BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

# Investigating the Structural Dependence of Protein Stabilization by Amino Acid Substitution

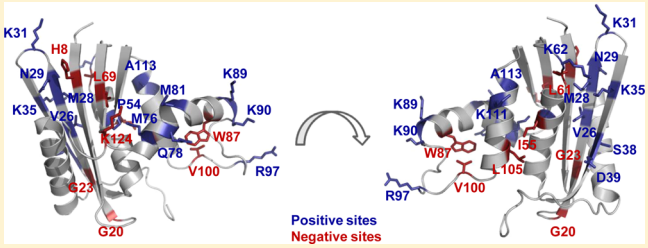
Takashi Tadokoro,[†,§] Hirotaka Kazama,[†] Yuichi Koga,[†] Kazufumi Takano,*[,†,‡] and Shigenori Kanaya[†]

[†]Department of Material and Life Science, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[‡]Department of Biomolecular Chemistry, Kyoto Prefectural University, 1-5 Hangi-cho, Shimogamo, Sakyo-ku, Kyoto 606-8522, Japan

Supporting Information

**ABSTRACT:** A goal of protein engineering technology is developing methods to increase protein stability. However, rational design of stable proteins is difficult because the stabilization mechanism of proteins is not fully understood. In this study, we examined the structural dependence of protein stabilization by introducing single amino acid substitution into ribonuclease H1 from the psychotropic bacterium *Shewanella oneidensis* MR-1 (So-RNase H1), which was our model protein. We performed saturation mutagenesis at various sites. Mutations that stabilized So-RNase H1 were screened using an RNase H-dependent temperature-sensitive *Escherchia coli* strain. Stabilizing mutations were identified by the suppressor mutagenesis method. This method yielded 39 stabilized mutants from 513 mutations at 27 positions. This suggested that more than 90% of mutations caused destabilization even in a psychotropic protein. However, 17 positions had stabilizing mutations, indicating that the stabilization factors were dispersed over many positions. Interestingly, the identified mutations were distributed mainly at exposed or nonconserved sites. These results provide a novel strategy for protein stabilization.




An important goal of protein engineering is designing variants that enhance the conformational stability of proteins. To date, extensive studies have demonstrated the stabilization mechanisms of individual model proteins.[1−4] Various interactions between amino acid side chains within protein molecules are reported to contribute to protein stability. These interactions include hydrogen bonds,[5] electrostatic interactions,[6−8] metal bindings,[9] and hydrophobic effects.[10,11] However, rationally designing a stable protein is difficult because the effects of amino acid changes on protein stability often differ depending on the surrounding environment, the substituted residues, and structural changes caused by mutations. Therefore, experimental analysis is needed to determine the influence of particular residues on the overall stability of a protein and to develop methods that can be used to stabilize proteins.

The technique of directed evolution mimics natural selection, enabling evolution and adaptation of enzymes under controlled conditions with well-defined selective pressures.[12−14] Directed evolution has been shown to be effective at modifying enzyme functions.[12,15] The suppressor mutation method is similar to this technique.[16] In the suppressor mutation method, a mutant enzyme with decreased activity or stability is constructed. Then, random mutagenesis is performed on the mutant enzyme, and the resultant large mutant enzyme library is screened for second-site revertants. Finally, the suppressor mutations identified in the second-site revertants are introduced into the wild-type enzyme. This method has been used to construct *Escherichia coli* ribonuclease H1 (Ec-RNaseH1) variants with increased stability,[16] *Thermus thermophilus* RNase H1 (Tt-RNase H1) variants with enhanced activity,[17] and *Shewanella oneidensis* MR-1 RNase H1 (So-RNase H1) variants with increased stability.[18] *E. coli* strain MIC3001 with *rnh-339::cat* and *recB270*(Ts), which shows an RNase H-dependent temperature-sensitive (ts) growth phenotype,[19] was used for these studies. In this study, we applied the suppressor mutation method to identify factors in protein stabilization.

So-RNase H1 is a monomeric protein from a psychrotrophic bacterium. It has 158 amino acid residues, and its amino acid sequence is 67% and 45% identical to those of its mesophilic and thermophilic counterparts, Ec-RNase H1 and Tt-RNase H1, respectively. RNase H is widely present in bacteria, archaea, eukaryotes, and retroviruses[20] and is involved in DNA replication, repair, and transcription.[21−26] The crystal structure of So-RNase H1 has been determined and resembles the structure of Ec-RNase H1 and Tt-RNase H1.[27] However, So-RNase H1 is much less stable than Ec-RNase H1, with a $T_m$ of 22.4 °C and $\Delta G(H_2O)$ of 12.5 kJ/mol.[27] Mutations that stabilize So-RNase H1 have been determined by structure-based design and random mutagenesis.[18,27,28] For these reasons and because of its thermolability, So-RNase H1 is a good model for determining factors for protein stabilization.

**Received:** January 20, 2013
**Revised:** March 31, 2013
**Published:** April 2, 2013

   dx.doi.org/10.1021/bi400076f | *Biochemistry* 2013, 52, 2839−2847

Table 1. Summary of the Comprehensive Mutagenesis and Screening

| positive sites for stabilization | | | | | negative sites | | | | |
|---|---|---|---|---|---|---|---|---|---|
| targeted site | secondary structure | solvent accessibility[a] | conservation[b] | mutation | targeted site | secondary structure | solvent accessibility[a] | conservation[b] | mutation |
| I26 | β | − | NC | V | H8 | β | + | NC | none |
| M28 | β | − | NC | L | G20 | loop | − | C | none |
| N29 | β | + | NC | R,I,Q,T,K,[c]H,S | G23 | β | − | C | none |
| K31 | loop | + | NC | R,Y,Q,G,N | I55 | α | − | C | none |
| K35 | β | + | C | C | L61 | loop | + | C | none |
| S38 | β | + | C | A,T | L69 | β | − | C | none |
| E39 | β | + | NC | G[c] | W87 | α | + | C | none |
| P54 | α | − | NC | A,V | V100 | loop | − | C | none |
| K62 | loop | + | C | M,P | L105 | α | − | C | none |
| M76 | α | − | NC | V,[c]L | K124 | loop | + | C | none |
| Q78 | α | + | NC | R,L,M | | | | | |
| M81 | α | + | NC | T[c] | | | | | |
| K89 | α | + | C | V | | | | | |
| K90 | α | + | NC | L,N,[c]W | | | | | |
| R97 | loop | + | NC | G[d] | | | | | |
| K111 | α | + | NC | R,V,Q,E | | | | | |
| A113 | α | − | NC | C,I | | | | | |

[a]Plus and minus indicate exposed and buried, respectively. Solvent accessibility was calculated using a probe radius of 1.4 Å. Amino acid residues with more than 10 Å$^2$ solvent accessibility were defined as exposed (+), whereas amino acid residues with less than 10 Å$^2$ accessibility were defined as buried (−). [b]C and NC indicate evolutionarily conserved and not conserved, respectively. Among these ten RNase H1 sequences (Figure 1A), residues identical in more than eight sources were counted as conserved residues (highlighted in black in Figure 1A); others were defined as nonconserved residues (not highlighted in Figure 1A). [c]Indicated mutations were also identified previously.[18] [d]Indicated mutation was also identified previously.[27]

In this study, we investigated the structural dependence of protein stabilization mechanisms using So-RNase H1 as a model protein. Using the suppressor mutation method, we identified mutations at 17 positions, including 10 new sites, that stabilized the wild-type protein. It is noted that the randomizing mutagenic primer actually examined all 20 amino acids, and selected mutants that complement the RNase H1-dependent ts phenotype of the mutant strain were more stable than the wild type. Our results and analyses suggested that the stabilizing mutations were distributed mainly at exposed or nonconserved sites.

## MATERIALS AND METHODS

**Cells and Plasmids.** *E. coli* MIC3001 [F$^-$, supE44, supF58, lacY1 or Δ(lacIZY)6, trpR55, galK2, galT22, metB1, hsdR14-(rK$^-$mK$^+$), rnhA339::cat, recB270][19] and *E. coli* MIC2067 [F$^-$, λ$^-$, IN(rrnD-rrnE)1, rnhA339::cat, rnhB716::kam][29] were kindly donated by Dr. M. Itaya. λDE3 lysogens of these strains, *E. coli* MIC3001(DE3) and MIC2067(DE3), were previously constructed in our laboratory.[30] Plasmid pET500MC with a unique ClaI site at the 3′-terminal region of the So-RNase H1 gene for overproduction and plasmid pET500MC153 with the same ClaI site of pET500MC encoding truncated So-RNase H1 (153-RNase H1) were constructed previously.[18]

**Site-Directed Saturation Mutagenesis and Screening.** Plasmid pET500MC153, encoding 153-RNase H1, a deletion mutant of So-RNase H1, was used as a template for mutagenesis. We selected 27 different sites within the 153 amino acids of 153-RNase H1 for comprehensive mutagenesis based on the secondary structure, solvent accessibility, and evolutionary conservation (Figure 1 and Table 1). A library of 153-RNase H1 derivatives was generated using a KOD-Plus-Mutagenesis Kit (Toyobo, Osaka, Japan). A total of 19 amino acid substitutions were randomly introduced into each selected position using mutagenic PCR primers containing an NNS sequence (N = A, T, G, or C and S = C or G). DNA oligos were synthesized by Hokkaido System Science (Sapporo, Japan) and Life Technologies Japan (Tokyo, Japan).

The resulting library of pET500MC153 derivatives was used to transform *E. coli* MIC3001(DE3) to screen for revertants that restored the ability of 153-RNase H1 to complement the ts growth phenotype of the *E. coli* strain. Transformants were spread on LB-agar plates with 50 mg/L ampicillin and 30 mg/L chloramphenicol, and plates were incubated at both 30 and 42 °C. Colonies grown at 42 °C were selected as revertants, and plasmid DNA was isolated from each clone. After confirming that each plasmid complemented the ts growth phenotype of MIC3001(DE3), we determined the nucleotide sequences of the mutant 153-RNase H1 genes.

**Overproduction and Purification.** Derivatives of pET500MC for overproduction of So-RNase H1 with suppressor mutations were constructed by replacing the *Nde*I-*Cla*I fragment of pET500MC containing the So-RNase H1 gene with corresponding fragments from the pET500MC153 derivatives with suppressor mutations, as described previously.[18] Overproduction and purification of mutant So-RNase H1 proteins were as described for the wild-type protein.[27] Protein concentration was determined from the UV absorption at 280 nm assuming that the absorption coefficient at this wavelength (2.1 for a 0.1% solution) was not changed by the mutations.

**Circular Dichroism.** Far-UV circular dichroism (CD) spectra were recorded on a J-725 spectropolarimeter (JASCO, Tokyo, Japan) at 4 °C. Protein was dissolved in 10 mM sodium acetate (pH 5.5) at 0.1−0.2 mg/mL with an optical path length of 2 mm. Mean residue ellipticity, θ, which

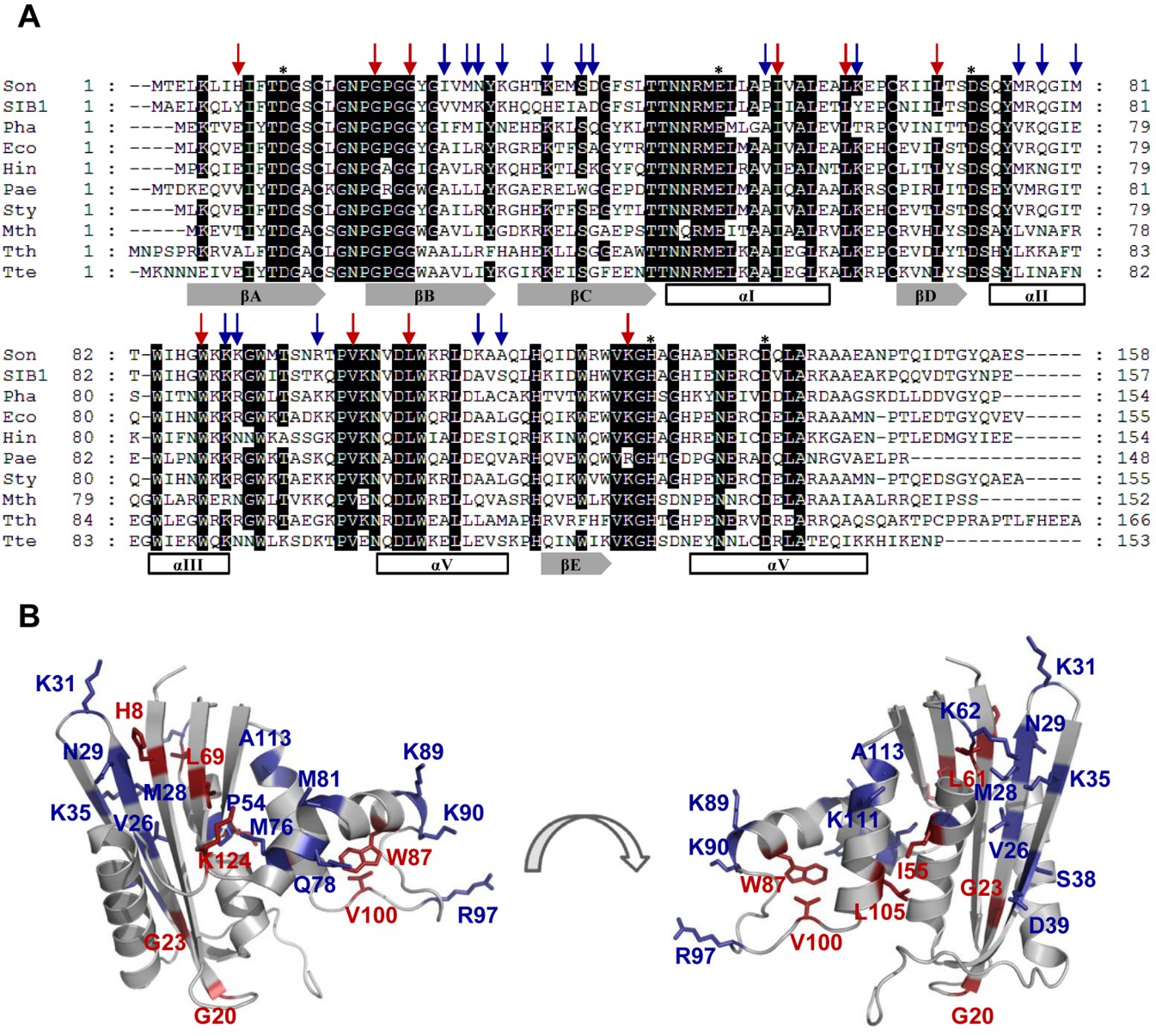



**Figure 1.** Selection of the target amino acids for comprehensive mutagenesis. (A) Alignment of amino acid sequences of RNases H1. The amino acid sequences of RNases H1 from Son, *S. oneidensis* MR-1; SIB1, *Shewanella* sp. SIB1; Pha, *P. haloplanktis*; Eco, *E. coli*; Hin, *H. influenzae*; Pae, *P. aeruginosa*; Sty, *S. typhimurium*; Mth, *M. thermoacetica*; Tth, *T. thermophilus*; Tte, *T. tengcongensis* are aligned. The ranges of the secondary structures of So-RNase H1 are shown below the sequences. The positions of the five acidic active site residues are denoted with asterisks. The amino acid residues that are conserved in more than eight proteins are highlighted in black. Mutated amino acid residues in this study are indicated by arrows: positive (stabilized mutants were obtained) by blue arrows and negative (no mutants were obtained) by red arrows. (B) Crystal structure of So-RNase H1. Side chains of the mutated amino acid residues in this study are indicated with stick models: positive in blue and negative in red.

has units of degrees square centimeter per decimole, was calculated using an average amino acid molecular weight of 110.

**Thermal Denaturation.** Thermal denaturation curves of So-RNase H1 and its derivatives were measured as described previously.[27] Proteins were dissolved in 10 mM sodium acetate (pH 5.5) with or without 1 M guanidine hydrochloride (GdnHCl) at 0.1−0.2 mg/mL. The optical path length was 2 mm. The temperature of the protein solution was linearly increased by approximately 1.0 °C/min. Thermal denaturation of proteins was reversible only in the presence of 1 M GdnHCl. The temperature of the midpoint of the transition, $T_m$, was calculated from curve fitting of CD values versus temperature data based on least-squares analysis.

## RESULTS

**Site-Directed Saturation Mutagenesis and Screening.** To examine the structural dependence of protein stability, we used So-RNase H1 as a model protein. First, we chose target amino acid residues within the So-RNase H1 sequence to introduce mutations according to three factors: (1) secondary structure, (2) solvent accessibility, and (3) evolutionary conservation. The three factors were chosen for the following reasons: Secondary structure is an important component for protein structure and stability. Indeed, it is thought that a reduced helix stability and a reduced number of proline residues at the loop regions are the structural features in a psychrophilic protein (unstable protein), suggesting that the stability of the secondary structure is one of the key factors for protein


dx.doi.org/10.1021/bi400076f | *Biochemistry* 2013, 52, 2839−2847

stability.[31] Solvent accessibility could be related to the possibility of interactions within the protein molecule. For example, hydrophobic effect is known to be one of the factors of protein stability for both mesophilic and thermo/hyperthermophilic proteins.[4,32] However, hydrophobic interactions could be restricted to the buried environment in the protein. On the other hand, the residues at the protein surface could have the chance to acquire different stabilization factors, such as hydrogen bonds and electrostatic interactions. Evolutionary conservation could be an important factor for protein stability because conserved amino acid residues in a protein have been selected to maintain protein structure and function through natural selective pressure. Conversely, nonconserved amino acid residues may still have the chance to be substituted to improve protein stability and function. On the basis of these hypotheses, we have defined each factor. Secondary structure was based on the three-dimensional structure of So-RNase H1 (PDB entry 2E4L). Solvent accessibility was calculated using a probe radius of 1.4 Å. Amino acid residues with more than 10 Å$^2$ solvent accessibility were defined as exposed residues ("+" in Table 1); amino acid residues with less than 10 Å$^2$ accessibility were defined as buried residues ("−" in Table 1). To evaluate evolutionary conservation, the amino acid sequence of So-RNase H1 was compared to nine bacterial RNases H1 (Figure 1A). Among these ten homologous sequences, residues identical in more than eight sources were counted as conserved residues (C in Table 1, Figure 1A); others were defined as nonconserved residues (NC in Table 1, Figure 1A). *S. oneidensis* MR-1, *Shewanella* sp. SIB1, and *Pseudoalteromonas haloplanktis* are psychrophiles/psychrotrophs; *E. coli*, *Haemophilus influenzae*, *Pseudomonas aeruginosa*, and *Salmonella typhimurium* are mesophiles; and *Moorella thermoacetica*, *T. thermophilus*, and *Thermoanaerobacter tengcongensis* are thermophiles. We selected 27 amino acid residues for mutagenesis: His8, Gly20, Gly23, Ile26, Met28, Asn29, Lys31, Lys35, Ser38, Glu39, Pro54, Ile55, Leu61, Lys62, Leu69, Met76, Gln78, Met81, Trp87, Lys89, Lys90, Arg97, Val100, Leu105, Lys111, Ala113, and Lys124 (Figure 1A,B, and Table 1).

At the 27 selected positions of So-RNase H1, we performed saturation PCR mutagenesis using mutagenic primers containing an NNS sequence that could encode all 20 amino acids. We screened for stabilizing mutations using RNase H-dependent ts *E. coli* MIC3001(DE3). We have used this strategy before to identify stabilizing mutations.[16−18] Because *E. coli* MIC3001-(DE3) with the wild-type 153-RNase H1 gene cannot grow at 42 °C,[18] transformants with 153-RNase H1 derivatives that grow at 42 °C were expected to contain stabilized (active) 153-RNase H1. At each mutation site, more than 1000 colonies (1000−10000) were examined. About ten colonies grew at 42 °C, indicating positive mutation sites, so approximately 0.01−0.1% of colonies complemented the ts phenotype of an RNase H mutant strain. We obtained revertant colonies that grew at 42 °C that had mutations at 17 different positions (Ile26, Met28, Asn29, Lys31, Lys35, Ser38, Glu39, Pro54, Lys62, Met76, Gln78, Met81, Lys89, Lys90, Arg97, Lys111, and Ala113) (Table 1, left). No colonies were obtained at 10 positions (His8, Gly20, Gly23, Ile55, Leu61, Leu69, Trp87, Leu105, and Lys124) (Table 1, right). Plasmid DNA was isolated from revertant colonies, and the 153-RNase H1 gene was sequenced. The results indicated that each position had at least one amino acid substitution, and at ten positions (Asn29, Lys31, Ser38, Pro54, Lys62, Met76, Gln78, Lys90, Lys111, and Ala113), multiple amino acid substitutions were obtained (Table 1). Notably, N29K, D39G, M76V, M81T, K90N, and R97G have been previously identified in our laboratory as stabilization variants of So-RNase H1.[18,28]

To evaluate the efficiency of comprehensive mutagenesis, we examined amino acid distributions in the mutant library. Theoretically, the NNS sequence encoded Arg, Leu, and Ser at the highest frequency, followed by Ala, Gly, Pro, Thr, and Val. The remaining amino acids showed lower frequencies, suggesting a small bias for mutagenic frequency (Figure 2A).

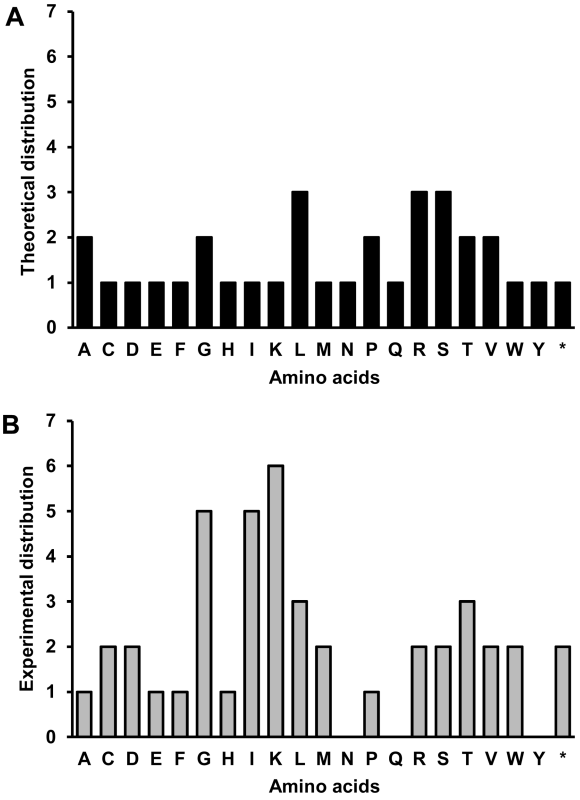



**Figure 2.** Experimental and theoretical amino acid distributions of NNS randomization. All 20 amino acids and stop codons are represented by one-letter notation and asterisks, respectively. (A) Theoretical amino acid distributions of NNS. (B) Experimental amino acid distributions of NNS. The mutations occurring at the position of Asn29 are evaluated. About 50 colonies grown at 30 °C were sequenced, and the colonies containing insertions or deletions are not included in the data.

We examined the amino acid distribution of the mutant library at the Asn29 position. About 50 colonies that grew at 30 °C were randomly picked; extracted plasmids were sequenced, and encoded amino acids were analyzed. In this library, 17 amino acids were determined, and two stop codons were obtained (Figure 2B). Among the sequenced colonies, Gly, Ile, Lys, and Thr were frequently observed (>3); Asn, Gln, and Tyr, theoretically low frequency amino acids, were missing. This result suggested a small bias for encoding amino acids. However, the missing amino acids in the Asn29 library were obtained in different libraries, such as Asn in the Lys31 and Lys90 libraries, Gln in the Asn29 and Lys111 libraries, and Tyr in the Lys31 library (Table 1). Thus, collectively, our mutagenesis using an NNS primer generated a mutant library in which the 20 amino acids were efficiently covered. Our screen showed that 39 mutations had increased stability from 513 mutations at 27 positions. This suggested that more than

90% of mutations cause destabilization even in psychotropic proteins.

**Stability of Mutant Proteins with Single Suppressor Mutations.** Next, to validate whether the suppressor mutations identified in the revertants of 153-RNase H1 enhanced the stability of the wild-type protein, mutant So-RNase H1 proteins with N29 derivatives were constructed (N29R, N29I, N29Q, N29T, N29K, N29H, and N29S). All mutant proteins were overproduced in *E. coli* in a soluble form and purified using the same methods as for the wild-type protein, giving a single band on sodium dodecyl sulfate−polyacrylamide gel electrophoresis (data not shown). The far-UV CD spectra were similar for mutant and wild-type proteins (data not shown), suggesting that the mutations did not significantly affect overall conformation. Mutant protein stability was analyzed by thermal denaturation. All mutant proteins reversibly unfolded with increasing temperature, similar to the wild type. These results indicated that all seven mutant proteins were more stable, with a $T_m$ increase of about 0.7−9.7 °C over that of the wild-type So-RNase H1 (Figure 3 and Table 2). These results suggested

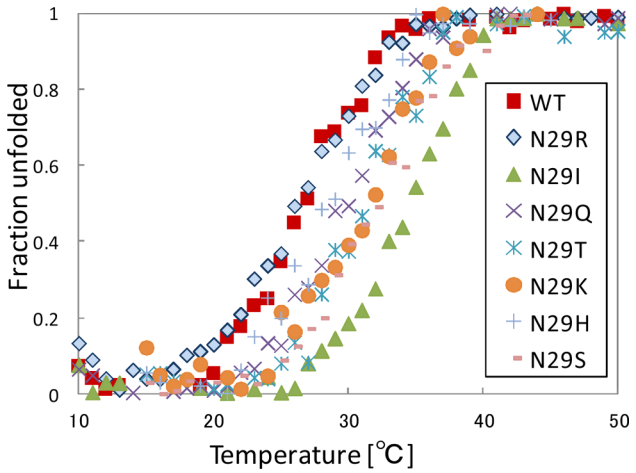



**Figure 3.** Thermal denaturation curves of the wild type (WT) and seven mutant proteins of So-RNase H1. These curves were measured at pH 5.5 in the presence of 1.0 M GdnHCl by monitoring the change in the CD value at 220 nm as described in Materials and Methods.

**Table 2. Melting Temperatures of the Wild Type and N29 Variants**

| protein | $T_m$ (°C) | $\Delta T_m$[a] (°C) |
|---|---|---|
| wild type | 25.8 | – |
| N29R | 26.5 | +0.7 |
| N29I | 35.5 | +9.7 |
| N29Q | 30.2 | +4.4 |
| N29T | 29.8 | +4.0 |
| N29K | 33.8 | +8.0 |
| N29H | 27.9 | +2.1 |
| N29S | 32.0 | +6.2 |

[a] $\Delta T_m = T_m$ (N29X) − $T_m$ (wild type).

that our screening system selected mutant proteins with increased stability. N29I was the most effective substitution (+9.7 °C), and N29R had the smallest effect on stabilization (+0.7 °C) (Table 2).

**Correlation between Primary Structure and Protein Stability.** We used a bioinformatics approach to evaluate structural characteristics and mutability using the primary structure of So-RNase H1, comparing the information with our experimental results. A Kyte−Doolittle scale was applied to delineate a protein's hydrophobic character,[33] and a relative mutability plot was generated using the ProtScale online server (http://web.expasy.org/protscale/) (Figure 4). To visualize

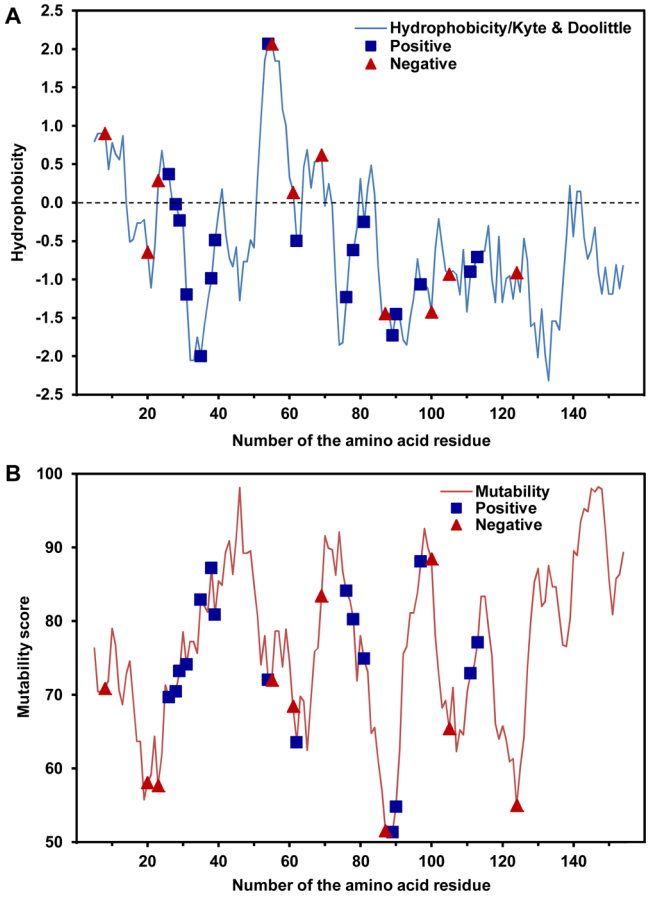



**Figure 4.** Hydrophobic character and relative mutability of So-RNase H1. In both plots, positive (stabilized mutants were obtained) and negative (no mutants were obtained) sites are indicated with blue squares and red triangles, respectively. (A) Hydropathy plot on the amino acids of So-RNase H1. The Kyte−Doolittle scale[33] was applied to generate the plots. Positive and negative values represent hydrophobicity and hydrophilicity, respectively. (B) Mutability plot on the amino acids of So-RNase H1. The plot was generated using the ProtScale online server (http://web.expasy.org/protscale/).

correlations, the sites mutated in this study were plotted on both diagrams. The hydropathy plot indicated that 14 of 17 positive sites were hydrophilic, one was neutral, and two were hydrophobic, suggesting that mutations at hydrophilic sites tended to stabilize the protein. Hydrophobic effects are a major factor in protein stability. Indeed, mutational analyses have demonstrated that deletion of hydrophobic side chains decreases stability, but mutations that fill a cavity with a hydrophobic side chain stabilize proteins.[34,35] Thus, hydrophobic amino acids are not suitable for stabilizing mutations, whereas hydrophilic amino acids are good targets for substitutions for protein stabilization.

Mutability scores increased as follows: K89 < W87 < K90 < K124 < G23 < G20 < K62 < L105 < L61 < I26 < M28 < H8 < P54, I55 < K111 < N29 < K31 < M81 < A113 < Q78 < E39 < K35 < L69 < M76 < S38 < R97 < V100. Of these residues,


dx.doi.org/10.1021/bi400076f | *Biochemistry* 2013, 52, 2839−2847

Table 3. Identified Mutations and Evolutionary Conservations[a]

| targeted site | identified mutations | thermophiles | mesophiles | no relatives | note |
|---|---|---|---|---|---|
| I26 | **V** | (A) | (A) | V | buried |
| M28 | **L** | L | L | | buried |
| N29 | **H,I,K,Q,R,S,T** | I, R | R (L) | H,K,Q,S,T | |
| K31 | **G,N,Q,R,Y** | G,R | R | G,N,Q,Y | |
| K35 | **C** | (K) | (K,R) | C | conserved |
| S38 | **A,T** | (S) | (W,S) | A,T | conserved |
| D39 | **G** | G | (A,E,G,K) | | |
| P54 | **A,V** | A | A,V | | buried |
| K62 | **M,P** | (K) | (K) | M,P | conserved |
| M76 | **L,V** | L | V (M) | | buried |
| Q78 | **L,M,R** | (K,N) | R (N,Q) | L,M | |
| M81 | **T** | T (R,N) | T | | |
| K89 | **V** | (K,R) | (K) | V | conserved |
| K90 | **L,N,W** | N (R) | N (R) | L,W | |
| R97 | **G** | G (K) | G (K) | | |
| K111 | **E,Q,R,V** | E,Q (L) | E (A) | R,V | |
| A113 | **C,I** | (A,M,S) | I (L,V) | C | buried |

[a]Evolutionary conservation is defined as described in Table 1 and Figure 1. Conserved amino acids, not identified in this study, are noted in parentheses (X1, X2, ...).

mutations stabilizing So-RNase H1 occurred at 11 of 13 sites with relatively high scores (>P54, I55), whereas 80% of negative sites (8/10) were in the group with lower scores (≤P54, I55). Taken together, these results suggest that using primary structure to predict sites to mutate to improve protein stability might be possible.

## DISCUSSION

Using a suppressor mutation method and an *E. coli* RNase H mutant strain, we identified 17 different mutation sites that stabilize So-RNase H1. Ten were new factors for the stabilization of So-RNase H1. This result indicated that stabilization factors are dispersed over many positions. In this study, we used So-RNase H1 as a model for isolating stabilizing mutations because of its low thermostability, 25 °C in $T_m$ (Figure 3 and Table 2). This suggests that our model protein is quite unstable and may have negative factors for protein stability. Cold-adapted proteins, like So-RNase H1, share several structural features related to their thermolability, such as a reduced number of ion pairs and hydrogen bonds, weakened hydrophobic interactions and packing at the core, an increased fraction of nonpolar surface area, a reduced surface hydrophilicity, a reduced helix stability, and a reduced number of proline residues at the loop regions.[31] Indeed, we have previously determined that the So-RNase H1 has an increased nonpolar surface area and a reduced number of ion pair networks as compared to Ec-RNase H1.[27] Thus, the removal or relaxation of the negative factors could also contribute to improving the stability of So-RNase H1.

Here we performed the experiment according to the three factors as mentioned above: (1) secondary structure, (2) solvent accessibility, and (3) evolutionary conservation. We found that the mutations in an $\alpha$-helix and $\beta$-sheet were more likely to stabilize the wild-type protein than mutations in a loop region and stabilizing mutations preferably occur at the molecular surface and nonconserved sites. These findings are provided because we focus on the three factors in this study. We analyzed each mutation in further detail, and on the basis of these analyses, we discuss each factor contributing to protein stabilization.

**Evolutionary Conservation and Protein Stability.** In this study, we found that the stabilized protein So-RNase H1 had mutations at four sites (Lys35, Ser38, Lys62, and Lys89) among 13 evolutionarily conserved positions tested, or about ∼30% (Table 1). In contrast, nine of ten negative sites, or 90%, were conserved amino acids (Table 1). This result suggested that conserved residues are not good targets for protein stabilization. Because stabilizing mutations were rarely obtained at conserved sites in the 19 amino acid substitutions that were examined, evolutionarily conserved amino acids might already contribute to protein stability to a large extent. Alternatively, conserved residues are essential for the maintenance of protein conformation and function, such as at catalytic sites. Several studies have shown that catalytic site residues, which are usually well-conserved across species, are not always optimized for conformational stability in activity−stability trade-offs.[36−41] Thus, conserved amino acids in proteins are not a crucial factor of protein stability but are necessary as the basic backbone of the protein structure and function.

Stabilizing mutations were obtained at 13 sites (Ile26, Met28, Asn29, Lys31, Glu39, Pro54, Met76, Gln78, Met81, Lys90, Arg97, Lys111, and Ala113) among 14 nonconserved positions, for about ∼90% (Table 1). Thus, nonconserved positions are likely good candidate sites for introducing mutations to increase protein stability. It should be noted that at 12 conserved sites, except Ile26, the amino acids were found to be replaced to thermophilic or mesophilic types of amino acid residues. Of them, 13 of a possible 24 amino acid substitutions to thermophilic types were observed, whereas 13 of a possible 26 amino acid substitutions to mesophilic types were observed (Table 3). Interestingly, at nonconserved and buried sites (Ile26, Met28, Pro54, Met76, and Ala113), except at Ala113, the stabilizing mutations are well-conserved with thermophilic and/or mesophilic counterparts (∼70%) despite the limited number of identified stabilizing mutations. We exclude Ala113 because the amino acid at this position is not conserved within mesophilic or thermophilic groups, whereas the residues at the other four sites are at least highly conserved within mesophilic or thermophilic groups (Table 3 and Figure S1 of Supporting Information). This may suggest that those sites in mesophilic

and thermophilic proteins are highly optimized for protein stability through natural selective pressure. On the other hand, the amino acid residues at exposed and nonconserved sites vary even in the mesophilic group or thermophilic group (Table 3). Moreover, we previously demonstrated that a combination of five point mutations at nonconserved positions enhanced the stability of So-RNase H1 to a level similar to that of a mesophilic counterpart, Ec-RNase H1.[28] Interestingly, only five amino acid substitutions in So-RNase HI were sufficient to stabilize the protein to the level of Ec-RNase H1, even though the proteins have about 50 amino acid differences. In general, focusing on evolutionarily nonconserved residues would be a guideline to efficiently improve protein stability.

**Buried/Exposed Sites and Protein Stability.** We found that mutations at five sites (Ile26, Met28, Pro54, Met76, and Ala113) among 11 buried amino acids, or 45%, enhanced the protein stability of the wild-type protein (Table 1), and the original amino acid residues were all substituted to hydrophobic residues (Ala, Val, or Leu). In the So-RNase H1 structure, Ile26, Met28, Pro54, and Met76 are involved in the formation of the hydrophobic core. Interestingly, upon comparison of these amino acids in So-RNase H1 with corresponding amino acids in RNase H1 homologues, the So-RNase H1 amino acids, except for Ile26, were replaced with amino acids conserved across mesophilic or thermophilic counterparts (Figure 1A and Figure S1 of Supporting Information). The hydrophobic core of hyperthermophilic or thermophilic proteins is rigid, whereas the core of psychrophilic protein is relatively soft, which yields loose molecular packing. Psychrophilic proteins are proposed to decrease their stability by acquiring a cavity within the protein core while maintaining the basic backbone of the protein to avoid disruption of the entire structure.[31,42] Therefore, substitutions at buried positions of thermophilic types of amino acids often alter protein stability. Stabilizing mutations at Met76 of So-RNase H1 were previously identified by suppressor mutagenesis.[18] We also demonstrated that cavity-filling mutations at Ala52 and Val74 in Ec-RNase H1, corresponding to Pro54 and Met76 in So-RNase H1, effectively increased the stability of Ec-RNase H1.[34] However, the buried residues, especially those in the hydrophobic core, might not be a good choice for rational design to obtain a stabilized protein because we identified either none or a limited number of mutations at conserved and buried sites (Table 1).

Mutations at 12 of 16 exposed sites, or 75%, stabilized the wild-type protein (Table 1). The frequency of the mutation at exposed sites is likely higher than that at buried sites. The stabilizing mutations at exposed sites may contribute to resolve a negative factor for stability of So-RNase H1, alteration of the fraction of nonpolar surface area or intramolecular interactions. Sites were substituted into a variety of amino acids with no correlations. Therefore, we are not able to propose any specific factors involved in protein stabilization from this result. Because the protein surface is prone to being influenced by the surrounding environment and is highly flexible at the molecular level, we might not be able to systematically design proper mutations. Stabilizing mutations were obtained at four sites (Lys35, Ser38, Lys62, and Lys89) in seven exposed and evolutionarily conserved sites (~57%). Comparing the crystal structures of So-, Ec-, and Tt-RNases H1 showed that the local conformations of these four sites varied, i.e., side chains were rotated or differently oriented (Figure S2 of Supporting Information). Thus, some of the conserved amino acids at the surface greatly contributed to the formation of the intramolecular network in one protein, but the importance of these residues in homologous protein might be decreased through environmental and evolutionary changes. This suggested that even conserved residues might be affected by a change in the surrounding environment if they are exposed to the environment. Nevertheless, we hypothesize that novel mutations to improve protein stability might occur at the molecular surface regardless of evolutionary conservation.

**Prediction of the Protein Stability from Primary Structure.** Bioinformatic approaches suggest the possibility of rational design of mutations based on primary structure. Our results suggested that hydrophilic amino acids, which are often exposed at the surface, were candidate mutation sites for improving protein stability, and that the mutability score was relatively reliable. This is important because the crystal structure of a target protein is not always available, so guidance is needed in designing mutations. Our analysis of the secondary structure of the mutation sites showed that eight of 11 sites in an $\alpha$-helix (~72%), six of nine sites in a $\beta$-sheet (~67%), and three of seven sites in a loop (43%) were positive for stabilization. Interestingly, although the number we examined was limited, mutations in an $\alpha$-helix and $\beta$-sheet were more likely to stabilize the wild-type protein than mutations in a loop region. This might be because loop structures are more flexible than other secondary structures. In $\alpha$-helix and $\beta$-sheet structures, amino acid main chains form hydrogen bonds, whereas such interactions between main chains are usually missing in loops. Therefore, an amino acid in a loop is difficult to replace with other amino acids without considering neighboring amino acids. We previously demonstrated that the replacement of a left-handed non-Gly residue in a loop structure with Gly stabilizes a wild-type protein.[28,43] Nonetheless, we cannot rule out the possibility that the secondary structure contributes to protein stability. Further research is needed.

**Tolerance to Random Amino Acid Mutations and Protein Stability.** Several studies found protein tolerance to random amino acid changes leading to an inactivation of the protein or disruption of protein conformation using rational mutational design. Consistent with our results, amino acid substitutions at conserved positions and/or buried positions often result in destabilization and/or inactivation, but mutations at nonconserved and/or exposed positions stabilize and/or activate proteins, including barnase,[44] T4 lysozyme,[45] and DNA glycosylase.[46] Because nonconserved residues at the protein surface are not usually crucial for enzymatic functions, mutations at these positions are expected to enhance protein stability.

**Conclusion.** In summary, we examined the effects of all 20 amino acids at 27 different sites of So-RNase H1 using randomizing mutagenic primers and an RNase H mutant strain to investigate the mechanisms of protein stabilization. So-RNase H1 is a good model for isolating stabilizing mutations because of its low thermolability. We identified at least one mutation at 17 different sites, including 10 new sites. Interestingly, the identified mutations were distributed mainly at exposed and nonconserved sites. Our findings provide guidance for designing stabilized mutant proteins.

## ■ ASSOCIATED CONTENT

### Supporting Information

Alignment of amino acid sequences of RNases H1 (A) and Ile26, Met28, Pro54, and Met76 (B) in the crystal structure of

So-RNase H1 (Figure S1), and side chain conformation of Lys35, Ser38, Lys62, and Lys89 in So-RNase H1, Ec-RNase H1, and Tt-RNase H1 (Figure S2). This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION


### Corresponding Author

*E-mail: takano@kpu.ac.jp. Tel/fax: +81-75-703-5654.

### Present Address

[§]T.T.: Laboratory of Molecular Gerontology, National Institute on Aging, 251 Bayview Blvd., Suite 100, Baltimore, MD 21224.



### Funding

This work was supported in part by a Grant (24380055) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and by an Industrial Technology Research Grant Program (09A03002) from the New Energy and Industrial Technology Development Organization (NEDO) of Japan.


### Notes

The authors declare no competing financial interest.

## ABBREVIATIONS

CD, circular dichroism; Ec-RNase H1, ribonuclease H1 from *E. coli*; GdnHCl, guanidine hydrochloride; So-RNase H1, ribonuclease H1 from *S. oneidensis* MR-1; ts, temperature-sensitive; Tt-RNase H1, ribonuclease H1 from *T. thermophilus*.

## REFERENCES


(1) Rozek, A., Powers, J. P., Friedrich, C. L., and Hancock, R. E. (2003) Structure-based design of an indolicidin peptide analogue with increased protease stability. *Biochemistry 42*, 14130−14138.

(2) Schweiker, K. L., and Makhatadze, G. I. (2009) Protein stabilization by the rational design of surface charge-charge interactions. *Methods Mol. Biol. (N.Y., NY, U.S.) 490*, 261−283.

(3) Spadiut, O., Leitner, C., Salaheddin, C., Varga, B., Vertessy, B. G., Tan, T. C., Divne, C., and Haltrich, D. (2009) Improving thermostability and catalytic activity of pyranose 2-oxidase from *Trametes multicolor* by rational and semi-rational design. *FEBS J. 276*, 776−792.

(4) Tanaka, S., Igarashi, S., Ferri, S., and Sode, K. (2005) Increasing stability of water-soluble PQQ glucose dehydrogenase by increasing hydrophobic interaction at dimeric interface. *BMC Biochem. 6*, 1.

(5) Alber, T., Sun, D. P., Wilson, K., Wozniak, J. A., Cook, S. P., and Matthews, B. W. (1987) Contributions of hydrogen bonds of Thr 157 to the thermodynamic stability of phage T4 lysozyme. *Nature 330*, 41−46.

(6) Chen, L., Cabrita, G. J., Otzen, D. E., and Melo, E. P. (2005) Stabilization of the ribosomal protein S6 by trehalose is counter-balanced by the formation of a putative off-pathway species. *J. Mol. Biol. 351*, 402−416.

(7) Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X., and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: Guidelines for protein engineering. *J. Mol. Biol. 327*, 1135−1148.

(8) Strickler, S. S., Gribenko, A. V., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T., Loladze, V. V., and Makhatadze, G. I. (2006) Protein stability and surface electrostatics: A charged relationship. *Biochemistry 45*, 2761−2766.

(9) Kuroki, R., Taniyama, Y., Seko, C., Nakamura, H., Kikuchi, M., and Ikehara, M. (1989) Design and creation of a $Ca^{2+}$ binding site in human lysozyme to enhance structural stability. *Proc. Natl. Acad. Sci. U.S.A. 86*, 6903−6907.

(10) Matsumura, M., Becktel, W. J., and Matthews, B. W. (1988) Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature 334*, 406−410.

(11) Takano, K., Yamagata, Y., and Yutani, K. (1998) A general rule for the relationship between hydrophobic effect and conformational stability of a protein: Stability and structure of a series of hydrophobic mutants of human lysozyme. *J. Mol. Biol. 280*, 749−761.

(12) You, L., and Arnold, F. H. (1996) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng. 9*, 77−83.

(13) Arnold, F. H., Wintrode, P. L., Miyazaki, K., and Gershenson, A. (2001) How enzymes adapt: Lessons from directed evolution. *Trends Biochem. Sci. 26*, 100−106.

(14) Brustad, E. M., and Arnold, F. H. (2011) Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol. 15*, 201−210.

(15) Pulido, M. A., Koga, Y., Takano, K., and Kanaya, S. (2007) Directed evolution of Tk-subtilisin from a hyperthermophilic archaeon: Identification of a single amino acid substitution responsible for low-temperature adaptation. *Protein Eng., Des. Sel. 20*, 143−153.

(16) Haruki, M., Noguchi, E., Akasako, A., Oobatake, M., Itaya, M., and Kanaya, S. (1994) A novel strategy for stabilization of Escherichia coli ribonuclease HI involving a screen for an intragenic suppressor of carboxyl-terminal deletions. *J. Biol. Chem. 269*, 26904−26911.

(17) Hirano, N., Haruki, M., Morikawa, M., and Kanaya, S. (2000) Enhancement of the enzymatic activity of ribonuclease HI from *Thermus thermophilus* HB8 with a suppressor mutation method. *Biochemistry 39*, 13285−13294.

(18) Tadokoro, T., Matsushita, K., Abe, Y., Rohman, M. S., Koga, Y., Takano, K., and Kanaya, S. (2008) Remarkable stabilization of a psychrotrophic RNase HI by a combination of thermostabilizing mutations identified by the suppressor mutation method. *Biochemistry 47*, 8040−8047.

(19) Itaya, M., and Crouch, R. J. (1991) A combination of RNase H (rnh) and recBCD or sbcB mutations in *Escherichia coli* K12 adversely affects growth. *Mol. Gen. Genet. 227*, 424−432.

(20) Ohtani, N., Haruki, M., Morikawa, M., and Kanaya, S. (1999) Molecular diversities of RNases H. *J. Biosci. Bioeng. 88*, 12−19.

(21) 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.

(22) Cerritelli, S. M., Frolova, E. G., Feng, C., Grinberg, A., Love, P. E., and Crouch, R. J. (2003) Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol. Cell 11*, 807−815.

(23) Haruki, M., Tsunaka, Y., Morikawa, M., and Kanaya, S. (2002) Cleavage of a DNA-RNA-DNA/DNA chimeric substrate containing a single ribonucleotide at the DNA-RNA junction with prokaryotic RNases HII. *FEBS Lett. 531*, 204−208.

(24) Qiu, J., Qian, Y., Frank, P., Wintersberger, U., and Shen, B. (1999) *Saccharomyces cerevisiae* RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease. *Mol. Cell. Biol. 19*, 8361−8371.

(25) Rydberg, B., and Game, J. (2002) Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proc. Natl. Acad. Sci. U.S.A. 99*, 16654−16659.

(26) Sparks, J. L., Chon, H., Cerritelli, S. M., Kunkel, T. A., Johansson, E., Crouch, R. J., and Burgers, P. M. (2012) RNase H2-initiated ribonucleotide excision repair. *Mol. Cell 47*, 980−986.

(27) Tadokoro, T., You, D. J., Abe, Y., Chon, H., Matsumura, H., Koga, Y., Takano, K., and Kanaya, S. (2007) Structural, thermodynamic, and mutational analyses of a psychrotrophic RNase HI. *Biochemistry 46*, 7460−7468.

(28) Rohman, M. S., Tadokoro, T., Angkawidjaja, C., Abe, Y., Matsumura, H., Koga, Y., Takano, K., and Kanaya, S. (2009) Destabilization of psychrotrophic RNase HI in a localized fashion as revealed by mutational and X-ray crystallographic analyses. *FEBS J. 276*, 603−613.

(29) 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.
(30) 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. 127*, 895−899.
(31) Feller, G., and Gerday, C. (2003) Psychrophilic enzymes: Hot topics in cold adaptation. *Nat. Rev. Microbiol. 1*, 200−208.
(32) Dong, H., Mukaiyama, A., Tadokoro, T., Koga, Y., Takano, K., and Kanaya, S. (2008) Hydrophobic effect on the stability and folding of a hyperthermophilic protein. *J. Mol. Biol. 378*, 264−272.
(33) Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol. 157*, 105−132.
(34) Akasako, A., Haruki, M., Oobatake, M., and Kanaya, S. (1997) Conformational stabilities of *Escherichia coli* RNase HI variants with a series of amino acid substitutions at a cavity within the hydrophobic core. *J. Biol. Chem. 272*, 18686−18693.
(35) Ishikawa, K., Nakamura, H., Morikawa, K., and Kanaya, S. (1993) Stabilization of *Escherichia coli* ribonuclease HI by cavity-filling mutations within a hydrophobic core. *Biochemistry 32*, 6171−6178.
(36) Beadle, B. M., and Shoichet, B. K. (2002) Structural bases of stability-function tradeoffs in enzymes. *J. Mol. Biol. 321*, 285−296.
(37) Kanaya, S., Oobatake, M., and Liu, Y. (1996) Thermal stability of *Escherichia coli* ribonuclease HI and its active site mutants in the presence and absence of the $Mg^{2+}$ ion. Proposal of a novel catalytic role for Glu48. *J. Biol. Chem. 271*, 32729−32736.
(38) Meiering, E. M., Serrano, L., and Fersht, A. R. (1992) Effect of active site residues in barnase on activity and stability. *J. Mol. Biol. 225*, 585−589.
(39) Mukaiyama, A., Haruki, M., Ota, M., Koga, Y., Takano, K., and Kanaya, S. (2006) A hyperthermophilic protein acquires function at the cost of stability. *Biochemistry 45*, 12673−12679.
(40) Shoichet, B. K., Baase, W. A., Kuroki, R., and Matthews, B. W. (1995) A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. U.S.A. 92*, 452−456.
(41) Zhi, W., Srere, P. A., and Evans, C. T. (1991) Conformational stability of pig citrate synthase and some active-site mutants. *Biochemistry 30*, 9281−9286.
(42) Kumar, S., and Nussinov, R. (2004) Different roles of electrostatics in heat and in cold: Adaptation by citrate synthase. *ChemBioChem 5*, 280−290.
(43) Kimura, S., Kanaya, S., and Nakamura, H. (1992) Thermostabilization of *Escherichia coli* ribonuclease HI by replacing left-handed helical Lys95 with Gly or Asn. *J. Biol. Chem. 267*, 22014−22017.
(44) Axe, D. D., Foster, N. W., and Fersht, A. R. (1998) A search for single substitutions that eliminate enzymatic function in a bacterial ribonuclease. *Biochemistry 37*, 7157−7166.
(45) Rennell, D., Bouvier, S. E., Hardy, L. W., and Poteete, A. R. (1991) Systematic mutation of bacteriophage T4 lysozyme. *J. Mol. Biol. 222*, 67−88.
(46) Guo, H. H., Choe, J., and Loeb, L. A. (2004) Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci. U.S.A. 101*, 9205−9210.