Backbone Conformational Constraints in a Microcrystalline U-¹⁵N Labeled Protein by 3D Dipolar-Shift Solid-State NMR Spectroscopy

W. Trent Franks, Benjamin J. Wylie, Sara A. Stellfox and Chad M. Rienstra*

Experimental

NMR Spectroscopy. The 3D experiment was performed as described,¹ modified only to ensure so that the total radio frequency duty cycle was constant throughout the experiment, on an 600 MHz Infinity Plus three-channel spectrometer (Varian, Inc., Palo Alto, California) with 3.2 mm ¹H-¹³C-¹⁵N T3 probe. The U-¹⁵N sample (~12 mg) was prepared according to previously published procedures.² The spinning rate was 8.333 kHz and temperature of the variable temperature gas 0 °C (with 100 scfh flow). Each 2D plane consisted of 64 scans per row, 192 real points by 240 μs dwell in t₁, a proton-driven spin diffusion mixing time of 2.0 s, 4096 complex points by 15 μs dwell in t₂, and 1.5 s pulse delay, resulting in a total experiment time of 12 hours per 2D plane. TPPM decoupling (~70 kHz) was employed during in the evolution and acquisition periods.³ The third, T-MREV dimension consisted of 17 points with a 60 μs (one-half rotor period) increment. The T-MREV symmetry was set to 4, resulting in a ¹H π/2 pulse width of 2.5 μs.⁴ CW decoupling (100 kHz) was utilized for the duration of the ¹⁵N echo period.

The spectra were processed with nmrPipe⁵ as follows: (1) two points back prediction; (2) –12.5 Hz Lorentzian to +25 Hz Gaussian apodization with a 0.0234 (3/128) Gaussian center shift; (4) zero filling to 16384 points; (5) t₁ Fourier transformation and phasing; (6) –12.5 Hz Lorentzian to +25 Hz Gaussian apodization; (7) zero filling to 2048 points; (8) Fourier transformation and phasing. The t₃ time domain was extracted and analyzed further as described below. The fully assigned 2D plane corresponding to t₃=0 is shown in Figure S1.

Crystal Growth and X-Ray Structure Determination. Single crystals of GB1 were grown by vapor diffusion with a reservoir solution of 150 mM NaCl, 50% MPD, 20% IPA in 25 mM sodium acetate (pH of 3.8), and a protein solution of 10 mg/ml GB1 in 50 mM sodium phosphate (pH of 5.5). Data was collected at the Brookhaven National Laboratory beam line X12C (0.979 Å wavelength), with the following parameters derived from the analysis. The refined structure will be deposited to the Protein Data Bank.

Table S1. Crystallographic Data

X-ray source	BNL-X12C
Wavelength	0.9791 Å
Space group	P2 ₁ 2 ₁ 2 ₁
a	25.197 Å
b	36.384 Å
c	50.274 Å
γ	90°
Total reflections	106,995

Data Analysis

Experimental SSNMR spectra (Figure S1) were analyzed using Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco) and quantitative cross-peak intensities extracted with the nmrPipe autofit.com procedure provided with nmrPipe. Simulated T-MREV dipolar spectra were generated with Fortran programs as previously described.^{1,6} A library of simulated spectra was produced with $\theta=0$, with a range of differential relaxation and scaling factors (0.250 to 8.0 ms⁻¹ and 0.385 to 0.585 respectively). Two protons, the H^N and H α , were modeled in each of the two T-MREV dephasing periods; the two H α protons in glycine were modeled as a single proton. Fits of the 13 resolved diagonal peaks resulted in an average values of 3.25 ms⁻¹ for the relaxation rate and 0.455 for the dipolar scaling factor.

Fits of the cross peaks were subsequently performed by varying only the pseudo-bond angle θ , with the globally optimal parameters for scaling factor and relaxation rate held fixed. A library of theta values was constructed from 0° to 90° in 1° increments. The best fit value of θ was determined, and the deviation between the best fit and experimental data used to normalize the noise factor for 10,000 Monte Carlo simulations. Six representative Monte Carlo distributions are shown as Figure S2. The results were assigned to a bin, grouped around local maxima in the histograms. Standard deviations were calculated for results within each bin.

In some cases, simulations with different pseudo-bond angles (e.g., ~28° and ~80°, as in Y3-K4) yielded satisfactory agreement with the experimentally observed line shape. This occurred in 14 of the N[i]-N[i+1] spectra, and 6 of the HN[i]-HN[i+2] spectra. This result is a fundamental consequence of spectral degeneracy, which does not necessarily imply dynamic averaging. We have previously demonstrated that the majority of the GB1 backbone amide sites exhibit minimal dynamic averaging.² To determine which solution is statistically favored, the Monte Carlo simulation results (histograms) were manually inspected and grouped into bins, with the borders between bins identified as the observed minima in the histograms. Typically, one solution is highly favored in this analysis of the population statistics; remaining ambiguities can be resolved, and the degeneracies broken, by utilizing other constraints such as TALOS. For example, for Y3-K4 the TALOS result implies a pseudo-bond angle near 20°, consistent with the solution of 28.4±3.2°, but far away from the degenerate HN[i]-HN[i+1] solution at 80.4±5.1°.

In protein structure calculations, rather than presuming a single solution based on limited data, both values should be incorporated as potential pseudo-energy terms in the simulated annealing algorithm. The use of cross-correlated relaxation constraints has previously been used in this manner to refine the structure of the protein rhodniin by restrained molecular dynamics simulations,^{7,8} but to our knowledge such information alone has not been demonstrated to be sufficient for a *de novo* structure determination. This remains a challenge due to the fact that standard protein structure calculation programs are not designed with this capability, although it should be possible to develop this in future work.

Figure S2. Representative histograms of best fit values from 10,000 Monte Carlo simulations as described in the text. For the bimodal distributions, two possible values for θ , with corresponding standard deviations, are reported in Table S2.

Table S2. Pseudo-Bond Angles for 39 HN[i]-HN[i+1] and 9 HN[i]-HN[i+2] Line Shapes.

i to i+1					
<i>Resonance</i>	Θ (°)	σ (°)	<i>Population</i>	TALOS Θ (°)	X-ray Θ (°)
Q2N-Y3N	56.9	4.5	99.8%	27.7	44.4
Y3N-K4N	28.4	3.2	64.3%	20.8	35.2
	80.4	5.1	9.0%		
K4N-L5N	30.6	3.9	47.3%	15.6	32.3
	73.8	8.1	25.1%		
I6N-L7N	17.5	3.9	76.8%	32.6	34.7
	86.0	2.1	2.2%		
L7N-N8N	27.6	2.0	89.4%	33.4	35.2
	75.5	2.9	8.7%		
N8N-G9N	53.2	6.2	99.6%	35.9	58.6
K10N-T11N	17.7	5.2	94.4%	67.2	26.0
T11N-L12N	56.4	2.4	98.5%	62.4	42.2
L12N-K13N	26.0	2.6	64.0%	22.5	30.1
K13N-G14N	8.7	6.0	99.8%	23.5	14.4
G14N-E15N	11.1	4.5	99.8%	10.6	11.3
E15N-T16N	10.2	3.2	99.7%	25.2	11.5
D22N-A23N	16.9	9.2	85.0%	26.2	19.3
	83.6	3.2	4.3%		
A23N-A24N	29.6	1.4	78.1%	29.5	28.1
	80.3	2.4	21.2%		
A24N-T25N	25.5	1.0	99.8%	25.6	26.2
T25N-A26N	22.5	2.2	94.8%	24.3	15.7
A26N-E27N	21.1	1.8	98.6%	24.8	22.9
E27N-K28N	29.2	1.2	88.2%	27.8	20.8
	76.8	1.7	11.2%		
K28N-V29N	21.9	1.4	99.1%	23.6	20.9
Y33N-A34N	17.5	1.3	99.9%	25.1	17.9
A34N-N35N	17.8	2.0	99.7%	20.5	17.0
N35N-D36N	18.6	0.9	100.0%	27.2	17.5
D36N-N37N	42.1	2.4	17.7%	32.3	40.0
	62.3	2.5	81.9%		
N37N-G38N	41.0	2.7	32.8%	70.2	74.8
	60.8	3.8	66.9%		
G38N-V39N	41.5	4.8	25.1%	47.2	49.3
	66.8	4.0	74.6%		
V39N-D40N	54.7	9.0	99.2%	38.3	43.7
D40N-G41N	27.4	5.7	78.7%	34.1	28.9
	67.5	3.4	14.8%		

i to i+1					
<i>Resonance</i>	Θ (°)	σ (°)	<i>Population</i>	TALOS Θ (°)	X-ray Θ (°)
G41N-E42N	15.5	4.4	14.2%	16.5	35.2
	48.5	2.5	59.0%		
	82.4	3.7	11.8%		
E42N-W43N	45.6	6.1	43.0%	29.2	48.1
	62.6	4.1	55.3%		
W43N-T44N	0.0	0.1	4.8%	40.5	32.2
	25.0	5.1	65.1%		
	65.3	5.4	27.6%		
Y45N-D46N	11.7	2.4	99.5%	18.1	13.1
D46N-D47N	25.1	2.3	86.5%	33.6	30.7
D47N-A48N	48.8	3.4	99.9%	71.1	35.6
A48N-T49N	26.8	1.7	93.0%	40.0	44.0
T49N-K50N	48.9	6.9	99.9%	72.7	75.7
K50N-T51N	27.0	1.5	95.0%	28.9	24.3
T53N-V54N	6.7	3.3	100.0%	18.0	18.8
V54N-T55N	10.7	2.8	99.1%	14.5	14.3
T55N-E56N	11.8	1.7	99.9%	30.5	14.7

i to i+2				
<i>Resonance</i>	Θ (°)	σ (°)	<i>Population</i>	X-ray Θ (°)
A23N-T25N	5.5	4.8	79.1%	34.5
	49.8	1.4	4.4%	
A24N-A26N	9.3	7.1	94.0%	22.5
E27N-V29N	19.2	7.0	99.4%	22.1
V29N-K31N	27.9	2.1	67.4%	27.6
	69.9	2.1	32.0%	
Q32N-A34N	29.6	2.6	47.2%	27.0
	73.2	3.9	52.2%	
D36N-G38N	38.4	8.0	35.5%	35.7
	66.8	6.7	51.6%	
N37N-V39N	55.4	8.7	93.5%	84.7
D47N-T49N	22.9	4.9	4.6%	42.2
	54.8	4.2	66.4%	
	74.4	5.1	24.3%	
T49N-T51N	16.7	11.8	58.4%	80.8
	61.9	9.5	30.8%	

A graphical comparison of our results with TALOS is presented in Figure S3a. In cases where the SSNMR result had a bimodal distribution, the value in better agreement with TALOS was selected. Nevertheless, the agreement is relatively poor, relative to the X-ray diffraction data in Figure S3b. The TALOS values are particularly inaccurate for loop and turn residues, such as G41, E42, K10, D47 and T49. In addition, several residues are not reported by TALOS, including some of the most critical for defining the overall fold of the protein. Among those values reported by TALOS, the overall RMSD versus the SSNMR dipolar spectra is 15.0° .

Figure S3. Comparison of θ values determined by SSNMR with (a) TALOS determined values (RMSD 15.0°) and (b) a 1.14 Å resolution crystal structure (RMSD 8.9°).

Supporting Information References

- (1) Reif, B.; Hohwy, M.; Jaroniec, C. P.; Rienstra, C. M.; Griffin, R. G. *J. Magn. Reson.* **2000**, *145*, 132-141.
- (2) Franks, W. T.; Zhou, D. H.; Wylie, B. J.; Money, B. G.; Graesser, D. T.; Frericks, H. L.; Sahota, G.; Rienstra, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 12291-12305.
- (3) Bennett, A. E.; Rienstra, C. M.; Auger, M.; Lakshmi, K. V.; Griffin, R. G. *J. Chem. Phys.* **1995**, *103*, 6951-6958.
- (4) Hohwy, M.; Jaroniec, C. P.; Reif, B.; Rienstra, C. M.; Griffin, R. G.,. *J. Am. Chem. Soc.* **2000**, *122*, 3218-3219
- (5) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.
- (6) Rienstra, C. M.; Hohwy, M.; Mueller, L. J.; Jaroniec, C. P.; Reif, B.; Griffin, R. G. *J. Am. Chem. Soc.* **2002**, *124*, 11908-11922.
- (7) Reif, B.; Hennig, M.; Griesinger, C. *Science* **1997**, *276*, 1230-1233.
- (8) Reif, B.; Diener, A.; Hennig, M.; Maurer, M.; Griesinger, C. *J. Magn. Reson.* **2000**, *143*, 45-68.