

COMPARATIVE ANALYSIS ON STRUCTURE BASED PROPERTIES OF MESOPHILIC AND THERMOPHILIC PROTEINS

S.C.Jeyakumar¹, J.Thampi Thanka Kumaran²

Abstract - The basis of high thermostability in thermophilic proteins is complex and not yet fully understood. It is clear that hydrophobic interactions, ion pair networks, disulfide bonds, and amino acid composition are important factors influencing thermal stability of proteins. Understanding the thermal stability mechanism will be useful for applications as industrial protein engineering. In this work, we have compared differences and similarities between various molecular interactions of by computing short range interactions, medium range interactions, long range interactions, ionic interactions and hydrophobic interactions and amino acid composition for 373 thermophilic and mesophilic pairs. Structural based properties like Contact order, Total contact distance, surface hydrophobicity and hydrophobic free energy were used for analysis. Surrounding hydrophobicity values of all residues of a particular type of protein were taken together for analysis. Similar treatment was given to Average gain in surrounding hydrophobicity values and Long Range Order values of all residues. Percentage of amino acid residues taking part in ionic interaction and hydrophobic interaction was found out to be different for both types of proteins. For both ionic and hydrophobic interactions, average value of LRO values and surrounding hydrophobicity values of interacting residues was found to be greater than the respective values of noninteracting residues. Thermostability of proteins is a complex property with contributions from different types of interactions. Understanding different mechanisms, used by proteins to achieve stability at high temperature would be useful for design of thermostable proteins.

Key words: Surrounding hydrophobicity; Long Range Order; ionic interactions; hydrophobic interactions; mesophilic; thermophilic.

I. INTRODUCTION

Proteins produced by thermophilic organisms have extreme thermal stability and they withstand up to the temperature of 120⁰. Understanding the reasons for thermal stability of thermophilic proteins is an important problem and it would help to design stable proteins. Several investigations have been carried out to understand the features influencing the stability of thermophilic proteins and are surveyed in detail.¹⁻⁹

Based on a dataset of 373 protein families, Gromiha et al¹⁰ reported that hydrophobic environment is the major factor for the stability of thermophilic proteins and reported that 80% of thermophilic proteins showed higher surrounding hydrophobicity values than their mesophilic counterparts. Hydrophobic free energy was also important and higher in 62% of thermophilic proteins. During the process of protein folding, the amino acid residues along the polypeptide chain interact with each other in a cooperative manner to form the stable native structure. The knowledge about the interactions in between the residues in protein structures is very helpful to understand the mechanism of protein folding and stability.^{11,12}

Understanding the relationship between protein structure and protein function is a longstanding goal in molecular and computational biology. The development of structure-based parameters has helped to relate the structure with the function of a protein. Using web-based tool, PDBparam¹³, more than 50 structure-based features for any given protein structure can be computed. In this work we analyzed the differences and similarities between 12 structural based properties of mesophilic and thermophilic proteins. Long range, medium range and short range interactions were calculated and the differences and similarities were analysed. Surrounding hydrophobicity, average gain in surrounding hydrophobicity and long range order of all residues of mesophilic and thermophilic proteins were used for analysis. Ionic interactions and hydrophobic interactions of mesophilic and thermophilic proteins were used for analysis. In addition to that differences in surrounding hydrophobicity, average gain in surrounding hydrophobicity and long range order between interacting and noninteracting residues classified based on both ionic interactions and hydrophobic interactions were analyzed.

II. MATERIALS AND METHODS

Dataset

¹ Department of Physics, Nesamony Memorial Christian College, Marthandam - 629 165, Kanyakumari District, Tamilnadu, India

² Department of Physics, Malankara Catholic College, Mariagiri, Kaliakkavilai - 629 153, Kanyakumari District, Tamilnadu, India

PDB coordinates of 373 mesophilic proteins and their thermophilic counterparts used by Gromiha et al¹⁰ to investigate thermal stability of thermophilic proteins was used to find the similarities and differences between structural based properties of both type of proteins.

Structure based properties used in this study are Short range interactions, Medium range interactions, Long range interactions, Contact order, Long range order, Total contact distance, Ionic interactions, Hydrophobic interactions, Surrounding hydrophobicity, Average gain in surrounding hydrophobicity, Surface hydrophobicity and Hydrophobic free energy.

Computational procedure

Clear description of Structure based properties and formulae needed to calculate them are available at the server at <http://www.iitm.ac.in/bioinfo/pdbparam/>, which can be freely accessed.

Formulae of Structure based properties used in the analysis are given below

1. Contact order $CO = \sum (S_{ij} / (L * N))$

L = total number of residues in the protein

N = total number of contacts.

S_{ij} = sequence separation between residue i and j

2. Surface hydrophobicity $\Phi_{\text{surface}} = \sum (s_i * \varphi_i / s_p)$

s_i = solvent accessible area occupied by amino acid 'i'

φ_i = hydrophobicity value assigned to amino acid

s_p = total solvent accessible area of protein

3. Long range order $LRO = \sum (n_{ij} / N)$; $n=1$ if $|i-j| > 12$; $n=0$ otherwise

Where i and j are two contacting residues within dist 8Å

N = total number of residues in the protein

4. Total Contact Distance $TCD = CO * LRO$

5. Surrounding hydrophobicity $Hp(i) =$

n_{ij} = total number of surrounding residues of type j around ith residue of the protein

h_j = hydrophobicity index (kcal/mol) given by Tanford and Jones(1971)

6. Average gain in surrounding hydrophobicity =

Surrounding hydrophobicity in folded state - Surrounding hydrophobicity in unfolded state

7. Short range interactions Residues lying within a distance of 2 residues from the central residue contribute to short range interactions

8. Medium range interactions Residues lying within a distance of 3 or 4 residues from the central residue contribute to medium range interactions.

9. Long range interactions Those residues that are 4 residues away from the central residue contribute to long range interactions.

10. Ionic Interactions: Ionic residue pairs(R,K,H) : (D,E) falling within a distance of 6Å contribute to ionic interactions

11. Hydrophobic interactions CB residues of A,V,L,I,M,F,W,P,Y show hydrophobic interactions when they fall within 5Å range.

12. Hydrophobic free energy $G_{hy} = \sum \Delta \sigma_i [Ai(\text{folded}) - Ai(\text{unfolded})]$
 $Ai(\text{folded})$ = ASA of each atom in folded state
 $Ai(\text{unfolded})$ = ASA of each atom in unfolded state
 σ_i = atomic salvation parameter

III. PRESENT STUDY

Long range order, Ionic interactions, Hydrophobic interactions, Surrounding hydrophobicity, Average gain in surrounding hydrophobicity of 373 mesophilic proteins and their thermophilic counterparts were calculated using PDB coordinates of 373 mesophilic proteins and their thermophilic counterparts.

Surrounding hydrophobicity of protein were calculated by calculating the average of surrounding hydrophobicity of aminoacid residues of the respective proteins. Similarly Average Gain in Surrounding hydrophobicity of protein is calculated by calculating the average of Average Gain in Surrounding hydrophobicity of aminoacid residues.

Total number of Short range interactions, Medium range interactions, Long range interactions, Ionic interactions, Hydrophobic interactions of a particular protein were normalised by dividing with total number of aminoacid residues of a particular protein.

PDB coordinates of 373 mesophilic proteins and their thermophilic counterparts were uploaded to www.iitm.ac.in/bioinfo/pdbparam/ to calculate Contact order, Total contact distance, Surface hydrophobicity and Hydrophobic free energy.

Values of structure based properties of mesophilic proteins and thermophilic proteins were compared using graphical method and correlation analysis method.

Comparison of Structural based properties of both mesophilic and thermophilic proteins

Average values of Structural based properties of both mesophilic proteins and thermophilic proteins are listed in table 1.

Table 1. Mean value of Structure based properties of mesophilic and thermophilic proteins

Structure based properties of proteins	Contact Order	Total Contact Distance	Surface hydrophobicity	Long Range Order	Surrounding Hydrophobicity	Average gain in Surrounding Hydrophobicity	Short Range Interactions per residue	Medium Range Interactions per residue	Long Range Interactions per residue	Ionic Interactions per residue	Hydrophobic Interactions per residue	Hydrophobic Free Energy per residue Kcal/mol
Mesophilic proteins	14.46	0.89	0.575	1.521	12.719	8.977	1.974	0.955	1.912	0.2	0.144	-0.469
Thermophilic proteins	14.543	0.899	0.555	1.531	13.402	9.454	1.975	0.955	1.93	0.243	0.158	-0.441

There no significant difference in averages of contact order, total contact distance and long range order of mesophilic proteins and thermophilic proteins. This may be due to the structural similarity between mesophilic proteins and thermophilic proteins.

By comparing the averages of structure based properties of mesophilic proteins and thermophilic proteins, we found that thermophilic proteins have higher average values of surrounding hydrophobicity, average gain in surrounding hydrophobicity, hydrophobic interaction per residue and ionic interaction per residue.

Thermophilic proteins have lower average values of surface hydrophobicity, which can be explained on the basis of published result. Amino acid composition of the interior and exterior of variety of mesophilic and thermophilic proteins available on the Protein Data Bank were compared by Fukuchi and Nishikawa¹⁴. They have reported that the frequency of polar uncharged residues (Asn, Ser, and Thr) on the surface of thermophilic proteins was much lower than mesophilic proteins. Conversely, the proportion of polar charged residues (Asp, Glu, Arg, and Lys) on the surface of thermophilic proteins was higher than mesophilic proteins. Due to the contribution of polar charged residues which have lower hydrophobicity, surface hydrophobicity of thermophiles became lower.

Table 2. Comparison of Structure based properties of mesophilic and thermophilic proteins

Structure based properties of proteins	Contact Order	Total Contact Distance	Surface hydrophobicity	Long Range Order	Surrounding Hydrophobicity	Average gain in Surrounding Hydrophobicity	Short Range Interactions per residue	Medium Range Interactions per residue	Long Range Interactions per residue	Ionic Interactions per residue	Hydrophobic Interactions per residue	Hydrophobic Free Energy per residue Kcal/mol	Total number of residues
Number and percentage of protein pairs in which mesophiles having higher property values	167 44.8%	159 42.6%	254 68.1%	169 45.3%	77 20.6%	91 24.4%	123 33%	174 46.7%	177 47.5%	124 33.2%	135 36.2%	142 38.1%	181 48.5%
Number and percentage of protein pairs in which thermophiles having higher property values	206 55.2%	208 55.8%	114 30.6%	198 53.1%	296 79.4%	282 75.6%	85 22.8%	187 50.1%	192 51.5%	239 64.1%	224 60.1%	229 61.4%	85 22.8%
Number and percentage of protein pairs in which property values are same	0	6 1.6%	5 1.3%	6 1.6%	0	0	165 44.2%	12 3.2%	4 1.1%	10 2.6%	14 3.8%	2 0.5%	107 28.7%

Table 2 gives the pair wise comparison between properties of mesophilic protein and thermophilic proteins. Table 2 shows that 79.36 % of thermophilic proteins of the data set have higher surrounding hydrophobicity compared to their mesophilic counterparts. Similarly 75.6 % of thermophilic proteins of the data set have higher average gain in surrounding hydrophobicity compared to their mesophilic counterparts. From above two important results it was clear that hydrophobic interactions play the most important part in thermal stabilization of thermophilic proteins.

Very high positive correlation was found between Contact order and Total contact distance (0.94773 & 0.94334) for mesophilic proteins and thermophilic proteins. Surrounding hydrophobicity and average gain in surrounding hydrophobicity also showed very high positive correlation (0.97715 & 0.97265) for mesophilic proteins and thermophilic proteins. Long range order of protein had very high positive correlation (0.96209 & 0.9348) for mesophilic proteins and thermophilic proteins with long range interaction per residue.

Comparison of number of interactions in mesophilic and thermophilic proteins

Long range interactions, Medium range interactions, Short range interactions, Ionic interactions and Hydrophobic interactions of 373 mesophilic proteins and their thermophilic counterparts were calculated using PDB coordinates of 373 mesophilic proteins and their thermophilic counterparts.

Table 3. Number of interactions in mesophilic and thermophilic proteins

Interaction type	Mesophilic proteins	Thermophilic proteins
Number of Long Range Interactions	129418	128838
Number of Medium Range Interactions	63985	63409
Number of Short Range Interactions	129234	127986
Total number of residues of 373 proteins	65287	64641
Number of ionic interactions	9903	10558
Number of hydrophobic interactions	13708	16406

Total number of residues of 373 mesophilic proteins was greater than total number of residues of 373 thermophilic proteins. So the total number of long range interactions, medium range interactions and short range interactions of mesophilic proteins were greater than thermophilic protein values.

Eventhough total number of residues of 373 thermophilic proteins was lesser than total number of residues of 373 mesophilic proteins, total number of ionic interactions and hydrophobic interactions of thermophilic proteins were greater than total number of ionic interactions and hydrophobic interactions of mesophilic proteins. This showed that ionic interactions and hydrophobic interactions were prominent in thermophilic proteins than in mesophilic proteins.

Comparison of different types of interactions

Long range interactions of 373 mesophilic proteins were calculated. 20 x 20 matrix of residues taking part in above interactions was formed by having first residue in row and second residue along column. From that matrix, type of residue pairs and their occurrence in interactions were analysed. Percentage of amino acid residues appearing as the first residue in long range interactions were calculated and tabulated. Above procedure was repeated for both medium range interactions and short range interactions of 373 mesophilic proteins. Same treatment was followed for 373 thermophilic proteins also.

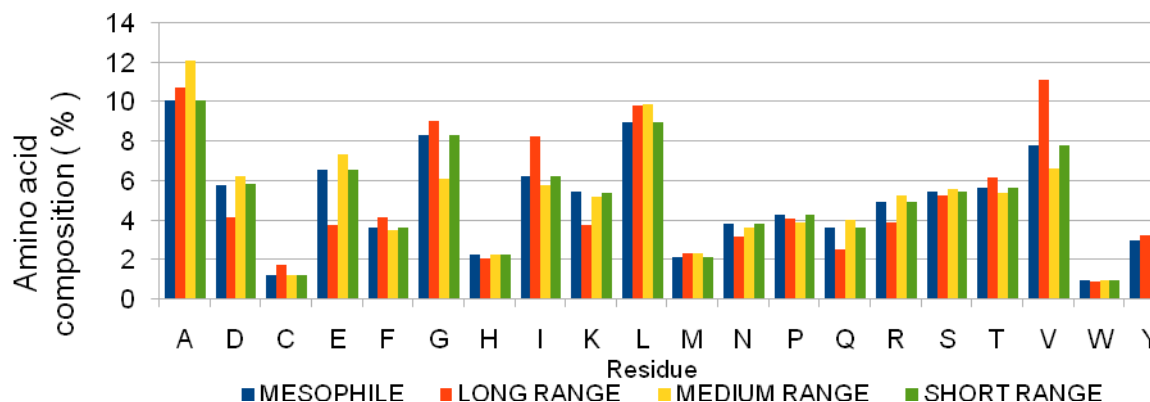


Figure 1. Amino acid composition of mesophilic protein and composition of amino acids in different types of interactions

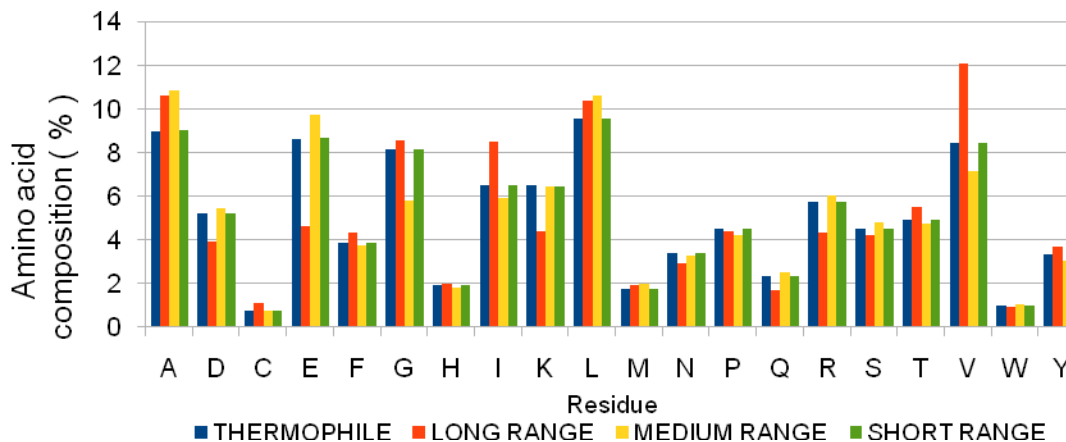


Figure 2. Amino acid composition of thermophilic protein and composition of amino acids in different types of interactions

Amino acid composition of mesophiles and thermophiles differ. Occurrence of CYS was less in thermophiles than mesophiles. On the other hand, the occurrence of VAL, ILE was higher in thermophiles than in mesophiles. Further, the charged residues, LYS, ARG, and GLU had significantly higher occurrence in thermophilic proteins than in mesophilic ones.¹⁵

For both mesophiles and thermophiles, percentage of amino acids taking part in short range interaction was found to be equal to their percentage in respective proteins.

For both mesophiles and thermophiles and percentages of ALA, LEU, GLY, ILE and VAL taking part in long range interaction was greater than their percentages in proteins. All the above five amino acids have aliphatic side groups. On contrast percentage of ASP, ASN, GLU, GLN (acidic amino acids and their amides) and LYS, ARG (basic amino acids) taking part in long range interaction was lower than their percentage in proteins.

For both mesophiles and thermophiles and percentage of GLY, ILE and VAL taking part in medium range interaction was lower than their percentage in proteins. On contrast, percentage of ASP, ASN, GLU, GLN, LYS, ARG taking part in medium range interaction were higher than their percentages in proteins.

Amino acids GLY, ILE and VAL prefer long range contacts and ASP, ASN, GLU, GLN, LYS, ARG prefer medium range contacts. ALA and LEU prefer both long range contacts and medium range contacts.

Analysis of surrounding hydrophobicity values of mesophiles and thermophiles

Surrounding hydrophobicity values of 65287 residues of 373 mesophiles 64641 residues of 373 thermophiles were analysed.

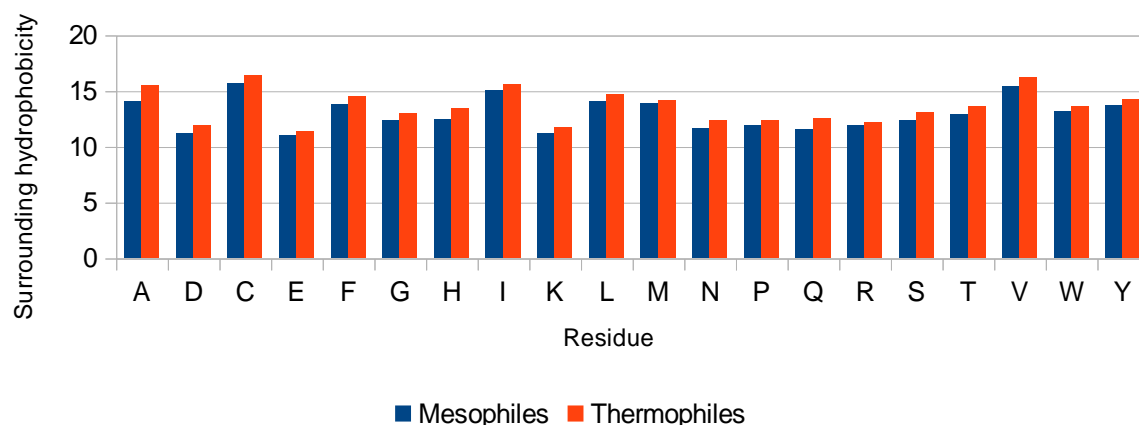


Fig 3. Average value of surrounding hydrophobicity different amino acid residues of mesophilic proteins and thermophilic proteins

Average value of surrounding hydrophobicity values of all types of amino acid residues of thermophilic proteins were greater than mesophilic proteins.

Analysis of average gain in surrounding hydrophobicity values of mesophiles and thermophiles

Average gain in surrounding hydrophobicity values of residues of 373 mesophiles and residues of 373 thermophiles were analysed.

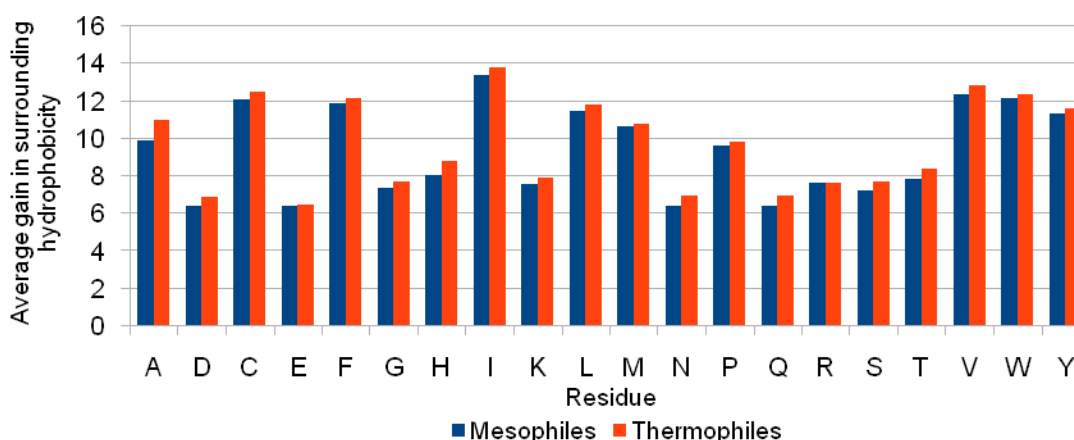


Fig 4. Average value of average gain in surrounding hydrophobicity values of different amino acid residues of mesophilic proteins and thermophilic proteins

Average gain in surrounding hydrophobicity values of different amino acid residues of thermophilic proteins were greater than mesophilic proteins.

Analysis of Long Range Order values of mesophiles and thermophiles

Long Range Order values of all residues of 373 mesophiles and 373 thermophiles were analysed.

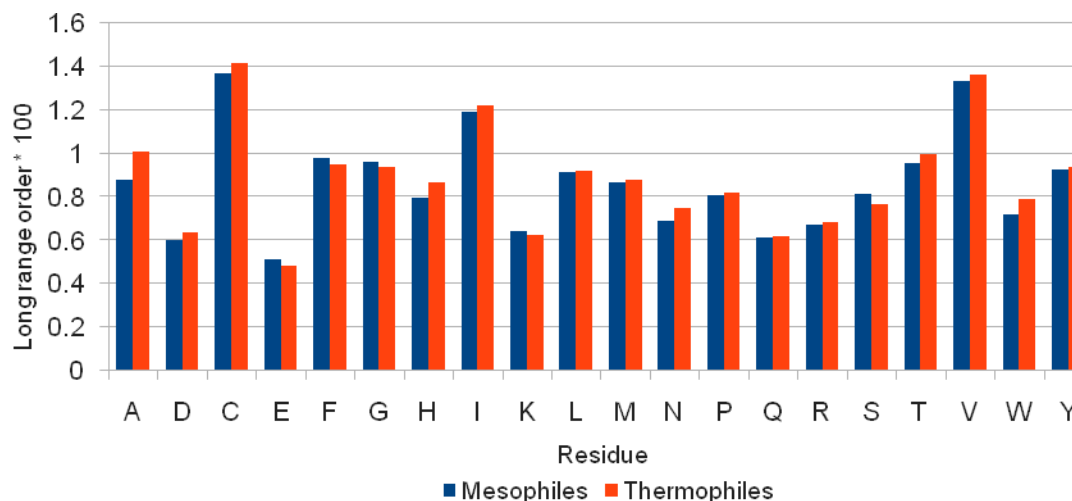


Fig 5. Average value of long range order values of different amino acid residues of mesophilic proteins and thermophilic proteins

For amino acid residues ALA, ASP, CYS, HIS, ILE, LEU, MET, ASN, PRO, GLN, ARG, THR, VAL, TRP, TYR Long Range Order values of thermophilic proteins were greater than mesophilic proteins. For amino acid residues, GLU, PHE, GLY, LYS, SER, Long Range Order values of thermophilic proteins were lesser than mesophilic proteins.

Analysis of Ionic interactions of mesophiles and thermophiles

In mesophilic proteins percentage of amino acid residues taking part in Ionic interaction was 13.38 %. In thermophilic proteins percentage of amino acid residues taking part in Ionic interaction was 15.95 %. For both mesophiles and thermophiles, average LRO values and average surrounding hydrophobicity values of ionic interacting residues were slightly greater than ionic noninteracting residues.

Analysis of Hydrophobic interactions of mesophiles and thermophiles

In mesophilic proteins percentage of amino acid residues taking part in hydrophobic interaction was 22.1 %. In thermophilic proteins percentage of amino acid residues taking part in ionic interaction was 23.47 %. For both mesophiles and thermophiles, average LRO values and average surrounding hydrophobicity values of Hydrophobic interacting residues were slightly greater than hydrophobic noninteracting residues.

IV. CONCLUSION

Surrounding hydrophobicity values of 79.36 % thermophilic proteins were higher compared to their mesophilic counterparts. Number of ionic and hydrophobic interactions was higher in thermophilic proteins than in mesophilic proteins. Percentage of amino acid residues taking part in Ionic interaction and hydrophobic interaction were higher in thermophilic proteins than in mesophilic proteins. For both ionic and hydrophobic interactions, average value of LRO values of interacting residues were greater than average value of LRO values of noninteracting residues for both mesophilic and thermophilic proteins. 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 noninteracting residues for both mesophilic and thermophilic proteins.

REFERENCES

- [1] Ladenstein R, Antranikian G. Proteins from hypothermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv Biochem Eng Biotechnol* 1998;61:37-85.
- [2] Jaenicke R, Bohm G. The stability of proteins in extreme environments. *Curr Opin Struct Biol* 1998;8:638-648.
- [3] Szilagyi A, Zavodszky P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of comprehensive survey. *Struct Fold des* 2000;8:493-504.
- [4] Kumar S, Nussinov R. How do thermophilic proteins deal with heat? *Cell Mol Life Sci* 2001;58:1216-1233.
- [5] Yano JK, Poulos TL. New understandings of thermostable and peizostable enzymes. *Curr Opin Biotechnol* 2003;14:360-365.
- [6] Razvi A, Scholtz JM. Lessons in stability from thermophilic proteins. *Proteins Sci* 2006;15:1569-1578.
- [7] Gromiha MM, Oobatake M, Sarai A. Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophys Chem* 1999;82:51-67.
- [8] Kumar S, Tsai CJ, Nussinov R. Factors enhancing protein thermostability. *Proteins Eng* 2000;13:179-191.
- [9] Saraboji K, Gromiha MM, Ponnuswamy MN. Importance of main chain hydrophobic free energy to the stability of thermophilic proteins. *Int J Biol Macromol* 2005;35:211-220.
- [10] Gromiha MM, Pathak MC, Saraboji K, Ortlund EA, Gaucher EA. Hydrophobic environment is a key factor for the stability of thermophilic proteins. *Proteins*. 2013; 81: 715-21.

- [11] Gromiha MM, Important inter-residue contacts for enhancing the thermal stability of thermophilic proteins. *Biophys Chem* 2001;91:71-77.
- [12] Gromiha MM, Selvaraj S. Inter-residue interactions in protein folding and stability. *Prog Biophys Mol Biol* 2004;86:235-277.
- [13] 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.
- [14] Fukuchi S and Nishikawa K. Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *J. of Mol. Bio.*, 2001, 309, 835–843.
- [15] Michael Gromiha M, Xavier Suresh M . Discrimination of mesophilic and thermophilic proteins using machine learning algorithms. *Proteins* 2008; 70: 1274 – 1279.