Solution Structure, Dynamics, and Thermodynamics of the Native State Ensemble of the Sem-5 C-Terminal SH3 Domain †

ARTICLE *in* BIOCHEMISTRY · JUNE 2003

Impact Factor: 3.02 · DOI: 10.1021/bi030005j · Source: PubMed

5 AUTHORS, INCLUDING:

David E Volk
University of Texas Health Science Center at …
62 PUBLICATIONS **786** CITATIONS
SEE PROFILE

David G Gorenstein
University of Texas Health Science Center at …
299 PUBLICATIONS **7,326** CITATIONS
SEE PROFILE

Vincent J Hilser
Johns Hopkins University
85 PUBLICATIONS **3,261** CITATIONS
SEE PROFILE

Solution Structure, Dynamics, and Thermodynamics of the Native State Ensemble of the Sem-5 C-Terminal SH3 Domain†

Josephine C. Ferreon, David E. Volk, Bruce A. Luxon, David G. Gorenstein, and Vincent J. Hilser*

Department of Human Biological Chemistry and Genetics and Sealy Center for Structural Biology, University of Texas Medical Branch at Galveston, Galveston, Texas 77555-1055

Received January 3, 2003; Revised Manuscript Received March 18, 2003

ABSTRACT: Although the high-resolution structure of a protein may provide significant insight into which regions are important for function, it is well-known that proteins undergo significant conformational fluctuations, even under native conditions. This suggests that the static structure alone may not provide sufficient information for elucidation of the thermodynamic determinants of biological function and that an accurate molecular-level description of function requires knowledge of the nature and energetics of the conformational states that constitute the native state ensemble. Here the native state ensemble of the C-terminal src homology domain-3 (C-SH3) from *Caenorhabditis elegans* Sem-5 has been studied using a variety of high-resolution biophysical techniques. In addition to determining the first solution structure of the unliganded protein, we have performed ^{15}N relaxation and native state hydrogen−deuterium exchange. It is observed that the regions of greatest structural variabilility also show low protection and order parameters, suggesting a higher degree of conformational diversity. These flexible regions also coincide with those regions of Sem-5 that have been predicted by the COREX algorithm to be unfolded in many of the most probable conformational states within the native state ensemble. The implications of this agreement and the potential role of conformational heterogeneity of the observed biophysical properties are discussed.

The existence of conformational excursions in proteins is well-established from hydrogen−deuterium (H/D)[1] exchange and NMR relaxation experiments (*1*−*5*). Despite knowledge of their presence, the extent of motion and the motional pathways have only been marginally characterized. In ensemble terms, little is known of the detailed structural and thermodynamic attributes of the relevant microscopic states comprising the native state ensemble. The importance of understanding the structural and thermodynamic details of the conformational states that exist under native conditions cannot be overstated. Because the observed biological activity of a particular protein results from the energy (or Boltzmann)-weighted contributions of the component microstates in the ensemble, knowledge of the structural and thermodynamic characteristics of these states is a prerequisite for a molecular-level understanding of protein function.

Most of what is known to date about the energetics of the native state conformational ensemble comes from native state amide H/D exchange experiments. According to these analyses, which have been performed on numerous proteins, most backbone amides within a particular protein become exposed to solvent, and therefore exchange competent, as a result of local unfolding reactions that are many times more probable than global unfolding under native conditions (*6*). While these experiments clearly demonstrate that the equilibrium under such conditions is more heavily influenced by local unfolding fluctuations than by global unfolding, a structural and thermodynamic characterization of the states comprising these fluctuations has proven to be problematic for several reasons. First, because hydrogen exchange does not provide the simultaneity of the exchange reactions, the cooperativity cannot be unambiguously assigned. Second, and equally important, the exchange rates of many amides are extremely high in typical H/D exchange experiments such that exchange is completed during the "dead time" of the experiment. As a consequence of this limitation, states with free energies of less than ∼3.0 kcal/mol (i.e., states with the highest probability) cannot be detected by standard H/D exchange experiments.

Here we utilize a wide range of high-resolution biophysical techniques to address the structure and energetics of the native state conformational ensemble of the C-terminal SH3 (C-SH3) domain of the *Caenorhabditis elegans* protein, Sem-5. In particular, we focus on the extent to which the component microstates in the macroscopic native state deviate both structurally and thermodynamically from the average structure. In addition to determining the ^{15}N relaxation rates, and H/D exchange rates, we have determined the first high-resolution structure of the unligated protein using multidimensional NMR. The experimental results

† Supported by National Science Foundation Grant MCB-9875689, National Institutes of Health Grant R01-GM13747, and Welch Award H-1461. J.C.F. is the recipient of a Sealy Center for Structural Biology Pre-Doctoral Fellowship.

* To whom correspondence should be addressed. Fax: (409) 747-6816. Telephone: (409) 747-6812. E-mail: vince@hbcg.utmb.edu.

[1] Abbreviations: SH3, src-homology domain 3; C-SH3, C-terminal SH3 domain; Sos, Son of Sevenless; mSos, murine Sos; 2D, two-dimensional; 3D, three-dimensional; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; HSQC, heteronuclear single-quantum coherence; NOESY, NOE spectroscopy; rmsd, root-mean-square deviation; H-bond, hydrogen bond; H/D, hydrogen−deuterium.

10.1021/bi030005j CCC: $25.00
Published on Web 04/25/2003

indicate that in addition to the high-energy states (>3.0 kcal/mol) revealed from H/D analysis, the native state of Sem-5 C-SH3 is characterized by various states whose summed probability may contribute substantially to the partition function and, therefore, to the observed biological function. We investigate the impact of these conformational fluctuations on the thermodynamics of the native state by using the COREX algorithm (*7*). The results indicate that conformational diversity (i.e., entropy) plays an important role in stabilizing the native state, and suggest that accurate structure-based models of biological function must account for this diversity.

MATERIALS AND METHODS

Overexpression and Purification of Sem-5 SH3 Domains. The C-terminal Sem-5 SH3 domain was cloned into the pET19b vector (Novagen) and overexpressed in BL21 (DE3) pLysS *Escherichia coli* cells as described previously (*8*, *9*). Cysteine residue 209 was mutated to alanine using standard PCR-directed mutagenesis (*10*) to prevent possible oxidation and intermolecular cross-linking. Cells expressing the protein were harvested and sonicated in 20 mL of 50 mM NaPhos, 300 mM NaCl buffer (pH 7.5) per liter of culture. The cell suspension was then centrifuged to obtain the supernatant, which contains the overexpressed protein. The protein, which contains a 10-histidine leader, was then purified using a nickel−NTA resin (Qiagen). The supernatant was loaded onto the Ni resin column where the His tag bound to the matrix. Subsequently, the column was washed and the protein eluted using 500 mM imidazole, in the same NaPhos buffer (pH 7.5). After final dialysis in enterokinase buffer [50 mM Tris, 1 mM $CaCl_2$, and 1% (w/v) Tween 20 (pH 8.0)], the protein was digested with enterokinase (Invitrogen) at 37 °C to cleave the His tag. Subsequently, the His tag was removed using a fast flow SP-Sepharose resin (Pharmacia) in 50 mM NaOAc, 10 mM $CaCl_2$, and 100 mM NaCl (pH 4.8) wherein the cleaved protein flows through while the His tag and uncleaved protein remain bound to the column. The purity of the eluted protein (>95%) was checked by electrospray ionization mass spectrometry (Mass Spectrometry Facility, University of Texas Medical Branch), reverse-phase HPLC, and SDS−PAGE. ^{15}N-labeled or ^{15}N- and ^{13}C-labeled protein was obtained from cells grown in M9 minimal medium containing either $^{15}NH_4Cl$ only or both $^{15}NH_4Cl$ and [U-^{13}C]glucose, respectively (Isotec and Cambridge Isotope Laboratories). The purified protein was concentrated to ~1 mM in 50 mM NaOAc, 100 mM NaCl, and 10 mM $CaCl_2$ (pH 4.8) (90% H_2O/10% D_2O) for structure determination by NMR. This protein has two extra residues at both the N-terminus and the C-terminus compared to the C-SH3 Sem-5 protein used for the X-ray structure determination in the complexed state with the polyproline peptide (*11*).

Sequential Assignments. All NMR spectra were recorded at 25 °C on a Varian UnityPlus 400 MHz instrument (with the exception of ^{13}C NOESY-HSQC spectra which were recorded on a Varian UnityPlus 600 MHz instrument) using a triple-resonance probe equipped with a pulsed field *z* gradient. Two-dimensional (2D) ^{1}H−^{15}N HSQC (*12*) and three-dimensional (3D) NMR CBCANH (*13*), CBCACONH (*14*), HNCO (*15*), HCCONH (*16*, *17*), CCONH (*16*, *17*), and HCCH-TOCSY (*18*) experiments were carried out. Carrier frequencies were 4.76 ppm for ^{1}H, 119.0 ppm for ^{15}N, 46 ppm for aliphatic ^{13}C, 56 ppm for $^{13}C_\alpha$, and 175 ppm for ^{13}C. The ^{1}H chemical shifts were referenced relative to DSS, and the ^{15}N and ^{13}C chemical shifts were referenced indirectly using the ^{1}H/X frequency ratios. All NMR data were processed in FELIX version 98.0 (MSI), and forward linear prediction was used to extend the time domain data in the indirectly detected dimensions. Sequential backbone resonance assignments were obtained by comparing the CBCANH and CBCACONH spectra, and verified by comparing the HNCO and HNCACO (*19*) spectra and/or by comparing the ^{15}N-edited TOCSY-HSQC and HCCONH spectra. Assignments for the carbonyl carbons were obtained from both the HNCO and HNCACO spectra, and side chain aliphatic proton assignments were obtained from the ^{15}N-edited TOCSY-HSQC and HCCONH spectra. Side chain aliphatic carbons were obtained from the CCONH spectrum, and the side chain protons were matched to their attached carbons in the HCCH-TOCSY spectrum. Finally, the terminal ends of Asn and Gln side chains were assigned by matching the α and β (Asn) or β and γ (Gln) carbons in the CBCANH (or CBCACONH) spectra with the terminal amide groups. Once the terminal amide groups were assigned, the side chain carbonyl carbons were assigned from the HNCACO spectrum. Aromatic resonance assignments were made using the (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE (*20*) pulse sequences obtained from L. Kay (University of Toronto, Toronto, ON).

Structure Calculation. Distance restraints were obtained from the ^{15}N-edited HSQC-NOESY (*12*, *21*) and ^{13}C-edited HSQC-NOESY spectra. Cross-peak volumes were integrated and converted to distance restraints by calibrating using the $\sim r^{-1/6}$ relationship. The highest cross-peak volumes obtained from the residues involved in the β-sheet are assumed to have a distance of 2.2 Å for sequential $d_{\alpha N}$ cross-peaks (*22*). Upper limits for the NOE volumes were divided into three categories: strong, <2.8 Å; medium, <3.5 Å; and weak, <5.0 Å. Lower limits were set to 1.8 Å in all cases. Additional limits (0.4−2.4 Å) were added to the upper limits of the distance restraints to account for pseudoatom correction (*22*). Backbone Φ torsion angle restraints were derived from the $^3J_{HN\alpha}$ coupling constants obtained from the 3D HNHA (*23*, *24*) experiment. The Φ dihedral angle restraints were restrained to $-120 \pm 30°$ for $^3J_{HN\alpha}$ values of >9 Hz, $-120 \pm 50°$ for $^3J_{HN\alpha}$ values between 8 and 9 Hz, and $-60 \pm 30°$ for $^3J_{HN\alpha}$ values of <6 Hz for helical regions only (*24*). Hydrogen bond restraints were obtained from the hydrogen−deuterium exchange. The ^{15}N-labeled protein was lyophilized and dissolved in 100% deuterated buffer [50 mM NaOAc, 100 mM NaCl, and 10 mM $CaCl_2$ (pH 4.8, uncorrected for isotope effects)]. The amides that are protected from exchange were determined from a ^{1}H−^{15}N HSQC spectrum after 24 h and were assigned as having hydrogen bonds. Two distance restraints were then used for each hydrogen bond: $d_{NH-O} < 2.3$ Å and $d_{N-O} < 3.3$ Å. There were 23 slowly exchanging amides identified (46 H-bond restraints), and an additional six H-bonds between amides and acidic residue side chains were added as evidenced by the initial structure calculations and the HSQC spectrum.

All structure calculations were carried out using CNSsolve version 1.0 (*25*). Initial embedded structures were generated from extended structures using the distance geometry/simulated annealing protocol. Default values were used for

all force constants and molecular parameters that were involved. High-temperature Cartesian molecular dynamics of with a 9 ps time step at 2000 K (3000 steps) were conducted. Cooling was then achieved by 4000 steps with a 20 ps time step. The 50 generated structures were then minimized with 2500 steps of Powell energy minimization. A final minimization of 100 cycles of 2500 steps with an NOE scale factor of 50 kcal mol^{-1} and a dihedral scale factor of 200 kcal mol^{-1} was run. The quality of the NMR structures was assessed using AQUA version 3.0 and Procheck-NMR version 3.4.4 (*26*) and visualized using Molmol version 2K.1 (*27*). The coordinates for the 16 lowest-energy structures and the minimized average structure have been deposited in the RCSB Protein Data Bank as entries 1kfz and 1k76, respectively. Chemical shift assignments have been deposited in the BioMagResBank as entry 5729 (http://www.bmrb.wisc.edu).

Hydrogen−Deuterium Exchange Experiments. The ^{15}N-labeled sample (~2.5 mM C-SH3 protein) was lyophilized from water, dissolved into 99.9% D_2O buffer [50 mM NaOAc and 100 mM NaCl (pH* 4.8, uncorrected)], and equilibrated for approximately 10 min. The dissolved protein was then immediately placed in the NMR tube and inserted into the NMR spectrometer. ^{1}H−^{15}N HSQC spectra were then recorded with identical acquisition parameters at different time intervals. The total measuring time for each HSQC spectrum was ~10 min. Cross-peak volumes were obtained from a volume integration routine in Felix. To calculate the exchange rates, the peak volumes corresponding to each amide peak acquired at different time intervals were fitted to a three-parameter single-exponential decay function:

$$I(t) = I_o e^{-kt} + C \quad (1)$$

where C is the baseline noise and k is the exchange rate using Microcal Origin (Microcal Software, Inc.). Random coil hydrogen exchange rate constants were obtained by the method of Bai et al. (*28*, *29*).

^{15}N Relaxation Experiments. ^{15}N T_1, ^{15}N T_2, and {^{1}H}−^{15}N NOE NMR experiments were conducted using sequences within ProteinPack (Varian, Inc.) which are based on pulse sequences described previously (*30*). R_1 (T_1^{-1}) data were acquired with 10 relaxation delays (10, 40, 80, 100, 150, 230, 320, 430, 590, and 840 ms), while R_2 (T_2^{-1}) data were obtained with 11 relaxation delays (10, 30, 50, 70, 90, 110, 150, 170, 190, 210, and 250 ms). Both R_1 and R_2 experiments employed 1 s recycle delays. The {^{1}H}−^{15}N NOEs were measured by recording HSQC spectra with and without proton saturation. The former spectra used a 3 s proton saturation and a 2 s delay, while the latter used a 5 s delay only. Duplicates were made for all spectra that were acquired. The spectra were apodigitized with a 90° shifted sine-bell window function and zero-filled to a 256 × 1024 matrix. Peak height intensities were measured by the automated routine in the software. The ^{15}N T_1 and T_2 time constants were obtained by fitting the data to a two-parameter single-exponential decay function using software obtained from N. A. Farrow. Appropriate model fitting of the relaxation data was performed as suggested by Palmer and co-workers (*31*, *32*), and using their software (Modelfree 4.01, http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer). The isotropic diffusion tensor was also determined to be the appropriate choice for the C-SH3 from statistical tests, also using other software from their group (pdbinertia, r2r1_diffusion). A detailed analysis of dynamics in the Sem-5 C-SH3 is described elsewhere (*33*).

Table 1: Summary of the Number of Restraints Used in Structure Calculations

total no. of NOE-derived restraints	1146
intraresidue	497
sequential	263
medium-range	68
long-range	318
ϕ dihedral restraints	34
H-bond restraints	52
total	1232

RESULTS

Solution Structure of the Unliganded Sem-5 C-SH3. As was the case with the Grb2 SH3 domain (*34*, *35*), Cys209 in the wild-type Sem-5 C-SH3 (*8*, *9*, *11*) was mutated to alanine to prevent oxidation and intermolecular cross-linking. The melting temperature (T_m), free energy of unfolding ($\Delta G°_{N-D}$), and dissociation constant of binding with the polyproline peptide (K_d) were found to be similar to those of the wild-type SH3 domain under the conditions used for structural analysis (*8*).

Complete backbone and side chain ^{1}H, ^{13}C, and ^{15}N resonance assignments have been made and deposited in the BioMagResBank (entry 5729). A total of 1232 restraints (Table 1) were obtained from NMR experiments and used in structure calculations using CNSsolve version 1.0 (*25*) as described in Materials and Methods. These include 1146 NOE-derived distance restraints (Figure 1A), 34 ϕ dihedral angle restraints, and 52 H-bond restraints. Panels A and B of Figure 2 show the NMR-derived ensemble of the 16 lowest-energy conformers. The structures have converged as indicated by the low backbone rmsd (Figure 1B). The N- and C-terminal residues are poorly defined due to the small number of medium- and long-range NOEs in this region. Table 2 shows the energetic and structural statistics for the 16 lowest-energy conformers of the NMR-derived Sem-5 C-SH3 ensemble. There are no distance restraint and dihedral restraint violations of greater than 0.2 Å and 2°, respectively, and more than 96% of the residues are located in most favored and additionally allowed regions. Nine of the 16 structures in the Sem-5 C-SH3 ensemble have Asn197 (rmsd of 0.74 Å) located in the disallowed region. This residue is the only non-glycine residue exhibiting a positive ϕ angle and is located in a turn in the distal loop, as observed in Grb2 N-SH3 domain (*36*). This residue falls at the edge of the generously allowed region in the crystal structure (PDB entry 1SEM) and is involved in crystal contacts with other asymmetric units. The distal, n-Src, and RT loops have also been found to have higher backbone rmsd values and *B*-factors than other residues in other SH3 domains (*37*, *38*).

The tertiary fold of the Sem-5 C-SH3 is similar to other known structures of SH3 domains, following the "classic" β-barrel-like SH3 topology consisting of two β-sheets of five antiparallel strands packed nearly perpendicular to each other. For the Sem-5 C-SH3, a single turn of conserved 3_{10}-helix connects the last two β-strands, β_d and β_e. Three less conserved yet important loops connect the β-strands. The RT loop (between strands β_a and β_b) and the n-Src loop

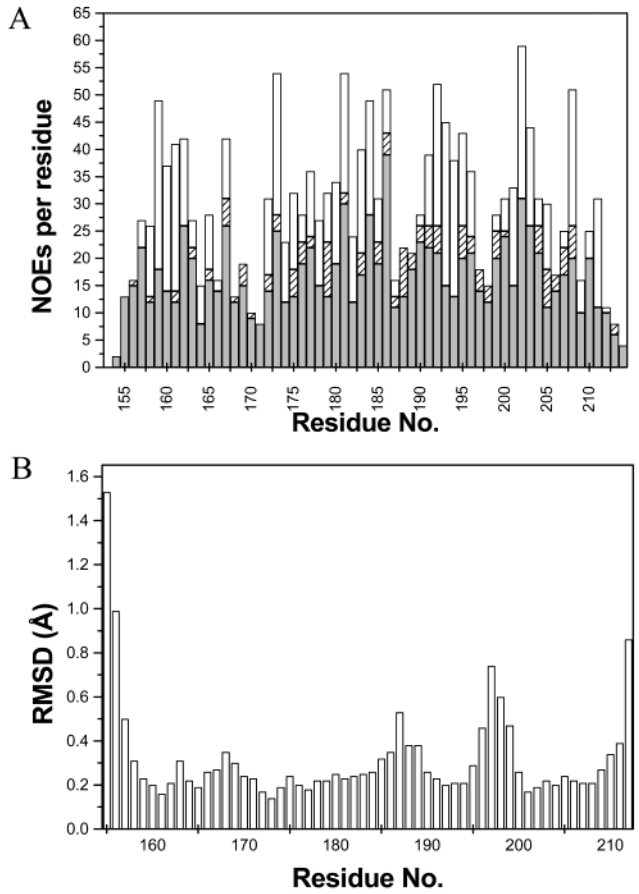

FIGURE 1: (A) Number of $^1H-^1H$ NOEs per residue used in the structure calculation. Light gray is for intraresidue and sequential NOEs ($|i - j| \leq 1$), hatched for medium-range NOEs ($2 \leq |i - j| \leq 4$), and white for long-range NOEs ($|i - j| \geq 5$). (B) Backbone rmsd values for the 16 lowest-energy conformers in the ensemble.

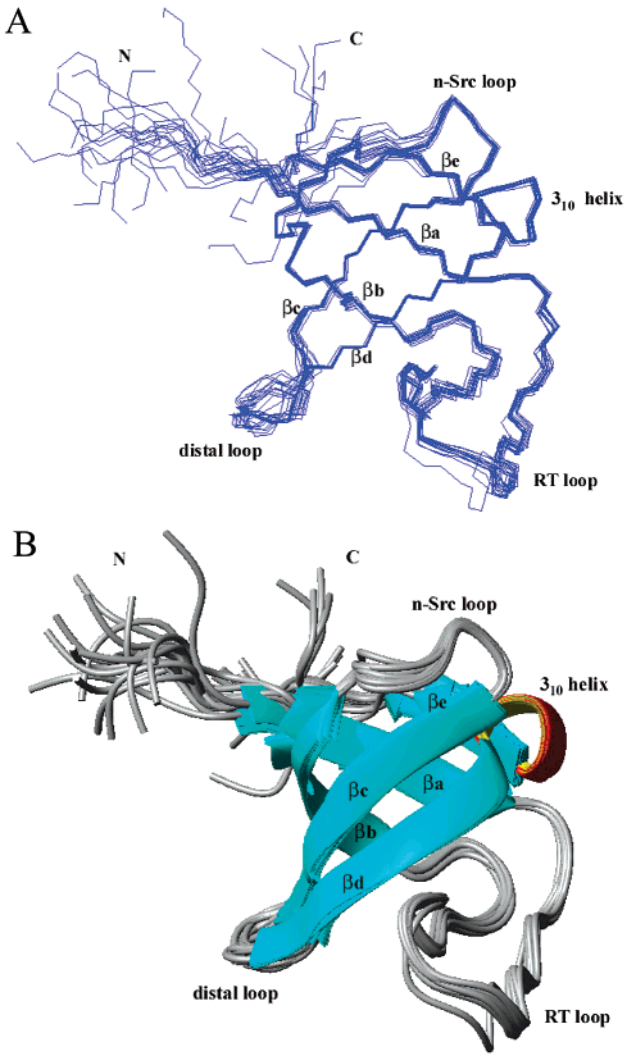

FIGURE 2: (A) Line representation of the backbone and (B) ribbon representation of the 16 lowest-energy conformers calculated for the Sem-5 C-SH3.

(between strands β_b and β_c) have been implicated as major subsites for binding and specificity. Acidic and polar residues in these loops are believed to play a role in determining the orientation for the docking of the peptide onto the SH3 domain surface (*11*).

Comparison of the Unliganded NMR-Derived Ensemble of the Sem-5 C-SH3 with the Liganded X-ray Crystal Structure. The crystal structure of the Sem-5 C-SH3 was among the first SH3 structures determined that revealed the minus binding orientation of the mSos-derived peptide sequence PPPVPPRRR (*11*). Figure 3A shows the superposition of the 16 lowest-energy NMR-calculated conformers with the crystal structure (PDB entry 1SEM). The NMR ensemble of structures shows similar packing of the β-sheets and an equivalent arrangement of the secondary structure elements. This high degree of similarity is manifested by a 0.61 Å backbone rmsd when excluding the loops. Inclusion of the loops results in an overall backbone rmsd of 1.12 Å. The deviation is largely influenced by the noticeable difference in the RT and distal loop (Figure 3A−C).

To ensure that the observed differences between the NMR and crystal structure are not simply artifacts of the structure calculations, the raw NMR data (NOESY spectra) were extensively analyzed to identify NOEs between proximal proton pairs in the crystal structure. More than 20 proton pairs with a distance of less than 5 Å in the crystal structure were identified where no NOEs were present in the NOESY spectra. Indeed, the structural differences observed here are consistent with differences observed for other SH3 domains. Work by Byrd and colleagues (*39*), as well as by Nicholson and colleagues (*40*), suggests that ligand binding to SH3 domains can be characterized as "induced fit".

In comparing the solution structure of Hck in the unliganded form to that of the protein in complex with its peptide ligand, Byrd and colleagues show conformational changes at the aromatic binding interface in the 3_{10}-helix (*39*). Similarly, in the Sem-5 C-SH3, substantial shifts occur among the side chains of Trp191, Pro204, Tyr207, Phe165, Phe163, Asn206, and Pro204 when comparing the unbound NMR-derived ensemble and the complexed crystal structure. Interestingly, the side chains of several residues occupy positions that are in van der Waals violation with the peptide in the NMR-derived ensemble but not in the crystal structure. These conformational changes were attributed to the absence of NOEs among key residues (*39*). Although several NOEs between H^{β}, H^{γ}, and H^{δ} of Pro204 and H^N, H^{β}, and H^{δ} of Tyr207 are observed in Sem-5, there is an absence of NOEs for other aromatic residues, namely, Phe165 H^{β} and Pro204 H^{γ} and H^{δ}, Phe163 H^{β} and Tyr207 H^{β} and H^{ϵ}, and Asn206 H^{δ} and Pro204 $H^{\beta 2}$, which correspond to distances of 2.67−4.53 Å in the crystal structure.

Table 2: Structural Statistics for the Sem-5 C-SH3 NMR Ensemble (16 Structures)

atomic rmsd (Å)[a]	
backbone atoms	0.403
heavy atoms	1.24
rmsd from experimental data[b]	
distance restraints (Å)	0.0137 ± 0.0012
dihedral angle restraints (deg)	0.1333 ± 0.0387
no. of violations[b]	
distance restraints (>0.2 Å)	0
dihedral angle restraints (>2°)	0
rmsd from idealized covalent geometry[b]	
bond lengths (Å)	0.0014 ± 0.00009
bond angles (deg)	0.3168 ± 0.0075
improper angles (deg)	0.1458 ± 0.0136
energetic statistics (kcal mol^{-1})[b,c]	
overall	59.71 ± 5.33
bond	2.01 ± 0.27
angle	27.65 ± 1.32
improper	1.84 ± 0.35
vdw	16.83 ± 2.09
NOE	11.34 ± 2.00
Ramachandran plot analysis[b]	
most favored regions	67.8% (564)
additional allowed regions	29.0% (241)
generously allowed regions	1.8% (15)
disallowed regions	1.4% (12)

[a] Calculated over residues 155−212. [b] Calculated over residues 153−214. [c] Energies as calculated by CNS.

The quality of the structural data, the lack of NOEs for proximal proton pairs, and the agreement with previously published results for other SH3 domains indicate that the solution structure determined here provides a suitable canonical representation of the unliganded Sem-5 C-SH3, and that the canonical structure of the unbound Sem-5 C-SH3 is significantly different from the canonical X-ray structure of the complex. The significance of the observed structural differences is addressed in the Discussion.

Native State Hydrogen−Deuterium (H/D) Exchange. Native state amide hydrogen−deuterium (H/D) exchange represents a unique tool for the study of protein folding, as it provides a means of monitoring the folding−unfolding reaction from a large number of sites throughout the protein structure (*29*, *41*, *42*). For a subset of amide protons, exchange with solvent protons can only occur through the following reaction:

$$NH_j(\text{closed}) \overset{K_{op}}{\longleftrightarrow} NH_j(\text{open}) \xrightarrow{k_{int}} \text{exchanged}$$

where K_{op} is the equilibrium constant for the opening reaction of residue j and k_{int} is the intrinsic exchange rate of amide j in the open form. Such protons are said to be "protected" from exchange, and are usually located in the interior of the protein or are involved in hydrogen bonding. In the limiting case, where the folding−unfolding equilibrium occurs on a much faster time scale than the exchange reaction (i.e., in the EX2 or bimolecular exchange regime), the measured exchange rate for these protons, k_{obs}, can be expressed as

$$k_{obs} = K_{op}k_{int} \quad (2)$$

Since k_{int} is known from exchange studies of peptides (*28*), the equilibrium for the closing−opening reaction can be evaluated directly from measured exchange rates. Under native conditions, wherein protected protons exchange only

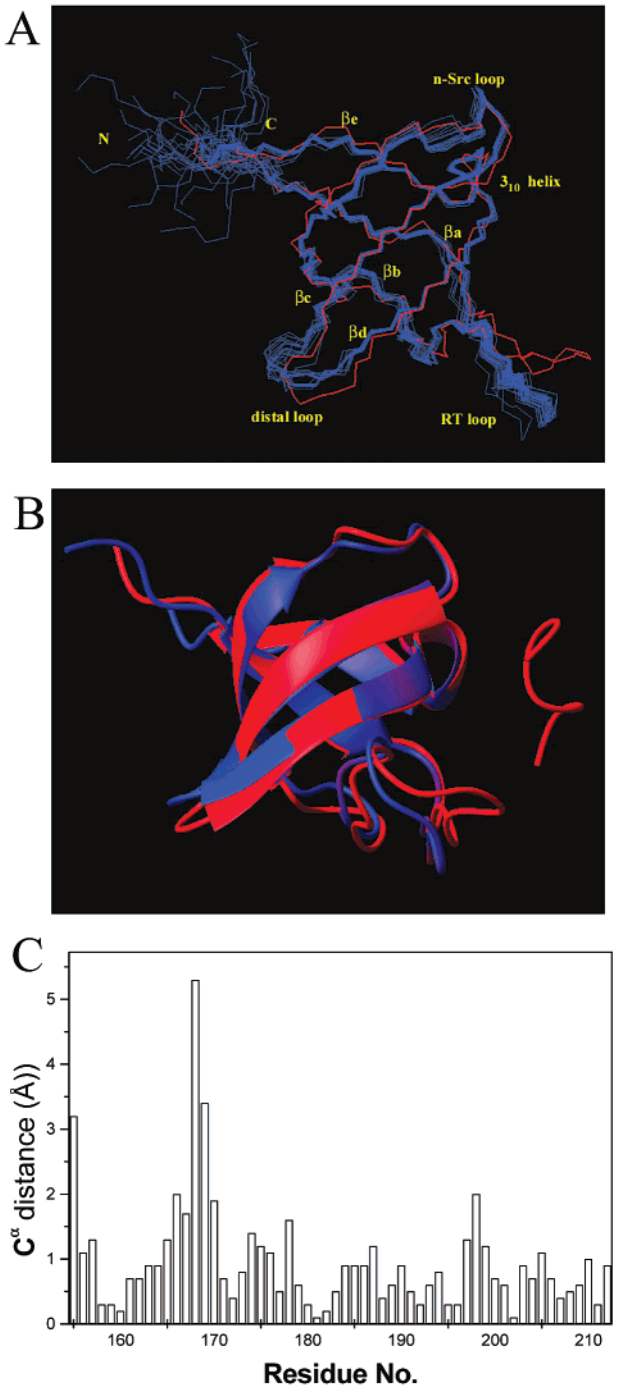

FIGURE 3: Difference between the canonical X-ray and NMR structures. (A) Overlay of the ensemble of 16 lowest-energy conformers (blue) with the X-ray structure (PDB entry 1SEM, chain A) of the Sem-5 C-SH3. (B) Overlay of the NMR-minimized average structure (blue) with the crystal structure (red) in the ribbon representation. (C) Absolute C_α deviation of the minimized NMR structure from the crystal structure.

as a result of partial or total unfolding of the protein, the equilibrium constant for the residue-specific folding−unfolding reaction is the same as that for the closing−opening reaction, and can be expressed as a protection factor PF (PF = $1/K_{op}$ = k_{int}/k_{obs}) for that residue. From the PF, an expression for the residue-specific free energy of unfolding ($\Delta G_{H/D,j}$) can be written

$$\Delta G_{H/D,j} = -RT \ln(PF_j) \quad (3)$$

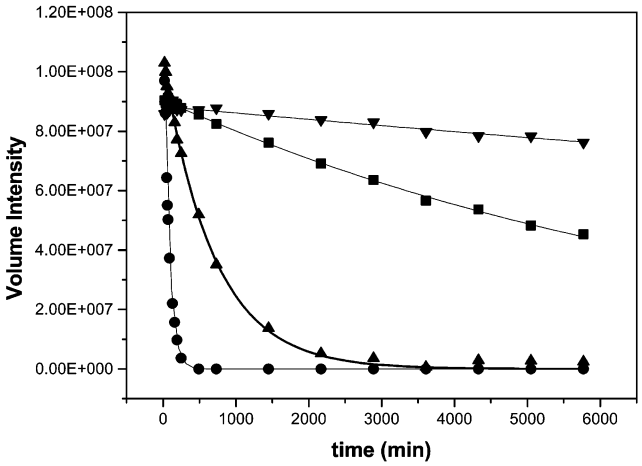

FIGURE 4: Selected residues [Glu172 (●), Asn185 (▲), Gly194 (▼), and Ser205 (■)] of the Sem-5 C-SH3 exhibiting different exchange rates measured by hydrogen−deuterium exchange. Glu172 belongs to the RT loop and Asn185 to the n-Src loop; Gly194 is on the core of the β-sheet sandwich, and Ser205 is part of the 3_{10}-helix.

Thus, hydrogen−deuterium exchange provides a means of determining the relative free energy difference between the closed and open subensemble for each residue.

Shown in Figure 4 are the exchange rates for representative amides of the Sem-5 C-SH3. Residues that exchange by the EX2 mechanism were identified by monitoring the change in exchange as a function of pH (data not shown). The quality of the fits to the data suggests that the exchange can be accurately determined over a wide range of values. Figure 5 shows the protection factors determined from eq 2. As is evident from Figure 5, residues in the core β-barrel exhibit the greatest protection, while residues in the loops exhibit weaker protection, indicating that the loops are exchange competent in many of the high-probability states in the native state ensemble.

Relaxation Data Analysis. ^{15}N relaxation experiments were performed and the data analyzed as described elsewhere (*33*). All ^{15}N T_1 and ^{15}N T_2 values were adequately fit to single-exponential two-parameter decays. The average value of R_1 is 2.84 ± 0.29 s^{-1}, of R_2 is 6.30 ± 0.86 s^{-1}, and of ^{15}N−H NOE is 0.53 ± 0.39 for the C-SH3 protein. Figure 6 shows the R_1, R_2, and NOE profiles for the C-SH3. The R_1 profile shows a distribution where there are significantly lower than average R_1 values in the loop regions, i.e., the RT, n-Src, and distal loops, as well as residues in the termini, indicating greater than average mobility in these regions. The R_2 and NOE distributions also show a similar profile, where most of the residues in the loop regions exhibit lower values.

As described previously (*33*), using the Lipari−Szabo model-free formalism, backbone amides were model-fitted to obtain the generalized order parameter (S^2, Figure 7). Consistent with R_1, R_2, and NOE data, the S^2 profile shows higher than average S^2 ($S^2 > 0.8$) values in β-strands I−V and in the 3_{10}-helix, while residues in the N- and C-termini or loop regions (i.e., RT, n-Src, and distal loop) have lower than average S^2 values. Among the different regions, the RT loop residues (Asn166−Ser171) have the lowest S^2 values (Glu169 was found to be the most mobile with an S^2 of 0.52). The results obtained for the Sem-5 C-SH3 are consistent with regional differences in dynamics observed for other SH3 domains.

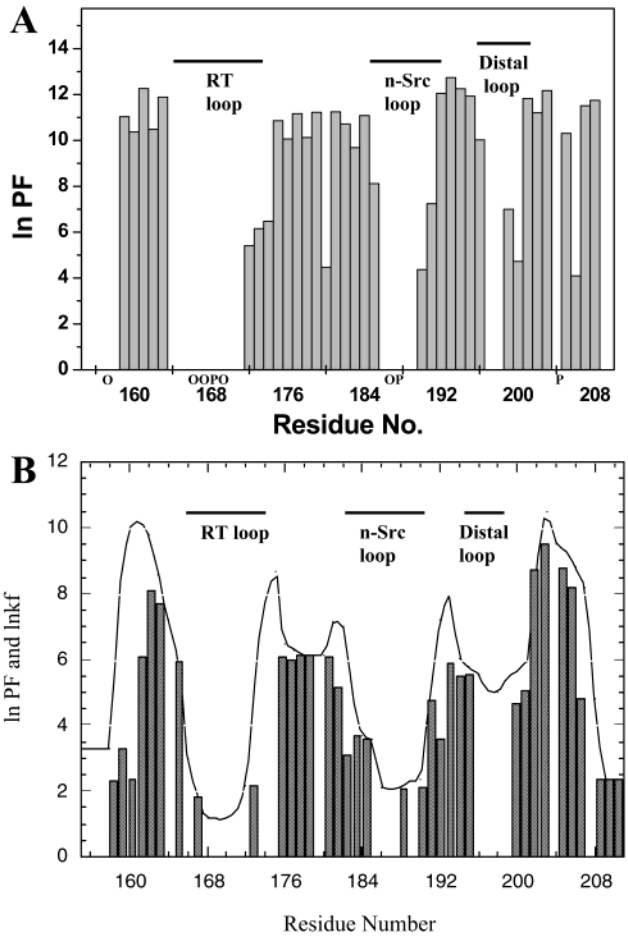

FIGURE 5: (A) Hydrogen−deuterium exchange protection factors (PFs) for the Sem-5 C-SH3. Residues with no PFs are either prolines (designated by P), residues that overlap in the HSQC spectra (O), or residues whose exchange occurs within the dead time of the experiment. (B) Calculated stability constants (κ_f, line) and protection factors (bars) for the Sem-5 C-SH3 using the COREX algorithm.

As discussed elsewhere, the general pattern of low order parameters in the loop regions, particularly the RT, n-Src, and distal loops, is common among SH3 domains (*33*, *43*). Interestingly, the relative flexibility of the loops within a particular SH3 domain varies. For the Hck SH3 domain, for example, the n-Src loop shows greater disorder than the RT or distal loops (*43*), while in the Btk SH3 domain, the RT and n-Src loops have similarly high levels of disorder (*38*). As indicated by Figure 7, both of these situations differ from the regional difference in dynamics seen in the Sem-5 C-SH3. Although the observed differences in dynamics between SH3 domains are significant, we have no explanation for the determinants of these differences.

DISCUSSION

Regional Differences in the Structure and Stability of the Sem-5 C-SH3. Inspection of Figure 3 reveals that the canonical solution structure of the unliganded Sem-5 differs from that of the Sem-5−Sos peptide complex, with the major difference being in the RT loop. This result indicates that the RT loop possesses a high degree of conformational plasticity and that significant structural deviations are not energetically prohibited in this region. This structural result is consistent with the energetic and dynamic information

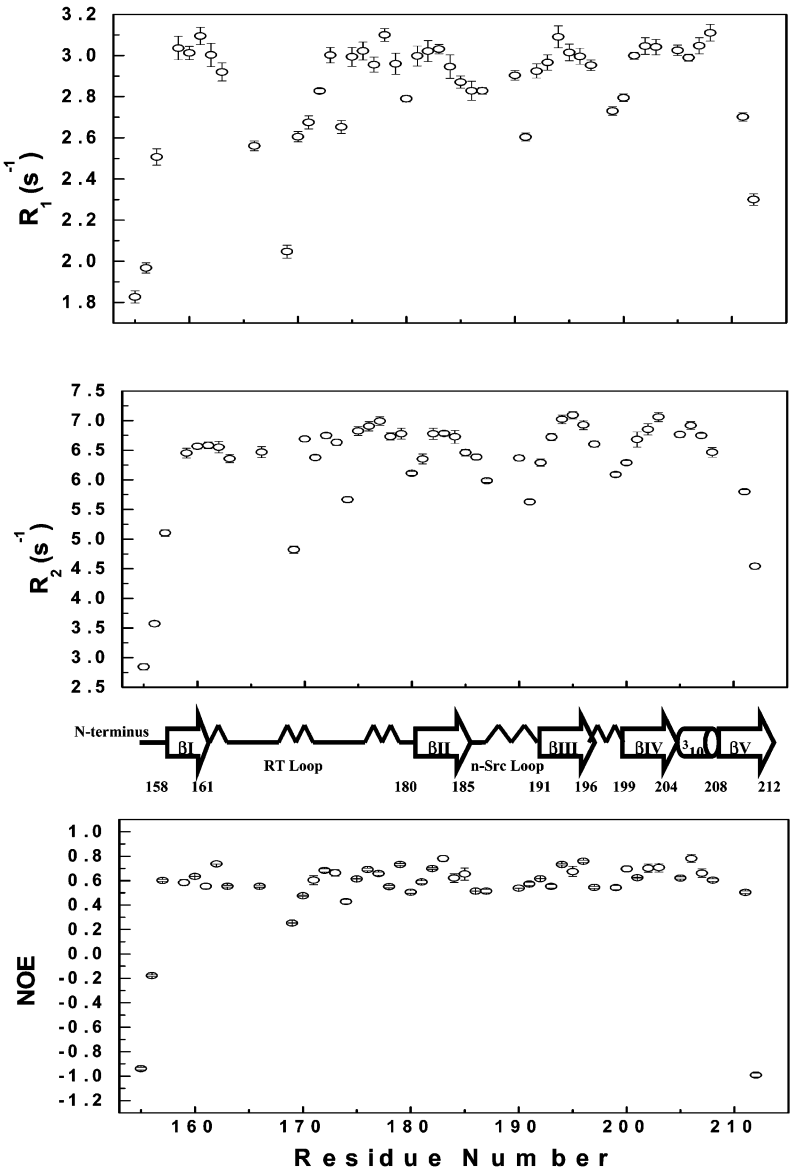

FIGURE 6: Relaxation data for the Sem-5 C-SH3 in the unliganded state. Elements of secondary structure are the same for both states. β-Sheets are represented as arrows; the 3_{10}-helix is represented by a cylinder, and loops are represented by straight black lines and turns by bent lines.

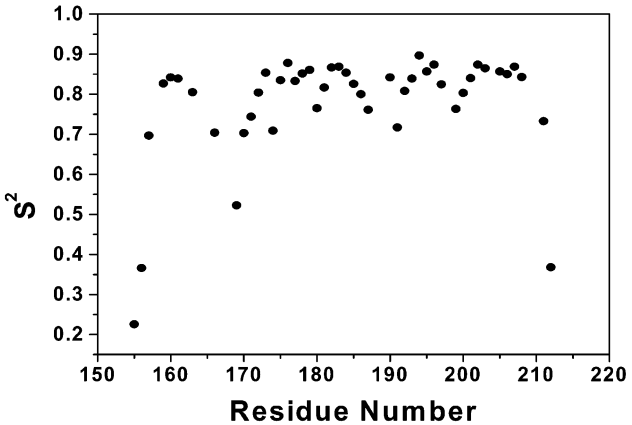

FIGURE 7: Generalized order parameter (S^2) profile for the unliganded Sem-5 C-SH3.

shown in Figures 5−7. Namely, although some regions of the Sem-5 C-SH3 display a wide variability in hydrogen exchange behavior, with no clear similarities between neighboring residues, the residues in the RT loop show trends in behavior that progress over many residues. From these trends, we can identify three classes of exchange. Residues 169−171 have exchange rates that are too rapid to be measured by H/D exchange and are categorized as "fast exchangers". Further along the sequence, residues 172−174 show slower H/D exchange, with the level of protection being lower than that for global unfolding, and are thus categorized as "moderate exchangers". "Slow exchangers" are those residues that exchange as a result of global unfolding (residues 175−177). These three classes of exchange behavior suggest that the RT loop is less stable than the core of the molecule, and that the tapering in the protection is monitoring the stability differences in the different regions of the loop. Indeed, the progression in exchange behavior described here is also seen in the order parameters as shown in Figure 7, and is suggestive of a correlation in their behavior. It should be noted that of the residues comprising the RT loop (162−173), the amide nitrogens of only Asp164, Asn166, Ser170, and Gly171 display accessible surface area

in either the NMR structure presented here or the X-ray structure of the complex. This indicates that hydrogen exchange at all other amides in the RT loop is dependent on conformational changes that either expose the amides to solvent or allow solvent to penetrate to the amide and facilitate exchange (*6*, *44*).

The combined structural, energetic, and dynamic data presented here reveal a native state ensemble somewhat different from that suggested by H/D exchange alone. Specifically, because H/D exchange experiments cannot monitor the most rapidly exchanging protons, states that are present at a level of greater than ~1 part in 1000 are effectively invisible to these studies. Such a picture of the native state ensemble is incomplete, and ^{15}N relaxation results suggest that states that are conformational excursions from the high-resolution structure are significant components of the native state ensemble, a scenario which, as shown below, could dramatically impact not only our ability to understand how proteins are stabilized but also how they function.

Structural Thermodynamic Analysis of the Sem-5 C-SH3 Native State Ensemble. To assess the potential energetic impact of conformational fluctuations on the native state of Sem-5, an ensemble of states was generated and analyzed using the COREX algorithm (*7*). Briefly, the COREX algorithm models the native state as an ensemble rather than as a discrete conformation using the high-resolution structure as a template (*45*). Through the combinatorial unfolding of a predefined set of folding units, the COREX algorithm exhaustively enumerates all possible partially unfolded states. By virtue of the parametrized energetics described previously (*46*−*48*), the free energy and thus probability of each state (P_i) are defined as

$$P_i = \frac{\exp -\Delta G^i/RT}{\sum \exp -\Delta G^i/RT} \quad (4)$$

where the summation in the denominator is over all states in the ensemble, and represents the partition function for the system. From the probability described by eq 4, a residue-specific descriptor of the equilibrium can be calculated. Known as the residue stability constant ($\kappa_{f,j}$), this quantity is the ratio of the summed probability of states in the ensemble in which a particular residue is folded ($\Sigma P_{f,j}$) to the summed probabilities of all states in which that residue is unfolded ($\Sigma P_{nf,j}$):

$$\kappa_{f,j} = \frac{\sum P_{f,j}}{\sum P_{nf,j}} \quad (5)$$

The importance of the residue stability constants is that they can be compared to the protection factors obtained from hydrogen exchange measurements (*7*, *45*, *49*, *50*), thus representing an experimentally verifiable description of the ensemble.

As seen in Figure 5B, the calculated stability constants for the Sem-5 C-SH3 agree both in location and relative magnitude with the protection factors obtained from H/D exchange. Of particular note are the residues in the RT loop. Although many residues (i.e., 165, 168, 169, and 172) have amides that are buried from solvent in the high-resolution structure, they experience no observed protection as noted above. Consistent with these results, the COREX algorithm predicts that the entire RT loop is significantly unstable. Given the experimental sampling time and the predicted stability of these residues, the COREX algorithm predicts that no significant protection should be observed, as is the case.

Table 3: Summary of Energetics of COREX-Derived States with Certain Residues in the RT Loop Unfolded

unfolded residues	ΔC_p [a]	ΔH[b]	$-T\Delta S$[b]	ΔG[b]	$P_i\Delta H_i$ [b]
167−170	41.8	1555.6	−856.3	699.3	23.3
167−171	34.7	2552.5	−3089.4	−536.9	314.1
167−172	82.9	389.6	−1498.3	−1108.7	125.8
168−171	−10.2	4730.2	−5405.4	−675.2	735.4
168−172	49.9	2040.2	−3016.4	−976.2	527.2
169−172	52.9	1364.8	−1606.1	−241.3	102.1
					$\langle \Delta H_{conf} \rangle$[c] = 1828.8

[a] ΔC_p is the heat capacity change in calories per mole per kelvin. [b] ΔH, $-T\Delta S$, ΔG, and $P_i\Delta H_i$ in calories per mole. [c] Excess conformational enthalpy ($\langle \Delta H_{conf} \rangle = \Sigma P_i \Delta H_i$) in calories per mole.

Of particular importance in Figure 5B is the fact that the magnitude of the stability constants in the RT loop indicates that many of the most probable states in the Sem-5 C-SH3 ensemble have some portion of the RT loop unfolded. In other words, if a single molecule were to be observed over a sufficient time, regions of the RT loop would undergo conformational fluctuations (i.e., folding−unfolding reactions) a significant fraction of the time. This result is consistent with the ^{15}N relaxation measurements of the Sem-5 C-SH3 (Figure 7), which shows significant disorder of the RT loop.

The available experimental and computational results for the Sem-5 C-SH3 suggest that the native state is actually an ensemble of states wherein regions of the protein, especially in the RT loop, undergo conformational fluctuations, and that these conformational fluctuations may in fact dominate the partition function (i.e., the majority of the states that contribute substantially to the denominator in eq 4 are conformational excursions of the canonical structure). An interesting question concerns the nature of these fluctuations. In ensemble terms, how much do the structure and energy vary among the different states, and how much does that impact the view that the native state can be quantitatively represented as a unique (canonical) structure?

In the context of the COREX algorithm, fluctuations in the RT loop are modeled as order to disorder (or local unfolding) transitions, which of course is only one of a number of possible fluctuation models. Nevertheless, it is considerably beneficial to obtain an estimate of the energetic impact of fluctuations using the COREX algorithm as COREX quantitatively captures the H/D exchange behavior (Figure 5). To estimate the quantitative impact of changes in conformational heterogeneity, the energy difference between the ensemble of partially folded states in the Sem-5 ensemble and the structure of the unliganded C-SH3 was determined. Shown in Table 3 are the dominant states, their relative enthalpies and entropies, and the corresponding statistical weights determined from a COREX analysis of the solution structure of the Sem-5 C-SH3 described above. As the statistical weights determine the relative contribution

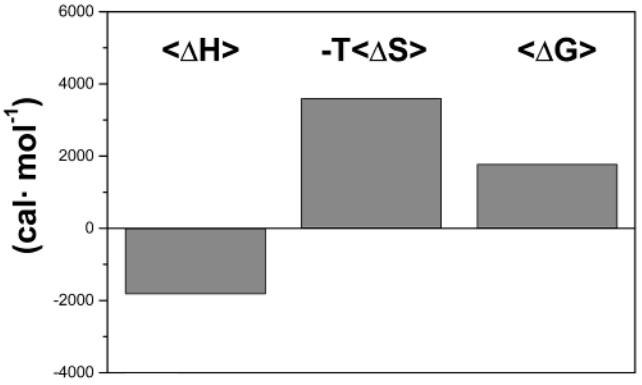

FIGURE 8: Residual enthalpy $\langle \Delta H \rangle$, entropy $T\langle \Delta S \rangle$, and free energy $\langle \Delta G \rangle$ of the native state ensemble, calculated from the states depicted in Table 3. Residual quantities are defined here as the energy difference between the canonical structure and the ensemble of states (e.g., $\langle \Delta H \rangle = 0 - \langle \Delta H_{conf} \rangle$) which represents the energy that must be added to the system to obtain the canonical structure.

of each state to the ensemble properties, the thermodynamic impact of each state can be calculated.

Shown in Figure 8 is the energy associated with the degeneracy in the native state as determined from the most probable states in the Sem-5 ensemble shown in Table 3. Although the time scale of exchange in the RT loop renders H/D exchange only slightly useful in experimentally verifying the calculated stability within the loop itself, the good agreement in the relative magnitude and location of protection factors throughout the rest of protein suggests that the energies shown in Figure 8 are reasonable representations of the actual energies. In addition, as the order parameters in the RT loop indicate significant disorder, the energies of the states with alternative conformations in the RT loop must contribute substantially to the partition function.

An important implication of Figure 8 is that the negative enthalpy and entropy, coupled with the positive free energy, reveal that constraining the ensemble (i.e., folding the loop) results in a significant release of heat. The values in Figure 8 correspond to the excess (or residual) enthalpy and entropy of the native state (relative to the high-resolution structure), and correspond to the energy difference between the canonical structure and the ensemble values determined in Table 3. The magnitude of the values in Figure 8 is striking, and suggests that the energetics of the high-resolution structure alone may not provide a sufficient representation, structurally or energetically, of the native state.

The most immediate implication of the values shown in Figure 8 is the possible effect that changes in the distribution of states could have on binding. Specifically, if binding results in a significant constriction of the ensemble, the experimentally observed enthalpy, entropy, and free energy of binding will reflect this process. As such, the measured values would be significantly lower (more negative) than those that might be estimated from inspection of the high-resolution structure (assuming a static model). The extent to which ensemble redistribution in the Sem-5 C-SH3 affects the binding to its putative peptide ligand is currently under investigation.

The magnitudes of the energies shown in Figure 8 are estimates and clearly depend on the local unfolding model of fluctuations that is used in the COREX algorithm. In this respect, the local unfolding model for fluctuations represents an extreme case, which can be used to provide an estimate of the upper limit for the energetic variations within the native state. Notwithstanding this point, the fact that COREX analysis is consistent with the H/D exchange results indicates that energetic variation within the native state of a protein can be significant. This means that the canonical structure may not provide a suitable function description of proteins (regardless of the resolution of the structure) and that a quantitative description of conformational fluctuations is a prerequisite for a molecular-level understanding of the biophysical and functional behavior.

NOTE ADDED AFTER ASAP POSTING

This article was posted prematurely on the Web on 04/25/03. The financial support for this work is now included in the title footnote. The correct version of this paper was posted on the Web on 04/29/03.

ACKNOWLEDGMENT

We thank Edward Ezell and Drs. Shanmin Zhang, Quinn Kleerekoper, Krishna Rajarathnam, Youxing Qu, and Jamie Wrabl for NMR assistance. We are also grateful to Dr. Robert Fox for providing the original clone of the wild type C-SH3 and for useful insights and discussion.

REFERENCES

1. Englander, S. W., and Kallenbach, N. R. (1983) *Q. Rev. Biophys. 16*, 521−655.
2. Wagner, G. (1983) *Q. Rev. Biophys. 16*, 1−57.
3. Clore, G. M., Driscoll, P. C., Wingfield, P. T., and Gronenborn, A. M. (1990) *Biochemistry 29*, 7387−7401.
4. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry 28*, 8972−8979.
5. Palmer, A. G., Rance, M., and Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371−4380.
6. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science 269*, 192−197.
7. Hilser, V. J., and Freire, E. (1996) *J. Mol. Biol. 262*, 756−772.
8. Lim, W. A., Fox, R. O., and Richards, F. M. (1994) *Protein Sci. 3*, 1261−1266.
9. Lim, W. A., and Richards, F. M. (1994) *Struct. Biol. 1*, 221−225.
10. McPherson, M. J., Hames, B. D., and Taylor, G. R. (1995) in *PCR2. A Practical Approach Series*, Oxford University Press, Oxford, U.K.
11. Lim, W. A., Richards, F. M., and Fox, R. O. (1994) *Nature 372*, 375−379.
12. Wang, C., Pawley, N. H., and Nicholson, L. K. (2001) *J. Mol. Biol. 313*, 873−887.
13. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
14. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 99*, 201−207.
15. Ikura, M., Kay, L. E., and Bax, A. (1990) *Biochemistry 29*, 4659.
16. Grzesiek, S., Anglister, J., and Bax, A. (1993) *J. Magn. Reson., Ser. B 101*, 114−119.
17. Lyons, B. A., and Montelione, G. T. (1993) *J. Magn. Reson., Ser. B 101*, 206−209.
18. Sattler, M., Schwendinger, M. G., Schleucher, J., and Griesinger, C. (1995) *J. Biol. NMR 6*, 11−22.
19. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. (1994) *J. Am. Chem. Soc. 116*, 11655−11666.
20. Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1993) *J. Am. Chem. Soc. 115*, 11054−11055.
21. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.
22. Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
23. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.

24. Zhang, W., Smithgall, T. E., and Gmeiner, W. H. (1997) *J. Biomol. NMR 10*, 263−272.
25. Brünger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
26. Laskowski, R. A., Rullman, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
27. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
28. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
29. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1994) *Proteins: Struct., Funct., Genet. 20*, 4−14.
30. 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) *Biochemistry 33*, 5984−6003.
31. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) *J. Mol. Biol. 246*, 144−163.
32. Mandel, A. M., Akke, M., and Palmer, A. G. (1996) *Biochemistry 35*, 16009−16023.
33. Ferreon, J. C., and Hilser, V. J. (2003) *Protein Sci.* (in press).
34. Bousquet, J. A., Garbay, C., Roques, B. P., and Mely, Y. (2000) *Biochemistry 39*, 7722−7735.
35. Vidal, M., Goudreau, N., Cornille, F., Cussac, D., Gincel, E., and Garbay, C. (1999) *J. Mol. Biol. 290*, 717−730.
36. Wittekind, M., Mapelli, C., Lee, V., Goldfarb, V., Friedrichs, M. S., Meyers, C. A., and Mueller, L. (1997) *J. Mol. Biol. 267*, 933−952.
37. Arold, S., O'Brien, R., Franken, P., Strub, M., Hoh, F., Dumas, C., and Ladbury, J. E. (1998) *Biochemistry 37*, 14683−14691.
38. Hansson, H., Mattsson, P. T., Allard, P., Haapaniemi, P., Vihinen, M., Smith, C. I. E., and Härd, T. (1998) *Biochemistry 37*, 2912−2924.
39. Horita, D. A., Baldisseri, D. M., Zhang, W., Altieri, A. S., Smithgall, T. E., Gmeimer, W. H., and Byrd, R. A. (1998) *J. Mol. Biol. 278*, 253−265.
40. Cordier, F., Wang, C., Grzesiek, S., and Nicholson, L. K. (2000) *J. Mol. Biol. 304*, 497−505.
41. Loh, A. P., Guo, W., Nicholson, L. K., and Oswald, R. E. (1999) *Biochemistry 38*, 12547−12557.
42. Mayo, S. L., and Baldwin, R. L. (1993) *Science 262*, 873−876.
43. Horita, D. A., Zhang, W., Smithgall, T. E., Gmeiner, W. H., and Byrd, R. A. (2000) *Protein Sci. 9*, 95−103.
44. Woodward, C., Simon, I., and Tuchsen, E. (1982) *Mol. Cell. Biochem. 48*, 135−160.
45. Hilser, V. J., Townsend, B. D., and Freire, E. (1997) *Biophys. Chem. 64*, 69−79.
46. Murphy, K. P., Xie, D., Garcia, K. C., and Amzel, L. M. (1993) *Proteins: Struct., Funct., Genet. 15*, 113−120.
47. Gomez, J., Hilser, V. J., Xie, D., and Freire, E. (1995) *Proteins: Struct., Funct., Genet. 22*, 404−412.
48. Hilser, V. J., Gomez, J., and Freire, E. (1996) *Proteins: Struct., Funct., Genet. 26*, 123−133.
49. Hilser, V. J., and Freire, E. (1997) *Proteins: Struct., Funct., Genet. 27*, 171−183.
50. Hilser, V. J., Dowdy, D., Oas, T. G., and Freire, E. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9903−9908.

BI030005J