ELSEVIER

Biophysical Chemistry 82 (1999) 51–67

Biophysical Chemistry

www.elsevier.nl/locate/bpc

Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins

M. Michael Gromiha, Motohisa Oobatake[1], Akinori Sarai*

Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Received 7 July 1999; received in revised form 8 September 1999; accepted 8 September 1999

Abstract

Understanding the role of various interactions in enhancing the thermostability of proteins is important not only for clarifying the mechanism of protein stability but also for designing stable proteins. In this work, we have analyzed the thermostability of 16 different families by comparing mesophilic and thermophilic proteins with 48 various physicochemical, energetic and conformational properties. We found that the increase in shape, s (location of branch point in side chain) increases the thermostability, whereas, an opposite trend is observed for Gibbs free energy change of hydration for native proteins, G_{hN}, in 14 families. A good correlation is observed between these two properties and the simultaneous increases of $-G_{hN}$ and s is necessary to enhance the thermostability from mesophile to thermophile. The increase in shape, which tends to increase with increasing number of carbon atoms both for polar and non-polar residues, may generate more packing and compactness, and the position of β and higher order branches may be important for better packing. On the other hand, the increase in $-G_{hN}$ in thermophilic proteins increases the solubility of the proteins. This tendency counterbalances the increases in insolubility and unfolding heat capacity change due to the increase in the number of carbon atoms. Thus, the present results suggest that the stability of thermophilic proteins may be achieved by a balance between better packing and solubility. © 1999 Elsevier Science B.V. All rights reserved.

[1] Present address: Faculty of Science and Technology, Meijo University, 1-501 Shiogamaguchi, Tenpaku-ku, Nagoya 468-8502, Japan.

* Corresponding author. Tel.: +81-298-36-9082; fax: +81-298-36-9080.
E-mail address: sarai@rtc.riken.go.jp (A. Sarai)

0301-4622/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0301-4622(99)00103-9

1. Introduction

The major intramolecular interactions, namely, hydrophobic, electrostatic, van der Waals and hydrogen bonds play an important role in the stability of protein structures [1–4]. Several investigations have been carried out to understand the mechanism for the thermostability of proteins. Argos et al. [5] observed that Gly, Ser, Lys and Asp in mesophiles are generally substituted by Ala, Thr, Arg and Glu, respectively, in thermophiles to enhance the stability. Ponnuswamy et al. [6] found a correlation between melting temperature and amino acid composition of a stabilizing and destabilizing group of amino acids and reported an empirical relation between thermostability and amino acid content of proteins. Recently, it has been shown that the thermostability can be calculated from amino acid contents of proteins [7].

Imanaka et al. [8] proposed a new approach to enhance the thermostability of enzymes by single amino acid substitutions based on increase in hydrophobicity and stabilization of helices. Menendez-Arias and Argos [9] reported that the decrease in flexibility and increase in hydrophobicity in α-helical regions are the major factors to enhance the thermostability of proteins. The importance of disulfide bonds, ion binding sites and salt bridges along with hydrophobic and hydrogen bonds, was stressed by Fontana [10]. Querol et al. [11] analyzed the conformational characteristics of proteins related to thermostability.

Recently, Argos et al. [12,13] examined the thermostability of 16 different families of mesophilic and thermophilic proteins and found a good correlation between the thermostability of the familial members and an increase in the number of hydrogen bonds as well as an increase in the fractional polar surface. Our recent studies on protein stability upon mutations show the importance of some specific amino-acid properties to the stability [14,15]. In this work, we analyzed the relation between the increase in thermostability and various physico–chemical, energetic and conformational properties for the meso and thermophilic proteins. We found that the properties, Gibbs free energy of hydration for native protein and shape (position of branch point in a side chain) play a dominant role in enhancing thermostability. We discuss the comparison of these properties and their physical meanings in relation to the stability of thermophilic proteins.

2. Materials and methods

2.1. Database

Recently, Vogt et al. [13] collected a set of 56 globular proteins from 16 different families to analyze the relation between thermostability and other interactions. We used the same set of data in our present analysis. The PDB codes for all the proteins along with the average environmental temperatures (T_{env}) are given in Table 1. The temperatures were assigned to each of the proteins according to those of optimal growth or normal living environment for the species involved or those from in vitro experiments for half-life stability as stated in the literature [9,16–20]. We observed a direct relationship between the average environmental temperature (T_{env}) and melting temperature (T_{m}) of proteins in each family and the correlation coefficient is 0.91; the corresponding regression equation is $T_{\mathrm{m}} = 24.4 + 0.93T_{\mathrm{env}}$. As the number of samples for T_{env} is more than T_{m}, and there is a strong correlation between T_{env} and T_{m}, we have used T_{env} for the present study. In Table 1, we also include the sequence identity (percentage of same amino acid residues) and rms deviation for C_{α} atoms between the two proteins of highest and lowest average environmental temperatures in each family. The amino acid sequence and three-dimensional structures were taken from the recent release of Protein Data Bank from the Brookhaven National Laboratory [21]. We used the 'bestfit' module available in the package GCG 8.1 (Genetics Computer Group, Inc., Madison, USA) to estimate the sequence identity and Sybyl 6.4 (Tripos Associates, St. Louis, MO, USA) to compute the rms deviation. The computed average sequence identity and rms deviation for the 16 families are, respectively, 47.1% and 3.0 Å,

Table 1
List of protein families used in the present study[a]

Family		PDB entry ID	T_{env} (°C)	% Secondary structure				Sequence identity (%)	rms deviation (Å)
No.	Name			Helix	Strand	Turn	Coil		
1	Malate dehydrogenase	4MDH	37.0	43	19	19	19		
		1BMD	72.5	46	19	19	17	54.8	2.24
2	Glycosyltransferase A (α-amylase)	1CDG	35.0	23	30	22	24		
		1CGT	35.0	23	31	24	23		
		1CYG	52.5	23	29	23	26		
		1CIU	60.0	21	30	25	24	68.0	3.52
3	Glyceraldehyde-3-phosphate dehydrogenase	4GPD	20.0	29	21	25	26		
		1GAD	37.0	29	28	25	18		
		3GPD	37.0	27	22	25	26		
		1GD1	52.5	29	30	24	16		
		1CER	71.0	30	27	24	19		
		1HDG	82.5	31	27	23	18	44.7	2.33
4	Lactate dehydrogenase	6LDH	20.0	44	17	18	21		
		1LLC	35.0	26	14	38	23		
		5LDH	37.0	39	10	26	25		
		9LDB	37.0	44	20	18	18		
		1LLD	39.0	45	19	18	18		
		1LDN	52.5	46	22	19	13	35.1	3.48
5	Thermolysin	1NPC	30.0	42	20	21	18		
		1LNF	52.5	41	17	24	17	73.0	0.91
6	Ribonuclease H	2RN2	37.0	35	30	19	15		
		1RIL	72.5	38	24	18	20	56.7	1.56
7	Subtilisin	1ST3	30.0	30	20	26	23		
		1SUP	35.0	30	20	24	25		
		1SCA	42.5	30	19	26	25		
		1THM	60.0	29	20	28	23	40.5	3.37
8	Ferredoxin	1FCA	28.0	13	18	27	42		
		1FDX	37.0	15	11	28	46		
		2FXB	52.5	20	17	32	31	30.9	6.58
9	Superoxide dismutase	3SDP	27.5	26	3	43	27		
		1ABM	37.0	60	13	14	13		
		1IDS	37.0	54	14	20	13		
		1ISA	37.0	51	12	22	15		
		3MDS	72.5	57	10	18	15	41.9	3.93
10	Phosphofructokinase	2PFK	37.0	46	20	17	17		
		3PFK	52.5	46	19	16	19	55.3	0.88
11	Phosphoglycerate kinase	3PGK	27.5	34	12	27	27		
		1PHP	52.5	42	17	21	20	41.4	5.48
12	Triose phosphate isomerase	1YPI	27.5	43	17	18	21		
		1HTI	37.0	42	16	21	21		
		1TIM	37.0	46	18	15	21		
		1TPE	41.0	45	16	17	23		
		1BTM	52.5	49	16	16	19	34.0	2.61
13	Rubredoxin	1RDG	35.5	17	23	29	31		
		6RXN	35.5	22	22	24	31		
		8RXN	35.5	17	23	29	31		
		5RXN	37.0	17	22	28	33		
		1CAA	110.0	17	26	30	26	66.0	0.63

Table 1 (*Continued*)

Family		PDB entry ID	T_{env} (°C)	% Secondary structure				Sequence identity (%)	rms deviation (Å)
No.	Name			Helix	Strand	Turn	Coil		
14	Hydrolase	1INO	37.0	18	33	27	22		
		2PRD	72.5	25	34	22	18	34.5	2.97
15	Glycosyltransferase B	2EXO	30.0	45	21	21	14		
	(β-glycanase)	1XYZ	60.0	45	20	21	14	37.0	2.61
16	Reductase	1LPF	27.5	35	29	19	17		
		1LVL	27.5	35	24	19	22		
		3LAD	37.0	34	26	21	20		
		1EBD	52.5	34	29	22	16	40.0	4.86

[a] T_{env}, average environmental temperature, taken from Vogt et al. [13]. Sequence identity is the percentage of same amino acid residues and rms deviation is the root mean square deviation for C_α atoms between the two proteins of highest and lowest average environmental temperatures in each family.

indicating high sequence identity and similar tertiary structures of proteins in each family. The DSSP algorithm [22] was used for the assignment of secondary structures and the percentage of secondary structural content for all the 56 proteins are given in Table 1.

2.2. Amino acid properties

It has been shown that there are specific cooperativities among amino acid residues, which due to similarities in their various physical, chemical, energetic and conformational properties, enable them to preserve their specific, preferred environments and spatial positions in the folded conformation of proteins [23]. We considered a set of 48 diverse amino acid properties, which fall into various clusters analyzed by Tomii and Kanehisa [24]. The numerical values of these properties are given in Table 2. Brief descriptions of these properties are available in our previous articles [7,25].

2.3. Computational procedure

The amino acid composition and the average property value for all the proteins were computed. For a given property, $p(i)$, the average value (P) was computed using the equation

$$P = \frac{\sum_{i=1}^{20} p(i) \cdot n(i)}{N} \tag{1}$$

where $p(i)$ and $n(i)$ are, respectively, the property value of the ith amino acid residue and the number of amino acids of ith type in a protein. N is the total number of residues in a protein.

We followed the method of Vogt et al. [13] to compute the score (S) for a specific property. The overall family trend for a specific property was found by summing the property changes over each unique family pair, weighted by the square of the difference of adapted living temperatures for the two compared members. The property differences were each divided by the associated living temperature differences for each pair in order to standardize the characteristics. The score (S_j) for a particular property per 10°C rise in thermostability for a given family j is given by:

$$S_j = \frac{\sum_{i=2}^{n} \sum_{k=1}^{i-1} (T_i - T_k)^2 \left[\frac{P_i - P_k}{T_i - T_k} \right] 10}{\sum_{i=2}^{n} \sum_{k=1}^{i-1} (T_i - T_k)^2} \tag{2}$$

where n is the number of constituents in family j. P_i and T_i are, respectively, property value and average environmental temperature (given in Table 1) of ith family members.

Table 2
Numerical values of 48 selected physico–chemical, energetic and conformational properties of the 20 amino acids/residues[a]

No	Property	Ala	Asp	Cys	Glu	Phe	Gly	His	Ile	Lys	Leu
1	K^0	−25.50	−33.12	−32.82	−36.17	−34.54	−27.00	−31.84	−31.78	−32.40	−31.78
2	H_t	0.87	0.66	1.52	0.67	2.87	0.10	0.87	3.15	1.64	2.17
3	Hp	13.05	11.10	14.30	11.41	13.89	12.20	12.42	15.34	11.01	14.19
4	P	0.00	49.70	1.48	49.90	0.35	0.00	51.60	0.10	49.50	0.13
5	pH_i	6.00	2.77	5.05	5.22	5.48	5.97	7.59	6.02	9.74	5.98
6	pK'	2.34	2.01	1.65	2.19	1.89	2.34	1.82	1.36	2.18	2.36
7	M_w	89.00	133.00	121.00	147.00	165.00	75.00	155.00	131.00	146.00	131.00
8	B_l	11.50	11.68	13.46	13.57	19.80	3.40	13.67	21.40	15.71	21.40
9	R_f	9.90	2.80	2.80	3.20	18.80	5.60	8.20	17.10	3.50	17.60
10	μ	14.34	12.00	35.77	17.26	29.40	0.00	21.81	19.06	21.29	18.78
11	H_{nc}	0.62	0.90	0.29	−0.74	1.19	0.48	−0.40	1.38	−1.50	1.06
12	E_{sm}	1.40	1.16	1.37	1.16	1.14	1.36	1.22	1.19	1.07	1.32
13	E_l	0.49	0.35	0.67	0.37	0.72	0.53	0.54	0.76	0.30	0.65
14	E_t	1.90	1.52	2.04	1.54	1.86	1.90	1.76	1.95	1.37	1.97
15	P_α	1.42	1.01	0.70	1.51	1.13	0.57	1.00	1.08	1.16	1.21
16	P_β	0.83	0.54	1.19	0.37	1.38	0.75	0.87	1.60	0.74	1.30
17	P_t	0.66	1.46	1.19	0.74	0.60	1.56	0.95	0.47	1.01	0.59
18	P_c	0.71	1.21	1.19	0.84	0.71	1.52	1.07	0.66	0.99	0.69
19	C_a	20.00	26.00	25.00	33.00	46.00	13.00	37.00	39.00	46.00	35.00
20	F	0.96	1.14	0.87	1.07	0.69	1.16	0.80	0.76	1.14	0.79
21	B_r	0.38	0.14	0.57	0.09	0.51	0.38	0.31	0.56	0.04	0.50
22	R_a	3.70	2.60	3.03	3.30	6.60	3.13	3.57	7.69	1.79	5.88
23	N_s	6.05	4.95	7.86	5.10	6.62	6.16	5.80	7.51	4.88	7.37
24	α_n	1.59	0.53	0.33	1.45	1.14	0.53	0.89	1.22	1.13	1.91
25	α_c	1.44	2.13	0.76	2.01	1.01	0.62	0.56	0.68	0.59	0.58
26	α_m	1.22	0.56	1.53	1.28	1.13	0.40	2.23	0.77	1.65	1.05
27	V^0	60.46	73.83	67.70	85.88	121.48	43.25	98.79	107.72	108.50	107.75
28	N_m	2.11	1.80	1.88	2.09	1.98	1.53	1.98	1.77	1.96	2.19
29	N_l	3.92	2.85	5.55	2.72	4.53	4.31	3.77	5.58	2.79	4.59
30	H_{gm}	13.85	11.61	15.37	11.38	13.93	13.34	13.82	15.28	11.58	14.13
31	ASA_D	104.00	132.20	132.50	161.90	182.00	73.40	165.80	171.50	195.20	161.40
32	ASA_N	33.20	62.40	17.90	81.00	33.10	29.20	57.70	28.30	107.50	31.10
33	ΔASA	70.90	69.60	114.30	80.50	148.40	44.00	107.90	142.70	87.50	129.80
34	ΔG_h	−0.54	−2.97	−1.64	−3.71	−1.06	−0.59	−3.38	0.32	−2.19	0.27
35	G_{hD}	−0.58	−6.10	−1.91	−7.37	−1.35	−0.82	−5.57	0.40	−5.97	0.35
36	G_{hN}	−0.06	−3.11	−0.27	−3.62	−0.28	−0.23	−2.18	0.07	−1.70	0.07
37	ΔH_h	−2.24	−4.54	−3.43	−5.63	−5.11	−1.46	−6.83	−3.84	−5.02	−3.52
38	$-T\Delta S_h$	1.70	1.57	1.79	1.92	4.05	0.87	3.45	4.16	2.83	3.79
39	ΔC_{ph}	14.22	2.73	9.41	3.17	39.06	4.88	20.05	41.98	17.68	38.26
40	ΔG_c	0.51	2.89	2.71	3.58	3.22	0.68	3.95	−0.40	1.87	−0.35

Table 2 (*Continued*)

41	ΔH_c	2.77	4.72	8.64	5.69	11.93	1.23	7.64	4.03	3.57	3.69
42	$-T\Delta S_c$	−2.25	−1.83	−5.92	−2.11	−8.71	−0.55	−3.69	−4.42	−1.70	−4.04
43	ΔG	−0.02	−0.08	1.08	−0.13	2.16	0.09	0.56	−0.08	−0.32	−0.08
44	ΔH	0.51	0.18	5.21	0.05	6.82	−0.23	0.79	0.19	−1.45	0.17
45	$-T\Delta S$	−0.54	−0.26	−4.14	−0.19	−4.66	0.31	−0.23	−0.27	1.13	−0.24
46	v	1.00	4.00	2.00	5.00	7.00	0.00	6.00	4.00	5.00	4.00
47	s	0.00	2.00	0.00	3.00	2.00	0.00	2.00	1.00	0.00	2.00
48	f	0.00	2.00	1.00	3.00	2.00	0.00	2.00	2.00	4.00	2.00

No	Property	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
1	K^0	−31.18	−30.90	−23.25	−32.60	−26.62	−29.88	−31.23	−30.62	−30.24	−35.01
2	H_t	1.67	0.09	2.77	0.00	0.85	0.07	0.07	1.87	3.77	2.67
3	$H\text{p}$	13.62	11.72	11.06	11.78	12.40	11.68	12.12	14.73	13.96	13.57
4	P	1.43	3.38	1.58	3.53	52.00	1.67	1.66	0.13	2.10	1.61
5	$\text{p}H_i$	5.74	5.41	6.30	5.65	10.76	5.68	5.66	5.96	5.89	5.66
6	$\text{p}K'$	2.28	2.02	1.99	2.17	1.81	2.21	2.10	2.32	2.38	2.20
7	M_w	149.00	132.00	115.00	146.00	174.00	105.00	119.00	117.00	204.00	181.00
8	B_l	16.25	12.82	17.43	14.45	14.28	9.47	15.77	21.57	21.61	18.03
9	R_f	14.70	5.40	14.80	9.00	4.60	6.90	9.50	14.30	17.00	15.00
10	μ	21.64	13.28	10.93	17.56	26.66	6.35	11.01	13.92	42.53	31.55
11	H_{nc}	0.64	−0.78	0.12	−0.85	−2.53	−0.18	−0.05	1.08	0.81	0.26
12	E_{sm}	1.30	1.18	1.24	1.12	0.92	1.30	1.25	1.25	1.03	1.03
13	E_l	0.65	0.38	0.46	0.40	0.55	0.45	0.52	0.73	0.83	0.65
14	E_t	1.96	1.56	1.70	1.52	1.48	1.75	1.77	1.98	1.87	1.69
15	P_α	1.45	0.67	0.57	1.11	0.98	0.77	0.83	1.06	1.08	0.69
16	P_β	1.05	0.89	0.55	1.10	0.93	0.75	1.19	1.70	1.37	1.47
17	P_t	0.60	1.56	1.52	0.98	0.95	1.43	0.96	0.50	0.96	1.14
18	P_c	0.59	1.37	1.61	0.87	1.07	1.34	1.08	0.63	0.76	1.07
19	C_a	43.00	28.00	22.00	36.00	55.00	20.00	28.00	33.00	61.00	46.00
20	F	0.78	1.04	1.16	1.07	1.05	1.13	0.96	0.79	0.77	1.01
21	B_r	0.42	0.15	0.18	0.11	0.07	0.23	0.23	0.48	0.40	0.26
22	R_a	5.21	2.12	2.12	2.70	2.53	2.43	2.60	7.14	6.25	3.03
23	N_s	6.39	5.04	5.65	5.45	5.70	5.53	5.81	7.62	6.98	6.73
24	α_n	1.25	0.53	0.00	0.98	0.67	0.70	0.75	1.42	1.33	0.58
25	α_c	0.73	0.93	2.19	1.20	0.39	0.81	1.25	0.63	1.40	0.72
26	α_m	1.47	0.93	0.00	1.63	1.59	0.87	0.46	1.20	0.46	0.52
27	V^0	105.35	78.01	82.83	93.90	127.34	60.62	76.83	90.78	143.91	123.60
28	N_m	2.27	1.84	1.32	2.03	1.94	1.57	1.57	1.63	1.90	1.67
29	N_l	4.14	3.64	3.57	3.06	3.78	3.75	4.09	5.43	4.83	4.93
30	H_{gm}	13.86	13.02	12.35	12.61	13.10	13.39	12.70	14.56	15.48	13.88
31	ASA_D	189.80	134.90	135.10	164.90	210.20	111.40	130.40	143.90	208.80	196.40
32	ASA_N	41.30	60.50	60.70	71.50	94.50	48.70	52.00	28.10	39.50	50.40

Table 2 (*Continued*)

33	ΔASA	147.90	74.00	73.50	93.30	116.00	62.80	78.00	115.60	167.80	145.90
34	ΔG_h	−0.60	−3.55	0.32	−3.92	−5.96	−3.82	−1.97	0.13	−3.80	−5.64
35	G_{hD}	−0.71	−6.63	0.56	−7.12	−12.78	−6.18	−3.66	0.18	−4.71	−8.45
36	G_{hN}	−0.10	−3.03	0.23	−3.15	−6.85	−2.36	−1.69	0.04	−0.88	−2.82
37	ΔH_h	−4.16	−5.68	−1.95	−6.23	−10.43	−5.94	−4.39	−3.15	−8.99	−10.67
38	$-T\Delta S_h$	3.56	2.13	2.27	2.31	4.47	2.12	2.42	3.28	5.19	5.03
39	ΔC_{ph}	31.67	3.91	23.69	3.74	16.66	6.14	16.11	32.58	37.69	30.54
40	ΔG_c	1.13	3.26	−0.39	3.69	5.25	3.42	1.74	−0.19	5.59	6.56
41	ΔH_c	7.06	3.64	1.97	4.47	6.03	5.80	4.42	3.45	13.46	14.41
42	$-T\Delta S_c$	−5.93	−0.39	−2.36	−0.78	−0.78	−2.38	−2.68	−3.64	−7.87	−7.95
43	ΔG	0.53	−0.30	−0.06	−0.23	−0.71	−0.40	−0.24	−0.06	1.78	0.91
44	ΔH	2.89	−2.03	0.02	−1.76	−4.40	−0.16	0.04	0.30	4.47	3.73
45	$-T\Delta S$	−2.36	1.74	−0.08	1.53	3.69	−0.24	−0.28	−0.36	−2.69	−2.82
46	v	4.00	4.00	3.00	5.00	7.00	2.00	3.00	3.00	10.00	8.00
47	s	0.00	2.00	0.00	3.00	5.00	0.00	1.00	1.00	2.00	2.00
48	f	3.00	2.00	0.00	3.00	5.00	1.00	1.00	1.00	2.00	2.00

[a]*Abbreviations:* K^0, compressibility; H_t, thermodynamic transfer hydrophobicity; H_p, surrounding hydrophobicity; P, polarity; pH_i, isoelectric point; pK', equilibrium constant with reference to the ionization property of COOH group; M_w, molecular weight; B_l, bulkiness; R_f, chromatographic index; μ, refractive index; H_{nc}, normalized consensus hydrophobicity; E_{sm}, short and medium range non-bonded energy; E_l, long-range non-bonded energy; E_t, total non-bonded energy ($E_{sm} + E_l$); P_α, P_β, P_t, and P_c are, respectively, α-helical, β-structure, turn and coil tendencies; C_a, helical contact area; F, mean rms fluctuational displacement; B_r, buriedness; R_a, solvent accessible reduction ratio; N_s, average number of surrounding residues; α_n, α_c and α_m are, respectively, power to be at the N-terminal, C-terminal and middle of α-helix; V^0, partial-specific volume; N_m and N_l are, respectively, average medium and long-range contacts; H_{gm}, combined surrounding hydrophobicity (globular and membrane); ASA_D, ASA_N and ΔASA are, respectively, solvent accessible surface area for denatured, native and unfolding; ΔG_h, G_{hD} and G_{hN} are, respectively, Gibbs free energy change of hydration for unfolding, denatured and native protein; ΔH_h, unfolding enthalpy change of hydration; $-T\Delta S_h$, unfolding entropy change of hydration; ΔC_{ph}, unfolding hydration heat capacity change; ΔG_c, ΔH_c and $-T\Delta S_c$ are, respectively, unfolding Gibbs free energy, unfolding enthalpy and unfolding entropy changes of chain; ΔG, ΔH and $-T\Delta S$ are, respectively, unfolding Gibbs free energy change, unfolding enthalpy change and unfolding entropy change; v, volume (number of non-hydrogen side chain atoms); s, shape (position of branch point in a side-chain); f, flexibility (number of side-chain dihedral angles). *Notes:* K^0 and V^0 [34]; H_t [35]; H_p [33]; P, pH_i, pK', B_l and R_f [36]; M_w and μ [37]; H_{nc} [38]; E_{sm}, E_l and E_t, [39]; P_α, P_β, P_t, $P_{c,}$ α_n, α_c and α_m [40]; C_a [41]; F [42]; B_r [43]; R_a [44]; N_s [45]; N_m and N_l [27]; H_{gm} [46]; ASA_D, ASA_N, ΔASA, ΔG_h, G_{hD}, G_{hN}, ΔH_h, $-T\Delta S_h$, ΔC_{ph}, ΔG_c, ΔH_c, $-T\Delta S_c$, ΔG, ΔH and $-T\Delta S$ [7]; v, s and f [26]. K^0 in m^3/mol/Pa ($\times 10^{-15}$); H_t, H_p, H_{nc}, H_{gm}, ΔG_h, G_{hD}, G_{hN}, ΔH_h, $-T\Delta S_h$, ΔG_c, ΔH_c, $-T\Delta S_c$, ΔG, ΔH and $-T\Delta S$ in kcal/mol; P in Debye; pH_i and pK' in pH units; E_{sm}, E_l and E_t in kcal/mol/atom; B_l, C_a, ASA_D, ASA_N and ΔASA in Å^2; F in Å; V^0 in m^3/mol ($\times 10^{-6}$); ΔC_{ph}, in cal/mol/K and the rest are dimensionless quantities.

Table 3
Number of families increasing/decreasing thermostability for 24 selected properties

Number of families	Property
Increased	
10	H_t, P, pH_i, M_w, B_l, μ, C_a, R_a, N_s, α_c, V^0, ASA_D, ΔASA, -$T\Delta S_h$, ΔC_{ph}, v
11	R_f, E_l, P_β, N_l
14	s
Decreased	
10	α_m, ΔH_h
14	G_{hN}

2.4. Computation of preferred residue exchanges from meso to thermophilic proteins

We estimated the preference of residues replaced from meso to thermophilic proteins by aligning their sequences in all 16 families. The proteins having the highest and lowest average environmental temperature have been selected in each family and these sequences have been aligned using the bestfit module of GCG 8.1. We selected all the 380 possible mutations and the total number of occurrences for each type of mutations has been computed. This preference of residue exchanges was normalized by dividing the total number of each type of the 20 amino acid residues present in mesophilic proteins. The preference between two residue pairs (e.g. Ala–Asp and Asp–Ala) was deduced by extracting the difference between these pairs.

3. Results and discussion

3.1. Amino acid properties that increase thermostability

Considering each of the 48 amino acid properties, we computed the score (S), using Eq. (2), for all the 16 families given in Table 1. A careful survey of the results shows that some families increase the average property value with increase in thermostability, whereas, some families decrease. The overall behavior of the 16 considered families for 24 selected amino acid properties is given in Table 3.

From this table, it is interesting to note that the shape, s (position of branch point in a side chain; e.g. 0 for Ala, 1 for Ile, 2 for Leu, 0 for Met and 2 for Phe, etc.) increases with an increase in thermostability for 14 out of the 16 families (87.5%). This result shows that the increase in β- ($s = 1$) or γ- ($s = 2$) branched amino acid residues enhances the thermostability. Surprisingly, s is one of the best properties to predict the activity and stability of mutant proteins in buried regions [26]. The two families with decreased thermostability by increased s are phosphofructokinase and reductase, both of which are oligomers with more than 300 residues per chain.

We observed an opposite trend for Gibbs free energy change of hydration for native protein (G_{hN}). G_{hN} decreased with increased thermostability in the same 14 families. The decrease in G_{hN} theoretically increases the stability by increasing the exposure of polar atoms for native proteins. This observation is consistent with the result obtained by Vogt et al. [13] that the net changes in accessible surface area of polar-atoms increase with an increase in stability. Note that our observation showed this behavior for 14 families while it was observed in 13 families by Vogt et al. [13].

Table 3 shows that the other properties, chromatographic index (R_f), long-range non-bonded energy (E_l), β-strand tendency (P_β) and average long-range contacts (N_l) increase with an increase in thermostability for more than two-thirds of the families tested. Interestingly, the properties, E_l, P_β and N_l are responsible for the long-range interactions and these interactions are important to stabilize proteins [27–30]. Furthermore, the set of properties R_f, E_l and P_β is

Table 4
Difference in preference of residue substitutions from mesophile to thermophile in all 16 families[a]

	Ala	Asp	Cys	Glu	Phe	Gly	His	Ile	Lys	Leu	Met
Ala	0.0	0.5	−10.1	−1.2	−3.9	0.1	−6.6	−1.8	−1.8	−3.5	−10.4
Asp	−0.5	0.0	0.0	−1.0	−1.2	2.7	−2.0	−0.3	−4.3	3.2	0.7
Cys	**10.1**	0.0	0.0	1.8	1.1	6.8	0.7	**7.6**	1.8	3.5	−1.2
Glu	1.2	1.0	−1.8	0.0	0.5	−1.0	−5.3	1.9	−3.7	1.8	−1.2
Phe	3.9	1.2	−1.1	−0.5	0.0	1.2	−0.4	5.4	2.1	6.9	1.2
Gly	−0.1	−2.7	−6.8	1.0	−1.2	0.0	−1.6	−0.1	−1.9	−0.9	0.2
His	6.6	2.0	−0.7	5.3	0.4	1.6	0.0	0.9	3.5	1.8	−4.0
Ile	1.8	0.3	−7.6	−1.9	−5.4	0.1	−0.9	0.0	−1.2	−3.3	−4.7
Lys	1.8	4.3	−1.8	3.7	−2.1	1.9	−3.5	1.2	0.0	0.7	−4.0
Leu	3.5	−3.2	−3.5	−1.8	−6.9	0.9	−1.8	3.3	−0.7	0.0	−8.6
Met	**10.4**	−0.7	1.2	1.2	−1.2	−0.2	4.0	4.7	4.0	**8.6**	0.0
Asn	−1.0	1.2	−4.8	−0.5	0.7	4.1	−1.2	1.0	3.8	0.9	−2.5
Pro	6.4	0.2	0.0	3.4	1.1	3.4	−2.5	0.0	−2.7	5.2	−0.7
Gln	5.0	0.7	−1.0	2.6	1.7	3.0	−2.3	0.4	3.1	2.4	0.2
Arg	2.5	0.3	0.7	−1.7	0.1	−0.8	−1.1	0.4	−2.7	2.7	0.1
Ser	1.1	−0.5	−3.9	4.0	−1.2	1.9	−1.0	1.0	1.7	2.1	1.4
Thr	1.6	0.7	−4.9	−1.6	−1.6	0.8	0.5	0.3	2.3	−0.8	−2.1
Val	2.2	−1.0	−8.3	0.1	−2.7	2.4	0.5	−4.1	−0.5	−4.0	−6.3
Trp	3.7	1.6	0.0	0.0	3.5	0.7	0.0	−1.2	0.0	2.9	3.2
Tyr	0.7	0.3	−3.5	−0.4	5.7	1.5	1.2	0.0	2.0	0.9	−1.2
Gap	21.9	0.0	−1.8	−5.2	2.7	−0.2	−1.0	7.9	2.5	19.7	4.5

	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr	Gap
Ala	1.0	−6.4	−5.0	−2.5	−1.1	−1.6	−2.2	−3.7	−0.7	−21.9
Asp	−1.2	−0.2	−0.7	−0.3	0.5	−0.7	1.0	−1.6	−0.3	0.0
Cys	4.8	0.0	1.0	−0.7	3.9	4.9	**8.3**	0.0	3.5	1.8
Glu	0.5	−3.4	−2.6	1.7	−4.0	1.6	−0.1	0.0	0.4	5.2
Phe	−0.7	−1.1	−1.7	−0.1	1.2	1.6	2.7	−3.5	−5.7	−2.7
Gly	−4.1	−3.4	−3.0	0.8	−1.9	−0.8	−2.4	−0.7	−1.5	0.2
His	1.2	2.5	2.3	1.1	1.0	−0.5	−0.5	0.0	−1.2	1.0
Ile	−1.0	0.0	−0.4	−0.4	−1.0	−0.3	4.1	1.2	0.0	−7.9
Lys	−3.8	2.7	−3.1	2.7	−1.7	−2.3	0.5	0.0	−2.0	−2.5
Leu	−0.9	−5.2	−2.4	−2.7	−2.1	0.8	4.0	−2.9	−0.9	−19.7
Met	2.5	0.7	−0.2	−0.1	−1.4	2.1	6.3	−3.2	1.2	−4.5
Asn	0.0	−0.5	−4.0	−3.1	−1.0	3.7	2.0	0.4	0.8	−4.1
Pro	0.5	0.0	−2.0	−1.5	−0.1	−0.1	0.7	−1.6	0.0	−8.5
Gln	4.0	2.0	0.0	3.6	−1.2	−0.3	2.2	−1.6	1.6	0.8
Arg	3.1	1.5	−3.6	0.0	0.0	−0.5	2.6	−2.5	−0.7	0.7
Ser	1.0	0.1	1.2	0.0	0.0	4.3	1.2	−1.2	3.1	0.7
Thr	−3.7	0.1	0.3	0.5	−4.3	0.0	2.8	−4.0	1.6	−7.6
Val	−2.0	−0.7	−2.2	−2.6	−1.2	−2.8	0.0	−2.7	−0.8	0.2
Trp	−0.4	1.6	1.6	2.5	1.2	4.0	2.7	0.0	**9.8**	−4.5
Tyr	−0.8	0.0	−1.6	0.7	−3.1	−1.6	0.8	−9.8	0.0	−3.2
Gap	4.1	8.5	−0.8	−0.7	−0.7	7.6	−0.2	4.5	3.2	0.0

[a] The residues in rows and columns represent, respectively, for meso and thermo. Positive sign indicates that the mutation from meso to thermo is favorable and the negative sign shows that the mutation from meso to thermo is unfavorable. The topmost six preferred mutations are shown in bold.

among one of the best sets to predict the stability of buried mutations in strand segments [14,15]. This shows that the increase in long-range interactions enhance the stability. It is evident that the enhancement in thermostability is achieved not only due to hydrogen bonds but also by long-range interactions. A striking feature is that the four properties R_f, E_l, P_β and N_l can explain the observation of Argos et al. [5] that the residues Gly, Ser, Lys and Asp in mesophiles are generally substituted by Ala, Thr, Arg and Glu, respectively, in thermophiles to enhance the stability, since these quantities are increased by the respective amino acid changes.

We normalized the score of each property obtained for all the 16 families by assigning the highest score for each property as ± 1 and observed that the highest scores for most of the properties are acquired by the family glycosyltransferase B. Further analysis shows that approximately one-third of the families have the highest score for the property, shape. The other eight properties, ASA_N, $-T\Delta S$, ΔG, H_{nc}, B_r, K^0, pH_i and P have the highest score in at least one of the families.

3.2. Physical interpretation for shape and $-G_{hN}$

Inspection of s values for 20 amino acids shows the following behavior: among the non-polar residues, s value increases with an increase in the number of carbon atoms; $p(L) > p(V) = p(I) > p(A)$. We observed a similar type of behavior for the other set of properties R_f, E_l, P_β and N_s that also increase the stability. Furthermore, this type of pattern is observed for a set of properties, which have significant correlation with stability upon buried mutations [14,15]. We observed a similar trend among the polar residues. The value for the property, s, increases with increase in the number of carbon atoms; $p(R) > p(K)$; $p(E) > p(D)$; $p(Q) > p(N)$. This shows that the thermostability may be ascribed to the increase in s, that generates more packing and compactness, etc., and the position of β- and higher order branches may be important for better packing. Moreover, the preference of residue substitutions from mesophilic to thermophilic proteins in Table 4 shows that the gaps in meso are filled by hydrophobic amino acids with higher preferences, yielding better packing of thermophilic proteins. In Fig. 1, we show the relationship between average shape and average environmental temperature for a set of six selected families. From this figure, we observe that the shape increases from meso to thermo with an increase in environmental temperature.

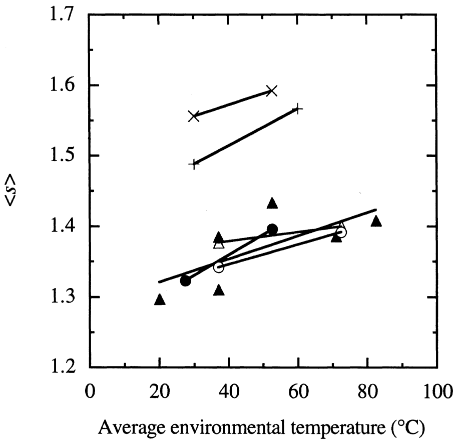

Fig. 1. Plot connecting average environmental temperature and average shape for a set of six selected families. ○: malate dehydrogenase; ▲: glyceraldehyde-3-phosphate dehydrogenase; X: thermolysin; ●: phosphoglycerate kinase; Δ: hydrolase; +: glycosyltransferase B.

The increased number of carbon atoms increases the hydrophobicity in proteins. On the other hand, it also increases the insolubility of native proteins and creates high heat capacity changes (ΔC_p) upon unfolding because ΔC_p is proportional to the number of carbon atoms in the protein interior [7,31]. This large ΔC_p destabilizes the proteins at a higher temperature than room temperature. The thermophilic protein may overcome this situation by increasing the hydration energy for native proteins. This increase in hydration energy for native proteins increases the solubility and reduces the magnitude of ΔC_p in thermophilic proteins.

In order to test this, we examined the relationship between solubility and thermostability of proteins. We assume that the solubility of native proteins is represented by their total hydration free energy, $-G_{hN}^{total} = \Sigma - G_{hN}$, and the normal-

Table 5
Hydration free energy as a measure of solubility for all the native proteins in 16 families[a]

Family no.	PDB entry ID	N	T_{env} (°C)	$-G_{\mathrm{hN}}^{\mathrm{total}}$ (kcal/mol)	$\langle -G_{\mathrm{hN}} \rangle$ (kcal/mol/residue)	Family no.	PDB entry ID	N	T_{env} (°C)	$-G_{\mathrm{hN}}^{\mathrm{total}}$ (kcal/mol)	$\langle -G_{\mathrm{hN}} \rangle$ (kcal/mol/residue)
1a	4MDH	333	37.0	387	1.162	8c	2FXB	81	52.5	90	1.111
1b	1BMD	327	72.5	429	1.312	9a	3SDP	195	27.5	235	1.205
2a	1CDG	686	35.0	834	1.216	9b	1ABM	198	37.0	259	1.308
2b	1CGT	684	35.0	824	1.205	9c	1IDS	207	37.0	265	1.280
2c	1CYG	680	52.5	890	1.309	9d	1ISA	192	37.0	233	1.214
2d	1CIU	683	60.0	861	1.261	9e	3MDS	203	72.5	282	1.389
3a	4GPD	333	20.0	357	1.072	10a	2PFK	320	37.0	426	1.331
3b	1GAD	330	37.0	383	1.161	10b	3PFK	319	52.5	415	1.301
3c	3GPD	334	37.0	370	1.108	11a	3PGK	415	27.5	489	1.178
3d	1GD1	334	52.5	414	1.240	11b	1PHP	394	52.5	507	1.287
3e	1CER	331	71.0	413	1.248	12a	1YPI	247	27.5	303	1.227
3f	1HDG	332	82.5	415	1.250	12b	1HTI	248	37.0	290	1.169
4a	6LDH	329	20.0	384	1.167	12c	1TIM	247	37.0	290	1.174
4b	1LLC	325	35.0	392	1.206	12d	1TPE	250	41.0	300	1.200
4c	5LDH	333	37.0	410	1.231	12e	1BTM	252	52.5	324	1.286
4d	9LDB	331	37.0	411	1.242	13a	1RDG	52	35.5	49	0.942
4e	1LLD	319	39.0	388	1.216	13b	6RXN	45	35.5	52	1.156
4f	1LDN	316	52.5	427	1.351	13c	8RXN	52	35.5	47	0.904
5a	1NPC	317	30.0	373	1.177	13d	5RXN	54	37.0	64	1.185
5b	1LNF	316	52.5	396	1.253	13e	1CAA	53	110.0	57	1.075
6a	2RN2	155	37.0	224	1.445	14a	1INO	175	37.0	215	1.229
6b	1RIL	166	72.5	243	1.464	14b	2PRD	174	72.5	230	1.322
7a	1ST3	269	30.0	273	1.015	15a	2EXO	312	30.0	398	1.276
7b	1SUP	275	35.0	241	0.876	15b	1XYZ	347	60.0	466	1.343
7c	1SCA	274	42.5	250	0.912	16a	1LPF	477	27.5	540	1.132
7d	1THM	279	60.0	267	0.957	16b	1LVL	458	27.5	570	1.245
8a	1FCA	55	28.0	48	0.873	16c	3LAD	476	37.0	519	1.090
8b	1FDX	54	37.0	40	0.741	16d	1EBD	455	52.5	526	1.156

[a] T_{env}, average environmental temperature; N, number of residues in a protein; $\langle -\mathrm{G}_{\mathrm{hN}} \rangle = -\mathrm{G}_{\mathrm{hN}}^{\mathrm{total}}/\mathrm{N}$.

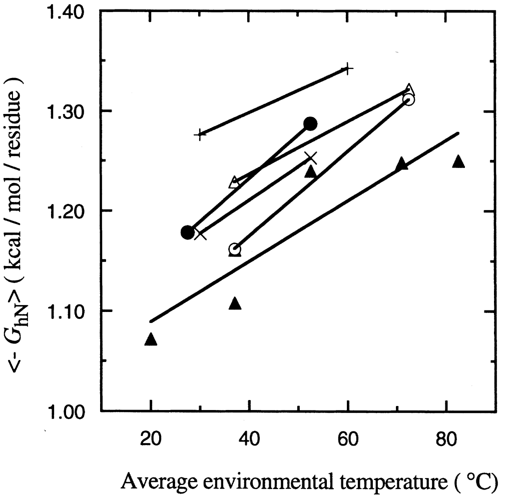

Fig. 2. Plot connecting average environmental temperature and normalized hydration energy per residue for a set of six selected families. The explanation to symbols are as in Fig. 1.

ized hydration energy per residue, $\langle -G_{\mathrm{hN}}\rangle = -G_{\mathrm{hN}}^{\mathrm{total}}/N$ (N is the total number of residues in a protein), which is a good measure to compare the solubilities of individual proteins in water. In Table 5, we present the solubility and average environmental temperature of all proteins in the 16 families. The direct relationship between solubility and average environmental temperature for a set of six selected families are shown in Fig. 2. Interestingly, from this figure and Table 5, we observe that the solubility increases from meso to thermo with an increase in environmental temperature. This may provide a physical meaning for the increase in $-G_{\mathrm{hN}}$ with increase of stability.

3.3. Relationship between ΔG_{hN} and Δshape

Ribonuclease H (RNase H) is one of the families in which more than 80% of the properties increase with increase in thermostability. We analyzed the changes in G_{hN} and shape between mesophilic (*E. coli*) and thermophilic (*Thermus thermophilus*) RNase H using a set of aligned sequences [32]. A plot showing the difference of G_{hN} and shape from meso to thermo RNase H is displayed in Fig. 3. From this figure, we observe that $-\Delta G_{\mathrm{hN}}$ is proportional to Δshape. A detailed analysis shows that the replacements from Arg in *Thermus* to residues L6 (coil), K91 (helix), K96 (coil), V106 (helix), K122 (strand) and A144 (helix) give largest changes in $-G_{\mathrm{hN}}$ (more than 5 kcal/mol) and shape. Similarly, the replacements from Arg in mesophile to C45 (strand), K79 (helix) and A111 (helix) shows the largest negative changes in these quantities. Surprisingly, all these residues are located on the surface but are not fully exposed to solvent; the fractional accessible surface area of these residues is between 23 and 63%. We note that these residues are present in all secondary structures, helix, strand and coil.

We observed a direct relationship between $-G_{\mathrm{hN}}$ and shape for all the 20 amino acid residues ($r = 0.77$) and the residue Arg is the most effective in replacement from mesophile to thermophile. This good correlation reveals that the simultaneous increases of $-G_{\mathrm{hN}}$ and shape are necessary to enhance thermostability from mesophile to thermophile. Also, this feature explains the observation of Argos et al. [5] that the residues Gly, Ser, Lys and Asp in mesophiles are generally substituted by Ala, Thr, Arg and Glu, respectively, in thermophiles, all of which tend to move closer to the regression line and towards the upper right direction. Furthermore, we speculate from this preference that Phe and Trp may be replaced by Tyr to enhance thermostability.

3.4. Comparison with other studies

Querol et al. [11] suggested that several properties, such as, increase in hydrogen bonds, hydrophobic packing, secondary structural propensity and helix-dipole stabilization increases the thermostability and the relative importance of these properties is still not clear. Vogt et al. [13] found that the increase in hydrogen bonds and increase in fractional polar surface increases the stability. In the present analysis, we found that the properties, shape and free energy of hydration are important for thermostability. Remarkably, instead of individual properties, the simultaneous increase of these two properties are very

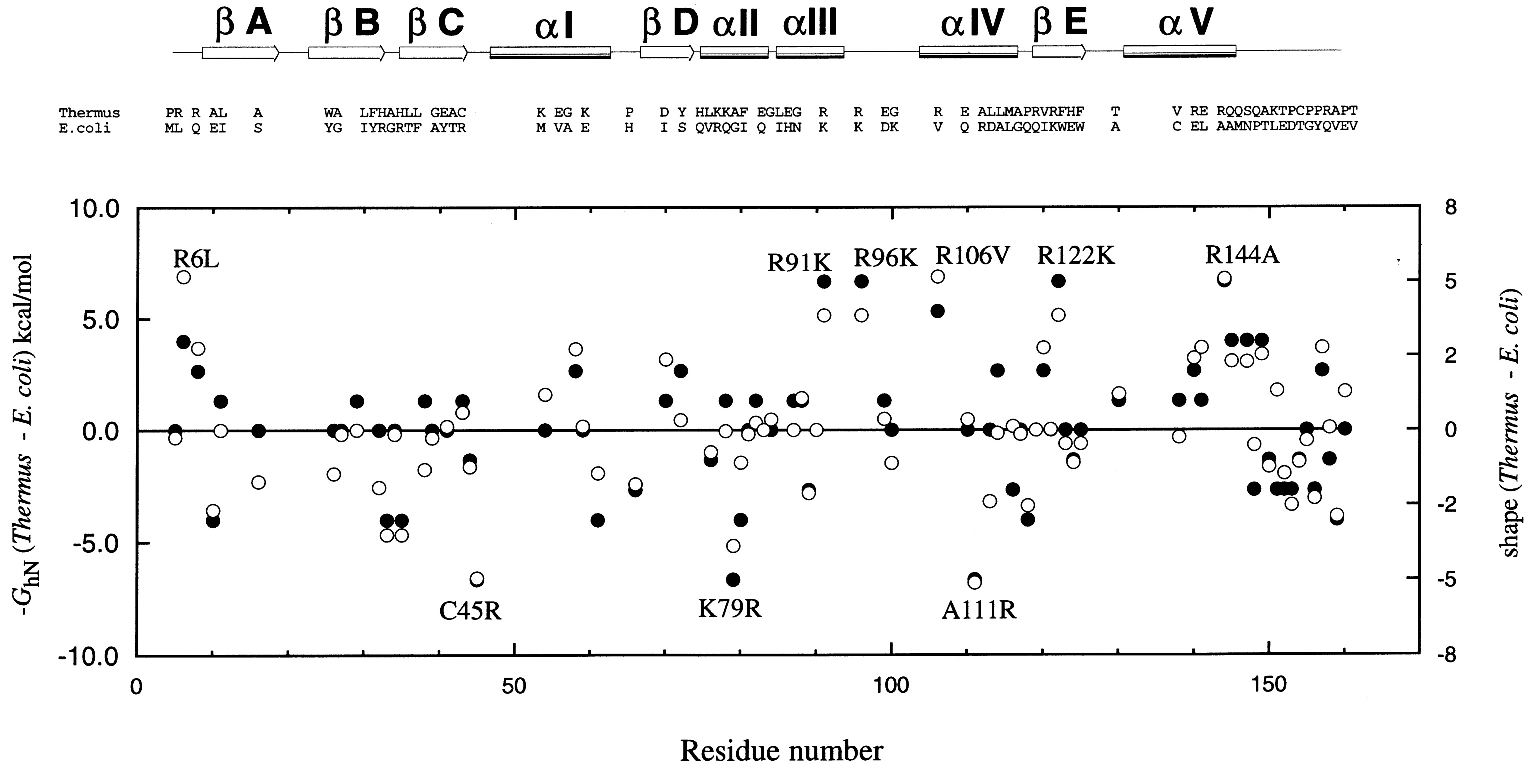

Fig. 3. Difference of $-G_{hN}$ and shape from meso to thermophilic RNase H (mutations with same residues are not shown). The amino acid residues at mutant sites of *Thermus* and *E. coli* are shown by single letter code. The helical and strand segments are indicated by cylinders (αI, αII, αIII, αIV and αV) and arrows (βA, βB, βC, βD and βE), respectively. The amino acid sequence of *Thermus* RNase H was used as a reference. ●: shape; ○: $-G_{hN}$.

important to enhance the thermostability from mesophile to thermophile, suggesting that the stability of thermophilic proteins may be achieved by a balance between better packing and solubility which show opposite effect on stability each other.

Furthermore, in comparing with Vogt et al. [13] we found that the enhancement in thermostability is achieved not only due to hydrogen bonds but also by long-range interactions. Interestingly, the properties responsible for long-range interactions, E_l, P_β and N_l can explain the Argos's replacements that have not been revealed so far. We also found that Phe and Trp may be replaced by Tyr to enhance thermostability, which has not been observed by Argos et al. [5].

3.5. Hydrophobicity and entropy show opposite effects

A detailed analysis on the effect of hydrophobicity on thermostability shows two different trends. One set of families increases stability with increase in hydrophobicity and another set of families decreases the stability with increase in hydrophobicity. The hydrophobic characters are represented by the properties H_t, H_p, E_l, N_s and ΔC_{ph} and the entropic effect is represented by the unfolding chain entropy change ($-T\Delta S_c$). Fig. 4 shows the opposite effect of $-T\Delta S_c$ [7] and H_p [33] on thermostability. From this figure and Table 3, we observe that these five properties increase the stability for nine to 11 families. When we carefully examine the details of the behavior of these properties (H_t, H_p, E_l, N_s and ΔC_{ph}) on thermostability, we find that the five families, malate dehydrogenase, lactate dehydrogenase, phosphofructokinase, phosphoglycerate kinase and reductase show a decrease in stability due to an increase in properties influenced by hydrophobicity and long-range interactions. Interestingly, we observed a common behavior among these five families. The entropic effect highly influences these families, which have positive values for $-T\Delta S_c$. This shows that the main contribution for stability to these five families are due to the chain entropic effect and not from hydrophobicity. This mechanism can directly be observed from Fig. 4 that the normalized score for $-T\Delta S_c$ is highly positive for these five families while H_p and other four properties show highly negative scores.

3.6. Influence of secondary structures on thermostability

We analyzed the influences of different secondary structures, helix, strand, turn and coil on thermostability by computing the score, S, using the secondary structural content (%) of proteins

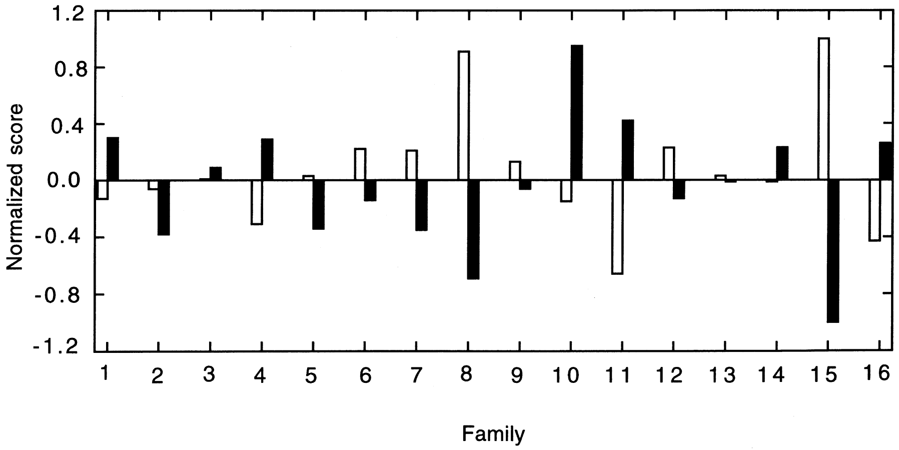

Fig. 4. Normalized scores for the properties, H_p and $-T\Delta S_c$ obtained for all the 16 families. white H_p; black: $-T\Delta S_c$.

in each family (Table 1). We observed an opposite trend between helical and turn content to the thermostability. While nine families increase the stability with increase in helical content, nine families decrease the stability with increase in the turn content. A similar behavior is also observed for strand content. The correlation coefficient between scores obtained for helix and turn in 16 families is -0.64. The increase in coil content decreases the thermostability for approximately 75% of the considered families.

3.7. Residue substitution preference from meso to thermophile

We examined our speculation, Phe and Trp may be replaced by Tyr to enhance thermostability, by aligning the sequences of mesophilic and thermophilic proteins in each family. We selected all the 380 possible mutations and analyzed the preference of each mutation by computing the total occurrences of all possible mutations in 16 families. We normalized the values, and the differences in preference of residue substitutions from mesophilic to thermophilic proteins in all 16 families are given in Table 4. From Table 4 we found that the thermophilic proteins favor mutation from Cys to Ala/Val/Ile, Trp → Tyr, Leu → Ala, etc. This clearly indicates the possibility of general substitution from Trp in mesophiles to Tyr in thermophiles to enhance the stability in addition to those proposed by Argos et al. [5]. Furthermore, from this table, strong preference of replacements is observed for the mutations, Met → Ala, Cys → Ala, Trp → Tyr, Met → Leu, Cys → Val and Cys → Ile. Interestingly, five of the mutations occur from Cys and Met (sulfur-containing residues) to other hydrophobic residues and such mutation removing sulfur atom are strongly favored from meso to thermophile. This may be due to the fact that the size of the sulfur atoms is so different from other atoms, and they might not pack easily among the other more similarly sized atoms or due to the higher reactivity of sulfur atoms than that of carbon atoms. We also found that the gaps in meso tend to be filled by Ala, Leu, Pro, Ile and Thr, which may be related to better packing in thermophilic proteins. Table 4 contains 30 preferred mutations with values greater than 4.0. Interestingly, 25 of them (83.3%) agree well with our suggestion that the simultaneous increase of $-G_{\mathrm{hN}}$ and shape is necessary to increase thermostability from mesophile to thermophile.

Thus, these results suggest that the stability of thermophilic proteins may be achieved by a balance between better packing and solubility which show an opposite effect of stability on each other.

4. Conclusions

The comparative analysis on the relation between thermostability and amino acid properties for a family of meso and thermophilic proteins revealed important factors for the stability of thermophilic proteins. We found that the properties, Gibbs free energy change of hydration for native protein and shape play a dominant role in the thermostability of proteins. The simultaneous increase of negative hydration and shape is necessary to increase the thermostability from mesophilic to thermophilic proteins. Also, the properties, long-range non-bonded energy, β-strand tendency and average long-range contacts enhance the thermostability. The properties that increase the stability are related to an increase in the number of carbon atoms within polar and non-polar amino acids. These results imply the importance of packing in the stability of thermophilic proteins. On the other hand, solubility increases in thermophilic proteins, which tends to counterbalance the increase in insolubility due to the increased number of carbon atoms of thermophilic proteins. Thus, the stability of thermophilic proteins may be achieved by a balance between better packing and solubility.

Acknowledgements

M.M.G. thanks the Japan International Science and Technology Center (JISTEC) for providing financial support in the form of STA fellowship.

References

[1] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.

[2] G.D. Rose, R. Wolfenden, Hydrogen bonding, hydrophobicity, packing and protein folding, Ann. Rev. Biophys. Biomol. Str. 22 (1993) 381–415.

[3] P.K. Ponnuswamy, M.M. Gromiha, On the conformational stability of folded proteins, J. Theor. Biol. 166 (1994) 63–74.

[4] C.N. Pace, B.A. Shirely, M. McNutt, K. Gajiwala, Forces contributing to the conformational stability of proteins, FASEB J. 10 (1996) 75–83.

[5] P. Argos, M.G. Rossman, U.M. Grau, H. Zuber, G. Frank, J.D. Tratschin, Thermal stability and protein structure, Biochemistry 18 (1979) 5698–5703.

[6] P.K. Ponnuswamy, R. Muthusamy, P. Manavalan, Amino acid composition and thermal stability of globular proteins, Int. J. Biol. Macromol. 4 (1982) 186–190.

[7] M. Oobatake, T. Ooi, Hydration and heat stability effects on protein unfolding, Prog. Biophys. Mol. Biol. 59 (1993) 237–284.

[8] T. Imanaka, M. Shibazaki, M. Takagi, A new way of enhancing the thermostability of proteases, Nature 324 (1986) 695–697.

[9] 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.

[10] A. Fontana, Analysis and modulation of protein stability, Curr. Opin. Biotech. 2 (1991) 551–560.

[11] E. Querol, J.A. Perez-Pons, A. Mozo-Villarias, Analysis of protein conformational characteristics related to thermostability, Protein Eng. 9 (1996) 265–271.

[12] G. Vogt, P. Argos, Protein thermal stability: hydrogen bonds or internal packing? Fold. Des. 2 (1997) S40–S46.

[13] G. Vogt, S. Woell, P. Argos, Protein thermal stability, hydrogen bonds, and ion pairs, J. Mol. Biol. 269 (1997) 631–643.

[14] M.M. Gromiha, M. Oobatake, H. Kono, H. Uedaira, A. Sarai, Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations, Protein Eng. 12 (1999) 549–555.

[15] M.M. Gromiha, M. Oobatake, H. Kono, H. Uedaira, A. Sarai, Relationship between amino acid properties and protein stability: buried mutations, J. Protein Chem. 18 (1999) 565–578.

[16] J. Barnett, R. Payne, D. Yarrow, Yeasts: Characteristics and Identification, Cambridge University Press, Cambridge, 1983.

[17] N.R. Krieg, J.G. Holt, Bergey's Manual of Systematic Bacteriology, William and William, Baltimore, MD, 1984.

[18] R.A. Herbert, R.J. Sharp, Molecular Biology and Biotechnology of Extremophiles, Baltimore, MD, New York, 1992.

[19] D. Iny, A. Pinsky, H. Malovani, The effect of cations on the thermophilic character of alkaline phosphatase from *Thermoactinomyces vulgaris*, Biochem. Mol. Biol. Int. 29 (1993) 729–737.

[20] R.J.M. Russell, G.L. Taylor, Engineering thermostability: lessons from thermophilic proteins, Curr. Opin. Biotech. 6 (1995) 370–374.

[21] F.C. Bernstein, T.F. Koetzle, G.J. Williams et al., The protein data bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.

[22] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometric features, Biopolymers 22 (1983) 2577–2637.

[23] M. Prabhakaran, P.K. Ponnuswamy, The spatial distribution of physical, chemical, energetic and conformational properties of amino acid residues in globular proteins, J. Theor. Biol. 80 (1979) 485–504.

[24] K. Tomii, M. Kanehisa, Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins, Protein Eng. 9 (1996) 27–36.

[25] M.M. Gromiha, P.K. Ponnuswamy, Relationship between amino acid properties and protein compressibility, J. Theor. Biol. 165 (1993) 87–100.

[26] W.F. van Gunsteren, A.E. Mark, Prediction of the activity and stability effects of site directed mutagenesis on a protein core, J. Mol. Biol. 227 (1992) 389–395.

[27] M.M. Gromiha, S. Selvaraj, Influence of medium and long range interactions in different structural classes of globular proteins, J. Biol. Phys. 23 (1997) 151–162.

[28] M.M. Gromiha, S. Selvaraj, Protein secondary structure prediction in different structural classes, Protein Eng. 11 (1998) 249–251.

[29] M.M. Gromiha, S. Selvaraj, Importance of long-range interactions in protein folding, Biophys. Chem. 77 (1999) 49–68.

[30] S. Selvaraj, M.M. Gromiha, Importance of long range interactions in $(\alpha/\beta)_8$ barrel fold, J. Protein Chem. 17 (1998) 691–697.

[31] P.L. Privalov, G.I. Makhatadze, Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects, J. Mol. Biol. 213 (1990) 385–391.

[32] S. Kimura, H. Nakamura, T. Hashimoto, M. Oobatake, S. Kanaya, Stabilization of *Escherichia coli* Ribonuclease HI by strategic replacement of amino acid residues with those from the thermophilic counterpart, J. Biol. Chem. 267 (1992) 21535–21542.

[33] P.K. Ponnuswamy, Hydrophobic characteristics of folded proteins, Prog. Biophys. Mol. Biol. 59 (1993) 57–103.

[34] M. Iqbal, R.E. Verrall, Implications of protein folding. Additivity schemes for volumes and compressibilities, J. Biol. Chem. 263 (1988) 4159–4165.

[35] D.D. Jones, Amino acid properties and side-chain orientation in proteins: a cross correlation approach, J. Theor. Biol. 50 (1975) 167–183.

[36] J.M. Zimmerman, N. Eliezer, R. Simha, The characteri-

zation of amino acid sequences in proteins by statistical methods, J. Theor. Biol. 21 (1968) 170–201.

[37] H.A. Sober, Handbook of Biochemistry, Selected Data for Molecular Biology, 2nd ed., The Chemical Rubber Co., Cleveland, Ohio, 1970.

[38] D. Eisenberg, R.M. Weiss, T.C. Terwilliger, W. Wilcox, Hydrophobic moments and protein structure, Faraday Symp. Chem. Soc. 17 (1982) 109–120.

[39] M. Oobatake, T. Ooi, An analysis of non-bonded energy of proteins, J. Theor. Biol. 67 (1977) 567–584.

[40] P.Y. Chou, G.D. Fasman, Prediction of the secondary structure of proteins from their amino acid sequence, Adv. Enzymol. 47 (1978) 45–148.

[41] T.J. Richmond, F.M. Richards, Packing of alpha-helices: geometrical constraints and contact areas, J. Mol. Biol. 119 (1978) 537–555.

[42] R. Bhaskaran, P.K. Ponnuswamy, Dynamics of amino acid residues in globular proteins, Int. J. Pept. Protein Res. 24 (1984) 180–191.

[43] C. Chothia, The nature of the accessible and buried surfaces in proteins, J. Mol. Biol. 105 (1976) 1–12.

[44] G.D. Rose, A.R. Geselowitz, G.J. Lesser, R.H. Lee, M.H. Zehfus, Hydrophobicity of amino acid residues in globular proteins, Science 229 (1985) 834–838.

[45] P. Manavalan, P.K. Ponnuswamy, Hydrophobic character of amino acid residues in globular proteins, Nature 275 (1978) 673–674.

[46] P.K. Ponnuswamy, M.M. Gromiha, Prediction of transmembrane helices from hydrophobic characteristics of proteins, Int. J. Pept. Protein Res. 42 (1993) 326–341.