

**On the physics of thermal-stability changes upon mutations of a protein**

Shota Murakami, Hiraku Oshima, Tomohiko Hayashi, and Masahiro Kinoshita

Citation: *The Journal of Chemical Physics* **143**, 125102 (2015); doi: 10.1063/1.4931814

View online: <http://dx.doi.org/10.1063/1.4931814>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/143/12?ver=pdfcov>

Published by the [AIP Publishing](#)

---

**Articles you may be interested in**

[Confinement in nanopores can destabilize  \$\alpha\$ -helix folding proteins and stabilize the  \$\beta\$  structures](#)

*J. Chem. Phys.* **135**, 125101 (2011); 10.1063/1.3641482

[Protein stability at a carbon nanotube interface](#)

*J. Chem. Phys.* **134**, 125101 (2011); 10.1063/1.3558776

[Predicting stability of alpha-helical, orthogonal-bundle proteins on surfaces](#)

*J. Chem. Phys.* **133**, 115102 (2010); 10.1063/1.3479039

[Molecular simulation of protein dynamics in nanopores. I. Stability and folding](#)

*J. Chem. Phys.* **128**, 115105 (2008); 10.1063/1.2894299

[Phosphorylation effect on the GSSS peptide conformation in water: Infrared, vibrational circular dichroism, and circular dichroism experiments and comparisons with molecular dynamics simulations](#)

*J. Chem. Phys.* **126**, 235102 (2007); 10.1063/1.2738472

---



**AIP** | APL Photonics

*APL Photonics* is pleased to announce  
**Benjamin Eggleton** as its Editor-in-Chief



## On the physics of thermal-stability changes upon mutations of a protein

Shota Murakami,<sup>1</sup> Hiraku Oshima,<sup>2</sup> Tomohiko Hayashi,<sup>2</sup> and Masahiro Kinoshita<sup>2,a)</sup>

<sup>1</sup>Graduate School of Energy Science, Kyoto University, Uji, Kyoto 611-0011, Japan

<sup>2</sup>Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan

(Received 8 June 2015; accepted 14 September 2015; published online 30 September 2015)

It is of great interest from both scientific and practical viewpoints to theoretically predict the thermal-stability changes upon mutations of a protein. However, such a prediction is an intricate task. Up to now, significantly many approaches for the prediction have been reported in the literature. They always include parameters which are adjusted so that the prediction results can be best fitted to the experimental data for a sufficiently large set of proteins and mutations. The inclusion is necessitated to achieve satisfactorily high prediction performance. A problem is that the resulting values of the parameters are often physically meaningless, and the physicochemical factors governing the thermal-stability changes upon mutations are rather ambiguous. Here, we develop a new measure of the thermal stability. Protein folding is accompanied by a large gain of water entropy (the entropic excluded-volume (EV) effect), loss of protein conformational entropy, and increase in enthalpy. The enthalpy increase originates primarily from the following: The energy increase due to the break of protein-water hydrogen bonds (HBs) upon folding cannot completely be cancelled out by the energy decrease brought by the formation of protein intramolecular HBs. We develop the measure on the basis of only these three factors and apply it to the prediction of the thermal-stability changes upon mutations. As a consequence, an approach toward the prediction is obtained. It is distinguished from the previously reported approaches in the following respects: The parameters adjusted in the manner mentioned above are not employed at all, and the entropic EV effect, which is ascribed to the translational displacement of water molecules coexisting with the protein in the system, is fully taken into account using a molecular model for water. Our approach is compared with one of the most popular approaches, FOLD-X, in terms of the prediction performance not only for single mutations but also for double, triple, and higher-fold (up to sevenfold) mutations. It is shown that on the whole our approach and FOLD-X exhibit almost the same performance despite that the latter uses the adjusting parameters. For multiple mutations, however, our approach is far superior to FOLD-X. Five multiple mutations for staphylococcal nuclease lead to highly enhanced stabilities, but we find that this high enhancement arises from the entropic EV effect. The neglect of this effect in FOLD-X is a principal reason for its ill success. A conclusion is that the three factors mentioned above play essential roles in elucidating the thermal-stability changes upon mutations. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4931814>]

### I. INTRODUCTION

Maintaining the function of a protein at a temperature above the thermal denaturation temperature  $T_m$  of its native state is a principal objective of researchers in a variety of fields related to biophysics and biochemistry.<sup>1</sup> This enhancement of the thermal stability, in general, can also lead to a higher stability against cooling, addition of chemical compounds, and change in pH. One of the methods of enhancing the thermal stability is the mutation. Though there are a number of possible mutations, only a small percentage of them actually lead to the enhancement with the protein function retained. It is strongly desired that the change in the thermal stability resulting from a mutation be predicted using a theoretical approach. The theoretical prediction is to be made on the condition that only the folded structure of the wild type is known (i.e., the folded structure of a mutant is unknown and there is no definite

information on unfolded states of the wild type and the mutant). In the present article, we are concerned with the physicochemical factors governing the thermal-stability changes upon mutations as well as the theoretical approaches toward the prediction based on chemical physics.

Up to now, significantly many approaches<sup>1-8</sup> for the theoretical prediction have been reported. They are concerned primarily with the free-energy difference between the folded and unfolded states  $\Delta G$  at a given temperature (e.g., 25 °C) and the change in  $\Delta G$  upon mutation  $\Delta\Delta G$ . When the performance of an approach is examined,  $\Delta\Delta G_{\text{cal}}$  is compared with  $\Delta\Delta G_{\text{exp}}$  (the subscripts “cal” and “exp” denote the calculated and experimental values, respectively). Though the most straightforward way of evaluating the thermal stability is to look at  $T_m$ ,  $\Delta G$  has preferentially been employed.<sup>1</sup> There are two principal reasons for this: The construction of reliable  $\Delta G$  is an important target in protein research and  $\Delta G$  is certainly related to  $T_m$ . In most of the previously reported approaches,  $\Delta G$  comprises the components relevant to the

<sup>a)</sup>Electronic mail: kinoshit@iae.kyoto-u.ac.jp

protein intramolecular energy, hydration free energy, and protein intramolecular conformational entropy, and each component is further decomposed into multiple terms. Weighting coefficients are multiplied to the constituent terms of  $\Delta G$  and determined in advance so that  $\Delta\Delta G_{\text{cal}}$  can be best fitted to  $\Delta\Delta G_{\text{exp}}$  for a sufficiently large set of proteins and mutations.<sup>2,4,6-8</sup> It has been pointed out that the performance of an approach is remarkably dependent on the set of proteins and mutations employed for the fitting.<sup>8</sup> Though the weighting coefficients determined should be close to unity in principle, some of the values obtained from the fitting are far from unity and even negative,<sup>2-4,6-8</sup> which is physically unreasonable. There is an approach<sup>3</sup> in which the weighting coefficients are not employed. However, it introduces two parameters in the modified Lennard-Jones (LJ) potential function defined between protein atoms. Moreover,  $\Delta\Delta G_{\text{cal}}$  is multiplied by an adjusting parameter  $\eta$ . A total of three parameters are thus introduced, and they are determined so that the root-mean-squared error between  $\Delta\Delta G_{\text{cal}}$  and  $\Delta\Delta G_{\text{exp}}$  can be minimized. The parameter  $\eta$  should be close to unity but the determined value is 0.2.<sup>3</sup> Recently, a different type of approach<sup>9</sup> has been reported. In this approach, electrostatics based on the Debye-Hückel (DH) theory and a simplified treatment of the hydration free energy are introduced into a Go-like model in which only the interactions present in the native structure are taken into account. The dielectric constant of water in the DH theory is set at 29: This value is far smaller than the rightful one 78. The interaction energy per native heavy-atom contact  $\xi$  is determined so that  $T_m$  of the wild type can be matched with the experimental value. (There are two more parameters fitted to the experimental data.)  $T_m$  of a mutant is then predicted and the change in  $T_m$  by the mutation,  $\Delta T_m$ , thus obtained is compared with the experimental value. The meanings of the Go-like model and the parameter  $\xi$  employed, which largely vary from protein to protein,<sup>9</sup> are not physically sound.

We have shown that the driving force of protein folding is a large gain of water entropy<sup>10-12</sup> (our physical picture of the folding is explained in Sec. II A). The water in the bulk makes a larger contribution to the gain than the water near the protein surface. Further, the protein-water-water triplet and higher-order correlations play critical roles in the gain. This entropic excluded-volume (EV) effect cannot be taken into account by a continuum model for water.<sup>10-12</sup> Using our theoretical method emphasizing the effect, which is combined with a molecular model for water, we have succeeded in reproducing the large water-entropy gain upon apoplastocyanin (apoPC) folding<sup>13</sup> experimentally estimated and in explicating the mechanisms of cold<sup>14</sup> and pressure<sup>15</sup> denaturing and of sugar-induced enhancement of the thermal stability.<sup>16</sup> (The importance of the entropic EV effect has been pointed out by other groups in different ways.<sup>17,18</sup>) The most serious drawback shared by all of the previously reported approaches for predicting the thermal-stability changes upon mutations<sup>1-9</sup> is that the entropic EV effect is not taken into account. In the present article, we report the results of the first attempt to examine a theoretical approach for predicting the thermal-stability changes upon mutations, which possesses the following features: It is completely free from the parameters fitted to the experimental data and it

accounts for the entropic EV effect to its full extent. It is not an empirical approach but a physics-based one.

In earlier works, we proposed a measure of the thermal stability of a protein. The measure was defined as the water-entropy gain upon folding at 25 °C divided by the number of residues.<sup>19,20</sup> A larger measure implied higher  $T_m$ . The validity of the measure was corroborated for homologous proteins: the yeast, bacterial, and human orthologues of frataxins<sup>19</sup> and four cytochromes *c* treated as models of mesophilic, moderately thermophilic, thermophilic, and hyperthermophilic proteins,<sup>20</sup> respectively. On the other hand, the structural differences among the wild type and mutants of a protein are much smaller than those among homologous proteins. Nevertheless, these slight differences give rise to denaturation-temperature changes in various magnitudes. We first applied the measure to the present subject but found that the result was not very successful. Therefore, we decided to construct a new measure on the basis of our previously developed free-energy function<sup>21,22</sup> including the enthalpic component as well. The function has been tested for the discrimination of the native fold from misfolded decoys: The success rate of the discrimination is almost 100%,<sup>21,22</sup> demonstrating its superiority over any of the previously reported functions in which the incorporation of the entropic EV effect is inadequate. However, the decoy structures are considerably different from the native structure. Further, it has only been shown that the function takes the lowest value for the native structure. Taken together, the subject tackled in the present study is significantly different from and much more delicate than the subjects treated in our earlier publications.<sup>19-22</sup>

In the present article, the performance of our approach is compared with that of FOLD-X<sup>2,23,24</sup> for 10 proteins and a total of 207 mutations including 18 double and 13 triple and higher-fold (up to sevenfold) mutations. The number of mutations changing the total charge of the protein is 81. FOLD-X is one of the most popular, successful approaches. In the version of FOLD-X employed by us, the number of the weighting coefficients is reduced to 1 but its recommended value is only 0.33.<sup>23,24</sup> It is still empirical in the sense that the terms in its free-energy function are adjusted using the data from protein-engineering experiments and the experimental mutational free-energy changes.<sup>23,24</sup> We find that our approach and FOLD-X exhibit almost the same performance. For multiple mutations, however, our approach is far superior to FOLD-X. This superiority is crucially important, because most of the mutations realizing remarkable enhancement of the stability are multiple mutations. Five multiple mutations for staphylococcal nuclease lead to highly enhanced stabilities. An important finding is that this high enhancement originates from the entropic EV effect. The neglect of this effect in FOLD-X is a principal reason for its ill success. The number of the factors taken into account in the new measure for our approach is still relatively small: They are the water-entropy gain, loss of the protein conformational entropy, break of protein-water hydrogen bonds (HBs), and formation of protein intramolecular HBs upon folding. The success mentioned above indicates that these physicochemical factors govern the thermal-stability changes upon mutations.

In our opinion, a variety of subjects regarding proteins (protein folding,<sup>13</sup> cold denaturation,<sup>14</sup> pressure

denaturation,<sup>15</sup> thermal denaturation,<sup>19,20</sup> effects of cosolvent addition on the structural stability,<sup>16</sup> effects of mutations on the structural stability, etc.) should be elucidated in a unified manner within the same theoretical framework: A theory which can elucidate a particular subject but fails to elucidate the others, for example, is not a good one. Our goal is to develop a theoretical method which is capable of explicating a number of protein-related issues and demonstrate its versatility. A pivotal factor in the method is the entropic effect originating from the translational displacement of water molecules coexisting with a protein. The present manuscript reports part of this development and demonstration.

## II. NEW MEASURE OF THERMAL STABILITY OF A PROTEIN

### A. Picture of protein folding

As illustrated in Fig. 1, a backbone and side chains of a protein generate excluded spaces which the centers of water molecules cannot enter. The volume of an excluded space is “EV.” Upon formation of  $\alpha$ -helix and  $\beta$ -sheet by the backbone (see Figs. 1(a) and 1(b)), the overlap of the EVs occurs and the total EV decreases by the overlapped volume.<sup>25</sup> Likewise, large reduction of the total EV occurs when the side chains are closely packed (see Fig. 1(c)).<sup>25</sup> A decrease in the total EV leads to an increase in the volume of the configurational phase space for water molecules, which is accompanied by a gain of water entropy. It is often claimed that intramolecular HBs and van der Waals attractive interactions within a protein play essential roles in driving a protein to fold. However, the folding undergoes serious dehydration comprising the loss of protein-water electrostatic and van der Waals attractive interactions and structural reorganization of water molecules near the protein surface. The importance of the loss of protein-water electrostatic attractive interactions was first pointed out by Honig and Nicholls.<sup>26</sup> Protein-water HBs form a pivotal component of protein-water electrostatic attractive interactions. Terazima

*et al.*<sup>27</sup> showed that apoPC folding at 25 °C exhibits a significantly large enthalpic increase, proving that the dehydration effect dominates.<sup>13</sup> (The experimental technique of Terazima *et al.* enables us to directly measure the enthalpic change upon protein folding at a given temperature.) Taken together, a large water-entropy gain surpasses the enthalpic increase and loss of the protein conformational entropy.<sup>10–13</sup>

### B. Free-energy function for a protein and free-energy change upon protein folding

The free-energy function  $F$  defined by<sup>21,22</sup>

$$F = (\Lambda - TS_{\text{VH}})/(k_{\text{B}}T_0), \quad T_0 = 298 \text{ K} \quad (1)$$

has recently been developed by us on the basis of the physical picture described in Sec. II A. Here,  $\Lambda$  is the sum of protein intramolecular energy and hydration energy,  $S_{\text{VH}}$  the hydration entropy,  $T$  the absolute temperature, and  $k_{\text{B}}$  Boltzmann’s constant.  $\Lambda$ ,  $S_{\text{VH}}$ , and  $F$  are functions of the protein structure.  $F$  was originally developed for the discrimination of the native fold from many misfolded decoys. Since only very compact structures were treated in the discrimination, the protein conformational entropy was not incorporated in  $F$ . In the present study, we introduce the free-energy change upon protein folding,  $\Delta\Phi = k_{\text{B}}T_0(F_{\text{F}} - F_{\text{U}}) - T\Delta S_{\text{C}}$  (the subscripts “F” and “U” denote the values of the folded (native) and unfolded (denatured) states, respectively, and  $\Delta S_{\text{C}}$  is the change in the conformational entropy) expressed by

$$\Delta\Phi/(k_{\text{B}}T_0) = (\Delta\Lambda - T\Delta S_{\text{VH}} - T\Delta S_{\text{C}})/(k_{\text{B}}T_0), \quad (2)$$

where  $\Delta X$  denotes the change in  $X$  upon folding:  $\Delta\Lambda > 0$ ,  $\Delta S_{\text{VH}} > 0$ , and  $\Delta S_{\text{C}} < 0$ .

The protein insertion can be considered under either isochoric (constant-volume) or isobaric (constant-pressure) condition. The hydration free energy  $\mu$  takes the same value irrespective of the insertion condition. However, this is not the case for the hydration energy  $E_{\text{VH}}$  and  $S_{\text{VH}}$ .<sup>28</sup> Fortunately,

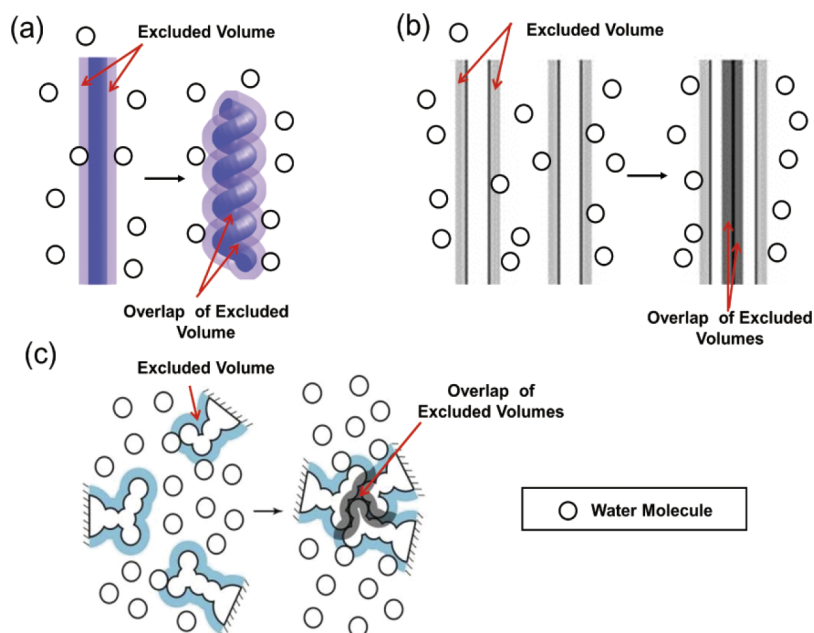


FIG. 1. (a) Formation of  $\alpha$ -helix by a portion of the backbone. (b) Formation of  $\beta$ -sheet by portions of the backbone. (c) Close packing of side chains. The total excluded volume decreases by the overlapped volume, which leads to a corresponding increase in the total volume available to the translational displacement of water molecules coexisting with the protein.



protein folding occurs with the system pressure and volume almost unchanged (the EV of a more compact structure is smaller but the partial molar volume is almost independent of the compactness).<sup>10,12</sup> It follows that the hydration energy and entropy under isochoric condition, respectively, are almost equal to the hydration enthalpy and entropy under isobaric condition. We adopt isochoric condition. Isochoric condition is free from compression or expansion of the bulk water by which the physical interpretation of a change in a thermodynamic quantity of hydration is made rather difficult.<sup>28</sup>

### C. Water-entropy gain upon protein folding

$\Delta S_{\text{VH}}$  is a function of the number density of water  $\rho$  and  $T$ . If  $\rho$  is kept constant,  $\Delta S_{\text{VH}}$  becomes larger as  $T$  increases.<sup>29</sup> If  $T$  is kept constant,  $\Delta S_{\text{VH}}$  becomes smaller as  $\rho$  decreases.<sup>29</sup> It is experimentally known that above 298 K  $\rho$  decreases progressively with increasing  $T$ : In this temperature region, the effect of  $\rho$  dominates and  $\Delta S_{\text{VH}}$  is a decreasing function of  $T$ .<sup>29</sup> We then consider the wild type and a mutant of a protein whose number of residues is  $N_r$ .  $S_{\text{VH,U}}/(k_B N_r)$  is not significantly dependent on the mutation if the total number of S–S bonds remains unchanged upon mutation. However, this is not true for  $S_{\text{VH,F}}/(k_B N_r)$  that is quite sensitive to the packing efficiency of the backbone and side chains in the folded state.<sup>19,20</sup> As a result,  $\Delta S_{\text{VH}}/(k_B N_r)$  ( $\Delta S_{\text{VH}} = S_{\text{VH,F}} - S_{\text{VH,U}}$ ;  $\Delta S_{\text{VH}}$  is the water-entropy gain upon folding) varies significantly from mutant to mutant. ( $\Delta S_{\text{VH}}/(k_B N_r)$  of the wild type is also significantly different from that of a mutant.) See Fig. 2 illustrating our physical picture of the thermal stability of the wild type and a mutant.

### D. Loss of protein conformational entropy upon protein folding

The protein conformational entropy is denoted by  $S_C$ . Since the folded state is under structural constraint on

account of its closely packed properties,  $S_{\text{C,F}}$  remains roughly constant against an increase in  $T$ .  $S_{\text{C,U}}$  is much more influenced by  $T$ .  $S_{\text{C,U}}$  is related to the ranges of dihedral angles allowed, which depend on the torsion energy and  $T$ .<sup>19,20</sup> At low  $T$ , angles giving only low torsion energy are accessible. As  $T$  increases, the allowed range of each angle is increasingly widened and  $S_{\text{C,U}}$  becomes larger. As  $T$  increases further, the enlargement of  $S_{\text{C,U}}$  is diminished due to the steric repulsion among atoms in a residue and in neighboring residues. Therefore,  $dS_{\text{C,U}}/dT > 0$  and  $d^2S_{\text{C,U}}/dT^2 < 0$  with the result of  $d|\Delta S_C|/dT > 0$  and  $d^2|\Delta S_C|/dT^2 < 0$  ( $\Delta S_C = S_{\text{C,U}} - S_{\text{C,F}}$ ).<sup>20</sup> This temperature dependence of  $\Delta S_C$  has been verified by Fitter's experimental study.<sup>30</sup> If the total number of S–S bonds remains unchanged upon mutation,  $S_{\text{C,U}}/(k_B N_r)$  is almost constant while  $S_{\text{C,F}}/(k_B N_r)$  is essentially zero, and  $|\Delta S_C|/(k_B N_r)$  can be considered independent of the mutation.

### E. Enthalpy change upon protein folding

Protein folding is accompanied by a decrease in the protein intramolecular energy (factor 1) and an increase in the hydration energy (factor 2).<sup>13</sup> Factor 2 originates from a loss of protein-water attractive (electrostatic and van der Waals) interactions and a gain of water-water attractive interactions ascribed to the structural reorganization of water near the protein surface. Factor 1 remains unchanged against an increase in  $T$ . Experimental results<sup>27</sup> have shown that factor 2 dominates at 298 K and protein folding gives rise to an enthalpy increase and that factor 2 becomes weaker as  $T$  increases. The gain of water-water attractive interactions, which is largely influenced by  $T$ , is also significant in factor 2.  $\Delta\Lambda$  is a strongly decreasing function of  $T$ : It takes large, positive and negative values at 298 K and in the vicinity of  $T_m$ , respectively.

### F. New measure of thermal stability of a protein

We define  $\Sigma(T)$  by

$$\Sigma(T) = -\Delta F/N_r = (T\Delta S_{\text{VH}} - \Delta\Lambda)/(k_B T_0 N_r), \quad (3)$$

where  $\Delta X$  denotes the change in  $X$  upon protein folding. As illustrated in Fig. 2 considering the wild type and a mutant of a protein, the thermal stability can be argued by the competition of  $\Sigma(T)$  and  $T|\Delta S_C|/(k_B T_0 N_r)$ . The argument is based on the free-energy change upon folding defined by Eq. (2).  $\Delta\Lambda$  is a strongly decreasing function of  $T$  whereas  $T\Delta S_{\text{VH}}$  remains roughly constant against an increase in  $T$ . Hence,  $\Sigma(T)$  is an increasing function of  $T$ . Below  $T_m$ , the folded state is more stable than the unfolded state because  $\Sigma(T)$  is larger than  $T|\Delta S_C|/(k_B T_0 N_r)$ . The inversion occurs above  $T_m$ , causing thermal unfolding.  $\Sigma_0 = \Sigma(T_0)$  ( $T_0 = 298$  K) can be a measure of the thermal stability. The larger  $\Sigma_0$  is, the higher  $T_m$  is. The slopes of the two solid lines for the wild type and the mutant can be different, but the difference is assumed to be not large enough to invert the order of the thermal stability,  $T_m^{\text{W}} > T_m^{\text{M}}$  (the superscripts, “M” and “W,” denote values for the mutant and the wild type, respectively) in Fig. 2. We examine the correlation between  $\Delta\Sigma_0$  and  $\Delta T_m$

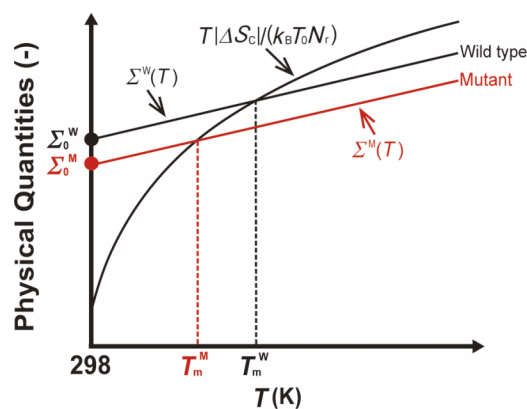


FIG. 2. Illustration of our physical picture for thermal stabilities of the wild type and a mutant of a protein.  $\Sigma(T)$  is defined by Eq. (3),  $\Delta S_C$  is the conformational-entropy loss upon protein folding,  $N_r$  the number of residues,  $k_B$  Boltzmann's constant,  $T_m$  the thermal denaturation temperature,  $\Sigma_0 = \Sigma(T_0)$  ( $T_0 = 298$  K), and the superscripts, “W” and “M,” denote values or quantities of the wild type and the mutant, respectively.  $\Sigma^{\text{W}}(T)$  and  $\Sigma^{\text{M}}(T)$  do not necessarily change linearly with  $T$ . The two solid lines of  $\Sigma^{\text{W}}(T)$  and  $\Sigma^{\text{M}}(T)$  are not necessarily parallel.

defined as

$$\Delta\Sigma_0 = \Sigma_0^M - \Sigma_0^W, \quad (4a)$$

$$\Delta T_m = T_m^M - T_m^W. \quad (4b)$$

### III. MODEL AND THEORETICAL METHOD

#### A. Models of water and proteins

A multipolar model<sup>31,32</sup> is employed for water. A point dipole and a point quadrupole of tetrahedral symmetry are embedded at the center of a hard sphere with diameter  $d_S = 2.8$  Å. We adopt the self-consistent mean field (SCMF) theory developed by Kusalik and Patey<sup>31,32</sup> to account for the effect of the molecular polarizability.  $S_{VH}$  is determined primarily by the geometric characteristics of a protein structure and rather insensitive to the protein-water interaction potentials. The insensitivity has been substantiated in our earlier studies.<sup>20,33</sup> For example, even when the protein-water electrostatic potentials are completely shut off,  $|S_{VH}|$  of a protein decreases only by  $\sim 5\%$ . Therefore, we model a protein as a set of fused hard spheres just for calculating  $S_{VH}$ .<sup>10-16</sup>

#### B. Angle-dependent integral equation theory (ADIET) for molecular liquids

The ADIET<sup>29,31,32</sup> is employed for calculating  $S_{VH}$  of a spherical solute (see step (1) described in Sec. III C). The water-water and solute-water potentials and correlation functions are dependent on the Euler angles representing the orientations of water molecules. The details of basic equations and numerical solution procedure were described in our earlier publications.<sup>29,34</sup>  $S_{VH}$  is evaluated via the temperature derivative of  $\mu$  calculated using the hypernetted-chain closure and the Morita-Hiroike formula adapted to molecular liquids.<sup>29</sup> By the ADIET combined with the multipolar water model, the dielectric constant of water is calculated to be 84 (this value is in close proximity with the experimental one 78).<sup>29</sup> Further,  $\mu$  of a nonpolar solute calculated is in perfect agreement with that obtained from a Monte Carlo simulation.<sup>29</sup>

#### C. Calculation of hydration entropy of a protein with a prescribed structure

When a structure of a protein is given, its  $S_{VH}$  is calculated by a hybrid of the ADIET and the morphometric approach (MA).<sup>35</sup> In the MA, the geometric characteristics of a solute molecule are represented by only the four measures,  $V_{ex}$ ,  $A$ ,  $X$ , and  $Y$ .<sup>35</sup>  $V_{ex}$  is the EV,  $A$  is the water-accessible surface area (ASA), and  $X$  and  $Y$  are the integrated mean and Gaussian curvatures of the accessible surface, respectively.  $S_{VH}$  is expressed by the linear combination of the four geometric measures,<sup>35</sup>

$$S_{VH}/k_B = C_1V_{ex} + C_2A + C_3X + C_4Y. \quad (5)$$

The four coefficients ( $C_1$ – $C_4$ ) are dependent only on the thermodynamic state of the solvent and independent of the solute shape. Hence, they can be evaluated in the simplest geometries: for hard-sphere solutes (isolated atoms) with various

diameters. We refer to the effect expressed by Eq. (5) as the EV effect. At normal temperature and pressure,  $C_1V_{ex}$  usually makes the largest contribution to  $S_{VH}/k_B$ .

The calculation consists of the four steps summarized below.<sup>13-16</sup>

- (1) Calculate  $S_{VH}$  of a hard-sphere solute with diameter  $d_U$  immersed in the multipolar-model water using the ADIET. Consider different values of  $d_U$  in the range,  $0.6 \leq d_U/d_S \leq 10$ , to obtain a sufficiently large set of data for  $S_{VH}$  and  $R$  ( $R = (d_U + d_S)/2$ ).
- (2) Determine  $C_1$ – $C_4$  by applying the least-squares method to the following equation:

$$S_{VH}/k_B = C_1(4\pi R^3/3) + C_2(4\pi R^2) + C_3(4\pi R) + C_4(4\pi). \quad (6)$$

Equation (6) is the linear combination of the four geometric measures for spherical solutes. The determined values are as follows:  $C_1 = -0.1968$  Å<sup>-3</sup>,  $C_2 = 0.0452$  Å<sup>-2</sup>,  $C_3 = 0.2567$  Å<sup>-1</sup>, and  $C_4 = -0.3569$ .

- (3) Calculate  $V_{ex}$ ,  $A$ ,  $X$ , and  $Y$  of a protein with a prescribed structure using an extended version<sup>35</sup> of Connolly's algorithm.<sup>36,37</sup> The  $x$ - $y$ - $z$  coordinates and diameters of the protein atoms are the input data. The diameter of each atom is set at the sigma value of the LJ potential parameters taken from CHARMM22.<sup>38</sup>
- (4) Obtain  $S_{VH}$  from Eq. (5) to which  $C_1$ – $C_4$  determined in step (2) are substituted.

The maximum value of  $d_U$  considered in step (1) must be sufficiently large so that the effects of the four geometric measures can fully be taken into account. We have verified that  $d_U = 10d_S$  is large enough: Altering  $10d_S$  to  $30d_S$  results in essentially no changes in  $C_1$ – $C_4$  determined. We emphasize that protein molecules are not assumed to be ideally spherical. Step (4) is applicable to a variety of protein structures including random coils or fully extended structures. More detailed descriptions are provided in our earlier publications,<sup>10-12,35,39,40</sup> two of which demonstrated the high accuracy of the MA.<sup>35,40</sup>

#### D. Calculation of energetic component for a protein with a prescribed structure

We briefly summarize the procedure for calculating the energetic component  $\Lambda$  (see the thermodynamic cycle illustrated in Fig. 3).<sup>21,22</sup> A fully extended structure is chosen as the reference one because it possesses the maximum number of HBs with water molecules and no intramolecular HBs:  $\Lambda = 0$ .

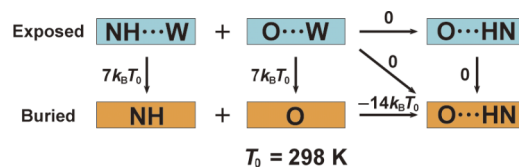


FIG. 3. Thermodynamic cycle for calculating the energetic component  $\Lambda$ . “W” and “...” represent a water molecule and a hydrogen bond, respectively, and  $k_B$  is Boltzmann's constant. This figure is illustrated for the case where N is the donor and O is the acceptor though there are four different donor-acceptor combinations: (N, O), (O, N), (O, O), and (N, N).

Let us consider a transition from the reference structure to a more compact one. A gain of protein intramolecular interactions, loss of protein-water interactions, and energy change caused by structural reorganization of water molecules near the protein surface are assumed to be cancelled out except for the following: When the break of HBs with water molecules is not compensated by the formation of intramolecular HBs, a serious energetic increase is caused and to be taken into account. The torsion energy, which is sufficiently low for any structure considered, can be neglected.

The structural transition defined above is followed by two cases: (i) a donor and an acceptor are buried in the protein interior after the break of HBs with water molecules, but they form an intramolecular HB and (ii) a donor or an acceptor is buried but it finds no partner for a HB. No penalty is imposed in case (i) but a penalty of  $7k_B T_0$  is imposed in case (ii).<sup>21,22</sup>  $7k_B T_0$  is based on the estimation that the free-energy lowering brought by hydrogen-bond formation between two formamide molecules in a nonpolar solvent is  $-14k_B T_0$ .<sup>41</sup> The nonpolar solvent mimics the environment of protein interior, and  $-14k_B T_0$  includes the entropic gain of the nonpolar solvent upon the hydrogen-bond formation. The ASA of each donor or acceptor is calculated using Connolly's algorithm.<sup>36,37</sup> The donor or acceptor is considered buried if its ASA is smaller than  $0.001 \text{ \AA}^2$ . On the basis of the criteria proposed by McDonald and Thornton,<sup>42</sup> we determine whether an intramolecular HB is formed or not. All of the donors and acceptors are examined for backbone-backbone, backbone-side chain, and side chain-side chain intramolecular HBs to calculate  $\Lambda$ .<sup>21,22</sup>

### E. Proteins and mutations considered

We choose the proteins treated in experiments under the condition which satisfies the following: The values of  $T_m$  and/or  $\Delta\Delta G_{\text{exp}}$  were measured in aqueous solution whose pH is in the range from 6 to 8; no denaturants were utilized in the determination of  $T_m$  and/or  $\Delta\Delta G_{\text{exp}}$ , and the folded structures of the wild types were determined using the X-ray crystallography. The structure models obtained from NMR are excluded for the following reasons. The NMR models are constructed by a structure calculation upon which the structural information experimentally obtained as a set of constraints is imposed.<sup>43</sup> Typical constraints are the nuclear Overhauser effect (NOE), residual dipolar coupling (RDC), hydrogen bonding, and dihedral angle restraints. Unless the amount of constraints is sufficiently large, the models constructed are substantially influenced by the structure calculation employed. About 20–40 candidate models are usually prepared, but they are often significantly different from one another.<sup>43</sup> The prediction of the thermal-stability changes upon mutations is made under the condition that the wild-type structure is given whereas a mutant structure is unknown. Since this is a subtle task, it is required that the wild-type structure be sufficiently certain. We exclude the NMR models to concentrate on the investigation of our thermal-stability measure. It is true, however, that some NMR models give successful results while others do not, but this is beyond the scope of our work. (Of course, when the amount of constraints is large enough, the models from NMR can be better than those from the X-ray

crystallography because the former accounts for the structure fluctuation in aqueous solution.<sup>43</sup>)

We test 10 proteins<sup>44–54</sup> listed in Table I. Four of them possess S–S bonds. The crystallization for the X-ray crystallography was made in aqueous solution whose pH was in the range from 6 to 8 with the exception of ribosomal protein L30e,<sup>55</sup> ribonuclease (RNase) HI,<sup>56</sup> and ribose binding protein.<sup>57</sup> Their pH-values were 5.6, 9.0, and 5.0, respectively. We could not find the pH-values in the crystallization for chicken lysozyme<sup>58</sup> and RNase A.<sup>59</sup> In order to confirm that the proteins considered are sufficiently diverse, we analyze the amino-acid sequence homology for all of the protein pairs using “CLUSTAL W.”<sup>60</sup> For any two of RNase A, RNase HI, RNase Sa, and RNase T1, the homology is in the range 3%–9%. For chicken lysozyme and T4 lysozyme, it is 11%. Among the 45 pairs, only 7 of them exhibit the homology exceeding 10%. Thus, the amino-acid sequences of the 10 proteins are quite different. The values of  $T_m$  and  $\Delta G_{\text{exp}}$  for the wild types of the 10 proteins are collected in Table II.  $\Delta G_{\text{exp}}$ , the free-energy change upon *unfolding* at 25 °C, was calculated using the experimental data (see Secs. IV B and IV C for  $\Delta\Delta G_{\text{exp}}$ ).

The mutants considered for each protein are also listed in Table I. A total of 207 mutations (176 are single, 18 are double, and 13 are triple and higher-fold) are chosen. It is observed in the table that we consider a variety of mutations including those changing the total charge (e.g., mutating from a nonpolar residue to a charged one): The number of such mutations is 81. The 5 multiple mutations in the case of staphylococcal nuclease<sup>51</sup> include the mutation from proline (Pro) to glycine (Gly). Chicken lysozyme,<sup>54</sup> RNase A,<sup>52</sup> RNase Sa,<sup>46</sup> and RNase T1<sup>53</sup> possess 4, 4, 1, and 2 S–S bonds, respectively, but the number of S–S bonds remains unchanged upon any mutation considered. By the mutation Pro  $\rightarrow$  Gly, the unfolded state becomes more extended. The presence of an S–S bond makes the unfolded state less extended. There is a trend that proteins with S–S bonds and mutations including Pro or Gly are avoided due to the resulting lower prediction performance,<sup>3,9</sup> but we challenge such proteins and mutations. (Our result is discussed in Sec. IV A.)

### F. Preparation of folded state for wild type and mutant

Models of the folded structures of the wild types are taken from Protein Data Bank (PDB). After giving hydrogen atoms to the models using the CHARMM<sup>61</sup> and MMTSB<sup>62</sup> programs, we slightly modify the models using the energy minimization described in our earlier publication<sup>20</sup> to remove unrealistic overlaps of protein atoms. The modification is carried out using the CHARMM22 parameters<sup>38</sup> with the CMAP correction<sup>63</sup> and the GBMV implicit solvent model.<sup>64,65</sup> We put positional restraints in the harmonic form on all of the heavy atoms during the energy minimization. The force constant for the restraints is  $2M \text{ kcal}/(\text{mol } \text{ \AA}^2)$  where  $M$  is the mass of each atom.

It is required that the thermal-stability change upon mutation be predicted without any experimentally determined structural data for the resultant mutant. Starting from the folded structure of the wild type taken from the PDB code as the template, we construct the folded-structure model of a mutant

TABLE I. Proteins and mutations considered. The multiple mutations are italicized.

Protein name (number of S–S bonds)	PDB code	Mutations (total number)
Barnase <sup>44</sup>	1BNI	L14A, I88A, I96A, I88V, I96V (5)
Chicken lysozyme <sup>54</sup> (4)	4LYZ	T40S, T40I, I55L, I55V, I55M, I55F, I55A, I55T, S91T, S91V, S91A, S91D, S91Y, ( <i>T40S/I55V</i> ), ( <i>T40S/S91T</i> ), ( <i>T40S/S91V</i> ), ( <i>T40S/S91A</i> ), ( <i>I55L/S91T</i> ), ( <i>I55V/S91T</i> ), ( <i>I55V/S91A</i> ), ( <i>I55V/S91V</i> ), ( <i>I55A/S91T</i> ), ( <i>T40S/I55V/S91T</i> ), ( <i>T40S/I55V/S91A</i> ) (24)
Ribosomal protein L30e <sup>49</sup>	1H7M	D2A, E6A, R8A, K9A, D12A, K15A, R21A, K22A, K28A, K33A, R39A, R42A, D44A, K46A, E47A, D48A, E50A, R54A, E62A, E64A, E69A, R76A, H78A, D87A, E90A, R92A (26)
Ribonuclease A <sup>52</sup> (4)	1RTB	F46L, F46V, F46A (3)
Ribonuclease HI <sup>47</sup>	2RN2	H62A, H83A, H114A, H124A, ( <i>H62A/H83A</i> ), ( <i>H124A/H127A</i> ), ( <i>H62A/H83A/H124A/H127A</i> ) (7)
Ribonuclease Sa <sup>46</sup> (1)	1RGG	D79F, D79Y, D79A, D79I, D79R, D79L, D79K, D79W, D79H, D79N, D79E, Q94K, D33A (13)
Ribonuclease T1 <sup>53</sup> (2)	1RN1	W59Y, Y24W, Y42W, Y45W, H40 T, H92A, ( <i>Y24W/W59Y</i> ), ( <i>Y42W/W59Y</i> ), ( <i>Y45W/W59Y</i> ), ( <i>H40T/W59Y</i> ), ( <i>W59Y/H92A</i> ) (11)
Ribose-binding protein <sup>48</sup>	1URP	S9A, N13A, F15A, F16A, N64A, D89A, S103A, I132A, F164A, N190A, F214A, D215A, Q235A, ( <i>S9A/I132A</i> ), ( <i>S9A/I132A/S103A</i> ), ( <i>S9A/I132A/S103A/N13A</i> ), ( <i>S9A/I132A/S103A/N13A/Q235A</i> ) (17)
Staphylococcal nuclease <sup>50,51</sup>	1EY0	V23T, V39T, V51T, V66T, V74T, V99T, V104T, V111T, V114T, V39S, V51S, V66S, V114S, Y27F, Y54F, Y85F, Y91F, Y93F, Y113F, Y115F, Y54L, Y85L, Y91L, Y113L, Y115L, S128A, T13S, T22S, T33S, T41S, T44S, T62S, T82S, T120S, T13V, T22V, T33V, T41V, T44V, T62V, T82V, T120V, T13C, T22C, T33C, T41C, T44C, T62C, T82C, T120C, T22I, T33I, T41I, T44I, T62I, T82I, T120I, D19F, K28F, K48F, K49F, E52F, E57F, M65F, E67F, E73F, Q80F, K84F, E101F, A112F, K116F, E122F, Q123F, K127F, S128F, E135F, K136F, T62A, T62F, T62G, T62H, T62K, T62L, T62M, T62N, T62Q, ( <i>P117G/H124L/S128A</i> ), ( <i>T41I/P117G/H124L/S128A</i> ), ( <i>T33V/T41I/P117G/H124L/S128A</i> ), ( <i>T41I/S59A/P117G/H124L/S128A</i> ), ( <i>T33V/T41I/S59A/P117G/H124L/S128A</i> ) (91)
T4 lysozyme <sup>45</sup>	1L63	N40A, K43A, S44A, E45A, L46A, D47A, K48A, ( <i>N40A/K43A/S44A/E45A/L46A/D47A/K48A</i> ), ( <i>N40A/S44A/E45A/D47A/K48A</i> ), ( <i>E45A/K48A</i> ) (10)

using Modeller<sup>66</sup> (Ver. 9.11). We generate 10 candidate models on the condition that the modification of the coordinates is limited to the protein atoms within the distance of 2 Å from the center of the mutated residue. The model with the lowest value of our free-energy function  $F$  at  $T = T_0$  (see Eq. (1)) is chosen as the best one. Since the 10 models are all compact, the protein conformational entropy need not be taken into consideration. When unrealistic overlaps of protein atoms are observed, they are removed in the manner mentioned above

TABLE II. Values of thermal denaturation temperature,  $T_m$ , and free-energy change upon *unfolding* at 25 °C,  $\Delta G_{\text{exp}}$ , for wild types of the 10 proteins listed in Table I.  $\Delta G_{\text{exp}}$  was calculated using the experimental data, and  $\Delta G_{\text{exp}}$  for staphylococcal nuclease was calculated at 20 °C.  $\Delta G_{\text{exp}}$  is not presented in the literature for chicken lysozyme, RNase (ribonuclease) HI, RNase Sa, ribose-binding protein, and T4 lysozyme.

Protein name	$T_m$ (°C)	$\Delta G_{\text{exp}}$ (kcal/mol)
Barnase <sup>44</sup>	53.9	9.5
Chicken lysozyme <sup>54</sup>	74.0	...
Ribosomal protein L30e <sup>49</sup>	93.8	11.4
Ribonuclease A <sup>52</sup>	59.7	9.30
Ribonuclease HI <sup>47</sup>	50.2	...
Ribonuclease Sa <sup>46</sup>	47.8	...
Ribonuclease T1 <sup>53</sup>	57.2	7.82
Ribose-binding protein <sup>48</sup>	57.5	...
Staphylococcal nuclease <sup>50,51</sup>	52.7	5.4
T4 lysozyme <sup>45</sup>	62.2	...

with the alteration that the minimization is terminated once the LJ potential energy becomes negative. This alteration is for preserving the original structure as much as possible.

### G. Preparation of unfolded state for wild type and mutant

Structural properties of the unfolded state of a protein are not exactly known. A prevailing method for modeling the unfolded state is to generate a sufficiently large set of random coils. In our earlier works,<sup>19,20</sup> it was shown that the previous measure of the thermal stability<sup>19,20</sup> retains its efficacy even when a small number of extended structures are employed as the model of the unfolded state.<sup>20</sup> Therefore, we adopt five extended structures. The new measure differs from the previous one in the respect that  $\Lambda$  is taken into account. Since  $\Lambda$  is calculated by regarding a fully extended structure as the reference one, the employment of a small number of extended structures is justifiable. All of the S–S bonds are broken in the preparation of the unfolded state. The effect of this treatment is discussed in Sec. IV A. The main-chain dihedral angles ( $\phi, \psi$ ) of the 5 extended structures are in the range from ( $-130^\circ, 130^\circ$ ) to ( $-170^\circ, 170^\circ$ ) with a step of ( $-10^\circ, 10^\circ$ ) ( $\omega$  is set at  $180^\circ$ ).<sup>20</sup> For Pro, however, ( $\phi, \psi$ ) is set at ( $-60^\circ, 40^\circ$ ). The most probable conformer for each side chain is taken from Dunbrack's Backbone-Dependent Rotamer Library.<sup>67,68</sup> The extended structures are modified using the energy minimization explained above with the alteration that we put the



position restraints only on  $C_\alpha$  atoms to allow for more structural flexibility.

## H. Performance measures defined for theoretical prediction method

In addition to the correlation coefficient between two quantities of interest (e.g.,  $\Delta\Sigma_0$  and  $\Delta T_m$  in our theoretical approach), a variety of measures are examined in the performance test. Here, mutations which lead to higher and lower thermal stabilities are referred to as “good mutation” and “bad mutation,” respectively. There are two cases: case (a) where a mutation is actually (i.e., experimentally) a good one and case (b) where a mutation is actually a bad one. In case (a), when it is predicted to be a good one, it is counted as a true positive (TP), and when it is predicted to be a bad one, it is counted as a false negative (FN).<sup>69</sup> In case (b), when it is predicted to be a bad one, it is counted as a true negative (TN), and when it is predicted to be a good one, it is counted as a false positive (FP).<sup>69</sup>

The performance measures for a prediction method are then defined as follows:<sup>69</sup>

$$\text{Accuracy} = M_A = (\text{TP} + \text{TN})/(\text{TP} + \text{FN} + \text{TN} + \text{FP}), \quad (7a)$$

$$\text{Recall} = M_R = \text{TP}/(\text{TP} + \text{FN}), \quad (7b)$$

$$\text{Precision} = M_P = \text{TP}/(\text{TP} + \text{FP}), \quad (7c)$$

$$\text{Specificity} = M_S = \text{TN}/(\text{TN} + \text{FP}), \quad (7d)$$

$$\text{Negative predictive value} = M_N = \text{TN}/(\text{TN} + \text{FN}). \quad (7e)$$

“Recall” is defined for the actual good mutations. It represents the proportion of those which are successfully predicted to be good mutations. “Precision” is defined for the predicted good mutations. It represents the proportion of those which are actually good mutations. “Specificity,” which is defined for the actual bad mutations, represents the proportion of those which are successfully predicted to be bad mutations. “Negative predictive value” is defined for the predicted bad mutations and it represents the proportion of those which are actually bad mutations. “Accuracy” is defined for all of the mutations and the proportion of the successful predictions. The measures are concerned with whether the thermal stability becomes higher or lower and irrelevant to the degree of the stability change. We remark that  $M_A$  is most frequently considered.<sup>8</sup>

## IV. RESULTS AND DISCUSSION

### A. Relation between $\Delta\Sigma_0$ and $\Delta T_m$

$\Delta T_m$  is plotted against  $\Delta\Sigma_0$  in Fig. 4(a). In a strict sense,  $\Delta\Sigma_0$  is not linearly correlated to  $\Delta T_m$ . However, negative  $\Delta\Sigma_0$  implies negative  $\Delta T_m$  and positive  $\Delta\Sigma_0$  implies positive  $\Delta T_m$ , and larger  $|\Delta\Sigma_0|$  leads to larger  $|\Delta T_m|$ . Hence, the performance can be evaluated through the plot of  $\Delta T_m$  against  $\Delta\Sigma_0$  and the measures, in particular,  $M_A$ . The correlation coefficient in Fig. 4(a) is 0.409 and  $M_A$  is 73.4%. Even the 5 multiple mutations including the mutation Pro  $\rightarrow$  Gly in the case of staphylococcal nuclease do not appreciably deviate from the plot. Further, the data points for the proteins possessing S–S bonds do not exhibit apparent deviation, either. When a protein

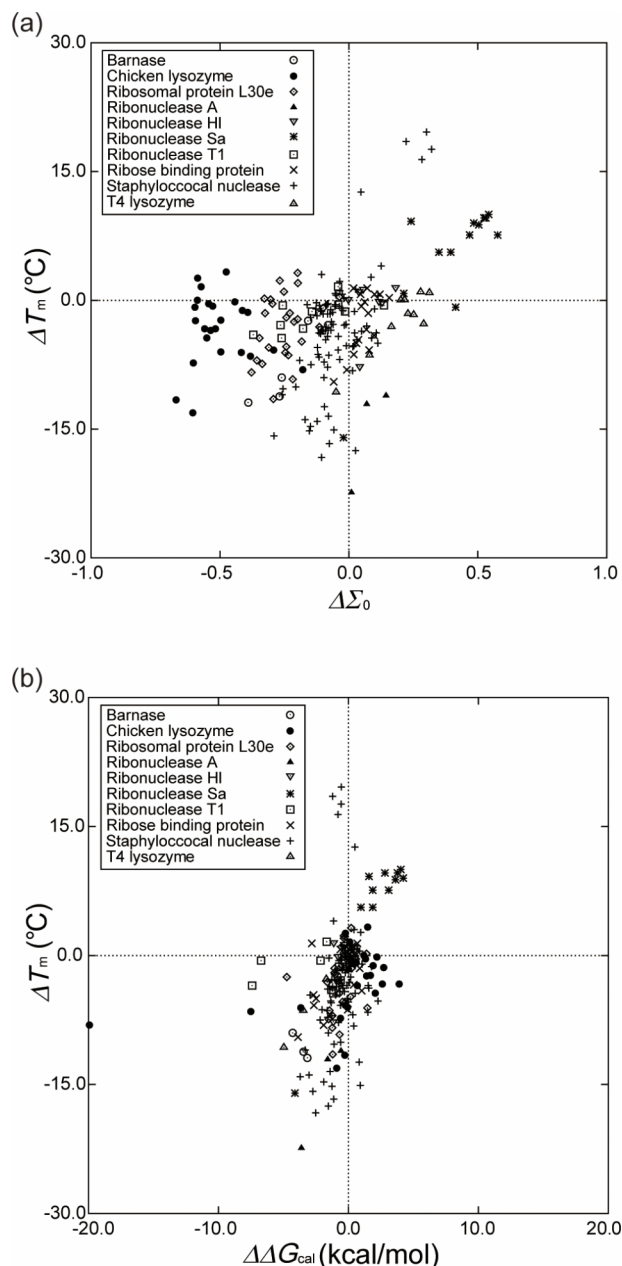


FIG. 4. (a) Relation between  $\Delta T_m$  and  $\Delta\Sigma_0$  calculated by our approach for the 207 mutants listed in Table I.  $T_m$  is the thermal denaturation temperature,  $\Sigma(T)$  is defined by Eq. (3), and  $\Sigma_0 = \Sigma(T_0)$  ( $T_0 = 298$  K). The unit of  $T_m$  is  $^\circ\text{C}$  here.  $\Delta$  denotes “the value for a mutant minus that for the wild type”. (b) Relation between  $\Delta T_m$  and  $\Delta\Delta G_{\text{cal}}$ .  $\Delta\Delta G$  is the change in  $\Delta G$  upon mutation ( $\Delta\Delta G = \Delta G^M - \Delta G^W$ ; the superscripts, “W” and “M,” denote values of the wild type and the mutant, respectively),  $\Delta G$  ( $\Delta G > 0$ ) is the free-energy change upon *unfolding*, and the subscript “cal” denotes the value calculated by FOLD-X.

possesses more S–S bonds, the unfolded state becomes less extended with the result that  $|\Delta S_C|$  decreases for all  $T$ . This effect (effect I) makes the folded state more stable. By contrast, the EV of the unfolded state and the decrease in the EV upon folding become smaller, which is followed by decreased  $\Delta S_{\text{VH}}$  for all  $T$ . This effect (effect II) makes the folded state less stable. It is not definite which of the two effects dominates. However, the number of S–S bonds remains unchanged upon any mutation considered, which should be responsible for the exhibition of no apparent deviation.

## B. Comparison with FOLD-X in terms of performance by looking at $\Delta T_m$

FOLD-X, which is one of the most popular, successful approaches, can freely be used.<sup>2,23,24</sup> We use Version 3 that is essentially the same as the newest one, Version 4 (Version 4 is just easier to use than Version 3). We then compare our approach and FOLD-X in terms of the performance. It should be noted that all of the previously reported approaches including FOLD-X share roughly the same prediction performance.<sup>5,9</sup> In FOLD-X, the force field employed is different from ours and the folded states of the wild type and a mutant are also differently prepared. We use FOLD-X in its original way without any modification. We plot  $\Delta T_m$  against  $\Delta\Delta G_{\text{cal}}$  calculated by FOLD-X in Fig. 4(b). Here,  $\Delta\Delta G$  is the change in  $\Delta G$  upon mutation ( $\Delta\Delta G = \Delta G^M - \Delta G^W$ ),  $\Delta G$  ( $\Delta G > 0$ ) is the free-energy change upon *unfolding* at 25 °C, and the subscript “cal” denotes the calculated value. Positive  $\Delta\Delta G$  implies that the stability is enhanced by the mutation. Since  $\Delta\Delta G$  in FOLD-X is defined as “ $\Delta G^W - \Delta G^M$ ,” the calculated value is multiplied by  $-1$ . The correlation coefficient in Fig. 4(b) is 0.395. In the previous tests, the correlation coefficient for FOLD-X was about 0.5 or better,<sup>5</sup> so the protein data set considered in the present study should be more difficult to theoretically treat. In Table III, we compare FOLD-X with our approach in terms of the five performance measures defined by Eq. (7). All of the measures are considerably higher in ours than in FOLD-X.  $M_A$ , for instance, is 73.4% in ours and 66.7% in FOLD-X.

## C. Comparison with FOLD-X in terms of performance by looking at $\Delta\Delta G_{\text{exp}}$

It is worthwhile to examine the relation between  $\Delta T_m$  and  $\Delta\Delta G_{\text{exp}}$ . The subscript “exp” denotes the experimentally determined value. There is no data of  $\Delta\Delta G_{\text{exp}}$  for chicken lysozyme, RNase HI, ribose binding protein, and single mutations of staphylococcal nuclease. There are 73 mutants with the data of  $\Delta\Delta G_{\text{exp}}$  available. Care must be taken in handling the data of  $\Delta\Delta G_{\text{exp}}$  because  $\Delta G$  was defined for folding in some proteins and for unfolding in others. Since the values of  $\Delta\Delta G_{\text{exp}}$  were calculated at 52.6 °C for RNase Sa<sup>46</sup> and at 62.2 °C for T4 lysozyme<sup>45</sup> in the references, we recalculated them at 25 °C using the thermodynamic data given.<sup>52</sup> For the multiple mutations of staphylococcal nuclease,<sup>51</sup> the values of  $\Delta\Delta G_{\text{exp}}$  were calculated at 20 °C in the reference. However, the thermodynamic data were not given. We decided to adopt the values at 20 °C because they should be very close to those at 25 °C. The relation between  $\Delta T_m$  and  $\Delta\Delta G_{\text{exp}}$  is plotted in Fig. 5. As expected, they are highly correlated: The correlation coefficient reaches 0.881.

TABLE III. Comparison between our approach and FOLD-X in terms of the performance measures defined by Eq. (7). The data points in Fig. 4 are considered.

Approach	$M_A$	$M_R$	$M_P$	$M_S$	$M_N$
Ours	0.734	0.623	0.485	0.773	0.856
FOLD-X	0.667	0.604	0.400	0.688	0.835

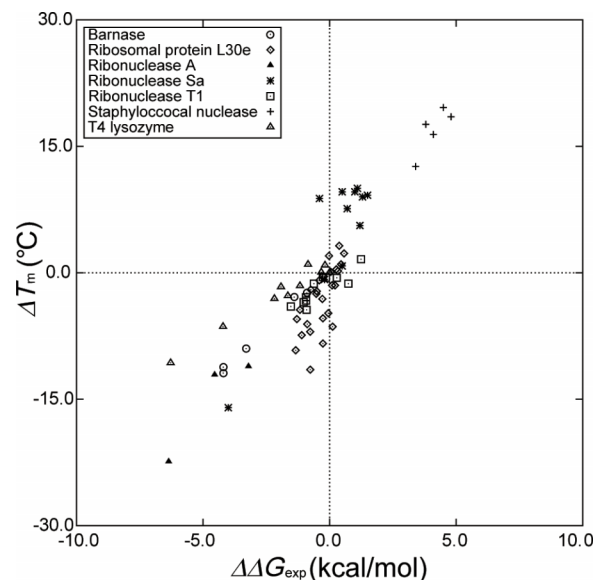


FIG. 5. Relation between  $\Delta T_m$  and  $\Delta\Delta G_{\text{exp}}$  for the 73 mutants with the data of  $\Delta\Delta G_{\text{exp}}$  available. See the caption of Fig. 4 for the notation. The subscript “exp” denotes an experimental value.

We then employ  $\Delta\Delta G_{\text{exp}}$  instead of  $\Delta T_m$  in the performance evaluation. Figure 6(a) shows the plot of  $\Delta\Delta G_{\text{exp}}$  against  $\Delta\Sigma_0$  for our approach and Fig. 6(b) shows the plot of  $\Delta\Delta G_{\text{exp}}$  against  $\Delta\Delta G_{\text{cal}}$ <sup>2,23,24</sup> for FOLD-X. The 73 mutants are considered (i.e., the other 134 mutants are excluded). The correlation coefficients in Figs. 6(a) and 6(b) are 0.314 and 0.473, respectively. The five performance measures for our approach and FOLD-X are compared in Table IV. Except for  $M_R$ , the measures are considerably higher in FOLD-X than in ours.  $M_A$  is 64.4% in ours and 71.2% in FOLD-X. Overall, the performance of FOLD-X is higher than ours. This result is not surprising because the weighting coefficients employed in FOLD-X are determined in advance so that  $\Delta\Delta G_{\text{cal}}$  can be best fitted to  $\Delta\Delta G_{\text{exp}}$  for a sufficiently large set of proteins and mutations. By contrast, our approach includes no such fitting procedure.

To confirm that the higher performance of FOLD-X mentioned above is attributed not to the exclusion of the 134 mutants but to the replacement of  $\Delta T_m$  by  $\Delta\Delta G_{\text{exp}}$ , we plot  $\Delta T_m$  against  $\Delta\Sigma_0$  for ours and against  $\Delta\Delta G_{\text{cal}}$  for FOLD-X only for the 73 mutants. The results are shown in Figs. 7(a) and 7(b) where the correlation coefficients are 0.583 and 0.540, respectively. FOLD-X exhibits lower performance with respect to the correlation coefficient. The five performance measures for ours and FOLD-X are compared in Table V:  $M_R$  and  $M_N$  are higher in ours than in FOLD-X, but the opposite is true for the other three measures ( $M_A$  is 75.3% in ours and 79.5% in FOLD-X). However, the differences are not large.

In summary, our approach and FOLD-X are better suited to  $\Delta T_m$  and  $\Delta\Delta G_{\text{exp}}$ , respectively, in the exhibition of higher performance. The reason why ours is better suited to  $\Delta T_m$  may be the following.  $\Delta T_m$  can directly be measured, whereas  $\Delta\Delta G_{\text{exp}}$  at 25 °C is evaluated using the enthalpy and heat-capacity changes upon unfolding *at the denaturation temperature* and the assumption that the heat-capacity change is

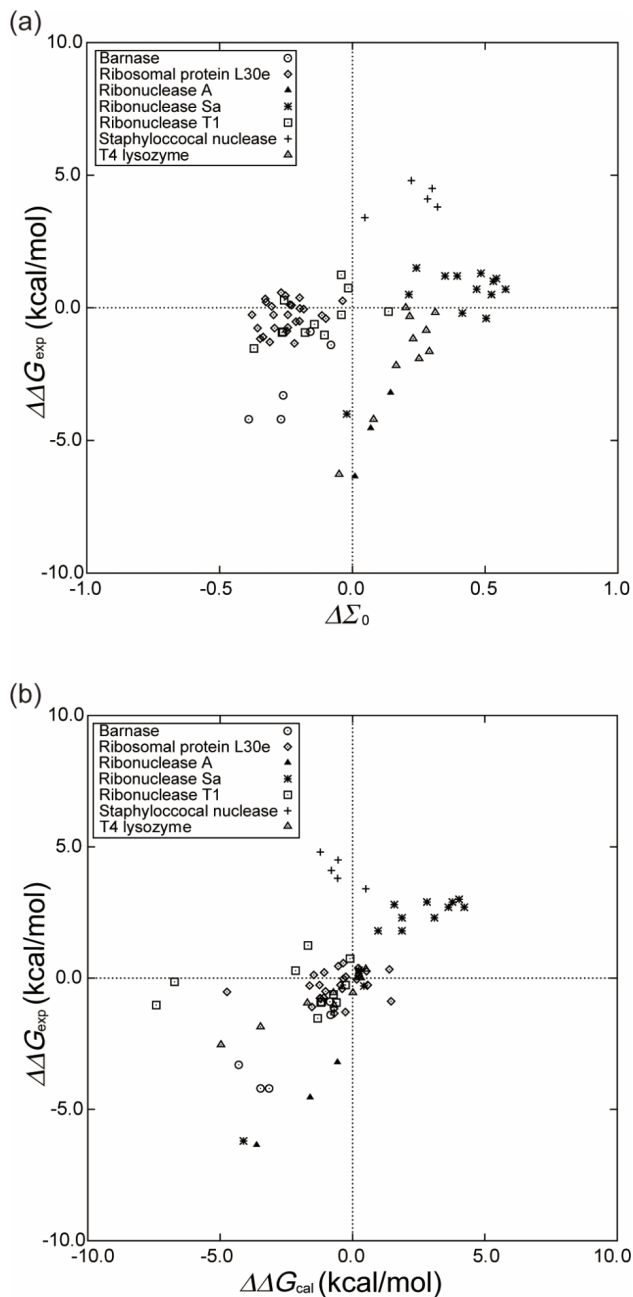


FIG. 6. (a) Relation between  $\Delta\Delta G_{\text{exp}}$  and  $\Delta\Sigma_0$  calculated by our approach for the 73 mutants with the data of  $\Delta\Delta G_{\text{exp}}$  available. (b) Relation between  $\Delta\Delta G_{\text{exp}}$  and  $\Delta\Delta G_{\text{cal}}$ . The subscript “cal” denotes the value calculated by FOLD-X. See the captions of Figs. 4 and 5 for the notation.

independent of  $T$ . Actually, the assumption is not quite correct because the heat-capacity change increases progressively as  $T$  becomes lower. (See our earlier publication<sup>11</sup> for a detailed discussion.)  $\Delta T_m$  may be more reliable than  $\Delta\Delta G_{\text{exp}}$ .

TABLE IV. Comparison between our approach and FOLD-X in terms of the performance measures defined by Eq. (7). The data points in Fig. 6 are considered.

Approach	$M_A$	$M_R$	$M_P$	$M_S$	$M_N$
Ours	0.644	0.571	0.533	0.689	0.721
FOLD-X	0.712	0.571	0.640	0.800	0.750

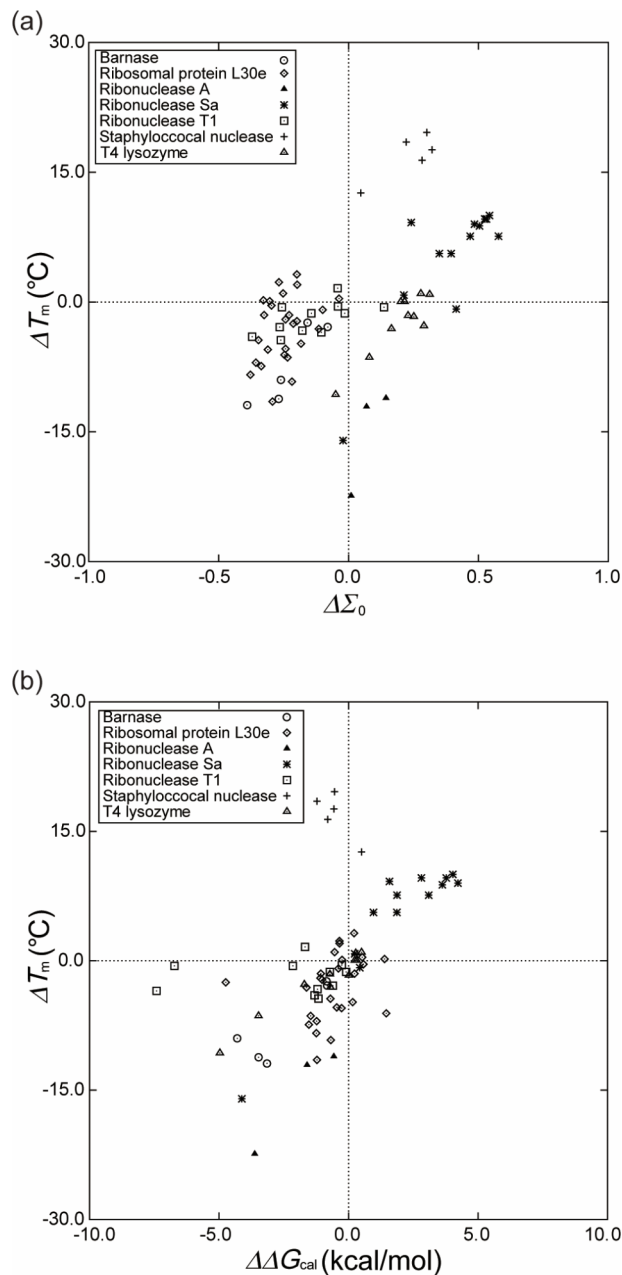


FIG. 7. (a) Relation between  $\Delta T_m$  and  $\Delta\Sigma_0$  calculated by our approach for the 73 mutants considered in Fig. 6. (b) Relation between  $\Delta T_m$  and  $\Delta\Delta G_{\text{cal}}$ . The subscript “cal” denotes the value calculated by FOLD-X. See the caption of Fig. 4 for the notation.

#### D. Comparison with FOLD-X in terms of performance for multiple mutations

Taking only the 31 multiple mutations, we plot  $\Delta T_m$  against  $\Delta\Sigma_0$  for our approach and against  $\Delta\Delta G_{\text{cal}}$  for FOLD-X

TABLE V. Comparison between our approach and FOLD-X in terms of the performance measures defined by Eq. (7). The data points in Fig. 7 are considered.

Approach	$M_A$	$M_R$	$M_P$	$M_S$	$M_N$
Ours	0.753	0.714	0.667	0.778	0.814
FOLD-X	0.795	0.679	0.760	0.867	0.813

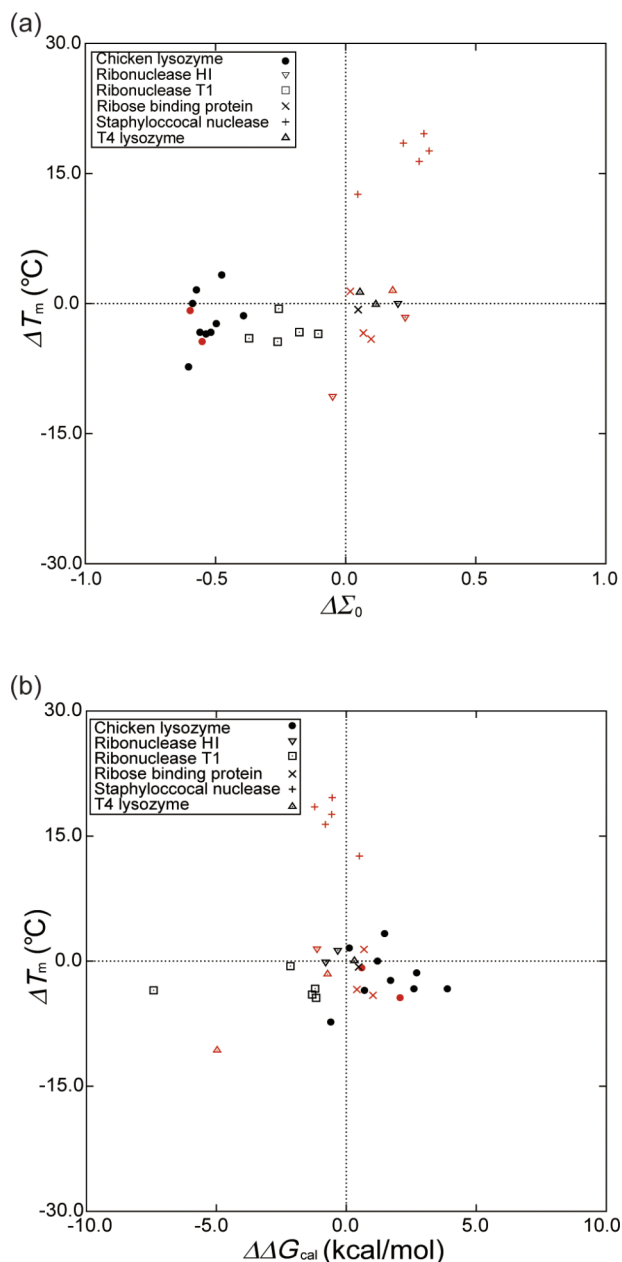


FIG. 8. (a) Relation between  $\Delta T_m$  and  $\Delta \Sigma_0$  calculated by our approach for the 31 multiple mutants in Table I. (b) Relation between  $\Delta T_m$  and  $\Delta \Delta G_{\text{cal}}$ . The subscript “cal” denotes the value calculated by FOLD-X. The keys for the triple and higher-fold mutations are drawn in red. See the caption of Fig. 4 for the notation.

in Figs. 8(a) and 8(b), respectively. The correlation coefficient in Fig. 8(a) is 0.549 whereas that in Fig. 8(b) is only 0.024. The correlation coefficient in FOLD-X for the 18 double mutations is 0.177 whereas that for the triple and higher-fold mutations is *negative*. There is a strong trend in FOLD-X that the prediction performance becomes progressively worse as the time of the mutations increases. In Table VI, we compare FOLD-X with our approach in terms of the five performance measures. All of the measures in ours are much higher than those in FOLD-X.  $M_A$  is 74.2% in ours and 48.4% in FOLD-X.

There are two possible reasons for the ill success of FOLD-X for multiple mutations. The first reason is that adjustment of the terms in its free-energy function and determination

TABLE VI. Comparison between our approach and FOLD-X in terms of the performance measures defined by Eq. (7). The data points in Fig. 8 are considered.

Approach	$M_A$	$M_R$	$M_P$	$M_S$	$M_N$
Ours	0.742	0.750	0.643	0.737	0.824
FOLD-X	0.484	0.500	0.375	0.474	0.600

of the weighting coefficient in the van der Waals term are performed only for single mutations.<sup>2,23,24</sup> The second one is much more important. As observed in Fig. 8,  $T_m$  of staphylococcal nuclease is made higher by 12.6–19.6 °C due to the 5 multiple mutations.<sup>51</sup> FOLD-X fails to reproduce these highly enhanced stabilities. By contrast, our approach is successful in reproducing them. We find the following: As shown in Fig. 9, this success is thanks to the entropic component of  $\Delta \Sigma_0$ ,  $\Delta S_{\text{VH}}/(k_B N_r)$ , and the energetic component,  $-\Delta \Lambda/(k_B T_0 N_r)$ , works to *lower* the stabilities. Thus, the highly enhanced stabilities are brought by the entropic EV effect.

The free-energy function of FOLD-X is parameterized using only the experimental data for single mutations. If it were re-parameterized using the experimental data including those for multiple mutations, the result from FOLD-X would certainly be improved. It should be emphasized, however, that the entropic EV effect is not incorporated in the free-energy function of FOLD-X. Without incorporating this crucial effect, even if the result is improved by the re-parameterization, the improvement is not physically sound.

### E. Cases where structural data for folded states of mutants are experimentally available

The performance of our approach will become higher if the preparation method for the folded state of a mutant is improved. As an illustration, we limit the prediction to the mutants whose structural data for folded states are experimentally available (a total of 22 mutants). The experimental data are used for these mutants. The method of preparing the folded structure for each mutant is the same as that for the wild type

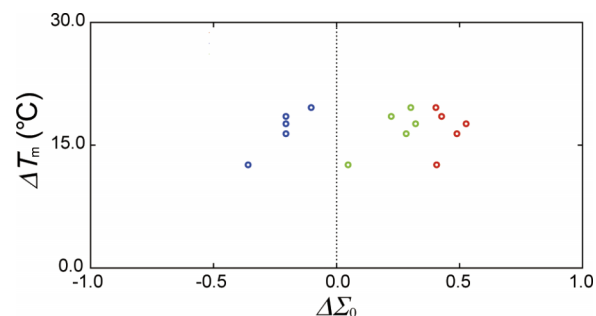


FIG. 9. Relation between  $\Delta T_m$  and  $\Delta \Sigma_0$  calculated by our approach for the 5 multiple mutations of staphylococcal nuclease. Red: Only the entropic component of  $\Delta \Sigma_0$  is considered;  $\Delta \Sigma_0 = \Delta S_{\text{VH}}/(k_B N_r)$ . Blue: Only the energetic component of  $\Delta \Sigma_0$  is considered;  $\Delta \Sigma_0 = -\Delta \Lambda/(k_B T_0 N_r)$ . Green: Both of the entropic and energetic components of  $\Delta \Sigma_0$  are considered;  $\Delta \Sigma_0 = \Delta S_{\text{VH}}/(k_B N_r) - \Delta \Lambda/(k_B T_0 N_r)$ . This figure should be compared with the data points for the five multiple mutations of staphylococcal nuclease in Fig. 8(b).



described in Sec. III F. We find that the correlation coefficient between  $\Delta T_m$  and  $\Delta \Sigma_0$  reaches 0.678.

There are two different data for chicken lysozyme: The PDB codes are 4LYZ<sup>54</sup> and 1HEL.<sup>70</sup> The two data give the results which are different in a quantitative sense. In all of the sections described above, we adopt 4LYZ leading to better performance-test results for both of our approach and FOLD-X. When the structural data experimentally determined are used for the mutants, however, 1HEL is adopted because the result for chicken lysozyme is quantitatively better for 1HEL.

## V. CONCLUDING REMARKS

We have examined a new measure of the thermal stability of a protein by applying it to the prediction of the thermal-stability changes upon mutations. In the measure, the entropic EV effect<sup>10-16</sup> is incorporated to its full extent using a molecular model for water. A unique approach has thus been obtained for the prediction using the measure. The performance of our approach would become much higher by the introduction of fitting parameters (e.g., by optimizing two weighting coefficients multiplied to  $T\Delta S_{VH}$  and to  $\Delta A$  in Eq. (3), respectively, by adjusting  $7k_B T_0$  or  $-14k_B T_0$  in Fig. 3), in which case, however, the thermal-stability measure or the free-energy function becomes somewhat physically ambiguous. Hence, the use of those parameters is avoided.

We have compared our approach with FOLD-X<sup>2,23,24</sup> in terms of the prediction performance for 10 proteins and a total of 207 mutations including 31 multiple mutations. Further, 4 of the 10 proteins possess S-S bonds and the 5 multiple mutations for staphylococcal nuclease<sup>51</sup> include Pro  $\rightarrow$  Gly, which presents much challenge. FOLD-X is one of the most popular, successful approaches. Our approach and FOLD-X exhibit almost the same performance on the whole, despite that ours employ no fitting parameters unlike FOLD-X. The factors taken into account in our thermal-stability measure are the water-entropy gain, loss of the protein conformational entropy, break of protein-water HBs, and formation of protein intramolecular HBs upon folding (i.e., the protein thermal stability is described by the competition of  $\Sigma(T)$  and  $T|\Delta S_C|/(k_B T_0 N_T)$  as illustrated in Fig. 2). In our approach, when the structures of folded states of the mutants are experimentally available and they are utilized, the correlation coefficient between  $\Delta T_m$  and  $\Delta \Sigma_0$  reaches 0.678. These successful results indicate that the three physicochemical factors govern the thermal-stability changes upon mutations.

For multiple mutants, ours is far superior to FOLD-X. This superiority is important, because most of the mutations realizing remarkable enhancement of the stability are multiple mutations. For example,  $T_m$  of staphylococcal nuclease could be made higher by 12.6–19.6 °C only by multiple mutations.<sup>51</sup> Moreover, even when two single mutations, (a) and (b), lead to higher thermal stability, the double mutation of (a) plus (b) does not necessarily result in further higher stability. It is thus crucial to successfully predict the stability change brought by a multiple mutation. We have argued that the highly enhanced stabilities for five multiple mutations of staphylococcal nuclease are realized by the entropic EV effect and that

a principal reason for the ill success of FOLD-X is the neglect of this effect. We believe that the entropic EV effect, which is quite sensitive to the protein structural changes, generally plays a pivotal role in achieving high enhancement of the stability by means of mutation. It is worthwhile to incorporate this effect in the free-energy function of FOLD-X and re-parameterize it using the experimental data including those for multiple mutations.

Khechinashvili *et al.*<sup>71-73</sup> made thermodynamic analyses on thermal denaturation of mesophilic and thermophilic proteins and draw the following conclusions: The thermal stability is not correlated with the protein intramolecular energy and it has entropic nature, and the conformational-entropy gain upon denaturation becomes more substantial as the temperature increases, which plays a pivotal role in the inducement of thermal denaturation. The first conclusion is consistent with our claim that the water-entropy effect is a crucially important factor in the thermal stability, and the second conclusion is in line with our physical picture of thermal denaturation illustrated in Fig. 2.

In the present article, we are interested in not only the physicochemical aspects of the thermal-stability changes upon mutations but also the prediction performance. A future study pursuing higher performance is of value: It can be obtained by improving the energetic and entropic components of the free-energy function. In particular, the protonation states of some residues might be changeable during the folding or unfolding process, and this effect should be examined. Our method of preparing the folded structure of a mutant seems to be cruder than that employed in FOLD-X. In fact, when the experimentally determined mutant structures are used (when they are available), the performance of our method is considerably improved as explained in Sec. IV E. The simplified model of the unfolded state (i.e., a set of five extended structures described in Sec. III H) is also to be reconsidered.

## ACKNOWLEDGMENTS

The computer program for the MA was developed with R. Roth and Y. Harano. This work was supported by JSPS (Japan Society for the Promotion of Science) Grant-in-Aid for Scientific Research (B) (No. 25291035: M. Kinoshita) and by Grant-in-Aid for JSPS fellows (H. Oshima).

<sup>1</sup>A. Mozo-Villiarías and E. Querol, *Curr. Bioinf.* **1**, 25 (2006).

<sup>2</sup>R. Guerois, J. E. Nielsen, and L. Serrano, *J. Mol. Biol.* **320**, 369 (2002).

<sup>3</sup>N. Pokala and T. M. Handel, *J. Mol. Biol.* **347**, 203 (2005).

<sup>4</sup>S. Yin, F. Ding, and N. V. Dokholyan, *Nat. Methods* **4**, 466 (2007).

<sup>5</sup>V. Potapov, M. Cohen, and G. Schreiber, *Protein Eng., Des. Sel.* **22**, 553 (2009).

<sup>6</sup>A. Benedix, C. M. Becker, B. L. de Groot, A. Caffisch, and R. A. Böckmann, *Nat. Methods* **6**, 3 (2009).

<sup>7</sup>Z. Zhang, L. Wang, Y. Gao, J. Zhang, M. Zhenirovskyy, and E. Alexov, *Bioinformatics* **28**, 664 (2012).

<sup>8</sup>L. Wickstrom, E. Gallicchio, and R. M. Levy, *Proteins* **80**, 111 (2012).

<sup>9</sup>A. N. Naganathan, *J. Phys. Chem. B* **117**, 4956 (2013).

<sup>10</sup>M. Kinoshita, *Front. Biosci.* **14**, 3419 (2009).

<sup>11</sup>M. Kinoshita, *Int. J. Mol. Sci.* **10**, 1064 (2009).

<sup>12</sup>M. Kinoshita, *Biophys. Rev.* **5**, 283 (2013).

<sup>13</sup>T. Yoshidome, M. Kinoshita, S. Hirota, N. Baden, and M. Terazima, *J. Chem. Phys.* **128**, 225104 (2008).

<sup>14</sup>T. Yoshidome and M. Kinoshita, *Phys. Chem. Chem. Phys.* **14**, 14554 (2012).

- <sup>15</sup>Y. Harano, T. Yoshidome, and M. Kinoshita, *J. Chem. Phys.* **129**, 145103 (2008).
- <sup>16</sup>H. Oshima and M. Kinoshita, *J. Chem. Phys.* **138**, 245101 (2013).
- <sup>17</sup>S.-H. Chong and S. Ham, *Chem. Phys. Lett.* **535**, 152 (2012).
- <sup>18</sup>G. Graziano, *Phys. Chem. Chem. Phys.* **16**, 21755 (2014).
- <sup>19</sup>K. Amano, T. Yoshidome, Y. Harano, K. Oda, and M. Kinoshita, *Chem. Phys. Lett.* **474**, 190 (2009).
- <sup>20</sup>K. Oda, R. Kodama, T. Yoshidome, M. Yamanaka, Y. Sambongi, and M. Kinoshita, *J. Chem. Phys.* **134**, 025101 (2011).
- <sup>21</sup>T. Yoshidome, K. Oda, Y. Harano, R. Roth, Y. Sugita, M. Ikeguchi, and M. Kinoshita, *Proteins* **77**, 950 (2009).
- <sup>22</sup>S. Yasuda, T. Yoshidome, Y. Harano, R. Roth, H. Oshima, K. Oda, Y. Sugita, M. Ikeguchi, and M. Kinoshita, *Proteins* **79**, 2161 (2011).
- <sup>23</sup>J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, *Nucleic Acids Res.* **33**, 382 (2005).
- <sup>24</sup>J. W. H. Schymkowitz, F. Rousseau, I. C. Martins, J. Ferkinghoff-Borg, F. Stricher, and L. Serrano, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10147 (2005).
- <sup>25</sup>S. Yasuda, H. Oshima, and M. Kinoshita, *J. Chem. Phys.* **137**, 135103 (2012).
- <sup>26</sup>B. Honig and A. Nicholls, *Science* **268**, 1144 (1995).
- <sup>27</sup>N. Baden, S. Hirota, T. Takabe, N. Funasaki, and M. Terazima, *J. Chem. Phys.* **127**, 175103 (2007).
- <sup>28</sup>N. M. Cann and G. N. Patey, *J. Chem. Phys.* **106**, 8165 (1997).
- <sup>29</sup>M. Kinoshita, *J. Chem. Phys.* **128**, 024507 (2008).
- <sup>30</sup>J. Fitter, *Biophys. J.* **84**, 3924 (2003).
- <sup>31</sup>P. G. Kusalik and G. N. Patey, *J. Chem. Phys.* **88**, 7715 (1988).
- <sup>32</sup>P. G. Kusalik and G. N. Patey, *Mol. Phys.* **65**, 1105 (1988).
- <sup>33</sup>T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, and F. Hirata, *J. Chem. Phys.* **125**, 024911 (2006).
- <sup>34</sup>M. Kinoshita and D. R. Bérard, *J. Comput. Phys.* **124**, 230 (1996).
- <sup>35</sup>R. Roth, Y. Harano, and M. Kinoshita, *Phys. Rev. Lett.* **97**, 078101 (2006).
- <sup>36</sup>M. L. Connolly, *J. Appl. Crystallogr.* **16**, 548 (1983).
- <sup>37</sup>M. L. Connolly, *J. Am. Chem. Soc.* **107**, 1118 (1985).
- <sup>38</sup>A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B* **102**, 3586 (1998).
- <sup>39</sup>T. Hayashi, H. Oshima, T. Mashima, T. Nagata, M. Katahira, and M. Kinoshita, *Nucleic Acids Res.* **42**, 6861 (2014).
- <sup>40</sup>H. Oshima and M. Kinoshita, *J. Chem. Phys.* **142**, 145103 (2015).
- <sup>41</sup>S. F. Sneddon, D. J. Tobias, and C. L. Brooks III, *J. Mol. Biol.* **209**, 817 (1989).
- <sup>42</sup>I. K. McDonald and J. M. Thornton, *J. Mol. Biol.* **238**, 777 (1994).
- <sup>43</sup>H. Mishima, S. Yasuda, T. Yoshidome, H. Oshima, Y. Harano, M. Ikeguchi, and M. Kinoshita, *J. Phys. Chem. B* **116**, 7776 (2012).
- <sup>44</sup>J. T. Kellis, Jr., K. Nyberg, and A. R. Fersht, *Biochemistry* **28**, 4914 (1989).
- <sup>45</sup>D. W. Heinz, W. A. Baase, and B. W. Matthews, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3751 (1992).
- <sup>46</sup>S. R. Trevino, K. Gokulan, S. Newsom, R. L. Thurlkill, K. L. Shaw, V. A. Mitkevich, A. A. Makarov, J. C. Sacchettini, J. M. Scholtz, and C. N. Pace, *J. Mol. Biol.* **354**, 967 (2005).
- <sup>47</sup>S. Kanaya, M. Oobatake, H. Nakamura, and M. Ikehara, *J. Biotechnol.* **28**, 117 (1993).
- <sup>48</sup>N. C. Vercillo, K. J. Herald, J. M. Fox, B. S. Der, and J. D. Dattelbaum, *Protein Sci.* **16**, 362 (2007).
- <sup>49</sup>C.-F. Lee, G. I. Makhatadze, and K.-B. Wong, *Biochemistry* **44**, 16817 (2005).
- <sup>50</sup>M. P. Byrne and W. E. Stites, *Biophys. Chem.* **125**, 490 (2007).
- <sup>51</sup>J. Chen, Z. Lu, J. Sakon, and W. E. Stites, *J. Mol. Biol.* **303**, 125 (2000).
- <sup>52</sup>T. Kadonosono, E. Chatani, R. Hayashi, H. Moriyama, and T. Ueki, *Biochemistry* **42**, 10651 (2003).
- <sup>53</sup>W.-D. Schubert, G. Schluckebier, J. Backmann, J. Granzin, C. Kisker, H.-W. Choe, U. Hahn, W. Pfeil, and W. Saenger, *Eur. J. Biochem.* **220**, 527 (1994).
- <sup>54</sup>P. Shih, D. R. Holland, and J. F. Kirsch, *Protein Sci.* **4**, 2050 (1995).
- <sup>55</sup>Y. W. Chen, M. Bycroft, and K.-B. Wong, *Biochemistry* **42**, 2857 (2003).
- <sup>56</sup>S. Kanaya, A. Kohara, M. Miyagawa, T. Matsuzaki, K. Morikawa, and M. Ikehara, *J. Biol. Chem.* **264**, 11546 (1989).
- <sup>57</sup>A. J. Björkman and S. L. Mowbray, *J. Mol. Biol.* **279**, 651 (1998).
- <sup>58</sup>R. Diamond, *J. Mol. Biol.* **82**, 371 (1974).
- <sup>59</sup>D. L. Birdsall and A. McPherson, *J. Biol. Chem.* **267**, 22230 (1992).
- <sup>60</sup>J. D. Thompson, D. G. Higgins, and T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
- <sup>61</sup>B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- <sup>62</sup>M. Feig, J. Karanicolas, and C. L. Brooks III, *J. Mol. Graphics Modell.* **22**, 377 (2004).
- <sup>63</sup>A. D. Mackerell, Jr., M. Feig, and C. L. Brooks III, *J. Comput. Chem.* **25**, 1400 (2004).
- <sup>64</sup>M. S. Lee, M. Feig, F. R. Salsbury, Jr., and C. L. Brooks III, *J. Comput. Chem.* **24**, 1348 (2003).
- <sup>65</sup>J. Chochołoušová and M. Feig, *J. Comput. Chem.* **27**, 719 (2006).
- <sup>66</sup>A. Šali and T. L. Blundell, *J. Mol. Biol.* **234**, 779 (1993).
- <sup>67</sup>R. L. Dunbrack, Jr. and F. E. Cohen, *Protein Sci.* **6**, 1661 (1997).
- <sup>68</sup>R. L. Dunbrack, Jr., *Curr. Opin. Struct. Biol.* **12**, 431 (2002).
- <sup>69</sup>H. Oshima, S. Yasuda, T. Yoshidome, M. Ikeguchi, and M. Kinoshita, *Phys. Chem. Chem. Phys.* **13**, 16236 (2011).
- <sup>70</sup>K. P. Wilson, B. A. Malcolm, and B. W. Matthews, *J. Biol. Chem.* **267**, 10842 (1992).
- <sup>71</sup>N. N. Khechinashvili, M. V. Fedorov, A. V. Kabanov, S. Monti, C. Ghio, and K. Soda, *J. Biomol. Struct. Dyn.* **24**, 255 (2006).
- <sup>72</sup>N. N. Khechinashvili, S. A. Volchkov, A. V. Kabanov, and G. Barone, *Biochim. Biophys. Acta* **1784**, 1830 (2008).
- <sup>73</sup>N. N. Khechinashvili, A. V. Kabanov, M. S. Kondratyev, and R. V. Polozov, *J. Biomol. Struct. Dyn.* **32**, 1396 (2014).