# Single-Step Determination of Protein Substructures Using Dipolar Couplings: Aid to Structural Genomics

Markus Zweckstetter* and Ad Bax

*Laboratory of Chemical Physics, NIDDK*
*National Institutes of Health, Bethesda, Maryland 20892-0520*

*Received June 26, 2001*
*Revised Manuscript Received August 13, 2001*

The wealth of genomic data that has recently become available with completion of the sequencing of both the human and a variety of other genomes[1] has created a need for rapid and efficient determination of three-dimensional structures of the corresponding proteins. So far, most effort in this so-called structural genomics project has focused on X-ray crystallography, but NMR also shows considerable potential.[2,3] Conventionally, NMR structure determination consists of a resonance assignment phase, which usually relies on analysis of an extensive set of triple resonance *J*-connectivity data, followed by a structure determination phase that relies on interpretation of NOE spectra and measurement of scalar and dipolar couplings.[3,4] Here, we describe and demonstrate a more integrated approach, where the assignment and structural data are derived from the same experiment. This offers the opportunity to greatly accelerate the NMR structure-determination process.

The method demonstrated here relies on the 3D (HA)CANH triple resonance experiment,[5] which correlates amide $^{1}H$ and $^{15}N$ chemical shifts with those of the intraresidue and preceding $^{13}C^{\alpha}$ nuclei. Because the $^{13}C^{\alpha}$ chemical shift is usually insufficiently unique, such a spectrum alone cannot be used for determining complete sequential assignments. Therefore, separate experiments that correlate the amide with intraresidue and preceding residue $^{13}C'$ or $^{13}C^{\beta}$ resonances are frequently used to complete the assignment process. Here, we demonstrate that if the (HA)CANH spectrum is recorded in the $^{1}H^{\alpha}$-coupled mode, both the size of the $J_{C\alpha H\alpha}$ splitting and the asymmetry in doublet intensity[6] resulting from relaxation interference between $^{13}C^{\alpha}$ chemical shift anisotropy (CSA) and $^{13}C^{\alpha}-^{1}H^{\alpha}$ dipolar coupling can be used to resolve the ambiguities caused by the non-uniqueness of the $^{13}C^{\alpha}$ chemical shift. When such an experiment is conducted in both an isotropic and a liquid crystalline medium, the resulting spectra not only yield complete assignment for the backbone $^{1}H^{N}$, $^{15}N$, and $^{13}C^{\alpha}$ resonances, but also contain important structural information in the form of $^{13}C^{\alpha}-^{1}H^{\alpha}$ dipolar coupling, $^{13}C^{\alpha}$ CSA, and deviations from random-coil $^{13}C^{\alpha}$ isotropic chemical shifts. Together, this provides sufficient information to obtain complete backbone $^{1}H^{N}$, $^{15}N$, and $^{13}C^{\alpha}$ assignments of small and medium-sized proteins, such as ubiquitin and calmodulin, and to determine the 3D structure of fragments of such proteins.

To avoid increased crowding in the $H^{\alpha}$-coupled (HA)CANH spectrum relative to the regular (HA)CANH spectrum, the two $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ doublet components are separated into two separate spectra in the usual manner[7] by calculating the sum and difference of an in-phase and an anti-phase $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ (HA)CANH spectrum. The pulse sequence for this so-called IPAP-(HA)CANH experiment is available as Supporting Information. The $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ coupling remains active during the 28-ms $^{13}C^{\alpha}$ constant-time evolution period, resulting in two highly resolved $^{13}C^{\alpha}$ doublet components that are split by the $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ coupling and have an intensity ratio that depends on the $^{13}C^{\alpha}$ CSA. Correlations from a given $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ atom pair to the intraresidue and sequential amide not only will exhibit identical $^{13}C^{\alpha}$ shifts, but also the same $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splitting and the same $^{13}C^{\alpha}$ CSA-induced intensity ratio of the two $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ doublet components. This greatly alleviates the problem of $^{13}C^{\alpha}$ chemical shift degeneracy

The method has been applied to samples of U-$^{13}C/^{15}N$ ubiquitin (3 mM), pH 6.5, 25 °C and U-$^{13}C/^{15}N$ C-terminal $Ca^{2+}$-ligated calmodulin (1 mM), pH 7, 25 °C, 100 mM NaCl, without and with 15.5 mg/mL Pf1 bacteriophage.[8] For the isotropic 3 mM ubiquitin sample perfect conditions are available, and the backbone can completely and fully automatically be assigned using the IPAP-(HA)CANH experiment, even in the absence of residual alignment. Owing to the high $^{13}C^{\alpha}$ resolution and the high signal-to-noise ratio, the isotropic $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splitting and the $^{13}C^{\alpha}$ CSA are sufficient for resolving any $^{13}C^{\alpha}$ chemical shift ambiguities (data not shown).

For less concentrated samples of α-helical proteins, the accuracy at which the small variation in vicinal $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splitting can be measured is generally insufficient to resolve $^{13}C^{\alpha}$ chemical shift degeneracy. For example, for the 1 mM calmodulin sample the $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splitting measured from the intraresidue correlation differs by a root-mean-square (rms) value of 2.4 Hz from the same splitting measured through the cross-peak to the sequential amide (this rmsd is only 0.4 Hz for the ubiquitin sample). Similarly, for isotropic calmodulin the rms difference in the reduced CSA ($CSA^{red}$),[6] as measured from the intraresidue and sequential amides, becomes rather large (11.3 ppm vs 5.6 ppm for ubiquitin). For aligned calmodulin, the signal-to-noise ratio is even lower, and the pairwise rms differences between the intraresidue and sequential measurements of the $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splitting and $CSA^{red}$ increase to 3.2 Hz and 14.8 ppm, respectively. Note that these high rmsd's primarily reflect the large uncertainty in the weak $C^{\alpha}_{i}-H^{N}_{i+1}$ correlation. (The uncertainty in the $CSA^{red}$ and $J_{CH}$ values, as measured for the intraresidue correlation, are much lower.) However, the introduction of a liquid crystalline medium, also increases the range of $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splittings from a few Hz to ±50 Hz (Figure 1). This greatly increased variation in splitting, together with the remaining dispersion in $CSA^{red}$, enables long contiguous stretches of residues to be assigned. Common assignment approaches rely heavily on $^{13}C^{\beta}$ shifts for positioning assigned fragments within the primary sequence. With the present approach, secondary structure is already available for many residues (see below) and $C^{\alpha}$ chemical shifts alone are then quite characteristic of the amino acid type.[9] Combination of this information with uniquely identified glycines (from the $C^{\alpha}$ chemical shift and the $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splitting) relaxes the need for $^{13}C^{\beta}$ chemical shifts when positioning assigned fragments within the primary sequence. This made it possible to completely assign the backbone of the 148-residue, α-helical protein calmodulin using only the IPAP-(HA)CANH data.

In addition to backbone assignment, the IPAP-(HA)CANH data can be used to define the secondary structure. First, α-helices

(1) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; et al. *Science* **2001**, *291*, 1304−1351. Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; et al. *Nature* **2001**, *409*, 860−921.

(2) Abbott, A. *Nature* **2000**, *408*, 130−132.

(3) Moseley, H. N. B.; Montelione, G. T. *Curr. Opin. Struct. Biol.* **1999**, *9*, 635−642. Fowler, C. A.; Tian, F.; Al-Hashimi, H. M.; Prestegard, J. H. *J. Mol. Biol.* **2000**, *304*, 447−460. Montelione, G. T.; Zheng, D.; Huang, Y. J.; Gunsalus, K. C.; Szyperski, T. *Nature Struct. Biol.* **2000**, *7*, Suppl., 982−985.

(4) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.

(5) Montelione, G. T.; Wagner, G. *J. Magn. Reson.* **1990**, *87*, 183−188. Kay, L. E.; Ikura, M.; Bax, A. *J. Magn. Reson.* **1991**, *91*, 84−92.

(6) Tjandra, N.; Bax, A. *J. Am. Chem. Soc.* **1997**, *119*, 9576−9577.

(7) Yang, D. W.; Nagayama, K. *J. Magn. Reson., Ser. A* **1996**, *118*, 117−121. Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373−378.

(8) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065−1074. Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2001**, *20*, 365−377.

(9) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490−5492. Grzesiek, S.; Bax, A. *J. Biomol. NMR* **1993**, *3*, 185−204.

10.1021/ja016496h CCC: $20.00 
Published on Web 08/30/2001

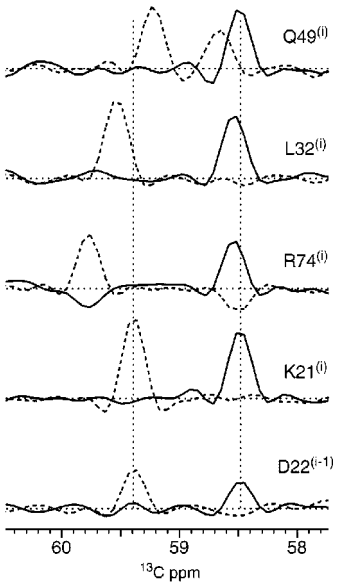


**Figure 1.** 1D cross-sections of the 3D IPAP-(HA)CANH measured on aligned calmodulin for four residues that match the $^{13}C^{\alpha}$ chemical shift of K21 observed for the D22-$^{1}H^{N}$-cross-peak (bottom trace). Up- (- - -) and downfield (–) doublet components, separated in two 3D subspectra by the IPAP element, are superimposed. Cross-sections are labeled with the amino acid code, where (i) and (i-1) indicate intraresidual and sequential peaks. The total measuring time was 2 days. In almost all cases (all for ubiquitin and calmodulin) the sequential peak is weaker than the intraresidual peak. Note that when the intraresidual peak is overlapped the $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splitting and $^{13}C^{\alpha}$ CSA can be extracted from the sequential peak.

are characterized by a contiguous set of residues with large, isotropic $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splitting [$J(C^{\alpha}-H^{\alpha}) > 145.5$ Hz].[10] These residues are then used for determination of an average value of the spectral density for dipolar−CSA cross-correlation, assuming an average $^{13}C^{\alpha}$ $CSA^{red}$ of 6 ppm in helices.[6] With the spectral density available, the $CSA^{red}$ for the remaining residues can be calculated, and a high $CSA^{red}$ [$CSA^{red} > 29$ ppm][6] in combination with an isotropic $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splitting of $140.5 \pm 1.8$ Hz allows unique identification of residues in extended structures. Furthermore, non-glycine residues with positive $\phi$ angles are identified by $J(C^{\alpha}-H^{\alpha}) < 137$ Hz.[10]

Thus, a wealth of NMR parameters that are sensitive to protein structure is available from the IPAP-(HA)CANH experiment: $^{13}C^{\alpha}$ and $^{15}N$ chemical shifts, energetically allowed $\phi/\psi$ ranges for residues in regular secondary structure elements, positive $\phi$ angles for appropriate residues, and $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ residual dipolar couplings (rdc's) obtained from the difference between isotropic and aligned $^{13}C^{\alpha}-\{^{1}H^{\alpha}\}$ splittings. In addition, $^{15}N-\{^{1}H^{N}\}$ rdc's are readily accessible from a simple 3D HNCO spectrum or can be obtained from the IPAP-(HA)CANH by keeping the $^{15}N-\{^{1}H^{N}\}$ coupling active during acquisition. These various constraints were used in a molecular fragment replacement (MFR) search for determination of protein backbone segments.[11]

Here, the MFR search is performed using an eight-residue reading frame. The 20 best eight-residue fragments found by searching a 478-protein subset of the protein databank are retained, and their $\phi/\psi$ angles are mapped in a Ramachandran plot. Ignoring the backbone angles for the terminal residues in the eight-residue fragments, 120 $\phi/\psi$ pairs are available for each residue, and

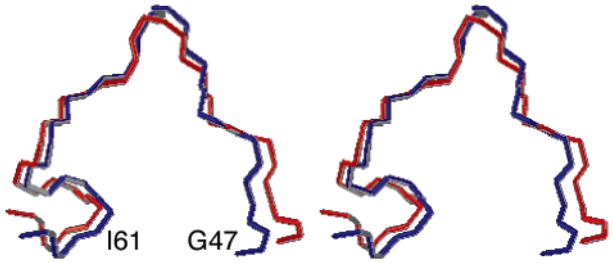


**Figure 2.** Stereodiagram of the backbone representation of the large loop formed by residues 47−61 of ubiquitin in the solution structure (PDB code: 1d3z; blue) and in the model (red) obtained from an MFR homology search. Only NMR constraints obtained from an IPAP-(HA)-CANH experiment measured in isotropic phase and in charged bicelles were used. In total five fragments (2−10,10−35,35−47,47−61,61−74) are obtained that differ by 0.63, 0.60, 1.30, 1.32, and 1.07 Å from 1d3z.

(provided the majority of these are present in a single cluster) a reliable estimate for the true backbone torsion angles is obtained by taking the median of the most populated region in the combined dipolar coupling/chemical shift Ramachandran map. Due to the small number of dipolar couplings, chemical shifts play a more important role for $\phi/\psi$ estimates than in a previous study.[11] Besides ubiquitin and calmodulin, the MFR approach with the minimal input data described above was tested for the proteins dinI (81 residues) and thioredoxin (105 residues). On average, fragments are obtained with a length of $15 \pm 6$ residues that differ by $1.6 \pm 0.9$ Å from a high-resolution reference structure (Figure 2).

This report demonstrates that it is possible to rapidly determine backbone assignment and substructures for proteins from a single set of IPAP-(HA)CANH spectra. These 3D backbone substructures can be useful for structural genomics applications, such as fold identification, and shift the focus of protein structure determination to the assembly of a limited number of well-determined backbone segments. In a first step, for example, individual fragments can be oriented with respect to each other, using residual dipolar couplings.[12,13] Alternatively, packing of fragments using molecular modeling is expected to be relatively rapid and straightforward, particularly with information on allowed relative orientations available. Moreover, the fragments can serve as valuable starting points for determining high-resolution 3D structures, including side-chains, when additional dipolar couplings and NOEs are available.

**Acknowledgment.** We thank Claude Klee and James Chou for the calmodulin sample, Georg Kontaxis and Frank Delaglio for help with the MFR homology search. M.Z. is the recipient of a DFG Emmy Noether-Fellowship. This work was supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the NIH.

**Supporting Information Available:** Pulse sequence of the 3D IPAP-(HA)CANH experiment; tables with ubiquitin and calmodulin NMR restraints, and $\phi$ and $\psi$ angles determined from the MFR search (PDF). This material is available free of charge via the Internet http://pubs.acs.org.

JA016496H

(10) Vuister, G. W.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **1992**, *114*, 9674−9675.
(11) Delaglio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142−2143.
(12) Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371−424.
(13) Hus, J. C.; Marion, D.; Blackledge, M. *J. Am. Chem. Soc.* **2001**, *123*, 1541−1542.