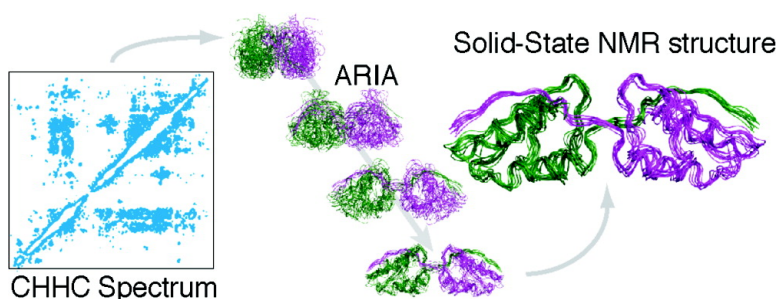**Article**

3D Structure Determination of the Crh Protein from Highly Ambiguous Solid-State NMR Restraints

Antoine Loquet, Benjamin Bardiaux, Carole Gardiennet, Christophe Blanchet, Marc Baldus, Michael Nilges, Thrse Malliavin, and Anja Bckmann

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

3D Structure Determination of the Crh Protein from Highly Ambiguous Solid-State NMR Restraints

Antoine Loquet,[‡] Benjamin Bardiaux,[#] Carole Gardiennet,[‡] Christophe Blanchet,[‡] Marc Baldus,[§] Michael Nilges,[#] Thérèse Malliavin,[#] and Anja Böckmann*,[‡]

Institut de Biologie et Chimie des Protéines, UMR 5086 CNRS Université Lyon 1, IFR128 BioSciences Lyon-Gerland, 7, passage du Vercors, 69367 Lyon, France, Unité de Bio-Informatique Structurale, URA 2185 CNRS, Institut Pasteur, 25-28 rue du docteur Roux, F-75015 Paris, France, and Department of NMR-based Structural Biology, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Received October 19, 2007; E-mail: a.bockmann@ibcp.fr

Abstract: In a wide variety of proteins, insolubility presents a challenge to structural biology, as X-ray crystallography and liquid-state NMR are unsuitable. Indeed, no general approach is available as of today for studying the three-dimensional structures of membrane proteins and protein fibrils. We here demonstrate, at the example of the microcrystalline model protein Crh, how high-resolution 3D structures can be derived from magic-angle spinning solid-state NMR distance restraints for fully labeled protein samples. First, we show that proton-mediated rare-spin correlation spectra, as well as carbon-13 spin diffusion experiments, provide enough short, medium, and long-range structural restraints to obtain high-resolution structures of this 2×10.4 kDa dimeric protein. Nevertheless, the large number of $^{13}C/^{15}N$ spins present in this protein, combined with solid-state NMR line widths of about 0.5−1 ppm, induces substantial ambiguities in resonance assignments, preventing 3D structure determination by using distance restraints uniquely assigned on the basis of their chemical shifts. In the second part, we thus demonstrate that an automated iterative assignment algorithm implemented in a dedicated solid-state NMR version of the program ARIA permits to resolve the majority of ambiguities and to calculate a de novo 3D structure from highly ambiguous solid-state NMR data, using a unique fully labeled protein sample. We present, using distance restraints obtained through the iterative assignment process, as well as dihedral angle restraints predicted from chemical shifts, the 3D structure of the fully labeled Crh dimer refined at a root-mean-square deviation of 1.33 Å.

Introduction

The application of the two main analytical methods for 3D structure determination, X-ray crystallography, and liquid-state nuclear magnetic resonance spectroscopy (NMR), remains challenging for membrane proteins and protein fibrils. Even if membrane proteins can often be solubilized and sometimes be crystallized in the presence of detergents, their structural analysis in phospholipid bilayers is difficult by these methods. With regard to protein fibers, up to now only very short stretches of amino acids were amenable to crystallization, and only few studies exist on structural features of fibers from full-length proteins by classical approaches. Solid-state NMR (SSNMR) spectroscopy can address questions on structure, dynamics, and interactions of insoluble proteins (for recent reviews see references 1−4). The use of fully or extensively isotopically labeled proteins provides the opportunity to study these features using a single sample.

3D structure determination by NMR requires the measurement of long-range restraints in order to determine the fold of the protein. These restraints can be measured between nitrogen-15, carbon-13, and/or proton spins. Correlations between carbon-13 spins can be recorded directly by several types of multidimensional experiments. In SSNMR, polarization transfer is however dominated by the strong dipolar couplings between covalently bonded carbon spins. It has been shown recently that the resulting dipolar truncation is less severe in proton-driven spin-diffusion experiments,[5] so that the long-range ($|i - j| > 4$) correlations which form the basis of protein fold determination can in principle be extracted from this type of spectra, even if the correlations observed largely originate from intra-residue and sequential connections. However, as the distribution of polarization within one amino acid proceeds fast compared to inter-residue polarization transfer, relayed mechanisms may have an important contribution to the observed cross-signals. This leads to less precision in the distance restraints derived from this type of spectra. Structure determination from carbon−carbon distances has been limited up to date to proteins labeled with strategic schemes, resulting in samples where most carbon-13 spins are neighbors to carbon-12 spins.[6−9]

‡ UMR 5086 CNRS Université Lyon 1.
# Institut Pasteur.
§ Max-Planck-Institute for Biophysical Chemistry.
(1) McDermott, A. E. *Curr. Opin. Struct. Biol.* **2004**, *14*, 554−561.
(2) Luca, S.; Heise, H.; Baldus, M. *Acc. Chem. Res.* **2003**, *36*, 858−865.
(3) Böckmann, A. *C. R. Chim.* **2006**, *9*, 381−392.
(4) Baldus, M. *Curr. Opin. Struct. Biol.* **2006**, *16*, 618−623.

(5) Grommek, A.; Meier, B. H.; Ernst, M. *Chem. Phys. Lett.* **2006**, *427*, 404−409.

3D structures of small peptides have also been probed by selective recoupling schemes involving spin pairs. These methods can require however a large amount of selective experiments to be performed. Band-selective experiments have been introduced to selectively recouple spin-pairs in fully labeled peptides to determine multiple distances.[10−13] A high-resolution structure of the model tripeptide f-MLF-OH[14] and of the TTR-[105-115] peptide[15] could be determined using this approach. Developed for small peptides, this technique remains difficult to apply to larger systems, where the poor spectral dispersion presents a major obstacle for chemical-shift selection, and where multispin effects lead to complex correlations between polarization exchange magnitude and distance.

Valuable information can also be obtained from measuring proton−proton distances. Recent work on spin-diluted proteins,[16−18] and also on fully protonated samples[19] at high spinning frequencies, points to the possibility to resolve and to directly detect them in near future. Another possibility, proposed by Baldus and co-workers in a series of experiments,[20−22] is to indirectly probe ^{1}H−^{1}H interactions by a detection in 2D homonuclear (CHHC/NHHN), heteronuclear (NHHC), or 3D (CCHHC/NHHCC) correlation spectra. In contrast to carbon correlations, where the dominant one-bond and two-bond correlations do not carry structural information, the large majority of correlations in proton-mediated spectra yield valuable distance restraints. It has recently been demonstrated that proton-mediated, rare-spin detected correlation spectra yield enough restraints for the determination of the overall fold and the characteristic secondary structure elements for Ubiquitin[23] and Kaliotoxin.[24]

One of the biggest challenges in both rare-spin and proton-mediated correlation experiments remains the unambiguous assignment of the peaks to distance restraints between two spins. All recent efforts for structure determination focused on relatively small proteins, while for larger proteins resonances could only be assigned assuming homology models,[23] or in combination with strategic labeling schemes which reduce the number of resonances.[8,9] In solution NMR, the reduced line-width leads to a high accuracy of peak positions and experimental chemical shift values and allows, in the case of folded proteins of a size similar to the size of the catabolite repression HPr-like protein (Crh), to assign a large fraction of the 2D-NOESY or 13C/15N-edited NOESY cross-peaks to a single and unique spin pair, the equivalent spins being considered as unique. In solid-state NMR, detection on rare spins (^{13}C or ^{15}N) yields line widths of about 0.5 to 1 ppm. Compared to solution NMR, the number of isolated cross-peaks is greatly reduced, and the size of the chemical shift tolerance window must be increased. As a consequence, the majority of cross-peaks in ^{13}C−^{13}C or ^{15}N−^{13}C correlation spectra remain ambiguously assigned. In solution NMR, the use of ambiguous assignments[25] has been developed for the study of proteins larger than 10 kDa, as well as for the study of multimers, to account for the possibility of intramonomer and intermonomer cross-peak assignments.[26] The use of ambiguous distance restraints (ADRs) is widespread today to handle ambiguities in solution NMR (for applications see for example references 27−29). In the automated structure calculation protocol ARIA,[30] an iterative procedure resolves ambiguous assignments by calculating sets of structures with ADRs, and applying a structure-based filter to discard the least probable assignments. The use of ADRs in the context of solid-state NMR data has been previously explored for the assignment of checkerboard-labeled proteins using a dedicated algorithm (SOLARIA).[31]

We use here as a model the 2 × 10.4 kDa Crh protein, which structure, determined by us using X-ray diffraction, shows a domain-swapped dimer.[32] For this protein, the structure determination problem consists in the monomer structure determination, as well as in the relative orientation of the two monomers to form the dimer. The latter presents a difficult problem in NMR,[26,33,34] and shall not be a subject discussed in this work; we here focus on the accurate determination of protein structure using SSNMR restraints.

The scope of the present work is twofold. First, we evaluate the information content of solid-state NMR proton-mediated, as well carbon-13, 2D correlation experiments and demonstrate that both contain enough distance restraints for the determination of a high-resolution 3D structure of the Crh model protein. In this first part, calculations are performed with distance restraints identified via the X-ray structure of the Crh dimer. In a second part, we evaluate whether de novo structure determination is possible from solid-state NMR data using the concept of

(6) Hong, M.; Jakes, K. *J. Biomol. NMR* **1999**, *14*, 71−74.
(7) Hong, M. *J. Magn. Reson.* **1999**, *139*, 389−401.
(8) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98−102.
(9) Zech, S. G.; Wand, A. J.; McDermott, A. E. *J. Am. Chem. Soc.* **2005**, *127*, 8618−8626.
(10) Ladizhansky, V.; Jaroniec, C. P.; Diehl, A.; Oschkinat, H.; Griffin, R. G. *J. Am. Chem. Soc.* **2003**, *125*, 6827−6833.
(11) Ladizhansky, V.; Griffin, R. G. *J. Am. Chem. Soc.* **2004**, *126*, 948−958.
(12) Jaroniec, C. P.; Filip, C.; Griffin, R. G. *J. Am. Chem. Soc.* **2002**, *124*, 10728−10742.
(13) Rienstra, C. M.; Hohwy, M.; Mueller, L. J.; Jaroniec, C. P.; Reif, B.; Griffin, R. G. *J. Am. Chem. Soc.* **2002**, *124*, 11908−11922.
(14) 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*, 10260−10265.
(15) 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*, 711−716.
(16) Paulson, E. K.; Morcombe, C. R.; Gaponenko, V.; Dancheck, B.; Byrd, R. A.; Zilm, K. W. *J. Am. Chem. Soc.* **2003**, *125*, 15831−15836.
(17) Chevelkov, V.; van Rossum, B. J.; Castellani, F.; Rehbein, K.; Diehl, A.; Hohwy, M.; Steuernagel, S.; Engelke, F.; Oschkinat, H.; Reif, B. *J. Am. Chem. Soc.* **2003**, *125*, 7788−7789.
(18) Zhou, D. H.; Shea, J. J.; Nieuwkoop, A. J.; Franks, W. T.; Wylie, B. J.; Mullen, C.; Sandoz, D.; Rienstra, C. M. *Angew. Chem., Int. Ed.* **2007**.
(19) Zhou, D. H.; Shah, G.; Cormos, M.; Mullen, C.; Sandoz, D.; Rienstra, C. M. *J. Am. Chem. Soc.* **2007**, *129*, 11791−801.
(20) Lange, A.; Luca, S.; Baldus, M. *J. Am. Chem. Soc.* **2002**, *124*, 9704−9705.
(21) Lange, A.; Seidel, K.; Verdier, L.; Luca, S.; Baldus, M. *J. Am. Chem. Soc.* **2003**, *125*, 12640−12648.
(22) Heise, H.; Seidel, K.; Etzkorn, M.; Becker, S.; Baldus, M. *J. Magn. Reson.* **2005**, *173*, 64−74.
(23) Seidel, K.; Etzkorn, M.; Heise, H.; Becker, S.; Baldus, M. *Chembiochem* **2005**, *6*, 1638−1647.
(24) Lange, A.; Becker, S.; Seidel, K.; Giller, K.; Pongs, O.; Baldus, M. *Angew. Chem., Int. Ed.* **2005**, *44*, 2−5.

(25) Linge, J. P.; O'Donoghue, S. I.; Nilges, M. *Methods Enzymol.* **2001**, *339*, 71−90.
(26) Nilges, M. *Proteins* **1993**, *17*, 297−309.
(27) Luh, F. Y.; Archer, S. J.; Domaille, P. J.; Smith, B. O.; Owen, D.; Brotherton, D. H.; Raine, A. R.; Xu, X.; Brizuela, L.; Brenner, S. L.; Laue, E. D. *Nature* **1997**, *389*, 999−1003.
(28) Aghazadeh, B.; Zhu, K.; Kubiseski, T. J.; Liu, G. A.; Pawson, T.; Zheng, Y.; Rosen, M. K. *Nat. Struct. Biol.* **1998**, *5*, 1098−1107.
(29) Mott, H. R.; Owen, D.; Nietlispach, D.; Lowe, P. N.; Manser, E.; Lim, L.; Laue, E. D. *Nature* **1999**, *399*, 384−388.
(30) Nilges, M.; Macias, M. J.; O'Donoghue, S. I.; Oschkinat, H. *J. Mol. Biol.* **1997**, *269*, 408−422.
(31) Fossi, M.; Castellani, F.; Nilges, M.; Oschkinat, H.; van Rossum, B. J. *Angew. Chem., Int. Ed.* **2005**, *44*, 6151−6154.
(32) Juy, M.; Penin, F.; Favier, A.; Galinier, A.; Montserret, R.; Haser, R.; Deutscher, J.; Böckmann, A. *J. Mol. Biol.* **2003**, *332*, 767−776.
(33) Bewley, C. A.; Clore, G. M. *J. Am. Chem. Soc.* **2000**, *122*, 6009−6016.
(34) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337−2338.

ambiguous distance restraints (ADRs)[26,35,36] for automated 3D structure calculations. We show that nearly exclusively ambiguous $^1H-^1H$ contacts detected on ^{13}C and ^{15}N spins in CHHC and NHHC spectra of the fully labeled $^{13}C/^{15}N$ Crh protein can be automatically assigned in an iterative manner, and can be used to obtain a high-resolution 3D structure of this 2×10.4 kDa protein. Here, calculations are based on peak lists derived from the SSNMR spectra. This set of calculations comprises two steps: (i) ARIA is used to iteratively determine unambiguous restraints from the SSNMR peak lists; (ii) the structure is calculated using XPLOR-NIH using the unambiguous restraints determined by ARIA.

Materials and Methods

Sample Preparation. Crh was overexpressed with a C-terminal LQ-(6xHis) extension as described previously.[37] ^{13}C and ^{15}N-enriched Crh was prepared by growing bacteria in >98% ^{13}C, ^{15}N labeled medium (Silantes). The protein was purified on Ni-NTA agarose (QUIAGEN) columns followed by anion exchange chromatography on a Resource Q column.[38] Crh-containing fractions were dialyzed against 20 mM NH_4HCO_3. The protein was crystallized as described previously[39] in the presence of 20% PEG 6000 in a crystallization plate over a 2 M NaCl solution. The microcrystals resulting from about 20 mg of protein were centrifuged directly into a 4 mm CRAMPS rotor, and the rotor cap was sealed.

NMR Spectroscopy. NMR experiments were performed on a Bruker AVANCE DSX 500 MHz wide-bore spectrometer, equipped with double/triple resonance Bruker magic angle spinning (MAS) probes, at a spinning frequency of 11 kHz. All experiments were carried out between 0 and 5 °C sample temperature. A ramped cross-polarization[40,41] was used in all experiments to transfer proton polarization to or from the ^{13}C or ^{15}N spins. High-power proton decoupling using the SPINAL-64 decoupling scheme[42] was applied during evolution and detection periods. The relaxation delay between scans was 2.5 s. Proton–proton contacts were encoded using two-dimensional (^{13}C, ^{13}C) (CHHC) or (^{15}N, ^{13}C) (NHHC) correlation experiments.[20,21,43] The 2D CHHC correlation spectrum was recorded using a first ^1H to ^{13}C 1 ms cross-polarization (CP) followed by two short CP steps of either 125 μs and 200 μs (^1H, ^1H) mixing time. Acquisition times were 15 ms in t_2 and 7.9 ms in t_1, respectively, corresponding to a total acquisition time of 46 h. The spectral width was 350 ppm in the acquisition dimension and 90 ppm in the indirect dimension, centered in the aliphatic region. The ^1H rf field during SPINAL-64 decoupling and CP was set to 71 and 56 kHz, respectively. The ^{13}C rf field during CP was 34 kHz. The 2D NHHC correlation spectrum was recorded using a first ^1H to ^{15}N transfer of 900 μs. The subsequent CP contact times t_{NH} and t_{CH} were, respectively, 200 and 100 μs long. A proton mixing time of 100 μs was chosen. Acquisition times were 30 ms in t_2 and 10.5 ms in t_1. The total acquisition time was 35 h. The spectral widths in the ^{13}C and ^{15}N dimensions were 350 and 60 ppm, respectively. The

^1H decoupling power was set to 71 kHz. The rf field strengths for ^{15}N and ^{13}C during CP were 38 and 34 kHz, respectively.

For the DARR[44] spectrum, a ramped CP[40,41] of 1 ms was used to transfer proton polarization to the ^{13}C spins. High power proton decoupling using the SPINAL-64 decoupling scheme[42] was applied during evolution and detection periods. The relaxation delay between scans was 2.5 s. During the mixing time of 200 ms, the radio frequency field on the proton channel matched the spinning frequency (11 kHz). Acquisition times were 20 ms in t_2 and 7.3 ms in t_1, respectively, corresponding to a total acquisition time of 21 h. The spectral width was 250 ppm in both dimensions.

Prediction of Dihedral Angle Restraints. The TALOS[45] software was used to predict torsion angles from N, $C\alpha$, $C\beta$, and C′ chemical shifts.[39] Dihedral angle predictions for 56 out of 85 residues were considered as "good" by TALOS and used as dihedral angle restraints with error margins given by the program.

Structure Calculations with SSNMR Distance Restraints Identified Using the X-ray Structure. $^1H-^1H$ and $^{13}C-^{13}C$ distance restraints were identified using the X-ray crystallographic structure of Crh as a homology model. Restraints for proton–proton distances were identified in proton detected, $^{13}C/^{15}N$ spin detected experiments (CHHC and NHHC). $^1H-^1H$ restraints were defined by a common lower bound of 1.8 Å, a target distance of 3.5 Å for proton–proton distances identified in the NHHC spectrum, and 3.8 Å for distances identified in the CHHC spectrum, with a common upper bound of 5.0 Å. No distance classes were established, as suggested by our previous analysis of the correlation between polarization transfer build-up curves and distances measured on the Crh X-ray structure.[46] $^{13}C-^{13}C$ distance restraints were identified in the DARR experiment, and bounds for these restraints were set to 2.5–7.0 Å.

Calculations of the monomer structure were realized using the program CNS (crystallography and NMR system) version 1.1.[47] A molecular dynamics simulated annealing protocol was used with torsion angles as internal degrees of freedom.[48,49] The structure calculation protocol consists of three stages: (i) 3000 steps of high-temperature torsion angle molecular dynamics at 50000 K; (ii) 5000 steps of slow-cooling annealing stage in torsion angle space from 50000 K to 0 K and (iii) 10 cycles of final conjugate gradient minimization, each cycle comprising 200 steps.

Assignments of Ambiguous Distance Restraints by an Automated Iterative Process. To assign cross-peaks from SSNMR spectra without the use of a homology model, we used a dedicated solid-state NMR version of the program ARIA 2.2, which can be obtained from Michael Nilges (nilges@pasteur.fr).

As in CHHC and NHHC experiments proton–proton interactions are indirectly detected on rare spins, this SSNMR version of ARIA allows the use of ^{13}C and ^{15}N chemical shifts, encoding ^1H-^1H contacts. Chemical shifts, as well as cross-peak lists from CHHC and NHHC spectra, were used as input for the program. Manual peak picking was realized using the program Sparky 3.1 (T. D. Goddard and D. G. Kneller, University of California). Additionally, a set of 25 intermonomer distance restraints[50] previously identified in a uniformly but heterogeneously [^{13}C:^{15}N] labeled protein Crh was used in these calculations, as well as the dihedral angle restraints from TALOS. In the ARIA program, the ambiguity in peak assignments is handled

(35) Nilges, M. *J. Mol. Biol.* **1995**, *245*, 645–660.

(36) Junius, F. K.; O'Donoghue, S. I.; Nilges, M.; Weiss, A. S.; King, G. F. *J. Biol. Chem.* **1996**, *271*, 13663–13667.

(37) Galinier, A.; Haiech, J.; Kilhoffer, M. C.; Jaquinod, M.; Stulke, J.; Deutscher, J.; Martin-Verstraete, I. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8439–8444.

(38) Penin, F.; Favier, A.; Montserret, R.; Brutscher, B.; Deutscher, J.; Marion, D.; Galinier, A. *J. Mol. Microbiol. Biotechnol.* **2001**, *3*, 429–432.

(39) 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.

(40) Metz, G.; Wu, X. L.; Smith, S. O. *J. Magn. Reson. Ser. A* **1994**, *110*, 219–227.

(41) Hediger, S.; Meier, B. H.; Ernst, R. R. *Chem. Phys. Lett.* **1995**, *240*, 449–456.

(42) Fung, B. M.; Khitrin, A. K.; Ermolaev, K. *J. Magn. Reson.* **2000**, *142*, 97–101.

(43) Heise, H.; Hoyer, W.; Becker, S.; Andronesi, O. C.; Riedel, D.; Baldus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15871–15876.

(44) Takegoshi, K.; Nakamura, S.; Terao, T. *Chem. Phys. Lett.* **2001**, *344*, 631–637.

(45) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289–302.

(46) Gardiennet, C.; Loquet, A.; Etzkorn, M.; Heise, H.; Baldus, M.; Böckmann, A., *J. Biomol. NMR*, in press.

(47) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr., D* **1998**, *54*, 905–921.

(48) Rice, L. M.; Brunger, A. T. *Proteins* **1994**, *19*, 277–290.

(49) Stein, E. G.; Rice, L. M.; Brunger, A. T. *J. Magn. Reson.* **1997**, *124*, 154–164.

(50) Etzkorn, M.; Böckmann, A.; Lange, A.; Baldus, M. *J. Am. Chem. Soc.* **2004**, *126*, 14746–14751.

through the use of ambiguous distance restraints (ADR).[30] As the Crh protein is a homodimer, ambiguous distance restraints were also used to deal with the ambiguity between intra- and intermonomer restraints, using an effective distance of $D = [^N\Sigma(d_{intra}^{-6} + d_{inter}^{-6})]^{-1/6}$, where d_{intra} and d_{inter} correspond to the intra- and intermonomer distance, respectively. During the Crh dimer calculations, a packing distance restraint was applied between the centers of mass of the monomers, and a pseudoenergy term, the noncrystallographic symmetry (NCS) restraint,[43] was used to keep the two monomer units superimposable by minimizing the atomic rmsd between the two monomer units which form the dimer. The 2-fold symmetry was enforced through distance symmetry restraints.[26] The CNS torsion angle dynamics protocol used in ARIA consists for each iteration in (i) 1000 steps of an initial high-temperature conformational search at 2000 K, followed by (ii) two cooling stages in which the temperature is linearly decreased from 2000 to 1000 K and then from 1000 to 50 K. Considering the high degree of ambiguity of the NMR-derived restraints,[51] we used slow cooling stages: 80000/64000 for the first run, and 60000/48000 for the second run.

No automated peak-picking routine is implemented in ARIA. Peak-picking was realized manually with the program SPARKY. The threshold used was 2 times the noise measured on the spectra. The iterative assignment of restraints is performed automatically by ARIA, starting from the peak list determined using SPARKY. At each iteration, each peak is converted to a distance restraint, which is the weighted sum of several contributions between isolated protons. During the iterative algorithm, each contribution is filtered by rejecting the systematically violated contributions. The input values of the algorithm are the tolerance added to the upper and lower bounds before checking that the contribution is violated (violation_tolerance varying from 1000 down to 0.1 Å), the threshold to determine whether a restraint is violated (violation_threshold = 0.5 Å), the threshold for the contribution weigth (weight_threshold varying from 1.0 down to 0.8), and the maximum number of contributions (max_contributions = 25).[52]

Structure Calculations from Unambiguous Restraints Determined Using ARIA. We used all unambiguous restraints assigned in the last iteration of the ARIA run as input for a final 3D structure calculation round using XPLOR-NIH.[53] Calculations started with the generation of a random monomer structure with good local geometry, followed by the duplication of the monomeric unit and a rotation of 180° around one of the internal axes to obtain a symmetric dimer. For each of the 200 random dimer structures generated, calculations proceeded through three stages: (i) a high-temperature searching phase at 2000 K (40000 steps), (ii) a annealing stage from 2000 to 100 K in temperature steps of 50 K and (iii) a final gradient minimization of 500 cycles of Powell minimization. During the Crh dimer calculations, a pseudoenergy term, the noncrystallographic symmetry (NCS) restraint,[43] was used to keep the two monomer units superimposable by minimizing the atomic rmsd between the two monomer units which form the dimer. The 10 lowest-energy conformers, as well as the NMR restraint data file, were deposited to the Protein Data Bank (PDB id: 2RLZ).

Analysis of Obtained Structures. The 10 selected conformers were aligned on the backbone atoms using MOLMOL 2K.2.[54] The software PROCHECK[55] was used to analyze the quality of the obtained conformers. The number of violations larger than 0.5, 0.3, and 0.1 Å, as well as the maximum violation and the rms of violations were also determined. We assessed the convergence by calculating the rmsd

(51) Fossi, M.; Oschkinat, H.; Nilges, M.; Ball, L. J. *J. Magn. Reson.* **2005**, *175*, 92−102.
(52) Habeck, M.; Rieping, W.; Linge, J. P.; Nilges, M. *Methods Mol. Biol.* **2004**, *278*, 379−402.
(53) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65−73.
(54) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51−55.
(55) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
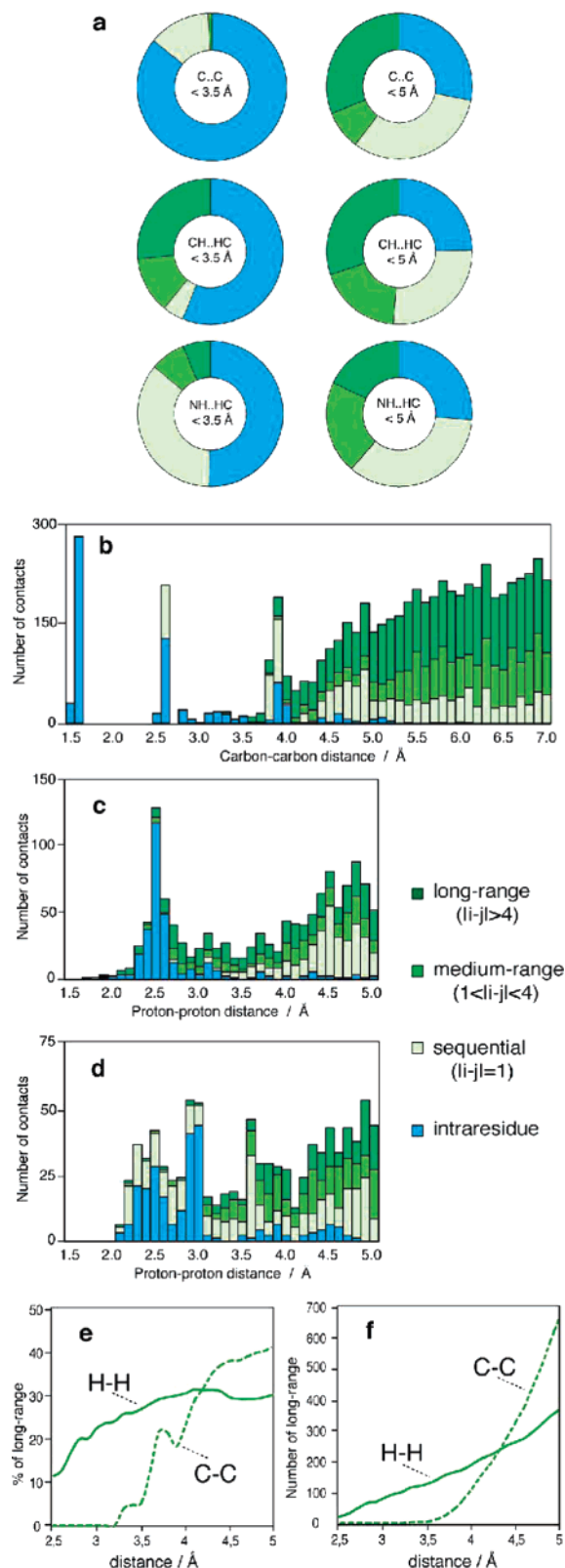
between the conformers superimposed either on a hypothetical monomer (residues 2−12 of chain A, and 13−80 of chain B) or the complete dimer. We compared the distance between the centers of mass of the monomers, as well as the angles between the two monomers, to the corresponding values in the crystallographic structure.

Results and Discussion

Analysis of Proton and Carbon Contacts in the Crh Dimer. We first evaluate the number and types of contacts found for different distance cut-offs in the Crh dimer, and compare them for protons and carbons, to assess which information content can be expected from the different spectra. Crh is 85 residues dimeric protein, with a tertiary structure formed by a four-stranded antiparallel β-sheet ($\beta1-4$), and three α-helices A, B, and C. In the hinge region, an intermolecular short $\beta1a$-sheet is formed. Solid-state NMR sequential assignments and a thereon-based structural analysis of a microcrystalline form of the protein[39] have revealed that the dimeric domain-swapped form is present in this preparation as well. We thus used atomic coordinates from the 1.8 Å crystal structure (PDB entry: 1MU4[32]) to extract heavy-atom internuclear distances. Distances between protons were obtained by adding hydrogen atoms to the crystal structure using the MOLMOL program.[54] We used a cutoff of 5 Å for proton−proton distances[20,21] and of 7 Å for carbon−carbon distances.[8]

Each signal in carbon, as well as in heteronuclei detected proton correlation spectra, yield information about the spatial proximities between two spins. If spectra are crowded by short-range contacts carrying little structural information, identification of the medium- and long-range contacts becomes a difficult task. Their percentage consequently determines the amount of structural information that can be extracted. Figure 1a shows the distribution of distances involving either carbon or proton spins between short, medium and long-range as well as intraresidue contacts, with a distance cutoff of 3.5 and 5 Å for C−C, CH−HC, and NH−HC internuclear distances. For carbon spins, one bond and two bond contacts are highly dominant below 3.5 Å, as they represent more than 95% of potentially observable correlations in the C−C correlation spectra for this cutoff. The distribution for CH−HC contacts is different: even for a cutoff of 3.5 Å, long and medium range contacts account for a high percentage (about 40%). In NHHC spectra, the percentage of intraresidue contacts is even smaller, and a large number of sequential contacts are present. One should notice that, in contrast to carbon correlations, where one- and two-bond transfers proceed extremely fast within one residue and hamper distance calibration, in proton-mediated spectra even intraresidue contacts carry structural information.[46] For longer distance cut-offs (Figure 1a, right panel), the proportion of intraresidue/sequential and of medium/long-range contacts is similar for the three types of spectra, with slightly more long- and medium-range contacts still found in CHHC spectra. Figures 1b−d show a quantitative analysis of the number of contacts observed in the Crh X-ray structure as a function of the internuclear distance. For carbon correlations (Figure 1b), a large number of contacts (about 500) are present for one- and two-bond transfers. For proton contacts, intraresidue correlations dominate up to 2.5 in CHHC (Figure 1c) and 3 Å in NHHC (Figure 1d) experiments, and long- and medium-range contacts show an onset at shorter distances than for carbon spins (Figure 1b). The percentage (Figure 1e) and the total number of long-

Figure 1. (a) Distribution between long-range (dark green), medium-range (green), sequential (light green), and intraresidue (blue) contacts involving carbon spins (C···C) and proton spins (CH−HC and NH−HC) in the Crh protein for distance cut-offs of 3.5 (left) and 5 Å (right). Number of contacts between carbon spin pairs (b), between proton spin pairs bonded to carbon atoms (c), and proton pairs with one proton bonded to a nitrogen spin and one proton bonded to a carbon spin (d), shown as a function of the internuclear distance in Å. (e) Fractions and (f) total numbers of long-range proton−proton and carbon−carbon contacts as a function of the distance in Å.

range contacts (Figure 1f) for carbon and proton correlations show that for distances smaller than 4 Å, proton contacts are dominant. For longer distances, the percentage of long-range contacts between protons increases only slightly; the total number of carbon contacts however increases rapidly. Nevertheless, for long distances, the precision of these contacts may be compromised by different factors like dipolar truncation, relayed polarization transfer, spectral crowding and relaxation; thus the most useful restraints arise from long-range contacts at short distances.

Analysis of Distance Restraints Experimentally Observed in the SSNMR Spectra. We first investigate whether the proton−proton and carbon−carbon distance restraints extracted from CHHC/NHHC[20−22] and DARR[44,56,57] spectra are sufficient to calculate high-resolution 3D structures of the Crh protein. This has been demonstrated using proton-mediated experiments for smaller molecules,[23,24] but no structures have yet been determined using carbon-13 distances measured in fully labeled proteins. We have recently shown for the Crh protein that CHHC and NHHC spectra contain high numbers of both local and long-range distance restraints.[46]

Using an assignment strategy based on the distances in the crystal structure, we prepared a list of ^{1}H-^{1}H distance restraints based on proton-mediated, rare-spin detected spectra. The strategy consists in considering the cross-peaks for which only one assignment is possible among the ^{1}H-^{1}H contacts with a distance cutoff of 5 Å, and within a chemical shift window of ± 0.25 ppm with respect to the Crh chemical-shift assignments. A total of 192 proton−proton distance restraints were extracted from the CHHC spectrum and the NHHC spectrum (see below), including 88 long-range restraints (Figure 2a). As here we only consider cross-peaks with one unique assignment possibility, this results in a lack of restraints for β1-strand involved in the domain swap, for which the majority of cross-peaks could be intra- or intermonomeric. The same assignment strategy was applied to the DARR[44,56,57] spectrum (shown in Supplementary Figure S1 in the Supporting Information). The carbon−carbon restraints identified in the same manner as described above for the CHHC spectra are shown in Figure 2b. A total of 196 restraints, including 88 long-range restraints have been extracted from DARR spectrum, which is about the same number as the restraints extracted from the CHHC spectrum. In comparison to the CHHC spectrum, the DARR spectrum shows fewer restraints in the hydrophobic core of the protein, and particularly between the three β-strands. Interestingly, more contacts between helices A and C are observed in ^{13}C−^{13}C correlation experiments than in proton-mediated experiments.

Figure 2c shows the contact plot of the restraints identified in the proton-mediated rare-spin detected experiments (red) and in the DARR experiment (blue). The graph indicates that the correlations present in these two experiments are able to connect the different secondary structure elements of the protein Crh, and also reveals the complementarities of the two approaches.

Calculation of the Crh 3D Structure from SSNMR Unambiguous Restraints. To evaluate whether the correlations extracted from the CHHC, NHHC, and the DARR spectra provide enough structural data for high-resolution 3D structure

(56) Takegoshi, K.; Nakamura, H.; Terao, T. *Chem. Phys. Lett.* **1999**, *307*, 295−302.
(57) Morcombe, C. R.; Gaponenko, V.; Byrd, R. A.; Zilm, K. W. *J. Am. Chem. Soc.* **2004**, *126*, 7196−7197.
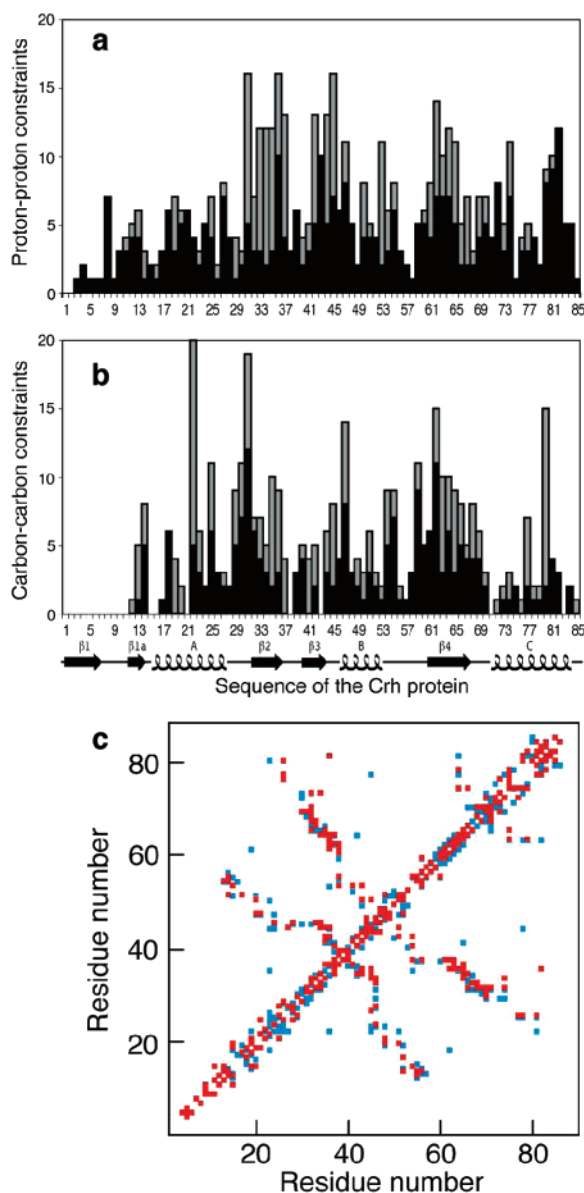
Figure 2. Number of distance restraints identified as a function of the amino acid sequence for the Crh protein: intraresidue, sequential, and medium-range distance restraints are shown in black, long-range in gray. (a) Proton−proton distance restraints identified in the CHHC/NHHC spectra. (b) $^{13}C−^{13}C$ restraints identified in the DARR spectrum. (c) Contact plot of the distance restraints identified from CHHC and NHHC spectra (red) and the DARR spectrum (blue).

Figure 3. NMR ensemble of the 10 lowest-energy conformers calculated from (a) carbon−carbon restraints, (b) carbon−carbon restraints and TALOS dihedral angle predictions, (c) proton−proton restraints, (d) proton−proton restraints and TALOS dihedral angle predictions, (e) carbon−carbon and proton−proton restraints, (f) carbon−carbon, proton−proton restraints and TALOS dihedral angle predictions.

Table 1. Backbone RMSD Values for the Crh Monomer Structures Shown in Figure 3

figure	restraints used	precision (Å)[a]	accuracy (Å)[b]
3a	$^{13}C−^{13}C$	2.4 ± 0.1	2.5 ± 0.3
3b	$^{13}C−^{13}C$, TALOS	1.3 ± 0.1	1.5 ± 0.1
3c	$^{1}H−^{1}H$	1.9 ± 0.2	2.5 ± 0.2
3d	$^{1}H−^{1}H$, TALOS	1.0 ± 0.1	1.4 ± 0.1
3e	$^{1}H−^{1}H$, $^{13}C−^{13}C$	1.0 ± 0.1	1.6 ± 0.1
3f	$^{1}H−^{1}H$, $^{13}C−^{13}C$, TALOS	0.6 ± 0.1	1.3 ± 0.1

[a] Backbone rmsd between the 10 lowest-energy conformers. [b] Backbone rmsd between the average of the 10 lowest-energy conformers and the X-ray structure (PDB entry 1MU4[32])

calculation of the Crh protein, we have used these as distance restraints in molecular-dynamics-based structure calculations. For the sake of simplicity we start out on one monomer unit of this dimeric domain-swapped protein. We used different sets of data to calculate 100 structures for the protein Crh: (i) restraints from CHHC and NHHC spectra; (ii) restraints from DARR spectra, and (iii) restraints both from CHHC, NHHC, as well as DARR spectra. We calculated structures with and without using the dihedral angle information predicted from chemical shifts using the TALOS program.[45]

Figure 3 shows the 10 lowest-energy conformers calculated from the different data sets. It is clear from the figure that calculations based only on carbon−carbon restraints (Figure 3a) or on proton−proton restraints (Figure 3c) are sufficient to provide the right protein fold. Both data sets yield structures
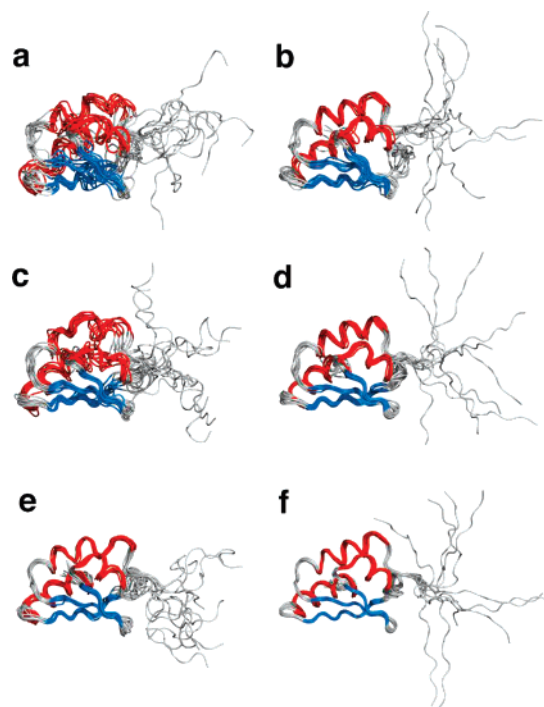
with an rmsd between the 10 lowest-energy conformers and the X-ray structure of 2.5 Å (Table 1). As the $^{1}H−^{1}H$ distance class is tighter than the one for $^{13}C−^{13}C$, the precision of the structures calculated from proton restraints is higher, with a backbone rmsd of 1.9 Å (Figure 3c), compared to 2.4 Å for the calculations from carbon restraints (Table 1). Using as input both $^{1}H−^{1}H$ and $^{13}C−^{13}C$ distance restraints, the rmsd decreases to 1 Å (Figure 3e), and the accuracy to 1.6 Å. The precision is improved by the incorporation of TALOS dihedral angle restraints, with a reduction of the backbone rmsd to approximately 1 Å for the calculations based on $^{1}H−^{1}H$ restraints and to 1.3 Å based on $^{13}C−^{13}C$ restraints. For the calculations including dihedral angle restraints and both $^{1}H−^{1}H$ and the $^{13}C−^{13}C$ restraints, the backbone rmsd decreases to 0.6 Å. As the TALOS dihedral angle prediction is most efficient for the definition of local conformations, the gain in the accuracy of the structure is smaller, with a decrease in rmsd from 1.6 to 1.3 Å (Table 1). A detailed analysis of the local rmsd of the different calculations is given in the Supporting Information (Supplementary Figure S2).
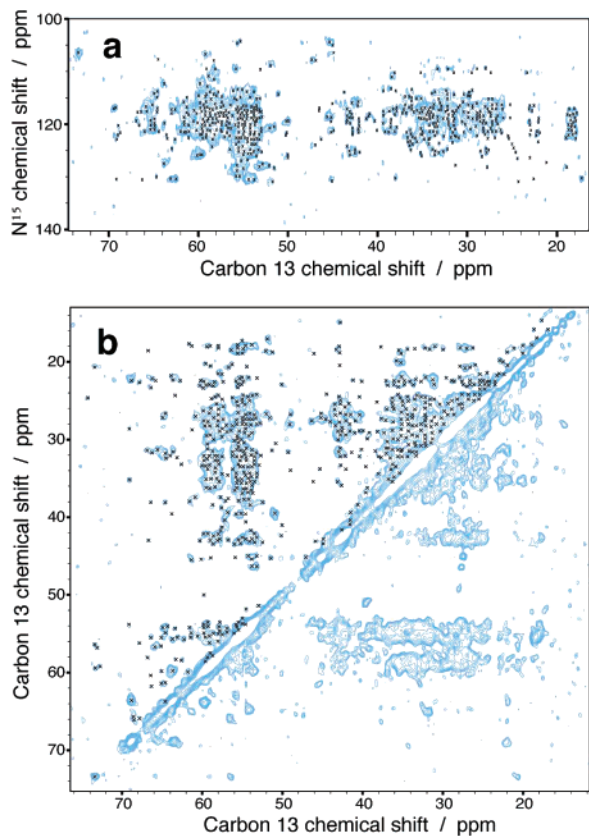
Figure 4. Peak-picking on spectra of the Crh protein: peaks picked (a) NHHC spectrum (mixing time of 100 μs), (b) CHHC spectrum (mixing time of 200 μs).

Analysis of Spectral Ambiguities in the SSNMR Spectra. In a second step we now resolve the high spectral ambiguities by using the concept of ambiguous distance restraints, which allows de novo protein structure determination, without the need of homology models. The iterative ARIA procedure[58] is used to assign the cross signals. As the default ARIA protocols are optimized for moderately ambiguous data, an optimization of certain parameters was necessary. Previous studies[51,59] have investigated the effects of critical parameters like the chemical shift tolerance, the maximum number of assignment possibilities allowed per peak, or the number of simulated annealing cooling steps for the solution NMR case. In the following we investigate the effect of these parameters on the efficiency of automated assignments of solid-state NMR data for the Crh protein. The data extracted from the solid-state NMR spectra consist of a peak list obtained by manual peak picking on CHHC/NHHC spectra (Figure 4). The resulting peak list contains 1267 entries. For the following analysis, the Crh protein was treated as a hypothetical monomer, which means that ambiguities due to the distinction between intra- and intermonomer possibilities are not taken into account.

The choice of the chemical shift tolerance window has a substantial impact on the number of assignment ambiguities of the signals present in the peak list. A chemical shift tolerance window of ± 0.1 ppm (Figure 5a) leads to approximately 150 unambiguous cross-peaks, i.e., cross-peaks which can be
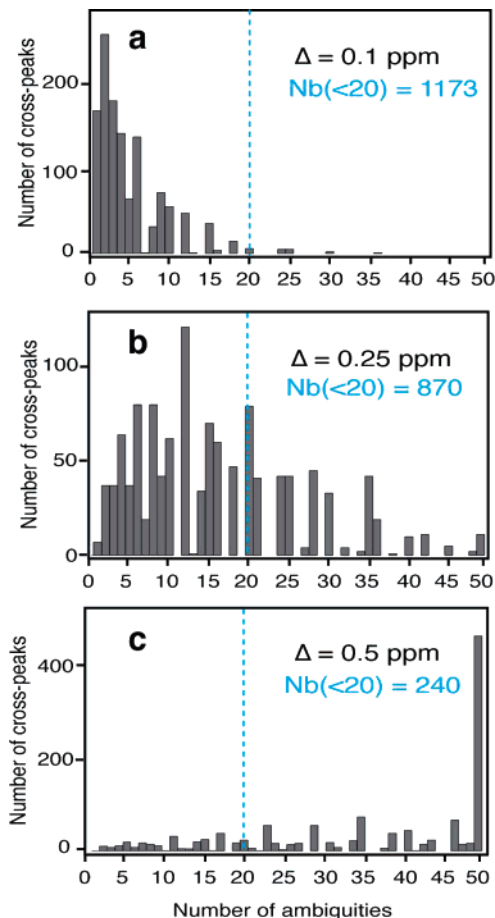
(58) Linge, J. P.; Habeck, M.; Rieping, W.; Nilges, M. *Bioinformatics* **2003**, *19*, 315−316.
(59) Fossi, M.; Linge, J.; Labudde, D.; Leitner, D.; Nilges, M.; Oschkinat, H. *J. Biomol. NMR* **2005**, *31*, 21−34.

Figure 5. Influence of the chemical shift tolerance window on the number of ambiguities. Shown are histograms, for different chemical shift tolerance windows (a = 0.1, b = 0.25, c = 0.5 ppm), of the number of proton−proton contacts as a function of possible assignments per cross-peak. The number of considered cross-peaks, with the number of assignment possibilities per peak set to 20 (Nb(<20)) is given in each frame. Cross-peaks which represent more than 50 possible assignments are plotted at number of ambiguities = 50.

uniquely assigned on the basis of chemical shifts. For the spectra used here, a value of ± 0.25 ppm seems more realistic considering a typical line width of 0.5 to 1.0 ppm, and reduces the number of unambiguous cross-peaks to seven (Figure 5b). The average number of assignment ambiguities per peak is then 16.3, which leads to a set of highly ambiguous restraints for this medium-size protein. For comparison, in solution NMR studies of folded proteins of medium size, the ^{1}H chemical shift tolerance is commonly set to values between 0.01 and 0.015 ppm in NOESY experiments, which typically leads to (i) a sufficient number of unambiguous restraints, which often allow a definition of the global fold of the protein, and (ii) a set of restraints showing only few ambiguities, often around 2−3, which can be easily assigned automatically by using the template structure calculated with the first set of restraints. The large average number of assignment ambiguities in the solid-state NMR case, combined with the small number of unambiguous restraints, illustrates well the difficulty to determine 3D structures of larger fully labeled proteins from SSNMR data without the use of automated iterative assignment software.

At a chemical shift tolerance window of 0.5 ppm (Figure 5c), no cross-peaks can be assigned unambiguously any more. A detailed analysis of the assignment statistics for different
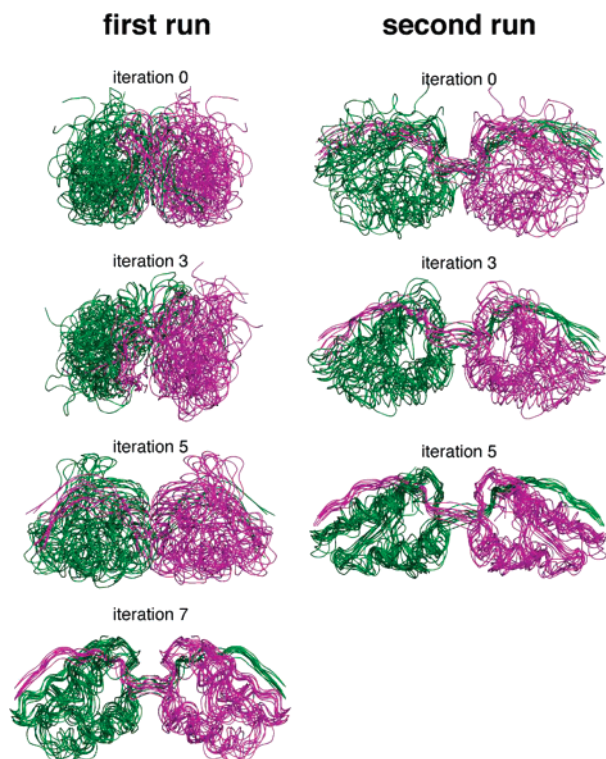
Figure 6. Comparison of conformers calculated in different iterations of ARIA runs: iterations 0, 3, 5, and 7 for the first run (left panel), iterations 0, 3, and 5 for the second run (right panel). Twenty-eight conformers are calculated for each iteration. The 10 lowest-energy conformers superimposed on the backbone atoms are shown.

chemical shift tolerance windows is given in the Supporting Information (Supplementary Figure S3).

Automatic Assignments of Highly Ambiguous Restraints Using ARIA. To determine the 3D structure of the Crh protein *without the need of a homology model*, we used the iterative assignment process implemented in ARIA program. The chemical shifts and the peak lists generated from the CHHC and NHHC spectra provided the input for the distance restraint assignments using ARIA. A chemical shift tolerance window of 0.25 ppm and a maximum number of assignment possibilities per peak of 25 (see Supplementary Figure S3 in the Supporting Information) were used, which results in a total number of 1002 cross-peaks, representing 32600 ambiguities, with an average number of 2×16.3 ambiguities per peak, the factor of 2 taking into account the dimeric nature of the protein.

The left panel of Figure 6 shows the 10 lowest energy conformers obtained after iterations 0, 3, 5, and 7 from the first run of ARIA. The quality of the structures calculated for each iteration was evaluated with respect to the average number of ambiguities per cross-peak (Figure 7a), the number of unambiguous and ambiguous restraints (Figure 7b), as well as the precision and accuracy of the structures (Figure 7c). These Figures reveal the good correlation between the convergence of the conformers and the progress of the automated assignment.

The automated assignments performed by the ARIA program highly reduce the average number of ambiguities per peak in each iteration (Figure 7a), from 16.3 in iteration 0 to 2.5 in the last iteration. After the last iteration, 397 cross-peaks represent more than one possible assignment, with a reduced ambiguity level (2.5 possible assignment). The automated assignment procedure has achieved unambiguous assignments for 593 ^{1}H−
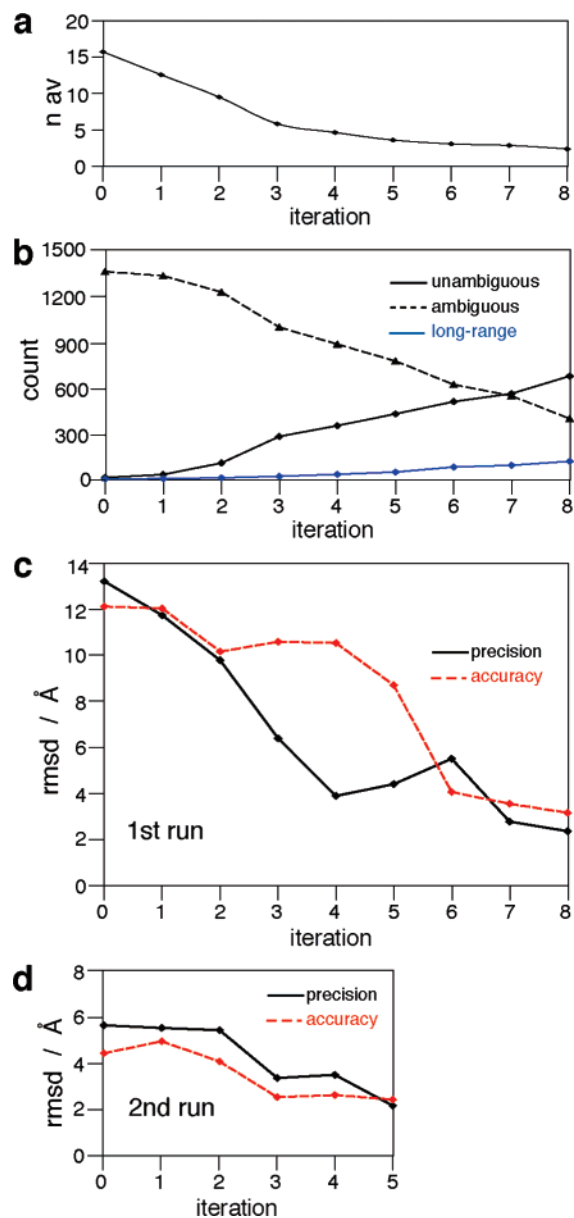

Figure 7. Analysis of the two runs performed using ARIA 2.2 for the structure calculations of the Crh dimer. Evolution as a function of the iteration number: (a) average number of possible assignments per cross-peak for the first run; (b) number of unambiguous, ambiguous, and long-range distance restraints for the first run; (c) backbone rmsd of the 10 lowest-energy conformers (precision) and backbone rmsd between the 10 lowest-energy conformers and the crystal structure (accuracy) for the first run; and (d) rmsd of the 10 lowest-energy conformers (precision) and rmsd between the 10 lowest-energy conformers and the crystal structure (accuracy) for the second run. Rmsd values were calculated for residues 2−81. In each iteration, 28 conformers were calculated, and the rmsd values is indicated for the 10 lowest-energy conformers.

^{1}H contacts after the last iteration of the first run (Figure 7b), including 115 long-range restraints. Iteration 7 is the first iteration where the number of unambiguous restraints is higher than the number of ambiguous restraints. The precision of the structures increases with each iteration (Figure 7c), to reach a value of 2.8 Å for the iteration 8, and the rmsd decreases steeply during the five first iterations down to 4 Å.

To improve the convergence of the calculated structure, we have realized a second round of ARIA iterations. We used the 7 lowest-energy conformers obtained after iteration 8 of the first ARIA run as template structures to filter the possible assign-

ments in the second ARIA run. The same solid-state NMR data was used for this second round (chemical shift and peak lists, TALOS dihedral angle restraints and intermonomer restraints), but the input of the template structures allowed to decrease the number of possible assignments already during the first iteration. We set the parameter of violation tolerance to 2 Å instead of the standard value of 1000 Å used for the first iteration, and reduced the number of iterations to 6. The right panel in Figure 6 shows the 10 lowest-energy conformers for iterations 0, 3, and 5 of the second ARIA run. The addition of the template structures highly increases the convergence of the first iteration resulting in a precision of 5.6 Å (Figure 7d) compared to 13.2 Å at iteration 0 of the first run. As shown in Figure 6, the global fold of the Crh protein is defined after the first iteration of the second run, and the calculations converge to reach a precision of 2.3 Å.

Calculation of the Crh 3D Structure from the Unambiguous Restraints Determined Using ARIA. The automated assignment procedure has achieved unambiguous assignments for 643 $^1H-^1H$ contacts in the last iteration of the second run, including 131 long-range restraints. Iterative assignment has thus been successful for about 65% of the initial 1002 peak entries. In the ARIA protocol, the structures were calculated with both unambiguous and ambiguous restraints. In the final structure calculation realized using XPLOR-NIH, we used only the 643 unambiguous $^1H-^1H$ restraints assigned by ARIA, as well as the TALOS dihedral restraints, to calculate a total of 200 conformers. A set of conformers was chosen on the basis of the absence of distance violations >0.5 Å. The 10 lowest-energy conformers are shown in Figure 8a and c, and are compared to the structure determined by X-ray crystallography (Figure 8b). These 10 structures were analyzed using PROCHECK[55] and show a good covalent geometry (Table 2); 74.2% of the residues have backbone conformations in favorable regions and 20.6% in allowed regions of the Ramachandran plot. The selected 10 conformers show a precision of 1.33 Å when superimposing the backbone atoms for the dimer and an rmsd of 1.20 Å when superimposing the heavy atoms in regular secondary structures elements: the precision is improved with respect to the one obtained after the second ARIA run. This improvement is because only unambiguous distance restraints were considered in this calculation round, as well as to the higher number of conformers calculated. If we consider a hypothetic monomer (residues 2−12 from chain A, and residues 13−80 from chain B) (Figure 8c), the rmsd decreases to 0.84, and to 0.76 Å if only regular secondary structure elements are taken into account. The 10 selected conformers were deposited to the Protein Data Bank (PDB id: 2RLZ). The backbone fold of the solid-state NMR structure shows an accuracy of 2.89 Å with respect to the single-crystal structure of the dimeric Crh protein (Figure 8a). The geometry of the obtained structure was also compared to the geometry of the crystallographic dimer using the mean distance of the monomer centers-of-mass, and the orientation between the monomers, calculated as the angle formed by the two vectors going from the center-of-mass of the interface region (residues 10 to 15) to the center-of-mass of each monomer. These two parameters display close mean values in the SSNMR as well as in the crystallographic structures (Table 2). The accuracy improves to 1.62 Å when considering the hypothetic monomer, as shown in Figure 8c. Besides showing that the
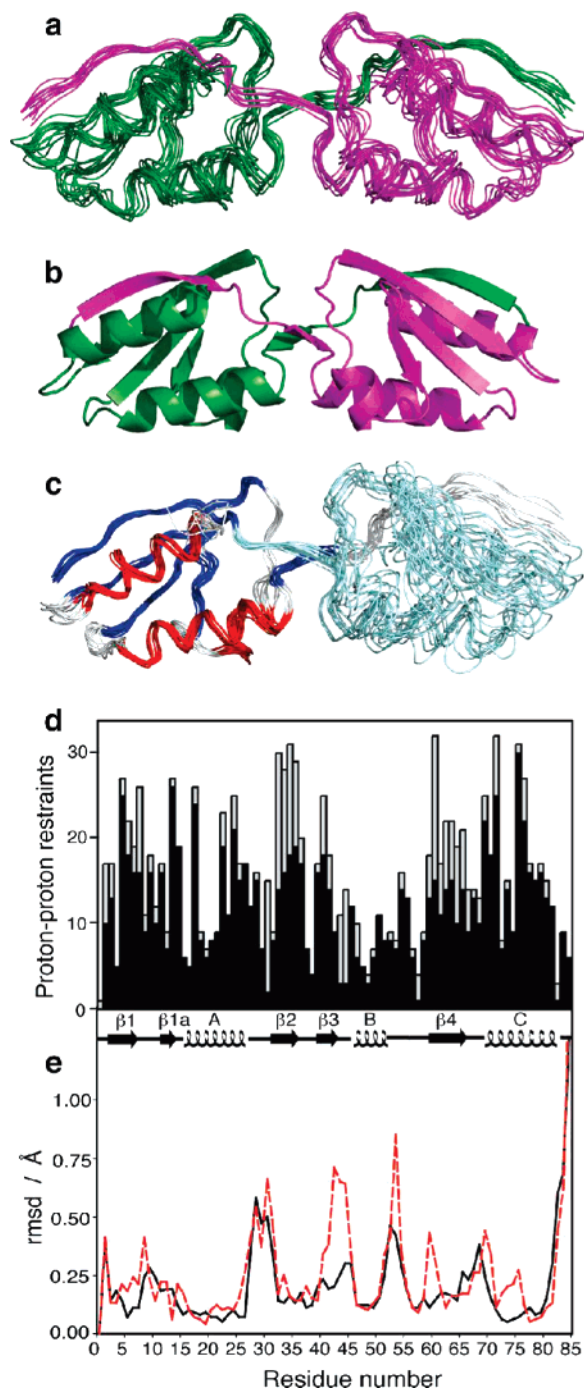

Figure 8. (a) Backbone ensemble of the 10 lowest-energy conformers of the dimer structure calculated from solid-state NMR data as described in the text; (b) dimer structure of Crh protein (PDB entry: 1MU4);[32] (c) backbone ensemble of the 10 lowest-energy conformers of the monomeric part of the dimer structure, with the superposition realized on one monomer (residues 2−12 from chain A, and residues 13−80 from chain B) (d) Proton−proton restraints used for the final structure calculations of the Crh dimer. Long-range restraints are show in gray. (e) Local rmsd for the 10 lowest-energy conformers as a function of the primary sequence of Crh protein, calculated for the NMR conformers (black) and between the NMR and the crystal structure (red).

accuracy obtained for Crh is comparable to the one achieved for other proteins using solid-state NMR methods (see below), this also underlines that multimeric protein structure determination by NMR remains a challenge.

Figure 8d shows the distribution, as a function of the sequence, of the unambiguous restraints used for the structure

Table 2. Structural Statistics for the 10 Lowest-Energy Conformers of the Crh Dimer Protein Calculated with XPLOR-NIH Using the Unambiguous Distance Restraints Assigned after the Second ARIA Run

no. of structures in the final set	10
no. of unambiguous distance restraints	
total	643
sequential ($|i$-$j\,|$ =1)	181
medium-range ($1 < |i$-$j\,| < 5$)	85
long-range ($|i$-$j\,| > 4$)	131
no. of backbone dihedral angle restraints	58
distance violations:	
> 0.50 Å	none
> 0.30 Å	0.40 ± 0.50
> 0.10 Å	2.20 ± 1.30
rmsd (Å)	0.012 ± 0.005
rmsd of the NMR ensemble[a] (Å)	
monomer, backbone	0.84 ± 0.12
monomer , SSE[b]	0.76 ± 0.09
dimer, backbone	1.33 ± 0.23
dimer, SSE	1.20 ± 0.24
dimer, SSE all heavy atoms	1.74 ± 0.21
rmsd $<10>$[c] versus X-ray structure[d] (Å)	
monomer	1.62
dimer	2.89
deviation from idealized covalent geometry	
rmsd bond (Å)	0.0011 ± 0.0007
rmsd angles (deg)	0.37 ± 0.03
rmsd impropers (deg)	0.25 ± 0.01
Ramachandran data[e]	
residues in most favored region (%)	74.2
residues in allowed regions (%)	20.6
residues in generously allowed regions (%)	4.0
residues in disallowed regions (%)	1.2
Ramachandran data for SSE	
residues in most favored region (%)	89.6
residues in allowed regions (%)	8.7
residues in generously allowed regions (%)	0.8
residues in disallowed regions (%)	0.9
angle between the monomers (deg)	
final set of SSNMR structure	164.3 ± 4.0
X-ray structure (1MU4)	164.8
distance versus the center of mass (Å)	
final set of SSNMR structure	22.1 ± 0.7
X-ray structure (1MU4)	20.7

[a] Calculated for residues 2−12 from chain A and residues 13−80 from chain B for the monomer, and residues 2−80 from both chains for the dimer. [b] SSE: secondary structure elements. [c] $<10>$ represents the average for the 10 selected conformers. [d] Calculated on secondary structure elements between $<10>$ and the X-ray structure (PDB entry 1MU4).[32] [e] From PROCHECK.[55]

calculation using XPLOR-NIH. Large numbers of long-range restraints are found in β-sheet regions, mainly due to numerous ^{1}H−^{1}H side-chain contacts present in these secondary structure elements. The helical regions in the protein are less well defined, because the methyl and aromatic contacts, which would help to better define these regions, are difficult to detect in CHHC/NHHC experiments. Figure 8e shows the local rmsd of the 10 lowest-energy conformers after the calculation with XPLOR-NIH. Low precision is observed in regions not corresponding to regular secondary structure elements, particularly the segment 27−32. The least restrained protein segments, which include helix B and beginning of loop 2 (residues 53−59), correspond to flexible regions of the protein[60,61] and show low accuracy.

Comparison of the Convergence of the Crh Structure with That of Other Protein Structures Obtained from SSNMR Data. The convergence of the Crh dimer structure calculated here was compared to the convergence of protein structures previously calculated from SSNMR data by calculating the backbone rmsd obtained on the NMR conformers. Two structures have been determined from extensively labeled protein samples: SH3 (62 residues) and Ubiquitin (76 residues). The SH3 structure shows a precision of 0.7 Å (calculated on the Cα atoms in the secondary structure elements) and an accuracy of 1.3 Å with respect to the crystal structure.[8,62] Ubiquitin shows a precision on the backbone atoms of about 1.0 Å; no accuracy is given.[9] Two structures were obtained for small fully labeled proteins by means of simple manual assignments: Kaliotoxin[24] (38 residues) and GB1[18,63] (56 residues). They show backbone rmsds of 0.81 and 0.82 Å, respectively, and an accuracy of 1.9 Å. The backbone rmsd of 1.33 Å (1.20 Å on secondary structure elements) and the accuracy of 2.89 Å obtained for the fully labeled Crh dimer (170 residues) is slightly higher than the values obtained for the above proteins; this increase can mainly be explained by the larger size of the protein, combined with the additional difficulty of the calculation of an elongated dimeric molecule. This is supported by the rmsd calculated for a hypothetic monomer, where the backbone rmsd for regular structure elements decreases to 0.76 Å, and which shows an accuracy of 1.62 Å, values which compare favorably to those of previously determined SSNMR structures.

Conclusion

We demonstrated how the structure of the fully ^{13}C−^{15}N labeled dimeric form of the Crh protein can been derived using solid-state NMR techniques. We could show that proton and carbon distances can be extracted from 2D CHHC/NHHC as well as from DARR correlation spectra, and that the information content of both types of experiments is sufficient to calculate high-resolution structures. Carbon and proton restraints are complementary, and the combination of both, together with dihedral angles derived from chemical shifts, yields the highest resolution.

We also could show, for the first time, how highly ambiguous proton-mediated, rare-spin detected solid-state NMR data sets of a fully labeled protein sample can be used for structure calculation through automated iterative assignments. The input of over 1000 cross-signals as ambiguous restraints, 25 inter-monomer restraints and chemical-shift derived dihedral angles yields a total of 643 unambiguously assigned distance restraints, including 131 long-range restraints. The Crh dimer structure calculated using this data shows a precision of 1.33 Å and an accuracy with respect to the crystal structure of 2.89 Å. Crh is the largest protein structure which has been determined so far from solid-state NMR data; our work shows that even complex structural features, like the dimeric and elongated nature of Crh, are not an obstacle to high-resolution structure determination by solid-state NMR.

This study is a further step toward the structure determination of insoluble proteins by the more general approach using fully labeled protein samples, and paves the way for the study of

(60) Favier, A.; Brutscher, B.; Blackledge, M.; Galinier, A.; Deutscher, J.; Penin, F.; Marion, D. *J. Mol. Biol.* **2002**, *317*, 131−144.
(61) Giraud, N.; Böckmann, A.; Lesage, A.; Penin, F.; Blackledge, M.; Emsley, L. *J. Am. Chem. Soc.* **2004**, *126*, 11422−11423.
(62) Castellani, F.; van Rossum, B. J.; Diehl, A.; Rehbein, K.; Oschkinat, H. *Biochemistry* **2003**, *42*, 11476−11483.
(63) Peng, X.; Libich, D.; Janik, R.; Harauz, G.; Ladizhansky, V. *J. Am. Chem. Soc.* **2007**.

larger molecules. For larger proteins, the use of 3D experiments is a means to increase spectral resolution and to considerably decrease the ambiguity level due to the additional frequency dimension; this is possible also by the use of higher magnetic fields, which equally benefit the limited signal-to-noise of proton-mediated rare-spin detected experiments. With higher resolution available, it is possible to decrease the chemical shift tolerance window, and room is left for the study of even larger molecules, without substantially modifying the ambiguity profile. The implementation of simultaneous iterative assignments of proton *and* carbon spectra, combined with the faster structure convergence using network anchoring,[64] should allow even higher resolution structures to be obtained.

(64) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **2002**, *24*, 171–189.

Supporting Information Available: The 2D DARR spectrum with assignments; analysis of the local rmsd for the structures calculated from unambiguous constraints; assignment statistics for different chemical-shift tolerance windows. This material is available free of charge via the Internet at http://pubs.acs.org.

JA078014T