

Atomic resolution protein structure determination by three-dimensional transferred echo double resonance solid-state nuclear magnetic resonance spectroscopy

Andrew J. Nieuwkoop, Benjamin J. Wylie, W. Trent Franks, Gautam J. Shah, and Chad M. Rienstra^{a)}

Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, USA

(Received 25 February 2009; accepted 3 August 2009; published online 1 September 2009)

We show that quantitative internuclear ^{15}N – ^{13}C distances can be obtained in sufficient quantity to determine a complete, high-resolution structure of a moderately sized protein by magic-angle spinning solid-state NMR spectroscopy. The three-dimensional ZF-TEDOR pulse sequence is employed in combination with sparse labeling of ^{13}C sites in the $\beta 1$ domain of the immunoglobulin binding protein G (GB1), as obtained by bacterial expression with 1,3- ^{13}C or 2- ^{13}C -glycerol as the ^{13}C source. Quantitative dipolar trajectories are extracted from two-dimensional ^{15}N – ^{13}C planes, in which ~ 750 cross peaks are resolved. The experimental data are fit to exact theoretical trajectories for spin clusters (consisting of one ^{13}C and several ^{15}N each), yielding quantitative precision as good as 0.1 Å for ~ 350 sites, better than 0.3 Å for another 150, and ~ 1.0 Å for 150 distances in the range of 5–8 Å. Along with isotropic chemical shift-based (TALOS) dihedral angle restraints, the distance restraints are incorporated into simulated annealing calculations to yield a highly precise structure (backbone RMSD of 0.25 ± 0.09 Å), which also demonstrates excellent agreement with the most closely related crystal structure of GB1 (2QMT, bbRMSD 0.79 ± 0.03 Å). Moreover, side chain heavy atoms are well restrained (0.76 ± 0.06 Å total heavy atom RMSD). These results demonstrate for the first time that quantitative internuclear distances can be measured throughout an entire solid protein to yield an atomic-resolution structure. © 2009 American Institute of Physics. [doi:10.1063/1.3211103]

I. INTRODUCTION

Magic-angle spinning (MAS) solid-state NMR (SS-NMR) has been used to solve several complete protein structures in recent years,^{1–6} and the precision of these structures is improving.⁵ These advances are important because SS-NMR is able to investigate samples with only local order and offers the most promising avenue toward routinely solving the structures of insoluble proteins that do not form single crystals. Specifically, substantial progress has been made in SSNMR structural investigations of both membrane proteins^{7–14} and protein aggregates.^{6,15–20} One major remaining goal is to solve structures of these systems at a rate and quality comparable to those produced by solution NMR and x-ray crystallography. Most SSNMR structures so far have been solved with a combination of semiquantitative ^{13}C – ^{13}C , ^{15}N – ^{15}N , and ^1H – ^1H distances, of precision comparable to nuclear Overhauser effects (NOEs) measured in solution. In addition, empirical dihedral angle restraints from TALOS assist in computational refinement of secondary structure elements.²¹ Until recently, the highest quality SS-NMR structures were reported with backbone RMSD (bbRMSD) values of ~ 0.8 – 1.4 Å.^{1–4} We have recently reported that with large numbers of distances, as well as high precision vector angle restraints, structures can be refined to

~ 0.2 Å bbRMSD.⁵ Likewise, we have shown that chemical shift tensors provide another avenue to atomic-resolution structure refinement.²²

Here we investigate the measurement of high precision heteronuclear (^{15}N – ^{13}C) distance restraints to refine protein structure to atomic resolution. Heteronuclear dipolar couplings, normally averaged by MAS, can be reintroduced with a train of rotor synchronized rf pulses, as in the rotational echo double resonance (REDOR) (Ref. 23) experiment and its analog transferred echo double resonance (TEDOR).²⁴ For spin pairs or clusters, these methods offer exquisitely high precision determination of internuclear distances by reporting dipolar dephasing (for REDOR) or buildup (for TEDOR) trajectories under the influence of a pulse sequence in which the multiple spin-pair interactions mutually commute with one another.^{25,26} Furthermore, for application to proteins with a large number of isotopic labels, the three-dimensional (3D) z-filtered TEDOR (ZF-TEDOR) (Ref. 27) pulse sequence was developed, enabling quantitative trajectories to be extracted from a series of two-dimensional (2D) spectra in which the ^{13}C and ^{15}N chemical shifts for each unique spin pair are encoded. Therefore many dozens of distances could be measured in uniformly ^{13}C , ^{15}N -labeled peptides.^{28,29}

Although the ZF-TEDOR experiment addresses the direct and indirect effects of ^{13}C – ^{13}C scalar couplings in the observed trajectories and peak line shapes, the modulations

^{a)}Electronic mail: rienstra@scs.uiuc.edu.

arising from one-bond couplings compromise the dynamic range of the measurement of weak dipolar couplings. This well-recognized challenge has previously been addressed by band-selective decoupling (BASE TEDOR),²⁷ as well as a semiconstant time TEDOR version.³⁰ Here we address this in a third, complementary way by preparing protein samples where the percentage of directly bonded ¹³C pairs is minimized, as derived from either 1,3-¹³C-glycerol (along with natural abundance carbonate) or 2-¹³C-glycerol (and ¹³C carbonate) as its sole sources of carbon. This expression scheme, originally developed by LeMaster *et al.*,³¹ and utilized previously in SSNMR,^{1,5,32} produces an anticorrelated “checkerboard” pattern of ¹³C and ¹²C labeling for most amino acids, enhancing resolution and sensitivity most notably in ¹³C–¹³C 2D and ¹⁵N–¹³C–¹³C 3D experiments.

We show here that for 3D ZF-TEDOR experiments, this glycerol-derived labeling pattern is particularly beneficial, enabling high-quality experimental data to be acquired with hundreds of resolved correlations. We then demonstrate a protocol for fitting the dipolar trajectories quantitatively, modeling the spin dynamics for each set of several ¹⁵N spins coupled to each resolved ¹³C site. This approach enables quantitative analysis for distances of up to at least ~5 Å, as well as detection of much longer distances (~8 Å) with moderate precision. These distance restraints prove useful in determining protein tertiary structure.

II. EXPERIMENTAL AND COMPUTATIONAL METHODS

A. Sample preparation

Samples of GB1, a 6 kDa streptococcal protein expressed from *E. coli*, were ¹³C and ¹⁵N isotopically labeled by bacterial overexpression in media containing ¹⁵N ammonium chloride as the sole nitrogen source and either (a) 2-¹³C-glycerol and calcium ¹³C-carbonate, or (b) 1,3-¹³C-glycerol and natural abundance carbonate, as the sole carbon sources.^{1,3,5,31,33} We refer to these preparations throughout the text as [2]-GB1 and [1,3]-GB1, respectively. In each case, ~18 mg of nanocrystalline protein was packed into the central 80% of a limited speed 3.2 mm Varian rotor (Varian, Inc., Fort Collins, CO). All experiments were performed using a 500 MHz InfinityPlus spectrometer (Varian, Inc., Palo Alto, CA and Fort Collins, CO) equipped with a 3.2 mm T3 Balun™ ¹H–¹³C–¹⁵N MAS probe. Pulse widths ($\pi/2$) for ¹H, ¹³C, and ¹⁵N were 2.6, 3.0, and 6.0 μ s, respectively. Spinning was controlled via a Varian MAS controller to $11\,111 \pm 2$ Hz.

B. NMR spectroscopy

A series of 2D ¹⁵N–¹³C planes was acquired according to the 3D ZF-TEDOR pulse sequence.²⁷ TPPM decoupling was used during acquisition with a ¹H field of ~70 kHz (6.7 μ s, 15.0° total phase difference). TPPM conditions during REDOR (5.0 μ s, 18.0°) were optimized with 5.76 ms mixing and a nominal ¹H field of ~100 kHz. ¹³C and ¹⁵N π -pulse widths during TEDOR were 6.0 and 15.2 μ s, respectively, the latter value adjusted to ensure that the ratio of ¹H to ¹⁵N nutation frequency was at least 3:1 to minimize decoupling interference.^{34,35} Spectra were acquired with the

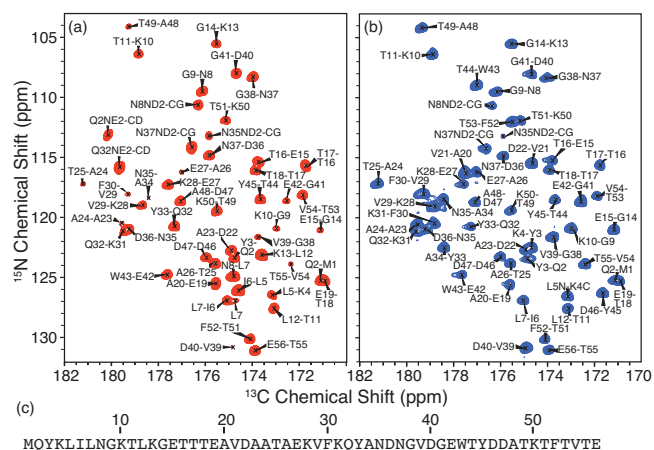


FIG. 1. Carbonyl ¹³C region of 2D ¹⁵N–¹³C TEDOR spectra (1.44 ms mixing time). Spectra of GB1 samples prepared from (a) 2-¹³C-glycerol, calcium ¹³C-carbonate, and ¹⁵NH₄Cl ([2]-GB1) and (b) 1,3-¹³C-glycerol, natural abundance carbonate, and ¹⁵NH₄Cl ([1,3]-GB1). The sequence of GB1 (c) is provided for reference. Spectra were acquired at 500 MHz ¹H frequency, except for side chain ¹⁵N sites which are explicitly indicated. Data were processed with 15 Hz net line broadening (Lorentzian-to-Gaussian apodization) in ¹³C and 5 Hz in ¹⁵N. Acquisition time was 9 h for each spectrum.

minimum phase cycle of 16 scans per row, resulting in 4.5 h blocks of measurement time with each 2D plane digitized to 640 points (TPPI) by 45 μ s per row in t_1 (¹⁵N, 28.8 ms maximum evolution time) and 3072 complex points (15 μ s dwell) in the t_2 acquisition dimension (¹³C, 46.1 ms acquisition time). Data processing was performed with 15 and 5 Hz net line broadening in the direct and indirect dimensions, respectively (using a 1:2.5 ratio of negative Lorentzian to positive Gaussian apodization) and zero filling to 8192 by 8192 complex points.

The ¹⁵N–¹³C dipolar trajectories were sampled with values of t_{mix} (according to Jaroniec *et al.*, Fig. 1) (Ref. 27) incremented from 1.44 to 14.40 ms in steps of 1.44 ms. Signal averaging times were increased for longer mixing times, with a minimum of two blocks (9 h) acquired with $t_{\text{mix}}=1.44$ ms and a maximum of nine blocks (40.5 h) with $t_{\text{mix}}=14.40$ ms. Spectra were normalized, according to the total number of scans per row, prior to further analysis.

C. Numerical simulations and data fitting

Spectra were processed in NMRPipe (Ref. 36) and peak intensities extracted from the 2D ¹⁵N–¹³C planes using the ninLS package, resulting in trajectories of integrated peak intensities as a function of mixing time, from 1.44 to 14.4 ms in 1.44 ms increments.

Simulated trajectories were generated using SPINEVOLUTION (Ref. 37) to model spin dynamics of ZF-TEDOR with the following assumptions: (1) All spin coherences derived from a single ¹³C site commute with all other spin operators involving other ¹³C spins; (2) each ¹³C site is coupled to n ¹⁵N sites, which are not coupled to other ¹⁵N spins; (3) the relative orientations of the heteronuclear dipoles could be ignored. (We show experimentally (*vide infra*) that these approximations are valid in the limit of weak couplings.) Thus,

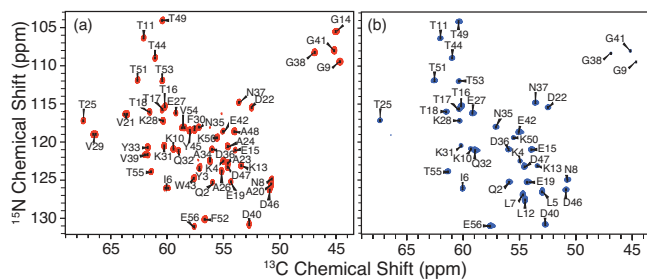


FIG. 2. CA region of 2D ^{15}N - ^{13}C TEDOR spectra (1.44 ms mixing time). (a) [2]-GB1; (b) [1,3]-GB1. Spectra were processed with 15 Hz net line broadening (Lorentzian-to-Gaussian apodization) in ^{13}C and 5 Hz in ^{15}N . Acquisition time was 9 h for each spectrum.

simulated trajectories for each of the n ^{15}N sites coupled to a single ^{13}C resonance were generated.

The dipolar couplings and other incidental parameters were then determined by minimizing the global difference between the simulated and experimental trajectories, using in-house FORTRAN code that called MINUIT minimization libraries and SPINEVOLUTION.³⁷ The distance between ^{13}C and ^{15}N spin pairs was simulated for each point on the trajectory and a scaling factor was applied to the result to account for the ^{13}C labeling percentage. In this scheme, y -scaling, corresponding to labeling percentage, is held fixed over the ensemble and T_2 relaxation, applied as an exponential decay, is allowed to vary by 10% over the ensemble, to account for possible higher order relaxation from ^{15}N CSA recoupling during the REDOR periods.

III. RESULTS AND DISCUSSION

A. 2D TEDOR heteronuclear correlation spectroscopy

As demonstrated previously,³⁸ the 2D ^{15}N - $^{13}\text{C}'$ spectra of [2]-GB1 [Fig. 1(a)] and [1,3]-GB1 [Fig. 1(b)] GB1 are well resolved, with natural ^{13}C linewidths of ~ 0.2 ppm and ^{15}N linewidths of ~ 0.5 ppm, principally limited by instrumental factors (such as the B_0 homogeneity and the maximum time for which high power ^1H decoupling can be applied). Spectra acquired at 1.44 ms ^{15}N - ^{13}C mixing, the shortest value utilized here, contain strong peaks from one-bond correlations (e.g., $\text{N}[i]\text{-C}'[i-1]$), with the relative intensity depending principally upon the efficiency of the ^{13}C labeling within each amino acid. For example, in the [2]-GB1 sample, carbonyl sites for Leu are nearly 100% labeled, whereas Ala, Gly, Val, and other amino acids from the glycolysis pathway are labeled only to a small extent; in the [1,3]-GB1 sample, a complementary pattern is observed. In both samples, carbonyl sites from amino acids in the citric acid cycle (most notably, Gln/Glu, Asp/Asn, Lys, and Thr) are observed with approximately equal intensity in the [2]-GB1 and [1,3]-GB1 samples, due to the scrambling of ^{13}C within the cycle. These signals serve as valuable internal controls to validate data analysis procedures.

Likewise, the TEDOR spectra are well resolved in the CA region for both the [2]-GB1 [Fig. 2(a)] and [1,3]-GB1 [Fig. 2(b)] samples, with nearly every expected cross peak resolved. Partial overlap of the N8, A20, and D46 resonances in the [2]-GB1 spectrum is alleviated by the reduction in A20

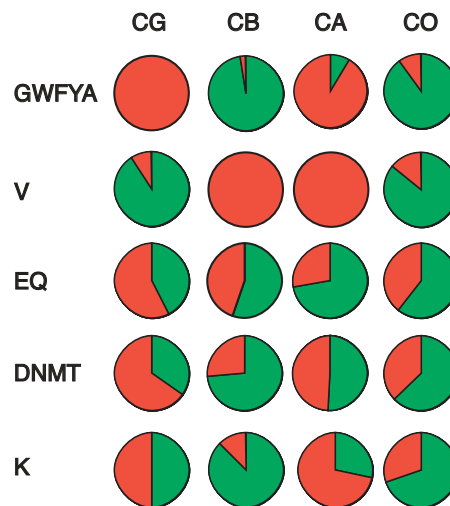


FIG. 3. ^{13}C labeling pattern measured by TEDOR in GB1. Red represents the [2]-GB1 and green [1,3]-GB1. Each circle represents the intensity of correlations to the carbon type indicated at the top in the residue group indicated to the left.

intensity in the [1,3]-GB1 spectrum. An additional benefit of the glycerol labeling scheme is the absence of most one-bond homonuclear J -couplings, with the exception the Val and Ile CA-CB, which lead to doublet patterns observed for 16 and the four Val residues (21, 29, 39, and 54). Moreover, broadening from off or near rotational resonance (R^2) is avoided.^{33,39-43} These factors together contribute to high observed signal-to-noise ratio (SNR) for both samples, approximately 500:1 in 9 h of data collection for directly bonded ^{15}N - ^{13}C pairs that were $\sim 100\%$ labeled and proportionately lower for the fractionally labeled sites. Thus it was also possible to observe weak peaks at natural abundance and/or with a small percentage of labeling; in several instances, sites that are nominally unlabeled appeared with approximately 5% of the maximum intensity [e.g., A48 C' in the [2]-GB1 sample, Fig. 1(a)].

The glycerol-derived isotopic incorporation on an amino acid specific basis is likely to vary among proteins, due to differences in the relative percentages of each amino acid; in the case of GB1, such effects are likely to be exaggerated by the very high levels of expression (>100 mg/L) and the fact that GB1 lacks several amino acids (Cys, His, Pro, Ser, Arg) and has a large number of others (such as Thr, Asp, and Glu). Therefore the NCA and NCO 2D spectra can be utilized to approximate the percentage of ^{13}C labeling in our samples. This percentage was incorporated as a parameter into the fitting procedure as an intensity scaling factor; this treatment is valid in the limit that the relaxation parameters are similar for all sites, which is satisfactory in the limit of short TEDOR mixing times. The results of this analysis are shown in Fig. 3. Overall patterns were very similar to those reported previously,^{1,3} but with substantial signal intensities ($\sim 10\%$ of the maximum) for sites that were not expected to be labeled, such as CO resonances for Gly, Trp, Phe, Tyr, and Ala in the [2]-GB1 sample. Despite the presence of these peaks, we consistently observed lineshapes lacking fine structure from $^1J_{\text{CC}}$ couplings, consistent with the anticorrelated labeling of neighboring carbons.

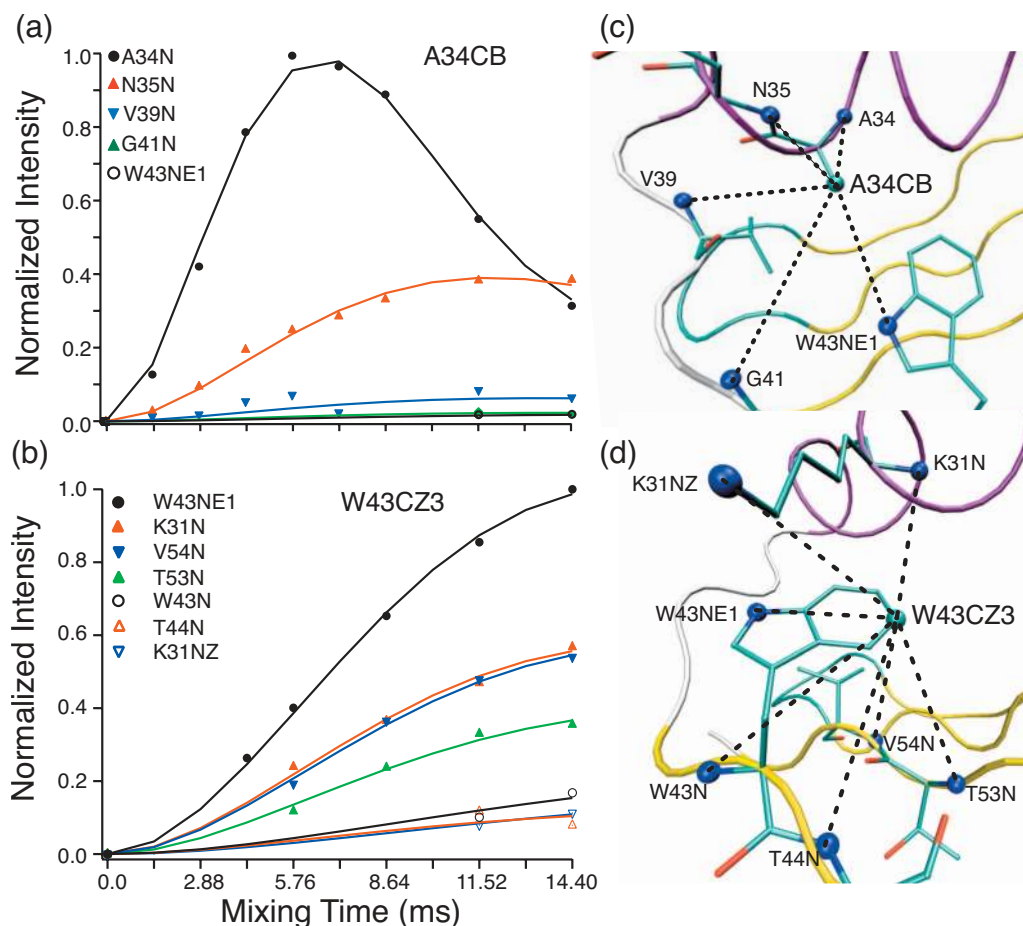


FIG. 6. TEDOR polarization transfer trajectories of selected ^{13}C - ^{15}N correlations in GB1. Trajectories for (a) A34CB in [1,3]-GB1 and (b) W43CZ3 in [2]-GB1. Correlations labeled on the 2QMT crystal structure for (c) A34B and (d) W43CZ3. Intensities in the trajectories are derived from integrated volumes of peaks in 2D ^{15}N - ^{13}C TEDOR experiments at the respective mixing times. Simulated trajectories were generated using SPINEVOLUTION simulations of the TEDOR pulse sequence as described in the text. Fit and x-ray distances for A34CB [(a) and (c)] are, respectively, as follows; A34N, 2.44 ± 0.10 Å vs 2.44 Å; N35N, 3.29 ± 0.10 Å vs 3.09 Å; V39N, 4.46 ± 0.86 Å vs 5.13 Å; G41N, 5.30 ± 1.79 Å vs 6.74 Å; W43NE1, 5.9 ± 3.2 Å vs 5.47 Å. Fit and x-ray distances for W43CZ3 [(b) and (d)] are, respectively, as follows; W43NE1, 4.10 vs 4.06 Å; K31N, 4.53 vs 4.61 Å; V54N, 4.55 vs 4.85 Å; T53N, 4.90 vs 5.18 Å; W43N, 6.02 vs 6.36 Å; T44N, 6.04 vs 6.15 Å; K31NZ, 6.38 vs 6.00 Å. The molecular graphic shows the backbone of helices in purple, strands in yellow, and loops in cyan; the dark blue circles are nitrogen atoms, and the distances to the labeled ^{13}C site are indicated with dotted lines.

DOR, however, the fit of the TEDOR trajectories is somewhat complicated by the fact that the maximum absolute intensity in the trajectory depends not only on heteronuclear dipolar coupling, but also the ^{13}C labeling percentage, as well as other factors (relaxation, presence of multiple dipoles, relative orientations, etc.).²⁷ In particular, the labeling percentage is a critical parameter that contributes to the normalization of the spectral intensity, which distinguishes between (a) short distances involving a fractionally labeled ^{13}C and (b) long distances involving a ^{13}C that is $\sim 100\%$ labeled. These factors must be explicitly considered in order to determine quantitative coupling constants. To first approximation, we considered the couplings, labeling percentage, and relaxation as fit parameters, assuming that the labeling percentage and relaxation must be the same for each ^{13}C site, and that all ^{15}N sites were labeled to 100%. The relative orientations of multiple ^{13}C - ^{15}N dipoles were ignored.

Figure 6 illustrates two typical scenarios encountered in fitting ensembles of ^{15}N sites correlated with a single ^{13}C . The A34CB resonance [Fig. 6(a)] has intrasidue, sequential, and long-range correlations. The intrasidue (A34N-A34CB) trajectory exhibits a maximum at 5.76 ms mixing,

from which the distance of 2.44 Å is fit directly, also uniquely constraining the intensity scaling factor. With accurate knowledge of this scaling factor, the remaining trajectories can be fitted unambiguously. In comparison to the crystal structure [2QMT, Fig. 6(c)], we find reasonable agreement for the short (A34N-A34CB, 2.44 ± 0.10 Å TEDOR versus 2.44 Å x ray), medium (N35N-A34CB, 3.29 ± 0.10 Å TEDOR versus 3.09 Å x ray), and long (V39N-A34CB, 4.46 ± 0.86 Å TEDOR versus 5.13 Å x ray; G41N-A34CB, 5.30 ± 1.79 Å TEDOR versus 6.74 Å x ray; W43NE1-A34CB, 5.9 ± 3.2 Å TEDOR versus 5.47 Å x ray) distances. Uncertainties in the TEDOR-determined distances (as determined by the MINUIT routine) are largest for the peaks observed only at the long mixing times, such as W43NE1-A34CB, which is at the borderline of sensitivity in this experiment.

In general, ^{15}N - ^{13}C distances of less than ~ 3.5 Å will yield trajectories exhibiting maxima at less than 14.4 ms, which can be uniquely fit to intensity scaling and heteronuclear coupling. In cases where no ^{15}N nuclei are within 3.5 Å of an observed ^{13}C resonance, this procedure becomes problematic and the final scaling factor, and therefore dis-

tances, becomes less reliable. For example, the W43CZ3 TEDOR trajectory [Fig. 6(b)] has not yet reached its maximum at 14.4 ms; trajectories in the initial rate regime exhibit a strong covariance between the intensity scaling factor and the dipolar coupling. Therefore the certainty of distance determination is limited by knowledge of the hypothetical maximum signal intensity. To address this problem, knowledge of known molecular geometry in this case can be utilized as an internal control to disambiguate the fit parameters. Specifically, the W43NE1-W43CZ3 distance is dictated by the geometry of the indole ring to be 4.06 Å. Thus in fitting this group of trajectories we constrained this distance to be 4.05 ± 0.05 Å, and the global fit for all six distances was found to be internally consistent. Distances calculated from this trajectory were systematically short without the control, but using this chemical restraint the ^{15}N to W43CZ3 distances converge very well to the those observed in the x-ray structure (K31N, 4.53 ± 0.10 versus 4.61 Å x ray; V54N, 4.55 ± 0.10 versus 4.85 Å; T53N, 4.90 ± 3.90 versus 5.18 Å; W43N, 6.02 ± 0.10 versus 6.36 Å; T44N, 6.04 ± 0.10 versus 6.15 Å; K31NZ, 6.38 ± 0.10 versus 6.00 Å). When fixing a distance within the ensemble, error determinations for the remaining distances become less reliable as the mingrad MINUIT algorithm cannot properly sample the solution space to determine the reliability of the fit, although a reasonable estimate of the uncertainty for all trajectories shown here is better than ± 0.4 Å (i.e., the range of distances from the mean required to generate simulations that encompass all observed experimental data points). A total of 726 distance restraints (340 from the [1,3]-GB1 and 386 from the [2]-GB1) were determined. Among these, 454 had distance uncertainties of 0.2 Å or less; another 95 had uncertainties between 0.2 and 0.5 Å, 48 had uncertainties between 0.5 and 1.0 Å, and the remaining 64 had uncertainties greater than 1.0 Å. In some instances (including most of the 64 reported to have greater than 1.0 Å uncertainty), MINUIT failed to report reasonable errors [for example, T53N-W43CZ3, Fig. 6(b)] although by manual inspection it was clear that the trajectories were well fit. Nevertheless, the large uncertainties were used in subsequent calculations (*vide infra*).

We identified and removed intermolecular correlations based on similarity to contacts observed previously.^{5,30,49} The remaining TEDOR-determined ^{13}C - ^{15}N distances showed good agreement with the x-ray structure (Fig. 7). Excluding the prochiral methyl carbons with ambiguous assignments and side chain ^{15}N sites, 240 out of 327 distances (73%) determined in the [2]-GB1 sample were within 10% of the x-ray determined distances, and 313 out of 327 (96%) were within 25% of the x-ray distance. For [1,3]-GB1, likewise, 168 out of 260 (65%) were within 10%, and 235 (90%) were within 25%. There is a trend evident in the plot, that outlying points with shorter-than-expected NMR distances are observed more frequently than longer-than-expected NMR distances; we attribute this to the fact that the short-than-expected NMR distances are derived from stronger-than-expected cross peaks, which are above the noise floor, whereas the weaker-than-expected cross peaks are not observed at all. In addition, variations between the TEDOR and

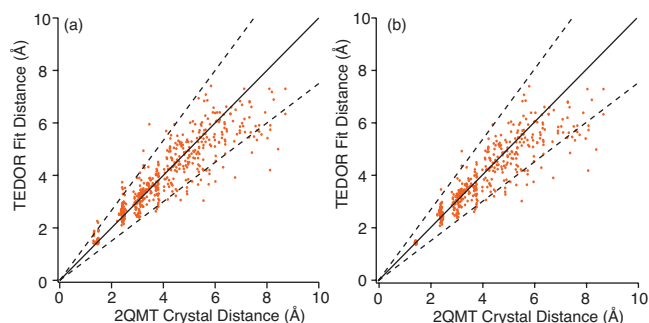


FIG. 7. Scatter plot of ^{15}N - ^{13}C distances determined by TEDOR experiments and x-ray crystallography (PDB entry 2QMT). (a) Plot including all ^{15}N sites. (b) Plot including only backbone amide sites. The solid line has a slope of 1, to guide the eye, and the dotted lines indicate variations of $\pm 25\%$ from the nominal distance.

x-ray distances are expected because the sample preparations are not identical. For example, we previously found significant differences in side chain conformation due to subtle differences in crystal polymorphism⁴⁵ which more detailed NMR studies could elucidate.

C. Protein structure calculations with heteronuclear distance restraints

We next calculated a *de novo* structure of GB1 in XPLOR-NIH (Ref. 50) using as experimental restraints the complete set of TEDOR-determined ^{15}N - ^{13}C distances. Experimental uncertainties were assumed to have a minimum of 0.1 Å, even in cases where MINUIT reported smaller uncertainties. In addition to distances, TALOS (Ref. 21) was used to determine backbone dihedral angle from the isotropic chemical shifts. Initial structures calculated this way converged to an ~ 0.6 Å backbone RMSD fold of GB1, with 36 ^{15}N - ^{13}C distances violating by more than 0.5 Å. In cases where distance restraints consistently violated the majority of lowest energy structures, the uncertainties were increased by 1.0 Å and the calculation repeated. Two restraints, T53N-I6CG1 from [1,3]-GB1 and K10NZ-CG from [2]-GB1 (out of the 36) that continued to violate, despite this increase in uncertainty, were removed altogether. The backbone RMSD converged to 0.35 Å after this round of calculations. The violation threshold was then reduced to 0.3 Å, and another 52 violating distance restraints were relaxed and three more, W43N-V54CG1 from both [1,3]-GB1 and [2]-GB1 and T54N-I6CG1 from [2]-GB1, were removed (86 in total). The process was repeated until no violations greater than 0.3 Å remained. Table I shows all restraints used in the calculation sorted by the distance in the primary structure. Long-range ($|i-j| > 3$), medium range ($1 < |i-j| < 4$), sequential ($|i-j| = 1$), and intraresidue correlations are listed along with their number of violations seen in each category. The majority of all restraints ($\sim 88\%$) never violated in any of the calculations.

The violating distance restraints were predominantly in one of two categories: (1) long-range restraints to ^{13}C or ^{15}N sites on flexible side chains and (2) distances involving ^{13}C nuclei with no ^{15}N within ~ 4 Å. Examples of the first case are I6CD1, I6CG1/2, L7CG, L7CD1/2, K13CE, K13CD, and

TABLE I. TEDOR restraints used in structure calculations.

Restraint type	Never violate ^a	Violate once ^b	Violate twice ^c	Total
Long ^d	108	23	4	135
Medium ^e	53	4	0	57
Sequential ^f	211	17	0	228
Intraresidue	268	42	1	311
Total	640	86	5	731
Percentage of total	87.6	11.8	0.7	

^aRestraints that did not violate by more than 0.3 Å in a majority of the lowest energy models.

^bRestraints that violated once and had 1 Å added to their error.

^cRestraints that violated after being lengthened and were removed from additional calculations.

^d $|i-j| > 4$.

^e $|i-j| \leq 3$.

^f $|i-j| = 1$.

several Lys NZ resonances. As shown by Ishii and Terao, molecular motion alters the effective dipolar couplings in SSNMR,⁵¹ in a manner that for strong couplings will generally result in larger values for r determined by $\langle 1/r^3 \rangle$ (the NMR observable) than from $\langle r \rangle$ (the x-ray diffraction observable). In the case of ¹³C nuclei with no nearby ¹⁵N, the γ -scaling factor is highly uncertain, and therefore the covariance between intensity and coupling in this initial rate regime gives the distance estimations a high uncertainty. These effectively serve as long-range distance restraints of quality similar to NOEs.

Once violating distance restraints were removed by this procedure, the resulting structure calculation converged to a family of structures (ten lowest energy out of 260 calculated) with a backbone RMSD of 0.25 ± 0.09 Å with an all atom RMSD of 0.76 ± 0.06 Å. The average structure showed good agreement with the most closely related crystal structure (2QMT), with a bbRMSD of 0.79 ± 0.03 Å. Figure 8 shows the family of best structures shown with (a) only

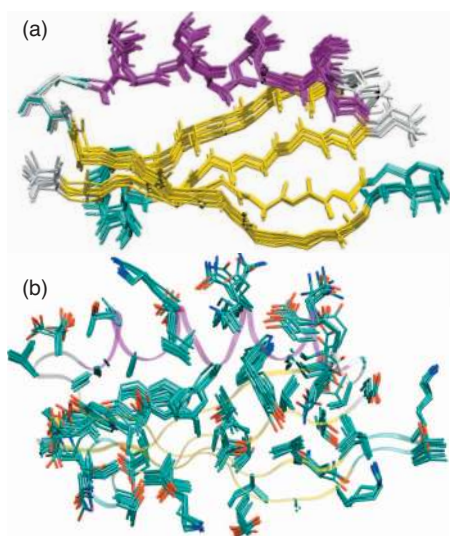


FIG. 8. GB1 structure calculated using TEDOR distance and TALOS dihedral constraints. (a) Backbone trace of the ten lowest energy structures (out of 260) calculated. (b) Side chains drawn in CPK coloring scheme. The backbone RMSD is 0.25 ± 0.09 Å within the family of the ten best NMR structures, and the RMSD in comparison to the crystal structure (2QMT) is 0.76 ± 0.06 Å. The all heavy atom RMSD is 0.79 ± 0.03 Å.

TABLE II. XPLOR calculation energies for the ten lowest energy structures.

Energies	Average (kcal mol ⁻¹)	Standard deviation
Total	228.78	3.52
Bonds	19.35	1.18
Angles	80.15	2.28
VDW	0.03	0.09
TEDOR	115.96	2.93
Dihedral	4.56	0.55
Improper	8.72	0.76

backbone atoms and (b) all side chain atoms. Most variation among the structures is observed in the turns at the beginning and end of the helix. The side chains of residues in the core of the protein are very well constrained leading to very high similarity as indicated by the low all atom RMSD. Thus a majority of the variance in the all atom RMSD is from external side chains that are less well constrained. Table II shows the average XPLOR energies for this family of structures, which indicate that a majority of the pseudoenergy in these structures arises from the TEDOR distance constraints and the XPLOR angle restraints.

IV. CONCLUSIONS

We have demonstrated that the TEDOR pulse sequence can be applied to a preparation of GB1 that is uniformly ¹⁵N and sparsely ¹³C labeled. This sparse ¹³C labeling allows for long-range TEDOR distances to be observed for every ¹³C site in GB1. Multiple ¹⁵N–¹³C mixing times were used to create dipolar coupling trajectories for carbon and nitrogen sites up to 6 Å apart and to detect correlations arising from pairs up to 8 Å apart. These trajectories were fit to exact numerical simulations of TEDOR allowing ¹⁵N–¹³C distances to be determined with both precision and accuracy. When incorporated into simulated annealing calculations these distance restraints, combined with dihedral restraints, produced a very high-resolution fold of GB1. Substantially fewer distances were required to determine this structure than in solution NMR studies with similar precision due to the fact that TEDOR short-to-medium distances can be determined with greater precision and accuracy than ¹H–¹H NOEs. Because SSNMR has no intrinsic upper molecular weight limit, this approach is promising for the study of larger nanocrystalline proteins, membrane proteins, and insoluble aggregates, in cases where high-resolution 2D ¹⁵N–¹³C spectra can be obtained.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health Grant No. R01-GM073770. A.J.N. thanks the Molecular Biophysics Training Program at the University of Illinois for financial support.

¹F. Castellani, B. van Rossum, A. Diehl, M. Schubert, K. Rehbein, and H. Oschkinat, *Nature (London)* **420**, 98 (2002).

²A. Lange, S. Becker, K. Seidel, K. Giller, O. Pongs, and M. Baldus, *Angew. Chem., Int. Ed.* **44**, 2089 (2005).

³S. G. Zech, A. J. Wand, and A. E. McDermott, *J. Am. Chem. Soc.* **127**,

- 8618 (2005).
- ⁴D. H. Zhou, J. J. Shea, A. J. Nieuwkoop, W. T. Franks, B. J. Wylie, C. Mullen, D. Sandoz, and C. M. Rienstra, *Angew. Chem., Int. Ed.* **46**, 8380 (2007).
- ⁵W. T. Franks, B. J. Wylie, H. L. F. Schmidt, A. J. Nieuwkoop, R. M. Mayrhofer, G. J. Shah, D. T. Graesser, and C. M. Rienstra, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4621 (2008).
- ⁶C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, and B. H. Meier, *Science* **319**, 1523 (2008).
- ⁷S. J. Opella and F. M. Marassi, *Chem. Rev. (Washington, D.C.)* **104**, 3587 (2004).
- ⁸A. J. Gammeren, F. B. Hulsbergen, J. G. Hollander, and H. J. de Groot, *J. Biomol. NMR* **31**, 279 (2005).
- ⁹M. Kobayashi, Y. Matsuki, I. Yumen, T. Fujiwara, and H. Akutsu, *J. Biomol. NMR* **36**, 279 (2006).
- ¹⁰H. L. Frericks, D. H. Zhou, L. L. Yap, R. B. Gennis, and C. M. Rienstra, *J. Biomol. NMR* **36**, 55 (2006).
- ¹¹M. Etzkorn, S. Martell, O. C. Andronesi, K. Seidel, M. Engelhard, and M. Baldus, *Angew. Chem., Int. Ed.* **46**, 459 (2007).
- ¹²A. Z. Kijac, Y. Li, S. G. Sligar, and C. M. Rienstra, *Biochemistry* **46**, 13696 (2007).
- ¹³M. Hiller, V. A. Higman, S. Jehle, B. J. van Rossum, W. Kuhlbrandt, and H. Oschkinat, *J. Am. Chem. Soc.* **130**, 408 (2008).
- ¹⁴Y. Li, D. A. Berthold, R. B. Gennis, and C. M. Rienstra, *Protein Sci.* **17**, 199 (2008).
- ¹⁵H. Heise, W. Hoyer, S. Becker, O. C. Andronesi, D. Riedel, and M. Baldus, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15871 (2005).
- ¹⁶R. Tycko, *Q. Rev. Biophys.* **39**, 1 (2006).
- ¹⁷K. D. Kloepper, W. S. Woods, K. A. Winter, J. M. George, and C. M. Rienstra, *Protein Expression Purif.* **48**, 112 (2006).
- ¹⁸A. B. Siemer, A. A. Arnold, C. Ritter, T. Westfeld, M. Ernst, R. Riek, and B. H. Meier, *J. Am. Chem. Soc.* **128**, 13224 (2006).
- ¹⁹K. D. Kloepper, K. L. Hartman, D. T. Lador, and C. M. Rienstra, *J. Phys. Chem. B* **111**, 13353 (2007).
- ²⁰K. Varga, L. Tian, and A. E. McDermott, *Biochim. Biophys. Acta* **1774**, 1604 (2007).
- ²¹G. Cornilescu, F. Delaglio, and A. Bax, *J. Biomol. NMR* **13**, 289 (1999).
- ²²B. J. Wylie, C. D. Schwieters, E. Oldfield, and C. M. Rienstra, *J. Am. Chem. Soc.* **131**, 985 (2009).
- ²³T. Gullion and J. Schaefer, *J. Magn. Reson. (1969-1992)* **81**, 196 (1989).
- ²⁴A. Hing, S. Vega, and J. Schaefer, *Magn. Reson. (1969-1992)* **96**, 205 (1992).
- ²⁵C. A. Michal and L. W. Jelinski, *J. Am. Chem. Soc.* **119**, 9059 (1997).
- ²⁶C. P. Jaroniec, J. C. Lansing, B. A. Tounge, M. Belenky, J. Herzfeld, and R. G. Griffin, *J. Am. Chem. Soc.* **123**, 12929 (2001).
- ²⁷C. P. Jaroniec, C. Filip, and R. G. Griffin, *J. Am. Chem. Soc.* **124**, 10728 (2002).
- ²⁸C. P. Jaroniec, C. E. MacPhee, N. S. Astrof, C. M. Dobson, and R. G. Griffin, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16748 (2002).
- ²⁹C. P. Jaroniec, C. E. MacPhee, V. S. Bajaj, M. T. McMahon, C. M. Dobson, and R. G. Griffin, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 711 (2004).
- ³⁰J. J. Helmus, P. S. Nadaud, N. Hofer, and C. P. Jaroniec, *J. Chem. Phys.* **128**, 052314 (2008).
- ³¹D. M. LeMaster, *J. Am. Chem. Soc.* **118**, 9255 (1996).
- ³²M. Hong, *J. Magn. Reson.* **139**, 389 (1999).
- ³³B. J. Wylie, L. J. Sperling, H. L. Frericks, G. J. Shah, W. T. Franks, and C. M. Rienstra, *J. Am. Chem. Soc.* **129**, 5318 (2007).
- ³⁴Y. Ishii, J. Ashida, and T. Terao, *Chem. Phys. Lett.* **246**, 439 (1995).
- ³⁵A. E. Bennett, C. M. Rienstra, J. M. Griffiths, W. G. Zhen, P. T. Lansbury, and R. G. Griffin, *J. Chem. Phys.* **108**, 9463 (1998).
- ³⁶F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, *J. Biomol. NMR* **6**, 277 (1995).
- ³⁷M. Veshort and R. G. Griffin, *J. Magn. Reson.* **178**, 248 (2006).
- ³⁸B. J. Wylie, L. J. Sperling, and C. M. Rienstra, *Phys. Chem. Chem. Phys.* **10**, 405 (2008).
- ³⁹D. P. Raleigh, M. H. Levitt, and R. G. Griffin, *Chem. Phys. Lett.* **146**, 71 (1988).
- ⁴⁰M. H. Levitt, D. P. Raleigh, F. Cruzet, and R. G. Griffin, *J. Chem. Phys.* **92**, 6347 (1990).
- ⁴¹M. Helmle, Y. K. Lee, P. J. E. Verdegem, X. Feng, T. Karlsson, J. Lugtenburg, H. J. M. de Groot, and M. H. Levitt, *J. Magn. Reson.* **140**, 379 (1999).
- ⁴²L. Duma, S. Hediger, A. Lesage, D. Sakellariou, and L. Emsley, *J. Magn. Reson.* **162**, 90 (2003).
- ⁴³T. I. Igumenova and A. E. McDermott, *J. Magn. Reson.* **164**, 270 (2003).
- ⁴⁴W. T. Franks, D. H. Zhou, B. J. Wylie, B. G. Money, D. T. Graesser, H. L. Frericks, G. Sahota, and C. M. Rienstra, *J. Am. Chem. Soc.* **127**, 12291 (2005).
- ⁴⁵H. L. Frericks Schmidt, L. J. Sperling, Y. G. Gao, B. J. Wylie, J. M. Boettcher, S. R. Wilson, and C. M. Rienstra, *J. Phys. Chem. B* **111**, 14362 (2007).
- ⁴⁶K. T. Mueller, T. P. Jarvie, D. J. Aurentz, and B. W. Roberts, *Chem. Phys. Lett.* **242**, 535 (1995).
- ⁴⁷K. T. Mueller, *J. Magn. Reson., Ser. A* **113**, 81 (1995).
- ⁴⁸J. Schaefer, *J. Magn. Reson.* **137**, 272 (1999).
- ⁴⁹X. H. Peng, D. Libich, R. Janik, G. Harauz, and V. Ladizhansky, *J. Am. Chem. Soc.* **130**, 359 (2008).
- ⁵⁰C. D. Schwieters, J. J. Kuszewski, N. Tjandra, and G. M. Clore, *J. Magn. Reson.* **160**, 65 (2003).
- ⁵¹Y. Ishii, T. Terao, and S. Hayashi, *J. Chem. Phys.* **107**, 2760 (1997).