Journal of Magnetic Resonance **153,** 246–253 (2001)
doi:10.1006/jmre.2001.2440, available online at http://www.idealibrary.com on IDEAL®

Use of the Carbonyl Chemical Shift to Relieve Degeneracies in Triple-Resonance Assignment Experiments

Eric W. Sayers and Dennis A. Torchia

Molecular Structural Biology Unit, National Institute of Dental and Craniofacial Research, National Institutes of Health, 30 Convent Drive, MSC 4307, Bethesda, Maryland 20892-4307

E-mail: dtorchia@dir.nidcr.nih.gov

Received May 22, 2001; revised August 30, 2001; published online October 25, 2001

We illustrate an approach that uses the backbone carbonyl chemical shift to relieve resonance overlaps in triple-resonance assignment experiments conducted on protein samples. We apply this approach to two cases of simultaneous overlaps: those of ($^1H^N$, ^{15}N) spin pairs and those of ($^1H^\alpha$, $^{13}C^\alpha$) spin pairs in residues preceding prolines. For these cases we employed respectively CBCACO(N)H and H(CA)CON experiments, simple variants of the commonly used CBCA(CO)NH and HCA(CO)N experiments obtained by replacing one of the indirect dimensions with a carbonyl dimension. We present data collected on ribosomal protein S4 using these experiments, along with overlap statistics for four other polypeptides ranging in size from 76 to 263 residues. These data indicate that the CBCACO(N)H, in combination with the CBCA(CO)NH, can relieve >83% of the ($^1H^N$, ^{15}N) and ($^1H^N$, $^{13}C'$) overlaps for these proteins. The data also reveal how the H(CA)CON experiment successfully completed the assignment of triply and quadruply degenerate X-Pro spin systems in a mobile, proline-rich region of S4, even when X was a glycine. Finally, we discuss the relative sensitivities of these experiments compared to those of existing sequences, an analysis that reinforces the usefulness of these experiments in assigning extensively overlapped and/or proline-rich sequences in proteins.

Key Words: **carbonyl chemical shift; CBCACO(N)H; H(CA)CON; proline; ribosomal protein S4; resonance overlap.**

INTRODUCTION

A combination of triple-resonance experiments, such as HNCACB (*1*), CBCA(CO)NH (*2*), HNCO (*3*), and HCACO (*4*), is an efficient means of assigning the NMR signals of protein backbone atoms. However, because several of these experiments depend upon correlating $^{13}C'$, $^{13}C^\alpha$, and $^{13}C^\beta$ resonances to those of a ($^1H^N$, ^{15}N) spin pair, assignment bottlenecks can arise in two cases: (a) when the amide proton and ^{15}N resonances of two (or more) residues simultaneously overlap, and (b) when an amino acid occurs immediately N-terminal to a proline, which lacks an amide proton. While isolated occurrences of simultaneous ($^1H^N$, ^{15}N) overlap are usually not a serious problem, when such cases are numerous, as found in unfolded proteins or large α-helical proteins, these overlaps can become a severe impediment. Likewise, isolated prolines rarely present assignment difficulties; however, cases of multiple adjacent prolines or several X-Pro elements in which the resonances of X are overlapped with one another or with other resonances present significant challenges. In this paper, we present approaches that overcome these problems by using the carbonyl carbon ($^{13}C'$) chemical shift to relieve degeneracies among the other backbone spins. We illustrate this general strategy with two 3D experiments: first, the CBCACO(N)H as an aid for multiple ($^1H^N$, ^{15}N) overlaps, and second, the H(CA)CON as an aid for overlaps in proline-rich sequences. We further show that the use of these experiments in conjunction with existing ($^1H^N$, ^{15}N) and ($^1H^\alpha$, $^{13}C^\alpha$) experiments is a simple, convenient, and yet powerful approach to overcome assignment bottlenecks in proteins.

The protein that motivated this work was the ribosomal protein S4 from *B. stearothermophilus* (200 aa, 23.2 kDa). The 159 C-terminal residues of S4, S4 Δ41, form a highly ordered tertiary structure in crystals (*5*) and in solution (*6, 7*). In contrast, the 41 N-terminal residues of S4, which are absent from S4 Δ41, are quite flexible in solution (*8*). The radical difference in dynamics between these two regions ($\tau_c \sim 1$ ns vs $\tau_c \sim 12$ ns) resulted in spectra that resembled an overlay of the individual spectra of a folded protein and a flexible peptide, causing numerous cases of resonance overlap, particularly in ($^1H^N$, ^{15}N) and ($^1H^\alpha$, $^{13}C^\alpha$) spin pairs. In addition, the N-terminus contains five prolines, two of which are adjacent, complicating the assignment of surrounding residues by conventional experiments that require amide protons. In response to these issues, we modified existing pulse sequences, taking advantage of the dispersion of the $^{13}C'$ chemical shifts to facilitate signal assignments.

Recent years have witnessed a growing interest in pulse sequences that use the $^{13}C'$ chemical shift as an assignment tool. These include the HCACO-CBHB (*9*), CBCACO(CA)HA (*10*), (HCA)CONH (*11, 12*), COHNNCA (*13*), CO-H(N)CACB (*14*), and 4D HNCO$_{i-1}$CA$_i$ (*15*) experiments. Taking a lead from this work, we found that a CBCACO(N)H experiment, a variant of the CBCA(CO)NH experiment previously suggested

by other authors (*14*), was a convenient solution to our assignment problem. Unlike the (HCA)CONH, COHNNCA, and CO_H(N)CACB experiments, which yield both intra- and interresidue connectivities, the CBCACO(N)H is advantageous in that it provides only sequential interresidue connectivities, eliminating potential ambiguities. In this way it is similar to the HCACO-CBHB and CBCACO(CA)HA, except that it establishes a sequential connectivity from the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ of residue i to the amide proton of residue $i+1$ rather than an intraresidue connectivity to the $^{1}H^{\alpha}$ of residue i.

Likewise, a growing number of pulse sequences for overcoming the problems associated with assigning proline-rich regions of polypeptides have been developed, and these include the CDCA(NCO)CAHA (*16*), HACAN (*17, 18*), CBCA(CO)N(CA)HA (*18*), and the family of four "proline neighbor selective" experiments (*19*). However, the former three of these experiments all label the resulting signals with the frequencies of the ($^{1}H^{\alpha}$,$^{13}C^{\alpha}$) spin pair, and therefore fail when these pairs simultaneously overlap those of other residues, as is extensively the case with S4. While the proline neighbor experiments avoid this problem, they use the amide protons of the amino acids immediately preceding and succeeding the target proline, and thus cannot be used for sequences of adjacent prolines. We therefore found that a 3D H(CA)CON experiment, a simple variant of the 3D HCA(CO)N (*20, 21*) and 4D HCACON (*21*) experiments, could successfully use the $^{13}C'$ spins to alleviate these problems of multiple degeneracies so that we could achieve a complete resonance assignment.

RESULTS AND DISCUSSION

Simultaneous ($^{1}H^{N}$, ^{15}N) Overlap

We first describe how the problem of simultaneous overlap of ($^{1}H^{N}$, ^{15}N) spin pairs is relieved by the CBCACO(N)H experiment, displayed in Fig. 1. Between points *a* and *b*, the CBCA(CO)NH experiment is duplicated, producing at point *b* terms $-2C_z^{\alpha}C_y'$ that are frequency labeled with the chemical shift of either the $^{13}C^{\alpha}$ or the $^{13}C^{\beta}$spin. Between points *b* and *c*, as in the CBCA(CO)NH experiment, $^{13}C^{\alpha}/^{13}C'$ and $^{13}C_i'/^{15}N_{i+1}$ scalar couplings are active for $2\tau_1$ and $2(\tau_1+\tau_2)$; however, the $^{13}C'$ chemical shift also evolves during this period in a constant-time fashion. Thus, at point *c* terms $-2C_z'N_y$ are generated that are frequency labeled with the $^{13}C'$ chemical shift. Following point *c*, the constant-time ^{15}N evolution period of the CBCA(CO)NH is replaced with a fixed delay during which the ^{15}N magnetization refocuses with respect to $^{13}C'$ and becomes antiphase with respect to its amide proton. This antiphase magnetization is then transferred to the amide proton and detected.

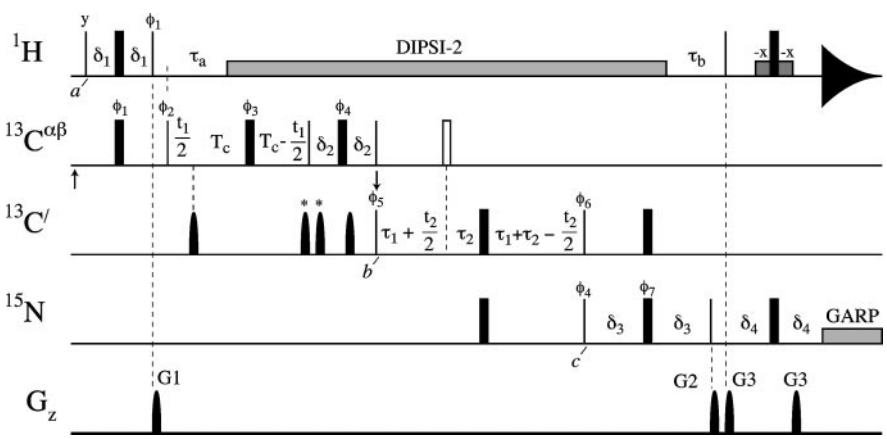

FIG. 1. Pulse sequence used to record the CBCACO(N)H experiment. Narrow and wide pulses were applied with flip angles of 90° and 180°, respectively along the *x* axis unless indicated otherwise. Proton pulses were applied using a field strength of 27 kHz, except for the water-selective pulses of the WATERGATE sequence (*27*), which were 1.0 ms in duration and applied using a 250-Hz field, and the pulses of the DIPSI-2 decoupling sequence (*28*), which were applied using a 5.0-kHz field. All ^{15}N pulses were applied using a 4.9-kHz field, except for those of the GARP decoupling sequence, which were applied using a 1.3-kHz field. $^{13}C^{\alpha\beta}$ pulses were applied at fields of $\Delta_{\alpha\beta}/\sqrt{15}$ for 90° pulses and $\Delta_{\alpha\beta}/\sqrt{3}$ for 180° pulses (filled bars), where $\Delta_{\alpha\beta}$ is the frequency difference in hertz between the centers of the aliphatic (43 ppm) and carbonyl (177 ppm) spectral regions. Arrows underneath the $^{13}C^{\alpha\beta}$ line indicate times when the ^{13}C frequency was shifted between 43 ($^{13}C^{\alpha\beta}$) and 177 ($^{13}C'$) ppm. The $^{13}C^{\alpha\beta}$ 180° pulse represented by the open bar was applied as a frequency-shifted rectangular pulse at 56 ppm using a field of $\Delta_{\alpha}/\sqrt{3}$, where Δ_{α} is the frequency difference in hertz between the centers of the $^{13}C^{\alpha}$ (56 ppm) and carbonyl (177 ppm) spectral regions. Rectangular $^{13}C'$ pulses (between *b* and *c*) were applied at fields of $\Delta_{\alpha}/\sqrt{15}$ for 90° pulses and $\Delta_{\alpha}/\sqrt{3}$ for 180° pulses. Shaped pulses (between *a* and *b*) were applied with a sinc profile for a duration of 300 μs. Pulses labeled with an asterisk were applied to compensate for Bloch–Siegert effects (*29*). The values of the delays were as follows: $\delta_1 = 1.5$ ms, $\delta_2 = 3.5$ ms, $\delta_3 = 13.0$ ms, $\delta_4 = 2.35$ ms, $\tau_a = 2.1$ ms, $\tau_b = 5.4$ ms, $\tau_1 = 4.5$ ms, $\tau_2 = 6.5$ ms, and $T_c = 3.1$ ms. The strengths and durations of the gradient pulses were as follows: G1 = (1.0 ms, 25 G/cm), G2 = (1.5 ms, 25 G/cm), and G3 = (0.4 ms, 25 G/cm). The phase cycle used was as follows: $\phi_1 = 8(x), 8(-x)$; $\phi_2 = x$ + States–TPPI(t_1); $\phi_3 = 2(x), 2(y), 2(-x), 2(-y)$; $\phi_4 = x, -x$; $\phi_5 = 4(x), 4(-x)$ + States–TPPI(t_2); $\phi_6 = 48°$; $\phi_7 = x, x, -x, -x$; and rec = $(x, -x, -x, x), 2(-x, x, x, -x), (x, -x, -x, x)$.

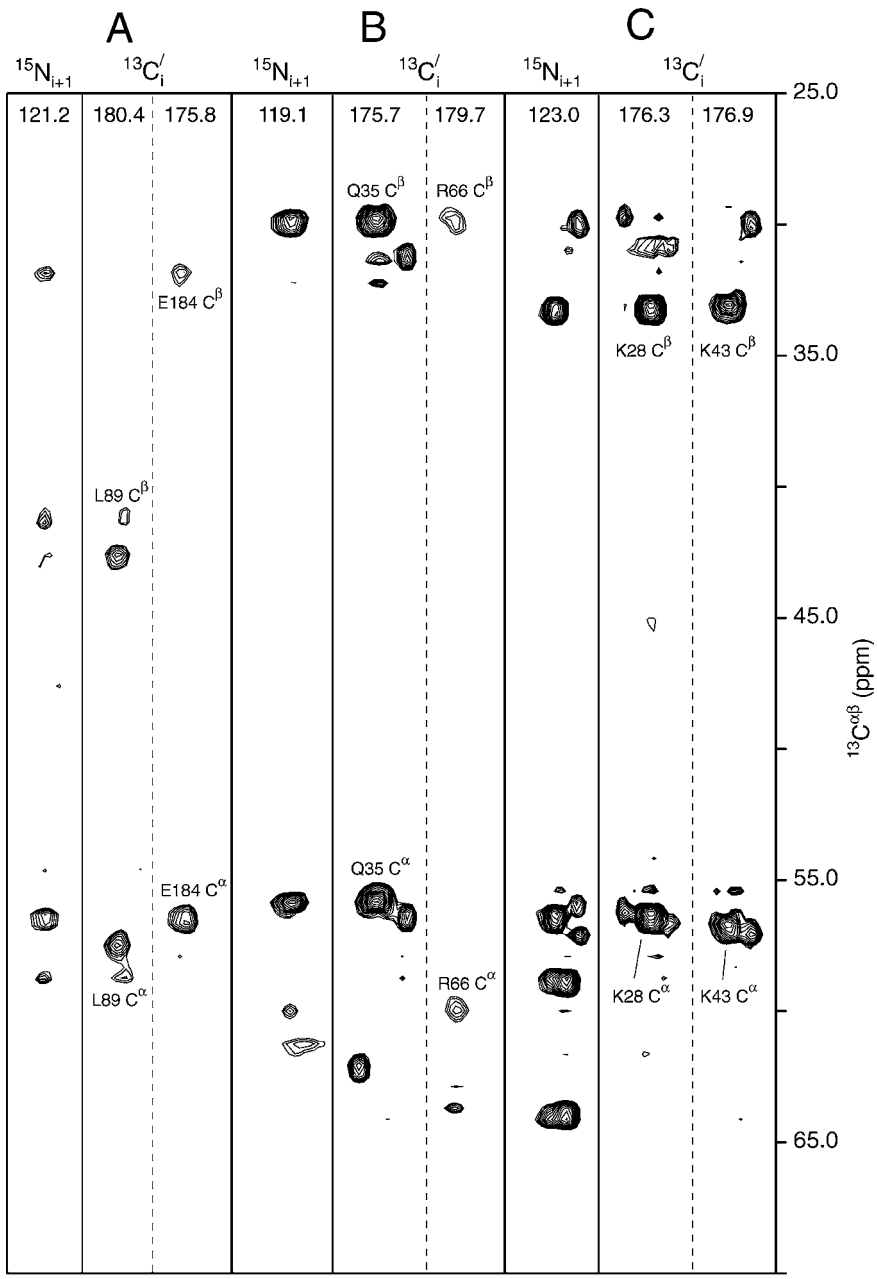

FIG. 2. F1–F3 strips taken from a 3D CBCA(CO)NH spectrum (strips labeled $^{15}N_{i+1}$) and a 3D CBCACO(N)H spectrum (strips labeled $^{13}C_i'$) at the common $^{1}H^{N}_{i+1}$ frequencies of three residue pairs of S4: (A) L89 and E184, (B) Q35 and R66, and (C) K28 and K43. The strips are further labeled with their ^{15}N or $^{13}C'$ chemical shift. Both spectra were recorded at 310 K on a sample containing 0.7 mM [*U*-^{13}C; *U*-^{15}N]-S4, 20 mM sodium acetate, pH 5.4, 250 mM KCl, and 1 mM EDTA in a volume of 300 μl 90% H_2O/10% D_2O. The CBCA(CO)NH spectrum was recorded on a Bruker DMX 750 MHz spectrometer, resulting in a 66 × 50 × 512 matrix corresponding to acquisition times of (6.6, 20.0, 54.1 ms) for (t_1, t_2, t_3), while the CBCACO(N)H spectrum was recorded on a Bruker DMX 500 MHz spectrometer, resulting in a 45 × 38 × 512 matrix with acquisition times of (6.7, 22.0, 80.5 ms).

Selected F1–F3 strips from 3D CBCA(CO)NH and CBCACO(N)H spectra of S4, Fig. 2, illustrate the progressively difficult kinds of overlap that are relieved by using both experiments. In all of the cases shown, both the ^{15}N and $^{1}H^{N}$ resonances of the two selected spin systems are simultaneously degenerate. In Fig. 2A, four peaks appear along a given ($^{1}H^{N}$, ^{15}N) strip, clearly announcing the presence of overlap, which is easily relieved by $^{13}C'$ chemical shifts. A somewhat more interesting case is depicted in Fig. 2B, in which only three peaks appear along the ($^{1}H^{N}$, ^{15}N) strip, two of which are much more intense than the third. This case arose from two residues with markedly different dynamic behavior (one from the flexible N-terminus and the other from the well-structured region) and with spin systems that are triply degenerate, having overlapping $^{13}C_i^{\beta}$

resonances in addition to $^{15}N_{i+1}$ and $^{1}H^{N}_{i+1}$. While this overlap is not conclusive from the CBCA(CO)NH alone, the CBCACO(N)H spectrum demonstrates it clearly, and allows a definitive assignment of the $^{13}C^{\beta}$ spins. Finally, in Fig. 2C the overlap of two quadruply degenerate ($^{13}C_i^{\beta}$, $^{13}C_i^{\alpha}$, $^{15}N_{i+1}$, and $^{1}H^{N}_{i+1}$) spin systems is almost undetectable in the CBCA(CO)NH spectrum. Again the CBCACO(N)H spectrum both relieves and confirms this degeneracy, enabling the accurate assignment of both spin systems.

From a more general perspective, the usefulness of a ($^{1}H^{N}_{i+1}$, $^{13}C'_i$) experiment such as the CBCACO(N)H in relieving overlap in ($^{1}H^{N}$, ^{15}N) experiments such as the CBCA(CO)NH depends upon the number of residues that have the same $^{1}H^{N}_{i+1}$, $^{15}N_{i+1}$, and $^{13}C'_i$ chemical shifts. Clearly, the HNCO experiment identifies such residue pairs directly, since the degeneracies of spin systems that are fully overlapped in 3D HNCO spectra cannot be relieved by using ($^{1}H^{N}$, ^{15}N) and ($^{1}H^{N}_{i+1}$, $^{13}C'_i$) experiments together. To illustrate these ideas, we have collected some statistics of relevant overlaps in Table 1 for three proteins: (1) human ubiquitin, representing a small, well-ordered protein; (2) *B. stearothermophilus* S4, representing a moderately sized (23 kDa) protein with both well-structured and highly flexible regions; and (3) *M. charantia* MAP30, representing a relatively large (30 kDa), well-ordered protein that also contains a short disordered region of 19 residues. In addition, we have listed statistics for S4 Δ41 and MAP30-C19, which correspond to fragments of S4 and MAP30 that lack the flexible regions of the full-length proteins. The first group of overlaps represents those that would be observed in a CBCA(CO)NH spectrum, the second group are those that would be observed in a CBCACO(N)H spectrum, and the third group are those that would be observed in both spectra. First, these data indicate that the resolving power of ($^{1}H^{N}$, ^{15}N) and ($^{1}H^{N}_{i+1}$, $^{13}C'_i$) experiments are essentially equal for these proteins. Second, the number of fully overlapped HNCO peaks ($^{1}H^{N}_{i+1}$–$^{15}N_{i+1}$–$^{13}C'_i$ overlaps) is no more than 17% of the total number of ($^{1}H^{N}$, ^{15}N) and ($^{1}H^{N}_{i+1}$, $^{13}C'_i$) overlaps, thereby allowing these degeneracies to be relieved for the great majority of these pairs by using the two experiments together. Indeed, this near uniqueness of their sets of overlapped peaks is the main advantage of using these experiments in combination. This uniqueness may result from the fact that the $^{13}C'$ and ^{15}N of the $^{1}H^{N}_{i+1}$–$^{15}N_{i+1}$–$^{3}C'_i$ spin system lie in different amino acids, minimizing any correlation between the shifts of these three spins. Making use of additional spins, particularly the $^{1}H^{\alpha}$ in the HCACO experiment, will often relieve the degeneracies of the remaining unresolved pairs. Notably, there are no cases of quintuply degenerate ($^{1}H^{\alpha}_i$, $^{13}C^{\alpha}_i$, $^{13}C'_i$, $^{15}N_{i+1}$, $^{1}H^{N}_{i+1}$) spin systems in any of the listed proteins.

We also note that the flexible regions of S4 and MAP30 contribute significantly to the overlapped sets. Even though the N-terminus comprises only 20% of the S4 sequence, it is responsible for 58 and 39% of the ($^{1}H^{N}$, ^{15}N) and ($^{1}H^{N}_{i+1}$, $^{13}C'_i$) overlaps, respectively. Similarly, the C-terminal 19 residues of MAP30 compose only 7.2% of the MAP30 sequence, yet contribute 29 and 22% of the ($^{1}H^{N}$, ^{15}N) and ($^{1}H^{N}_{i+1}$, $^{13}C'_i$) overlaps, respectively. In addition, for these proteins the size of the flexible region as a percentage of the total sequence seems to scale quite well with the number of new overlaps it contributes, expressed as a percentage of the number of overlaps in the well-structured region. The evidence thus suggests both that flexible regions contribute more overlap difficulties than ordered regions and that these difficulties increase simply in proportion to the fractional size of the flexible region in a given protein. However, if only the ordered regions of S4 and MAP30 are considered (S4 Δ41 and MAP30-C19), one finds that S4 Δ41 (a predominately α-helical protein) generates a significantly higher percentage of high multiplicity overlaps involving three or more spins than does MAP30-C19 (a mixed α-helical/β-sheet protein). For example, 39% of the ($^{1}H^{N}$, ^{15}N) overlaps in S4 Δ41 involve three or more spins, while only 12% do the same for MAP30-C19.

TABLE 1
Overlapping Spin Systems Observed in CBCA(CO)NH and CBCACO(N)H Experiments[a]

Spin system	Ubiquitin 76 aa	S4 Δ41 159 aa	S4 200 aa	MAP30-C19 239 aa	MAP30 263 aa[b]
H^{N}_{i+1}–N_{i+1} only	2	11	29	37	51
H^{N}_{i+1}–N_{i+1}–C^{α}_i	1	5	7	3	3
H^{N}_{i+1}–N_{i+1}–C^{β}_i	0	1	3	1	3
H^{N}_{i+1}–N_{i+1}–$C^{\alpha/\beta}_i$	0	0	0	1	1
H^{N}_{i+1}–N_{i+1}–C^{α}_i–C^{β}_i	0	1	4	0	0
Total H^{N}_{i+1}–N_{i+1} overlaps	**3**	**18**	**43**	**41**	**58**
H^{N}_{i+1}–C'_i only	4	17	28	48	61
H^{N}_{i+1}–C'_i–C^{α}_i	1	4	5	5	6
H^{N}_{i+1}–C'_i–C^{β}_i	0	5	5	1	3
H^{N}_{i+1}–C'_i–$C^{\alpha/\beta}_i$	0	0	0	0	0
H^{N}_{i+1}–C'_i–C^{α}_i–C^{β}_i	0	2	8	1	1
Total H^{N}_{i+1}–C'_i overlaps	**5**	**28**	**46**	**55**	**71**
H^{N}_{i+1}–N_{i+1}–C'_i only	0	1	1	6	7
H^{N}_{i+1}–N_{i+1}–C'_i–C^{α}_i	0	1	1	1	1
H^{N}_{i+1}–N_{i+1}–C'_i–C^{β}_i	0	1	1	0	1
H^{N}_{i+1}–N_{i+1}–C'_i–C^{α}_i–C^{β}_i	0	0	1	0	0
Total H^{N}_{i+1}–N_{i+1}–C'_i overlaps	**0**	**3**	**4**	**7**	**9**

[a] A given spin system of n spins, $(A_1, A_2, \ldots, A_n)$, in residue A was considered to be overlapped with the same spin system, $(B_1, B_2, \ldots, B_n)$, of residue B if $|(\delta A_i - \delta B_i)| < 2\Delta_i$ $(i = 1, \ldots, n)$ for all n spins, where δ values are chemical shifts in ppm taken from chemical shift tables and Δ_i is the final digital resolution of spin i after processing. Values of Δ_i were as follows (in ppm): $\Delta H^{N} = 0.02$, $\Delta^{15}N = 0.25$, $\Delta^{13}C^{\alpha} = 0.40$, and $\Delta^{13}C' = 0.10$. The numbers in the table indicate the total number of such overlaps between residue pairs observed for each listed spin system. Numbers in bold are sums of the overlaps of spin systems sharing the indicated spin pair or triplet in common.

[b] Signals for residues 240–245 were weak or absent (*26*).

These percentages are almost identical for ($^1H^N_{i+1}$, $^{13}C'_i$) overlaps involving three or more spins: 39% for S4 Δ41 and 13% for MAP30-C19. Thus, while the combined use of the two experiments clearly seems to benefit the analysis of flexible regions of proteins, this strategy may also be useful for a number of well-structured proteins that have significant inherent signal overlap.

Simultaneous ($^1H^\alpha$, $^{13}C^\alpha$) Overlap in Residues Preceding Prolines

We next turn to how the H(CA)CON experiment (Fig. 3) can be used to relieve the problem of simultaneous overlap of ($^1H^\alpha$,$^{13}C^\alpha$) spin pairs, particularly in proline residues and amino acids immediately N-terminal to proline residues. The H(CA)CON experiment was derived from both the 3D constant-time HCA(CO)N (*4*, *22*) and the 4D HCACON (*21*) experiments. Between points *a* and *b*, the experiment is fundamentally equivalent to the HCA(CO)N experiment, except that the constant-time $^{13}C^\alpha$ evolution period is replaced with a fixed delay period during which the antiphase $^{13}C^\alpha$ magnetization refocuses with respect to $^1H^\alpha$ and becomes antiphase with respect to its directly bonded $^{13}C'$. This antiphase $^{13}C^\alpha$ magnetization is then transferred to the $^{13}C'$, generating at point *b* the term $2C^\alpha_z C'_x$. At this point the $^{13}C'$ chemical shifts are allowed to evolve during a constant-time period whose length was chosen to maximize antiphase $^{13}C'$ magnetization with respect to the ^{15}N of the following residue. Thus at point *c* the INEPT transfer generates the term $-4C^\alpha_z C'_z N_y$, labeled with the chemical shift of the $^{13}C'$ spin. We note that this term is an antiphase ^{15}N single-quantum coherence, in contrast to the two-spin $^{13}C'$–^{15}N multiple-quantum coherences generated by the HMQC-style transfer employed in the 3D HCA(CO)N and 4D HCACON experiments. The ^{15}N chemical shift evolves during t_2, after which the ^{15}N magnetization is transferred back to the $^{13}C'$. The remainder of the sequence essentially reverses the magnetization transfers accomplished to this point, ultimately returning the magnetization to the original $^1H^\alpha$, which is then detected.

Selected F1–F3 strips from a 3D H(CA)CON spectrum recorded on S4 are shown in Fig. 4 alongside F2–F3 strips from a 3D HCACO spectrum. Figure 4A illustrates the kind of situation where resolving spin systems using spins other than the $^{13}C^\alpha$ can be useful. The resonances of four proline residues of the N-terminus are displayed, all of which have simultaneously degenerate ($^1H^\alpha$,$^{13}C^\alpha$) spin pairs. In addition, the spin systems of P7, P33, and P38 are triply degenerate, also sharing the same $^{13}C'$ chemical shift. The H(CA)CON was partly able to relieve this degeneracy, separating P7 from P33/P38 on the basis of the $^{15}N_{i+1}$ shift. Notably, these data highlight the fact that P7 and P30 are also triply degenerate, with equivalent $^1H^\alpha_i$, $^{13}C^\alpha_i$, and $^{15}N_{i+1}$ shifts, indicating that these spin systems would be completely overlapped in an HCA(CO)N spectrum. Also revealed is the quadruple degeneracy of P33 and P38, which have equivalent $^1H^\alpha_i$, $^{13}C^\alpha_i$, $^{13}C'_i$, and $^{15}N_{i+1}$ shifts. Fortunately, the $^1H^N$ shifts of G34 and G39 differ by 0.05 ppm, allowing the assignment of P33 and P38. However, in the general case of a flexible, proline-rich peptide with numerous adjacent prolines, amide

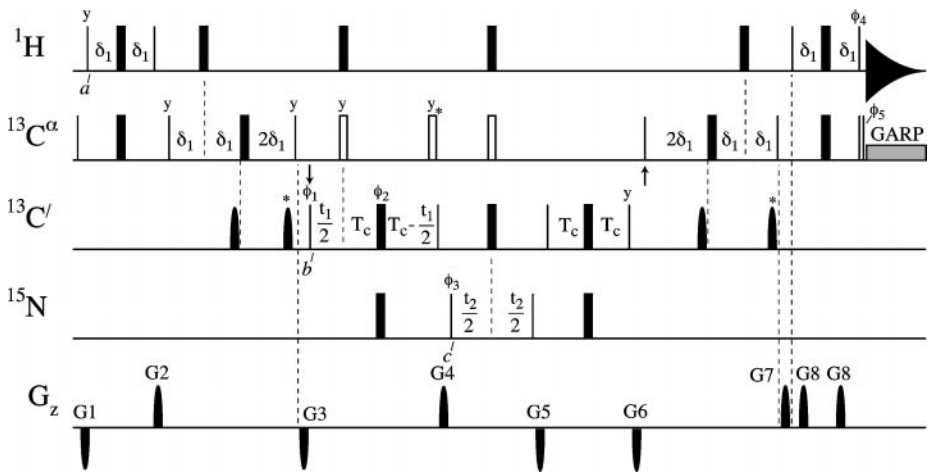

FIG. 3. Pulse sequence used to record the H(CA)CON experiment. Narrow and wide pulses were applied with flip angles of 90° and 180°, respectively, along the *x* axis unless indicated otherwise. Proton and ^{15}N pulses were applied using field strengths of 27 and 4.9 kHz, respectively. Rectangular $^{13}C^\alpha$ pulses were applied at fields of $\Delta_\alpha/\sqrt{15}$ for 90° pulses and $\Delta_\alpha/\sqrt{3}$ for 180° pulses (filled bars), where Δ_α is the frequency difference in hertz between the centers of the $^{13}C^\alpha$ (56 ppm) and $^{13}C'$ (177 ppm) spectral regions. Pulses for the GARP decoupling sequence were applied using a 2.9-kHz field. Arrows underneath the $^{13}C^\alpha$ line indicate times when the ^{13}C frequency was shifted between 56 ($^{13}C^\alpha$) and 177 ($^{13}C'$) ppm. The $^{13}C^\alpha$ 180° pulses represented by the open bars were applied as frequency-shifted rectangular pulses at 56 ppm using a field of $\Delta_\alpha/\sqrt{3}$. Rectangular $^{13}C'$ pulses were applied at fields of $\Delta_\alpha/\sqrt{15}$ for 90° pulses and $\Delta_\alpha/\sqrt{3}$ for 180° pulses. Shaped pulses were applied with a sinc profile for a duration of 240 μs. Pulses labeled with an asterisk were applied to compensate for Bloch–Siegert effects (*29*). The values of the delays were as follows: $\delta_1 = 1.7$ ms and $T_c = 12.5$ ms. The strengths and durations of the gradient pulses were as follows: G1 = (0.75 ms, –25 G/cm), G2 = (2.1 ms, 25 G/cm), G3 = (0.5 ms, –25 G/cm), G4 = (0.3 ms, 25 G/cm), G5 = (0.45 ms, –25 G/cm), G6 = (0.5 ms, –25 G/cm), G7 = (0.8 ms, 25 G/cm), G8 = (0.3 ms, 25 G/cm). The phase cycle used was as follows: $\phi_1 = y, -y$ + States–TPPI(t_1); $\phi_2 = 4(x), 4(y), 4(-x), 4(-y)$; $\phi_3 = x, x, -x, -x$ + States–TPPI(t_2); $\phi_4 = -x, x$; $\phi_5 = 8(x), 8(-x)$; and rec = $x, -x, -x, x, -x, x, x, -x$.

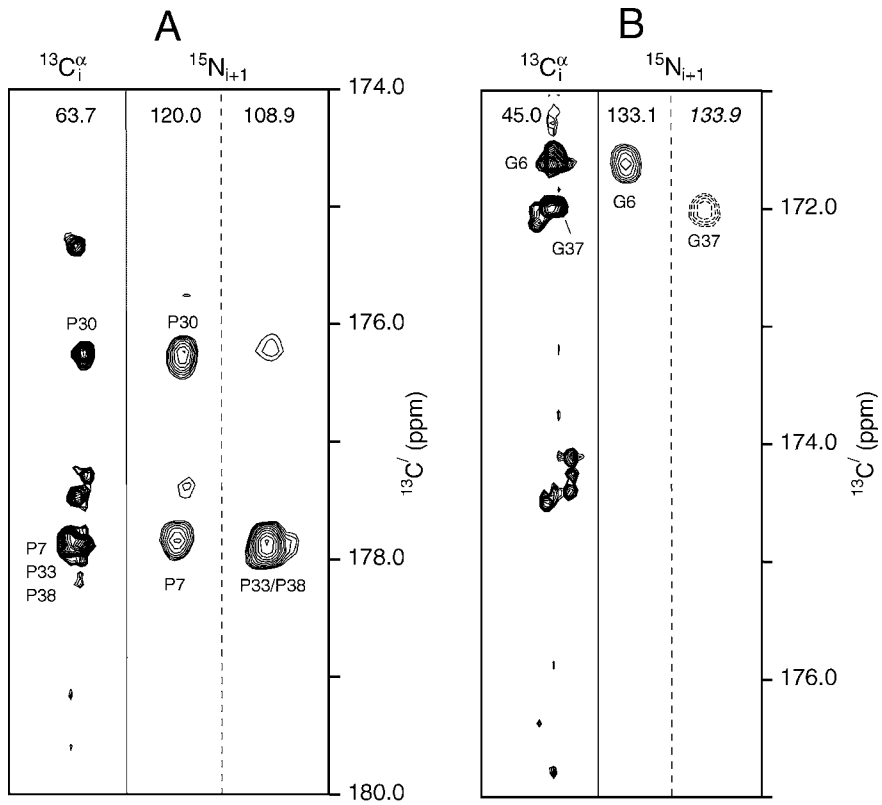

FIG. 4. F2–F3 strips taken from a 3D HCACO spectrum (strips labeled $^{13}C_i^{\alpha}$) alongside F1–F3 strips from a 3D H(CA)CON spectrum (strips labeled $^{15}N_{i+1}$) at the common $^1H^{\alpha}$ frequencies of two residue sets of S4: (A) P7, P30, P33, and P38 and (B) G6 and G37. The strips are further labeled with their $^{13}C^{\alpha}$ or ^{15}N chemical shift, with italic type indicating folding in the ^{15}N dimension. Negative contours resulting from such folding are indicated by dashed lines. Both spectra were recorded at 310 K using a Bruker DMX 500 MHz spectrometer, and the sample conditions were the same as those described in the legend to Fig. 2, except that the solvent was 100% D_2O. The HCACO spectrum was collected as a $23 \times 128 \times 512$ matrix corresponding to acquisition times of (7.0, 84.1, 102.4 ms) for (t_1, t_2, t_3), while the H(CA)CON spectrum was collected as a $33 \times 60 \times 384$ matrix with acquisition times of (21.7, 36.0, 76.8 ms).

protons may well be unavailable, suggesting the need to use the resolving power of all available spins. A related case is shown in Fig. 4B, which involves another pair of residues, G6 and G37, which both have simultaneously degenerate ($^1H^{\alpha}$,$^{13}C^{\alpha}$) spin pairs and are N-terminal to a proline. Therefore, the only available i to $i+1$ connectivity for these residues involves the proline ^{15}N, and since the $^{13}C'$ spins of G6 and G37 are well resolved, this connectivity is clearly provided by the H(CA)CON experiment. Unfortunately, this did not complete the assignment, since the available $i-1$ to i connectivities for G6 and G37 involve their ^{15}N and $^1H^N$ spins, not their $^1H^{\alpha}$ or $^{13}C'$ spins. Ultimately a careful comparison of the HNCACB, HCACO, and HNHA spectra to the H(CA)CON spectrum was able to finalize these assignments. While the recently proposed version of the HACAN experiment (*18*) would in principle be a straightforward solution to this problem, given that it would correlate the ^{15}N spins of both G6 and P7 to the G6 ($^1H^{\alpha}$,$^{13}C^{\alpha}$) pair (and equivalently for G37 and P38), in this and similar cases the experiment would be difficult to interpret due to the simultaneous ($^1H^{\alpha}$,$^{13}C^{\alpha}$) degeneracy. Moreover, as is the case with the CBCA(CO)N(CA)HA experiment (discussed below), these connectivities would be severely attenuated in the HACAN spectrum because the magnetization in question originates on glycine residues (*18*).

The general usefulness of ($^1H^{\alpha}$,$^{13}C'$) experiments such as the H(CA)CON in relieving ($^1H^{\alpha}$,$^{13}C^{\alpha}$) overlaps can be seen from the data in Table 2, where relevant overlap statistics are presented for ubiquitin, S4, S4 Δ41, MAP30, and MAP30-C19. Again the data suggest that the resolving power of the ($^1H^{\alpha}$,$^{13}C^{\alpha}$) and ($^1H^{\alpha}$,$^{13}C'$) experiments is very similar for these proteins. What is notable is that unlike the data for the ($^1H^N$, ^{15}N)

TABLE 2
Overlapping Spin Systems Observed in HCA(CO)N and H(CA)CON Experiments[a]

Spin system	Ubiquitin	S4 Δ41	S4	MAP30-C19	MAP30
H_i^{α}–C_i^{α} overlaps	7	55	116	60	81
H_i^{α}–C_i' overlaps	7	47	88	59	80
H_i^{α}–C_i^{α}–C_i' overlaps	1	5	24	11	11

[a] Spin systems are considered overlapped as defined in the footnote to Table 1. Values of Δ are also the same, with the addition that $\Delta^1H^{\alpha} = 0.02$ ppm.

and ($^1H^N_{i+1}$, $^{13}C'_i$) experiments, in which the number of overlaps increased with the size of the protein, S4 has the greatest number of overlaps for the ($^1H^\alpha$,$^{13}C^\alpha$) and ($^1H^\alpha$,$^{13}C'$) experiments. Moreover, the flexible regions of S4 (and MAP30) contribute significantly to these overlaps, accounting for 52 and 46% of the ($^1H^\alpha$,$^{13}C^\alpha$) and ($^1H^\alpha$,$^{13}C'$) overlaps in the case of S4, respectively, and 26% of both classes of overlaps in the case of MAP30. Because many proline-rich regions in proteins tend to be similarly flexible in solution, we expect that ($^1H^\alpha$,$^{13}C'$) experiments such as the H(CA)CON should prove quite useful in overcoming the extensive overlap in such regions.

Experimental Considerations

In an attempt to compare the CBCACO(N)H and H(CA)CON experiments to existing pulse sequences, we collected a series of 2D planes from relevant experiments on a sample containing [*U*-^{13}C; *U*-^{15}N]-human ubiquitin. We then quantitated the intensities of as many well-resolved peaks as possible in each spectrum and determined the average signal-to-noise ratio. Turning first to the CBCACO(N)H and CBCA(CO)NH experiments, these measurements on $^1H^N$–$^{13}C^{\alpha\beta}$ planes indicated that for ubiquitin the sensitivities of these experiments were statistically equivalent (data not shown), as expected given that the experiments share essentially the same magnetization transfer pathway. This equivalence appears to extend to larger proteins as well, based on the data acquired on S4. Similarly, the sensitivities of the CBCA(CO)N(CA)HA, HCA(CO)N, and H(CA)CON experiments as measured from $^1H^\alpha$–$^{13}C^\alpha$ or $^1H^\alpha$–$^{13}C'$ planes were also found to be statistically equivalent, each being roughly one-third as sensitive as the HCACO (data not shown). However, when analyzing $^1H^\alpha_i$–$^{15}N_{i+1}$ planes from these latter experiments, we noted that the ^{15}N linewidths in the HCA(CO)N were uniformly broader than those of the H(CA)CON and CBCA(CO)N(CA)HA, likely resulting in part from the enhanced relaxation of the multiple-quantum terms produced in the HCA(CO)N during t_2. If we used the peak volumes from $^1H^\alpha_i$–$^{15}N_{i+1}$ planes rather than the peak intensities to measure the sensitivity, we again found that the sensitivities of the HCA(CO)N and H(CA)CON were statistically equivalent. Nevertheless, when applied to larger proteins, the improved ^{15}N lineshape of the H(CA)CON experiment may well be advantageous.

In general, it should be noted that the ^{15}N lineshapes in the H(CA)CON (and HCA(CO)N) would also be improved by amide ^{2}H decoupling during the ^{15}N evolution time for samples in D_2O. Such decoupling was not used in the case of the S4 N-terminus because efficient amide ^{2}H spin–lattice relaxation nearly eliminates the ^{15}N–^{2}H *J*-coupling. A straightforward calculation (*17*, *23*, *24*) showed that for $\tau_c \sim 1$ ns (the approximate correlation time of the S4 N-terminus) and $T_1(^2H) \sim 2.5$ ms, the ^{15}N line broadening due to the residual ^{15}N–^{2}H *J*-coupling is less than ~5 Hz. However, when the H(CA)CON experiment is applied to proteins having amide ^{2}H T_1 values >7 ms, e.g., when $\tau_c > \sim 8$ ns at $B_o = 12$ T, ^{2}H decoupling during t_2 is highly recommended. Indeed, the sensitivity for the well-ordered residues of S4 ($\tau_c \sim 12$ ns) was markedly reduced in the H(CA)CON, supporting the use of ^{2}H decoupling. If circumstances require, the sensitivity of the H(CA)CON could be further improved by gradient enhanced detection (*25*).

The H(CA)CON experiment offers additional advantages as well. First, the H(CA)CON offers improved resolution relative to the ($^1H^\alpha$,$^{13}C^\alpha$) experiments that arises simply from the greater length of the constant-time $^{13}C'$ dimension of the H(CA)CON (21.7 ms) compared to that of the $^{13}C^\alpha$ dimension of the HCA(CO)N and CBCA(CO)N(CA)HA (7.0 ms). Therefore, the resolution in ppm/pt of the $^{13}C'$ dimension (before any post-acquisition processing) will be higher than that of the $^{13}C^\alpha$ dimensions by a factor of 3.1 (at 500 MHz, 0.37 ppm/pt for $^{13}C'$ and 1.1 ppm/pt for $^{13}C^\alpha$). Second, we note that the H(CA)CON and the HCA(CO)N experiments both provide sequential connectivities for all dipeptide units, unlike the CBCA(CO)N(CA)HA experiment, which severely attenuates magnetization originating on glycines (*18*). This is particularly important for assigning occurrences of the Gly–Pro dipeptide, which is quite common in many proteins.

Finally, the strategy presented here of using 3D experiments with $^{13}C'$ dimensions, either alone or paired with related 3D experiments, offers advantages over other approaches. While one might consider that 4D CBCACONH or HCACON experiments would provide an equivalent solution to the overlap problems addressed here, such spectra would suffer a significant loss of resolution and would require more experimental time than their 3D analogues. For example, the 3D CBCA(CO)NH and CBCACO(N)H each require ~36 h of instrument time for 0.5–1.0 mM samples, so that both experiments can be acquired in less time than one 4D CBCACONH experiment. One might also change sample conditions (temperature, pH) to relieve overlap problems; however, such changes may be problematic or impossible for some samples, and in any case may create new overlap problems that did not otherwise exist. Our approach of using the $^{13}C'$ spin requires only a single sample, and thus is a simple and convenient method for relieving overlap problems in protein spectra.

CONCLUSIONS

We have presented data showing that two modified pulse sequences, the CBCACO(N)H and H(CA)CON, can successfully employ the carbonyl chemical shift to relieve simultaneous ($^1H^N$, ^{15}N) and ($^1H^\alpha$,$^{13}C^\alpha$) degeneracies in spectra of S4. Based on the overlaps found in ubiquitin, S4, and MAP30, the data suggest that the benefits of using the carbonyl chemical shift may well extend to many proteins that either are large or contain flexible regions. Moreover, the utility of the carbonyl chemical shift is general, and extends to other related experiments, suggesting the benefits of additional variants such as the HCCO(N)H,

CCO(N)H, and HBHACO(N)H. These experiments should be useful additions to the current armament of assignment experiments, particularly as proteins with increasing amounts of resonance overlap are studied.

ACKNOWLEDGMENTS

This work was supported in part by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health, along with a Pharmacology Research Associate Award from the National Institute of General Medical Sciences to E.W.S. We thank Frank Delaglio for NMRPipe and NMRDraw and Dan Garrett for PIPP. We also thank Michelle Markus, Rieko Ishima, and Ad Bax for reviewing the manuscript and for helpful discussions.

REFERENCES

1. M. Wittekind and L. Mueller, HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins, *J. Magn. Reson. B* **101,** 201–205 (1993).
2. S. Grzesiek and A. Bax, Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR, *J. Am. Chem. Soc.* **114,** 6291–6293 (1992).
3. S. Grzesiek and A. Bax, Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein, *J. Magn. Reson.* **96,** 432–440 (1992).
4. R. Powers, A. M. Gronenborn, G. M. Clore, and A. Bax, Three-dimensional triple-resonance NMR of $^{13}C/^{15}N$-enriched proteins using constant time evolution, *J. Magn. Reson.* **94,** 209–213 (1991).
5. C. Davies, R. B. Gerstner, D. E. Draper, V. Ramakrishnan, and S. W. White, The crystal structure of ribosomal protein S4 reveals a two domain molecule with an extensive RNA-binding surface: One domain shows structural homology to the ETS DNA-binding motif, *EMBO J.* **17,** 4545–4558 (1998).
6. M. A. Markus, R. B. Gerstner, D. E. Draper, and D. A. Torchia, The solution structure of ribosomal protein S4 delta41 reveals two subdomains and a positively charged surface that may interact with RNA, *EMBO J.* **17,** 4559–4571 (1998).
7. M. A. Markus, R. B. Gerstner, D. E. Draper, and D. A. Torchia, Refining the overall structure and subdomain orientation of ribosomal protein S4Δ41 with dipolar couplings measured by NMR in uniaxial liquid crystalline phases, *J. Mol. Biol.* **292,** 375–387 (1999).
8. E. W. Sayers, R. B. Gerstner, D. E. Draper, and D. A. Torchia, Structural preordering in the N-terminal region of ribosomal protein S4 revealed by heteronuclear NMR spectroscopy, *Biochemistry* **39,** 13,602–13,613 (2000).
9. L. E. Kay, M. Ikura, A. A. Grey, and D. R. Muhandiram, Three-dimensional NMR experiments for the separation of side-chain correlations in proteins via the carbonyl chemical shift, *J. Magn. Reson.* **99,** 652–659 (1992).
10. L. E. Kay, Pulsed-field gradient-enhanced three-dimensional NMR experiment for correlating $^{13}C^{\alpha/\beta}$, $^{13}C'$, and $^{1}H^{\alpha}$ chemical shifts in uniformly ^{13}C-labeled proteins dissolved in H_2O, *J. Am. Chem. Soc.* **115,** 2055–2057 (1993).
11. R. Bazzo, D. O. Cicero, and G. Barbato, A new three-dimensional pulse sequence for correlating intraresidue NH, N, and CO chemical shifts in C-13, N-15-labeled proteins, *J. Magn. Reson. B* **110,** 65–68 (1996).
12. F. Löhr and H. Rüterjans, A new triple-resonance experiment for the sequential assignment of backbone resonances in proteins, *J. Biomol. Nucl. Magn. Reson.* **6,** 189–197 (1995).
13. T. Szyperski, D. Braun, C. Fernández, C. Bartels, and K. Wüthrich, A novel reduced-dimensionality triple-resonance experiment for efficient polypeptide backbone assignment, 3D COHNNCA, *J. Magn. Reson. B* **108,** 197–203 (1995).
14. N. Astrof, C. Bracken, J. Cavanagh, and A. G. Palmer, CO_H(N)CACB experiments for assigning backbone resonances in $^{13}C/^{15}N$-labeled proteins, *J. Biomol. Nucl. Magn. Reson.* **11,** 451–456 (1998).
15. R. Konrat, D. W. Yang, and L. E. Kay, A 4D TROSY-based pulse scheme for correlating (HNi)-H-1, N(i)-15, C-13(i) alpha, C′-13 (i-1) chemical shifts in high molecular weight, N-15, C-13, H-2 labeled proteins, *J. Biomol. Nucl. Magn. Reson.* **15,** 309–313 (1999).
16. M. J. Bottomley, M. J. Macias, Z. Liu, and M. Sattler, A novel NMR experiment for the sequential assignment of proline residues and proline stretches in $^{13}C/^{15}N$-labeled proteins, *J. Biomol. Nucl. Magn. Reson.* **13,** 381–385 (1999).
17. A. C. Wang, S. Grzesiek, R. Tschudin, P. J. Lodi, and A. Bax, Sequential backbone assignment of isotopically enriched proteins in D_2O by deuterium-decoupled HA(CA)N and HA(CACO)N, *J. Biomol. Nucl. Magn. Reson.* **5,** 376–382 (1995).
18. V. Kanelis, L. Donaldson, D. R. Muhandiram, D. Rotin, J. D. Forman-Kay, and L. E. Kay, Sequential assignment of proline-rich regions in proteins: Application to modular binding domain complexes, *J. Biomol. Nucl. Magn. Reson.* **16,** 253–259 (2000).
19. M. Schubert, L. J. Ball, H. Oschkinat, and P. Schmieder, Bridging the gap: A set of selective ^{1}H-^{15}N-correlations to link sequential neighbors of prolines, *J. Biomol. Nucl. Magn. Reson.* **17,** 331–335 (2000).
20. L. E. Kay, M. Ikura, R. Tschudin, and A. Bax, Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins, *J. Magn. Reson.* **89,** 496–514 (1990).
21. L. E. Kay, M. Ikura, G. Zhu, and A. Bax, Four-dimensional heteronuclear triple-resonance NMR of isotopically enriched proteins for sequential assignment of backbone atoms, *J. Magn. Reson.* **91,** 422–428 (1991).
22. A. G. Palmer, W. J. Fairbrother, J. Cavanagh, P. E. Wright, and M. Rance, Improved resolution in 3-dimensional constant-time triple resonance NMR-spectroscopy of proteins, *J. Biomol. Nucl. Magn. Reson.* **2,** 103–108 (1992).
23. A. C. LiWang and A. Bax, Solution NMR characterization of hydrogen bonds in a protein by indirect measurement of deuterium quadrupole couplings, *J. Magn. Reson.* **127,** 54–64 (1997).
24. J. A. Pople, The effect of quadrupole relaxation on nuclear magnetic resonance multiplets, *Mol. Phys.* **1,** 168–174 (1958).
25. L. E. Kay, P. Keifer, and T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc.* **114,** 10,663–10,665 (1992).
26. Y. X. Wang, N. Neamati, J. Jacob, I. Palmer, S. J. Stahl, J. D. Kaufman, P. L. Huang, H. E. Winslow, Y. Pommier, P. T. Wingfield, S. Lee-Huang, A. Bax, and D. A. Torchia, Solution structure of anti-HIV-1 and anti-tumor protein MAP30: Structural insights into its multiple functions, *Cell* **99,** 433–442 (1999).
27. M. Piotto, V. Saudek, and V. Sklenar, Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, *J. Biomol. Nucl. Magn. Reson.* **2,** 661–665 (1992).
28. A. J. Shaka, C. J. Lee, and A. Pines, Iterative schemes for bilinear operators: Application to spin decoupling, *J. Magn. Reson.* **77,** 274–293 (1988).
29. G. W. Vuister and A. Bax, Resolution enhancement and spectral editing of uniformly ^{13}C-enriched proteins by homonuclear broadband ^{13}C decoupling, *J. Magn. Reson.* **98,** 428–435 (1992).