

Proton-Detected Solid-State NMR Spectroscopy of Natural-Abundance Peptide and Protein Pharmaceuticals**

Donghua H. Zhou, Gautam Shah, Charles Mullen, Dennis Sandoz, and Chad M. Rienstra*

Over 80% of pharmaceutical products on the market, including many biopharmaceuticals, are in various solid forms.^[1–4] Solid-state NMR (SSNMR) provides atomic-resolution structural insight to such materials, and therefore is recognized as a powerful analytical approach to analyze formulation chemistry in pharmaceutical products.^[1] Dramatic progress in SSNMR over the last decade has enabled the determination of structures for several proteins, with the assistance of ^{13}C and ^{15}N isotope enrichment.^[5,6] However, owing to the limited experimental sensitivity inherent in the low natural abundance of ^{13}C (1.07%) and ^{15}N (0.368%), the application in pharmaceutical SSNMR research has been limited to small molecules.^[1] Furthermore, the one-dimensional ^{13}C spectra commonly acquired for small-molecule drugs have insufficient resolution for larger peptide and protein pharmaceuticals.

Herein we report a solution to these sensitivity and resolution problems by demonstrating resolved two-dimensional (^{15}N – ^1H and ^{13}C – ^1H) spectra of the therapeutic proteins aprotinin and insulin, acquired in half to one day with milligram quantities of sample. This result is remarkable because one-dimensional ^{15}N cross-polarization magic-angle spinning (CP-MAS) experiments require measurement times of at least tens of hours, and more likely several days to obtain similar signal-to-noise ratios. Because proteins are susceptible to physical and chemical degradation under many conditions (e.g., exposure to heating, freeze–thaw cycles, light, moisture, shaking, and hydrophobic surfaces), the ability to monitor changes efficiently and precisely will enable an improved understanding of the molecular basis of formulation toxicity and/or loss of efficacy.^[7] Improved SSNMR methods in this context enable analysis of detailed conformational changes (e.g. denaturation and aggregation) and chemical modifica-

tions (e.g. deamination, oxidation, and hydrolysis), and therefore offer a fundamentally new analytical tool for biochemists in the pharmaceutical industry. More generally, the higher throughput capability enables the study of natural-abundance proteins (that is, proteins that have not been enriched with isotopes, such as ^{13}C , ^{15}N) to gain insight into a variety of structure–function problems in modern molecular science.

We achieve a major sensitivity gain by employing proton detection in combination with the recently developed fast magic-angle spinning (MAS) technique^[8,9] and a highly optimized solvent suppression method.^[10] Proton detection, which is used almost ubiquitously in solution NMR spectroscopy,^[11] has traditionally been less often utilized in the solid state because of the broad resonance lines arising from strong proton dipolar couplings. One approach to address this challenge is proton dilution, in which all non-exchangeable protons are (bio)synthetically replaced with deuterons.^[9,12,13] This approach is inconvenient and cost-prohibitive for many chemical problems, and is undesirable for pharmaceutical analyses since it alters drug substance sameness.^[7,14,15] Alternatively, the fast (40 kHz) MAS approach has enabled us to demonstrate significant resolution and sensitivity enhancements for fully protonated small-molecule pharmaceuticals^[16] and proteins.^[8] However in that protein study ^{13}C and ^{15}N isotope enrichment was used,^[8] which suffers from the aforementioned shortcomings. Herein we demonstrate that the fast MAS approach can be practically applied to natural-abundance proteins within reasonable experimental times.

As a first demonstration, we examine lyophilized powders of bovine pancreatic trypsin inhibitor (BPTI, also known as aprotinin) from two suppliers (see Experimental Section). BPTI is a protease inhibitor that for many years has been administered (brand name Trasylol from Bayer AG, Germany) to reduce blood loss in high-risk cardiac surgeries until its recent temporary withdrawal from the market owing to a suspected increase in mortality.^[17] Most biopharmaceuticals, especially when initially marketed, are often lyophilized to maintain the long-term potency and reconstituted with aqueous diluent prior to administration.^[18] Although Trasylol is pre-reconstituted, lyophilized aprotinin powders are commonly sold for laboratory use. ^{15}N – ^1H and ^{13}C – ^1H heteronuclear correlation (HETCOR) spectra are shown in Figures 1a and b respectively for sample A and B; several slices are shown in Supporting Information Figure S1. Around 40 peaks for sample A (and around 50 for sample B) can be identified for this 58-residue protein in the ^{15}N – ^1H spectra. In rare cases where proton chemical shifts are assigned in the solid state (with enriched samples), such as protein GB1,^[8] HETCOR peaks for natural abundance can be specifically identified by

[*] Dr. D. H. Zhou,^[†] G. Shah, Prof. Dr. C. M. Rienstra
Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801 (USA)
Fax: (+1) 217-244-4655
E-mail: rienstra@scs.uiuc.edu
Homepage: <http://www.scs.uiuc.edu/~rienstra>

C. Mullen
Varian, Inc., Fort Collins, CO (USA)

D. Sandoz
Varian, Inc., Palo Alto, CA (USA)

[†] Current address: Department of Physics, Oklahoma State University, OK 74078 (USA)

[**] This research was supported by the National Institutes of Health (R01 GM-75937 to C.M.R.). We thank John A. Stringer and Mircea Cormos (Varian, Inc.) for advice on probe performance.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.200801029>.

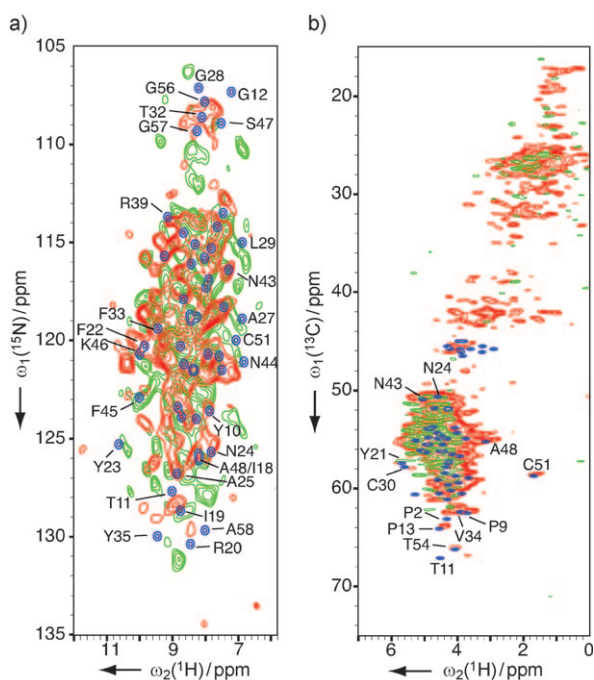


Figure 1. Solid-state HETCOR NMR spectra of natural-abundance lyophilized aprotinin sample A (red; 7.2 mg) and B (green; 3.5 mg) from different suppliers. Spectra in blue are simulated from solution chemical shifts (BMRB database No. 5359). a) ^{15}N - ^1H 2D spectra acquired in 26 (31) h with 640 (1024) scans per row, $t_{1\text{max}}(^{15}\text{N}) = 10$ (7.5) ms and linear-predicted to 20 (15) ms, $t_{2\text{max}}(^1\text{H}) = 3.4$ (3.4) ms for sample A (B). Lorentzian-to-Gaussian apodization was applied (150-to-300 Hz for ^1H , 15-to-30 Hz for ^{15}N) followed by sine bell apodization (54° shift) for each dimension. b) ^{13}C - ^1H 2D spectra acquired in 19 h with 48 scans per row for both samples. $t_{1\text{max}}(^{13}\text{C}) = 11$ ms and linear-predicted to 22 ms, $t_{2\text{max}}(^1\text{H}) = 3.4$ ms. Lorentzian-to-Gaussian apodization was applied (150-to-300 Hz ^1H , 40-to-80 Hz for ^{13}C) followed by sine bell apodization (54° shift) for each dimension. Contours are drawn starting at six times the noise level, with a spacing factor 1.15.

direct comparison (Supporting Information Figure S2). Since such information is not available for BPTI, hypothetical HETCOR spectra were constructed from solution chemical shifts (in blue, Figure 1). For both samples, the overall distribution of peaks agrees well with those predicted from solution chemical shifts, indicating the protein is still folded in the lyophilized powders. Moreover for sample A, several peaks including A25, A48, I18, N24, Y10, N43, A58 in Figure 1 a and C51, A48, Y21, T54 in Figure 1 b seem to have very similar chemical shifts to solution state. For sample B, Y23, F45, A27, C51, N44, A58 in Figure 1 a also have very similar chemical shifts to the solution state. However, sample B displays clear difference from sample A and solution state in these spectra. For instance, the ^{15}N high-field region ($\delta = 106$ to 112 ppm) that contains glycines, T32 and S47 is much more dispersed than the solution state and sample A (Figure 1 a). This result implies that the structure of sample B deviates more than sample A does from solution structure. We therefore expect HETCOR NMR spectroscopy to complement infrared, Raman, and mass spectroscopies in the study of protein lyophilization,^[4] as well as the processes

(e.g. milling, spray drying) used to prepare dry inhalable protein powders.^[19]

As a second demonstration of the technique, experiments were performed on three recombinant human insulin samples: lyophilized powders (sample C), fibrils (sample D), and crystalline suspension formulation (sample E). Insulin, the hormone used to treat diabetes which afflicts over 170 million people worldwide, consists of two peptide chains (A and B having 21 and 30 residues, respectively) covalently linked by two disulfide bonds. During prolonged storage, insulin is known to form insoluble fibrillar aggregates.^[20] The neutral protamine Hagedorn (NPH) insulin is an important formulation that provides longer action than regular insulin. It is a suspension of microcrystals co-crystallized in the presence of the positively charged peptide protamine as well as zinc.^[21] Figure 2 shows ^{15}N - ^1H and ^{13}C - ^1H spectra for the lyophilized (C) and fibril (D) insulin samples. Also shown are

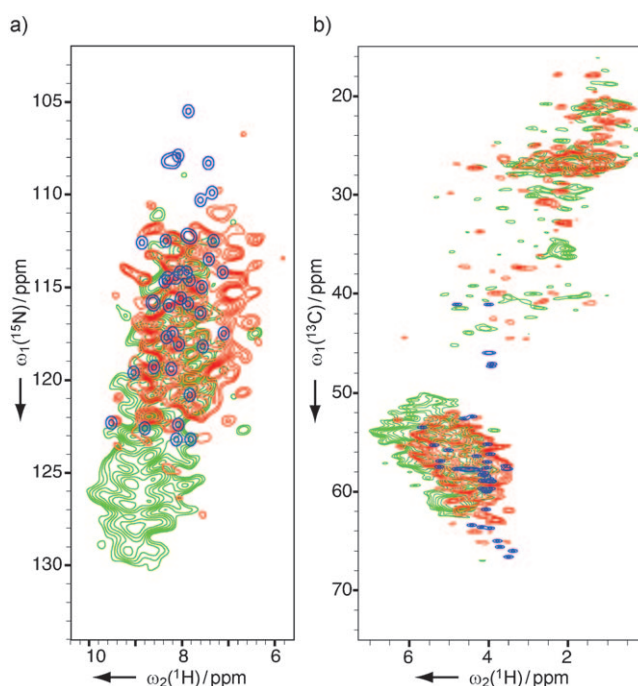


Figure 2. Solid-state HETCOR NMR spectra of natural-abundance recombinant human insulin samples C (red; 4.8 mg lyophilized powder) and green; 5.7 mg fibrils). Spectra in blue are simulated from solution chemical shifts (details in text). a) ^{15}N - ^1H 2D spectra acquired in 24 h with 640 scans per row. $t_{1\text{max}}(^{15}\text{N}) = 10$ ms and linear-predicted to 20 ms, $t_{2\text{max}}(^1\text{H}) = 3.4$ ms. b) ^{13}C - ^1H 2D spectra acquired in 19 h with 48 scans per row. $t_{1\text{max}}(^{13}\text{C}) = 11$ ms and linear-predicted to 22 ms, $t_{2\text{max}}(^1\text{H}) = 3.4$ ms. Apodization and contour display parameters are identical to Figure 1.

simulated spectra from solution chemical shifts for a mutant monomer (replacing F1, H10, Y16, T27 of the B chain with Glu and deleting T30 of B chain to avoid aggregation; for ^{15}N and amide ^1H shifts)^[22] and a phenol-stabilized R6 hexamer (for ^{13}C and $^1\text{H}^\alpha$).^[23] For the lyophilized sample C, most peaks in the ^{15}N - ^1H spectrum (Figure 2a) are at least partially resolved for this 51-residue protein. This sample displays a similar overall signal distribution with the solution data in

both the ^{15}N - ^1H and ^{13}C - ^1H spectra, indicating a similar folding. In contrast, the fibril sample D displays dramatically different spectra. Its ^{15}N - ^1H spectrum has strong low-field intensities (126 to 130 ppm in ^{15}N) that are not observed in sample C. Its ^1H in Figure 2b displace to lower field and ^{13}C to higher field, in agreement with the fact that insulin fibrils consist of mainly β -sheet structures.^[24]

Figure 3 shows HETCOR spectra for crystalline suspension formulation insulin (sample E). The extreme challenge of solvent suppression for this hydrated sample that is

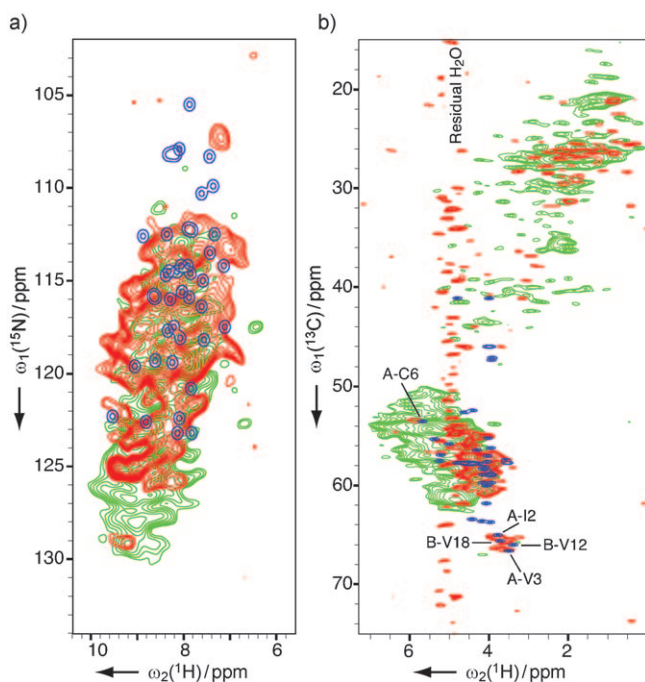


Figure 3. Solid-state HETCOR NMR spectra of formulated insulin sample E (red; containing 5 mg protein). a) ^{15}N - ^1H 2D spectrum acquired in 27 h with 896 scans per row, $t_{1\text{max}}(^{15}\text{N}) = 7.5$ ms and linear-predicted to 15 ms, $t_{2\text{max}}(^1\text{H}) = 3.4$ ms. b) ^{13}C - ^1H 2D spectrum acquired in 7 h with 16 scans, $t_{1\text{max}}(^{13}\text{C}) = 11$ ms and linear-predicted to 22 ms, $t_{2\text{max}}(^1\text{H}) = 3.4$ ms. No post-acquisition solvent suppression was applied. Apodization and contour display parameters are identical to Figure 1. Also shown are the insulin fibril spectra (sample D, green) as well as the simulated solution spectra (blue) from Figure 2.

amplified by the low natural abundance has been satisfactorily addressed by our recent MISSISSIPPI scheme;^[10] its performance is compared with a previous method in the Supporting Information Figure S2a and S2b. The ^{15}N - ^1H peak distribution in Figure 3a agrees in general with the monomer solution spectrum, meanwhile differences are evident in both low ($\delta = 124$ to 130 ppm) and high ($\delta = 105$ to 111 ppm) fields in the ^{15}N dimension. The ^{13}C - ^1H correlations in Figure 3b appear to agree very well with the solution spectrum. For example, residues C6, I2, V3 of the A chain and V12 and V18 of the B chain have very similar peak positions in the two states. This result is consistent with the fact that the protein exists as hexamers in both the crystalline formulation (zinc-stabilized T_6 hexamers^[25]) and the solution sample (phenol-stabilized R_6 hexamers^[26]). The formulation sample is further

compared with the fibril sample in Figure 3. Potential partial fibrillization in a formulation can be spotted from $\delta = 126.5$ to 128.5 ppm in the ^{15}N dimension of ^{15}N - ^1H spectra, these spectra are much more sensitive than one-dimensional ^{15}N spectra. Fibrillization can also be detected from $\delta = 5.5$ to 7 ppm in the ^1H dimension of ^{13}C - ^1H spectrum.

Although the chemical shifts of both insulin and BPTI cannot be uniquely assigned from these 2D spectra alone, the site resolution inherent to SSNMR enables a variety of statistical analysis methods to assess sameness, and as desired, full chemical shifts can be derived from separate, isotopically labeled samples. We anticipate further that homonuclear correlation schemes (e.g., ^1H - ^1H 2D or 3D spectra), as well as technical advances—such as higher magnetic field, faster MAS spinners, and digital receivers with enhanced dynamic range and digital filtering capabilities—will further enhance resolution and sensitivity of natural abundance SSNMR protein spectra.

Progress in this regard is particularly timely in the context of a rapidly expanding market for peptide and protein pharmaceuticals. Although there were only 40 synthetic peptide and 50 recombinant protein drugs on the market in 2001,^[3] the total number of peptide drugs and candidates reached 720 in 2004.^[27] In contrast to small-molecule drugs, the therapeutic peptides and proteins are usually more potent and specific but less stable.^[27] Thus the ability to assess chemical changes directly in the solid form is a particularly powerful tool for the pharmaceutical industry in development, manufacturing, and storage. Beyond pharmaceuticals, this approach can be applied to study all synthetic peptides, natural fibrous proteins (e.g. silk fibroins, hair keratin),^[28,29] as well as polymers.^[30] Finally, in pushing the limits of high-sensitivity proton detection of solid proteins, we envision that such 2D spectra could be acquired in the same amount of time with less than 50 micrograms of isotopically enriched material (10 nanomoles of a 5 kDa protein).

Experimental Section

Sample Preparation: Two lyophilized aprotinin samples were used in this study as received. Sample A (7.2 mg packed in NMR sample rotor) is from Sigma-Aldrich (St. Louis, MO) and sample B (3.5 mg) from Calbiochem (San Diego, CA). Three recombinant insulin samples (C, D, E) were used in this study. Sample C is 4.8 mg lyophilized recombinant insulin (ProSpec-Tany Techonogene Ltd., Rehovot, Israel) as received. Sample D is a fibril sample prepared by heating insulin solution (2 mM, pH 2.6) at 70 °C for 26 h according to literature.^[24] The gel formed was pelleted and dried under gentle N_2 flow; 5.7 mg was packed for NMR experiments. Sample E is Humulin N (Eli Lilly, Indianapolis, IN) purchased over the counter with 100 unit mL^{-1} (ca. 4 mgmL^{-1}) concentration in a 10 mL vial, from which 2 mL was drawn with a syringe and subsequently pelleted by centrifugation (3000 g for 3 min). After the supernatant solution was carefully removed by pipette, slightly more than half of the pellet was transferred to an NMR rotor. The 8.5 mg sample roughly consisted of 5 mg insulin, 0.5 mg protamine, 0.05 mg zinc, and 3 mg water.

NMR Experiments: The SSNMR experiments were performed at 7 °C (sample temperature calibrated with ethylene glycol)^[10] on a 750 MHz Varian INOVA spectrometer (Varian, Inc., Palo Alto, CA) with a FastMAS ^1H - ^{13}C - ^{15}N probe (Varian, Inc., Fort Collins, CO).

Probe specifications have been published earlier along with the heteronuclear correlation pulse sequence that employed MISSIS-SIPPI solvent suppression.^[10] Contact times were 1.6 and 0.6 ms for ^1H to ^{15}N and ^{15}N to ^1H cross polarization (CP) steps, respectively. Contact times were 1.2 and 0.25 ms for the ^1H to ^{13}C and ^{13}C to ^1H CP steps, respectively. The longer initial contact time maximizes total signal intensity on the ^{15}N or ^{13}C site, whereas the shorter second contact time primarily polarizes directly bonded protons. Experiments were performed at 36 kHz MAS rate.

Received: March 3, 2008

Revised: September 25, 2008

Published online: January 7, 2009

Keywords: insulin · magic-angle spinning · NMR spectroscopy · proteins · structure elucidation

- [1] P. A. Tishmack, D. E. Bugay, S. R. Byrn, *J. Pharm. Sci.* **2003**, *92*, 441–474.
- [2] S. Byrn, R. Pfeiffer, M. Ganey, C. Hoiberg, G. Poochikian, *Pharm. Res.* **1995**, *12*, 945–954.
- [3] A. Loffet, *J. Pept. Sci.* **2002**, *8*, 1–7.
- [4] W. Wang, *Int. J. Pharm.* **2000**, *203*, 1–60.
- [5] F. Castellani, B. V. Rossum, A. Diehl, M. Schubert, K. Rehbein, H. Oschkinat, *Nature* **2002**, *420*, 98–102.
- [6] C. Ritter, M. L. Maddelein, A. B. Siemer, T. Luhrs, M. Ernst, B. H. Meier, S. J. Saupe, R. Riek, *Nature* **2005**, *435*, 844–848.
- [7] A. K. Banga, *Therapeutic Peptides and Proteins: Formulation, Processing, and Delivery Systems*, 2nd ed., CRC Press/Taylor & Francis, New York, **2006**, pp. 67–83.
- [8] D. H. Zhou, G. Shah, M. Cormos, C. Mullen, D. Sandoz, C. M. Rienstra, *J. Am. Chem. Soc.* **2007**, *129*, 11791–11801.
- [9] D. H. Zhou, J. J. Shea, A. J. Nieuwkoop, W. T. Franks, B. J. Wylie, C. Mullen, D. Sandoz, C. M. Rienstra, *Angew. Chem.* **2007**, *119*, 8532–8535; *Angew. Chem. Int. Ed.* **2007**, *46*, 8380–8383.
- [10] D. H. Zhou, C. M. Rienstra, *J. Magn. Reson.* **2008**, *192*, 167–172.
- [11] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, *69*, 185–189.
- [12] V. Chevelkov, B. J. van Rossum, F. Castellani, K. Rehbein, A. Diehl, M. Hohwy, S. Steuernagel, F. Engelke, H. Oschkinat, B. Reif, *J. Am. Chem. Soc.* **2003**, *125*, 7788–7789.
- [13] E. K. Paulson, C. R. Morcombe, V. Gaponenko, B. Dancheck, R. A. Byrd, K. W. Zilm, *J. Am. Chem. Soc.* **2003**, *125*, 15831–15836.
- [14] A. S. Raw, M. S. Furness, D. S. Gill, R. C. Adams, F. O. Holcombe, L. X. Yu, *Adv. Drug Delivery Rev.* **2004**, *56*, 397–414.
- [15] T. Layloff, *Pharm. Technol.* **1991**, *15*, 146–148.
- [16] D. H. Zhou, C. M. Rienstra, *Angew. Chem.* **2008**, *120*, 7438–7441; *Angew. Chem. Int. Ed.* **2008**, *47*, 7328–7331.
- [17] A. Sedrakyan, T. Treasure, J. A. Elefteriades, *J. Thorac. Cardiovasc. Surg.* **2004**, *128*, 442–448.
- [18] J. J. Schwegman, L. M. Hardwick, M. J. Akers, *Pharm. Dev. Technol. Pharm. Dev. Tech.* **2005**, *10*, 151–173.
- [19] K. A. Johnson, *Adv. Drug Delivery Rev.* **1997**, *26*, 3–15.
- [20] J. Brange, L. Langkjoer, *Pharm. Biotechnol.* **1993**, *5*, 315–350.
- [21] M. Norrman, F. Hubalek, G. Schluckebier, *Eur. J. Pharm. Sci.* **2007**, *30*, 414–423.
- [22] H. B. Olsen, PhD thesis, Roskilde University (Denmark), **1996**.
- [23] X. Chang, A. M. M. Jorgensen, P. Bardrum, J. J. Led, *Biochemistry* **1997**, *36*, 9409–9422.
- [24] M. Bouchard, J. Zurdo, E. J. Nettleton, C. M. Dobson, C. V. Robinson, *Protein Sci.* **2000**, *9*, 1960–1967.
- [25] T. Karlsson, J. M. Popham, N. O. J. R. Long, G. P. Drobny, *J. Am. Chem. Soc.* **2003**, *125*, 7394–7407.
- [26] N. C. Kaarsholm, H. C. Ko, M. F. Dunn, *Biochemistry* **1989**, *28*, 4427–4435.
- [27] V. Marx, *Chem. Eng. News* **2005**, *83*(11), 17–24.
- [28] N. Nishikawa, Y. Horiguchi, T. Asakura, I. Ando, *Polymer* **1999**, *40*, 2139–2144.
- [29] A. Shoji in *Modern Magnetic Resonance, Part 1* (Ed.: G. A. Webb), Springer, New York, **2006**, pp. 587–600.
- [30] Y. Ishii, J. P. Yesinowski, R. Tycko, *J. Am. Chem. Soc.* **2001**, *123*, 2921–2922.