Thermostability of the N-Terminal RNA-Binding Domain of the SARS-CoV Nucleocapsid Protein: Experiments and Numerical Simulations

Huey-Jen Fang,[†‡] Yong-Zhong Chen,[†] Mai Suan Li,[§] Ming-Chya Wu,[†¶‖] Chun-Ling Chang,[†‖] Chung-ke Chang,[**] Yen-lan Hsu,[**] Tai-huang Huang,[**] Hueih-Min Chen,[‡‡‡] Tian-Yow Tsong,[‡‡] and Chin-Kun Hu[†§§*]

[†]Institute of Physics, [‡]Agricultural Biotechnology Research Center, and [**]Institute of Biomedical Sciences, Academia Sinica, Nankang, Taipei, Taiwan; [§]Institute of Physics, Polish Academy of Sciences, Warsaw, Poland, and Institute for Computational Science and Technology, Ho Chi Minh City, Vietnam; [¶]Research Center for Adaptive Data Analysis, [‖]Department of Physics, National Central University, Chungli, Taiwan; [‡‡‡]Nano Biosystem Technology Division, National Nano Device Laboratories, Hsinchu, Taiwan; [‡‡]Institute of Physics, Academy of Sciences, Taipei, Taiwan, and National Chiao Tung University, Hsinchu, Taiwan; and [§§]Center for Nonlinear and Complex Systems and Department of Physics, Chung Yuan Christian University, Chungli, Taiwan

ABSTRACT   Differential scanning calorimetry, circular dichroism spectroscopy, nuclear magnetic resonance spectroscopy, and numerical simulations were used to study the thermostability of the N-terminal RNA-binding domain (RBD) of the SARS-CoV nucleocapsid protein. The transition temperature of the RBD in a mixing buffer, composed of glycine, sodium acetate, and sodium phosphate with 100 mM sodium chloride, at pH 6.8, determined by differential scanning calorimetry and circular dichroism, is 48.74°C. Experimental results showed that the thermal-induced unfolding-folding transition of the RBD follows a two-state model with a reversibility >90%. Using a simple Gō-like model and Langevin dynamics we have shown that, in agreement with our experiments, the folding of the RBD is two-state. Theoretical estimates of thermodynamic quantities are in reasonable agreement with the experiments. Folding and thermal unfolding pathways of the RBD also were experimentally and numerically studied in detail. It was shown that the strand β_1 from the N-terminal folds last and unfolds first, while the remaining β-strands fold/unfold cooperatively.

INTRODUCTION

The SARS-CoV nucleocapsid protein (NP) is a major virion structural protein in severe acute respiratory syndrome (SARS) (1,2). It has 422 amino-acid residues, a molecular weight of 46 kDa, and shares 20–30% homology with the NP of other coronaviruses (1,2). Its major function is to assemble the viral RNA genome of coronavirus to form the helical ribonucleoprotein core. In vitro, the NPs, in the absence of genomic nucleic acid, exist as dimers at a low concentration, and may form higher oligomers at higher concentrations (3). The NP does not have any cysteine residues, thus it contains no disulfide bond, and may be less stable than other structural proteins of SARS-CoV. Experiments showed that it is most stable at pH 9.0 (3). Thermal and chemical denaturant-induced denaturation analysis (3) showed that the unfolding and refolding of oligomeric NP can be described by a two-state model. Chemical unfolding is found to be reversible (3,4), while both irreversibility (3) and reversibility (4) of the thermal-induced unfolding in different buffers have been reported. Recently, Wang et al. (4) reported that the NP in phosphate-buffered saline buffer starts to unfold at 35°C and is completely denatured at 55°C. The stability of the NP is critical for the function of the SARS virus, which is completely noninfectious at ~56°C (5).

Quite recently, Chang et al. (6) found that the NP consists of two independent structural domains—the N-terminal RNA-binding domain (RBD) (residues 45–181) and the

C-terminal dimerization domain (DD) (residues 248–365), surrounded by flexible linkers. Fig. 1 shows a model for the overall structure of SARS-CoV NP (6). The two structural domains carry out two distinct functions. The N-terminal domain is able to bind RNA, while the C-terminal domain acts as a dimerization domain. Hsieh et al. (7) showed that the DD could also bind to an RNA packing signal, and Chen et al. (8) proposed an octamer model (note that the structure of the octamer can be accessed at the Protein Data Bank, PDB code: 2CJR) to demonstrate a possible framework for helical packing of viral RNA. Whereas the two structural domains do not interact with each other, they could act in concert to carry out biological functions with the aid of the long flexible linker between the two domains (6). In this context, two structural domains are logical candidates for further studies. We then investigated the thermostability of the two domains separately, instead of using the full-length of the NP. In this article, we focus on studying the RBD, and the study of the DD domain will be reported elsewhere.

The ability of the RBD to bind RNA is closely related to its structure. Fig. 2 shows the cartoon representation of the structure of the RBD (PDB code: 1SSK) (10). An additional 21 residues have been added to the N-terminal in the structure determination, so there is a total of 158 residues in this PDB structure, in which residues 22–158 correspond to residues 45–181 of the SARS-CoV NP. Nuclear magnetic resonance (NMR) studies indicate that the RBD has a five-stranded β-sheet, and all residues are solvent-exposed (6), which implies that it is essentially a loosely packed protein.
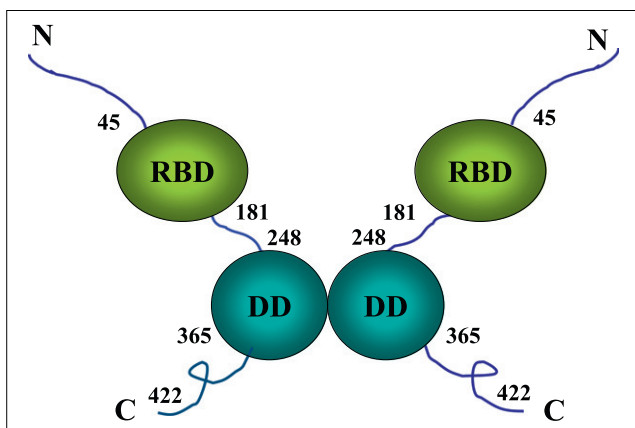
FIGURE 1   A model of the overall structure of SARS-CoV NP. The RBD and DD are structural domains, and lines represents disordered segments.

We studied the thermostability of the RBD by differential scanning calorimetry (DSC) (11,12), circular dichroism (CD) spectroscopy, and NMR spectroscopy. Our experiments show that the transition midpoint of thermal denaturation of the RBD at pH 6.8 is 48.74°C, and the transition curve of the RBD can be fitted nicely with a two-state model.
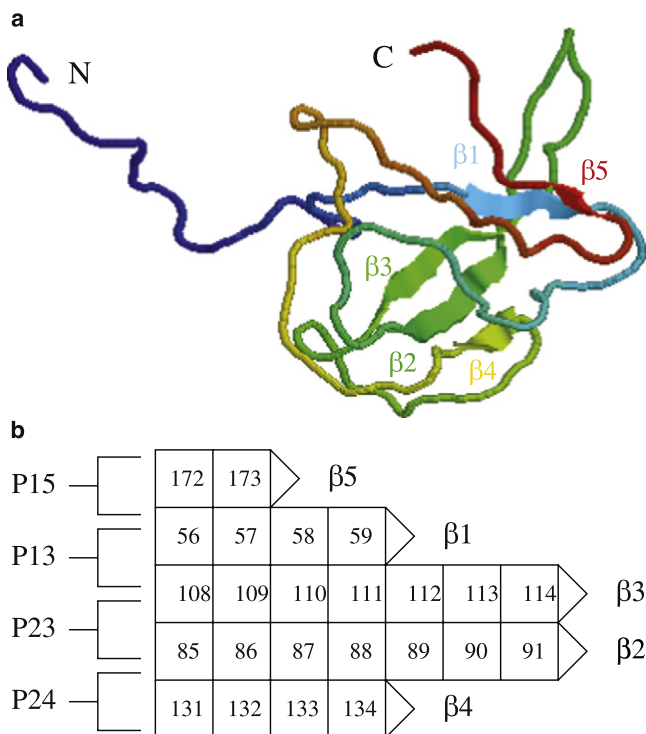

FIGURE 2   (*a*) Cartoon representation of the structure of the RBD. The Protein Data Bank accession code is 1SSK. There are five β-strands: β_1(aa 56–59, numbered according to the full length of NP), β_2(aa 108–114), β_3(aa 85–91), β_4(aa 131–134), and β_5(aa 172–173). In the native state, these strands have 30, 49, 48, 24, and 11 contacts, respectively. (*b*) A schematic plot for four pairs of strands P_{13}, P_{15}, P_{23}, and P_{24}. Here P_{13}, for example, means the pair between strands 1 and 3. In the native state, there exists only the interactions between strands of these pairs. The numbers of native contacts of these pairs are equal to 10, 6, 22, and 13, respectively.

Further, NMR spectroscopy experiments conducted under equilibrium condition suggests that the N-terminus folds last at low temperature and unfolds first at high temperature. The remaining four β-sheets in the (un)folding process are highly cooperative.

We have also performed numerical studies of the folding thermodynamics and kinetics of RBD using the simple Gō modeling approach (13,14). Using the Langevin dynamics and the reweighting technique (15), relevant thermodynamic quantities were computed and compared with the experiments. From the double-minimum structure of the free energy, plotted as a function of the number of native contacts, we conclude that the RBD is indeed a two-state folder. The reversibility of thermal folding of RBD was theoretically investigated by considering folding and unfolding pathways of five β-strands. The results are substantially congruent with the experimental results, while the folding of β-strands is less cooperative.

The rest of the article is organized as follows. In Materials and Methods, we briefly present sample preparation, experimental procedures, and simulation details. The experimental and simulation results are summarized in the Results and, finally, our results are addressed in the Discussion.

MATERIALS AND METHODS

Sample preparation

The SARS-CoV TW1 strain cDNA sequencing clones were provided by Dr. P.-J. Chen of National Taiwan University Hospital (16). The RBD of the SARS CoV NP (residues 45–181) was cloned into the pET6H vector as described in Chang et al. (17). The fragment was expressed in *Escherichia coli* BL21(DE3) cells overnight at 37°C in Luria-Broth media without inducing agents. For purification, 10 amino-acid residues (MHHHHHHAMG) have been added to the N-terminal of RBD as His-tag, such that the final RBD sample has a total of 147 amino-acid residues, in which residues 11–147 correspond to residues 45–181 of the SARS-CoV NP. The protein was purified with Ni-NTA affinity column (Qiagen, Valencia, CA) in buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.4) containing 7 M urea. The protein was then allowed to refold by gradually lowering the denaturant concentration through dialysis in liquid chromatography buffer (50 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA, 0.01% NaN$_3$, pH 7.4). Renatured protein was loaded onto an AKTA-EXPLORER fast performance liquid chromatography system equipped with a HiLoad 16/60 Superdex 75 column (Amersham Pharmacia Biotech, Uppsala, Sweden). Complete Protease Inhibitor cocktail (Roche, Mannheim, Germany) was added to the purified protein. Protein concentration was determined with the Bio-Rad Protein Assay kit as per instructions from the manufacturer (Bio-Rad, Hercules, CA). The correct molecular weights of the expressed proteins were confirmed by mass spectroscopy.

Differential scanning calorimetry

The conformational stability of the RBD is investigated by analyzing the unfolding transition induced by elevating temperature. The experiments were carried out with a VP-DSC calorimeter from MicroCal (Northampton, MA) at a scan rate of 1.0°C/min. The samples were degassed for 15 min at room temperature before the calorimetric experiments. The thermograms were measured in the temperature range 10–80°C. The sample was in a mixing buffer, which was composed of glycine, sodium acetate, and sodium phosphate with 100 mM sodium chloride, at pH 6.8. Fitting of the

thermal unfolding/refolding transitions were carried out by the computer software MicroCal Origin Ver. 7.0.

Circular dichroism

The CD spectra were measured on a Jasco (Tokyo, Japan) J-715 spectropolarimeter at different temperatures with wavelength scanning taken from 250 to 200 nm. All experiments were carried out at 2-nm bandwidth in a 1-mm quartz square cuvette, thermostated to $\pm 0.1°C$. Protein concentration was ~20 μM and was determined by the BCA Protein Assay Kit (Pierce, Rockford, IL). Protein was dissolved in 30 mM mixing buffer containing glycine, sodium acetate, and sodium phosphate with 100 mM sodium chloride at pH 7.0. Temperature control was achieved by using a circulating water bath system and the equilibrium time was 10 min for each temperature point.

NMR spectroscopy

^{13}C, ^{15}N-labeled samples were prepared as previously described. Samples were exchanged to NMR buffer (10 mM sodium phosphate, pH 6.0, 50 mM NaCl, 1 mM EDTA, 0.01% NaN$_3$, 1 mM DSS, 10% D2O) with an Amicon Ultra-15 concentrator (Millipore, Billerica, MA). Final sample concentration was ~1.5 mM. Variable temperature ^{15}N-edited heteronuclear single-quantum coherence (HSQC) spectra were acquired on a Bruker AV600 spectrometer (Bruker, Rheinstetten, Germany) equipped with a normal temperature probe. Spectra were acquired using spectral widths of 12 ppm × 34 ppm and 1024 × 128 complex points for the ^1H and ^{15}N dimensions, respectively. The sample was allowed to equilibrate at the target temperature for at least 30 min before experiments were conducted. These spectra were processed with iNMR (Nucleomatica, Molfetta, Italy), which was also used to extract peak intensities. Graphs were generated with Microsoft Excel (Microsoft, Redmond, WA). Projection-reconstruction three-dimensional HNCO spectra were acquired on a Bruker AV500 spectrometer equipped with a cryoprobe. Spectra were acquired using spectral widths of 12 ppm × 34 ppm × 12 ppm and 1024 × 64 × 64 complex points on the ^1H, ^{15}N, and ^{13}C dimensions, respectively. Spectra were processed with TopSpin 2.1 (Bruker) and the reconstructed spectra were analyzed with CARA (18). Digital resolution on the ^{13}C dimension after processing was 0.046 ppm. Secondary shifts were calculated from the random coil shifts tabulated by Schwarzinger et al. (19,20), and corrected for sequence composition (19,20). The secondary shifts were graphed with GraphPad Prism (GraphPad Software, La Jolla, CA).

Simulation details

To simulate the thermal folding and unfolding of the RBD, we utilize the Gō-like model (14), in which amino-acid residues are represented by point particles located at positions of C$_\alpha$ atoms. This model is defined through the experimentally determined native structure, and it captures essential aspects of the important role played by the native structure (14,21). In our simulations, the lowest energy native structure was taken from the PDB (PDB ID: 1SSK), and only the positions of 158 C$_\alpha$-carbons are considered. The energy of the Gō-like model has the following form (14):

$$E = \sum_{\text{bonds}} K_r (\Delta r_i)^2 + \sum_{\text{angles}} K_\theta (\Delta \theta_i)^2 + \sum_{\text{dihedral}} \left[K_\phi^{(1)} (1 - \cos\Delta\phi_i) \right.$$
$$\left. + K_\phi^{(3)} (1 - \cos 3\Delta\phi_i) \right]$$
$$+ \sum_{i>j+3}^{NC} \epsilon_H \left[5 \left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{0ij}}{r_{ij}} \right)^{10} \right] + \sum_{i>j+3}^{NNC} \epsilon_H \left(\frac{C}{r_{ij}} \right)^{12}.$$

(1)

Here $\Delta r_i = r_{i, \, i+1} - r_{0i, \, i+1}$, $\Delta \theta_i = \theta_i - \theta_{0i}$, $\Delta \phi_i = \phi_i - \phi_{0i}$, and $r_{i, \, i+1}$ is the distance between residues i and $i + 1$, θ_i is the bond angle between bonds

($i - 1$) and i, ϕ_i is the dihedral angle around the ith bond, and r_{ij} is the distance between the ith and j^{th} residues. Subscripts 0, NC, and NNC refer to the native conformation, native contacts, and nonnative contacts, respectively. Residues i and j are in native contact if r_{0ij} is less than a cutoff distance $d_c = 7.5$ Å, where r_{0ij} is the distance between the residues in the native conformation. The first harmonic term in Eq. 1 accounts for chain connectivity and the second term represents the bond-angle potential. The potential for the dihedral angle degrees of freedom is given by the third term in Eq. 1. The interaction energy between residues that are separated by at least three residues is given by 10-12 Lennard-Jones potential, given by the fourth term. A soft sphere repulsive potential, represented by the fifth term in Eq. 1, disfavors the formation of nonnative contacts. We choose $K_r = 100 \ \epsilon_H/Å^2$, $K_\theta = 20\epsilon_H/rad^2$, $K_\phi^{(1)} = \epsilon_H$, and $K_\phi^{(3)} = 0.5\epsilon_H$, where ϵ_H is the characteristic hydrogen-bond energy (14) and $C = 4$ Å.

We assume that the dynamics of the polypeptide chain obeys the Langevin equation (22). To compute thermodynamic quantities, we performed simulations at the friction $\zeta = 2\frac{m}{\tau_L}$, where the folding is fast. Here $\tau_L = (ma^2/\epsilon_H)^{1/2} \approx 3$ ps, m is the typical mass of amino acids, and a is the distance between two neighboring residues. For this value of friction, the equations of motion (see (23–25) for more details) were integrated using the velocity form of the Verlet algorithm (26) with the time step $\Delta t = 0.005\tau_L$. The equilibrium simulations have been carried out at $T = 0.45$, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, and 1.1 ϵ_H/k_B. For each temperature, 15 independent trajectories of $1.2 \times 10^6\tau_L$ were generated to have good statistics. The first period of $1.2 \times 10^5\tau_L$ was spent on equilibration and it was discarded from the data collection. Histograms for the energy and native contacts have been collected. Then the reweighting technique (15) was used to obtain the thermodynamic quantities at arbitrary temperature.

Since folding and unfolding pathways are kinetic in nature, they may depend on the friction. Thus, to mimic real experiments, kinetic simulations were carried out for the friction $\zeta = 50\frac{m}{\tau_L}$, which is believed to correspond to the viscosity of water (27). In this overdamped limit, we use the Euler method for integration and the time step $\Delta t = 0.1\tau_L$. To probe folding pathways, for i^{th} trajectory we introduce the progressive variable $\delta_i = t/\tau^i_F$, where τ^i_F is the folding time (28). One can then average the fraction of native contacts over many trajectories in a unique time window $0 \leq \delta_i \leq 1$, and monitor the folding sequence with the help of the progressive variable δ. The thermal unfolding pathways may be probed using a similar progressive variable, but instead of folding times we use unfolding ones, i.e., $\delta_i = t/\tau^i_U$, where τ^i_U values are unfolding times. We define τ_U as an average of first passage times to reach a conformation, which has no native contacts. Different trajectories start from the same native conformation but with different random number seeds.

RESULTS

DSC experiments

A typical curve for the specific heat C_p of the RBD is shown in Fig. 3. The transition temperature at pH 6.8 in the presence of 100 mM NaCl is $T_M = 48.74°C$. The reversibility of thermal denaturation was estimated by the ratio of the maximum value of C_p^{max} of subsequent scans and found to be >90%, suggesting that the thermal transition of the RBD is reversible. The sharp peak of the specific heat curve suggests that RBD is a two-state folder. Whereas the fitting of the transition may be carried out on the basis of the two-state model, here we assume a generalized situation instead of specific fitting functions (of a two-state folding model) and calculate thermodynamic quantities directly from definitions of such quantities.
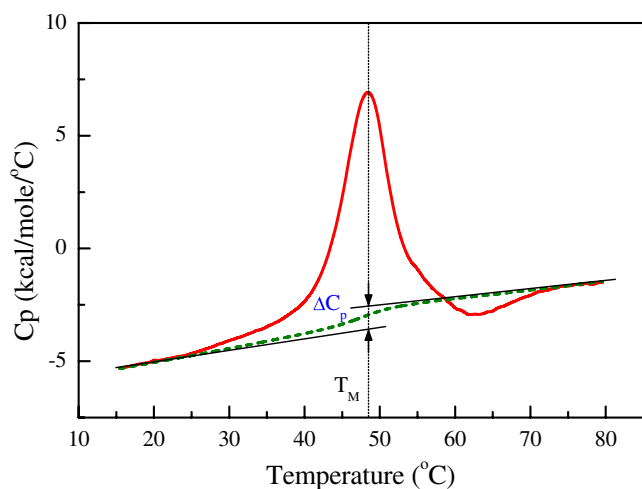
FIGURE 3   Temperature dependence of the RBD partial heat capacity in a mixing buffer of glycine, sodium acetate, and sodium phosphate, with 100 mM sodium chloride, at pH 6.8 in the presence of 100 mM NaCl. The transition temperature T_M is determined to be 48.74°C. The dashed line indicates the baseline curve and $\Delta C_p \approx 1.01$ kcal/mol/K is the difference between extrapolations of the baseline at T_M.

Table 1 presents the thermodynamic parameters obtained from the DSC measurements. The calorimetric enthalpy ΔH_M was calculated by integrating the area enclosed by the specific heat curve and a proper baseline curve (dashed line in Fig. 3). The increment of specific heat ΔC_p was measured by calculating the difference between the extrapolations of the specific heat curves at two ends. The entropy ΔS_M at T_M, the temperature of maximum stability T_S, the temperature with zero enthalpy T_H, and the conformational stability ΔG_S at T_S were determined by

$$\Delta S_M = \frac{\Delta H_M}{T_M}, \qquad (2)$$

$$\ln\frac{T_M}{T_S} = \frac{\Delta H_M}{T_M \Delta C_p}, \qquad (3)$$

$$T_H = T_M - \frac{\Delta H_M}{\Delta C_p}, \qquad (4)$$

$$\Delta G_S = \Delta C_p(T_S - T_H). \qquad (5)$$

CD spectra

The CD spectra for the RBD at pH 7.0 in the presence of 100 mM NaCl are presented in Fig. 4. The profiles of the CD curves peak at wavelength at ~225 nm. We determine the folding fraction from the CD curves at wavelengths 208 nm, 212 nm, and 225 nm, using the equation

$$f = \frac{\theta - \theta_D}{\theta_N - \theta_D}, \qquad (6)$$

where θ represents the observed mean residue ellipticity under the given conditions, and θ_N and θ_D are the corresponding ellipticities in the native (N) and denatured (D) states, respectively. The results are shown in Fig. 5. The melting temperature determined by the CD spectra is 48.5°C, consistent with the DSC measurement.

NMR spectroscopy

The peak intensity of the backbone amides of a protein in a ^{15}N-edited HSQC spectrum is a function of the exchange rate with the solvent, the correlation time of the molecule, and the population of the state. Increased tumbling of the molecule results in stronger peak intensities, whereas increased exchange rate lowers the intensities. When the temperature is raised, both the tumbling rate and exchange rates increase. Residues that are involved in hydrogen bonding, e.g., secondary structure formation, usually have slower exchange rates and are less prone to lose intensity when the temperature is raised. Thus, a plot of peak intensities versus temperature provides a qualitative view of the (un)folding process under equilibrium conditions. Such a plot for the RBD is shown in Fig. 6 for the β-sheet residues. The residues comprising the β-sheet fall into three categories: Q59 (numbered according to the full length of NP), which shows extreme loss of intensity even at moderate temperatures; T58 and Y110, which lose intensity at an intermediate rate; and the rest, which lose intensity only at higher temperatures. Our results suggest that the β_1 strand, which includes T58 and Q59, has a lower thermal stability compared to the other strands in the β-sheet. Thus, the strand β_1 likely unfolds first and folds last in the (un)folding process. The fact that most residues exhibit a transition at approximately the same temperature implies a high degree of cooperativity of the (un)folding process.

Another way to analyze secondary structures is through the use of secondary shifts, which is defined as the difference between the observed chemical shift and the average random-coil chemical shift. This approach requires triple-resonance NMR experiments. We opted to observe only the carbonyl shifts because it employs a highly sensitive experiment and is sensitive to both the α- and β-structures. Positive values imply α-propensity, whereas negative values

TABLE 1   Thermodynamic parameters obtained from the DSC measurements and the Gō model simulations

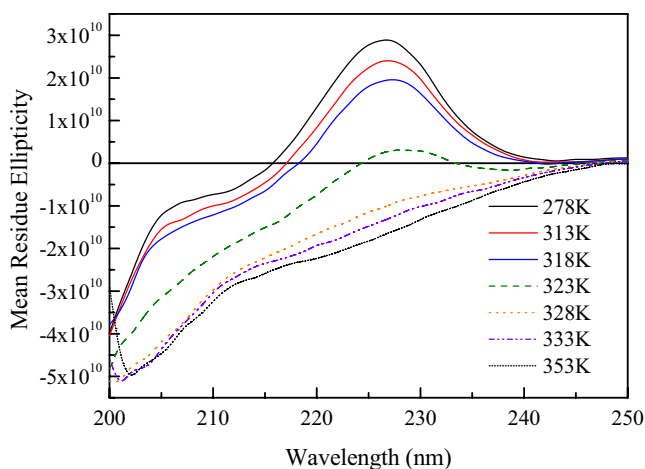| Domain	T_M (K)	ΔH_M (kcal/mol)	ΔC_p (kcal/mol/K)	ΔS_M (cal/mol/K)	T_S (K)	T_H (K)	ΔG_S (kcal/mol)
RBD (exp)	321.9 ± 0.27	81.99	1.01	254.71	250.15	240.72	9.52
RBD (sim)	321.9	69.02	0.92	214.41	254.98	246.88	7.45

FIGURE 4   CD spectra of the RBD in a mixing buffer of glycine, sodium acetate, and sodium phosphate, at pH 7.0 in the presence of 100 mM NaCl.

imply β-propensity. We carried out this experiment at three different temperatures with the results summarized in Fig. 6. Attempts to conduct the experiments at higher temperature were not feasible due to low signal/noise ratio. In this experiment, Y110 shows marginal β-propensity across all temperatures tested. Coupled with the results shown in Fig. 7, it seems that Y110 could serve as a starting point for (un)folding of the $\beta_2/\beta_3/\beta_4$ sheet. Surprisingly, a number of residues located at the β_2 (Y87, Y88, and R89) and β_4 (I131 and W133) strands showed non-beta propensities at the temperatures tested. Higher temperatures led to a more random coil-like propensity for these residues. Whether there is a correlation between the secondary structure propensity of these residues and their thermal stability requires further investigation.
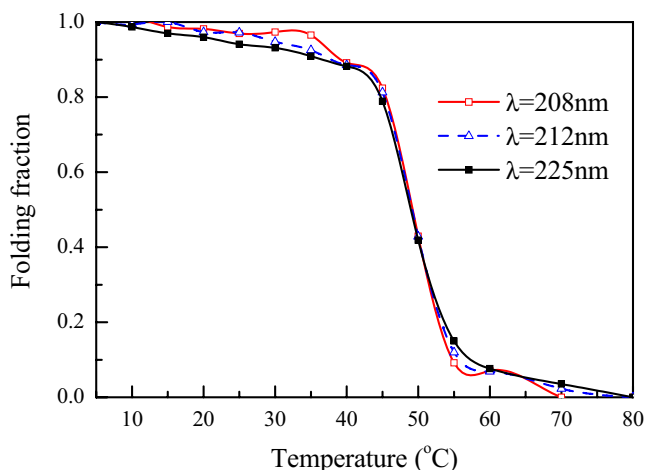

FIGURE 5   Folding fractions of the RBD at pH 7.0 with respect to different temperature, determined by the CD spectra of Fig. 4 at $\lambda = $ 208 nm, 212 nm, and 225 nm.

Simulation results

Thermodynamic properties

The specific heat C_p was calculated using the formula $C_p(\langle E^2 \rangle - \langle E \rangle^2)/T^2$, where E is the total energy of the system, and $\langle \cdots \rangle$ stands for thermodynamic averaging. The averaging procedure was carried with the help of the reweighting technique (see Materials and Methods). As followed from our simulations, the maximum of the specific heat locates at the melting temperature $T_M = 0.8\varepsilon_H/k_B$. Identifying this value with the experimental value $T_M = 321.9$ K we obtained the characteristic energy in our model $\varepsilon_H = 0.73$ kcal/mol. The simulated temperature dependence of C_p (Fig. 8 a) is shown in the same temperature interval as for the experimental curve in Fig. 3. The maximum value $C_p^{max} \approx 6.68$ kcal/mol/K is not far from the experimental result (Fig. 3).

The jump of specific heat at the melting temperature ΔC_p was defined as the difference between the base lines from two ends (Fig. 8 a). We obtain $\Delta C_p \approx 0.92$ kcal/mol/K, which is also closed to the experimental estimate 1.01 kcal/mol/K. Due to the uncertainty in determination of baselines (Fig. 3 and Fig. 8 a), both the theoretical and experimental value of ΔC_p should be considered as rough estimates. Therefore, it is not clear whether the agreement between the Gō modeling and experiments is just fortuitous for the RBD or robust for other proteins. Elucidating this issue requires further investigation.

Defining the calorimetric enthalpy ΔH_M as the area enclosed by the specific heat curve and a proper baseline curve (Fig. 8 a), we obtained $\Delta H_M = 69.02$ kcal/mol. Using $T_M = 321.9$ K, Eq. 2 gives the jump of entropy at the denaturation transition $\Delta S_M \approx 214.41$ cal/mol/K, which is a bit lower than the experimental one (Table 1). Substituting ΔH_M, T_M, ΔS_M, and ΔC_p to thermodynamic relations given by Eqs. 3–5, T_S, T_H, and ΔG_S were also calculated and listed in Table 1. Clearly, the agreement between the theory and experiments is reasonable.

The sharpness of the folded→unfolded transition might be quantitatively characterized via the cooperativity index Ω_c defined by (29,30,31)

$$\Omega_c = \frac{T_M^2}{\Delta T}\left(\frac{df_N}{dT}\right)_{T=T_M}, \qquad (7)$$

where ΔT is the transition width, defined as the full width at half-maximum of $|df_N/dT|$. Using the experimental and theoretical values of ΔH_M (Table 1), and the two-state formula for f_N, we obtain the temperature dependence of $|df_N/dT|$ shown in Fig. 8 b. Then, Eq. 7 gives $\Omega_c = 4456$ and 3195 for DSC experiments and the Gō model simulation, respectively. We can also estimate Ω_c from its dependence on the number of amino acids N (30,32). From the data set collected for 34 proteins (32), we have $\Omega_c \approx 0.0456 \times N^\mu$, where exponent μ is universal and expressed via the random walk susceptibility exponent γ as $\mu = 1 + \gamma$. Using $\mu \approx 2.22$
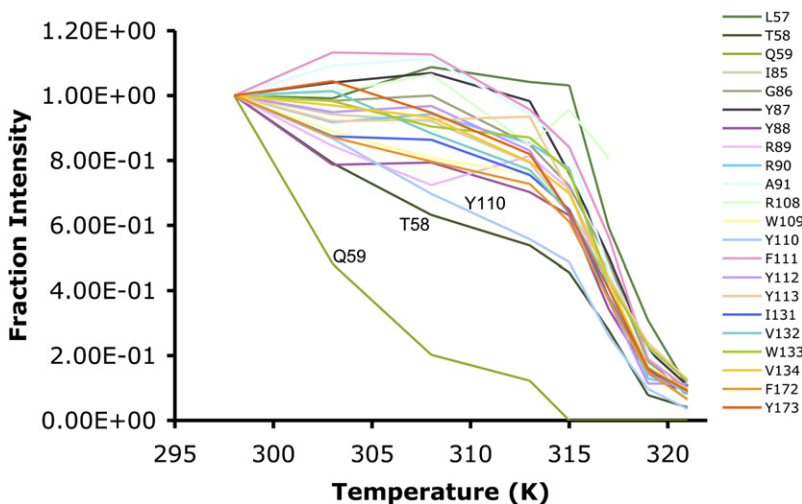
FIGURE 6 The peak intensity of the backbone amides in a ^{15}N-edited HSQC spectrum as a function of temperature for β-sheet residues, L57, T58, Q59 (β_1), I85, G86, Y87, Y88, R89, R90, A91 (β_2), R108, W109, Y110, F111, Y112, Y113 (β_3), I131, V132, W133, V134 (β_4), and F172, Y173 (β_5).

($\gamma \approx 1.22$) and $N = 158$, we obtain $\Omega_c \approx 3467$, which is in reasonable agreement with the simulation and experimental estimate. Thus, one can use the scaling law to roughly estimate the cooperativity index of RBD.

RBD folds without intermediates

Assuming the number of native contacts as a good reaction coordinate, we constructed the free energy landscape (Fig. 9 a). Since two minima, which correspond to the denatured and folded states, are separated by only one transition state, the RBD is a two-state folder. This is in qualitative agreement with our experimental finding. However, some remarks are in order. First, the two-state behavior is obvious for the β-protein RBD. Second, the all-or-none folding of the RBD is not a mere consequence of Gō modeling, because

this model correctly captures not only the two-state folding of proteins CI2 and SH3 (more two-state Gō folders may be found in (33)) but also intermediates of the three-state folder barnase, RNase H, and CheY (14). The reason for this is that the simple Gō model ignores the energetic frustration but it still takes the topological frustration into account. Therefore, it can capture intermediates that occur due to topological constraints but not those emerging from the frustration of the contact interactions. The fact that the simple Gō model correctly captures the two-state behavior as was observed in our experiments, suggests that the energetic frustration ignored in this model plays a minor role compared to the topological frustration (14). Even though the Gō model by itself cannot be employed to ascertain the two-state behavior of proteins, one can use it in conjunction with experimental methods.

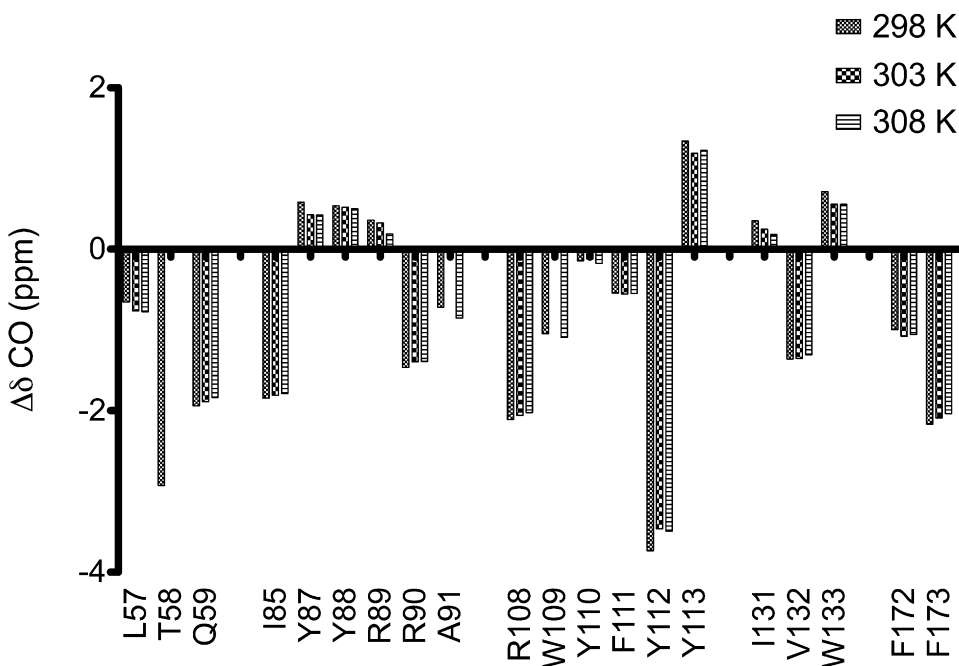FIGURE 7 Carbonyl shifts at three different temperatures, 298 K, 303 K, and 308 K, for β-sheet residues L57, T58, Q59 (β_1), I85, G86, Y87, Y88, R89, R90, A91 (β_2), R108, W109, Y110, F111, Y112, Y113 (β_3), I131, V132, W133, V134 (β_4), and F172, F173 (β_5).
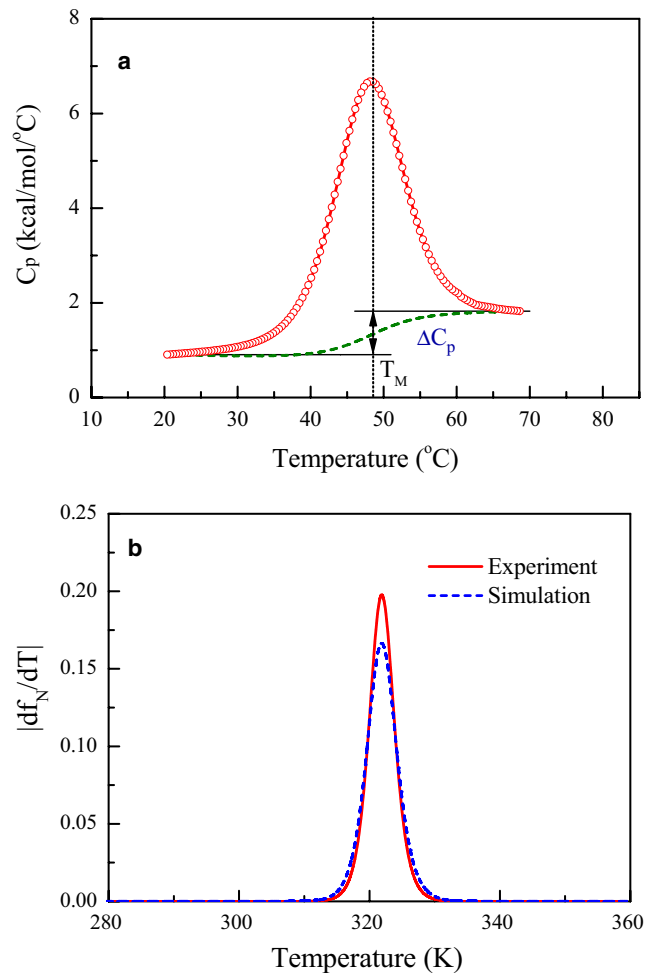
FIGURE 8 (*a*) Temperature dependence of the specific heat obtained in the Gō model. We use the same temperature interval as in the experiment (Fig. 3). We have $\Delta C_p \approx 0.92$ kcal/mol/K. (*b*) The temperature dependence of df_N/dT obtained from the experiments and simulations. This quantity was calculated using $f_N = 1/\{1 + \exp[-\frac{\Delta H_M}{T}(1 - \frac{T}{T_M})]\}$, where ΔH_M and T_M are taken from Table 1.

As evident from Fig. 9 *a*, at the denaturation transition point the folding/unfolding barrier is rather low (≈ 1 kcal/mol). This presumably reflects shortcomings of the Gō modeling. Account of nonnative contacts would make a protein energetically frustrated (in addition to the topological frustration), leading to the barrier enhancement.

Using Fig. 9 *a*, we define transition state (TS) conformations as those that have the number of native contacts $130 \leq Q \leq 180$. Then one can sort out all of these conformations for a structural analysis. Fig. 9 *b* shows the distribution of the root-mean-square deviation (RMSD) for the denatured state (DS), TS, and folded state (FS). Representative snapshots of these states are shown in Fig. 9 *a*. The average RMSD of TS $\langle RMSD \rangle_{TS} \approx 7$ Å, and the average gyration radius $\langle R_g \rangle_{TS} = 21.67$ Å. For the PDB structure the gyration radius is $R_g^{PDB} = 19.92$ Å. Since $\langle R_g \rangle_{TS}/R_g^{PDB} = 1.088$ is close to unity, TS structures are relatively compact compared to the folded ones. It is also clear from snapshots shown in Fig. 9 *a*.
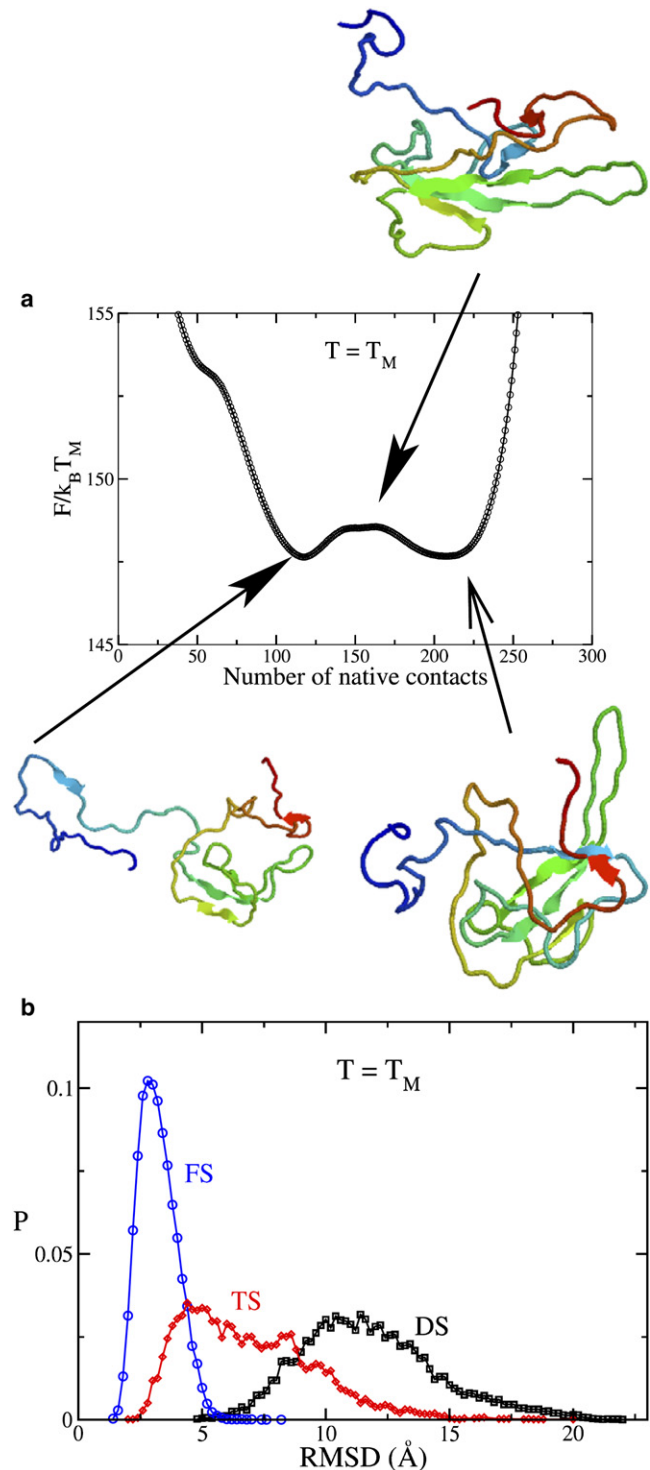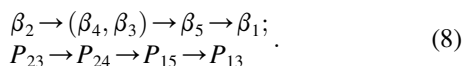

FIGURE 9 (*a*) Free energy as a function of the number of native contacts at $T = T_M$. Typical snapshots are shown. (*b*) The distribution of RMSD for the denatured state (DS), transition state (TS), and folded state (FS) at $T = T_M$. The average values of RMSD are $\langle RMSD \rangle_{DS} = 11.9$ Å, $\langle RMSD \rangle_{TS} = 7.0$ Å, and $\langle RMSD \rangle_{FS} = 3.3$ Å.
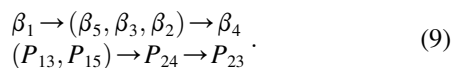
Reversibility of thermal folding

As discussed above, our experiments showed that the thermal folding of the RBD is highly reversible. Here, we present simulation results obtained at the water viscosity $\zeta = 50\frac{m}{\tau_L}$, to check this experimental observation. To see whether the thermal unfolding is the reverse of folding, we study the folding and unfolding pathways using the progressive variable δ as a reaction coordinate (see Simulation Details). Fig. 10 a shows the dependence of native contact fractions on δ for five β-strands at $T = 285$ K, which is $<T_M$. The same as in Fig. 10 a, but for four-strand pairs, is presented in Fig. 10 b. Clearly, we have the following folding sequencing:

$$\begin{aligned} \beta_2 &\rightarrow (\beta_4, \beta_3) \rightarrow \beta_5 \rightarrow \beta_1; \\ P_{23} &\rightarrow P_{24} \rightarrow P_{15} \rightarrow P_{13} \end{aligned} \qquad (8)$$

Thus, the folding initiates from the core by forming native contacts between strands β_2 and β_3. The native contacts, which belong to strands β_3 and β_4, form almost at the same timescales. The N-terminus is expected to fold last, when all contacts between strands β_1 and β_3 occur.

The time dependence of native contact fractions at $T = 450$ K for the β-strands and their pairs is shown in Fig. 11. In this case, we have the following unfolding pathways:

$$\begin{aligned} \beta_1 &\rightarrow (\beta_5, \beta_3, \beta_2) \rightarrow \beta_4 \\ (P_{13}, P_{15}) &\rightarrow P_{24} \rightarrow P_{23} \end{aligned} \qquad (9)$$

It should be noted that the pathways given by Eqs. 8 and 9 are valid in the statistical sense. Our detailed analysis shows that strand β_1, e.g., unfolds first in 48% of events (Eq. 9), while β_5 and β_4 unfold first in 42% and 10% of events, respectively. Comparing Eq. 8 and Eq. 9, one can see that in terms of strand pairs, the unfolding is the reverse of folding. In this sense, our simulations capture the experimental fact that the folding of the RBD is highly reversible. However, it is not entirely true if one considers individual strands. In fact, assuming that the thermal unfolding is reverse of folding, then from Eq. 8 it would follow that strand β_2 unfolds last, but β_4 is supposed to unfold last according to Eq. 9. Notably, Eq. 8 and Eq. 9 unambiguously show that statistically β_1 folds last at low temperatures, and it unfolds first at high temperatures. These results are congruent with the experimental results.

DISCUSSION

We have carried out the DSC, CD, and NMR spectroscopy, and Gō-like model simulation to investigate the thermostability of the RBD. We find that even though the thermal-induced unfolding of the full-length NP is irreversible in the same buffer condition in our DSC measurements (results not shown), the RBD is a highly reversible segment. The transition midpoint is determined to be 48.74°C located in the range of the unfolding temperature of the NP (35–55°C
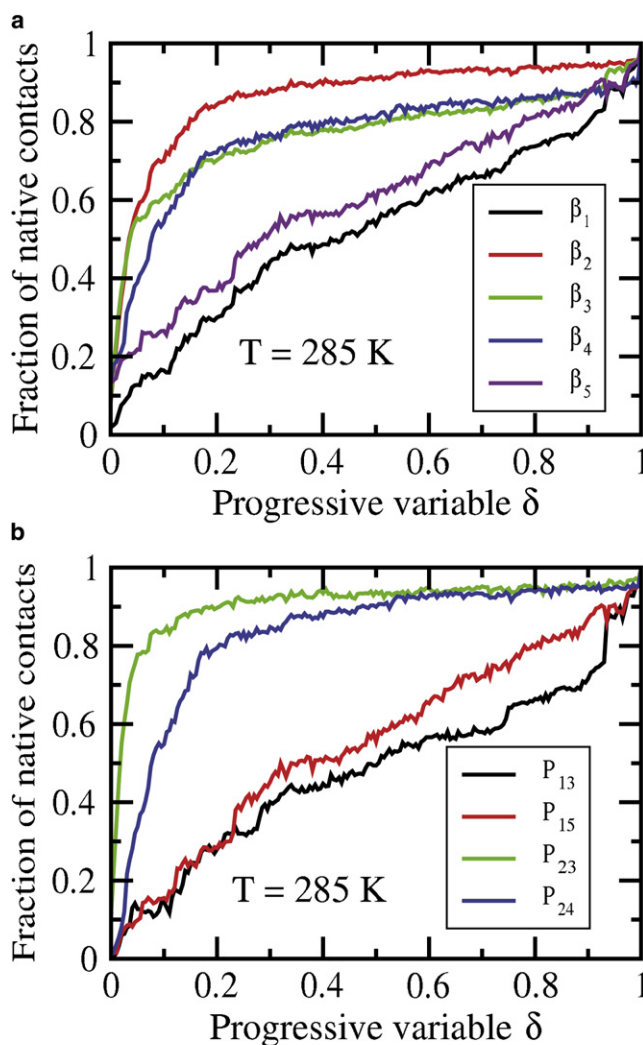

FIGURE 10 (*a*) Time dependence of the fraction of native contacts of five β-strands at $T = 285$ K. The definition of these strands is given in the caption of Fig. 2. (*b*) The same as in panel *a*, but for four pairs of β-strands (see Fig. 2 for definitions).

with a T_M of 43°C (4)). The thermal unfolding of the full-length NP was studied by fluorescence technique that monitors primarily the structural changes at the vicinities of the five tryptophan residues, three located in the RBD domain and two in the DD domain. Thus, the results are the composite effects of the melting of the two domains and as such, the observed melting is broad. The observed lower T_M for the NP also suggests that the DD appears to be less stable and more crucial to the properties of the NP, and predominantly determines the stability of the framework of the DD octamer (8) in forming helical ribonucleoprotein core. The RBD and DD do not interact with each other. It is unclear how the two domains contribute to the high irreversibility. This issue requires further investigation.

Our DSC and CD spectra analyses and the free-energy landscape constructed by the Gō-like model suggest that the folding-unfolding of the RBD can be treated as two-state
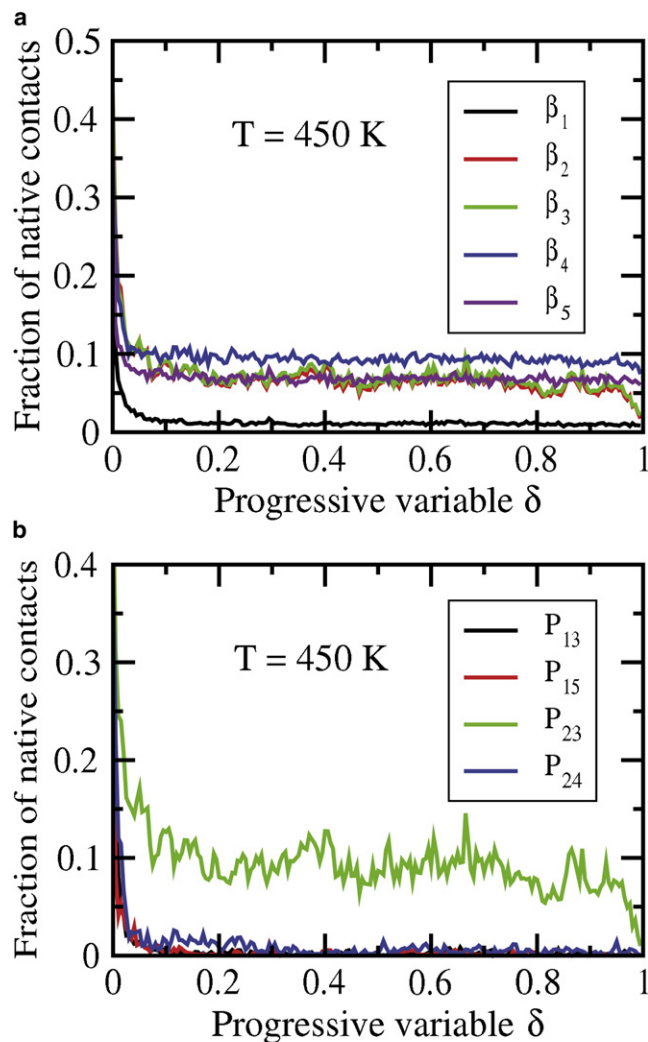
FIGURE 11 (*a*) Time dependence of the fraction of native contacts of five β-strands at $T = 450$ K. The definition of these strands is given in the caption of Fig. 2. Results were averaged over 100 trajectories. (*b*) The same as in panel *a*, but for four pairs of β-strands (see Fig. 2 for definitions).

within the experimental timescale. The Gō model by itself can not be employed to ascertain the two-state behavior of proteins. However, one can use it in conjunction with experiments to probe the two-state folding because this model does not always provide the two-state behavior as have been clearly shown in the work of Clementi et al. (14). It can capture intermediates that occur due to topological constraints but not those emerging from the frustration of the contact interactions. Although the folding process in the Gō model is less cooperative compared to what observed experimentally, we have obtained good estimates for other thermodynamic quantities. Given the simplicity of the Gō modeling, our theoretical results should be considered as reasonable. In the future, it is of interest to simulate RBD with all-atom models (34–38) to see whether such simulated results can be more cooperative. In addition, our present simulations show that the folding of the RBD initiates

from the core and the N-terminus folds last. Our NMR spectroscopy experiment conducted under equilibrium conditions suggests that the N-terminus unfolds first and folds last, while the (un)folding of remaining β-strands is highly cooperative. Due to the sensitivity of the solution pH and hydrogen exchange rate with respect to temperature variations, our analysis based on the NMR data for the folding sequence of the remaining four β-strands has a limited resolution.

As evident from our study that the unfolding-folding process of the RBD is highly reversible, it is appealing to investigate the pH dependence of the folding of the RBD. Furthermore, our experiments show that a number of residues located in the β_2 and β_4 strands showed non-beta propensities at the tested temperatures (298–308 K) and a more random coil-like propensity for higher temperatures. Whether there is a correlation between the secondary structure propensity of these residues and their thermal stability implied by the simulations remains to be seen. We are working on these problems, and related studies will be reported in a future article.

REFERENCES

1. Rota, P. A., M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, et al. 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*. 300:1394–1399.

2. Marra, M. A., S. J. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, et al. 2003. The genome sequence of the SARS-associated coronavirus. *Science*. 300:1399–1404.

3. Luo, H., F. Ye, T. Sun, L. Yue, S. Peng, et al. 2004. In vitro biochemical and thermodynamical characterization of nucleocapsid protein of SARS. *Biophys. Chem.* 112:15–25.

4. Wang, Y., X. Wu, Y. Wang, B. Li, H. Zhou, et al. 2004. Low stability of nucleocapsid protein in SARS virus. *Biochemistry*. 43:11103–11108.

5. Duan, S. M., X. S. Zhao, R. F. Wen, J. J. Huang, G. H. Pi, et al. 2003. Stability of SARS coronavirus in human specimen and environment and its sensitivity to heating and UV irradiation. *Biomed. Environ. Sci.* 16:105–111.

6. Chang, C.-k., S.-C. Sue, T.-h. Yu, C.-M. Hsieh, C.-K. Tsai, et al. 2006. Modular organization of SRAS coronavirus nucleocapsid protein. *J. Biomed. Sci.* 13:59–72.

7. Hsieh, P. K., S. C. Chang, C. C. Huang, T. T. Lee, C. W. Hsiao, et al. 2005. Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. *J. Virol.* 79:13848–13855.

8. Chen, C.-Y., C.-k. Chang, Y.-W. Chang, S.-C. Sue, H.-I. Bai, et al. 2007. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* 368:1075–1086.

9. Reference deleted in proof.

10. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, et al. 2000. The Protein Data Bank. *Nucl. Acid Res.* 28:235–242.

11. Tsong, T.-Y., R. P. Hearn, D. P. Wrathall, and J. M. Sturtevant. 1970. A calorimetric study of thermally induced conformational transitions of ribonuclease A and certain of its derivatives. *Biochemistry.* 9:2666–2677.

12. Tsong, T.-Y., C.-K. Hu, and M.-C. Wu. 2008. Hydrophobic condensation and modular assembly model of protein folding. *Biosystems.* 93:78–89.

13. Gō, N. 1983. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* 12:183–210.

14. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural detail of transition state ensemble and en-route intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.

15. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198.

16. Yeh, S.-H., H.-Y. Wang, C.-Y. Tsai, C.-L. Kao, J.-Y. Yang, et al. 2004. Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: molecular epidemiology and genome evolution. *Proc. Natl. Acad. Sci. USA.* 101:2542–2547.

17. Chang, C.-k., S.-C. Sue, T.-H. Yu, C.-M. Hsieh, C.-K. Tsai, et al. 2005. The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive viruslike structure. *FEBS Lett.* 579:5663–5668.

18. Keller, R. L. J. 2004. The Computer Aided Resonance Assignment Tutorial. CANTINA Verlag, Goldau, Switzerland.

19. Schwarzinger, S., G. J. Kroon, T. R. Foss, P. E. Wright, and H. J. Dyson. 2000. Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR.* 18:43–48.

20. Schwarzinger, S., G. J. Kroon, T. R. Foss, J. Chung, P. E. Wright, et al. 2001. Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* 123:2970–2978.

21. Takada, S. 1999. Gō-ing for the prediction of protein folding mechanisms. *Proc. Natl. Acad. Sci. USA.* 96:11698–11700.

22. Allen, M. P., and D. J. Tildesley. 1987. Computer Simulations of Liquids. Oxford Science Publications, Oxford, UK.

23. Kouza, M., C.-F. Chang, S. Hayryan, T.-h. Yu, M. S. Li, et al. 2005. Folding of the protein domain hbSBD. *Biophys. J.* 89:3353–3361.

24. Li, M.-S., C.-K. Hu, D. K. Klimov, and D. Thirumalai. 2006. Multiple stepwise refolding of immunoglobulin domain I27 upon force quench depends on initial conditions. *Proc. Natl. Acad. Sci. USA.* 103:93–98.

25. Kouza, M., C.-K. Hu, and M. S. Li. 2008. New force replica exchange method and protein folding pathways probed by force-clamp technique. *J. Chem. Phys.* 128:045103.

26. Swope, W. C., H. C. Andersen, P. H. Berens, and K. R. Wilson. 1982. Computer simulation method for the calculation of equilibrium constants for the formation of physical clusters and molecules: application to small water clusters. *J. Chem. Phys.* 76:637–649.

27. Veitshans, T., D. K. Klimov, and D. Thirumalai. 1997. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Des.* 2:1–22.

28. Li, M. S., M. Kouza, and C.-K. Hu. 2007. Refolding upon force quench and pathways of mechanical and thermal unfolding of ubiquitin. *Biophys. J.* 92:547–561.

29. Li, M. S., D. K. Klimov, and D. Thirumalai. 2004. Thermal denaturation and folding rates of single domain proteins: size matters. *Polymer (Guildf.).* 45:573–579.

30. Li, M. S., D. K. Klimov, and D. Thirumalai. 2004. Finite size effects on thermal denaturation of globular proteins. *Phys. Rev. Lett.* 93:268107.

31. Li, M. S., D. K. Klimov, and D. Thirumalai. 2005. Finite size effects on calorimetric cooperativity of two-state proteins. *Physica A.* 350:38–44.

32. Kouza, M., M. S. Li, C.-K. Hu, E. P. O'Brien, Jr., and D. Thirumalai. 2006. Effect of finite size on cooperativity and folding rates of proteins. *J. Phys. Chem. A.* 110:671–676.

33. Koga, K., and S. Takaga. 2001. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *J. Mol. Biol.* 313:171–180.

34. Eisenmenger, F., U. H. E. Hansmann, S. Hayryan, and C.-K. Hu. 2001. SMMP: a modern package for simulation of proteins. *Comput. Phys. Commun.* 138:192–212.

35. Eisenmenger, F., U. H. E. Hansmann, S. Hayryan, and C.-K. Hu. 2006. An enhanced version of SMMP—open-source software package for simulation of proteins. *Comp. Phys. Commun.* 174:422–429.

36. Lin, C.-Y., C.-K. Hu, and U. H. E. Hansmann. 2003. Parallel tempering simulations of HP-36. *Proteins: Struct. Func. Gen.* 52:436–445.

37. Nguyen, P. H., G. Stock, E. Mittag, C.-K. Hu, and M. S. Li. 2005. Free energy landscape and folding mechanism of a β-hairpin in explicit water: a replica exchange molecular dynamics study. *Proteins: Struct. Funct. Bioinf.* 61:795–808.

38. Ghulghazaryan, R. G., S. Hayryan, and C.-K. Hu. 2007. Efficient combination of Wang-Landau and transition matrix Monte Carlo methods for protein simulations. *J. Comp. Chem.* 28:715–726.