Article

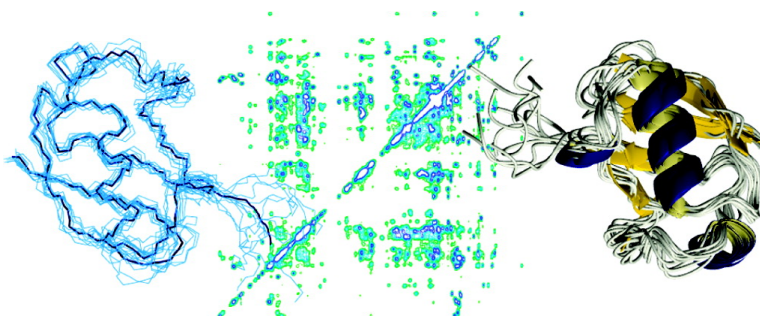# Protein Structure Determination from C Spin-Diffusion Solid-State NMR Spectroscopy

Theofanis Manolikas, Torsten Herrmann, and Beat H. Meier

More About This Article

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

- Supporting Information
- Links to the 3 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

ACS Publications
High quality. High impact.

Protein Structure Determination from ^{13}C Spin-Diffusion Solid-State NMR Spectroscopy

Theofanis Manolikas,[†] Torsten Herrmann,[‡] and Beat H. Meier*,[†]

Physical Chemistry, ETH Zurich, CH-8093 Zurich, Switzerland, and Institute of Molecular Biology and Biophysics, ETH Zurich, CH-8093 Zurich, Switzerland

Received October 19, 2007; E-mail: beme@ethz.ch

Abstract: Proton-driven ^{13}C spin diffusion (PDSD) is a simple and robust two-dimensional NMR experiment. It leads to spectra with a high signal-to-noise ratio in which cross-peaks contain information about internuclear distances. We show that the total information content is sufficient to determine the atomic-resolution structure of a small protein from a single, uniformly ^{13}C-, ^{15}N-labeled microcrystalline sample. For the example of ubiquitin, the structure was determined by a manual procedure followed by an automatic optimization of the manual structure as well as by a fully automated structure determination approach. The relationship between internuclear distances and cross-peak intensities in the spectra is investigated.

1. Introduction

While first atomic structures from (microcrystalline) proteins have been obtained,[1−4] an established protocol for structure determination by solid-state NMR is not yet available and many open questions remain. The determination of an unknown structure from NMR data is still a formidable task. To date, published records describe only relatively small systems (for examples, see refs 3, 5, and 6).

Establishing a robust structure-determination protocol for solid-state NMR, preferably using a single, uniformly ^{13}C-, ^{15}N-labeled protein, will enable solid-state NMR to play a crucial role in structural biology, in particular because the solid-state NMR method is not restricted to crystalline systems. It has already been demonstrated that amyloids[7] as well as membrane proteins[8] can give rise to well-resolved NMR spectra. Therefore, structure-determination protocols developed for and tested with microcrystalline proteins can also be applied to those systems, where diffraction methods are of limited use. In the following, we *de novo* determine the three-dimensional (3D) structure of ubiquitin from proton-driven spin-diffusion (PDSD) data and compare the results with the known X-ray structure. We discuss

the approximations involved and the features of the experimental approach in order to identify limitations of the procedure and possible improvements.

As for the established protocols for liquid-state NMR,[9] the sequence-specific resonance assignment is a prerequisite for structure determination. Thanks to recent technical advances including the availability of higher static magnetic fields, improved decoupling techniques, advanced sample preparation, and higher magic-angle spinning (MAS) frequencies, the resonance assignment of microcrystalline proteins[10−17] is nowadays well-established for ^{13}C and ^{15}N resonances on which structure determination mostly relies, since the proton resonances of solid proteins are not sufficiently well-resolved.

The next step in structure determination is the collection of distance restraints from NMR spectra. According to our assessment, this is a crucial bottleneck in structure determination by solid-state NMR, and we address this aspect in the following. We explore the case of PDSD,[18] where ^{13}C spin diffusion is observed under a residual or recoupled[19] dipolar interaction with the protons. Alternatively, spin diffusion between protons,

† Physical Chemistry.
‡ Institute of Molecular Biology and Biophysics.
(1) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420* (6911), 98−102.
(2) Castellani, F.; van Rossum, B. J.; Diehl, A.; Rehbein, K.; Oschkinat, H. *Biochemistry* **2003**, *42* (39), 11476−11483.
(3) Lange, A.; Becker, S.; Seidel, K.; Giller, K.; Pongs, O.; Baldus, M. *Angew. Chem., Int. Ed.* **2005**, *44* (14), 2089−2092.
(4) Zech, S. G.; Wand, A. J.; McDermott, A. E. *J. Am. Chem. Soc.* **2005**, *127* (24), 8618−8626.
(5) Rienstra, C. M.; Tucker-Kellogg, L.; Jaroniec, C. P.; Hohwy, M.; Reif, B.; McMahon, M. T.; Tidor, B.; Lozano-Perez, T.; Griffin, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (16), 10260−10265.
(6) Jaroniec, C. P.; MacPhee, C. E.; Bajaj, V. S.; McMahon, M. T.; Dobson, C. M.; Griffin, R. G. *Proc. Natl. Acad. Sci. U.S.A* **2004**, *101* (3), 711−716.
(7) Siemer, A. B.; Ritter, C.; Ernst, M.; Riek, R.; Meier, B. H. *Angew. Chem., Int. Ed.* **2005**, *44* (16), 2441−2444.
(8) Etzkorn, M.; Martell, S.; Andronesi, O. C.; Seidel, K.; Engelhard, M.; Baldus, M. *Angew. Chem., Int. Ed.* **2007**, *46* (3), 459−462.

(9) Wuthrich, K. *J. Biomol. NMR* **2003**, *27* (1), 13−39.
(10) Detken, A.; Hardy, E. H.; Ernst, M.; Kainosho, M.; Kawakami, T.; Aimoto, S.; Meier, B. H. *J. Biomol. NMR* **2001**, *20* (3), 203−221.
(11) Pauli, J.; Baldus, M.; van Rossum, B.; de Groot, H.; Oschkinat, H. *ChemBioChem* **2001**, *2* (4), 272−281.
(12) van Rossum, B. J.; Castellani, F.; Pauli, J.; Rehbein, K.; Hollander, J.; de Groot, H. J. M.; Oschkinat, H. *J. Biomol. NMR* **2003**, *25* (3), 217−223.
(13) 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* (4), 323−339.
(14) Igumenova, T. I.; Wand, A. J.; McDermott, A. E. *J. Am. Chem. Soc.* **2004**, *126* (16), 5323−5331.
(15) Igumenova, T. I.; McDermott, A. E.; Zilm, K. W.; Martin, R. W.; Paulson, E. K.; Wand, A. J. *J. Am. Chem. Soc.* **2004**, *126* (21), 6720−6727.
(16) Gammeren, A. J. v.; Hulsbergen, F. B.; Hollander, J. G.; Groot, H. J. M. d. *J. Biomol. NMR* **2005**, *31* (4), 279−293.
(17) Pintacuda, G.; Giraud, N.; Picratelli, R.; Bockmann, A.; Bertini, I.; Emsley, L. *Angew. Chem., Int. Ed.* **2007**, *46* (7), 1079−1082.
(18) Meier, B. H. Polarization Transfer and Spin Diffusion in Solid-State NMR. In *Advances in Magnetic and Optical Resonance*; Warren, W. S., Ed.; Academic Press, New York, 1994; Vol. 18, pp 1−116.
(19) Takegoshi, K.; Nakamura, S.; Terao, T. *Chem. Phys. Lett.* **2001**, *344* (5−6), 631−637.

detected indirectly via the ^{13}C resonances (CHHC[20]), can be used. In this contribution, we concentrate on the PDSD experiment with proton recoupling (DARR[19]). PDSD is a particularly simple and sensitive experiment because, neglecting T_1 relaxation, PDSD preserves the total ^{13}C magnetization as a constant of the motion.[18]

2. Theoretical Background

To sketch the problem of obtaining distance restraints from MAS solid-state NMR, it is instructive to compare the situation with liquid-state NMR and discuss the various assumptions needed to arrive at a master-equation approach, similar to the one used in liquid-state NMR[21] where the cross-peak intensity between N spins (e.g., in nuclear Overhauser effect spectroscopy (NOESY)) is described by a kinetic master equation for the polarizations $p_i = \langle S_{iz} \rangle$:

$$\frac{d}{dt}\begin{bmatrix} p_1 \\ p_2 \\ : \\ : \\ : \\ p_N \end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} & : & : & W_{1N} \\ W_{21} & : & : & : & : \\ : & : & : & : & : \\ : & : & : & : & : \\ W_{N1} & : & & : & W_{NN} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ : \\ : \\ : \\ p_N \end{bmatrix} \quad (1)$$

The matrix elements W_{ij} are proportional to r_{ij}^{-6}, where r_{ij} is the internuclear distance between spins i and j. Using the initial-rate approximation the cross-peak intensity between i and j becomes proportional to W_{ij} and independent of all other elements in the kinetic matrix **W**.[21] For longer mixing times, intensities are evaluated by taking the matrix exponential $\exp(-\mathbf{W}t)$.

In solids, the nature of the processes leading to polarization transfer is fundamentally different. In the absence of MAS, the dynamics of the spin system is described by a Hamiltonian which explicitly contains a secular dipolar coupling. Polarization transfer is therefore a quantum-mechanical many-body problem which cannot in general be solved, and approximations must be used. The most popular and successful approach is the spin-diffusion concept originally formulated by Bloembergen,[22] which also yields a master equation approach (for a review, see ref 23). The elements of the kinetic matrix W_{ij} depend again on the internuclear distance as r_{ij}^{-6}, but additionally on the second-order Legendre polynomial of the angle θ_{ij} between the internuclear vector $\mathbf{r}_{ij}$ and the magnetic field, which is different for each crystallite in a powder sample. Furthermore, the matrix elements depend on the zero-quantum line shape function $f_{ij}(\omega)$,[24] evaluated at $\omega = 0$, as well as on the magnetic constant μ_0 and the gyromagnetic ratio γ of the nuclei i and j:

$$W_{ij} = \frac{(\mu_0 \hbar \gamma_i \gamma_j)^2}{32\pi} \frac{1}{r_{ij}^6} (P_2(\cos \theta_{ij}))^2 f_{ij}(0) \quad (2)$$

The function $f_{ij}(\omega)$ has its maximum at the difference frequency $|\omega_i - \omega_j|$ and has a width determined by the

interactions with the spins other than i and j. In the case of PDSD, the dominant interaction for $f_{ij}(\omega)$ is the interaction with the strongly coupled proton bath. Making the crude approximation that $f_{ij}(\omega)$ is a constant, independent of i and j, and neglecting the dependence on the angle θ_{ij}, the basic relationship present in the NOESY experiment is reestablished approximately as

$$W_{ij} \cong \frac{k}{r_{ij}^6} \quad (3)$$

where k contains the constants of eq 2 as well as the powder average of the square of the Legendre polynomial.

MAS adds an additional complication to the description of the polarization transfer. Slow MAS does not interfere with the basic spin-diffusion mechanism[25] but fast spinning drastically changes the mechanism of spin diffusion because the usual flip-flop terms of the dipolar interaction, $I_i^+ I_j^- + I_i^- I_j^+$, become averaged by MAS and the leading terms for polarization transfer (using an average Hamiltonian description) are three-spin terms. Accordingly, the relevant rate constants in the master-equation ansatz of eq 1 are no longer just the internuclear distances between two spins but the geometry of a three-spin system, which becomes relevant.[26] Typical spin-diffusion experiments are performed at spinning frequencies between 10 and 20 kHz where neither of the two limiting cases are fully justified.

The master equation for solids is obtained by perturbation theory.[27,28] When the conditions of this approach are violated, dipolar truncation effects[18,26,29] are observed and weak couplings can be almost entirely suppressed by the presence of strong couplings, preventing the measurement of long-distance restraints. It has been however demonstrated that dipolar truncation is weak for spin-diffusion experiments.[26]

Despite all the assumptions entering the master-equation approach, it should be emphasized that a defined, albeit complex, relation between cross-peak intensity and local geometry exists. Obviously, there is no mechanism that could lead to sizable cross-peaks between nuclei separated more than an upper distance u_{ij} depending on the experiment performed and the mixing time used. Typically u_{ij} can be chosen between 2.5 Å (e.g., for CHHC with very short mixing time[20]) and 10 Å (e.g., PDSD with long mixing time, *vide infra*). Therefore, we accept in the following the validity of the master equation and assume that the cross-peak intensity between resonances i and j is indeed determined by

$$I_{ij} = k/u_{ij}^6 \quad (4)$$

where u_{ij} is the experimentally determined NMR distance. The proportionality constant k can be determined experimentally from known internuclear distances. The justification for using the crude approximation of eq 4 for the structural analysis lies in the fact that no precise restraints are indeed necessary if their number is sufficiently large and leads to an overdetermined problem. A similar situation is often encountered in structure

(20) Lange, A.; Luca, S.; Baldus, M. *J. Am. Chem. Soc.* **2002**, *124* (33), 9704−9705.
(21) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71* (11), 4546.
(22) Bloembergen, N. *Physica* **1949**, *15*, 386.
(23) Ernst, M.; Meier, B. H. Spin Diffusion in Solids. In *Solid State NMR of Polymers*; Ando, I., Asakura, T., Eds.; Elsevier: Amsterdam, 1998; Vol. 84.
(24) Suter, D.; Ernst, R. R. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1985**, *32* (9), 5608.

(25) Gan, Z.; Ernst, R. R. *Chem. Phys. Lett.* **1996**, *253*, 13−19.
(26) Grommek, A.; Meier, B. H.; Ernst, M. *Chem. Phys. Lett.* **2006**, *427* (4−6), 404−409.
(27) Abragam, A. *The Principles of Nuclear Magnetism*; Clarendon Press: Oxford, 1961.
(28) Fermi, E. *Nuclear Physics*; University of Chicago Press: Chicago, 1950.
(29) Hohwy, M.; Rienstra, C. M.; Jaroniec, C. P.; Griffin, R. G. *J. Chem. Phys.* **1999**, *110* (16), 7983−7992.

determination by liquid-state NMR and has been shown to lead to reliable structures.[9]

3. Materials and Methods

3.1. Spectroscopy. Microcrystals of uniformly ^{13}C-, ^{15}N-labeled ubiquitin (purchased from VLI Research Inc.; Malvern, PA) have been grown as described previously.[30] Crystals (5–6 mg of protein) were transferred after 14 days to a 2.5 mm rotor. The experiments were carried out using a triple 2.5 mm Chemagentics/Varian T3 probe on a Bruker AV600 spectrometer operating at a static field strength of 14.09 T. Three two-dimensional (2D) ^{13}C–^{13}C correlation spectra were recorded with 100, 250, and 400 ms mixing times at 12 kHz MAS and a sample temperature of 269 K. After an initial 100 kHz 90° pulse on protons, cross-polarization fields were adjusted to 72 kHz on the proton channel and 60 kHz on the carbon channel. An adiabatic-passage cross polarization with a tangential shape on the proton channel with $\Delta_{rf} = 8$ kHz (initial deviation of the RF field from its average value), $d_{est} = 2$ kHz (shape parameter), and a contact time of 1 ms was used.[31]

During the mixing time, carbon magnetization was stored along B_0, and proton irradiation with a field amplitude matching the MAS frequency (DARR[19]) was applied. During t_1 and t_2, SPINAL64 proton decoupling[32] was employed at 110 kHz. All spectra were recorded with 800 t_1 increments and 128 scans resulting in a total measurement time of 75 h for the 100 ms mixing time spectrum, 79.5 h for the 250 ms mixing time spectrum, and 84 h for the 400 ms mixing time spectrum.

All spectra were processed using the XWINNMR software package (Bruker-Biospin) and analyzed with CARA.[33] A cosine-squared window function was applied as well as zero filling to 4096 data points in both dimensions.

3.2. Protocol for Automated Spectral Analysis Using ATNOS/ CANDID. The ATNOS/CANDID protocol for automated collection of conformational restraints, originally developed for the evaluation of [^1H, ^1H] NOESY spectra, was modified for the analysis of 2D ^{13}C spin-diffusion experiments. The input for the iterative ATNOS/ CANDID procedure consists of the amino-acid sequence of the protein, the chemical-shift lists from the previous sequence-specific resonance assignment, and one or several 2D ^{13}C-correlation spectra. The standard protocol with seven cycles of peak picking with ATNOS, cross-peak assignment with CANDID, and structure calculation with CYANA was performed.[34,35] During the first six ATNOS/CANDID cycles, ambiguous distance restraints[36] were used. At the outset of the spectral analysis, ATNOS/CANDID uses highly permissive criteria with respect to signal-to-noise ratio (SNR) and local-extrema detection as detailed in ref 34 to identify and assign a comprehensive set of peaks in the ^{13}C-correlation spectra. The ranking criteria for the initial assignments is described in ref 35. The procedure is identical to the one for liquid-state spectroscopy with the assignment tolerance for the chemical shift set to 0.25 ppm, slightly more generous than that for the manual procedure. In cycle 1, ambiguous peaks that can be interpreted as short range as well as intermediate or long range will always be assigned to a short-range restraint. In the second and subsequent cycles, the intermediate protein 3D structures are used as an additional guide for the interpretation of the spin-diffusion spectra. The output in each ATNOS/CANDID cycle consists of assigned cross-peak lists for each input spectrum and a final set of meaningful upper distance restraints

which constitute the input for the CYANA[37] 3D structure calculation algorithm. In addition, torsion angle restraints for the backbone dihedral angles ϕ and ψ derived from $C\alpha$ and $C\beta$ chemical shifts[38,39] were added to the input for each cycle of structure calculation. The 20 conformers with the lowest residual CYANA target function (representing violations, e.g., in upper distance restraints, van der Waals repulsion, and dihedral angles[40]) values obtained from cycle 7 of ATNOS/CANDID were energy-refined in a water shell with the program OPALp,[41,42] using the AMBER force field.[43] The program MOLMOL[44] was used to analyze the protein structure and to prepare the figures.

3.3. Protocol for Semiautomated Spectral Analysis Using AT-NOS/CANDID. The manually obtained 3D structure was used as starting point for an automated refinement procedure using ATNOS/ CANDID. To this end, the ATNOS/CANDID protocol was applied starting from the second cycle onward, thus substituting the Cartesian coordinates—usually provided by cycle 1—of the fully automated protocol (see Section 3.2) by a manually obtained bundle of conformers (*vide infra*). Otherwise the same protocol as described in Section 3.2 was applied.

3.4. Conversion of Peak Intensities into Upper Distance Bounds in the Automatic and Semiautomatic Procedures. The intensity of a PDSD signal attributed to a pair of atoms is interpreted as an upper bound distance on the interatomic distance rather than as a precise distance. The upper bound $u_{ij} \equiv u_n$ on the distance between two atoms, i and j, are derived from the corresponding signal intensity, $I_{ij} \equiv I_n$, according to eq 4 for all assigned cross-peaks, $n = 1...N$. The calibration constant k was determined in analogy to the procedures applied to analyze liquid-state NOESY spectra on the basis of known distances derived from the fact that the fixed bond lengths, bond angles, and chiralities of the covalent structure of the protein impose upper limits on intraresidual and sequential ^{13}C–^{13}C distances.[45–47] These conformation-independent distances d_{ij}^{cov} are computed for all pairs of atoms, i and j, where the two spins are related by a number of torsion angles, thereby defining an upper and lower limit for the spatial separation of i and j

$$d_{ij}^{cov,min} \leq d_{ij}^{cov} < d_{ij}^{cov,max} \tag{5}$$

While for liquid-state applications only spins separated by one or two torsion angles are usually included, this scheme was extended here to perform a systematic analysis of the local conformation of the covalent polypeptide structure. For fragments of the polypeptide chain (up to 7–12 amino acids), an exhaustive search of the accessible conformational space is feasible if the conformation space is discretized in the form of a multidimensional grid, where each dimension corresponds to an independent torsion angle, and the size of the grid mesh is set to 20°. The conformational space is further limited by only considering angles consistent with the observed secondary ^{13}C chemical shifts of the corresponding residues.[38] Secondary shifts greatly restrict the local conformation of a residue in the α-helical and β-sheet regions

(30) Schubert, M.; Manolikas, T.; Rogowski, M.; Meier, B. H. *J. Biomol. NMR* **2006**, *35*, 167–173.
(31) Hediger, S.; Meier, B. H.; Ernst, R. R. *Chem. Phys. Lett.* **1995**, *240*, 449.
(32) Fung, B. M.; Khitrin, A. K.; Ermolaev, K. *J. Magn. Reson.* **2000**, *142* (1), 97–101.
(33) Keller, R. L. J. *The Computer Aided Resonance Assignment Tutorial* (see also http://www.nmr.ch); Cantina Verlag: Goldau, Switzerland, 2004.
(34) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **2002**, *24* (3), 171–189.
(35) Herrmann, T.; Guntert, P.; Wuthrich, K. *J. Mol. Biol.* **2002**, *319* (1), 209–227.
(36) Nilges, M. *Folding Des.* **1997**, *2* (4), S53–S57.
(37) Guntert, P.; Mumenthaler, C.; Wuthrich, K. *J. Mol. Biol.* **1997**, *273* (1), 283–298.
(38) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490–5492.
(39) Luginbuhl, P.; Szyperski, T.; Wuthrich, K. *J. Magn. Reson., Ser. B* **1995**, *109* (2), 229–233.
(40) Güntert, P. Automated NMR Structure Calculation With CYANA. In *Methods in Molecular Biology*; Downing, A. K., Ed.; Humana Press: Totowa, NJ, 2004; Vol. 278, pp 353–378.
(41) Luginbuhl, P.; Guntert, P.; Billeter, M.; Wuthrich, K. *J. Biomol. NMR* **1996**, *8* (2), 136–146.
(42) Konradi, R.; Billeter, M.; Günthert, P. *Comput. Phys. Commun.* **2000**, *124*, 139–147.
(43) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.
(44) Konradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graph.* **1996**, *14*, 51–55.
(45) Guntert, P.; Braun, W.; Wuthrich, K. *J. Mol. Biol.* **1991**, *217* (3), 517–530.
(46) Wuthrich, K.; Billeter, M.; Braun, W. *J. Mol. Biol.* **1983**, *169* (4), 949–961.
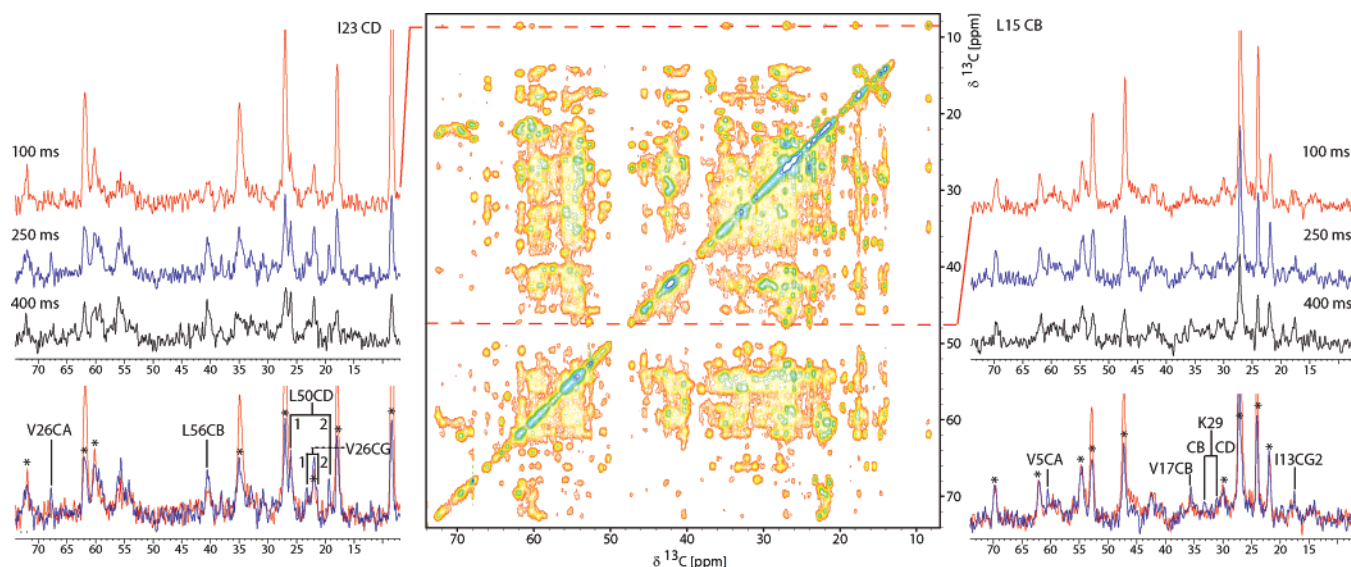(47) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

Figure 1. 2D $^{13}C-^{13}C$ PDSD spectrum of uniformly labeled ubiquitin recorded at 12 kHz MAS and a mixing time of 100 ms. The dashed lines indicate the region of one-dimensional (1D) slices extracted from 2D PDSD spectra at 100, 250, and 400 ms mixing time for I23CD (on the left) and L15CB (on the right). Beneath the 1D slices the overlay of the slices for 100 and 250 ms is given, indicating medium- and long-range correlations. Intraresidual and sequential contacts are marked with an asterisk.

of the protein. Consequently, the exhaustive search employed yielded conformational upper distance limits between nuclei in regular secondary structure elements which can be used to determine the calibration constant in eq 4 by identifying relevant pairs of nuclei, a and b, which satisfy

$$d_{a,b}^{\text{cov,max}} \equiv d_i^{\text{cov,max}} \le d^{\text{max}} \tag{6}$$

The parameter d^{max} is a user-defined parameter set to a sufficiently small value ($d^{\text{max}} = 7.5$ Å in our case) to ensure that the corresponding atom pair a and b is expected to yield an observable signal in the 2D PDSD spectrum. Following this, the calibration constant k is set to a value such that the arithmetic average distance is satisfying

$$\bar{u} = \frac{1}{M}\sum_{i=1}^{M} d_i^{\text{cov,max}} = \frac{k}{M}\sum_{i=1}^{M} \frac{1}{\sqrt[6]{I_i}} \tag{7}$$

4. Results

The structure of microcrystalline ubiquitin was determined from uniformly labeled ubiquitin (U-[^{13}C,^{15}N]) using three PDSD spectra recorded at mixing times of 100, 250 and 400 ms. The spectrum at 100 ms mixing is shown in Figure 1, together with representative cross sections at all three mixing times. The spectral assignment was taken from Schubert et al.[30] and includes residues 1−7 and 12−70. The remaining resonances are not observed in the spectra, probably because of their molecular dynamics. The corresponding chemical shifts are available from BioMagResBank under accession number 7111 and contain stereospecific information for valine and leucine sidechains.

4.1. Structure Determination Using a Simple Manual Analysis. The 2D carbon−carbon correlation spectra used for the analysis have a carbon-resonance line width exceeding 0.1 ppm, which is typical for spectra from uniformly labeled proteins. Consequently, rather severe spectral overlap is observed in each 1D projection of the 2D spectra, and only a few cross-peaks are *spectrally unambiguous* in the sense that their presence provides a distance restraint between two nuclei that

Table 1. Upper Distance Restraints Used for the Structure Calculations

meaningful UDRs[a]		manual I 35	manual II 55[b]/4[c]	semiautom 996	autom 983
intra	$r = 0$			15	18
sequential	$r = 1$			219	214
medium	$r < 5$	4	6	372	379
long	$r \ge 5$	31	49	390	372

[a] UDR: upper distance restraint. [b] Number includes the unambiguous restraints from manual I and the additional restraints that are unambiguous in the context of the fold (with a maximum of three assignment possibilities but only one assignment that conforms to the structure of manual I using 7.5 Å as a distance limit). [c] Number of additional ambiguous restraints (with two assignment possibilities and both assignments accord to the structure of manual I using 7.5 Å as a distance limit).

are unambiguously identified from the ^{13}C *resonance frequencies only*. As a criterion for unambiguousness, we request that the cross-peak resonance frequency in the protein spectrum fits an assigned resonance within ±0.1 ppm and, at the same time, is more than ±0.2 ppm away from any alternative resonance frequency. Reducing the tolerance value of ±0.2 ppm to ±0.15 ppm for resonances with a line width narrower than 0.2 ppm (full width at half-height) added a few more peaks. With the use of these criteria, 26 unambiguous peaks have been identified in the 250 ms PDSD spectrum and 9 additional unambiguous peaks in the 400 ms spectrum (for a summary, see Table 1). The detailed list of all spectrally unambiguous assignments is given in the Supporting Information (Table 1).

The manually unambiguous picked distance restraints as well as TALOS[48] dihedral restraints for all assigned residues (omitting ubiquitin from the TALOS library) were used to calculate a structure model for ubiquitin with the software CYANA 2.1.[37] The obtained long-range correlations in the 250 ms spectrum were set to an upper distance limit of 5.5 Å, and the ones from the 400 ms spectrum were set to an upper distance limit of 7.5 Å. The 10 lowest-energy structures from 200 calculated structures have been selected and are plotted in Figure 2a

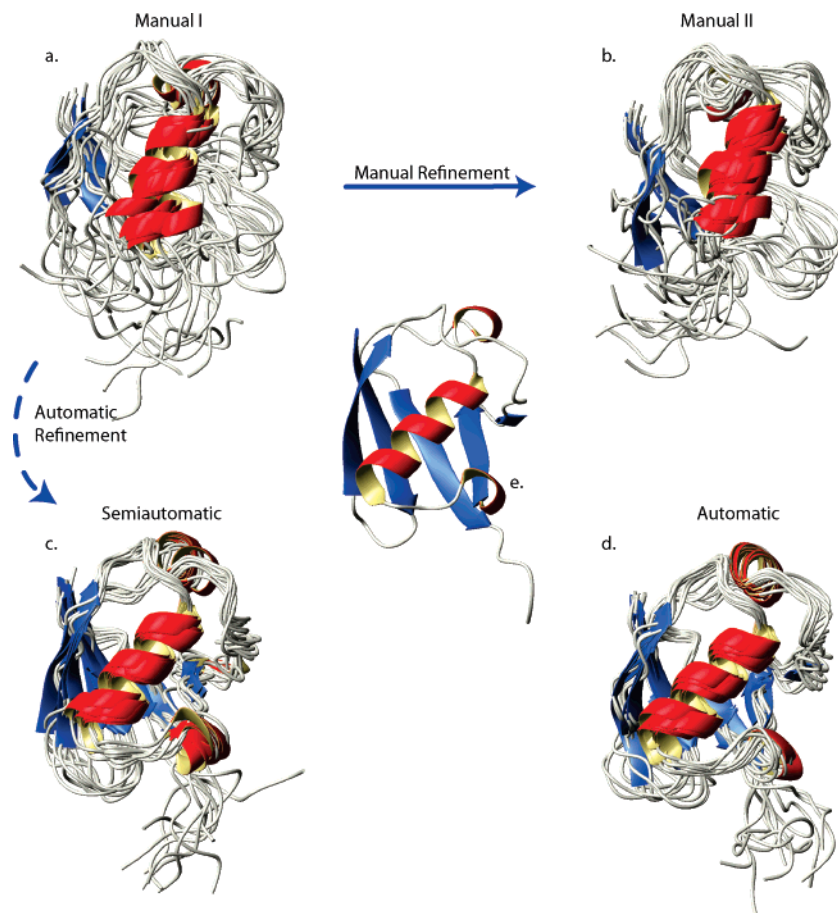(48) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13* (3), 289−302.

Figure 2. Structures derived from the manual (a), refined manual (b), semiautomatic (c), and fully automatic (d) approach. The 10 lowest-energy backbone structures are displayed in a ribbon cartoon representation showing the secondary-structure elements as they are automatically recognized by the program MOLMOL. For comparison, the crystal structure of ubiquitin (pdb code: 1UBQ) is depicted (e).

Table 2. RMSD Values of Backbone and All Heavy Atoms for Each Structure Calculation and Bias to the X-ray Crystal Structure for Residues 1−7 and 12−70

	manual I	manual II	semiautom	autom
Mean rmsd				
bb	1.72 ± 0.45	1.21 ± 0.18	0.65 ± 0.07	0.68 ± 0.08
heavy	2.43 ± 0.45	1.83 ± 0.18	1.14 ± 0.08	1.16 ± 0.09
Bias to X-ray				
bb	3.60 ± 0.21	2.95 ± 0.22	1.46 ± 0.11	1.64 ± 0.15
heavy	4.63 ± 0.21	4.06 ± 0.16	2.34 ± 0.15	2.57 ± 0.15

together with the X-ray structure of ubiquitin (pdb code: 1UBQ).[49] The obtained structure is of rather low resolution with a heavy atom root-mean-square deviation (rmsd) of 2.43 ± 0.45 Å and a backbone rmsd of 1.72 ± 0.45 Å (residues 1−7 and 12−70) and rather inaccurate with a bias to the X-ray backbone of 3.60 ± 0.21 Å (Table 2).

Despite the low number of unambiguous distance restraints, the fold is reproduced correctly. To obtain a more accurate and precise structure, spectrally ambiguous restraints must be taken into account. Here, spectrally ambiguous restraints (by the above criteria) having a maximum of three alternative assignment possibilities were added as ambiguous restraints to the CYANA input. If only one of the assignments was possible based in the

(49) Vijay-Kumar, S.; Bugg, C. E.; Wilkinson, K. D.; Vierstra, R. D.; Hatfield, P. M.; Cook, W. J. *J. Biol. Chem.* **1987**, *262* (13), 6396−6399.

fold obtained above (using 7.5 Å as the distance limit), however, the restraint was treated as unambiguous in the CYANA calculation.

The structure obtained by this manual refinement has a heavy atom rmsd of 1.83 ± 0.18 Å, and the bias to the X-ray backbone is 2.95 ± 0.22 Å (Table 2, Figure 2b). The complete list of restraints used for the calculation is given in the Supporting Information (Table 1). For a further structure refinement, more permissive assignment criteria (e.g., allowing ambiguous restraints with more than three assignment possibilities) need to be applied. While this could be done manually, we will, in the following, use an automated procedure to reach this goal.

4.2. Structure Determination Using Semiautomated Spectral Analysis. The ATNOS/CANDID procedure was used for an automated refinement procedure of the manually obtained 3D structure, derived from the spectrally unambiguous assignments only. ATNOS/CANDID assigned, in cycle 7, 1562 cross-peaks in the 100 ms PDSD spectrum, 1891 in the 250 ms spectrum, and 1883 in the 400 ms spectrum using a chemical-shift tolerance value of ±0.25 ppm. These translate into 996 meaningful upper distance limits as input for the final structure calculation in ATNOS/CANDID cycle 7. Details are given in Table 1 as well as in Supporting Information Table 2, and the structure obtained is displayed in Figure 2c. The low residual CYANA target function value of 0.42 ± 0.12 Å^2 (Table 3) and the average global rmsd value relative to the mean coordinates of 0.65 ± 0.07 Å calculated for the backbone and 1.14 ± 0.08

Table 3. Structure Calculation Input and Characterization of the Energy-Minimized NMR Structures

	semiautom	autom
meaningful UDRs	996	983
intra $r = 0$	15	18
sequential $r = 1$	219	214
medium $r < 5$	372	379
long $r \geq 5$	390	372
dihedral angle restraints	70	70
residual target function, Å^2	0.42 ± 0.12	1.28 ± 0.23
residual UDR violations^a		
no. ≥ 0.2Å	0 ± 1	4 ± 2
avg, Å	0.18 ± 0.03	0.34 ± 0.08
residual angle violations^a		
no. ≥ 5°	0 ± 0	0 ± 0
avg, °	0.82 ± 0.66	1.06 ± 1.21
Amber energies, kcal/mol^a		
total	−3033 ± 92	−2922 ± 138
van der Waals	−187 ± 14	−153 ± 15
electrostatic	−3503 ± 89	−3454 ± 126
rmsd from ideal geometry^a		
bond lengths, Å	0.0077 ± 0.0002	0.0079 ± 0.0003
bond angles, °	2.087 ± 0.061	2.215 ± 0.073
rmsd to the mean coordinates, Å^b		
bb (1−7,12−70)	0.65 ± 0.07	0.68 ± 0.08
ha (1−7, 12−70)	1.14 ± 0.08	1.16 ± 0.09
Ramachandran plot statistics^c		
most-favored regions (%)	67	65
additional allowed regions (%)	23	24
generously allowed regions (%)	5	6
disallowed regions (%)	5	5

^{a} The average value for the 20 energy-minimized conformers with the lowest residual CYANA target function values and the standard deviation among them are given. ^{b} bb indicates the backbone atoms N, Ca, C′; ha stands for "all heavy atoms". ^{c} As determined by *PROCHECK* (Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. *Proteins* **1992**, *12*, 345−364. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283−291).

Å for the heavy atoms of residues 1−7 and 12−70 (Table 2) mark a significant improvement over the manually obtained structures. Details of the structure are specified in Table 3. The bias to the X-ray structure is 1.46 ± 0.11 Å calculated for the backbone and 2.34 ± 0.15 Å for the heavy atoms, still indicating minor systematic distortions in the structure. Nevertheless, the precision as well as the accuracy of the obtained structure are very satisfying.

4.3. Structure Determination with a Fully Automated Spectral Analysis By ATNOS/CANDID. The ATNOS/CANDID program was next employed for a fully automated analysis of the 2D PDSD spectra (see Materials and Methods) without the need of manual peak picking. The ATNOS/CANDID protocol in conjunction with the CYANA torsion angle dynamic program yielded a total of 1557 assigned cross-peaks in ATNOS/CANDID cycle 7 for the 100 ms PDSD spectrum, 1882 assigned cross-peaks in the 250 ms spectrum, and 1869 assigned cross-peaks in the 400 ms spectrum using a chemical-shift tolerance value of ±0.25 ppm. These translate into 983 meaningful upper distance restraints detailed in Table 2. These restraints were the input for the final structure calculation in ATNOS/CANDID cycle 7. The structure obtained is displayed in Figure 2d, and the detailed characterization of the calculated structure is given in Table 3. The low residual CYANA target function value of 1.3 ± 0.2 Å^2 (Table 3) and the average global rmsd value relative to the mean coordinates of 0.68 ± 0.08 Å calculated for the backbone and of 1.16 ± 0.09 Å calculated for the heavy atoms of residues 1−7 and 12−70 are indicative of a well-defined NMR structure. The bias to the X-ray structure is 1.64 ± 0.15 Å calculated for the backbone and 2.57 ± 0.15 Å for the heavy atoms. Overall, the structure obtained has a very similar precision as well as accuracy as the

one obtained by the semiautomated procedure. The 20 lowest-energy conformers as well as the NMR restraint data file are deposited at the Protein Data Bank (PDB id: 2jzz).

5. Discussion

5.1. Obtained Structures. The structure calculation described in the preceding chapter has led to the correct fold for ubiquitin in manual, semiautomatic, or fully automatic mode. The application of the modified ATNOS/CANDID scheme (automated or semiautomated mode) has yielded well-defined structures with low rmsd values. The bias to the X-ray structure is about twice the rmsd of the NMR bundle. Assuming that the X-ray structure of the molecule is indeed identical to the structure in our microcrystallites, this indicates that some, albeit minor, systematic distortions are observed which could come from the approximations discussed above.

It is interesting to note that the correct fold is already obtained using only 35 restraints from spectrally unambiguous peaks. The structure can be improved to a backbone rmsd value of 1.72 ± 0.45 Å using only cross-peak intensities obtained manually. This structure has, however, a bias from the X-ray structure of 3.60 ± 0.21 Å. The semiautomatic and automatic procedures lead to the identification of a large number of spectrally ambiguous peaks in the spectrum that support the structure. This significantly improves the rmsd (by a factor of 2), and maybe more importantly reduces the bias to the X-ray structure (for the backbone) to about 1.5 Å for both structures, indicating that the automated procedure can not only obtain the structure without manual intervention but can also identify structural information not readily available manually.

We conclude that the accuracy and precision of the structures obtained from uniformly labeled ubiquitin by the semiautomatic
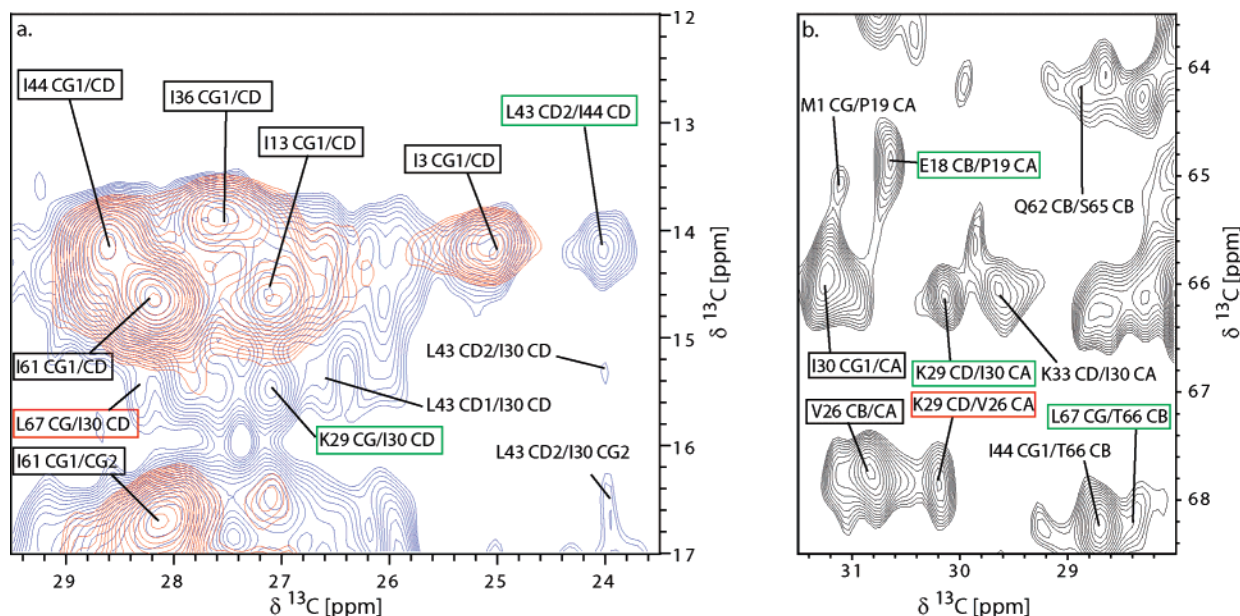
Figure 3. (a) Overlay of 2D ¹³C−¹³C PDSD spectra recorded at 100 ms (red) and 250 ms (blue) mixing time. Intraresidual contacts are observed in both spectra (black boxes). In the 250 ms spectrum, sequential contacts (L43/I44 and K29/I30 in green boxes) as well as medium- and long-range contacts appear. The appearing unambiguous long-range contact L67/I30 is marked with a red box. Contacts indicating ambiguous but structural meaningful medium- or long-range cross-peaks are not marked (e.g., a, L43/I30; b, M1/P19). (b) 2D ¹³C−¹³C PDSD spectrum recorded at 400 ms mixing time. The only direct medium-range contact from the unambiguous manually assigned list is given here (K29/V26, marked with a red box).

and automatic approaches are indeed very satisfactory. Since the modified ATNOS/CANDID scheme was only tested on ubiquitin itself, more experimental work and results from other proteins will be needed before a universal automatic approach can be postulated.

5.2. Cross-Peak Intensities and Internuclear Distances. The large number of long-range restraints identified during the structure-calculation procedure (for a summary, see Supporting Information Table 2) already indicates that the dipolar-truncation effects are indeed small in PDSD spectra of uniformly labeled proteins as predicted from small model systems[26] but not previously verified for proteins. For the 2D 250 ms PDSD spectrum, a total of 1882 peaks were assigned automatically. Of these, 1245 (66%) are intraresidual ($r = 0$) or sequential ($r = 1$) contacts while 357 are medium-range restraints ($r < 5$) and 280 are long-range ($r \geq 5$) restraints. Very similar numbers were obtained in the semiautomatic approach (Supporting Information Table 2).

The absence of significant truncation can also be directly seen by inspection of the slices displayed in Figure 1 and the 2D spectra shown in Figure 3 which show ambiguous intra-, short-, medium- and long-range correlations picked and assigned manually as well as automatically. Most of the medium- and long-range correlations depicted in the overlay of the cross sections in Figure 1 appear primarily in the 250 ms spectrum (unambiguously picked contacts, e.g., Figure 1, I23CD−V26CA, I23CD−L50CD2, L15CB−V5CA, and L15CB−K29CD; as well as Figure 3a, C67CG/I30CD). The signal intensity of medium- and long-range contacts, appearing already in the 100 ms spectrum, increases in the 250 ms spectrum (contacts: I23CD−L56CB, L15CB−V17CB, and L15CB−I13CG2, Figure 1). In the 400 ms 2D spectrum, additional correlations were found. In Figure 3b one additional unambiguous contact K29CD−V26CA and four ambiguous but structural meaningful

peaks (M1CG−P19CA, K33CD−I30CA, I44CG1−T66CB, and Q62CB−S65CB) are shown.

The presence of these long-range correlations clearly demonstrates that polarization transfer over longer distances takes place and that dipolar truncation is not dominant in PDSD spectra. However, the fact that these peaks are long range does not mean that relayed transfer plays no role in the transfer. In fact, all of the contacts from Figures 1 and 3, as discussed above, except K29CD−V26CA, can also be described by a relayed transfer over one intermediate partner for which the longer of the two distances is shorter than the direct contact.

5.3. Comparison of Experimental Cross-Peak Intensity and Internuclear Distance. To investigate the validity of the proportionality of the cross-peak intensity with the inverse sixth power of the internuclear distance (eq 4), we have plotted in Figure 4 the upper distance restraints u_{ij} for each assigned cross-peak as a function of the actual distance known from the X-ray structure using the calibration from the fully automatic structure determination. The longest distances represented by cross-peaks are about 10 Å. Ideally, the cross-peak intensity would directly reflect the internuclear distance and all data points would lie on the diagonal shown as a solid line in Figure 4. Because the distance restraints are to be interpreted as upper distance restraints only, all experimental points should be above (or on) the diagonal because NMR limits may be larger but not smaller than real distances. Figure 4a clearly shows that there is a significant number of data points (212 distance restraints) below the diagonal, representing 22% of all 983 cross-peaks that correspond to meaningful upper distance restraints. Of these, 108 are, however, within 0.5 Å from the diagonal, indicating only a weak violation. Eighty-nine percent of the signals lie above or close to the diagonal (within 0.5 Å). This finding, together with the overdetermination of the problem, explains the empirical finding that correct structures are found using the
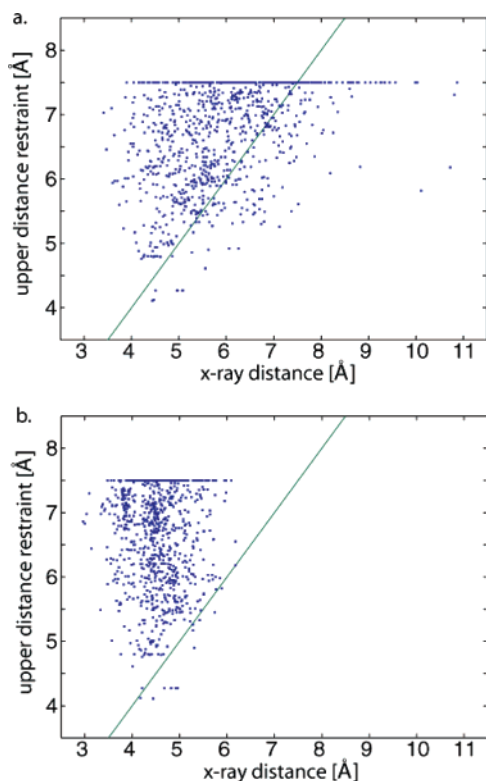
Figure 4. Distances of meaningful restraints in the X-ray structure correlated to the calibrated NMR upper distance limits used for structure calculation (a) and X-ray distances adjusted to the longer distance of a two-step relay transfer (b).

approximations discussed. The data also clearly indicate that the cross-peak intensity contains distance information, if interpreted as upper distance restraints.

We have investigated the effect of the calibration procedure and the assumption of an u_{ij}^{-6} dependence of the cross-peak intensity for our example and have found them particularly important for the first cycle in ATNOS/CANDID. The backbone rmsd value of the structure determined in cycle 1 amounted to 4.65 ± 1.70 Å without calibration (setting all distance restraints to a uniform 7.5 Å) and improved to 1.80 ± 0.28 Å with calibration. Without calibration, the procedure often failed to converge. For the final structure calculaton in cycle 7, the improvement obtained by calibration is small (0.77 ± 0.12 Å and 0.68 ± 0.08 Å, without and with calibration, respectively). Additional studies using different molecules and data sets will be needed to determine if this finding can be generalized.

As mentioned earlier, we suspect that the presence of relayed transfer, as a consequence of the violated initial-rate approximation, is responsible for obtaining too short upper distance restraints for some cross-peaks (data points below the diagonal in Figure 4a). To test this hypothesis, we have used a simple correction procedure based on the X-ray structure. For 857 out of the 983 cross-peaks, it was indeed possible to find a relayed transfer pathway. Instead of one long transfer step, two short steps, with the longer of the two being shorter than the direct transfer step, can be identified from the structure. Replacing

the direct internuclear distance by the longer of the two distances in the relay chain leads to the result shown in Figure 4b. Here, only three restraints have an upper distance restraint which is too small by >0.5 Å. This supports our hypothesis and indicates that, instead using the initial-rate approximation, it might be more favorable to use the exact solution of the master equation ("full matrix approach").[50-52]

6. Conclusions

We have demonstrated, using ubiquitin as an example, that the information from a set of three PDSD spectra from a uniformly ^{13}C-, ^{15}N-labeled microcrystalline protein (76 amino acids) is sufficient to *de novo* determine its structure without using any additional information. The structure has been calculated manually, semiautomatically, and by a fully automated approach. The three spectra employed in the analysis were obtained from a sample containing 5−6 mg of protein within a total measurement time of 9 days using a 600 MHz spectrometer.

A detailed comparison of the quality of the distance restraints determined from the experimental cross-peak intensities indicates that the correlation between internuclear distance and cross-peak intensity is significant but relatively weak. Nevertheless, the dense network of restraints allows one to obtain the correct structure with a backbone rmsd value of about 0.7 Å. The analysis indicates that approaches going beyond the initial-rate approximation should significantly improve mapping of cross-peak intensities to internuclear distances.

The approach described is a simple procedure for 3D structure determination from a single sample of a uniformly labeled protein. The size of proteins amenable to this scheme is mainly restricted by the spectral resolution. Extension to additional spectral dimensions, however, should allow structure determination in significantly larger proteins. Because of the high SNR value, PDSD experiments are particularly suited as an element of 3D and 4D pulse sequences.

Supporting Information Available: List of manually picked peaks used as upper distance restraints in the manual structure calculations and a table summarizing indicating picked and assigned peaks for each PDSD experiment in the automatic approach. This material is available free of charge via the Internet at http://pubs.acs.org.

JA078039S

(50) Boelens, R.; Koning, T. M. G.; Kaptein, R. *J. Mol. Struct.* **1988**, *173*, 299−311.
(51) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1990**, *87* (3), 475−487.
(52) Linge, J. P.; Habeck, M.; Rieping, W.; Nilges, M. *J. Magn. Reson.* **2004**, *167* (2), 334−342.