Eur Biophys J (2001) 30: 378–384
DOI 10.1007/s002490100161

BIOPHYSICS LETTER

Akira R. Kinjo · Ken Nishikawa

Comparison of energy components of proteins from thermophilic and mesophilic organisms

Received: 13 December 2000 / Revised version: 10 April 2001 / Accepted: 12 April 2001 / Published online: 7 June 2001

Abstract In order to infer the energetic determinants of thermophilic proteins, molecular mechanics calculations were applied to five proteins from thermophilic eubacteria and their mesophilic homologs. The energy function includes a hydration term as well as the electrostatic contribution from the solvent in addition to the usual conformational energy terms. We calculated energy values for three different states of each protein: the native, near-native, and unfolded structures. The energy difference and its components between pairs of these states were compared. The hypothetical near-native structures have almost the same backbone conformation as the native structure but with largely distorted side-chain packing, thus enabling us to extract the energy components important for stabilizing the native backbone topology itself, irrespective of structural details. It was found that the sum of the electrostatic and hydration energies, although of large positive values, were consistently lower for the thermophilic proteins than for their mesophilic counterparts. This trend was observed in the energy difference not only between the native and unfolded structures, but also between the near-native and unfolded structures. In contrast, the energy components regarding side-chain packing did not show any clear tendency. These results suggest that the thermophilic proteins are stabilized so that the precise packing of the native structure does not significantly affect the stability. Implications of this conclusion are also discussed.

Keywords Molecular mechanics · Electrostatic interaction · Hydration · Packing · Thermostability

Introduction

The stabilization mechanisms of proteins from thermophilic organisms have attracted interest for many years. Recently, the accumulation of both sequence and structural data of thermophilic proteins as well as their mesophilic homologs has made it possible to study structural properties that may stabilize the thermophilic proteins (e.g., Karshikoff and Ladenstein 1998; Kumar et al. 2000; Szilágyi and Závodszky 2000). Although comprehensive, most of these studies are limited to statistical analysis of some structural parameters such as amino acid composition, the number of ion pairs, the number of cavities, etc. Thus, in these studies, the energetic bases of the thermostability were not directly addressed. However, since early statistical studies indicated the importance of ion pairs in some thermophilic proteins, the electrostatics of thermophilic proteins were studied intensively (e.g., Elcock1998; Xiao and Honig 1999). These studies confirmedthe importance of the electrostatic interactions in thermophilic proteins, although other energetic factors remained to be clarified. Lazaridis et al. (1997) investigated the thermostability of the rubredoxin from the hyperthermophilic archaeon *Pyrococcus furiosus* and the one from mesophilic *Desulfovibrio vulgaris* by unfolding molecular dynamics simulation at various temperatures. Although the computational complexity of the molecular dynamics simulation inhibited extensive analysisto draw definitive conclusions, some interesting insights were obtained regarding the energetic factors and the rigidity of the hyperthermophilic protein (Lazaridis et al. 1997).

In the present article, we study various energycomponents of thermophilic and mesophilic proteins by the

A.R. Kinjo[1] · K. Nishikawa (✉)
Center for Information Biology,
National Institute of Genetics,
Mishima, Shizuoka 411-8540, Japan
E-mail: knishika@genes.nig.ac.jp
Tel.: +81-559-816859
Fax: +81-559-816889

A.R. Kinjo · K. Nishikawa
Department of Genetics, The Graduate University
for Advanced Studies, Mishima, Shizuoka 411-8540, Japan

Present address:
[1]Department of Chemistry, Faculty of Science, Kobe University,
1-1 Rokkodai, Nada, Kobe 657-8501, Japan

molecular mechanics method. In order to find the factors that are common to different thermophilic proteins, we have employed a computationally more feasible method than Lazaridis et al. (1997), and have investigated five protein families of reasonable sizes. Basically, we have analyzed the energy difference between the native and unfolded structures of proteins. In addition, we also calculated the energy of artificially constructed near-native structures (Kinjo et al. 2001). The near-native structures are so called because they have almost the same backbone conformation as the native one, but their side-chain packing is significantly distorted. From our previous study, the near-native structures are shown to be useful to extract the dominant factors that stabilize the native fold (i.e., backbone topology), irrespective of the detailed side-chain packing. To summarize our previous results, the sum of local geometry and van der Waals terms is the dominant factor that stabilizes the native structure, which is defined by both the backbone topology and precise side-chain packing, whereas the electrostatic and hydration terms are the factors that stabilize the native topology or fold relatively independent of the details of the side-chain packing (Kinjo et al. 2001). In the present study, we divide the energy difference between the native and unfolded structures into two parts: one between the native and near-native, the other between the near-native and unfolded structures. As will be shown later, this strategy made it possible to understand more clearly the stabilization mechanism of thermophilic proteins.

From statistical analyses, it is suggested that hyperthermophilic proteins from archaea are stabilized by different mechanisms than thermophilic eubacteria (Szilágyi and Závodszky 2000). Moreover, the distant evolutionary relationship between archaea andother organisms would make the comparison difficult. Therefore, in the present study we limit our analysis to the proteins from thermophilic eubacteria and their homologs from mesophilic organisms.

Materials and methods

Preparation of structures

All the protein structures were obtained from the Protein Data Bank (PDB; Berman et al. 2000). We selected proteins from thermophilic eubacteria and their mesophilic homologs that were solved by X-ray crystallography with 2.5 Å or better resolution, and that were composed of less than 200 residues. Proteins from thermophilic eubacteria were selected by database search with the keyword "THERM" in the SOURCE field of the PDB file. The thermophilic proteins thus found were used as queries for the BLAST (Altschul et al. 1990) search against the PDB sequence database. Only those targets whose appropriate mesophilic homologs existed in the PDB were retained. After this, we had five thermophilic proteins (Table 1) and their mesophilic counterparts (Table 2). Note that, in the present study, thermophilic proteins from archaea are not treated, but all the target thermophilic proteins are from eubacteria.

For the analysis of the energy components, we calculated energies for three states of each protein: the native, near-native, and unfolded structures. The native structure was obtained from the X-ray structure and its energy was minimized by 500 steps of the conjugate gradient method with the distance geometry force field followed by 500 steps of the conjugate gradient method with the AMBER force field (Weiner et al. 1986) without the 1–5 electrostatic term to remove close contacts. The distance geometry force field includes only the local geometry terms and non-local soft repulsion terms but no attractive term (Nakai et al. 1993). The positional restraints on the backbone atoms were imposed so that

Table 1 Target proteins from thermophilic eubacteria

PDB[a]	Name	Species	T_{opt}[b]	Resol[c]
1cz3A	Dihydrofolate reductase	*Thermotoga maritima*	80	2.10
1dz3A	Stage 0 sporulation protein A	*Bacillus stearothermophilus*	52.5	1.65
1srvA	GroEL apical domain	*Thermus thermophilus*	75	1.70
1tmy	CheY protein	*Thermotoga maritima*	80	1.9
2prd	Pyrophosphate phosphohydrolase	*Thermus thermophilus*	75	2.0

[a] PDB codes and chain identifier
[b] Optimum growth temperature in °C according to Szilágyi and Závodszky (2000)
[c] Resolution of the crystal structure in Å

Table 2 Homologs of the target proteins

Thermophile[a]	Mesophile[a]	Species	ID (%)[b]	Resol[c]
1cz3A (164)	1aoeA (192)	*Candida albicans*	52.7	1.60
	1dyjA (159)	*Escherichia coli*	53.5	1.85
1dz3A (123)	1srrA (119)	*Bacillus subtilis*	64.6	1.9
1srvA (145)	1kid (193)	*Escherichia coli*	82.6	1.7
1tmy (119)	2chf (128)	*Salmonella typhimurium*	60.7	1.8
	3chy (128)	*Escherichia coli*	59.5	1.66
2prd (174)	1obwA (175)	*Escherichia coli*	66.5	1.9

[a] PDB code with chain identifier. Number of residues in parentheses
[b] Percent sequence identity with the thermophile
[c] Resolution of the crystal structure in Å

the minimized structure did not deviate much from the experimental coordinates.

Near-native structures introduced in this study have almost the same backbone conformations as the native structure, but their side-chain conformations are significantly distorted. Although the near-native structures are only hypothetical ones, they have been shown to be useful to extract the energetic factors important for the formation of the approximate native fold of a protein irrespective of the details of side-chain packing (Kinjo et al. 2001). To generate the near-native structure, first a random coil was generated which was minimized for 500 steps of the conjugate gradient method. Then the random coil was subject to 55,000 steps of a simulated annealing molecular dynamics in four-dimensional space using the computer program EMBOSS (Nakai et al. 1993). The temperature was set to 500 K for the first 5000 steps, then cooled exponentially to 1 K. Next, 3000 steps of conjugate gradient minimization was applied with an increased weight for the fourth dimensional energy to compress the fourth dimension and to obtain the three-dimensional structure. Finally, 500 steps of conjugate gradient minimization with the AMBER force field without the 1–5 electrostatic energy term completed the generation of a near-native structure. Except for the final stage, the distance geometry force field was used. Throughout the stages, the positional restraints were imposed on the backbone atoms so that the near-native model converges to the backbone conformation of the native structure. Since the main optimization stages employ the distance geometry force field which does not include any attractive term, side-chain conformations are determined solely by steric hindrance. Other characteristics of the near-native structure are presented in our previous paper (Kinjo et al. 2001). Twenty near-native structures were generated with different initial conditions (random coils and initial velocities) for each protein and the average energy and energy components were used for the analysis given below.

In order to compare the energetic properties of several families of proteins at once, we employed a rough approximation of the denatured state of the proteins, that is, we regard a set of unfolded structures as the denatured state. The unfolded structures were generated by almost the same procedure as the near-native structures. The differences are that the starting structure was the native structure, that the simulated annealing molecular dynamics was performed in the three-dimensional space, that the minimization following the simulated annealing lasted only for 1000 steps, and that the restraints were imposed so that the structure deviated from the native one by the root mean square deviation (RMSD) between 8 Å and 12 Å (Ferrara et al. 2000). The unfolded structure thus constructed shows a global chain topology somewhat similar to the native structure, but is significantly expanded and contains no residual secondary structures (data not shown). Again, 20 structures were generated with different initial velocities for each protein, and the average energy and its components were used for the analysis below.

When the optimization involved the distance geometry force field, the cutoff distance of non-local interactions was set to 6 Å, while in the minimization with the AMBER force field it was set to 12 Å. In the final evaluation of the total energy, no distance cutoff was applied (see the next section).

Results

Energy function and decomposition of energy difference

The energy function for the final evaluation of structures is composed of the AMBER all-atom force field (Weiner et al. 1986) together with the hydration term of Ooi et al. (1987) and the electrostatic contribution from the solvent. In this calculation, no distance cutoff was applied. The hydration term of Ooi et al.(1987) is based on the accessible surface area of protein atoms which was calculated by the analytical method of Richmond (1984). The electrostatic contribution from the solvent (reaction field) was calculated based on a continuum dielectric model of the protein-solvent system (Nakamura and Nishida 1987; Nakamura 1996). The dielectric constants were set to 4 in the protein region, and to 80 for the solvent and boundary regions. The ionic strength was set to zero. The grid size of the system was set to 1 Å. With these conditions, the Poisson equation was numerically solved to obtain the reaction field energy of the continuum protein-solvent system. Other details of the calculation procedure are given in our previous paper (Kinjo et al. 2001).

In order to make the comparison of proteins of different sizes easier, the energy values were normalized by the molecular weight of each protein. Hence the energy unit cal/g is used instead of more often used kcal/mol.

The calculated energy values were analyzed as follows. Let E_{n} be the energy of the native structure of a protein, and E_{u} be the average energy of the 20 unfolded structures of the same protein. The first quantity we investigate is the energy change between these states, $\Delta E_{\mathrm{n-u}} = E_{\mathrm{n}} - E_{\mathrm{u}}$, which corresponds to the enthalpy change during the folding transition. Next, we define E_{m} as the average energy of the 20 near-native structures of the protein of interest, and $\Delta E_{\mathrm{n-m}} = E_{\mathrm{n}} - E_{\mathrm{m}}$, $\Delta E_{\mathrm{m-u}} = E_{\mathrm{m}} - E_{\mathrm{u}}$. Accordingly, we decompose $\Delta E_{\mathrm{n-u}}$ into two terms: $\Delta E_{\mathrm{n-u}} = \Delta E_{\mathrm{n-m}} + \Delta E_{\mathrm{m-u}}$. We can regard $\Delta E_{\mathrm{m-u}}$ as the energy change associated with the formation of an approximate native fold, and $\Delta E_{\mathrm{n-m}}$ as the one with the formation of specific packing to reach the precise native structure.

Energy difference between the native, near-native, and unfolded structures

Table 3 summarizes the average radius of gyration (R_{g}) calculated for all heavy atoms of each protein together with the RMSD of the near-native and unfolded structures from the native one. The near-native structures have slightly larger R_{g} because of their imperfect packing. The unfolded structures show much larger R_{g} than the native structure, indicating that they are indeed unfolded. The RMSD of the unfolded structures are all close to 12 Å, which is the upper limit of the RMSD restraints imposed (see above).

The upper block of Table 4 shows $\Delta E_{\mathrm{n-u}}$ (the energy difference between the native and unfolded structures) and its components, and some combinations of the components. Thermophilic proteins show lower total energy changes than their mesophilic counterparts in four out of five families of proteins (the only exception is 1tmy). To the contrary, only one thermophilic protein, 2prd, shows lower energy in vacuo than its mesophilic homolog. These observations suggest that the solvent effect may be crucial for the stabilization of thermophilic proteins. The electrostatic energy has been shown to be important for the stabilization of thermophilic proteins

Table 3 Radius of gyration calculated for all heavy atoms (in Å)

Protein[a]	Native	Near-native[b]	Unfolded[c]
t1cz3A	15.3	15.5 (0.16, 1.81)	22.5 (11.8)
m1aoeA	16.2	16.3 (0.14, 1.88)	23.4 (11.9)
m1dyjA	15.1	15.3 (0.17, 1.99)	21.9 (11.7)
t1dz3A	16.5	16.6 (0.11, 1.66)	22.9 (11.7)
m1srrA	13.0	13.1 (0.16, 1.69)	20.5 (11.8)
t1srvA	13.8	13.9 (0.15, 1.59)	21.1 (11.7)
m1kid	16.5	16.6 (0.13, 1.60)	23.9 (11.8)
t1tmy	12.7	12.8 (0.14, 1.52)	20.3 (11.7)
m2chf	13.1	13.4 (0.13, 1.76)	21.0 (11.8)
m3chy	13.1	13.3 (0.12, 1.65)	21.0 (11.8)
t2prd	15.0	15.2 (0.15, 1.75)	22.3 (11.9)
m1obwA	15.1	15.2 (0.13, 1.74)	22.2 (11.8)

[a] The PDB code and chain identifier with the first letter "t" or "m" indicating that the protein is either thermophilic or mesophilic, respectively

[b] The numbers in the parentheses are RMSD (Å) from the native structure; the first number was calculated for backbone atoms, the second for all heavy atoms

[c] RMSD (Å) from the native structure calculated for backbone atoms

(Xiao and Honig 1999). In our case, the electrostatic energy is defined as the sum of the Coulomb and reaction field terms in Table 4. Although most of the thermophiles show lower electrostatic energy than their mesophilic homologs, 1cz3A does not. However, $\Delta E_{\text{n–u}}$(EH) (the sum of the electrostatic and hydration terms in $\Delta E_{\text{n–u}}$) is consistently lower for all the thermophiles than in the mesophiles. This observation confirms the importance of the solvent effect.

In the previous study (Kinjo et al. 2001), we have shown that the "packing energy", i.e., the sum of bond length, bond angle, torsion angle, and van der Waals terms (Vorobjev et al. 1998), is the dominant factor for determining the precise native structure. Table 4 shows that the packing energy, $\Delta E_{\text{n–u}}$ (pack), of each protein is of a large negative value but only two thermophiles, 1cz3A and 2prd, show lower $\Delta E_{\text{n–u}}$ (pack) than their mesophilic counterparts. Furthermore, the difference in $\Delta E_{\text{n–u}}$ (pack) of these two thermophiles from their corresponding homologs are small (–0.1 and –0.15 cal/g, respectively) compared to the difference in $\Delta E_{\text{n–u}}$ (EH) (–1.5 to –3.8 cal/g, respectively). Thus, the relative stability of the thermophilic proteins is dominated by the difference in $\Delta E_{\text{n–u}}$ (EH). The electrostatic and hydration energies have been shown to be important for determining the approximate native fold regardless of the specific side-chain packing (Kinjo et al. 2001). Therefore, the trend of $\Delta E_{\text{n–u}}$ (pack) and $\Delta E_{\text{n–u}}$ (EH) in Table 4 suggests that the relative stability of the thermophilic proteins is associated with the approximate backbone topology rather than the precise side-chain packing. This point is further discussed in the following paragraphs.

The middle and lower blocks of Table 4 show $\Delta E_{\text{m–u}}$ (the energy difference between the near-native and unfolded structures) and $\Delta E_{\text{n–m}}$ (the energy difference between the native and near-native structures), respectively. Also shown are their components and some combinations of the components. As mentioned above, $\Delta E_{\text{m–u}}$ is associated with the enthalpy change during the formation of the approximate native fold, whereas $\Delta E_{\text{n–m}}$ is associated with the formation of the specific side-chain packing to reach the precise native structure.

Thermophilic proteins from four out of the five families show lower total $\Delta E_{\text{m–u}}$ values than the corresponding mesophiles. The thermophile 1tmy has lower $\Delta E_{\text{m–u}}$ than any other homolog, although it has higher $\Delta E_{\text{n–u}}$ than its mesophilic homolog 3chy (the upper block of Table 4). The opposite trend is found for 1srvA, which has higher $\Delta E_{\text{m–u}}$ than 1kid. As implied in the analysis of $\Delta E_{\text{n–u}}$, $\Delta E_{\text{m–u}}$ (EH) is consistently lower for the thermophiles than for the mesophiles. None of the thermophiles shows $\Delta E_{\text{m–u}}$ (pack) lower than the mesophiles. Therefore the relative stability of the thermophiles associated with the formation of the approximate fold is also dominated by $\Delta E_{\text{m–u}}$ (EH).

The stabilization by the formation of the specific side-chain packing can be seen in $\Delta E_{\text{n–m}}$. The stability of thermophilic proteins is not so conspicuous in $\Delta E_{\text{n–m}}$ as in $\Delta E_{\text{n–u}}$ and $\Delta E_{\text{m–u}}$. Three out of five thermophiles show lower total $\Delta E_{\text{n–m}}$ than their mesophilic homologs. Only two thermophiles, 1cz3A and 2prd, show lower $\Delta E_{\text{n–m}}$ (pack) than their mesophilic homologs. The $\Delta E_{\text{n–m}}$ (pack) of the thermophilic 1tmy is significantly higher than those of 2chf and 3chy. This observation shows that although good packing stabilizes some thermophilic proteins, it is not the universal mechanism for the thermostability of thermophiles, which confirms the statistical analysis by Karshikoff and Ladenstein (1998). $\Delta E_{\text{n–m}}$ (EH) are lower for the thermophiles than for the mesophiles, except for 1dz3A. The $\Delta E_{\text{n–m}}$ (EH) of 1dz3A is less stable than the $\Delta E_{\text{n–m}}$ (EH) of 1srrA by 1.35 cal/g. However, this difference is overwhelmed by the difference in $\Delta E_{\text{m–u}}$ (EH) (–5.07 cal/g). Two thermophiles, 1srvA and 1tmy, show significantly lower $\Delta E_{\text{n–m}}$ (EH) than the corresponding mesophiles (–2.02 cal/g and –2.78 cal/g, respectively). For these thermophilic proteins, the solvent effect also helps to stabilize the specific side-chain packing.

The side-chain packing of the near-native structures are largely distorted, and the electrostatic and hydration

Table 4 Total energy and its components (cal/g). The energy value of thermophiles is typed in boldface if it is lower than the corresponding values of any other mesophiles

Protein[a]	Total[b]	Local[c]	vdW[d]	Coul[e]	HB[f]	Oo[g]	Rea[h]	Vac[i]	Pack[j]	EH[k]
$\Delta E_{\mathrm{n-u}}$										
t1cz3A	**–13.53**	**0.05**	–28.91	–18.03	**–2.65**	**8.29**	27.72	–49.54	**–28.85**	**17.98**
m1aoeA	–10.30	0.83	–28.16	–25.08	–2.49	10.60	34.00	–54.90	–27.33	19.52
m1dyjA	–10.81	0.40	–29.10	–17.41	–2.56	11.11	26.76	–48.68	–28.70	20.45
t1dz3A	**–15.79**	**0.25**	–25.43	–12.68	–2.61	**6.01**	**18.67**	–40.47	–25.18	**12.00**
m1srrA	–14.88	1.94	–29.68	–19.56	–2.86	7.61	27.67	–50.16	–27.74	15.72
t1srvA	**–15.57**	1.11	**–29.23**	**–21.66**	–2.69	**7.03**	**29.87**	–52.47	–28.12	**15.24**
m1kid	–14.72	–0.50	–29.08	–21.64	–2.90	8.79	30.61	–54.12	–29.58	17.76
t1tmy	–15.60	1.17	–27.66	**–25.35**	–3.03	**7.03**	32.25	–54.88	–26.49	**13.93**
m2chf	–12.87	0.65	–30.89	–22.41	–3.05	9.16	33.66	–55.69	–30.24	20.41
m3chy	–18.04	–1.18	–32.03	–21.72	–3.23	9.19	30.93	–58.16	–33.21	18.40
t2prd	**–12.36**	**–0.39**	–30.29	**–22.26**	**–2.76**	10.16	33.18	**–55.70**	**–30.68**	**21.08**
m1obwA	–8.19	–0.22	–30.35	–9.36	–2.53	9.86	24.42	–42.46	–30.57	24.92
$\Delta E_{\mathrm{n-u}}$										
t1cz3A	**–3.19**	5.41	**–23.13**	–13.44	**–2.30**	**6.87**	23.39	–33.46	–17.72	**16.83**
m1aoeA	–0.66	5.75	–22.43	–17.27	–2.04	8.65	26.69	–35.99	–16.68	18.07
m1dyjA	–2.33	5.01	–22.81	–12.07	–2.23	8.58	21.19	–32.11	–17.80	17.70
t1dz3A	**–8.59**	**3.29**	–21.75	–9.93	–2.39	**4.86**	**17.33**	–30.78	–18.46	**12.26**
m1srrA	–5.62	4.84	–25.32	–13.97	–2.48	6.47	24.83	–36.92	–20.47	17.33
t1srvA	–5.30	5.41	–24.94	–16.21	–2.53	**6.67**	**26.30**	–38.28	–19.53	**16.76**
m1kid	–6.30	4.63	–25.53	–17.64	–2.66	7.68	27.23	–41.20	–20.90	17.27
t1tmy	**–8.03**	**3.13**	–24.85	**–17.79**	–2.69	6.52	27.64	**–42.19**	–21.72	**16.37**
m2chf	–5.85	4.83	–25.52	–12.52	–2.60	6.10	23.87	–35.81	–20.69	17.44
m3chy	–6.90	4.07	–26.29	–14.07	–2.73	7.01	25.12	–39.02	–22.22	18.05
t2prd	**–2.19**	4.70	–24.58	**–14.07**	**–2.39**	**7.32**	26.82	**–36.33**	–19.88	**20.07**
m1obwA	–0.24	3.67	–24.67	–8.82	–2.23	8.15	23.65	–32.04	–20.99	22.98
$\Delta E_{\mathrm{n-u}}$										
t1cz3A	**–10.34**	**–5.36**	**–5.78**	–4.59	–0.35	**1.41**	**4.33**	–16.08	**–11.14**	**1.15**
m1aoeA	–9.64	–4.93	–5.72	–7.81	–0.45	1.95	7.30	–18.90	–10.65	1.45
m1dyjA	–8.48	–4.61	–6.29	–5.34	–0.33	2.53	5.57	–16.57	–10.90	2.76
t1dz3A	–7.20	**–3.04**	–3.68	–2.75	–0.22	1.15	**1.34**	–9.68	–6.72	–0.26
m1srrA	–9.26	–2.90	–4.36	–5.59	–0.38	1.14	2.84	–13.24	–7.26	–1.61
t1srvA	**–10.27**	–4.30	**–4.28**	**–5.45**	–0.16	**0.35**	3.57	**–14.19**	–8.58	**–1.52**
m1kid	–8.42	–5.13	–3.55	–4.00	–0.23	1.11	3.38	–12.92	–8.68	0.50
t1tmy	–7.56	–1.97	–2.81	–7.56	–0.35	**0.51**	**4.61**	–12.68	–4.78	**–2.44**
m2chf	–7.03	–4.18	–5.37	–9.89	–0.45	3.06	9.80	–19.88	–9.55	2.97
m3chy	–11.15	–5.25	–5.74	–7.65	–0.50	2.18	5.81	–19.14	–11.00	0.34
t2prd	**–10.17**	**–5.09**	**–5.72**	**–8.19**	**–0.37**	2.84	6.36	**–19.37**	**–10.81**	**1.01**
m1obwA	–7.95	–3.89	–5.69	–0.54	–0.30	1.70	0.77	–10.42	–9.58	1.94

[a] See the caption of Table 3. For the names of the proteins, see Tables 1 and 2

[b] Total energy

[c] Local energy, which is the sum of bond length, bond angle, torsion, and improper torsion terms

[d] The sum of 1–4 and 1–5 van der Waals terms

[e] The sum of 1–4 and 1–5 Coulomb interaction terms

[f] Hydrogen bond term

[g] Hydration free energy term of Ooi et al. (1987)

[h] Electrostatic contribution from the solvent, i.e. reaction field energy

[i] Total energy in vacuum, which is equal to Total–Oo–Rea

[j] Packing energy which is the sum of Local and vdW terms

[k] The sum of Coul, Rea, and Oo terms

energy components are known to stabilize the native backbone topology even in the absence of the specific side-chain packing (Kinjo et al. 2001). Thus we can conclude that the thermophilic proteins are stabilized by the force that stabilizes the backbone topology rather than by the force that stabilizes the specific and precise side-chain packing of the native structure.

Discussion

From Table 4, we can see that $\Delta E_{\mathrm{n-u}}$ (pack) and $\Delta E_{\mathrm{m-u}}$ (pack) are large negative values whereas $\Delta E_{\mathrm{n-u}}$ (EH) and $\Delta E_{\mathrm{m-u}}$ (EH) are large positive values. This means that while the packing energy stabilizes the native and near-native structures, the EH energy destabilizes them. Therefore, the conspicuous tendency that $\Delta E_{\mathrm{n-u}}$ (EH) and $\Delta E_{\mathrm{m-u}}$ (EH) are always lower for the thermophilic proteins than for the mesophilic ones indicates that the relative energetic stability of the thermophilic proteins is due to destabilization of their unfolded state. This tendency is not seen in $\Delta E_{\mathrm{n-m}}$ (EH) and not all the thermophiles show lower $\Delta E_{\mathrm{n-m}}$ (EH). Therefore the main difference of the thermophilic from the mesophilic proteins resides in between the near-native and unfolded states. The near-native structures are a set of hypothetical structures which represent a group of structures somewhere between the native and unfolded

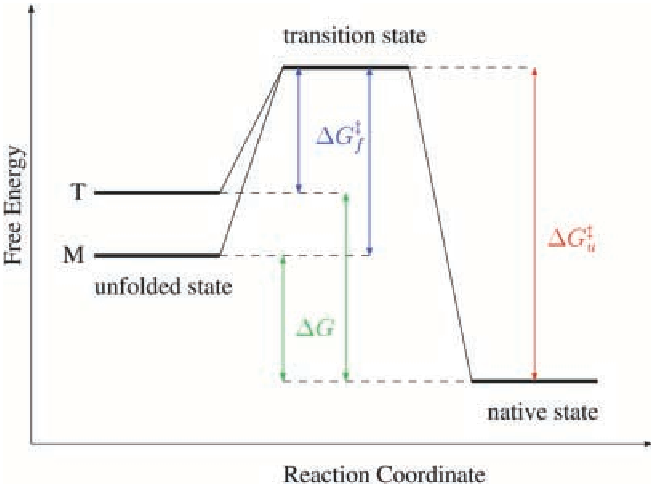

Fig. 1 The suggested stabilization mechanism of thermophilic proteins. $\Delta G^{\ddagger}_{\mathrm{f}}$ and *blue arrows*, the activation free energy of folding; $\Delta G^{\ddagger}_{\mathrm{u}}$ and *red arrows*, the activation free energy of unfolding; ΔG and *green arrows*, the free energy of folding and unfolding. "*T*" and "*M*" stand for the unfolded state of thermophilic and mesophilic proteins, respectively

states, in which overall backbone topology or fold is well formed but the side-chain packing is not fully achieved. On the one hand, since the overall native topologyis well formed in the near-native structure, there should be a large entropy loss with respect to the unfolded structures. On the other hand, the energetic stabilization of the native structure is mostly attained by the precise packing, $\Delta E_{\mathrm{n-m}}$ (pack), which is not realized in the near-native structure. Therefore the near-native structures would be thermodynamically unstable. These characteristics of the near-native structures are qualitatively similar to the transition state structure of proteins (Shakhnovich and Finkelstein 1989; Fersht 2000). If we assume that the near-native structures are a representative set of the transition state structures, then the stability of the thermophilic proteins can be explained as follows (Fig. 1).

First, the less positive values of $\Delta E_{\mathrm{m-u}}$ (EH) of the thermophilic proteins (the middle block of Table 4) indicate less stable unfolded states, hence lower $\Delta G^{\ddagger}_{\mathrm{f}}$ (the activation free energy of folding). Second, since there is not a significant difference in $\Delta E_{\mathrm{n-m}}$ or $\Delta E_{\mathrm{n-m}}$ (pack) (the lower block of Table 4), $\Delta G^{\ddagger}_{\mathrm{u}}$ (the activation free energy of unfolding) should not be different between thermophiles and mesophiles. These result in a larger negative value of the overall folding free energy ΔG of the thermophilic proteins. This scheme for the stabilization of thermophilic proteins implies that these proteins fold faster. The faster folding of thermophilic proteins is actually suggested by a study of amino acid compositions (S. Fukuchi and K. Nishikawa, unpublished results).

That the thermophilic proteins show less stable $\Delta E_{\mathrm{m-u}}$ (pack) suggests their near-native structures are structurally more flexible or entropically more stable compared to the mesophilic proteins. The entropic stabilization of a thermostable protein was observed by hydrogen exchange and neutron scattering experiments (Fitter and Heberle 2000). At a glance, this conclusion contradicts the results of Lazaridis et al. (1997). However, the protein they investigated was from an archaeon. It is suggested that one of the factors that stabilizes the proteins from archaea is better packing, but it is not the case for the proteins from thermophilic eubacteria (Szilágyi and Závodszky 2000). In any case, our conclusion is limited to the proteins from thermophilic eubacteria. Since the thermal fluctuations at high temperature make it difficult to maintain the precise side-chain packing, it may be possible that the thermophilic proteins have evolved so that they can rapidly re-cross the transition state to recover the native state. Perl et al. (1998) compared the folding and unfolding kinetics of thermophilic cold-shock proteins with their mesophilic homolog. In that study, the large difference between the thermophiles and the mesophile was observed in the unfolding rather than in the folding, which contradicts our prediction. However, the transition state of the cold-shock protein is known to be extremely anomalous (Plaxco et al. 1998). Similar experimental studies on other proteins are required to validate our results.

Acknowledgements We thank Drs. Satoshi Fukuchi and Takeshi Kawabata for extensive discussions. We are also grateful to Dr. Haruki Nakamura and the Biomolecular Engineering Research Institute (Osaka, Japan) for providing the program for electrostatics calculation and EMBOSS, respectively. A.R.K. was a predoctoral research fellow of the Japan Society for the Promotion of Science. All the computations were carried out on a VPP500 (Fujitsu) at the National Institute of Genetics, Mishima, Japan.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242

Elcock AH (1998) The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. J Mol Biol 284:489–502

Ferrara P, Apostolakis J, Calflisch A (2000) Computer simulation of protein folding by targeted molecular dynamics. Proteins 39:252–260

Fersht AR (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. Proc Natl Acad Sci USA 97:1525–1529

Fitter J, Heberle J (2000) Structural equilibrium fluctuations in mesophilic and thermophilic α-amylase. Biophys J 79:1629–1636

Karshikoff A, Ladenstein R (1998) Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. Protein Eng 11:867–872

Kinjo AR, Kidera A, Nakamura H, Nishikawa K (2001) Physicochemical evaluation of protein folds predicted by threading. Eur Biophys J 30:1–10

Kumar S, Tsai CJ, Nussinov R (2000) Factors enhancing protein thermostability. Protein Eng 13:179–191

Lazaridis T, Lee I, Karplus M (1997) Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci 6:2589–2605

Nakai T, Kidera A, Nakamura H (1993) Intrinsic nature of the three-dimensional structure of proteins as determined by distance geometry with good sampling properties. J Biomol NMR 3:19–40

Nakamura H (1996) Roles of electrostatic interaction in proteins. Q Rev Biophys 29:1–90

Nakamura H, Nishida S (1987) Numerical calculations of electrostatic potentials of protein-solvent systems by the self consistent boundary method. J Phys Soc Jpn 56:1609–1622

Ooi T, Oobatake M, Nemethy G, Scheraga HA (1987) Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 84:3086–3090

Perl D, Welker C, Schindler T, Schröder K, Marahiel MA, Jaenicke R, Schmid FX (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. Nat Struct Biol 5:229–235

Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 277:985–994

Richmond TJ (1984) Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. J Mol Biol 178:63–89

Shakhnovich EI, Finkelstein AV (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. Biopolymers 28:1667–1680

Szilágyi A, Závodszky P (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure 8:493–504

Vorobjev YN, Almagro JC, Hermans J (1998) Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. Proteins 32:399–413

Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force field for simulations of proteins and nucleic acids. J Comput Chem 7:230–252

Xiao L, Honig B (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. J Mol Biol 289:1435–1444