

# Important inter-residue contacts for enhancing the thermal stability of thermophilic proteins

M. Michael Gromiha\*

*RIKEN Tsukuba Institute, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan*

Received 17 January 2001; received in revised form 20 March 2001; accepted 20 March 2001

---

## Abstract

Proteins from thermophilic organisms exhibit high thermal stability, but have structures that are very similar to their mesophilic homologues. In order to gain insight into the basis of thermostability, we have analyzed the medium- and long-range contacts in mesophilic and thermophilic proteins of 16 different families. We found that the thermophiles prefer to have contacts between residues with hydrogen-bond-forming capability. Apart from hydrophobic contacts, more contacts are observed between polar and non-polar residues in thermophiles than mesophiles. Residue-wise analysis showed that Tyr has good contacts with several other residues, and Cys has considerably higher long-range contacts in thermophiles compared with mesophiles. Furthermore, the residues occurring in the range of 31–34 residues apart in the sequence contribute significant long-range contacts to the stability of thermophilic proteins. © 2001 Elsevier Science B.V. All rights reserved.

**Keywords:** Long-range interactions; Hydrogen bond; Thermal stability; Residue contacts

---

## 1. Introduction

Thermophilic organisms produce proteins of exceptional stability [1,2]. Understanding the structural basis for the enhanced stability of proteins from thermophilic organisms relative to their mesophilic counterparts is a challenging problem.

Recently, Ladenstein and Antranikian [3] reviewed the possible factors in the mechanism of protein stabilization under different conditions, such as high temperature, salt and extreme pH, and the contributions of different interactions to enhance the thermostability. Jaenicke and Bohm [4] explained the significant enhancement in the understanding of extremophilic adaptation from four viewpoints: the discovery of more extreme thermophiles; complete genome sequences; new genetic methods; and the rapidly increasing num-

---

\*Tel.: +81-298-36-9082; fax: +81-298-36-9080.

E-mail address: gromiha@rtc.riken.go.jp (M. Michael Gromiha).

ber of high-resolution three-dimensional structures.

The analysis of three-dimensional structures of proteins from both mesophiles and thermophiles has provided several insights into the structural basis of thermal stability. Querol et al. [5] studied the relationship between thermal stability and conformational characteristics of proteins. The thermostability of 16 different families of mesophilic and thermophilic proteins has been examined by Vogt et al. [6], who found a good correlation between the thermostability of the familial members and an increase in the number of hydrogen bonds, as well as in the fractional polar surface. Our previous analysis showed that the increase in negative Gibbs free-energy change for hydration of native proteins ( $-G_{\text{hN}}$ ) increases the stability of thermophiles, and this stability may be achieved by a balance between better packing and solubility [7]. Recently, Hasegawa et al. [8] increased the stability of mesophilic cytochrome through five substitutions and observed that the  $-G_{\text{hN}}$  may contribute to the stability, consistent with the suggestion of Gromiha et al. [7]. Furthermore, it has been reported that the increase in helical propensity of amino acid residues [9], salt bridges and side-chain–side-chain hydrogen bonds [10], ion pairs [11] and van der Waals' contacts [12] enhances the stability of thermophilic proteins.

On the other hand, Hollien and Marquese [13] determined protein stability from mesophilic and thermophilic ribonuclease H and Gromiha [14] analyzed the factors influencing the stability of  $\alpha$ -helices and  $\beta$ -strands in thermophilic ribonuclease H. Furthermore, Perl et al. [15] showed that the surface-exposed residues are responsible for the thermostability of cold shock proteins.

In spite of these studies, the important inter-residue contacts for enhanced thermostability of thermophilic proteins have not yet been fully explored. Our recent studies showed that medium- and long-range interactions are important for the structure and stability of proteins [16,17] and for understanding the stability of proteins upon mutations [18–20]. Hence, in this work, we have analyzed the inter-residue contacts in mesophilic and thermophilic proteins of 16 different families.

We found that an increase in contacts between residues of hydrogen-bond (H-bond)-forming capability increases the stability. Furthermore, the role of different amino acid residues in forming inter-residue contacts and the contribution of long-range contacts in terms of residue distance to the stability of thermophiles will be discussed.

## 2. Materials and methods

### 2.1. Database

In our previous work [7], we have used a set of 56 globular proteins from 16 different families to analyze the important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. We have selected the proteins of highest and lowest average environmental temperature in each family for the present analysis. The Protein Data Bank (PDB) codes for the mesophilic proteins are: 4MDH, 1CDG, 4GPD, 6LDH, 1NPC, 2RN2, 1ST3, 1FCA, 3SDP, 2PFK, 3PGK, 1YPI, 1RDG, 1INO, 2EXO and 1LPF, and the PDB codes for their respective thermophilic proteins are: 1BMD, 1CIU, 1HDG, 1LDN, 1LNF, 1RIL, 1THM, 2FXB, 3MDS, 3PFK, 1PHP, 1BTM, 1CAA, 2PRD, 1XYZ and 1EBD. The three-dimensional structures of all these proteins have been taken from the Protein Data Bank [21].

### 2.2. Computation of surrounding residues and long-range contacts

The computation of surrounding residues in a protein molecule has been described in our earlier articles [16,17]. The residues in a protein molecule are represented by their  $C_{\alpha}$  atoms. Using the  $C_{\alpha}$  coordinates, a sphere of radius 8 Å is fixed around each residue and the composition of surrounding residues associated with all the residues is calculated. It has been shown that the influence of each residue over the surrounding medium extends effectively only up to 8 Å [22].

For a given residue, the composition of surrounding residues is analyzed in terms of the location at the sequence level and the contribu-

tions from  $< \pm 3$  residues are treated as short-range,  $\pm 3$  or  $\pm 4$  residues as medium-range, and  $> \pm 4$  residues are treated as long-range contacts [23]. The long-range contacts are further classified into several intervals with a step of one residue, and the results were analyzed.

### 2.3. Preference of surrounding residues influenced by medium- and long-range contacts

The residues coming within a sphere of 8 Å for each residue in all proteins were computed, and the residues which contribute towards medium and long-range contacts were selected as described above. The average preference of surrounding residues is computed using the expression:

$$\langle N \rangle_{ij} = \frac{\sum N_{ij}}{\sum N_i + \sum N_j}$$

where  $N_{ij}$  is the number of surrounding residues of type  $j$  around residue  $i$ .  $N_i$  and  $N_j$  are the total number of residues of type  $i$  and  $j$ , respectively.

## 3. Results and discussions

### 3.1. Medium- and long-range contacts in mesophilic and thermophilic proteins

The total number of medium- and long-range contacts in the three-dimensional structures of mesophilic and thermophilic proteins showed that both these proteins have similar contacts, indicating their structural similarities. Hence, we analyzed the influence of each amino acid residue in forming the contacts with other residues. The average medium- and long-range contacts in mesophilic and thermophilic proteins are presented in Table 1. We found that most of the residues have approximately similar contacts, and the residues Cys, Arg, Asn and Trp have different behaviors in mesophilic and thermophilic proteins. Cys has more long-range contacts in thermophiles than mesophiles, while an opposite trend is observed for medium-range contacts. Arg and

Table 1  
Medium- and long-range contacts in mesophilic and thermophilic proteins

Residue	Medium			Long		
	Meso	Thermo	Ratio	Meso	Thermo	Ratio
Ala	2.32	2.38	1.03	4.50	4.84	1.08
Asp	1.75	1.71	0.98	3.47	3.54	1.02
Cys	1.75	1.42	<b>0.81</b>	5.98	6.74	<b>1.13</b>
Glu	2.03	2.13	1.05	3.14	2.83	0.90
Phe	1.99	2.01	1.01	4.74	4.71	0.99
Gly	1.56	1.61	1.03	4.74	4.66	0.98
His	2.03	2.08	1.02	3.75	3.49	0.93
Ile	1.76	1.81	1.03	5.76	5.88	1.02
Lys	2.07	2.07	1.00	2.90	3.14	1.08
Leu	2.12	2.21	1.04	4.90	4.63	0.94
Met	2.27	2.38	1.05	4.36	4.61	1.06
Asn	1.78	1.81	1.02	3.99	3.64	<b>0.91</b>
Pro	1.39	1.36	0.98	4.25	4.43	1.04
Gln	2.11	2.26	1.07	3.21	3.10	0.97
Arg	1.96	2.14	<b>1.09</b>	4.01	3.93	0.98
Ser	1.63	1.67	1.02	4.03	4.18	1.04
Thr	1.57	1.52	0.97	4.72	4.78	1.01
Val	1.50	1.57	1.05	6.28	6.31	1.00
Trp	1.68	1.69	1.01	5.02	5.53	<b>1.10</b>
Tyr	1.91	1.82	0.95	4.73	4.70	0.99

Trp have more medium- and long-range contacts, respectively, in thermophilic proteins. The higher number of medium-range contacts exhibited by Arg might be due to the presence of a greater number of ion pairs and salt bridges in thermophilic proteins, which are reported to be important factors for enhancing protein thermostability [10,11].

### 3.2. Long-range contacts in different residue intervals

The number of long-range contacts for different residue intervals in mesophilic and thermophilic proteins have been computed and the difference between these two sets of proteins is depicted in Fig. 1. In this figure, the dominating long-range contacts in thermophiles have significant positive values and high long-range contacts in mesophiles are below zero. Interestingly, we found a stretch of high long-range contacts at residue distances 31–34 in the sequence, which might contribute to the stability of thermophilic proteins. This interval is close to the limit of

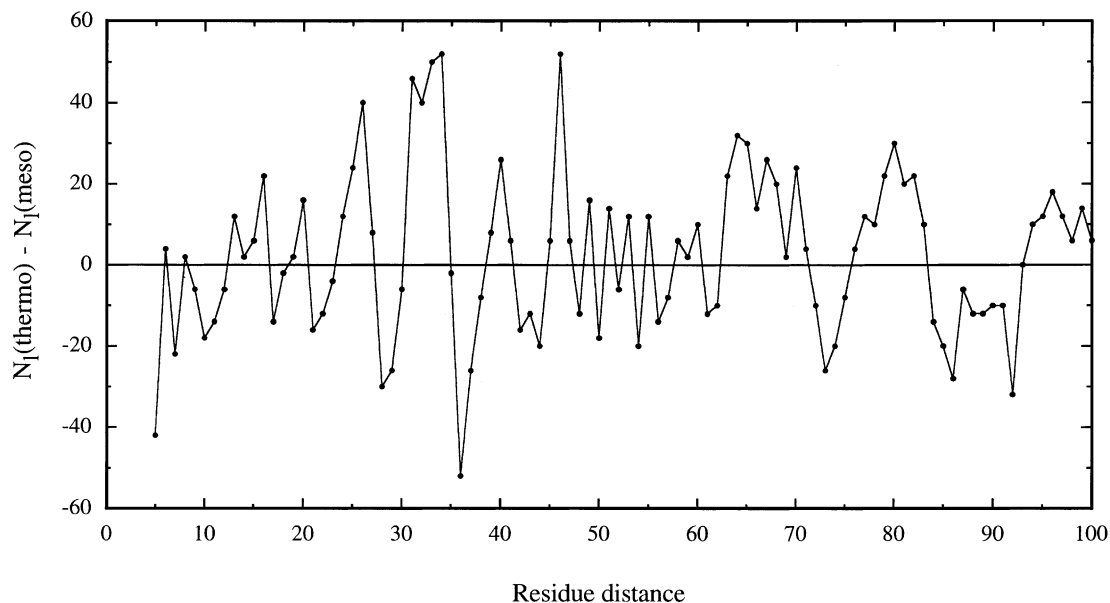


Fig. 1. Plot connecting residue distance and the difference in number of long-range contacts between mesophilic and thermophilic proteins.

residue distances (20–30) to form long-range contacts for maintaining the  $(\alpha/\beta)_8$  barrel fold [24]. Furthermore, we observed high long-range contacts in the C-terminal domain of thermophilic proteins for residue intervals 63–71, 76–83 and 94–100, and in that of mesophilic proteins for residue intervals 72–75 and 84–92. These residue distances for the formation of long-range contacts may influence the stability of thermophilic proteins.

### 3.3. Number of long-range contacts versus number of residues

The percentage of residues for different numbers of long-range contacts in mesophilic and thermophilic proteins are displayed in Fig. 2. From the figure, we observed high peaks at 3–5 long-range contacts in thermophilic proteins, whereas the corresponding number was 2–4 in mesophilic proteins. This result indicates the importance of long-range contacts to the stability of thermophilic proteins.

### 3.4. Residue pairs influenced by medium- and long-range contacts

We have computed the preference of each amino acid residue to be surrounded by each of the 20 amino acid residues due to medium- and long-range contacts, and the average preference. The residue pairs with preference above the average have been selected for further analysis.

#### 3.4.1. Medium-range contacts

An appreciable number of residue pairs between two hydrophobic residues has been observed in both mesophilic and thermophilic proteins. We selected the residue pairs which are mainly observed in thermophiles, but not in mesophiles, and vice versa; the results are given in Table 2a. We found that the thermophiles have more residue pairs influenced by medium-range contacts than mesophiles. This result demonstrates the feature of forming a wide range of residue pairs (inter-residue contacts) in ther-

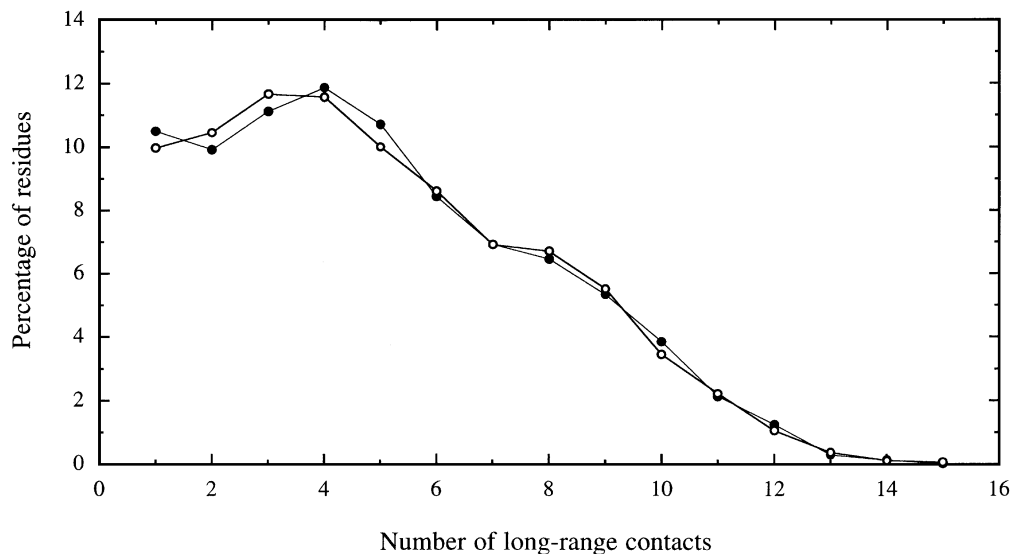


Fig. 2. Number of long-range contacts vs. percentage of residues in mesophilic and thermophilic proteins: open circles, mesophilic; and filled circles, thermophilic.

mesophilic proteins other than specific interactions, such as hydrophobic clusters, which are observed in both mesophilic and thermophilic proteins. Classification of residues into polar, hydrophobic and H-bond-forming capability revealed that the thermophilic proteins have a significant number of inter-residue contacts between residues of H-bond-forming capability. Recent structural analysis on a set of mesophilic and thermophilic proteins also showed that an increase in the number of hydrogen bonds increases the stability [6]. Furthermore, we observed that, apart from hydrophobic residue pairs, thermophilic proteins prefer to form contacts between hydrophobic and polar residues. The Tyr residue prefers to have contacts with residues Asp, Ile, Lys, Ser and Val in thermophiles, while its contacts in mesophiles are fewer.

### 3.4.2. Long-range contacts

The residue pairs influenced by long-range contacts observed in thermophiles, but not in mesophiles, and vice versa, are presented in Table 2b. We found that the number of preferred long-range contacts is fewer in both mesophiles and

thermophiles compared with medium-range contacts. However, we derived the same conclusion, that the thermophiles have more residue pairs and a wider range of interactions with other residues than mesophiles, influenced by long-range contacts. Furthermore, significant contacts are observed between residues of H-bond-forming capability and between hydrophobic and polar residues.

The statistical significance of the residue pairs observed in enhancing the thermal stability of thermophilic proteins has been assessed using resampling techniques. Several sets of proteins belonging to different structural classes have been used to represent the mesophilic proteins, and the residue–residue contacts for each set were computed. The results showed that the thermophiles have more residue pairs than all sets of proteins, influenced by both medium- and long-range interactions. Furthermore, the residue pairs observed for enhanced thermostability are similar to those obtained from the comparison of mesophilic and thermophilic proteins in 16 different families. This result verifies the significance of the observation presented in the study.

Table 2  
Residue pairs influenced by inter-residue contacts

Mesophilic only	Thermophilic only
<i>(a) Medium-range contacts</i>	
D–M, D–S, E–F, E–H, E–T, E–Y, F–G, F–I, F–K, F–T, H–S, K–N, L–Q, L–S, M–R, N–P, N–Q, P–P, S–V	A–F, D–I, D–Q, D–Y, E–I, E–M, E–Q, E–R, F–M, F–P, F–Q, G–R, H–K, H–T, I–M, I–R, I–T, I–Y, K–Q, K–Y, L–M, L–N, L–P, N–T, Q–R, R–T, R–V, S–Y V–Y
<i>(b) Long-range contacts</i>	
D–Y, F–N, I–K, K–S, L–N, L–S, N–P, N–V, P–P, R–Y	A–R, D–P, E–R, E–T, F–K, F–P, G–Q, I–P, L–P, L–R, L–Y, N–Q, P–R, P–V, R–V, T–Y, Y–Y

### 3.5. Comparison with other studies

Vogt et al. [6] reported that an increase in hydrogen bonding and fractional polar surface increase the stability of thermophilic proteins. Xiao and Honig [25] found that electrostatic interactions are more favorable in thermophiles. Recently, Kumar et al. [10] showed that salt bridges and side-chain–side-chain hydrogen bonds increase the stability in most of the thermophilic proteins. In this work, we have delineated the preferred residue pairs for enhancing the thermal stability of thermophilic proteins. Furthermore, we found that, apart from hydrophobic contacts, thermophiles have an appreciable number of contacts between polar and non-polar residues compared to mesophiles. The higher occurrence of contacts between residues of H-bond-forming capability reflects the importance of H bonds in enhancing the stability of thermophilic proteins [6,10].

## 4. Conclusions

The analysis of inter-residue contacts in mesophilic and thermophilic proteins provides an insight into the stability of thermophilic proteins. We found that the thermophilic proteins have a wide variety of inter-residue contacts other than hydrophobic interactions. The higher occurrence of residue pairs between (i) residues of H-bond-forming capability and (ii) hydrophobic and polar

residues may enhance the stability. Furthermore, the residues in thermophilic proteins have an average of three–five long-range contacts, whereas two–four long-range contacts are observed in mesophilic proteins.

## Acknowledgements

The author wishes to thank Ms Bettyma for encouragement and Dr P. Prabakaran for critical reading of the manuscript.

## References

- [1] M.M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, A. Sarai, ProTherm: thermodynamic database for proteins and mutants, *Nucleic Acids Res.* 27 (1999) 286–288.
- [2] M.M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, P. Prabakaran, A. Sarai, ProTherm, version 2.0: thermodynamic database for proteins and mutants, *Nucleic Acids Res.* 28 (2000) 283–285.
- [3] R. Ladenstein, G. Antranikian, Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water, *Adv. Biochem. Eng. Biotechnol.* 61 (1998) 37–85.
- [4] R. Jaenicke, G. Bohm, The stability of proteins in extreme environments, *Curr. Opin. Struct. Biol.* 8 (1998) 738–748.
- [5] E. Querol, J.A. Perez-Pons, A. Mozo-Villarias, Analysis of protein conformational characteristics related to thermostability, *Protein Eng.* 9 (1996) 265–271.

- [6] G. Vogt, S. Woell, P. Argos, Protein thermal stability, hydrogen bonds, and ion pairs, *J. Mol. Biol.* 269 (1997) 631–643.
- [7] M.M. Gromiha, M. Oobatake, A. Sarai, Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins, *Biophys. Chem.* 82 (1999) 51–67.
- [8] J. Hasegawa, S. Uchiyama, Y. Tanimoto et al., Selected mutations in a mesophilic cytochrome c confer the stability of a thermophilic counterpart, *J. Biol. Chem.* 275 (2000) 37824–37828.
- [9] A.M. Facciano, G. Colonna, R. Ragone, Helix stabilizing factors and stabilization of thermophilic proteins: an X-ray-based study, *Protein Eng.* 11 (1998) 753–760.
- [10] S. Kumar, C.J. Tsai, R. Nussinov, Factors enhancing protein thermostability, *Protein Eng.* 13 (2000) 179–191.
- [11] A. Szilagy, P. Zavodszky, Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey, *Structure Fold. Des.* 8 (2000) 493–504.
- [12] J. Chen, Z. Lu, J. Sakon, W.E. Stites, Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability, *J. Mol. Biol.* 303 (2000) 125–130.
- [13] J. Hollien, S. Marquese, A thermodynamic comparison of mesophilic and thermophilic ribonuclease H, *Biochemistry* 38 (1999) 3831–3836.
- [14] M.M. Gromiha, Factors influencing the stability of  $\alpha$ -helices and  $\beta$ -strands in thermophilic ribonuclease H, *Prep. Biochem. Biotech.*, in press.
- [15] D. Perl, U. Mueller, U. Heinemann, F.X. Schmid, Two exposed amino acid residues confer thermostability on a cold shock protein, *Nat. Struct. Biol.* 7 (2000) 380–383.
- [16] M.M. Gromiha, S. Selvaraj, Importance of long-range interactions in protein folding, *Biophys. Chem.* 77 (1999) 49–68.
- [17] M.M. Gromiha, S. Selvaraj, Inter-residue interactions in the structure, folding and stability of proteins, *Recent Res. Dev. Biophys. Chem.* 1 (2000) 1–14.
- [18] 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.
- [19] 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.
- [20] M.M. Gromiha, M. Oobatake, H. Kono, H. Uedaira, A. Sarai, Importance of surrounding residues for protein stability of partially buried mutations, *J. Biomol. Struct. Dyn.* 18 (2000) 281–295.
- [21] H.M. Berman, J. Westbrook, Z. Feng et al., The Protein Data Bank, *Nucleic Acids Res.* 28 (2000) 235–242.
- [22] P.K. Ponnuswamy, Hydrophobic characteristics of folded proteins, *Prog. Biophys. Mol. Biol.* 59 (1993) 57–103.
- [23] 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.
- [24] S. Selvaraj, M.M. Gromiha, Importance of long-range interactions in  $(\alpha/\beta)_8$  barrel fold, *J. Protein Chem.* 17 (1998) 691–697.
- [25] L. Xiao, B. Honig, Electrostatic contributions to the stability of hyperthermophilic proteins, *J. Mol. Biol.* 289 (1999) 1435–1444.