ELSEVIER

Biophysical Chemistry 119 (2006) 256 – 270

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Effective factors in thermostability of thermophilic proteins

M. Sadeghi [a], H. Naderi-Manesh [b,*], M. Zarrabi [b], B. Ranjbar [b]

[a] National Institute of Genetic Engineering and Biotechnology, P.O. Box 14155-6343,Tehran, Iran
[b] Faculty of Science, Tarbiat Modares University, P.O. Box 14115-175, Tehran, Iran

Received 14 June 2005; received in revised form 13 September 2005; accepted 14 September 2005
Available online 25 October 2005

Abstract

Thermostability of proteins in general and especially thermophilic proteins has been subject of a wide variety of studies based on theoretical and experimental investigation. Thermostability seems to be a property obtained through many minor structural modifications rather than certain amino acids substitution. In comparison with its mesophile homologue in a thermostable protein, usually a number of amino acids are exchanged. A wide variety of theoretical studies are based on comparative investigation of thermophilic proteins characteristics with their mesophilic counterparts in order to reveal their sequences, structural differences and consequently, to relate these observed differences to the thermostability properties. In this work we have compared a dataset of thermophilic proteins with their mesophilic homologues and furthermore, a mesophilic proteins dataset was also compared with its mesophilic homologue. This strategy enabled us first, to eliminate noise or background differences from signals and moreover, the important factors which were related to the thermostability were recognized too. Our results reveal that thermophilic and mesophilic proteins have both similar polar and nonpolar contribution to the surface area and compactness. On the other hand, salt bridges and main chain hydrogen bonds show an increase in the majority of thermophilic proteins in comparison to their mesophilic homologues. In addition, in thermophilic proteins hydrophobic residues are significantly more frequent, while polar residues are less. These findings indicate that thermostable proteins through evolution adopt several different strategies to withstand high temperature environments.

Keywords: Thermostability; Thermophile; Mesophile; Hydrogen bond; Ion pair; Surface accessibility

1. Introduction

It is an ongoing interest in understanding the stability mechanism of the proteins in the organism which are living in the so-called harsh environments such as high pressure, high temperature and non-physiological pH [1]. Despite a series of intense studies on the molecular mechanisms of protein thermostability in many theoretical and experimental research efforts, this subject is still not fully understood [2–6]. Therefore, further investigations of the thermostability mechanisms are not only essential for the theoretical description of protein folding and stability, but also is important for designing efficient enzymes that are functional at high temperature.

Since the sequence of thermophilic and mesophilic proteins are usually composed of the similar distribution of amino acids and similar protein structure, thus comparison of sequence and structure of thermophilic and mesophilic proteins has formed the basis of theoretical efforts. At least four different approaches are used to study thermophilic protein stability. First, comparing a single thermophilic protein structure with one or more of its mesophilic homologues [7–12]. Second, mutational approach that uses protein structure to locate the differences between high and low temperature proteins and thereby propose hypotheses for the bases of thermal adaptation [13–20]. Third, computational studies that utilize data from a dataset of thermophilic proteins and their mesophilic homologues to analyze various features in order to reach to a general conclusion [21–28] and finally, whole genome sequence comparison of thermophilic species with their homologues from mesophilic species. The ability to sequence whole genomes can provide the necessary quantities of data [29–31]. From this diverse collection of studies, it is difficult to draw a general conclusion about the features underlying the increased thermal stability of proteins from thermophilic microorganisms. Several reasons have been attributed to the

* Corresponding author. Tel./fax: +98 21 8800 9730.
E-mail address: naderman@modares.ac.ir (H. Naderi-Manesh).

0301-4622/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2005.09.018

greater stability of thermophilic proteins including better hydrogen bonding, better hydrophobic internal packing, enhanced secondary structure propensity, helix dipole stabilization, burying hydrophobic accessible area, improved electrostatic interaction, better Van der Waals interaction, better affinity for calcium, decrease glycines and enhanced prolins in loops, shorter loops, decrease in number and volume of internal cavities, and increase in aromatic interaction [32].

In this study we constructed a large high-quality dataset containing 60 structures of thermophilic proteins and their mesophilic homologues with structures of equally high quality. The rapidly growing number of structures in Protein Data Bank (PDB) [33] enables us to extend the analysis of important properties in a large dataset of thermophilic proteins and their mesophilic homologues. We determined hydrogen bonds, ion pairs, buried and exposed surface area, amino acid composition and substitution from the sequence and atomic coordination and performed a comparison of properties using the methods of statistical analysis. In addition, we applied a new approach to concluding the results obtained from this comparison to extract important features for thermal adaptation from the background or random differences. For this reason, we constructed a dataset of mesophilic homologues of mesophilic proteins and performed all comparisons and analysis between the two mesophilic homologues dataset to determine which factors differed between thermophilic and mesophilic dataset in order to find the meaningful differences.

2. Materials and methods

A dataset containing thermophilic proteins constructed for this work was chosen from Protein Data Bank (PDB), by searching for the word “thermo”. This search yielded 710 proteins most of them from thermophilic organisms. At the first refinement step, the entries in which protein structures were determined by nuclear magnetic resonance (NMR) were discarded.

All the entries were then examined and the dataset size was reduced by eliminating multiple structures of the same protein and choosing the wild type structure with the highest resolution. The resulting proteins were compared all against all so that none of them have more than 20% identity with all other sequences. Finally, non-homologue thermophilic proteins were retained if there was at least one high resolution crystal structure for their corresponding mesophilic homologues. Through this screening, 60 non-redundant families were remained. For each thermophilic protein, corresponding mesophilic homologue with more than 20% identity and same topology or function was selected. Table 1 shows the list of thermophilic and mesophilic homologues with their PDB entry ID and sequence identity. Temperatures were assigned to each protein according to those of optimal growth or normal living environment for the species involved. Another dataset is the mesophilic homologues of the mesophilic proteins. To determine which factors are specific between thermophilic and mesophilic proteins, we constructed a dataset of mesophilic homologues of mesophilic dataset. These proteins were selected based on more than 25% identity with the mesophilic proteins and high resolution crystal structures and same topology or functions. In this step, we Constructed a dataset for 42 mesophilic proteins listed in Table 2.

3. Amino acid composition

The ratio between the total number of occurrence of each amino acid in each protein and the length of protein for all proteins in each dataset were calculated. Amino acids were grouped into three classes based on their relative polarities. These consisted of hydrophobic residues (Ala, Val, Ile, Leu, Met, Pro, Trp, Phe, Tyr), charged residues (Arg, Lys, His, Glu, Asp) and polar (non-charged) residues (Ser, Thr, Gln, Asn, Cys). For each dataset, mean, variance and standard deviation of amino acid composition were calculated. The statistical test on the thermophilic and mesophilic homologues and mesophilic and mesophilic homologues carried out to determine which amino acids and groups are showing meaningful differences in the thermophile and mesophile sets.

4. Amino acid replacement

Each type of amino acid replacement was counted in the BLASTP alignment [34]. Amino acid replacement was considered in direction of mesophile to thermophile in consistency with other studies. The same calculation was done for replacement in the mesophilic proteins and their mesophilic homologues. In addition to the single amino acid replacement, calculations were done on each group of amino acids (polar, hydrophobic, charged).

5. Exposed and buried surface area

The surface area was calculated by the program written in our group [35]. A probe radius of 1.4 Å was used as the standard water radius. N and O atoms were considered polar and the rest nonpolar. The buried surface for each residue was obtained by subtracting the values calculated for the folded chain from those calculated in the fully extended form. Different ratios between total, polar or nonpolar exposed and buried area were calculated for the all atoms and side chain atom cases. When two homologue proteins have identical subunits, surface area were calculated for the oligomeric state and the section buried between the two monomers were assigned as buried area. But in the case of different oligomeric states, calculation was performed on the common subunits of the two proteins and the coordination of different subunits was eliminated. Statistical test were carried out on the both sample datasets.

6. Hydrogen bonds and ion pairs

The numbers of hydrogen bonds were determined using DSSP program [36]. This program determines the number and the percentage of the total main chain hydrogen bonds, the hydrogen bonds in parallel and antiparallel beta structures and between position i, i+2, i, i+3 and i, i+4 for the different helix

Table 1
Thermophilic proteins and their mesophilic homologues

Thermophile PDB entry ID	Thermophile organism	Subunits	Organism growth temperature	Function	Subunit sequence length	Homolog mesophile PDB entry ID
1A2Z	*Thermococcus litoralis*	a,b,c,d	83	Pyrrolidone carboxyl peptidase	220	1 Aug
1A53	*Sulfolobus solfataricus*	Null	70–85	Indole-3-glycerolphosphate synthase	247	1PII
1A5Z	*Thermotoga maritima*	Null	80	Lactate dehydrogenas	312	9LDT
1A7W	*Methanothermus fervidus*	Null	83	Histone hmfb	68	1EQZ(H,D)
1A8H	*Thermus thermophilus*	Null	75	Methionyl-tRNA synthetase	500	1QQT(a)
1AJ8(a)	*Pyrococcus furiosus*	a,b	97–100	Citrate synthase	371	1A59
1B67	*Methanothermus fervidus*	a,b	83	Histone hmfa	68	1AOI(b)
1B8Z	*Thermotoga maritima*	a,b	80	Histonelike protein hu	67	1IHF(a)
1BJW	*Thermus thermophilus*	a,b	75	Aspartate aminotransferase	382	1BW0
1BMD	*Thermus aquaticus flavus*	a,b	70	Malate dehydrogenase	327	1B8P(a)
1BQC(a)	*Thermobifida fusca*	a	50	Beta-mannanase	302	1A3H
1BRF(a)	*Pyrococcus furiosus*	a	97–100	Rubredoxin	53	4RXN
1BRW(a)	*Geobacillus stearothermophilus*	a,b	55	Pyrimidine nucleoside phosphorylase	433	1OTP
1BVU	*Thermococcus litoralis*	a,b,c,d,e,f	83	Glutamate dehydrogenase	416	1HRD
1BXB	*Thermus thermophilus*	a,b,c,d	75	Xylose isomerase	387	1XIF
1C3U	*Thermotoga maritima*	a,b	80	Adenylosuccinate lyase	423	1AUW
1C9S	*Geobacillus stearothermophilus*	a,b,c,. . .,u,v	55	RNA binding protein	69	1WAP
1CAA	*Pyrococcus furiosus*	Null	97–100	Electron transport	53	2RDV(A)
1COJ	*Aquifex pyrophilus*	a	85	Superoxide dismutase	211	1VAR
1CZ3	*Thermatoga maritima*	a,b	80	Dihydrofolate reductase	164	1AI9
1D3U(a)	*Pyrococcus woesei*	a,b	97–100	Tata-binding protein	181	1TBP
1DR0	*Thermus thermophilus*	a,b	75	3-isopropylmalate dehydrogenase	346	1CNZ(a)
1DT1(a)	*Thermus thermophilus*	a	75	Cytochrome *c* 552	129	1C53
1DV7(a)	*Methanothermobacter thermoautot*	a	65	Orotidine 5′-phosphate decarboxylase	212	1EIX(a)
1DZ3(A)	*Geobacillus stearothermophilus*	a	55	Sporulation response regulator spo0a	123	1NAT
1EFT	*Thermus aquaticus*	Null	70	Elongation factor tu	405	1EFC
1EG5	*Thermotoga maritima*	a,b	80	Aminotransferase. Chain	364	1C0N(a)
1FFH	*Thermus aquaticus*	Null	70	gtpase domains of the signal sequence	287	1FTS
1GLN	*Thermus thermophilus*	Null	75	Glutamyl-tRNA synthetase	468	1EUQ(a)
1GTM	*Pyrococcus furiosus*	a,b,c	97–100	Glutamate dehydrogenase	417	1AUP
1HYT	*Geobacillus thermoproteolyticus*	Null	60	Thermolysin	316	1EZM
1KTQ	*Thermus aquaticus*	Null	70	DNA polymerase	534	2KFN(a)
1MRO	*Methanobacterium thermoautotrop*	a,b,c,d,e,f	65	Methyl-coenzyme m reductase	548	1E6Y(a,d)
1NOX	*Thermus thermophilus*	Null	75	Nadh oxidase	200	1VFR
1OBR	*Thermoactinomyces vulgaris*	Null	50	Carboxypeptidase t	323	2CTC
1OSJ	*Thermus thermophilus*	a,b	75	3-isopropylmalate dehydrogenase	345	1A05
1PHP	*Geobacillus stearothermophilus*	Null	55	3-phosphoglycerate kinase	394	3PGK
1QHO	*Geobacillus stearothermophilus*	a	55	alpha-amylas	686	1CGT
1QMP	*Geobacillus stearothermophilus*	a,b,c,d	55	sporulation response regulator	126	1B00
1QQ0	*Methanosaricina thermophilus*	a	50	Carbonic anhydrase	208	1LXA
1QVB	*Thermosphaera aggregans*	a,b	85	Beta-glycosidase	481	1BGA
1SRV	*Thermus thermophilus*	a	75	Groel (hsp60 class)	145	1DK7
1SSS	*Sulfolobus solfataricus*	a,b	70–85	Iron superoxide dismutase	205	1QNN
1TFE	*Thermus thermophilus*	Null	75	Elongation factor ts	142	1EFU(b,d)
1THM	*Thermoactinomyces vulgaris*	Null	50	Hydrolase(serine protease)	279	1C9J
1TML	*Thermo monospora fusca*	Null	50	Endo-1,4-beta-d-glucanase	286	1BVW(a)
1TMY	*Thermotoga maritima*	Null	80	Chey protein	118	2FSP
1TUX	*Thermoascus aurantiacus*	Null	40	Xylanase	301	1BG4
1VPE	*Thermotoga maritima*	Null	80	Phosphoglycerate kinase	398	16PK
1XGS(a)	*Pyrococcus furiosus*	a,b	97–100	Methionine aminopeptidase	295	1MAT
1XWL	*Geobacillus stearothermophilus*	Null	55	DNA polymerase i	580	1KFD
1XYZ(a)	*Clostridium thermocellum*	a,b	55–62	Glycosyltransferase	322	1XAS
1YKF	*Thermoanaerobium brockii*	a,b,c,d	65	NADP-dependent alcohol dehydrogenase	352	1KEV
1YNA	*Thermomyces lanuginosus*	Null	48	Endo-1,4-beta-xylanase	197	1XND
1ZIN	*Geobacillus stearothermophilus*	Null	55	Adenylate kinase	217	1AK2
2BTM	*Geobacillus stearothermophilus*	a,b	60	Triosephosphate isomerase	256	1TPF
2FXB	*Bacillus thermoproteolyticus*	Null	60	Ferredoxin	81	1FXD
3DAA	Thermophilic *Geobacillus* SP	a,b	60	D-amino acid aminotransferase	277	1A3G
3MDS	*Thermus thermophilus*	a,b	79	Manganese superoxide dismutase	207	3SDP
4PFK	*Geobacillus stearothermophilus*	Null	55	Phosphofructokinase	319	1PFK(a)

Mesophile organism	Subunits	Organism growth temperature	Function	Subunit sequence length	ID %
Bacillus Amylloliquefacious	a,b,c,d	30	Pyroglutamyl peptidase-1	210	35
E. coli	Null	37	*N*-(5′phosphoribosyl)anthranilate isomerase	452	36
Pig	a,b	37	Lactate dehydrogenase	332	40
Gallas gallus	a,b,c,d,e,f,g,h	37	Histone h2a	89	33
E. coli	a	37	Methionyl-tRNA synthetase	546	24
Antarction bacterium	Null	20	Citrate synthase	377	39
Xenopus laevis	a,b,c,d,e,f,g,h	37	Histone h3	83	32
E. coli	a,b	37	Integration host factor	96	37
Trypanosoma cruzi	a,b		Tyrosine aminotransferase	412	28
Aquaspirillum arction	a	4	Malate dehydrogenase	327	61
Bacillus agaradherans	Null	30	Endoglucanase	300	21
Clostridium pasteuriaum	Null	37	Rubredoxin	54	59
E. coli	Null	37	Thymidine phosphorylase	440	44
Clostridium symbiosum	a,b,c	37	Glutamate dehydrogenase	449	35
Streptomyces rubiginosus	Null	28	D-xylose isomerase	388	59
Anas platyrhychos	a,b,c,d	25	Delta 2 crystallin	447	21
Bacillus subtilis	a,b,c,. . .,u,v	30	Trp RNA-binding attenuation protein	68	81
Desulfovibrio vulgaris	a,b,c	37	Rubredoxin	52	72
Human	a,b	37	Manganese superoxide dismutase	198	34
Candida albicans	a,b	37	Dihydrofolate reductase	192	29
Sacharomyces cerevisia	a,b	25	Tata-binding protein	180	42
Salmonelle tiphimorium	a,b	37	3-isopropylmalate dehydrogenase	363	51
Desulfovibrio vulgaris	Null	37	Cytochrome *c*553	79	28
E. coli	a,b,c,d	37	Orotidine 5′-monophosphate decarboxylase	231	22
Bacillus subtilis	Null	30	Sporulation response regulatory protein	119	31
E. coli	a,b	37	Elongation factor	386	70
E. coli	a	37	Csdb protein	404	24
E. coli	Null	37	Ftsy	295	38
E. coli	a	37	Glutaminyl-tRNA synthetase	529	26
Clostridium symbiosum	Null	37	NAD-specific glutamate dehydrogenase	427	35
Pseudomonas aeruginosa	Null	30–37	Elastase	298	32
E. coli	a	37	Klenow fragment	601	49
Methanosarcina barkeri	a,b,c,d,e,f	30–37	Methyl-coenzyme m reductase i alpha subunit	568	65
Vibrio fischeri	a,b	20–25	NAD(p)h\:fmn oxidoreductase	217	27
Bos taurus	Null	37	Carboxypeptidase	307	31
Thiobacillus ferrooxidans	a,b	30	3-isopropylmalate dehydrogenase	357	53
Sacharomyces cerevisia	Null	25	Phosphoglycerate kinase	416	49
Bacillus circulans	Null	30	Cyclodextrin glycosyltransferase	684	48
E. coli	a,b	37	Phosphate regulon transcriptional protein	122	28
E. coli	Null	37	Udp *n*-acetylglucosamine *o*-acyltransferase	262	24
Bacillus polymyna	a,b,c,d	30	Beta-glucosidase	447	24
E. coli	a,b	37	Groel	146	69
Porphyromonas gingivalis	a,b	37	Superoxide dismutase	191	39
E. coli	a,b,c,d	37	Elongation factor tu	282	37
Bacillus lentus	a	26	Serine protease	269	49
Humicola insolens	a	40	Cellobiohydrolase ii	360	27
Bacillus subtilisis	Null	30	Stage 0 sporulation protein f	124	33
Penicilium simplicissimum	Null	24	Endo-1,4-beta-xylanase	302	72
Trypanosoma brucei	Null	37	3-phosphoglycerate kinase	415	47
E. coli	Null	37	Methionine aminopeptidase	263	30
E. coli	Null	37	DNA polymerase i	560	44
Streptomyces lividans	Null	28	1,4-beta-d-xylan xylanohydrolase	295	41
Clostridium beijeninckii	a,b,c,d	37	NADP-dependent alcohol dehydrogenase	351	75
Trichoderma harzianum	Null	22	Xylanase	190	61
Bos taurus	Null	37	Adenylate kinase isoenzyme-2	220	43
Trypanosoma brucei brucei	a,b		Triosephosphate isomerase	250	42
Desulfovibrio gigas	Null	30–37	Ferredoxin Ii	58 r	29
E. coli	a,b,c	37	Amino acid aminotransferase	295	28
Pseudomonas putida	a,b	26	Iron superoxide dismutase	186	42
E. coli	a,b	37	Phosphofructokinase	320	57

Table 2
Mesophilic proteins and their mesophilic homologues

Mesophile PDB entry ID	Mesophile organism	Subunits	Function	Subunit sequence length	E.C. number	Homolog mesophile PDB entry ID	Mesophile organism	Subunits	Function	Subunit sequence length	E.C. number	ID%
16PK	*Trypanosoma brucei*	Null	3-phosphoglycerate kinase	415	2.7.2.3	3PGK	*Sacharomyces cerevisia*	Null	Phosphoglycerate kinase	416	2.7.2.3	43
1A05	*Thiobacillus ferrooxidans*	a,b	3-isopropylmalate dehydrogenase	357	1.1.1.85	1 cm7	*E. coli*	a,b	3-isopropylmalate dehydrogenase	363		52
1A3H	*Bacillus agaradherans*	Null	Endoglucanase	300	3.2.1.4	1EGZ(a,c)	*Erwinia chrysanthemi*	a,b,c	Endoglucanase z	291	3.2.1.4	47
1A59	*Antarction bacterium*	Null	Citrate synthase	377		6CSC(a)	Chicken	a,b	Citrate synthase	437	4.1.3.7	25
1AI9	*Condida albicans*	a,b	Dihydrofolate reductase	192	1.5.1.3	1D8R(a)	*Pneumocystis carinii*	a	Dihydrofolate reductase	205	1.5.1.3	36
1AK2	*Bos taurus*	Null	Adenylate kinase isoenzyme-2	220	2.7.4.3	1AKY	*Saccharomyces cerevisia*	Null	Adenylate kinase	218	2.7.4.3	57
1AUP	*Clostridium symbiosum*	Null	NAD-specific glutamate dehydrogenase	427	1.4.1.2	1HWX	Bovine	a,b,c,d,e,f	Glutamate dehydrogenase	501	1.4.1.3	27
1AUW	*Anas platyrhychos*	a,b,c,d	Delta 2 crystallin	447	4.3.2.1	1AOS	Human	a,b	Argininosuccinate lyase	434	4.3.2.1	70
1AW2	*Vibrio marinus*	a,b,d,e,g,h,j	Triosephosphate isomerase	255	5.3.1.1	1TRE(a)	*E. coli*	a,b	Triosephosphate isomerase tim	255	5.3.1.1	64
1B00	*E. coli*	a,b	Phosphate regulon transcriptional regulatory protein	122		2chf	*Salmonella typhimurium*	Null	Chey	128		30
1B8P(a)	*Aquaspirillum arction*	a	Malate dehydrogenase	327		4MDH	Pig	a,b	Cytoplasmic malate dehydrogenase	334	1.1.1.37	51
1BG4	*Penicilium simplicissimum*	Null	Endo-1,4-beta-xylanase	302	3.2.1.8	1XYF	*Streptomyces olivaceoviridis*	a,b	Endo-1,4-beta-xylanase	427	3.2.1.8	47
1BGA	*Pacillus polymyna*	a,b,c,d	Beta-glucosidase	447	3.2.1.21	1QOX	*Bacillus circulans*	a,b,c,. . .n,o,p	Beta-glucosidase	449	3.2.1.21	65
1BQB	*Staphilococcus aureus*	a	Aureolysin	301	3.4.24.27	1EZM	*Pseudomonas aeruginosa*	Null	Elastase	298	3.4.24.26	31
1BVW(a)	*Humicola insolens*	a	Cellobiohydrolase ii	360	3.2.1.91	3CBH	*Trichoderma reesei*	Null	Cellobiohydrolase II core protein	365	3.2.1.91	64
1C53	*Desulfovibrio vulgaris*	Null	Cytochrome *c* 553	79		1DVH	*Desulfovibrio vulgaris*	Null	Cytochrome *c*553	79		79
1C9J	*Bacillus lentus*	a	Serine protease	269	3.4.21.62	1MEE(a)	*Bacillus pumilus*	a,I	Mesentericopeptidase	275	3.4.21.62	60
1CGT	*Bacillus circulans*	Null	Cyclodextrin glycosyltransferase	684	2.4.1.19	1CXL(a)	*Bacillus circulans*	a	Cyclodextrin-glycosyltransferase	686	2.4.1.19	73
1CNZ(a)	*Salmonelle tiphimorium*	a,b	3-isopropylmalate dehydrogenase	363		1CM7	*E. coli*	a,b	3-isopropylmalate dehydrogenase	363		94
1EFC	*E. coli*	a,b	Elongation factor	386		1d2e	*Bos taurus*	a,b,c,d	Elongation factor tu	397		57
1EIX(a)	*E. coli*	a,b,c,d	Orotidine 5′-monophosphate decarboxylase	231	4.1.1.23	1DBT(a)	*Bacillus subtilis*	a,b,c	Orotidine 5′-phosphate decarboxylase	237	4.1.1.23	41
1EZM	*Pseudomonas aeruginosa*	Null	Elastase	298	3.4.24.26	1ESP	*Bacillus cereus*	Null	Neutral protease mutant e144s	317	3.4.24.28	33
1FXD	*Desulfovibrio gigas*	Null	Ferredoxin Ii	58 r		7FD1(a)	*Azotobacter vinelandii*	a	7-fe ferredoxin i	106		43
1HRD	*Clostridium symbiosu*m	a,b,c	Glutamate dehydrogenase	449	1.4.1.2	1HWX	Bovine	a,b,c,d,e,f	Glutamate dehydrogenase	501	1.4.1.3	28
1IDS	*Mycobacterium tuberculosis*	a,b,c,d	Iron-dependent superoxide dismutase	198	1.15.1.1	1BSM	*Propionibacterium freudenreichi*	a,b	Superoxide dismutase	201	1.15.1.1	63
1IHF(a)	*E. coli*	a,b	Integration host factor	96		1IHF(b)	*E. coli*	a,b	Integration host factor	94		32
1KEV	*Clostridium beijeninckii*	a,b,c,d	NADP-dependent alcohol d ehydrogenase	351	1.1.1.2	1CDO	*Gadus callarios*	a,b	Alcohol dehydrogenase	374	1.1.1.1	28
1MAT	*E. coli*	Null	Methionine aminopeptidase	263	3.4.11.18	1AZ9	*E. coli*	Null	Aminopeptidase p	440	3.4.11.9	23
1NAT	*Bacillus subtilis*	Null	Sporulation response regulatory protein	119		1DC7(a)	*Salmonelle typhimorium*	a	Nitrogen regulation protein	124		37
1QNN	*Porphyromonas gingivalis*	a,b	Superoxide dismutase	191	1.15.1.1	1ISA	*E. coli*	a,b	Iron(ii) superoxide dismutase	192	1.15.1.1	52
1TBP	*Sacharomyces cerevisia*	a,b	Tata-binding protein	180		1QNE(b)	*Arabidopsis thaliana*	a,b	Transcription initiation factor	187		84
1VAR	Human	a,b	Manganese superoxide dismutase	198	1.15.1.1	1IDS	*Mycobacterium tuberculosis*	a,b,c,d	Iron-dependent superoxide dismutase	198	1.15.1.1	52
1VFR	*Vibrio fischeri*	a,b	NAD(p)h\:fmn oxidoreductase	217	1.6.8.1	1DS7(a)	*E. coli*	a,b	Fmn-dependent nitroreductase	217	1.6.99.7	34
1XAS	*Cellulomonas fimi*	Null	1,4-beta-d-xylan xylanohydrolase	295	3.2.1.8	1FH7	*Cellulomonas fimi*	a	Beta-1,4-xylanase	312	3.2.1.91	49
1XIF	*Streptomyces rubiginosus*	Null	D-xylose isomerase	388	5.3.1.5	1BHW	*Actinoplanes missouriensis*	a,b,c,d	Xylose isomerase	392	5.3.1.5	66
1XND	*Trichoderma harzianum*	Null	Xylanase	190	3.2.1.8	2BVV(a)	*Bacillus circulans*	a	Endo-1,4-beta-xylanase	185	3.2.1.8	54
2CTC	*Bos taurus*	Null	Carboxypeptidase	307	3.4.17.1	1PCA	Pig (porcine sus scrofa) pancrea	Null	Procarboxypeptidase a	402	3.4.17.1	87
2FSP	*Bacillus subtilisis*	Null	Stage 0 sporulation protein f	124		1NTR	*Salmonella typhimurium*	Null	Ntrc receiver domain	124		37
2RDV(A)	*Desulfovibrio vulgaris*	a,b,c	Rubredoxin	52		6RXN	*Desulfovibrio desulfuricans*	Null	Rubredoxin	46		53
3SDP	*Pseudomonas putida*	a,b	Iron superoxide dismutase	186	1.15.1.1	1ISA	*E. coli*	a,b	Iron(ii) superoxide dismutase	192	1.15.1.1	69
4RXN	*Clostridium pasteuriaum*	Null	Rubredoxin	54		1RB9	*Desulfovibrio vulgaris*	Null	Rubredoxin	53		72
9LDT	Pig	a,b	Lactate dehydrogenase	332	1.1.1.27	5LDH	Pig (sus scrofa)	Null	Lactate dehydrogenase	334	1.1.1.27	72

structures. Furthermore, the percentage of the total and each type of hydrogen bonds for each of the protein and their mean in each dataset were obtained and compared. Ion pairs were inferred when two opposite charged atoms (carbonyl oxygen atoms of Asp and Glu and nitrogen atoms of Arg, Lys, and His) were closer than 4.0 Å to each others. Since some proteins have more than one subunit, ion pairs were calculated as intra- and inter-chains per subunit and to compare proteins with different sizes, ion pairs per residue were also calculated. We also determined the ratio of the free and the bonded charge residues in each datasets and the results were analyzed.

7. Results and discussions

In all cases to find factors responsible for the protein stability based on the comparison of thermophilic–mesophilic homologues, either one protein or the whole dataset of the thermophilic and mesophilic proteins are compared. Since energy difference between a thermophile and its homologue mesophile proteins is small, we do not expect to include every observed difference between these proteins as a thermostable related difference. Consequently, the comparative studies of the mesophile–mesophile homologues provide the possibility of assigning stability related differences. This strategy was applied to all the considered parameters.

8. Amino acid composition

Table 3 shows the amino acid composition of the 60 thermophilic proteins and their mesophilic homologues and also the amino acid composition of the 42 mesophilic proteins (selected from mesophilic homologues of thermophilic proteins) and their mesophilic homologues. The bold scores are significantly different at the 5% level of significance. Of the individual amino acids the propensity of His, Ser, Asn, Gln, and Cys are decreased while Glu and Pro are more in the thermophilic proteins relative to their mesophilic counterpart. The thermophilic proteins also show an increase in charged and hydrophobic amino acids contents and in the other hand a decrease in uncharged polar amino acids. In addition, the comparison of amino acids content in mesophilic homologues reveal a decrease in Glu, Met and charged amino acids and an increase in Trp. Consequently, results obtained from the two dataset groups, show that observed increase in Glu and the charged amino acid may be unrelated to the thermostability of the thermophilic proteins. The amino acid content comparison in studies based on the protein family or the genome comparison reveal different results specially in the propensity of charged and hydrophobic amino acids [2,37,24,29].

There are both similarities and differences between our finding and those reported previously. The main difference is an increase in the Glu and charged amino acid content in both mesophilic datasets. This will eliminate their role as a possible mean to enhance the protein thermostability. Although it is generally believed that an increase in charged residue content is probably due to the enhanced occurrence of salt bridges and ion pairs in thermophilic proteins [38], but our results show that this increase could be an indirect consequence of a decrease in the uncharged polar residues. The replacement of

Table 3
Comparison of amino acid composition in thermophile–mesophile and mesophile–mesophile homologues

AA	Mean ± St. dev			Mean ± St. dev		
	Thermophile	Homolog mesophile	Paired P ($T \leq t$) two tail	Mesophile	Homolog mesophile	Paired P ($T \leq t$) two tail
His	1.91 ± 1.21	2.26 ± 1.26	***0.0166***	2.2 ± 1.31	2.32 ± 1.05	0.479877
Glu	8.05 ± 2.67	6.79 ± 2.56	***0.0003***	6.6 ± 2.8	5.85 ± 2.43	0.015595
Lys	6.69 ± 3.28	6.14 ± 2.78	0.1142	5.92 ± 2.78	6.05 ± 2.6	0.730499
Arg	5.21 ± 2.58	4.8 ± 2.66	0.1852	4.15 ± 2.28	3.81 ± 1.97	0.218976
Asp	5.88 ± 2.59	6.14 ± 2.53	0.3497	6.59 ± 2.84	6.32 ± 2.22	0.374612
Ser	4.34 ± 2.26	5.44 ± 2.53	***0.0005***	5.73 ± 2.79	6.24 ± 2.28	0.08362
Asn	3.81 ± 2.18	4.31 ± 1.95	***0.0167***	4.86 ± 2.22	4.69 ± 2.2	0.486698
Gln	2.64 ± 1.62	3.13 ± 1.34	***0.0422***	3.1 ± 1.4	3.4 ± 1.33	0.296564
Cys	0.92 ± 1.5	1.22 ± 1.85	***0.0326***	1.33 ± 2.15	1.5 ± 2.31	0.268476
Thr	4.7 ± 2.08	5.16 ± 1.79	0.0705	5.04 ± 1.97	5.16 ± 2.21	0.641969
Pro	4.79 ± 1.7	4.32 ± 2.11	***0.0148***	4.6 ± 2.19	4.49 ± 2.02	0.594191
Ala	9.52 ± 3.18	9.31 ± 3.04	0.6257	9.29 ± 3.62	10.15 ± 3.2	0.093046
Gly	8.18 ± 2.26	8.49 ± 2.69	0.3326	8.38 ± 2.71	8.26 ± 2.36	0.625749
Ile	6.7 ± 2.58	6.24 ± 2.38	0.246	5.83 ± 2.63	5.19 ± 2.06	0.067652
Leu	8.3 ± 2.97	7.91 ± 2.67	0.1943	7.71±2.86	7.96±2.71	0.299113
Met	2.29 ± 1.39	2.57 ± 1.31	0.0996	2.61 ± 1.45	2.27 ± 1.23	***0.046214***
Phe	3.43 ± 1.24	3.67 ± 1.22	0.1968	3.77 ± 1.34	3.79 ± 1.4	0.912037
Trp	1.19 ± 1.14	1.18 ± 1.05	0.9185	1.43 ± 1.17	1.74 ± 1.29	***0.010712***
Tyr	3.6 ± 2.03	3.33 ± 1.69	0.2401	3.63 ± 1.9	3.72±1.87	0.576613
Val	7.77 ± 2.11	7.51 ± 1.86	0.4482	7.16 ± 1.96	7 ± 2.04	0.57194
Charged	27.74 ± 5.24	26.13 ± 5.21	***0.0003***	25.45 ± 5.52	24.36 ± 4.89	***0.035835***
Polar	16.4 ± 5.62	19.27 ± 4.67	***2E−07***	20.06 ± 5.28	20.98 ± 4.92	0.076487
Hydrophobe	55.77 ± 3.59	54.52 ± 3.65	***0.0177***	54.41 ± 3.93	54.57 ± 3.04	0.771213

Bold and italic numbers show amino acids that have significant difference.

polar residues by charged ones would provide less labile residues while retains the hydrogen bonding capacity. The decrease in the content of uncharged polar residues is in good agreement with the other studies [29,26–28,31]. This decrease is likely to minimize deamination, oxidation and backbone cleavage involving Asn, Gln and Cys that are induced by temperature [39]. It is also assumed that a reduction in uncharged polar residues will diminish the hydrogen bonding capability, while some could be retained by switching to charged amino acids. The mesophile–mesophile homologue comparison shows that the reduction of the uncharged polar residues has specific role in the thermostability of thermophilic proteins. Pro is a rigid amino acid which has been used to increase the protein stability in the several mutational studies [40,19] and hence an increase of Pro content may be due to the increase of the thermophilic protein rigidity. Even though individual hydrophobic residues do not show significant differences between the thermophilic proteins but overall hydrophobic group shows an increase in thermophilic proteins. This could be due to the role of the hydrophobic effect that destabilizes the unfolded forms and it will increase with the temperature [41,42].

9. Amino acid replacements

To investigate the effect of directional replacements of different type of amino acids number of replacements both in forward and reverse direction were calculated. Since frequency of amino acids in different proteins is not the same, number of replacements was normalized. This was done by dividing the number of replaced amino acid in aligned sequences to the total number of that amino acid. The directional replacement was defined based on the ratio and difference of forward and reverse replacements. Forward replacement means the replacement in the direction of mesophilic to thermophilic homologues. To eliminate the noise due to the evolution, the same comparison was done between the mesophile–mesophile homologues. The final alignments between 60 thermophilic proteins and their mesophilic homologues comprised of 26,307 aligned pair of amino acids of which 5342 (20%) were uncharged polar, 14,400 (55%) were hydrophobic and 6559 (25%) were charged amino acids. From 20,573 paired amino acids of mesophile–mesophile homologues, 4350 (21%) were polar, 11,218 (55%) were hydrophobic and 5005 (24%) were charged amino acids.

Table 4 shows the percentage of replacements between the polar, charged and hydrophobic groups of amino acids in the forward and the reverse direction within the thermophilic–mesophilic and mesophilic–mesophilic datasets. in the forward and reverse direction and their ratio and difference for the polar, hydrophobic and charged groups between mesophile–thermophile and mesophile–mesophile homologues. Results show that the observed replacements cause a decrease in the uncharged polar group content and an increase in the hydrophobic ones. Although it seems that the replacement ratio of charged to hydrophobic amino acids in the mesophile–thermophile homologues is significant, but the comparison of this replacement with the mesophile–mesophile homologues show that this replacement is not significant and probably with no specific role in thermostability. Tables 5 and 6 report all 380 replacement types between mesophile–thermophile and mesophile–mesophile homologues, respectively. We are presenting our results similar to the results obtained by P.J. Haney for mesophilic and extremely thermophilic Metanococcus species [29]. The bold entries indicate pairs of amino acids that have a directional replacement bias with the relative ratio of forward to reverse replacement more than one. Tables 7, 8 and 9 show 30 amino acid replacements that are most biased in the ratio, most frequent and most biased in the number between the mesophile–thermophile and the mesophile–mesophile homologues, respectively. Tables 5 and 6 show that from 75 possible replacements of polar amino acids with charged and hydrophobic ones in mesophiles and their thermophile homologues, 56 pairs are directional, but in mesophiles and their mesophile homologues, 44 pairs are directional. This comparison shows 30 and 31 directional replacements in mesophiles–thermophiles and mesophiles–mesophiles, respectively from 100 possible replacements of hydrophobic amino acids and 36 and 45 directional replacements from 75 possible replacements of charged amino acids, respectively. The replacements with the most biased ratio situation in Table 7 are very rare or we witness similar

Table 4
Amino acid groups replacements in aligned homologous proteins in forward and reverse direction and their ratio and difference

		Mesophile–thermophile homologues					Mesophile–mesophile homologues		
	Polar		Hydrophobe	Charged		Polar		Hydrophobe	Charged
Polar	Forward	47.913	28.704	20.74	Polar	Forward	66.368	18.989	14.644
	Reverse		8.145	13.076		Reverse		7.372	12.647
	Ratio		3.524	1.586		Ratio		2.576	1.158
	Difference		20.559	7.664		Difference		11.617	1.997
Hydrophobe	Forward	8.145	80.243	8.983	Hydrophobe	Forward	7.372	86.415	6.213
	Reverse	28.704		19.651		Reverse	18.989		12.907
	Ratio	0.284		0.457		Ratio	0.388		0.481
	Difference	−20.559		−10.668		Difference	−11.617		−6.694
Charged	Forward	13.076	19.651	65.066	Charged	Forward	12.647	12.907	74.446
	Reverse	20.74	8.983			Reverse	14.644	6.213	
	Ratio	0.63	2.188			Ratio	0.864	2.077	
	Difference	−7.664	10.668			Difference	−1.997	6.694	

Table 5
Percent of amino acid replacements from mesophile to thermopile proteins in forward direction and their ratio and difference

Homolog mesophile		Thermophile amino acid																			
		Uncharged polar					Nonpolar										Charged				
		Ser	Gln	Asn	Thr	Cys	Gly	Ala	Met	Tyr	Phe	Val	Leu	Pro	Ile	Trp	His	Asp	Glu	Lys	Arg
Uncharged polar																					
Ser	Forward	27.098	3.139	**5.637**	**9.673**	**1.089**	**6.342**	**12.236**	0.705	**1.602**	0.833	**2.306**	**2.05**	**2.755**	**2.178**	**0.192**	1.153	**5.317**	**5.83**	**3.33**	**12.947**
	Ratio		0.997	**1.057**	**1.475**	**0.223**	**1.934**	**2.206**	0.433	**1.175**	0.872	**1.193**	**3.233**	**1.151**	**1.679**	**>0.192**	0.438	**1.227**	**2.579**	**1.12**	**1.067**
Gln	Forward	**3.149**	27.47	**4.669**	**3.583**	0.109	**3.366**	**5.863**	**1.846**	**2.063**	0.651	**3.257**	**3.257**	**2.714**	**1.629**	0	**2.063**	**4.343**	**12.704**	**9.446**	**6.08**
	Ratio	**1.003**		**1.19**	**1.499**	0.29	**3.08**	**2.438**	**1.03**	**1.969**	0.568	**3.287**	**3.595**	**1.965**	**5.272**	0	**1.11**	**1.166**	**2.738**	**2.563**	**1.514**
Asn	Forward	5.333	3.922	39.137	4.627	**0.392**	**7.059**	**4.863**	0.314	1.49	**1.412**	**1.804**	**2.196**	1.255	**1.176**	0.471	2.745	**6.902**	**5.882**	**4.549**	**2.902**
	Ratio	0.946	0.84		0.996	**0.348**	**2.273**	**1.99**	0.386	0.618	**2.464**	**1.733**	**3.728**	0.649	**2.381**	0.908	0.985	**1.119**	**2.746**	**1.234**	**1.084**
Thr	Forward	6.557	2.391	**4.645**	37.842	0.683	**2.732**	**7.719**	1.23	**2.186**	0.751	**7.036**	**2.664**	1.913	**3.484**	0.205	0.546	**3.689**	**3.825**	**4.167**	2.732
	Ratio	0.678	0.667	**1.004**		0.091	**2.84**	**2.397**	0.581	**1.158**	0.393	**1.632**	**1.278**	0.693	**1.128**	0.197	0.353	**1.726**	**2.009**	**1.653**	0.907
Cys	Forward	**4.887**	**0.376**	**1.128**	**7.519**	22.932	**5.639**	**12.782**	**3.008**	**2.632**	3.008	**14.286**	**4.887**	**0.752**	**5.639**	**0.752**	0	**2.256**	**1.128**	**2.256**	**0.752**
	Ratio	**4.488**	**3.45**	**2.878**	**11.009**		**64.816**	**16.472**	**4.621**	**12.533**	4.496	**32.031**	**11.978**	**4.087**	**8.305**	**>0.752**	0	**6.164**	**>1.128**	**8.71**	**8.952**
Nonpolar																					
Gly	Forward	3.279	1.093	3.105	0.962	0.087	66.025	6.209	0.306	0.962	**0.875**	**0.875**	**0.831**	2.536	0.525	0.35	0.831	2.143	1.88	2.23	1.224
	Ratio	0.517	0.325	0.44	0.352	0.015		**1.28**	0.209	0.918	**1.308**	**1.471**	**1.222**	0.918	0.944	0.269	0.268	0.524	0.854	0.676	0.563
Ala	Forward	5.547	2.405	2.444	3.22	0.776	4.849	42.32	1.202	**1.202**	1.009	5.896	**4.616**	4.267	2.521	0.194	0.892	2.211	4.538	3.685	2.832
	Ratio	0.453	0.41	0.503	0.417	0.061	0.781		0.273	**1.433**	0.258	0.684	**1.158**	0.927	0.609	0.149	0.23	0.787	0.771	0.663	0.664
Met	Forward	**1.629**	1.792	**0.814**	**2.117**	0.651	**1.466**	4.397	29.642	2.606	3.42	**9.446**	**15.635**	0.977	**11.401**	0.814	1.303	**0.814**	2.932	**1.954**	**2.932**
	Ratio	**2.311**	0.971	**2.592**	**1.721**	0.216	**4.791**	3.658		**1.554**	1.377	**3.344**	4.209	0.964	**3.482**	0.448	0.765	**1.668**	4.483	**1.44**	**3.894**
Tyr	Forward	1.363	1.048	**2.411**	1.887	0.21	**1.048**	0.839	1.677	52.516	11.426	2.411	**4.822**	0.734	**3.04**	2.411	2.83	1.258	**1.782**	**1.677**	**3.459**
	Ratio	0.851	0.508	**1.618**	0.863	0.08	**1.089**	0.698	0.644		**1.031**	0.973	**2.217**	0.664	**2.051**	0.265	0.59	0.859	**1.664**	**2.161**	**2.066**
Phe	Forward	**0.955**	**1.146**	0.573	**1.91**	0.669	0.669	**3.916**	2.483	11.079	40.21	**5.731**	**11.079**	1.146	**7.545**	3.629	1.719	**0.669**	1.242	**0.86**	**1.528**
	Ratio	**1.146**	**1.76**	0.406	**2.543**	0.222	0.765	**3.881**	0.726	0.97		**2.754**	**2.629**	1.555	**2.656**	0.517	0.529	**1.371**	2.986	**1.108**	**1.141**
Val	Forward	1.933	0.991	1.041	4.311	0.446	0.595	**8.622**	2.825	**2.478**	2.081	40.139	8.523	1.338	14.321	0.149	0.297	0.545	2.081	**2.775**	1.586
	Ratio	0.838	0.304	0.577	0.613	0.031	0.68	**1.462**	0.299	**1.028**	0.363		0.904	0.559	0.678	0.036	0.107	0.425	0.874	**1.101**	0.729
Leu	Forward	0.634	0.906	0.589	2.084	0.408	0.68	3.987	3.715	2.175	4.214	**9.425**	46.896	1.042	13.049	0.589	0.816	1.133	1.45	1.541	2.175
	Ratio	0.309	0.278	0.268	0.782	0.083	0.818	0.864	0.238	0.451	0.38	**1.106**		0.539	0.855	0.108	0.377	0.977	0.641	0.582	0.743
Pro	Forward	2.394	1.381	**1.934**	**2.762**	0.184	**2.762**	4.604	1.013	1.105	0.737	**2.394**	1.934	60.589	1.381	0.276	0.46	2.67	**4.788**	2.21	**2.302**
	Ratio	0.869	0.509	**1.541**	**1.444**	0.245	**1.089**	1.079	1.037	1.505	0.643	**1.789**	1.856		0.771	>0.276	0.198	0.994	**1.963**	0.9	**1.1**
Ile	Forward	1.297	0.309	0.494	3.088	0.679	**0.556**	4.138	3.274	1.482	2.841	**21.124**	15.256	1.791	35.516	0.803	0.556	0.556	0.926	1.668	1.544
	Ratio	0.596	0.19	0.42	0.886	0.12	**1.059**	1.641	0.287		0.488	0.377	**1.475**	1.169		0.343	0.239	0.607	0.623	0.697	0.971
Trp	Forward	0	0.519	**0.519**	**1.039**	0	**1.299**	1.299	1.818	9.091	7.013	4.156	5.455	0	**2.338**	59.221	**1.558**	0.779	0.26	**1.299**	**1.558**
	Ratio	0	>0.519	**1.102**	**5.068**	0	**3.711**	6.696	2.233	3.771	1.932	27.893	9.261	0	**2.912**		**2.517**	4.257	1.092	**19.985**	**4.651**
Charged																					
His	Forward	**2.632**	1.858	**2.786**	**1.548**	**1.393**	**3.096**	**3.87**	**1.703**	**4.799**	**3.251**	**2.786**	**2.167**	**2.322**	**2.322**	0.619	**48.297**	2.632	2.477	**3.87**	**4.644**
	Ratio	**2.283**	0.901	**1.015**	**2.835**	**1.393**	**3.726**	**4.339**	**1.307**	**1.696**	**1.891**	**9.38**	**2.656**	**5.048**	**4.176**	0.397		**2.536**	1.666	**4.607**	**2.921**
Asp	Forward	4.335	3.724	6.166	2.137	0.366	**4.09**	**2.808**	0.488	**1.465**	0.488	**1.282**	**1.16**	**2.686**	**0.916**	0.183	1.038	**48.168**	11.538	3.053	2.442
	Ratio	0.815	0.857	0.893	0.579	0.162	**1.909**	**1.27**	0.6	**1.165**	0.729	**2.352**	**1.024**	**1.006**	**1.647**	0.235	0.394		**1.197**	0.694	0.768
Glu	Forward	2.261	4.64	2.142	1.904	0	**2.201**	**5.889**	0.654	1.071	0.416	**2.38**	**2.261**	**2.439**	**1.487**	0.238	1.487	9.637	**44.081**	7.02	4.402
	Ratio	0.388	0.365	0.364	0.498	0	**1.171**	**1.298**	0.223	0.601	0.335	**1.144**	**1.559**	**0.509**	**1.606**	0.915	0.6	0.835		0.734	0.674
Lys	Forward	2.973	3.685	3.685	2.521	0.259	**3.297**	**5.559**	1.357	0.776	0.776	2.521	**2.65**	**2.456**	**2.392**	0.065	0.84	**4.396**	9.567	37.233	10.666
	Ratio	0.893	0.39	0.81	0.605	0.115	**1.478**	**1.509**	0.694	0.463	0.902	0.908	**1.72**	**1.111**	**1.434**	0.05	0.217	**1.44**	1.363		0.973
Arg	Forward	2.762	4.017	2.678	**3.013**	0.084	**2.176**	**4.268**	0.753	1.674	1.339	**2.176**	**2.929**	**2.092**	**1.59**	0.335	1.59	**3.18**	6.527	**10.962**	43.766
	Ratio	0.937	0.661	0.923	**1.103**	0.112	**1.778**	**1.507**	0.257	0.484	0.876	**1.372**	**1.347**	**0.909**	**1.03**	0.215	0.342	**1.302**	1.483	**1.028**	

Bold numbers show amino acids that have significant difference.

Table 6
Percent of amino acid replacements from mesophile to mesophile proteins in forward direction and their ratio and difference

Homolog mesophile		Mesophile amino acid																			
		Uncharged polar					Nonpolar										Charged				
		Ser	Gln	Asn	Thr	Cys	Gly	Ala	Met	Tyr	Phe	Val	Leu	Pro	Ile	Trp	His	Asp	Glu	Lys	Arg
Uncharged polar																					
Ser	Forward	48.531	1.884	**4.521**	7.913	0.528	**4.295**	**8.666**	0.452	0.904	**0.528**	**1.884**	**1.432**	1.809	**0.904**	0.151	0.904	3.316	**3.24**	**3.24**	1.356
	Ratio		0.607	**1.04**	0.969	0.287	**1.313**	**1.405**	0.94	0.633	**1.033**	**1.173**	**1.236**	0.688	**1.04**	0.505	0.416	0.893	**1.287**	**1.035**	0.748
Gln	Forward	**3.103**	43.865	**5.36**	**3.103**	0.141	**1.128**	**4.09**	1.128	0.423	**1.269**	**2.257**	**2.68**	**2.116**	**0.705**	0.141	1.975	**3.385**	**10.155**	**7.052**	**4.372**
	Ratio	**1.647**		**1.828**	**1.475**	0.192	**1.725**	**2.739**	0.361	0.362	**1.986**	**2.344**	**2.434**	**2.146**	**1.116**	0.472	0.833	**1.714**	**1.736**	**1.959**	**1.134**
Asn	Forward	4.348	2.932	55.308	**3.842**	0.202	**3.842**	**3.64**	0.506	0.809	**0.809**	**0.303**	**1.82**	0.506	**1.314**	0.202	1.719	**7.078**	**2.73**	**3.64**	1.416
	Ratio	0.962	0.547		**1.268**	0.549	**1.762**	**2.523**	0.351	0.692	**1.266**	**1.179**	**1.653**	0.66	**1.663**	0.337	0.414	**1.4**	**1.769**	**1.257**	0.657
Thr	Forward	**8.165**	2.104	3.03	53.451	0.589	**2.104**	**4.461**	0.673	0.758	**0.926**	**4.714**	**2.02**	1.599	**2.441**	0	0.673	2.104	2.694	2.778	1.263
	Ratio	**1.032**	0.678	0.789		0.16	**2.031**	**1.228**	0.255	0.449	**1.209**	**1.596**	**1.584**	0.695	**1.065**	0	0.31	0.951	0.829	0.986	0.696
Cys	Forward	**1.838**	**0.735**	**0.368**	**3.676**	64.338	**1.838**	**6.618**	**0.735**	**1.103**	**1.471**	**5.515**	**2.941**	0.368	**2.206**	0.735	0	**2.206**	0.368	**0.735**	0
	Ratio	**3.481**	**5.213**	**1.822**	**6.241**		**11.207**	**7.386**	**1.528**	**2.125**	**3.841**	**8.59**	**7.244**	0.368	**9.308**	0.735	0	**27.924**	0.368	**9.423**	0
Nonpolar																					
Gly	Forward	3.272	0.654	2.181	1.036	0.164	75.245	**4.035**	0.382	0.327	0.273	**0.927**	0.491	0.654	**0.327**	0	0.545	2.181	0.872	1.363	0.709
	Ratio	0.762	0.58	0.568	0.492	0.089		**1.158**	0.227	0.36	0.713	**1.804**	0.652	0.427	**1.035**	0	0.212	0.767	0.596	0.581	0.569
Ala	Forward	6.169	1.493	1.443	3.632	0.896	3.483	53.234	0.945	1.144	0.697	4.03	**2.488**	1.841	1.592	0.149	0.547	2.338	3.98	2.935	1.592
	Ratio	0.712	0.365	0.396	0.814	0.135	0.863		0.281	0.587	0.39	0.826	**1.101**	0.4	0.672	0.498	0.277	0.74	0.778	0.834	0.638
Met	Forward	**0.481**	**3.125**	**1.442**	**2.644**	0.481	**1.683**	**3.365**	45.192	**0.721**	3.125	**6.01**	**14.423**	**0.962**	**6.25**	0.24	0.481	**0.721**	**1.923**	**1.923**	**1.202**
	Ratio	**1.064**	**2.77**	**2.85**	**3.929**	0.654	**4.406**	**3.561**		**1.389**	2.72	**3.9**	**3.83**	**4.393**	**1.364**	0.803	0.811	**4.563**	**4.736**	**2.732**	**2.12**
Tyr	Forward	**1.429**	**1.169**	**1.169**	**1.688**	0.519	**0.909**	**1.948**	0.519	61.818	8.701	**1.948**	**3.247**	0.519	**2.078**	2.078	2.857	**1.039**	**1.429**	**1.688**	0.779
	Ratio	**1.581**	**2.764**	**1.445**	**2.227**	0.471	**2.78**	**1.703**	0.72		0.908	**1.895**	**1.698**	0.519	**2.923**	0.315	0.688	**2.63**	**5.857**	**3.08**	0.491
Phe	Forward	0.511	0.639	0.639	0.766	0.383	**0.383**	**1.788**	1.149	**9.579**	62.58	**3.704**	**6.386**	0.255	**2.81**	2.81	1.022	**0.894**	0.383	0.383	**0.639**
	Ratio	0.968	0.504	0.79	0.827	0.26	**1.403**	**2.565**	0.368	**1.101**		**3.204**	**1.75**	0.582	**1.186**	0.348	0.646	**2.263**	0.786	0.544	**2.815**
Val	Forward	1.606	0.963	0.257	2.954	0.642	0.514	**4.881**	1.541	1.028	1.156	56.904	**7.001**	0.706	12.267	0.257	0.193	0.257	0.899	1.413	1.092
	Ratio	0.852	0.427	0.848	0.627	0.116	0.554	**1.211**	0.256	0.528	0.312		**1.151**	0.645	0.765	0.107	0.195	0.407	0.791	0.752	0.688
Leu	Forward	1.159	1.101	1.101	1.275	0.406	**0.753**	2.26	3.766	1.912	3.65	6.083	60.545	0.463	8.401	0.232	0.406	0.348	1.043	1.448	0.869
	Ratio	0.809	0.411	0.605	0.631	0.138	**1.534**	0.908	0.261	0.589	0.572	0.869		0.423	0.724	0.097	0.147	0.551	0.802	0.661	0.766
Pro	Forward	**2.629**	0.986	**0.767**	**2.3**	0	**1.533**	**4.6**	0.219	0	**0.438**	**1.095**	**1.095**	73.384	**0.986**	0.11	0.219	**1.862**	**2.3**	**2.519**	0.986
	Ratio	**1.453**	0.466	**1.516**	**1.438**	0	**2.344**	**2.499**	0.228	0	**1.718**	**1.551**	**2.365**		**1.387**	0.368	0.369	**1.386**	**1.573**	**2.146**	0.791
Ile	Forward	0.869	0.632	0.79	2.291	0.237	0.316	**2.37**	4.581	0.711	2.37	**16.035**	**11.611**	0.711	50.632	0.474	0.079	0.316	**1.106**	1.027	0.632
	Ratio	0.961	0.896	0.601	0.939	0.107	0.966	**1.489**	0.733	0.342	0.843	**1.307**	**1.382**	0.721		0.226	0.079	**0.667**	1.134	0.772	0.557
Trp	Forward	**0.299**	**0.299**	**0.599**	0.898	0	0.898	**0.299**	0.299	**6.587**	**8.084**	**2.395**	**2.395**	**0.299**	**2.096**	70.958	**0.599**	**0.299**	0.299	0.299	**1.198**
	Ratio	**1.98**	**2.121**	**2.965**	0.898	0	0.898	**2.007**	1.246	**3.17**	**2.877**	**9.319**	**10.323**	**2.718**	**4.422**		**3.025**	**1.892**	0.299	0.299	**10.602**
Charged																					
His	Forward	**2.174**	**2.372**	**4.15**	**2.174**	0.791	**2.569**	**1.976**	**0.593**	**4.15**	**1.581**	**0.988**	**2.767**	**0.593**	0	0.198	62.451	**1.186**	**2.372**	**2.372**	**2.964**
	Ratio	**2.405**	**1.201**	**2.414**	**3.23**	0.791	**4.714**	**3.612**	**1.233**	**1.453**	**1.547**	**5.119**	**6.815**	**2.708**	0	0.331		**1.365**	**2.246**	**2.02**	**2.906**
Asp	Forward	**3.712**	1.975	5.055	**2.212**	0.079	**2.844**	**3.16**	0.158	0.395	0.395	**0.632**	**0.632**	1.343	**0.474**	0.158	0.869	61.848	7.741	2.607	1.106
	Ratio	**1.119**	0.583	0.714	**1.051**	0.036	**1.304**	**1.352**	0.219	0.38	0.442	**2.459**	**1.816**	0.721	**1.5**	0.528	0.733		0.982	0.775	0.813
Glu	Forward	2.518	5.849	1.543	**3.249**	0	**1.462**	**5.118**	0.406	0.244	**0.487**	**1.137**	**1.3**	1.462	0.975	0	1.056	**7.88**	53.94	5.768	1.625
	Ratio	0.777	0.576	0.565	**1.206**	0	**1.677**	**1.286**	0.211	0.171	**1.272**	**1.265**	**1.246**	0.636	0.882	0	0.445	**1.018**		0.933	0.512
Lys	Forward	3.13	3.599	2.895	**2.817**	0.078	**2.347**	**3.521**	0.704	0.548	**0.704**	**1.878**	**2.191**	1.174	**1.33**	0	1.174	**3.365**	**6.182**	51.252	7.512
	Ratio	0.966	0.51	0.795	**1.014**	0.106	**1.722**	**1.2**	0.366	0.325	**1.838**	**1.329**	**1.513**	0.466	**1.295**	0	0.495	**1.291**	**1.072**		0.808
Arg	Forward	**1.814**	3.855	**2.154**	**1.814**	0.227	**1.247**	**2.494**	0.567	**1.587**	0.227	**1.587**	**1.134**	**1.247**	**1.134**	0.113	1.02	**1.361**	**3.175**	**9.297**	61.451
	Ratio	**1.338**	0.882	**1.521**	**1.436**	0.227	**1.759**	**1.567**	0.472	**2.037**	0.355	**1.453**	**1.305**	**1.265**	**1.794**	0.094	0.344	**1.231**	**1.954**	**1.238**	

Bold numbers show amino acids that have significant difference.

Table 7
Amino acid replacements that are most biased in ratio between mesophile–thermophile and mesophile–mesophile homologues

Replacement	Mesophile–thermophile replacement most biased in ratio				Replacement	Mesophile–mesophile replacement most biased in ratio			
	Forward	Reverse	Ratio	Difference		Forward	Reverse	Ratio	Difference
Cys–Gly	5.639	0.087	64.816	5.552	Cys–Asp	2.206	0.079	27.924	2.127
Cys–Val	14.286	0.446	32.031	13.84	Cys–Gly	1.838	0.164	11.207	1.674
Trp–Val	4.156	0.149	27.893	4.007	Trp–Arg	1.198	0.113	10.602	1.085
Trp–Lys	1.299	0.065	19.985	1.234	Trp–Leu	2.395	0.232	10.323	2.163
Cys–Ala	12.782	0.776	16.472	12.006	Cys–Lys	0.735	0.078	9.423	0.657
Cys–Tyr	2.632	0.21	12.533	2.422	Trp–Val	2.395	0.257	9.319	2.138
Cys–Leu	4.887	0.408	11.978	4.479	Cys–Ile	2.206	0.237	9.308	1.969
Cys–Thr	7.519	0.683	11.009	6.836	Cys–Val	5.515	0.642	8.59	4.873
His–Val	2.786	0.297	9.38	2.489	Cys–Ala	6.618	0.896	7.386	5.722
Trp–Leu	5.455	0.589	9.261	4.866	Cys–Leu	2.941	0.406	7.244	2.535
Cys–Arg	0.752	0.084	8.952	0.668	His–Leu	2.767	0.406	6.815	2.361
Cys–Lys	2.256	0.259	8.71	1.997	Cys–Thr	3.676	0.589	6.241	3.087
Cys–Ile	5.639	0.679	8.305	4.96	Tyr–Glu	1.429	0.244	5.857	1.185
Trp–Ala	1.299	0.194	6.696	1.105	Cys–Gln	0.735	0.141	5.213	0.594
Cys–Asp	2.256	0.366	6.164	1.89	His–Val	0.988	0.193	5.119	0.795
Gln–Ile	1.629	0.309	5.272	1.32	Met–Glu	1.923	0.406	4.736	1.517
Trp–Thr	1.039	0.205	5.068	0.834	His–Gly	2.569	0.545	4.714	2.024
His–Pro	2.322	0.46	5.048	1.862	Met–Asp	0.721	0.158	4.563	0.563
Met–Gly	1.466	0.306	4.791	1.16	Trp–Ile	2.096	0.474	4.422	1.622
Trp–Arg	1.558	0.335	4.651	1.223	Met–Gly	1.683	0.382	4.406	1.301
Cys–Met	3.008	0.651	4.621	2.357	Met–Pro	0.962	0.219	4.393	0.743
His–Lys	3.87	0.84	4.607	3.03	Met–Thr	2.644	0.673	3.929	1.971
Cys–Phe	3.008	0.669	4.496	2.339	Met–Val	6.01	1.541	3.9	4.469
Cys–Ser	4.887	1.089	4.488	3.798	Cys–Phe	1.471	0.383	3.841	1.088
Met–Glu	2.932	0.654	4.483	2.278	Met–Leu	14.423	3.766	3.83	10.657
His–Ala	3.87	0.892	4.339	2.978	His–Ala	1.976	0.547	3.612	1.429
Trp–Asp	0.779	0.183	4.257	0.596	Met–Ala	3.365	0.945	3.561	2.42
Met–Leu	15.635	3.715	4.209	11.92	Cys–Ser	1.838	0.528	3.481	1.31
His–Ile	2.322	0.556	4.176	1.766	His–Thr	2.174	0.673	3.23	1.501
Cys–Pro	0.752	0.184	4.087	0.568	Phe–Val	3.704	1.156	3.204	2.548

Bold entries show amino acids that have significant difference.

situation in the mesophil–mesophil homologous and therefore, cannot be a major contributor to the protein thermal adaptation. Table 8 lists 30 replacements with the highest frequency in forward direction. Most of these replacements are in the list of 30 most replacements of mesophile–mesophile homologues. Only Val–Ala, Thr–Ala, Phe–Ile and Asn–Gly contribute an mean of about one net replacement per typical 300-aa protein.

Table 9 shows the amino acid pairs with the largest difference between forward and reverse replacements. From comparative study between mesophile–mesophile homologues some of the replacements from the list were eliminated and only 17 remained show significant ratio between forward and reverse. The high frequency of these substitutions reveals their importance in the thermal adaptation. The observed replacement for individual amino acid and polar, charged and hydrophobic groups are in agreement with the other lists [37,43,29]. These replacements decrease uncharged polar and increase hydrophobic amino acids.

Although hydrogen bonds and polar surfaces have important roles in thermostsbility [24], but some of these hydrogen bonds are compensated by the charged residues and most of the polar residues are replaced with the hydrophobic ones that will increase buried nonpolar surface and core hydrophobicity. In addition, decrease in the number of polar residues is used to avoid deamination and backbone cleavage in proteins. The substitutions in the secondary structures are important factors [37]. In this study, we did not show the position of replacements, but most of replaced amino acids have higher propensities for secondary structures.

10. Polar and nonpolar exposed and buried surface area

Table 10 shows the differences between the polar and nonpolar surface area ratio for the exposed and buried surface of each dataset. Protein surfaces were divided into the buried and the exposed parts and contribution of the polar and nonpolar atoms were calculated for all thermophilic proteins and their homologues and mesophilic homologues of mesophilic proteins. All atoms and side chain atoms were calculated separately. Results show that distribution of exposed polar and nonpolar atoms are the same for homologue proteins, but buried polar and nonpolar atoms show significant differences and suggest that the core of thermophilic proteins are more hydrophobic. This conclusion is supported by the thermophilic and mesophilic homologues comparison that indicate the hydrophobic effect has contributed to the stability at the higher temperature [24,44,45,27,28]. The study of the amino acid composition and replacement showed that hydrophobic amino acids content is increased in thermophilic proteins and

Table 8
Amino acid replacements that are most biased in frequency between mesophile–thermophile and mesophile–mesophile homologues

Replacement	Mesophile–thermophile most frequent replacement				Replacement	Mesophile–mesophile most frequent replacement			
	Forward	Reverse	Ratio	Difference		Forward	Reverse	Ratio	Difference
Ile-Val	21.124	14.321	1.475	6.803	Ile-Val	16.035	12.267	1.307	3.768
Met-Leu	15.635	3.715	4.209	11.92	Met-Leu	14.423	3.766	3.83	10.657
Ile-Leu	15.256	13.049	1.169	2.207	Val-Ile	12.267	16.035	0.765	−3.768
Val-Ile	14.321	21.124	0.678	−6.803	Ile-Leu	11.611	8.401	1.382	3.21
Cys-Val	14.286	0.446	32.031	13.84	Gln-Glu	10.155	5.849	1.736	4.306
Leu-Ile	13.049	15.256	0.855	−2.207	Phe-Tyr	9.579	8.701	1.101	0.878
Cys-Ala	12.782	0.776	16.472	12.006	Arg-Lys	9.297	7.512	1.238	1.785
Gln-Glu	12.704	4.64	2.738	8.064	Tyr-Phe	8.701	9.579	0.908	−0.878
Ser-Ala	12.236	5.547	2.206	6.689	Ser-Ala	8.666	6.169	1.405	2.497
Asp-Glu	11.538	9.637	1.197	1.901	Leu-Ile	8.401	11.611	0.724	−3.21
Tyr-Phe	11.426	11.079	1.031	0.347	Thr-Ser	8.165	7.913	1.032	0.252
Met-Ile	11.401	3.274	3.482	8.127	Trp-Phe	8.084	2.81	2.877	5.274
Phe-Leu	11.079	4.214	2.629	6.865	Ser-Thr	7.913	8.165	0.969	−0.252
Phe-Tyr	11.079	11.426	0.97	−0.347	Glu-Asp	7.88	7.741	1.018	0.139
Arg-Lys	10.962	10.666	1.028	0.296	Asp-Glu	7.741	7.88	0.982	−0.139
Lys-Arg	10.666	10.962	0.973	−0.296	Lys-Arg	7.512	9.297	0.808	−1.785
Ser-Thr	9.673	6.557	1.475	3.116	Asn-Asp	7.078	5.055	1.4	2.023
Glu-Asp	9.637	11.538	0.835	−1.901	Gln-Lys	7.052	3.599	1.959	3.453
Lys-Glu	9.567	7.02	1.363	2.547	Val-Leu	7.001	6.083	1.151	0.918
Met-Val	9.446	2.825	3.344	6.621	Cys-Ala	6.618	0.896	7.386	5.722
Gln-Lys	9.446	3.685	2.563	5.761	Trp-Tyr	6.587	2.078	3.17	4.509
Leu-Val	9.425	8.523	1.106	0.902	Phe-Leu	6.386	3.65	1.75	2.736
Trp-Tyr	9.091	2.411	3.771	6.68	Met-Ile	6.25	4.581	1.364	1.669
Val-Ala	8.622	5.896	1.462	2.726	Lys-Glu	6.182	5.768	1.072	0.414
Val-Leu	8.523	9.425	0.904	−0.902	Ala-Ser	6.169	8.666	0.712	−2.497
Thr-Ala	7.719	3.22	2.397	4.499	Leu-Val	6.083	7.001	0.869	−0.918
Phe-Ile	7.545	2.841	2.656	4.704	Met-Val	6.01	1.541	3.9	4.469
Cys-Thr	7.519	0.683	11.009	6.836	Glu-Gln	5.849	10.155	0.576	−4.306
Asn-Gly	7.059	3.105	2.273	3.954	Glu-Lys	5.768	6.182	0.933	−0.414
Thr-Val	7.036	4.311	1.632	2.725	Cys-Val	5.515	0.642	8.59	4.873

Bold entries show amino acids that have significant difference.

furthermore, the surface accessibility study showed that this increase has occurred in the buried core of thermophilic proteins. The comparative study of the mesophilic dataset and its mesophilic homologues show no significant differences in any of the states. Our results of the exposed and buried area of the side chains atoms are the same as in the case of all atoms.

11. Hydrogen bonds and ion pairs

The hydrogen bonds and ion pairs have been compared between the thermophilic proteins and their mesophilic homologues and also the mesophilic proteins and their mesophilic homologues. Table 11 presents the mean of the total hydrogen bonds, bonds at positions (i, i+2), (i, i+3), (i, i+4), and also parallel and antiparallel beta sheets. The calculated hydrogen bonds in this table are only from the main chain–main chain class. This class of the hydrogen bonds is the main factor in the secondary structure formation. Our results reveal that although the total number of hydrogen bonds in thermophilic proteins is increased, but only the increase at position i, i+4 is significant. In the comparison between the mesophilic proteins and their mesophilic homologues there is not any significant change in the total hydrogen bonds and at each position and this suggests that increase of hydrogen bonds play an important role in the stability of thermophilic proteins. Our detailed analysis of the data revealed that in 63% of the investigated thermophilic protein families total hydrogen bond have increased and in 64% of them, hydrogen bonds of position i, i+4 have increased. Detailed analysis showed that in the 63% and 64% of the examined thermophilic protein families, total hydrogen bonds and bonds at position i, i+4 were increased, respectively. In the analysis of mesophilic proteins and their mesophilic homologues, 48% of protein families show an increase of total hydrogen bonds and 54% show an increase at position i, i+4. These are significantly less than thermophilic and mesophilic homologues while an increase of hydrogen bonds at other positions have the same number as thermophilic protein families.

The role of the hydrogen bonds in the stability of thermophilic proteins is controversial [46,16,47,23,24]. Explanation for the enhanced thermal stability based on the thermophilic and mesophilic studies or mutagenesis is inconsistent, but majority of them show that hydrogen bonds have dominant role in the stability [32]. Our results would also point to the importance of hydrogen bonds. The data from amino acid substitution in this study and also from other studies [48] show that replacements with high rates are always in favor of an increase in the number of amino acids desired for the main chain hydrogen bonds formation in alpha helix structures.

Table 9
Amino acid replacements that are most biased in difference between mesophile–thermophile and mesophile–mesophile homologues

Replacement	Mesophile–thermophile replacement most biased in difference				Replacement	Mesophile–mesophile replacement most biased in difference			
	Forward	Reverse	Ratio	Difference		Forward	Reverse	Ratio	Difference
Cys–Val	14.286	0.446	32.031	13.84	Met–Leu	14.423	3.766	3.83	10.657
Cys–Ala	12.782	0.776	16.472	12.006	Cys–Ala	6.618	0.896	7.386	5.722
Met–Leu	15.635	3.715	4.209	11.92	Trp–Phe	8.084	2.81	2.877	5.274
Met–Ile	11.401	3.274	3.482	8.127	Cys–Val	5.515	0.642	8.59	4.873
Gln–Glu	12.704	4.64	2.738	8.064	Trp–Tyr	6.587	2.078	3.17	4.509
Phe–Leu	11.079	4.214	2.629	6.865	Met–Val	6.01	1.541	3.9	4.469
Cys–Thr	7.519	0.683	11.009	6.836	Gln–Glu	10.155	5.849	1.736	4.306
Ile–Val	21.124	14.321	1.475	6.803	Ile–Val	16.035	12.267	1.307	3.768
Ser–Ala	12.236	5.547	2.206	6.689	Gln–Lys	7.052	3.599	1.959	3.453
Trp–Tyr	9.091	2.411	3.771	6.68	Ile–Leu	11.611	8.401	1.382	3.21
Met–Val	9.446	2.825	3.344	6.621	Cys–Thr	3.676	0.589	6.241	3.087
Gln–Lys	9.446	3.685	2.563	5.761	Pro–Ala	4.6	1.841	2.499	2.759
Cys–Gly	5.639	0.087	64.816	5.552	Phe–Leu	6.386	3.65	1.75	2.736
Cys–Ile	5.639	0.679	8.305	4.96	Gln–Ala	4.09	1.493	2.739	2.597
Trp–Leu	5.455	0.589	9.261	4.866	Phe–Val	3.704	1.156	3.204	2.548
Phe–Ile	7.545	2.841	2.656	4.704	Cys–Leu	2.941	0.406	7.244	2.535
Thr–Ala	7.719	3.22	2.397	4.499	Ser–Ala	8.666	6.169	1.405	2.497
Cys–Leu	4.887	0.408	11.978	4.479	His–Asn	4.15	1.719	2.414	2.431
Trp–Val	4.156	0.149	27.893	4.007	Gln–Asn	5.36	2.932	1.828	2.428
Asn–Gly	7.059	3.105	2.273	3.954	Met–Ala	3.365	0.945	3.561	2.42
Cys–Ser	4.887	1.089	4.488	3.798	His–Leu	2.767	0.406	6.815	2.361
Asn–Glu	5.882	2.142	2.746	3.74	Asn–Ala	3.64	1.443	2.523	2.197
Phe–Val	5.731	2.081	2.754	3.65	Trp–Leu	2.395	0.232	10.323	2.163
Ser–Glu	5.83	2.261	2.579	3.569	Trp–Val	2.395	0.257	9.319	2.138
Gln–Ala	5.863	2.405	2.438	3.458	Cys–Asp	2.206	0.079	27.924	2.127
Trp–Phe	7.013	3.629	1.932	3.384	His–Gly	2.569	0.545	4.714	2.024
Met–Ala	4.397	1.202	3.658	3.195	Asn–Asp	7.078	5.055	1.4	2.023
Ser–Thr	9.673	6.557	1.475	3.116	Met–Gln	3.125	1.128	2.77	1.997
Ser–Gly	6.342	3.279	1.934	3.063	Met–Phe	3.125	1.149	2.72	1.976
His–Arg	4.644	1.59	2.921	3.054	Met–Thr	2.644	0.673	3.929	1.971

Bold entries show amino acids that have significant difference.

Table 12 shows the results obtained from the comparison of ion pairs. Since proteins have different sizes and oligomeric state, therefore the numbers of ion pairs per subunit and per residue were calculated. As shown in Table 12, the thermophilic proteins tend to make more ion pairs than their mesophilic counterpart and the differences for intra-chain ion pairs per

Table 10
Differences between accessible and buried polar and nonpolar atoms in each data set separately for all atoms and side chain atoms of residues

	Mean ± St. dev.			Mean ± St. dev.		
	Thermophile	Homolog mesophile	Paired P ($T \leq t$) two	Mesophile	Homolog mesophile	Paired P ($T \leq t$) two
All atoms						
Frac-pol-exp-area[1]	0.373 ± 0.03	0.372 ± 0.05	0.832	0.374 ± 0.06	0.37 ± 0.04	0.696
Frac-nonpol-exp-area[2]	0.627 ± 0.03	0.628 ± 0.05	0.832	0.626 ± 0.06	0.63 ± 0.04	0.696
Frac-pol-buried-area[3]	0.402 ± 0.02	0.406 ± 0.02	**0.045**	0.405 ± 0.02	0.408 ± 0.02	0.313
Frac-nonpol-buried-area[4]	0.598 ± 0.02	0.594 ± 0.02	**0.045**	0.595 ± 0.02	0.592 ± 0.02	0.313
Frac-exp-tot-area[5]	0.229 ± 0.06	0.236 ± 0.07	0.296	0.237 ± 0.07	0.237 ± 0.06	0.994
Side chain atoms						
Frac-pol-exp-area	0.324 ± 0.03	0.318 ± 0.04	0.279	0.317 ± 0.05	0.314 ± 0.04	0.681
Frac-nonpol-exp-area	0.676 ± 0.03	0.682 ± 0.04	0.279	0.683 ± 0.05	0.686 ± 0.04	0.681
Frac-pol-buried-area	0.161 ± 0.02	0.168 ± 0.02	**0.039**	0.17 ± 0.03	0.174 ± 0.03	0.389
Frac-nonpol-buried-area	0.839 ± 0.02	0.832 ± 0.02	**0.039**	0.83 ± 0.03	0.826 ± 0.03	0.389
Frac-exp-tot-area	0.297 ± 0.07	0.3 ± 0.09	0.719	0.297 ± 0.09	0.3 ± 0.07	0.808

Bold numbers show amino acids that have significant difference.
1—Fraction of total accessible surface area occupied by polar atoms.
2—Fraction of total accessible surface area occupied by nonpolar atoms.
3—Fraction of total buried surface area occupied by polar atoms.
4—Fraction of total buried surface area occupied by nonpolar atoms.
5—Fraction of total accessible surface area in folded state to extended chain.

Table 11
Mean of the total hydrogen bonds and at positions (i, i+2), (i, i+3), (i, i+4) and also parallel and antiparallel beta sheets in each homolog data sets

	Mean±St. dev.			Mean±St. dev		
	Thermophile	Homolog mesophile	Paired P ($T \leq t$) two tail	Mesophile	Homolog mesophile	Paired P ($T \leq t$) two tail
% H bond	72.1±5.3	70.1±6.3	***0.003085***	70±6	69.1±6.6	0.262696
% parallel	7.6±6.7	7.5±6.4	0.635422	7.2±5.9	6.6±5.3	0.102273
% antiparallel	9.3±10.1	9.9±10.8	0.201655	10±9.7	9.5±9.3	0.553877
% H i–i+2	6.4±2.8	6.9±2.2	0.250433	6.9±2.4	6.6±2.3	0.67176
% H i–i+3	10.7±2.9	10.3±3.1	0.388953	10.7±2.9	11±4.2	0.726518
% H i–i+4	32.9±12.4	30.3±11.5	***0.001436***	30.2±11.3	30.5±11.2	0.728509

Bold and italic numbers show amino acids that have significant difference.

residue are significant too. Comparative study of ion pairs for each thermophile–mesophile homologue, showed that in 38 (63%), 3 (5%) and 19 (32%) homologues, number of intra-chain ion pairs per residue increased, remained same and decreased, respectively. Similar comparison between mesophilic proteins and their mesophilic homologues showed no significant difference in the mean number of ion pairs, but from 42 pairs of mesophile–mesophile homologues, 23 (55%), 1 (2%) and 18 (43%) pairs showed increase, no difference and decrease of intra chain ion pairs, respectively. Comparison of inter-chain ion pairs in 14 pairs of oligomer thermophiles and their homologue oligomer mesophiles showed no significant difference (data not shown). Our results also showed that this increase of ion pairs is not due to the increase of charged residues. The bonded charged residues with positive and negative charge show that in thermophilic proteins percentage of charged residues that have involved in ion pair formation are increased.

The role of ion pairs in thermostability is controversial. The majority of experiments have shown that ion pairs will induce thermosatbility and comparison between thermophilic and mesophilic proteins support this idea [49,8,24,50,38,27,28]. On the other hand, some experimental and theoretical studies have indicated that salt bridges usually destabilize the native state of proteins [51,52,53]. The major reason for destabilzation effect of ion pair is that the formation ion pair in the protein core has a desolvation penalty and is energetically unfavorable and is not fully compensated for by interaction of charged residues.

Theoretical studies have indicated that at elevated temperature, energy penalty for desolvation of charged residues in folded protein is reduced [54]. Thus, at high temperature, electrostatic contribution may have stabilizing role in thermophilic proteins. Our results confirm that in the statistical analysis between a dataset of thermophilc and mesophilic homologues, formation of extra ion pairs in thermophilic proteins is significant and this is supported by the observation that no difference between mesophilic and mesophilic homologues dataset are seen. It seems that due to the desolvation penalty of charged residues in folded state, numbers of charged residues are not increased and thermophilic proteins prefer to involve other free charged residues to make ion pairs. Although the mean of ion pairs and hydrogen bonds in thermophilic proteins dataset is more than mesophilic ones, but this is not a general path for all of the thermophilic proteins to achieve the extra thermostability. As mentioned above, only in 63% of thermophilic proteins we saw an increase in ion pairs and hydrogen bonds and in the remained pairs, ion pairs and hydrogen bonds do not change or in some case are even decreased. On the other hand, in 55% and 48% of mesophilic homologues of mesophilic proteins, ion pairs and hydrogen bonds are increased, respectively, but the mean of ion pairs and hydrogen bonds show no significant difference.

Finally, we can conclude that although these interactions have some roles in the thermostability but based on the protein structure, the role and effect of different factors will vary due to

Table 12
Comparison of mean of ion pairs in each homolog data sets

	Mean±St. dev.			Mean±St. dev.		
	Thermophile	Homolog mesophile	Paired P ($T \leq t$) two tail	Mesophile	Homolog mesophile	Paired P ($T \leq t$) two tail
Intra-ion-pair/chain[1]	16.35±11.07	14.5±9.8	0.11	12.52±0.02	12.79±0.024	0.83
Intra-ion-pair/residue[2]	0.053±0.023	0.046±0.019	**0.05**	0.044±3.103	0.039±1.85	0.19
% bond neg. a.a./chain[3]	38.46±15.99	36.13±14.71	0.33	33.34±15.35	32.77±18.86	0.82
% bond pos. a.a./chain[4]	39.35±16.4	34.69±14.58	**0.07**	34.24±16.12	31.35±18.55	0.29
% bond charged a.a./chain[5]	38.13±15.49	34.2±13.57	**0.09**	32.6±14.76	31.67±18.41	0.7

Bold numbers show amino acids that have significant difference.
1—Intrachain ion pairs per chain.
2—Intrachain ion pairs per residue.
3—Percent of residues with negative charge in ion pairs per chain.
4—Percent of residues with positive charge in ion pairs per chain.
5—Percent of charged residues in ion pairs per chain.

the evolutionary pathway that particular family of the protein choose to improve its thermostability.

12. Conclusion

As a general conclusion, our results suggest that the results obtained from computational studies that utilize data from a dataset of thermophilic proteins and their mesophilic homologues as an approach to analyze various factors affecting the thermostability, must be regarded with care. Although these studies show that there is no general strategy for the thermophilic protein thermostability, but by investigating the large data sets, it is possible to show that some factors are generally involved in the thermostability. In other hand, in many of homologous proteins, we can see differences in amino acids that have no effect on protein structure or stability. Hence, to demonstrate that observed differences in thermophile–mesophile datasets are involved in the thermostability, the mesophilic protein dataset must be compared with their mesophilic homologues to eliminate noises or background. This strategy can be used as a model in all of the studies that use the results of comparison between two datasets to extract effective factors on the stability or other features of proteins.

Acknowledgment

Authors are grateful to Tarbiat Modares University and its molecular modeling lab. in biophysics department for financial support and computer facilities.

References

[1] R. Huber, H. Huber, K.O. Stetter, Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties, FEMS Microbiol. Rev. 24 (2000) 615–623.
[2] P. Argos, R.G. Rossman, U.M. Grau, H. Zuber, G. Franck, J.D. Tratschin, Thermal stability and protein structure, Biochemistry 18 (1979) 5698–5703.
[3] R. Jaenick, Protein stability and molecular adaptation to extreme conditions, Eur. J. Biochem. 202 (1991) 725–728.
[4] R.J. Russel, G.L. Taylor, Engineering thermostability: lessons from thermophilic proteins, Curr. Opin. Biotechnol. 6 (1995) 370–374.
[5] R. Jaenick, G. Bohm, The stability of proteins in extreme environments, Curr. Opin. Struct. Biol. 8 (1998) 738–748.
[6] M. Lehmann, L. Pasamontes, S.F. Lassen, M. Wyss, The consensus concept for thermostability engineering of proteins, Biochem. Biophys. Acta 1543 (2000) 408–415.
[7] G.J. Davies, S.J. Gamblin, J.A. Littlechild, H.C. Watson, The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent, Proteins 15 (1993) 283–289.
[8] K.S.P. Yip, T.J. Stillman, K.L. Britton, P.J. Artymiuk, P.J. Baker, S.E. Sedelnikova, P.C. Engel, A. Pasquo, R. Chiaraluce, V. Consalvi, R. Scandurra, D.W. Rice, The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures, Structure 3 (1995) 1147–1158.
[9] D.W. Rice, K.S. Yip, T.J. Stillman, K.L. Britton, A. Fuentes, I. Connerton, A. Pasquo, R. Scandura, P.C. Engel, Insights into the molecular basis of thermal stability from the structure determination of *Pyrococcus furiosus* glutamate dehydrogenase, FEMS Microbiol. Rev. 18 (1996) 105–117.
[10] G.W. Harris, R.W. Pickersgill, I. Connerton, P. Debeire, J.P. Touzel, C. Breton, S. Perez, Structural basis of the properties of an industrially relevant thermophilic xylanase, Proteins 29 (1997) 77–86.
[11] G. Wallon, G. Kryger, S.T. Lovett, T. Oshima, D. Ringe, G.A. Petsko, Crystal structures of *Escherichia coli* and *Salmonella typhimurium* 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from *Thermus thermophilus*, J. Mol. Biol. 266 (1997) 1016–1031.
[12] R.J.M. Russel, J.M.C. Ferguson, D.W. Hough, M.J. Danson, G.L. Taylor, The crystal structure of citrate synthase from the hyperthermophilic Archaeon *Pyrococcus furiosus* at 1.9 Å resolution, Biochemistry 36 (1997) 9983–9994.
[13] B. Matthews, H. Nicholson, W. Becktel, Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding, Proc. Natl. Acad. Sci. U. S. A. 84 (1987) 6663–6667.
[14] H. Nicholson, W.J. Becktel, B.W. Matthews, Enhanced protein stability from designed mutations that interact with helix dipoles, Nature 336 (1988) 651–655.
[15] D.E. Anderson, W.J. Becktel, F.W. Dahlquist, pH-induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme, Biochemistry 29 (1990) 2403–2408.
[16] A.R. Fersh, L. Serrano, Principles of protein stability derived from protein engineering experiments, Curr. Opin. Struct. Biol. 3 (1993) 75–83.
[17] I. Korndorfer, B. Steipe, R. Huber, A. Tomschy, R. Jaenicke, The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution, J. Mol. Biol. 246 (1995) 511–521.
[18] A. Szilagyi, P. zavodszky, Structural basis for the extreme thermostability of D-glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: analysis based on homology modeling, Protein Eng. 8 (1995) 779–789.
[19] B. Van den Burg, G. Vriend, O.R Veltman, G. Venema, V.G.H. Eijsink, Engineering an enzyme to resist boiling, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 2056–2060.
[20] S. Spector, M. Wang, S.A. Carp, J. Robblee, Z.S. Hendsch, R. Fairman, B. Tidor, D.P. Raleigh, Rational modification of protein stability by the mutation of charged surface residues, Biochemistry 39 (2000) 872–879.
[21] G. Bohm, R. Jaenick, Relevance of sequence statistics for the properties of extremophilic proteins, Int. J. Pep. Protein Res. 43 (1994) 97–106.
[22] G.L. Warren, G.A. Petsko, Composition analysis of alpha-helices in thermophilic organisms, Protein Eng. 8 (1995) 905–913.
[23] C.N. Pace, B.A. Shirley, M. McNutt, K. Gajiwala, Forces contributing to the conformational stability of proteins, FASEB J. 10 (1996) 75–83.
[24] G. Vogt, S. Woell, P. Argos, Protein thermal stability, hydrogen bonds, and ion pairs, J. Mol. Biol. 269 (1997) 631–643.
[25] A. Karshikoff, R. Ladenstein, Proteins from thermophilic and mesophilic organisms essentially do not differ in packing, Protein Eng. 11 (1998) 867–872.
[26] A. Szilagyi, P. zavodszky, Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey, Structure 8 (2000) 493–504.
[27] S. Kumar, B. Ma, C.J. Tsai, R. Nussinov, Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers, Proteins 38 (2000) 368–383.
[28] S. Kumar, C.J. Tsai, R. Nussinov, Factors enhancing protein thermostability, Protein Eng. 13 (2000) 179–191.
[29] P.J. Haney, J.H. Badger, G.L. Buldak, C.I. Reich, C.R. Woese, G.J. Olsen, Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 3578–3583.
[30] J. Thompson, D. Eisenberg, Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability, J. Mol. Biol. 290 (1999) 595–604.
[31] S. Chakravarty, R. Varadarajan, Elucidation of determinants of protein thermal stability through genome sequence analysis, FEBS Lett. 470 (2000) 65–69.
[32] E. Querol, J.A. Perez-Pons, A. Mozo-Villarias, Analysis of protein conformational characteristics related to thermostability, Protein Eng. 9 (1996) 265–271.

[33] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanovichi, M. Tasumi, The Protein Data Bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.
[34] S.F. Altschul, T.L. Madden, A.A. Schaffer, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[35] H. Naderi-Manesh, M. Sadeghi, S. Arab, A.A. Moosavi-Movahedi, Prediction of protein surface accessing with information theory, Proteins 42 (2001) 452–459.
[36] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[37] L. Menendez-Arias, P. Argos, Engineering protein thermal stability. Sequence statistics point to residue substitutions in alpha-helices, J. Mol. Biol. 206 (1989) 397–406.
[38] L. Xiao, B. Honig, Electrostatic contributions to the stability of hyperthermophilic proteins, J. Mol. Biol. 289 (1999) 1435–1444.
[39] S.J. Tomazic, A.M. Klibanov, Mechanisms of irreversible thermal inactivation of *Bacillus* alpha–amylases, J. Biol. Chem. 263 (1988) 3086–3091.
[40] O.R. Veltman, G. Vriend, P.J. Middelhoven, B. Van Den Burg, G. Venema, V.G.H. Eijsink, Protein Eng. 9 (1996) 1181–1189.
[41] A. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (1980) 1895–1898.
[42] K.L. Britton, P.J. Baker, K.M. Borges, P.C. Engel, A. Pasquo, D.W. Rice, F.T. Robb, R. Scandurra, T.J. Satillman, K.S. Yip, Insights into thermal stability from a comparison of the glutamate dehydrogenases from *Pyrococcus furiosus* and *Thermococcus litoralis*, Eur. J. Biochem. 229 (1995) 688–695.
[43] R. Scandurra, V. Consalvi, R. Chiaraluce, L. Politi, P.C. Engel, Protein thermostability in extremophiles, Biochimie 80 (1998) 933–941.
[44] C.J. Tsai, S.L. Lin, H.J. Wolfson, R. Nussinov, Studies of protein–protein interfaces: a statistical analysis of the hydrophobic effect, Protein Sci. 6 (1997) 53–64.
[45] G. Vogt, P. Argos, Protein thermal stability: hydrogen bonds or internal packing?, Fold. Des. 2 (1997) S40–S46.
[46] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.
[47] T.E. Creighton, Protein Structures and Molecular Properties, 2nd ed., W.H. Freeman and Company, New York, 1993.
[48] N.T. Marbet, A. Van den broeck, I. Van den brande, P. Stanssens, W.J. Quax, I. Lasters, M. De Maeyer, S.J. Wodak, Arginine residues as stabilizing elements in proteins, Biochemistry 31 (1992) 2239–2253.
[49] A. Tomschy, G. Bohm, R. Jaenicke, The effect of ion pairs on the thermal stability of D-glyceraldehyde 3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritime*, Protein Eng. 7 (1994) 1471–1478.
[50] C. Vetriani, D.L. Maeder, N. Tolliday, K.S.P. Yip, T.J. Stillman, K.L. Britton, D.W. Rice, H.H. Klump, F.T. Robb, Protein thermostability above 100 degrees C: a key role for ionic interactions, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 12300–12305.
[51] Z.S. Hendsch, B. Tidor, Do salt bridges stabilize proteins? A continuum electrostatic analysis, Protein Sci. 3 (1994) 211–226.
[52] C.D. Waldburger, J.F. Schildbach, R.T. Sauer, Are buried salt bridges important for protein stability and conformational specificity, Nat. Struct. Biol. 2 (1995) 122–128.
[53] F.B. Sheinerman, R. Norel, B. Honig, Electrostatic aspects of protein–protein interactions, Curr. Opin. Struct. Biol. 10 (2000) 153–159.
[54] A.H. Elcock, The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins, J. Mol. Biol. 284 (1998) 489–502.