Cavity-Filling Mutations Enhance Protein Stability by Lowering the Free Energy of Native State

Minoru Saito,*,† Hidetoshi Kono,‡,§ Hisayuki Morii,|| Hatsuho Uedaira,§ Tahir H. Tahirov,⊥ Kazuhiro Ogata,⊥ and Akinori Sarai*,§

Faculty of Science and Technology, Hirosaki University, 3 Bunkyo-cho, Hirosaki, Aomori 036-8561, Japan, Tsukuba Life Science Center, The Institute of Physical & Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan, National Institute of Bioscience and Human-Technology, 1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan, and Kanagawa Academy of Science and Technology (KAST), 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan, and Department of Structural Biology, Yokohama City University School of Medicine, 3−9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan

Received: May 25, 1999; In Final Form: December 13, 1999

The stabilization of proteins due to cavity-filling mutations are thought to be attributable to removal of hydrophobic residues from solvent exposure in denatured (D) state and formation of close packing in native (N) state. However, it is still unclear which contribution is dominant to stabilize proteins, because experiments can probe only the free energy difference between the two states (N and D). To address this question, we carried out molecular dynamics simulations, circular dichroism (CD) measurements, and X-ray crystallographic experiments on the cavity-filling mutations of the DNA-binding domain of the Myb transcriptional regulator. The cavity size was altered by systematic natural and nonnatural amino acid substitutions at a fixed site. The stability free energy change ($\Delta\Delta G$(N→D;W→M)) and the cavity-size change (ΔV) calculated for the mutations agreed with the experimental data observed by urea-titration/CD measurements and crystallographic structure analysis, respectively. We found that the experimental $\Delta\Delta G$ values correlate well with the calculated native-state free energy change due to mutations ΔG(N;W→M) and with ΔV (their correlation coefficients are larger than 0.9) but not with the denatured-state ΔG(D;W→M). These results demonstrated that the decrease in cavity size increases the protein stability by lowering the free energy of native state for this protein. We discussed physicochemical meanings of our calculation results for ΔG(N;W→M) and ΔG(D;W→M).

1. Introduction

The thermodynamic stability of proteins has been extensively investigated to clarify the stabilizing mechanisms of their tertiary structures and to create artificially stabilized proteins for bioreactors. Among the various strategies developed to stabilize native protein structures,[1,2] a cavity-filling strategy has been applied for various proteins.[3−6] Two major factors in these strategies have been proposed to contribute to the protein stability: removal of hydrophobic residues from exposure to solvent in denatured (D) state[7,8] and formation of close packing in the interior of protein in native (N) state.[9,10]

It is difficult to assess which factor, exposure in the D state or packing in the N state, is dominant to stabilize proteins, because these factors originate from the different states and experimental stability studies measure relative free energy differences between the N and D states. As shown in Figure 1, the experimental studies measure the free energy differences between the native (N) and denatured (D) states for the wild type (W), ΔG (N→D;W), and for a mutant (M), ΔG(N→D;M),

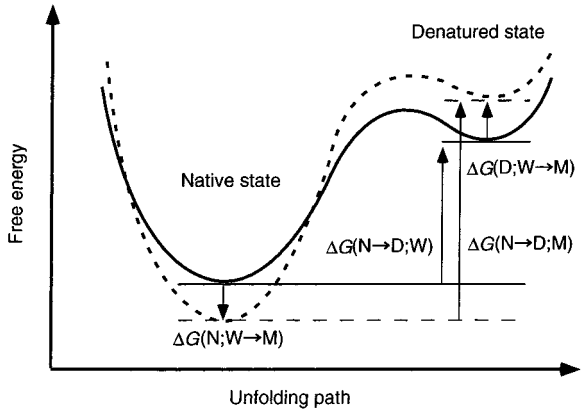

Figure 1. Free energy profile of the thermodynamic stability of proteins. The thermodynamic stability of proteins is determined by a delicate balance between the free energies of the native (N) and denatured (D) states. Experimental methods directly measure the unfolding free energies of wild type (W) and mutant proteins (M), ΔG(N→D;W) and ΔG(N→D;M). On the other hand, theoretical methods calculate free energy changes caused by a mutation for the individual native and denatured states, ΔG(N;W→M) and ΔG(D;W→M), respectively, and estimate the unfolding free energy difference $\Delta\Delta G$(N→D;W→M) = ΔG(N→D;M) − ΔG(N→D;W) = ΔG(D;W→M) − ΔG(N;W→M).

* Corresponding authors.
† Hirosaki University. E-mail: msaito@si.hirosaki-u.ac.jp. FAX: +81-172-39-3656.
‡ Present address; Dept. of Chemistry, Univ. of Pennsylvania, 231 S. 34 th. Philadelphia PA19104.
§ The Institute of Physical & Chemical Research (RIKEN). E-mail: hkono@rtc.riken.go.jp, sarai@rtc.riken.go.jp, uedaira@rtc.riken.go.jp. FAX: +81-298-36-9080.
|| National Institute of Bioscience and Human-Technology. E-mail: morii@nibh.go.jp. FAX: +81-298-54-6194.
⊥ Kanagawa Academy of Science and Technology (KAST) and Yokohama City University School of Medicine. E-mail: ogata@med.yokohama-cu.ac.jp. FAX: +81-45-781-6763.

10.1021/jp991717f CCC: $19.00
Published on Web 03/18/2000

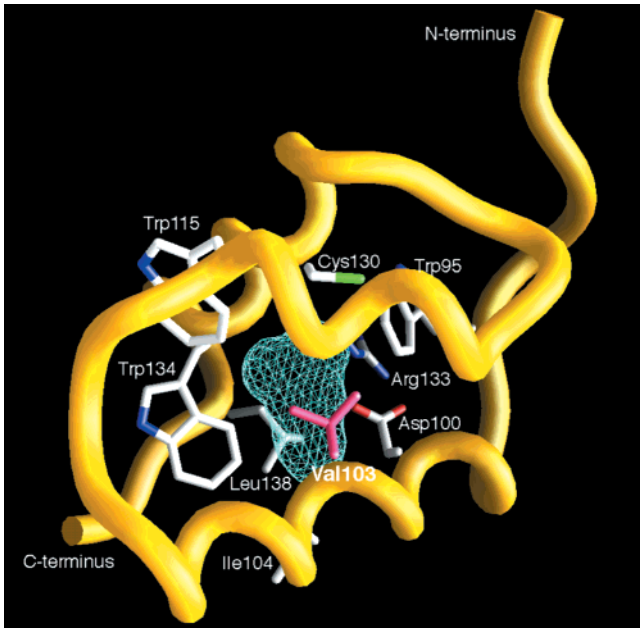

Figure 2. Three-dimensional structure of the wild-type Myb R2 domain resolved by NMR-structural studies.[12] For simplicity, only the main-chain skeleton is drawn. A cavity surrounded by Val103 and eight other residues is shown in cyan. The cavity was produced by a Connolly surface with a 1.2 Å probe. This figure was drawn by GRASP.[34]

but not the free energy changes for each state caused by the mutation, ΔG(N;W→M) and ΔG(D;W→M). In contrast, the free energy perturbation method based on molecular dynamics (MD) simulations permits a direct calculation of ΔG(N;W→M) and ΔG(D;W→M).

Our intent is to clarify the mechanism of protein stability for cavity-filling mutations by a combination of experimental stability measurements, structural analyses, and theoretical calculations. To this end, we probed the stability change by systematic natural and nonnatural mutations at a fixed site in the DNA-binding domain of the c-Myb transcriptional regulator (Figure 2). We will not investigate mutations to significantly larger amino acids than the cavity at least in this study, because severe steric hindrances unquestionably destabilize the protein.

The Myb DNA-binding domain warranted special attention because of the intimate relationship among its cavity, stability, and function. The DNA-binding domain of Myb consists of three homologous repeat units (R1, R2, and R3) of 51 to 52 amino acids each. The melting temperature (T_m) of R2 is 15−20 °C lower than those of R1 and R3.[11] An NMR analysis of each repeat unit has shown that R2 has a cavity formed by Val103 and other amino acids in its hydrophobic core.[12] A V103L mutation of R2, which fills the cavity, was found to stabilize the protein considerably but decrease its affinity for DNA and transactivation capacity.[13] Thus, the cavity in R2 is considered to play an important functional role.

First, Val103 in the Myb R2 domain was substituted by the following natural and nonnatural amino acids with alkyl side chains: Leu, norvaline (Nva), allo-isoleucine (Ail), Ile, α-amino-*n*-butyric acid (Abu), and Ala, all of which can systematically change the size and shape of the cavity at the mutation site. Second, the changes in ΔG for the unfolding process ($\Delta\Delta G$), resulting from the mutations, were measured by circular-dichroism (CD) urea-titration studies; the cavity-volume changes (ΔV) caused by the mutations were measured by X-ray crystallography.

ΔG(N;W→M), ΔG(D;W→M), $\Delta\Delta G$, and ΔV were simultaneously calculated by the free energy perturbation method for the protein with the explicit inclusion of water, incorporating long-range Coulomb interactions, and all degrees of freedom.[14,15] Free energy changes were calculated by the free energy perturbation method based on MD simulations, in the same manner that the ΔT_m and the pK_a shift of RNaseHI were previously calculated.[16−18] To our knowledge, this is the first systematic single-site study for cavity-filling mutations including nonnatural amino acids based on simulations and experiments.

We compared the calculation values of ΔG(N;W→M), ΔG(D;W→M), $\Delta\Delta G$, and ΔV with the experimental values and answered the question: which factor, solvent exposure in the D state or packing in the N state, is dominant to stabilize the Myb protein? We discussed physicochemical meanings of our calculation results for ΔG(N;W→M) and ΔG(D;W→M).

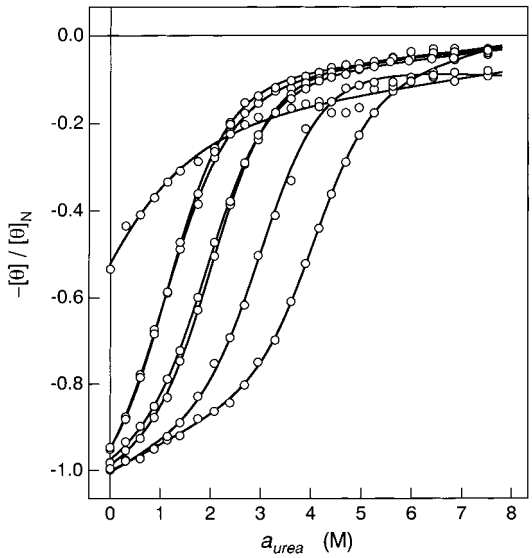

Figure 3. Urea titration curves of wild-type and mutants of Myb R2 domain. The abscissa and the ordinate respectively represent the activity of urea and the relative molar ellipticity, which was normalized by the molar ellipticity for each native state ($[\theta]_N$) in the absence of urea at 300 K. The open circles and the solid curves denote the observed and the fitted data, respectively. The calculation for curve-fitting was based on nonlinear expression of ΔG as described in methods section. The linear expression of ΔG gave similar curves showing a good fit with nonlinear ones (results not shown). The fitted curves are in the order of Ala, Abu, Val, Ile, Ail, Nva, and Leu from left to right at −0.4 of the ordinate level.

2. Experimental and Computational Methods

2.1. Synthesis of Mutants. The wild-type and the mutants of the Myb R2 domain, which is mouse c-Myb (90−141) including the amide group at the C-terminus, were chemically synthesized by the usual Fmoc-solid-phase method. The R2 domain proteins were purified by means of reversed-phase HPLC and were lyophilized to be a fluffy powder.

2.2. Urea Titration and CD Measurements. The experimental values for the free energy change associated with unfolding were obtained by means of urea titration (Figure 3). The synthetic R2 fragments (about 0.01 mM) in a 50 mM potassium phosphate buffer solution (pH 7.5) containing 50 mM potassium chloride and 4 mM dithiothreitol were titrated with urea in the concentration range from 0 to 8 M at 300 K.

The unfolding was monitored by CD molar ellipticity using a light of wavelength 230 nm; the technique is sensitive to changes in secondary structure. The observed data were analyzed on the basis of two different models, linear and nonlinear. The linear extrapolation model is expressed by $\Delta G = \Delta G(H_2O) -$

m[Urea], where [Urea] is the molar concentration of urea and *m* is a constant, and the nonlinear denaturant-binding model by $\Delta G = \Delta G(H_2O) - \Delta nRT\ln(1 + \kappa a)$, where the binding constant (κ) was fixed at 0.1 and the activity of urea (a) was calculated from its concentration.[19] On the basis of two-state transition model, the titration curves can be theoretically expressed with the equation, $Y = \exp(-\Delta G/RT)(p_D C + q_D - p_N C - q_N)/\{1 + \exp(-\Delta G/RT)\} + (p_N C + q_N)$, where C denotes the concentration ([Urea]) or activity (a) of urea for linear or nonlinear model, respectively. In this experiment, the value Y is a molar ellipticity, which is the function of the variable C. The parameters, p_N, q_N, p_D, and q_D, represent the slopes (p) and intercepts (q) of the lines corresponding to the native (N) and denatured (D) states. Thus, the above expression of Y includes six parameters, i.e., p_N, q_N, p_D, q_D, $\Delta G(H_2O)$, and either m or Δn. These parameters were determined by the nonlinear least-squares method. Exceptionally, the parameter of the slope, p_N, was fixed to that obtained for the Leu mutant, because the plateau regions for native states of the other mutants were narrow.

2.3. X-ray Crystallograpy. Crystallization and preliminary X-ray diffraction data collection from wild type and V103L mutant R2 have been described.[20] Crystal structures of R2 domain mutants (V103L and V103I) have been determined by molecular replacement method using a solution structure of R2 as a search model and refined at resolution of 1.53 Å using the program XPLOR. The crystallographic R values for both current models are 17%. Full structure description at atomic resolution (~1.00 Å) is in progress and will be reported elsewhere. Therefore, the structures for the wild type and two mutants (V103L and V103I) have not yet acquired the PDB code number.

2.4. MD Simulations and Free Energy Calculations. The MD simulations were carried out for the protein in water with long-range Coulomb interactions and all degrees of freedom.[14−18] The native state was prepared by immersing Myb R2 in a 25 Å radius sphere of water. The entire system was equilibrated using a 150 ps MD simulation. The denatured state was represented by a short peptide of five amino acids (from the 101st to the 105th). Atomic force fields used were AMBER4.1/all atom type[21] for the protein and SPC for water.[22]

The unfolding free energy difference $\Delta\Delta G$(N→D;W→M) was derived from the free energy changes for the native state and denatured state, ΔG(N;W→M) and ΔG(D;W→M), as described in the caption of Figure 1. ΔG(N;W→M) and ΔG(D;W→M) were defined by subtracting the free energy changes of single amino acids in a vacuum ΔG(vacuum;W→M), which represents the free energy differences between the amino acids themselves. These free energy values do not affect $\Delta\Delta G$ because they are the same for the both states. The subtractions were necessary to compare values of ΔG(N;W→M) and ΔG(D;W→M) for different mutants, as shown in Figure 6a,b. Computational mutations and free energy calculations for ΔG(N;W→M), ΔG(D;W→M), and ΔG(vacuum;W→M) were performed by the same procedure, as follows.

The computational mutations were performed for both states by varying the atomic force fields of the 103rd residue for 10 intermediate stages after the equilibration process. Each stage consisted of a 10 ps MD simulation for equilibration and 14 ps for sampling. Free energy differences between the stages were estimated by the acceptance-ratio method.[23,38] The reliability of this methodology had been previously checked by calculating the pK_a shift and ΔT_m for RNaseHI and by comparing these values with the experimental ones.[16−18]

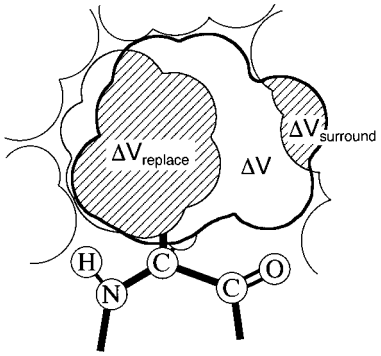

Figure 4. Schematic explanation of a calculation method for the cavity volume ΔV. The cavity is filled by the Leu residue. The Leu is substituted into smaller residues, Ala, Val, Abu, Ail, Ile, and Nva. According to the mutation from Leu to these small residues, the surrounding residues around Leu close to the mutation site and overlap with the space filled by the Leu residue V_{Leu} (this overlap volume with V_{Leu} is denoted by $\Delta V_{surround}$). The small residues also overlap partly with V_{Leu} (this overlap volume is denoted by $\Delta V_{replace}$). Then, the cavity volume ΔV is defined as a remaining volume of the space filled by the Leu residue, that is, $\Delta V = V_{Leu} - \Delta V_{surround} - \Delta V_{replace}$.

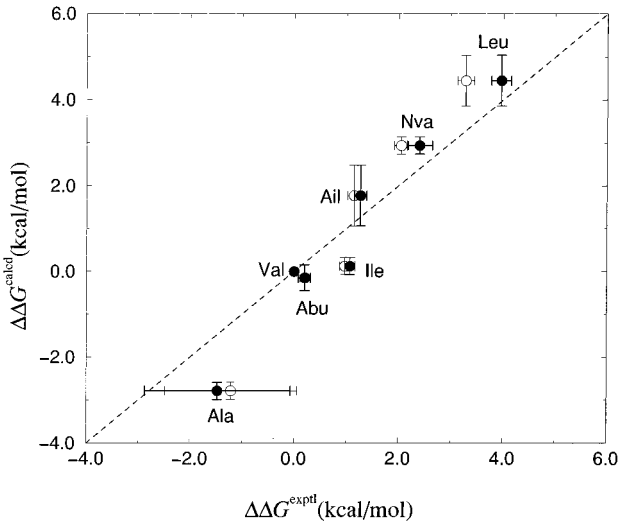

Figure 5. Correlation between the unfolding free energy changes observed (horizontal axis) and calculated (vertical axis) for the six mutants (Ala, Abu, Ile, Ail, Nva, and Leu). The experimental values were obtained from the different data-fitting models, nonlinear (closed circle) and linear (open circle), as described in the methods section. The horizontal error bars denote the error in the fitting of experimental denaturation curves. The large error for Ala results from the difficulty in determining the baseline for the native state, because the Ala mutant is mostly unfolded at room temperature (T_m is 2.8 °C). The calculation errors obtained from different mutation paths and some reverse paths are denoted by the vertical error bars. The dotted line denotes a linear function $\Delta\Delta G^{Calcd} = \Delta\Delta G^{Exptl}$.

The entropy and enthalpy differences, $T\Delta S$ and ΔU, were estimated from the following equations,

$$\Delta U = \sum_{i=0}^{n-1}(\langle u_{i+1}\rangle_{i+1} - \langle u_i\rangle_i) \quad (1)$$

$$T\Delta U = -\Delta G + \Delta U \quad (2)$$

where u_i and u_{i+1} denote the potential energies of the whole system at stage i and i+1, respectively. $\langle\ \rangle_i$ denotes the statistical average over configurations sampled by the MD simulation of the stage i. In the present simulation, eq 1 is an approximation, because ΔU is strictly the internal energy for a constant volume simulation. At the present time, it is still difficult for this kind

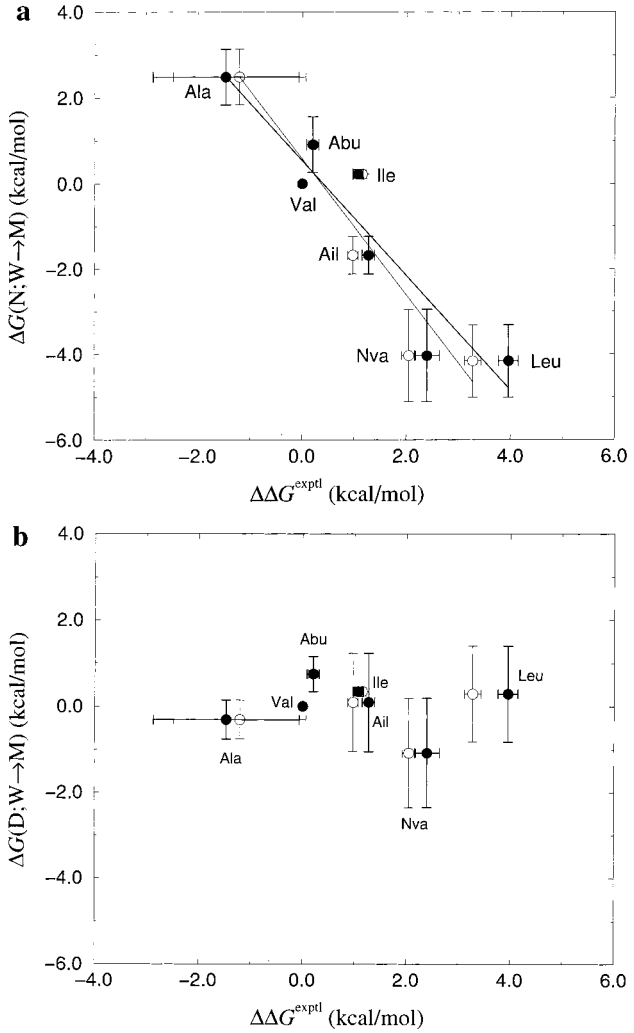

Figure 6. Correlations between the experimental values for $\Delta\Delta G$ and the calculated values for (a) ΔG(N;W→M) and (b) ΔG(D;W→M). ΔG(N;W→M) and ΔG(D;W→M) are the free energy changes due to mutations for the native state and the denatured state, respectively, as shown in Figure 1. Symbols have the same meanings as Figure 5.

of system (a protein in water) to discuss the difference in the thermodynamic quantities between the statistical ensembles. ΔU calculated from eq 1 has a large statistical deviation asymptotically proportional to a square root of N (N: number of atoms), as pointed out by Levy and Gallicchio in their recent review.[35] In fact, we recognized that ΔU and $T\Delta S$ were accompanied by large calculation errors than ΔG.

2.5. Mutation Paths. The computational mutations were performed for three initial proteins: the wild type, the Leu mutant, and the Abu mutant proteins. We used the tertiary structure of the wild-type protein, which was solved by NMR in a previous study (PDB code:1MBG).[12] We prepared the tertiary structures of the Leu and Abu mutants from the wild-type structure by molecular modeling and from the NMR information. The model structure for the Leu mutant was almost the same as the preliminary X-ray structure.[20] We considered the following mutation paths: Leu↔Val, Leu→Ail, Leu→Nva, Val→Ail, Val↔Ile, Val→Nva, Val↔Abu, Val→Ala, Abu→Ala, and Abu→Nva. $\Delta\Delta G$ was estimated from the direct paths connecting the wild type and mutant proteins and a few indirect paths connected by a third amino acid. Different mutation paths and some reverse paths were used to estimate calculation errors in the $\Delta\Delta G$ estimation. These calculation errors are denoted by the vertical error bars in Figures 5, 6a,b, and 7. Furthermore, to test whether our calculation methodology can predict the thermal stability of mutant proteins and whether different types of mutations affect the stability differently, we also performed free energy calculations for Leu→Met and Val→Thr mutations for which the experimental values are yet to be determined.

2.6. Calculation of ΔV. Changes in cavity volume (ΔV) were evaluated from the 14 ps mutant trajectory of the last stage of each mutation, using a method based on the fact that the Leu residue of the V103L mutant completely fills the cavity around the Val103 residue. This permitted us to estimate the cavity-volume changes for the other mutants in the following way (Figure 4). First, a virtual Leu residue whose volume is denoted by V_{Leu} was created at the 103rd site of the mutants. Next, the overlapping volumes of the virtual Leu residue with the mutant residue and the residues surrounding the 103rd site, $\Delta V_{\mathrm{replace}}$ and $\Delta V_{\mathrm{surround}}$, were estimated, where atomic vdW radii[21] were used to calculate the volumes. The cavity volume was obtained by subtracting the overlapping volume from the virtual Leu volume. This cavity volume represents the volume in which the residue can move about without strong vdW collisions. Our method does not include the narrow corner spaces and does not depend on variables such as the probe radius used in the Connolly method.[24] The Connolly method result is very sensitive to the probe-sphere radius; the resulting cavity can even vanish or become part of the solvent depending on the probe size. Our method gives a better correlation between the cavity volume and $\Delta\Delta G$ for the T4 lysozyme than does the Connolly method, the correlation coefficients being 0.93 compared to 0.83 for L99 mutants and 0.94 compared to 0.66 for F153 mutants[25] (details will be described elsewhere).

2.7. Program Packages. All MD simulations were carried out using a program package COSMOS90[14,38] on Fujitsu VPP700. This program was developed at Protein Engineering Research Institute (PERI) in 1990 by one of authors (M.S.). Details of the methods implemented in the programs were described in the refs 14 and 38. The amino acid mutations were performed using a program PERTURB, and the free energy calculations for the mutations were performed using a program FENE. These programs were also developed by one of the authors (M.S.) at PERI in 1994. PERTURB changes linearly force field parameters of an amino acid to be mutated to a mutant amino acid with a similar topology. FENE estimates a free energy difference between two continuous, intermediate stages by the acceptance ratio method (ARM). Details of this program were described in refs 23 and 38.

Long-range Coulomb interactions were calculated using the particle−particle and particle-cell (PPPC) method.[14,15,38] The accuracy parameters of PPPC were kept to the commonly used values. That is, the accuracy parameters for dividing space to hierarchical cells, α and β, were 0.25×10^{-3} and 1.0×10^{-3}, respectively. The radius separating space to the inner and outer regions was 7 Å. The update frequencies of the hierarchical cells and forces from the outer region were every 50 and 10 steps, respectively. The step length was 0.5 fs. Other parameter values in setting up MD simulations and free energy calculations were almost the same as those used in the previous studies.[14−18]

On a web site (http://www.si.hirosaki-u.ac.jp/~msaito), we will present all information and data necessary for readers to reproduce our calculations using other program packages as well as ours. Furthermore, one of the authors (M.S.) is preparing executable modules of the program packages (COSMOS90, PERTURB, and FENE) for widely distributed computers such as COMPAQ alpha and SGI IRIX workstations. Instructions

for readers to obtain these program packages will be described on the web site.

2.8. Data Known, Observed, and Calculated. $\Delta\Delta G^{Exptl}$ values for the wild type (Val) and 6 mutants (Ala, Abu, Ile, Ail, Nva, and Leu) were observed by the urea denaturation experiments in this study and also by the thermal denaturation experiments for melting temperature (T_m). The $\Delta\Delta G^{Exptl}$ values obtained from the transition enthalpy and T_m values highly correlated with those of the urea denaturation experiments and the calculated values $\Delta\Delta G^{Calcd}$ (correlation coefficients were 0.98 and 0.97, respectively). The details of the thermal denaturation experiments were reported in another paper.[37]

The experimental data of $\Delta\Delta G$ were not used to calculate $\Delta\Delta G$. The tertiary structure of the wild type was clarified by the previous NMR study (1MBG).[12] This structure was used as a starting structure of the mutations from the wild type (Val) to Ala, Abu, Nva, Ile, Leu, and Ail mutants. We also performed mutations from Leu and Abu mutants, whose structures were prepared by modeling. The tertiary structures of the wild type, Leu, and Ile mutants were preliminary solved by X-ray studies after performing the MD simulations. Therefore, these structures were not used as starting structures of MD simulations but used to obtain experimental values of the cavity volume difference.

3. Results

The agreement between the calculated and experimental values of $\Delta\Delta G$ for the six mutants is good (Figure 5), with a correlation coefficient of 0.98 even though the calculation method did not incorporate the experimental data for the mutants or include any adjustable parameters. This agreement suggests that the present calculations have enough reliability to analyze details of the simulation results at the atomic level.

To clarify which state, D or N, dominantly contributes to stabilize the protein, the free energy changes due to mutations for the D and N states, ΔG(N;W→M) and ΔG(D;W→M), were directly compared with the experimental $\Delta\Delta G$ (Figure 6a,b). We found that the experimental $\Delta\Delta G$ correlates well with ΔG(N;W→M) but not with ΔG(D;W→M). In other words, the differences in stability among the cavity-filling mutants mainly originate from the native state but not the denatured state. These results suggest that the decreases of cavity size due to the cavity-filling mutations stabilize the native state.

To shed additional light on the role of the cavity-filling mechanism in enhancing protein stability, we evaluated the cavity size ΔV from MD simulations and compared with ΔG(N;W→M). To ensure the reliability of the ΔV derived from the free energy simulation, we compared the calculated ΔV values with the experimental values for the two mutants (Leu and Ile), for which preliminary X-ray structure analysis was performed.[20] The ΔV values derived from the X-ray mutant structures and NMR wild-type structure[12] (30.3 Å^3 for Val and 24.2 Å^3 for Ile) were in good agreement with those derived from the simulated structures (31.2 ± 3.4 Å^3 for Val and 28.3 ± 3.5 Å^3 for Ile). This agreement supports that the cavity-size changes of the mutant proteins estimated by the simulations are reliable.

Figure 7 shows a plot of the calculated values for ΔG(N;W→M) and ΔV. It is apparent that the two quantities are highly correlated. This result demonstrates that the cavity filling around the 103rd site decreases the native-state free energy. The cavity volume change ΔV also correlated with the stability free energy change $\Delta\Delta G$ (a correlation coefficient of −0.97). Therefore, the cavity volume is the major factor controlling the thermodynamic stability of the Myb R2 domain.

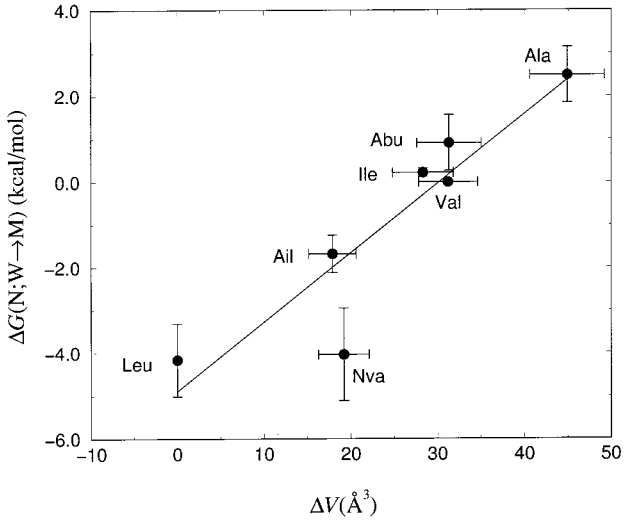

Figure 7. Correlation between the calculated values for ΔG(N;W→M) and ΔV for the six mutants. Error bars for ΔV denote the standard deviation of ΔV values over the 14 ps sampling of the last stage of each mutation.

Furthermore, the free energy change for the native state ΔG(N;W→M) was divided into contributions from various interactions, bond, angle, 1−4 nonbond, charge, van der Waals (vdW), and Intra, where Intra denotes nonbonded interactions among perturbed atoms. We found that ΔG(N;W→M) was highly correlated with the contribution from the vdW interaction ΔG_{vdw}(N;W→M) between the 103rd residue and the surrounding residues (a correlation coefficient is 0.86), but not with the contributions from other interactions (0.29, −0.06, −0.45, 0.40, and −0.05 for bond, angle, 1−4 nonbond, charge, and Intra, respectively). Thus, the decrease in the cavity volume causes a favorable vdW interaction and decreases the native-state free energy, showing an important role of the packing in protein stabilization.

It is also an interesting question which residue surrounding the mutation site mainly contributes to stabilize the protein through the vdW interaction. We might answer this question by decomposing ΔG_{vdw}(N;W→M) into contributions from the residues.[17] However, the analysis of the reason of stability and instability for each mutant is beyond the purpose of this paper. Additional details describing the energetics involved in the cavity-volume changes associated with the mutations of Myb R2 will be reported elsewhere. We will concentrate to discuss the physicochemical meanings of ΔG(N;W→M) and ΔG(N;W→M) and the validity of our results.

4. Discussion

The significant differences in stability between the Myb R2 domain mutants are mainly attributed to the native state but not to the denatured state, as shown in Figure 6a,b. ΔG(N;W→M) has a wide range of values from about −4 to +2 kcal/mol depending on the mutations. On the other hand, ΔG(D;W→M) has a narrow range of values from about −0.3 to +0.4 kcal/mol. We discuss the physicochemical origin of the different behavior of ΔG(N;W→M) and ΔG(D;W→M) for the mutations.

In the present study, ΔG(N;W→M) and ΔG(D;W→M) are defined by subtracting the free energy difference in a vacuum ΔG(vacuum;W→M) from the free energy changes in the native state and the denatured state. ΔG(vacuum;W→M) means a difference in the free energy intrinsic to the amino acids themselves. Since ΔG(vacuum;W→M) is subtracted from ΔG(N;W→M) and ΔG(D;W→M), ΔG(N;W→M) and

TABLE 1: Comparison of the Calculated Values for ΔG(N;W→M) and ΔG(D;W→M) with the Experimental Values for ΔG(Gas→Water; W→M)[a] and ΔG(Gas→Alcohol; W→M)[b] (in kcal/mol)

amino acid	ΔG(gas→water)	ΔG(D;W→M)[c]	ΔG(gas→alcohol)[c]	ΔG(N;W→M)[c]
Leu	0.29	0.29(1.11)	−0.20(0.25)	−4.16(0.85)
Ile	0.16	0.34(0.10)	−0.89(0.39)	0.22(0.11)
Val	0.0	0.0	0.0	0.0
Ala	−0.05	−0.31(0.45)	0.94(0.05)	2.49(0.65)

[a] Reference 29. [b] ΔG(gas→alcohol) was obtained from the transfer free energies ΔG(gas→water; W→M)[29] and ΔG(water→octanol, ethanol; W→M).[26−28] [c] Values in the parentheses denote the standard deviations.

ΔG(D;W→M) correspond to the differences in the transfer free energy from gas to protein interiors ΔG(gas→protein; W→M) and from gas to water ΔG(gas→water; W→M), respectively. Since the latter ΔG(gas→water; W→M) means the hydration free energy difference, ΔG(D;W→M) also corresponds to the hydration free energy difference between amino acid monomers, although the amino acid in water accompanies adjacent four amino acids. On the other hand, since the alcohol solution mimics the protein environment for an amino acid,[26−28] ΔG(N;W→M) is comparable to a difference in the transfer free energy from gas to alcohol between W and M, ΔG(gas→alcohol; W→M).

In Table 1, we compared ΔG(D;W→M) and ΔG(N;W→M) with ΔG(gas→water;W→M)[29] and ΔG(gas→alcohol;W→M), where ΔG(gas→alcohol; W→M) was estimated from ΔG(gas→water;W→M) and ΔG(octanol,ethanol→water;-W→M).[26−28] ΔG(D;W→M) and the hydration free energy difference ΔG(gas→water;W→M) have close values and slightly increase as a function of the residue size (Table 1). The increments are very small: 0.34 for ΔG(gas→water; W→M) and 0.7 for ΔG(D;W→M).

To clarify the origin of this small difference of ΔG(D;W→M) among the amino acids, we divided ΔG(D;W→M) into enthalpy and entropy. We found that the small difference was caused by the compensation between enthalpy (ΔU) and entropy ($-T\Delta S$), i.e., 0.47 kcal/mol and −0.72 for Ala, −4.40 and 4.50 for Ile, and −2.70 and 2.97 for Leu. This means that water molecules surrounding a hydrophobic side chain tend to tolerate the difference in the residue size by flexibly rearranging the configurations. The enthalpy−entropy compensation effect was clearly demonstrated for small molecules in solution.[35,36] The difference in ΔG(D;W→M) is slightly larger than that in ΔG(gas→water; W→M), probably because some water molecules around the mutation site were replaced by the four adjacent amino acids of the model peptide in the denatured state whereas ΔG(gas→alcohol; W→M) shows the difference between two single amino acids.

In contrast, ΔG(gas→alcohol; W→M) and ΔG(N;W→M) approximately decrease as a function of the residue size. This means that the hydrophobic side chains are favorable in the alcohol solution and also the core region of protein, as we usually expected. However, the variation range is only 1.1 kcal/mol for ΔG(gas→alcohol; W→M) and in contrast is very large 6.65 kcal/mol for ΔG(N;W→M). We divided ΔG(N;W→M) into enthalpy and entropy and found that the large differences in ΔG(N;W→M) were caused by the small compensation effect between enthalpy and entropy, i.e., 6.91 kcal/mol and −4.59 for Ala, −4.09 and 4.31 for Ile, and −4.63 and 0.48 for Leu. These small compensations probably originate from the fact that the protein residues surrounding the mutation site are too rigid to relieve the change in the residue size but rather remain the difference as a cavity.

In this study, the amino acid residues surrounding the mutation site were common and gave the same environmental condition for all the mutation. The surrounding residues may change ΔG(N;W→M) and ΔG(D;W→M), even if the mutation W→M is performed between the same W and M. For example, the same mutation Val→Ile causes different values of ΔG(N;W→M) and ΔG(D;W→M) depending on the kind of protein and the location of Val in a protein.[30]

There may be cases where some secondary structures remain in the thermally denatured state. Recently, Sugita et al.[30,31] calculated the free energy changes for four Ile→Val mutations of Human lysozyme (Ile23→Val, Ile56→Val, Ile89→Val, and Ile106→Val) by using native-like structures as well as extended peptides for the models of denatured state. They showed that the native-like model for the denatured state gave closer values to the experimental data[32] for the three mutations except for Ile23→Val than the extended model. However, the differences in $\Delta\Delta G$ between the models are smaller than the calculation error for the Ile89→Val and Ile106→Val mutations. The native-like model for Ile23→Val gave the different $\Delta\Delta G$ value from the experimental values. On the other hand, their free energy calculations for chymotrypsin inhibitor[33] using the extended peptide model showed an agreement with an experimental value. Several years ago, Saito and Tanimura calculated $\Delta\Delta G$ for three mutants of RNaseHI using the extended models and showed a good agreement with the experimental data.[16,17]

We preliminary calculated ΔG(D;W→M) for two mutations Leu→Val and Val→Ala of Myb using the native-like structure (i.e., α-helix consisting of 19 amino acids) for the denatured state instead of the extended peptide. We obtained almost the same ΔG(D;W→M) value for Leu→Val (4.80 kcal/mol vs 4.66) but a different value for Val→Ala (18.13 vs 19.28), which caused a disagreement with the experimental $\Delta\Delta G$ value. These results support the extended peptide model for the denatured state but not the native-like model, since the extended model for the denatured state is the only assumption in the present calculation study. This assumption is also supported by the CD spectra that show the absence of an α-helix in the denatured state (Figure 3).

All the mutations discussed above were performed for the hydrophobic amino acids with smaller or equal sizes of side chains than/to that of the cavity in the wild type. In addition to these mutants, we performed calculations for two mutations Leu→Met and Val→Thr, although experimental measurements for the mutants are yet to be done. These mutations have the following different features from the above mutations. The Met residue has a longer side chain than the Leu residue penetrating through the original cavity wall observed in the wild type. The Val→Thr mutation introduces an O−H group into the cavity region. We found that the Met mutant was less stable (−3.58 kcal/mol) than the Leu mutant but slightly more stable than the wild type ($\Delta\Delta G$(N→D;W→M) = 0.87). On the other hand, the Thr mutant was 1.83 kcal/mol more stable than the wild type.

We found the following stability mechanisms which are different from the cavity-filling stability mechanism. The Val→Thr mutation changed the C_γ methyl group to the O−H group, which formed a hydrogen bond (distance is 2.74 ± 0.13 Å) with the oxygen atom of the Glu99 main chain. This hydrogen bond stabilized the Thr mutant, although the Thr mutant had almost the same size of cavity as the wild type (Val). On the other hand, the longer side chain of the Met mutant pushed away the cavity wall, especially Trp115, Cys130, and

Trp134. Furthermore, the Met mutant had a lower free energy in the denatured state than the Val and Leu mutants, because of the large hydration free energy of the Met residue. Therefore, the stability of these kinds of mutations, i.e., introducing an O−H group and a long S−CH_3 side chain inside the cavity region, cannot be simply explained by the remaining cavity volume. The calculation results for Met and Thr mutations will be experimentally validated in near future.

5. Conclusion

In general, the thermodynamic stability of mutant proteins is determined by the delicate balance between the free energy changes due to mutation for the native and denatured states, ΔG(N;W→M) and ΔG(D;W→M). These free energy changes are affected by the two factors, i.e., the amino acid at the mutation site and the surrounding amino acids. In this study for Myb R2 domain, we varied only the former factor and fixed the latter factor by substituting an amino acid at the single site to various natural and nonnatural amino acids. Furthermore, the restriction of the substitution to amino acids with alkyl side chains simplified the interactions of the amino acid at the mutation site with the surrounding residues to the vdW interaction. Then, we found the following stabilization mechanisms for the Myb R2 domain. The thermodynamic stability of cavity-filling Myb mutants is controlled by the free energy of the native state but not the denatured state. The free energy of the denatured state negligibly changes because of the compensation effect between enthalpy and entropy. In contrast, the free energy of the native state changes largely depending on the cavity volume around the mutation site. The decrease of the cavity volume decreases the free energy of the native state mainly through the vdW interactions with the surrounding residues. We expect that this study will become a starting point for investigating more general cases of cavity filling stability, in which mutations are performed at various sites of a protein and to various amino acids with hydrophilic side chains.

Supporting Information Available: Supporting information about the calculation method can be obtained free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Fersht, A. R.; Serrano, L. *Curr. Opin. Struct. Biol.* **1993**, *3*, 75−83.
(2) Matthews, B. W. *Adv. Protein Chem.* **1995**, *46*, 249−78.
(3) Karpusas, M.; Baase, W. A.; Matsumura, M.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8237−41.
(4) Mendel, D.; Ellman, J. A.; Chang, Z.; Veenstra, D. L.; Kollman, P. A.; Schultz, P. G. *Science* **1992**, *256*, 1798−802.
(5) Ishikawa, K.; Nakamura, H.; Morikawa, K.; Kanaya, S. *Biochemistry* **1993**, *32*, 6171−6178.
(6) Eijsink, V. G.; Dijkstra, B. W.; Vriend, G.; van der Zee, J.; Veltman, O. R.; van der Vinne, B.; van den Burg, B.; Kempe, S.; Venema, G. *Protein Eng.* **1992**, *5*, 421−426.
(7) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1−64.
(8) Chothia, C. *Nature* **1974**, *248*, 338−9.
(9) Chothia, C. *Nature* **1975**, *254*, 304−8.
(10) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−76.
(11) Sarai, A.; Uedaira, H.; Morii, H.; Yasukawa, T.; Ogata, K.; Nishimura, Y.; Ishii, S. *Biochemistry* **1993**, *32*, 7759−64.
(12) Ogata, K.; Morikawa, S.; Nakamura, H.; Hojo, H.; Yoshimura, S.; Zhang, R.; Aimoto, S.; Ametani, Y.; Hirata, Z.; Sarai, A. *Nat. Struct. Biol.* **1995**, *2*, 309−20.
(13) Ogata, K.; Kanei-Ishii, C.; Sasaki, M.; Hatanaka, H.; Nagadoi, A.; Enari, M.; Nakamura, H.; Nishimura, Y.; Ishii, S.; Sarai, A. *Nat. Struct. Biol.* **1996**, *3*, 178−87.
(14) Saito, M. *Mol. Simulation* **1992**, *8*, 321−333.
(15) Saito, M. *J. Chem. Phys.* **1994**, *101*, 4055−4061.
(16) Saito, M.; Tanimura, R. *Chem. Phys. Lett.* **1995**, *236*, 156−161.
(17) Tanimura, R. and Saito, M. *Mol. Simulation* **1996**, *16*, 75−85.
(18) Saito, M. *J. Phys. Chem.* **1995**, *99*, 17043−17048.
(19) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266.
(20) Tahirov, T. H.; Morii, H.; Uedaira, H.; Sarai, A.; Ogata, K. *Acta Crystallogr. D* **1999**, *55*, 1345−1347.
(21) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K. J.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(22) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; p 331.
(23) Saito M.; Nakamura H. *J. Comput. Chem.* **1990**, *11*, 76−81.
(24) Connolly, M. L. *Science* **1983**, *221*, 709−713
(25) Eriksson, A. E.; Baase, W. A.; Matthews, B. W. *J. Mol. Biol* **1993**, *229*, 747−69.
(26) Guy, H. R. *Biophys. J.* **1985**, *47*, 61−70.
(27) Fauchere, J.-L.; Pliska, V. *Eur. J. Med. Chem.* **1983**, *18*, 369−375.
(28) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1971**, *246*, 2211−2217.
(29) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849−855.
(30) Sugita, Y.; Kitao, A.; Go, N. *Fold. Des.* **1998**, *3*, 173−81.
(31) Sugita, Y.; Kitao, A. *Proteins* **1998**, *30*, 388−400.
(32) Takano, K.; Ogasahara, K.; Kaneda, H.; Yamagata, Y.; Fujii, S.; Kanaya, E.; Kikuchi, M.; Oobatake, M.; Yutani, K. *J. Mol. Biol.* **1995**, *254*, 62−76.
(33) Sugita, Y.; Kitao, A. *Biophys. J.* **1998**, *75*, 2178−87.
(34) Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins* **1991**, *11*, 281−96.
(35) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531−567.
(36) Kubo, M. M.; Gallicchio, E.; Levy, R. M. *J. Phys. Chem. B* **1997**, *101*, 10527−10534.
(37) Morii, H.; Uedaira, H.; Ogata, K.; Ishii, S.; Sarai, A. *J. Mol. Biol.* **1999**, 292, 909−920.
(38) Saito, M. In *Structure and function of biological systems under extreme conditions*; Taniguchi, Y., Stanley, H. E., Ludwig, H., Eds.; Springer-Verlag: Berlin, 2000.