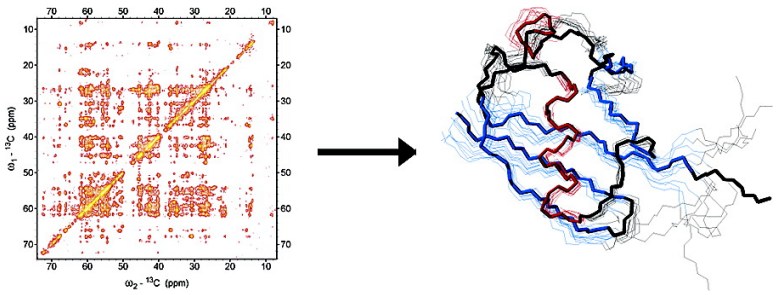
Article

Protein Structure Determination by High-Resolution Solid-State NMR Spectroscopy: Application to Microcrystalline Ubiquitin

Stephan G. Zech, A. Joshua Wand, and Ann E. McDermott

J. Am. Chem. Soc., **2005**, 127 (24), 8618-8626• DOI: 10.1021/ja0503128 • Publication Date (Web): 28 May 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 41 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 05/28/2005

Protein Structure Determination by High-Resolution Solid-State NMR Spectroscopy: Application to Microcrystalline Ubiquitin

Stephan G. Zech,*,†,§ A. Joshua Wand,‡ and Ann E. McDermott*,†

Contribution from the Department of Chemistry, Columbia University, 3000 Broadway Mail Code 3113, New York, New York 10027, and Department of Biochemistry and Biophysics, University of Pennsylvania, The Johnson Research Foundation, Philadelphia, Pennsylvania 19104

Received January 17, 2005; E-mail: Stephan.Zech@web.de; aem5@columbia.edu.

Abstract: High-resolution solid-state NMR spectroscopy has become a promising method for the determination of three-dimensional protein structures for systems which are difficult to crystallize or exhibit low solubility. Here we describe the structure determination of microcrystalline ubiquitin using 2D $^{13}C-^{13}C$ correlation spectroscopy under magic angle spinning conditions. High-resolution ^{13}C spectra have been acquired from hydrated microcrystals of site-directed ^{13}C-enriched ubiquitin. Interresidue carbon−carbon distance constraints defining the global protein structure have been evaluated from 'dipolar-assisted rotational resonance' experiments recorded at various mixing times. Additional constraints on the backbone torsion angles have been derived from chemical shift analysis. Using both distance and dihedral angle constraints, the structure of microcrystalline ubiquitin has been refined to a root-mean-square deviation of about 1 Å. The structure determination strategies for solid samples described herein are likely to be generally applicable to many proteins that cannot be studied by X-ray crystallography or solution NMR spectroscopy.

Introduction

Over the past few years, there have been remarkable advances in solid-state NMR (ssNMR) experiments for characterization of protein structure and function. In a wide variety of systems the protein's insolubility made X-ray crystallography or solution NMR unsuitable, while questions on structure and dynamics can be addressed with ssNMR.[1−5] These efforts typically involve structurally homogeneous samples and utilize recently developed pulse sequences for sequential correlation of resonances, detection of tertiary contacts, and characterization of torsion angles.

Excellent NMR line widths can be achieved for microcrystalline or precipitated hydrated globular systems using magic angle spinning (MAS) methods. More recent advances in high-field instrumentation and pulse sequences for chemical shift correlation experiments led to more efficient methods of assigning solid-state proteins with extensive isotopic enrichment and a rapid succession of studies on small globular proteins has become evident, including BPTI,[6] the α-spectrin SH3 domain,[7,8,9] the catabolite repression phosphocarrier protein (Crh),[10] human ubiquitin,[11,12] thioredoxin,[13] the immunoglobulin binding domain β1 of streptococcal protein G (GB1),[14] and peptides such as neurotensin,[15] mastoparan-X,[16] and a fibrillar peptide fragment of transthyretin.[17]

For tertiary structure determination, or fold definition, long-range constraints have typically served a critical role. It has been recently demonstrated that simple spin diffusion experiments, if combined with strategic labeling schemes, are sufficient to determine a moderate resolution structure of a protein. This was first shown for the α-spectrin SH3 domain using site-directed selective carbon labeling schemes and a 'proton-driven spin diffusion' (PDSD) pulse sequence.[8]

† Columbia University.
‡ University of Pennsylvania.
§ Present address: EPIX Pharmaceuticals, Inc., 67 Rogers St., Cambridge, MA 02142.

(1) Thompson, L. K. *Curr. Opin. Struct. Biol.* **2002**, *12*, 661−669.
(2) Williamson, P. F.; Ernst, M.; Meier, B. H. MAS solid-state NMR of isotopically enriched biological samples. In *BioNMR in Drug Research*; Zerbe, O., Ed.; Wiley-VCH: Weinheim, 2003; pp 243−282.
(3) Opella, S. J.; Marassi, F. M. *Chem. Rev.* **2004**, *104*, 3587−3606.
(4) Tycko, R. *Curr. Opin. Struct. Biol.* **2004**, *14*, 96−103.
(5) McDermott, A. E. *Curr. Opin. Struct. Biol.* **2004**, *14*, 554−561.
(6) McDermott, A.; Polenova, T.; Bockmann, A.; Zilm, K. W.; Paulson, E. K.; Martin, R. W.; Montellione, G. T. *J. Biomol. NMR* **2000**, *16*, 209−219.
(7) Pauli, J.; Baldus, M.; van Rossum, B.; de Groot, H.; Oschkinat, H. *Chembiochem.* **2001**, *2*, 272−281.
(8) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98−102.
(9) Castellani, F.; van Rossum, B.; Diehl, A.; Rehbein, K.; Oschkinat, H. *Biochemistry* **2003**, *42*, 11476−11483.
(10) Böckmann, A.; Lange, A.; Galinier, A.; Luca, S.; Giraud, N.; Juy, M.; Heise, H.; Montserret, R.; Penin, F.; Baldus, M. *J. Biomol. NMR* **2003**, *27*, 323−339.
(11) Igumenova, T.; Wand, A. J.; McDermott, A. *J. Am. Chem. Soc.* **2004**, *126*, 5323−5331.
(12) Igumenova, T. I.; McDermott, A. E.; Zilm, K. W.; Martin, R. W.; Paulson, E. K.; Wand, A. J. *J. Am. Chem. Soc.* **2004**, *126*, 6720−6727.
(13) Marulanda, D.; Tasayco, M. L.; McDermott, A.; Cataldi, M.; Arriaran, V.; Polenova, T. *J. Am. Chem. Soc.* **2004**, *125*, 16608−16620.
(14) Franks, W. T.; Zhou, D. H.; Money, B. G.; Graesser, D. T.; Sahota, G.; Rienstra, C. M. *J. Am. Chem. Soc.* **2005**, in press.
(15) Luca, S.; White, J. F.; Sohal, A. K.; Filippov, D. V.; van Boom, J. H.; Grisshammer, R.; Baldus, M. *Proc. Natl. Acad. Sci., U.S.A.* **2003**, *100*, 10706−10711.
(16) Fujiwara, T.; Todokoro, Y.; Yanagishita, H.; Tawarayama, M.; Kohno, T.; Wakamatsu, K.; Akutsu, H. *J. Biomol. NMR* **2004**, *28*, 311−325.
(17) Jaroniec, C. P.; MacPhee, C.; Bajaj, V. S.; McMahon, M. T.; Dobson, C. M.; Griffin, R. G. *Proc. Natl. Acad. Sci., U.S.A.* **2004**, *101*, 711−716.

10.1021/ja0503128 CCC: $30.25

Advances in fast MAS-based magnetization transfer schemes allow for directed, highly efficient homonuclear ($^{13}C-^{13}C$) recoupling during the mixing times of chemical shift correlation spectra.[2] Several homonuclear methods have been developed recently which achieve broadband (i.e. nonspectrally selective) efficient recoupling. The '$^{1}H-^{13}C$ dipolar-assisted rotational resonance' (DARR) experiment[18] or the related 'radiofrequency field-assisted diffusion' sequence[19] have been shown to exhibit an excellent efficiency in recoupling of distant carbon atoms.

In this study we use the DARR carbon−carbon correlation experiments on selectively ^{13}C-enriched microcrystals of human ubiquitin to obtain a large number of interresidue constraints defining secondary structure and global fold of the protein. In line with previous studies,[11,12,20,21] hydrated ubiquitin microcrystals give rise to highly resolved ssNMR spectra, facilitating site-specific assignment of resonances. The line width can be further reduced if selective ^{13}C labeling schemes are involved.[22,8]

Furthermore, significant information is available from analysis of chemical shifts alone. Cα and Cβ isotropic shifts provide information on the secondary structure context. This information can be crucial in defining the fold of a protein or validating the sample conditions. Here we apply the database provided by the program TALOS[23] to obtain backbone dihedral angle restraints. Together with the distance restraints gathered from $^{13}C-^{13}C$ correlation spectroscopy, the 3D structure of ubiquitin is refined to a great level of molecular detail.

Materials and Methods

Sample Preparation. Uniformly [^{13}C,^{15}N] labeled human ubiquitin (referred to as [U-^{13}C]-Ubq) has been prepared as described previously.[24] Selectively ^{13}C-enriched and uniformly ^{15}N-enriched ubiquitin largely devoid of adjacent ^{13}C nuclei (referred to as [2-^{13}C]-Ubq) has been isolated from *E. coli* grown on [2-^{13}C]-glycerol (Cambridge Isotopes) as carbon source and $^{15}NH_4Cl$ as nitrogen source.[25] In contrast to the original paper, the *E. coli* strain used here did not lack succinate dehydrogenase and malate dehydrogenase. Furthermore, $NaH^{13}CO_3$ has not been added to the growth medium in our study. Isolation and purification of the protein has been done as described previously.[24]

About 8 mg of lyophilized ubiquitin has been dissolved to 25 mg/mL in a sodium citrate buffer (20 mM, pH 4.1) and crystallized as described previously[11] by slow addition of 2-methyl-2,4-pentanediol (~60% final volume). A white precipitate appeared within minutes and showed needle-shaped microcrystals under the microscope. The emulsion has been stored overnight at 4 °C and centrifuged at 10^4 rpm for 20 min. Most of the supernatant has been removed, and the hydrated solid (ca. 25−30 mg) has been filled into a 4 mm Bruker MAS rotor. The sample has been confined to the center of the rotor using Teflon spacers. For the diluted sample, 6 mg of [2-^{13}C]-Ubq and 12 mg of isotopically unenriched ubiquitin (Aldrich) have been cocrystallized, yielding about 55 mg of hydrated microcrystals. All rotors have been stored at −80 °C until use with no changes in the spectra observable for various experiments recorded over a period of several months.

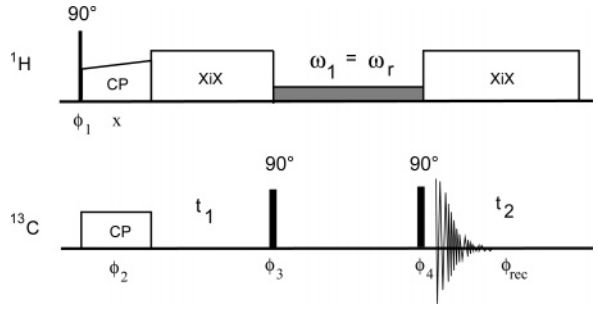


Figure 1. Pulse sequence for carbon−carbon correlation experiments using the DARR scheme for magnetization transfer during the mixing time. The parameters for RF field strengths and pulse durations are described in the text. The following phase cycle has been used: $\phi_1 = (y\bar{y})_8$, $\phi_2 = (xx\bar{x}\bar{x})_4$, $\phi_3 = (y_4\bar{y}_4)_2$, $\phi_4 = x_8\bar{x}_8$, $\phi_{rec.} = x\bar{x}\bar{x}x\ (\bar{x}xx\bar{x})_2\ x\bar{x}\bar{x}x$.

NMR Experiments. All experiments have been performed on a Bruker Avance DRX-750 spectrometer, operating at 750.22 MHz proton and 188.65 MHz carbon frequency. A double resonance (H−F/X) widebore MAS probe with 4.0 mm rotor diameter has been used. The sample has been cooled by a flow of dry air (ca. 2000 L/h) with an inlet temperature between 233 and 245 K as indicated in the figure captions. The spinning frequency has been adjusted to 10.0 kHz. 1D CP-MAS experiments have been acquired with a ramped RF field (80−100%) on the proton channel, applying an RF field strength of 76 kHz on the proton and 45 kHz on the carbon channel during the cross-polarization time of 2.0 ms. The proton RF field has been adjusted to 100 kHz during the initial 90° pulse and 95 kHz during the acquisition time of about 25 ms using the XiX sequence for proton decoupling.[26]

For the 2D carbon−carbon correlation spectra, the DARR pulse sequence[18] as depicted in Figure 1 has been used. RF power levels have been identical to 1D CP-MAS experiments with a RF strength of 50 kHz for the 90° pulses on the carbon channel. During the mixing time (10−500 ms), the RF field on the proton channel matched the spinning speed of 10 kHz. The spectral width has been set to 377 ppm in ω_2 and 150 ppm in ω_1 with the transmitter frequency at 75 ppm. Typically 350 points in t_1 with 80−240 averages per point have been acquired. The TPPI method has been used for phase sensitive detection in t_1 by incrementing ϕ_2.[27] The total acquisition time has been about 10 min for 1D spectra and between 20 and 72 h for the 2D data sets.

All data have been processed with NMRpipe.[28] A phase-shifted sinebell function has been used for apodization before zero filling to 2048 points in t_2 and 1024 in t_1 prior to FT in each dimension. The magnetic field has been referenced to DSS, using the ^{13}C methylene peak in solid adamantane as an external standard.[29] Analysis and assignment of the 2D data sets has been carried out using Sparky version 3.1.[30]

Structure Calculations. The structure of ubiquitin has been calculated using CNS version 1.1[31] by applying a molecular dynamics simulated annealing protocol with torsion angles as internal degrees of freedom. All parameters were identical to the default protocol (see 'NMR Tutorials' section of CNS) which involves a high-temperature annealing at 50 000 K for 1000 steps with 15 ps per step and a subsequent cooling stage with decreasing temperature from 50 000 K to 0 K in 1000 steps. Experimental carbon−carbon distance constraints and backbone dihedral angle restraints obtained by TALOS[23] have been

(18) Takegoshi, K.; Nakamura, S.; Terao, T. *Chem. Phys. Lett.* **2001**, *344*, 631−637.
(19) Morcombe, C. R.; Gaponenko, V.; Byrd, R. A.; Zilm, K. W. *J. Am. Chem. Soc.* **2004**, *126*, 7196−7197.
(20) Paulson, E. K.; Morcombe, C. R.; Gaponenko, V.; Dancheck, B.; Byrd, R. A.; Zilm, K. W. *J. Am. Chem. Soc.* **2003**, *125*, 14222−14223.
(21) Paulson, E. K.; Morcombe, C. R.; Gaponenko, V.; Dancheck, B.; Byrd, R. A.; Zilm, K. W. *J. Am. Chem. Soc.* **2003**, *125*, 15831−15836.
(22) Hong, M. *J. Magn. Reson.* **1999**, *139*, 389−401.
(23) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289−302.
(24) Wand, A. J.; Urbauer, J. L.; McEvoy, R. P.; Bieber, R. J. *Biochemistry* **1996**, *35*, 6116−6125.
(25) LeMaster, D. M.; Kushlan, D. M. *J. Am. Chem. Soc.* **1996**, *118*, 9255−9264.
(26) Detken, A.; Hardy, E. H.; Ernst, M.; Meier, B. H. *Chem. Phys. Lett.* **2002**, *356*, 298−304.
(27) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. *Protein NMR Spectroscopy − Principles and Practise*; Academic Press: San Diego, 1996.
(28) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(29) Morcombe, C.; Zilm, K. W. *J. Magn. Reson.* **2003**, *162*, 479−486.
(30) Goddard, T. D.; Kneller, D. G. *Sparky 3.106*; University of California, San Francisco, 2002.
(31) Brunger, A.; Adams, P.; Clore, G.; Delano, W.; Gros, P.; Grosse-Kunstleve, R.; Liang, J.-S.; Kuszewski, J.; Nilges, N.; Pannu, N.; Read, R.; Rice, L.; Simonson, T.; Warren, G. *Acta Crystallogr. D* **1998**, *54*, 905−921.

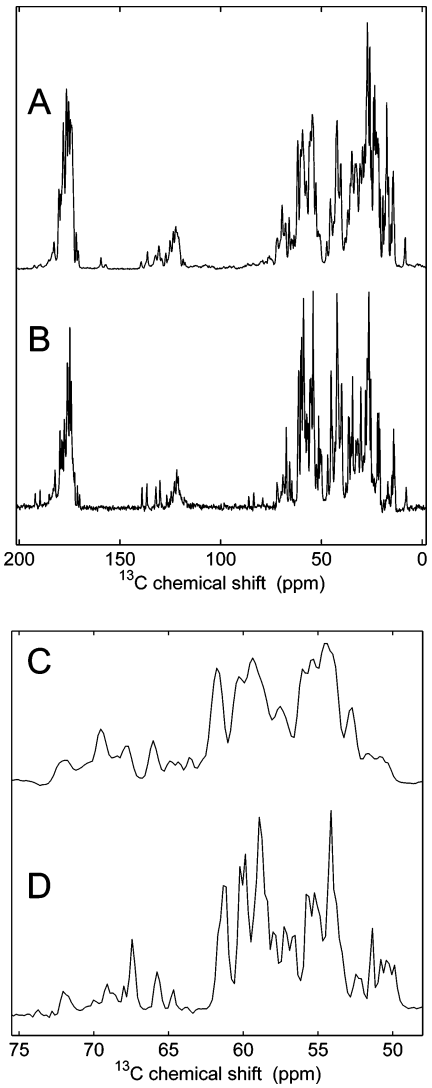


Figure 2. 1D ^{13}C-CP-MAS spectra acquired at 245 K from ubiquitin microcrystals with uniform ^{13}C enrichment (A) and with selective ^{13}C enrichment (B) obtained with [2-^{13}C]-glycerol as carbon source. Experimental conditions are described in the text. Panels C and D show the expanded aliphatic region of the spectra in A and B, respectively. Spectrum C exhibits considerable narrower lines for most Cα atoms due to the absence of one-bond ^{13}C−^{13}C couplings in the selective labeling scheme.

used as input for the structure refinement. All structures have been energy minimized and the 10 lowest energy structures have been selected and analyzed.

Results and Discussion

Improved Resolution in Sparsely Labeled Samples. Figure 2 compares the 1D CP-MAS spectra of [U-^{13}C]-Ubq (A) and [2-^{13}C]-Ubq (B). The spectrum for [U-^{13}C]-Ubq is almost identical to the one obtained previously for a similar preparation of microcrystals recorded under different experimental conditions.[12] The ^{13}C line width of this sample can be estimated to be in the 0.5−0.6 ppm range. The spectrum of the [2-^{13}C]-Ubq sample (Figure 2B) shows reduced signal intensities in the methyl group region (5−25 ppm) and in the carbonyl region (∼174 ppm) due to the low amount of ^{13}C enrichment predicted for this labeling scheme. More importantly, the spectral resolution of the [2-^{13}C]-Ubq sample is considerably improved mainly due to removal of one-bond *J* and *D* couplings.[8,25] This is easily seen by comparison of the Cα region of both spectra (Figure 2C and D). The labeling scheme predicts a high degree of ^{13}C enrichment in the Cα atoms of most amino acids (except Glu, Arg, Pro, Leu), whereas the C*β* and carbonyl atoms are expected to have low ^{13}C content. Due to the resulting absence of *J* and *D* couplings, the line width of the [2-^{13}C]-Ubq sample decreases to about 0.2−0.3 ppm (Figure 2D).

Identification of Mobile Residues at Low Temperature. Homonuclear ^{13}C chemical shift correlation spectra of microcrystals of the [U-^{13}C]-Ubq sample have been measured at 233 and 245 K with mixing times of 10, 15, and 20 ms. Figure 3A shows the spectrum of the [U-^{13}C]-Ubq sample at 233 K with a mixing time of 20 ms. It turned out that the spectra taken between 233 and 245 K exhibit many peaks that have not been observed in the previous studies which had been conducted at higher temperatures (∼278−288 K).[12] The ^{13}C line width in the 2D spectra, however, appears to be nearly independent of temperature within the range investigated. This indicates that the temperature dependent observation of some residues is influenced by their mobility rather than due to disorder or sample inhomogeneity.

Identification of the spin system of the additional resonances has been achieved by a side chain walk as described previously.[6,7,12] Site-specific assignments have been initially guided by solution NMR shifts and subsequently corroborated utilizing sequential two- and three-bond couplings observed at 20 ms mixing time (Figure 3A). For most residues, the site specific assignment has been further confirmed by sequential contacts in the DARR spectra of the [2-^{13}C]-Ubq sample described below. Thereby, several residues which have been missing in the previous experiments[12] could be identified. Those comprise Leu8(C_α, C_β, $C_{\delta 1}$, CO), Thr9(C_α, C_β, $C_{\gamma 2}$), Gly10(C_α, CO), and Lys11(C_α, C_β, C_γ, CO) located in a mobile loop region as well as Arg72(C_α, C_β, C_δ) and Leu73(C_α, C_β, C_γ) near the flexible C-terminus of the protein. For many residues for which only partial assignment has been obtained previously,[12] more atoms comprising the spin system could be identified (e.g. for Glu24, Glu34, Arg42, Gln49, Asp52).

A likely reason for the temperature dependence is that the initial proton−carbon magnetization transfer is not efficient for mobile residues. For example, the rate of relaxation or dissipation of proton magnetization, which can be measured through the line intensity in the ^{13}C spectrum after CP, is controlled by thermal motion and can therefore be used to probe the mobility of individual regions within a complex structure. This demonstrates that the experimental condition, and in particular the temperature, have a profound influence on the number of observable amino acids especially for mobile residues. While lower temperatures are more demanding on the long time spectrometer stability and potentially lead to line broadening,[32] experiments over a range of temperatures might be required to identify amino acids in mobile regions of the protein.

Observation of Long-Range Contacts for Selectively ^{13}C-Labeled Ubiquitin. Structure determination by NMR spectroscopy relies to a large degree on the observation of long-range contacts which provide a large number of interresidue restraints. While with solution state NMR methods, mainly proton−proton restraints are obtained, for the first 3D protein structure determination based on MAS solid-state NMR techniques predominantly carbon−carbon contacts and a number of ^{15}N−

(32) Martin, R. W.; Zilm, K. W. *J. Magn. Reson.* **2003**, *165*, 162−174.

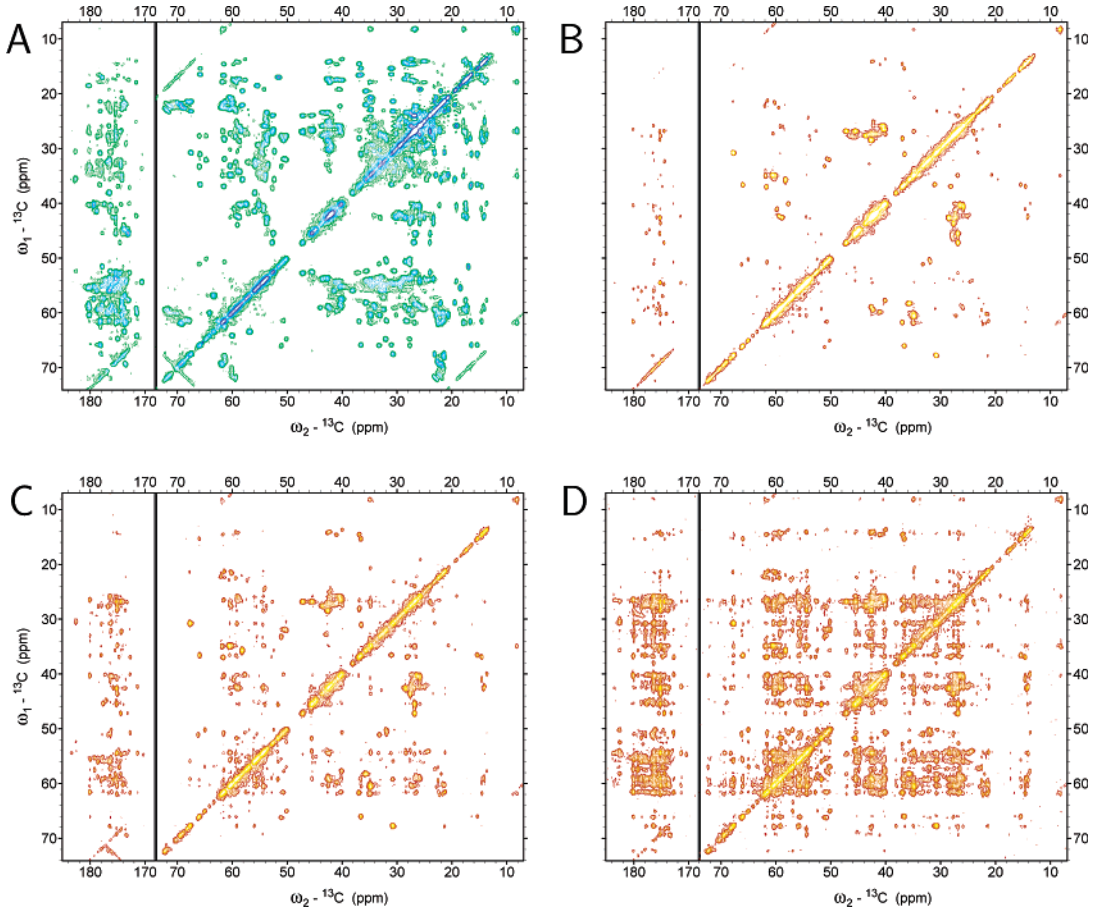


Figure 3. Aliphatic and carbonyl regions of the ^{13}C−^{13}C DARR spectra of uniformly labeled microcrystalline ubiquitin (A) and selectively labeled [2-^{13}C]-Ubq (B−D). Spectrum A has been obtained at 233 K with a mixing time of 20 ms and shows mainly one-bond intraresidue correlations. Spectra B−D have been measured at 245 K with mixing times of 50, 200, and 500 ms, respectively. With increasing mixing time, a larger number of long-range correlations is observed.

^{15}N contacts have been used for structure refinement.[8] To recouple distant carbon atoms in ubiquitin, the DARR experiment has been applied to the [2-^{13}C]-Ubq sample. Since in this sample, adjacent ^{13}C atoms are present only for few residues, dominant one-bond dipolar couplings are sufficiently suppressed and dipolar truncation effects are reduced.[8] The ability of the DARR sequence to efficiently recouple distant carbon atoms (up to ∼6 Å) has been recently demonstrated on sparsely labeled samples of the anti-apoptotic protein Bcl-xL (∼20 kDa)[33] and for extensively ^{2}H and uniformly ^{13}C-enriched samples of ubiquitin.[19]

Various DARR spectra of [2-^{13}C]-Ubq have been obtained at 245 K with mixing times ranging between 10 ms and 500 ms. Figure 3B−D show the spectra obtained at 50 ms (B), 200 ms (C), and 500 ms (D). The removal of one-bond couplings due to site-directed selective labeling of the [2-^{13}C]-Ubq sample is obvious by comparison of Figure 3A and B. Since at short mixing times (∼10 ms) mainly one-bond carbon−carbon correlations are observed, the DARR spectrum of [2-^{13}C]-Ubq appears greatly simplified compared to the [U-^{13}C]-Ubq sample. In line with the proposed labeling scheme,[8,25] the peaks in th aliphatic region can be assigned to intraresidue correlation for Val, Leu and Ile which exhibit adjacent ^{13}C-enriched atoms.

Comparison of the carbonyl peak integrals shown Figure 2A and B reveal an amount of ^{13}C enrichment in the carbonyl groups of the [2-^{13}C]-Ubq sample that is roughly in agreement with the proposed labeling scheme.[8,25] However, closer inspection of Figures 3B−D shows that carbon atoms without proposed ^{13}C enrichment[25] are observed at mixing times ≥50 ms. This suggests that the ^{13}C enrichment present in our sample is not identical to the one proposed previously.[25] A likely explanation for this might be the use of a different strain of *E. coli* and the lack of ^{13}C-enriched sodium carbonate which leads to increased scrambling of ^{13}C labels. As shown below, the larger ^{13}C enrichment for carbonyl groups in some amino acid types (Asn, Thr, Lys) allows us to obtain significant structural information also from the carbonyl atoms in this sample.

However, the absence of dominant Cα-CO correlations at short mixing times indicates a very low degree of *simultaneous* labeling of Cα and CO. In fact, the peaks observed in the Cα-CO region in the spectra at mixing times of 10 ms (not shown) and 50 ms (Figure 3B) are primarily assigned to sequential contacts (∼2.5 Å). These can be used to confirm the site-specific assignments which have been obtained previously by 3D ^{15}N−^{13}C−^{13}C correlation experiments.[11] Apart from the CO groups, fractional ^{13}C enrichment is also found for the methyl groups of Val, Leu, and Ile(Cγ2).

At larger mixing times of 200 and 500 ms (Figure 3C−D), an increasing number of correlations are observed due to recoupling of distant carbons. More than 700 peaks can be resolved in the 500 ms spectrum. A summary of connectivities observed as function of the mixing time is given in the

(33) Zech, S. G.; Olejniczak, E.; Hajduk, P.; Mack, J.; McDermott, A. E. *J. Am. Chem. Soc.* **2004**, *126*, 13948−13956.

Supporting Information as Supplemental Figure 2. In principle, from the analysis of the cross-peak buildup curves (i.e. peak volume as function of the mixing time) it is possible to obtain (semi-)quantitative distance information. However, many problems such as fractional labeling and peak overlap have to be overcome to derive reliable information from buildup curves. An alternative strategy for assignment of the long-range peaks and distance classification is given below with the resulting structure refinement based on the interresidue constraints.

Comparison of DARR and PDSD Experiments. In the structure determination of the α-spectrin SH3 domain, the simple PDSD experiment has been very successful in the observation of long-range distance correlation up to about 7 Å.[8] This experiment has lenient power requirements for recoupling carbon atoms, i.e., no RF irradiation is applied during the mixing time. In contrast, other sequences such as RFDR[34] require high-power proton decoupling during the mixing time and cause sample heating and potentially degradation if small dipolar couplings need to be observed.

In $^{13}C-^{13}C$ spin diffusion experiments, the two major factors governing the magnetization transfer are the dipolar coupling between the carbon atoms and the spectral overlap between the spins involved. One way of increasing the spectral overlap between ^{13}C spins is to recouple the $^{1}H-^{13}C$ dipolar interaction which then assists the magnetization exchange between carbon atoms. This can be facilitated by applying a RF field on the proton channel which is matched with multiples of the spinning speed (ω_r or $2\omega_r$).[18] Since an efficient magnetization transfer is crucial for the detection of small dipolar couplings between distant spins, it is important to compare the performance of the DARR sequence with the previously used PDSD experiment.

Figure 4 shows spectral regions for the DARR experiment (blue) overlayed with the PDSD experiment (magenta) both measured at 500 ms mixing time. The two spectra are shown with the same contour levels and have been measured back-to-back under the same experimental conditions. They differ only with respect to the proton irradiation during the mixing time which has been matched with the spinning speed in case for the DARR experiment ($\omega_1^H = \omega_r = 2\pi \times 10$ kHz) but has been omitted in the PDSD experiment ($\omega_1^H = 0$). As expected, both experiments give very similar results with respect to the peak positions and intensities of the diagonal peaks. However, for many cross-peaks, the intensity is stronger in the DARR experiment than for the PDSD experiment. Peaks with considerable difference in volume between the two experiments have been assigned in Figure 4.

The distances in the X-ray structure (PBD entry 1ubq[35]) between the assigned carbon atoms in Figure 4 range between 4.5 Å (Ser20-Cα to Pro19-Cβ) to 6.7 Å (Val70-Cβ to Leu71-Cγ). Distances in that range are very important for the identification of tertiary contacts and secondary structure elements. In fact, at 500 ms mixing the tertiary contact between Ala28-Cα and Pro38Cβ (5.5 Å) is much more pronounced under DARR conditions and is already present at 200 ms mixing time. Many other tertiary contacts have been identified only in the DARR experiment, but are not present under PDSD conditions. We therefore conclude that under the experimental conditions

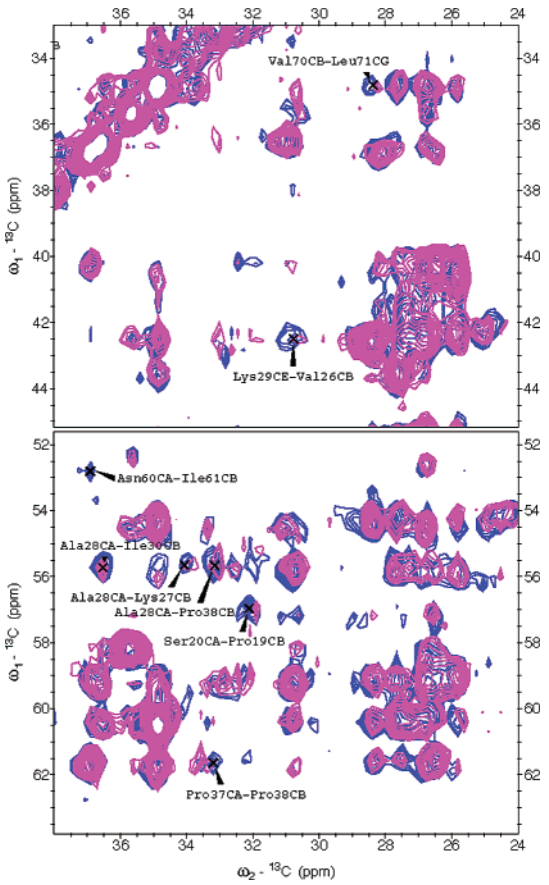


Figure 4. Comparison of magnetization transfer efficiency for the DARR experiment (blue) with the PDSD experiment (magenta). Both spectra are measured on [2-^{13}C]-Ubq microcrystals at 245 K under identical conditions, except for the ^{1}H-RF field applied during the 500 ms mixing time which was 10 kHz for the DARR and 0 for the PDSD experiment. Both spectra have been processed identically and are shown with the same contour levels. Peaks with considerable difference in volume are shown with assignment.

chosen here, the DARR experiment has a better performance in recoupling distant carbon atoms.

It is important to note that simply increasing the mixing time in a PDSD experiment beyond 500 ms is unlikely to yield the same correlations than observed with the DARR sequence at 500 ms, because both experiments behave similarly with respect to the relaxation parameters for ^{13}C spins. Therefore, a PDSD experiment with longer mixing time would yield significantly lower overall signal intensity with many important correlation remaining below the S/N level. Furthermore, extensive mixing times would diffuse the spin magnetization randomly over larger distances, but not necessarily reveal more structurally important long-range correlations.

The improved recoupling efficiency of the DARR experiment is expected to become even more pronounced at higher spinning speeds where the $^{1}H-^{1}H$ dipolar interactions are averaged out more efficiently. However, since only low power ^{1}H irradiation is required during the mixing time, sample heating is not a severe problem for the DARR experiment, making it also a practical choice for temperature sensitive protein samples, even if long mixing times are required.

Assignment Strategy for Long-Range Correlations. With the assignments obtained on the [U-^{13}C]-Ubq sample, it is straightforward to assign the spectrum of the [2-^{13}C]-Ubq sample at short mixing times (~10 ms). Since [2-^{13}C]-Ubq contains

(34) Bennett, A. E.; Ok, J. H.; Griffin, R. G.; Vega, S. *J. Chem. Phys.* **1992**, *96*, 8624−8627.
(35) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.

only a small number of residues with adjacent ^{13}C-enriched atoms, only intraresidue peaks are observed at this mixing time, including e.g. Leu Cβ-Cγ, Val Cβ-Cγ, and Ile Cα-Cβ. In addition, a number of two-bond couplings is already observed at 10 ms mixing time. The 50 ms spectrum shown in Figure 3B allowed us to identify most of the amino acids on the basis to their chemical shifts due to intraresidue (one to three bond) couplings between highly ^{13}C-enriched atoms. The 50 ms spectrum exhibits also sequential contacts (see supplemental Figure 2 in Supporting Information), particularly in the Cα-CO region which are used to confirm site specific assignment for the residues observed only at low temperature (see above).

A statistical analysis of chemical shifts obtained at mixing times of 10 and 50 ms revealed that for the majority of ^{13}C atoms the chemical shifts are very well defined with rms deviations in the order of 0.03−0.08 ppm. This has been confirmed for various samples prepared from [2-^{13}C]-Ubq and hence demonstrates the reproducibility of the preparation protocol. For some atoms, however, a deviation of the ^{13}C chemical shift observed for the [2-^{13}C]-Ubq sample compared to the uniformly labeled sample has been observed and is interpreted as residual dipolar shift due to the presence of one bond ^{13}C−^{13}C couplings in [U-^{13}C]-Ubq.

The intraresidue and sequential assignments obtained at 10 and 50 ms have been copied to the 200 and 500 ms spectra. For the assignment of the remaining peaks in the long mixing time spectra, we used the following guidelines: (i) The agreement of the chemical shifts must be better than $\pm$0.1 ppm, which is range slightly larger than the average rms deviation of the ^{13}C atoms and also closely resembles the line width for the [2-^{13}C]-Ubq sample. (ii) Only atoms which are labeled to high degree in [2-^{13}C]-Ubq are considered. (iii) Since sequential contacts are usually shorter than medium or long-range contacts, preference is given to the assignment of sequential peaks over a medium or long-range contacts if both fulfill points i and ii.

This strategy allowed us to assign about 122 interresidue peaks in the 500 ms spectrum unambiguously. Medium and long range distances are depicted in supplemental Figure 1 in Supporting Information. It turns out, that more than half of the remaining peaks can be assigned unambiguously in one dimension. In many cases, the ambiguity in the second dimension can be addressed by an analysis of the secondary structure elements which are obtained separately e.g. from secondary chemical shifts analysis. Furthermore, the assignment can be assisted by 'network anchoring techniques' which evaluate the self-consistency of long-range assignment independently of knowledge of the three-dimensional structure.[36,37] Contacts between secondary structure elements which are close in the 3D structure can thus be identified by generating a connectivity matrix as shown in supplemental Figure 2 in Supporting Information.

Moreover, paralleling many solution NMR structure refinement procedures, ambiguities can be resolved by taking distance information from a preliminary three-dimensional structure model into account.[36−38] The peak list obtained so far contained already a significant number of unambiguous medium (35) and long-range (28) contacts defining the global fold of the protein. These are depicted in supplemental Figure 1 in Supporting Information. We decided to use these restraints in a structure calculation with the program CNS (see below for details). This preliminary structure already showed the correct fold of the protein within secondary structure elements, although many regions were flexible due to lack of restraints. Nonetheless, most of the one-dimensional ambiguities and many two-dimensional could be resolved with the help of the preliminary structure obtained in the first refinement step. A very similar strategy for resolving ambiguities is used by the programs ARIA[38] or CYANA[36] in an more elegant, automated fashion. This iterative process has been repeated twice with the structure showing a great improvement in the backbone rmsd values. For the last cycle, also atoms with lower ^{13}C enrichment such as the CO groups have been taken into account. Thereby, it was possible to assign the wast majority of peaks in the 200, 300 and 500 ms spectra. A total of 336 unambiguous interresidue contacts have been identified and were used for the subsequent structure refinement.

A critical milestone for this procedure is the identification of a sufficient set of unambiguous long-range correlations defining the global protein fold. If the initial assignment contains false constraints, the result of successive cycles could be biased by the erroneous global fold of structures generated in the first calculation. In this respect, the novel PASD algorithm[39] recently implemented in the structure determination package Xplor-NIH,[40] provides a highly error-tolerant approach for automated peak-picking, constraint identification, and structure calculation cycles. Once modified for the specific needs of ssNMR applications (e.g. including ^{13}C−^{13}C or ^{15}N−^{15}N constraints, calibration of distance ranges from dipolar recoupling experiments, and proper handling of fractional labeling and intermolecular contacts), this algorithm could dramatically speed up the protein structure refinement from ssNMR data.

Identification of Intermolecular Contacts. In a highly ordered system such as microcrystalline material, it should be possible to observe contacts between ubiquitin molecules at the crystal contact sites. If not identified properly, these could potentially lead to a misinterpretation of the intramolecular restraints and subsequently lead to problems in the structure refinement. As demonstrated previously, intermolecular contacts can be suppressed by diluting the isotopically enriched sample with natural abundance material, mostly at the expense of reduced signal intensity.[6,8] As a compromise between signal-to-noise ratio and suppression of crystal contacts, we prepared a sample containing 1/3 labeled and 2/3 unlabeled material. Intermolecular contacts can thus be identified by comparing signal intensities between the fully labeled and the 'diluted' sample.

As a representative example, the Cα-Cα region of the spectrum of the diluted sample (green) is shown in Figure 5A overlayed on the spectrum of the labeled sample (red), both measured in a DARR experiment at 500 ms mixing time. Relative peak intensities for all sequential and medium range distances assigned in this spectral region are very similar, except

(36) Güntert, P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, *43*, 105−125.
(37) Güntert, P. Structure calculation using automated techniques. In *BioNMR in Drug Research*; Zerbe, O., Ed.; Wiley-VCH: Weinheim, 2003; pp 39−66.
(38) Linge, J.; Habeck, M.; Rieping, W.; Nilges, M. *Bioinformatics* **2003**, *19*, 315−316.
(39) Kuszewski, J.; Schwieters, C. D.; Garrett, D. S.; Byrd, R. A.; Tjandra, N.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, *126*, 6258−6273.
(40) Schwieters, C.; Kuszewski, J.; Tjandra, N.; Clore, G. *J. Magn. Reson.* **2003**, *160*, 66−74.

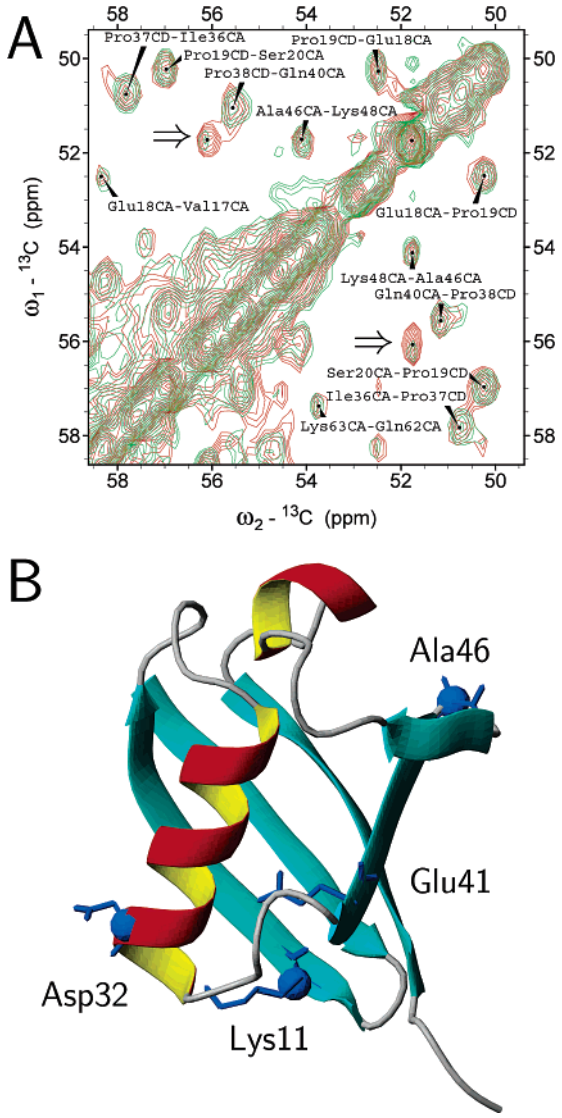


Figure 5. (A) Comparison of peaks intensities at 500 ms mixing time for a sample containing 100% [2-^{13}C]-Ubq (red) with a diluted sample containing 33% [2-^{13}C]-Ubq and 67% natural abundance ubiquitin (green). Peaks with considerable difference in intensity are likely to represent intermolecular contacts. The peaks marked with an arrow are assigned to Cα of Ala46 in one dimension. (B) Location of the Cα atoms (spheres) of Ala46, Asp32, Lys11. and Glu41 in the X-ray structure of ubiquitin (1ubq). With respect to Ala46, the Cα atoms of Asp32 and Lys11 are located on the opposite site of the protein, making a crystal contact with Ala46 possible.

for the pair of cross-peaks at 56.1/51.7 ppm (indicated by arrows). This pair shows a considerable differences in peak intensities on both sides of the diagonal with the signal being hardly above the noise level for the diluted sample. It turns out that those peaks can be unambiguously assigned in one dimension to the Cα atom of Ala46. In the second dimension, only three atoms (all Cα) fall into a ±0.1 ppm range with their distances to the Cα atom of Ala46 given in parentheses: Asp32 (24.7 Å), Glu41 (16.6 Å), and Lys11 (19.3 Å).

Figure 5B shows the location of these residues in the X-ray structure (1ubq). Indeed, Ala46 is located within a loop region with its Cα facing to the outside of the protein, hence making intermolecular contacts feasible. The residues Asp32 and Lys11 are located on opposite sides of the protein and can act as potential coupling partner by means of intermolecular contacts. On the other hand, a contact with Glu41 seems to be unlikely since (i) the Cα of Glu is not ^{13}C enriched to a high degree in the [2-^{13}C]-Ubq sample and (ii) Glu41 faces to the inside of the protein making the Cα atom difficult to access from the outside. Along those lines, several other crystal contacts have been identified. Indeed, a large number of weak peaks which could not be assigned in the spectra of the fully labeled sample are not present in the diluted sample and are therefore likely to represent intermolecular contacts. Consequently, those have been excluded from the restraint list.

It is worth noting that problems with intermolecular contacts may not arise if amorphous precipitates are used, i.e., if no long range order is present. In this case intermolecular contacts are not expected to be predominant. Nonetheless, hydrated protein precipitates can show line widths similar to those observed for microcrystals,[32,33] indicating a very defined environment (*local* order) for both microcrystals and amorphous precipitates. On the other hand, the observation of well-defined crystal contacts proves that even under tough experimental conditions such as high *g*-force due to fast MAS, or periodically heating of the sample due to high power proton decoupling and subsequent cooling, the microcrystals stay intact over a long period of time with a large degree of long-range order present.

Structure Refinement of Microcrystalline Ubiquitin. Using the iterative procedure described above, it was possible to identify 336 interresidue contacts for the [2-^{13}C]-Ubq sample. Out of those, 149 were sequential, 74 were medium range ($1 < |i - j| \leq 4$) and 113 were long range contacts. For the calculation of the protein structure, these restraints need to be assigned to specific distance ranges preferentially on an experimental basis. In principle, the relative cross-peak volumes are expected to correlate with the internuclear distances and with the mixing time. Unfortunately, a comparison of the peak volumes in the spectra at long mixing times with the distances in the X-ray structure (1ubq) showed only poor correlation. Even (short) sequential or intraresidues peaks were found to have a wide range of peak volumes in the DARR spectra. We ascribe this phenomenon to fractional labeling or insufficient suppression of dipolar couplings as discussed previously.[8]

We therefore decided to assign distance ranges on the basis of the well-defined secondary structure elements[8,9] and the first appearance of the peaks at particular mixing times. Sequential contacts are restraint to the appropriate distances given by the peptide geometry which include e.g. sequential CO−Cα: 2.1−2.9 Å, sequential Cα−Cα: 2.5−5.0 Å, sequential Cα−Cβ: 2.5−5.0 Å, sequential Cβ−Cβ: 2.5−6.0 Å. Distances in secondary structure elements are assigned according to their average values as analyzed previously[8,9] and include e.g. the Cα$_i$−C α$_{i+3}$ distances in an α-helix which are restrained to 2.5−5.0 Å. All other medium and long range peaks in the 500 ms spectrum are initially assigned to a range between 2.5 and 7.5 Å. If one of those peaks is identified already at 300 ms or 200 ms mixing time, the distance range is reduced to 2.5−6.5 Å and 2.5−5.5 Å, respectively. These upper limits for the distance bins have been motivated by histograms (see supplemental Figure 3 in Supporting Information) depicting the number of identified contacts within a certain X-ray distance range as function of the mixing time. Nonetheless, even without information from the X-ray structure, a refinement would have been possible with uncalibrated distances ranges, although with likely larger rmsd's for the resulting NMR structure. A more quantita-

tive range for the distance bins could then be obtained by detailed, thus time-consuming, analysis of the restraint and van der Waals violations.

Interestingly, many of the contacts close to the upper distance limits involve amino acids with two adjacent enriched ^{13}C-atoms (i.e. Val, Leu, Ile). Is is likely that these contacts represent a two-step magnetization transfer, i.e., a 'fast' one-bond transfer followed by a long-range transfer over a shorter distance than the indicated one. This raises an additional concern for the quantitative analysis of cross-peak buildup curves, since very detailed experiments would be required to distinguish between one-step and multistep transfer. It should be noted that the upper limit of observable distances depends not only on the mixing time, but particularly on the S/N of the spectra (i.e. sample amount, labeling scheme, magnetic field, spectrometer sensitivity, etc.) and might therefore vary from one experiment or sample to another.

The resulting list of interresidue distance restraints has been subjected to a structure calculation using the default protocols of CNS and a linear conformer as the initial structure.[31] More than 600 structures were calculated, which showed no more than one violation in any of the following categories: bond lengths (>0.1 Å), bond angles (>5°), improper dihedral angles (>5°) and van der Waals contacts (<1.6 Å). The 10 lowest energy structures have been selected for further analysis. The rmsd for the backbone heavy atoms of this structure ensemble is shown in Figure 6C as dashed line. It turned out that the structure bundle obtained sorely from distance constraints exhibits the correct global fold of the protein, but was still flexible even in secondary structure elements with respective backbone rmsd's between 1 and 3 Å.

To further improve the quality of the structure, we included restraints for the backbone dihedral angles, θ and ϕ, obtained from secondary chemical shift analysis performed by TALOS.[23] The TALOS database used here contains 78 proteins with more than 9000 residue triplets. To avoid a bias, we deleted the ubiquitin entry from the database. Chemical shifts obtained from the ssNMR assignment for backbone ^{15}N, Cα, Cβ, and CO were used as input.[11,12] TALOS generated restraints classified as 'good' for 61 out of 76 residues. Accordingly, 122 restraints for θ and ϕ including respective error margins given by TALOS have been used for the subsequent structure refinement. For this final step, both distance and dihedral angle restraints have been used as input with the average structure obtained from distance restraints only used as starting structure. The resulting bundle of 10 structures lowest in energy is shown in Figure 6A and B as represented by the backbone heavy atoms. For comparison, the X-ray structure (1ubq) is depicted as thick line.

The structure based on distance and dihedral angle constraints shows very well-defined secondary structure elements such as the α-helices with rms deviations in the backbone heavy atoms between 0.7 and 1.1 Å (Figure 6C). It furthermore closely resembles the X-ray structure (1ubq). Larger flexibility is observed only for regions with small number of restraints such as the loop between Leu8 and Lys11, and the residues near the C-terminus. This correlates well with the temperature-dependent observation of residues discussed above. Residues in mobile regions often lack intensity due to inefficient cross-polarization, making their observation by ssNMR methods more difficult. Deviations between X-ray and ssNMR structure are therefore likely due to lack of experimental restraints rather than caused by different crystal packing.

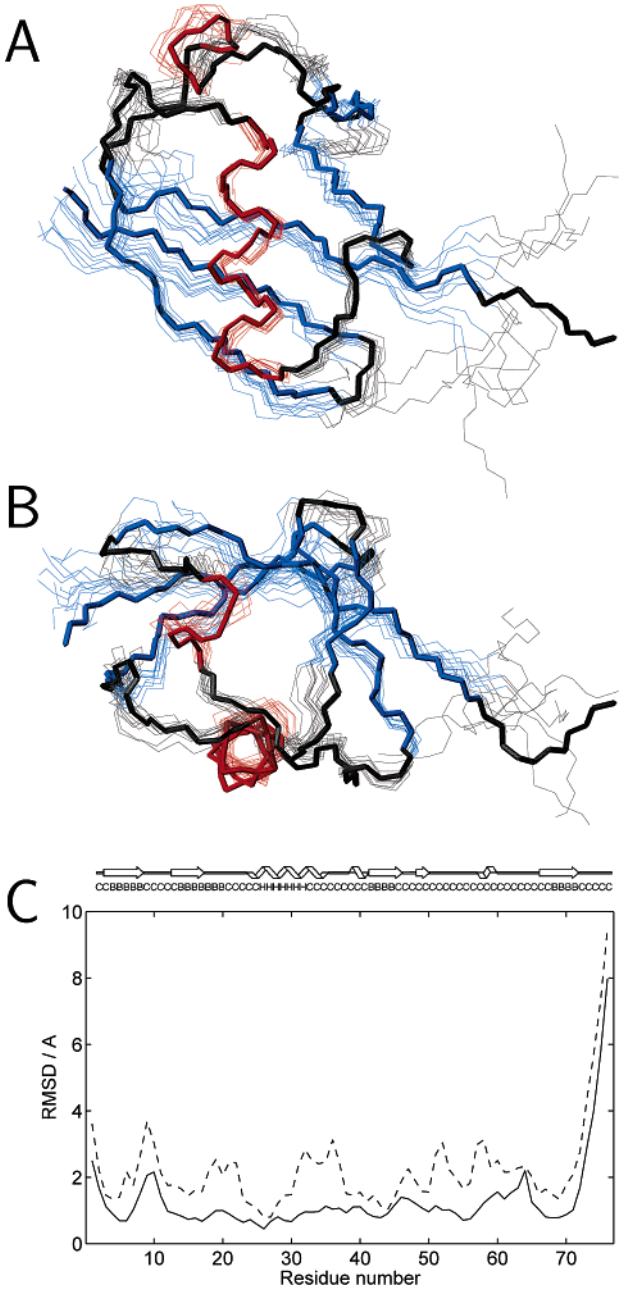


Figure 6. (A) Side and (B) top view of the structure ensemble for the 10 lowest-energy structures of ubiquitin calculated using distance constraints obtained from the DARR experiments and backbone dihedral angle constraints provided by TALOS. The thick line represents the backbone of the X-ray structure 1ubq for comparison. Secondary structure elements such as α-helices and β-sheets are shown in red and blue, respectively. The backbone atoms of well-defined regions of secondary structure have been fitted to minimum rmsd using MOLMOL 2K.1.[44] (C) Backbone heavy atom rmsd values obtained for the 10 lowest energy structures as function of residue number. The dashed line represents the structure obtained from distance constraints only. The solid line shows the rmsd's obtained for the structure bundle shown in A and B based on both distance and dihedral angle constraints. Regions with well-defined secondary structures show rmsd's between 0.7 and 1.1 Å while mobile regions such as the C-terminus are poorly defined.

Conclusion

The 3D structure of microcrystalline ubiquitin has been obtained from solid-state MAS NMR data providing a large

number of interresidue carbon−carbon restraints. The quality of the structure could be improved by an analysis of secondary chemical shifts providing backbone dihedral angle restrains. Essentially, only two separate sample preparations with different ^{13}C labeling schemes and a small number of experiments have been used for both resonance assignment[11,12] and structure determination of ubiquitin.

This study represents the second example of a protein structure derived ab initio from high-resolution ssNMR data obtained on hydrated microcrystals. Our approach shows large similarities to the structure determination of the SH3 domain with respect to labeling schemes and experiments used.[8] In the initial SH3 structure refinement, about 300 distance restraints were obtained mainly from 2D carbon correlation experiments on selectively ^{13}C-enriched SH3 microcrystals, grown with either [2-^{13}C]-glycerol or [1,3-^{13}C]-glycerol as a carbon source. Both samples give rise to complementary labeling schemes, which simplifies the identification of long-range contacts.[8]

In contrast, in our study ambiguities in the assignment of carbon−carbon restraints have been resolved using an iterative approach between assignment and structure refinement. This represents an alternative route which can be easily automated using recently developed software. Eventually, 336 interresidue distance restraints complemented by 122 dihedral angle restraints obtained from chemical shifts analysis, gave rise to a 3D structure with rmsd's of ∼1 Å or less for secondary structure elements. Hence, the quality of the ubiquitin structure presented here is similar to that of the SH3 domain with backbone rmsd's of ∼1.6 Å for secondary structure elements.[8] This demonstrates the general applicability of this method to determine 3D structures of proteins with moderate size.

Clearly, our study represents only a first step toward refinement of protein structures on the basis of ssNMR data. As demonstrated recently, increasing the spectral resolution by means of 3D $^{15}N-^{13}C-^{13}C$ correlation experiments will increase the number of distance restraints dramatically.[9] Furthermore, by using novel experiments and advanced labeling schemes, it is feasible to complement the $^{13}C-^{13}C$ restraints by $^{15}N-^{15}N$ and $^{1}H-^{1}H$ and backbone dihedral angle restraints.[8,20−22,41] Higher dimensional spectra and novel labeling schemes will increase the number of unambiguous restraints necessary for the iterative assignment/structure calculation approach discussed above and will therefore extend the applicability of this method to more complex systems.

Soon, the combination of those techniques will enable a refinement of 3D structures for solid protein formulations to a high quality. This has been demonstrated already for the structure of peptides which could be resolved at truly atomic detail.[42,17] Furthermore, it is likely that the complexity of the systems accessible to ssNMR studies at high-resolution will increase beyond the 12−15 kDa range soon. Since the techniques discussed above are independent of molecular tumbling times, ssNMR can be ideally applied to large molecular weight systems of medium complexity, such as smaller membrane bound proteins in their natural environment, or for the characterization of protein−protein and protein−peptide interactions in high molecular weight matrixes.

The experiments described here can be performed on solid sample formulations, such as microcrystalline material or hydrated precipitates and are therefore not limited to highly soluble globular proteins. Although still in its infancy, high-resolution MAS solid-state NMR spectroscopy provides an exciting opportunity to solve structures of proteins not amenable to solution NMR studies or X-ray crystallography, such as insoluble proteins, protein aggregates, and membrane proteins for which considerable progress has been reported recently.[43]

Acknowledgment. We thank T. Igumenova (U. Pennsylvania) for providing the assignments of microcrystalline ubiquitin prior to publication. S.G.Z. acknowledges the Alexander-von-Humboldt Foundation for financial support. This work was supported by grants from NSF, MCB 0316248 (to A.E.M.) and NIH, DK 39806 (to A.J.W.). A.E.M. is member of the New York Structural Biology Center. Support for the NYSBC has been provided by NIH/NIGMS through grant P41 GM66354.

Supporting Information Available: Supplemental Figures 1−3. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0503128

(41) Reif, B.; van Rossum, B. J.; Castellani, F.; Rehbein, K.; Diehl, A.; Oschkinat, H. *J. Am. Chem. Soc.* **2003**, *125*, 1488−1489.
(42) Rienstra, C. M.; Tucker-Kellogg, L.; Jaroniec, C. P.; Hohwy, M.; Reif, B.; McMahon, M. T.; Tidor, B.; T. L.-P.; Griffin, R. G. *Proc. Natl. Acad. Sci., U.S.A.* **2002**, *99*, 10260−10265.
(43) Krabben, L.; van Rossum, B.-J.; Castellani, F.; Bocharov, E.; Schulga, A. A.; Arseniev, A. S.; Weise, C.; Hucho, F.; Oschkinat, H. *FEBS Lett.* **2004**, *564*, 319−324.
(44) Konradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51−55.