BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(23)

Distribution of amino acids in functional sites of proteins with high melting temperature

Amutha Selvaraj Maheshwari[1, 2] & Govindaraju Archunan[1]*

[1]Department of Animal Science, School of Life Sciences, Bharathidasan University, Tiruchirappalli – 620 024, Tamil Nadu, India; [2]Department of Biotechnology, Anna University – BIT campus, Tiruchirappalli – 620 024, Tamil Nadu, India; Govindaraju Archunan – Email: archunan@bdu.ac.in; Phone: +91-431-2407040; Fax: +91-431-2407045 / 2412750; *Corresponding author

Received September 27, 2012; Accepted October 26, 2012; Published November 23, 2012

Abstract:
The stability of proteins in its native state has an important implication on its function and evolution. The functional site analysis may lead to better understanding of how these amino acid distributions influence the melting temperature of proteins. It has been reported that increasing the fraction of hydrophobic contacts in a protein tends to raise melting temperature; increasing the fraction of repulsive charge contacts decrease the melting temperature and consistent with a destabilizing effect. The role of amino acid distribution as hydrophobic, charged and polar residues in proteins and mainly in its functional sites has been studied. Due to limited data availability, redundancy check and controlled environment parameters, the study was carried out with ten single chain-wild proteins having melting temperature above 80°C at pH 7. The analysis depicts that, the entire protein, hydrophobic residues are more frequent in single chain proteins and charged residues are more frequent in multi-chains proteins. In functional sites of these proteins, hydrophobic and charged residues are equally frequent in single chain proteins and charged residues are very high in multi-chains proteins. But, the polar residue distribution remains same for both single chain and multi-chain proteins and its functional sites.

Key words: Classified amino acids, Functional residues, Melting temperature, Single and Multi-chain, Thermo-stable proteins.

Background:
The melting temperature is an important characteristic feature of a sequence in the prediction of thermal stability of proteins. Increasing the fraction of hydrophobic contacts in a protein tends to raise melting temperature; increasing the fraction of repulsive charge contacts decrease the melting temperature and consistent with a destabilizing effect. It has been revealed that thermophilic and mesophilic proteins have both similar polar and non-polar contribution to the surface area and compactness. Salt bridges and main chain hydrogen bonds show an increase in the majority of thermophilic proteins than their mesophilic homologues. In thermophilic proteins, hydrophobic residues are significantly more frequent and polar residues are less frequent **[1].** It has been described that thermophiles prefer to have contacts between residues with hydrogen-bond-forming capability. The contact density is not significantly correlated with protein melting temperature. More contacts are observed between polar and non-polar residues in thermophiles than mesophiles. Tyr has good contacts with several other residues, and Cys has considerably higher long-range contacts in thermophiles compared with mesophiles **[2].** It has been reported that both thermophilic and mesophilic proteins have similar hydrophobicities, compactness, oligomeric states, polar non-polar contribution to surface areas, main-chain and side-chain hydrogen bonds. Salt bridges and side-chain hydrogen bonds increase in the majority of the thermophilic proteins. Arg and Tyr are more frequent, Cys, Ser are less frequent in thermophilic proteins. Thermophiles have a larger fraction of their residues in the α-helices and avoid Pro in its α-helices to a greater extent than the mesophiles. It may be the cause to withstand high temperature **[3].**

The stabilizing group consists of polar-charged residues and non-polar residues that possess high surrounding hydrophobicity. It has been described as the polar uncharged residues destabilize the molecule against temperature and Ser being the most destabilizing residue. A very high co-operativity exists among the stabilizing non-polar residues. In small

globular proteins, the melting temperature remains mainly a function of amino acid composition and in complex molecules it depends on other factors also **[4].** The maximum melting points of proteins with respect to pH are reported that the correlation coefficient of hydrophobic index versus melting point was 0.622; average residue volumes versus melting points was 0.960; the average residue volume versus hydrophobic index was 0.697 **[5].** In proteins, the strongest contributors to thermo-stability are increased in ion pairs on the surface and the very strong hydrophobic interior **[6].** Identifying the protein regions that are most likely to be involved in function is useful to gain information about its potential role. The combination of experimental and in-silico approaches will help us to interpret the genetic information in functional terms and be the final goal of the so-called post-genomic era **[7].** It has been analyzed that multiple peptide motifs are capable of different temperature dependent transitions **[8].** In the present work, we have studied the distribution of amino acids in both thermophilic proteins and in its functional sites. In this amino acids distribution studies, the classification of amino acids as hydrophobic, charged and polar residues is followed. Then, the distribution of such classified amino acids and the melting temperature transition among proteins is analyzed.

Methodology:

Datasets

The database ProTherm **[9]** available at: http://gibk26.bio.kyutech.ac.jp/jouhou/Protherm/protherm.html is used for our study. The melting temperature of wild proteins measured by different techniques such as DSC, CD, Adsorption, Fluorescence, and NMR, for both single chain and multi-chains proteins is mainly focused in this study. Certain conditions such as measurement, state, chain type, pH etc. is set to derive a proper dataset for the analysis. To avoid bias in prediction of result, the size of the derived dataset was reduced by redundancy check.

Amino acid composition (%)

The percentage of amino acids present in the protein sequences were collected using Protparam tool **[10]** available at http://web.expasy.org/protparam/. The percentages of collected amino acids were classified as hydrophobic, Hp (A, I, L, M, F, P, V), charged (R, D, E, K) and polar (N, C, Q, G, H, S, T, W, Y) residues.

Non-polar residues (hydrophobic)

Hydrophobic amino acids are found at the core of many proteins and they mainly composed of non-polar residues. A significant amount of non-polar residues are also found on the surface of proteins. These are not favorably interact with water but stabilized by numerous van der Waals interactions.

Charged residues

Charged amino acids are found on the surface of the protein as well as seldom buried in the interior of a folded protein. They can interact with water and other important biological molecules. The positively and negatively charged amino acids form salt bridges.

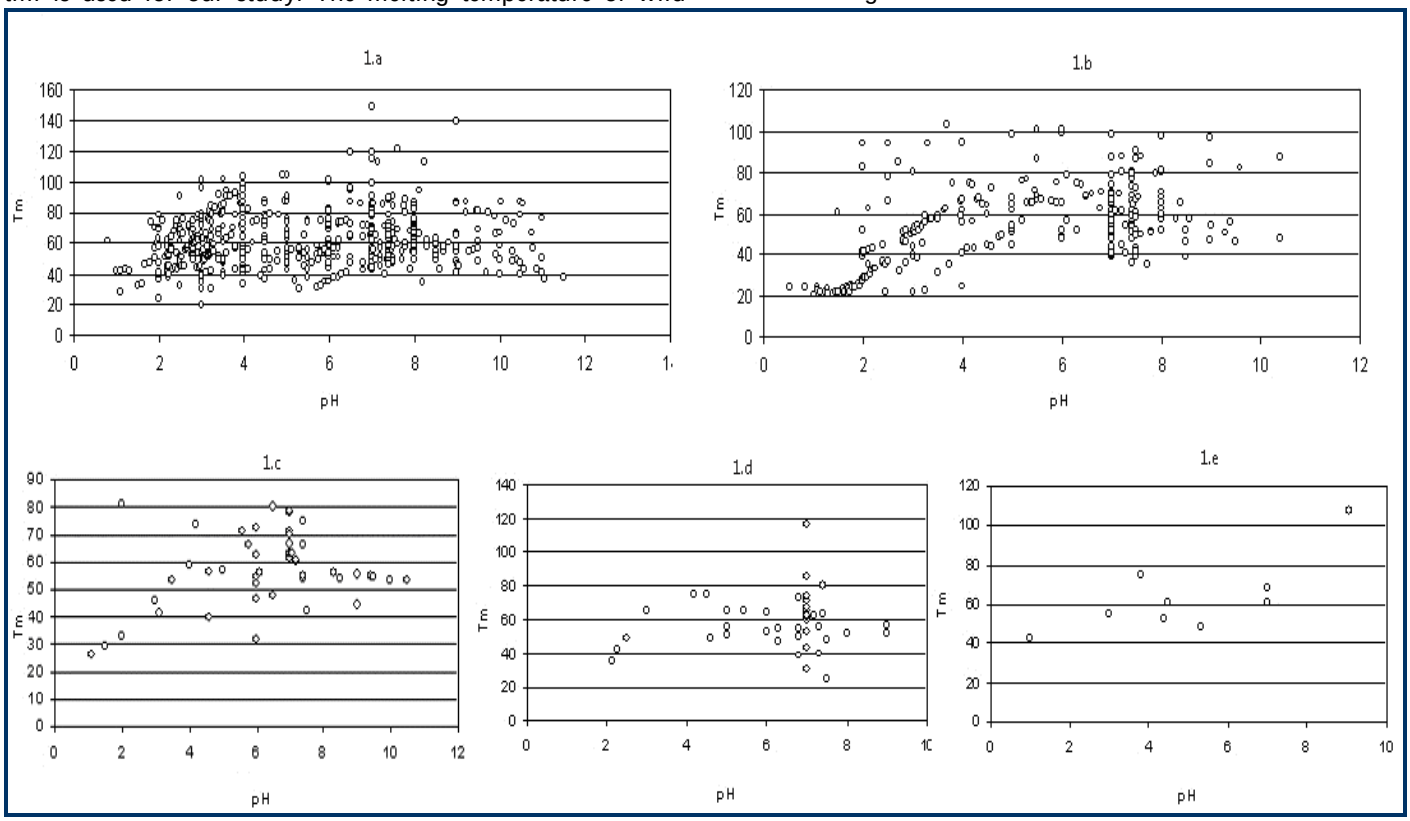

Figure 1.a: The plot depicts the melting temperature of various proteins obtained by DSC technique at different pH; **1.b):** The plot depicts the melting temperature of various proteins obtained by CD technique at different pH; **1.c);** The plot depicts melting temperature of various proteins obtained by Absorption technique at different pH; **1.d):** The plot depicts melting temperature of various proteins obtained by Fluorescence technique at different pH; **1.e):** The plot depicts melting temperature of various proteins obtained by NMR technique at different pH.

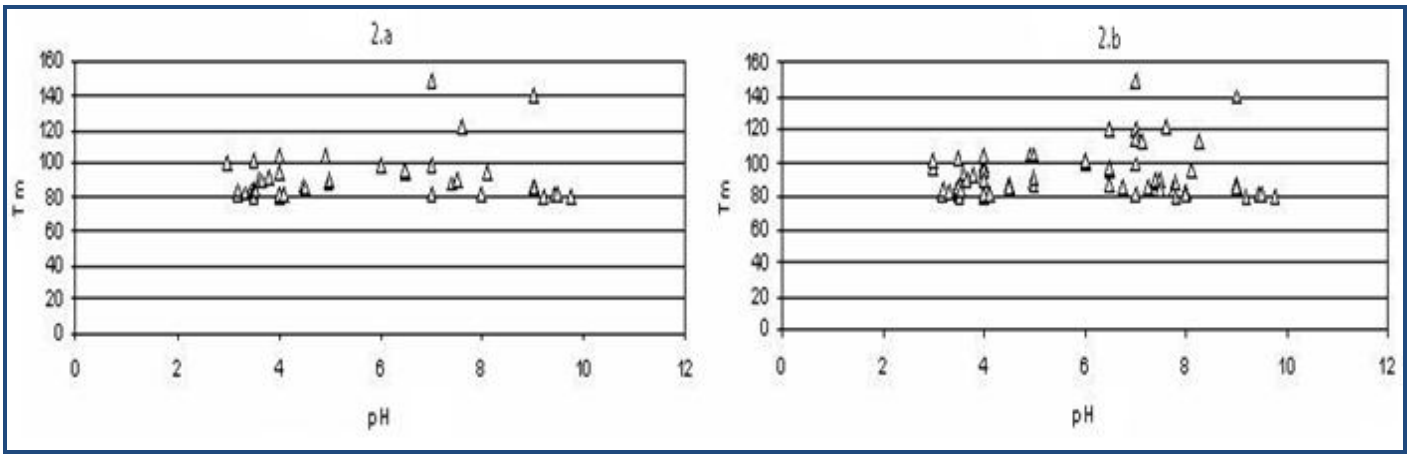

Figure 2.a: The plot depicts the single chain proteins having melting temperature above 80°C at different pH **[16]. 2. b)** The plot depicts the multi-chains proteins having melting temperature above 80°C at different pH **[16].**

Polar residues (hydrophilic)

Polar amino acids are found both at buried position as well as on the surface of proteins and posses hydrophilic groups which either form hydrogen bonds with other polar residues in the protein or with water. Polar amino acid side chains may form side chain-side chain or side chains-main chain hydrogen bonds. These interactions are often crucial for the stabilization of protein tertiary structure and are normally conserved.

Functional/ Active sites

The CASTp server **[11, 18]** available at http://sts-fw.bioengr.uic.edu/castp/ is used. Using this server, the amino acids present in the active sites of single chains proteins were identified. The POOL program (Partial Order Optimal Likelihood) available at: http://www.pool.neu.edu:8080/wPOOL/, which is a machine learning method that identifies the functionally important residues in a given protein structure **[12]** is also used in this research.

Multiple sequence alignment

Analysis of multiple alignments was performed using the software MAFFT. It is a multiple alignment program for amino acid or nucleotide sequences **[13, 14]** available at http://mafft.cbrc.jp/alignment/software/. This program is used to understand the phylogenetic relationships among wild proteins particularly with single chain and Tm > 80˚C at pH7. Jalview version-2 was used for editing and for obtaining phylogenetic tree **[14].**

Discussion:

As per previous report, in complex molecules the melting temperature depends on other factors also **[4].** So, the dataset with defined conditions such as pH, state, number of chains etc is generated. Due to these limitations, the study is carried out with limited data. The overview of available Tm measured by different techniques at different pH is given in **(Figure 1a, 1b, 1c, 1d & 1e).** These scatter plots depicting the size of Tm dataset measured by DSC and CD measurements is larger than the datasets derived through adsorption, florescence, NMR measurements. Since the reliability of result prediction attained by large dataset, the research is focused on Tm obtained through DSC. The distribution of single chain and multi-chains proteins with Tm > 80˚C at different pH are shown in **(Figure 2a, 2b).** The study is to find out the type of amino acids significant for maintaining the temperature in thermophilic proteins. While considering the whole protein with Tm > 100˚C at pH 7, hydrophobic residues are more frequent followed by polar residues in single chain proteins (4 data) and hydrophobic/ charged residues are more frequent and polar residues are very less frequent in multi-chains proteins (4 data) which illustrated in (**Figure 3a & 3b)**. **As** per the previous report the hydrophobic residues are significantly more frequent and polar residues which are less frequent in thermophilic proteins **[1].** The dataset of 16 proteins with many single chains as well as few multi-chains having Tm > 40˚C at pH 7 make it clear that almost 50% are shared by polar residues and the remaining 50% are shared by mostly hydrophobic residues which followed by charged residues **(Figure 3.2**). This dataset is taken just for the comparison of different melting temperature transition.

While considering the whole protein with Tm > 80˚C at pH 7, not much variations is observed in polar residue distribution for both single chain and multi-chain proteins. In single chain proteins (11 data), hydrophobic residues are more frequent and in multi-chains proteins (7 data), charged residues are more frequent (**Figure 3.3 & 3.4) Figure 3.5a & 3.5b)** illustrates the top ten residues

Which constitute the functional site of proteins having Tm > 80˚C at pH 7. While considering the functional sites of these proteins, not much variation is observed in polar residue distribution for both single chain and multi-chain proteins. It seems polar residues remain same in protein as well as in functional sites for both single chain proteins. In single chain proteins (10 data), hydrophobic and charged residues are equally frequent and in multi-chains proteins (7 data), charged residues are high frequent **(Figure 3.5a & 3.5b).** As reported earlier more contacts have been observed between polar and non-polar (Hp) residues in thermophiles than in mesophiles **[2].** The individual amino acid distribution of single and multi-chain proteins having Tm>100˚C at different pH is shown in **(Figure 4).** In which, the composition of Hp (A, I, L) and charged (E, K) are high; charged (R, D), polar (G) and Hp (V) are of medium composition; polar (N, Q, H, P, S, T, W, Y) and Hp (M, F) are less in composition. Previous studies demonstrated that Arg and Tyr are more frequent and Cys, Ser are less frequent in thermophilic proteins **[3].**

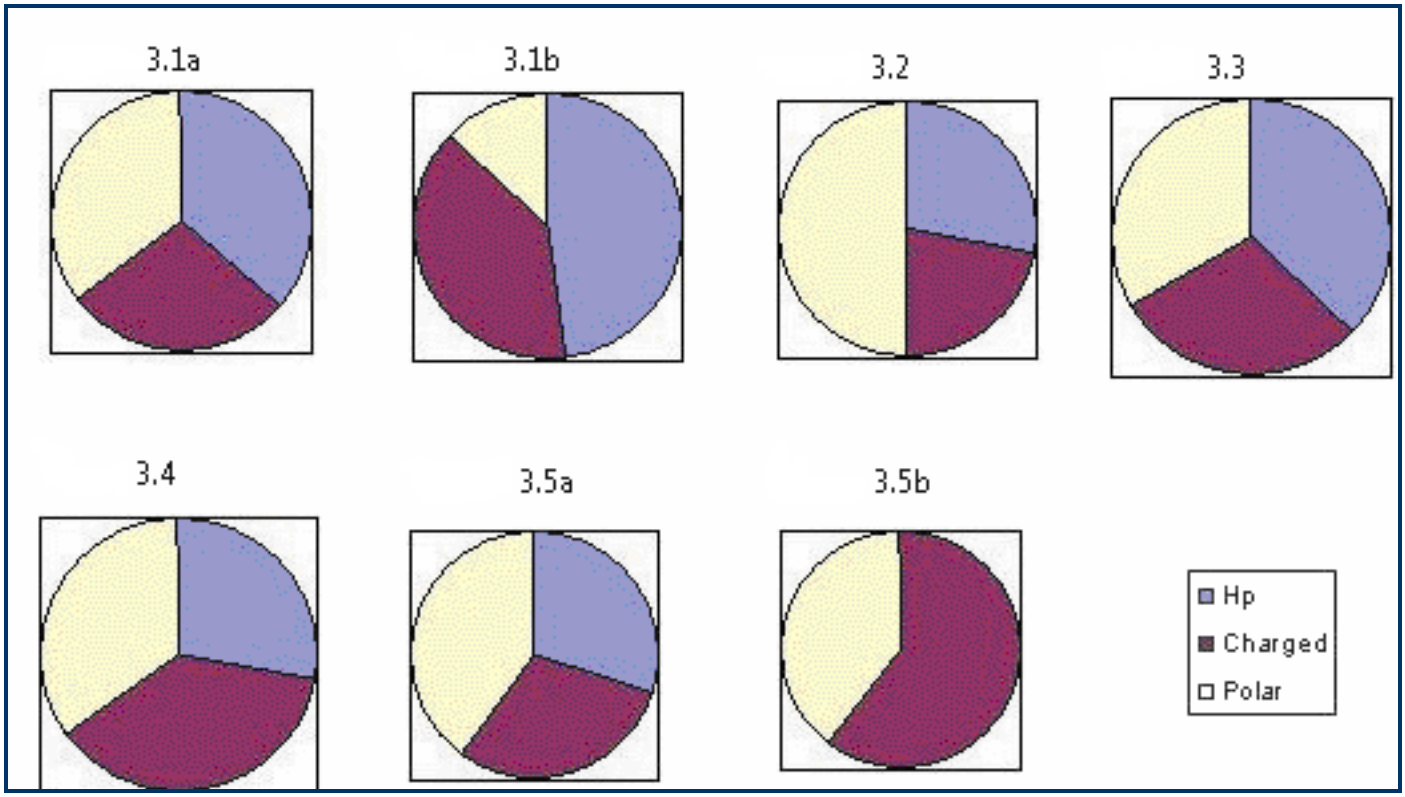

Figure 3.1a: Illustration of amini acids distribution among single-chains proteins having Tm> 100°c at different pH; **3.1b)** Illustration of amini acids distribution among multi-chains proteins having Tm> 100°C at different pH; **3.2)** Illustration of Proteins with many single chains as well as few multi-chains having Tm> 40°C at pH7; **3.3)** Illustration of single chain two state proteins having Tm> 80°C at pH 7; **3.4)** Illustration of multi-chain two state proteins having Tm> 80°C at pH 7; **3.5a)** Illustration of functional site residues of single chain proteins having Tm> 80°C at pH 7; **3.5b)** Illustration of functional site residues of multi-chains proteins having Tm> 80°C at pH 7

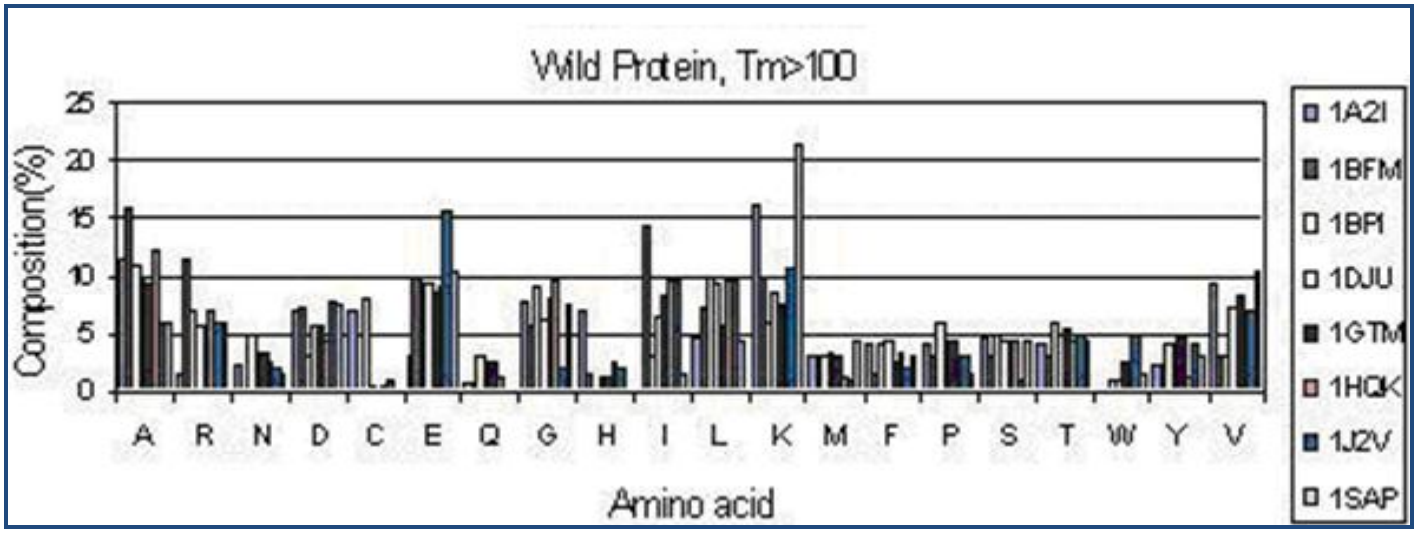

Figure 4: plot depicting the amino acid distribution (%) of single and multi-chain proteins having Tm>100°C and at different pH **[19].** (Note: Different proteins used for our study are highlighted in VIBGYOR coloring scheme by its PDB ID)

Identifying the protein regions that are most likely to be involved in function may give the direction to melting temperature transition. So, the study is focused on the amino acid distribution in its functional sites. As reported by Pool server, either because of atoms with undefined parameters or because of high-potential regions in the protein, the functional site residues for our entire dataset could not be collected. And also, it is given as the top 8-10% of the residues in the rank order gives the best performance in this server. Hence, the top 10 residues were considered for work **Table 1 & 2 (see supplementary material).** The Accession number and PDB ID of proteins used in the multiple sequence alignment for the dataset generated using the conditions such as: single chain, two state, pH7, Tm>80°C **Table 3 (see supplementary**

material). The phylogenetic tree shown in **(Figure 5)** depicted that 1BSQ, 1QG5 are similar proteins; 1ONC and 1J2V/1NZA might have another internal ancestor. In this, 1J2V and 1NZA belong to same family. 1NZA, 1ONC and 1J2V are of almost same length and the percentage of top 10 functional site residues are also same for single chain proteins with Tm>80˚C at pH7.

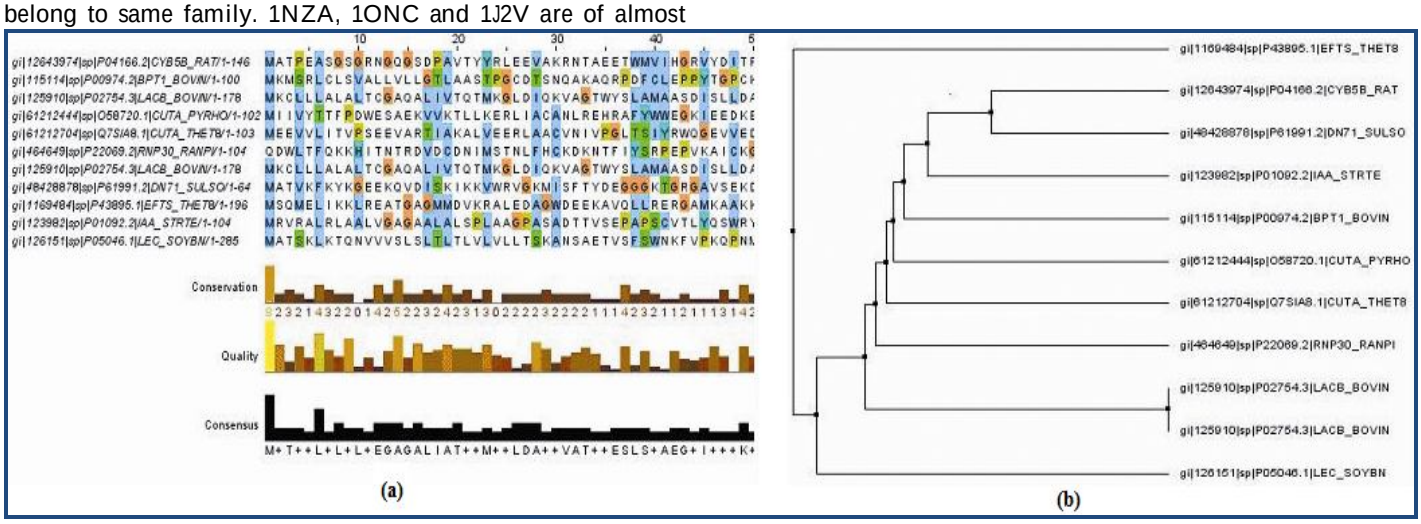

Figure 5.a: Pictorial representation of the output of MAFFT program which depicts the multiple sequence alignment of user-submitted sequences (note:cinsensus positions of amino acids are highlighted in VIBGYOR coloring scheme); 5.**b):** Snapshot for the phylogentic tree of user-submitted multiple sequence alignment for wild proteins particularly with single chain having Tm > 80°was performed using the software MAFFT **[20].**

Conclusion:

Previous report described whether it is a single chain or multi-chain protein, the polar amino acids composition is often crucial for the thermal stabilization of proteins. The present study concludes that while considering the entire protein, hydrophobic residues are more frequent in single chain proteins and charged residues are more frequent in multi-chains proteins. While considering the functional sites of these proteins, hydrophobic and charged residues are equally frequent in single chain proteins and charged residues are very high in multi-chains proteins. But, the polar residue distribution remains same for both single chain and multi-chain proteins and its functional sites. From the phylogenetic tree of single chain, two state proteins having Tm>80˚C at pH7, it is clear that the thermophilic proteins having similar amino acid length might have same internal ancestor. This may lead to the conservation of functional site residues and its chemical nature is same which may also responsible for high melting temperature. Instead of considering the entire protein, the functional site analysis leads to better understanding of how these distributions of amino acids may affect the melting temperature of proteins.

Acknowledgement:

We thank Bharathidasan University and Anna University – BIT campus, Tiruchirappalli for providing the facility to carry out our research work. We also thank Ms. A. Sheeba Princess, Assistant Professor, Department of English, Bishop Heber College, Tiruchirappalli for helping the language corrections. This work was partially supported by grants from University Grants commission and UGC – SAP, Government of India.

References:

[1] Sadeghi M *et al. Biophys Chem*. 2006 **119**: 256 [PMID: 16253416]
[2] Gromiha MM *et al. Biophys Chem.* 2001 **91:** 71 [PMID: 11403885]
[3] Kumar S *et al. Protein Eng.* 2000 **13**: 179 [PMID: 10775659]
[4] Ponnuswamy PK *et al. Int J Biol Macromol*. 1982: 4.
[5] Bull HB *et al. Arch Biochem Biophys.* 1973 **158**: 681 [PMID: 4782528]
[6] Taylor TJ *et al. BMC Struct Biol.* 2010 **10**: Suppl 1 [PMID: 20487512]
[7] Pazos F & Bang JW, *Current Bioinformatics.* 2006 **1:** 15
[8] Mackay JA & Chilkoti A. *Int J Hyperthermia*. 2008 **24**: 483 [PMID: 18608590]
[9] Kumar MD *et al. Nucleic Acids Res*. 2006 **34**: 204
[10] Gasteiger E *et al. The Proteomics Protocols Handbook*. 2005: 571
[11] Dundas J *et al. Nucleic Acid Research*. 2006 **34**: 116
[12] Somarowthu S & Ondrechen MJ, *Bioinformatics*. 2012 **28**: 2078 [PMID: 22661648]
[13] Katoh K & Toh H, *Bioinformatics*. 2010 **26**: 1899 [PMID: 20427515]
[14] Waterhouse AM *et al. Bioinformatics.* 2009 **25**: 1189 [PMID: 19151095]
[15] Berman HD *et al. Structure*. 2012 **20**: 391 [PMID: 22404998]
[16] http://gibk26.bio.kyutech.ac.jp/jouhou/Protherm/protherm.html
[17] http://web.expasy.org/protparam/
[18] http://sts-fw.bioengr.uic.edu/castp/
[19] http://www.pool.neu.edu:8080/wPOOL/
[20] http://mafft.cbrc.jp/alignment/software/

Edited by P Kangueane

Citation: Maheshwari & Archunan, Bioinformation 8(23): 1176-1181 (2012)

BIOINFORMATION

open access

Supplementary material:

Table 1: Functional site residues of different single chain proteins having Tm>80˚C **[12, 19]**

Rank	1B5M	1BPI	1BSQ	1J2V	1NZA	1ONC	1QG5	1SSO	1TFE	2AIT
1	HIS:A 63	TYR:A23	TYR:A 42	GLU:A50	GLU:A52	HIS:A 10	TYR:A102	TYR:A 33	TYR:A 50	HIS:A 64
2	ASP:A 53	TYR:A10	HIS:A 161	ASP:A 48	GLU:A 1	LYS:A 9	LYS:A 135	ARG:A42	TYR:A 25	ASP:A 40
3	HIS:A 39	ARG:A 0	PHE:A 05	TYR:A75	CYS:A 30	HIS:A 97	PHE:A105	LYS:A 12	GLU:A 21	LEU:A74
4	TYR:A 30	ARG:A 9	VAL:A 41	TYR:A39	ASP:A 50	ILE:A 11	ASP:A 85	TYR:A 7	GLU:A 53	VAL:A 4
5	ALA:A54	TYR:A 5	GLN:A120	GLU:A67	LYS:A 58	LYS:A 31	LEU:A103	LYS:A 20	LYS:A 149	THR:A 3
6	GLU:A56	LYS:A 46	ARG:A 40	HIS:A 35	HIS:A 75	PHE:A 6	TYR:A 42	GLU:A 5	ARG:A106	ASP:A 39
7	ILE:A 25	CYS:A30	MET:A 07	GLU:A 3	PRO:A 0	ASP:A67	ASP:A 96	THR:A32	ARG:A187	SER:A 5
8	MET:A23	ASN:A 4	VAL:A 43	HIS:A 73	TYR:A 77	PHE:A6	GLU:A 89	LYS:A 21	ARG:A146	ASP:A 1
9	ARG:A28	LYS:A 41	CYS:A 121	THR:A 7	LEU:A37	PHE:A98	ASN:A 90	ILE:A 19	ASP:A 95	CYS:A 73
10	HIS:A 26	ILE:A 19	GLU:A 44	ASP:A 60	LEU:A53	GLN:A 7	GLU:A131	VAL:A22	HIS:A 99	SER:A 63
11	SER:A 57	PRO:A 9	GLU:A108	ALA:A37	GLN:A51	ARG:A5	MET:A107	GLU:A11	VAL:A186	GLU:A6
12	ASP:A 31	GLU:A 7	LEU:A122	GLU:A46	THR:A 8	LYS:A 8	CYS:A 121	ASP:A 49	HIS:A 65	THR:A41
13	LEU:A46	THR:A11	ASP:A 96	PRO:A 9	GLY:A36	ASN:A34	ILE:A 84	ASP:A 15	LYS:A 170	GLY:A65
14	THR:A55	TYR:A21	CYS:A 106	LYS:A 66	GLU:A49	HIS:A 29	LEU:A122	LYS:A 4	GLU:A117	GLU:A38
15	GLU:A44	VAL:A34	GLU:A 89	PHE:A 8	VAL:A 9	LYS:A 33	GLU:A134	ALA:A 1	VAL:A185	GLY:A62
16	ALA:A52	LEU:A29	ILE:A 162	ASP:A 10	THR:A38	THR:A12	ASP:A 129	GLU:A10	LEU:A157	LYS:A 34
17	VAL:A24	PHE:A33	LEU:A103	VAL:A51	TYR:A 94	LYS:A 80	ASP:A 98	ASP:A 34	ARG:A191	THR:A 2
18	PHE:A58	LYS:A 15	TYR:A102	GLU:A64	SER:A 39	ASN:A13	ALA:A 86	LYS:A 6	TYR:A 107	ARG:A72
19	GLU:A43	GLY:A28	ASP:A129	GLU:A59	VAL:A57	THR:A4	ASN:A 88	LYS:A 18	ARG:A 56	PRO:A 7
20	LYS:A 85	ASP:A 3	LEU:A 39	TYR:A 5	TYR:A 41	LEU:A27	VAL:A 92	PHE:A31	LYS:A 168	PRO:A 9
21	PRO:A40	ARG:A 1	LEU:A 1	TRP:A 62	LEU:A56	THR:A 5	GLU:A108	GLY:A 3	GLU:A158	GLU:A42
22	TYR:A 75	PHE:A22	LYS:A 135	CYS:A 29	LEU:A54	VAL:A96	GLN:A120	GLN:A13	TYR:A 63	TYR:A 37
23	GLY:A41	LYS:A 26	SER:A 21	TYR:A92	GLU:A 3	PHE:A28	LEU:A104	THR:A40	ARG:A188	ALA:A 8
24	GLU:A59	THR:A32	ASP:A 98	LYS:A 56	HIS:A 62	TYR:A64	GLU:A 74	VAL:A14	GLU:A152	TYR:A 69
25	VAL:A29	GLY:A12	LYS:A 101	GLU:A71	VAL:A34	ILE:A 22	LYS:A 101	SER:A 30	ILE:A 100	ASP:A 58

Table 2: Functional site residues of different proteins with multi-chains and Tm> 80˚C **[14].**

Rank	1AZP	1C9O	1CVN	1HQK	1TUP	1WPW	1Y4Y
1	LYS:A 13	LYS:A 39	LYS:D 114	GLU:A 95	CYS:C 141	ARG:B 116	GLU:B 66
2	TYR:A 8	GLU:A 36	LYS:A 114	LYS:A 98	CYS:B 176	ASP:B 224	GLU:C 36
3	TYR:A 34	GLU:B 36	HIS:D 127	LYS:E 98	TYR:B 236	TYR:B 123	GLU:C 66
4	ASP:A 36	LYS:B 39	HIS:B 127	LYS:D 98	HIS:B 179	TYR:A 123	LYS:C 62
5	ARG:A 42	ALA:B 32	LYS:B 114	LYS:C 98	CYS:B 238	LYS:A 169	GLU:A 72
6	GLU:A 11	GLU:A 50	HIS:A 127	LYS:B 98	CYS:B 242	ASP:A 224	ASP:A 69
7	LYS:A 39	THR:B 31	HIS:C 127	GLU:D 95	ARG:B 175	ARG:A 116	GLU:B 32
8	GLY:A 41	HIS:B 29	GLU:B 192	GLU:C 95	TYR:A 236	LYS:B 169	GLY:B 67
9	LYS:A 9	ALA:B 61	GLU:D 192	GLU:B 95	GLU:C 198	ASP:A 200	ALA:A 71
10	THR:A 33	ALA:A 61	LYS:C 114	GLU:E 95	TYR:C 236	GLU:B 219	ALA:B 65
11	ASP:A 16	ALA:A 32	GLU:C 192	ASP:C 90	HIS:B 214	GLU:A 219	GLU:C 3
12	GLU:A 12	THR:A 31	LYS:D 116	ASP:B 90	HIS:C 233	ASP:B 121	TYR:C 37
13	GLY:A 43	GLN:A 34	LYS:A 116	ASP:A 90	CYS:B 182	ASP:A 121	THR:C 41
14	LYS:A 7	GLU:B 50	HIS:D 51	ASP:D 90	CYS:A 176	ASP:B 228	ALA:A 2
15	GLU:A 14	HIS:A 29	HIS:C 51	TYR:E 91	CYS:B 135	ASP:B 200	THR:C 64
16	THR:A 40	ALA:B 60	HIS:B 51	TYR:D 91	CYS:A 238	ASP:A 228	ASP:B 69
17	GLU:A 53	ASN:A 62	LYS:B 116	TYR:C 91	TYR:A 163	ASP:B 177	ALA:C 65
18	PHE:A 32	ILE:A 33	GLU:A 192	TYR:B 91	HIS:B 178	ASP:A 177	ALA:A 70
19	ILE:A 20	PHE:A 38	HIS:D 121	ASP:E 90	GLU:A 171	TYR:B 222	ALA:B 2
20	ASP:A 35	GLN:B 59	HIS:B 121	TYR:A 91	CYS:C 135	GLU:A 229	ALA:A 73
21	ASP:A 56	GLN:B 34	HIS:A 121	GLU:C 58	CYS:C 124	LYS:B 139	LYS:A 28
22	GLU:A 64	GLY:A 35	HIS:C 121	HIS:B 88	TYR:C 234	GLU:B 229	ALA:A 75
23	PHE:A 6	GLY:A 37	THR:D 112	GLU:B 58	HIS:A 179	TYR:A 222	GLU:C 32
24	LYS:A 5	ASN:B 62	THR:B 112	GLU:D 58	ARG:B 213	LYS:A 139	GLN:C 34
25	LYS:A 21	ALA:A 60	GLU:D 8	GLU:E 58	GLY:B 244	GLU:B 117	ALA:A 68

Table 3: Accession number and PDB Id of proteins used in the multiple sequence alignment for the dataset (Conditions: Single chain, 2 state, pH7, Tm>80˚C) **[9, 15]**

Protein	PDB	Sequence Description
Cytochrome b5 type B	1B5M	gi | 12643974 | sp | P04166.2 | CYB5B_RAT
Pancreatic trypsin inhibitor	1BPI	gi | 115114 | sp | P00974.2 | BPT1_BOVIN
Beta-lactoglobulin	1BSQ	gi | 125910 | sp | P02754.3 | LACB_BOVIN
Divalent-cation tolerance protein CutA	1J2V	gi | 61212444 | sp | O58720.1 | CUTA_PYRHO
Divalent-cation tolerance protein CutA	1NZA	gi | 61212704 | sp | Q7SIA8.1 | CUTA_THET8
Protein P-30	1ONC	gi | 464649 | sp | P22069.2 | RNP30_RANPI
Beta-lactoglobulin	1QG5	gi | 125910 | sp | P02754.3 | LACB_BOVIN
DNA-binding protein	1SSO	gi | 48428878 | sp | P61991.2 | DN71_SULSO
Elongation factor Ts	1TFE	gi | 1169484 | sp | P43895.1 | EFTS_THET8
Alpha-amylase inhibitor HOE-467A	2AIT	gi | 123982 | sp | P01092.2 | IAA_STRTE
Lectin	2SBA	gi | 126151 | sp | P05046.1 | LEC_SOYBN