京都府立大学博士論文

# Structural dynamics of c-Myb DNA-binding domain and its correlation with function

（**c-Myb DNA** 結合ドメインの動的構造と機能との相関）

**March, 2016**

**Satomi Inaba**

# Table of contents

## Abbreviations

| | |
|---|---|
| BMRB | biological magnetic resonance bank |
| CD | circular dichroism |
| CPMG | Carr–Purcell–Meiboom–Gill |
| CLEANEX-PM | Phase-Modulated CLEAN chemical EXchange |
| c-Myb | *c-myb* protoongene product |
| DBD | DNA-binding domain |
| DSC | differential scanning calorimetry |
| HSQC | heteronuclear single quantum correlation |
| IDP | intrinsically disordered protein |
| ITC | isothermal titration calorimetry |
| $K_a$ | equilibrium association constant |
| $k_{int}$ | intrinsic hydrogen exchange rate |
| $k_{off}$ | dissociation rate constant |
| $k_{on}$ | association rate constant |
| $k_{ex}$ | exchange rate |
| KPB | potassium phosphate buffer |
| MBS-I | DNA fragment containing the Myb-binding DNA sequence |
| NRD | negative regulation domain |
| $n$ | stoichiometry of binding |
| $R_{transverse}$ | $^{15}N$ transverse relaxation rate |
| PBS | phosphate buffered saline |
| PDB | protein data bank |

| | |
|---|---|
| SPR | surface plasmon resonance |
| TAD | trans-activation domain |
| TFE | trifluoroethanol |
| $T_d$ | denaturation temperature |
| $T_m$ | transition temperature |
| $\tau_m$ | mixing time in CLEANEX-PM experiments |
| $\Delta C_p$ | heat capacity change |
| $\Delta G$ | Gibbs free energy change |
| $\Delta H_{cal}$ | calorimetric enthalpy change |
| $\Delta H$ | enthalpy change |
| $\Delta H_{vH}$ | van't Hoff enthalpy change |
| $\Delta S$ | entropy change |

## Chapter 1 General introduction

A great number of protein structures have been determined at atomic resolution using the techniques such as X-ray and nuclear magnetic resonance (NMR) spectroscopies (http://pdbj.org/info/statistics). However, most are static or time-averaged structures and are lack information about the structural dynamics. A protein in solution generally equilibrates among multiple conformational sub-states that are critical for the expression of function (Haliloglu and Bahar, 2015). Thus, analysis of the structural dynamics of protein is important for understanding its function, as discussed below in more detail. In order to understand the structural dynamics of proteins under the physiological conditions, it is necessary to detect the protein structure world from the basic folded structures to those beyond them. Apart from the rapid non-biological fluctuations of atoms that are anticipated for any molecules in solution, proteins are also endowed with slow conformational fluctuations involving concerted motions of their atoms (Karplus, 1984; Dyson and Wright, 1996). The slow fluctuations or rare events in protein conformational dynamics are much less studied but could be decisively important in protein function, because they are more likely to be evolutionarily designed as “biological fluctuations” (Akasaka and Matsuki ed., 2015). Therefore, dynamics have been implicated in many aspects of function, including molecular recognition, signaling, folding, enzyme activities, and functional regulation (Henzler-Wildman *et al*., 2007). Many of the structural changes are involve conversion to low-populated and transient conformers, which are extremely difficult to study using conventional biophysical analysis. Recently, some time-resolution experiments have been developed to observe excited, transient, intermediate, and unfolded states.

However, these have limits for obtaining information about "biological fluctuation" which should be analyzed using multiple approaches in biophysical methods. Some researchers have considered Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR to be powerful tool for revealing low-populated state (including low-lying excited state), excited state, high-energy conformer, and transient conformers, which are closely correlated with protein function (Sugase *et al.*, 2007; Bouvignies *et al*., 2011; Tzeng and Kaldomos, 2012; Gagne *et al*., 2012). This method can detect millisecond time-scale conformational transitions of proteins at atomic resolution, which are regarded as a conformational exchange for biological events (Korzohnev and Kay, 2008; Ishima, 2012). As another method, variable-pressure experiments using on-line cells have been developed and can provide residue-specific and/or atom-based information on conformers beyond the basic folded structure, termed "high-energy conformers" (Akasaka and Yamada, 2001; Peterson and Wand, 2005). On the other hand, thermodynamic studies have traditionally been used for analysis of the stability, folding, interaction, and dynamics of several proteins (Sturtevant, 1977; Privalov, 1979; Murphy and Freire, 1992; Perozzo *et al*., 2004). The thermal and calorimetric analyses can give the thermodynamic parameters directly, such as Gibb's free energy ($\Delta G$), enthalpy ($\Delta H$), entropy ($\Delta S$), and heat capacity ($\Delta C_p$) changes for individual events. There are many reports about molecular interactions between a protein and its target molecules (viz., another protein, DNA, carbohydrate and synthetic ligand) using calorimetric methods, providing the thermodynamics properties which are useful information such as protein *de novo* design, engineering, and drug design (Wiseman *et al*., 1989; Sagawa *et al*., 2003; Oda *et al*., 2007; Chaires, 2008). In particular, the entropy term would include the contribution derived from structural dynamics and biological fluctuation.

The *c-myb* gene is the cellular homolog of the viral oncogene *v-myb* carried by the chicken avian leukemia viruses E26 and avian myeloblastosis virus (AMV), and is involved in the transformation of hemopoietic cells (Ness, 1999; Prouse and Campbell, 2012). The *c-myb* protoongene product protein (c-Myb), which is well-known as a transcriptional factor, has three functional domains reading from the N-terminus: DNA-binding domain (DBD), a transcriptional activation domain (TAD), and a negative regulation domain (NRD) (Gonda *et al*., 1985; Sakura *et al*., 1989, Fig. 1-1A). The components of the v-Myb domain are almost similar to those of c-Myb, but parts of the N- and C- terminus domains are deleted or mutated in the v-Myb protein (Weston, 1992).

The DBD consists of three imperfect tandem repeats, denoted as R1, R2 and R3, for the N-terminus. The Fig. 1-1B shows the complex structure of R2R3 and DNA-containing consensus sequence, as determined by the NMR-distance method (Ogata *et al*., 1994; PDB code: 1MSE). Although the amino acid sequences of the three repeats are quite different (Fig. 1-1C), three conserved Trp residues in each repeat participate in a hydrophobic core (Kanei-Ishii *et al*., 1990). Each repeat has a very similar folding architecture containing three well-defined helices (Ogata *et al*., 1995). Portions of the second and third helices in each repeat form a variant of the helix-turn-helix motif. The last two repeats, R2 and R3, are necessary for specific binding to the DNA containing consensus sequence, $P_yAAC^G/_TG$, where $P_y$ indicates a pyrimidine (Biedenkapp *et al*., 1988). R1 is not essential for the specific DNA recognition of c-Myb, and is considered to enhance the stability of the R2R3 complex with DNA (Tanikawa *et al*., 1993). In addition, crystal structures of the complexes have shown that it is comprised of the c-Myb or AMV v-Myb DBD, the C/EBPβ DBD, and

the promoter DNA fragment, indicating that the target DNA recognition mechanism and the cooperativity of v-Myb are different from that of c-Myb (Tahirov *et al.*, 2002).

Some of the structural features for c-Myb DBD have been revealed by previous studies. One is that the stability of R2 is lower than those of R1 and R3, because only R2 has a large internal cavity in the hydrophobic core (Sarai *et al*., 1993), in which Val-103 is located. The cavity-filling mutation, V103L, stabilizes the R2 structure and reduces transcriptional activation, indicating that the conformational fluctuation of R2 induced by the cavity would be critical for the specific DNA-binding function (Ogata *et al.*, 1996; Morii *et al*., 1999). On the other hand, in R3, the residue corresponding to R2's Val-103 is Ile-155, which also has unique features. A previous NMR study using isolated R3 with two of the buried Ile residues (Ile-155 and Ile-181) substituted to Leu, I155L/I181L, revealed multiple conformations, as monitored by two-dimensional $^{15}N/^{1}H$-HMQC and three-dimensional HNCA spectoscopies (Furukawa *et al*., 1996). The results implied that the side-chain conformational entropy would be increased upon the mutation. Although R2 and R3 have in individual sequences and structural properties, both R2 and R3 bind to DNA in a cooperative recognition manner (Ogata *et al*., 1996). Both repeats are connected by a short linker, which is composed of Asn-139, Pro-140, and Glu-141. Previous studies have shown that this structural uniqueness of the linker region is also important for the DNA-binding function (Hegvold and Gabrielsen, 1996; Oda *et al*., 1998). Similarly, mutation analysis of the c-Myb DBD also supports the notion that unique conformational fluctuations are critical for the DNA-binding function. Therefore, correlation of the structural dynamics of the c-Myb DBD with its function would be critical for understanding this important for life science phenomenon.

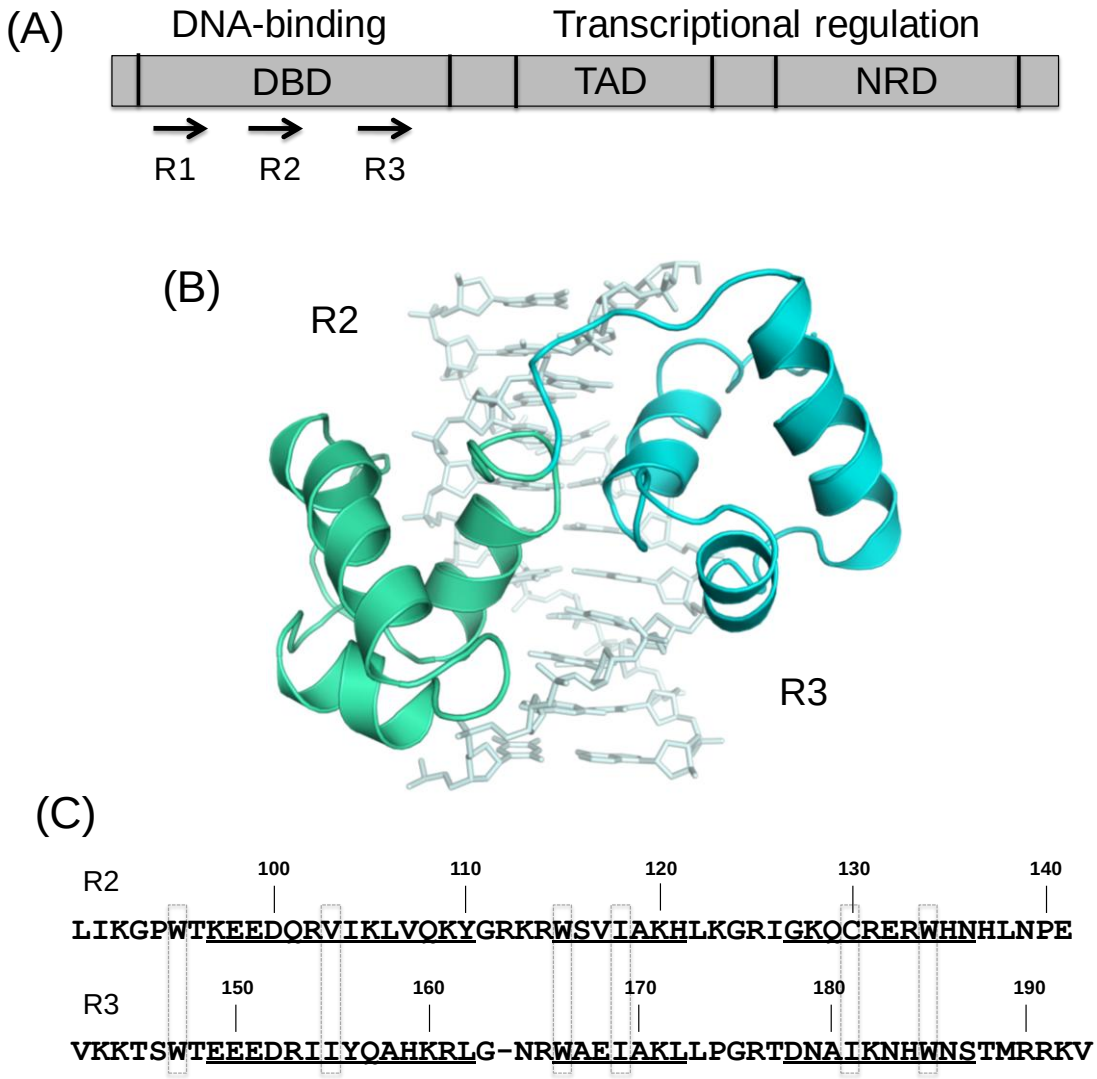



Fig. 1-1. (A) The three functional domains of c-Myb are shown schematically. From N-terminus, DNA-binding (DBD) containing three imperfect tandem repeats, transcriptional activation (TAD), and negative regulation (NRD) domains. (B) The complex structure of c-Myb R2R3 with DNA containing consensus sequences determined by solution NMR (Ogata *et al.*, 1994; PDB code: 1MSE). R2 and R3 are colored in green and blue, respectively. (C) Amino acid sequences of R2 and R3 are aligned with one-letter amino acid codes. The underlines indicate the helical regions of each repeat. The residues involved in the hydrophobic core are shown by gray boxes.

In this study, I have focused on “biological fluctuation” and characterized the structural dynamics of the c-Myb DBD, which closely correlate with its function. In Chapter 2, the functional conformer of the c-Myb R2R3 was analyzed by temperature perturbation experiments and its dynamics behavior under the physiological temperature was discussed. In Chapter 3, the folding thermodynamics of c-Myb R2R3 were determined from its pH-dependent thermal stability measurements, and correlated with its structural information. In Chapters 4 and 5, the dynamics of some mutants that showed unique structural features in the previous studies were analyzed using the functional unit of R2R3 and were quantified as entropy terms. In Chapter 6, the results are summarized and the correlation of structural dynamics or biological fluctuation with function, and biological evolutionary strategies are discussed.

Prior to the current analyses, Cys-130 in R2, which is the only cysteine residue in the c-Myb R2R3 and is located at a position equivalent to an isoleucine in R3, was replaced with Ile, to facilitate protein purification and characterization. It has been shown that the affinity and the specificity of mutant C130I are similar to those of the wild-type (Myset *et al*., 1993; Oda *et al*., 1997a). Therefore, the C130I protein was used as the standard R2R3 which is denoted as R2R3* in this study.

## Chapter 2 Functional conformer of c-Myb R2R3 revealed from temperature-dependent studies

### Introduction

A comparison of the solution structures of R2R3 in its free and specific DNA complexed forms determined by NMR spectroscopy showed that, upon binding to DNA, the indole ring of Trp95 slides into the cavity mentioned above (Ogata *et al*., 1996). This cavity is therefore considered to be crucial for the DNA binding and stability of the complex. One unique structural feature of the protein is the presence of a cavity of ~33 $Å^3$ in size located in the hydrophobic core of R2 (Kono *et al*., 2000). In fact, the cavity-filled mutation in the R2 hydrophobic core, created by mutating Val103 to Leu, was shown to decrease the DNA-binding affinity, while the stability of R2 was increased (Ogata *et al*., 1996; Morii *et al*., 1999; Lassalle *et al*., 2001). The role of the cavity in the conformational flexibility of c-Myb R2R3 thus appears to be crucially important for its DNA binding and its biological activity. To correlate these findings with the function of mouse c-Myb R2R3 *in vivo*, the conformational flexibility must be elucidated at physiological temperature, namely 37ºC.

The main purpose of this study was to assess the dynamic conformational state of the minimal functional DBD of c-Myb, R2R3, under conditions closely resembling physiological temperature and pH. Such an assessment is crucially important to elucidating the functional strategy of this group of proteins based on their unique structures consisting of imperfect 51- or 52-residue repeats. A previous NMR analysis of the structural dynamics of the free and DNA-complexed forms of R2R3 was carried out at 17ºC (Ogata *et al*., 1996). The thermodynamic stability of R2R3, including the differential stability of the two repeats, has been amply discussed based on the results of

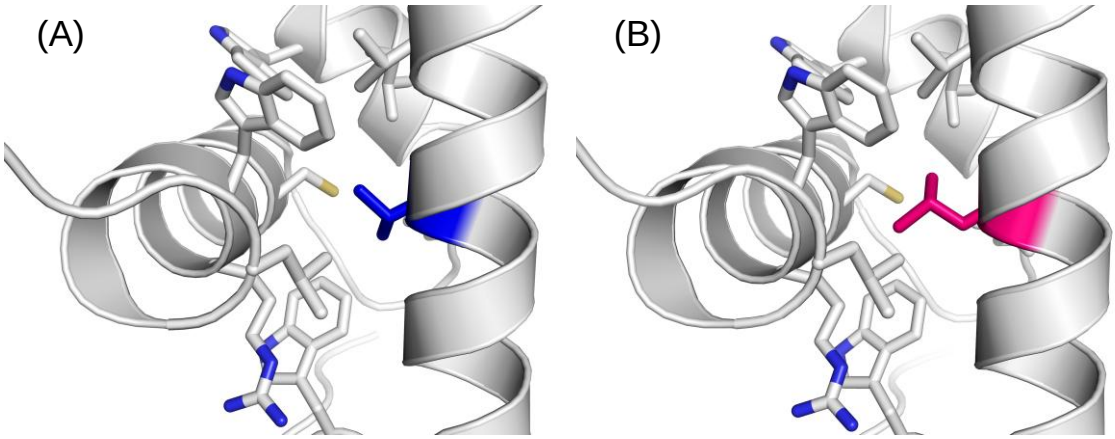



Fig. 2-1. The three-dimension structure of isolated R2 (A) and R2 V103L mutant (B) determined by X-ray crystallography (Tahirov *et al*., 2002, PDB code; 1GV5 and 1GVD, wild-type and V103L mutant, respectively). Val103 and Leu 103 are colored in blue and pink, respectively.

differential scanning calorimetry (DSC) and circular dichroism (CD) analyses (Ogata *et al*., 1996; Morii *et al*., 1999; Oda *et al*., 1998). However, the conformational fluctuations of R2R3 under physiological conditions have not been assessed. Such site-specific structural analyses at atomic resolution for site-specific detail can only be achieved using NMR spectroscopy.

For this NMR analysis, the conformational fluctuations of R2R3* and its cavity-filled mutant V103L were assessed at pH 7.4 as a function of temperature (5 - 70ºC), using various 1D and 2D NMR measurements, as well as CD for global transition analysis. The conformational fluctuations of R2R3 were found to be exceptionally sensitive to temperature variations, showing profound fluctuations under physiological conditions. The focus of this study was on the role of the cavity and the differential fluctuation between the two repeats. The dynamic features of the c-Myb

R2R3 protein were found to be somewhat unique. I furthermore discuss in this report how these features may be used in DNA binding.

## Materials and Methods

*Protein expression and purification*

A DNA fragment encompassing the second and third repeats in the DBD of c-Myb was amplified by polymerase chain reaction (PCR), using R2R3 as the template and the primer pairs. V103L mutant plasmid was obtained in site-directed mutagenesis. The conditions of PCR reaction were as fellows: 94˚C for 2 min for 1 cycle, 94˚C for 15 sec, 59˚C for 30 sec, and 68˚C for 30 sec for 30 cycles. R2R3* and V103L were over-expressed in *E. coli* and purified. The expression and purification methods of these R2R3 proteins were described previously (Oda *et al*., 1997b). Briefly, the overexpressed proteins in *Escherichia coli* BL21 (DE3) were purified by using phosphocellulose (P11, Whatman) and gel filtration (Sephacryl S-100, GE healthcare) columns. The purified proteins were concentrated and the buffer was exchanged to phosphate buffered saline (PBS, pH 7.4) or NMR sample buffer using Amicon Ultra-4 (Millipore). The purity of each protein was determined to be about 95% by SDS/PAGE analysis. For the NMR experiment, the $^{13}$C-, $^{15}$N-uniformly labeled R2R3* and V103L were over-expressed in *E. coli* in the M9 minimum medium containing 1.0 g $L^{-1}$ $^{15}NH_4Cl$ and/or 4.0 g $L^{-1}$ $^{13}C$-$U_6$-glucose. The protein concentrations were determined from UV absorption at 280 nm and were calculated by using the molar absorption coefficient of $3.7 \times 10^4$ $M^{-1}$ $cm^{-1}$ (Oda *et al*., 1997a), determined by the amino acid composition analysis.

*CD measurements*

Far-UV CD spectra were recorded on a Jasco J-820 spectropolarimeter at 20˚C equipped with Peltier-type temperature control system. The measurements were carried out for the 0.02 mg $ml^{-1}$ proteins in PBS (pH 7.4) using quartz cell with 1.0 cm path-length. CD spectra were obtained using scanning speed of 20 nm $min^{-1}$, a time response of 1 s, a bandwidth of 1 nm, and an average over 4 scans. The melting curves were recorded in variable temperature mode at 222 nm, from 10 to 80ºC with a heating rate of 1.0ºC $min^{-1}$. The analysis of the transition curves obtained by temperature-scanning CD measurements was performed on the basis of two-state transition model, as described previously (Morii *et al*., 1999).

The molar ellipticity data of temperature-scanning CD ($Y$) were fitted with the weighted-mean combination of two linear functions for the folded and the unfolded states, that is,

$$y_n = A_n (T - T_m) + B_n \qquad \text{(Eq. 2-1)}$$

and

$$y_d = A_d (T - T_m) + B_d, \qquad \text{(Eq. 2-2)}$$

from Eq. (1) and (2), as shown below:

$$Y = y_n (1 - f_D) + y_d f_D \qquad \text{(Eq. 2-3)}$$

Here, $f_D$ represents the molar fraction of the unfolded state and be expressed with the function of $\Delta G_{01}$, the free energy of the transition:

$$f_D = \exp(-\Delta G_{01} / RT) / \{1 + \exp(-\Delta G_{01} / RT)\} \qquad \text{(Eq. 2-4)}$$

$\Delta G_{01}$ is also the function of the temperature involving three parameter, which are heat capacity change ($\Delta C_p$), transition temperature ($T_m$), and van't Hoff enthalpy change ($\Delta H_{vH}$):

$$\Delta G_{01} = \Delta C_p T \ln(T / T_m) + \{\Delta C_p - (\Delta H_{vH} / T_m)\} (T - T_m) \qquad \text{(Eq. 2-5)}$$

Substituting Eqs. (4) and (5) into (3), $Y$ results in the function containing six parameters, that is $A_n$, $A_d$, $B_n$, $B_d$, $T_m$ and $\Delta H_{vH}$, to be determined. The curve fitting was carried out by the non-linear least-square method on Origin 5.0 software.

*SPR measurements*

The DNA-binding of R2R3* and V103L was analyzed using Biacore 2000 (GE healthcare), as described previously (Oda *et al*., 1999). Briefly, DNA fragment containing the Myb-binding DNA sequence (MBS-I) with one strand biotinylated at the 5´-end was immobilized on the sensor-chip, SA (GE healthcare), through streptavidin pre-immobilized, and various concentrations of R2R3* and V103L were applied to the sensor chip surface at a rate of 20 μl $min^{-1}$ during 3 min. Data were analyzed using the BIAevaluation program 3.2 to determine the kinetic rate constants, associate rate constant ($k_{on}$) and dissociation rate constant ($k_{off}$). The equilibrium association constant ($K_a$) values were calculated from the two rate constants as $K_a = k_{on}/k_{off}$.

*NMR measurements*

The $^{1}H$ one-dimensional and $^{15}N/^{1}H$ two-dimensional heteronuclear single quantum correlation (HSQC) NMR experiments were performed between 5 and 70ºC on Bruker AVANCE III 600 spectrometer operating at a proton resonance frequency of 600 MHz. The $^{1}H$ one-dimensional measurements were carried out with the 0.45 mM protein in 20 mM potassium phosphate buffer containing 20 mM NaCl, 1 mM $NaN_3$, and 10% $^{2}H_2O$ (pH 7.4), and the $^{15}N/^{1}H$ two-dimensional HSQC measurements were carried out with the 1 mM protein in 20 mM Tris-HCl buffer containing 20 mM NaCl, 1 mM $NaN_3$, and 10% $^{2}H_2O$ (pH 7.4). The $^{13}C/^{1}H$ two-dimensional HSQC measurements at the natural abundance of $^{13}C$ were carried out under the same solution conditions of

the $^{15}N/^{1}H$ HSQC measurement. In addition to the information reported previously (Ogata *et al*., 1994), the $^{15}N/^{1}H$ cross peaks of R2R3* in the present experimental conditions were assigned using $^{15}N/^{13}C$ labeled samples from three-dimensional spectra, including HNCO, HN(CA)CO, HNCACB, and CBCA(CO)NH on Bruker AVANCE III 950 spectrometer at a proton resonance frequency of 950.33 MHz equipped with a z-axis gradient and triple-resonance TCI cryogenic probe. In addition to these information, the $^{15}N/^{1}H$ cross peaks of V103L were assigned using HNCA spectrum. The measurements for the $R_{transverse}$ values of the $^{15}N$ nuclei of backbone amides were carried out with the pulse sequence described by Farrow *et al*. using a AVANCE III 600 spectrometer equipped with a z-axis gradient and triple-resonance TXI probe (Farrow *et al*., 1994). The experiment includes a series of 10 experiments with transverse decay times ranging from 17.6 ms to 176 ms. The data were processed and analyzed using NMRpipe (Delaglio *et al*., 1995) and Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Chemical shifts and peak lists of R2R3* and V103L have been deposited in the Biological Magnetic Resonance Bank (www.bmrb.wisc.edu) as entries 11584 and 11585, respectively.

$^{15}N$ $R_2$ relaxation dispersion was measured on Bruker a AVANCE III 600 spectrometer equipped with a z-axis gradient and triple-resonance TXI probe at 25 °C using relaxation-compensated constant time Carr–Purcell–Meiboom–Gill (rcCPMG) pulse sequences with a constant relaxation time of $T_{CPMG}$ = 20 ms (Tollinger *et al*., 2001). $R_2$ dispersion profiles were generated by measuring effective transverse relaxation rates ($R_{2,eff}$) as a function of $1/\tau_{CP}$ (= $\nu_{CP}$), where $\tau_{CP}$ is the duration time for two successive CPMG pulse train units. Two-dimensional data sets with 1,024 × 64 ($t_2$ × $t_1$) complex points were acquired at $\tau_{CP}$ = 10, 5, 3.33, 2.0, 1.43, 1.0, 0.66, 0.50, and

0.4 ms. Reference spectra were acquired by omitting the CPMG period. Each spectrum was collected with 64 scans and the same measurements were repeated 3 times. The $R_2$ dispersion profiles of non-overlapping peaks were globally fitted to the equation derived for a two-state conformational exchange model (A ⇄ B) under the fast conformational exchange regime (Loria *et al*., 1999; Ishima, 2012):

$$R_{2,\mathrm{eff}} = R_{2,0} + R_{\mathrm{ex}} = R_{2,0} + \Phi_{\mathrm{ex}} [1 - (4 / k_{\mathrm{ex}}\tau_{\mathrm{CP}}) \tanh (k_{\mathrm{ex}}\tau_{\mathrm{CP}} / 4)] / k_{\mathrm{ex}} \quad \text{(Eq. 2-6)}$$

where $R_{2,0}$ is the population-average intrinsic relaxation rate and $k_{\mathrm{ex}}$ (= $k_{\mathrm{AB}}$ + $k_{\mathrm{BA}}$) is the rate of exchange between the states. $\Phi_{\mathrm{ex}}$ is a scaling factor and expressed as $p_{\mathrm{A}}p_{\mathrm{B}}\Delta\omega^2$, where $p_{\mathrm{A}}$ and $p_{\mathrm{B}}$ are the fractional populations of the A and B states, respectively, and $\Delta\omega$ is the chemical shift difference between these two states. I fitted the data on R2R3* to Eq. (2-6) with $k_{\mathrm{ex}}$ as a global fitting parameter, $R_{2,0}$ and $\Delta\omega$ as individual fitting parameters, and fixed $p_{\mathrm{A}}$ and $p_{\mathrm{B}}$ values according to the assumed values in the line shape simulation mentioned below. In the case of V103L, as there was no information about $p_{\mathrm{A}}$ and $p_{\mathrm{B}}$, the data of individual residues were fitted with $k_{\mathrm{ex}}$, $\Phi_{\mathrm{ex}}$ and $R_{2,0}$ as fitting parameters. Then, the $R_{\mathrm{ex}}$ value for each residue was obtained from the relation $R_{\mathrm{ex}} = \Phi_{\mathrm{ex}} / k_{\mathrm{ex}}$. The significances of the $R_2$ relaxation dispersion for each residue were judged by whether the largest deviation of $R_{2,\mathrm{eff}}$ superiors to all of the error bars for individual $R_{2,\mathrm{eff}}$ points.

*Line shape analysis of $^1$H one-dimensional NMR spectra*

The obtained $^1$H NMR spectra were analyzed with a multi-Lorentzian function. The function used for the analysis is as follows;

$$S(\delta) = \sum_i h_i \frac{w_i^2}{(\delta - \delta_i)^2 + w_i^2} + baseline$$

where $h_i$, $w_i$ and $\delta_i$ are peak height, half-height width and the chemical shift at the peak center for *i*th signal, respectively. A third-order polynomial function was used for the baseline. I fitted the obtained spectra with this function using IgorPro (Wavemetrics). The area of the *i*th signal was calculated from $h_i$ multiplied by $w_i$.

*Validation of the measured NMR proton spectra by using computer simulation*

The simulations of the $^1$H NMR spectra were performed using WINDNMR-Pro (Reich, 1995). The NMR spectra were simulated as a function of the populations of respective states (the basic folded state (N), the intermediary state (I) and the globally unfolded state (U)), and exchange rates between N and I ($k_{ex,NI}$), and I and U ($k_{ex,IU}$), assuming a three-state exchange model (Scheme 2-1). The population of each state ($p_N$, $p_I$ and $p_U$) at each temperature point (5 - 70$^o$C at an intervals of 5$^o$C) was calculated considering the temperature dependence of the equilibrium constant between N and I state ($K_{NI}$) and between N and U state ($K_{NU}$) as follows;

$K_{NI} = \exp\{-\Delta H_{NI} / R\,(1/T - 1/T_{m,NI})\}$

$K_{NU} = \exp\{-\Delta H_{NU} / R\,(1/T - 1/T_{m,NU})\}$

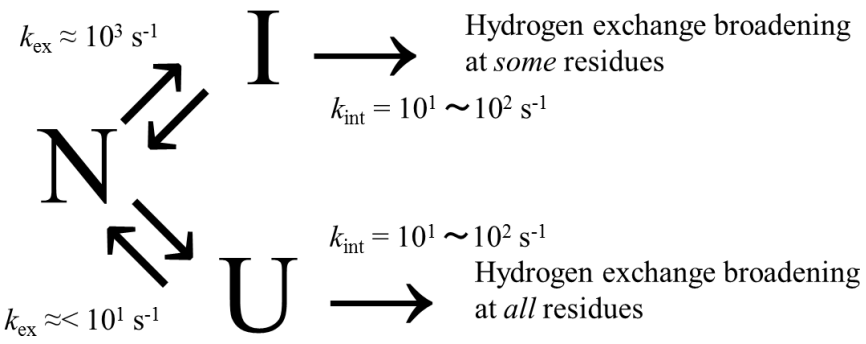
$p_N = 1 / (1 + K_{NI} + K_{NU})$

$p_I = K_{NI} / (1 + K_{NI} + K_{NU})$

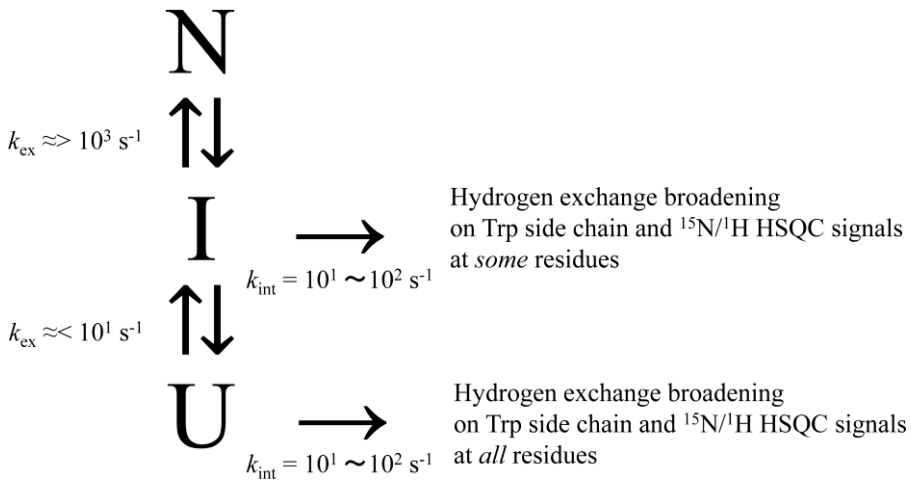
$p_U = K_{NU} / (1 + K_{NI} + K_{NU})$

where $\Delta H$, $R$, $T$, $T_m$ are the enthalpy change, gas constant, temperature and melting temperature, respectively. The chemical shifts for the N state was assumed to be –0.5 ppm (–300 Hz) from the $^1$H spectra at 5$^o$C whereas that of the U state was to be 0.8 ppm (480 Hz) from the typical value for unfolded protein. That for the I state was set as 0.04 ppm (25 Hz) as a results of the manual iterative fittings for a plausible value. In addition, $R_{2,0}$, which is the intrinsic transverse relaxation rate constant, was estimated as 10 s$^{-1}$.

(A)




(B)




Scheme 2-1. Putative thermal unfolding model of c-Myb R2R3. (A) Off-pathway and (B) on-pathway.

## Results

*Secondary structure and thermal denaturation by CD*

R2R3* and V103L were over-expressed in *E. coli* and purified. The purity of each protein was determined to be ~ 95% by SDS/PAGE analysis. The far-UV CD spectra of R2R3* and V103L were similar, and showed typical α-helical structures (Fig. 2-2A). While the far-UV CD spectrum at 85°C showed the denatured structure, that at 20°C after heating reverted to the native structure (Fig. 2-2B). Assuming a two-state transition, native-state (N) ↔ unfolded-state (U), the thermal stabilities of R2R3* and V103L were analyzed by monitoring the CD values at 222 nm (Fig. 2-2C and Table 2-1), and the $T_m$ of R2R3* was found to be 1.7ºC lower than that of V103L. The CD experiments indicate that both proteins exhibit native α-helical structures in the range of 10 to ~35ºC, and transform into denatured structures at temperatures of > 40ºC. These findings are consistent with the results of previous DSC experiments which demonstrated that little change in the heat capacity of R2R3 occurs at temperatures of < 30ºC (Oda *et al.*, 1998). The ratios of R2R3* in the native-state at 20 and 37ºC were calculated to be 99.7% and 90.4%, respectively, while those of V103L at either temperature were 100%.

Table 2-1. Thermodynamic parameters for thermal unfolding of R2R3* and V103L.

| | CD | | $^1$H NMR | |
|---|---|---|---|---|
| | $T_m$ /°C | $\Delta H_{vH}$ /kJ mol$^{-1}$ | $T_m$ /°C | $\Delta H_{vH}$ /kJ mol$^{-1}$ |
| R2R3* | 48.6 ± 0.1 | 161 ± 3 | 47.4 ± 0.2 | 233 ± 10 |
| V103L | 50.3 ± 0.2 | 154 ± 2 | 50.6 ± 0.3 | 191 ± 9 |

The S.D. were based upon fitting errors.

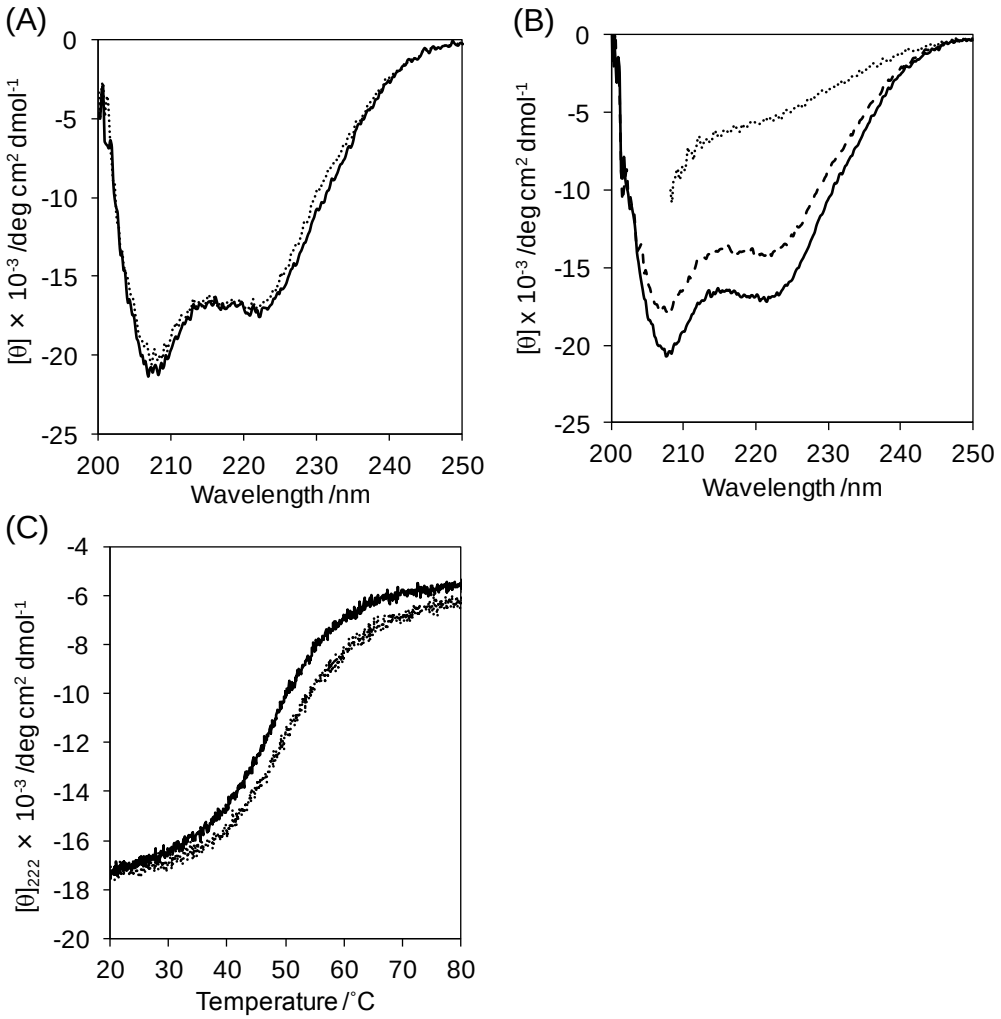



Fig. 2-2. (A) The far-UV spectra of R2R3* (solid line) and V103L (dotted line) at 20ºC. (B) The far-UV CD spectra of R2R3* at 20ºC (solid line), 85ºC (dotted line), and 20ºC after heating (broken line). (C) The thermal denaturation curves of R2R3* (solid line) and V103L R2R3* (dotted line).

*DNA-binding analysis by SPR biosensor*

The DNA-binding abilities of R2R3* and V103L were analyzed and compared using SPR at 20ºC for both proteins in the native state (Table 2-2). The binding affinity of R2R3* to cognate DNA, DNA fragment containing MBS-I, was found to be approximately three times higher than that of V103L, which is in agreement with

previously reported affinities determined using a filter-binding assay (Ogata *et al*., 1996). The decreased $K_a$ of the protein with the cavity-filled mutation can be attributed mainly to the increased $k_{off}$, not to the decreased $k_{on}$.

Table 2-2. Kinetic parameters of interactions of R2R3* and V103L to MBS-I.

| | $k_{on}$ /$M^{-1}$ $s^{-1}$ [a] | $k_{off}$ /$s^{-1}$ [a] | $K_a$ /$M^{-1}$ [b] |
|---|---|---|---|
| R2R3* | 1.73 (± 0.04) × $10^6$ | 1.29 (± 0.21) × $10^{-3}$ | 1.34 × $10^9$ |
| V103L | 1.95 (± 0.02) × $10^6$ | 4.65 (± 0.38) × $10^{-3}$ | 4.20 × $10^8$ |

The S.D. were based upon fitting errors.

[a] The $k_{on}$ and $k_{off}$ are the average of four experiments for protein concentrations from 3.1 to 25 nM.

[b] The $K_a$ are calculated from the two rate constants: $K_a = k_{on}/ k_{off}$.

*Overall $^1$H NMR spectral change with temperature*

Temperature-dependent residue-specific structural properties of R2R3* and V103L were analyzed using NMR. Figs. 2-3A and 3B show the one-dimensional $^1$H NMR spectra for the proteins at 5 - 70ºC. Above 50ºC, fairly sharp signals at 0.8 - 0.9 ppm, typical for side-chain methyl signals of Val, Leu, and Ile residues of unfolded proteins, were observed for both proteins. These signals grew with increasing temperatures, indicating that both proteins undergo thermal transitions to fully unfolded polypeptide chains at higher temperatures. The plot of the combined area of signals at 0.8 - 0.9 ppm showed a significant increase at temperatures of > 40ºC (Fig. 2-3C), indicating that global denaturation only takes place at temperatures of > 40ºC. From the curve fittings, the $T_m$ values for R2R3* and V103L were estimated to be ~48 and ~51ºC, respectively (Table 2-1). These $T_m$ values are comparable to those obtained in the CD measurements mentioned above.

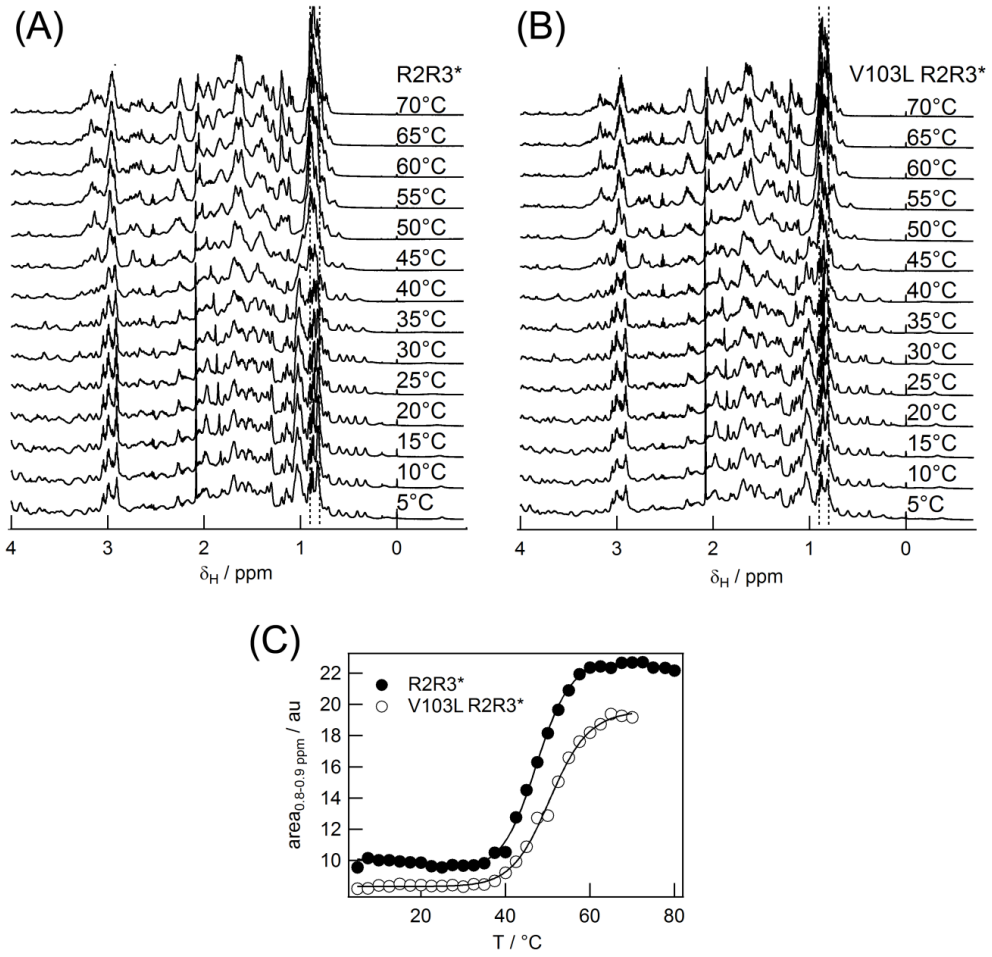



Fig. 2-3. $^{1}$H NMR spectral changes of R2R3* (A) and V103L (B) at temperatures between 5 and 70ºC are shown for the entire spectral region. (C) Plot of the combined area of signals observed at 0.8 - 0.9 ppm. The continuous lines indicate the results of the fitting analyses, which used the same theory as those used in the CD data analysis (Fig. 2-2C). The fittings showed apparent transitions at ~48ºC for R2R3* and ~51ºC for V103L.

*Spectral changes of individual side-chain protons with temperature*

The $^{1}$H NMR spectral changes were examined carefully in this study, with a focus on the behavior of side-chain signals. Fig. 2-4 shows the expansion of the low-field (six Trp rings in addition to others) and high-field (methyl signals of Val, Leu, and Ile in the hydrophobic core) regions of the spectra for R2R3* and V103L. Most

assigned side-chain signals in these regions were well-dispersed, representing their individual microenvironment within the folded structure. The positions and widths of the signals were changed with increases in temperature. I fitted the NMR peak shapes of the six tryptophan and several isolated methyl protons with a multi-Lorentzian function to evaluate the peak widths, positions, and areas at respective temperatures. Figs. 2-5A, 5D, 5G, and 5J (left column) show the signal areas of the proton signals as a function of temperature. The areas of most signals began to decline at around 20ºC and became zero at around 40ºC. These transitions already occurred even at temperature lower than that at which the global unfolding occurs as mentioned above, indicating that there might be another transition distinct from the global unfolding. Figs. 2-5B, 5E, 5H, and 5K (middle column) show the plot of the peak positions of the signals as a function of temperature. Most signals showed slight upfield shifts up to ~40ºC, reflecting the thermal expansion of the native structure. At the higher temperatures, each signal showed respective shifts, indicating the global unfolding. However, it should be noted that some methyl signals from R2R3* showed distinct temperature-dependent peak shifts. For example, the γ1 methyl signal of Val-107 of R2R3* began to show a large shift at ~20ºC (Fig. 2-5B). These observations also indicate that there is a distinct conformational change from the global unfolding and these conformational changes occur at a fast time scale in terms of the NMR time scale.

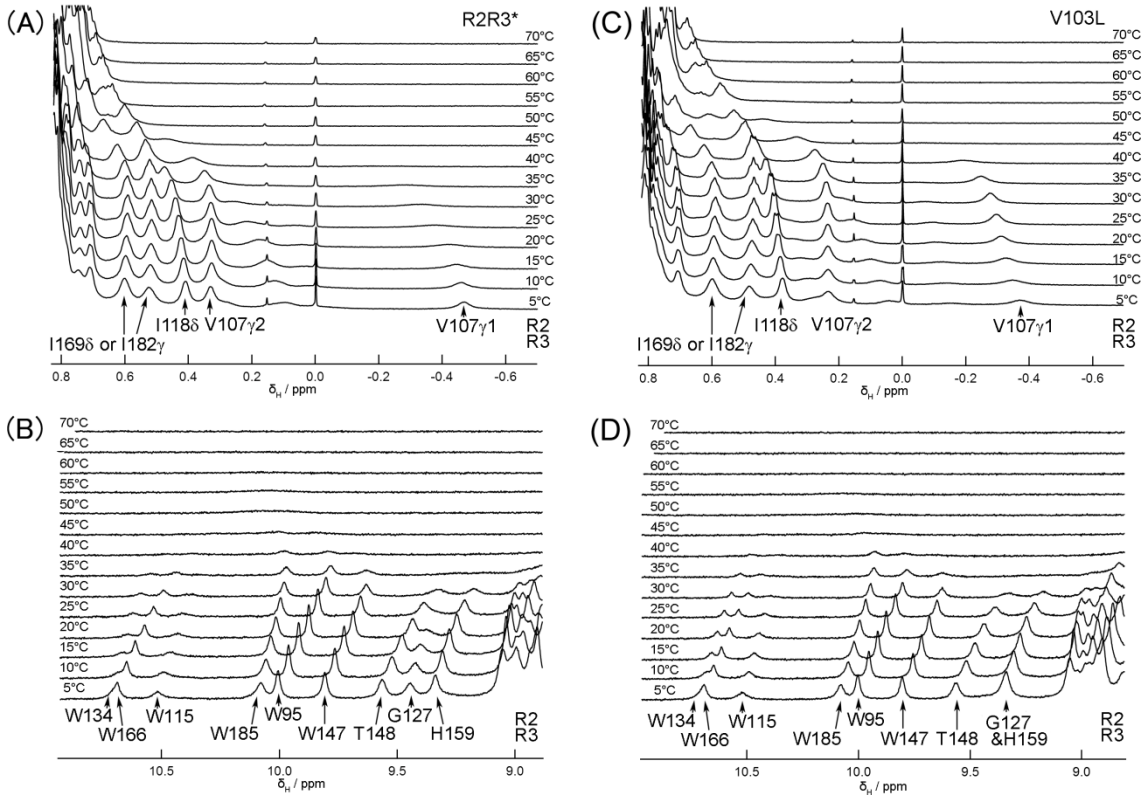



Fig. 2-4. Temperature-dependent $^{1}H$ NMR spectra for R2R3* (A, B) and V103L (C, D). Panels A and C show the high-field methyl proton region and panels B and D show the low-field (Trp $^{1}H^{\varepsilon 1}$) region. Assignments are given for most signals, where the labels in upper and lower rows refer to signals from R2 and R3, respectively.

*The effect of cavity on the side-chain dynamics*

Another finding was that the γ1 methyl proton signal of Val-107 and the ε proton signals of Trp-115 and Trp-134, which are located in close proximity of the cavity in R2R3*, broadened out even at temperatures of < 35ºC (Figs. 2-5C and 5I), while the corresponding signals of the cavity-filled mutant V103L changed less at temperatures of < 40ºC (Figs. 2-5F and 5L). $^{13}C/^{1}H$ HSQC spectra at 15 - 35ºC were also measured and checked their Val-107γ1 methyl signals. I found that the signal was broadened out for R2R3* over the measurement temperature range (Fig. 2-6A) whereas

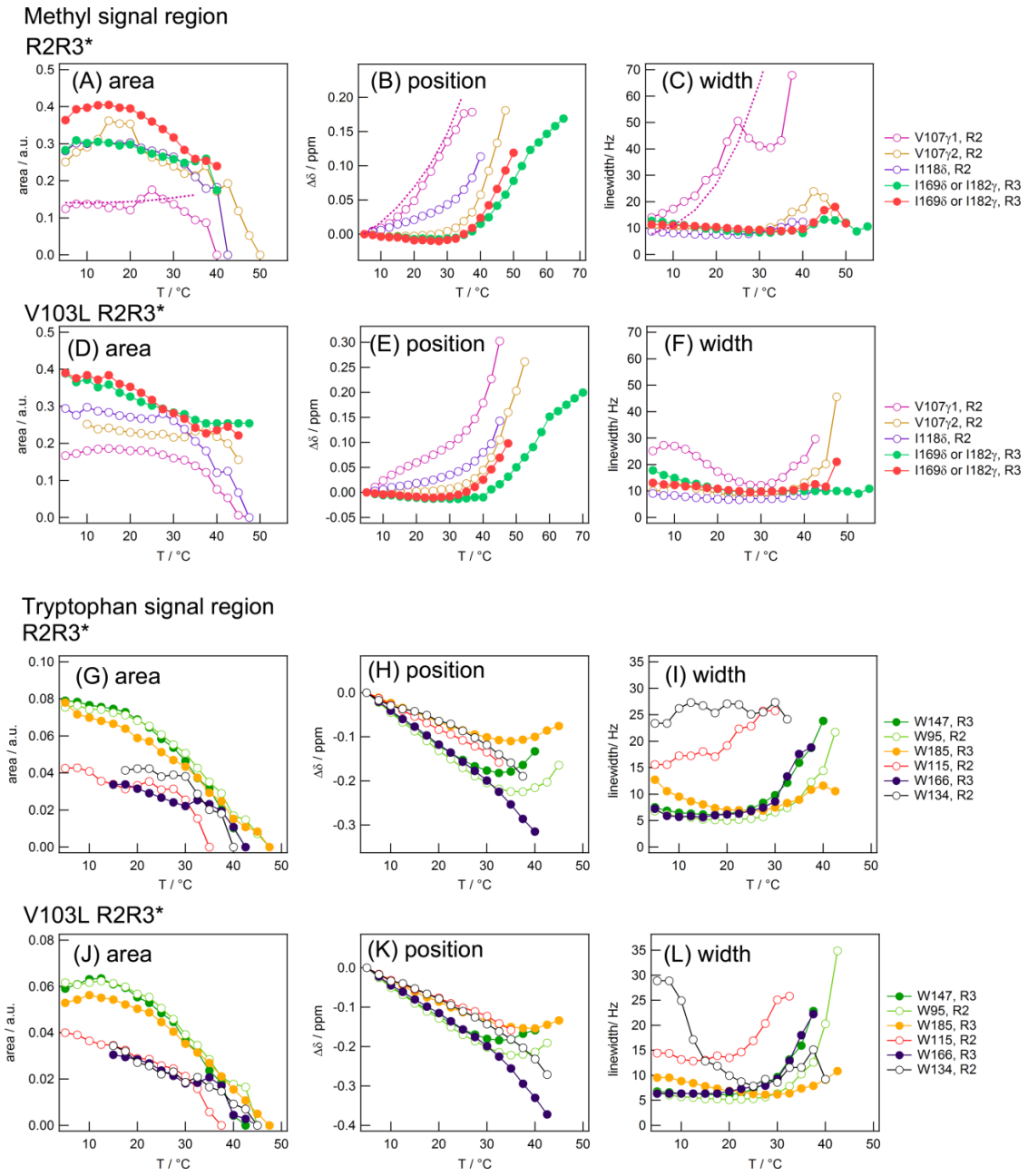



Fig. 2-5. Plots of the temperature-dependent peak areas (left column), positions (middle column), and widths (right column) based on the multi-Lorentzian fit. The upper 6 panels (A-F) and lower 6 panels (G-L) show the results of several high-field methyl proton and low-field tryptophan $^{1}H^{\varepsilon 1}$ signals, respectively. The three upmost panels (A-C) and the three panels in the second row (D-F) are fitting results from R2R3* and V103L, respectively. The lower 6 panels are arranged in the same manner. The broken lines in panels A-C indicates the results of the spectral simulation for the Val-107γ1 methyl proton signal based on the three-state transition model (see text).

the signal was observable for V103L at up to 30ºC (Fig. 2-6B, indicated by an arrow). The observed broadenings also indicate that there is another transition occurring at temperatures lower than that of the global unfolding. I will discuss the mechanism of the signal broadenings in more detail below.

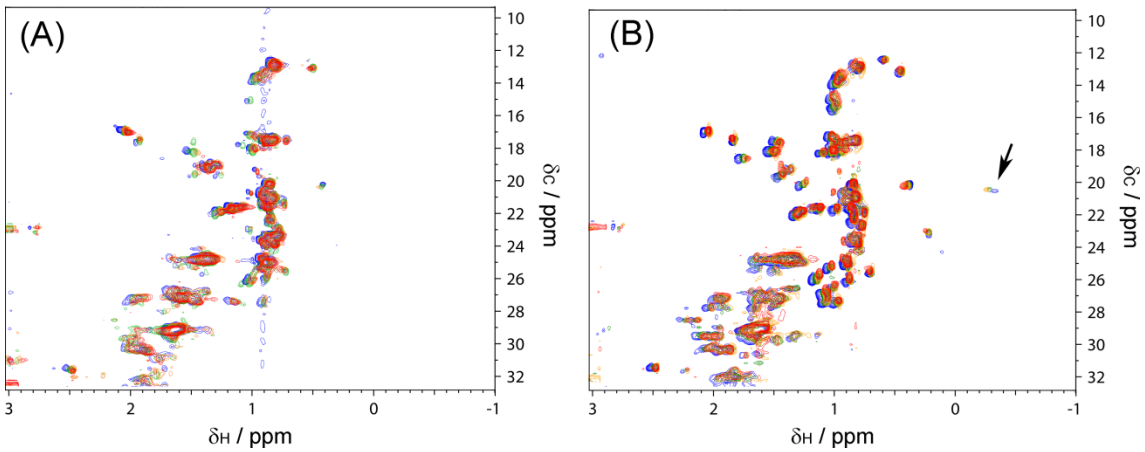



Fig. 2-6. $^{13}C/^{1}H$ HSQC spectra for R2R3* (A) and V103L (B) obtained at 15 (blue), 25 (green), 30 (orange), and 35ºC (red) were superimposed. The arrow in (B) indicates the signal from Val-107γ1 methyl group. The corresponding signal of R2R3* was not observed.

*Temperature-dependent $^{15}N/^{1}H$ HSQC experiments*

Fig. 2-7 shows the superimposition of $^{15}N/^{1}H$ HSQC spectra from the main-chain part of the uniformly $^{15}N$-labeled R2R3* and V103L measured at five different temperatures (5, 15, 25, 35 and 45ºC). With increasing temperature, all cross peaks shifted in position and decreased in intensity, becoming weaker and almost diminished at 35ºC. At 45ºC, all signals completely disappeared except for one signal at the C-terminus, Val-193. This disappearance was kept even at 65ºC (data not shown), where the proteins were assumed to be unfolded completely. Typically, completely

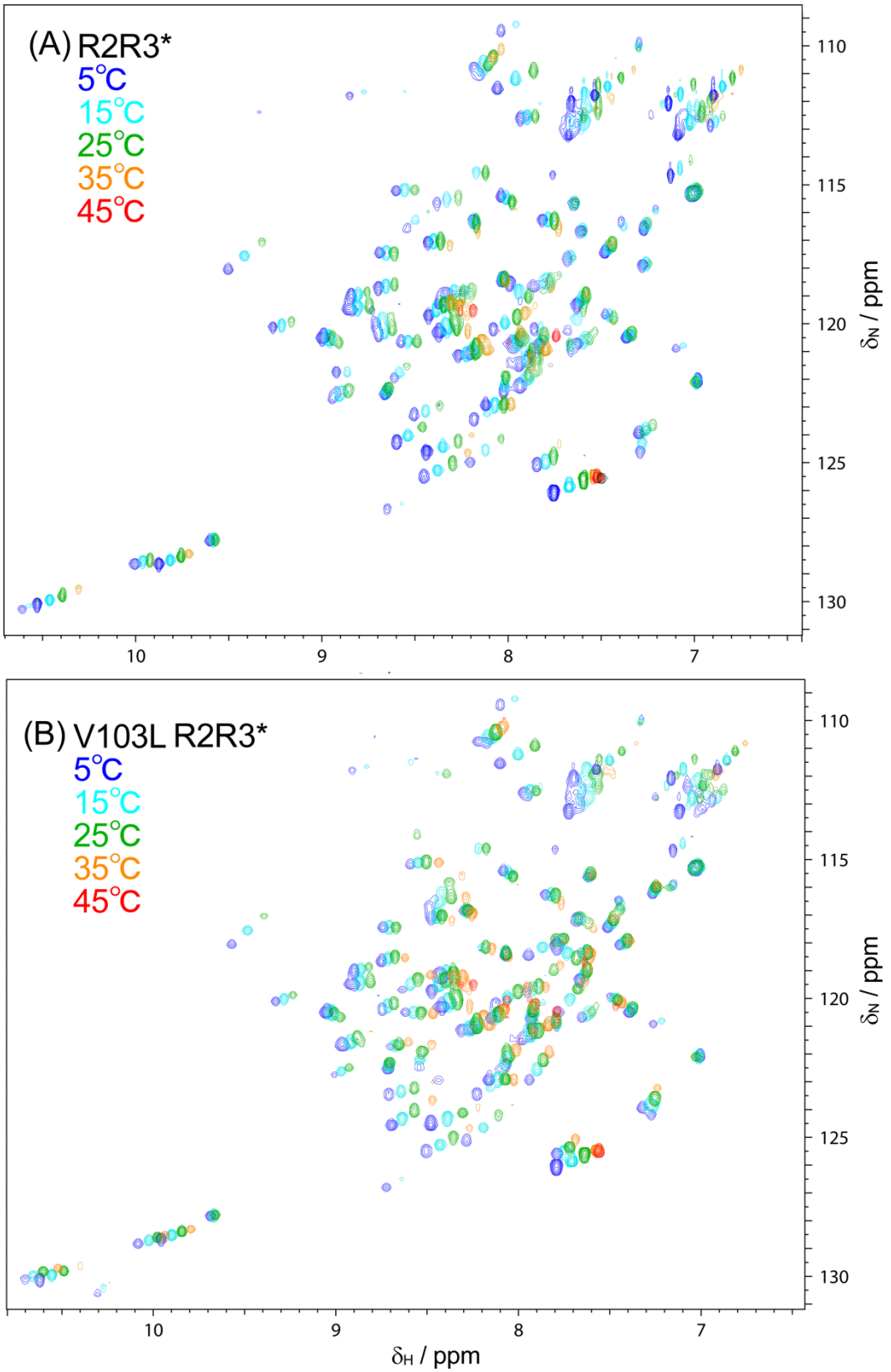



Fig. 2-7. $^{15}N/^{1}H$ HSQC spectra at various temperatures superimposed for R2R3* (A) and V103L (B).

denatured polypeptides show relatively sharp signals because of the increased averaging of the local magnetic field. Considering the present pH condition (pH 7.4) for the measurement, the disappearance of the signals from the denatured state is thought to be rather caused by amide proton exchange with bulk water. Indeed, the estimated intrinsic hydrogen exchange rates ($k_{int}$) for each residue are as high as 34 ± 30 $s^{-1}$, 51 ± 46 $s^{-1}$,74 ± 68 $s^{-1}$ and 103 ± 98 $s^{-1}$ at 15, 25, 35 and 45ºC, respectively.

*Temperature-dependent $^{15}N$ transverse relaxation rates*

$^{15}N$ transverse relaxation rates ($R_{transverse}$) were also measured. The transverse relaxation rates for individual amide $^{15}N$ signals are shown in Fig. 2-8 for a temperature range of 20 - 35ºC. It is prominent that, in the case of R2R3*, the $^{15}N$ spin apparent $R_{transverse}$ values from R2 and R3 progressively increased as the temperature increased (Fig. 2-8A). It is a clear contrast to V103L, which did not show such increases in the corresponding regions in R2 (Fig. 2-8B). These observations indicate that the cavity contributes to local unfolding or fluctuations of R2 in R2R3*, but does not affect the stability of R3.

It is noted that some residues in R2R3*, such as residues around 150, showed constant $^{15}N$ $R_{transverse}$ values of around 10 $s^{-1}$ up to 30ºC (Fig. 2-8A). However, most other residues showed increase in the $R_{transverse}$ values from ~10 $s^{-1}$ to ~30 $s^{-1}$ with increasing temperature, indicating that these residues experience local conformational exchanges, which caused excessive contribution ($R_{ex}$) to their apparent $R_{transverse}$ values.

In order to assess the effect of local conformational changes to the $R_{transverse}$ values, I performed $^{15}N$-spin $R_2$ relaxation dispersion experiment at 25ºC with 600 MHz NMR apparatus (Fig. 2-9). For R2R3*, 9 of 29 analyzable signals showed clear

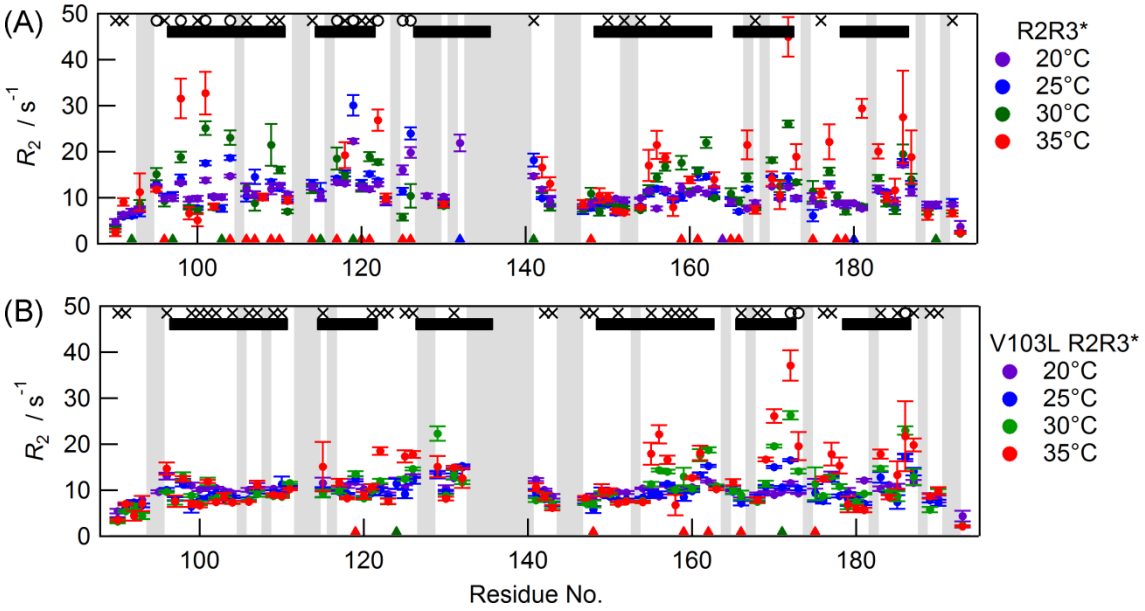



Fig. 2-8. Temperature-dependent transverse relaxation rates for R2R3* (A) and V103L (B). The error bars indicate the fitting errors. Triangles on the horizontal axis indicate the temperature where the signal became broadened beyond detection. The horizontal black bars at the top of each panel indicate the position of α-helices. The markers (circles and crosses) above the horizontal bars indicate the position of residue whose $R_{2,\mathrm{eff}}$ vs $\nu_{CP}$ profile were obtained. Among them, the residues that showed significant $R_2$ relaxation dispersion are indicated by circles. The grey back grounds indicate the position of unassigned residues.

relaxation dispersions, whereas, for V103L, only 3 of 42 analyzable signals showed dispersions (Fig. 2-8, top of each panel, and Fig. 2-9). It should be noted that, for R2R3*, all residues that showed dispersions belong to R2, whereas no R2 residues showed significant dispersion for V103L. The fact that a group of residues showed dispersion also supports the occurrence of a local structural transition. Global curve fittings to the data of individual residues of R2R3* gave an exchange rate of 2,658 (± 192) $s^{-1}$, and residue-specific chemical shift variances ($\Delta\delta_{N,CPMG}$) (Fig. 2-10).

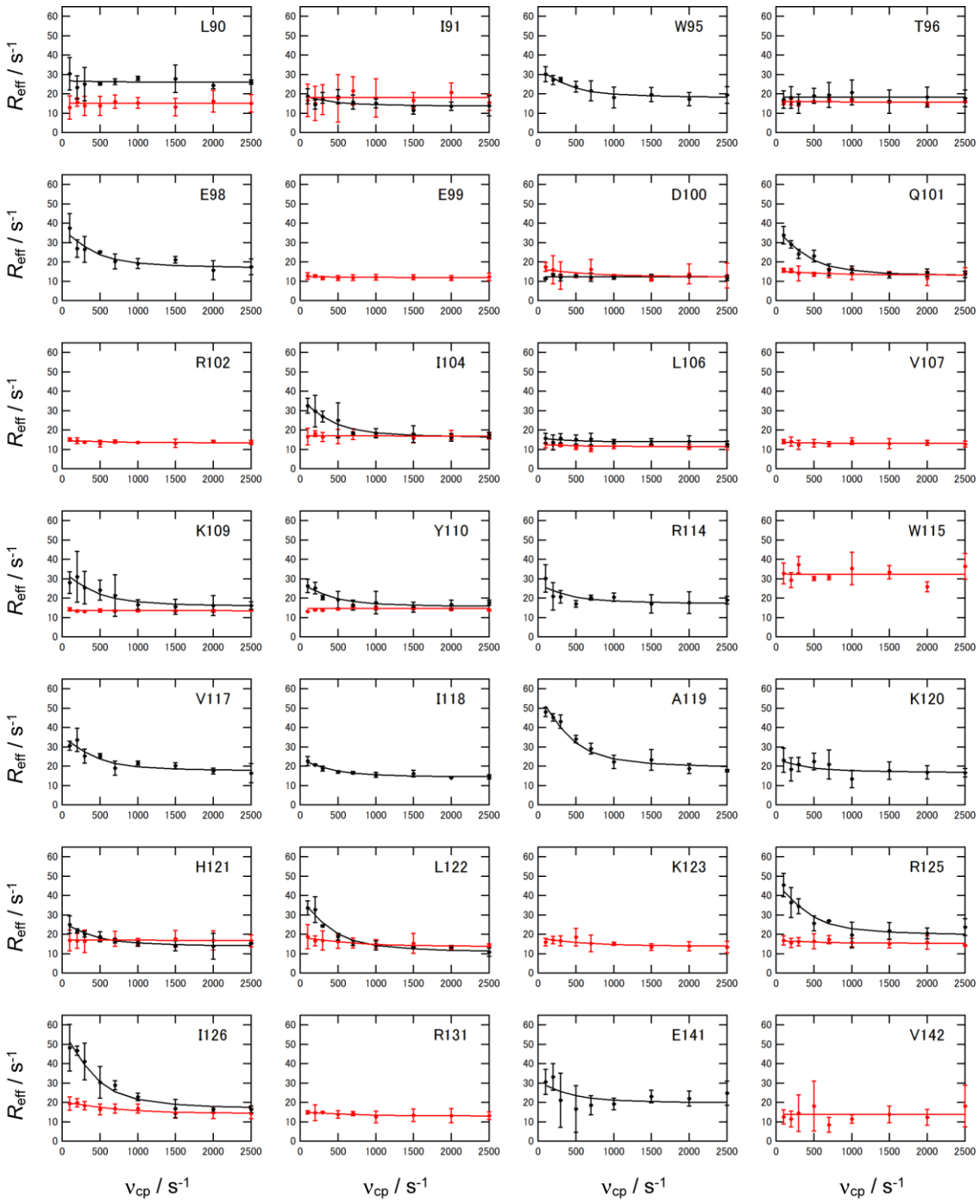



Fig. 2-9. The data of $R_2$ relaxation dispersion experiments. The effective transverse relaxation rates ($R_{2,eff}$) are plotted as a function of $\nu_{CP}$ for only separate signals in R2R3* (black) and V103L (red). The data were globally fitted to equation 5 in the main text. The continuous lines are the theoretical curves. $R_{ex}$ values obtained for Leu-172, Leu-173, and Asn-186 of V103L were 17.4 $\pm$ 2.9 s$^{-1}$, 7.94 $\pm$ 1.59 s$^{-1}$, 12.2 $\pm$ 2.2 s$^{-1}$, respectively.

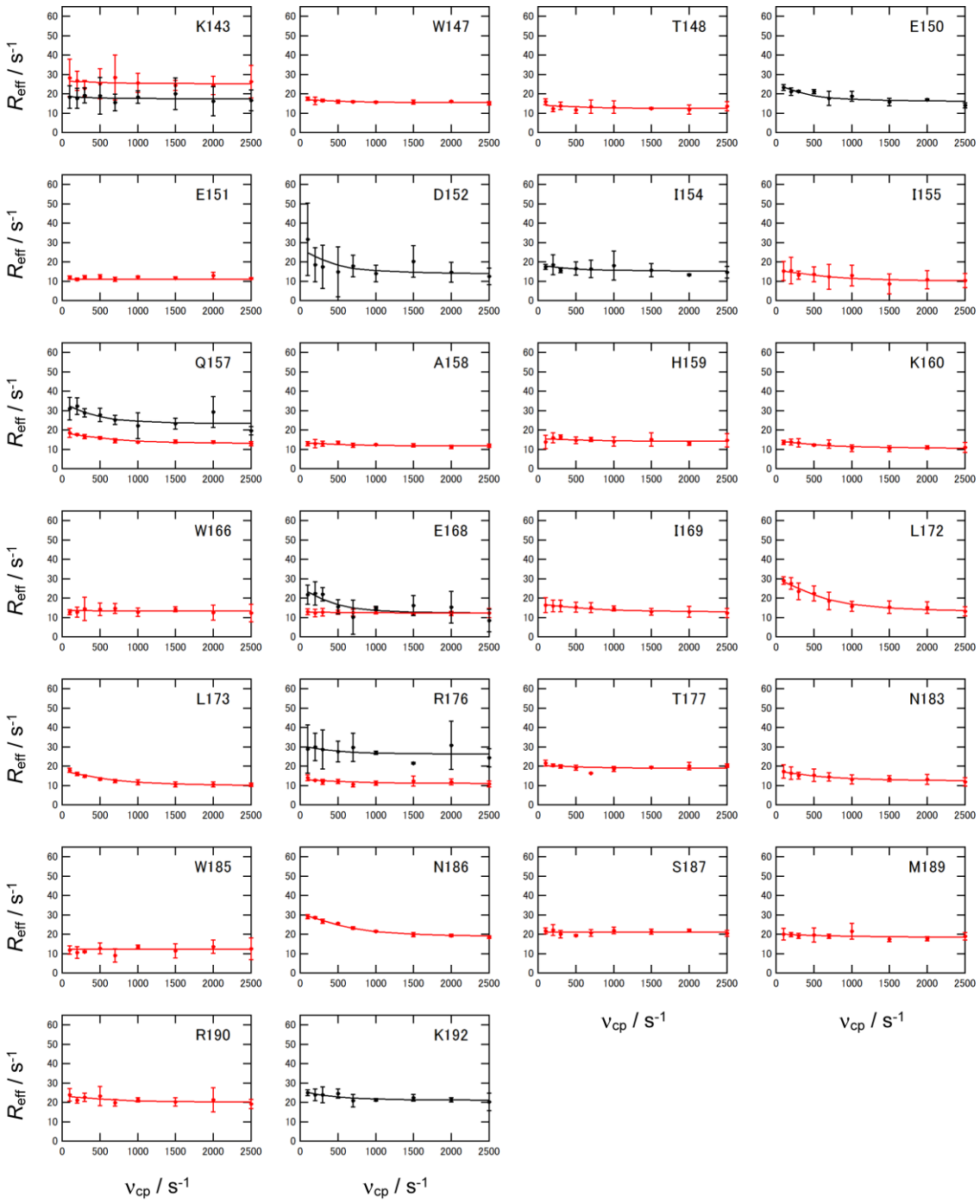



Fig. 2-9. (continue)

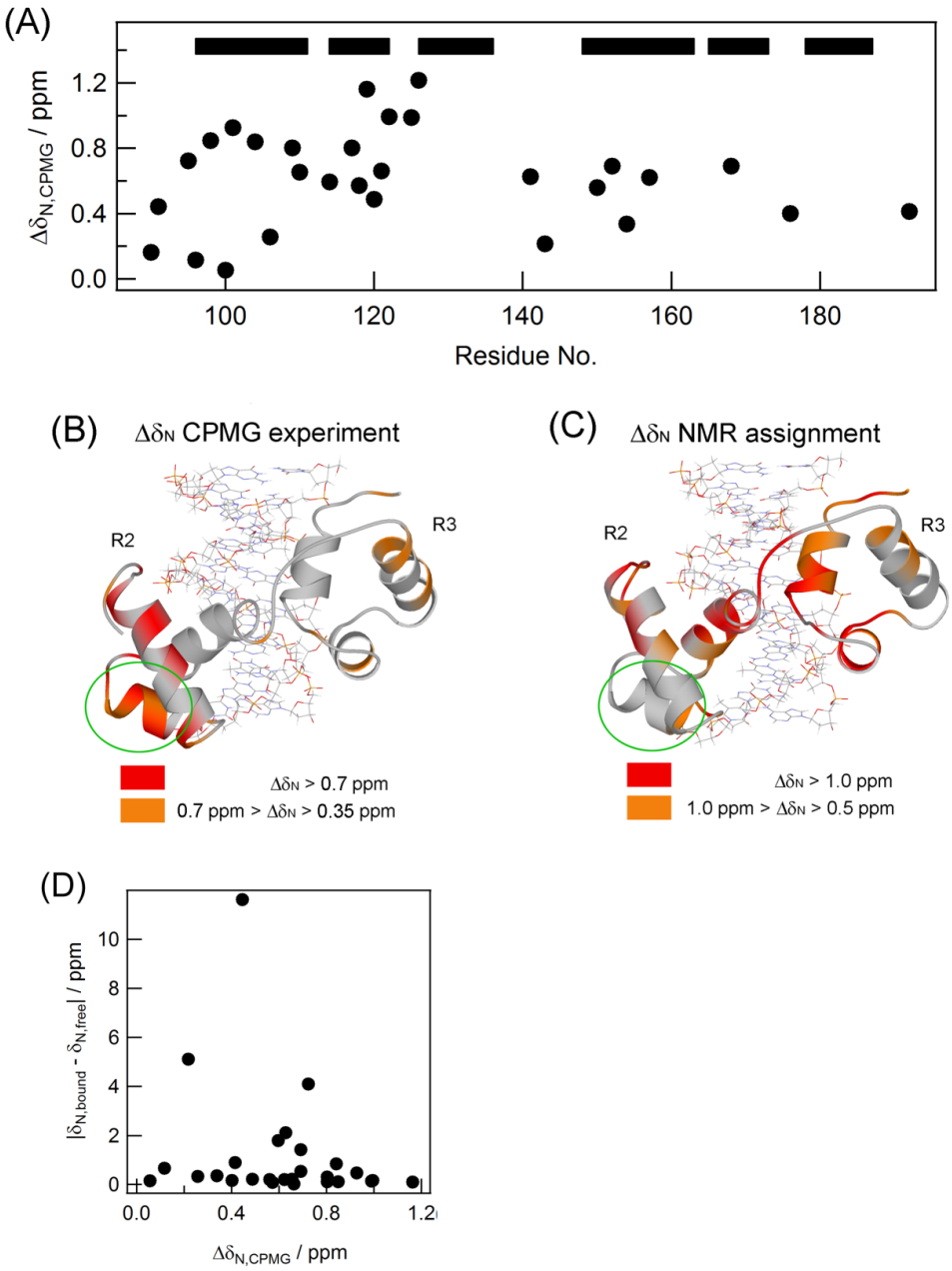



Fig. 2-10. Results of the analysis of the $R_2$ relaxation dispersion data. (A) Obtained residue-specific chemical shift variances ($\Delta\delta_{N,CPMG}$) for R2R3*. The horizontal black bars at the top of each panel indicate the position of α-helices. (B, C) The mapping of the residues with significant $\Delta\delta_{N,CPMG}$ from the analysis of the $R_2$ relaxation dispersion data (B) and $\Delta\delta_{N,free\text{-}bound}$, the chemical shift differences in the free and DNA-bound states (C) on the crystal structure. The colors of the residues indicate the magnitude of $\Delta\delta_N$ values according to the gauge below the structures. (D) Correlation plot between $\Delta\delta_{N,CPMG}$ and $\Delta\delta_{N,free\text{-}bound}$ values.

*Effect of cavity on the global fluctuation of the protein: the $^{15}N/^{1}H$ HSQC cross peak intensity*

Figs. 2-11A and B show the site-specific intensity changes in $^{15}N/^{1}H$ HSQC cross peaks for the two proteins. Progressive decreases in individual cross peak intensities with increasing temperature were observed. It is, therefore, expected that the intensities of HSQC cross peaks are attenuated by hydrogen exchange as well as by conformational exchange broadenings instead of by intrinsic $R_{transverse}$. In the case of R2R3*, the signal intensities from R2 interestingly diminished more quickly than those from R3 with increasing temperature (Fig. 2-11A); however, this tendency was lost in the cavity-filled mutant V103L, particularly at 35ºC (Fig. 2-11B). This was evident from the correlation plot of relative intensities separately for R2 and R3 (Figs. 2-11C, 11D, 11E, and 11F), in which it was even clear that the difference in cross peak intensities of R2 increased with increasing temperature, while that of R3 was not so significant. These results also support that the cavity contributes to local unfolding or fluctuations of R2 in R2R3*, but does not affect the stability of R3, that is, the two repeats are not dynamically coupled, consistent to the results of $R_2$ relaxation dispersion experiment (Ogata *et al*., 1994).

## Discussion

The present experiments carried out using CD as well as $^{1}H$ and $^{15}N$ NMR consistently showed that the global stability of c-Myb R2R3 is significantly low at 37ºC. Furthermore, the precise analysis of the obtained NMR data suggested that there is a state distinct from the N and U states. This is a locally fluctuating state and here denoted as I. Considering the obtained results altogether, I suggest that the thermal transition of

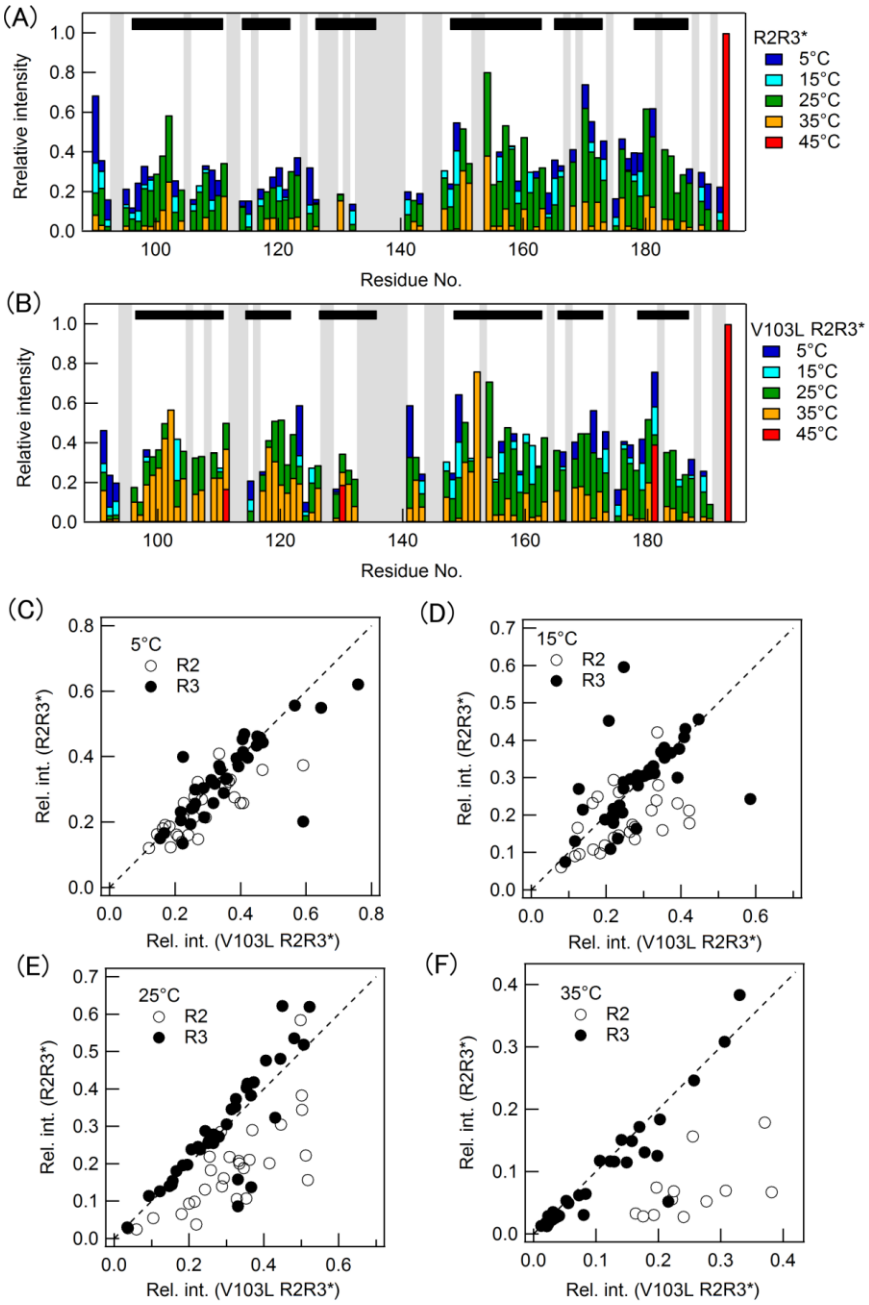



Fig. 2-11. Temperature-dependent $^{15}N/^{1}H$ HSQC signal intensities and their correlations between R2R3* and V103L. (A, B) Intensities of HSQC cross peaks are shown in histograms for R2R3* (A) and V103L (B). The intensities are normalized with respect to the intensity of the C-terminal residue, Val193. The black bars at the top of each panel indicate the position of α-helices and the grey backgrounds indicate the position of unassigned residues. (C-F) Correlations of the relative intensities between R2R3* and V103L at 5°C (C), 15°C (D), 25°C (E), and 35°C (F) are shown. The open and solid circles represent data from residues of R2 and R3, respectively.

c-Myb R2R3 proceeds through Scheme 2-1. In this model, the N state is in equilibrium with the I state as well as the U state. These equilibria are temperature-dependent. At low temperature, the N state is dominant. At physiological temperature, the I state accumulates to some extent, and, at temperature higher than 45ºC, the U state becomes dominant. It is noted that the time scales of the exchange rate ($k_{ex}$) of the transitions between the N and I states and between the N and U states are supposed to be approximately $10^3$ $s^{-1}$ and less than $10^1$ $s^{-1}$, respectively. Considering the experimental pH of 7.4, the hydrogen exchange reaction is assumed to be significantly rapid ($k_{int}$ = $10^1$ ~ $10^2$ $s^{-1}$). Thus, some residues in the I state and all residues in the U state will be subjected to the hydrogen exchange, causing the reductions of the signal intensities. Based on this model, I will discuss the present NMR results below.

When integrated as a function of temperature, the area of NMR signals at 0.8 - 0.9 ppm exhibited a thermal transition at ~50ºC similar to that observed in the CD experiments (Figs. 2-2C and 3C), confirming that the U state became dominant at these temperatures. The signal disappearances at almost all residues in the temperature-dependent HSQC measurements (Fig. 2-7) were caused by hydrogen exchange, confirming that the residues are exposed to the solvent.

Some NMR signal broadenings indicate the existence of the I state. The enhancements of apparent $R_{transverse}$ observed for some residues (Fig. 2-8) were attributed to increased local conformational exchanges with the I state. The broadening at the γ1 methyl signal of Val-107 (Fig. 2-5C) is also attributed to the exchange dynamics between N and I. Furthermore, disappearance of the methyl signal in $^{13}C/^{1}H$ HSQC spectra (Fig. 2-6) might result from the exchange between multiple conformations, because the chemical shifts of methyl $^{13}C$ spins are the weighted average

of those from each rotameric state and are, therefore, sensitive to the population change among these states (Smith *et al*., 2015). On the other hand, the signal broadenings of the ε protons of Trp-115 and Trp-134 observed at temperatures of < 35ºC (Figs. 2-5I and 5L) can be attributed to the hydrogen exchange with the bulk water upon local exposures in the I state besides the contribution from the local conformational changes. At higher temperatures, additional hydrogen exchange occurred from the U state, leading to the complete broadenings on both proteins at similar temperatures.

To check the suggested model, I simulated the temperature dependence of Val-107γ1 methyl proton signals for R2R3* as a representative example based on Scheme 2-1 (Fig. 2-12). For the simulations, $\Delta H_{NU}$, $\Delta H_{NI}$, $T_{m\,NU}$ and $T_{m\,NI}$ were assumed to be 233 kJ mol$^{-1}$, 60 kJ mol$^{-1}$, 47$^{o}$C and 37$^{o}$C, respectively. In addition, $k_{ex,NI}$ and $k_{ex,IU}$ were assumed to be 1,000 s$^{-1}$ and 10 s$^{-1}$, respectively. In Fig. 2-12, the simulation showed that the signal appeared at the chemical shift of the N state (~ –0.5 ppm) at 5$^{o}$C and significantly shifted toward low-magnetic field with a severe line broadening in the temperature range from 5 to 40$^{o}$C (dashed line in Fig. 2-5C). Above 40$^{o}$C, new NMR signal appeared at 0.8 ppm, corresponding to the U conformers, with the simultaneous disappearance of the N state signal. The simulation substantially reproduced the experimental spectral changes (Fig. 2-4A), confirming the validity of Scheme 1 and the orders of the values assumed for respective parameters. The assumed $k_{ex,NI}$ value (≈ 1,000 s$^{-1}$) was comparable with that estimated from the $^{15}$N-spin $R_2$ relaxation dispersion experiment (2,658 s$^{-1}$) described above.

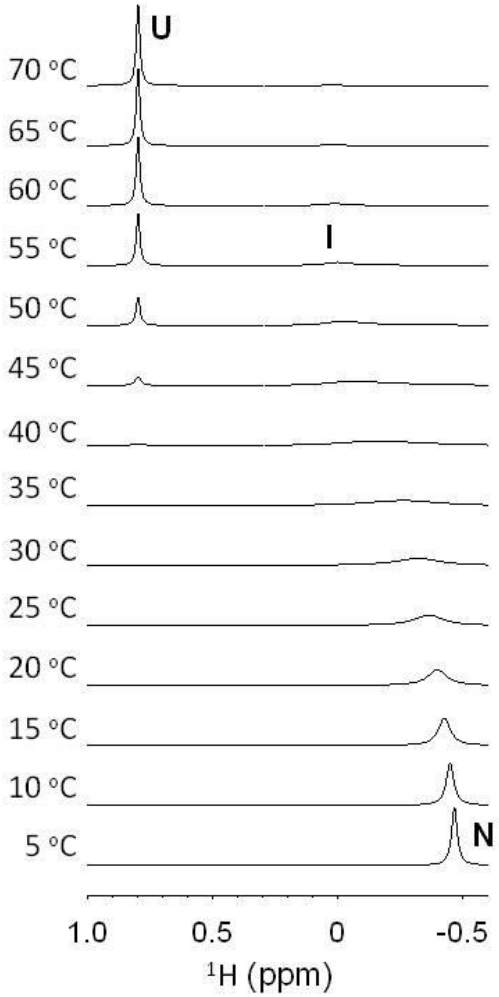



Fig. 2-12. Line-shape simulation for the methyl proton peaks of R2R3*, in accordance with the three-state exchange model between folded state (N), intermediary state (I) and unfolded state (U). For spectral simulations, $\Delta H_{NU}$, $\Delta H_{NI}$, $T_{m\,NU}$ and $T_{m\,NI}$ were assumed to be 233 kJ mol$^{-1}$, 60 kJ mol$^{-1}$, 47$^{o}$C and 37$^{o}$C, respectively. $k_{ex,IU}$ was assumed to be 10 s$^{-1}$, while $k_{ex,NI}$ was 1,000 s$^{-1}$. The chemical shifts for the N, I and U states were assumed to be –0.5 ppm (–300 Hz), 0.04 ppm (25 Hz) and 0.8 ppm (480 Hz), respectively.

All the present data support the three-state transition mechanism of c-Myb R2R3* rather than the two-state mechanism. Assuming the two-state model, the normalized positions of the signal would be proportional to the population of the N (or U) state. However, when I plotted the temperature-dependent positions of several

methyl signals (Fig. 2-13), the $T_m$ values were estimated to be significantly different from the $T_m$ values obtained from the CD data (Tables 2-1 and 2-3). This result indicates that the destination of the temperature-dependent signal shift does not represent the denatured state. In other words, there should be another state distinct from both the N and U states.

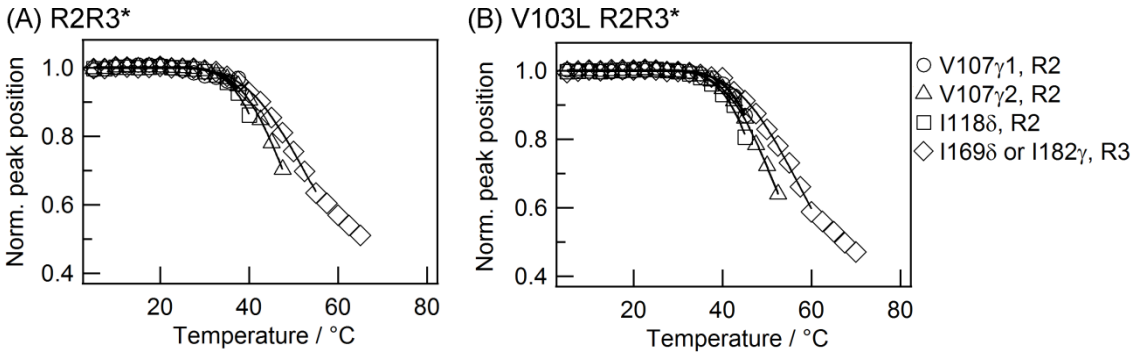



Fig. 2-13. Evidence to controvert the possibility of two-state mechanism. Plots of the normalized peak positions of several methyl proton signals of Figs. 2-4 and 2-5 as a function of temperature. For the normalization, the initial and final chemical shifts were assumed to be those obtained at 5°C and those for the corresponding methyl protons of typical unfolded peptides, respectively. The solid lines are the best-fit to the data when two-state transitions are assumed. See Table 2-3 for the assumed chemical shift values and the obtained melting temperatures, which were inconsistent with those obtained by fitting to the CD melting curves.

Table 2-3. Limiting shifts assumed for the denatured state ($\delta_{\text{random coil}}$) and melting temperatures ($T_m$) deduced from NMR signals of some methyl protons of R2R3* and V103L based on the two-state model.

| | | V107 δ1 | V107 δ2 | I118 δ | I169 δ or I182 γ2 |
|---|---|---|---|---|---|
| $\delta_{\text{random coil}}$ /ppm | | 0.94 | 0.94 | 0.89 | 0.89 |
| $T_m$ /°C | R2R3* | 53.0 | 53.8 | 50.2 | 61.2 |
| | V103L | 55.1 | 56.9 | 53.3 | 64.2 |

I also examined an off-pathway model represented as I ⇄ N ⇄ U (Scheme 2-1A). The simulations of the $^1$H NMR spectra based on the off-pathway model showed almost the same line shape change as that based on Scheme 2-1B (data not shown). However, considering the large gap between the time scales of the transitions to the I and U states, the former is $k_{ex}$ = 1,000 $s^{-1}$ and the latter is $k_{ex}$ = 10 $s^{-1}$, it is likely that the N and I states lie in the same basin and the U state is in another basin on the energy landscape and these states are mutually exchangeable within a basin (N ⇄ I) or between basins (I ⇄ U and N ⇄ U). In such situations, it is practically impossible to define clearly whether the intermediary state is on-pathway or off-pathway. However, the I state significantly populated at temperatures slightly below $T_m$. It is likely that the local unfolding is a preceding event to the global unfolding, and the I state can access to the U state more easily than the N state can. Therefore, in the present result, I tentatively suggested the on-pathway scheme (Scheme 2-1A).

It was suggested that the I state is in equilibrium with the N state at physiological temperature. The results of the CPMG experiment gave some information about the conformation of the I state. Fig. 2-10A shows residue-specific chemical shift variances obtained from the present CPMG experiment ($\Delta\delta_{N,CPMG}$). In Fig. 2-10B, the residues with significant $\Delta\delta_{N,CPMG}$ were mapped onto the protein structure. For a comparison, the chemical shift differences in the free and DNA-bound states ($\Delta\delta_{N,free\text{-}bound}$) quoted from Ogata *et al*. were also mapped (Fig. 2-10C, Ogata *et al*., 1994; Ogata *et al*., 1995). Fig. 2-10D shows the correlation plot between the $\Delta\delta_{N,free\text{-}bound}$ values and $\Delta\delta_{N,CPMG}$ values. The plot showed little correlation between them. For example, the residues of the second helix of R2 (indicated by green circles) showed no significant $\Delta\delta_{N,free\text{-}bound}$, whereas they showed significant $\Delta\delta_{N,CPMG}$ values,

indicating that these residues in the I state are experiencing different conformations from both the free and DNA-bound states. Other residues might also be in similar situations, resulting in a wide dispersion in the $\Delta\delta_{N,free\text{-}bound}$ vs $\Delta\delta_{N,CPMG}$ plot (Fig. 2-10D). The present interpretation is consistent with the finding that the cavity-containing R2 exhibits a high degree of water permeability to its internal Trp residues (Puthenprackal *et al*., 2012). The present results of 1D and 2D NMR measurements clearly indicated additional cavity-based fluctuations inherent to the wild-type c-Myb protein.

Taking a conformational ensemble prior to a DNA binding is well-known characteristics for intrinsically disordered proteins (Ward *et al*., 2004). Conformational disorder is often advantageous in binding to a macromolecule (Dyson, 2012). Since c-Myb is just a DNA-binding protein, it is natural to consider that c-Myb takes such a conformational ensemble prior to its DNA binding. Thus, it is plausible that the I state is in an ensemble of several conformations, one of which might be the DNA-bound form, and these conformations are quickly interconvertible.

In contrast, the cavity-filled mutant, V103L, would have relatively higher population of the N state, which is rigid and "unfavorable" for the DNA-binding. Thus, the cavity of R2 hydrophobic core is relevant to the DNA-binding affinity. The SPR experiments revealed that the reduced DNA-binding affinity for V103L (Table 2-2) was mainly due to the increase in $k_{off}$, not to the decrease in $k_{on}$. This may happen because the initial binding, which influences the $k_{on}$ value, is driven by long-range electrostatic interactions. After the initial complex is made, it proceeds to a tighter binding by adjusting the conformation of the binding site. The cavity would increase the local conformational fluctuation and would facilitate the conformational change required for

the tight binding with DNA, which can affect $k_{off}$. Because long-range electrostatic interactions would mainly contribute to $k_{on}$ in protein and DNA interactions (Oda and Nakamura, 2000), the conformational change would not be the rate-determining step for the initial binding. The lower stability of the protein-DNA complex upon V103L mutation could increase the $k_{off}$. Previous NMR relaxation experiments previously carried out at 17ºC also revealed that the slow conformational motion is suppressed by the V103L mutation (Ogata *et al*., 1996).

The present experiments showed that the global stability of c-Myb R2R3* is basically low, allowing a significant fraction of more dynamic conformer to be present at 37ºC. Among the two repeats, the R2 mobility is brought by the presence of cavity. The low stability and high fluctuation of the folded conformation of R2R3* might be attributable to the small size of the subunits, each of which consists of only 51 - 52 amino acids lacking disulfide bonds and mutual stabilizing cooperativity. Additional effects from the cavity on R2 would allow thermal fluctuations to occur on the wild-type DBD, leading to a sufficient mobility of the polypeptide chain at physiological temperature. The resulting flexibility is likely to allow specific binding sites to be found with more favorable collisions with target DNA (Oda and Nakamura, 2000), although such disorder has a large entropic penalty for binding. In the case of the c-Myb DBD, the I state is not the ground state, but exists as an excited-state conformer at physiological temperature. The concept of the functional state of a protein being the rarely populated excited-state rather than the ground-state itself has long ago been put forward and has since been widely recognized in many proteins (Akasaka, 2006; Fourme *et al*., 2012). In the case of c-Myb R2R3, the functional state could be observed under the physiological conditions, possibly due to its relatively larger population.

# Chapter 3 Folding thermodynamics of R2R3 in correlation with its α-helical contents

## Introduction

Tertiary structures of proteins are maintained by intramolecular interactions including covalent and non-covalent interactions such as disulfide bonds, salt bridges, hydrogen bonds, and hydrophobic interactions. Owing to protein architecture, some proteins, such as protease inhibitors, are rigid and highly stable, while others such as enzymes fluctuate largely in solution. In other words, there exists a close correlation between the protein stabilities and its functions. “Rigidness” and “softness” are usually balanced in every protein in addition to its individual domains. These properties are regulated by intramolecular interactions, especially by non-covalent bonds, which can be subtly or drastically altered by environmental factors. Non-covalent interactions can be influenced by physical perturbation like temperature and pressure, and chemical perturbations such as pH, ionic strength, denaturant, and buffer components. The ionization state of residues is one of the important factors responsible for maintenance of the protein tertiary structure and is gravely affected by solvent conditions such as pH and ionic strength (Phelan *et al*., 2002; Grey *et al*., 2006; Pace *et al*., 2009; Lau *et al*., 2010; Robinson *et al*., 2014). Structural and physical properties of model peptides which form typical α-helix and/or β-sheet structures, under different pH conditions have been reported previously (Scholtz *et al*., 1993; Huyghues-Despointes and Baldwin, 1997; Metrick *et al*., 2013). In most of these analyses, the protein stability was analyzed only with respect to Gibbs free energy change (Δ*G*). Precise folding thermodynamics including enthalpy (Δ*H*) and entropy changes (Δ*S*) in correlation with tertiary structure can provide critical information on protein folding, including the strength of

intramolecular interactions and the protein flexibility. In this study, I used the functional unit of protein with a helix-turn-helix motif, to analyze the pH-dependent changes on tertiary structure and folding thermodynamics.

Using the functional unit of c-Myb DBD, R2R3, I found clear difference in secondary structure and thermal stability with pH ranging from 4.0 to 7.5. This unique property of R2R3, which enables large fluctuations, makes it possible to analyze the correlation of helical content with folding thermodynamics precisely with greater precision.

## Materials and Methods

*Protein expression and purification*

The expression and purification methods of R2R3 proteins were described at Chapter 2. The purity of each protein was determined to be about 95% by SDS/PAGE analysis.

*CD measurements*

Far-UV (200-250 nm) and near-UV CD (250-340 nm) spectra were recorded on a Jasco J-820 and J-720 spectropolarimeter at 20˚C equipped with Peltier-type temperature control system, as described at Chapter 2. The spectra of R2R3* were obtained in 20 mM sodium acetate (pH range from 4.0 to 5.5) or potassium phosphate (pH range from 6.0 to 7.5) buffer, using quartz cell with 0.2 cm path-length. The protein concentrations were 0.1 mg $ml^{-1}$ and 1.0 mg $ml^{-1}$ for far- and near-UV respectively. CD spectra were obtained using scanning speed of 20 nm $min^{-1}$, a time response of 1 s, a bandwidth of 1 nm, and an average over 4 or 8 scans, for the far- and near-UV range,

respectively. In order to assess the effects of protein concentration on secondary structures, far-UV CD spectra were measured in the different concentration of R2R3*, 0.02 and 0.5 mg $ml^{-1}$. The α-helical contents, $f_H$ were calculated as reported previously (Chen *et al*., 1972).

$$f_H /\% = - \{([\theta]_{222} + 2340) / 30300\} \times 100 \qquad \text{(Eq. 3-1)}$$

The melting curves were recorded in temperature mode at 222 nm, from 20 to 80˚C with a heating rate of 1.0˚C $min^{-1}$.The protein concentrations were 0.02 mg $ml^{-1}$ and the quartz cell with 1 cm path-length was used. The analysis of the transition curves obtained by temperature-scanning CD measurements was performed on the basis of two-state transition model, as described at Chapter 2.

*DSC measurements*

DSC experiments were carried out on a Nano-DSC calorimeter (TA instruments). The data were collected by heating the solution from 5 to 90˚C at a rate of 1˚C $min^{-1}$ and the sample concentrations were 0.5 or 2.0 mg $ml^{-1}$ in 20 mM sodium acetate (pH range from 4.0 to 5.5) or potassium phosphate (pH range from 6.0 to 7.5) buffer. The outer buffer solution recovered from final exchanged experiment was used in the reference cell for each case. The data were analyzed using CpCalc software supplied by the manufacture, in order to subtract the baseline from heat capacity data and convert to molar heat capacity. The fitting analyses were performed by the least-squares methods on Origin 5.0 software. The $\Delta C_p$ was calculated from the difference in the $C_p$ values between unfolded and folded states at $T_d$. $\Delta H_{cal}$ was calculated by integrating the area in each heat capacity curve. The $\Delta H_{vH}$ was calculated by the next equation for assuming the two-state transition.

$$\Delta H_{vH}(T_d) = 4RT_d^2 \{C_p(T_d)/\Delta H_{cal}(T_d)\} \quad \text{(Eq. 3-2)}$$

where $R$ is the gas constant. Using the $\Delta C_p$ values with the $\Delta H_{cal}$ and $T_d$ values, the thermodynamic parameters of unfolding as a function of temperature, $\Delta G(T)$, $\Delta H(T)$, and $\Delta S(T)$, could be calculated from the following equation, on the assumption that the $\Delta C_p$ is constant within the temperature range analyzed.

$$\Delta G(T) = \Delta H_{cal}(T_d)\{1 - (T/T_d)\} - \Delta C_p T \ln(T/T_d) - \Delta C_p(T_d - T) \quad \text{(Eq. 3-3)}$$

$$\Delta H(T) = \Delta H_{cal}(T_d) + \Delta C_p(T - T_d) \quad \text{(Eq. 3-4)}$$

$$\Delta S(T) = \Delta S(T_d) + \Delta C_p \ln(T/T_d) \quad \text{(Eq. 3-5)}$$

## Results

*Effects of pH on the secondary and tertiary structures of R2R3**

The far- and near-UV CD spectra of R2R3* were measured at 20˚C and pH ranging between 4.0 and 7.5. In concordance with previous results (Oda *et al*., 1997b; Oda *et al*., 1998; see Chapter 2), the far-UV CD spectra of R2R3* in the pH range analyzed in this study indicated typical α-helical structures, but the intensities of CD values were different (Fig. 3-1A). The spectrum at the protein concentration of 0.1 mg $ml^{-1}$ was similar to those of 0.02 mg $ml^{-1}$ and 0.5 mg $ml^{-1}$ (Fig. 3-1B), indicating that the secondary structure is independent to the concentration. The α-helical contents, $f_H$, were calculated from Eq. 3-1, and plotted as a function of pH (Fig. 3-1C). The α-helical content of R2R3* at pH 6.5 was 60%, and was the highest compared to the α-helical content at other pHs. As shown in the pH decreased from 6.5 to 4.0 or increased from 6.5 to 7.5, the α-helical content was also reduced. As the previous NMR study, the α-helical content of c-Myb R2R3 wild-type was 59% at pH 6.8 (Ogata *et al*., 1994; PDB: 1MSE), which was close to the α-helical content of 60% at pH 6.5 determined in

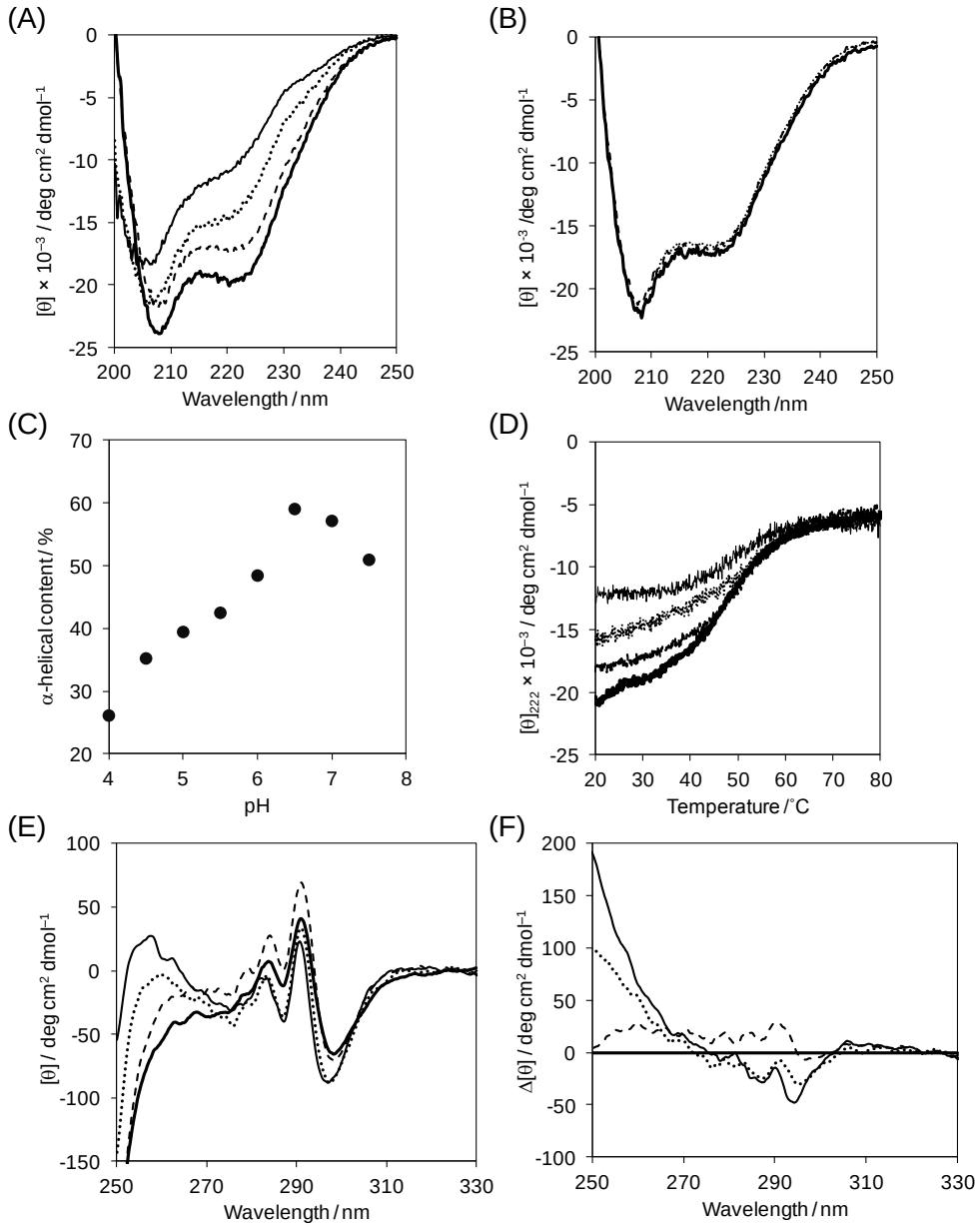



Fig. 3-1. The secondary and tertiary structure analysis by CD. (A) Far-UV CD spectra of R2R3* at pH 4.0, 5.0, 6.0, and 7.0. (B) Far-UV CD spectra of R2R3* at pH 6.0 under different protein concentrations, 0.1 (solid line), 0.5 (dotted line), and 0.02 (broken line) mg $ml^{-1}$. (C) Plots of α-helical contents calculated from eq. 1 as a function of pH. (D) Thermal denaturation analysis by CD. Transition curves of R2R3* at pH 4.0, 5.0, 6.0, and 7.0. (E) Near-UV CD spectra of R2R3* at pH 4.0, 5.0, 6.0, and 7.0. (F) The difference spectra obtained by subtracting the spectrum at pH 7.0 from those at respective pHs. The thin solid, dotted, broken, and thick solid lines indicate the spectra at pH 4.0, 5.0, 6.0, and 7.0, respectively.

this study. Another NMR study for isolated R3 showed α-helical content of ~ 80% at pH 5.0 (Furukawa *et al*., 1996; PDB: 1IDY). Taken together, my results suggest that pH change from 6.5 to 4.0 and 7.5 can lead to serious structural fluctuations of R2R3*, especially R2, and partial distortion of the α-helices. Furthermore, secondary structures were changed and α-helical contents were also increased by the addition of TFE (Figs. 3-2A and 2B).

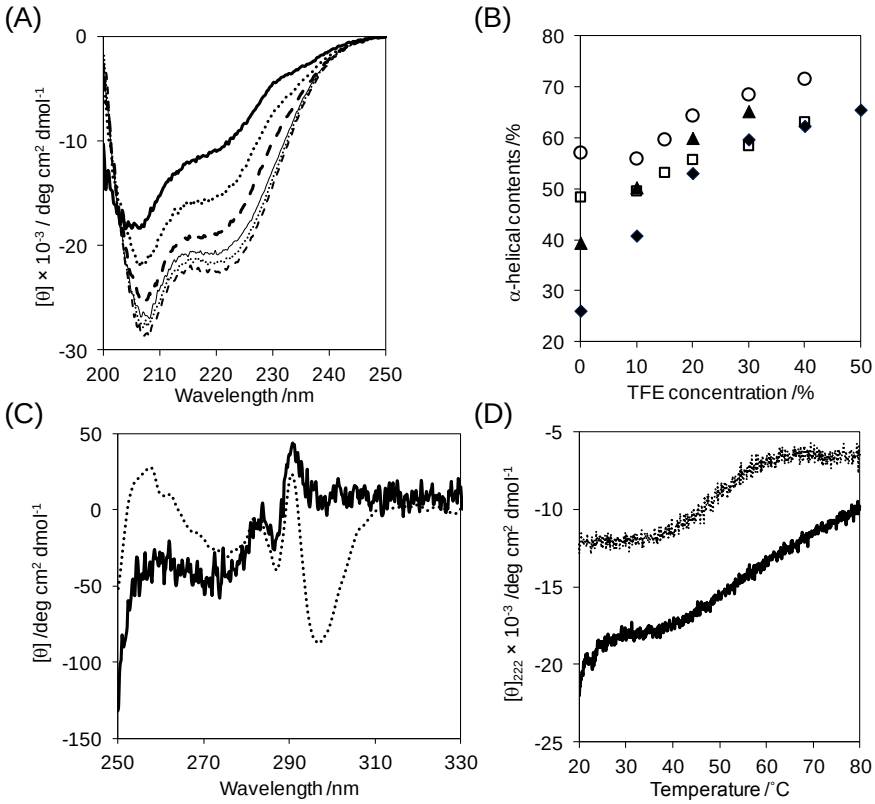



Fig. 3-2. (A) Far-UV CD spectra of R2R3* at pH 4.0 in the absence (solid line) or presence of 10% (dotted line), 20% (dash line), 30% (thin solid line), 40% (thin dotted line), 50% (thin dash line) TFE. (B) Plots of α-helical contents calculated from eq. 1 as a function of TFE concentration. The close diamond, close triangle, open square, and open circle indicate pH 4.0, 5.0, 6.0, and 7.0, respectively. (C) Near-UV CD spectra of R2R3* at pH 4.0 in the absence (dotted line) or presence (solid line) of 30% TFE. (D) Thermal transition curves of R2R3* at pH 4.0 in the absence (dotted line) or presence (solid line) of 30% TFE.

Fig. 3-1E shows the near-UV CD spectra of R2R3* at pH ranging from 4.0 to 7.0. The results indicate that though the tertiary structure is affected by pH changes, native structure is maintained, and differs from the unfolded or molten globule structure. On comparison of respective spectra with that at pH 7.0, the variations at wavelengths of 250 to 260 nm, 270 to 300 nm were clearly observed (Fig. 3-1F). Considering that the variations around 280 nm is mainly due to a change in Trp, the local structure around Trp residues located in the hydrophobic cores of R2 and/or R3 would be perturbed by pH changes. Similarly, since the variations around 260 nm were also observed in the mutant proteins in which the residues in the hydrophobic core of R2 were substituted (Oda *et al*., 1997b), the structure around the hydrophobic core could be changed by decreasing pH. On the other hands, tertiary structures were affected by the in addition TFE, suggesting the disordered or molten globule structure (Figs. 3-2C and 2D).

*Effects of pH on the thermal stability of R2R3**

The thermal stabilities of R2R3* at different pHs were analyzed using CD, and thermodynamic parameters for thermal unfolding were determined. The thermal unfolding process was quantitatively reversible under the pH conditions analyzed, as described at Chapter 2 (Fig. 2-2B). Fig. 3-1D shows the thermal transition curves monitored by CD values at 222 nm, and the values were plotted as a function of temperature. The CD values at 20˚C were different between neutral and acidic pHs, while those at 80˚C were identical, indicating that thermally unfolded structures at all pHs analyzed are similar. The $T_m$ and $\Delta H_{vH}$ were obtained by curve fitting based on the two-state transition model, and are summarized in Table 3-1. The $T_m$ value at pH 7.5 in

20 mM potassium phosphate buffer, 46.9˚C, was lower than that at pH 7.4 in PBS, 48.6˚C (see Chapter 4, Table 4-1). The difference would be due to the ionic strength, which suggests that the hydrophobic interactions can be strengthened, as reported previously (Del Vecchio *et al*., 2008; Beauchamp and Khajehpour, 2012; Blumlein and McManus, 2013). The $T_m$ value was highest at pH 5.0, and was reduced as pH increased and decreased. There seems to be no correlation between the thermal stabilities of R2R3* and the α-helical contents determined by far-UV CD spectra (Fig. 3-1C).

Table 3-1. Thermodynamic parameters for denaturation of R2R3* at different pHs analyzed using CD.

| pH | $T_m$ /˚C | $\Delta H_{vH}$ /kJ mol$^{-1}$ |
|---|---|---|
| 4.0 | 49.1 ± 0.5 | 157 ± 11 |
| 4.5 | 53.3 ± 0.3 | 165 ± 7 |
| 5.0 | 55.6 ± 0.4 | 177 ± 12 |
| 5.5 | 54.1 ± 0.3 | 146 ± 6 |
| 6.0 | 51.0 ± 0.3 | 151 ± 6 |
| 6.5 | 47.7 ± 0.2 | 155 ± 4 |
| 7.0 | 48.4 ± 0.2 | 196 ± 6 |
| 7.5 | 45.0 ± 0.3 | 141 ± 4 |

The error values indicated are derived from fitting of the transition.

In order to determine the thermodynamic parameters for thermal unfolding in further detail, DSC measurements were carried out. Fig. 3-3A show the heat capacity curves of R2R3* from pH 4.0 to 7.0, analyzed at a heating rate of 1˚C min$^{-1}$, and protein concentration of 2.0 mg ml$^{-1}$. The $T_d$, $\Delta H_{cal}$, and $\Delta C_p$ values were determined using the model of two-state transition. The data could be well fitted on the assumption that the ratio of $\Delta H_{cal}/\Delta H_{vH}$ is 1 (Fig. 3-3A), and the results are summarized in Table 3-2. The ratio of $\Delta H_{cal}/\Delta H_{vH}$ was almost 1 under the low ionic conditions, and was increased

with increasing the ionic strength (see Chapter 4), possibly due to the decrease of inter-repeat interactions. The thermodynamic parameters were similar to those obtained from the DSC experiments at the protein concentration of 0.5 mg $ml^{-1}$ (Table 3-3). Within the concentration range, at least, between 0.5 and 2.0 mg $ml^{-1}$ analyzed for both CD structural and DSC thermodynamic analyses, there is little concentration dependence, and thus the two-state model would be appropriate to determine the thermodynamic parameters. The $T_d$ value of R2R3* was maximal at pH 5.0 and 56.5˚C, similar to the $T_m$ determined using CD. The $\Delta H_{cal}$ and $\Delta C_p$ values were maximal at pH 6.5; 225 kJ $mol^{-1}$ and 4.93 kJ $mol^{-1}$ $K^{-1}$, respectively. The $\Delta C_p$ values at acidic pHs were lesser than those at neutral pHs, indicating that the heat capacities in folded state at acidic pH are similar to those in thermal unfolded state. Considering the results of far-UV CD spectra at acidic pH, the conformational fluctuations would become larger, and subsequently, that the heat capacity in folded state would become higher than that at neutral pH. Figs. 3-3B, 3C, and 3D show the plots of $T_d$, $\Delta H_{cal}$, and $\Delta C_p$ values as a function of pH, respectively. The $T_d$ values were changed linearly as the pH increased from 5.0 to 7.5, and so were the $T_m$ values from CD. In contrast, the $\Delta H_{cal}$ increased in apparent two apparent phases, from pH 4.0 to 5.5 and from pH 5.5 to 6.5, followed by decrease, till pH 7.5.

Table 3-2. Thermodynamic parameters for denaturation of R2R3* at different pHs analyzed using DSC (Protein concentration of 2.0 mg $ml^{-1}$).

| pH | $T_d$ /°C | $\Delta H_{cal}$ /kJ $mol^{-1}$ | $\Delta C_p$ /kJ $mol^{-1}$ $K^{-1}$ |
|---|---|---|---|
| 4.0 | 51.9 ± 0.4 | 146 ± 6 | 2.03 ± 0.18 |
| 4.5 | 55.6 ± 0.2 | 147 ± 9 | 1.26 ± 0.39 |
| 5.0 | 56.5 ± 0.4 | 156 ± 2 | 1.27 ± 0.91 |
| 5.5 | 54.3 ± 0.4 | 161 ± 8 | 0.78 ± 0.70 |
| 6.0 | 53.2 ± 0.2 | 192 ± 14 | 3.70 ± 0.88 |
| 6.5 | 51.5 ± 0.2 | 225 ± 3 | 4.93 ± 0.73 |
| 7.0 | 50.4 ± 0.2 | 214 ± 11 | 4.61 ± 0.53 |
| 7.5 | 49.6 ± 0.3 | 188 ± 20 | 3.44 ± 1.19 |

Averaged values from the three experiments with standard errors.

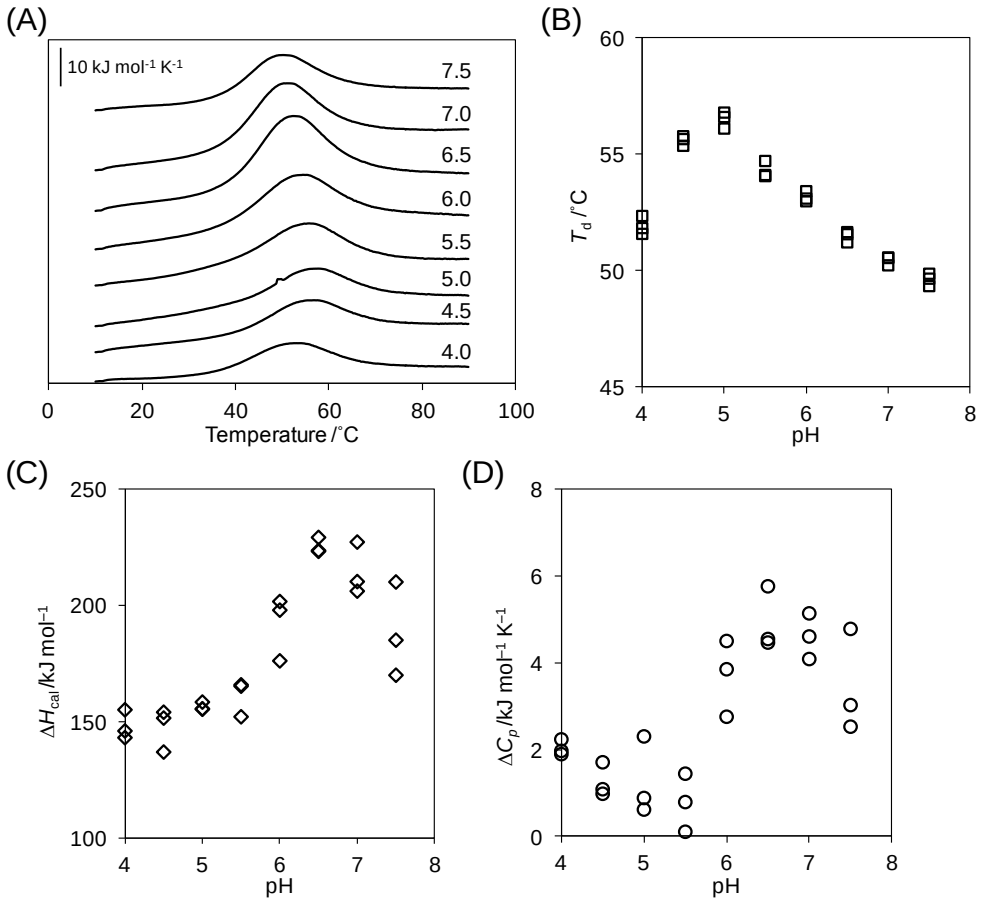



Fig. 3-3. Thermal denaturation analysis by DSC. (A) The heat capacity curves of R2R3* at pH 4.0, 4.5, 5.0, 5.5, 6.0 6.5, 7.0 and 7.5. Plots of $T_d$ (B), $\Delta H_{cal}$ (C)

and $\Delta C_p$ (D) as a function of pH.

Table 3-3. Thermodynamic parameters for denaturation of R2R3* at different pHs analyzed using DSC (Protein concentration of 0.5 mg $ml^{-1}$).

| pH | $T_d$ /°C | $\Delta H_{cal}$ /kJ $mol^{-1}$ | $\Delta C_p$ /kJ $mol^{-1}$ $K^{-1}$ |
|---|---|---|---|
| 4.0 | 51.5 | 155 | 1.97 |
| 5.0 | 56.7 | 159 | 2.31 |
| 6.0 | 53.4 | 193 | 4.50 |
| 7.0 | 50.6 | 210 | 5.14 |

## Discussion

The secondary structure analysis at different pHs showed that the α-helical content of R2R3* was largely perturbed by changing pH and significantly decreased at acidic pHs. Although the secondary structure of R2R3* at acidic pH was partially disordered, the near-UV CD spectra indicated that the tertiary structure was maintained and was different from the unfolded or molten globule structure (Figs. 3-1C and 3D). Indeed, the DNA binding ability at pH 4.0 was clearly observed by using isothermal titration calorimetry (data not shown). Similar to other helical proteins (Arunkumar *et al*., 1997; Okada *et al*., 2003; Banerjee and Kishore; 2005; Dave *et al*., 2011), the α-helical content of R2R3* was increased on addition of TFE. In the presence of 30% TFE at pH 4.0, the near-UV CD spectrum of R2R3* and its thermal transition was indicative of the typical molten globule structure, despite the increased α-helical content (Fig. 3-2A). Taken together, the functional structure of R2R3* can be retained at all pH conditions analyzed, but is largely perturbed by solution pH. This is also supported by the previous structural and physical experiments, including temperature and pressure perturbation experiments (Sarai *et al*., 1993; Morii *et al*., 1999; Puthenpurackal *et al*., 2012; Lassalle *et al*., 2001; see Chapter 2). With an increase in temperature (up to 40˚C), the fluctuations became larger and the α-helical content became smaller, as shown in

the previous NMR and the present CD experiments (Fig. 3-1D). To detect the pH-dependent structural changes of R2R3* more clearly, the present structural analysis was carried out at 20˚C. The structural properties of R2R3* are very unique and can be altered within the framework of the native structure, which enables analysis of the structure and folding thermodynamics correlation, as discussed below. As shown by the previous NMR experiments (Ogata *et al*., 1996; Puthenpurackal *et al*., 2012), the R2 region of R2R3* fluctuates more than the R3 region, and would be mainly perturbed by pH, temperature, and pressure, probably due to the cavity in the hydrophobic core of R2. This is supported by the recent NMR experiments that the HSQC signals were broadened and disappeared as pH decreased, and the degree was larger in the residues of R2 in comparison with those of R3 (data not shown). In the protein architecture, the "soft" region, R2, would be connected to the "rigid" region, R3, by the short linker. This would be advantageous for its function, by balancing the stability under physiological conditions.

Two different buffers were used to analyze folding thermodynamics in correlation with structural properties in the range of neutral and acidic pHs. Both buffers, sodium acetate (from pH 4.0 to 5.5) and potassium phosphate (from pH 6.0 to 7.5) have small enthalpy for proton dissociation, as reported previously (Fukada and Takahashi; 1998). Thus, the solution pH was changed minimally during the experimental temperature, and the enthalpy changes for protonation/deprotonation were small. In order to analyze the thermodynamic origin of stability differences, the thermodynamic parameters at 50.4˚C, the denaturation temperature of R2R3* at pH 7.0, were calculated using the $\Delta C_p$ values of respective pHs, as described previously (Saito *et al*., 2003). Within the narrow range of temperature around the $T_d$ values, the error of the calculated

$\Delta G$ and $\Delta H$ values at the reference temperature should be small, even if the $\Delta C_p$ values obtained have errors. The calculated parameters are summarized at Table 3-3, and Figs. 3-4A, 4B, and 4C shows the plots of these parameters as a function of pH. Similar to the $T_m$ from CD and $T_d$ from DSC, the $\Delta G$ value of R2R3* was maximum at pH 5.0 (Fig. 3-4A), indicating that the thermal stability at pH 5.0 is higher than those at other pHs. In contrast, the $\Delta H$ and $\Delta S$ values of R2R3* were maximum at pH 6.5 (Figs. 3-4B and 4C).

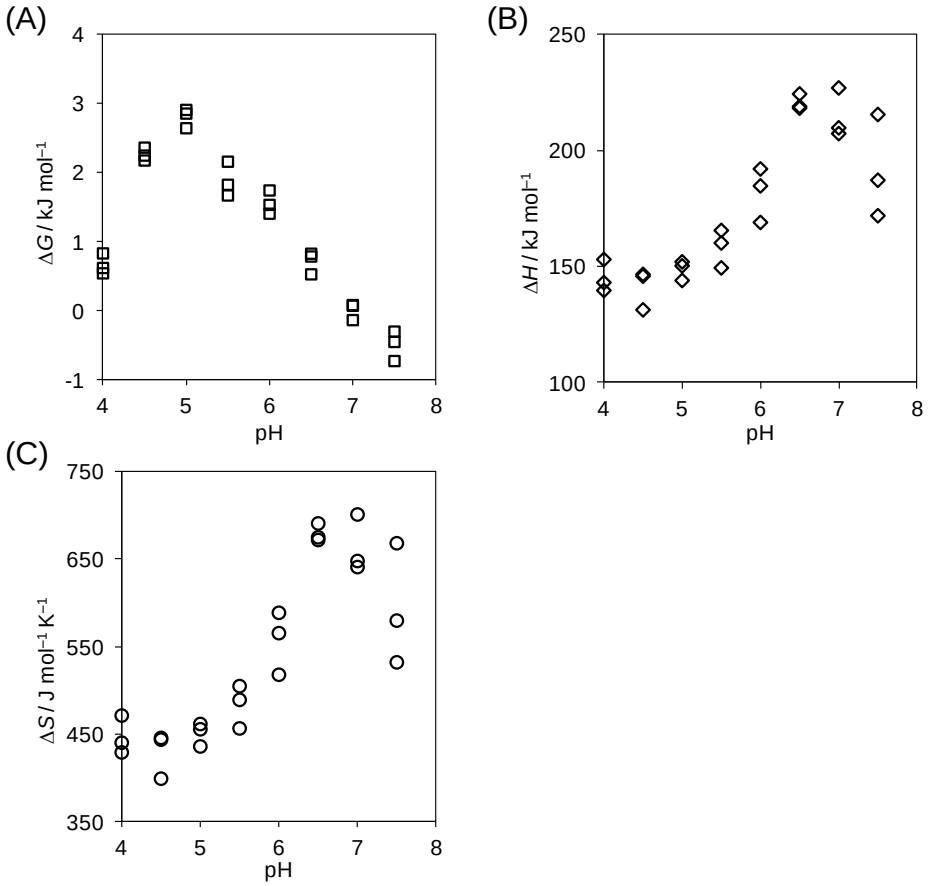



Fig. 3-4. Plots of thermodynamic parameters, $\Delta G$ (A), $\Delta H$ (B), and $\Delta S$ (C), at 50.4˚C as a function of pH.

In order to assess the correlation of folding thermodynamics with the secondary structure of R2R3*, the α-helical contents analyzed using CD were plotted as a function of thermodynamic parameters listed in Table 3-4 (Figs. 3-5A and 5B). While no specific correlation was observed for Δ*G*, a linear correlation was observed for Δ*H* in the pH range between 4.5 and 7.5 (Fig. 3-5B). The results suggest that R2R3* at pH 5.0 is enthalpically destabilized, but entropically stabilized in the native state, resulting in maximal thermal stability in the pH range analyzed. In contrast, R2R3* at pH 6.5 is enthalpically stabilized but entropically destabilized in comparison with R2R3* at pH 5.0. It is reasonable to draw correlation well between Δ*H* and α-helical content, because intramolecular interactions in the folded state would contribute to Δ*H* favorably and Δ*S* unfavorably, resulting in a stabilized structure with reduced fluctuations. Some groups, using model proteins or peptides, have reported that the major effects of pH and ionic strength in solution on thermal stability would be due to intramolecular salt bridges and hydrogen bonds (Scholtz *et al*., 1993; Huyghues-Despointes and Baldwin, 1997; Phelan *et al*., 2002; Pace *et al*., 2009; Robinson *et al*., 2014). Table 3-5 summarizes the side-chain interactions of Asp and Glu residues in R2R3* which can be affected by pH. At neutral pH, owing to the different ionization states of His, Δ*H* and α-helical content would be changed. R2R3* has five His residues, and two of them form ion pairs (Tahirov *et al*., 2002; PDB: 1GV2, Fig. 3-6). The p$K_a$ value of His side chain is 6.0 in the unfolded state, and could shift to a higher value in the folded state, due to the interaction with negatively charged residues (Anderson *et al*., 1990; Huyghues-Despointes and Baldwin, 1997; Hendsch and Tidor 1994; Pace *et al*., 2009; Yang *et al*., 2012). At acidic pH, the ionization states of Asp and Glu side chains, in addition to those of carbonic acids of main chains, can be changed, leading to change in

the conformational and thermal stabilities. As reported previously (Pace *et al*., 2009), when the carbonic groups form individual ion pairs within 3.0 Å, their p$K_a$ values would be different between folded and unfolded states. Considering that the p$K_a$ of Glu side chain as 4.3, the ionization state of Glu in the unfolded state at pH 4.0 is different from those above pH 4.5. The protonation/deprotonation of carbonic acid should have an effect on $\Delta H$, but this would be relative small for the $\Delta H$ values determined in this study (Fukada and Takahashi, 1998). Therefore, the structural differences in the native state would mainly lead to variations in the folding thermodynamics.

Table 3-4. Thermodynamic parameters for denaturation of R2R3* at different pHs at denaturation temperature of pH 7.0 (50.4°C).

| pH | $T_d$ [a] /°C | $\Delta T_d$ [b] | $\Delta G$ [c] /kJ mol$^{-1}$ | $\Delta H$ [c] /kJ mol$^{-1}$ | $\Delta S$ [c] /J mol$^{-1}$ K$^{-1}$ |
|---|---|---|---|---|---|
| 4.0 | 51.9 ± 0.4 | 1.5 ± 0.4 | 0.66 ± 0.15 | 145 ± 7 | 446 ± 22 |
| 4.5 | 55.5 ± 0.2 | 5.2 ± 0.2 | 2.3 ± 0.09 | 141 ± 9 | 428 ± 26 |
| 5.0 | 56.5 ± 0.4 | 6.1 ± 0.4 | 2.8 ± 0.14 | 149 ± 4 | 451 ± 13 |
| 5.5 | 54.3 ± 0.4 | 3.8 ± 0.4 | 1.9 ± 0.25 | 158 ± 8 | 483 ± 25 |
| 6.0 | 53.2 ± 0.2 | 2.7 ± 0.2 | 1.6 ± 0.17 | 182 ± 12 | 557 ± 36 |
| 6.5 | 51.5 ± 0.2 | 1.0 ± 0.2 | 0.71 ± 0.16 | 221 ± 3 | 680 ± 10 |
| 7.0 | 50.4 ± 0.2 | 0.0 ± 0.2 | 0.0 ± 0.1 | 215 ± 11 | 663 ± 33 |
| 7.5 | 49.6 ± 0.3 | −0.8 ± 0.3 | −0.50 ± 0.22 | 191 ± 22 | 593 ± 69 |

[a] The data were taken from Table 3-2.

[b] $\Delta T_d$ values were calculated from the equation, $\Delta T_d = T_d$ (at pH x.x) − $T_d$ (at pH 7.0)

[c] The values were calculated using the $\Delta C_p$ values.

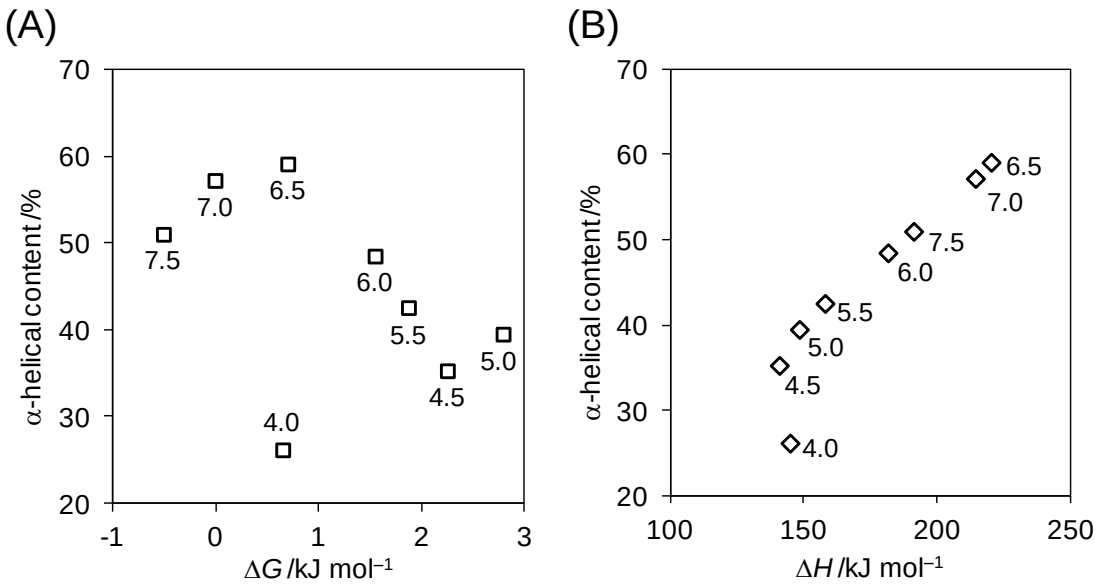



Fig. 3-5. The correlation of α-helical contents with thermodynamic parameters, Δ*G* (A) and Δ*H* (B) at 50.4˚C.

Table 3-5. Intramolecular interactions of acidic residues in R2R3.

| | NMR (complex)[a] | | X-ray (free)[b] | |
|---|---|---|---|---|
| Acidic | Basic | Distance /Å[c] | Basic | Distance /Å[c] |
| Glu-98 | Lys-97 | 2.26 | | |
| | Arg-102 | 2.70 | Arg-102 | 2.97 |
| Glu-99 | Arg-125 | 2.80 | Arg-125 | 2.74 |
| Asp-100 | Arg-133 | 2.81 | Arg-133 | 2.80 |
| Glu-132 | Arg-131 | 2.69 | Arg-131 | 3.38 |
| Glu-141 | | | | |
| Glu-149 | Arg-153 | 2.74 | | |
| Glu-150 | Arg-153 | 2.68 | | |
| Glu-151 | Arg-176 | 2.65 | Arg-176 | 3.12 |
| Asp-152 | | | His-184 | 2.66 |
| Glu-168 | Arg-165 | 2.72 | Lys-171 | 2.97 |
| Asp-178 | Arg-131 | 2.88 | His-137 | 2.92 |

[a] PDB code; 1MSE (Solution NMR structure of c-Myb R2R3 in the presence of DNA).

[b] PDB code; 1GV2 (X-ray crystal structure of c-Myb R2R3 in the absence of DNA).

[c] Distance between heavy atoms, O and N.

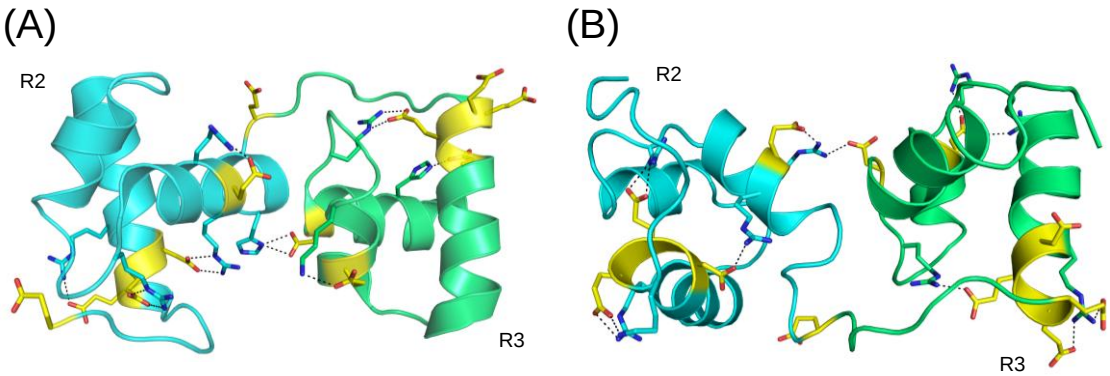



Fig. 3-6. The Coulomb (electrostatic) interaction between Asp/Glu and basic residues. (A) DNA-free form (PDB code; 1GV2), and (B) DNA-bound form (PDB code; 1MSE). The side-chain of Asp/Glu and the residues of basic amino acid whose interact to those acidic animo acid show stick lines. The dash lines indicate that the carbonic groups form individual ion pairs within 3.0 Å (see Table 3-5).

The loss or reduction of intramolecular interactions has favorable effects on conformational flexibilities of proteins, resulting in favorable effects on entropy terms of folding thermodynamics. In this study, I found close correlation between α-helical content and $\Delta H$, but not $\Delta G$. Since the α-helical content of R2R3* is variable under different pHs, the relation could be determined clearly. This is not just specific to R2R3*, and could be applied for other proteins as well. All proteins can functionally evolve for certain amount of time, before obtaining a specific structure. To a given extent, the structural architecture might be tolerant to changes in the environment, thus retaining the functionality. In case of c-Myb DNA-binding protein, under physiological conditions, around pH 7.4, some levels of fluctuations would be required to express its function, resulting in that c-Myb R2R3 exists in the enthalpically unstable but entropically stable state. Even at pH 4.0, R2R3* can form the native structure with larger fluctuations, that might be the outcome of structural tolerance. The extrapolated

$\Delta H$ value at 50.4˚C (approximately 140 kJ $mol^{-1}$) might be low to maintain the folded structure of R2R3*.

# Chapter 4 Thermodynamic effects multiple conformations of R3 on stability and DNA binding of R2R3

## Introduction

Proteins fluctuate in solution, and this dynamic behavior is critical for protein function (Wolynes *et al*., 1995; Tsai *et al*., 1999; Oda *et al*., 2010). Although protein structures have been determined at atomic resolution, most of them are stable structures that can be crystallized, and the averaged NMR structures that can be influenced by the major population among the various conformations in solution. In order to understand the real view of a protein and its function, the dynamic behavior should be added to the X-ray or NMR structure. A protein in the native state has multiple conformations in solution in general, including the most stable and low-lying excited ones, both of which are in thermodynamic equilibrium, although the minor populated conformation is difficult to detect. The degree of protein fluctuation among the multiple conformations can be determined by thermodynamic analysis as the conformational entropy. The protein fluctuation can also be analyzed by using NMR methods such as spin relaxation and $^{1}H/^{2}H$-exchange experiments (Palmer, 2004; Kay, 2005).

The side-chain conformations of amino acids in the hydrophobic core are important for protein folding, design, and function (Shank *et al*., 2010; Doyle *et al*., 2015; Brunette *et al*., 2015). Three Ile residues, namely, Ile-155, Ile-169, and Ile-181, are also involved in the hydrophobic core of R3 (Figs. 4-1A and 1B, Ogata *et al*., 1992). A previous NMR study has shown that a mutant protein of c-Myb R3, I155L/I181L, has multiple conformations, attributed to increased side-chain conformational entropy due to the mutation of Ile to Leu (Furukawa *et al*., 1996). The two distinguishable conformations detected by NMR made it possible to analyze the effect of protein

dynamic behavior on structure and function, especially in the view of conformational entropy. In the NMR analysis of R3 Ile mutant proteins, the unique structure of the I155L mutant could be determined, showing that the overall structure is similar to that of the wild type, while that of the I155L/I181L mutant could not, because many minor peaks appeared in the two-dimensional $^{1}H/^{15}N$ heteronuclear multiple quantum coherence spectroscopy spectra, due to the multiple conformations (Furukawa *et al*., 1996). The lower protection factors of the I155L/I181L mutant determined by the NMR $^{1}H/^{2}H$-exchange experiments supported the notion that protein fluctuation, at least between the two conformations, would be increased upon the I181L mutation (Furukawa *et al*., 1996). Although the previous study showed the effects of conformational entropy only qualitatively, it could not identify the quantitative effects on the total free-energy balance on the protein stability. In addition, since R3 alone was used, it was still unclear whether the mutation affected the function, i.e., DNA binding.

In this study, I analyzed the thermodynamic effects of multiple conformations on stability and DNA-binding function using the I155L/I181L R2R3 mutant and its control proteins. Considering that conformational entropy increases upon the I181L mutation, the $\Delta S$ value of folding thermodynamics, F (folded state) ↔ U (unfolded state), becomes smaller, and that of binding thermodynamics, protein (DNA-free state) + DNA ↔ complex (DNA-bound state), becomes smaller, assuming that the conformational entropies in the unfolded and the DNA-bound states are unchanged upon the mutation. The increased conformational fluctuation upon the I181L mutation in R2R3 at neutral pH could be observed by the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) NMR experiments, which detects $^{1}H^{N}$ exchange rates with the solvent $^{1}H$ (Hwang *et al*., 1998).

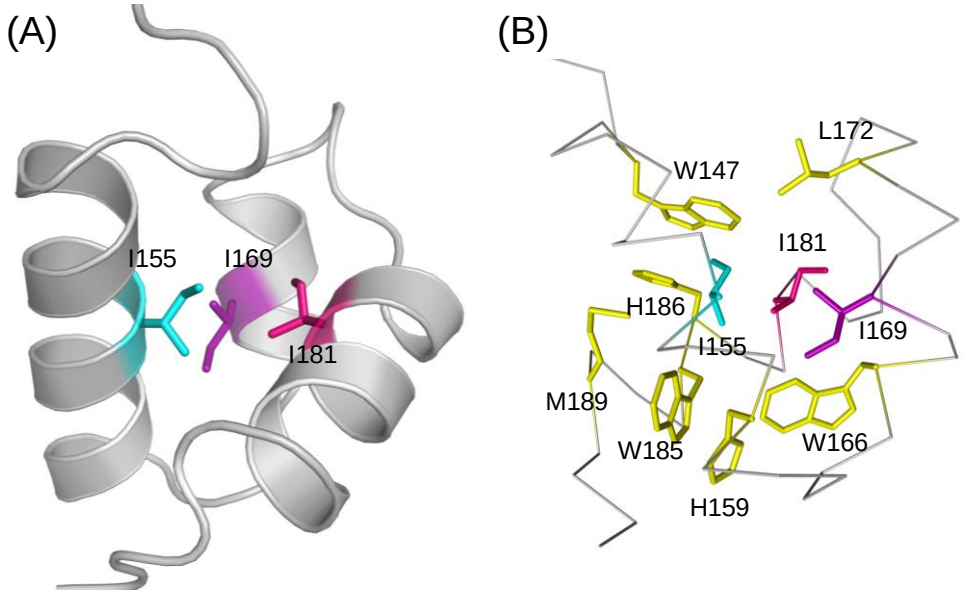



Fig. 4-1. The three-dimensional structure of R3 (PDB code; 1IDZ). (A) Cartoon model. The side-chains of Ile-155 (cyan), Ile-169 (purple), and Ile-181 (pink) are indicated by stick lines with the one-letter amino acid codes. (B) Stick model. In addition of three Ile residues, the side-chains of Trp-147, His-159, Trp-166, Leu-172, Trp-185, His-186, and Met-189 are indicated by yellow stick lines with the one-letter amino acid codes. Those residues are involved in the hydrophobic core.

## Materials and methods

*Preparation of proteins and DNA*

The expression and purification methods of R2R3 proteins were described at Chapter 2. For the NMR experiment, the $^{15}$N-uniformly labeled R2R3* and I155L/I181L were over-expressed in *E. coli* in the M9 minimum medium containing 1.0 g $L^{-1}$ $^{15}NH_4Cl$ and 4.0 g $L^{-1}$ glucose. The purity of each protein was determined to be about 95% by SDS/PAGE analysis. In this study, the C130I protein is denoted as R2R3*, and C130I/I155L and C130I/I155L/I181L are simply denoted as I155L and I155L/I181L, respectively.

The 22-mer oligonucleotide, CACCCTAACTGACACACATTCT, was synthesized by Operon Biotechnologies, Inc., and purified by high performance liquid chromatography with a $C_{18}$ reversed-phase column. The purity was determined to be about 95%. The complementary strands were annealed in PBS, generating the cognate DNA, MBS-I. The DNA concentrations were determined from UV absorption at 260 nm using the molar absorption coefficient of $1.3 \times 10^4$ M basepair$^{-1}$ cm$^{-1}$.

*CD measurements*

Far-UV (200-250 nm) and near-UV (250-320 nm) CD spectra were recorded on a Jasco J-820 spectropolarimeter at 20ºC equipped with Peltier-type temperature control system, as described at Chapter 2. The spectra were obtained in PBS (pH 7.4), using quartz cell with 1.0 cm path-length. The protein concentrations were 0.02 mg ml$^{-1}$ and 0.2 mg ml$^{-1}$ for far- and near-UV respectively. CD spectra were obtained using scanning speed of 20 nm min$^{-1}$, a time response of 1 s, a bandwidth of 1 nm, and an average over 4 or 8 scans, for the far- and near-UV ranges, respectively. The melting curves were recorded in temperature mode at 222 nm, from 20 to 85ºC with a heating rate of 1.0ºC min$^{-1}$. The analysis of the transition curves obtained by temperature-scanning CD measurements was performed on the basis of two-state transition model, as described at Chapter 2. The curve fitting was carried out by the non-linear least-squares method on Origin 5.0 software.

*DSC measurements*

DSC experiments were carried out on a Nano-DSC II calorimeter (TA instruments), as described at Chapter 3. The data were collected by heating the solution from 5 to 80ºC at a rate of 1ºC min$^{-1}$. The sample was reheated without exchanging the

solution in the cells to check the reversibility. The outer buffer solution recovered from final dialysis experiment was used in the reference cell for each case. The protein concentrations were 0.5 mg $ml^{-1}$ in PBS buffer (pH 7.4). The data were analyzed using CpCalc software and fitting analysis was performed using Origin software, as described at Chapter 3.

*NMR measurements*

The CLEANEX-PM fast HSQC NMR experiments were performed on Bruker AVANCE III 950 spectrometer at a proton resonance frequency of 950.33 MHz equipped with a z-axis gradient and triple-resonance TCI cryogenic probe at 293K in 20 mM potassium phosphate containing 20 mM NaCl, 1 mM $NaN_3$, and 10% $^2H_2O$ (pH 7.4). The mixing times ($\tau_m$) were set at 0.21, 2.94, 5.88, 8.82, 11.8, 14.7, 17.6, 20.6, 23.5, 26.5, and 29.8 ms. The sampled complex point numbers and acquired spectral widths were 128 and 1,832 Hz and 1,536 and 18,939 Hz for the $^{15}N$ and $^1H$ dimensions, respectively. For each free induction decay (FID) 16 scans were accumulated. The protein concentrations were 0.8 and 1.0 mM for R2R3* and I155L/I181L, respectively. The HSQC crosspeak signals were partially assigned using the information reported previously (Ogata *et al*., 1996). The water amide hydrogen exchange rates, $k_{ex}$, were calculated from the following equation, as described previously (Hwang *et al*., 1998).

$$I/I_0 = k_{ex}/ (R_{1A} + k_{ex} - R_{1B}) \times [\exp (-R_{1B}\, \tau_m) - \exp \{- (R_{1A} + k_{ex})\, \tau_m \}] \quad \text{(Eq. 4-1)}$$

$R_{1A}$ and $R_{1B}$ are the amide hydrogen relaxation rate (combination of longitudinal and transverse relaxation rates), and the solvent relaxation rate (0.5 $s^{-1}$ was used in this experiments), respectively. $I$ and $I_0$ are the peak intensities in the CLEANEX-PM and the reference spectra, respectively. The fraction factor of water that recovers during the

repetition delay (2s) was assumed as unity because it showed a negligible difference from the factor measured at a repetition delay of 30s. The data were processed and analyzed using NMRpipe (Delaglio *et al*., 1995) and Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). The fitting analysis was performed on Excel's Solver. Uncertainty of each $k_{ex}$ value was estimated by means of Monte-Carlo simulations (300 times) with the reduced $\chi^2$ value obtained as the result of the corresponding optimization.

*ITC measurements*

ITC experiments were performed on $iTC_{200}$ (GE healthcare). The DNA solution (110 μM) was titrated into the protein solution (10 μM) using 40 μl syringe. Each titration consisted of a preliminary 0.5 μl injection followed by 19 subsequent 2 μl additions. The heat for each injection was subtracted from the heat of dilution of the injectant, which was measured by injecting the DNA solution into the experimental buffer. Each corrected heat was divided by the mole of DNA injected and was analyzed on the basis of "One Set of Sites" model with the Microcal Origin 5.0 software supplied by the manufacturer.

## Results

R2R3*, I155L, and I155L/I181L were over-expressed in *E. coli* and purified. The purity of each protein was determined to be about 95% by SDS/PAGE analysis. The far-UV CD spectra of purified I155L and I155L/I181L are similar to that of R2R3*, which shows typical α-helical structures (Fig. 4-2A). In contrast, the near-UV CD spectra of I155L and I155L/I181L are different from that of R2R3* (Fig. 4-2B),

possibly due to the conformational change of Trp residues, located close to the mutated Ile residues in the hydrophobic core (Fig. 4-1), induced by the Ile mutation.

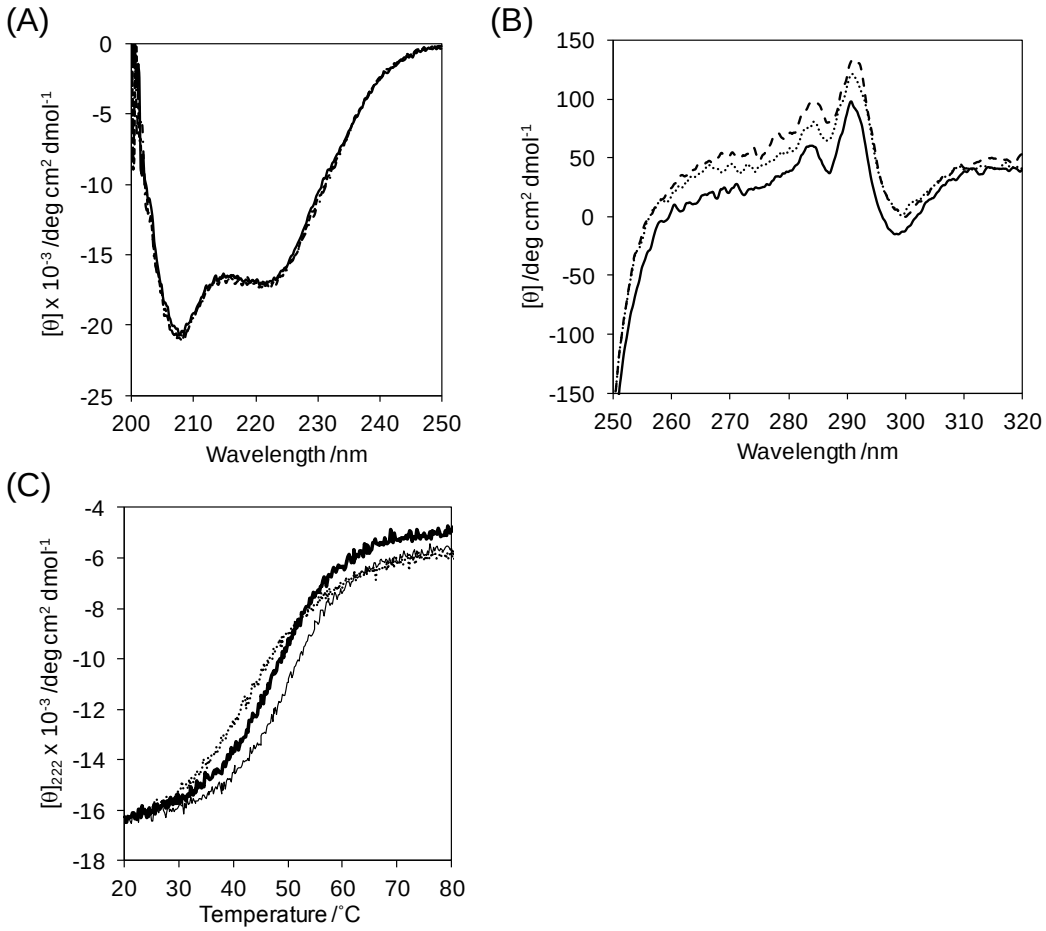



Fig. 4-2. (A) The far- and (B) near-UV spectra of R2R3* (solid line), I155L (dotted line), and I155L/I181L (dashed line) at 20ºC. (C) The thermal denaturation curves of R2R3* (solid thick line), I155L (solid thin line), and I155L/I181L (dotted line).

The thermal stabilities of R2R3* and its mutants were first analyzed using CD. The thermal denaturation was observed by monitoring the CD value at 222 nm, and each fraction unfolded was plotted as a function of the temperature (Fig. 4-2C). The $T_m$ and $\Delta H_{vH}$ values were obtained by the curve fitting on the basis of two-state transition

model, as described in Materials and methods, and are summarized in Table 4-1. The $T_m$ value of I155L is slightly higher than that of R2R3*, while that of I155L/I181L is much lower than those of R2R3* and I155L, indicating that the mutation of I181L destabilizes the protein. This is also supported by the results that the yield of purified I181L R2R3 was much lower than those of R2R3*, I155L, and I155L/I181L (data not shown), and expression in *E. coli* of I181L R3 was low, as reported previously (Furukawa *et al.*, 1996).

Table 4-1. Thermodynamic parameters for denaturation of R2R3* and its mutants analyzed using CD

| | $T_m$ /ºC | $\Delta H_{vH}$ /kJ mol$^{-1}$ |
|---|---|---|
| R2R3* | 48.5 ± 0.1 | 161 ± 3 |
| I155L | 50.8 ± 0.1 | 176 ± 3 |
| I155L/I181L | 42.9 ± 0.4 | 119 ± 3 |

The error values indicated are derived from the fitting of the transition curves.

The thermal stabilities of R2R3* and its mutants were next analyzed using DSC. Fig. 4-3A shows the heat capacity curves of proteins analyzed at a heating rate of 1°C min$^{-1}$. The DSC measurement showed that the thermal unfolding process was quantitatively reversible, although it was not fully reversible (Fig. 4-3B), which was supported to previous CD results (see to Chapter 2). In the fitting analysis using Origin software, two-state model was better than three-state model, and the $T_d$, $\Delta H_{cal}$, $\Delta H_{vH}$, and $\Delta C_p$ values determined are summarized in Table 4-2. Decreased thermal stability was observed in response to the I181L mutation, similar to the CD experiments. The thermodynamic origin is discussed below. The ratio of $\Delta H_{cal}/\Delta H_{vH}$ was clearly different

from 2, indicating that cooperative transition between the two repeats, R2 and R3, occurred.

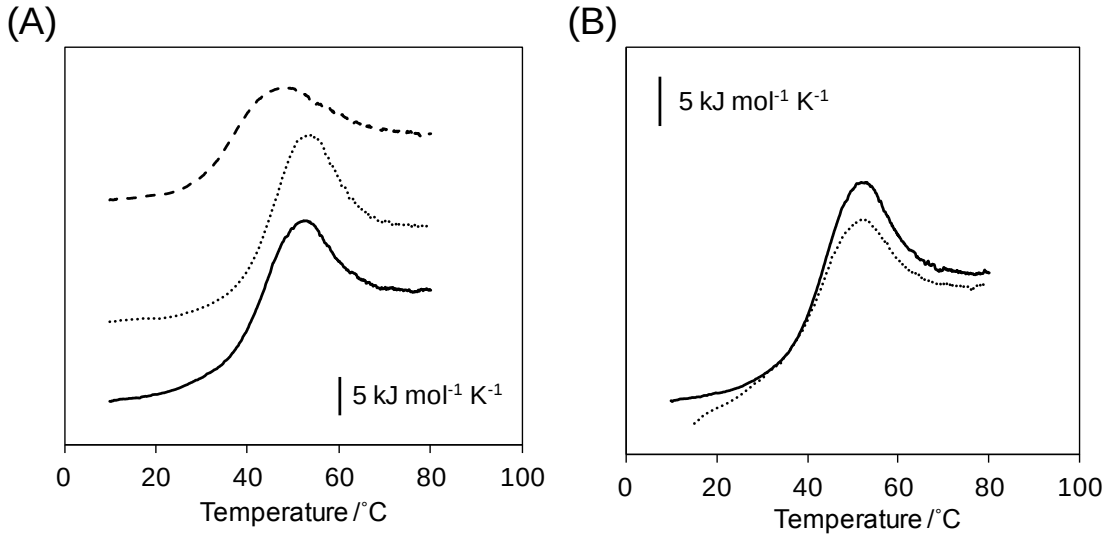



Fig. 4-3. (A) The heat capacity curves of R2R3* (solid line), I155L (dotted line), and I155L/I181L (broken line). (B) The heat capacity curves of R2R3* at the first scan (solid line) and second scan (dotted line).

Table 4-2. Thermodynamic parameters for denaturation of R2R3* and its mutants analyzed using DSC

| | $T_d$ /°C | $\Delta H_{cal}$ /kJ mol$^{-1}$ | $\Delta H_{vH}$ /kJ mol$^{-1}$ | $\Delta H_{cal}/\Delta H_{vH}$ | $\Delta C_p$ /kJ mol$^{-1}$K$^{-1}$ |
|---|---|---|---|---|---|
| R2R3* | 50.3 ± 0.02 | 266 ± 5 | 182 | 1.46 | 8.66 ± 0.10 |
| I155L | 51.2 ± 0.03 | 282 ± 4 | 193 | 1.46 | 12.6 ± 0.19 |
| I155L/I181L | 45.8 ± 0.1 | 188 ± 5 | 154 | 1.22 | 6.25 ± 0.42 |

The error values indicated are derived from the fitting.

The change of conformational fluctuation upon the I181L mutation in R2R3 at neutral pH was analyzed using NMR. Multiple conformations of R3 I155L/I181L mutant was observed in the previous NMR study, while in the case of R2R3

I155L/I181L mutant, any minor peaks are not observed on the HSQC spectra (Fig. 4-4A). First, the $^{1}H/^{2}H$-exchange experiments were carried out. However, even at pH 6.0 and 10ºC, most of HSQC signals were disappeared within 6 min from the start of the measurement (data not shown). Therefore, NMR CLEANEX-PM experiments were carried out, and provided useful information as described below. Some $^{15}N/^{1}H$ HSQC signals of I155L/I181L could be assigned based on the information of signal assignment of wild-type R2R3 reported previously (Ogata *et al*., 1994), and the changes of signal intensities in the CLEANEX-PM spectra were compared with those of R2R3*. The exchange rates between amide and water protons, $k_{ex}$, of Asn-183 and Asn-186 located close to the mutation site in R3 of I155L/I181L were larger than those of R2R3*, while those of Trp-115 and Lys-123 located in R2 were similar for both I155L/I181L and R2R3* (Figs. 4-4B, 4C, 4D, and 4E, Table 4-3). The exchange rates of amide protons with water protons often increase when the corresponding amide groups are more exposed to solvent due to their increased flexibility. The results indicate that the conformational fluctuation of I155L/I181L R2R3, at least those of the residues around the mutation site, is increased upon the mutation, which is in good agreement with the previous NMR results using $^{1}H/^{2}H$-exchange experiments at acidic pH (Furukawa *et al*., 1996).

Table 4-3. The hydrogen exchange rates, $k_{ex}$ ($s^{-1}$), of R2R3* and I155L/I181L

| | Trp115 | Lys123 | Asn183 | Asn186 |
|---|---|---|---|---|
| R2R3* | 7.12 ± 0.13 | 2.36 ± 0.07 | n.d.[a] | n.d.[a] |
| I155L/I181L | 7.12 ± 0.16 | 2.18 ± 0.08 | 2.08 ± 0.08 | 2.43 ± 0.13 |

The error values indicated are derived from the fitting of the plot curves.

[a] Not determined because of the weak peak intensity.

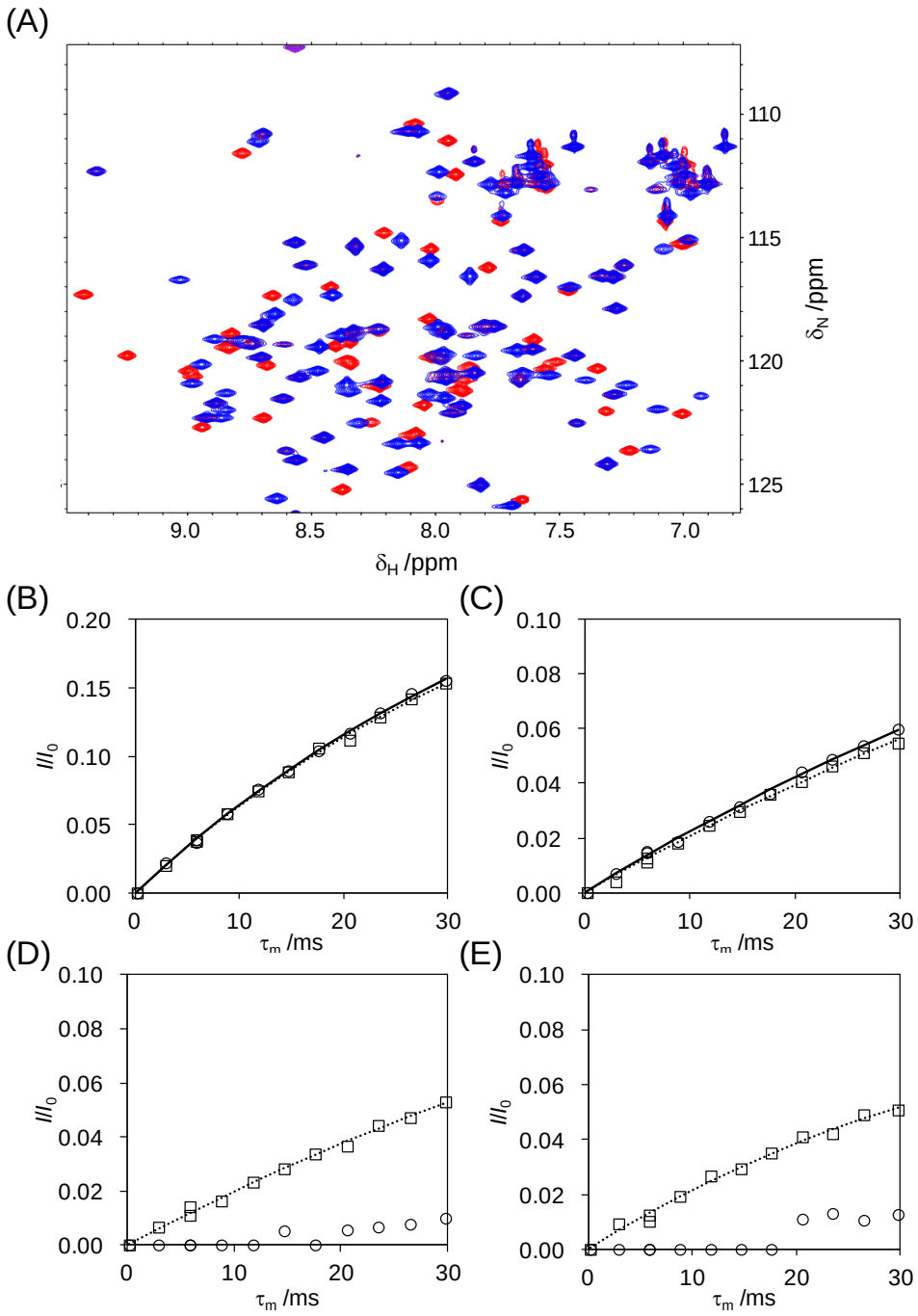



Fig. 4-4. (A) The $^{15}N/^{1}H$ HSQC spectra of R2R3* (red) and I155L/I181L (blue). Plots of $I/I_0$ as a function of $\tau_m$ for the residues (B) Trp-115, (C) Lys-123, (D) Asn-183, and (E) Asn-186 of R2R3* (open circle) and I155L/I181L (open square). The solid and dotted lines represent the fitting curves for the plots of R2R3* and I155L/I181L on the basis of equation (6), respectively.

The DNA binding of R2R3* and its mutants were analyzed using ITC. An exothermic heat pulse was observed after each injection of a cognate DNA fragment containing the Myb-binding site in the simian virus 40 enhancer sequence (MBS-I) into the protein (Fig. 4-5, Nakagoshi *et al*., 1990). Each exothermic peak was integrated, and the heat of dilution of MBS-I was subtracted from the integrated values. The corrected heat was divided by the moles of MBS-I injected, and the resulting values were plotted as a function of the molar ratio (Fig. 4-5). The resultant data were best fit to a model for one binding site using the nonlinear least-squares method. The stoichiometry of binding ($n$), the binding constant ($K_a$), and $\Delta H$ were obtained from the fitted curve. The Gibbs free energy change, $\Delta G$, and the entropy change, $\Delta S$, were calculated from the equation $\Delta G = -RT \ln K_a = \Delta H - T\Delta S$. The thermodynamic parameters determined are summarized in Table 4-4. The $n$ values indicate well the 1:1 complex formation. The thermodynamic parameters of R2R3* are similar to those reported previously (Oda *et al*., 1998). It should be noted that the I181L mutation lowered the DNA binding affinity, owing to the unfavorable entropy change.

Table 4-4. Thermodynamic parameters of specific DNA binding to R2R3* and its mutants at 20°C

| | $n$ | $K_a$ ($\times 10^7$ /$M^{-1}$) | $\Delta G$ /kJ $mol^{-1}$ | $\Delta H$ /kJ $mol^{-1}$ | $-T\Delta S$ /kJ $mol^{-1}$ |
|---|---|---|---|---|---|
| R2R3* | 0.96 ± 0.02 | 3.03 ± 0.30 | −42.0 | −43.3 ± 0.4 | 1.3 |
| I155L | 1.00 ± 0.02 | 2.52 ± 0.29 | −41.6 | −43.5 ± 0.6 | 1.9 |
| I155L/I181L | 1.00 ± 0.02 | 1.67 ± 0.01 | −40.5 | −45.7 ± 1.3 | 5.2 |

Averaged values from the three experiments with standard errors.

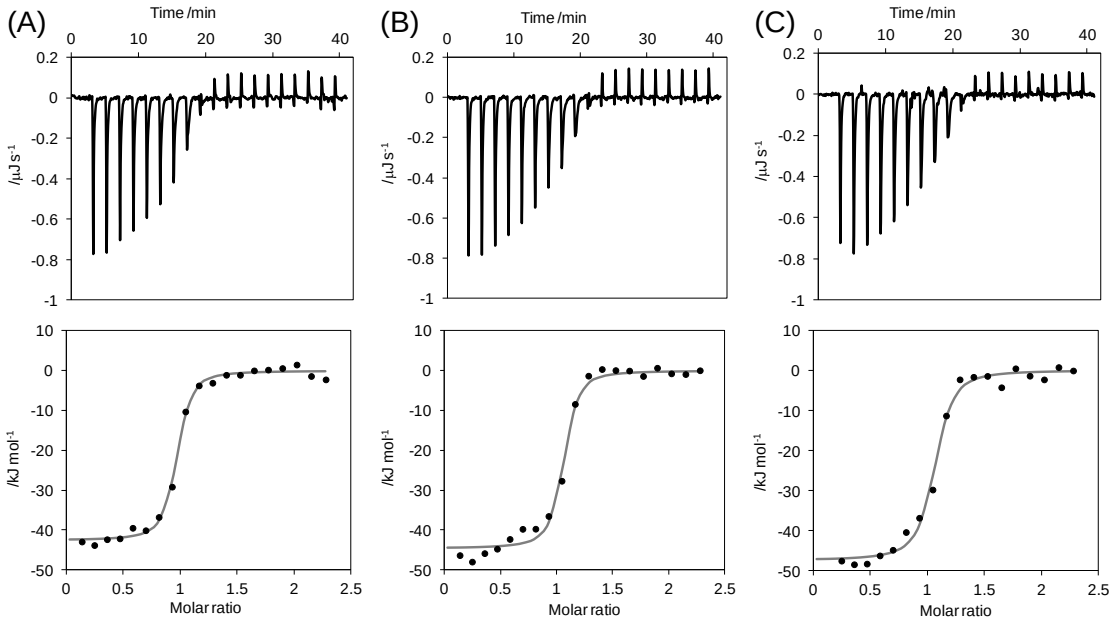



Fig. 4-5. ITC profiles of specific DNA-binding by (A) R2R3*, (B) I155L, and (C) I155L/I181L. (upper) Raw data. (lower) The data points were obtained by integration of the peaks in the raw data, corrected for the dilution heat, and plotted against the molar ratio, MBS-I/protein.

## Discussion

To analyze the thermodynamic origin of the stability difference, the thermodynamic parameters at 50.3°C, the denaturation temperature of R2R3*, were calculated using the $\Delta C_p$ values of respective mutants (Table 4-5). Within the narrow range of temperature around the $T_d$ values, the errors of the calculated $\Delta G$ and $\Delta H$ values at the reference temperature should be small, even if the $\Delta C_p$ values obtained have errors. The results clearly indicate that the decreased stability due to the I181L mutation is because of the enthalpic contribution, which is partially compensated by the favorable entropy change. The favorable entropy change can be explained by the increased entropy of mutant I155L/I181L in the folded state. This can also explain the unfavorable entropy change for DNA binding (Table 4-4). Assuming that the

conformational entropies in the unfolded and DNA-bound states are unchanged upon the I181L mutation, the schematic diagram shown in Fig. 4-6B could be considered. The decreased entropy change for either unfolding or binding was actually observed in the I155L/I181L mutant, which has larger fluctuation in comparison with R2R3* and I155L mutant (Furukawa *et al*., 1996). The decreased entropy change was compensated by the decreased enthalpy change, as a general characteristic of the enthalpy-entropy compensation in either intramolecular or intermolecular interactions of biomolecules (Dunitz, 1995; Fenley *et al*., 2012). Upon the I181L mutation, the degree of decreased enthalpy change for unfolding was larger than that of decreased entropy change, resulting in decreased stability. The decreased enthalpy change upon the I181L mutation should be due to the weakened hydrophobic contacts in the R3 core region in the folded state. Similar to the folding thermodynamics, the more fluctuated character upon the I181L mutation should increase the entropy level in the DNA-free state, and decrease the entropy change for DNA binding. The degree of decreased enthalpy change for DNA binding was smaller than that of decreased entropy change, resulting in decreased binding affinity.

Table 4-5. Thermodynamic parameters for denaturation of R2R3* and its mutants at the denaturation temperature of R2R3* (50.3°C)

| | $T_d$ /°C[a] | $\Delta T_d$[b] | $\Delta G$ /kJ mol$^{-1}$ | $\Delta H$ /kJ mol$^{-1}$[c] | $T\Delta S$ /kJ mol$^{-1}$[c] |
|---|---|---|---|---|---|
| R2R3* | 50.3 ± 0.02 | 0 | 0 | 266 ± 5 | 266 ± 5 |
| I155L | 51.2 ± 0.03 | 0.9 | 0.77 | 271 ± 4 | 270 ± 4 |
| I155L/I181L | 45.8 ± 0.1 | −4.5 | −2.8 | 216 ± 5 | 218 ± 6 |

[a] The data were taken from Table 4-2.

[b] $\Delta T_d$ values were calculated from the equation, $\Delta T_d = T_d$ (mutant) − $T_d$ (R2R3*)

[c] The error values indicated are based on the propagation of uncertainty.

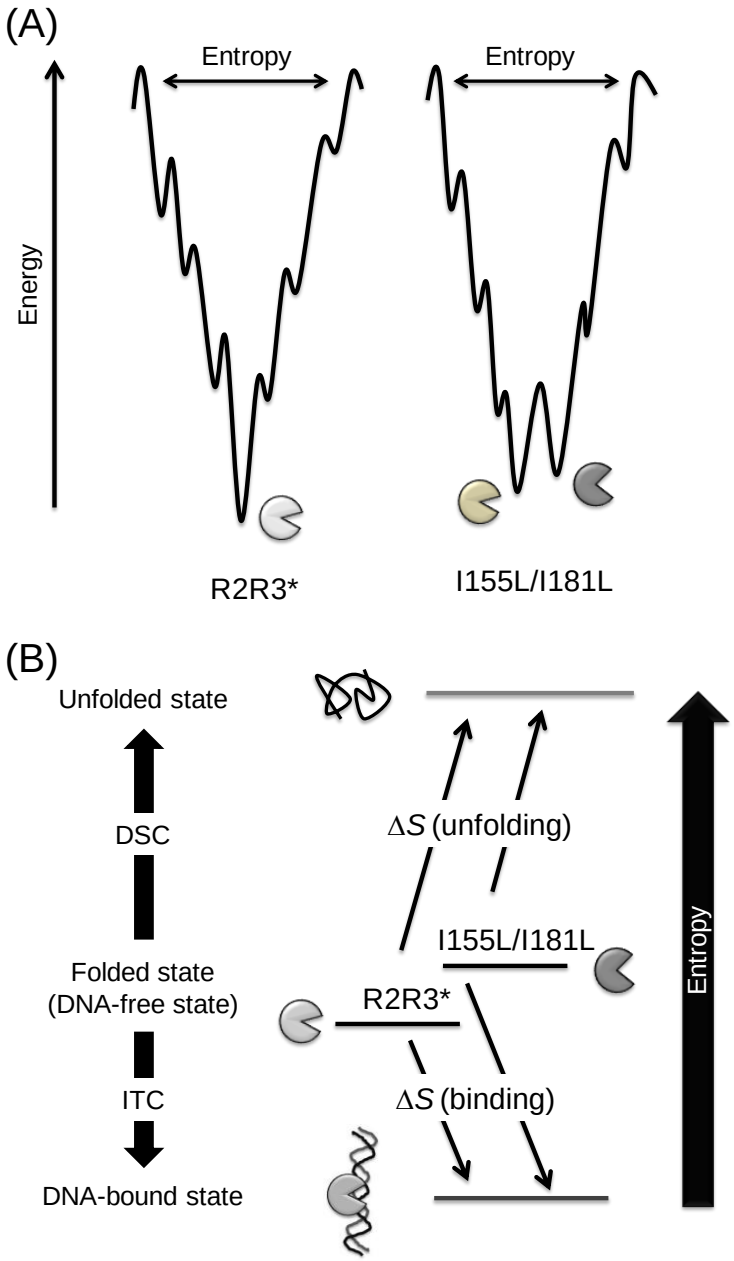



Fig. 4-6. (A) Schematic diagram for folding funnels of wild-type and I155L/I181L mutant. One averaged structure of wild-type was observed, while two or more averaged structures of I155L/I181L mutant were observed in the previous NMR study (Furukawa *et al*., 1996). (B) Schematic diagram representing the entropy levels of folded or DNA-free, DNA-bound, and unfolded states.

Proteins fluctuate in solution, and have various conformations with different energy states, not only the most stable but also the low-lying excited states, which are usually difficult to detect, mainly due to their low population. In general, the population of protein conformations in the most stable state is much larger than in the sub-states, which makes it difficult to correlate the structural dynamics with the conformational thermodynamics. As reported previously, the I155L/I181L R3 mutant has several different conformations with similar low energies, and the structural exchange among the different conformations must be very slow (Furukawa *et al*., 1996; Fig. 4-6A). In addition, NMR $^{1}H/^{2}H$-exchange experiments have shown that the mutant has relatively low protection factors in helices-1 and 2, and the protection almost disappears in helix-3, which is a part of the helix-turn-helix motif and contributes to the specific DNA binding. In this study, in order to elucidate the correlation of structural fluctuation with not only stability but also function, I used the functional unit of c-Myb, R2R3, and analyzed the effect of more fluctuated character of I155L/I181L R2R3 mutant on its stability and DNA-binding at neutral pH. Similar to the previous NMR $^{1}H/^{2}H$-exchange experiments, the increased fluctuation of Asn-183 and Asn-186 on helix-3 of R3 was observed upon the I155L/I181L mutation on R2R3* in the present CLEANEX-PM experiment (Fig. 4-4). The increased protein fluctuation could contribute to the increase in conformational entropy, especially at the site on helix-3 of R3, and could also contribute to the unfavorable entropy change for DNA binding.

There have been some reports on increased protein stability by fluctuation regulation, e.g., by the addition of disulfide bonds and Pro residues (Betz, 1993; Jaenicke, 2000). In most of these studies, the increased stability is due to the decreased entropy of unfolded state. Similar mechanisms could be applied to the design of

increased binding affinity. I actually showed that the Pro residue in the linker region between R2 and R3 of c-Myb contributes to the increased DNA-binding affinity due to the decrease in conformational entropy in the DNA-free state, with little change in the DNA-bound state (Oda *et al*., 1998). The present investigation also showed that the thermodynamics could well evaluate the protein fluctuation for either folding or binding function. Recently, multiple conformations have been reported in various proteins (Kitahara *et al*., 2005; Thickman *et al*., 2007; Naber *et al*., 2011; Wong *et al*., 2012), and the low-lying excited state of c-Myb R2 was also analyzed using high-pressure fluorescence and NMR (Puthenpurackal *et al*., 2012). Among the multiple conformations, a unique conformation in the sub-state should exist, which is critical for the function. Although thermodynamic analysis could not determine the conformation at atomic resolution, it could detect the effect of the minor populated conformer, especially the conformational entropy. In the present investigation, using the I155L/I181L R2R3 mutant, I succeeded to show the effect of increased conformational entropy on not only the folding thermodynamics but also the binding thermodynamics, both of which are well correlated.

# Chapter 5 Thermodynamic effects of a linker region between R2 and R3 on stability and structural dynamics of R2R3

## Introduction

Proteins evolve to suit their function via structural alterations, including amino acid sequence, ternary structure, and fluctuation. Protein function is closely correlated with both fluctuation and static ternary structure, the latter can be determined using several methods, such as X-ray and NMR; however, most of these methods can only be used to determine the most stable or averaged structure. In order to understand how proteins behave in solution and to elucidate structure-activity relationships, it is necessary to analyze the protein structural dynamics (Naganathan *et al*., 2006; Wand, 2013). Because the structural dynamics can alter the entropy term, thermodynamic analysis has widely been used. Recent improvements of the instruments used to detect protein folding and unfolding, especially in their sensitivity, have made it possible to analyze the subtle but the critical structural dynamics. In addition to the effect of protein fluctuation in terms of entropy, the effects such as hydration should also be considered. By comparing the thermodynamic properties of a protein with those of mutant proteins, using site-directed amino acid replacement, the effect of mutation of specific sites on protein fluctuation can be determined.

The last two repeats of c-Myb, R2 and R3, are connected by linker residues: Asn-139, Pro-140, and Glu-141 (Figs. 5-1A and 5B). One of the unique features of R2R3 is that the linker is quite short. In solution, the ternary structures of the specific DNA complex and the free form have been determined using NMR (Ogata *et al*., 1994; Ogata *et al*., 1995). It was observed that upon DNA binding, the orientation of the Trp side-chains in the hydrophobic cores of R2 and R3 gets slightly altered, while the

main-chain conformations remain almost identical. Various mutational analyses were also performed to determine their effects on protein stability and its DNA-binding function (Oda *et al*., 1997a; Oda *et al*., 1997b; Oda *et al*., 1999; Morii *et al*., 1999; see Chapter 4). Mutation of the Pro in the linker region decreased its DNA-binding affinity, due to a decrease in $\Delta S$ (Oda *et al*., 1998). This provides clear evidence of local folding of the linker, which accompanies specific DNA-binding, as observed in other DNA-binding proteins (Spolar and Record; 1994). Because the cyclic structure of the Pro side-chain locks the dihedral angle, φ, at approximately −60°, protein fluctuation decreases. In this study, I analyzed the thermodynamic effects of linker residues, including Pro140, on stability. Considering that conformational entropy increases with mutation of the Pro, the $\Delta S$ value for folding thermodynamics, F (folded state) ↔ U (unfolded state), decreases.

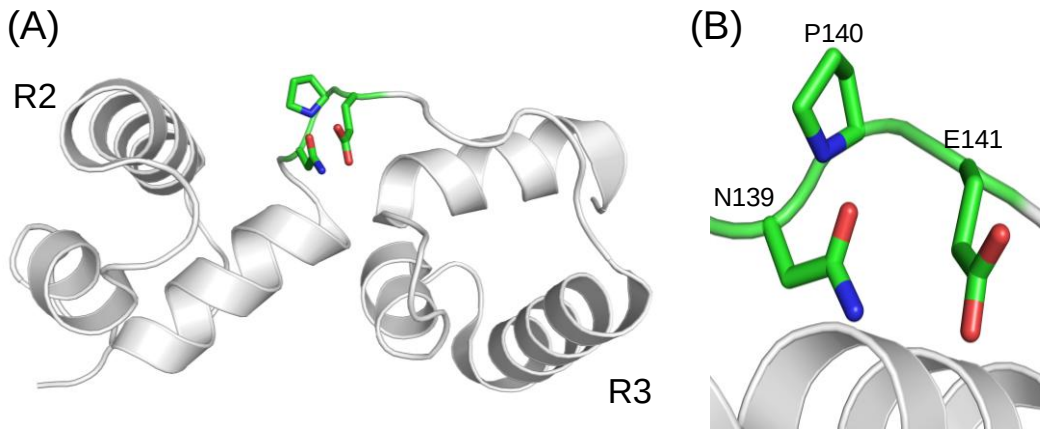



Fig. 5-1. The crystal structure of R2R3 wild-type (PDB code; 1GV2). (A) Over all and (B) linker region models. The side-chains of Asn-139, Pro-140, and Glu-141 are indicated by green stick lines with the one-letter amino acid codes.

## Materials and Methods

*Protein expression and purification*

The expression and purification methods of R2R3 proteins were described at Chapter 2. The purity of each protein was determined to be about 95% by SDS/PAGE analysis. In this study, the C130I protein is denoted as R2R3*, and C130I/P140G and C130I/P140A are simply denoted as P140G and P140A, respectively. In addition, C130I/N139G/P140G/E141G mutant denoted as 3G was used.

*CD measurements*

Far-UV CD spectra were recorded on a Jasco J-820 spectropolarimeter, as described at Chapter 2. The measurements were carried out for the 0.02 mg $ml^{-1}$ proteins in PBS using quartz cell with 1.0 cm path-length. Fitting analysis for thermal denaturation was performed using Origin 5.0 software, as described at Chapter 2.

*DSC measurements*

DSC experiments were carried out on a Nano DSC calorimeter (TA Instruments), as described at Chapter 3. The data were collected by heating the solution from 5 to 80ºC at a rate of 1ºC $min^{-1}$. The protein concentrations were 2.0 mg $ml^{-1}$ in PBS. In the fitting analysis using Origin software, two-state model was better than three-state model, and the denaturation temperature ($T_d$), calorimetric enthalpy change ($\Delta H_{cal}$), and heat capacity change ($\Delta C_p$) values were determined.

*$^1$H-one dimensional NMR measurements*

The one-dimensional $^1$H-NMR measurements of R2R3* and G3 were performed on Bruker AVANCE III 600 spectrometer at a proton resonance frequency of

600.13 MHz equipped with a z-axis gradient and triple-resonance TXI room temperature probe at 5˚C in 20 mM potassium phosphate containing 20 mM NaCl, 1 mM $NaN_3$, and 10% $^2H_2O$ (pH 7.4). A 3-9-19 plus sequence for water suppression (watergate) is used for recording all the $^1H$ NMR spectra. For each FID 1024 scans were accumulated. The $^1H$ chemical shift were referenced to the Sodium 2,2-dimethyl-2-silapentane-5-Sulfonate (DSS) signal added internally ($\delta$ = 0.00 ppm for $^1H$). The data were processed with Topspin 2.1 (Bruker Biospin). The $^1H$ signals of side-chains were partially assigned using the previous information for signal assignments (Ogata *et al*., 1994).

## Results

R2R3* and its mutants were over-expressed in *E. coli* and purified, and the purity of each protein was determined to be approximately 95% by SDS/PAGE analysis. The far-UV CD spectra of these proteins are similar, and show typical $\alpha$-helical structures (Fig. 5-2A). These results indicate that mutation of the linker region does not alter the secondary structure. The thermal denaturation was observed by monitoring the CD value at 222 nm (Fig. 5-2B), and the $T_m$ and $\Delta H_{vH}$ were determined and summarized in Table 5-1. The $T_m$ values of linker mutants were slightly lower than that of R2R3* and the proteins were more destabilized by P140G mutation than that of P140A. In order to assess the thermodynamic parameter for thermal unfolding in more detail, the thermal stabilities of R2R3* and its mutants were analyzed using DSC. Fig. 5-3 shows the heat capacity curves of proteins analyzed at a heating rate of 1°C $min^{-1}$. As reported at Chapter 2, the thermal unfolding process was quantitatively reversible, although it was not fully reversible. The thermodynamic parameters determined are summarized in

Table 5-2. Decreased thermal stability occurred in response to mutation in the linker region.

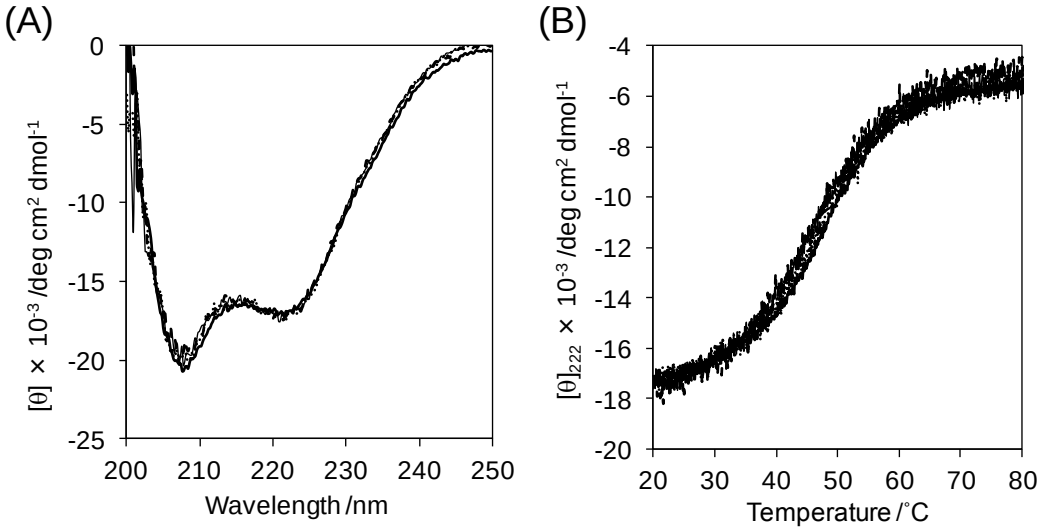



Fig. 5-2. (A) The far-UV spectra and of R2R3* (solid thick line), P140A (dotted line), P140G (solid thin line), and G3 (broken line) at 20ºC. The vertical scale is normalized by the molar concentration. (B) The thermal denaturation curves of R2R3* (solid thick line), P140A (dotted line), P140G (solid thin line), and G3 (broken line).

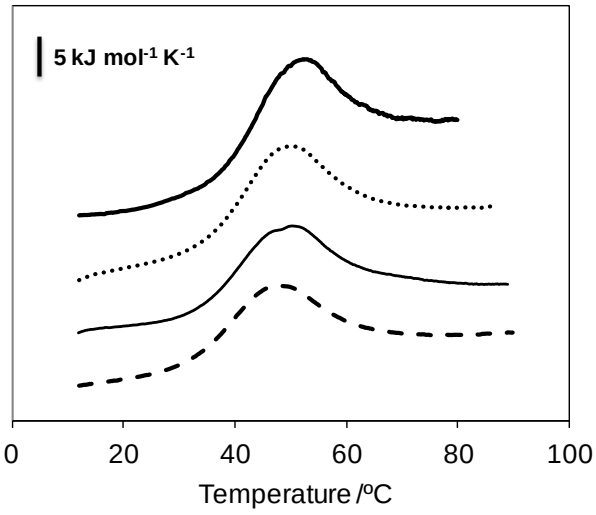



Fig. 5-3. The heat capacity curves of R2R3* (solid thick line), P140A (dotted line), P140G (solid thin line), and G3 (broken line).

Table 5-1. Thermodynamic parameters for denaturation of R2R3* and its mutants analyzed using CD.

| | $T_m$ /ºC | $\Delta H_{vH}$ /kJ mol$^{-1}$ |
|---|---|---|
| R2R3* | 48.5 ± 0.1 | 161 ± 3 |
| P140A | 47.6 ± 0.2 | 135 ± 5 |
| P140G | 46.4 ± 0.3 | 157 ± 5 |
| G3 | 46.8 ± 0.3 | 145 ± 5 |

Table 5-2. Thermodynamic parameters for denaturation of R2R3* and its mutants analyzed using DSC.

| | $T_d$ /°C | $\Delta H_{cal}$ /kJ mol$^{-1}$ | $\Delta C_p$ /kJ mol$^{-1}$ K$^{-1}$ |
|---|---|---|---|
| R2R3*[a] | 50.3 ± 0.02 | 266 ± 5 | 8.66 ± 0.10 |
| P140A[b] | 48.9 ± 0.1 | 193 ± 4 | 3.81 ± 0.59 |
| P140G[b] | 47.6 ± 0.6 | 177 ± 2 | 4.90 ± 0.37 |
| G3[b] | 47.1 ± 0.2 | 169 ± 3 | 2.77 ± 0.64 |

[a] Data were taken from Chapter 4.

[b] The error values indicated are derived from the values determined in three measurements.

In this study, 1H-NMR experiments were carried out for more in-depth analysis of residue-specific structural properties of the linker mutant, G3. Fig. 5-4 shows a comparison of $^1$H-NMR spectra of R2R3* and G3 mutant at 5°C. Some signals were partially assigned using previously reported assignment information (Ogata *et al*., 1994; Ogata *et al*., 1995). Most assigned side-chain signals in low-magnetic (Trp ring, in addition to other atoms, Fig. 5-4D) and high-magnetic (methyl signals of Val, Leu, and Ile in the hydrophobic core, Fig. 5-4E) regions were well-dispersed, representing their individual microenvironments within the folded structures. Most of $^1$H-NMR signals of G3 mutant were more broadened than those of R2R3* (Fig. 5-4A). In addition, the

positions and widths (peak area) of some signals differed between the R2R3* and G3 mutant, as described below. For both low and high magnetic field, signals from 6.5 to 9 ppm, corresponding to the main chain amide proton of G3 mutant, were more broadened than those of R2R3*. (Figs. 5-4B and 4C). These differences indicate that the rates of chemical exchange, namely, conformational fluctuation of backbone amide and carbonic groups, were greatly affected by mutation in the linker regions. In addition, some assigned signals derived from side-chains also exhibited changes in chemical shifts and/or peak intensities. While the six peaks for the ε1 Trp iodole proton appeared clearly in the spectrum of R2R3*, the signal derived from Trp-115 was absent and signal intensities of Trp-134 and/or Trp-166 and Trp-185 were lower for the G3 mutant than for R2R3* (Fig. 5-4D). Instead, the three split peaks appeared at chemical shifts of 10.3 ppm in the spectrum of the G3 mutant. The γ1 methyl proton signal of Val-107 appeared at a high magnetic field in the R2R3* spectrum, but showed a clear shift toward a low magnet field and was broadened in the G3 mutant. Likewise, the γ2 and δ methyl proton signals of Val-107, Ile-118, Ile-169, and Ile-181 shifted toward a high magnetic field and were slightly broaden for the G3 mutant (Fig. 5-4E). Those broadening signals were confirmed by temperature-dependent NMR studies, indicating that there was a definitive in the conformational fluctuation of the protein. (see Chapter 2). The results emphasized that the linker mutation affected both R2 and R3 in the hydrophobic core, in addition to the backbone structure.

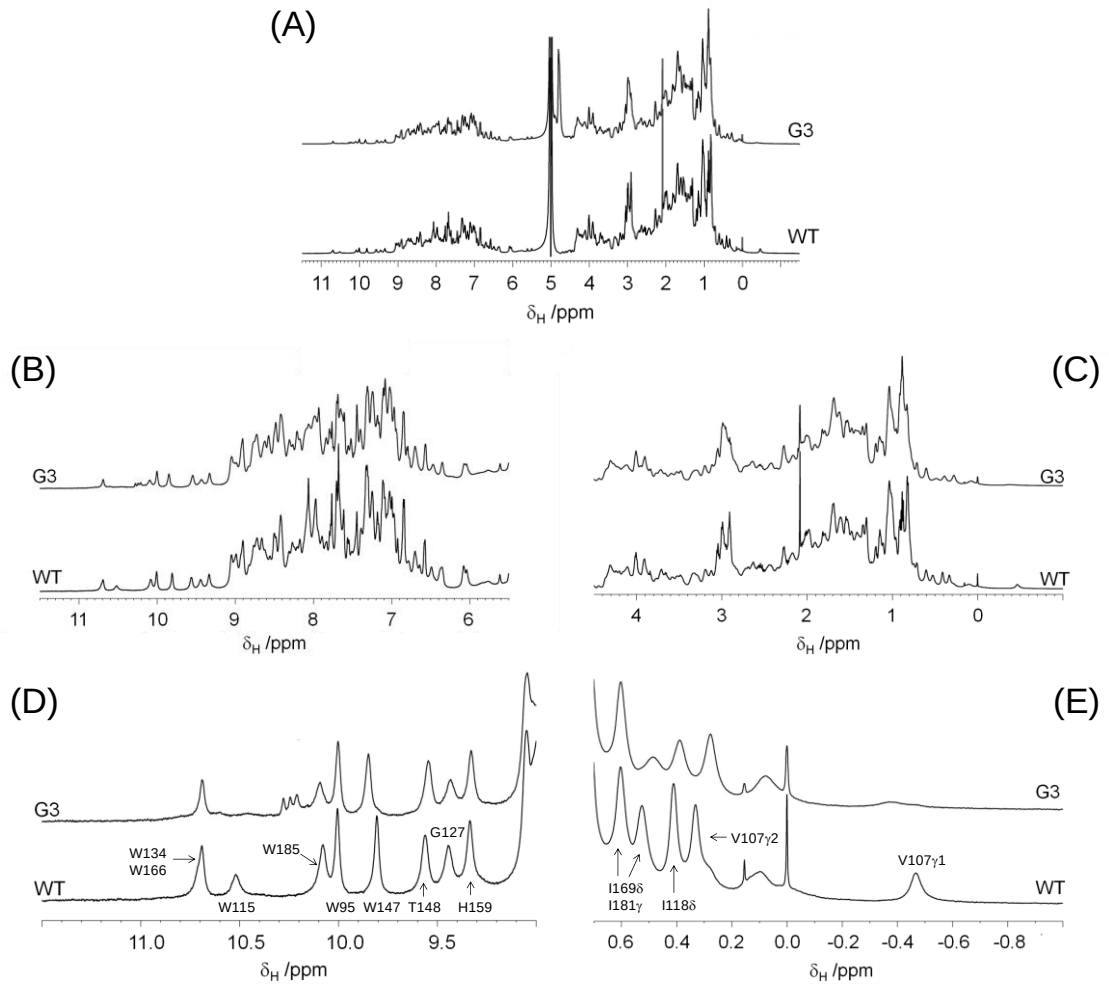



Fig. 5-4. The $^{1}$H-NMR spectra of R2R3* and G3 mutant. The spectra of (A) overall, (B) low and (C) high magnetic fields, (D) iodole and (E) methyl regions.

## Discussion

To analyze the stability difference in terms of thermodynamics, the thermodynamic parameters were calculated at 50.3°C, the denaturation temperature of R2R3*, using the respective $\Delta C_p$ values of mutants (Table 5-3). Within the narrow range of temperature around the $T_d$ values, the errors of the calculated $\Delta G$ and $\Delta H$ values at the reference temperature should be small, even if the $\Delta C_p$ values obtained have errors. The results clearly indicate that the decrease in stability is due to enthalpy, which is

partially compensated by a favorable entropy change. Enthalpy-entropy compensation is a phenomenon typical of either intramolecular or intermolecular interactions in biomolecules (Dunitz, 1995). The favorable entropy change can be explained by the increased entropy of mutants in the folded state. This can also explain the unfavorable entropy change for DNA binding reported previously (Oda *et al*., 1998). The $\Delta S$ of P140G was smaller than that of P140A (Table 5-3), indicating that Gly increases protein fluctuation more than Ala. In addition, mutation of Asn139-Pro140-Glu141 into Gly139-Gly140-Gly141 would make the fluctuation, especially in the folded state, larger. After the mutation of Pro into Gly or Ala, the conformational entropy in the unfolded state would also be increased (Fig. 5-5B). Some investigations have shown that the stability of a protein can be increased by the incorporation of Pro, due to the resulting decrease in the configurational entropy of the unfolded state (Matthews *et al*., 1987; Yutani *et al*., 1991). In the case of present study, the increased entropy in the folded state of R2R3 containing proline residue in the linker region would be smaller than that in the unfolded state, resulting in increased $\Delta S$ compared with linker mutants (Fig. 5-5B). In addition to the contribution of locking of the dihedral angle, the Pro residue in the linker region would have other specific effects on the structural dynamics in the folded state, including the interaction between R2 and R3. This is also supported by NMR results that showed that the $^{1}$H-NMR signals of G3 are more broaden than those of R2R3* (Figs. 5-4A, 4B, 4C, 4D, and 4E). Broadening of the NMR signal under physiological conditions strongly indicates that protein fluctuation in the folded and unfolded conformations equilibrated is closely correlated with its biological function (see Chapter 2). The temperature-dependent NMR spectra of G3 mutant were also measured and showed that most peaks of $^{1}$H signals were more broaden and disappeared

as temperature increased, similar to those of R2R3* (data not shown). Additional mutations of Asn-139 to Gly and Glu-141 to Gly would increase protein fluctuation in the folded state, possibly due to disruption of intramolecular interactions. In fact, several NMR signals derived from nuclear Overhauser effect were observed between the side-chain protons of Asn-139 and the protons of other R2 residues in the NMR structures (Ogata *et al*., 1994; Ogata *et al*., 1995).

Table 5-3. Thermodynamic parameters for denaturation of R2R3* and its mutants at the denaturation temperature of R2R3* (50.3°C).

| | $T_d$[a] /°C | $\Delta T_d$[b] | $\Delta G$[c] /kJ mol$^{-1}$ | $\Delta H$[c] /kJ mol$^{-1}$ | $\Delta S$[c] /J mol$^{-1}$ K$^{-1}$ |
|---|---|---|---|---|---|
| R2R3* | 50.3 ± 0.02 | 0 | 0 | 266 | 822 |
| P140A | 48.9 ± 0.1 | −1.8 ± 0.1 | −0.84 ± 0.08 | 198 ± 4 | 615 ± 11 |
| P140G | 47.6 ± 0.6 | −2.7 ± 0.6 | −1.5 ± 0.3 | 190 ± 1 | 591 ± 3 |
| G3 | 47.1 ± 0.2 | −3.2 ± 0.2 | −1.7 ± 0.1 | 178 ± 4 | 555 ± 14 |

[a] The data were taken from Table 5-1.

[b] $\Delta T_d$ values were calculated from the equation, $\Delta T_d = T_d$ (mutant) − $T_d$ (R2R3*).

Proteins fluctuate in solution, and this dynamic behavior is critical for protein function. Most of the methods to determine protein ternary structure at an atomic resolution can only confirm the most stable or averaged structure, similar to taking a snapshot. In the case of proteins that undergo different fluctuations, as shown in the schematic view of structural dynamics (Fig. 5-5A), although the averaged structures are similar, the structural dynamics are completely different. The increased fluctuation can be detected by the entropy term, as described in this study. I have found that R2R3 fluctuates greatly in solution and that their unfolded conformation is estimated to be present at level of >10% under physiological conditions (see Chapter 2). As ultimate

fluctuations, intrinsically disordered properties have been identified in many cell-signaling and DNA-binding proteins (Dyson, 2012; Chu *et al*., 2014). This flexibility would allow identification of the specific site while minimizing unfavorable collisions with the target DNA (Oda and Nakamura, 2000). In contrast, some proteins, such as protein inhibitors, have rigid structures with little fluctuation, probably to aid in resistance to protease digestion. Protein function can be attributed to the combination of this flexibility and rigidity. In the case of c-Myb R2R3, previous structural and physical analyses have shown that the fluctuation of R2 is larger than that of R3, mainly due to the cavity in the hydrophobic core of R2 (Ogata *et al*., 1996; Morii *et al*., 1999; Puthenpurackal *et al*., 2012; see Chapter 2). By connecting the "soft" region, R2, to the "rigid" region by the linker, the functional DBD would be formed. The linker region would have a role to restrict the flexibility of R2, resulting in the increased stability and DNA-binding affinity (Oda *et al*., 1998).

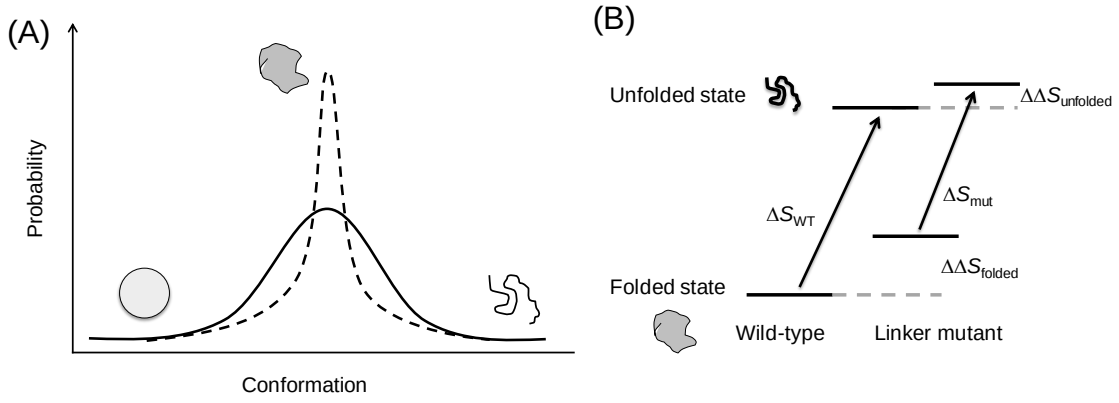



Fig. 5-5. (A) Schematic view of protein structural dynamics with large (solid line) and small (broken line) fluctuations. (B) Schematic diagram representing the entropy levels of folded and unfolded states.

## Chapter 6 General discussion and conclusion

In this study, using the several biophysical methods, structural dynamics of the c-Myb DBD (R2R3) were analyzed and correlated with its function. Although three-dimensional structural determination is a popular and powerful method for understanding a protein's function, it is difficult to discuss structure-function relationships because those structures have little information about dynamic behaviors. Recently, the conformational fluctuations of a protein have been considered to be critical for its individual function and tried to be explored using a wide variety of methods (Bongini *et al*., 2004; Fourme *et al*., 2012; Ishima, 2015). On the other hands, there are few reports about 'biological fluctuation' which should be determined under physiological conditions. Therefore, I have attempted to study the conformational fluctuation of proteins, with specific focus on "biological fluctuation", using the c-Myb DBD. A summary of each chapter is given below.

In Chapter 2, the conformational fluctuation of c-Myb R2R3 was studied under close to physiological conditions. A global unfolding transition, involving both the main chain and the side chains, was found to take place at the temperature range of ~30ºC to ~70ºC, with a transition temperature of around ~50ºC. In addition, the observation of simultaneous shift change and the broadening of NMR signals in both $^{1}H$ one-dimensional and $^{15}N/^{1}H$ two-dimensional NMR spectra indicated the occurrence of a locally fluctuating state at the physiological temperature. In the wild-type protein containing a cavity in R2, the local fluctuation of R2 was more prominent than that of R3, whereas it was suppressed in the cavity-filled mutant V103L. This indicates that the cavity in R2 contributes significantly to the conformational instability and transition into the locally fluctuating state. For the wild-type R2R3 protein, the more dynamic

conformer is estimated to be present to some extent at 37ºC, and is likely to be beneficial for its biological function, DNA-binding. This result is in agreement with the concept of an excited-state conformer that exists in equilibrium with the dominant ground-state conformer and acts as the functional conformer of the protein. From the findings of this study, it appears that the tandem repeats of two small domains with no disulfide bond and with a destabilizing cavity function as the evolutionary strategy of the wide-type c-Myb DBD, in order to produce just the appropriate fraction of the locally fluctuating state at 37ºC, which is more amenable to DNA-binding. I showed that extensive conformational fluctuations within the folded manifold of the R2R3* protein were activated at the physiological temperature. The results of the variable temperature NMR study described here indicate that ample fluctuations prevail heterogeneously over the entire polypeptide chain within the native ensemble of the protein.

In Chapter 3, the conformational and thermal stabilities of c-Myb R2R3 were analyzed under different pH conditions, ranging from 4.0 to 7.5, using CD and DSC. I showed that the conformational stability of the protein (vis., its secondary structure) was largely affected by the solution pH. Of all the conditions analyzed, the α-helical content was maximal at pH 6.5 and the thermal stability was highest at pH 5.0. Thermodynamic parameters for thermal unfolding of R2R3 were determined using DSC, and the origin of folding the thermodynamics at the different pHs and their correlation with the α-helical content were further analyzed. It should be noted that the α-helical content correlated well with the enthalpy change in the pH range from 4.5 to 7.5, suggesting that the strength of the hydrogen bonds and salt bridges needed for maintenance of the helical structure is related to enthalpy in the native state. Under physiological pH

conditions, c-Myb R2R3 exists in the enthalpically unstable but entropically stable state. Owing to the loss of its rigid structure and high stability, the protein could obtain structural flexibility, befitting its function. The conformational flexibility of R2R3* resulting from the reduction of intramolecular interactions would be needed to express its function (DNA-binding) despite of its decreased stability.

In Chapter 4, I analyzed the thermodynamic effects of the I155L and I181L mutations, using R2R3 that encompasses the minimum specific DNA-binding region, in order to quantitatively correlate the structural fluctuation with stability and function. A previous NMR study has shown that a mutant protein of I155L/I181L R3 has multiple conformations and increase fluctuation in comparison with the wild-type (Furukawa *et al.*, 1996). CD and DSC measurements showed that the mutation of I155L mutation had little effect on stability, whereas the I181L mutation significantly destabilized the protein. It is noteworthy that the decreased stability resulting from the I181L mutation was mainly due to a decreased enthalpy change, which was partially compensated by a decreased entropy change. On the other hand, the specific DNA-binding affinity was decreased owing to the I181L mutation, which was due to a decreased binding entropy change. Entropy in the folded state, which corresponds to the DNA-free state, increases due to the I181L mutation because of the increased conformational fluctuation observed in I155L/I181L mutant of R2R3 by CLEANEX-PM NMR analysis, which in turn results in decreased folding entropy and DNA-binding entropy changes.

In Chapter 5, I have analyzed the structural dynamic properties of the DBD of c-Myb R2R3. Earlier reports state that a substitution at the linker between R2 and R3 resulted in a significant loss of affinity towards the cognate DNA, mainly due to increased entropy in the DNA-unbound state. In this study, I analyzed the effects of the

linker on folding stability, showing that the mutation of Pro-140 to Gly or Ala in the linker region of c-Myb R2R3 caused decreased stability. The thermodynamic origin was found to be an unfavorable enthalpy change and favorable entropy change, which are mainly due to the increased protein fluctuation in the folded state. This is in good agreement with the previous findings that the same mutation resulted in decreased DNA-binding affinity, mainly due to the increase in conformational entropy in the DNA-unbound state, as in the folded state (Oda *et al*., 1998). Taken together, the effects of increased fluctuation could be shown on not only the binding thermodynamics but also the folding thermodynamics, both of which are well correlated.

In conclusion, I succeeded in observing the significantly conformational fluctuation of c-Myb R2R3, which included the important 'biological fluctuation'. This study emphasized that the conformational stability of wild-type c-Myb R2R3 is not maximal at physiological conditions (pH 7.4, 37˚C), and the degree of conformational fluctuation revealed by the thermodynamics studies would be programmed in the process of biological evolution. The balance between rigidness and softness would be critical for the function. In order to achieve the main function of c-Myb R2R3 (viz., the specific DNA-binding), flexibility would be needed to attain the precise recognition mechanism. There are some reports for inartistically disordered proteins (IDPs), which have no particular structures and are often observed in the transcriptional factors (Sugase *et al*., 2007; Jensen *et al*., 2014; Wright and Dyson, 2015; Sherry *et al*., 2015). A previous study showed that the KIX domain of c-Myb represented the typical IDP profile, and was folded by binding to its target molecules (Zor *et al*., 2004; De Guzman *et al*., 2006; Giri, *et al*., 2013). In the case of the c-Myb DBD, partially instead of fully unfolded states should exist under the physiological conditions. The protein fluctuation

for function should be defined as “biological fluctuation”, and we should consider it to be the most important phenomenon for the structure-function relationship.

## References


Akasaka, K., and Yamada, H. (2001). On-line cell high-pressure nuclear magnetic resonance technique: application to protein studies. Methods Enzymol. *338*, 134-158.

Akasaka, K. (2006). Probing conformational fluctuation of proteins by pressure perturbation. Chem. Rev. *106*, 1814-1835.

Akasaka, K., and Matsuki, H. ed. (2015). High Pressure Bioscience - Basic Concepts, Applications and Frontiers.

Anderson, D.E., Becktel, W.J., and Dahlquist, F.W. (1990). pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 Lysozyme. Biochemistry *29*, 2403-2408.

Arunkumar, A.I., Kumar, T.K.S., and Yu, C. (1997). Specificity of helix-induction by 2,2,2-trifluoroethanol in polypeptides. Int. J. Biol. Macromol. *21*, 223-230.

Banerjee, T., and Kishore, N. (2005). 2,2,2-Trifluoroethanol-induced molten globule state of concanavalin A and energetics of 8-anilinonaphthalene sulfonate binding: calorimetric and spectroscopic investigation. J. Phys. Chem. B *109*, 22655-22662.

Beauchamp, D.L., and Khajehpour, M. (2012). Studying salt effects on protein stability using ribonuclease t1 as a model system. Biophys. Chem. *161*, 29-38.

Betz, S.F. (1993). Disulfide bonds and the stability of globular proteins. Protein Sci. *2*, 1551-1558.

Biedenkapp, H., Borgmeyer, U., Sippel, A.E., and Klempnauer, K.H. (1988). Viral myb oncogene encodes a sequence-specific DNA-binding activity. Nature *335*, 835-837.

Blumlein, A., and McManus, J.J. (2013). Reversible and non-reversible thermal denaturation of lysozyme with varying pH at low ionic strength. Biochem. Biophys. Acta. *1834*, 2064-2070.

Bongini, L., Fanelli, D., Piazza, F., De Los Rios, P., Sandin, S., and Skoglund, U. (2004). Freezing immunoglobulins to see them move. Proc. Natl. Acad. Sci. USA *101*, 6466-6471.

Bouvignies, G., Vallurupalli, R., Hansen, D.F., Correia, B.E., Lange, O., Bah, A., Vernon, R.M., Dahlquist, F.W., Baker, D., and Kay, K.E. (2011). Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. Nature *477*, 111-114.

Brunette, T.J., Parmeggiani, F., Huang, P.-S., Bhabha, G., Ekiert, D.C., Tsutakawa, S.E., Hura, G.L., Tainer, J.A., and Baker, D. (2015). Exploring the repeat protein universe thorough computational protein design. Nature *528*, 580-584.

Chen, Y.H., Yang, J.T., and Martinez, H.M. (1972). Determination of secondary structure of protein by circular dichroism and optical rotator dispersion. Biochemistry *11*, 4120-4131.

Chaires, J.B. (2008). Calorimetry and thermodynamics in drug design. Annu. Rev. Biophys. *37*, 135-51.

Chu, H.-L., Chen, T.-H., Wu, C.-Y., Yang, Y.-C., Tseng, S.-H., Cheng, T.-M., Ho, L.-P., Tsai, L.-Y., Li, H.-Y., Chang, C.-S., and Chang, C.-C. (2014). Thermal stability and folding kinetics analysis of disordered protein, securing: J. Therm. Anal. Calorim. *115*, 2171-2178.

Dave, S., Mahajan, S., Chandra, V., and Gupta, P. (2011). Trifluoroetanol stabilizes the molten globule state and induces non-amyloidic turbidity in stem bromelain near its isoelectric point. Int. J. Biol. Macromol. *49*, 536-542.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277-293.

De Guzman, R.N., Goto, N.K., Dyson, H.J., and Wright, P.E. (2006). Structural basis for cooperative transcription factor binding to the CBP coactivator. J. Mol. Biol. *355*, 1005-1013.

Del Vecchio, P., Carullo, P., Barone, G., Pagano, B., Graziano, G., Iannetti, A., Acquaviva, R., Leonardi, A., and Fromisano, S. (2008). Conformational stability and DNA-binding energetic of the rat thyroid transcription factor 1 homeodomain. Proteins: Struct. Funct. Genet. *70*, 748-760.

Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B.L., and Bradley, P. (2015). Rational design of α-helical tandem repeat proteins with closed architectures. Nature *528*, 585-589.

Dunitz, J.D. (1995). Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. Chem. Biol. *2*, 709-712.

Dyson, H.J., and Wright, P.E. (1996). Insights into protein folding from NMR. Annu. Rev. Phys. Chem. *47*, 369-95.

Dyson, H.J. (2012). Roles of intrinsic disorder in protein–nucleic acid interactions. Mol. Biosyst. *8*, 97-104.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by $^{15}$N NMR relaxation. Biochemistry *33*, 5984-6003.

Fenley, A.T., Muddana, H.S., and Gilson, M.K. (2012). Entropy-enthalpy transduction caused by conformational shifts can obscure the forces driving protein-ligand binding. Proc. Natl. Acad. Sci. USA *109*, 20006-20011.

Fourme, R., Girard, E., and Akasaka, K. (2012). High-pressure macromolecular crystallography and NMR: status, achievements and prospects. Curr. Opin. Struct. Biol. *22*, 636-642.

Fukada, H., and Takahashi, K. (1998). Enthalpy and heat capacity changes for the proton dissociation of various buffer components in 0.1 M potassium chloride. Proteins: Struct. Funct. Genet. *33*, 159-166.

Furukawa, K., Oda, M., and Nakamura, H. (1996). A small engineered protein lacks structural uniqueness by increasing the side-chain conformational entropy. Proc. Natl. Acad. Sci. USA *93*, 13583-13588.

Gange, D., Charest, L.A., Morin, S., Kovrigin, E.L., and Doucet, N. (2012). Conservation of flexible residue clusters among structural and functional enzyme homologues. J. Biol. Chem. *287*, 44289-44300.

Giri, R., Morrone, A., Toto, A., Brunori, M., and Gianni, S. (2013). Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. Proc. Natl. Acad. Sci. USA *110*, 1492-1497.

Grey, M.J., Tang, Y., Alexov, E., McKnight, C.J., Raleigh, D.P., and Plamer, A.G.III. (2006). Characterizing a partially folded intermediate of the villin headpiece domain under non-denaturing conditions: contribution of His41 to the pH-dependent stability of the N-terminus subdomain. J. Mol. Biol. *355*, 1078-1094.

Gonda, T.J., Gough, N.M., Dunn, A.R., and de Blaquiere, J. (1985). Nucleotide sequence of cDNA clones of the murine myb proto-oncogene. EMBO J. *4*, 2003-2008.

Haliloglu, T., and Bahar, I. (2015). Adaptability of protein structures to enable functional interactions and evolutionary implications. Curr. Opin. Struct. Biol. *35*, 17-23.

Hegvold, A.B., and Gabrielsen, O.S. (1996). The importance of the linker containing the repeats of the c-Myb oncoprotein may be due to a positioning function. Nucleic Acid. Res. *24*, 3990-3995.

Hendsch, Z.S., and Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci. *3*, 211-226.

Henzler-Wildman, K.A., Lei, M., Thai, V., Kerns, S.J., Karplus, M., Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature *450*, 913-916.

Huyghues-Despointes, B.M.P. and Baldwin, R.L. (1997). Ion-pair and changed hydrogen-bond interactions between histidine and aspartate in a peptide helix. Biochemistry *36*, 1965-1970.

Hwang, T.-L., van Zijl, P.C., and Mori, S. (1998). Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. J. Biomol. NMR *11*, 221-226.

Ishima, R. (2012). Recent developments in $^{15}$N NMR relaxation studies that probe protein backbone dynamics. Top Curr. Chem. *326*, 99-122.

Ishima, R. (2015). Protein-inhibitor interaction studies using NMR. Appl. NMR Spectrosc. *1*, 143-181.

Jaenicke, R. (2000). Stability and stabilization of globular proteins in solution. J.

Biotechnol. *79*, 193-203.

Jensen, M.R., Zweckstetter, M., Huang, J.R., and Blackledge, M. (2014). Exploring free energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. Chem. Rev. *114*, 6632-6660.

Kanei-Ishii, C., Sarai, A., Sawazaki, T., Nakagoshi, H., He, D-N., Ogata, K., Nishimura, Y., and Ishii, S. (1990). The tryptophan cluster: a hypothetical structure pf the DNA-binding domain of the c-myb protooncogene product. J. Biol. Chem. *265*, 19990-19995.

Karplus, M. (1984). Dynamics of proteins. Adv. Biophys. *8*, 165-90.

Kay, L.E. (2005). NMR studies of protein structure and dynamics. J. Magn. Reson. *173*, 193-207.

Kitahara, R. Yokoyama, S. and Akasaka, K. (2005). NMR snapshots of a fluctuating protein structure: ubiquitin at 30 bar-3 kbar. J. Mol. Biol. *347*, 277-285.

Kono, H., Saito, M., and Sarai, A. (2000). Stability analysis for the cavity-filling mutations of the Myb DNA-binding domain utilizing free-energy calculations. Proteins: Struct. Funct. Genet. *38*, 197-209.

Korzhnev, D.M., and Kay, L.E. (2008) Probing invisible low-populated state of protein molecule by relaxation dispersion NMR spectroscopy: an application to protein folding. Accounts Chem. Rev. *41*, 442-451.

Lassalle, M.W., Yamada, H., Morii, H., Ogata, K., Sarai, A., and Akasaka, K. (2001). Filling cavity dramatically increases pressure stability of the c-Myb R2 subdomain. Proteins: Struct. Funct. Genet. *45*, 96-101.

Lau, W.L., DeGrado, W.F., and Roder, H. (2010). The effects of $pK_a$ tuning on the thermodynamics and kinetics of folding: design of a solvent-shielded carboxylate pair at the a-position of a coiled-coil. Biophys. J. *99*, 2299-2308.

Loria, J.P., Rance, M., and Palmer, A.G.III. (1999). A relaxation-compensated Carr–Purcell–Meiboom–Gill sequence for characterizing chemical exchange by NMR spectroscopy. J. Am. Chem. Soc. *121*, 2331–2332.

Matthews, B.W., Nicholson, H., and Becktel, W.J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc. Natl. Acad. Sci. USA *84*, 6663-6667.

Metrick, M.A., Temple, J.E., and MacDonald, G. (2013). The effects of buffers and pH on the thermal stability, unfolding and substrate binding of RecA. Biophys. Chem. *184*, 29-36.

Morii, H., Uedaira, H., Ogata, K., Ishii, S., and Sarai, A. (1999). Shape and energetic of a cavity in c-Myb probed by natural and non-natural amino acid mutations. J. Mol. Biol. *292*, 909-920.

Murphy, K.P., and Freire, E. (1992). Thermodynamics of structural stability and cooperative folding behavior in proteins. Adv. Protein Chem. *43*, 313-361.

Myrset, A.H., Bostad, A., Jamin, N., Lirsac, P.N., Toma, F., and Gabrielsen, O.S. (1993). DNA and redox state induced conformational changes in the DNA-binding domain of the Myb oncoprotein. EMBO J. *12*, 4625-4633.

Naber, N., Larson, A., Rice, S., Cooke, R., and Pate, E. (2011). Multiple conformations of the nucleotide site of Kinesin family motors in the triphosphate state. J. Mol. Biol. *408*, 628-642.

Naganathan, A.N., Doshi, U., Fung, A., Sadqi, M., and Muñoz. V. (2006). Dynamics, energetics, and structure in protein folding. Biochemistry *45*, 8466-8475.

Nakagoshi, H., Nagase, T., Kanei-Ishii, C., Ueno, Y., and Ishii, S. (1990). Binding of the c-myb proto-oncogene product to the simian virus 40 enhancer stimulates transcription. J. Biol. Chem. *265*, 3479-3483.

Ness, S.A. (1999). Myb binding proteins: regulators and cohorts in transformation. Oncogene *18*, 3039-3046.

Oda, M., Furukawa, K., Ogata, K., Sarai, A., Ishii, S., Nishimura, Y., and Nakamura, H. (1997a). Investigation of the pyrimidine preference by the c-Myb DNA-binding domain at the initial base of the consensus sequence. J. Biol. Chem. *272*, 17966-17971.

Oda, M., Furukawa, K., Ogata, K., Sarai, A., Ishii, S., Nishimura, Y., and Nakamura, H. (1997b) Identification of indispensable residues for specific DNA-binding in the imperfect tandem repeats of c-Myb R2R3. Protein Eng. *10*, 1407-1414.

Oda, M., Furukawa, K., Ogata, K., Sarai, A., and Nakamura, H. (1998). Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. J. Mol. Biol. *276*, 571-590.

Oda, M., Furukawa, K., Sarai, A., and Nakamura, H. (1999). Kinetic analysis of DNA binding by the c-Myb DNA-binding domain using surface plasmon resonance. FEBS Lett. *454*, 288-92.

Oda, M., and Nakamura, H. (2000). Thermodynamic and kinetic analyses for understanding sequence-specific DNA recognition. Genes to Cells *5*, 319-326.

Oda, M., Ito, N., Tsumuraya, T., Suzuki, K., Sakakura, M., and Fujii, I. (2007). Thermodynamics and structural basis for transition-state stabilization in antibody-catalyzed hydrolysis. J. Mol. Biol. *369*, 198-209.

Oda, M., Kitai, A., Murakami, A., Nishimura, M., Ohkuri, T., Abe, Y., Ueda, T., Nakamura, H., and Azuma, T. (2010). Evaluation of the conformational equilibrium of reduced hen egg lysozyme by antibodies to the native form. Arch. Biochem. Biophys. *494*, 145-150.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., and Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core, Proc. Natl. Acad. Sci. USA *89*, 6428-6432.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. Cell *79*, 639-648.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimoto, S., Ametani, Y., Hirata, Z., Sarai, A., Ishii, S., and Nishimura, Y. (1995). Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. Nat. Struct. Biol. *2*, 309-320.

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Nakamura, H., Nishimura, Y., Ishii, S., and Sarai, A. (1996). The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. Nat. Struct. Biol. *3*, 178-187.

Okada, A., Miura, T., and Takeuchi, H. (2003). Zinc- and pH dependent conformational transition in a putative interdomain linker region of the influenza virus matrix protein M1. Biochemistry *42*, 1978-1984.

Pace, C.N., Grimsley, G.R., and Scholtz, J.M. (2009). Protein ionizable groups: p*K* values and their contribution to protein stability and solubility. J. Biol. Chem. *284*, 13285-13289.

Perozzo, R., Folkers, G., and Scapozza, L. (2004). Thermodynamics of protein-ligand interactions: history, presence, and future aspects. J. Recept. Signal Transduct. Res. *24*, 1-52.

Phelan, P., Gorfe, A.A., Jelesarov, I., Marti, D.N., Warwicker, J., and Bosshard, H.R. (2002). Salt bridge destabilize a leucine zipper designed for maximized ion paring between helices. Biochemistry *41*, 2998-3008.

Palmer, A.G.III. (2004). NMR characterization of the dynamics of biomacromolecules. Chem. Rev. *104*, 3623-3640.

Peterson, R.W., and Wand, A.J. (2005). Self-contained high-pressure cell, apparatus, and procedure for the preparation of encapsulated proteins dissolved in low viscosity fluids for nuclear magnetic resonance spectroscopy. Rev. Sci. Instrum. *76*, 1-7.

Prouse, M.B., and Campbell, M.M. (2012). The interaction between proteins and their target DNA binding site. Biochem. Biophys. Acta. *1819*, 67-77.

Privalov, P.L. (1979). Stability of proteins: small globular proteins. Adv. Protein Chem. *33*, 167-241.

Puthenpurackal, S.N., Maeno, A., Matsuo, H., Oda, M., Morii, H., and Akasaka, K. (2012). Extensively hydrated but folded: a novel state of globular proteins stabilized at high pressure and low temperature. Biophys. J. *102*, L08-L10.

Reich, H.J. (1995). WinDNMR: Dynamic NMR spectra for Windows. J. Chem. Educ. *72*, 1086.

Robinson, A.C., Castaneda, C.A., Schlessman, J.L., and Garcia-Moreno, B.E. (2014). Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. Proc. Natl. Acad. Sci. USA *111*, 11685-11690.

Sagawa, T., Oda, M., Ishimura, M., Furukawa, K., and Azuma, T. (2003). Thermodynamics and kinetics aspects of antigen evolution during the immune response to hapten. Mol. Immunol. *39*, 801-808.

Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T.J,, and Ishii, S. (1989). Delineation of three functional domains of the transcriptional activator encoded by the c-myb protooncogene. Proc. Natl. Acad. Sci. USA *86*, 5758-5762.

Saito, K., Sarai, A., Oda, M., Azuma, T., and Kozono, H. (2003). Thermodynamic analysis of the increased stability of major histocompatibility complex class II molecule I-$E^{\kappa}$ complexed with an antigen peptide at an acidic pH. J. Biol. Chem. *278*, 14732-14738.

Sarai, A., Uedaira, H., Morii, H., Yasukawa, T., Ogata, K., Nishimura, Y., and Ishii, S. (1993). Thermal stability of the DNA-binding domain of the oncoprotein. Biochemistry *32*, 7759-7764.

Scholtz, J.M., Quan, H., Robbins, V.H., and Baldwin, R.L. (1993). The energetic of ion-pair and hydrogen-bonding interaction in a helical peptide. Biochemistry *32*, 9668-9676.

Shank, E.A., Cecconi, C., Dill, J.W., Marqusee, S., and Bustamante, C. (2010). The folding cooperativity of a protein is controlled by its chain topology. Nature *465*, 637-640.

Sherry, K.P., Johnson, S.E., Hatem, C.L., Majumdar, A., and Barrick, D. (2015). Effects of linker length and transient secondary structure elements in the intrinsically disordered Notch RAM region on Notch signaling. J. Mol. Biol. *427*, 3587-3597.

Smith, C.A., Ban, D., Giller, S.P.K., Schwiegk, C., de Groot, B.L., Becker, S., Griesinger, C., and Lee, D. (2015). Population shuffling of protein conformations. Angew. Chem. Int. Ed. *54*, 207-210.

Spolar, R.S., and Record, M.T.Jr. (1994). Coupling of local folding to site-specific binding of proteins to DNA. Science *263*, 777-784.

Stertevant, J.M. (1977). Heat capacity and entropy changes in processes involving proteins. Proc. Natl. Acad. Sci. USA *74*, 2236-2240.

Sugase, K., Dyson, H.J., and Wright, P.E. (2007). Mechanism of coupled folding and binding of a intrinsically disordered protein. Nature *447*, 1021-1025.

Tahirov, T.H., Sato, K., Ichikawa-Iwata, E., Sasaki, M., Inoue-Bungo, T., Shiina, M., Kimura, K., Takata, S., Fujikawa, A., Morii, H., Kumasaka, T., Yamamoto, M.,

Ishii, S., and Ogata, K. (2002). Mechanism of c-Myb-C/EBPβ cooperation from separated sites on a promoter. Cell *108*, 57-70.

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y, Ishii, S., and Sarai, A. (1993). Recognition of specific DNA sequences by the *c-myb* protooncogene product: Role of three repeat unit in the DNA-binding domain. Proc. Natl. Acad. Sci. USA *90*, 9320-9324.

Thickman, K.R., Sickmier, E.A., and Kielkopf, C.L. Alternative conformations at the RNA-binding surface of the N-terminal U2AF$^{65}$ RNA recognition motif. J. Mol. Biol. *366*, 703-710.

Tollinger, M., Skrynnikov, N.R., Mulder, F.A.A., Forman-Kay, J.D., and Kay, L.E. (2001). Slow dynamics in folded and unfolded states of an SH3 domain. J. Am. Chem. Soc. *123*, 11341-11352.

Tsai, C.J., Kumar, S., Ma, B., and Nussinov, R. (1999) Folding funnels, binding funnels, and protein function. Protein Sci. *8*, 1181-1190.

Tzeng, S.R., and Kaldimos, C.G. (2012). Protein activity regulation by conformational entropy. Nature *488*, 236-240.

Yang, J., Wu, J., Steichen, J.M., Kornev, A.P., Deal, M.S., Li, S., Sankaran, B., Woods, V.L.Jr., Taylor, S.S. (2012). A conserved Glu-Arg salt bridge connects coevolved motifs that define the eukaryotic protein kinase fold. J. Mol. Biol. *415*, 666-679.

Yutani, K., Hayashi, S., Sugisaki, Y., and Ogasahara, K. (1991). Role of conserved proline residues in stabilizing tryptophan synthase subunit: analysis by mutants with alanine or glycine. Proteins: Struct. Funct. Genet. *9*, 90-98.

Wand, A.J. (2013). The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. Curr. Opin. Struct. Biol. *23*, 75-81.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. *337*, 635-645.

Weston, K. (1992). Extension of the DNA binding consensus of the chicken c-Myb and v-Myb protein. Nucleic Acid. Res. *20*, 3043-3049.

Wiseman, T., Williston, S., Brandts, J.F., and Lin, L-N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal. Biochem. *179*, 131-137.

Wolynes, P.G., Onuchic, J.N., and Thirumalai, D. (1995). Navigating the folding routes. Science *267*, 1619-1620.

Wong, L.E., Li, Y., Pillay, S., Frolova, L., and Pervushin, K. (2012). Selectivity of stop codon recognition in translation termination is modulated by multiple conformations of GTS loop in eRF1. Nucleic Acids Res. *40*, 5751-5765.

Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signaling and regulation. Nat. Rev. Mol. Cell. Biol. *16*, 18-29.

Zor, T., De Guzman, R.N., Dyson, H.D., and Wright, P.E. (2004). Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. J. Mol. Biol. *337*, 521-534.

**List of publications**

1. Satomi Inaba, Harumi Fukada, Takahisa Ikegami, and Masayuki Oda (2013) Thermodynamic effects of multiple protein conformations on stability and DNA binding, Archives of Biochemistry and Biophysics, 537(2), 225-232.

2. Satomi Inaba, Harumi Fukada, and Masayuki Oda (2016) Thermodynamic effects of a linker region between two repeats of a protein, c-Myb R2R3, on its stability and structural dynamics, Journal of Thermal Analysis and Calorimetry, 123 (3), 1763-1767.

3. Satomi Inaba, Akihiro Maeno, Kazumasa Sakurai, Sunilkumar Puthenpurackal Narayanan, Takahisa Ikegami, Kazuyuki Akasaka, and Masayuki Oda (2015) Functional conformer of c-Myb DNA-binding domain revealed by variable temperature studies, The FEBS Journal, 282 (23), 4497-4514.

4. Satomi Inaba, Harumi Fukada, and Masayuki Oda (2016) Folding thermodynamics of c-Myb DNA-binding domain in correlation with its α-helical contents, International Journal of Biological Macromolecules, 82, 725-732.

The other supplementary papers

1. Makoto Nakabayashi, Yoshito Tsukahara, Yukiko Iwasaki-Miyamoto, Mika Mihori-Shimazaki, Sachiko Yamada, Satomi Inaba, Masayuki Oda, Masato Shimizu,

Makoto Makishima, Hiroaki Tokiwa, Teikichi Ikura, Nobutoshi Ito (2013)

Crystal structures of hereditary vitamin D-resistant rickets-associated vitamin D receptor mutants R270L and W282R bound to 1,25-dihydroxyvitamin $D_3$ and synthetic ligands, Journal of Medical Chemistry, 56(17), 6745-6760.

2. Ayako Miki*, Satomi Inaba*, Takayuki Baba, Koji Kihira, Harumi Fukada, Masayuki Oda (2015)

   Structural and physical properties of collagen extracted from moon jellyfish under neutral pH conditions, Bioscience Biotechnology and Biochemistry, 79(10), 1603-1607.

   (*These authors contributed equally to the work)

**Acknowledgements**

The author would like to express my grateful acknowledgment to Dr. Masayuki Oda (Kyoto Prefectural University) for continuous encouragement, timely guidance, and helpful discussion. The author thanks Drs. Kazufumi Takano, Kunihiko Watanabe, Akito Ishida (Kyoto Prefectural University) for their valuable discussion and critical reading of this manuscript. The author also wishes to thank Dr. Kazuyuki Akasaka (Kyoto Prefectural University) for valuable discussions. The author is deeply grateful Dr. Harumi Fukada (Osaka Prefectural University) for her support of DSC measurements, Drs. Kazumasa Sakurai (Kinki University) and Akihiro Maeno (Wakayama Medical University) for their support of NMR and high-pressure experiments, Drs. Takahisa Ikegami (Yokohama City University) and Kenji Kanaori (Kyoto Institute Technology) for their support of NMR measurements and signal assignments, Dr. Yuji Kobayashi and Mr. Takahiro Maruno (Osaka University) for their supports of ITC measurements, Ms. Shigenori Nishimura (Osaka Prefecture University) for his supports of CD measurements. The author also appreciates Drs. Kazuhiro Ogata (Yokohama City University), Hisayuki Morii (National Institute of Advanced Industrial Science and Technology), and Haruki Nakamura (Osaka University) for valuable discussions. The author expresses grateful acknowledgements to Drs. Hiroshi Sekiguchi (JASRI/SPring-8), Nobutaka Numoto, Tekichi Ikura, Nobutishi Ito (Tokyo Medical and Dental University), Rieko Ishima (University of Pittsburgh), Takashi Yonetani (University of Pennsylvania), Yuji O. Kamatari (Gifu University), and Minoru Kato (Ritsumeikan University) for their kind supports. The author also thanks laboratory member of biophysical chemistry and my friends for their helpful assistance. Finally, I

am thankful to my parents and sisters for encouraging and supporting me all the time.

This work was supported in part of Grant-in-Aid for Japan Society for the Promotion of Science fellows (JSPS KAKENHI: Grant number 15J03576), the research grant for young scientist from Kyoto Municipal University Corporation, and the graduate student grant from Ueda Ayako Science Research Promotion Fund. Part of the NMR study was performed under the Cooperative Research Program of Institute for Protein Research, Osaka University (Prof. Haruki Nakamura).

March, 2016

Satomi Inaba