Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Catalysis B: Enzymatic 26 (2003) 257–264
JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC
www.elsevier.com/locate/molcatb

Protein thermostability: structure-based difference of residual properties between thermophilic and mesophilic proteins

Seung Pil Pack, Young Je Yoo*

School of Chemical Engineering, Seoul National University, Seoul 151-742, South Korea

Received 7 July 2003; received in revised form 7 July 2003; accepted 7 July 2003

Abstract

Structure-based differences of residual properties between 20 pairs of thermophilic and mesophilic proteins were statistically analyzed to elucidate the factors governing protein thermostability. This study analyzed the distributions of outer residues, inner residues, flexible residues, rigid residues, hydrogen bonds, salt bridges, cation–pi interactions, and disulfide bonds in each protein in terms of residual structural states, which were determined as five kinds of states under residual packing value. Their structural patterns found in thermophilic protein groups were compared with those of mesophilic protein groups for showing distinctive difference of residual properties. The results of statistical tests (*t*-test) revealed that flexible residues in fully-exposed state and boundary state, salt bridges in exposed state, and hydrogen bonds in well-buried state could be critical factors related with protein thermostability. Such structure-based differences of residual properties would help to develop a strategy for enhancing protein thermostability.

Keywords: Flexibility; Salt bridges; Structural analysis; Systematical analysis; Thermostability

1. Introduction

Protein thermostability has been vigorously studied in the biophysical and biotechnological research areas, because protein instability at high temperature is main bottleneck in extending the application of protein. Several researches have suggested that the factors that may contribute to enhanced thermostability include improved hydrogen bonding, better hydrophobic packing, enhanced secondary structure propensity, helix dipole stabilization, removal of residues sensitive to oxidation or deamination, and improved electrostatic interactions [1].

Proteins from thermophilic organisms show substantially higher intrinsic thermal stabilities than their counterparts from mesophilic organisms, while retaining the basic fold characteristic of the particular protein family [2]. The comparison of structures and sequences of homologous proteins from thermophilic and mesophilic organisms could provide the important clues to stabilize proteins [3–7]. Although proteins could be engineered to achieve stability by introduction of residues or structural elements found in their homologous counterparts isolated from thermophilic or hyperthermophilic organisms, it is difficult to derive general ‘rule’ of protein thermostability from such comparative studies, because the stabilization modes found in these studies are different and various according to protein families [8].

As an alternative approach to get the general pattern of protein thermostability, systematical analyses,

* Corresponding author. Tel.: +82-2-880-7411; fax: +82-2-887-1659.
E-mail address: yjyoo@snu.ac.kr (Y.J. Yoo).

1381-1177/$ – see front matter © 2003 Published by Elsevier B.V.
doi:10.1016/j.molcatb.2003.07.004

which can statistically investigate the difference between thermophilic and mesophilic protein groups, have been tried. Systematical analyses have reported the common findings of intrinsic or extrinsic factors to underlie the increased stability of thermostable protein groups through comparing sequences and properties in several protein families [9–15]. Although these statistical studies, to some extent, succeeded to propose the characteristic factors of thermophilic proteins, their results were so general that they could not provide practical helps to stabilize protein structure or critical information to elucidate structure stability relationship [8]. Moreover, since most of the systematic studies have concentrated on the simple investigations such as the comparison of total numbers of residual properties and the observance of their relationship with temperature, they have shown a limitation to analyzing the detailed characteristics of protein thermostability.

This study seems to be a kind of a systematical analysis, which would suggest the thermostabilizing factors by investigating the general difference between thermophilic and mesophilic protein groups. However, it is distinguished from other previous systematic works in that it scrutinized the characteristic properties of model protein group in light of residual structure state, which might be closely correlated with the stabilization modes of thermophilic proteins. This study considered five kinds of residual structure states classified by residual packing value to find out the structural distributions of residual properties. Then, the structure-based differences of residual properties were investigated statistically between thermophilic and mesophilic proteins to elucidate critical factors related with protein thermostability.

2. Experimental

2.1. Protein model system

All the protein structural data used in this study were found in Protein Data Bank (PDB) atomic coordinate database at the Research Collaboratory for Structural Bioinformatics (RCSB) [16]. The protein model system of thermophilic and mesophilic protein structure sets was constructed as following procedures. At first, all the thermophilic protein structures were investigated in PDB database using “thermo” or “pyro” as search keyword. Secondly, related mesophilic protein structures were found by comparing and aligning with the sequences of the thermophilic proteins. All the mesophilic protein structures showed a residue match identity of more than 35% in pairwise alignments with their related thermophilic ones. Thirdly, the paired proteins were divided into structural families based on structural classification of protein (SCOP) database provided by MRC Laboratory of Molecular Biology and Center for Protein Engineering at Cambridge, UK [17]. Fourthly, all the protein structures were made to be non-redundant or representative of their sequences. Fifthly, one thermophilic protein was paired with its counter part mesophilic protein in each family set. When several pairs are available in same protein structure family, the best match pair was selected as representative set. Finally, if any protein set showed low resolution in protein structure, contained hetero compounds such as substrate or inhibitor, or composed of abnormal residues such as ASX or GLX, the set was not used as model system. The selected structures have at least a resolution $\leq$2.0 Å and R-factor $\leq$0.25. The designed 20 sets of thermophilic and mesophilic protein structures pairs were presented in Table 1.

2.2. Residual property calculation

Outer residues and inner residues were selected through comparing their residual surface area with one another. In this study, outer residues were defined as the top 10% residues showing the highest surface area among all the residues in a protein. Inner residues were defined as the top 10% residues showing the lowest surface area. Residual surface area was calculated by the Getarea 1.1 program, a web service provided by the Sealy Center for Structural Biology at the University of Texas Medical Branch [18]. The residual surface area computed is the locus of the center of a solvent molecule, as it rolls along the protein making maximum permitted contact. The default value of a solvent molecule size is 1.4 Å, being representative of the size of a water molecule.

Flexible residues and rigid residues were also selected through comparing their α-carbon flexibility with one another. In this study, flexible residues were defined as the top 10% residues showing the highest α-carbon flexibility among all the residues in a protein. Rigid residues were defined as the top 10% residues

Table 1
20 pairs set of thermophilic and mesophilic proteins; organism: the source organism of protein, temperature: the optimum growth temperature of the source organism

Protein name	Thermophilic proteins	Mesophilic proteins
	PDB code/organism/temperature (°C)	PDB code/organism/temperature (°C)
Adenylate kinase	1zin/*Bacillus stearothermophilus*/40–65	1aky/*Saccharomyces cerevisiae*/25–30
Che Y	1tmy/*Thermotoga maritima*/80–85	3chy/*Escherichia coli*/37
Citrate synthase	1aj8/*Pyrococcus furiosus*/100	1csh/*Chicken heart*/37
EF-TS and EF-TU-TS	1tfe/*Thermus thermophilus*/70–75	1efu_b/*Escherichia coli*/37
Endo-1.4-b xylanase	1yna/*Thermomyces lanuginosus*/50	1xnb/*Bacillus circulans*/30–40
Glutamate dehydrogenase	1gtm/*Pyrococcus furiosus*/75–100	1hrd/*Clostridium symbiosum*/30–37
Glyceraldehyde-3-phosphate dehydrogenase	1hdg/*Thermotoga maritima*/80–85	1gad/*Escherichia coli*/37
Inorganic pyrophosphatase	2prd/*Thermus thermophilus*/70–75	1ino/*Escherichia coli*/37
Lactate dehydrogenase	1ldn/*Bacillus stearothermophilus*/40–65	1ldg/*Plasmodium falciparum*/37
Malate dehydrogenase	1bdm/*Thermus flavus*/70–75	4mdh/*Porcine*/37
Manganese superoxide dismutase	3mds/*Thermus thermophilus*/70–75	1qmn/*Homo sapiens*/37
Methionine aminopeptidase	1xgs/*Pyrococcus furiosus*/100	1mat/*Escherichia coli*/37
Phosphofructokinase	3pfk/*Bacillus stearothermophilus*/40–65	2pfk/*Escherichia coli*/37
3-Phospho glycerate kinase	1php/*Bacillus stearothermophilus*/40–65	1qpg/*Saccharomyces cerevisiae*/25–30
Reductase	1ebd/*Bacillus stearothermophilus*/40–65	1lpf/*Pseudomonas fluorescens*/25–30
Ribonuclease H	1ril/*Thermus therimophilus*/70–75	2rn2/*Escherichia coli*/37
Rubredoxin	1caa/*Pyrococcus furiosus*/100	8rxn/*Desulfovibrio vulgaris*/34–37
Subtilisin	1thm/*Thermoactinomyces vulgaris*/55–65	1st3/*Bacillus lentus*/30
Thermolysin	1lnf/*Bacillus thermoproteolyticus*/53	1npc/*Bacillus cereus*/30
Triose phosphate isomerase	1btm/*Bacillus stearothermophilus*/40–65	1ypi/*Saccharomyces cerevisiae*/25–30

showing the lowest α-carbon flexibility. The residual α-carbon flexibility is calculated by obtaining the temperature B value of the α-carbon atoms in the PDB data [19].

The number of hydrogen bonds was calculated by the Protable module of Biopolymer on SYBYL. The number of hydrogen bonds is calculated by counting the number of non-hydrogen atoms in each residue involved donor and acceptor hydrogen bonds, which are determined by the distance of donors and acceptors within 4.0 Å [20]. The number of salt bridges, the number of cation–pi interactions and the number of disulfide bonds were calculated by the Protein Explorer package 1.9 provided by the Department of Microbiology, University of Massachusetts, Amherst. Salt bridges are assigned to two atoms of opposite charge, when the atoms were observed to be within 4.0 Å. Positively charged atoms include side chain N atoms in LYS, ARG, and HIS while negatively charged atoms include side chain O atoms in ASP and GLU [21]. Cation–pi interactions are assigned to aromatic residues, when a cationic side chain of ARG or LYS is near an aromatic side chain of PHE, TRP, or TYR. Ninety-nine percent of significant cation–pi interactions occur within a distance of 6.0 Å [22].

2.3. Residual structure index

The residual structure index was used as a standard index for describing residual conformational state in protein structure. The residual packing value for each residue was calculated by an extension of the occluded surface algorithm [23]. Residual packing value equals 0.0 if there is no occluding van der Waals surface within 2.8 Å of the molecular surface; equals 1.0 if 100% of the molecular surface were in contact with van der Waals surface of other atoms. Residues were endowed with structure index determined under the range of its residual packing value as arranged in Table 2. All the residues in a protein can be categorized to five classes according to structure index 1–5.

2.4. Statistical analysis

X_{ij} trait was defined in this study for describing the frequency of each residual property X in particular

Table 2
Relationship between structure index and residual packing value

Structure index	Residual packing value	Physical meaning
1	0.00–0.15	Fully-exposed state
2	0.15–0.30	Exposed state
3	0.30–0.45	Boundary state
4	0.45–0.60	Buried state
5	0.60–0.75	Well-buried state

structure index i of particular protein structure j. Therefore, X_{i}, average trait means the average frequency of each residual property (X) in particular structure index i of protein groups. X_i, average trait and S_i^2, its deviation are calculated as described below:

$$X_i = \frac{\sum X_{ij}}{\sum j} \tag{1}$$

$$S_i^2 = \frac{\sum (X_{ij} - X_i)^2}{\sum j - 1} \tag{2}$$

For obtaining an average trait (X_i) to have a higher value in thermophilic proteins than in mesophilic proteins, the appropriate null hypothesis would be that the average values of the traits in structure index i are equal in both groups as described below:

$$\mathrm{H}_0 : X_{i\text{-Th}} = X_{i\text{-Me}}, \tag{3}$$

where $X_{i\text{-Th}}$ is the average value of trait in structure index i of thermophilic protein groups, and $X_{i\text{-Me}}$ the average value of trait in structure index i of mesophilic protein groups. Therefore, an alternative hypothesis against the null hypothesis would be as follows:

$$\mathrm{H}_1 : X_{i\text{-Th}} > X_{i\text{-Me}} \tag{4}$$

Then, test statistic of average trait (X_i) of both groups could be carried out by calculation of t value, t_i as follows:

$$t_i = \frac{X_{i\text{-Th}} - X_{i\text{-Me}}}{\sqrt{(S^2_{i\text{-Th}}/N_{\mathrm{Th}} + S^2_{i\text{-Me}}/N_{\mathrm{Me}})}} \tag{5}$$

where $S^2_{i\text{-Th}}$ and $S^2_{i\text{-Me}}$ are the deviations of average traits, X_i in structure index i of thermophilic protein groups and mesophilic protein groups, respectively. N_{Th} and N_{Me} are the total number (20 proteins) of thermophilic protein groups and mesophilic protein groups investigated in this study, respectively.

The degrees of freedom, $\mathrm{d.f.} = N_{\mathrm{Th}} + N_{\mathrm{Me}} - 2$ are 38, which value is large enough to be considered as infinite sample sets. For a one tailed t-test (with $\mathrm{d.f.} =$ infinite), critical levels of t values are as follows [24]:

t value	d.f. = infinite (>30)
$t_{0.1}$	1.282
$t_{0.05}$	1.645
$t_{0.025}$	1.960
$t_{0.01}$	2.326
$t_{0.005}$	2.576

For a one tailed t-test at a 1% level of significance, H_0 is rejected for $t > 2.326$ ($\mathrm{t}_{0.01}$) or $t < -2.326$ ($-t_{0.01}$). If $t_i > 2.326$ then the probability that average traits, X_i of thermophilic protein groups are greater than X_i of mesophilic protein groups in structure index i is >0.99. In contrast, If $t_i < -2.326$ then the probability that average traits, X_i of thermophilic protein groups are less than X_i of mesophilic protein groups in structure index i is >0.99.

3. Results and discussion

3.1. Structure-based differences of outer residues and inner residues

To investigate the distribution of the residues showing the distinctive values of surface area, outer residues (the top 10% residues showing the highest surface area) and inner residues (the top 10% residues showing the lowest surface area) were used in this study. Under 99.5% level of significance, outer and inner residues considered in such manner were found in the outer and the inner part, respectively, which are determined by Fraczkiewicz and Braun's concept [18]. This result means that the defined residues could be expected as the representative residues showing the distinctive values of surface area. As given in Table 3, outer residues and inner residues in both of thermophilic and mesophilic proteins were investigated in terms of their average frequencies according to structure index. Regardless of thermophilic and mesophilic proteins, outer residues were found prevalently in exposed state (structure index 2), while inner ones were prevalently in buried state (structure index 4). The results of t-test

Table 3
Structural distributions of outer residues and inner residues; Thermo: average frequency (%) of thermophilic protein groups, Meso: average frequency (%) of mesophilic proteins groups

Structure index	Outer residues		Inner residues	
	Thermo	Meso	Thermo	Meso
1	3.2125 ± 0.3369	3.6544 ± 0.2729	0.0000	0.0000
2	6.8329 ± 0.3000	6.4298 ± 0.2419	0.0000	0.0000
3	0.2093 ± 0.0707	0.1475 ± 0.0548	0.1391 ± 0.0447	0.1128 ± 0.0447
4	0.0000	0.0000	8.9607 ± 0.4868	8.4356 ± 0.4712
5	0.0000	0.0000	3.4019 ± 0.4626	2.6597 ± 0.3450

showed how much the structural distributions of outer residues and inner residues (average frequencies of the residues according to structure index) are different between thermophilic and mesophilic proteins, as presented in Table 7. A positive value in the *t*-test result indicates that the structural property has a higher numerical value in the thermophilic proteins than in the mesophilic ones while a negative value indicates the opposite result. All the *t*-test values in the structural distribution of outer residues are below the critical value of $t_{0.10}$ (1.282) and above $-t_{0.10}$ (−1.282). This result means that the structural patterns of outer residues have no distinctive difference between thermophilic and mesophilic proteins. On the other hand, the inner residues in thermophilic proteins showed significant difference in structure index 5 compared with those of mesophilic ones. The *t*-test value of structure index 5 is 1.2862. It implies that the probability that the inner residues showing structure index 5 are found more frequently in thermophilic proteins than in mesophilic ones is near to 0.90. This result indicates that the inner residues in thermophilic proteins have a tendency to have better buried and packed conformation than those of mesophilic ones. This trend agreed with the finding that better internal packing would be dominant cause of protein thermostability [13].

3.2. Structure-based differences of flexible and rigid residues

It has been widely accepted that the residues with higher flexibility could evoke local mobility of protein structure, whereas the ones with lower flexibility serve local rigidity of protein structure [1]. To investigate the distribution of the residues showing the distinctive values of flexibility, flexible residues (the top 10% residues showing the highest flexibility) and rigid residues (the top 10% residues showing the lowest flexibility) were used in this study. Under 99.5% level of significance, flexible and rigid residues considered in such manner belonged to the flexible and the rigid regions, respectively, which are determined by Thorpe and co-workers' theory [25]. This result means that the defined residues could be expected the representative residues showing the distinctive values of flexibility. As given in Table 4, flexible residues and rigid residues in both of thermophilic and mesophilic proteins were investigated in terms of

Table 4
Structural distributions of flexible residues and rigid residues; Thermo: average frequency (%) of thermophilic protein groups, Meso: average frequency (%) of mesophilic proteins groups

Structure index	Flexible residues		Rigid residues	
	Thermo	Meso	Thermo	Meso
1	1.3946 ± 0.1565	2.0419 ± 0.2202	0.1038 ± 0.0592	0.1387 ± 0.0500
2	4.9970 ± 0.3661	5.1465 ± 0.3557	1.0810 ± 0.2225	1.0478 ± 0.2168
3	3.1204 ± 0.2291	2.2135 ± 0.2398	2.9461 ± 0.3384	2.6281 ± 0.2550
4	0.7044 ± 0.2037	0.7528 ± 0.1746	5.3009 ± 0.3715	5.7065 ± 0.4117
5	0.0562 ± 0.0447	0.0769 ± 0.0316	0.9714 ± 0.2408	0.9473 ± 0.2636

their average frequencies according to structure index. Flexible residues were found usually in exposed state (structure index 2) and rigid ones were found in buried state (structure index 4) although their distributions were displayed in all around structure states. Most of structural biologists have agreed that thermophilic proteins have more rigid conformation than mesophilic ones [26]. However, Lazaridis et al. argued that there is no fundamental reason for stability and rigidity to be correlated [27]. Flexibility implies the increased conformational entropy of folded state, and it should be favorable to thermodynamic stability. Therefore, more studies on protein flexibility are demanded for better understanding the relationship of conformational rigidity and stability [26]. This study showed how much the structural distribution of flexible residues is different between thermophilic and mesophilic proteins, as presented in Table 7. Especially, lower frequency in structure index 1 (fully-exposed state) and higher frequency in structure index 3 (boundary state) were observed to be distinctive patterns of flexible residues in thermophilic proteins compared with the patterns of mesophilic ones. The *t*-test values of structure index 1 and 3 are −2.3958 and 2.7344, respectively. These results mean that the probability that the average frequency of flexible residues in thermophilic proteins is lower in structure index 1 or higher in structure index 3 than that of mesophilic proteins is over 0.99. That is to say, the boundary state (structure index 3) is observed to be more preferable state of the flexible residues in thermophilic proteins than the fully-exposed state (structure index 1). It has been reported that protein flexibility and rigidity would be related with several molecular interactions such as hydrophobic interaction, hydrogen bonding, salt bridge, or disulfide bond [28]. In boundary state, flexible residues have higher chance to interact with neighboring residues and to be stabilized by molecular interactions and forces. These results inferred how the flexible residues in thermophilic proteins would be distributed to enhance the thermostability. On the other hand, the average frequencies of rigid residues did not show any characteristic difference between thermophilic and mesophilic proteins as arranged in Table 7. Their *t* values were investigated to be below the critical value of $t_{0.10}$ (1.282) and above $-t_{0.10}$ (−1.282).

3.3. Structure-based differences of hydrogen bonds and salt bridges

Hydrogen bonds and salt bridges in both of thermophilic and mesophilic proteins were investigated in terms of their average frequencies according to structure index as shown in Table 5. Hydrogen bonds and salt bridges have been considered as important molecular interactions to play critical roles in stabilizing protein structures [21]. The structure-based differences of hydrogen bonds and salt bridges between thermophilic and mesophilic proteins are presented in Table 7.

Hydrogen bonds in structure index 5 were found to show the characteristic difference between thermophilic and mesophilic proteins since the *t*-test value is 1.4903 above the critical value of $t_{0.10}$ (1.282). This result indicated that hydrogen-bonding residues in thermophilic proteins would be more packed and buried residues than those of mesophilic proteins. It was reported that better hydrogen bonding would be dominant cause of protein thermostability [13]. This result specified that the hydrogen bonds especially in well-buried state (structure index 5) would play more significant roles of protein thermostability than the hydrogen bonds in other states.

Table 5
Structural distributions of hydrogen bonds and salt bridges; Thermo: average frequency (%) of thermophilic protein groups, Meso: average frequency (%) of mesophilic proteins groups

Structure index	Hydrogen bonds		Salt bridges	
	Thermo	Meso	Thermo	Meso
1	1.7641 ± 0.2711	2.0700 ± 0.2864	0.0759 ± 0.0447	0.1293 ± 0.0447
2	30.2723 ± 1.3253	30.5112 ± 1.6773	3.0737 ± 0.3742	2.0630 ± 0.2480
3	57.6807 ± 2.0316	57.4965 ± 1.6670	4.0756 ± 0.4006	3.4571 ± 0.3536
4	64.0355 ± 2.9790	62.0848 ± 2.9207	1.4051 ± 0.2345	1.2837 ± 0.1517
5	8.5852 ± 1.1950	6.4256 ± 0.8198	0.0672 ± 0.0316	0.0447 ± 0.0224

Table 6
Structural distributions of cation–pi interactions and disulfide bonds; Thermo: average frequency (%) of thermophilic protein groups, Meso: average frequency (%) of mesophilic proteins groups

Structure index	Cation–pi interactions		Disulfide bonds	
	Thermo	Meso	Thermo	Meso
1	0.0000	0.0000	0.0000	0.0000
2	0.0342 ± 0.0316	0.0391 ± 0.0387	0.0000	0.0000
3	0.2432 ± 0.0922	0.3540 ± 0.1432	0.0369 ± 0.0316	0.0106
4	0.4150 ± 0.1323	0.4796 ± 0.1414	0.0369 ± 0.0316	0.0106
5	0.1202 ± 0.0975	0.0000	0.0000	0.0000

Salt bridges in structure index 2 were observed to show the characteristic difference between thermophilic and mesophilic proteins since the *t*-test value is 2.2516 above the critical value of $t_{0.025}$ (1.960). This result indicated that salt bridges in thermophilic proteins would be important factors especially in exposed part of proteins. A significant increase in the number of salt bridges has been reported for most structures of thermostable proteins [6,29,30]. Several comparative studies between proteins from thermophilic and mesophilic organisms investigated that the number of salt bridges increases in thermostable proteins [5,29–32]. However, all the comparative studies did not report a significant increase in the number of salt bridges in thermophilic proteins. Unlike other comparison studies, this work showed common occurrence of increased salt bridges throughout thermophilic proteins, especially in exposed state (structure index 2). Salt bridges could be expected to stabilize the exposed part of a protein structure, which might be more flexible and less stable than buried part.

3.4. Structure-based differences of cation–pi interactions and disulfide bonds

Cation–pi interactions and disulfide bonds in both of thermophilic and mesophilic proteins were investigated in terms of their average frequencies according to structure index as shown in Table 6. Cation–pi interactions and disulfide bonds have been also considered as important molecular interactions to play critical roles in stabilizing protein structure [21]. The structure-based differences of cation–pi interactions and disulfide bonds between thermophilic and mesophilic proteins are presented in Table 7. Cation–pi interactions and disulfide bonds did not show a meaningful difference between thermophilic and mesophilic proteins in terms of structure states. However, the residues with cation–pi interactions in thermophilic proteins were observed to have better packed forms than those of mesophilic ones. The *t*-test value is 1.232, which is near to the critical value of $t_{0.10}$ (1.282).

Table 7
t-test results showing the structure-based difference of residual properties between thermophilic and mesophilic proteins; the bold values are significant at the 10% level ($t_{0.10}$ (1.282)), the bold and underlined values are significant at 2.5% ($t_{0.025}$ (1.960)), and the bold italic and underlined values are significant at 1% ($t_{0.01}$ (2.326))

Residual property	Residual structure state				
	Index 1; fully-exposed state	Index 2; exposed state	Index 3; boundary state	Index 4; buried state	Index 5; well-buried state
Outer residues	−1.0192	1.0460	0.6909	0.0000	0.0000
Inner residues	0.0000	0.0000	0.4158	0.7751	**1.2862**
Flexible residues	**−2.3958**	−0.2929	**2.7344**	−0.1804	−0.3779
Rigid residues	−0.4506	0.1069	0.7506	−0.7314	0.0675
Hydrogen bonds	−0.7757	−0.1118	0.0701	0.4676	**1.4903**
Salt bridges	−0.8443	**2.2516**	1.1575	0.4347	0.5809
Cation–pi interactions	0.0000	−0.0980	−0.6506	−0.3336	1.2332
Disulfide bonds	0.0000	0.0000	0.8317	0.8317	0.0000

4. Conclusion

Up to now, there have been several systematical analyses to obtain thermostabilizing factors by investigating the difference of residual properties in thermophilic and mesophilic protein groups. This study is distinguished from other systematic works in that it considered three-dimensional structure states of residues when it analyze the information of residual properties related with protein thermostability. Structural distributions of residual properties were statistically compared between 20 pairs of thermophilic and mesophilic proteins. This study revealed several structural patterns, which prevailed in thermophilic proteins compared with mesophilic proteins: (1) lower frequency in fully-exposed state and higher frequency in boundary state of flexible residues; (2) higher frequency in exposed state of salt bridges; (3) higher frequency in well-buried state of hydrogen bonds; (4) higher frequency in well-buried state of inner residues. Such structure-based differences of residual properties can be considered as general and critical patterns related with protein thermostability. These results agreed with the previous findings identified from other studies, in addition, specified the findings in terms of their structural patterns. Therefore, this present work might help to develop a strategy for enhancing protein thermostability.

References

[1] T.E. Creighton, Proteins: structures and molecular properties, second ed., Freeman & Company, New York, 1997.
[2] R. Jaenicke, FASEB J. 10 (1996) 84.
[3] K. Hecht, A. Wrba, R. Jaenicke, Eur. J. Biochem. 183 (1990) 69.
[4] B.W. Matthews, Ann. Rev. Biochem. 62 (1993) 139.
[5] R.J. Russel, D.W. Hough, M.J. Danson, L.T. Garry, Structure 2 (1994) 1157.
[6] K.S. 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, Structure 3 (1995) 1147.
[7] P.C. Michels, D. Hei, D.S. Clark, Adv. Protein Chem. 48 (1996) 341.
[8] B.V. Burg, V.G. Eijsink, Curr. Opin. Biotech. 13 (2002) 333.
[9] P. Argos, M.G. Rossmann, U.M. Garu, H. Zuber, G. Frank, J.D. Tratschin, Biochemistry 18 (1979) 5698.
[10] V.Z. Spassov, A.D. Karshikoff, R. Ladenstein, Protein Sci. 4 (1995) 1516.
[11] G.L. Warren, G.A. Petsko, Protein Eng. 8 (1995) 905.
[12] A. Karshikoff, R. Landenstein, Protein Eng. 11 (1998) 867.
[13] G. Vogt, P. Argos, Fold. Des. 2 (1997) S40.
[14] G. Vogt, S. Woell, P. Argos, J. Mol. Biol. 269 (1997) 631.
[15] A. Szilagyi, P. Zavodszky, Structure 8 (2000) 493.
[16] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235.
[17] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, J. Mol. Biol. 247 (1995) 536.
[18] R. Fraczkiewicz, W. Braun, J. Comp. Chem. 19 (1998) 319.
[19] S. Parthasarathy, M.R. Murthy, Protein Sci. 6 (1997) 2561.
[20] I.K. McDonald, J.M. Thornton, J. Mol. Biol. 238 (1994) 777.
[21] K.A. Dill, Biochemistry 29 (1990) 7133.
[22] J.P. Gallivan, D.A. Dougherty, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 9459.
[23] P.J. Fleming, F.M. Richard, J. Mol. Biol. 299 (2000) 487.
[24] W. Mendenhall, R.J. Beaver, Introduction to Probability and Statistics, PWS-KENT, Boston, 1991.
[25] D.J. Jacobs, A.J. Rader, M.F. Thorpe, L.A. Kuhn, Proteins 44 (2001) 150.
[26] C. Vielle, S.J. Zeikus, Micobiol. Mol. Biol. Rev. 65 (2001) 1.
[27] T. Lazaridis, I. Lee, M. Karplus, Protein Sci. 6 (1997) 2589.
[28] P.A. Fields, Compara. Biochem. Physiol. 129 (2001) 417.
[29] I. Korndörfer, B. Steipe, R. Huber, A. Tomschy, R. Jaenicke, J. Mol. Biol. 246 (1995) 511.
[30] C.A. Kelly, M. Nishiyama, Y. Ohnishi, T. Beppu, J.J. Birktoft, Biochemistry 32 (1993) 3913.
[31] S. Korolev, M. Nayal, W.M. Barnes, E. Di Cera, G. Waksman, Proc. Natl Acad. Sci. U.S.A. 92 (1995) 9264.
[32] M. Perutz, H. Raidt, Nature 255 (1975) 256.