
2. Basu, S., A. Pan, C. Dutta and J. Das, 1997. Chaos game representation of protein structures. J.Mol. Graphics and Modelling, 15a: 279-289.
3. Deschavanne, P.J., A. Giron, J. Villain, G. Fagot and B. Fertil, 1999. Genomic Signature: Characterization and Classification of Species Assessed by Chaos Game Representation of Sequences. Mol. Biol. Evol., 16: 1391-1399.
4. Dutta, C. and J. Das, 1992. Mathematical characterization of chaos game representation: New algorithms for nucleotide sequence analysis. J. Mol. Biol., 228: 715-729.
5. Fiser, A., G.E. Tusn´ady and I. Simon, 1994. Chaos game representation of protein structures. J. Mol. Graphics, 12: 302-304.
6. Goldman, N., 1993. Nucleotide, dinucleotide and trinucleotide frequencies explain patterns observed in chaos game representation of DNA sequences. Nucleic Acids Research, 21: 2487-2491.
7. Jeffrey, H.J., 1990. Chaos game representation of gene structure. Nucleic Acids Research, 18: 2163-2170.
8. Jeffrey, H.J., 1992. Chaos game visualization of sequences. Computer Graphics, 16: 25-34.
9. Nandy, A., M. Harle and S.C. Basak, 2006. Mathematical descriptors of DNA sequences: development and applications. ARKIVOC (Gaines Ville, FL, USA) 9: 211-238.
10. Randi´cM, LeršN, Plavšsi´cD, S.C. Basak and A.T. Balaban, 2005. Four-color map representation of DNA or RNA sequences and their numerical characterization. Chem. Phys. Lett., 407: 205-208.
11. Manikandakumar, K., S. Muthu Kumaran and R. Srikumar, 2009. Matrix Frequency Analysis of *Oryza Sativa* (japonica cultivar-group) Complete Genomes. Journal of Computer Science & Systems Biology, 2: 159-166.
12. Manikandakumar, K., S. Muthukumaran, R. Srikumar, K. Gokulraj and S. Santhosh Baboo, 2009. Analysis of *Homo sapiens* (Human) Chromosomes Complete Genome Using Matrix Frequency. *nst* Life Sciences and Bioinformatics, 1: 57-66.
13. Manikandakumar, K., K. Gokulraj, R. Srikumar and S. Muthu Kumaran, 2010. Matrix Frequency Analysis of Genome Sequences: Pattern Identification of *Turfgrass* Species. World Applied Sciences Journal, 11 (3): 315-320.
14. Manikandakumar, K., K. Gokulraj, R. Srikumar and S. Muthu Kumaran, 2010. Analysis of parity ratio of protein sequences: A new approach based on Chargaff's rule, Romanian Journal of Biophysics, 20: 183-191.
15. Manikandakumar, K., K. Gokul Raj, R. Srikumar and S. Muthukumaran, 2010. Classification of Protein Structural Classes using Isoluecine and Lysine Amino Acids, Journal of Proteomics and Bioinformatics, 3(7): 221-229.