Folding and Stability of Sweet Protein Single-chain Monellin

AN INSIGHT TO PROTEIN ENGINEERING*

Yoon-Hui Sung‡, Hee-Deok Hong‡, Chaejoon Cheong§, Jin Hwan Kim¶‖, Joong Myung Cho¶‖, Yong-Rok Kim**, and Weontae Lee‡ ‡‡

*From the ‡Department of Biochemistry and Protein Network Research Center, the **Department of Chemistry, College of Science, Yonsei University, Seoul 120-740, §Magnetic Resonance Group, Korea Basic Science Institute, Taejon 305-333, and ¶Biotech Research Institute, LG Chemicals, Research Park, P. O. Box 61, Yu-Sung, Science Town, Taejon 305-380, Korea*

Engineered single-chain monellin (SCM) proteins were constructed by recombinant technology without disrupting the topology and sweet activity of native protein. Data from 8-anilinonaphthalene-1-sulfonic acid fluorescence, size-exclusion chromatography, and heteronuclear NMR strongly suggest the presence of a folding intermediate at 1.5 M GdnHCl for SCM protein. The structural feature of the folding intermediate from NMR data reveals that the secondary structures became mostly unstable, and protein experiences a dynamic equilibrium between native and unfolded state. All backbone amide protons exchange within 10 min, which imply that no stable hydrogen bonds exist in the secondary structural regions in the folding intermediate. From equilibrium unfolding and mutagenesis studies, the unfolding transition midpoints of mutant proteins gradually shifted toward lower denaturant concentration, indicating stability reductions of mutant proteins. Our results suggest that stability and folding pathways of SCM proteins could be regulated by a combined study of spectroscopy and mutagenesis, and these studies will provide useful information for understanding the folding kinetics of novel engineered proteins.

Understanding the fundamental folding mechanism for three-dimensional protein structures is considered as one of the most challenging areas in structural biology research. In recent years, studies of the refolding of chemically denatured proteins *in vitro* have provided insight into the fundamental nature of the folding process (1–3). Although in some cases the protein-folding mechanism has been found to be highly cooperative, partially structured species have also played an important role during early stages of refolding prior to the formation of the native state. A number of reports and reviews have already suggested that the folding intermediate could have extensive native-like structural elements even in the absence of tertiary structural data, which are known as molten globular states (4–6).

The sweet protein, monellin, was originally isolated from the berries of the west African plant *Dioscoreophyllum cumminsii* (7, 8). Native monellin consists of two separate polypeptide chains, a 45-residue A-chain and a 50-residue B-chain. Two-chain monellin is ~70,000 times sweeter than sucrose and 300 times sweeter than the dipeptide sweetener aspartame (9, 10). The crystal structures of two-chain monellin showed that the two chains are packed closely by both intermolecular hydrogen bonds and hydrophobic interactions (11). Recent biochemical studies of two-chain monellin indicated that the A-chain can convert from β-sheet to α-helix in 50% ethanol and 50% trifluoroethanol solution. They also suggested that the molten globule state of monellin could exist in a solution of ethanol and trifluoroethanol mixture (12). In addition, a conformational study for both native and mutant monellin by a non-peptide analogue demonstrated that the three-dimensional structures of monellin and two thiol proteinase inhibitors, cystatin and stefin B, are very similar (13), suggesting that monellin might have some other biological role besides its sweet activity. An engineered single-chain monellin (SCM),[1] constructed by fusing the two chains without disrupting its topology and sweet activity, has been proven more stable and sweet than the two-chain monellin under both high temperature and acidic conditions (14).

We reported (15) that engineered SCM exists as a monomer conformation in solution state, and the NMR data revealed that a long helix was folded into the concave side of a six-stranded antiparallel β-sheet, showing that the residues responsible for sweet activity are mostly solvent-exposed. Interestingly, most of the residues involved in the sweet taste were found on the same surface of the molecule. In addition, the structure suggested that the relative orientation of the long α-helix is important to stabilize the global topology of SCM (15, 16). However, it is still not clear whether the folding stability of SCM is important for its activity or not, even though it is suggested that the relative orientation of receptor-binding sites is responsible for sweet receptor binding, resulting in sweet activity.

For folding studies of the engineered protein, SCM could be used as an excellent model for the following two reasons. First, it consists of two regular structural elements (α-helix and β-sheet) assembled perpendicularly. Second, as an engineered fusion protein of two molecules, SCM can provide the engineering effects of two separate units during the folding process. We

‖ Current address: CrystalGenomics, Inc., Daeduck Biocommunity, Junmin-Dong, Yusung-Gu, Taejon 305-390, Korea.

‡‡ To whom correspondence should be addressed: Dept. of Biochemistry, College of Science, Yonsei University, Seodaemoon-Gu, Shinchon-Dong, Seoul 120-749, Korea. Tel.: 82-2-2123-2706; Fax: 82-2-362-9897; E-mail: wlee@spin.yonsei.ac.kr.

[1] The abbreviations used are: SCM, single-chain monellin; GdnHCl, guanidine hydrochloride; ANS, 8-anilinonaphthalene-1-sulfonic acid; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlated spectroscopy; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect.

report that the equilibrium unfolding pathways and stability of the engineered SCM proteins were studied by fluorescence, circular dichroism, heteronuclear NMR spectroscopy, and gel filtration chromatography. The detailed structural characterization of the folding intermediate in the unfolding process will be presented by heteronuclear NMR spectroscopy. In addition, we will discuss the stabilizing effect of each residue important for the engineered protein based on mutagenesis and spectroscopic techniques.

<div align="center">EXPERIMENTAL PROCEDURES</div>

Expression and Purification of SCM—The recombinant SCM proteins were expressed and purified as described (16). Transformed yeast cells in *Saccharomyces cerevisiae* AB110 were propagated in yeast nitrogen base containing 5% glucose and 0.5% ammonium sulfate at 30 °C for 2 h and grown in M9 media containing 2% glucose and 0.1% ammonium sulfate at 30 °C for 48 h. ^{15}N-Labeled ammonium sulfate was used as the sole source of the nitrogen for uniformly ^{15}N-labeled monellin proteins. The cells were harvested by centrifugation at 3500 rpm for 25 min. Cells were stored at −80 °C until used for purification procedures. Cell pastes were disrupted by a bead beater in 25 mM sodium phosphate, 5 mM EDTA, 150 mM NaCl, and 1 mM phenylmethylsulfonyl fluoride, pH 7.0. Protein concentration was determined using the Bradford method. GdnHCl induced unfolding transitions at equilibrium were determined by 24-h incubation of samples at various GdnHCl concentrations. The protein samples for NMR measurements were prepared by dissolving in 90% H$_2$O, 10% D$_2$O solution at a pH value of 7.0 with 50 mM sodium phosphate buffer. The final protein concentration was adjusted to 4–6 mg/ml.

Fluorescence Spectroscopy—Fluorescence spectra of wild type and mutant SCM proteins were measured in 50 mM aqueous potassium phosphate buffer, pH 7.0, at 25 °C on a F-4500 fluorescence spectrophotometer. Fluorescence emission spectra were recorded from 270 to 450 nm at each GdnHCl concentration using two different excitation wavelengths, 280 and 295 nm. The protein concentration in the cuvette was 30 μM, and the path length of the cuvette used was 1 cm. GdnHCl unfolding experiments were carried out after incubation of the protein in solutions containing different concentrations of the denaturant for 24 h at 25 °C. Controls of the reversibility of the folding reaction were performed at each condition by diluting out the denaturant. The average emission wavelength, $<\lambda>$, was calculated using Equation 1 (17),

$$<\lambda> = \frac{\sum_{i=\lambda_i}^{\lambda_N}(F_i\lambda_i)}{\sum_{i=\lambda_i}^{\lambda_N}F_i} \qquad (\text{Eq. 1})$$

where F is the fluorescence intensity, and λ is the wavelength. This parameter reflects changes in the shape of the spectrum as well as the position. The $<\lambda>$ of Trp invariably changes to longer wavelengths upon unfolding.

To monitor the accessibility of hydrophobic regions to the solvent, wild type SCM (20 μM) was exposed to 3 mM ANS in various concentrations of GdnHCl at 25 °C. The fluorescence spectra of the samples were measured at an excitation wavelength of 375 nm. The fluorescence intensities at 483 nm are plotted as a function of GdnHCl concentration.

CD Spectroscopy—CD spectra of wild type and mutants were measured in 50 mM aqueous potassium phosphate buffer, at pH 7.0 and 25 °C on a Jasco 720 spectropolarimeter. Far-UV CD spectra were monitored from 190–250 nm using a protein concentration of 30 μM with a path length of 0.1 mm, 20 millidegree sensitivity, a response time of 1 s, and a scan speed of 50 nm/min. Spectra were recorded as an average of 6 scans. Near-UV CD spectra were monitored from 240 to 310 nm using a protein concentration of about 0.13 mM with a bandwidth of 2.0 nm, a response time of 1 s, and a scan speed of 10 nm/min.

Size-exclusion Chromatography—20 μl of each sample in 0–5 M GdnHCl, at a final SCM concentration of 10 μM, was loaded onto a Superdex 75 PC 3.2/30 column equilibrated with the same buffer as the sample. The elution was carried out isocratically at a flow rate of 0.015–0.021 ml/min. The flow rate was monitored by absorbance at 280 nm. The spectroscopic properties for collected peaks were recorded, and protein samples were then loaded again under the same conditions. In all cases the time elapsed between separation and chromatography of the peaks was longer than 2 h.

NMR Spectroscopy—All NMR spectra were acquired on Bruker DMX600 and DRX500 spectrometers in quadrature detection mode equipped with a triple resonance probe with an actively shielded pulsed field gradient coil. All experiments were performed at 25 °C. Pulsed field gradient techniques with a WATERGATE pulse sequence (18) were used for all H$_2$O experiments, resulting in good suppression of the solvent signal. ^{15}N edited three-dimensional NOESY-HSQC (19) with a mixing time of 100 ms, and ^{15}N-edited three-dimensional TOCSY-HSQC (20) with a mixing time of 69.668-ms spectra for uniformly ^{15}N-labeled SCM of both native and folding intermediate (1.5 M GdnHCl) states were recorded.

^{1}H-^{15}N HSQC (21, 22) spectra for different concentrations of GdnHCl (0–3.4 M) were also recorded with 2048 complex data points in the t_2 and 64 increments in the t_1 dimensions. The protein samples were used after a 24-h incubation in the presence of GdnHCl. Slowly exchanging amide protons with solvent water were readily identified from a series of ^{1}H-^{15}N HSQC spectra in a freshly prepared D$_2$O solution after lyophilization of an H$_2$O sample for 24 h. The nitrogen-offset frequency was set to 119 ppm, and spectral widths were used as 7002.8 Hz in t_2 and 2027.68 Hz in t_1. 64 scans were recorded for each t_1 increment. Longitudinal (R_1) and transversal (R_2) relaxation data for the backbone ^{15}N nuclei for wild type SCM were recorded as 2048 × 64 data sets with 64 scans per point, using a relaxation delay of 1 s. Seven different values for the relaxation time were recorded with seven different durations of the T delay, $T_1 = 5, 65, 145, 246, 366, 527,$ and 757 ms; and $T_2 = 8.3, 25.1, 41.8, 58.6, 75.3, 108.8,$ and 142.3 ms (23). To permit estimation of noise levels, duplicate spectra were recorded for $T = 246$ ms (T_1 spectra) and $T = 56.8$ ms (T_2 spectra). To remove the cross-correlation effects between ^{15}N-^{1}H dipolar and ^{15}N chemical shift anisotropy relaxation mechanisms, ^{1}H 180° pulses were inserted during the T relaxation times (24, 25). ^{15}N-{^{1}H} steady-state heteronuclear NOE (26, 27) data were obtained using a relaxation delay of 5 s, yielding data sets of 2048 × 128 data sets after accumulation of 128 scans per point.

NMR Data Processing and Analysis—The NMR data were processed and analyzed with the nmrPipe/NMRView software packages (28, 29) and the Sparky program, respectively. Heteronuclear dimensions were extended by linear prediction and zero-filled to give 2048 × 512 matrices and processed using Gaussian multiplication and a shifted (π/3) sine bell function prior to Fourier transformation. The intensities of the peaks in the two-dimensional spectra were described by peak heights using Sparky. The heteronuclear NOE value for a given residue was calculated as the intensity ratio (I/I_0) of the ^{15}N-^{1}H correlation peak in the presence (I) and absence (I_0) of proton saturation for 3 s. The standard deviations of these values measured background noise levels. Relaxation rates were determined by nonlinear fits of the time dependence of the peak intensities, and Monte Carlo simulations were performed to estimate the uncertainty of the relaxation parameters.

The analysis of the overall tumbling of wild type SCM from the T_1/T_2 ratios of N-H groups without significant internal motion contributions was carried out with the program Quadric Diffusion 1.11 from A. G. Palmer III, Columbia University (30). The inertial tensor parameters ($\{D_{\text{iso}}\}$, $\{D, D, \Phi, \theta\}$ and $\{D_{zz}, D_{yy}, D_{xx}, \Phi, \Theta, \Psi\}$) for isotropic, axially symmetric, and fully anisotropic models were fit to the experimental data.

Hydrogen-Deuterium Exchange Experiments—Amide proton exchanges of native SCM were measured at pH 7.0 and 298 K by dissolving lyophilized protein in 99% D$_2$O and following hydrogen-deuterium exchange with a series of ^{15}N-^{1}H HSQC spectra. ^{1}H-^{15}N HSQC spectra were acquired with 2048 complex data points in t_2 and 64 t_1 increments. The pulse sequence was adapted to allow the collection of 64 scans per increment, and this gave a total experiment time of 2 h and 2 min. A total of 12 spectra was acquired sequentially with the final time point collected at 24 h and 24 min. A second amide proton exchange was measured after 24 h from the 12th exchange experiment. To identify hydrogen-deuterium exchange rate of folding intermediate, a series of ^{1}H-^{15}N HSQC spectra was acquired on a freshly prepared D$_2$O solution after lyophilization of an H$_2$O sample under 1.5 M GdnHCl environments. In order to check the reliability of the exchange data from the folding intermediate state, the exchange experiments were repeated three times using independent sample preparation procedures.

Hydrophobicity Plots—Hydrophobicity scores for SCM sequences were calculated with the ProScale module of the ExPASy molecular biology server (expasy.proteome.org.au/cgi-bin/protscale.pl). The amino acid scale was used, averaging over a window of seven residues (31–33).
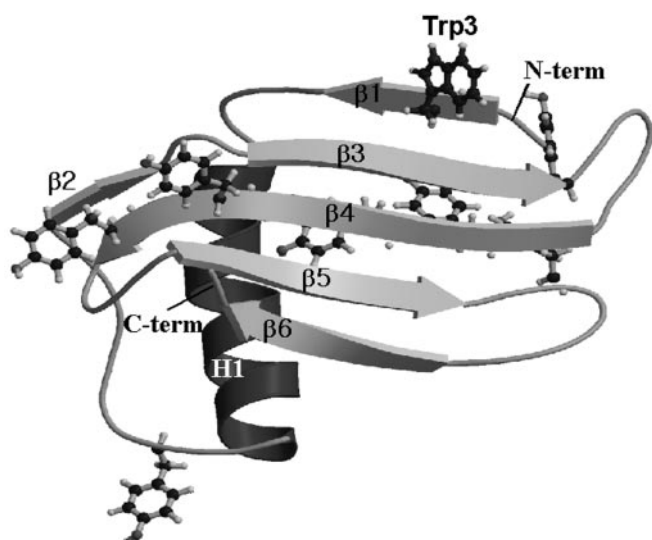
FIG. 1. **Ribbon diagram of the restrained energy minimized average structure of native SCM displaying ordered secondary structure elements and relative orientation of secondary structures.** Tryptophan and tyrosine residues are presented in *ball* and *stick model.*
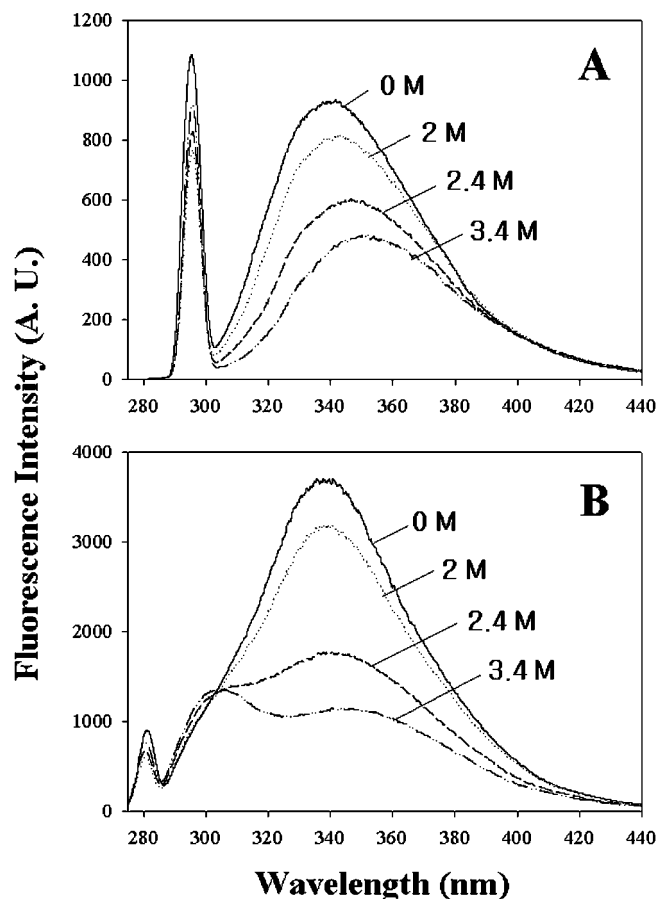

FIG. 2. *A,* fluorescence spectra at 0–3.4 M GdnHCl concentration for SCM. The excitation wavelength was 295 nm, and Trp[3] was the unique fluorescent probe. *B,* fluorescence spectra for 280 nm wavelength which was induced by the resonance energy transfer between tyrosine (donor) and tryptophan (acceptor).
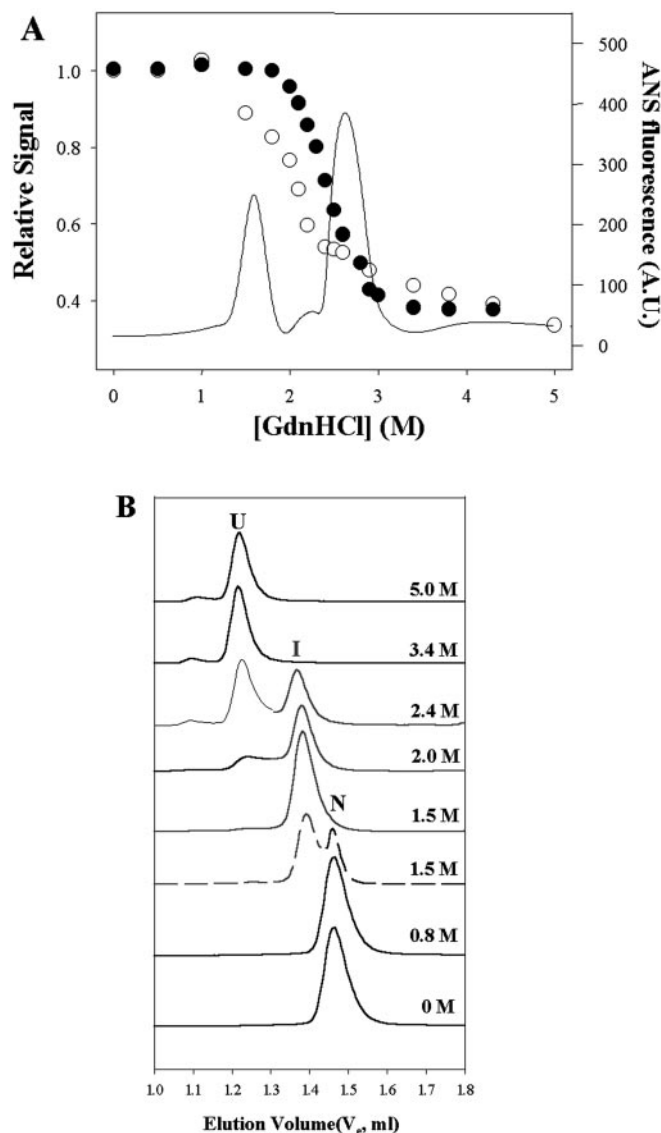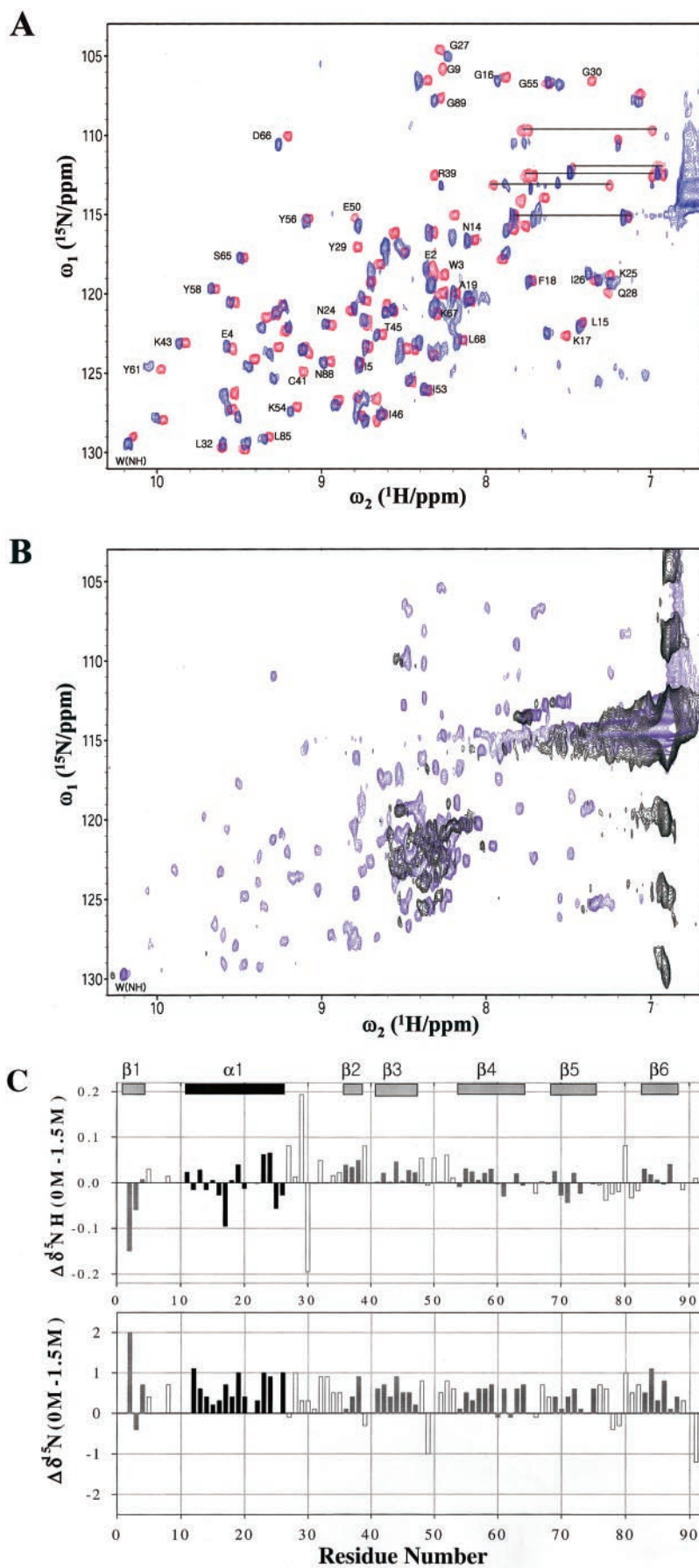

FIG. 3. *A,* ANS fluorescence intensity at 483 nm (*solid line*) and relative signals of CD at 222 nm (*open circles*) and Trp fluorescence (*filled circles*) are plotted for various concentrations of GdnHCl. The spectra were taken with a 150:1 molar ratio of ANS:wild type SCM, 20 μM protein in 50 mM potassium phosphate, and they are smoothed. *B,* size-exclusion HPLC profiles of wild type SCM as a function of GdnHCl concentration. Wild type SCM was incubated in various concentrations of GdnHCl for 24 h at 25 °C, and 20 μl of the sample containing 10 μM SCM was applied to a Superdex 75 PC 3.2/30 size-exclusion column that had been equilibrated with buffer containing an identical concentration of GdnHCl. Native (*N*), intermediate (*I*), and unfolding (*U*) states are marked, respectively.

RESULTS AND DISCUSSION

Dynamic Unfolding Process of Engineered SCM—Measurement of intrinsic fluorescence was made in the environment of aromatic residues, particularly tryptophan residues. Since SCM has only one Trp residue in the first β-strand region (Fig.

1), the characteristic fluorescence spectra are easily monitored from a single Trp with a maximum intensity at 339 nm. Fluorescence emission spectra at various concentrations of GdnHCl indicated that the unfolding process begins at 2.0 M GdnHCl, based on changes in fluorescence intensity with a subtle red-shift in the maximal wavelength of tryptophan (Fig. 2*A*). The emission spectrum was dominated by the Trp fluorescence at an excitation wavelength of 280 nm for different GdnHCl concentrations. The protein unfolding process induced by GdnHCl was also observed by two distinct emission peaks based on the contribution of tyrosine-tryptophan energy transfer mechanisms (Fig. 2*B*). There are negligible differences in the transition position between the results monitored by Trp fluorescence compared with those by fluorescence energy transfer, indicating that the two experimental probes are detecting the identi-
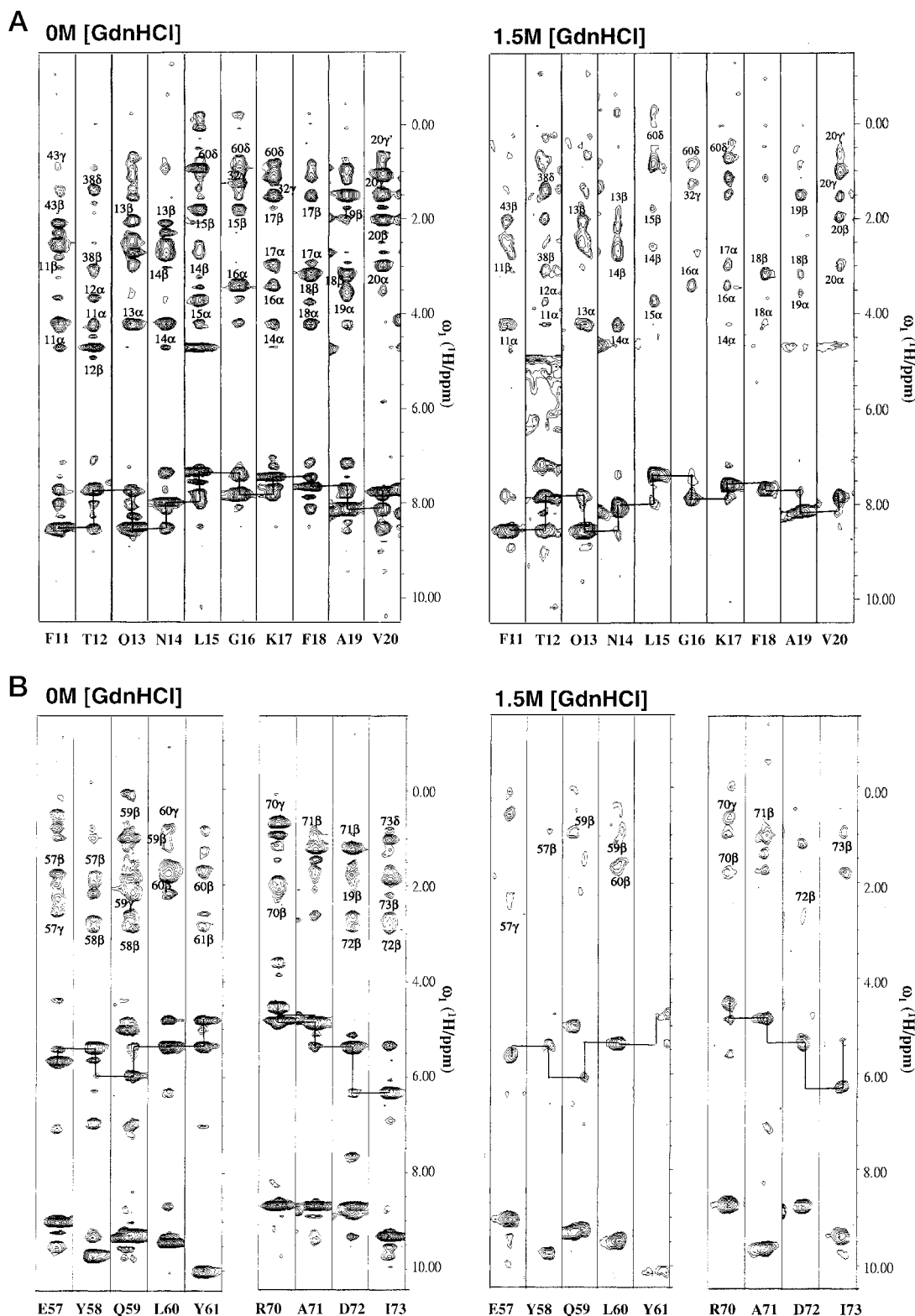
Folding and Stability of Sweet Protein SCM

FIG. 4. **The two-dimensional ^{1}H-^{15}N HSQC spectra of uniformly ^{15}N-labeled wild type SCM in 0 M (in *red*), 1.5 M (in *blue*) (*A*), and 2.4 M (in *purple*), and 3.4 M (in *black*) (*B*) GdnHCl concentration.** Indole NH cross-peak from Trp[3] is marked as *W(NH)*, and cross-peaks connected by *dotted lines* correspond to side chains of Gln and Asn residues. *C,* chemical shift differences of backbone amide proton and nitrogen atoms between 0 and 1.5 M [GdnHCl] from two-dimensional ^{1}H-^{15}N HSQC spectra.

FIG. 5. **NOE strip plots from ^{15}N-^{1}H three-dimensional NOESY-HSQC spectrum of both native (0 M GdnHCl) and folding intermediate (1.5 M GdnHCl) state.** NOE cross-peaks of α-helical (*A*) and β-sheet (*B*) regions are displayed with assignments.

cal conformational properties of SCM protein.

Characterization of Folding Intermediate—The equilibrium unfolding transition of intrinsic fluorescence and circular dichroism was monitored to identify the folding intermediate of SCM protein as a function of GdnHCl, since SCM protein does not demonstrate a typical two-state transition model for the unfolding process (Fig. 3). The structural transitions occur at

two conditions of both 1.5 and 2.8 M of GdnHCl, suggesting the presence of folding intermediate state during this unfolding process. The reversibility of the unfolding reaction was confirmed independently by obtaining refolding patterns through dilution of GdnHCl. Both refolding and unfolding curves were demonstrated to be nearly identical. In addition, we performed fluorescence experiments using a hydrophobic probe, 8-anilino-
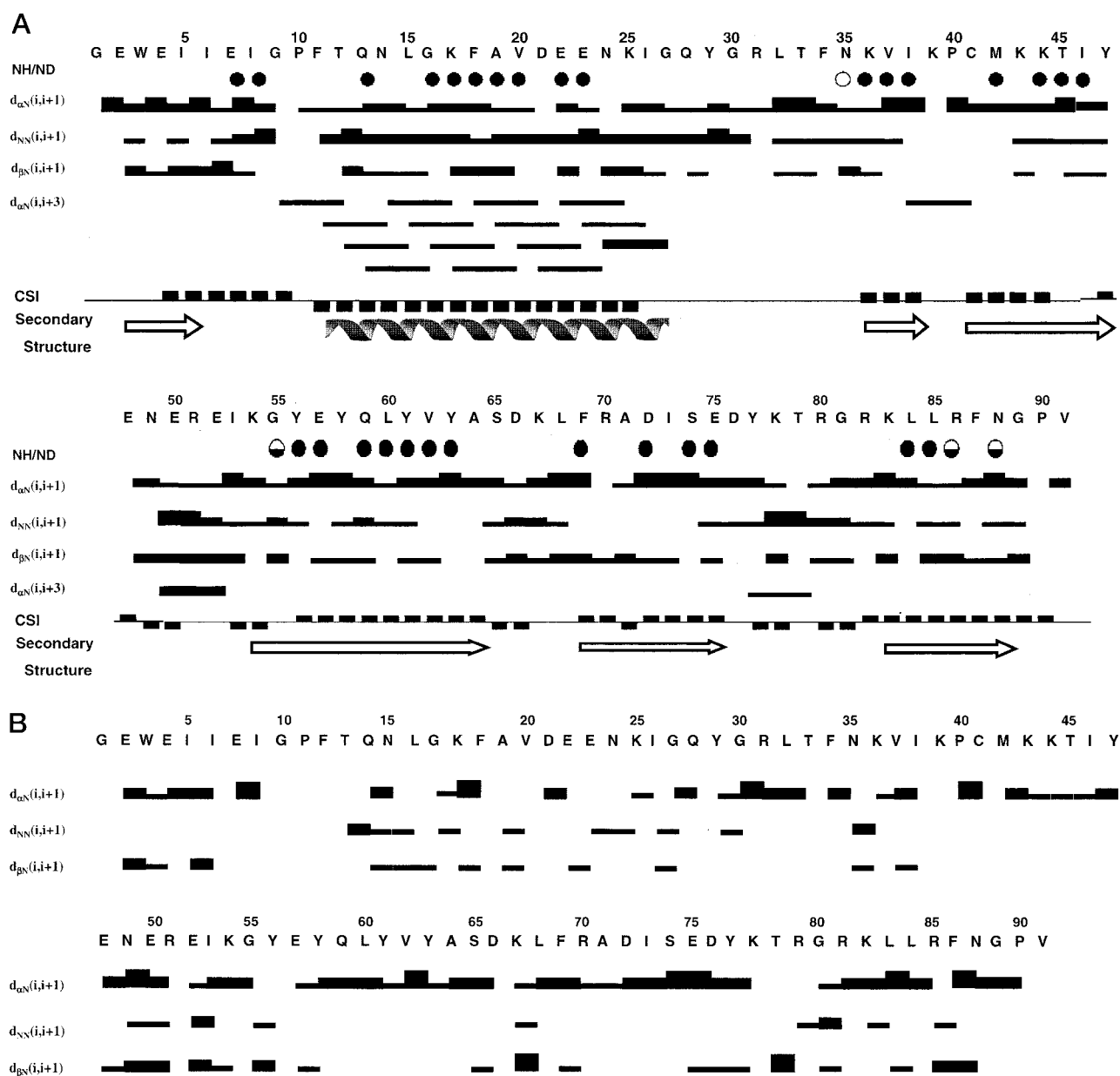
FIG. 6. **Summary of NMR data for both native (*A*) and folding (*B*) intermediate state showing the sequential and short range NOE connectivities.** Chemical shift indices (*CSI*) and secondary structural information are also displayed.

1-naphthalenesulfonic acid (ANS), which is a transient binder to hydrophobic patches of protein. It is known that the ANS molecule binds to hydrophobic regions of partially folded intermediates, reporting a substantial change in fluorescence emission intensity from a blue shift in λ_{max} (34). As shown in Fig. 3*A*, a significant increase in ANS fluorescence intensity was observed at two points (from 1.5 to 1.8 and 2.6 to 2.8 M concentration of GdnHCl), whereas the intensity decreased at other concentrations of denaturant. The increase at 1.5 M GdnHCl condition strongly suggested the presence of a hydrophobic intermediate for ANS binding.

Data from size-exclusion chromatography in the presence of GdnHCl also provided excellent agreement with ANS binding, showing tertiary structural changes at 1.5 M GdnHCl condition. Since the elution volume of the native state is normally larger than that of the denatured state and if the exchange between the two states occurs at the time scale of chromatography, both states can be separated as individual elution peaks. If the exchange rate between two states is fast enough, a single peak

which is a weighted average value of the population ratio of the folded and unfolded forms would be observed with an elution volume. After SCM protein was incubated in eight different concentrations of GdnHCl, those samples were loaded to size-exclusion chromatography. As expected, wild type SCM was eluted as a single peak with a molecular mass of ~10 kDa in the absence of GdnHCl. The elution profile with increased denaturant concentration gradually shifted to that of the unfolded state. The elution profile at 1.5 M condition was clearly different from that of 5.0 M, which is strong evidence of a folding intermediate during the unfolding process (Fig. 3*B*). The far-UV CD spectra at 1.5 M GdnHCl condition also support the existence of the folding intermediate (Fig. 3*A*). However, the fluorescence intensity at 1.5 M GdnHCl did not change much compared with that of CD spectrum. Since Trp[3] is mostly exposed to solvent even though the protein is under its native conditions, the fluorescence intensity does not change dramatically at 1.5 M GdnHCl.
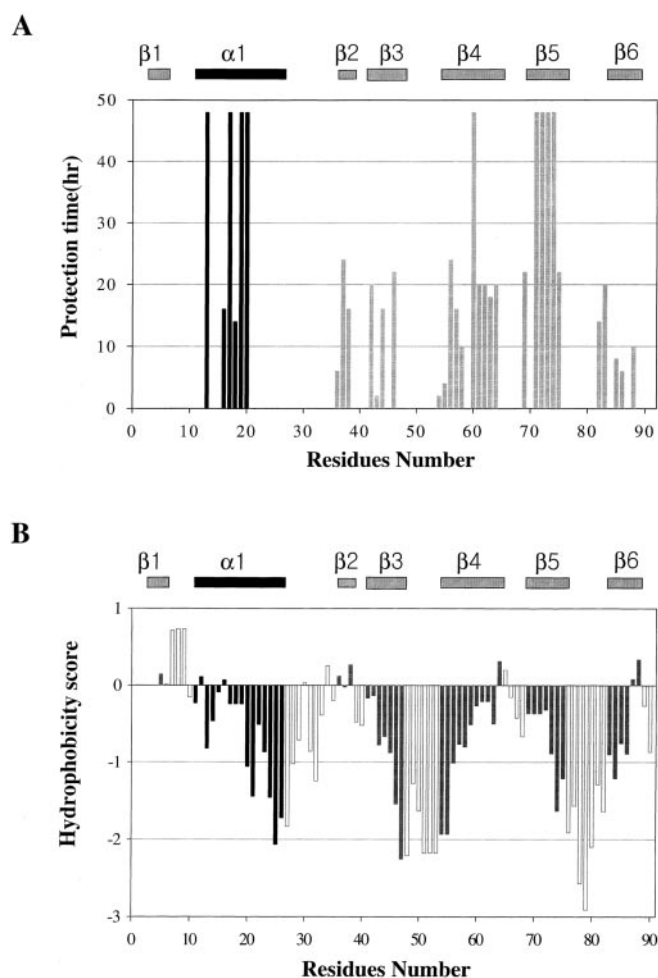
Residue-specific Characteristics in the Unfolding Proc-

A

B

FIG. 7. *A,* a plot of the deuterium-hydrogen exchange times as a function of residue number. Whereas amide hydrogens in the loop and β1 regions exchanged fast rates ($\tau_{ex} < 10$ min), those of the α1, β4, and β5 regions showed slow deuterium-exchange rates. *B,* Kyte-Doolittle hydrophobicity scores as a function of the SCM sequences. The plots were generated as described under "Experimental Procedures."

ess—To study residue-specific details of the unfolding process, a series of two-dimensional ^{15}N-^{1}H HSQC spectra was collected at various concentrations of GdnHCl (Fig. 4, *A* and *B*). At the 1.5 M GdnHCl environment, most residues show chemical shift perturbation, suggesting that most secondary structural regions experience conformational changes. In addition, the spectrum demonstrated a distinct line broadening effect for most loops and some secondary structural regions (Fig. 4*A*). The residues involved in the β1 strand showed a relatively large perturbation of ^{1}H chemical shifts (Fig. 4*C*). Interestingly, for some residues, two sets of resonance were observed on the spectrum at 2.4 M GdnHCl, which is indicative of the existence of both intermediate and unfolded state above 2.4 M GdnHCl concentration (Fig. 4*B*). Above a 3.4 M GdnHCl concentration, the protein became fully denatured, displaying a standard pattern of denatured protein in ^{15}N-^{1}H HSQC spectra.

Structural Characterization of a Folding Intermediate—For a detailed characterization of the folding intermediate during the unfolding pathway, data from ^{15}N-edited three-dimensional NOESY-HSQC, three-dimensional TOCSY-HSQC, and ^{15}N-^{1}H HSQC experiments have been performed at 1.5 M GdnHCl conditions. Even though most resonance patterns in ^{15}N-^{1}H HSQC spectra of the intermediate state are similar to those of the native form, both NOESY and TOCSY spectra of the intermediate state exhibit very poor sensitivities compared

with those of the native state. In ^{15}N-edited three-dimensional TOCSY-HSQC spectra, only NH/C$^\alpha$H and a few NH/C$^\beta$H scalar couplings were observed (data not shown). Fig. 5 shows the secondary structural regions of NOESY spectra of both native and intermediate states. NOE intensities of the intermediate state were relatively weak, and most NOE cross-peaks important for tertiary structural information were not observed in the intermediate state. Fig. 6 summarizes sequential and medium range NOEs detected in both native and intermediate state of SCM protein. Based on some NH(*i*)/NH(*j*) NOEs in the α-helical region as well as a few long range NOEs including NH(*i*)/NH(*j*) and C$^\alpha$H(*i*)/NH(*j*) in the β-sheet region, we expect that the secondary structures became a mostly unstable state in this environment (Fig. 6*B*). The hydrogen exchange data from hydrogen-deuterium exchange experiments show that all backbone NH protons exchanged quickly (within 10 min), which suggests that no stable hydrogen bonds exist in the secondary structural regions at 1.5 M GdnHCl conditions. All these NMR data are very well supported by CD data, showing a subtle change of absorbance of SCM protein in the presence of 1.5 M GdnHCl (Fig. 3*A*). The analysis of the ^{15}N-edited three-dimensional NOESY-HSQC spectra provided a number of NOEs; however, since most of these are intra-residue and sequential, it was not enough to define the tertiary structure of the intermediate state (Fig. 6*B*).

Structures, Activity, and Stability of Engineered Mutant Proteins—To identify residual stability of native SCM, we also performed deuterium-hydrogen exchange (Fig. 7*A*), Kyte-Doolittle hydrophobicity plot (Fig. 7*B*), and ^{15}N backbone dynamics experiments (Fig. 8). The data agreed that most residues from loops have the intrinsic characteristics of flexibility as shown by ordering parameters for these regions (Fig. 8*E*). We calculated hydrophobicity scores to correlate residual stability and sequential hydrophobicity of the protein, showing that most residues are clustered in three major hydrophilic regions (Fig. 7*B*). For specific structure-activity correlations, four engineered mutant proteins have been used to probe residues responsible for activity and folding stability (Fig. 9*A*).

The mutant proteins, SCM^{DR} (D7E/R39K), SCM^{D7R} (D7R), and SCM^{D7N} (D7N), show significantly lower sweet activity than that of wild type, whereas the triple mutant SCM^{GED} (G1M/E2M/D66T) demonstrates 1.3 times higher activity than that of wild type (16). The structural differences of wild type and mutant proteins have been compared by intrinsic fluorescence and far-UV CD data to examine alterations in both secondary and tertiary structures (Fig. 9*B*). Taken together, the spectral patterns clearly show that the global structures of all mutant and native proteins are almost identical to each other. However, fluorescence emission spectra revealed that band positions for the triple mutant SCM were red-shifted, indicating that the Trp^3 residue might be more flexible than that of native SCM as shown in Fig. 9*A*. The far-UV CD spectra for all SCM proteins showed a minima near 215 nm, showing that their structures are also similar (Fig. 9*B*). However, wild type SCM exhibits a different CD profile than that of the mutants, suggesting that the wild type SCM might be slightly different in the population of secondary structural elements. This is already proven from our recent NMR structures of both wild type and double mutant proteins (15). Denaturation curves for wild type and mutant proteins in the presence of GdnHCl were monitored by Trp fluorescence and far-UV CD (Fig. 9*C*). The equilibrium unfolding transitions of intrinsic fluorescence for mutant proteins exhibited reversible monophasic transition curves, which is the same as that of the wild type. However, the transition midpoints of mutant SCM proteins indicate that those are less stable than the wild type. In addition, GdnHCl-

FIG. 8. **^{15}N NMR relaxation parameters and model-free analysis of wild type SCM.** *A, T_1; B, T_2; C, R_2/R_1; and D,* heteronuclear NOE (*XNOE*) values plotted against residue number. *E,* generalized order parameter S^2; *F,* effective correlation times, τ_e, derived from model free analysis. The α-helical region is represented as *black bars*, β-strands as *dotted bars,* and loops as *open bars*, respectively.
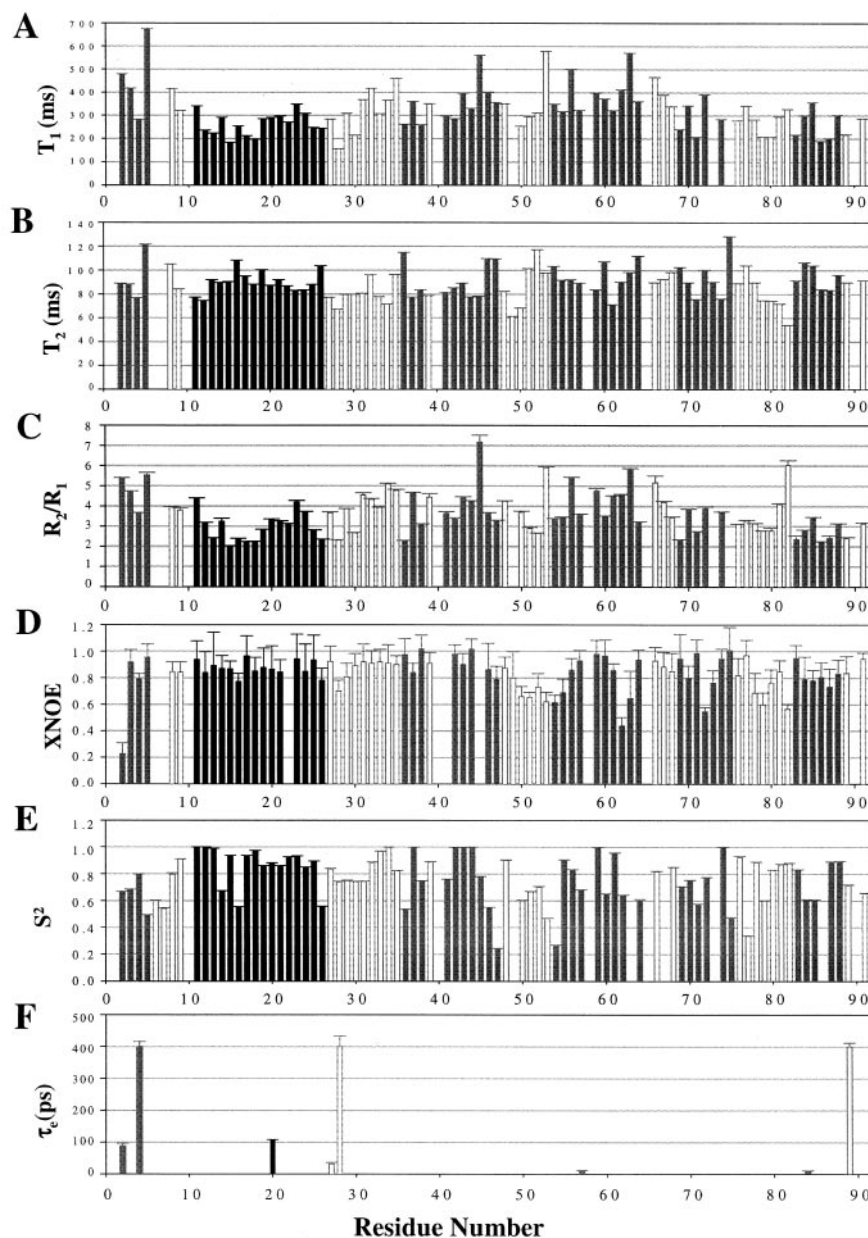
induced unfolding curves for wild type and mutant proteins clearly demonstrated that wild type SCM could have a folding intermediate in the unfolding process (Fig. 9*D*).

Folding Kinetics of SCM Proteins—We have performed the GdnHCl-induced reversible refolding reaction for wild type and mutant SCM proteins. Our experimental results of far-UV CD, size-exclusion chromatography, and ANS fluorescence presented here clearly showed that the unfolding pathway of wild type SCM may be represented by a three-state mechanism having an intermediate state at 1.5 M GdnHCl environments. Transition curves monitored by far-UV CD suggest that the folding intermediate of SCM in 1.5 M GdnHCl has different secondary structural organization in both α-helix and β-strands. Results obtained by size-exclusion chromatography and two-dimensional ^{15}N-^{1}H HSQC experiments also proposed that since equilibrium between the two states, native and folding intermediate or intermediate and denatured state, is on a slower time scale than the time of chromatography and NMR, both states are observable by these techniques (Fig. 3*B*, Fig. 4). In addition, chemical shift changes of both backbone amide

proton and nitrogen between native and intermediate states are observed in both α-helical and β-strand regions (Fig. 4*C*). The particularly broad peaks that belonged to loops or edges of secondary structural regions in two-dimensional ^{1}H-^{15}N HSQC spectrum at 1.5 M GdnHCl concentration indicate that proteins experience dynamic fluctuations of unstable regions. These results strongly suggest that the folding intermediate induced by 1.5 M GdnHCl could be considered as a dynamic intermediate state, which is an unstable native-like structure containing a population of dynamic secondary structures but lacking the side chain-side chain interactions for tertiary structure.

To provide insight into the folding and sweet activity of an engineered protein SCM, we took advantage of four mutant proteins to study protein stability, activity, and folding mechanisms. SCM proteins mutated at the potential active site were used for this study and are shown in Fig. 10. Since most residues are hydrophilic, these regions are mostly exposed to solvent (Fig. 10), suggesting that these residues may interact directly with taste receptors. The profile of transition midpoint shifts toward lower denaturant concentrations in transition
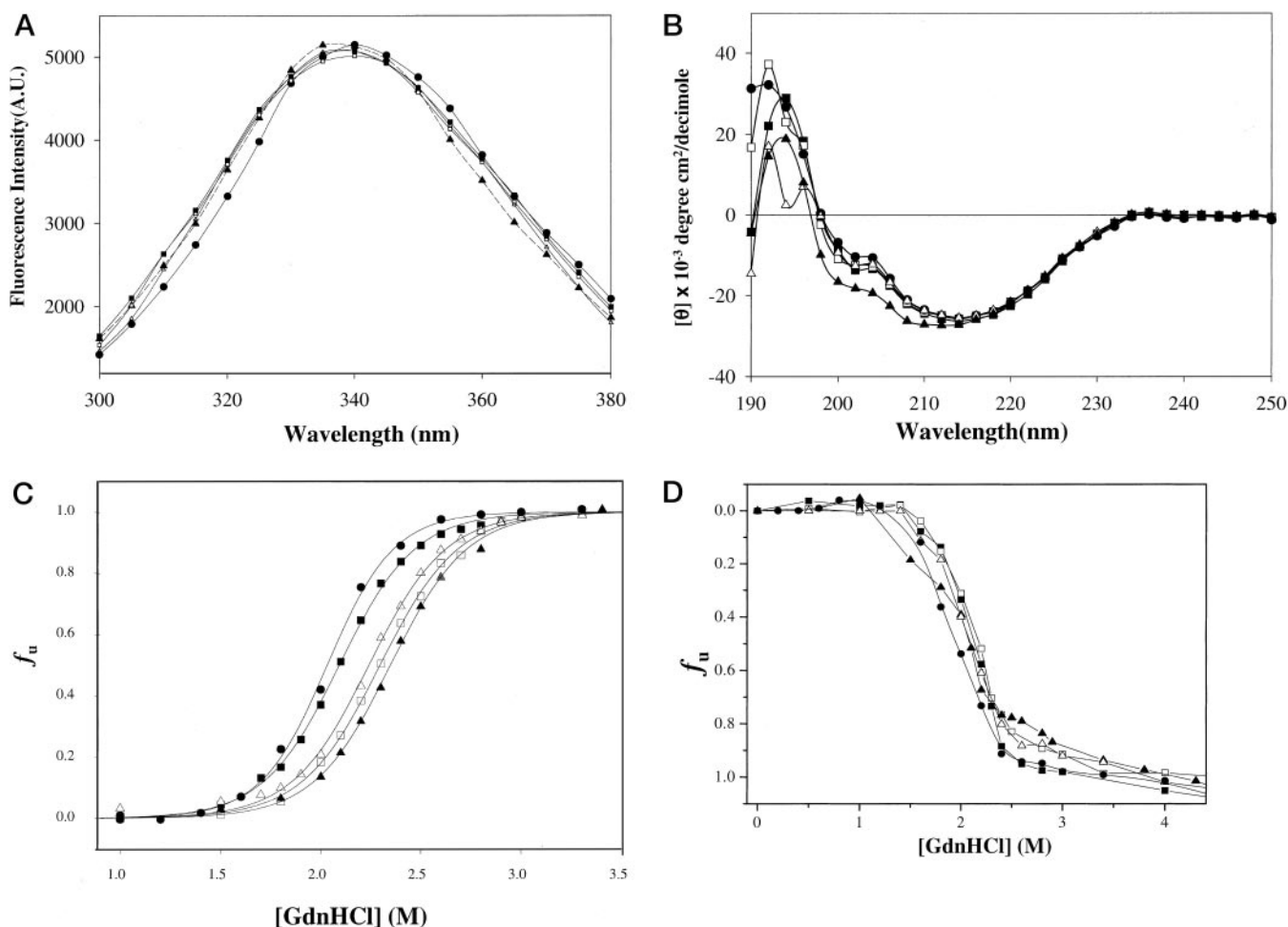
FIG. 9. Trp fluorescence spectra (*A*) and far-UV circular dichroism for wild type SCM (*closed triangles*) and its mutants SCM^DR (*closed squares*), SCM^D7N (*open squares*), SCM^DR (*open triangles*), and SCM^GED (*closed circles*) (*B*). *C,* equilibrium unfolding transition curve for GdnHCl-induced denaturation monitored by Trp fluorescence, and *D,* far-UV circular dichroism at 222 nm for wild type SCM (*closed triangles*) and its mutants SCM^D7R (*closed squares*), SCM^D7N (*open squares*), SCM^DR (*open triangles*), and SCM^GED (*closed circles*).
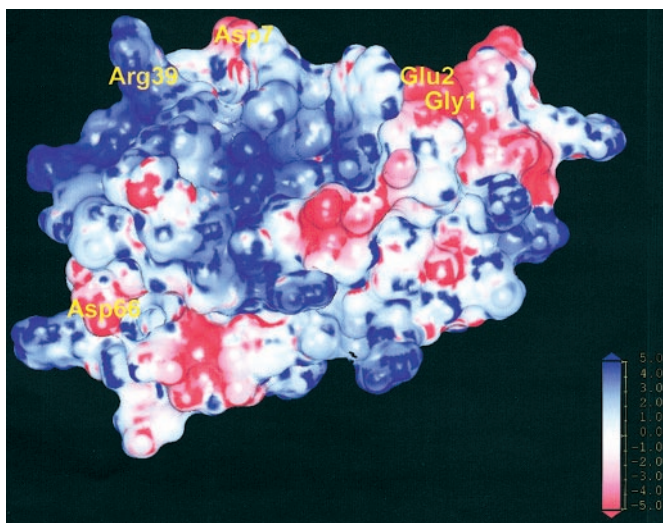

FIG. 10. **Electrostatic surface potential of SCM, displaying putative receptor binding and active site residues for mutant SCM proteins.** The negative electrostatic potential is represented in *red*, the positive in *blue*, and the neutral in *white*. The potential surface was calculated using Delphi program (Biosym/Molecular Simulation Inc.).

curves as monitored by fluorescence measurements indicates reduction of protein stability caused by mutagenesis (Fig. 9*C*). Unlike the wild type, the transition curve of mutants by CD spectroscopy does not show an apparent three-state transition (Fig. 9*D*). The equilibrium unfolding transitions monitored by far-UV CD showed that the equilibrium transition pattern of wild type SCM exhibits a biphasic transition, whereas mutant proteins do not exhibit apparent biphasic transitions. We can conclude that the stability and folding pathway of engineered SCM proteins could be regulated by a combined study of spectroscopy and mutagenesis, and those studies will provide useful information for understanding folding kinetics in engineered proteins.

REFERENCES

1. Shakhnovidh, E., and Fersht, A. R. (1998) *Curr. Opin. Struct. Biol.* **8,** 65–67
2. Dobson, C. M., Sali, A., and Karplus, M. (1998) *Angew. Chem. Int. Edit.* **37,** 868–893
3. Samuel, D., Kumar, T. K., Srimathi, T., Hsieh, H., and Yu, C. (2000) *J. Biol. Chem.* **275,** 34968–34975
4. Ptitsyn, O. B. (1981) *FEBS Lett.* **131,** 197–202
5. Ptitsyn, O. B. (1995) *Curr. Opin. Struct. Biol.* **5,** 74–78
6. Dobson, C. M. (1995) *Nat. Struct. Biol.* **2,** 523–527
7. van der Wel, H. (1976) *Biochem. Sensory Funct.* **197,** 235–242
8. Morris, J. A., and Cagan, R. H. (1972) *Biochim. Biophys. Acta* **261,** 114–122
9. Brouwer, J. N., Hellekant, G., Kasahara, Y., van der Wel, H., and Zotterman, Y. (1973) *Acta Physiol. Scand.* **89,** 550–557
10. Hough, C. A. M., and Edwardson, J. A. (1978) *Nature* **271,** 381–383
11. Ogata, C., Hatada, M., Tomlinson, G., Shin, W. C., and Kim, S. H. (1987) *Nature* **328,** 739–742
12. Fan, P., Bracken, C., and Baum, J. (1993) *Biochemistry* **32,** 1573–1582
13. Murzin, A. Z. (1993) *J. Mol. Biol.* **230,** 689–694
14. Kim, S. H., Kang, C. H., Kim, R., Cho, J. M., Lee, Y. B., and Lee, T. K. (1989) *Protein Eng.* **2,** 571–575
15. Lee, S. Y., Lee, J. H., Chang, H. J., Cho, J. M., Jung, J. W., and Lee, W. (1999) *Biochemistry* **38,** 2340–2346
16. Sung, Y. H., Shin, J., Chang, H. J., Cho, J. M., and Lee, W. (2001) *J. Biol.*

Folding and Stability of Sweet Protein SCM

Chem. **276,** 19624–19630

17. Agashe, V. R., and Udgaonkar, J. B. (1995) *Biochemistry* **34,** 3286–3299
18. Piotto, M., Ssaudek, V., and Skenar, V. (1992) *J. Biomol. NMR* **2,** 661–665
19. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc.* **111,** 1515–1517
20. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry* **28,** 6150–6156
21. Bodenhausen, G., and Ruben, D. J. (1980) *Chem. Phys. Lett.* **69,** 185–189
22. Kay, L. E., Keifer, P., and Saainen, T. (1992) *J. Am. Chem. Soc.* **114,** 10663–10665
23. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., and Torchia, D. A. (1992) *J. Magn. Reson.* **97,** 359–375
24. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry* **28,** 8972–8979
25. Skelton, N. J., Palmer, A. G., Akke, M., Kordel, J., Rance, M., and Chazin, W. J. (1993) *J. Magn. Reson.* **102,** 253–264

26. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. E., Pawson, T., Forman-kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33,** 5984–6003
27. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc.* **115,** 12593–12594
28. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6,** 277–293
29. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR* **4,** 603–614
30. Lee, L. K., Rance, M., Chazin, W. J., Palmer, A. G., III (1997) *J. Biomol. NMR* **9,** 287–298
31. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
32. Chothia, C. (1984) *Annu. Rev. Biochem.* **53,** 537–572
33. Hopp, T. P., and Woods, K. R. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 579–582
34. Ali, V., Prakash, K., Kulkarni, S., Ahmad, A., Madhusudan, K. P., Bhakuni, V. (1999) *Biochemistry* **38,** 13635–13642