J|A|C|S
COMMUNICATIONS

Published on Web 05/31/2007

Rapid Data Collection for Protein Structure Determination by NMR Spectroscopy

Yingqi Xu, Dong Long, and Daiwen Yang*

Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543

Received March 1, 2007; E-mail: dbsydw@nus.edu.sg

Traditionally, relatively long (at least several weeks) data collection time is needed for protein structure determination by NMR. This partly results from the intrinsic low sensitivity of NMR. However, with the availability of high-field magnets and cryogenic probes, sensitivity is no longer the bottleneck in some cases. Thus several fast sampling methods, such as GFT,[1] projection reconstruction,[2] and sparse sampling,[3] have been introduced during the last several years to speed up multidimensional NMR experiments to avoid the so-called "sampling limitation". Such methods are quite useful in cases where proteins under investigation are very soluble and experimental sensitivity is not an issue. For relatively large proteins and relatively low concentration protein samples, however, most experiments that are commonly used for the assignments of backbone and side-chain resonances (such as HNCACB, CBCA(CO)NH, and C(CO)NH-TOCSY-type experiments) are still limited by sensitivity. In these cases, the total experimental time cannot be shortened by the fast sampling methods. Another method to reduce experimental time is simultaneous acquisition of several spectra, which has been applied to both 3D and 4D NOESY.[4] Regardless of protein concentration, the time-shared method can reduce the total experimental time by a factor of 2 or more.

Recently, we introduced a strategy which uses only 3D HNCA, 3D MQ-CCH-TOCSY, and 4D $^{13}C,^{15}N$-edited NOESY to assign backbone and side-chain resonances of large proteins without deuteration.[5] For this strategy, a 4D $^{13}C,^{13}C$-edited NOESY is also needed for high-resolution structure determination. In order to obtain reasonable spectral resolutions in indirect dimensions, each 4D NOESY experiment needs at least 4−7 days (depending on spectrometer frequency) due to the steps of the phase cycle. If the total time for the NOESY experiments can be reduced greatly, this strategy should provide an alternative way to rapidly determine protein structures when it is applied to small and medium-sized proteins since the strategy uses a minimal number of experiments.

Here we present a time-shared 4D NOESY experiment that records 4D $^{13}C,^{13}C$-edited NOESY, $^{13}C,^{15}N$-edited NOESY, and $^{15}N,^{15}N$-edited NOESY spectra (Figure 1). The pulse scheme is similar to time-shared 3D experiments proposed previously.[4a−c] To optimize sensitivity, we employed (i) an HMQC-NOESY-HSQC scheme, (ii) acquisition of the first ^1H dimension during the INEPT transfer and refocus periods, and (iii) optimal water flip-back. If an HSQC-NOESY-HSQC scheme is used, the INEPT transfer periods cannot be utilized for acquisition and need an extra acquisition period that leads to ∼10% of signal loss for proteins with an overall correlation time of 8 ns. The loss would be much more significant for larger proteins. Due to the radiation damping effect, the apparent longitudinal relaxation time of water magnetization is often $<$40 ms for a protein sample with 90% H_2O in the absence of pulsed field gradients.[6] Placing gradient g6 just at the end of the mixing period (instead of before the mixing) allows optimal recovery of water magnetization, leading to sensitivity enhancement. Due to the application of one 180° ^{15}N pulse just after the second t_{2a} period, diagonal peaks resulting from NH and NH_2 groups have opposite signs to those from CH, CH_2, and CH_3 groups. In this way, we can discriminate NH−NH from $C_{aro}H_{aro}-C_{aro}H_{aro}$ NOEs, where subscript aro denotes aromatic, even without separating the $^{15}N,^{15}N$-edited spectrum from the $^{13}C,^{13}C$-edited spectrum. To obtain high resolution for ^{15}N resonances in F_3, extra acquisition periods (t_{3a}) for ^{15}N spins are used. During these periods, ^{13}C magnetization is aligned along the *Z*-axis, and thus it does not decay significantly. The resultant spectrum can be split into four sub-spectra according to proton chemical shifts: HN-NOESY-NH, HC-NOESY-NH, HN-NOESY-CH, and HC-NOESY-CH spectra. It is noteworthy that the NOEs between aromatic CH groups appear in the HN-NOESY-NH sub-spectrum.

The strategy was applied to ubiquitin (UB, 76 residues, ∼1 mM, 25 °C), human liver fatty acid binding protein whose NMR assignment and solution structure are presented here for the first time (LFABP, 129 residues, ∼0.5 mM, 25 °C), and a cell−cell adhesion protein (DdCAD-1, 214 residues, ∼0.8 mM, 30 °C). The HNCA experimental times were ∼2.5, 4, and 4 h for UB, LFABP, and DdCAD-1, respectively. For each sample, the experimental times were 8.5 and 96 h for 3D MQ-CCH-TOCSY and time-shared 4D NOESY, respectively. The NOE mixing times were 100, 100, and 75 ms for UB, LFABP, and DdCAD-1, respectively. A 3D HNCO was recorded for DdCAD-1 with an experimental time of 2.5 h. All of the experiments were performed on a 500 MHz spectrometer equipped with a cryoprobe. The acquisition and processing parameters for the 4D NOESY are listed in Table S1 of the Supporting Information.

The HC-NOESY-NH sub-spectra were used together with HNCA and CCH-TOCSY to derive backbone and side-chain sequence-specific assignments as described previously.[5] Representative data and assignment procedures are shown in Figure S1 of the Supporting Information. For UB, LFABP, and DdCAD-1, peaks in HNCA and HC-NOESY-NH could be easily grouped based on ($^1H_N,^{15}N$) chemical shifts because of well-dispersed $^1H-^{15}N$ correlations. For DdCAD-1, six clusters contained correlations from two amides, whereas there were two such clusters when the 3D spectra were collected on an 800 MHz spectrometer. In order to confirm the number of amides in each cluster, a 3D HNCO spectrum was used. Compared to larger proteins, such as maltose binding protein, the assignment was rather easy because of the reduced complexity and better quality in the NOESY and TOCSY spectra. Nearly complete backbone and side-chain assignments were obtained for all three proteins (Table 1). G53 of UB displayed a very weak $^1H-^{15}N$ HSQC peak and was not assigned at the beginning, but could be assigned later based on NOESY. Three amides (S2, S58, and N99) displayed no HSQC peaks for LFABP and were not assigned. Although the 3D HNCA and MQ-CCH-TOCSY spectra of DdCAD-1 were collected on a 500 MHz spectrometer, we could obtain similar assignments to those obtained previously with the 3D spectra recorded on an 800 MHz machine.

 10.1021/ja071442e CCC: $37.00

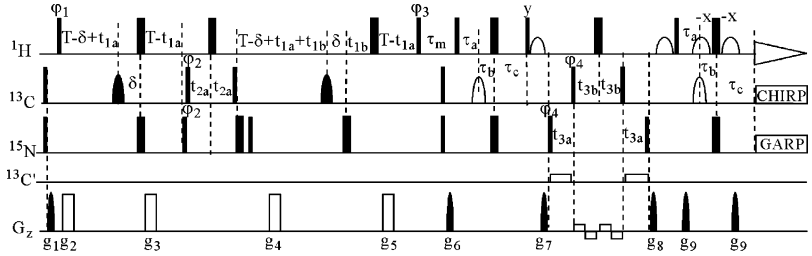

Figure 1. Pulse sequence for recording 4D time-shared $^{13}C/^{15}N$, $^{13}C/^{15}N$-edited NOESY. All narrow (wide) bars represent 90° (180°) rectangular pulses. The carriers are centered at 4.7, 56, and 119 ppm for ^{1}H, ^{13}C, and ^{15}N, respectively. Rectangular ^{1}H, ^{13}C, and ^{15}N pulses are applied with field strengths of 25, 17, and 6 kHz, respectively. The ^{1}H shaped 90° pulses have a rectangular profile (1.6 ms, water-selective). The ^{13}C shaped 180° pulses are ca-WURSTs, which sweep from low to high fields (filled shape: 400 μs; peak rf, 13.2 kHz; bandwidth, 32 kHz; open shape: 600 μs; peak rf, 10.3 kHz; bandwidth, 36 kHz). ^{15}N-decoupling is achieved with use of a 0.78 kHz GARP field, while ^{13}C-decoupling is achieved by using adiabatic CHIRP with a peak rf of 2.9 kHz on a 500 MHz machine. ^{13}CO-decoupling during t_{3a} uses WALTZ-16 with the seduce-1 shape of each element (centered at 178 ppm, peak field strength of 3.3 kHz). The delays used are $T = 2.1$ (1.91) ms for proteins with a correlation time <8 ns (>8 ns), $\delta = T - 1.7$ ms, $\tau_a = 1.8$ ms, $\tau_b = \tau_c - \tau_a$, $\tau_c = 2.2$ ms, $t_1 = 4 \times t_{1a} + 2 \times t_{1b}$ (t_{1b} was set to 0 in our experiments), $t_2 = 2t_{2a}$, $t_3(^{13}C) = 2t_{3b}$, $t_3(^{15}N) = 2t_{3a} + 2t_{3b}$. The durations and strengths of gradients are $g1$ = (1 ms, 15 G/cm), $g2$ = (0.2 ms, 10 G/cm), $g3$ = (0.2 ms, 17.5 G/c,), $g4$ = (0.2 ms, 32.5 G/cm), $g5$ = (0.2 ms, 25 G/cm), $g6$ = (2 ms, 20 G/cm), $g7$ = (1.5 ms, 25 G/cm), $g8$ = (0.5 ms, 22.5 G/cm), $g9$ = (0.4 ms, 25 G/cm); $g2-g5$ are rectangle-shaped gradients, while others are sine-shaped. Weak bipolar gradients (1.5 G/cm) are used during t_{3b}. The phase cycling employed is $\phi_1 = x$; $\phi_2 = x, -x$; $\phi_3 = 45°$; $\phi_4 = x, x, -x, -x$; $\phi_{ref} = x, -x, -x, x$. Quadrature detections in F_1, F_2, and F_3 are achieved by States-TPPI of ϕ_1, ϕ_2, and ϕ_4.

Table 1. Summary of Assignments and Structures of UB, LFABP, and DdCAD-1

	UB	LFABP	DdCAD-1
amide[a]	71/72	124/127	200/202
CH_n[a]	258/258	393/413	609/634
distance restraints[b]	160/209/177/ 320	722/412/123/ 322	1268/843/312/1016
dihedral restraints	102	170	212
backbone rmsd[c]	0.48 ± 0.08	1.07 ± 0.16	1.30 ± 0.16
heavy atom rmsd[c]	0.95 ± 0.07	1.46 ± 0.10	1.79 ± 0.15
ϕ/ψ space[d]	89.4/10.4/0.2/0.0	84.8/12.7/1.8/0.7	70.5/25.3/3.2/1.0

[a] Assigned/expected. [b] Intraresidue/sequential/medium range/long range. [c] To mean structure; all residues were used in the rmsd calculation except for UB (1−70). [d] Most favored/additionally allowed/generally allowed/disallowed.

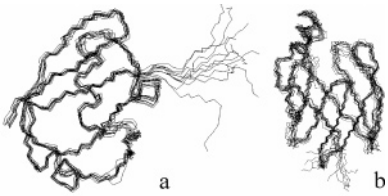

Figure 2. Superimposition of the 10 lowest-energy structures of UB (a) and LFABP (b).

Assignments of NOEs from 4D spectra were found to be easier than from 3D ^{15}N- and ^{13}C-edited spectra due to reduced ambiguities in 4D NOESY, especially for medium-sized proteins such as DdCAD-1. Many long-range NOEs were assigned with only chemical shifts without referring to initial structures. An initial structure with a backbone rmsd <2.5 Å was observed for each protein tested here. With the initial structures, more long-range NOE peaks, including some weak peaks, were assigned in an iterative manner. All of the final structures are well-defined (Table 1, Figure 2) and are very similar to those determined previously.[7,8]

The results demonstrate that the strategy introduced here can be used as an alternative to traditional approaches for small and medium-sized proteins. The total experimental time for our strategy can be significantly shorter than that for the traditional methods. In addition, our strategy can facilitate NOE assignments by using 4D instead of 3D NOESY, especially for proteins with >150 residues. Rapid data collection and easy NOE assignment will open an avenue to rapid protein structure determination and will also allow us to determine the structures of proteins with low stability in a more cost-effective way. One disadvantage of this method is the lower digital resolution of the 4D NOESY caused by the limited data points in the indirect dimensions. This drawback can be overcome by using a sparse 4D MDD-NOESY.[3c] In the case where the 4D NOESY does not yield enough long-range NOEs due to insufficient sensitivity, one can record a more sensitive 3D time-shared $^{13}C/^{15}N$-edited NOESY. With the 3D NOESY, more long-range NOEs can be assigned based on the structure obtained from the 4D NOESY experiment. Resonance assignments can be expedited using semiautomatic software, which is in the process of being developed by our group.

Acknowledgment. This research was supported by a grant from the Biomedical Research Council (BMRC), and Agency for Science, Technology and Research, A*Star of Singapore. D.Y. is the recipient of a BMRC Young Investigator Award. The authors thank Dr. Lin Zhi for preparing the DdCAD-1 sample.

Supporting Information Available: One table showing the experimental and processing details. One figure showing representative 3D and 4D slices and illustrating the assignment strategy. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Atreya, H. S.; Szyperski, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9642.

(2) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2003**, *125*, 13958.

(3) (a) Rovnyak, D.; Frueh, D. P.; Sastry, M.; Sun, Z.-Y. J.; Stern, A. S.; Hoch, J. C.; Wagner, G. *J. Magn. Reson.* **2004**, *170*, 15. (b) Jaravine, V.; Ibraghimov, I.; Orekhov, V. Y. *Nat. Methods* **2006**, *3*, 605. (c) Luan, T.; Jaravine, V.; Yee, A.; Arrowsmith, C. H.; Orekhov, V. Y. *J. Biomol. NMR* **2005**, *33*, 1.

(4) (a) Farmer, B. T.; Mueller, L. *J. Biomol. NMR* **1994**, *4*, 673. (b) Pascal, S. M.; Muhandiram, D. R.; Yamazaki, T.; Forman-Kay, J. D.; Kay, L. E. *J. Magn. Reson.* **1994**, *103*, 197. (c) Xia, Y. L.; Yee, A.; Arrowsmith, C. H.; Gao, X. L. *J. Biomol. NMR* **2003**, *27*, 193. (d) Frueh, D. P.; Vosburg, D. A.; Walsh, C. T.; Wagner, G. *J. Biomol. NMR* **2006**, *34*, 31. (e) Lin, Z.; Xu, Y.; Yang, S.; Yang, D. W. *Angew. Chem., Int. Ed.* **2006**, *45*, 1960.

(5) Xu, Y. Q.; Zheng, Y.; Fan, J.-S.; Yang, D. W. *Nat. Methods* **2006**, *3*, 931.

(6) Chen, J.-H.; Cutting, B. *J. Chem. Phys.* **2000**, *112*, 6511.

(7) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.

(8) Lin, Z.; Sriskanthadevan, S.; Huang, H.; Siu, C.-H.; Yang, D. W. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1016.

JA071442E