J|A|C|S
ARTICLES
Published on Web 05/22/2002

# Residue-Specific Real-Time NMR Diffusion Experiments Define the Association States of Proteins during Folding

Alexei V. Buevich† and Jean Baum*

*Contribution from the Department of Chemistry, Rutgers University, Piscataway, New Jersey 08854*

Received December 17, 2001. Revised Manuscript Received April 2, 2002

***Abstract:*** Characterizing the association states of proteins during folding is critical for understanding the nature of protein-folding intermediates and protein-folding pathways, protein aggregation, and disease-related aggregation. To study the association states of unfolded, folded, and intermediate species during protein folding, we have introduced a novel residue-specific real-time NMR diffusion experiment. This experiment, a combination of NMR real-time folding experiments and 3D heteronuclear pulsed field gradient NMR diffusion experiments (LED-HSQC), measures hydrodynamic properties, or molecular sizes, of kinetic species directly during the folding process. Application of the residue-specific real-time NMR diffusion experiments to characterize the folding of the collagen triple helix motif shows that this experiment can be used to determine the association states of unfolded, folded, and kinetic intermediates with transient lifetimes simultaneously. The ratio of the apparent translational diffusion coefficients of the unfolded to the folded form of the triple helix is 0.59, which correlates very well with a theoretical ratio for monomer to linear trimer. The apparent diffusion coefficients of the kinetic intermediates formed during triple helix folding indicate the formation of trimer-like associates which is consistent with previously published kinetic and relaxation data. The residue-specific time dependence of apparent diffusion coefficients of monomer and trimer peaks also illustrates the ability to use diffusion data to probe the directionality of triple helix formation. NMR diffusion experiments provide a new strategy for the investigation of protein-folding mechanisms, both to understand the role of kinetic intermediates and to determine the time-dependent aggregation processes in human diseases.

## Introduction

Characterizing the association states of protein-folding species is critical for understanding protein-folding pathways,[1,2] protein aggregation,[3] and disease-related aggregation[4−6] such as prion, Alzheimer's, and connective tissue diseases. In particular, defining the association states of unfolded, intermediate, and folded states can provide valuable insight into the nature of the free energy landscape during the folding process. However, the association states of kinetic folding intermediates are often difficult to determine due to their transient lifetimes and low intensities. To study association states of proteins directly during the folding process, we introduce a novel residue-specific real-time NMR diffusion experiment.

The real-time NMR diffusion experiment is a combination of multidimensional pulsed field gradient (PFG) NMR diffusion experiments and NMR real-time folding experiments. NMR diffusion experiments have been previously used to obtain molecular size, shape, and aggregation states in small molecules, polymers,[7,8] and recently proteins.[9−20] Real-time NMR folding experiments have been developed to follow in "real time" structural or dynamic changes taking place during protein folding and to obtain a detailed kinetic picture of folding events.[21,22] Our approach is a novel modification of the recently proposed one-dimensional homonuclear real-time diffusion experiment.[20] The superior spectral resolution of two-dimen-

* To whom correspondence should be addressed. Phone: 732/445-5666. Fax: 732/445-5312. E-mail: baum@rutchem.rutgers.edu.
† Present address: Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033. E-mail: alexei.buevich@spcorp.com.

(1) Jamin, M.; Baldwin, R. L. *J. Mol. Biol.* **1998**, *276*, 491−504.
(2) Silow, M.; Tan, Y. J.; Fersht, A. R.; Oliveberg, M. *Biochemistry* **1999**, *38*, 13006−13012.
(3) Fink, A. L. *Folding Des.* **1998**, *3*, R9−R23.
(4) Lansbury, P. T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3342−3344.
(5) Perutz, M. *Trends Biochem. Sci.* **1999**, *24*, 58−63.
(6) Dobson, C. M. *Trends Biochem. Sci.* **1999**, *24*, 329−332.
(7) Stejskal, E. O.; Tanner, J. E. *J. Chem. Phys.* **1965**, *42*, 288−292.
(8) Stilbs, P. *Prog. Nucl. Magn. Reson. Spectrosc.* **1987**, *19*, 1−87.
(9) Altieri, A. S.; Hinton, D. P.; Byrd, R. A. *J. Am. Chem. Soc.* **1995**, *117*, 7566−7567.
(10) Daragan, V. A.; Ilyina, E.; Mayo, K. H. *Biopolymers* **1993**, *33*, 521−533.
(11) Dingley, A. J.; Mackay, J. P.; Chapman, B. E.; Morris, M. B.; Kuchel, P. W.; Hambly, B. D.; King, G. F. *J. Biomol. NMR* **1995**, *6*, 321−328.
(12) Lin, M.; Larive, C. K. *Anal. Biochem.* **1995**, *229*, 214−220.
(13) Jones, J. A.; Wilkins, D. K.; Smith, L. J.; Dobson, C. M. *J. Biomol. NMR* **1997**, *10*, 199−203.
(14) Pan, H.; Barany, G.; Woodward, C. *Protein Sci.* **1997**, *6*, 1985−1992.
(15) Ilyina, E.; Roongta, V.; Pan, H.; Woodward, C.; Mayo, K. H. *Biochemistry* **1997**, *36*, 3383−3388.
(16) Krishnan, V. V. *J. Magn. Reson.* **1997**, *124*, 468−473.
(17) Lin, M.; Shapiro, M. J.; Wareing, J. R. *J. Am. Chem. Soc.* **1997**, *119*, 5249−5250.
(18) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. *Biochemistry* **1999**, *38*, 16424−16431.
(19) Yao, S.; Howlett, G. J.; Norton, R. S. *J. Biomol. NMR* **2000**, *16*, 109−119.
(20) Balbach, J. *J. Am. Chem. Soc.* **2000**, *122*, 5887−5888.
(21) Liu, X.; Siegel, D. L.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1996**, *35*, 4306−4313.
(22) Dobson, C. M.; Hore, P. J. *Nat. Struct. Biol.* **1998**, *5*, 504−507.


**10.1021/ja012699u CCC: $22.00** 

sional heteronuclear NMR spectroscopy over one-dimensional proton NMR spectroscopy, in combination with real-time folding experiments, provides a new, more powerful, approach to measuring hydrodynamic properties of proteins under nonequilibrium conditions as they undergo the folding process.

We apply real-time NMR diffusion experiments to the folding of the collagen triple helix to develop a detailed picture of the association states of the species that arise in this monomer to trimer transition. The linear $(Gly\text{-}X\text{-}Y)_n$ repetitive triple-helical collagen presents a rodlike folded form with the Gly residues all buried near a central axis, and the residues in the X and Y positions largely exposed to solvent. The X and Y positions are frequently occupied by proline and hydroxyproline, respectively, which are important to triple helix stability, and Gly-Pro-Hyp is the most common and most stabilizing tripeptide sequence. Connective tissue diseases have been found to result from collagen mutations leading to defective molecular or higher order folding.[23−25] Misfolding has been implicated in cases of *osteogenesis imperfecta* ("brittle bone disease") where a single Gly in type I collagen has been mutated to another residue, breaking the $(Gly\text{-}X\text{-}Y)_n$ repeating pattern.[26]

Model collagen-like peptides with the characteristic $(Gly\text{-}X\text{-}Y)_n$ amino acid sequence of collagen have been used to characterize folding of the triple helix motif by NMR spectroscopy and circular dichroism.[27−29] The triple helical peptide studied here consists of combining the C-terminal $(GPO)_4$ region and an 18 residue sequence from residues 892 to 910 of α-1 chain of collagen type I. This peptide, denoted T1-892, is proposed to be a good model of the collagen chain as it incorporates the C-terminal (Gly-Pro-Hyp) rich environment and a $(Gly\text{-}X\text{-}Y)_n$ sequence that includes the 901 site of a Gly to Ser *osteogenesis imperfecta* mutation.[30] The folding involves a transition from the single chain monomer to the supercoiled helical trimer via a nucleation-propagation mechanism and has been shown to be slow on the order of minutes to hours.[27−29] Real-time 2D heteronuclear NMR folding experiments on the $^{15}N$-enriched model triple helical peptide T1-892 provide a method of directly monitoring the formation of trimer peaks and transient kinetic intermediate peaks in real time, and for probing site-specific kinetic orders and rate constants. These kinetic experiments result in a detailed view of the residue-specific mechanism of both the nucleation and the C- to N-terminal propagation steps of triple helix folding. However, the nature of the association states of the intermediate and final forms cannot be obtained from these experiments and is critical to defining the folding mechanisms in more detail.[31] In this paper, we applied NMR real-time diffusion experiments to the folding of the triple helix motif to show that they can be used to measure simultaneously the association states of unfolded, folded, and kinetic intermediates with transient lifetimes. In addition, the residue-specific values of the apparent diffusion coefficients can be used to confirm the C- to N-terminal directionality of triple helix folding.

## Materials and Methods

**Sample Preparation.** T1-892 peptide containing different $^{15}N$-labeled positions was purchased from Synpep Corp. (Dublin, CA). The labeled positions Ala6, Gly10, Gly13, and Gly25 are shown in boldface below.

```
                1    5    10   15   20   25   30
T1-892:  GPAGPAGPV G PA GA R GPAG P OGPO G POGP O GV
```

The 9.1 mM solution of T1-892 peptide used in the NMR experiments was prepared in 10% $D_2O$/90% $H_2O$, and the pH (=2.5) was adjusted by titration with 5 mM HCl. The HPLC purity of the sample was 95−99%, and its identity was confirmed by mass spectrometry.

**NMR Spectroscopy.** NMR experiments were performed on a Varian INOVA 600 MHz spectrometer operating at 599.88 MHz for $^1H$ observation, and equipped with a triple resonance probe and pulsed field gradients. Heteronuclear single quantum coherence spectroscopy (HSQC) spectra were recorded at 10 °C. All the pulse sequences employed sensitivity enhanced pulsed field gradient techniques.[32,33] NMR data sets were processed on a Silicon Graphics workstation using the Felix97 program (Molecular Simulation Inc.).

**Real-Time Folding Experiments.** The NMR folding experiments were performed as previously described.[31] The T1-892 peptide was denatured outside of the NMR spectrometer by heating for 15 min at 55 °C and then immersed in an ice/water bath at 0 °C for 20 s to quickly reduce the temperature to 5−10 °C (melting point of T1-892 triple helix is 26 °C). After that the sample was transferred into a precooled 10 °C NMR probe in the spectrometer. 2D $^1H$-$^{15}N$ HSQC spectra were acquired every 3 min for the first 5−6 h and every hour for the remaining 32−40 h. The dead time of the experiment measured from the time of folding initiation to the start of the NMR experiment was exactly 60 s. Kinetics of folding were monitored by measuring cross-peak volumes of resolved monomer, trimer, and intermediate peaks for the labeled amino acids as a function of time. The folding experiment was repeated three times, and average intensities of each peak were normalized as described elsewhere.[31] Kinetic analysis of the time-dependent monomer, trimer, and intermediate intensities was performed using the KinFit program (D. V. Dearden, Brigham Young University, Provo, UT) as previously described.[31]

**Real-Time NMR Diffusion Experiments.** The experiment presented in this paper is the first residue-specific diffusion experiment applied to biomolecules, and represents a 3D diffusion encoded HSQC experiment (termed LED-HSQC) on isotopically labeled proteins. The diffusion-encoding part of the experiment was accomplished by a longitudinal encoding-decoding approach (LED)[34] (Figure 1a) which has the advantage of minimal loss of magnetization in the case of large biological macromolecules.[9] A bipolar pulse scheme,[35] found to be superior in sensitivity to the original monopolar gradient sequence, was also used to refocus chemical shift and heteronuclear coupling evolution of amide protons created during encoding-decoding time delays (Figure 1a). The sensitivity enhanced HSQC pulse sequence with gradient selection and minimal water perturbation[32,33] was used to ensure the highest sensitivity.

For diffusion experiments at high temperature (≥30 °C), the convection compensation pulse scheme (CC-LED) was implemented[36]

(23) Kuivaniemi, H.; Tromp, G.; Prockop, D. *Hum. Mutat.* **1997**, *9*, 300−315.
(24) Prockop, D. J.; Kivirikko, K. I. *Annu. Rev. Biochem.* **1995**, *64*, 403−434.
(25) Byers, P. H. In *Osteogenesis Imperfecta*; Royce, P. M., Steinmann, B., Eds.; Wiley Liss: New York, 1993; pp 317−350, 351−407.
(26) Bächinger, H. P.; Morris, N. P.; Davis, J. M. *Am. J. Med. Genet.* **1993**, *45*, 152−162.
(27) Baum, J.; Brodsky, B. In *Folding and Misfolding of the Collagen Triple Helix: Biophysical Studies of Model Peptides*; Pain, R. H., Ed.; Oxford University Press: New York, 1999; pp 330−347.
(28) Baum, J.; Brodsky, B. *Curr. Opin. Struct. Biol.* **1999**, *9*, 122−128.
(29) Buevich, A.; Baum, J. *Philos. Trans. R. Soc. London, Ser. B* **2001**, *356*, 159−168.
(30) Yang, W.; Battinenni, M.; Brodsky, B. *Biochemistry* **1997**, *36*, 6930−6935.
(31) Buevich, A. V.; Dai, Q.-H.; Liu, X.; Brodsky, B.; Baum, J. *Biochemistry* **2000**, *39*, 4299−4308.
(32) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663−10665.
(33) Zhang, O.; Kay, L. E.; Olivier, J. P.; Forman-Kay, J. D. *J. Biomol. NMR* **1994**, *4*, 845−858.
(34) Gibbs, S. J.; Johnson, C. S., Jr. *J. Magn. Reson.* **1991**, *93*, 395−402.
(35) Morris, K. F.; Johnson, C. S., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 3139−3141.

***Table 1.*** Residue-Specific Diffusion Coefficients of T1-892 Peptide Measured during Refolding Experiment at 10 °C[a]

| | $D$ ($10^{-7}$, cm$^2$/s) | | | |
|---|---|---|---|---|
| peak type | Ala6 | Gly10 | Gly13 | Gly25 |
| monomer | 8.09 ± 0.09 | 8.95 ± 0.09 | 9.27 ± 0.09 | 10.09 ± 0.10 |
| trimer | 5.99 ± 0.11 | 6.16 ± 0.13 | 6.20 ± 0.16 | 6.72 ± 0.10 |
| intermediate-1 | 6.35 ± 0.30 (20%) | | 5.91 ± 0.42 (07%) | |
| intermediate-2 | 5.74 ± 0.35 (05%) | | 6.14 ± 0.18 (33%) | |
| intermediate-3 | 6.18 ± 0.17 (21%) | | 5.87 ± 0.39 (16%) | |
| intermediate-4 | 5.89 ± 0.19 (15%) | | 6.17 ± 0.31 (21%) | |
| intermediate-5 | 5.91 ± 0.03 (16%) | | 6.06 ± 0.83 (05%) | |
| intermediate-6 | 5.46 ± 0.62 (07%) | | 6.46 ± 0.88 (07%) | |
| intermediate-7 | 6.02 ± 0.30 (14%) | | 5.86 ± 0.51 (10%) | |

[a] Error analysis is from the covariance matrix analysis. The numbers in parentheses for the intermediates represent the relative intensities of the peaks. Intermediate peaks of G10 and Gly25 were indistinguishable from trimer peaks.

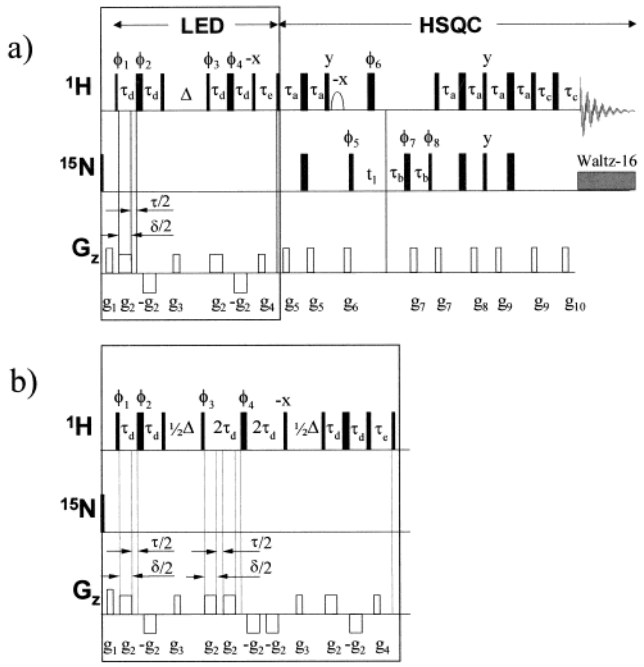


***Figure 1.*** (a) LED-HSQC pulse sequence with the diffusion encoded LED part shown in box. (b) Convection compensated diffusion encoding part of the CC-LED-HSQC pulse sequence.

to improve sensitivity losses that may result from convection within the sample (Figure 1b). In Figure 1, solid thin rectangles are hard 90° pulses, solid thick rectangles are hard 180° pulses, open rectangles are gradient pulses, the hatched rectangle is a WALTZ-16 decoupling pulse sequence, and the open semi-sphere is the low power 90° selective pulse on the water resonance. The pulse phase cycles were $\phi_1 = x, y$; $\phi_2 = y, -x$; $\phi_3 = y, -x$; $\phi_4 = -y, x$; $\phi_5 = x$; $\phi_6 = x, -x$; $\phi_7 = 2(x), 2(y), 2(-x), 2(-y)$; $\phi_8 = x$; $\phi_{rec} = 2(x), 2(-x)$. Pulses for which the phases are not indicated were applied along the $x$-axis. Delays $\Delta$, $\delta$, $\tau$, $\tau_a$, $\tau_b$, $\tau_c$, $\tau_d$, $\tau_e$ were 100, 4.5, 0.3, 2.5, 0.5, 2.4, 0.65 ms, respectively. For each increment of $t_1$, the $\phi_8$ phase was incremented by 180°, and the phase of the gradient pulse $g_9$ was inverted. Postacquisition processing of the N- and P-type data was accomplished with Felix97 software. The gradient durations and strengths were $g_1$ = (1 ms, 18 G/cm), $g_2$ = (2.25 ms, variable), $g_3$ = (0.5 ms, 22.5 G/cm), $g_4$ = (0.5 ms, 33.75), $g_5$ = (0.5 ms, 18 G/cm), $g_6$ = (0.5 ms, 33.75 G/cm), $g_7$ = (1.25 ms, 67.5 G/cm), $g_8$ = (0.15 ms, 33.75 G/cm), $g_9$ = (0.125 ms, 67.3 G/cm). The CC-LED-HSQC diffusion experiment at high temperature (55 °C) was performed with the following $g_2$-gradient strengths (G/cm): 2.12, 6.70, 9.45, 11.83, 13.39, 14.99, 16.43, 17.74, 18.96, 20.12, 21.20. The LED-HSQC diffusion experiment at low temperature (10 °C) was performed with the following $g_2$-gradient strengths (G/cm): 2.12, 14.98, 21.18, 25.94, 29.95, 33.47, 36.64, 39.61, 42.36. The spectral widths were 6000 Hz ($^1$H) and 1500 Hz ($^{15}$N). 2 K acquisition points, 64 complex $t_1$-points, 64 scans per each $t_1$-point, and 4 s recycling delay were used in LED-HSQC and CC-LED-HSQC experiments. The gradient strength was calibrated from a diffusion experiment on water[9] using the water apparent diffusion coefficient $D = 1.2 \times 10^{-5}$ cm$^2$/s at 10 °C. Real-time NMR diffusion experiments are a combination of the LED-HSQC experiment and the real-time folding experiment. The LED-HSQC experiment begins 30 min after initiation of protein folding and lasts 10 h. The combined kinetic and LED-HSQC experiments are repeated nine times with different gradient strengths to measure the apparent diffusion coefficients of the folding species.

The signal attenuation in the presence of pulsed field gradients is given by

$$I = I(0) \exp[-D\gamma^2G^2\delta^2(\Delta + 2\delta/3 + 3\tau/4)] \qquad (1)$$

where $D$ is the translational diffusion coefficient, $\gamma$ is $^1$H gyromagnetic ratio, $G$ is the gradient strength, and delays $\Delta$ (100 ms), $\delta$ (4.5 ms), and $\tau$ (0.3 ms) are explained in Figure 1.

One of the potential problems associated with measurements of diffusion of amide protons through heteronuclear NMR experiments is related to the exchange of the amide protons with water during the diffusion encoding delay $\Delta$.[37] This process can bias the apparent diffusion coefficients toward larger values particularly for the monomer form. However, considering the low temperature (10° C) and low pH (2.5) at which current experiments were done, this exchange process is likely to be slow on the order of minutes and therefore can be neglected. A control experiment with twice as short a delay $\Delta$ (50 ms) revealed essentially the same apparent diffusion coefficients as compared to those measured with 100 ms delay indicating that the exchange with water is slow and did not affect the diffusion experiment.

**Error Analysis.** The combined folding-diffusion experiment for T1-892 was repeated nine times with different $g_2$-gradient strengths from 2.12 to 42.36 G/cm. The apparent diffusion coefficients for T1-892 presented in Table 1 were estimated by fitting nine experimental volume integrals to eq 1 by a least-squares minimization algorithm. The standard deviations of the diffusion coefficients were obtained by two different methods. One was from the covariance matrix for the least-squares optimization, and the second was obtained by Monte Carlo simulations. In the Monte Carlo simulations, a random Gaussian noise (Numerical Technologies Random Generator for Excel, (NTRand)) with zero mean value and standard deviation equal to the experimental uncertainty ($1.6 \times 10^5$ au) was used. The experimental uncertainty was estimated as the standard deviation of the volume integration of the experimental data in the regions free of signals. For each estimate of the standard deviation of the apparent diffusion coefficients, 100 sets of nine experimental data points with added noise were generated and then subsequently fit to eq 1 by a least-squares minimization approach. The

(36) Jerschow, A.; Muller, N. *J. Magn. Reson.* **1997**, *125*, 372−375.

(37) Tillett, M. L.; Lian, L. Y.; Norwood, T. J. *J. Magn. Reson.* **1998**, *133*, 379−384.

***Table 2.*** Covariance Matrix and Monte Carlo Analysis of the Standard Deviations of the Diffusion Coefficients of T1-892 Peptide (Kinetic Conditions, 10 °C)

| | $\sigma$ ($10^{-7}$, cm$^2$/s) | |
|---|---|---|
| residue, peak type | covariance matrix analysis | Monte Carlo simulations |
| Ala6-M | 0.09498 | 0.02074 |
| Gly10-M | 0.08547 | 0.03984 |
| Gly13-M | 0.09166 | 0.04647 |
| Gly25-M | 0.09581 | 0.04788 |
| Ala6-T | 0.11356 | 0.07171 |
| Gly10-T | 0.12831 | 0.09417 |
| Gly13-T | 0.16324 | 0.14529 |
| Gly25-T | 0.10002 | 0.01826 |
| Ala6-I-1 | 0.29551 | 0.18827 |
| Ala6-I-2 | 0.35327 | 0.68600 |
| Ala6-I-3 | 0.16641 | 0.18157 |
| Ala6-I-4 | 0.17830 | 0.24735 |
| Ala6-I-5 | 0.03622 | 0.22937 |
| Ala6-I-6 | 0.62132 | 0.50577 |
| Ala6-I-7 | 0.29938 | 0.29278 |
| Gly13-I-1 | 0.42816 | 0.87407 |
| Gly13-I-2 | 0.18237 | 0.19258 |
| Gly13-I-3 | 0.38700 | 0.38540 |
| Gly13-I-4 | 0.30624 | 0.30483 |
| Gly13-I-5 | 0.82910 | 1.46810 |
| Gly13-I-6 | 0.88294 | 0.89252 |
| Gly13-I-7 | 0.51252 | 0.58587 |

result of these calculations was a set of 100 diffusion coefficients, with a Gaussian distribution. The width of this distribution was considered to be an estimate of the Monte Carlo standard deviation of the experimental apparent diffusion coefficient. Standard deviations obtained by the covariance matrix analysis and the Monte Carlo simulations of the apparent diffusion coefficients acquired under kinetic conditions are presented in Table 2. The covariance matrix analysis gave larger estimates of the standard deviations for monomer and trimer species as compared with those predicted by the Monte Carlo simulations. The average values of the monomer and trimer uncertainties by the covariance matrix approach are $0.109 \times 10^{-7}$ cm$^2$/s, whereas they are only $0.049 \times 10^{-7}$ cm$^2$/s by the Monte Carlo simulations (Table 2). The standard deviations of the diffusion coefficients of intermediate species are similar by both methods. The average values are $0.391 \times 10^{-7}$ cm$^2$/s and $0.502 \times 10^{-7}$ cm$^2$/s by the covariance matrix analysis and the Monte Carlo simulations, respectively. Because the covariance matrix analysis gave, in general, more conservative estimations of the diffusion uncertainties, we have used this approach in Table 1.

## Results and Discussion

**Real-Time Diffusion Experiments.** Real-time LED-HSQC diffusion experiments were performed on peptide T1-892 to obtain the association states of the kinetic species in solution during the refolding experiment from monomer to trimer. For NMR studies, peptides were made with $^{15}$N-enriched residues at specific positions, including positions Ala6, Gly10, Gly13, and Gly25. These positions were chosen to represent folding at the C- and N-terminal ends as well as the central region of the peptide. The real-time NMR folding experiments of peptide T1-892 described earlier monitor folding from the monomer to trimer species. For the Ala6 position of the triple helical peptide T1-892, NMR spectra as a function of time showed the disappearance of monomer peaks, the appearance of trimer peaks, and the appearance, during folding, of a number of transient kinetic intermediate species (Figure 2).[31] The other labeled positions in the peptide have similar kinetic profiles, although it is not possible to detect kinetic intermediates for

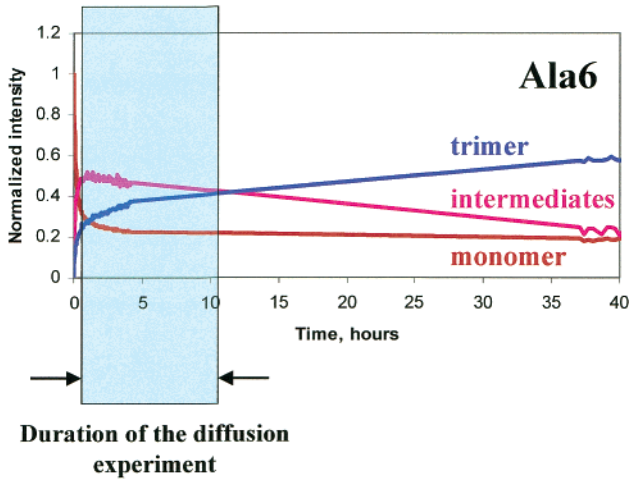


***Figure 2.*** Time dependence of the monomer, trimer, and intermediate intensities for Ala6 during folding of T1-892 at 10 °C using real-time NMR experiments.[31] Boxed area indicates the length of the LED-HSQC diffusion experiment for one gradient strength starting 30 min after folding initiation. The combined kinetic and LED-HSQC experiments were repeated nine times with different gradient strengths to measure the apparent diffusion coefficients of the folding species.

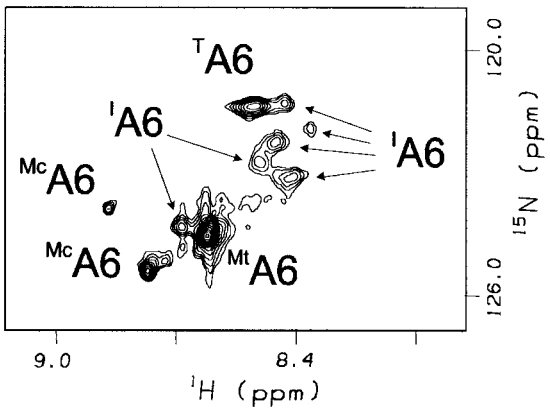


***Figure 3.*** Portion of the real-time $^{15}$N-$^1$H LED-HSQC spectra of T1-892 with a $g_2$-gradient strength of 2.12 G/cm acquired during a 10 h period starting 30 min after folding initiation. The spectrum shows the Ala6 residue obtained at 10 °C on a Varian INOVA 600. Labels Mc, Mt, T, I stand for monomer resonances with neighboring *cis*-Pro, monomer resonances with neighboring *trans*-Pro, trimer, and intermediate peaks, respectively.[31]

Gly10 and Gly25. The absence of intermediates for these two residues does not reflect a difference in mechanism between the residues but rather just reflects the fact that the chemical shifts of the intermediate peaks for Gly 10 and Gly 25 overlap with the trimer peaks. The LED-HSQC experiment was started 30 min after initiation of folding of T1-892 when the intermediate species had the highest peak intensity (Figure 2).[31] The LED-HSQC spectrum at a single gradient strength was obtained over a period of 10 h during which the intensity of intermediate peaks remained relatively high between 50 and 40%, monomer intensities quickly reached an equilibrium value, and the trimer peak grew from 25 to 40% (Figure 2). Figure 3 represents a portion of the real-time LED-HSQC diffusion experiment and shows the monomer, intermediate, and trimer peaks for Ala6. The NMR assignments for individual residues of T1-892 have been described earlier.[31] Ala6 has one major monomer peak and one well-defined minor peak that reflects trans and cis isomers respectively at the Gly4−Pro5 or Gly7−Pro8 bonds. Ala6 also shows six intermediate peaks and a single trimer peak

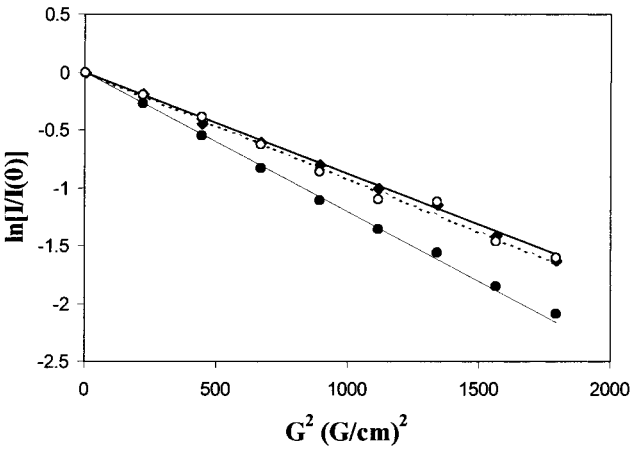


***Figure 4.*** Logarithmic (normalized) intensities of Ala6 trimer (●), monomer (▲), and intermediate (○) versus the square of the $g_2$-gradient strength. The experimental uncertainty was estimated as the standard deviation of the volume integration of the experimental data in the regions free of signals. The error bars are within the range of the symbols depicted in the figure. Solid line best fits of the experimental peak volumes with eq 1 and the corresponding apparent diffusion coefficients are listed in Table 1.

during folding. Repeating the LED-HSQC diffusion experiment nine times at different gradient strengths enables the measurement of the apparent diffusion coefficients of intermediate, trimer, and monomer peaks.

The length of the acquisition time for a single real-time diffusion experiment was dictated by the relatively low intensities of the intermediate species. Given the very slow folding kinetics of peptide T1-892, the duration of the diffusion experiment was not a limitation in the current study. To monitor faster kinetic processes, the experimental times can be reduced substantially to minutes or even seconds given acquisition of one-dimensional NMR spectra and acceptable one-dimensional resolution.[20] Another approach to shortening the experimental time is to reduce the recycle delay between successive pulses. Control experiments show that even with a recycle delay as short as 0.5 s (a steady-state regime), it is still possible to measure with reasonable accuracy the apparent diffusion coefficients of the most populated peaks.

The requirement of high protein concentration for diffusion studies by NMR in the present study is not strict, but was dictated mainly by the low population of the transient intermediate species. In general, NMR diffusion experiments require the same concentration as other bio-NMR experiments, which can be as low as sub-millimolar to micromolar concentrations owing to progress in NMR probe technology. Moreover, the high concentration of the diffusion by PFG NMR experiments does not pose a problem as opposed to other techniques, such as SAXS (small-angle X-ray scattering), where high concentrations may result in artificial scattering peaks. This then requires concentration-dependent studies and extrapolation of scattering curves to infinite dilution.[38] On the contrary, NMR diffusion experiments are lacking these systematic errors caused by high protein concentrations, and therefore changes in apparent diffusion coefficients as a function of concentration would indicate aggregation or association processes.

**Residue Specificity of the Diffusion Coefficients.** Figure 4 represents the plot of the intensities for the monomer, trimer, and intermediate peaks obtained for Ala6 as a function of gradient strength during the real-time diffusion experiment. Circles and triangles in Figure 4 show the experimental values, and the best fits to the experimental values (solid lines) were used to estimate the apparent diffusion coefficients. It is clear that the apparent diffusion coefficients for the trimer and intermediate peaks are quite similar, whereas that of the monomer differs from the other two. In Table 1, the apparent diffusion coefficients for T1-892 measured at 10° C under nonequilibrium conditions are presented for all labeled residues, Ala6, Gly10, Gly13, and Gly25. All labeled residues had multiple monomer peaks, and the most intense *trans*-proline conformation was chosen for the analysis. For residues with multiple trimer peaks (Gly10, Gly13), each peak was analyzed, characterized by an apparent diffusion coefficient, and the average value was presented. All intermediate peaks observed in the spectra of Ala6 and Gly13 were analyzed separately. For all labeled residues, the values for the intermediate and trimer peaks are quite similar, whereas monomer diffusion is clearly much faster. The apparent diffusion coefficients (Table 1) are residue specific with monomer values gradually increasing from $8.09 \times 10^{-7}$ cm$^2$/s for Ala6 at the N-terminal end to $10.09 \times 10^{-7}$ cm$^2$/s for Gly25 at the C-terminal end. Trimer apparent diffusion coefficients are smaller overall and increase only slightly from $5.99 \times 10^{-7}$ at the N-terminal end to $6.72 \times 10^{-7}$ cm$^2$/s at the C-terminal end. The residue-specific values of the diffusion coefficients reflect the nonequilibrium kinetic conditions under which the diffusion coefficients were measured. Measurement of residue-specific apparent diffusion coefficients under equilibrium conditions shows that they are substantially more similar to one another (Table 3).

***Table 3.*** Residue-Specific Diffusion Coefficients of T1-892 Peptide Measured after 36 h at 10 °C

| | $D$ ($10^{-7}$, cm$^2$/s) | | | |
|---|---|---|---|---|
| peak type | Ala6 | Gly10 | Gly13 | Gly25 |
| monomer | $9.22 \pm 0.15$ | $9.47 \pm 0.12$ | $9.69 \pm 0.04$ | $10.32 \pm 0.10$ |
| trimer | $6.15 \pm 0.09$ | $6.21 \pm 0.13$ | $6.11 \pm 0.15$ | $6.64 \pm 0.13$ |

The residue-specific diffusion coefficients measured under nonequilibrium folding conditions reflect the population weighted average over different kinetic species over the 10 h experiment. Therefore, interpretation of the residue-specific diffusion coefficients requires knowledge of the kinetics of folding which can be obtained from NMR real-time folding experiments. NMR kinetics experiments showed that the folding starts with C-terminal nucleation of the three chains followed by C- to N-terminal zipper-like propagation of the triple helix (Figure 5).[27,28] Because folding begins at the C-terminal end, the Ala6 monomer resonance reflects a mixture of pure monomer form and partially folded trimer forms, whereas the Gly25 monomer peak reflects primarily the pure monomer population (Figure 5). Therefore, the C-terminal Gly25 monomer peak has the fastest diffusion consistent with the interpretation that this value represents essentially only monomer population.

The apparent diffusion coefficients for the trimer form are on average smaller than the monomer apparent diffusion coefficients as expected for the higher molecular weight trimer molecule. In addition, the residue specificity for the trimer reflects a population average over different trimer shapes that are formed during the folding process (Figure 5). Hydrodynamic

(38) Doniach, S. *Chem. Rev.* **2001**, *101*, 1763−1778.

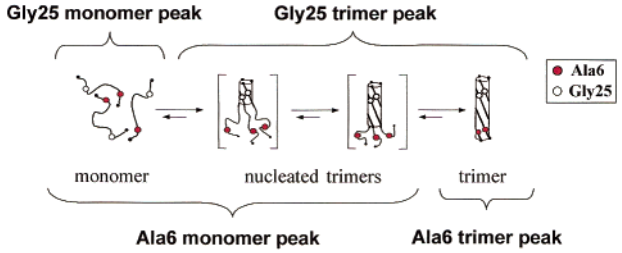


***Figure 5.*** Schematic representation of the folding pathway of T1-892 peptide. ● indicate the position of Ala6, and ○ indicate the position of Gly25. Population averaging of the apparent monomer and trimer peaks for Ala6 and Gly25 is indicated. Thus, for example, the Ala6 monomer resonance represents a mixture of pure monomer form and partially folded trimer forms, whereas the Gly25 monomer peak reflects primarily the pure monomer population.

calculations suggest that for molecules of equal weight the smallest diffusion coefficient is had by the molecule with the most anisotropic shape.[39] The N-terminal Ala6 trimer has the slowest diffusion consistent with a folding mechanism in which Ala6 is involved in the final stage of folding. Therefore the apparent diffusion coefficient for Ala6 represents the fully folded most anisotropic rodlike trimer (Figure 5).

The apparent diffusion coefficients that are obtained under equilibrium conditions at 10 °C, 36 h after initiation of folding (Table 3), are different from those obtained during the folding kinetics at the initial stages of folding (Table 1). In general, the apparent diffusion of the monomer peaks at equilibrium is faster than that of those measured under kinetic conditions, whereas the diffusion of the trimer peaks is similar under both equilibrium and kinetic conditions. The difference in diffusion properties between equilibrium and kinetic conditions is consistent with the folding mechanism of T1-892. As opposed to kinetic conditions, at equilibrium, the trimer population dominates the conformational ensemble, whereas the monomer and, in particular, the intermediates are present in minor quantities. Therefore, under equilibrium conditions, the monomer peak at the N-terminal end is not significantly populated with trimer-like intermediate species, and therefore its apparent diffusion more closely resembles the diffusion of the monomer at the C-terminal end (Figure 5). More quantitatively, the progress of folding can be followed by the change of the diffusion dispersion, which we define as the difference between apparent diffusion coefficients for the same association state at two termini (Ala6 and Gly25). Thus, the diffusion dispersion of monomers measured under kinetic conditions is $2.00 \times 10^{-7}$ versus $1.10 \times 10^{-7}$ cm$^2$/s under equilibrium conditions, and the trimer diffusion dispersion correspondingly changes from $0.73 \times 10^{-7}$ to $0.49 \times 10^{-7}$ cm$^2$/s. As expected, the diffusion dispersion for both monomer and trimer becomes smaller as the system reaches equilibrium. However, it is noteworthy that the diffusion dispersion at equilibrium does not vanish and remains experimentally significant. This observation is consistent with our previous studies, which show that even at equilibrium (10° C, after 40 h) the two terminal ends have different percentages of monomer and trimer.[31] This difference in percentage of the trimer conformation at the two termini in T1-892 at equilibrium may arise from the sequence-dependent difference of triple helix stability,[40] and can explain the difference in the apparent diffusion coefficients at two terminal ends.

**Ratio of the Diffusion Coefficients of Monomer and Trimer.** The apparent diffusion coefficients that are obtained from the real-time NMR diffusion experiments can be used to evaluate the association states of the unfolded form, folded form, and kinetic intermediates that are formed during the refolding experiment. According to the Stokes−Einstein equation, the apparent diffusion coefficients are sensitive to the conditions of the experiment:

$$D = (k_B T)/(6\pi\eta r) \qquad (2)$$

where $D$ is the diffusion coefficient, $k_B$ is the Boltzman constant, $T$ is the temperature in Kelvin, $\eta$ is the viscosity of the solution, and $r$ is the effective hydrodynamic (Stokes) radius. Therefore, it has been suggested that the ratio of diffusion coefficients measured under identical conditions is a good estimate of the degree of association. The hydrodynamic calculations show that the ratio of the diffusion constants of a n-mer, $D$(n), to that of a monomer, $D$(m), can be expressed as follows:[39]

$$R = D(\mathrm{n})/D(\mathrm{m}) = F_n n^{-1/3} \qquad (3)$$

where $F_n$ is a geometric factor. Because of differences in hydrodynamic calculations, slightly different estimates for $R$ ratios of trimers are found in the literature. For example, for linear trimerization, the predicted $R$ values are 0.598[39] and 0.586,[41] and for equilateral trimerization, $R$ ratios are 0.662[39] and 0.621.[41] The $R$ ratio for T1-892 calculated from the $D$ values that represent the highest population of trimer (Ala6) to the $D$ value that represents the highest population of monomer (Gly25) is $0.593 \pm 0.016$, which is in very good agreement with hydrodynamic calculations found for the most anisotropic linear trimer. Interestingly, this ratio remains essentially the same under equilibrium conditions (0.596), indicating that the selection of the representative monomer and trimer diffusion coefficients under nonequilibrium conditions based on the C- to N-terminal propagation mechanism of folding was correct.

**Diffusion Coefficients of Folding Intermediates.** Comparison of the apparent diffusion coefficients of the folding intermediates (Table 1) to the apparent diffusion coefficients of the monomer and trimer peaks indicates that they are very different from the monomer values and quite similar to the trimer apparent diffusion coefficient values. The diffusion experiments have shown that the association state of the intermediate species is a trimer. The precision of the apparent diffusion coefficients found for the intermediate peaks is somewhat lower than those of the trimer and monomer peaks, and varies among intermediate peaks. This difference in precision arises from the low overall intensity of the intermediate peaks and from differences in intensities within the family of intermediate peaks (from 5 to 33%). The overall intensity of the intermediate peaks for Ala6 is twice as large as that of Gly13;[31] therefore, the Ala6 intermediate peaks, in general, have better accuracy. Despite the lower accuracy of the intermediate peaks, the data conclusively indicate that the most dominant kinetic intermediates are

(39) Teller, D. C.; Swanson, E.; De Haen, C. *Methods Enzymol.* **1979**, *61*, 103−124.
(40) Persikov, A.; Ramshaw, J. A.; Brodsky, B. *Biopolymers* **2000**, *55*, 436−450.
(41) Garsia de la Torre, J.; Bloomfield, N. A. *Rev. Biophys.* **1981**, *14*, 81−139.

neither monomer nor dimer species but may represent a heterogeneous mixture of trimers.

The nature of the trimer-like intermediate is complex and heterogeneous. Possibilities for the intermediate conformation may include mis-staggering, altered supercoiling, or partial incorporation of the *cis*-proline conformations into a triple helix. Peptide T1-892 contains 13 prolines and hydroxyprolines that undergo cis−trans isomerization in the monomer form, and are required to be all trans to form the fully folded triple helix. Although the population of the *cis*-proline conformation at each X-Pro site of the monomer chain is only approximately 15%,[42,43] the presence of a large number of prolines and hydroxyprolines in the monomer chain leads to 60% of the monomer population containing at least one *cis*-proline. Folding of the triple helix requires interconversion of all *cis*-prolines to trans conformations, and misfolded trimer-like intermediate species may arise from incorporation of the *cis*-proline. The association state of the folding intermediates determined from the diffusion experiment correlates well with previous results obtained by heteronuclear NMR relaxation experiments. The $^{15}$N-$^{1}$H NOE relaxation parameter, being sensitive to rates of the overall correlation times, and therefore to the size of the molecular aggregates, showed that the relaxation properties of the intermediates closely resembled those of the trimer, but not the monomer.[31] The diffusion measurements confirm this similarity between intermediates and trimers and also compliment the relaxation data by ruling out the possibility that the observable intermediates in the real-time NMR folding experiments could be a dimer.

In summary, residue-specific real-time NMR diffusion experiments allow direct monitoring of association states of all NMR detectable folding species in solution under refolding conditions. The time resolution of real-time diffusion experiments is slower than that of real-time kinetic experiments and therefore cannot be used as a substitute for direct kinetic measurements. However, residue-specific apparent diffusion coefficients obtained from real-time NMR diffusion experiments reflect the stepwise folding process and the mixture of states that exist during the refolding experiment and serve as a complement to the kinetic mechanisms deduced from real-time NMR approaches. Kinetic real-time NMR diffusion measurements can also compliment existing equilibrium techniques, such as light scattering, osmometry, X-ray diffraction, ultracentrifugation, and one-dimensional PFG NMR diffusion experiments. Real-time NMR diffusion experiments provide a new strategy to study protein-folding mechanisms, to understand the role of kinetic intermediates, to determine the slow time-dependent aggregation processes in human neurodegenerative diseases, and to study defective collagen folding which has been implicated in various connective tissue diseases.

**Acknowledgment.** This work was supported by NIH Grant GM45302. We thank B. Brodsky for helpful discussions.

JA012699U

(42) Grathwohl, C.; Wüthrich, K. *Biopolymers* **1976**, *15*, 2025−2041.
(43) Grathwohl, C.; Wüthrich, K. *Biopolymers* **1981**, *20*, 2623−2633.