J. Biochem. **142,** 507–516 (2007)
doi:10.1093/jb/mvm157

Protein Thermostabilization Requires a Fine-tuned Placement of Surface-charged Residues

Dong-Ju You[1], Satoshi Fukuchi[2], Ken Nishikawa[2], Yuichi Koga[1], Kazufumi Takano[1,3,*] and Shigenori Kanaya[1]

[1]Department of Material and Life Science, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan; [2]Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan; and [3]CREST, JST, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Received June 14, 2007; accepted July 23, 2007; published online August 30, 2007

Using the information from the genome projects, recent comparative studies of thermostable proteins have revealed a certain trend of amino acid composition in which polar residues are scarce and charged residues are rich on the protein surface. To clarify experimentally the effect of the amino acid composition of surface residues on the thermostability of *Escherichia coli* Ribonuclease HI (RNase HI), we constructed six variants in which five to eleven polar residues were replaced by charged residues (5C, 7Ca, 7Cb, 9Ca, 9Cb and 11C). The thermal denaturation experiments indicated that all of the variant proteins are 3.2–10.1°C in T_m less stable than the wild proteins. The crystal structures of resultant protein variants 7Ca, 7Cb, 9Ca and 11C closely resemble that of *E. coli* RNase HI in their global fold, and several different hydrogen bonding and ion-pair interactions are formed by the mutations. Comparison of the crystal structures of these variant proteins with that of *E. coli* RNase HI reveals that thermal destabilization is apparently related to electrostatic repulsion of the charged residues with neighbours. This result suggests that charged residues of natural thermostable proteins are strictly posted on the surface with optimal interactions and without repulsive interactions.

Key words: amino acid composition, genome information, RNase HI, surface-charge residue, thermostability.

Abbreviations: CD, Circular dichroism; GdnHCl, guanidine hydrochloride; and RNase H, Ribonuclease H.

The rapid progress of genome projects has provided a large amount of information on the sequence of the genes, and therefore proteins, of various organisms. Systematic statistical analyses of the differences between the proteins of thermophiles and mesophiles have been made possible (*1–3*). Recent comparative studies have revealed a certain trend of the amino acid composition of thermophilic proteins. The difference between the amino acid composition of thermophilic proteins and that of mesophilic proteins has been found to follow a universal trend across a large number of protein families. Charged residues (e.g. Glu, Asp, Lys, and Arg) are rich, whereas polar residues (e.g. Ser, Thr, Asn, and Gln) are more scarce on thermophilic proteins than on mesophilic proteins (*3*). Additionally, the difference between the amino acid composition of mesophilic proteins and that of thermophilic proteins is much greater on the protein surface than in the interior, suggesting a strategy for adaptation to the environment (*4*). These observations lead to the proposal that only the introduction of charged residues on the protein surface can be a viable tactic to increase thermostability of protein. If so, it is a simple strategy to stabilize proteins in natural selection and even in protein engineering. However, if thermophiles closely select the formation of ion pairs, more systematic steps involving analyses based on sequence and structural information for a group of proteins are required. Therefore, it is necessary to conduct an experiment study examining the effect of amino acid composition of the surface on the stability of a protein.

Ribonuclease H (RNase H) endonucleolytically cleaves RNA of the RNA/DNA hybrid in the presence of divalent cation (*5*). RNase H is ubiquitously present in various organisms (*6*) and is involved in DNA replication, repair and transcription (*7–10*). *Escherichia coli* RNase HI provides a good model for the study of thermostabilization of thermophilic proteins because the thermal denaturation of *E. coli* RNase HI is reversible and its thermodynamics have been characterized in detail (*11*, *12*). In addition, the 3D-structure of RNase HI from two sources of different growth temperatures, *E. coli* (mesophile) and *Thermus thermophilus* (extreme thermophile), have been solved (*13–15*). The melting temperature of *T. thermophilus* RNase HI, which was determined from thermal denaturation experiments in the presence of 1 M guanidine hydrochloride (GdnHCl) at pH 5.5, was 30.9°C higher than that observed for *E. coli* RNase HI (*16*). Despite unusual differences in thermostability, these proteins share a high degree of homology in primary sequence and a high degree of similarity in 3D-structure. The sequence and structural similarities between their enzymes enable comparison for

*To whom correspondence should be addressed. Tel/Fax: +81-6-6879-4157, E-mail: ktakano@mls.eng.osaka-u.ac.jp

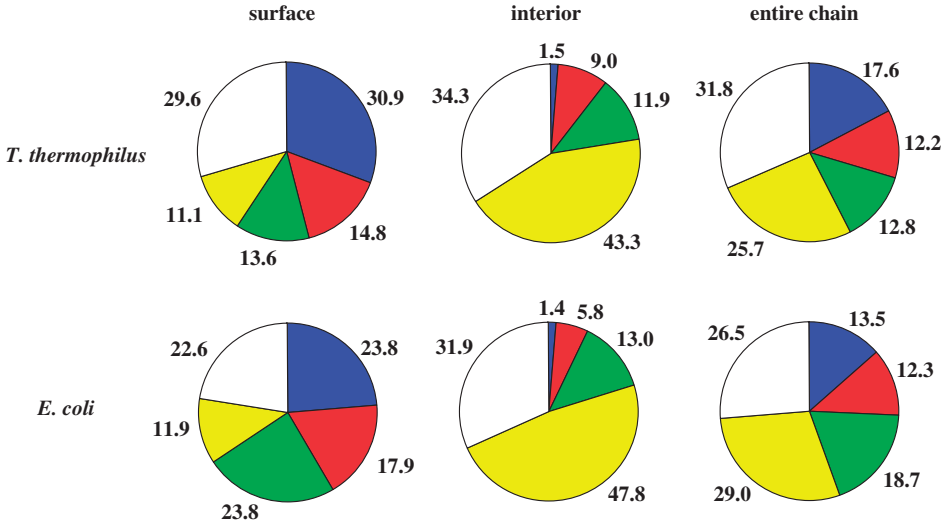

Fig. 1. **Pie diagrams representing the fractions of charged, polar and apolar residues by RNase HI from *T. thermophilus* and *E. coli*.** Blue denotes positively charged residues (Arg and Lys); red denotes negatively charged (Asp and Glu); green indicates polar (Asn, Gln, Ser, and Thr); yellow indicates apolar (Ile, Leu, Met, Phe, Trp, Tyr, and Val); and white denotes other residues (Ala, Cys, Gly, His, and Pro). In both the structures, amino acid residues with relative solvent accessibility greater than 25% were regarded as residues exposed to solvent.

thermal stability. Analysis of amino acid compositions in the RNase HI from *E. coli* and that from *T. thermophilus* suggests that the difference between amino acid compositions of these enzymes is much greater on the protein surface than on the protein interior (Fig. 1). RNase HI from *T. thermophilus* is rich in charged residues, with a reduced number of polar residues.

To determine the effect of surface-charged residues on protein thermostability, we constructed six variant proteins of *E. coli* RNase HI, in which five to eleven polar residues are replaced by charged ones. We confirmed that all the variant proteins were less stable than the wild protein. The crystal structure of four variants was then solved, and the relationship between thermostability and molecular structure was analysed. We found that repulsive interactions greatly disturb protein stabilization.

MATERIALS AND METHODS

Cells, Plasmids and Materials—Plasmid pJAL600 for the overproduction of the wild protein was previously constructed (*17*). The *E. coli* MIC2067 is an *rnhA* and *rnhB* double-mutant strain, and the MIC2067 (DE3) carrying a cloned T7 RNA polymerase gene has already been reported (*18*). The *E. coli* MIC2067 (DE3) was used as a host strain for overexpression of the variant protein with the pET system. The plasmid pET-25b was purchased from Novagen (Madison, WI, USA). The *E. coli* MIC2067 (DE3) transformants were grown in NZCYM medium (Novagen) containing 50 μg/ml ampicillin and 50 μg/ml chloramphenicol. All DNA oligomers for PCR were synthesized by Hokkaido System Science (Sapporo, Japan). Restriction and modifying enzymes were from TaKaRa Bio (Kyoto, Japan). Other chemicals were of reagent grade.

Mutagenesis, Overproduction and Purification—The genes encoding the *E. coli* RNase HI variants were constructed by PCR-based, site-directed mutagenesis (*19*). Plasmid pJAL600 was used as a template. The plasmids for overproduction of the variant proteins were constructed by ligating the DNA fragments containing these genes into the *Nde*I–*Sal*I sites of pET-25b. The oligonucleotides used for the mutations are listed subsequently, with the exchanged bases underlined.

Q4R: 5′-ATGCTTAAACGTGTAGAAATTTTCACCGATGG-3′

T40E: 5′-GCTGGCTACGAACGCACCACCAACAACCG-3′

Q72H, Q76K and Q80E: 5′-GAGTACCGACAGCCATTACGTACGCAAAGGTATCACCGA ATGGATCC-3′

T92K: 5′-GGCTGGAAAAAAGCAGACAAAAAACCAG-3′

Q105K: 5′-CTCTGGAAACGTCTAGATGCTGCATTGGGGC-3′

Q113R and Q115K: 5′-CGTCTAGATGCTGCATTGGGGCGCCATAAAATCAAATGGGAA TGGG-3′

N143K and T145K: 5′-GTGCCGCGGCGATGAAACCCAAACTCGAGGATACAGGCTACC AAGTT-3′

Overproduction and purification of the wild-type protein were performed as described previously (*17*). Transformation of *E. coli* MIC2076 (DE3) with the recombinant plasmids and overproduction of the recombinant proteins were performed according to the procedure described previously (*20*) with a slight modification. The transformed cells were cultured in NZCYM medium containing 50 μg/ml ampicillin and 50 μg/ml chloramphenicol, at 30°C. When the D_{600} of the culture reached 0.5, isopropyl-β-D-thiogalactoside was added to the culture medium (final concentration, 1 mM), and induction was continued for an additional 4 h. Cells were collected by centrifugation, suspended in buffer A (10 mM Tris–HCl pH 7.5, 1 mM EDTA, 0.2 M NaCl), disrupted by

a French press, and centrifuged at 30,000*g* for 30 min. The soluble fraction was loaded onto a HiTrap SP HP column (Pharmacia/GE Healthcare, Sweden) equilibrated with buffer A. The protein was eluted from the column with a linear gradient of 0–0.5 M NaCl. The fractions containing the protein were combined, dialysed against buffer B (10 mM Tris–HCl pH 7.5, 1 mM EDTA), and loaded onto a HiTrap Heparin HP column (Pharmacia/GE Healthcare, Sweden) equilibrated with buffer B. The protein was eluted from the column with a linear gradient of 0–0.5 M NaCl. The fractions containing the protein were combined and dialysed against 10 mM sodium acetate buffer (pH 5.5). The protein concentration was determined from UV absorption using the A_{280} value of 2.02 for the wild protein and six variant proteins. These values were calculated by using $\varepsilon = 1576\,M^{-1}cm^{-1}$ for Tyr and $5225\,M^{-1}cm^{-1}$ for Trp at 280 nm (*21*). The purity of the protein was confirmed by SDS–PAGE (*22*), followed by staining with Coomassie brilliant blue.

Circular Dichroism (CD) Spectra—The CD spectra were measured on a J-725 spectropolarimeter (Japan Spectroscopic Co., Ltd, Tokyo, Japan). Spectra were obtained at 25°C in 10 mM sodium acetate buffer (pH 5.5). The protein concentration was 0.15 mg/ml, and the optical path length was 2 mm.

Thermal Denaturation—Thermal denaturation curves and the temperature of the mid-point of the transition (T_m) were determined by monitoring the CD value at 220 nm, as described previously (*16*). Proteins were dissolved in 20 mM sodium acetate buffer (pH 5.5) containing 1 M GdnHCl. The protein concentration was 0.15 mg/ml, and the optical path length was 2 mm. The temperature of the cuvette containing the sample solution was raised at a rate of 1°C/min. Parameters characterizing the thermal denaturation of the mutant proteins were determined as described previously (*12*).

Crystallization, Data Collection and Refinement—Crystals of the *E. coli* RNase HI variants 7Ca, 7Cb, 9Ca and 11C were grown by the sitting-drop method at 20°C. The conditions were 100 mM HEPES–NaOH (pH 7.0–8.0) containing 15–25% PEG-3350, and 6–8 mg/ml protein. Crystals were flash-frozen in liquid nitrogen directly from the drop conditions.

X-ray diffraction data were collected at the synchrotron radiation facility at BL38B1, BL41XU and BL44XU (SPring-8, Hyogo, Japan). Processing of the diffraction images and scaling of the integrated intensities were conducted with the program HKL2000 (*23*). The structures of all variant proteins were solved by molecular replacement with Molrep using the coordinates of the wild-type protein. The model was rebuilt with the program O (*24*) and refined with CNS (*25*). The R-free set contained 5% of the reflections chosen at random.

Protein Data Bank Accession Numbers—The coordinates of the structures of the variant proteins have been deposited in the Protein Data Bank under ID codes 2Z1G for 7Ca, 2Z1H for 7Cb, 2Z1I for 9Ca and 2Z1J for 11C.

RESULTS

Preparation of E. coli RNase HI Variants—In order to examine whether increasing the number of charged residues on protein surface can improve protein thermostability, six variant proteins of *E. coli* RNase HI, 5C, 7Ca, 7Cb, 9Ca, 9Cb and 11C, were designed, so that the polar residues were replaced by the charged ones at eleven positions with different combinations. These positions were chosen because they are exposed to solvent in *E. coli* RNase HI structure and they are occupied by charged residues in RNases HI from several thermophilic bacteria (e.g. *Moorella thermoacetica*, *T. thermophilus*, and *Thermoanaerobacter tengcongenisi*), with the exception of Q72H (Fig. 2). This strategy was based on the results shown by Fukuchi and Nishikawa (*4*), in which it was shown that the differences between thermophilic proteins and mesophilic proteins on the protein surface are mainly caused by the presence or absence of single-charged residues. The mutations introduced into the variant proteins are summarized in Table 1 and Fig. 2C. All of these proteins are accumulated in the *E. coli* cells in a soluble form upon induction for overproduction and purified to give a single band on SDS–PAGE (data not shown).

Stability of Six Variants of RNase HI—The secondary structures of the proteins were estimated by CD measurement. Far-UV CD spectra of variants were indistinguishable from that of the wild-type protein (data not shown), implying that the mutations did not significantly affect the overall structure of the protein.

To compare the stability of these proteins, the thermal denaturation of the wild-type and variant proteins was monitored by measuring CD at 220 nm. Thermal denaturation processes of the wild-type and variant proteins were reversible under this condition. Typical thermal denaturation curves for the wild-type protein and six variants are depicted in Fig. 3. The result for the wild-type protein agreed with that of previous studies. We found that all of the variant proteins were 3.2–10.1°C in T_m less stable than the wild-type protein. The thermodynamic parameters characterizing the folding of the variants are presented in Table 1. Mutation of the polar residues to charged residues all yielded significant decrease in conformational stability, as indicated by the negative $\Delta\Delta G_m$ values, caused by decrease in ΔH_m. These results suggest that the *E. coli* RNase HI variants are destabilized due to unfavourable electrostatic interactions caused by the increased number of charges on the surface.

Three-dimensional Structure of 7Ca, 7Cb, 9Ca and 11C—The aforementioned results indicate that the stability of six variants is lower than that of the wild-type protein. To determine if any structural changes might account for decreased thermostability, we solved the structures of the variant proteins 7Ca, 7Cb, 9Ca and 11C. Data collection and refinement statistics for the variant proteins are summarized in Table 2. The structures of the these variant proteins seemed to be essentially the same as that of the wild-type protein (0.95Å RMSD for 151 Cα atoms on the average) except for the positions of the loops between βB and βC strands (residues 27–30), between αIII and αIV helices (residues 89–99), and at the mutation sites. The conformation of the mutation site was similar among the

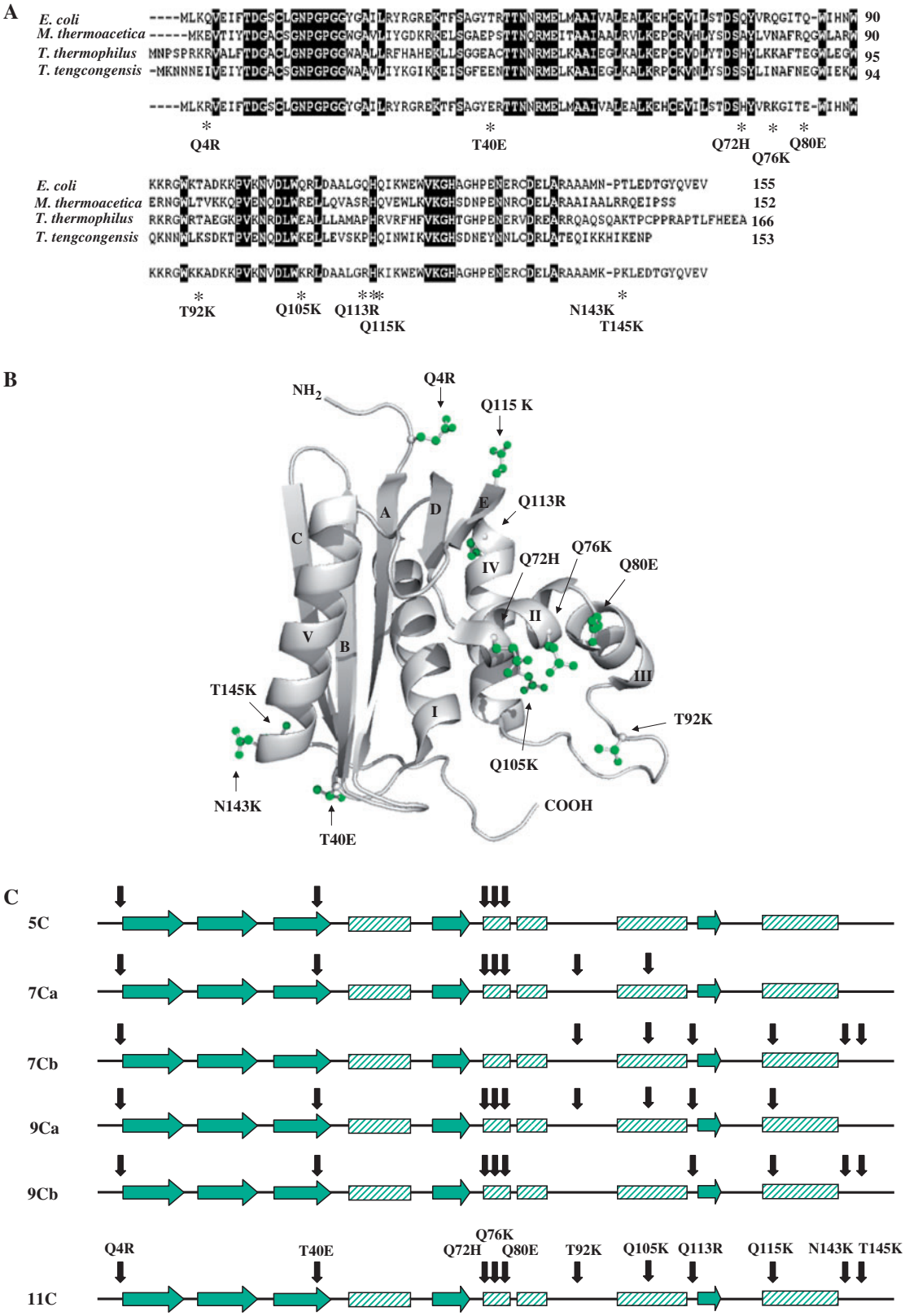

Fig. 2. **(A) Multiple alignment of RNase HI from *E. coli* and thermophilic bacterium, as *M. thermoacetica*, *T. thermophilus* and *T. tengcongenisi*.** The eleven residues of *E. coli* RNase HI substituted in the present study are indicated subsequently. (B) Cartoon representation of the 3D-structure of *E. coli* RNase HI (PDB entry 2RN2). The side chains of substitutions are depicted in ball-and-stick representation. (C) Schematic illustration for the mutation sites (black arrows). The secondary structure elements are represented by the arrow for β-strand and box with the stripes for α-helix.

Table 1. **The parameters characterizing the thermal denaturation of wild-type and variant RNase HI proteins.**

Proteins	Mutations	T_m (°C)	ΔT_m (°C)	ΔH_m (kcal mol^{-1})	ΔS (kcal mol^{-1} K^{-1})	$\Delta\Delta G_m$ (kcal mol^{-1})
Wild-type		52.0		80.6	0.248	
5C	Q4R/T40E/Q72H/Q76K/Q80E	48.8	− 3.2	76.5	0.238	− 0.8
7Ca	Q4R/T40E/Q72H/Q76K/Q80E/T92K/Q105K	45.2	− 6.8	64.8	0.204	− 1.7
7Cb	Q4R/T92K/Q105K/Q113R/Q115K/N143K/T145K	47.6	− 4.4	62.7	0.196	− 1.1
9Ca	Q4R/T40E/Q72H/Q76K/Q80E/T92K/Q105K/Q113R /Q115K	48.3	− 3.7	70.9	0.221	− 0.9
9Cb	Q4R/T40E/Q72H/Q76K/Q80E/Q113R/Q115K/N143K /T145K	46.2	− 5.8	64.7	0.203	− 1.4
11C	Q4R/T40E/Q72H/Q76K/Q80E/T92K/Q105K/Q113R /Q115K/N143K/T145K	41.9	− 10.1	53.0	0.168	− 2.5

The melting temperature (T_m) is the temperature of the mid-point of the thermal denaturation transition. The difference in the melting temperature between the wild-type and variant protein (ΔT_m) is calculated as T_m (variant) − T_m (wild-type). ΔH_m is the enthalpy change of unfolding at T_m calculated by van't Hoff analysis. The difference between the free energy change of unfolding of the variant proteins and that of the wild-type protein at the T_m of the wild-type protein ($\Delta\Delta G_m$) was estimated, $\Delta\Delta G_m = \Delta T_m \Delta S_m$(wild-type), where ΔS_m(wild-type) is the entropy change of the wild-type protein at T_m.

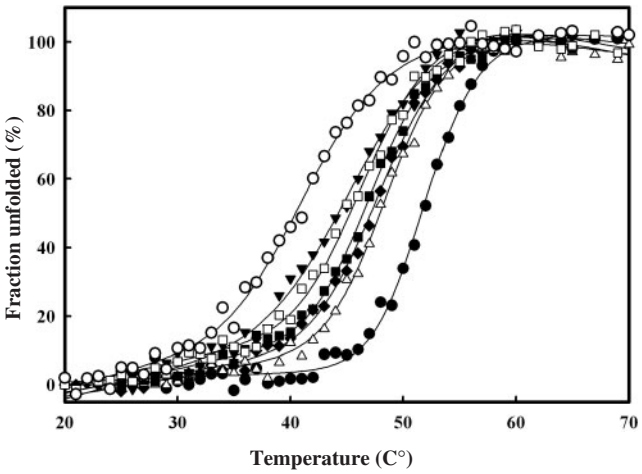

Fig. 3. **Thermal denaturation curves of the wild and variant proteins.** The apparent fraction of unfolded protein is presented as a function of temperature. Filled circle, wild type; open triangle, 5C; filled inverse triangle, 7Ca; filled square, 7Cb; filled diamond, 9Ca; open square, 9Cb; open circle, 11C. Thermal denaturation curves were determined in the presence of 1.0 M GdnHCl at pH 5.5 by monitoring the change in the CD value at 220 nm, as described under MATERIALS AND METHODS.

variant proteins. The details of the structures are described in the DISCUSSION section.

DISCUSSION

Relationship Between Stability and Structure—In order to reveal the relationship between stability and structure, we examined the effect of eleven amino acid substitutions on the structural characteristics of the variant proteins by comparing the structures of the wild-type protein and four variant proteins. The following results were obtained.

T40E, Q72H, Q76K and Q80E

The variant protein 11C is less stable than 7Cb by 5.7°C in T_m, indicating that the quadruple substitutions of T40E, Q72H, Q76K and Q80E destabilize the protein. Thr40 is located on the loop between the βC strand and αI helix regions. The hydroxyl group of Thr40 does not have any hydrogen-bonding partners in the wild-type protein and is located near Glu147 at the C-terminal region (Fig. 4A). The side chain of Thr40 substituted with Glu relocated near the side chain of Glu147 in 7Ca, 9Ca and 11C. Figure 4B illustrates the structure of 11C around Glu40.

From the structures of 7Ca, 9Ca and 11C, however, we found a new ion pair between Lys76 and Glu80, which were introduced in these variants (Fig. 5B and C). This ion pair is also conserved in the structure of *T. thermus* RNase HI (*13*) As a result, the quadruple substitutions produced one ion pair and one unfavourable interaction. An ion pair generally stabilizes protein, but 11C was less stable than 7Cb. These results suggest that electrostatic repulsion between Glu40 and Glu147 lead to a decrease in thermal stability of the protein.

T92K/Q105K

When we generated T92K and Q105K substitutions in the 5C and 9Cb backgrounds, the substitutions of T92K and Q105K destabilized the variants 7Cb and 11C by a similar extent (3.6–4.3°C in T_m). These results indicate that the substitutions of T92K and Q105K destabilize the protein.

In the wild-type structure, Thr92 forms a hydrogen bond with Asp94 (Fig. 5A). In 11C, the substitution of Thr92 to Lys revealed that Lys92 forms an ion pair with Asp94, and that Asp94 also has an ion pair with Lys96 (Fig. 5B). The substitution of T92K eliminates the hydrogen bond but forms an ion-pair network. Lys105 does not have any interaction in the 11C structure (Fig. 5B). The same results were obtained in the structures of the other variant proteins (data not shown).

It is generally acknowledged that introducing an ion pair at the position of hydrogen bond enhances stability. An individual hydrogen bond could stabilize a protein structure at an average of 1.3 kcal/mol in RNase T1 (*26*), and a salt bridge or ion pair could significantly stabilize a protein structure at 3–5 kcal/mol in T4 lysozyme (*27*). However, T92K and Q105K substitutions destabilize the protein.

Table 2. **Data collection and refinement statistics.**

	7Ca	7Cb	9Ca	11C
Wavelength (Å)	1.0	0.9	0.9	1.0
Space group	$P3_221$	$P3_221$	$P2_12_12_1$	$P3_221$
Unit-cell dimensions				
a (Å)	a = 88.1	a = 88.2	a = 52.5	a = 85.3
b (Å)	b = 88.1	b = 88.2	b = 129.6	b = 85.3
c (Å)	c = 55.0	c = 55.7	c = 43.5	c = 51.0
Resolution (Å)[a]	20.0–2.0 (2.07–2.0)	38.0–2.5 (2.6–2.5)	50.0–2.0 (2.07–2.0)	50.0 –2.3 (2.38–2.3)
R_{merge} (%)[a,b]	7.1 (31.8)	3.2 (18.0)	9.8 (23.1)	8.3 (37.1)
Completeness (%)[a]	99.9 (99.9)	99.9 (99.9)	98.4 (91.4)	99.2 (92.5)
Average I/σ (I) [a]	39.5 (5.1)	20.0 (3.9)	21.5 (4.9)	23.8 (1.9)
No. molecules a. u.	1	1	2	1
V_M (Å^3 Da^{-1})	3.5	3.7	2.1	3.0
R factor (R/R_{free}) [c]	21.4/24.1	21.4/25.7	21.1/26.0	20.8/23.8
No. of atoms (protein/solvent)	1245/144	1247/66	2430/289	1218/43
RMSD from ideal				
bond lengths (Å)	0.006	0.007	0.005	0.01
bond angles (deg.)	1.1	1.2	1.1	1.3

[a]Values in parentheses are for the highest resolution shell.
[b]$R_{merge} = \sum|I_{hkl} - < I_{hkl} >|/\sum I_{hkl}$, where I_{hkl} is the intensity measurement for reflection with indices hkl and $< I_{hkl} >$ is the mean intensity for multiply recorded reflections.
[c]R_{free} was calculated using 5% of the total reflections chosen randomly and omitted from refinement.

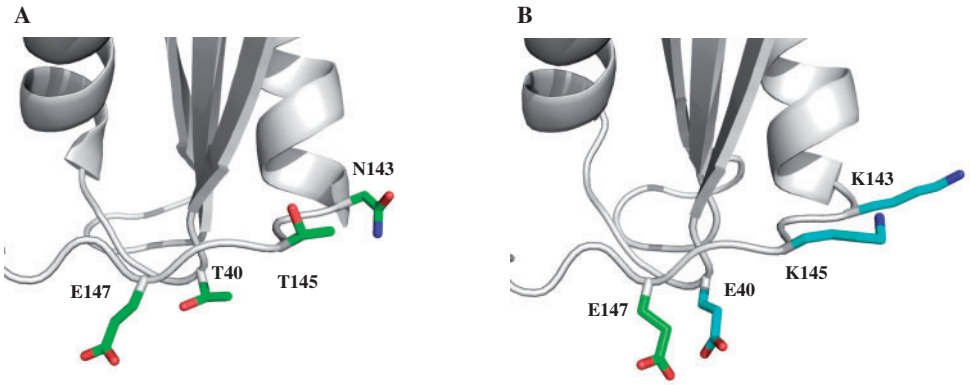

Fig. 4. **Structures of the regions containing Thr40, Asn143 and Thr145 in wild protein and those containing Glu40, Lys143 and Lys145 in 11C variant.** Wild, A; 11C variant, B. Introduced residues are indicated in cyan, and original residues are indicated in green.

The location of Thr92 in the basic protrusion region is important for RNA/DNA hybrid recognition. Gln105 is near this region. The surface of the basic protrusion region presents a patch of basic residues (K86, K87, R88, K91, K95 and K96) with high potential for participating in substrate binding (*28*). It has been reported that ionic interactions can sometimes destabilize an enzyme, due to the distribution of other charged groups in the native or non-native state (*29–33*). We suspect that the substitution of T92K and Q105K may result in unfavourable interactions distributed by the charged groups on this region in the native or non-native states, leading to a decrease in thermal stability.

Q113R/Q115K

In this study, all variant proteins exhibited decreased stability. In fact, the variant protein 9Ca, in which both Gln113 and Gln115 are replaced by Arg and Lys in the 7Ca background, was less stable than the wild-type protein by 3.7°C in T_m. However, it was more stable than 7Ca by 3.1°C in T_m (Table 1).

When the crystal structures of the variants are compared with that of the wild-type protein, an ionic interaction is observed at the position of Lys115 (Fig. 6C), which is located at the N-terminal region of the βE strand; however, Gln115 does not have any interaction partners in the structure of the wild-type protein (Fig. 6A). In the structures of the variant proteins, Lys115 is involved in an ion-pair network with Arg4 and Glu64 on the βA and βD strands (Fig. 6C). It has been reported that ion-pair networks contribute to the stability of thermophilic and hyperthermophilic proteins (*34–36*), and are energetically more favourable than an equivalent number of isolated ion pairs (*37*, *38*). Interestingly, in the *T. thermophilus* RNase HI structure, a five-member ion-pair network

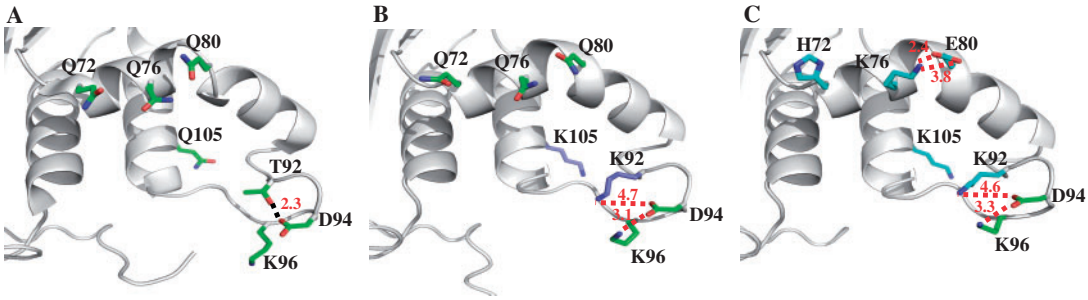

Fig. 5. **Structures of the regions containing Gln72, Gln76, Gln80, Thr92 and Gln105 in wild protein; those containing Gln72, Gln76, Gln80, Lys92 and Lys105 in the 7Cb variant; and those containing His72, Lys76, Glu80, Lys92 and Lys105 in the 11C variant.** Wild, A; 7Cb variant, B; 11C variant, C. The dotted lines represent hydrogen bonds (black) or ion pairs (red) with distance (Å).

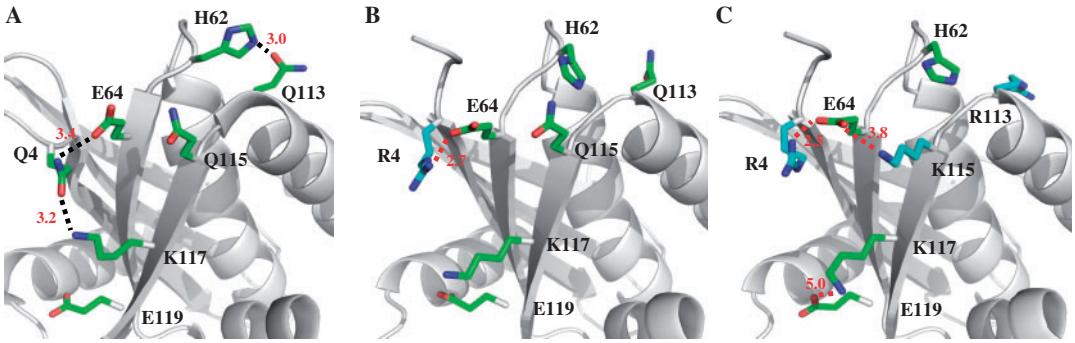

Fig. 6. **Structures of the regions containing Gln4, Gln113 and Gln115 in wild protein; those containing Arg4, Gln113 and Gln115 in the 7Ca variant; and those containing Arg4, Arg113 and Lys115 in the 11C variant.** Wild, A; 7Ca variant, B; 11C variant, C. The dotted lines represent hydrogen bonds (black) or ion pairs (red) with distance (Å).

involving Arg4-Asp66-Arg117-Glu64-Arg115 is located on the surface of the βA, βD and βE strands. The position and electrostatic properties of Arg4, Arg117, Glu64 and Arg115 in *T. thermophilus* RNase HI, with the exception of Asp66, are similar to those of Arg4, Lys117, Glu64 and Lys115 in the variants of *E. coli* RNase HI.

In the structure of the wild-type protein, Gln113 is located at the C-terminal region of αIV helix and forms a hydrogen bond with His62 (Fig. 6A). In 11C, the result from the substitution of Gln113 to Arg eliminates the hydrogen bond with His62. Arg113 is exposed in the solvent environment without any interaction (Fig. 6C). The same results are obtained in the structures of other variant proteins (data not shown).

These results indicate that the increase in the stability in the 7Ca background upon Q113R and Q115K substitutions is due to the formation of the new ion-pair network. Furthermore, these charged residues have few unfavourable interactions with neighbours.

N143K/T145K

When both Asn143 and Thr145 are replaced by Lys in the 9Ca background, the substitutions of N143K and T145K significantly decrease the stability of 9Ca by 6.4°C in T_m. These residues are exposed in the solvent environment and do not have any interaction partners in the structures of the wild-type and variant proteins (Fig. 4A and B). These results imply that the repulsive electrostatic interaction between Lys143 and Lys145 causes destabilization.

Q4R

The effect of Q4R substitution on thermostability could not be directly estimated by the data of the variants. Comparison of the structure of the wild-type and 5C proteins indicates that Q4R might have little effect on stability or might increase it slightly, because the quadruple substitution of T40E, Q72H, Q76K, and Q80E leads to a decrease in stability (5.7°C in T_m between 7Cb and 11C), as described above. In the structure of the wild-type protein, Gln4, which is located within the N-terminal βA strand, forms charged-neutral hydrogen bonds with Glu64 and Lys117 on βD and βE strands (Fig. 6A). In the structures of the variant proteins, the substitution of Q4R disrupts the hydrogen bonds but forms ion pairings between Arg4 and Glu64, and between Lys117 and Glu119 (Fig. 6B and C). Thus, these structural changes might not cause much destabilization by Q4R.

Charged Residues on the Surface of Thermophilic Proteins—We have designed the biased amino acid composition of the protein surface in *E. coli* RNase HI based on the amino acid sequence of thermophilic bacterial RNases HI, expecting to enhance stability of

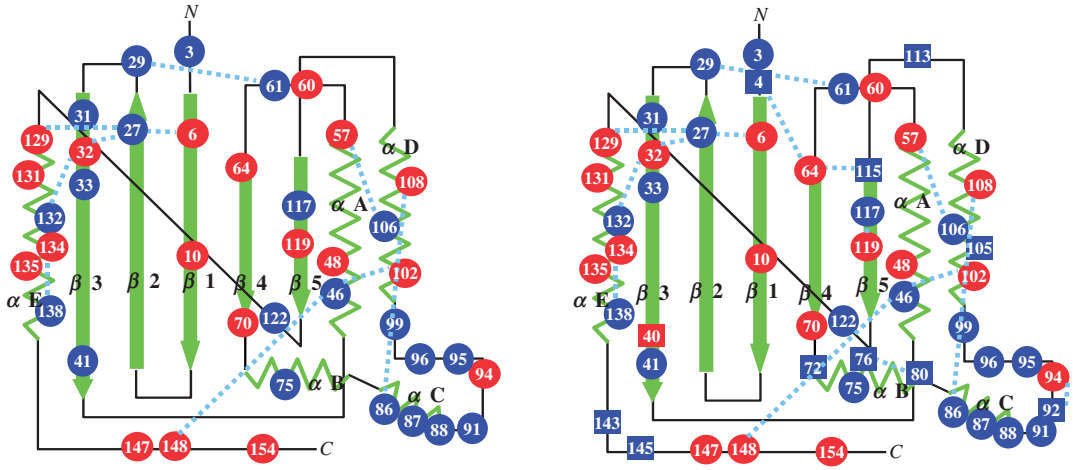

Fig. 7. **Schematic views of the charged residues in wild protein and 11C variant.** The green arrows, the green spirals and the black bars represent β-strands, α-helices and other segments, respectively. Original residues are shown in circles and introduced residues are shown in squares. Red and blue including a numeral are positive and negative ionizable residues, respectively, at the indicated position. Light blue dotted lines represent ion pairs within 5.0 Å between positive and negative ionizable residues on protein surface.

the *E. coli* enzyme. However, a decrease in stability was observed for all variants, although the mutated residues were successfully introduced on the protein surface with no significant overall structural change and a 4.7–11.8% increase of charge residues on the surface was confirmed, as depicted in Fig. 7.

Kumar and Nussinov (*39*) reported that citrate synthases from psychrophile, mesophile, and hyperthermophile share similar packing, burial of non-polar surface area, and main-chain hydrogen bonding. However, both psychrophilic and hyperthermophilic citrate synthases contain more charged residues, salt bridges and salt-bridge networks than the mesophilic protein. In the hyperthermophilic protein, salt bridges and their networks are largely clustered in the active-site region and at the dimer interface. In contrast, in the psychrophilic protein, they are more dispersed throughout the structure. This trend is important for both the heat and cold adaptation of citrate synthase. It has also been reported that the optimization of electrostatic properties at the surface is important for the high stability of thermophilic proteins (*30*, *40*, *41*). The present study clearly confirms that the omnipresence of charged residues on the protein surface alone is inadequate for stabilization. Furthermore, we reveal here that unfavourable interactions by charged residues on the surface are more critical for protein stability. These results indicate that the surface of thermophilic proteins is better optimized electrostatically, and that, in particular, the repulsive electrostatic interactions are reduced, adapting them to their environment. Therefore, protein stabilization requires a fine placement of surface-charged residues.

Protein Stabilization—Phylogenetic comparison of protein families has already been used to improve stability by consensus design (*42–44*). In this work, we have designed six variants of *E. coli* RNase HI based on the difference in the amino acid sequences between *E. coli* RNase HI and its thermophilic counterpart, so that the content of the surface-charged residues of *E. coli* RNase HI increases. The stability of all variants, however, is lower than that of the wild-type protein.

A previous study of *in vivo* molecular evolution, allowing adaptation to thermostability, indicated that only a few mutants, representing less than 1% of the possible missense mutations, were observed (*45*). Thus, the number of accessible mutational pathways toward increased fitness is surprisingly small. Moreover, the mutants isolated from the evolution experiment would not have been predicted by phylogenetic comparison of the thermophilic and mesophilic proteins (*45*). These findings correspond with the present results. Therefore, the sequence of a protein must critically code for structural characteristics other than amino acid composition to adapt to the environment.

We would like to thank SPring-8 for X-ray diffraction experiments (2005A0784, 2005B0017 and 2005BC05A44 XU-7209-N). This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas 'Systems Genomics'; by a Grant-in-Aid for the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; and by an Industrial Technology Research Grant Program from the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

REFERENCES

1. Chakravarty, S. and Varadarajan, R. (2000) Elucidation of determinants of protein stability through genome sequence analysis. *FEBS Lett.* **470**, 65–69
2. Haney, P.J., Badger, J.H., Buldak, G.L., Reich, C.I., Woese, C.R., and Olsen, G.J. (1999) Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely *thermophilic Methanococcus* species. *Proc. Natl Acad. Sci. USA* **96**, 3578–3583
3. Kumar, S., Tsai, C.J., and Nussinov, R. (2000) Factors enhancing protein thermostability. *Protein Eng.* **13**, 179–191

4. Fukuchi, S. and Nishikawa, K. (2001) Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *J. Mol. Biol.* **309**, 835–843
5. Crouch, R.J. and Dirksen, M.-L. (1982) Ribonucleases H. in *Nucleases* (Roberts, S.M. and Linn, R.J., eds.) pp. 211–241, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
6. Ohtani, N., Haruki, M., Morikawa, M., Crouch, R.J., Itaya, M., and Kanaya, S. (1999) Identification of the genes encoding Mn^{2+}-dependent RNase HII and Mg^{2+}-dependent RNase HIII from *Bacillus subtilis*: classification of RNases H into three families. *Biochemistry* **38**, 605–618
7. Arudchandran, A., Cerritelli, S., Narimatsu, S., Itaya, M., Shin, D.Y., Shimada, Y., and Crouch, R.J. (2000) The absence of ribonuclease H1 or H2 alters the sensitivity of *Saccharomyces cerevisiae* to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair. *Genes Cells* **5**, 789–802
8. Itaya, M., Omori, A., Kanaya, S., Crouch, R.J., Tanaka, T., and Kondo, K. (1999) Isolation of RNase H genes that are essential for growth of *Bacillus subtilis* 168. *J. Bacteriol.* **181**, 2118–2123
9. Kogoma, T. and Foster, P.L. (1998) Physiological functions of *E. coli* RNase HI. in *Ribonucleases H* (Crouch, R.J. and Toulme, J.J., eds.) pp. 39–66, INSERM, Paris
10. Qiu, J., Qian, Y., Frank, P., Wintersberger, U., and Shen, B. (1999) *Saccharomyces cerevisiae* RNase H functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease. *Mol. Cell. Biol.* **19**, 8361–8371
11. Kanaya, S., Katsuda, C., Kimura, S., Nakai, T., Kitakuni, E., Nakamura, H., Katayanagi, K., Morikawa, K., and Ikehara, M. (1991) Stabilization of *Escherichia coli* ribonuclease H by introduction of an artificial disulfide bond. *J. Biol. Chem.* **266**, 6038–6044
12. Kimura, S., Oda, Y., Nakai, T., Katayanagi, K., Kitakuni, E., Nakai, C., Nakamura, H., Ikehara, M., and Kanaya, S. (1992) Effect of cavity-modulating mutations on the stability of *Escherichia coli* ribonuclease HI. *Eur. J. Biochem.* **206**, 337–343
13. Ishikawa, K., Okumura, M., Katayanagi, K., Kimura, S., Kanaya, S., Nakamura, H., and Morikawa, K. (1993) Crystal structure of ribonuclease H from *Thermus thermophilus* HB8 refined at 2.8 A resolution. *J. Mol. Biol.* **230**, 529–542
14. Katayanagi, K., Miyagawa, M., Matsushima, M., Kimura, S., Kanaya, S., Nakamura, H., and Morikawa, K. (1990) Three-dimensional structure of ribonuclease H from *E. coli*. *Nature* **347**, 306–309
15. Yang, W., Hendrickson, W.A., Crouch, R.J., and Satow, Y. (1990) Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science* **249**, 1398–1405
16. Kanaya, S. and Itaya, M. (1992) Expression, purification, and characterization of a recombinant ribonuclease H from *Thermus thermophilus* HB8. *J. Biol. Chem.* **267**, 10184–10192
17. Kanaya, S., Oobatake, M., Nakamura, H., and Ikehara, M. (1993) pH-dependent thermostabilization of *Escherichia coli* ribonuclease HI by histidine to alanine substitutions. *J. Biotechnol.* **28**, 117–136
18. Ohtani, N., Haruki, M., Muroya, A., Morikawa, M., and Kanaya, S. (2000) Characterization of ribonuclease HII from *Escherichia coli* overproduced in a soluble form. *J. Biochem. (Tokyo)* **127**, 895–899
19. Horton, R.M., Cai, Z.L., Ho, S.N., and Pease, L.R. (1990) Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques* **8**, 528–535
20. Chon, H., Nakano, R., Ohtani, N., Haruki, M., Takano, K., Morikawa, M., and Kanaya, S. (2004) Gene cloning and biochemical characterizations of thermostable ribonuclease HIII from *Bacillus stearothermophilus*. *Biosci. Biotechnol. Biochem.* **68**, 2138–2147
21. Goodwin, T.W. and Morton, R.A. (1946) The spectrophotometric determination of tyrosine and tryptophan in proteins. *Biochem. J.* **40**, 628–632
22. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
23. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326
24. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta. Crystallogr. A.* **47**(Pt 2), 110–119
25. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta. Crystallogr. D. Biol. Crystallogr.* **54**, 905–921
26. Shirley, B.A., Stanssens, P., Hahn, U., and Pace, C.N. (1992) Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **31**, 725–732
27. Anderson, D.E., Becktel, W.J., and Dahlquist, F.W. (1990) pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29**, 2403–2408
28. Haruki, M., Noguchi, E., Kanaya, S., and Crouch, R.J. (1997) Kinetic and stoichiometric analysis for the binding of *Escherichia coli* ribonuclease HI to RNA-DNA hybrids using surface plasmon resonance. *J. Biol. Chem.* **272**, 22015–22022
29. Koide, A., Jordan, M.R., Horner, S.R., Batori, V., and Koide, S. (2001) Stabilization of a fibronectin type III domain by the removal of unfavorable electrostatic interactions on the protein surface. *Biochemistry* **40**, 10326–10333
30. Loladze, V.V., Ibarra-Molero, B., Sanchez-Ruiz, J.M., and Makhatadze, G.I. (1999) Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry* **38**, 16419–16423
31. Marshall, S.A., Morgan, C.S., and Mayo, S.L. (2002) Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol.* **316**, 189–199
32. Sanchez-Ruiz, J.M. and Makhatadze, G.I. (2001) To charge or not to charge? *Trends Biotechnol.* **19**, 132–135
33. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K., and Baker, D. (2005) Progress in modeling of protein structures and interactions. *Science* **310**, 638–642
34. Bogin, O., Levin, I., Hacham, Y., Tel-Or, S., Peretz, M., Frolow, F., and Burstein, Y. (2002) Structural basis for the enhanced thermal stability of alcohol dehydrogenase mutants from the mesophilic bacterium *Clostridium beijerinckii*: contribution of salt bridging. *Protein Sci.* **11**, 2561–2574
35. Hennig, M., Darimont, B., Sterner, R., Kirschner, K., and Jansonius, J.N. (1995) 2.0 A structure of indole-3-glycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinants of protein stability. *Structure* **3**, 1295–1306
36. Tanaka, T., Sawano, M., Ogasahara, K., Sakaguchi, Y., Bagautdinov, B., Katoh, E., Kuroishi, C., Shinkai, A., Yokoyama, S., and Yutani, K. (2006) Hyper-thermostability of CutA1 protein, with a denaturation temperature of nearly 150 degrees C. *FEBS Lett.* **580**, 4224–4230
37. Horovitz, A. and Fersht, A.R. (1990) Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. *J. Mol. Biol.* **214**, 613–617

38. Yip, K.S., Stillman, T.J., Britton, K.L., Artymiuk, P.J., Baker, P.J., Sedelnikova, S.E., Engel, P.C., Pasquo, A., Chiaraluce, R., and Consalvi, V. (1995) The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**, 1147–1158
39. Kumar, S. and Nussinov, R. (2004) Different roles of electrostatics in heat and in cold: adaptation by citrate synthase. *Chembiochem.* **5**, 280–290
40. Chakravarty, S. and Varadarajan, R. (2002) Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry* **41**, 8152–8161
41. Makhatadze, G.I., Loladze, V.V., Gribenko, A.V., and Lopez, M.M. (2004) Mechanism of thermostabilization in a designed cold shock protein with optimized surface electrostatic interactions. *J. Mol. Biol.* **336**, 929–942
42. Forrer, P., Binz, H.K., Stumpp, M.T., and Pluckthun, A. (2004) Consensus design of repeat proteins. *Chembiochem.* **5**, 183–189
43. Van den Burg, B., Vriend, G., Veltman, O.R., Venema, G., and Eijsink, V.G. (1998) Engineering an enzyme to resist boiling. *Proc. Natl. Acad. Sci. USA* **95**, 2056–2060
44. Watanabe, K., Ohkuri, T., Yokobori, S., and Yamagishi, A. (2006) Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. *J. Mol. Biol.* **355**, 664–674
45. Counago, R., Chen, S., and Shamoo, Y. (2006) In vivo molecular evolution reveals biophysical origins of organismal fitness. *Mol. Cell.* **22**, 441–449