

Protein thermal stability: insights from atomic displacement parameters (B values)

S.Parthasarathy and M.R.N.Murthy¹

Molecular Biophysics Unit, Indian Institute of Science,
Bangalore 560 012, India

¹To whom correspondence should be addressed.
Email: mnrn@mbu.iisc.ernet.in

The factors contributing to the thermal stability of proteins from thermophilic origins are matters of intense debate and investigation. Thermophilic proteins are thought to possess better packed interiors than their mesophilic counterparts, leading to lesser overall flexibility and a corresponding reduction in surface-to-volume ratio. These observations prompted an analysis of B values reported in high-resolution X-ray crystal structures of mesophilic and thermophilic proteins. In this analysis, the following aspects were addressed: (1) frequency distribution of normalized B values (B' factors) over all the proteins and for individual amino acids; (2) amino acid compositions in high B value regions of polypeptide chains; (3) variation in the B values from core to the surface of proteins in terms of their radius of gyration; and (4) degree of dispersion of normalized B values in spheres around the $C\alpha$ atoms. The analysis revealed that (1) Ser and Thr have lesser flexibility in thermophiles than in mesophiles, (2) the proportion of Glu and Lys in high B value regions of thermophiles is higher and that of Ser and Thr is lower and (3) the dispersion of B values within spheres at $C\alpha$ atoms is similar in mesophiles and thermophiles. These observations reflect plausible differences in the dynamics of thermophilic and mesophilic proteins and suggest amino acid substitutions that are likely to change thermal stability.

Keywords: protein dynamics/protein stability/temperature factors/thermophiles/X-ray structures

Introduction

Increasing the thermal stability of proteins is one of the goals of protein engineering studies. This goal is of commercial importance for industries where biocatalysts are used in extreme conditions to achieve higher solubility of substrates. It is, therefore, of primary importance to understand the factors that contribute to thermal stability. Knowledge regarding these factors has been accumulated from both experimental [mutational studies, especially on T4 lysozyme (Matthews, 1995)] and computational methods. Proteins from thermophilic origins function optimally at temperatures where most of their mesophilic counterparts will undergo denaturation. In spite of this dramatic increase in temperature optimum, sequential, structural, functional and chemical characteristics of thermophilic proteins are comparable to those of their respective mesophilic counterparts.

Various factors have been shown to contribute to the stability of proteins from thermophiles (Russell and Taylor, 1995; Querol *et al.*, 1996; Jaenicke and Bohm, 1998; Ladenstein and

Antranikian, 1998). Analysis of the amino acid composition of helices in thermophilic proteins appears to indicate that Tyr, Gly and Gln are enhanced whereas Val is suppressed compared with those of mesophilic proteins (Warren and Petsko, 1995). It has also been suggested that Lys→Arg and Ser→Ala are the most frequent mutations in mesophilic to thermophilic substitutions (Arias and Argos, 1989). The importance of electrostatic interactions (Goldman, 1995; Hennig *et al.*, 1995; Yip *et al.*, 1995; Xiao and Honig, 1999), increased compactness, shortening of loops, increased hydrophobicity and decreased flexibility of α -helical segments and subunit interfaces (Kelly *et al.*, 1993; Russell *et al.*, 1997) have been proposed as important factors conferring thermal stability. In the case of Che Y protein from *Thermotoga maritima*, thermal stability appears to be achieved by factors leading to the lowering of the entropy of unfolding (Usher *et al.*, 1998). Analysis of complete genome sequences has suggested loop deletion as a mechanism for thermal stability (Thompson and Eisenberg, 1999). Surface and volume analysis has indicated that proteins from mesophilic and thermophilic origins cannot be distinguished in terms of packing criteria (Karshikoff and Ladenstein, 1998). All these studies suggest that in thermophilic proteins stability is achieved through cooperative optimization of several subtle factors rather than any one predominant interaction.

The atomic displacement parameters (B values) determined by high-resolution X-ray crystallographic studies represent smearing of atomic electron densities around their equilibrium positions due to thermal motion and positional disorder. Analysis of B values, therefore, is likely to provide newer insights into protein dynamics, flexibility of amino acids and protein stability. Molecular dynamics studies have suggested that protein unfolding might be initiated at sites that are prone to large thermal fluctuations (Daggett and Levitt, 1992; Lazaridis *et al.*, 1997). Therefore, the pattern of B values determined by high-resolution X-ray crystallographic studies might contain information regarding protein stability. The correlation between experimentally observed B values and stability, unlike the contributions of various other interactions, has not been examined in any great detail.

The distribution of B values in high-resolution crystal structures has been shown to fit accurately the sum of two Gaussian functions (Parthasarathy and Murthy, 1997, 1998). Flexibility indices of individual amino acids derived from the fitted curve reflect the dynamics of the respective amino acids. Examination of the correlation between average main-chain and side-chain B values reveals the effect of restraints used by the crystallographers for the refinement of B values and has brought out the need to have better restraints on B values (Parthasarathy and Murthy, 1999). It has also been demonstrated that the distribution of B values reflects the special dynamic properties associated with some proteins and could possibly be used as a validation tool. In this paper, we report the analysis of B values obtained from the crystal

structures of thermophilic proteins. The degrees of dispersion in the B' factors (normalized B values) associated with atoms in spheres placed at each $C\alpha$ atomic position in mesophilic and thermophilic proteins are comparable. Similarly, the variation of B values from the centroid towards the surface, which is likely to depend on packing density, does not appear to be significantly different in the two sets of proteins. Although the overall frequency distribution is similar, the distributions for some amino acids, especially for Ser and Thr, are different, reflecting the role played by these residues in imparting stability to thermophilic proteins. Examination of regions of high temperature factors shows that the compositions of some residues are significantly different between thermophiles and mesophiles. These observations might be related to the role played by some key residues in imparting thermal stability.

Methods

Selection of protein structures

Ninety-three mesophilic protein coordinates used in this analysis were chosen from the representative list (Hobohm and Sander, 1994) of PDB (Bernstein *et al.*, 1977) entries released in November 1996. These structures have resolutions better than 2.0 Å and R factors <0.2 (Table I). Twenty-one thermophilic structures with resolution better than 2.5 Å were used for the analysis (Table I). Between any two of the mesophilic structures the maximum sequence similarity was 25%, while no sequence similarity criterion was applied for thermophilic structures. However, except for two structures, the thermophilic data are also non-redundant (Karshikoff and Ladenstein, 1998).

Frequency distribution of B factors

The B values at $C\alpha$ atoms of each selected protein were replaced by normalized B' factors defined as, $B' = (B - \langle B \rangle) / \sigma(B)$, where $\langle B \rangle$ is the mean B value at $C\alpha$ atoms and $\sigma(B)$ is their standard deviation. Frequencies of residues in 0.5 unit ranges in B' factors were counted. Various frequency distributions for individual amino acids and individual proteins and the overall distribution were counted and fitted analytically by least-squares minimization to the sum of

two Gaussian functions (Parthasarathy and Murthy, 1997, 1998), $f = k_1 \exp[-k_2(B'-B_1)^2] + k_3 \exp[-k_4(B'-B_2)^2]$, where k_1, k_2, k_3, k_4, B_1 and B_2 are parameters defining the two Gaussians. The constants for the second Gaussian were determined after fitting the first Gaussian to stabilize the minimization. The areas under the two Gaussians, A_1 and A_2 , are given by $A_1 = k_1(\pi/k_2)^{1/2}$ and $A_2 = k_3(\pi/k_4)^{1/2}$. The fractional areas under the second Gaussian, $p = A_2/(A_1 + A_2)$, for individual amino acids, representing the propensity to occur with high B value, were calculated.

Distribution of high B values

Amino acids with B values greater than $\langle B \rangle + 0.5\sigma(B)$ were considered as high B value residues. The amino acid compositions of residues with high B values in mesophilic and thermophilic proteins were determined and compared. The numbers of stretches of consecutive high B value residues of length 1–5 were counted. A similar analysis was also performed with a high B value threshold of $\langle B \rangle + 0.75\sigma(B)$.

Dispersion of B values in spheres placed at $C\alpha$ positions

Spheres of radius 5.0 and 7.5 Å were placed at each $C\alpha$ atom. The relatedness of the B values of atoms in these spheres was analysed by calculating the r.m.s. deviation of their B' factors. The frequency distributions of these r.m.s. values were determined. Also, a plausible correlation that might exist between the mean B' factors in these spheres and the corresponding atomic packing was examined.

Variation of B values with distance from centroid

For each protein, the radius of gyration was calculated as $R_g = (\sum r_i^2/n)^{1/2}$, where r_i and r_c represent the positions of $C\alpha$ atoms and the centroid of the molecule, respectively. Mean B' factors in spherical shells of radius expressed in terms of R_g were computed for thermophilic and mesophilic proteins.

Results

B' factor frequency distribution

Figure 1 shows the overall frequency distribution of B' factors for mesophilic and thermophilic proteins. The plots represent curves fitted as the sum of two Gaussian functions and

Table I. PDB codes for the high-resolution mesophilic and thermophilic structures selected for the analysis

Structure	PDB codes						
Mesophilic	131L	153L	1AMP	1ARB	1ARV	1ATL	
	1BAM	1BP2	1CCR	1CHD	1CHM	1CNS	
	1CSE	1CSH	1DAA	1DTS	1DUP	1DYR	
	1EDE	1FKJ	1FNC	1GOF	1GPR	1HMT	
	1IAE	1ISC	1KPT	1LCP	1LEN	1LTS	
	1MOL	1NFP	1NHK	1NAR	1OVA	1PBE	
	1PDA	1PGS	1PHG	1PTX	1REC	1REG	
	1RSY	1SAT	1SBP	1SNC	1SRI	1TAG	
	1TAH	1TCA	1THV	1TRY	1TTB	1TYS	
	1XNB	2ACQ	2ALP	2AYH	2AZA	2CBA	
	2CCY	2CDV	2CPL	2CTC	2END	2ER7	
	2GST	2HMZ	2HTS	2MNR	2NAC	2OLB	
	2PHY	2POR	2PRK	2SIL	2TGL	3CHY	
	3CLA	3COX	3DFR	3GRS	3PTE	3SIC	
	3TGI	4ENL	4GCR	4FGF	5RUB	5TIM	
	8ABP	8FAB	9RNT				
	Thermophilic	1AIS	1AJ8	1AOR	1BMD	1CAA	1CIU
		1GD1	1GTM	1IGS	1NOX	1NSJ	1PHP
		1RIS	1THM	1XGS	1XYZ	1ZIP	2PRD
			3MDS	3PFK	8TLN		

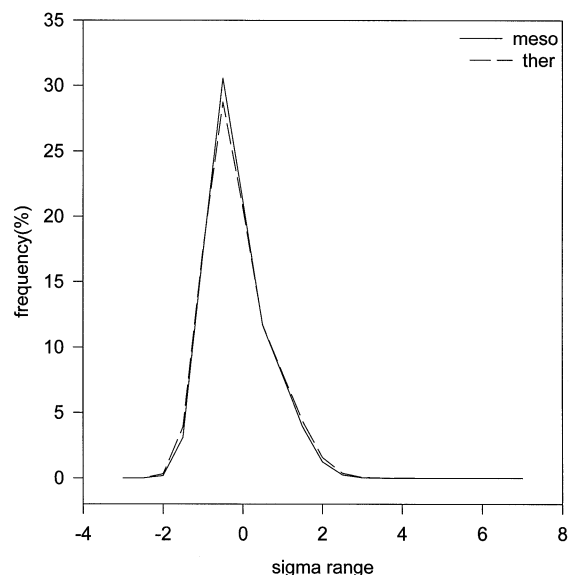


Fig. 1. Frequency distribution of B' factors for mesophilic and thermophilic protein structures used in this analysis

correspond to 30 960 amino acids for mesophiles and 10 469 amino acids for thermophiles. The six parameters characterizing the double Gaussian function are very similar for the two curves. The fractional areas under the second Gaussian, the *p*-values, are 0.357 and 0.361 for mesophilic and thermophilic proteins, respectively. Table II gives the *p*-values for individual amino acids. It can be seen that Glu, Leu, Tyr and Gln have higher *p*-values in thermophiles than mesophiles whereas Cys, Asn, Pro, Arg and Ser have lower *p*-values.

Length distribution of high B value stretches

The frequency of occurrence of consecutive high *B* values [defined as $\langle B \rangle + 0.5\sigma(B)$], stretches of length 1–5, were determined for mesophilic and thermophilic proteins. Table III lists the relevant statistics. It can be seen that there is no substantial difference in terms of the length distribution or frequency of amino acids found in these stretches between thermophiles and mesophiles. This observation suggests that the thermophilic and mesophilic proteins do not differ in the occurrence of segments of high or low mobilities.

Amino acid composition of residues with high B values

Figure 2a shows the scatter plot of overall amino acid composition in mesophiles and thermophiles. Table IV gives the corresponding statistics. Figure 2b and c show scatter plots of high *B* value residues. As can be seen in Figure 2a and Table IV, the overall amino acid compositions are very similar in mesophiles and thermophiles (correlation coefficient 0.89).

Table II. *p*-Values of amino acids for a bin size of 0.5

Residue	Mesophilic	Thermophilic
All	0.357	0.361
Ala	0.343	0.320
Cys	0.319	0.198
Asp	0.429	0.424
Glu	0.405	0.514
Phe	0.292	0.314
Gly	0.389	0.374
His	0.307	0.299
Ile	0.296	0.306
Lys	0.429	0.446
Leu	0.287	0.340
Met	0.293	0.313
Asn	0.409	0.384
Pro	0.432	0.354
Gln	0.395	0.436
Arg	0.353	0.327
Ser	0.416	0.376
Thr	0.362	0.339
Val	0.307	0.294
Trp	0.268	0.291
Tyr	0.227	0.287

The average correlation coefficient, however, between compositions of high *B* value residues in mesophiles and thermophiles is 0.77. The residues Glu, Lys, Ser and Thr are outliers in these plots (Figure 2b and c and Table IV). Notably, the percentage Glu residues in high *B* value regions is nearly twice and that of Lys is nearly 1.5 times higher in thermophiles than mesophiles. In contrast, the percentages of Ser and Thr in high *B* value regions of thermophiles are decreased by half (Table IV). These are also related to larger *p*-values (Table II) for Glu, Lys and smaller *p*-values for Ser and Thr in thermophiles. Similar observations were made when the *B* values of whole residues instead of those associated with C α alone were examined (data not shown).

Dispersion of B values within spheres at C α atoms

It is expected that in the well packed interior of proteins, atoms in close proximity will have correlated displacements, which will be reflected in the *B* values. If thermophiles have better packed interiors than mesophiles, atoms in spherical volumes around a large number of C α positions will have low

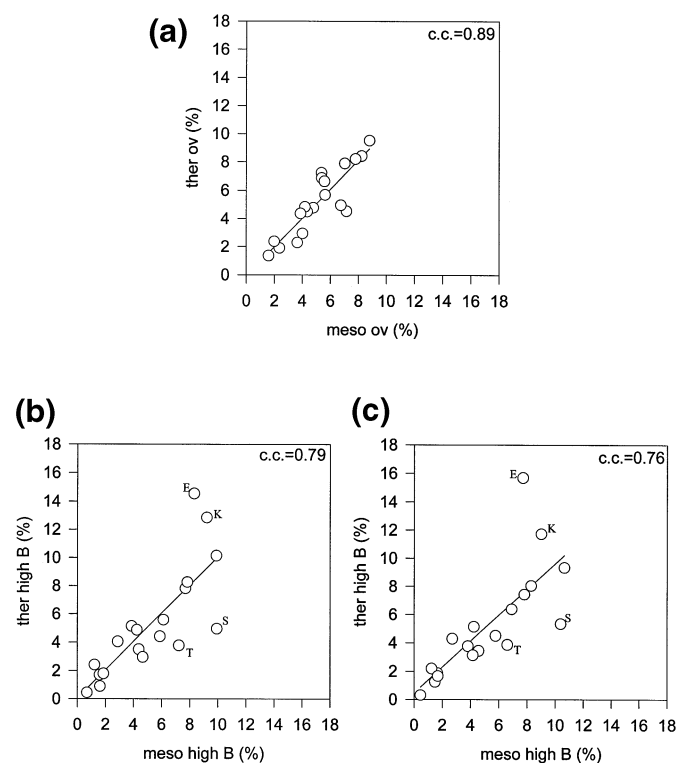


Fig. 2. Scatter plots showing amino acid compositions in mesophiles and in thermophiles. (a) Overall composition; (b) composition of residues with $B > \langle B \rangle + 0.5\sigma(B)$; (c) composition of residues with $B > \langle B \rangle + 0.75\sigma(B)$. Glu, Lys, Ser and Thr are marked in high *B* value plots.

Table III. Frequency of occurrence of consecutive high *B* value stretches of length 1–5

Order	No. of stretches	No. of aa in this group	% of segments	% of aa residues	No. of stretches	No. of aa in this group	% of segments	% of aa residues
1	971	971	53.03	27.44	290	290	50.61	26.03
2	390	780	21.30	22.04	131	262	22.86	23.52
3	202	606	11.03	17.12	75	225	13.09	20.20
4	158	632	8.63	17.86	48	192	8.38	17.24
5	110	550	6.01	15.54	29	145	5.06	13.02

Column 2 gives frequency; column 3 lists the number of residues in these stretches; columns 4 and 5 express columns 2 and 3 as percentages; columns 6–9 represent the corresponding statistics for thermophiles; aa = amino acid(s).

Table IV. Amino acid composition of residues with high B values

	1	2	3	4	5	6	7	8	9	10	11	12
Ala	2723	8.80	997	9.52	272	7.69	87	7.81	227	7.80	71	7.43
Cys	439	1.42	67	0.64	33	0.93	0	–	27	0.93	0	–
Asp	1740	5.62	596	5.69	277	7.83	92	8.26	241	8.28	77	8.05
Glu	1659	5.36	763	7.23	294	8.31	162	14.54	224	7.70	150	15.69
Phe	1243	4.01	308	2.94	56	1.58	19	1.71	48	1.65	18	1.88
Gly	2552	8.24	881	8.42	350	9.89	113	10.14	310	10.65	95	9.34
His	732	2.36	199	1.90	57	1.61	10	0.90	43	1.48	12	1.26
Ile	1666	5.38	719	6.87	102	2.88	45	4.04	79	2.71	41	4.29
Lys	1728	5.58	695	6.64	326	9.21	143	12.84	262	9.00	112	11.72
Leu	2416	7.80	860	8.21	137	3.87	57	5.12	123	4.23	49	5.13
Met	617	1.99	248	2.37	43	1.22	27	2.42	36	1.24	21	2.20
Asn	1480	4.78	500	4.78	217	6.13	62	5.57	201	6.91	61	6.38
Pro	1373	4.34	471	4.50	208	5.88	49	4.40	168	5.77	43	4.50
Gln	1127	3.64	240	2.29	155	4.38	39	3.50	133	4.57	33	3.45
Arg	1293	4.18	507	4.84	150	4.24	54	4.85	111	3.81	36	3.77
Ser	2215	7.15	474	4.53	351	9.92	55	4.94	302	10.38	51	5.33
Thr	2091	6.75	519	4.96	256	7.23	42	3.77	192	6.60	37	3.87
Val	2176	7.03	826	7.89	165	4.66	33	2.96	121	4.16	30	3.14
Trp	491	1.59	141	1.35	24	0.68	5	0.45	13	0.45	3	0.31
Tyr	1199	3.87	458	4.37	66	1.86	20	1.80	49	1.68	16	1.67

The first four columns give the numbers of residues and their composition (%) in mesophiles and thermophiles, respectively. Columns 5 and 6 represent the number and composition (%) of residues with $B > \langle B \rangle + 0.5\sigma(B)$ in mesophiles. Columns 9 and 10 represent these numbers for residues with $B > \langle B \rangle + 0.75\sigma(B)$. Columns 7, 8, 11 and 12 give the values corresponding to columns 5, 6, 9 and 10 for thermophiles. Large differences between mesophiles and thermophiles are highlighted in bold type.

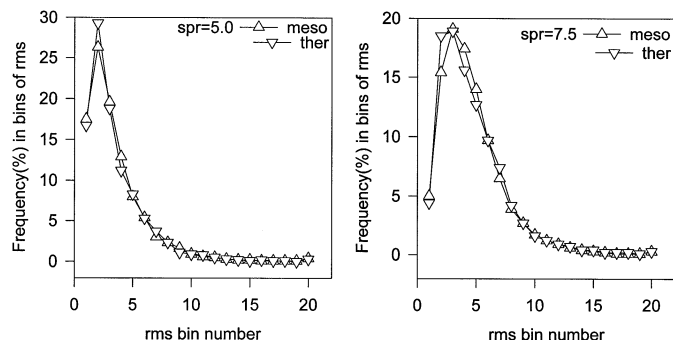


Fig. 3. Frequency distribution of r.m.s. of B' factors within spheres placed at $C\alpha$ atomic positions for mesophilic and thermophilic proteins (a) for sphere radius 5.0 Å; (b) for sphere radius 7.5 Å.

r.m.s. B values in thermophiles than mesophiles. Figure 3a and b show the frequency distribution of r.m.s. B' factors in these spherical regions for mesophiles and thermophiles, for radii of 5.0 and 7.5 Å, respectively. There is no significant difference in peak position or frequency in each bin between mesophilic and thermophilic proteins. This agrees with an earlier report that mesophilic and thermophilic proteins do not differ significantly with respect to packing interactions (Karshikoff and Ladenstein, 1998).

B' factor distribution in spherical shells around the molecular centre

It is known that B values tend to increase continuously from the core of the protein to its surface (Bhaskaran and Ponnuswamy, 1988). It is of interest, therefore, to compare the increment in B values from the core of the protein to the exterior between mesophiles and thermophiles. Hence the increment in B' factors in spherical shells of radius corresponding to a specified fraction of the radius of gyration of the molecule was examined. If the interiors of thermophiles are better packed than those of mesophiles, the slope of this

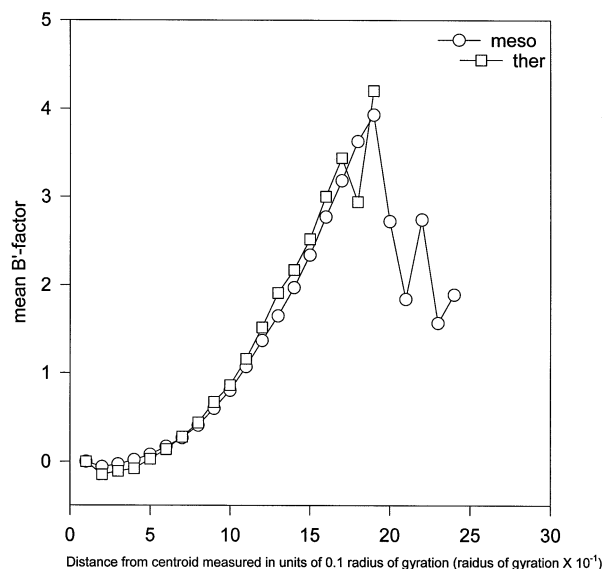


Fig. 4. Plot of mean B' factor in bins of fractions of radius of gyration for mesophilic and thermophilic proteins.

incremental curve will be smaller for thermophiles (Figure 4). It is clear from the plot that the two curves are almost identical except for fluctuations observed in shells near the outer surface of mesophiles. This fluctuation at outer shells shown by the mesophiles may be due to loops extending out from the rest of the protein.

Discussion

Increasing the thermal stability of proteins has been one of the primary goals of protein engineering. Sequence comparisons between proteins from mesophilic and thermophilic origins have been performed in order to gain information about the possible mutations leading to increased thermal stability. Based

on such studies, 'traffic rules' have been discussed by several researchers. Lys→Arg and Ser→Ala appear to be the top two amino acid substitutions (in helical segments) from mesophilic to thermophilic protein sequences (Arias and Argos, 1989). Mutations of Ser to Ala and Thr to Ala in mesophilic lactate dehydrogenase have been reported to enhance the stability of the mutant by 20°C or more compared with the wild-type enzyme (Kotik and Zuber, 1993). Bohm and Jaenike (1994), however, have concluded that the proposed 'traffic rules' of molecular adaptation cannot be used to predict extremophilic behaviour. An overall increase in conformational rigidity is also a factor contributing to the adaptation of proteins to extreme environments. It is likely that the dynamic features of thermophilic and mesophilic proteins have significant and recognizable differences. The degree to which such differences, if they exist, could be inferred from critical examination of crystallographic refinement of high-resolution structures needs to be evaluated. The preliminary analysis presented here does indeed suggest that the crystallographically determined atomic displacement parameters might actually convey information regarding thermal stability and highlight the differences between the dynamics of thermophilic and mesophilic proteins.

Better hydrogen bonding networks, electrostatic interactions and better internal packing have been suggested by various investigators as factors contributing to the stability of thermophilic proteins (Russell and Taylor, 1995; Querol *et al.*, 1996; Jaenike and Bohm, 1998; Ladenstein and Antranikian, 1998). Karshikoff and Ladenstein (1998) have analysed the packing density in mesophilic and thermophilic proteins and concluded that packing density is not a dominant factor contributing to the thermal stability. However, contradictory results were obtained when glyceraldehyde-3-phosphate dehydrogenase and glutamate dehydrogenase from mesophilic and thermophilic organisms were compared for their accessible surface area (Korndorfer *et al.*, 1997; Knapp *et al.*, 1997). Analyses of variations in B' factors around spheres at $C\alpha$ atoms and increment of B' factors from the core to the surface of proteins, presented here, show that thermophilic proteins are not very different from mesophilic proteins in these aspects. Packing differences, if any, are thus not reflected in the atomic displacement parameters.

The most significant observation in the present analysis is that Glu and Lys are enhanced whereas Ser and Thr are suppressed in high B value regions of thermophiles in comparison with mesophiles. The juxtaposition of these four residues is perhaps important in imparting thermal stability. These residues may be suitable candidates for site-specific mutations leading to enhanced stability. This also suggests that mutation of high B value Ser and Thr could lead to an improvement of thermal stability. The mutational experiments on lactate dehydrogenase by Kotik and Zuber (1993) and streptococcal protein G β_1 domain (Malakauskas and Mayo, 1998) involving Ser and Thr residues have led to considerable increases in thermal stability, suggesting that the conclusions drawn from the present analysis are likely to be significant. These results also suggest that the 'traffic rules' of amino acid replacements need to be revised with reference to amino acid flexibility.

Acknowledgements

This work was supported by a grant from the Council of Scientific and Industrial Research, India, to M.R.N.M. S.P. thanks the same agency for a research fellowship.

References

- Arias,L.M. and Argos,P. (1989) *J. Mol. Biol.*, **206**, 397–406.
 Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1997) *J. Mol. Biol.*, **112**, 535–542.
 Bhaskaran,R. and Ponnuswamy,P.K. (1988) *Int. J. Pept. Protein Res.*, **32**, 241–255.
 Bohm,G. and Jaenike,R. (1994) *Int. J. Pept. Protein Res.*, **43**, 97–106.
 Daggett,V. and Levitt,M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5142–5146.
 Goldman,A. (1995) *Structure*, **3**, 1277–1279.
 Hennig,M., Darimont,B., Sterner,R., Kirschner,K. and Jansonius,J.N. (1995) *Structure*, **3**, 1295–1306.
 Hobohm,U. and Sander,C. (1994) *Protein Sci.*, **3**, 522–524.
 Jaenike,R. and Bohm,G. (1998) *Curr. Opin. Struct. Biol.*, **8**, 738–748.
 Karshikoff,A. and Ladenstein,R. (1998) *Protein Engng*, **11**, 867–872.
 Knapp,S., de Vos,W.M., Rice,D. and Ladenstein,R. (1997) *J. Mol. Biol.*, **267**, 916–932.
 Kelly,C.A., Nishiyama,M., Ohnishi,Y., Beppu,T. and Birkofit,J.J. (1993) *Biochemistry*, **32**, 3913–3922.
 Korndorfer,I., Steipe,B., Huber,R., Tomschy,A. and Jaenike,R. (1997) *J. Mol. Biol.*, **246**, 511–521.
 Kotik,M. and Zuber,H. (1993) *Eur. J. Biochem.*, **211**, 267–280.
 Ladenstein,R. and Antranikian,G. (1998) *Adv. Biochem. Engng Biotechnol.*, **61**, 37–85.
 Lazaridis,T., Lee,I. and Karplus,M. (1997) *Protein Sci.*, **6**, 2589–2605.
 Malakauskas,S.M. and Mayo,S.L. (1998) *Nature Struct. Biol.*, **5**, 470–475.
 Matthews,B.W. (1995) *Adv. Protein Chem.*, **46**, 249–278.
 Parthasarathy,S. and Murthy,M.R.N. (1997) *Protein Sci.*, **6**, 2561–2567.
 Parthasarathy,S. and Murthy,M.R.N. (1998) *Protein Sci.*, **7**, 525.
 Parthasarathy,S. and Murthy,M.R.N. (1999) *Acta Crystallogr.*, **D55**, 173–180.
 Querol,E., Perez-Pons,J.A. and Villarias,A.M. (1996) *Protein Engng*, **9**, 265–271.
 Russell,R.J.M. and Taylor,G.L. (1995) *Curr. Opin. Biotechnol.*, **6**, 370–374.
 Russell,R.J.M., Ferguson,J.M.C., Hough,D.W., Danson,M.J. and Taylor,G.L. (1997) *Biochemistry*, **36**, 9983–9994.
 Thompson,M.J. and Eisenberg,D. (1999) *J. Mol. Biol.*, **290**, 595–604.
 Usher,K., De la Cruz,A., Dahlquist,F., Swanson,R., Simon,M. and Remington,S. (1998) *Protein Sci.*, **7**, 403–412.
 Warren,G.L. and Petsko,G.A. (1995) *Protein Engng*, **9**, 905–913.
 Xiao,L. and Honig,B. (1999) *J. Mol. Biol.*, **289**, 1435–1444.
 Yip,K.S.P., Stillman,T.J., Britton,K.L., Artymuick,P.J., Baker,P.J., Sedelniova,S.E., Engel,P.C., Pasquo,A., Chiaraluce,R., Consavi,V., Scandurra,R. and Rice,D.W. (1995) *Structure*, **3**, 1147–1158.

Received April 23, 1999; revised September 12, 1999; accepted October 1, 1999