International Standard Serial Number (ISSN): 2249-6807

International Journal of Institutional Pharmacy and Life Sciences 2(2): March-April 2012

INTERNATIONAL JOURNAL OF INSTITUTIONAL PHARMACY AND LIFE SCIENCES

Life Sciences **Research Article……!!!**

Received: 12-04-2012; Accepted: 21-04-2012

RELATIVE SOLVENT ACCESSIBILITY AS A PARAMETER FOR COMPARATIVE ANALYSIS OF MESOPHILIC AND THERMOPHILIC PROTEIN SEQUENCES

Sandhya Verma, Abhigyan Nath, Radha Chaube*

Department of Zoology/Bioinformatics, Mahila Mahavidyalaya, Banaras Hindu University, Varanasi-221005, U.P., India

Keywords:
Temperature adaptation, relative solvent accessibility, amino acid property groups

For Correspondence:

Dr. Radha Chaube

Department of Zoology/Bioinformatics, Mahila Mahavidyalaya, Banaras Hindu University, Varanasi-221005, U.P., India

E-mail:
chauberadha@rediffmail.com

ABSTRACT

Thermophilic organisms are able to survive at extremes of temperature. To thrive at the extremes of environment these organisms show adaptation at the molecular level. In this study we did a comparative analysis of differentially exposed regions of thermophilic and mesophilic proteins by taking solvent accessibility as the parameter. We found that there are no significant differences between the % of residues belonging to buried, intermediate and exposed regions between the two categories. We also confirmed the results of previous studies on compositional bias in terms of amino acid composition and amino acid property group composition existing between mesophilic and thermophilic proteins.

INTRODUCTION

Microorganisms requiring extreme environments for growth are called extremophiles and the enzymes they produce are extremozymes (1), the extreme conditions i.e high or low temperature, high or low pH, high salinity, high metal concentrations, very low nutrient content, very low water activity, high radiation, high pressure and low oxygen tension. Extremophiles are structurally adapted at the molecular level to withstand these harsh conditions, the biocatalysts, called extremozymes, produced by these microorganisms and are proteins that function under extreme conditions, extremophiles that have been identified to date belong to the domain of the archaea, extremophiles from the eubacterial and eukaryotic kingdoms have also been identified recently and characterized. The usual proteinaceous biocatalysts are functional under mild conditions of pH, temperature and pressure of the aquatic medium. The consideration that different strategies for thermal adaptation might have been exploited by organisms would be evolutionarily distant and that merging results obtained from thermophilic archaea and eubacterial might have hindered the previous attempts to identify the determinants of protein stability.

Understanding the determinants and properties of protein thermal stability is still an unsolved problem in protein biochemistry. A large number of investigations have been carried out in the past two decades in order to understand the factors that contribute to the thermal stability of thermophilic proteins. Proteins come in a wide variety of shapes and folds and possess a wide range of thermal stabilities. Proteins from thermophilic organisms usually exhibit substantially higher intrinsic thermal stabilities than their counterparts from mesophilic organisms (2; 3) while retaining the basic fold characteristic of the particular protein family. Protein thermostability has been vigorously studied in the biophysical and biotechnological research areas (4; 5), because protein instability at high temperature is one of the main bottlenecks in extending the application of protein (6).

MATERIALS AND METHODS

Construction of a data set

We used the protein structures of 16 pairs of mesophilic and thermophillic proteins as taken from [7], the structural data was downloaded from the protein data bank.

PDB_ID	ACTUAL TYPE	PDB_ID	ACTUAL TYPE
1tmy	T	1aky	M

1tfe	T	3chy	M
1yna	T	1csh	M
1gtm	T	1efu	M
1hdg	T	1xnb	M
2prd	T	1hrd	M
1ldn	T	1gad	M
1bdm	T	1ino	M
1xgs	T	1ldg	M
3pfk	T	4mdh	M
1php	T	1qmn	M
1ebd	T	1mat	M
1caa	T	2pkf	M
1thm	T	1qpg	M
3mds	T	1lpf	M
1btm	T	2rn2	M

Table 1- showing the pdb ids of thermophilic and mesophilic protein structures

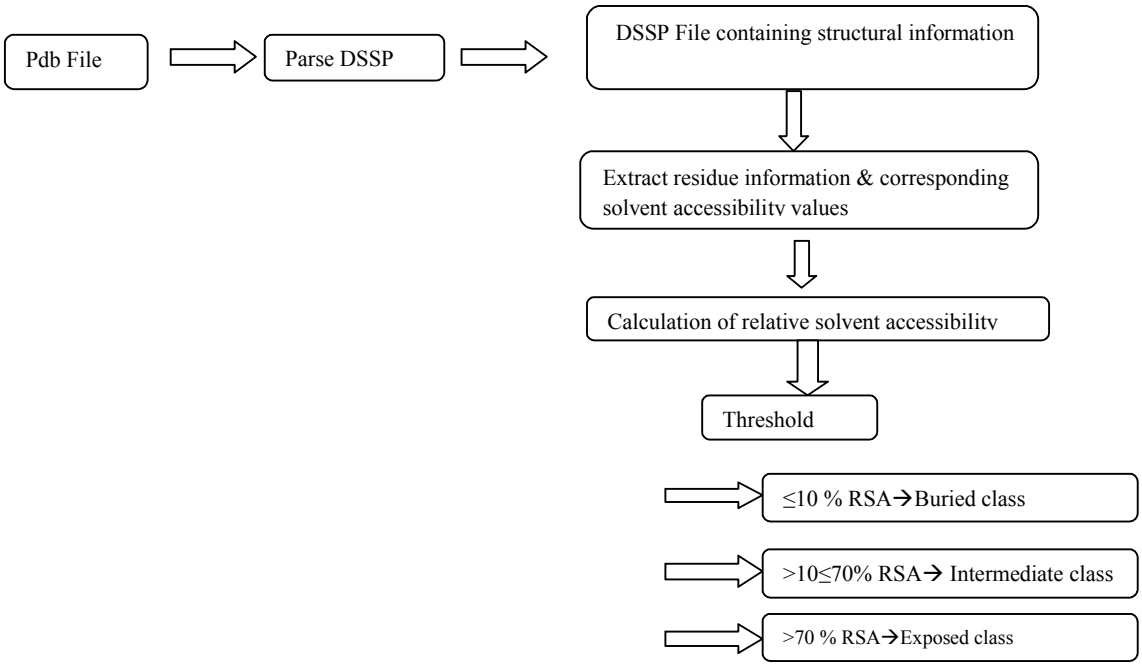

Fig 1. Flowchart of the experiment

After obtaining the protein structures we used DSSP program [8] to get the structural information for each structure including solvent accessibility values for each residue. A sample DSSP output is shown in fig 1. Solvent accessibility [9], is defined as the degree to which a residue in a protein is accessible to a solvent molecule. It is one of the important structural properties which can be used to gain insight into the tertiary structures of proteins. We extracted only the amino acid residue and the corresponding solvent accessibility values for each sequence from the DSSP output. For a comparative purpose we converted the solvent accessibility values into relative solvent accessibility by using the formula given below:

Relative solvent Accessibility of amino acid =accessibility from DSSP / Max accessibility [from G-X-G Tripeptide]*100 (1)

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30     *** HISTOGRAMS OF ***          .
 0  0  0  3  0  2  0  0  0  2  2  2  0  1  1  3  2  0  0  0  0  1  2  0  0  0  0  0  0  0    RESIDUES PER ALPHA HELIX        .
 0  0  4  2  0  4  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    PARALLEL BRIDGES PER LADDER     .
 5  4  2  2  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    ANTIPARALLEL BRIDGES PER LADDER .
 2  2  0  0  0  0  0  0  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    LADDERS PER SHEET               .
#  RESIDUE AA STRUCTURE BP1 BP2  ACC     N-H-->O    O-->H-N    N-H-->O    O-->H-N    TCO  KAPPA ALPHA  PHI   PSI    X-CA   Y-CA   Z-CA
  1    0 A M              0   0  175      0, 0.0    33,-0.1     0, 0.0    32,-0.0   0.000 360.0 360.0 360.0 153.7   16.3   13.7   54.5
  2    1 A K        -     0   0  121     31,-1.5     3,-0.1     1,-0.1    32,-0.0  -0.022 360.0 -91.9 -68.2 172.8   13.8   14.2   51.9
  3    2 A A        -     0   0   73      1,-0.1    -1,-0.1    31,-0.1     0, 0.0  -0.689  58.5 -87.2 -79.6 153.6   13.3   11.6   49.2
  4    3 A P        -     0   0   65      0, 0.0     2,-0.3     0, 0.0    32,-0.2  -0.203  35.8-144.9 -62.5 147.1   15.5   12.2   46.2
  5    4 A V  E     -a   36   0A  20     30,-1.6    32,-3.2    -3,-0.1     2,-0.6  -0.826  14.9-129.3-104.6 144.7   14.2   14.5   43.3
  6    5 A R  E     -a   37   0A  89     74,-0.3    76,-1.9    -2,-0.3    77,-1.6  -0.883  23.1-168.6 -96.1 126.5   15.2   13.6   39.7
  7    6 A V  E     -ab  38  83A   0     30,-3.5    32,-2.2    -2,-0.6     2,-0.4  -0.961   4.3-159.0-117.7 118.3   16.6   16.6   37.9
  8    7 A A  E     -ab  39  84A   0     75,-3.0    77,-2.7    -2,-0.6     2,-0.4  -0.801   9.3-174.6 -96.6 142.0   17.0   16.3   34.2
  9    8 A V  E     -ab  40  85A   0     30,-2.7    32,-2.4    -2,-0.4     3,-0.4  -0.990   6.9-159.7-135.8 125.4   19.5   18.6   32.4
 10    9 A T  E     S+a  41   0A   2     75,-2.3    77,-0.4    -2,-0.4    32,-0.2  -0.609  77.0  34.8 -95.1 166.0   19.9   18.8   28.7
 11   10 A G  S >   S+     0   0   12     30,-0.9     3,-2.0    -2,-0.2     6,-0.6   0.790  75.9 157.3  65.1  24.8   23.1   20.2   27.0
 12   11 A A  T 3   +      0   0    0     29,-1.2     9,-0.1    -3,-0.4    30,-0.1   0.624  61.7  54.3 -63.3 -21.7   25.0   18.7   29.9
 13   12 A A  T 3   S+     0   0   27     28,-0.2    -1,-0.3     4,-0.1     -2,-0.1   0.436  95.4  92.2 -96.9   3.6   28.3   18.4   28.2
 14   13 A G  S <>  S-     0   0   33     -3,-2.0     4,-2.2     1,-0.0     5,-0.1   -0.074  93.2 -86.8 -85.7-169.1   28.4   22.1   27.2
 15   14 A Q  H  > S+     0   0  126      2,-0.2     4,-1.1     1,-0.2     5,-0.1   0.920 125.7  41.0 -67.9 -50.6   29.8   25.3   28.8
 16   15 A I  H >> S+     0   0   40      1,-0.2     3,-1.0     2,-0.2     4,-0.9   0.948 114.1  56.1 -61.7 -43.6   26.8   26.2   31.0
 17   16 A G  H >> S+     0   0    0     -6,-0.6     4,-2.1     1,-0.3     3,-0.7   0.876 103.5  53.3 -56.2 -46.3   26.4   22.5   31.9
 18   17 A Y  H 3< S+     0   0   29     -4,-2.2     4,-0.3     1,-0.2    -1,-0.3   0.806 114.7  40.0 -63.7 -29.2   29.9   22.2   33.1
 19   18 A S  H << S+     0   0   19     -4,-1.1    -1,-0.2    -3,-1.0     4,-0.2   0.468 119.1  49.9 -91.6  -7.5   29.5   25.2   35.5
 20   19 A L  H <X S+     0   0    0     -4,-0.9     4,-1.9    -3,-0.7     3,-0.2   0.841  97.9  59.2 -95.9 -45.6   25.9   24.1   36.5
```

Fig 2. A sample output of DSSP giving structural details

On the basis of relative Solvent accessibility of residues, we divided the protein sequences into Buried, Intermediate and Exposed regions. All those amino acids which were having relative solvent accessibility value below 10% are grouped under Buried category, residues having relative solvent accessibility greater than equal to 10 and below 70% are grouped under intermediate category. The last Exposed category consisted of those residues which were having Relative solvent accessibility value more than 70%. The property groups of amino acids used in this study are shown in the table below.

Residue Group	**Residues in the Specific Group**
Tiny amino Residues	Ala, Cys, Gly, Ser, Thr
Small Residues	Ala, Cys, Asp, Gly, Asn, Pro, Ser, Thr and Val
Aliphatic Residues	Ile, Leu and Val.
Non-polar Residues	Ala, Cys, Phe, Gly, Ile, Leu, Met, Pro, Val, Trp and Tyr
Aromatic Residues	Phe, His, Trp and Tyr
Polar Residues	Asp, Glu, His, Lys, Asn, Gln. Arg, Ser, and Thr
Charged Residues	Asp, Glu, His, Arg, Lys
Basic Residues	His, Lys and Arg
Acidic Residues	Asp and Glu
Hydrophobic Residues	Ala, Cys, Phe, Ile, Leu, Met, Val, Trp, Tyr
Hydrophilic Residues	Asp, Glu, Lys, Asn, Gln

Table 2 Showing the distribution of amino acids into different property groups.

T-test: For statistical analysis we used independent sample t test. T-test is usually a good tool to assess the hypothesis that the difference in the means of a variable between two populations is significant or not. Various researchers have successfully used t-test for comparative analysis.

$$t = \frac{F_{M,S} - F_{T,S}}{\sqrt{\left(Var_M / n_M\right) + \left(Var_T / n_T\right)}} \quad (2)$$

$$t = \frac{F_M - F_T}{\sqrt{\left(Var_M / n_M\right) + \left(Var_T / n_T\right)}} \quad (3)$$

Where

$F_{M,S}$ is mean of percentage buried, intermediate and exposed regions in mesophilic proteins

$F_{T,S}$ is mean of percentage buried,intermediate and exposed regions in thermophilic proteins

F_M is the mean frequency of amino acid or amino acid property group of mesophilic proteins

F_T is the mean frequency of amino acid or amino acid property group of mesophilic proteins

Var_M and Var_T are the variance of residues or property groups of mesophilic and thermophilic proteins respectively.

nM and nT are the total number of mesophilic and thermophilic proteins

RESULT AND DISCUSSION

On the basis of relative accessibility protein structure was divided into buried, intermediate, and exposed regions. We performed the independent sample T –test taking % Buried, % Intermediate, and % Exposed regions between thermophilic and mesophilic proteins and we found no significant differences for these regions. It indicates that the proteins belonging to both thermophilic and mesophilic organisms are having comparable portions of their protein chains in buried, intermediate and exposed regions. This study indicates that adaptation of proteins at the molecular level does not involve the variation of number of residues into the core or surface of the proteins.

Without the differentiation of the protein sequences into the three regions we found that among property groups non-polar, polar ,hydrophobic were the significant difference between the thermophilic and mesophilic proteins and also that there is a significant compositional difference of amino acids –Argnine, Isolucine, Serine, Threonine . This result is in coherence with the previously published results showing existing compositional bias between the two categories [10, 11]. These significant differences between Thermophilic & Mesophilic proteins which we found in the current analysis can be used for better understanding of the adaptive changes at molecular level; also these differences can be used for creating feature vectors for discriminating Mesophilic & Thermophilic proteins.

Pdb_Id	Actual Type	% Buried	%Intermedeate	%Exposed	Pdb_Id	Actual Type	% Buried	%Intermedeate	%Exposed
1tmy	T	29.6610	60.1695	10.2712	1aky	M	30.7339	60.0917	9.2661
1tfe	T	23.2394	62.6761	14.2281	3chy	M	28.1250	60.1563	11.8359
1yna	T	38.3420	54.9223	6.8031	1csh	M	38.1609	51.4943	10.4483
1gtm	T	47.3264	47.3264	5.2394	1efu	M	35.6757	8.2239	5.7714
1hdg	T	37.8947	53.6842	8.353	1xnb	M	35.6757	56.2162	8.1892
2prd	T	27.5862	62.6437	9.8678	1hrd	M	45.0704	49.6664	5.1660
1ldn	T	51.6765	44.3393	3.7452	1gad	M	44.1755	47.0499	8.7095
1bdm	T	43.1889	49.8452	6.7229	1ino	M	29.7143	58.8571	11.5429
1xgs	T	44.1624	49.2386	6.4941	1ldg	M	36.8254	56.1905	7.0540
3pfk	T	39.1850	53.9185	6.9655	4mdh	M	40.3298	51.1244	8.4798
1php	T	39.0863	54.3147	6.6650	1qmn	M	37.6022	52.8610	9.0817
1ebd	T	39.3494	53.9349	6.5708	1mat	M	39.5437	53.9924	6.5285
1caa	T	22.6415	64.1509	13.3396	2pkf	M	40	64.6281	5.1752
1thm	T	45.1613	48.0287	6.8781	1qpg	M	38.5542	51.0843	10.4651
3mds	T	39.3132	53.0713	7.4447	1lpf	M	39.1534	54.6032	3.7452
1btm	T	46.9185	48.1509	4.6183	2rn2	M	29.0323	58.0645	13.0323

Table 3. Percentage distribution of protein regions into buried, intermediate and exposed regions in thermophillic and mesophilic sequences

CONCLUSION

Thermophilic proteins have the ability to withstand extremes of temperature and still remain biologically active. These proteins shows adaptation at the molecular level .In the present study we did a comparative study on the Zhang et al. (7) dataset by analyzing the differentially exposed regions of protein chains of thermophilic and mesophilic protein structures. The % of residues belonging to buried, intermediate and exposed regions did not show any significant differences between the two categories. Further analysis can shed more light into the packing of residues in protein chains. This study also confirmed the compositional differences in terms of amino acid composition and amino acid property group composition existing between mesophilic and thermophilic proteins as deduced in previous researches. More in depth statistical analysis can shed light on the molecular basis of these environmental adaptations.

REFERENCES

1. Gomes J and Steiner W, "The biocatalytic potential of extremophiles and extremozymes". Food Technol. Biotechnol. 2004; 42(4):223-225.
2. Jaenicke,R, "Protein stability and molecular adaptation to extreme conditons". European Journal of Biochemistry 1991; 202(3):715:728.
3. Jaenicke,R and Bohm,g.,"Thermostability of proteins from Thermotoga maritime". Methods Enzymol 2001;334,438-469.
4. McCarty T , "Evolution of three thermostable fungal endo-beta-glucocanases from *Talaromyces emersonii* for brewing,food application".ProcessBiochem2005;40(5)1741-8.
5. Sharma R,"Purification and characterization of a thermostable alkaline lipase from a new thermophillic Bacillus" sp. RSJ-1. Process Biochem 2002; 37(10): 1075-84.
6. Fagalin CO., "Understanding protein stability". Biochem Biophys Acta 1995;1252:1-14.
7. Zhang, G. and Fang, B., "Discrimination of thermophilic and mesophilic proteins via pattern recognition Method, Process Biochemistry". 2006;41:552–556.
8. Kabsch, W. Sander, C., "Dictionary of protein secondary structure:pattern recognition of hydrogen –bonded and geometrical features". Biopolymers 1983;22:2577-637.
9. Rost, B. and Sander, C., "Conservation and prediction of solvent accessibility in protein families, Proteins: Structure, Function, and Genetics".1994;20:216-226.
10. Zhou, X.X., "Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins.Amino Acids". 2008; 34(1):25-33.
11. J. Paul "Thermal adaptation analyzed by comparison of protein sequences from extremely thermophilic Methanococcus species"Proc Natl Acad Sci. 1999;96(7):3578-83.