|| Volume 2 || Issue 4 || OCTOBER 2017 || ISSN (Online) 2456-0774 ||
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Comparative Study on Structure Based Properties in Different Structural Classes of DNA Binding Proteins

S.C.Jeyakumar[1], J.Thampi Thanka Kumaran[2]
Department of Physics, Nesamony Memorial Christian College, Marthandam 629 165, Tamilnadu, India[1]
Principal, Malankara Catholic College, Mariagiri, Kaliakkavilai 629 153, Tamilnadu, India[2]
scjeyakumar@gmail.com[1], jthampi1558@yahoo.com[2]

Abstract:-**During the process of protein folding, amino acid residues along the primary sequence interact with each other in a cooperative manner to form the stable native structure. Knowledge about inter-residue interactions in protein structures is very helpful to understand the mechanism of protein folding and stability. Understanding the binding mechanism between DNA-binding domains and DNA will be useful for applications in industrial protein engineering. In this comparative study, we have systematically analyzed aminoacid composition and various structure based properties of molecular interactions in different classes of DNA binding proteins based on protein structure. Parameters used in the study are aminoacid composition, long range order, surrounding hydrophobicity, long range interactions, medium range interactions, accessible surface area, ionic interactions, hydrophobic interactions and protein-DNA binding interactions. Structural based properties of different types of DNA binding proteins based on protein structure were statistically analyzed. The results obtained in this work highlight the low value of long range order of all alpha proteins and high value of long range order of all beta proteins. Accessible surface area of polar residue was found to be greater than non-polar residues. There is marked difference in structural based properties of binding and non-binding residues. Ionic interacting residues have difference in structural based properties, compared to ionic non-interacting residues. Similarly difference in structural based properties of hydrophobic interacting residues and hydrophobic non-interacting residues was noticed. Hence DNA binding interactions, ionic interactions and hydrophobic interactions are influenced by the environment in which residues are present.**

Keywords: *Surrounding hydrophobicity, long range order, ionic interaction, hydrophobic interactions, binding residue*

I INTRODUCTION

Maintenance and care of DNA molecules, which are the predominant hereditary material of life, is the sole function DNA binding proteins. DNA binding proteins have specialized functions in DNA metabolism and allow the cell to maintain and replicate its genome. Gene regulation, DNA repair, DNA replication and DNA packaging are some of the biological processes, in which Protein–DNA interactions play vital roles. For understanding the recognition mechanism of protein–DNA complexes, knowledge about DNA-binding residues and structural based properties of the different types of DNA-binding proteins would be very helpful. Understanding protein–DNA recognition mechanism is possible with the availability of experimental data on binding specificity [1] and 3D structures of protein–DNA complexes [2]. Amino acid properties, conservation of residues, contribution of non-covalent interactions and conformational changes of DNA [3][4][5][6][7][8] are some of the areas focused by different researchers.

Theoretical investigations were of great use to understand about DNA binding proteins. Several investigators have stressed the importance of hydrogen bonds, electrostatic, hydrophobic and van der Waals interactions along with weak interactions. To understand the recognition mechanism of protein-DNA complexes, the contribution of energetic terms along with physical and chemical features were used. For understanding the protein-DNA recognition mechanism, Gromiha and his group [9] combined both inter and intramolecular interactions. Amino acid residues along the polypeptide chain interact with each other in a cooperative manner to form the stable native structure, during the process of protein folding. To

understand the mechanism of protein folding and stability, the knowledge about inter-residue interactions in protein structures is very helpful.[10] In the formation of stable secondary structures and a unique tertiary structure for a protein, interactions between amino acid residues of the protein and with the surrounding solvent molecules play an important role. These interactions are usually non-covalent and include hydrogen bonds, ion pairs, van der Waals interactions, and hydrophobic interactions.

Long range order highlights the importance of long-range contacts, which are made by residues that are far in sequence and closer in the 3D structure. Surrounding hydrophobicity provides valuable information with regard to hydrophobic domains, nucleation sites, surface domains, loop sites and the spatial positions of residues in protein molecules. Medium range interactions and long range interactions are required to stabilize the conformation uniquely. Ionic and hydrophobic interactions are also needed for biological activity of proteins. Knowledge about the similarities and differences between structural based properties of the different types of DNA-binding proteins will help to understand about protein-DNA binding mechanism.

In this work, we have used the protein-DNA complexes which were systematically classified into six groups based on protein structure. An attempt was made to find the similarities and differences between structural based properties of DNA-binding proteins, which are grouped on the basis of structure. Structure based properties used in this study are long range order, medium range interactions, long range interactions, surrounding hydrophobicity, average number of 8Å neighbors, average accessible surface area of polar and non-polar residues, ionic interactions, hydrophobic interactions and DNA binding interactions. Structure based properties of protein residues were calculated and from that structure based properties of protein chains were estimated.

All alpha proteins had low value of long range order and all beta proteins have high value of long range order. Structure based properties of binding and non-binding residues have marked difference. Binding residues have lesser value of surrounding hydrophobicity and lesser value of neighbors within 8 Å, compared to non-binding residues. This type of environment in protein favors protein DNA-binding. Ionic non-interacting residues have lower value of surrounding hydrophobicity and lower value of neighbors within 8 Å, compared to ionic interacting residues. Hydrophobic non-interacting residues have lower value of surrounding hydrophobicity and lower value of neighbors within 8A, compared to hydrophobic interacting residues. Hence the environment in which residues are present, influence DNA binding interactions, ionic interactions and hydrophobic interactions.

II MATERIALS AND METHODS

A. Data set

To evaluate the performance of 11 different methods in which online services or standalone programs were available for predicting DNA-binding sites of proteins Nagarajan and co-workers used data sets which were culled as non-redundant with sequence identities of 25%. They have used the SCOP database for structural classification of protein based on their structural classes.PDB codes and chain information for the different types of DNA-binding proteins are available at http://www.iitm.ac.in/bioinfo/DNA-protein/. [11].We have used that PDB codes and chain information for this analysis. Our final data set contains 170 protein chains from six classes with the sequence identity of <25%.

B. Computational Procedure

Clear description of Structure based properties like Medium range interactions, Long range interactions, Long range order, Surrounding hydrophobicity, number of 8Å neighbors & formulae needed to calculate them are available at the server at http://www.iitm.ac.in/bioinfo/pdbparam/, [12] which can be freely accessed. Procedure to calculate Ionic interactions and Hydrophobic interactions are also explained in the same web server.

B.1) Medium and long-range interactions: For a given residue, the surrounding residues within a sphere of 8 Å radii are analyzed in terms of their sequence position. Residues within a window between three and four residues contribute to medium-range interactions and those more than four residues apart contribute to long-range interactions. Both medium range and long range interactions play an important role in the formation of protein structure.

B.2) Number of 8Å contacts: The contacts between amino acid residues in the crystal structure are computed with cutoffs of 8 Å using Cα . Number of residues within 8Å of a particular aminoacid residue gives number of 8Å contacts of that residue.

B.3) Long-range order: LRO is derived from long-range contacts (contacts between two residues that are close in space and far in the sequence) in the protein structure. It is defined as

$$LRO = \sum (n_{ij} / N)$$

$n = 1$ if $i - j > 12$;

n = 0 otherwise

where i and j are the two contacting residues within a distance of 8 Å, and N represents the total number of residues in the protein.

B.4) Surrounding hydrophobicity: The sum of hydrophobic indices assigned to the residues that appear within a distance of 8 Å from the central residue can be used to characterize the hydrophobic behavior of each amino acid residue in the protein environment. It is defined as

$$Hp(i) = \sum_{j=0}^{20} n_{ij} * h_j$$

where n_{ij} is the total number of surrounding residues of type j around the i^{th} residue of the protein, and hj is the hydrophobicity index (kcal/mol) obtained from thermodynamic transfer experiments.

B.5) Accessible surface area: Accessible surface areas of all residues of proteins were calculated using PDB atomic coordinates and NACESS program. From that average accessible surface areas of all residues of different proteins were calculated. Average accessible surface areas of polar residues of a protein was calculated by dividing total accessible surface areas of all polar residues of a protein by total number of polar residues of that protein.

Similarly average accessible surface area of non-polar residues of a protein was calculated by dividing total accessible surface areas of all non-polar residues of a protein by total number of non-polar residues of that protein.

B.6) Ionic interactions: Ionic interactions is contributed by ionic residue pairs Arginine(R), Lysine(K), Histidine(H) : Aspartic Acid(D) Glutamic Acid(E) falling with in a distance of 6Å.

B.7) Hydrophobic interactions: CB atoms of residues of Alanine(A), Valine(V), Leucine(L), Isoleucine(I), Methionine(M), Phenylalanine(F), Tryptophan(W), Proline(P) and Tyrosine(Y) show hydrophobic interactions when they fall within 5Å range.

B.8) DNA binding interaction: The binding sites for a protein–DNA complex can be identified using the following distance criteria. An amino acid residue within a protein is designated as a binding site residue if its side chain or backbone atoms are within a cutoff distance 3.5 Å from any atom in DNA.

III PRESENT STUDY

Aminoacid composition, Long range order, Surrounding hydrophobicity, Medium range interactions, Long range interactions, number of 8 Å neighbors, Accessible surface areas, Ionic interactions, Hydrophobic interactions and DNA binding interactions were calculated using PDB atomic coordinate data files.

A. Computation of amino acid composition

The amino acid composition for each protein has been computed using the number of amino acids of each type and the total number of residues. Amino acid composition is defined as:

$$Comp(i) = \sum_{j=0}^{20} n_i / N$$

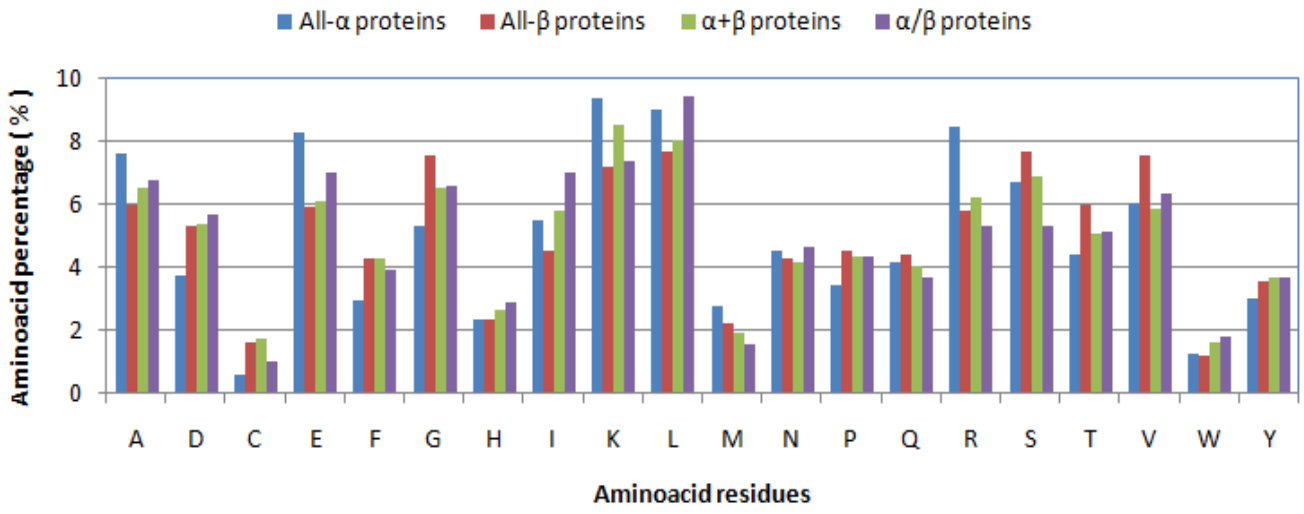

Figure 1. Aminoacid composition of All-α, All-β, α+β and α/β type DNA binding proteins

where j stands for the 20 amino acid residues. n_i is the number of residues of each type and N is the total number of residues. The summation is through all the residues in the particular protein. We have repeated the calculation for all the proteins in all six structural class types of DNA binding proteins.Average of the composition of each amino acid residue in different types of DNA binding proteins are shown in Fig. 1.

Percentage of Basic, Acidic, Neutral, Aromatic and Non-polar groups of aminoacids were calculated for the six types of DNA binding proteins. The result is tabulated in Table 1.

Table 1. Percentage Of Different Groups Of Aminoacid Residues

Protein type	Aminoacid residue group				
	Acidic	**Basic**	**Polar**	**Aromatic**	**Aliphatic**
All-α proteins	12.083	20.239	23.234	7.265	37.062
All-β proteins	11.26	15.407	26.277	9.146	37.91
α+β proteins	11.536	17.495	23.95	9.643	37.145
α/β proteins	12.726	15.627	21.419	9.404	40.531
Multi domain proteins	13.952	16.944	19.099	9.075	40.823
Small proteins	10.202	25.48	28.453	5.956	29.909

Amino acid residues were classified into acidic, basic, polar and non-polar and the composition of different groups of amino acid residue in different types of DNA binding proteins are shown in Figure 2.

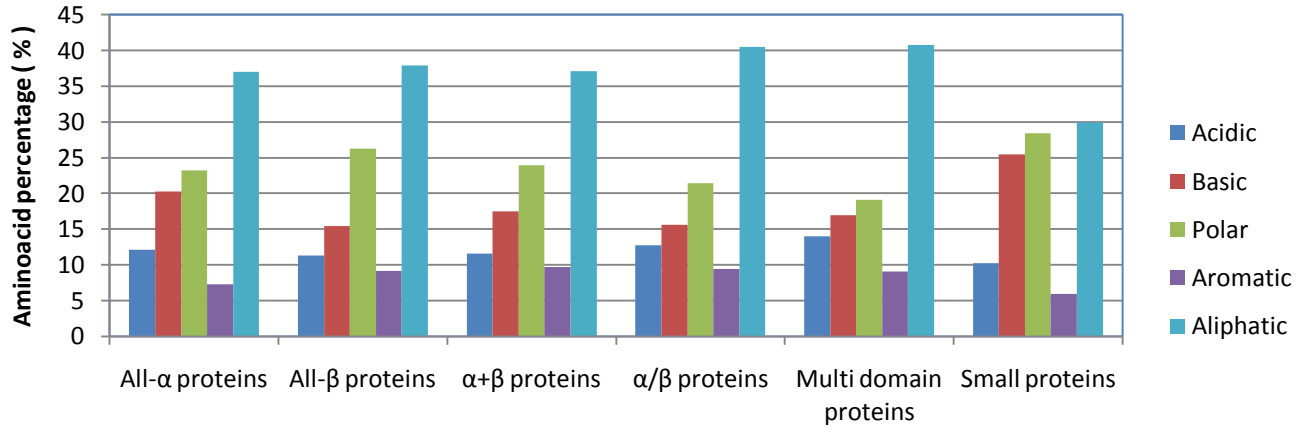

Figure 2. Composition of different groups of Aminoacids in different types DNA binding proteins

From the above graph it is clear that the composition of different groups of aminoacid residues of small proteins is different from other types of DNA binding proteins. Composition of basic and polar groups of aminoacid residue was higher in small proteins whereas the composition of aliphatic group of aminoacid residues was lower in small proteins.

B. Computation of protein properties

Using structure based properties of aminoacid residues, structure based properties of proteins were calculated using the following procedure.[14]

1) Long range order of a protein (LRO) = Sum of long range order of all aminoacid residues of that protein.

2) Ratio of total number of medium range interactions in a protein to total number of residues of a protein (MRR) = Total number of medium range interactions in a protein / Total number of residues of that protein.

3) Ratio of total number of long range interactions in a protein to total number of residues of a protein

(LRR) =Total number of long range interactions in a protein / Total number of residues of that protein.

4) Surrounding hydrophobicity of a protein (Hp) = Average of surrounding hydrophobicity of all aminoacid residues of that protein.
5) Average value of accessible surface area of residues of a protein (ASA) = Sum of accessible surface area of all residues of a protein /Total number of residues of that protein.
6) Average value of accessible surface area of polar residues of a protein (ASAp) = Sum of accessible surface area of all polar residues of a protein /Total number of polar residues of that protein.
7) Average value of accessible surface area of non-polar residues of a protein (ASAnp) = Sum of accessible surface area of all non-polar residues of a protein /Total number of non-polar residues of that protein.
8) Ratio of ionic interacting residues of a protein (RIR) = Total number of ionic interacting residues in a protein /Total number of (R,K,H,D,E) residues of that protein.
9) Ratio of hydrophobic interacting residues (RHR) = Total number of hydrophobic interacting residues in a protein /Total number of (A,V,L,I,M,F,W,P,Y) residues of that protein.
10) 8 A contact number of a protein (n8Ar) = Average of 8 Å contact number of residues
11) Percentage of binding residues of a protein
(PBR)= Total number of binding residues in a protein / Total number of residues of that protein.

Values of structure based properties of DNA binding proteins were tabulated and compared.Correlation analysis method was also used to find the relation between different protein properties.

IV RESULTS & DISCUSSION

Protein-DNA complexes have been classified into all alpha proteins, all beta proteins, alpha plus proteins, alpha by beta proteins, multi domain proteins and small proteins based on protein structure. Average values of structure based properties o protein chain of different proteins were calculated and tabulated below.

Table 2. Average Values Of Structure Based Properties Of Protein Chains

Protein type	**Structure based properties of protein chains**										
	LRO	**MRR**	**LRR**	**Hp**	**n8AR**	**ASA**	**ASAp**	**ASAnp**	**RIR**	**RHR**	**PBR**
All-α proteins	***0.516+ /-0.274***	**2.623+ /-0.399**	***1.749+ /-0.712***	11.216+ /-1.397	9.292+/ -0.604	54.953+ /-7.443	72.785+ /-8.516	33.400+ /-8.877	0.445+ /-0.160	0.332+ /-0.112	14.614+ /-7.940
All-β proteins	**1.684+ /-0.390**	***1.032+ /-0.311***	**4.250+ /-0.709**	12.339+ /-1.165	10.228+ /-0.747	48.930+ /-5.478	65.617+ /-9.878	30.561+ /-4.552	0.488+ /-0.134	0.414+ /-0.092	9.034+/ -7.060
α+β proteins	1.255+ /-0.419	1.716+ /-0.374	3.418+ /-0.784	12.459+ /-1.456	10.079+ /-0.700	49.715+ /-8.617	66.124+ /-9.658	31.792+ /-9.506	0.451+ /-0.143	0.395+ /-0.104	13.278+ /-8.932
α/β proteins	1.370+ /-0.247	1.996+ /-0.287	3.481+ /-0.503	13.286+ /-0.666	10.447+ /-0.378	***44.098+ /-4.058***	***62.056+ /-5.272***	***26.329+ /-4.592***	0.555+ /-0.094	0.438+ /-0.056	7.019+/ -4.145
Multi domain proteins	1.435+ /-0.172	2.040+ /-0.269	3.614+ /-0.394	**13.528+ /-0.698**	**10.638+ /-0.313**	45.834+ /-4.485	65.155+ /-6.158	26.698+ /-2.613	**0.568+ /-0.074**	**0.445+ /-0.040**	***4.740+/ -3.269***
Small proteins	0.646+ /-0.521	1.812+ /-0.443	2.407+ /-0.977	***10.252+ /-2.069***	***9.130+/ -1.022***	**64.990+ /-14.768**	**73.924+ /-12.794**	**49.389+ /-17.972**	***0.368+ /-0.192***	***0.310+ /-0.172***	**24.672+ /-13.470**
Complete set	1.047+ /-0.558	2.027+ /-0.636	2.899+ /-1.163	12.081+ /-1.585	9.867+/ -0.819	51.065+ /-9.212	68.031+ /-9.697	31.886+ /-9.961	0.476+ /-0.150	0.381+ /-0.111	12.125+ /-9.031

Statistical significance of the data was analyzed by calculating P value. For all cases $P < 0.001$, and highly statistical significant nature of the data was established.

Average value of LRO of all alpha type proteins and small proteins was less. Average value of LRO of all beta type proteins and multi domain proteins was high. Already it has been reported[13] that LRR of all beta type protein was in the range 3-8 and all the other class proteins was in the range 1-4.Multi domain proteins have highest average value of Surrounding hydrophobicity of a protein (Hp), 8 Å contact number of a protein, ratio of ionic interacting residues of a protein, ratio of hydrophobic interacting residues of a protein and lowest average value of percentage of binding residues. Small proteins have lowest average value of Surrounding hydrophobicity of a protein (Hp), 8 Å contact number of a protein, ratio of ionic interacting residues of a protein, ratio of hydrophobic interacting residues of a protein and highest average value of percentage of binding residues.

For all types of proteins, average value of accessible surface area of residues of a protein (ASA) was found to be greater than average value of accessible surface area of non-polar residues of a protein (ASAnp) and less than average value of accessible surface area of polar residues of a protein (ASAp). Above result explains the hydrophobic nature of non-polar residues and hydrophilic nature of polar residues.

A. General trend in average values of protein properties

Type of proteins having lower average value of long range order, have lower average value of ratio of total number of long range interactions in a protein to total number of residues of that protein.

Type of proteins having lower average value of long range order and lower average value of ratio of total number of long range interactions in a protein to total number of residues of that protein, have higher value of ratio of total number of medium range interactions in a protein to total number of residues of that protein. This result shows the complementary nature of long range interactions and medium range interactions.

Type of proteins having higher average value of Hp, have higher average value of number of 8Å neighbors. So the regions of proteins having highest packing of atoms have highest surrounding hydrophobicity.

B. General trend in correlation between average values of protein properties

Correlation between values of long range order (LRO), ratio of total number of medium range interactions in a protein to total number of residues of that protein (MRR), ratio of total number of long range interactions in a protein to total number of residues of that protein (LRR), surrounding hydrophobicity (Hp), percentage of binding residues (PBR), ratio of ionic interacting residues (RIR), ratio of hydrophobic interacting residues (RHR) of different types of proteins were found out .

For all types of proteins correlation between LRO and LRR was very high. LRO has high correlation with Hp and value of average number of 8Å neighbors.Significant negative correlation between MRR and LRO and between MRR and LRR was noticed.

LRR had very high correlation with value of average number of 8Å neighbors. Significant correlation between LRR and Hp was noticed.

Average number of 8Å neighbors and Hp have significant negative correlation with percentage of binding residues. High value of hydrophobicity and average 8Å neighbors of protein, may discourage binding between protein and DNA domain.

C. Relation between Surrounding hydrophobicity and other protein properties

For the complete set of 170 DNA-binding proteins, linear regression equation connecting Surrounding hydrophobicity and other protein properties was setup. Using linear regression equation, Surrounding hydrophobicity values of 170 DNA-binding proteins were predicted. Correlation between actual and predicted values of 170 DNA-binding proteins were found out to be maximum (0.839), for the following regression equation

$$Hp = 1.6494 * (MRR) + 1.5029 * (LRR) - 0.1809 * ASA + 0.1089 * ASAp + 0.0776 * ASAnp + 3.74$$

Graph connecting actual value of surrounding hydrophobicity and predicted value of surrounding hydrophobicity is shown in Fig.3.

Percentage error in predicted value of surrounding hydrophobicity in DNA-binding proteins was found to be less than 10% in 144 proteins out of 170 DNA-binding proteins used for the analysis.

Above results show the strong relation between surrounding hydrophobicity, medium range interactions, long range interactions and accessible surface areas.

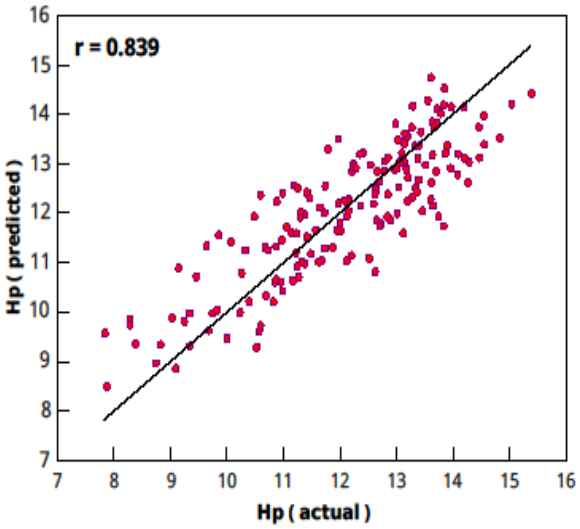

Figure 3. Actual value of surrounding hydrophobicity and predicted value of surrounding hydrophobicity in DNA binding proteins

D. Difference in residue properties

Properties of residues such as percentage of non-zero LRO values, average LRO values, average number of MRI residues, average number of LRI residues, average surrounding hydrophobicity and average number of 8Å neighbors of DNA binding resides and non- binding resides were computed.

Difference between properties of DNA-binding resides and non-binding resides were compared. Difference between properties of ionic interacting resides and ionic non-interacting resides were compared.

Similarly difference between properties of hydrophobic interacting resides and hydrophobic non-interacting resides were compared.

E. Calculation of percentage of binding in different types of aminoacid residues

Aminoacid composition of 20 aminoacids of all residues of 170 DNA-binding proteins were calculated. Similarly aminoacid composition of 20 aminoacids of binding residues of 170 DNA-binding proteins were calculated. Using aminoacid composition of binding residues and aminoacid composition of all residues, percentage binding of all aminocid residues were calculated.

Percentage of binding residues of 20 aminoacid residues are plotted in Figure 4.

Figure 4 graph shows that, the percentage of binding is higher than 10% in Arginine, Lysine, Histidine which are large and basic. Large and polar aminoacids Asparagine and Glutamine also have percentage of binding higher than 10%. Aminoacids Serine and Threonine with hydroxyl side groups also have percentage of binding higher than 10%. Large and non-polar aminoacids Tryptophan and Tyrosine have percentage of binding higher than 10%.

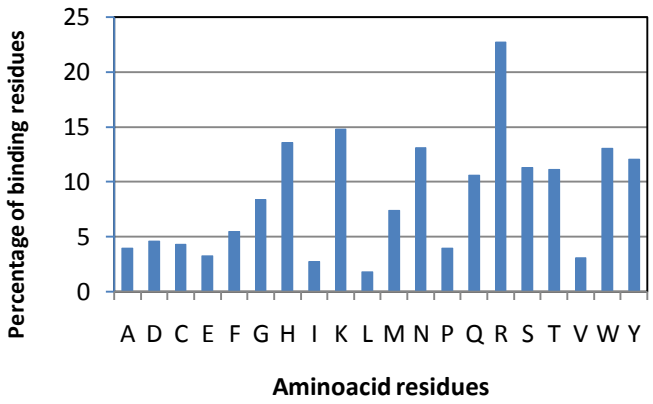

Figure 4.Percentage of binding residues for 20 amino acids in DNA binding proteins

Above result shows that, the nucleic acids prefer the association of basic residues. In addition to that easy association of aminoacids which are bigger in size with nucleic acid is explained.

Acidic aminoacids Aspartic acid and Glutamic acid have percentage of binding lower than 5%. Small aminoacids Alanine, Cysteine, Isoleucine, Leucine, Proline and Valine also have percentage of binding lower than 5%.

Above result shows that the nucleic acids do not prefer the association of acidic residues. In addition to that it is explained that the aminoacids which are smaller in size are unsuitable for easy association of with nucleic acid.

F. Comparison of properties of binding and non-binding residues in DNA binding proteins

Percentage of residues having non-zero long range order value was lower in binding residues than in non-binding residues.

Average value of medium range interactions and long range interactions was lower in binding residues than in non-binding residues.

Average surrounding hydrophobicity values and number of 8Å neighbors of binding residues were lesser than non-binding residues.

Above results show that protein DNA binding is favored in regions were atomic packing of proteins is less.

G. Comparison of properties of Ionic interacting (R,K,H,D,E) residues and Ionic non-interacting (R,K,H,D,E) residues of DNA binding proteins

For all types of DNA binding proteins, percentage of non-zero LRO values was higher in Ionic interacting (R,K,H,D,E) residues compared to Ionic non-interacting (R,K,H,D,E) residues.

Average LRO value was higher in Ionic interacting (R,K,H,D,E) residues compared to non-interacting (R,K,H,D,E) residues.

Average surrounding hydrophobicity and number of 8Å neighbors was higher in Ionic interacting (R,K,H,D,E) residues compared to non-interacting (R,K,H,D,E) residues.

Average value of MRI and LRI was higher in Ionic interacting (R, K, H, D, E) residues than in non-interacting (R,K,H,D,E) residues.

Above results show that ionic interactions are favored in regions were atomic packing of proteins is high.

H. Comparison of properties of Hydrophobic interacting (A,V,L,I,M,F,W,P,Y) residues and Hydrophobic non-interacting (A,V,L,I,M,F,W,P,Y) residues of DNA binding proteins

For all types of DNA binding proteins, percentage of non-zero LRO values was higher in Hydrophobic interacting (A,V,L,I,M,F,W,P,Y) residues compared to non-interacting (A,V,L,I,M,F,W,P,Y) residues.

Average LRO value was higher in Hydrophobic interacting (A,V,L,I,M,F,W,P,Y) residues compared to non-interacting (A,V,L,I,M,F,W,P,Y) residues.

Average LRI, surrounding hydrophobicity and number of 8Å neighbors was higher in Hydrophobic interacting (A,V,L,I,M,F,W,P,Y) residues compared to non-interacting (A,V,L,I,M,F,W,P,Y) residues.

Average MRI was lower in Hydrophobic interacting (A,V,L,I,M,F,W,P,Y) residues compared to non-interacting (A,V,L,I,M,F,W,P,Y) residues.Above results show that ionic interactions are favoured in regions were atomic packing of proteins is high.

I. Comparison of aminoacid percentage of buried residues and all residues

An aminoacid residue is considered as buried residue if the accessible surface area of that residue is less than 7. Buried residues occur at the interior of proteins.

To probe the interior of proteins, composition of aminoacids of buried residues were found out. From that percentages of aminoacids of buried residues were found out. Similarly composition of aminoacids of all residues were found out. From that percentages of aminoacids of all residues were found out.

To compare the composition of buried residues with the composition of all residues, a bar chart was plotted for percentages of aminoacids of buried residues and all residues.

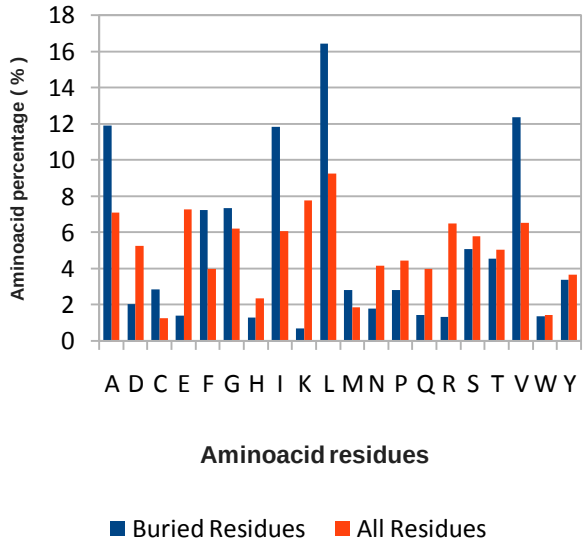

Figure 5. Percentages of aminoacids of buried residues and all residues in DNA binding proteins

From the above chart it is found out that the percentages of non-polar aminoacids Alanine, Phenylalanine, Isoleucine, Leucine and Valine are greater in buried regions. These aminoacids prefer interior of proteins. Negatively charged amino acids Aspatric acid, Glutamic acid and positively charged aminoacids Lysine, Arginine are lesser in buried regions. These aminoacids want to avoid interior of proteins.

V. CONCLUSION

Structure based properties of different types of DNA binding proteins were found out and tabulated. Correlation coefficient, between different structure based properties, were found out.

Average value of Surrounding hydrophobicity values of DNA binding residues were lesser than average value of surrounding hydrophobicity values of non-binding residues. This shows that protein DNA binding is favored in regions were atomic packing of proteins is less.

For both ionic and hydrophobic interactions, average value of Surrounding hydrophobicity values of interacting residues were greater than average value of surrounding hydrophobicity values of non-interacting residues for six types of DNA-binding proteins, classified on the basis of protein structure. This shows that ionic and hydrophobic interactions are favored in regions were atomic packing of proteins is high.

REFERENCES

[1] Prabakaran,P., An,J., Gromiha,M.M., Selvaraj,S., Uedaira,H.,Kono,H. and Sarai,A. Thermodynamic database for protein-nucleic acid interactions (ProNIT). Bioinformatics, 2001; 17, 1027–1034.

[2] Berman,H.M., Kleywegt,G.J., Nakamura,H. and Markley,J.L. The protein data bank at 40: reflecting on the past to prepare for the future. Structure, 2012; 20, 391–396.

[3] Gromiha,M.M., Munteanu,M.G., Simon,I. and Pongor,S. The role of DNA bending in Cro protein-DNA interactions. Biophys. Chem., 1997; 69, 153–160.

[4] Gromiha,M.M. Influence of DNA stiffness in protein-DNA recognition. J. Biotechnol., 2005; 117, 137–145.

[5] Ahmad,S., Keskin,O., Sarai,A. and Nussinov,R. Protein-DNA interactions: structural, thermodynamic and clustering patterns of conserved residues in DNA-binding proteins. Nucleic Acids Res., 2008; 36, 5922–5932.

[6] Prabakaran,P., Siebers,J.G., Ahmad,S., Gromiha,M.M., Singarayan,M.G. and Sarai,A. (2006) Classification of protein-DNA complexes based on structural descriptors. Structure, 2006; 14, 1355–1367.

[7] Ahmad,S., Gromiha,M.M. and Sarai,A. Analysis and Prediction of DNA-binding proteins and their binding residues based on composition, sequence and structure information. Bioinformatics, 2004; 20, 477–486.

[8] Ahmad,S. and Sarai,A. PSSM based prediction of DNA-binding sites in proteins. BMC Bioinformatics, 2005; 6, 33.

[9] Gromiha,M.M., Siebers,J.G., Selvaraj,S., Kono,H. and Sarai,A. Intermolecular and intramolecular readout mechanisms in protein-DNA recognition. J. Mol. Biol., 2004; 337, 285–294.

[10] Gromiha MM, Selvaraj S. Inter-residue interactions in protein folding and stability. Prog Biophys Mol Biol. 2004;86:235–77

[11] Nagarajan,R., Shandar Ahmad and Michael Gromiha,M. Novel approach for selecting the best predictor for identifying the binding sites in DNA binding proteins. Nucleic Acids Research, 2013 ;41, 7606–7614.

[12] Nagarajan R, Archana A, Mary Thangakani A, Jemimah S, Velmurugan D, Michael Gromiha M. PDBparam: Online Resource for Computing Structural Parameters of Proteins. Bioinform Biol Insights. 2016; 10: 73–80.

[13] Michael Gromiha M and Selvaraj S. Comparison between Long-range Interactions and Contact Order in Determining the Folding Rate of Two-state Proteins: Application of Long-range Order to Folding Rate Prediction J. Mol. Biol. (2001) 310, 27-32

[14] Jeyakumar S.C., and Thampi Thanka Kumaran J. Study on Different types of Structural based Properties in Human Membrane Proteins. IJSRST; VOL. 3; NO. 7,2017