J|A|C|S
ARTICLES
Published on Web 05/25/2005

(4,2)D Projection−Reconstruction Experiments for Protein Backbone Assignment: Application to Human Carbonic Anhydrase II and Calbindin D_{28K}

Ronald A. Venters,*,† Brian E. Coggins,‡ Doug Kojetin,# John Cavanagh,# and Pei Zhou*,‡

Contribution from the Duke University NMR Center, Duke University Medical Center, Durham, North Carolina 27710, Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, and Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina 27695

Received February 14, 2005; E-mail: venters@windsurf.mc.duke.edu; peizhou@biochem.duke.edu

Abstract: Projection−reconstruction NMR experiments have been shown to significantly reduce the acquisition time required to obtain protein backbone assignment data. To date, this concept has only been applied to smaller $^{15}N/^{13}C$-labeled proteins. Here, we show that projection−reconstruction NMR techniques can be extended to larger protonated and perdeuterated proteins. We present a suite of (4,2)D triple-resonance experiments for protein backbone assignment and a Hybrid Backprojection/Lower-Value algorithm for reconstructing data with relatively weak signal-to-noise ratios. In addition, we propose a sampling theorem and discuss its implication on the choice of projection angles. We demonstrate the efficacy of this approach using the 29 kDa protein, human carbonic anhydrase II and the 30 kDa protein, calbindin D_{28K}.

Introduction

Advances in molecular biology, isotopic labeling strategies, pulse sequence design, and instrument sensitivity have all contributed to decreasing the time required to complete assignments and determine the three-dimensional structures of proteins by NMR. Additionally, these advances have dramatically increased the size of proteins amenable to study. However, the assignment process continues to rely upon 3D and 4D NMR data that, when collected in the traditional way by systematically sampling each of the evolution dimensions independently, result in the need for substantial amounts of spectrometer time. Four-dimensional experimental data, which are often required to provide the requisite resolution in studies involving larger perdeuterated proteins, acquired by this conventional approach, often take 5−10 days each using the highest field instruments and highest sensitivity probes available, even with severely curtailed resolution. In attempts to reduce this time limitation, several groups have developed ideas for fast multidimensional NMR, including reduced dimensionality and G-matrix Fourier transform NMR,[1−11] filter diagonalization,[12−14] nonlinear sampling and maximum entropy reconstruction,[15,16] single-scan multidimensional NMR,[17,18] and Hadamard NMR.[19−21] In addition to publishing an excellent review of these methodologies,[22] Kupče and Freeman have recently introduced the concept of projection−reconstruction (PR) to the protein NMR community,[23−30] a method widely used in disparate fields, including X-ray tomography, medical imaging, and astronomy. They have elegantly demonstrated that it is possible to recon-

† Duke University NMR Center, Duke University Medical Center.
‡ Department of Biochemistry, Duke University Medical Center.
# Department of Molecular and Structural Biochemistry, North Carolina State University.

(1) Atreya, H. S.; Szyperski, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (26), 9642−9647.
(2) Bersch, B.; Rossy, E.; Coves, J.; Brutscher, B. *J. Biomol. NMR* **2003**, *27* (1), 57−67.
(3) Brutscher, B.; Simorre, J. P.; Caffrey, M. S.; Marion, D. *J. Magn. Reson. B* **1994**, *105* (1), 77−82.
(4) Ding, K.; Gronenborn, A. M. *J. Magn. Reson.* **2002**, *156* (2), 262−268.
(5) Kim, S.; Szyperski, T. *J. Am. Chem. Soc.* **2003**, *125* (5), 1385−1393.
(6) Kozminski, W.; Zhukov, I. *J. Biomol. NMR* **2003**, *26* (2), 157−166.
(7) Lohr, F.; Ruterjans, H. *J. Biol. NMR* **1995**, *6*, 189−197.
(8) Simorre, J. P.; Brutscher, B.; Caffrey, M. S.; Marion, D. *J. Biomol. NMR* **1994**, *4* (3), 325−333.
(9) Szyperski, T.; Wider, G.; Bushweller, J. H.; Wüthrich, K. *J. Am. Chem. Soc.* **1993**, *115* (20), 9307−9308.
(10) Szyperski, T.; Wider, G.; Bushweller, J. H.; Wüthrich, K. *J. Biomol. NMR* **1993**, *3* (1), 127−132.
(11) Szyperski, T.; Yeh, D. C.; Sukumaran, D. K.; Moseley, H. N.; Montelione, G. T. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (12), 8009−8014.
(12) Chen, J.; Mandelshtam, V. A.; Shaka, A. J. *J. Magn. Reson.* **2000**, *146* (2), 363−368.
(13) Hu, H.; De Angelis, A. A.; Mandelshtam, V. A.; Shaka, A. J. *J. Magn. Reson.* **2000**, *144* (2), 357−366.
(14) Mandelshtam, V. A. *J. Magn. Reson.* **2000**, *144* (2), 343−356.
(15) Orekhov, V. Y.; Ibraghimov, I.; Billeter, M. *J. Biomol. NMR* **2003**, *27* (2), 165−173.
(16) Rovnyak, D.; Frueh, D. P.; Sastry, M.; Sun, Z. Y.; Stern, A. S.; Hoch, J. C.; Wagner, G. *J. Magn. Reson.* **2004**, *170* (1), 15−21.
(17) Frydman, L.; Scherf, T.; Lupulescu, A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (25), 15858−15862.
(18) Frydman, L.; Lupulescu, A.; Scherf, T. *J. Am. Chem. Soc.* **2003**, *125* (30), 9204−9217.
(19) Kupče, E.; Freeman, R. *J. Magn. Reson.* **2003**, *163* (1), 56−63.
(20) Kupče, E.; Freeman, R. *J. Magn. Reson.* **2003**, *162* (2), 300−310.
(21) Kupče, E.; Freeman, R. *J. Magn. Reson.* **2003**, *162* (1), 158−165.
(22) Freeman, R.; Kupče, E. *J. Biomol. NMR* **2003**, *27* (2), 101−113.
(23) Freeman, R.; Kupče, E. *Concepts Magn. Reson.* **2004**, *23A* (2), 63−75.
(24) Kupče, E.; Freeman, R. *J. Am. Chem. Soc.* **2003**, *125* (46), 13958−13959.
(25) Kupče, E.; Freeman, R. *J. Biomol. NMR* **2003**, *27* (4), 383−387.
(26) Kupče, E.; Freeman, R. *J. Biomol. NMR* **2003**, *25* (4), 349−354.
(27) Kupče, E.; Freeman, R. *J. Am. Chem. Soc.* **2004**, *126* (20), 6429−6440.
(28) Kupče, E.; Freeman, R. *J. Biomol. NMR* **2004**, *28* (4), 391−395.
(29) Kupče, E.; Freeman, R. *Concepts Magn. Reson.* **2004**, *22A* (1), 4−11.
(30) Kupče, E.; Freeman, R. *J. Magn. Reson.* **2005**, *172* (2), 329−332.

struct full 3D and 4D spectra by collecting only a small set of 2D spectra at various tilt angles.[24,25,28] Using the terminology of reduced dimensionality, these can generally be referred to as PR-(3,2)D and PR-(4,2)D data collections.

We have shown that it is possible to generalize Kupče and Freeman's approach for *n*D spaces and have demonstrated that the same concepts can be applied for reconstruction from 3D projections, projections of even higher dimensionality, and using projections of mixed dimensionality.[31] Additionally, our laboratories have described the link between projection−reconstruction NMR (PR−NMR) and G-matrix Fourier transform NMR (GFT).[31] In this previous study, we presented the Lower-Value (LV) reconstruction of a 5D HACACONH spectra obtained using a 1 mM $^{13}C/^{15}N$-labeled sample of the 56 residue protein G B1 domain from four orthogonal 2D planes ($H_\alpha-H_N$, $C_\alpha-H_N$, $C'-H_N$, $N-H_N$) collected in four separate experiments and eight additional tilted planes collected in a single experiment by coevolution of H_α, C_α, C', and N. The data for this 5D reconstruction were collected in 87 min.[31] The pulse sequence used for this study was modified from the original by incorporating the ability to coevolve each of the indirect dimensions (H_α, C_α, C', and N in this case) either independently or simultaneously at any desired tilt angle (α). The time increment required in each evolution dimension in order to achieve a desired tilt angle is determined by:

$$\Delta t_i = \cos \alpha_i / sw_{\text{tilt}} \quad (1)$$

where sw_{tilt} is the maximum spectral window of the tilted vector calculated from the desired angle (α_i) and the spectral windows (sw_i) in each of the indirect dimensions involved:

$$sw_{\text{tilt}} = \sum_{i=1}^{n-1} sw_i \cos \alpha_i \quad (2)$$

If the set of projections for a given experiment is sufficient to define a unique reconstruction, the LV algorithm will position all cross-peaks correctly, with no spurious signals. It has been well documented, however, when using the Lower-Value reconstruction, that the intensities of the reconstructed peaks are limited to those found on the weakest of the projections, and that the signal-to-noise ratio of the reconstructed spectrum does not benefit from the cumulative signal recorded across all of the projections.[27] Thus, the LV algorithm is poorly suited for those cases when the signal-to-noise ratio on the individual projections is low, as could often be the case when studying larger proteins.

To date, the projection−reconstruction technique has only been applied to small- and medium-sized proteins.[24,25,28,31] Here, we extend the PR−NMR methodology to the backbone assignment of larger proteins. We present a suite of (4,2)D pulse sequences optimized for use with projection−reconstruction techniques and larger $^1H/^{13}C/^{15}N$- and $^2H/^{13}C/^{15}N$-labeled proteins. Additionally, we propose an alternative reconstruction algorithm for overcoming the sensitivity limitations inherent in the Lower-Value reconstruction scheme and discuss the signal-to-noise ratio in the reconstructed spectrum. Finally, we provide a sampling theorem and discuss its implications in the choice of projections.

Materials and Methods

NMR Sample. A 0.94 mM sample of $^2H/^{13}C/^{15}N$-methyl protonated (Leu, Val, and Ile $\delta1$) labeled human carbonic anhydrase II (HCA II) has been used to test the efficacy of the TROSY-based HNCACB, HN(CO)CACB, Intra-HNCACB, HNCACO, HNCOCA, and $HNCO_{i-1}CA_i$ pulse sequences presented herein. HCA II is a 29 kDa (260 aa) zinc−metalloenzyme that catalyzes the interconversion of CO_2 and bicarbonate. The expression and purification methods used to produce the sample have been described elsewhere.[32] The sample is at pH 6.8 in 100 mM phosphate buffer containing 10% D_2O.

A 0.9 mM sample of $^1H/^{13}C/^{15}N$-labeled calbindin D_{28K} has been used to test the efficacy of the HACA(CO)NH and HACANH experiments. Calbindin D_{28K} is a 30 kDa (261 aa) calcium-binding protein of unknown tertiary structure. It is widely distributed in various tissues that are vitally dependent upon intracellular Ca^{2+} concentrations for their function, including the brain, peripheral nervous system, kidney, and intestine.[33,34] The expression and purification methods used to produce the sample have been presented elsewhere.[33] The calcium-loaded calbindin D_{28K} NMR sample is at pH 6.2 in 10 mM Tris buffer containing 1 mM dithiothreitol (DTT), 0.02% sodium azide, 6 mM $CaCl_2$, and 10% D_2O.

NMR Spectroscopy. NMR data were collected at 25 °C for HCA II and at 50 °C for calbindin D_{28K} on either an 800 or 600 MHz Varian Inova spectrometer equipped with triple-resonance Varian cold-probes.

Results and Discussion

NMR Experiments. We have converted a collection of commonly used protein backbone assignment sequences to the (4,2)D PR−NMR format (Figure 1 and Supporting Information Figures S1−S3). Except for the HACA(CO)NH and HACANH experiments, the sequences presented below are all TROSY-based[35] and contain a deuterium decoupling option and are, therefore, designed to be used on large perdeuterated proteins. The high resolution available from these 4D experiments makes them particularly useful for the assignment of high molecular weight proteins if adequate signal-to-noise ratios can be achieved in the individual projections. Only features different from the conventional sequences are discussed in detail below.

TROSY−HNCACB/HN(CO)CACB/Intra-HNCACB. The magnetization transfer steps for the (4,2)D PR−NMR TROSY-based HNCACB, HN(CO)CACB, and Intra-HNCACB experiments, respectively, are presented below.

$$\mathrm{HN}_{(i)} \xrightarrow{J_{\mathrm{HN}}} {}^{15}\mathrm{N}_{(i)} \xrightarrow{J_{\mathrm{NC}\alpha}} {}^{13}\mathrm{C}^{\alpha}_{(i,i-1)} \xrightarrow{J_{\mathrm{C}\alpha\mathrm{C}\beta}} {}^{13}\mathrm{C}^{\beta}_{(i,i-1)}[t_1] \rightarrow {}^{13}\mathrm{C}^{\alpha}_{(i,i-1)}[\mathrm{CT}-t_2] \xrightarrow{J_{\mathrm{NC}\alpha}} {}^{15}\mathrm{N}_{(i)}[\mathrm{CT}-t_3] \xrightarrow{J_{\mathrm{HN}}} \mathrm{HN}_{(i)}[t_4] \quad (3)$$

$$\mathrm{HN}_{(i)} \xrightarrow{J_{\mathrm{HN}}} {}^{15}\mathrm{N}_{(i)} \xrightarrow{J_{\mathrm{NC}'}} {}^{13}\mathrm{C}'_{(i-1)} \xrightarrow{J_{\mathrm{C}'\mathrm{C}\alpha}} {}^{13}\mathrm{C}^{\alpha}_{(i-1)} \xrightarrow{J_{\mathrm{C}\alpha\mathrm{C}\beta}} {}^{13}\mathrm{C}^{\beta}_{(i-1)}[t_1] \rightarrow {}^{13}\mathrm{C}^{\alpha}_{(i-1)}[\mathrm{CT}-t_2] \rightarrow {}^{13}\mathrm{C}'_{(i-1)} \xrightarrow{J_{\mathrm{NC}'}} {}^{15}\mathrm{N}_{(i)}[\mathrm{CT}-t_3] \xrightarrow{J_{\mathrm{HN}}} \mathrm{HN}_{(i)}[t_4] \quad (4)$$

$$\mathrm{HN}_{(i)} \xrightarrow{J_{\mathrm{HN}}} {}^{15}\mathrm{N}_{(i)} \xrightarrow{J_{\mathrm{NC}'},J_{\mathrm{NC}\alpha}} {}^{13}\mathrm{C}^{\alpha}_{(i)} \xrightarrow{J_{\mathrm{C}\alpha\mathrm{C}\beta}} {}^{13}\mathrm{C}^{\beta}_{(i)}[t_1] \rightarrow {}^{13}\mathrm{C}^{\alpha}_{(i)}[\mathrm{CT}-t_2] \xrightarrow{J_{\mathrm{NC}\alpha}} {}^{15}\mathrm{N}_{(i)}[\mathrm{CT}-t_3] \xrightarrow{J_{\mathrm{HN}}} \mathrm{HN}_{(i)}[t_4] \quad (5)$$

In the corresponding conventional experiments,[36,37] the glycine resonances are of opposite sign compared to signals from

(31) Coggins, B. E.; Venters, R. A.; Zhou, P. *J. Am. Chem. Soc.* **2004**, *126* (4), 1000−1001.
(32) Venters, R. A.; Huang, C. C.; Farmer, B. T., II; Trolard, R.; Spicer, L. D.; Fierke, C. A. *J. Biomol. NMR* **1995**, *5* (4), 339−344.

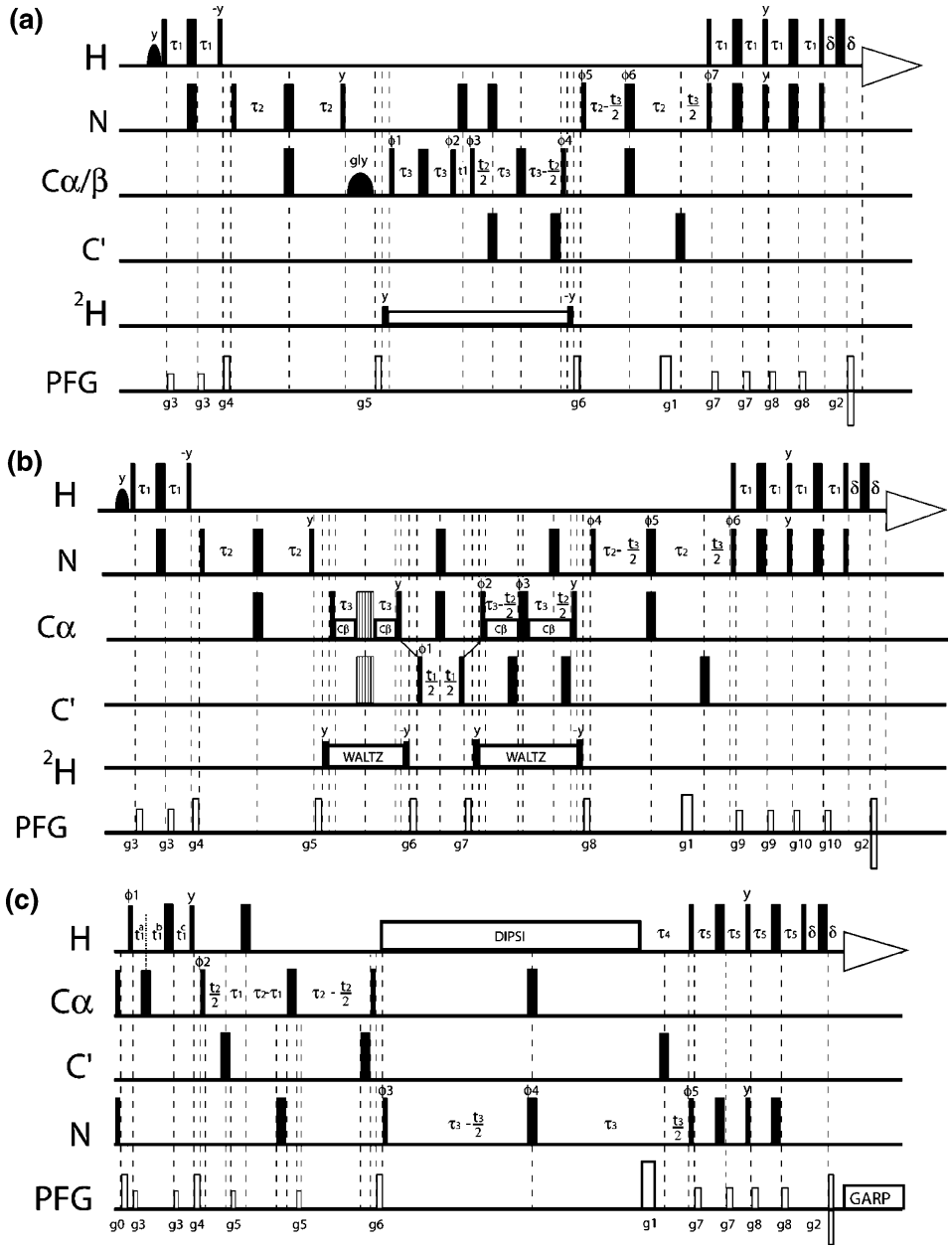

Figure 1. (4,2)D PR−NMR (a) TROSY−HNCACB, (b) TROSY−HNCACO, and (c) HACANH pulse sequences for the assignment of larger proteins. All pulses are applied along the *x*-axis unless otherwise indicated. (a) **HNCACB**: Selective inversion of glycine C_α nuclei has been included in the sequences in order to ensure that the signs of these resonances are the same as those of the remainder of the residues. The selective glycine inversion is accomplished with an on resonance 4.63 ms IBURP2[38] pulse ($B1_{max}$ of 1077 Hz) on our Inova 800 spectrometer. The carrier frequencies are 4.75 ppm in ^{1}H, 45.5 ppm in ^{13}C, and 119.4 ppm in ^{15}N. ^{2}H decoupling is achieved with a 2000 Hz WALTZ16[39] field. The water selective pulse is accomplished with a 1.14 ms sinc pulse. The delays are $\tau1 = 2.4$ ms, $\tau2 = 12.5$ ms, $\tau3 = 7.1$ ms, $\delta = 250$ μs. The phase cycle is $\phi1 = x, -x$; $\phi2 = y$; $\phi3 = 4(y), 4(-y)$; $\phi4 = 2(x), 2(-x)$; $\phi5 = y$; $\phi6 = x$; $\phi7 = x$; rec $= 2(x, -x, -x, x)$. Quadrature detection in C_β and C_α is achieved using States-TPPI[40] phase cycling of $\phi1$ and $\phi4$, respectively. Quadrature detection in N is accomplished using the sensitivity-enhanced pulse field gradient technique and $\phi7$.[41] The gradients are $g1 = 2.5$ ms at 26.5 G/cm; $g2 = 0.25$ ms at 26.5 G/cm; $g3 = 0.5$ ms at 14.3 G/cm; $g4 = 1.0$ ms at 22.5 G/cm; $g5 = 1.0$ ms at 34.7 G/cm; $g6 = 1.0$ ms at 29.2 G/cm; $g7 = 0.5$ ms at 8.2 G/cm; $g8 = 0.5$ ms at 15.3 G/cm. (b) **HNCACO**: Selective C_β decoupling has been added to the sequence during those periods when C_α magnetization is in the transverse plane. This is accomplished with a three-band WURST2 decoupling scheme.[42] The three bands cover the ranges of 69.5−65.5 ppm for threonines, 17.5−13.5 ppm for alanines, and 40−20 ppm for all other C_β nuclei. The carrier frequencies are 4.75 ppm in ^{1}H, 55.2 ppm in $^{13}C_\alpha$, 173.1 ppm in $^{13}C'$, and 119.4 ppm in ^{15}N. The water selective pulse is accomplished with a 1.14 ms sinc pulse. ^{2}H decoupling is achieved with a 573 Hz GARP1[43] field. The delays are $\tau1 = 2.2$ ms, $\tau2 = 13.0$ ms, $\tau3 = 4.5$ ms, $\delta = 250$ μs. The phase cycle is $\phi1 = 2(x), 2(-x)$; $\phi2 = x$; $\phi3 = x, y, -x, y$; $\phi4 = y$; $\phi5 = 2(x), 2(-x)$; $\phi6 = x$; rec $= x, -x, -x, x$. Quadrature detection in C′ and C_α is achieved using States-TPPI[40] phase cycling of $\phi1$ and $\phi2$, respectively. Quadrature detection in N is accomplished using the sensitivity-enhanced pulse field gradient technique and $\phi6$.[41] The gradients are $g1 = 1.25$ ms at 26.5 G/cm; $g2 = 0.125$ ms at 25.3 G/cm; $g3 = 0.4$ ms at 10.2 G/cm; $g4 = 0.8$ ms at 15.3 G/cm; $g5 = 0.9$ ms at 10.2 G/cm; $g6 = 0.6$ ms at 13.3 G/cm; $g7 = 0.5$ ms at 12.3 G/cm; $g8 = 0.4$ ms at 16.3 G/cm; $g9 = 0.4$ ms at 6.3 G/cm; $g10 = 0.4$ ms at 10.9 G/cm. (c) **HACANH**: The carrier frequencies are 5.17 ppm in ^{1}H, 57.9 ppm in $^{13}C_\alpha$, and 118.6 ppm in ^{15}N. ^{1}H decoupling is achieved with a 7500 Hz DIPSI2[44] scheme. The delays are $\tau1 = 1.7$ ms, $\tau2 = 12.7$ ms, $\tau3 = 11.5$ ms, $\tau4 = 5.4$ ms, $\tau5 = 2.4$ ms, $\delta = 250$ μs. The phase cycle is $\phi1 = x$; $\phi2 = x, -x$; $\phi3 = x$; $\phi4 = 2(x), 2(y)$; $\phi5 = x$; rec $= x, -x, -x, x$. Quadrature detection in H_α and C_α is achieved using States-TPPI[40] phase cycling of $\phi1$ and $\phi2$, respectively. Quadrature detection in N is accomplished using the sensitivity-enhanced pulse field gradient technique and $\phi5$.[41] The gradients are $g0 = 0.5$ ms at 16.3 G/cm; $g1 = 2.5$ ms at 26.5 G/cm; $g2 = 0.25$ ms at 26.5 G/cm; $g3 = 0.5$ ms at 19.0 G/cm; $g4 = 0.5$ ms at −26.5 G/cm; $g5 = 0.5$ ms at 21.8 G/cm; $g6 = 0.5$ ms at −30.6 G/cm; $g7 = 0.5$ ms at 4.1 G/cm; $g8 = 0.5$ ms at 6.1 G/cm.

all other residue types. This causes problems for some of the reconstruction methodologies, particularly the Lower-Value approach. If the glycine resonances are negative in any of the projected spectra, overlap of the glycine resonances with other positive resonances would lead to severe attenuation or even complete cancellation of both resonances, resulting in the disappearance of both signals in the reconstructed spectrum. To overcome this problem, we have inserted a selective glycine inversion pulse (Figure 1a) so that the signs of glycine resonances are the same as the remainder of the residues. This selective glycine inversion is accomplished with an on resonance 4.63 ms IBURP2[38] pulse ($B1_{max}$ of 1077 Hz) on an Inova 800 spectrometer. The spin state at the point of application of this glycine selective pulse is $H^zN^zC_\alpha{}^z$. We have determined that the lifetime of this particular spin state is 663 ms for HCA II on an 800 MHz instrument; therefore, this pulse does not adversely affect the sensitivity of these experiments. Experimentally, no difference in the signal-to-noise ratio could be measured for the (4,2)D PR−NMR HNCACB experiment, with and without the application of the glycine selective pulse using HCA II at 800 MHz.

TROSY−HNCACO/HNCOCA/HNCO$_{i-1}$CA$_i$. These 4D experiments have been adapted for PR−NMR from the sequences originally proposed by Yang and Kay[45] and Konrat et al.[46] with few modifications. Selective C_β decoupling is accomplished with a three-band WURST2 decoupling scheme (Figure 1b).[42] The three bands cover the ranges of 69.5−65.5 ppm for threonines, 17.5−13.5 ppm for alanines, and 40−20 ppm for all other C_β nuclei. The C_β nuclei for serine residues are not decoupled by this scheme since they overlap significantly with the C_α chemical shift range. The magnetization transfer steps for the (4,2)D PR−NMR TROSY-based HNCACO, HNCOCA, and the HNCO$_{i-1}$CA$_i$ experiments, respectively, are presented below.

$$\mathrm{HN}_{(i)} \xrightarrow{J_{\mathrm{HN}}} {}^{15}\mathrm{N}_{(i)} \xrightarrow{J_{\mathrm{NC}\alpha}} {}^{13}\mathrm{C}^{\alpha}{}_{(i-1,\,i)} \xrightarrow{J_{\mathrm{C'C}\alpha}} {}^{13}\mathrm{C'}_{(i-1,\,i)}[t_1] \rightarrow {}^{13}\mathrm{C}^{\alpha}{}_{(i-1,\,i)}[\mathrm{CT}-t_2] \xrightarrow{J_{\mathrm{NC}\alpha}} {}^{15}\mathrm{N}_{(i)}[\mathrm{CT}-t_3] \xrightarrow{J_{\mathrm{HN}}} \mathrm{HN}_{(i)}[t_4] \quad (6)$$

$$\mathrm{HN}_{(i)} \xrightarrow{J_{\mathrm{HN}}} {}^{15}\mathrm{N}_{(i)} \xrightarrow{J_{\mathrm{NC'}}} {}^{13}\mathrm{C'}_{(i-1)} \xrightarrow{J_{\mathrm{C'C}\alpha}} {}^{13}\mathrm{C}^{\alpha}{}_{(i-1)}[t_1] \rightarrow {}^{13}\mathrm{C'}_{(i-1)}[\mathrm{CT}-t_2] \xrightarrow{J_{\mathrm{NC'}}} {}^{15}\mathrm{N}_{(i)}[\mathrm{CT}-t_3] \xrightarrow{J_{\mathrm{HN}}} \mathrm{HN}_{(i)}[t_4] \quad (7)$$

$$\mathrm{HN}_{(i)} \xrightarrow{J_{\mathrm{HN}}} {}^{15}\mathrm{N}_{(i)} \xrightarrow{J_{\mathrm{NC'}},J_{\mathrm{NC}\alpha}} {}^{13}\mathrm{C'}_{(i-1)}[t_1]{}^{13}\mathrm{C}^{\alpha}{}_{(i)}[t_2] \rightarrow {}^{15}\mathrm{N}_{(i)}[\mathrm{CT}-t_3] \xrightarrow{J_{\mathrm{NC'}},J_{\mathrm{NC}\alpha}} \xrightarrow{J_{\mathrm{HN}}} \mathrm{HN}_{(i)}[t_4] \quad (8)$$

The TROSY−HNCOCA experiment can be collected with either constant time or nonconstant time evolution in the C_α dimension. We chose to collect the data using the nonconstant time version of the experiment in order to maximize sensitivity and to utilize fully the resolution advantage offered by the PR−NMR technique, which cannot be achieved by conventional data acquisition methodology within a short acquisition period.

HACANH/HACA(CO)NH. The 4D HACANH and HACA(CO)NH[47−51] experiments are designed to assign the backbone nuclei of fully protonated $^{13}C/^{15}N$-labeled proteins up to ~30 kDa in molecular weight. The magnetization transfer steps in the (4,2)D PR−NMR HACANH and HACA(CO)NH experiments, respectively, are presented below.

$$\mathrm{H}^{a}{}_{(i,\,i-1)}[t_1] \xrightarrow{J_{\mathrm{HC}\alpha}} {}^{13}\mathrm{C}^{\alpha}{}_{(i,\,i-1)}[t_2] \xrightarrow{J_{\mathrm{NC}\alpha}} {}^{15}\mathrm{N}_{(i)}[\mathrm{CT}-t_3] \xrightarrow{J_{\mathrm{HN}}} \mathrm{HN}_{(i)}[t_4] \quad (9)$$

$$\mathrm{H}^{a}{}_{(i-1)}[t_1] \xrightarrow{J_{\mathrm{HC}\alpha}} {}^{13}\mathrm{C}^{\alpha}{}_{(i-1)}[t_2] \xrightarrow{J_{\mathrm{C}\alpha\mathrm{C'}}} {}^{13}\mathrm{C'}_{(i-1)} \xrightarrow{J_{\mathrm{NC'}}} {}^{15}\mathrm{N}_{(i)}[\mathrm{CT}-t_3] \xrightarrow{J_{\mathrm{HN}}} \mathrm{HN}_{(i)}[t_4] \quad (10)$$

Originally, we reported a 5D version of the HACACONH experiment.[31] Here, we have chosen to use a 4D version of this experiment, making it more compatible with its companion sequence, the 4D HACANH (Figure 1c). If desirable, selective C_β decoupling can be used in the HACA(CO)NH experiment whenever C_α magnetization is in the transverse plane. This is accomplished with a three-band WURST2[42] decoupling scheme as described above. The use of C_β decoupling is not recommended in the HACANH experiment as resonances not covered by C_β decoupling are of the opposite sign to those covered by the C_β decoupling.

PR−NMR. All of the pulse sequences described above share common features unique to PR−NMR.

(1) Control of Projections. We have included in the pulse sequences the ability to control the increments of all indirectly detected evolution times (t_1, t_2, t_3) at any user-specified ratio by specifying the desired projection angles for (n − 1) out of the (n) indirectly detected evolution dimensions (t_1 and t_2 in the case of (4,2)D PR−NMR experiments). The angle of projection (α) for the third indirectly detected dimension (t_3) is determined by the relationship:

(33) Venters, R. A.; Benson, L. M.; Craig, T. A.; Bagu, J.; Paul, K. H.; Kordys, D. R.; Thompson, R.; Naylor, S.; Kumar, R.; Cavanagh, J. *Anal. Biochem.* **2003**, *317* (1), 59−66.
(34) Lutz, W.; Frank, E. M.; Craig, T. A.; Thompson, R.; Venters, R. A.; Kojetin, D.; Cavanagh, J.; Kumar, R. *Biochem. Biophys. Res. Commun.* **2003**, *303* (4), 1186−1192.
(35) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (23), 12366−12371.
(36) Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 11655−1666.
(37) Nietlispach, D.; Ito, Y.; Laue, E. D. *J. Am. Chem. Soc.* **2002**, *124* (37), 11199−11207.
(38) Geen, H.; Freeman, R. *J. Magn. Reson.* **1991**, *93* (1), 93−141.
(39) Shaka, A. J.; Keeler, J.; Frenkiel, T.; Freeman, R. *J. Magn. Reson.* **1983**, *52* (2), 335−338.
(40) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1989**, *85* (2), 393−399.
(41) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114* (26), 10663−10665.
(42) Matsuo, H.; Kupče, E.; Li, H.; Wagner, G. *J. Magn. Reson. B* **1996**, *113* (1), 91−96.
(43) Shaka, A. J.; Barker, P. B.; Freeman, R. *J. Magn. Reson.* **1985**, *64* (3), 547−552.
(44) Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *77* (2), 274−293.
(45) Yang, D.; Kay, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 2571−2575.
(46) Konrat, R.; Yang, D.; Kay, L. E. *J. Biomol. NMR* **1999**, *15*, 309−313.
(47) Feng, W.; Rios, C. B.; Montelione, G. T. *J. Biomol. NMR* **1996**, *8* (1), 98−104.
(48) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1990**, *89*, 496−514.
(49) Kay, L. E.; Ikura, M.; Bax, A. *J. Magn. Reson.* **1991**, *91*, 84−92.
(50) Palmer, A. G., III; Fairbrother, W. J.; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Biomol. NMR* **1992**, *2* (1), 103−108.
(51) Powers, R.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *J. Magn. Reson.* **1991**, *94*, 209−213.

$$\omega_{\text{tilt}} = \sum_{i=1}^{n-1} \omega_i \cos \alpha_i, \quad \text{with} \sum_{i=1}^{n-1} \cos^2 \alpha_i = 1 \qquad (11)$$

For instance, in the (4,2)D PR−NMR TROSY−HNCACB experiment, setting the angles for C_β to 0° and C_α to 90° will also set the angle for N to 90° and will result in the acquisition of the $C_\beta-H_N$ 2D plane (often referred to as an orthogonal plane). Setting angles for C_β to 15.5° and C_α to 75.0° would result in the angle for N being calculated and set to 86.0° (Table 1) and the acquisition of four tilted projections. The spectral window in the indirectly detected dimension for tilted projections is given by eq 2.

(2) Evolution Delay and Phase. Since all indirect dimensions are evolved simultaneously in the projection experiments, the post-Fourier transformation phase correction for each of these coevolving dimensions should be identical. We have thus taken extra care to ensure that there is no need for zero- or first-order phase correction in any of the indirectly detected dimensions. A problem unique to the projection experiment is the treatment of very small delays in any of the indirect dimensions (the case when the projection angle approaches 90°). The three conditions for delay adjustment in conventional experiments (zero delay, a delay larger than $2 \times pwX/\pi$, or a delay larger than decoupling pulses in addition to $2 \times pwX/\pi$) are not adequate for PR−NMR experiments and can lead to severe baseline distortion when the tilt angle approaches 90° in any dimension. This can be remedied by adding evolution delays before the 180° refocusing pulse such that the very small delay can be accommodated.

(3) Number of Scans. All of the reconstruction algorithms assume the same or similar signal-to-noise ratios in all of the projected spectra. This can be achieved by setting the appropriate number of scans for tilted versus orthogonal projections. For a sensitivity-enhanced experiment collected in (4,2)D PR−NMR mode, a ratio of 2:4:8 should be used for the orthogonal projections with sensitivity enhancement (nitrogen dimension): orthogonal projections without sensitivity enhancement (the other two indirect dimensions):generic tilted projections in all three dimensions, respectively.

Data Acquisition and Processing. We have collected the (4,2)D TROSY-based HNCACB, HN(CO)CACB, Intra-HNCACB, HNCACO, HNCOCA, and $HNCO_{i-1}CA_i$ PR−NMR data using a 0.94 mM sample of $^2H/^{13}C/^{15}N$-methyl protonated HCA II and the HACANH and HACA(CO)NH data using a 0.9 mM sample of $^1H/^{13}C/^{15}N$-labeled D_{28K} following the same general strategy, with the *a priori* notion that the data would be reconstructed using the new HBLV algorithm presented below. Eight data sets were collected for each of the (4,2)D experiments: the three orthogonal planes (HN−N, HN−CA, and HN−CB for the TROSY−HNCACB, for example) and five tilt angles (Table 1). On the NMR instruments used in this study, the acquisition was fully automated with a simple macro (Supporting Information). In theory, it is possible to use the orthogonal planes collected for one reconstruction in the reconstruction of other data sets, providing additional savings in spectrometer time. However, in this work, all of the data sets for each reconstruction were collected independently.

As an example, Table 1 presents the data acquisition parameters for the (4,2)D TROSY−HNCACB experiment. Data for the other experiments were collected using similar parameters. Tables for these other data sets are given as Supporting Information. It should be noted that 64 increments were collected for each of the projections, translating directly into a higher resolution than can reasonably be obtained using the conventional approach to the collection of 4D data. A total of eight 4D data sets were reconstructed, each at a $64 \times 64 \times 128 \times 738$ resolution, using 12.8 days of data collection time, the same amount of time required to collect just one or, at the most, two standard 4D data sets at significantly lower overall resolution.

Table 1. Data Acquisition Parameters for the (4,2)D TROSY−HNCACB Experiment[a]

projection	*ni*	phase C	phase N	angle N	angle Ca	angle Cb	nt	SW_{tilt}	time (min)
HN−N	64	1	1,2	0.0	90.0	90.0	8	2900	24
HN−CA	64	1,3	1	90.0	0.0	90.0	16	11400	48
HN−CB	64	1,2	1	90.0	90.0	0.0	16	12400	48
tilt 1	64	1,2,3,4	1,2	86.0	15.5	75.0	32	15092	376
tilt 2	64	1,2,3,4	1,2	73.9	33.7	61.3	32	16598	376
tilt 3	64	1,2,3,4	1,2	54.7	54.7	54.7	32	15415	376
tilt 4	64	1,2,3,4	1,2	33.7	73.9	61.3	32	11328	376
tilt 5	64	1,2,3,4	1,2	15.5	86.0	75.0	32	6595	376

[a] Total time = 33.3 h.

Following data acquisition, all 2D projections were processed with NMRPipe software[52] using one simple conversion and one simple processing script (Supporting Information). Pure constant-time evolution domains of all spectra were doubled using mirror image linear prediction,[53] while pure nonconstant time evolution domains were doubled using forward−backward linear prediction.[54] When a projection was acquired involving the coevolution of a constant time and a nonconstant time dimension, the time domains were doubled using forward−backward linear prediction. Cosine or cosine squared window functions were used in the processing of all dimensions after linear prediction.

Reconstruction. (1) Background. The Lower-Value algorithm introduced by Kupče and Freeman is distinct from the many reconstruction algorithms widely used in the tomography community and is best suited for data containing few, well dispersed signals.[27] This algorithm determines the value for a point in a reconstruction by finding where that point projects onto each of the projections and keeping the lowest value from those found on the individual projections (Figure 2a). If the point in the reconstruction is called *S*, and its points of projection are called P_θ for projections of varying angle θ, then the LV reconstruction is described as:

$$S = \min_\theta P_\theta \qquad (12)$$

The LV approach works best for sparse data with a relatively high signal-to-noise ratio. However, this method suffers a serious impediment, which becomes apparent when one attempts to apply PR−NMR to experiments with low sensitivity or to samples with limited concentration. Because any given point in the final reconstructed spectrum generated by the LV approach retains the lowest value found when comparing all of the points of projections (P_θ), a cross-peak's height in the final spectrum is always limited by the height found on a single and, more importantly, the weakest projection. Thus, if a cross-peak

(52) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6* (3), 277−293.
(53) Zhu, G.; Bax, A. *J. Magn. Reson.* **1990**, *90* (2), 405−410.
(54) Zhu, G.; Bax, A. *J. Magn. Reson.* **1992**, *100* (1), 202−207.

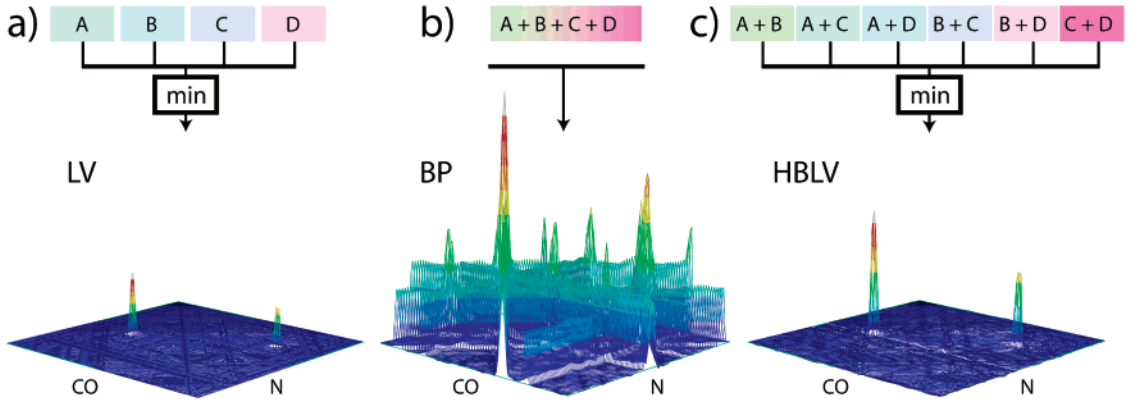

Figure 2. (a) Lower-Value (LV) reconstruction. The value of any given point in the final reconstructed matrix is determined by retaining the smallest of the values found among the points of projection (P_θ), resulting in no signal accumulation. (b) Backprojection (BP) reconstruction. Projection values (P_θ) are summed to give the reconstruction intensity. The signal accumulation is evident by the increased peak intensity; the drawback is the generation of ridges well above the noise level across the entire reconstructed spectrum. (c) The Hybrid Backprojection/Lower-Value algorithm (HBLV), demonstrated for the reconstruction from four projections, A−D, with a bin size k of 2. Backprojection reconstructions are computed for each possible combination of two projections, and these are then compared by the Lower-Value approach. HBLV reconstruction achieves signal accumulation for a bin size of k (compared to LV reconstruction) and also cleanly removes all of the reconstruction ridges in the BP reconstruction, generating a perfect baseline. These reconstructed CO/N planes were extracted from (3,2)D HNCO data collected on a 1 mM $^{13}C/^{15}N$-labeled sample of protein G B1 domain.

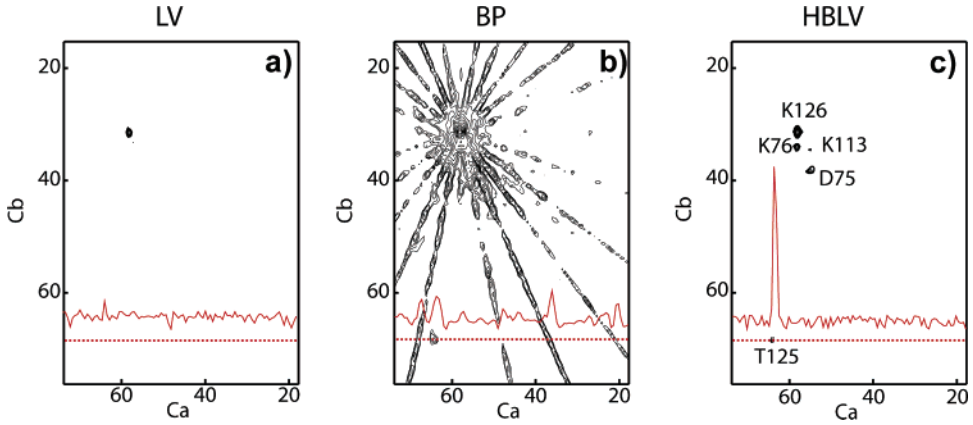

Figure 3. (a) LV, (b) BP, and (c) HBLV reconstructions of HCA II (4,2)D HNCACB data. The data were collected on a 0.94 mM sample of $^{2}H/^{13}C/^{15}N$-methyl protonated HCA II using the parameters given in Table 1. A bin size of eight ($k = 8$) was used in the HBLV reconstruction algorithm (see below). The C_α/C_β planes presented were extracted at the $^{1}H_N/^{15}N$ chemical shifts of residue K126. There is an overlap of the $^{1}H_N/^{15}N$ chemical shifts of K126 with those of K76 and, to a lesser degree, K113. The LV reconstruction only retains the strong intra-residue C_α/C_β correlation for residue K126. The BP reconstruction retains both the intra- and the inter-residue (T125) C_α/C_β correlations; however, the inter-residue correlation is the same intensity as the backprojection ridges. In addition, all of the cross-peaks near the intra-residue C_α/C_β correlation for K126 have broadened and are obscured by the backprojection ridges. The HBLV reconstruction provides the intra- and inter-residue C_α/C_β correlations for both K126 and K76. The cross-section through the T125 inter-residue correlation clearly shows the advantage of the HBLV algorithm.

does not appear above the noise on just one of the projections, that cross-peak will be lost. It is apparent that, in at least some cases, it would be desirable to have a reconstruction approach which could accumulate intensity across projections.

Recently, Kupče and Freeman proposed a Backprojection (BP) method as a solution to sensitivity problems associated with PR−NMR, demonstrating it with the reconstruction of a 3D HNCO spectrum from 18 projections.[29] Like LV, BP is also a simple procedure, and can be expressed using the same notation as above:

$$S = \sum_{\theta} P_\theta \tag{13}$$

BP involves the simple addition of contributions from all of the projections (Figure 2b). This method, therefore, enjoys the benefit of signal accumulation and retains as much signal as would be available from a conventional spectrum recorded for the same length of time as that used to collect the entire set of projections. Backprojection has its roots in the theory of projection−reconstruction; specifically, it is one component in the analytical equation for inverting the act of projection.[55−57]

Unfortunately, while addressing the issue of sensitivity and of missing peaks, backprojection can produce significant artifacts which can interfere with the proper interpretation of a spectrum. These problems quickly become apparent when one calculates PR−NMR reconstructions with a limited set of projections. Figure 2b depicts one such case for the (3,2)D HNCO BP reconstruction of data collected on a 1 mM $^{13}C/^{15}N$-labeled sample of protein G B1 domain. Likewise, Figure 3b presents an even more severe case for a plane extracted from the (4,2)D HNCACB BP reconstruction of data collected on a 0.94 mM sample of $^{2}H/^{13}C/^{15}N$-methyl protonated HCA II. BP generates

(55) Kak, A. C.; Slaney, M. *Principles of Computerized Tomographic Imaging*; IEEE Press: New York, 1999.
(56) Herman, G. T. *Image Reconstruction from Projections*; Springer-Verlag: Berlin, 1979.
(57) Radon, J. *Ber. Sächsischen Ges. Wiss. Math.-Phys. Kl.* **1917**, *69*, 262−277.

characteristic backprojection ridges, and when those ridges intersect, they reinforce each other, sometimes producing false cross-peaks. If real cross-peaks in a spectrum have significantly lower intensities than other cross-peaks in the same data set, as often occurs for the inter-residue correlations in the HNCACB experiment, the false cross-peaks formed from BP ridges can be taller than some of the real cross-peaks (Figure 3b). One solution to this problem is to collect more projections. However, with increasing numbers of projections, one sees another type of BP artifact, a broadening of cross-peaks. These two effects become particularly pernicious when several cross-peaks are close together; the cross-peaks are broadened and thus begin to overlap, and their backprojection ridges intersect to add false cross-peaks, eventually leading to an intensity cluster (Figure 3b).

Here, we present a *Hybrid Backprojection/Lower-Value* (HBLV) algorithm, which accumulates signal intensity without producing artifacts. We also describe a theory for the number of projections needed to obtain an accurate reconstruction using the LV and HBLV algorithms. The HBLV algorithm is particularly useful for the accurate reconstruction of sensitivity-limited data.

(2) Theory. To obtain the proper reconstruction of weak cross-peaks with signal accumulation and without the introduction of artifacts, we propose combining the BP and the LV algorithms to give a hybrid BP/LV algorithm. The concept for the HBLV algorithm arises from the recognition that all of the artifacts seen in BP reconstructions are in the form of a point in reconstruction space having a higher intensity than it ought to. These artificially higher intensities can manifest themselves in several ways: as intensity surrounding cross-peaks and broadening them, as false cross-peaks, or as ridges. The challenge for developing a correction to the BP method is to remove these areas of artificially increased intensity, while retaining intensity at the center of real cross-peaks. The key to solving this problem comes from the realization that a true cross-peak arises when *all* of the available projections intersect and reinforce. In contrast, BP artifacts are caused either by a single projection or by the intersection of a subset, but not all, of the projections. The HBLV algorithm solves this problem by using a two-step reconstruction process, namely, a BP step followed by an LV step (Figure 2c). In the first step, the n total projections are grouped into bins of size k to produce ${}_nC_k$ combinations:

$$ {}_nC_k = \frac{n!}{k!(n-k)!} \tag{14} $$

For example, with four projections (A, B, C, and D) and bins of size 2, there are ${}_4C_2 = 6$ different possible bins: AB, AC, AD, BC, BD, and CD (Figure 2c). For each bin, the BP reconstruction is computed. In this example, six BP reconstructions will be computed, one for each combination. All BP reconstructions are then compared by LV, and the lowest of the reconstructed intensities is retained. The whole process can be described mathematically as:

$$ S = \min_{i=1}^{{}_nC_k}\left[\sum_{P_\theta \in A_i} P_\theta\right] \tag{15} $$

where A_i denotes a projection set containing k projections chosen from n total projections.

HBLV offers true signal accumulation over k projection spectra (Figure 3). Thus, increasing the bin size, k, offers better signal accumulation so that the weaker cross-peaks emerge from the noise. This provides a lower constraint on the value of k, in that k must be large enough so that weak cross-peaks emerge from the noise upon backprojection. Although the LV step significantly reduces the cross-peak broadening associated with the BP step, some broadening will still occur as k increases. This provides an upper constraint on the value of k, as k must be small enough that broadening does not become excessive. Increasing k could be thought of as shifting the result from an LV-dominated character to a BP-dominated character, with a resulting shift in the quality of the reconstruction. Naturally, HBLV with $k = 1$ is the same as an LV reconstruction, and HBLV with k set to the full number of projections is the same as a BP reconstruction.

(3) Apparent Noise and Signal-to-Noise Ratio. With any reconstructed spectrum, the apparent noise (the "background" of the reconstructed spectrum at points other than true cross-peaks) will depend on the reconstruction algorithm. The apparent noise in the final reconstructed spectrum includes both the consequences of the spectral noise found on the projections, as well as any "artifacts" (such as backprojection ridges) produced by the algorithm as it reconstructs the signal.

In LV reconstructions, there is no net gain in signal intensity over that found on individual projections; however, the apparent noise in the final reconstructed spectrum is reduced significantly from the levels measured on individual projections (Figures 2a and 3a). This is caused by the LV process of noise comparison, which retains, for a noise point, the weakest value found in the projections. This decrease in noise has long been observed in the symmetrization of homonuclear spectra by Lower-Value approaches. The noise following an LV reconstruction can be estimated by assuming that the noise on all projections follows a uniform distribution between $-a$ and a, centered on zero. Lower-Value treatment of n projections would thus generate noise statistically distributed between $-a/n$ and a/n, leading to an n-fold gain in the signal-to-noise ratio (SNR). It is important to recognize that this n-fold SNR gain with LV is only an apparent gain and does not reflect signal accumulation because only signals which already stand out above the noise in every projection will experience this gain.

In the BP case, because each data point in the reconstructed spectrum is determined by the addition (backprojection) of the projected data, there is a true signal accumulation benefit (Figures 2b and 3b). Although it is easy to conclude that the rate of signal accumulation is linear, while the noise accumulation would be proportional to the square root of the number of projections, giving a net SNR gain of $n^{1/2}$ over the value found on any single projection,[29] this argument would ignore backprojection ridges, which can easily dominate and overshadow the noise in a reconstruction. With many projections, the background noise is formed entirely by these backprojection ridges, thus leading to a smaller apparent signal-to-noise ratio than would be predicted by theory.

For the HBLV approach, as described above, the apparent SNR is determined by the combined effects of the BP and LV characters of the algorithm (Figures 2c and 3c). Thus, for a bin size of k, the BP step provides a true k-fold gain in signal intensity in each bin over the signal intensity found on the

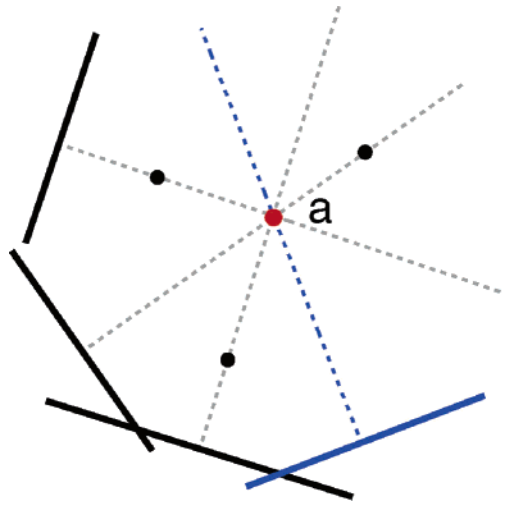

Figure 4. Analysis of the number of projections needed for a 2D plane reconstructed from 1D vectors by the LV method. In the worst possible case for three true peaks (black points) and three projections (black, heavy lines), the projections and peak positions would be such that three of the projection lines (dotted line) for the three peaks would intersect at point a (red). This point would then contain a false peak. Adding a single additional projection, for example, the one shown in blue, will eliminate the false peak at point a.

individual projections, including peaks buried in the noise, as well as a $k^{1/2}$-fold gain in noise level (ignoring backprojection ridges), resulting in a net $k^{1/2}$-fold increase in SNR in each bin over a single projection. The LV step then operates on ${}_nC_k$ bins, giving a $1/{}_nC_k$-fold reduction in the apparent noise provided that the noise is distributed randomly. However, since the noise points are computed from ${}_nC_k$ combinations, our experience indicates that a small number of noise points deviate from this noise estimation, producing isolated single point spikes on a largely flattened background. Recognizing this fact, we propose a modification to the HBLV algorithm that makes use of a conservative reconstruction noise level estimated as $k^{1/2}N$, where N is the noise level on the individual projections. If any of the ${}_nC_k$ bins for a reconstruction point has intensity at or below this level, that point can be identified as containing only noise, and it is not necessary to consider further bins. By stopping the calculation at this point, the apparent noise variation more closely reflects the true noise level, and a significant savings in computation time is realized.

Compared to conventional 4D NMR, the suite of (4,2)D PR−NMR experiments offers superior resolution, with significant saving in instrument time. Taking the (4,2)D HNCACB data collected on the HCA II sample (Table 1) as an example case, to obtain the same resolution as that found in the HBLV reconstruction, the conventional data would have to be collected with a full 64 increments in each of the indirectly detected dimensions. Presuming that eight transients are collected for each FID (as is the case for the ${}^1H_N-{}^{15}N$ orthogonal projection), this 4D HNCACB data set would take 273 days to collect. However, when it comes to sensitivity, it is more realistic to compare the conventional 4D data collected with the same acquisition time as the (4,2)D PR−NMR data, albeit at a much lower digital resolution. Under these conditions, since the HBLV reconstruction algorithm achieves true signal accumulation for the duration of k projections, the SNR of the final 4D HBLV reconstructed spectrum would be approximately $k^{1/2}/n^{1/2}$ of the final spectrum collected by the conventional approach. For a bin size (k) of eight and a total number of projections (n) of 23, this would translate into a SNR of 0.59 for the HBLV reconstruction as compared to the conventional data.

(4) Number of Projections. One of the most important considerations in the setup of a PR−NMR experiment is deciding how many projections to collect and at what projection angles. Here, we provide a theoretical analysis for the number of projections needed to ensure an artifact-free reconstruction using the LV and HBLV algorithms. Consider first the reconstruction of a single 2D plane from 1D vectors, with m peaks on the plane, ignoring peak overlap and line shape issues, and assuming, for the moment, that peaks are infinitely sharp. For a cross-peak to occur in an LV reconstruction, the projected value on *every* projection must be above the noise level. For a point of interest (Figure 4, point a), one could draw m projection lines, corresponding to m projected 1D vectors at specific angles, each passing through a real cross-peak and the point of interest (Figure 4). Under this situation, a false peak would occur at point a. It should be clear that it is not possible to select any other projection angle such that an *additional* projection line through the point of interest will also pass through a real cross-peak; therefore, adding one additional projection is guaranteed to remove the ambiguity and eliminate the false cross-peak at point a. Thus, we can write that if the condition

$$n \geq m + 1 \tag{16}$$

is met, under the assumptions given above, the LV reconstruction will not contain false peaks.

The same analysis can be applied to the HBLV algorithm. Consider the case of three peaks and three projections as shown in Figure 4. If the bin size k is set to two, there will be three (${}_3C_2$) bins. For any single bin, the backprojection ridges for two of the cross-peaks will pass through the point a, and thus all three bins will contain a false cross-peak. An LV comparison of these three bins will then retain a false cross-peak at point a. To resolve the ambiguity, one must include k (in this case, two) additional projections. There will then be one bin containing only the new projections, and hence, eliminating the false cross-peak at point a. The requirement for projection data with HBLV is then

$$n \geq m + k \tag{17}$$

This analysis can be extended to a d-dimensional space containing m peaks. For any point in that space, there exists a maximum of m lines connecting this point and the m peaks, with each line perpendicular to $d - 1$ projection vectors provided that none of these vectors can be represented as a linear combination of the remaining $d - 2$ vectors. Thus, there are a maximum of $m(d - 1)$ projections to produce an artifact at this point, and any additional projection would remove the artifact using the LV algorithm. Analogously, for a bin size of k, k additional projections are needed. Reconstructing a d-D space from p-D projections is equivalent to reconstructing a $(d - p + 1)$-D space from 1D vectors. Thus, one needs at least n projections, where n is given by

$$n \geq m(d - p) + k \tag{18}$$

In practice, peaks have finite width, and overlap can occur, which would provide more opportunities for ambiguity. At the

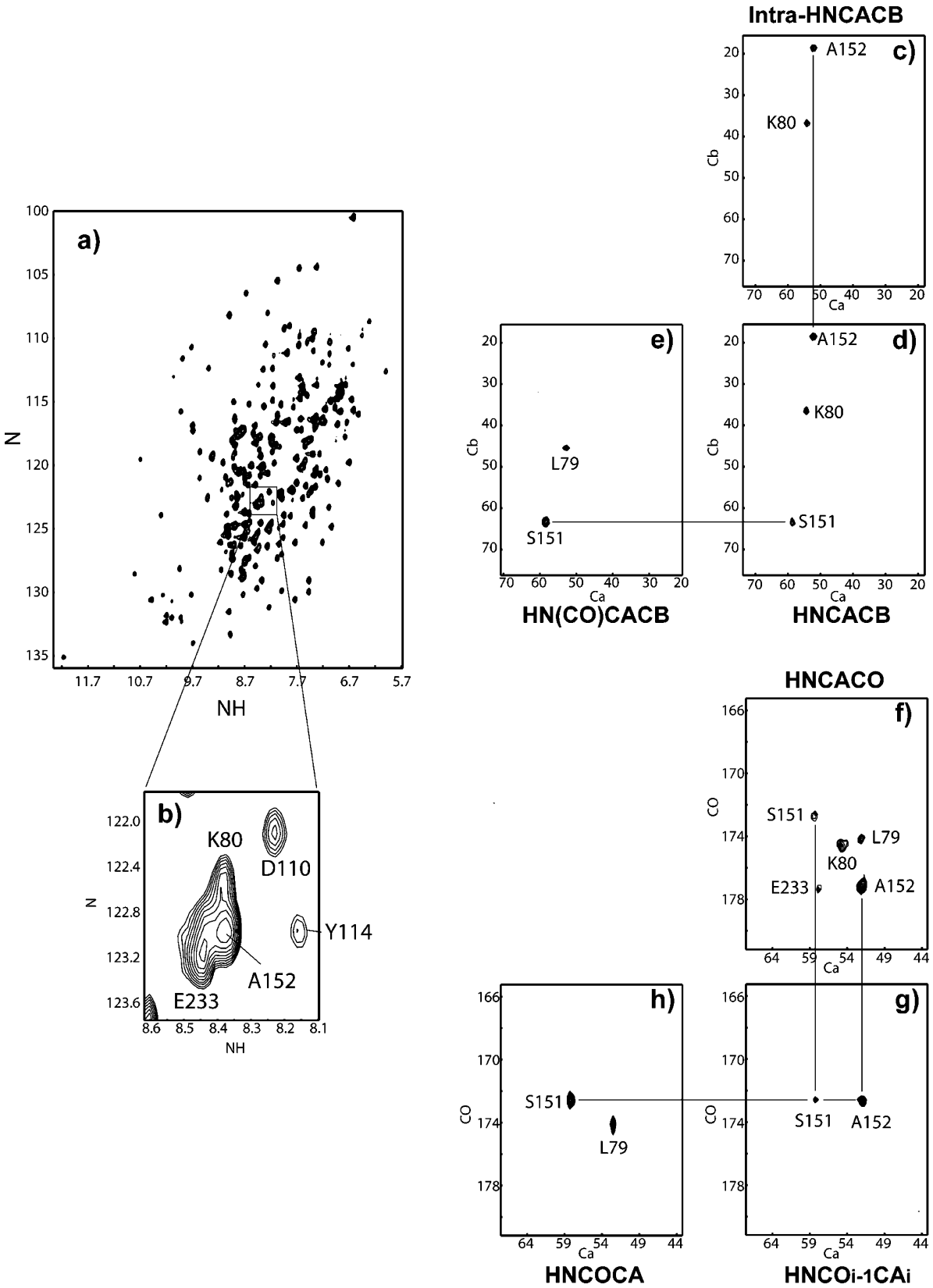

Figure 5. (a) 2D $^{1}H,^{15}N$-TROSY[35] spectrum collected using a 0.94 mM sample of $^{2}H/^{13}C/^{15}N$-methyl protonated (Leu, Val, and Ile $\delta 1$) labeled HCA II. The expanded region shown in (b) shows three residues, namely, A152, K80, and E233, that have similar ^{1}H and ^{15}N chemical shifts. The planes shown are extracted from the (c) Intra-HNCACB, (d) HNCACB, (e) HN(CO)CACB, (f) HNCACO, (g) $HNCO_{i-1}CA_i$, and (h) HNCOCA 4D spectra at the ^{1}H and ^{15}N chemical shifts (8.37 ppm, 122.9 ppm) of residue A152. The data were all collected on a Varian INOVA 800 MHz spectrometer using a triple-resonance cold probe. All of the spectra were reconstructed using the Hybrid Backprojection/Lower-Value method with a bin size of eight with the exception of the TROSY−HNCACO spectrum. Since the signal-to-noise ratio of the TROSY−HNCACO data was somewhat lower than that of the other data sets, a bin size of 12 was used in the reconstruction of this 4D spectrum. At the higher bin size, the signal intensities increase but also broaden leading to a "bleed through" of E233 from an adjacent plane (f).

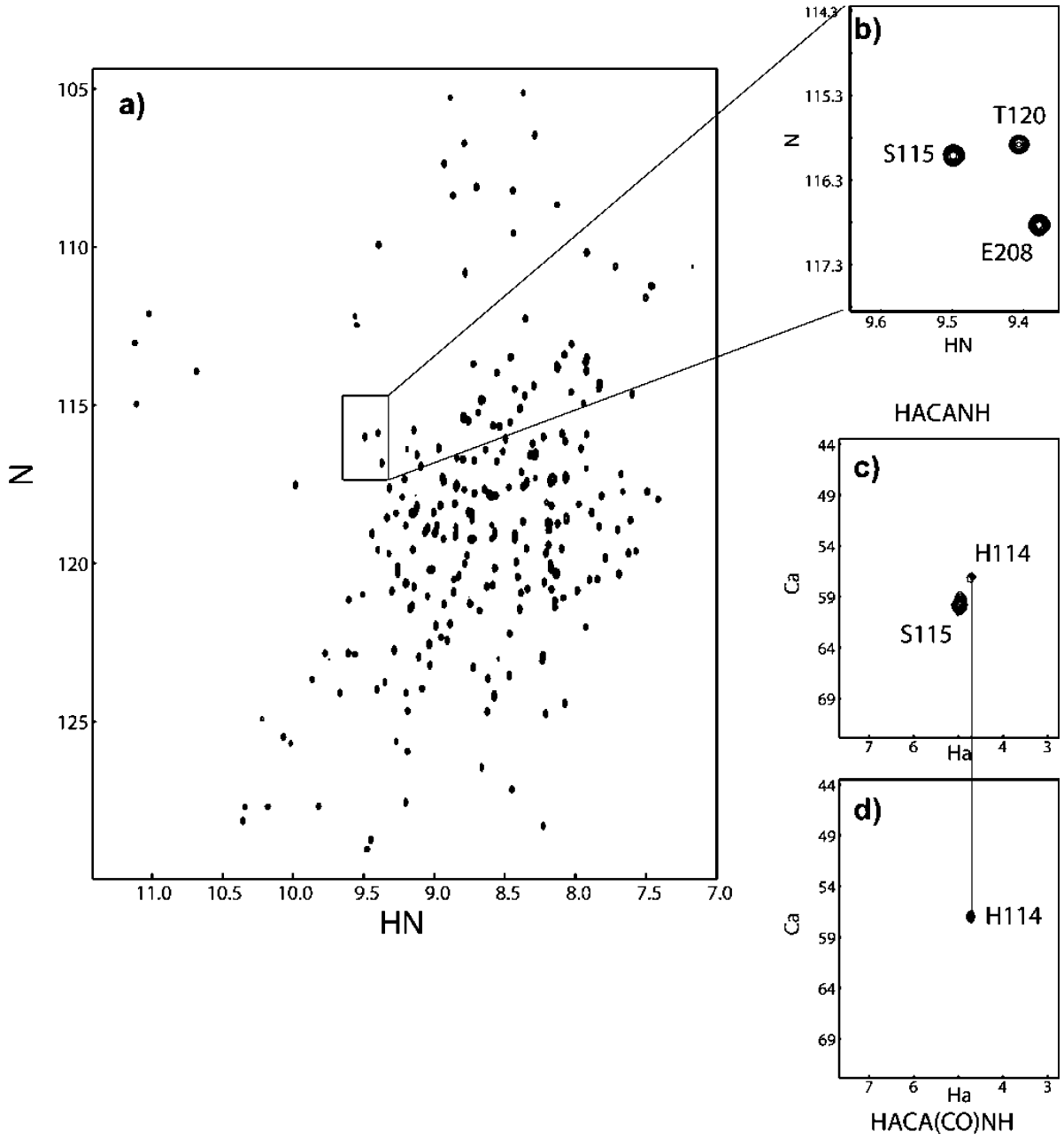

Figure 6. (a) 2D $^{1}H,^{15}N$-TROSY[35] spectrum and (b) expansion collected using a 0.9 mM sample of $^{1}H/^{13}C/^{15}N$-labeled calbindin D_{28K}. The planes shown are extracted from the (c) HACANH and (d) HACA(CO)NH 4D spectra at the ^{1}H and ^{15}N chemical shifts (9.49 ppm, 116.1 ppm) of residue S115. The data were all collected on a Varian INOVA 600 MHz spectrometer using a triple-resonance cold probe. All of the spectra were reconstructed using the Hybrid Backprojection/Lower-Value method with a bin size of eight.

same time, it is important to remember that with $n = m$, or even sometimes with $n < m$, false peaks need not occur, as the projections and peaks may be arranged such that there is no location that co-projects with cross-peaks on all projections. For these reasons, eqs 16−18 must be viewed as guidelines, rather than exact answers. It should be emphasized here that m refers to the total number of cross-peaks at a given position in the directly observed proton dimension, rather than the total number of cross-peaks in the entire spectrum.

As this paper was being submitted, Kupče and Freeman introduced an iterative fitting approach[58] aimed at removing backprojection artifacts while retaining optimum sensitivity. The hybrid approach that we present here achieves these same goals and represents an alternative scheme. An advantage of the approach presented here is that neither prior information about the reconstructed spectrum nor peak fitting procedures for the projected spectrum is required. This is likely to be beneficial when working with more complex spectra of larger molecules.

Data Reconstruction. Following NMRPipe processing, the 2D data sets were imported into the C program, PR−CALC, for reconstruction. The program, written in our laboratory and to be described in detail elsewhere, can accommodate the reconstruction of spectra of arbitrary dimensionality from projections of arbitrary dimensionality and orientation. The spectra generated by PR−CALC are output in NMRView[59] format for further analysis. The data collected herein were

(58) Kupče, E.; Freeman, R. *J. Magn. Reson.* **2005**, online ASAP.

(59) Johnson, B. A.; Belvins, R. A. *J. Biomol. NMR* **1994**, *4*, 603−614.

reconstructed into 4D spectra using the HBLV algorithm with a bin size of eight ($k = 8$). Empirical analysis determines that this bin size provides a good balance among three key factors: artifact suppression, signal strength, and line broadening. A final matrix size of 64 × 64 × 128 × 738 results in a 1.5 GB NMRView file. Although a large data matrix could be viewed as a drawback compared to the GFT approach, we found the ability to flip through planes and examine the relative distribution of peaks crucial for the analysis of large proteins.

It should be noted here that it is not necessary to reconstruct a full 4D matrix. The PR−CALC program allows for the reconstruction of any user-specified subset of the data. For instance, it may be advantageous, from the point of view of data storage and processing speed, to reconstruct a series of 2D $^{13}C_\alpha/^{13}C_\beta$ planes from a 4D HNCACB data set at $^1H_N/^{15}N$ chemical shifts predetermined from the cross-peaks observed in a 2D $^1H-^{15}N$ correlation map. This selective reconstruction is straightforward with PR−CALC.

Data Analysis. We have examined all of the reconstructed 4D spectra in detail using in-house NMRView scripts. Since the backbone assignments of HCA II[60] and D_{28K}[61] have previously been determined by conventional NMR experimentation, we have verified these assignments with the reconstructed spectra. For HCA II, the TROSY-based HNCACB, HN(CO)-CACB, and Intra-HNCACB spectra provided 94% of the C_α and C_β assignments, while the TROSY-based HNCACO, HNCOCA, and $HNCO_{i-1}CA_i$ spectra provided 94% of the C_α and C′ assignments. The 4D HACANH and HACA(CO)NH reconstructions of D_{28K} data provided 95% of the H_α/C_α pairs. Of note, since the conventional data were collected in a 3D mode and the PR−NMR data were reconstructed as 4D spectra with very high resolution, we were able to make some additional backbone assignments for HCA II which were not possible previously because of spectral overlap in the conventional data. With the 4D PR−NMR data, we were able to assign 6 new C_α/C_β pairs and 35 new C′ shifts. However, residues very close to the HCA II zinc binding sites at H94 (I91, Q92, and H94) and H119 (L118-V121) were missing in the PR−NMR data because of line broadening.

Figures 5 and 6 further illustrate the advantages offered by these high-resolution 4D data sets toward the assignment of higher molecular weight proteins and/or proteins which lack spectral dispersion. As can be seen in the TROSY spectrum of HCA II (Figure 5a and b), the ^1H and ^{15}N chemical shifts of residue A152 overlap significantly with those of residues E233 and K80; however, when all of the 4D data sets are considered, the assignment of the backbone nuclei for these three residues is straightforward. Figure 6 further illustrates the excellent quality and resolution offered by the PR−NMR data from calbindin D_{28K}. These figures also clearly demonstrate that the HBLV algorithm is capable of reconstructing artifact-free spectra even for complex data.

Conclusions

We have demonstrated that projection−reconstruction NMR techniques can be used successfully to assign larger protonated and perdeuterated proteins up to 30 kDa in molecular weight. The suite of 4D triple-resonance experiments presented herein provides data with exceptional resolution and sufficient signal intensity using an order of magnitude less spectrometer time than would be required if conventional 4D data were acquired. The added resolution and dimensionality should prove invaluable for protein studies. In addition, we have demonstrated that the Hybrid Backprojection/Lower-Value algorithm is superior for reconstructing data with weak signal-to-noise ratios and few tilted projections.

Acknowledgment. The Duke University NMR Center gratefully acknowledges instrumentation support from NIH, NSF, Duke University, and The North Carolina Biotechnology Center. We gratefully acknowledge Richele Thompson and David Kordys for providing us with the calbindin D_{28K} sample. This work was supported by the Whitehead Institute and by NIH Grants AI055588-01 to P.Z., DK58546 and GM55769 to J.C. B.E.C. is the recipient of an NSF Graduate Research Fellowship.

Supporting Information Available: Additional (4,2)D PR−NMR pulse sequences, data collection parameters, macros for automatic data collection, and NMRPipe processing scripts. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0509580

(60) Venters, R. A.; Farmer, B. T., II; Fierke, C. A.; Spicer, L. D. *J. Mol. Biol.* **1996**, *264* (5), 1101−1116.
(61) Helgstrand, M.; Vanbelle, C.; Thulin, E.; Linse, S.; Akke, M. *J. Biomol. NMR* **2004**, *28* (3), 305−306.