ELSEVIER

Journal of Molecular Structure 602–603 (2002) 275–292

Journal of
MOLECULAR
STRUCTURE

www.elsevier.com/locate/molstruc

Current approaches for the study of large proteins by NMR ☆

Ronald A. Venters[a], Richele Thompson[b], John Cavanagh[b,*]

[a]*Duke University NMR Center, Duke University Medical Center Durham, NC 27710, USA*
[b]*Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695-8301, USA*

Received 5 March 2001; accepted 27 March 2001

Abstract

An overview of current methods employed for characterizing larger (>25 kDa) proteins by NMR is presented. These techniques include: the attenuation of T_2 relaxation effects by offsetting dipole–dipole and chemical shift anisotropy relaxation mechanisms (TROSY); the extraction of residual dipolar couplings from partially oriented molecules; the elimination of relaxation pathways by incorporating deuterium nuclei into protein samples; the easing of resonance overlap by isotopically labeling only specific protein segments; and the decrease of rotational correlation times by dissolving proteins in low viscosity solvents.

Keywords: NMR; Large proteins; Structure; Methods

1. Introduction

Resolution and sensitivity concerns have always been at a premium in the analysis of the structural and dynamic characteristics of biomolecules using NMR spectroscopy. Until just recently the upper size limit for the complete characterization of proteins and their complexes was ~25 kDa. Such studies relied upon the highest possible fields available, the incorporation of $^{13}C/^{15}N$ isotopic labels and the use of the most sophisticated multidimensional NMR techniques. As Mother Nature would have it, the lion's share of interesting proteins are larger than this limit. Consequently, there is an enduring need to examine systems of increasing complexity and size. Such increases continue to present challenging problems in terms of excessive resonance congestion, slower tumbling times (longer rotational correlation times) and concomitantly shorter transverse relaxation times (T_2). The corresponding, and interrelated, increase in linewidth and decrease in sensitivity severely impedes the effectiveness of standard triple resonance NMR approaches. This article provides a brief overview of current methods employed in overcoming these limitations, therefore, allowing the study of relatively large proteins and complexes.

2. Deuteration and segmental labeling of proteins

2.1. Protein deuteration and selective protonation

By the incorporation of $^{13}C/^{15}N$ labels, improved spectrometer performance and a dizzying array of three and four dimensional multinuclear NMR methods, the protein molecular weight 'barrier' was

☆ Dedicated to Professor Graham A. Webb on the occasion of his 65th birthday.

* Corresponding author. Tel.: +1-919-513-4349; fax: +1-919-515-2047.

E-mail address: john_cavanagh@ncsu.edu (J. Cavanagh).

0022-2860/02/$ - see front matter
PII: S0022-2860(01)00690-1

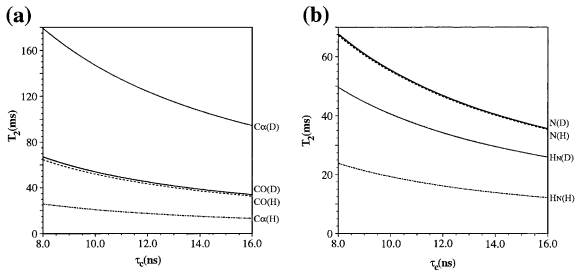

Fig. 1. Calculated T_2 relaxation times are plotted against the isotropic rotational correlation time (τ_c) for all backbone nuclei in both a fully protonated protein (:H) and its perdeuterated equivalent (:D). (a) C_α(D), ⋯; C_α(H), - - -; CO(D), —; CO(H), – – – (b) H_N(D), ⋯; H_N(H), - - -; N(D), —; N(H), – – –. T_2(N) relaxation times were calculated with the following parameters: $\Delta\sigma_N = -160$ ppm, $r_{N-HN} = 1.02$ Å, $r_{N-H_\alpha} = 2.12$ Å, $r_{N-CO} = 1.49$ Å and $r_{N-C_\alpha} = 1.49$ Å. T_2(CO) and T_{1calc}(CO) relaxation times were calculated with the following parameters: $\Delta\sigma_{CO} = 102$ ppm, $r_{CO-C_\alpha} = 1.54$ Å, $r_{CO-N} = 1.49$ Å, $r_{CO-H_\alpha} = 2.16$ Å and $r_{CO-HN} = (2.24, 3.30$ Å). $T_2(C_\alpha)$ relaxation times were calculated with the following parameters: $\Delta\sigma_{C_\alpha} = 29.5$ ppm for Gly and 21.5 ppm for Ala, $r_{C_\alpha-D_\alpha} = 1.05$ Å, $r_{C_\alpha-H_\alpha} = 1.09$ Å, $r_{C_\alpha-H_\beta} = 2.16$ Å, $r_{C_\alpha-CO} = 1.54$ Å, $r_{C_\alpha-N} = 1.49$ Å, $r_{C_\alpha-C_\beta} = 1.54$ Å (Ala only) and $r_{C_\alpha-HN} = (2.24, 3.00$ Å). $T_2(H_N)$ relaxation times represent an average of values from α-helices and parallel β-sheets. For an α-helix, each amide proton was assumed to interact with 4 other amide protons at the following distances: 2.8, 2.8, 4.2 and 4.2 Å. For a parallel β-sheet, each amide proton was assumed to interact with other three amide protons at the following distances: 4.2, 4.2 and 4.0 Å. In both cases, CSA contributions to the amide proton linewidth for $\tau_c = 11.4$ ns were estimated at 3.8 Hz. Reprinted with permission from [9].

raised to ~25 kDa in the mid 1990s [1]. However the increased resolution afforded by these new methods came with a price. Multiple transfer steps and evolution periods cause a reduction in the amount of magnetization that eventually is detected, and sensitivity suffers enormously. Sensitivity in critical backbone assignment/correlation experiments is most affected by decreases in carbon and $H_N T_2$ relaxation times. Relaxation of non-carbonyl carbon nuclei is dominated by dipolar interaction with any attached proton(s) [2]; relaxation of amide protons has a significant component (40%) arising from dipolar interaction with the surrounding aliphatic protons [3]. Clearly, removal of these offending relaxation pathways offers a significant advantage. This can be achieved by perdeuteration of the protein.

Recently, the combined use of perdeuteration and $^{13}C/^{15}N$ isotopic labeling, in conjunction with heteronuclear 3D and 4D experiments, has yielded another leap in the size of proteins for which NMR can provide both chemical-shift assignments and tertiary structural information [4–8]. Because the deuteron gyromagnetic ratio (γ_D) is 6.5-fold smaller than the proton γ_H, a deuteron is 45-fold, $(\gamma_H/\gamma_D)^2$, less effective than a proton at causing dipolar relaxation of both the directly attached X nucleus and the surrounding proton nuclei. Therefore, the incorporation of ^2H into $^{13}C/^{15}N$-labeled proteins dramatically increases both the directly attached carbon (notably the $^{\alpha}C$) and $H_N T_2$ relaxation times.

Fig. 1 displays the calculated T_2 relaxation times as a function of τ_c for all backbone nuclei in both a fully protonated protein and its perdeuterated equivalent. The calculations include both CSA and dipolar contributions [9–11]. The theoretical data in Fig. 1 indicate that perdeuteration dramatically increases the T_2 relaxation times both for C_α nuclei [8] and for H_N nuclei [3].

Table 1
Calculated and average experimental nuclear T_2 relaxation times in $^2H/^{13}C/^{15}N$-labeled (:D) and calculated T_2 times in $^1H/^{13}C/^{15}N$-labeled (:H) HCA II (Experimental values are reported as Mean ± SD (n), with all individual measurements lying within 2.5 SD of the mean. C_α and CO experimental T_2 relaxation times were calculated from $T_{1\rho}$ measurements. T_2 values were calculated for an isotropic rotational correlation time (τ_c) of 11.4 ns and a proton frequency of 600 MHz)

Nucleus	Experimental (:D)	Calculated (:D)	Calculated (:H)
C_α (Ala)	124 ± 12 (9)	131	18
C_α (Gly)	104 ± 11 (13)	93	
CO	47 ± 3 (161)	48	45
N	52 ± 7 (184)	49	49
$N(H_z)$	43 ± 5 (187)		
H_N	29 ± 7 (186)	31	
H_N (α-helix)	24 ± 3 (20)	25	
H_N (β-sheet)	29 ± 5 (36)	36	

To confirm the validity of these theoretical data, Farmer and Venters [9] calculated and measured the C_α T_2 values specifically for both protonated and deuterated HCA II (29.1 kDa). In these calculations, an isotropic rotational correlation time (τ_c) of 11.4 ns was used; both internal motion and all types of exchange were neglected. For HCA II, the Ala $T_2(C_\alpha H)$ and $T_2(C_\alpha D)$ are calculated to be, respectively, 18 and 131 ms; the $T_2(H_N H)$ and $T_2(H_N D)$ are calculated to be, respectively, 12.8 and 23.2 ms. T_2 relaxation times for ^{13}CO and ^{15}N nuclei have been similarly calculated for both ^{1}H-HCA II and ^{2}H-HCA II. The calculated T_2 relaxation data for HCA II are summarized in Table 1. The data in Fig. 1 and Table 1 indicate that perdeuteration does not affect to any significant extent the T_2 relaxation of inphase ^{15}N and ^{13}CO magnetization. This stands in stark contrast to the significant effect that perdeuteration has on the T_2 relaxation of $C_\alpha$and H_N magnetization. Table 1 also includes experimental T_2 relaxation data obtained on ^{2}H-HCA II. The experimental data for each type of nuclear magnetization agree well with the calculated data.

An important question arises concerning the use of fully deuterated proteins versus proteins containing random fractional deuteration. Random 50% fractional deuteration has been proposed as optimum for obtaining sidechain ^{13}C and ^{1}H resonance assignments in large proteins [12]. For assigning sidechain ^{13}C resonances, the completely deuterated approach has two main disadvantages: the $^{13}C(D)$ T_1 is notably longer than the $^{13}C(H)$ T_1 and magnetization transfer starts on carbon, which is intrinsically 4-fold less sensitive than proton. On the other hand one must also consider that random fractional deuteration has its own set of disadvantages: broad ^{13}C resonance linewidths arising from the distributed ^{2}H isotope effect; decreased magnetization transfer efficiency during isotropic mixing, relaxation during $H_C \rightarrow C$ INEPT and $C_\alpha \rightarrow CO$ INEPT magnetization transfers; decreased sensitivity in the terminal $^{15}N-^{1}H$ reverse INEPT subsequence due to decreased $^{1}H_N T_2$ values. From the standpoint of obtaining as many sidechain ^{1}H assignments and subsequent NOE connectivities for the structure determination process as possible, it appears that random fractional deuteration has a distinct advantage. Clearly, complete deuteration precludes the direct assignment of sidechain ^{1}H resonances whereas random fractional deuteration does not.

However, as we move to ever larger proteins an advantage we consider very important for complete deuteration, is that a homogeneous ^{2}H isotopic environment for ^{13}C and ^{15}N nuclei is established. Perdeuteration minimizes losses in both sensitivity and resolution arising from the multi-bond ^{2}H isotope effect [13,14]. In contrast, random fractional deuteration of uniformly ^{13}C-labeled proteins gives rise to all possible CH_nD_m isotopomers for each aliphatic carbon group. For a particular aliphatic carbon, its ^{13}C chemical shift depends not only on its isotopomeric state (n and m values) but also on the isotopomeric state of aliphatic carbon groups both one and two bonds removed. Because of the multi-bond ^{2}H isotope effect, random fractional deuteration produces a broad distribution of ^{13}C chemical shifts for aliphatic carbons, especially methylene and methyl groups, thereby decreasing both sensitivity and resolution in the heteronuclear NMR experiments used for assignment purposes.

This is a compeling argument for using full deuteration for heteronuclear backbone and sidechain carbon assignments. But how can ^{1}H sidechain assignments be achieved? In our opinion there are two appropriate strategies, dictated by the protein size. For 'smaller' larger proteins (25–35 kDa), ^{13}C assignments can be made from a fully deuterated sample and then the commonly used HCCH–TOCSY experiment can be employed to establish ^{1}H sidechain assignments, using a fully *protonated* sample [15]. This approach has been found to work well in the case of HCA II (29.1 kDa) where spectral congestion was an issue but not prohibitively so.

Larger proteins require different isotopic labeling strategies. One such example is the technique of *segmental labeling*, discussed in detail later. Here only a fragment of the protein is isotopically labeled and only *it* is subsequently detected, thereby reducing spectral overlap. Another approach is selective protonation [16–18] which can be extremely powerful. To date, the majority of selective protonation has involved methyl protons. Methyl groups, although somewhat limited in terms of proton chemical shift dispersion, retain excellent T_2 relaxation characteristics even in the presence of ^{13}C labeling [19]; therefore, ^{13}C labeling has accompanied selective

protonation of methyl groups in otherwise perdeuterated proteins in order to increase the available chemical-shift dispersion. The attractiveness of methyl groups is augmented by their abundance in the hydrophobic core of a protein, allowing substantially more critical NOE connectivities to be identified, providing important restraints for structure calculations.

Rosen et al. [17] developed a protocol, using $^{13}C/^{1}H$ pyruvate as the sole carbon source in D_2O media, for the generation of triply labeled proteins with selective protonation at the methyl groups of alanine, valine, leucine and γ2-isoleucine. Currently the most efficient and cost-effective method for the selective protonation of leucine δ, valine γ and isoleucine δ1 methyl groups has been provided by Goto et al. [20]. In this approach [3-^{2}H], $^{13}C_{\alpha}$-ketoisovalerate is used to supplement the $^{13}C/^{2}H$-glucose, $^{15}NH_4Cl$, D_2O bacterial growth media producing $^{13}C/^{2}H/^{15}N$ labeled proteins selectively protonated at the methyls of Leu and Val. [3,3-$^{2}H_2$]. $^{13}C_{\alpha}$-ketobutyrate is also added to the media, generating the methyl protonated isoleucine. The level of methyl protonation achieved in this way is $>90\%$.

Selectively protonated perdeuterated samples can then be subject to a variety of NOESY based experiments establishing the following NOE connections: H_N–H_N, H_N–methyl and methyl–methyl. Using these restraints and dihedral angle information, structures can be produced that are of sufficient quality to identify ligand-binding sites, map sites of interaction and to identify homologous proteins. As alluded to above, global folds can be used to *unambiguously* identify additional NOEs generated from fully protonated, segmentally labeled or randomly fractionally deuterated samples. This strategy, in conjunction with extra restraints provided by dipolar couplings (see below), can substantially refine global folds of large proteins. A comprehensive application of this process comes from Kay and co-workers in the determination of the global fold of the maltose-binding protein (MBP) in complex with the substrate β-cyclodextrin [21]. MBP is a 370 residue, 42 kDa protein. Distance restraints came from H_N–H_N (826), H_N–methyl (769) and methyl–methyl (348) NOEs. The number in parentheses refers to the number of NOEs for each type of interaction. Using these NOEs along with 48 hydrogen bond restraints and 555 dihedral angle restraints, the global RMSD of the 10 lowest energy structures was 5.5 Å. The RMSDs of the N- and C-terminal domains, respectively, were 2.4 and 3.8 Å. Importantly, the NOEs involving methyl groups tend to extend to greater distances thanks to the reduced possibility of spin diffusion and the concomitant increase in mixing times. Following this strategy, the inclusion of dipolar couplings reduced the global RMSD significantly. This is discussed in detail later.

Another method for removing clustering resonances is to produce a protein that is 'reverse isotope labeled'. Bax and co-workers produced proteins that were uniformly ^{13}C labeled except for the aromatic residues phenylalanine and tyrosine [22]. Using appropriate editing methods, NOEs between these residues and others close by can be exclusively obtained. This strategy works well for proteins below a ~30 kDa limit where aromatic ^{1}H resonances are still resolved. Oschkinat and co-workers followed a similar theme when incorporating completely unlabeled phenylalanine, tyrosine, threonine, isoleucine and valine into highly deuterated, ^{15}N labeled proteins [23]. A variety of isotope editing methods can be then used to determine NOEs from carbon-bound protons to either the amide protons of these unlabeled residues or the amide protons of other, labeled residues.

2.2. Segmental labeling

Another recent and intriguing approach to reducing spectral congestion is *to simply look at only a segment of the intact protein*. This can be accomplished by the joining of unlabeled and labeled recombinant proteins in vitro. Although there is no reduction in linewidths and resolution *per se* is not enhanced, the number of peaks per unit spectral region is reduced, making spectral interpretation and assignments easier.

The initial approach of Perler [24] was adapted by Yamazaki et al. and employs the self-catalytic splicing process of intein excision from a polypeptide chain [25–27]. It is well known that a variety of prokaryotic and lower eukaryotic genes contain an in-frame open reading frame that encodes for an intein. Inteins (*in*ternal pro*tein* fragments) are runs of 'useless' polypeptide chain removed by the post-translational editing operation of splicing. The remaining ends of the now broken polypeptide

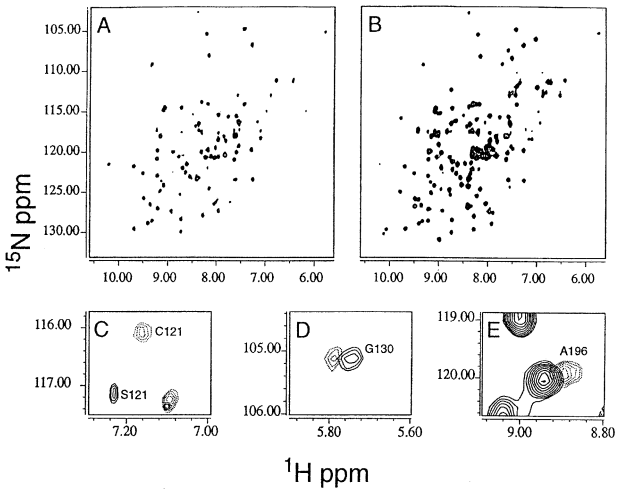

Fig. 2. $^{1}H-^{15}N$ correlation spectra (500 MHz) of: (a) segmentally labeled Abl-[$G^{120}C^{121}$][SH2-^{15}N]SH(32); (b) wild-type Abl-SH(32) with uniform ^{15}N labeling. The peaks in (a) are the SH2-associated subset of those in (b). (c–e) show peaks with detectable chemical shift changes away from their position in the wild-type spectrum (solid lines). (c) S^{121} in the wild-type is mutated to C^{121} in the segmentally labeled protein. Residues G^{130} and A^{196} are spatially close to the SH3–SH2 junction and are very slightly perturbed. Reprinted with permission from Ref. [28].

chain, the exteins (*ex*ternal pro*teins*) are able to ligate to form a continuous chain once again [24].

For the purpose of our discussion, consider that the two exteins to be ligated are designed to be the segmented target domains of a full length protein of interest. If one segment was isotopically labeled while the other was not, the removal of the intein and subsequent ligation would result in a 'segmentally labeled' protein. NMR methods employing simple isotope filters would then select only resonances from the segmentally labeled region, thereby eliminating a large number of peaks in the resulting spectra and increasing the accessibility to information. This, of course, is particularly important in the spectra of large proteins where resonance density is always problematic.

This method requires that specific residues are present at the splicing junctions: the N-terminus of the intein should end in a Cys/Ser sequence; the C-terminus of the intein should end in a His-Asn sequence; the N-terminus of the C-terminal segmental target domain (extein) should be a Cys/Thr/Ser — this will join with the C-terminus of the N-terminal segmental target domain (extein) to form the intact protein. For different types of labeling on the N- and C-terminal segments of the extein (e.g. unlabeled and $^{15}N/^{13}C$ labeled) the precursor intein is cut in the middle and the fragments produced independently in media containing the appropriate NMR active isotopes. Subsequently they are combined, refolded and then ligated. This method requires that the splicing junction site be in a flexible loop region of the protein. Such flexibility is ensured by introducing a string of glycines at the appropriate site.

A very elegant method for segmental labeling was introduced by Muir an co-workers [28]. This approach removes the requirement for both a denaturation/refolding step and also removes the need for the inclusion of a run of flexible glycines at the splicing site. An N-terminal target segment (with a Cys at the end) is produced with an intein at its C-terminus. The intein is attached to a chitin binding domain which in turn fixes the whole assembly to an affinity resin. The linkage between the N-terminal fragment and the fused intein protein-splicing domain is in equilibrium between an amide and a thioester. Addition of ethanethiol at pH 6 removes the intein and produces an independent ethyl-α-thioester derivative of the N-terminal target fragment. This is then combined with the required C-terminal target fragment containing the essential N-terminal cysteine for ligation. The domain of choice can be isotopically labeled prior to the ligation resulting in the required labeled fragment. Muir and co-workers have developed and improved their

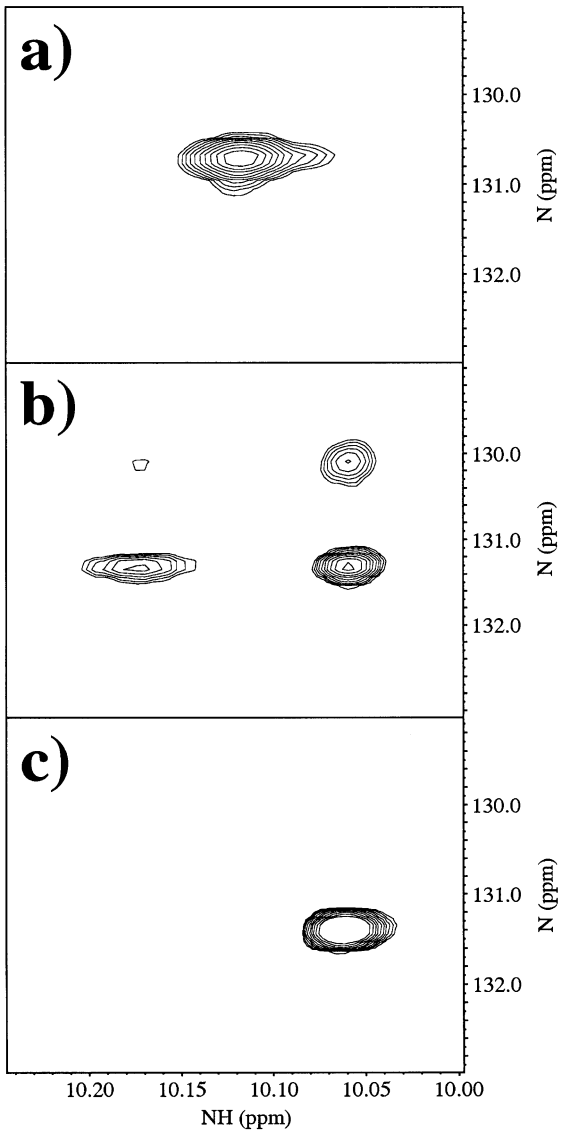

Fig. 3. Contour plots showing the ^{15}N, ^{1}H correlation for residue F146 of human carbonic anhydrase. These data were collected on a 0.94 mM $^{2}H/^{13}C/^{15}N$-labeled HCA II sample at 5°C using a Varian Inova 800 MHz spectrometer. The correlation time (τ_c) for HCA II at this temperature was determined to be 26 ns. (a) Sensitivity-enhanced gradient-enhanced $^{1}H_N$, ^{15}N HSQC spectrum with broad-band decoupling applied during the t_2 acquisition time. The $^{1}H_N$, ^{15}N scalar coupling during t_1 was refocused by applying a 180° proton pulse in the center of the evolution time. (b) Sensitivity-enhanced gradient-enhanced $^{1}H_N$, ^{15}N HSQC spectrum without decoupling in t_1 and t_2. (c) $^{1}H_N/^{15}N$ — TROSY correlation recorded with the improvements to the original sequence as described in Pervushin et al., 1998.

methodology to both improve yields and, importantly, to allow multiple protein segments to be regioselectively linked, allowing internal protein domains to be isotopically labeled [29]. Fig. 2b shows an $^{1}H-^{15}N$ HSQC spectrum of the fully ^{15}N labeled Abelson protein tyrosine kinase-(SH32) domain pair. This protein consists of an SH3 and an SH2 domain. Fig. 2a shows an $^{1}H-^{15}N$ HSQC spectrum of the same protein with only the SH2 domain ^{15}N labeled using the method of Muir. Further spectral details are provided in the figure caption.

In addition to removing signal overlap in larger proteins, segmental labeling provides some interesting options. The ability to label a single domain of a homomultimer allows for studies of the effects of inter-monomer interactions/influences by using appropriate mutations in monomer units prior to ligation. Also, the enhancement of dipolar coupling effects (see later) is possible by the ability to incorporate magnetically anisotropic regions in a protein that can assist in molecular orientation.

3. Transverse relaxation-optimized spectroscopy

As discussed above, perdeuteration can significantly increase the tractable size of macromolecules that can be studied when used in combination with optimized multi-dimensional multi-nuclear NMR experiments. Recently, Pervushin and coworkers reported a new class of NMR experiments that offer yet another significant increase in both sensitivity and resolution [30]. These so-called transverse relaxation-optimized spectroscopy (TROSY) experiments are based upon the mutual cancellation of relaxation contributions from dipole–dipole (DD) and chemical shift anisotropy (CSA) offering a dramatic reduction in T_2 relaxation.

In the ubiquitous ^{15}N-^{1}H HSQC experiment, decoupling of $^{1}H_N$ during t_1 and ^{15}N during t_2 generates a single cross-peak for each $^{1}H_N-^{15}N$ pair in the protein. Recording the same spectrum without decoupling produces a four cross-peak multiplet pattern as seen in Fig. 3b for residue F146 of human carbonic anhydrase II (HCA II). It has long been known that cross-correlated spin relaxation introduced by DD and CSA interference leads to differential relaxation rates for the individual transitions in a system containing two

coupled spin 1/2 nuclei [31]. In the standard HSQC experiment, application of decoupling pulses interchanges the slowly and rapidly relaxing components of the multiplet resulting in an averaging of the relaxation rates for the observable cross-peak (Fig. 3a). Such mixing of relaxation rates becomes increasingly destructive to the overall signal-to-noise ratio at higher magnetic field strengths and longer rotational correlation times. In the fully-coupled $^{15}N-^{1}H_N$ HSQC spectrum presented in Fig. 3b, differential relaxation is clearly evident and is manifest by differing linewidths for the multiplet components. These data were collected on a 0.94 mM $^{2}H/^{13}C/^{15}N$-labled HCA II sample at 5°C using a Varian Inova 800 MHz spectrometer. At 29 kDa, the correlation time (τ_c) for this protein was determined to be 26 ns. The cross-peak at $^{15}N = 131.35$ ppm and $^{1}H_N = 10.06$ ppm displays the narrowest linewidth in both dimensions since it originates from the spins where the two competing relaxation mechanisms partially cancel. The TROSY sequence, as was originally designed [30], minimizes exchange between multiplet components and measures only the narrowest component by not refocusing the one-bond $^{1}H_N-^{15}N$ scalar coupling during the t_1 and t_2 evolution times and by phase cycling out the remaining faster relaxing elements of the N–H cross-peak (Fig. 3c).

The optimum cancellation of transverse relaxation effects for the narrow component of a $^{1}H_N-^{15}N$ spin-pair is achieved at ^{1}H frequencies near 1.1 GHz [30]. At this field residual transverse relaxation caused by DD and CSA interactions of the $^{1}H_N-^{15}N$ spin-pair are almost independent of the molecular size leaving relaxation to be dominated, in this case, by DD interactions with remote protons and perhaps conformational exchange. Furthermore, 95% of the residual $^{1}H_N T_2$ relaxation and 75% of the residual ^{15}N T_2 relaxation is a result of DD interactions with remote protons [30]. These relaxation mechanisms can be almost entirely removed by perdeuteration. This is a particularly striking example of the synergy of these methodologies.

TROSY also yields spectacular improvements in sensitivity for aromatic CH groups [32]. In this case, $^{1}H-^{13}C$ DD coupling is of a comparable magnitude to the ^{13}C CSA at ^{1}H resonance frequencies between 600 and 800 MHz, whereas the aromatic protons have a small CSA value making them unsuitable for TROSY methodologies. Experimentally this translates into recording the ^{13}C chemical shifts using the TROSY method and the ^{1}H chemical shifts while applying broadband ^{13}C decoupling.

Not surprisingly, improvements to the original TROSY sequences appeared quickly. Pervushin and coworkers introduced the concept of single transition-to-single transition polarization transfer (ST2-PT) [33]. Essentially the improvement consists of changing the second half of the experiment to include a different phase cycling scheme in combination with the application of an echo/anti-echo quadrature detection method in the ^{15}N dimension. In the original TROSY experiment the evolution of the density operator can be represented by

$$I_y^{13} + I_y^{24} \rightarrow S_x^{34}\cos(\omega_s^{34} t_1) \rightarrow I_x^{24}\cos(\omega_s^{34} t_1) \quad (1)$$

Where I^{13} represents the single transition operators for the $1 \rightarrow 3$ transition when considering the common energy level diagram, I^{24}, the $2 \rightarrow 4$ transition, and so on [33]. The first half of the schematic represents ^{1}H to ^{15}N coherence transfer and ^{15}N chemical shift evolution during t_1. The second half of the equation represents the coherence transfer of ^{15}N back to ^{1}H. In order to obtain phase-sensitive data in the ^{15}N dimension, the $S_x^{34}\cos(\omega_s^{34} t_1)$ and $S_x^{34}\sin(\omega_s^{34} t_1)$ components were recorded in alternate scans in the original version of the TROSY experiment. Utilization of the improved scheme allows the simultaneous detection of the $S_x^{34}\cos(\omega_s^{34} t_1)$ and $S_x^{34}\sin(\omega_s^{34} t_1)$ components. With the improved sequence only one of the four frequency components is again selected, however, a full 50% of the magnetization is transferred from ^{15}N to $^{1}H_N$ spins instead of just 25% as was the case in the original TROSY pulse sequence [30]. Thus the improved ST2-PT version of the sequence results is a maximum $2^{1/2}$ sensitivity enhancement [33,34]. In addition, the water flip back technique was introduced in order to minimize the signal loss due to saturation transfer with solvent water. Furthermore, the phase cycle was reduced to two steps in order to prepare the TROSY sequence to be used as an element of more complex pulse sequences [33]. TROSY was then quickly incorporated into a host of experiments, most notably 3D heteronuclear protein assignment sequences, as described below [35–37].

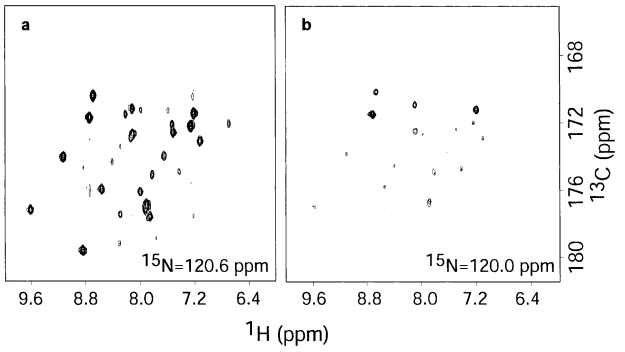

Fig. 4. HNCO spectra of TIM (54 kDa). (a) ^{1}H(F3)/^{13}CO(F1) 2D slice of a TROSY–HNCO spectrum; (b) ^{1}H(F3)/^{13}CO(F1) 2D slice of a conventional, gradient enhanced-HNCO spectrum. Reprinted with permission from [36].

As discussed previously, the standard procedure for determining the sequential assignment of a large protein or protein complex requires a ^{2}H/^{13}C/^{15}N labeled sample and a series of deuterium-optimized triple-resonance experiments that rely on the transfer of magnetization amongst scalar-coupled nuclei [13]. These experiments correlate either inter- or both intra- and inter-residue ^{13}C nuclei with a given $^{1}H_N$, ^{15}N spin-pair and contain long delays in which ^{13}C and ^{15}N magnetization evolves in the transverse plane. Fast transverse relaxation during these delays and the ^{1}H evolution times limits the size of proteins for which assignments can be successfully obtained even in perdeuterated systems. Triple resonance 3D experiments that incorporate the TROSY concept have recently been introduced [35,36]. These sequences suppress transverse relaxation of the ^{15}N nuclei during ^{15}N evolution times and of the $^{1}H_N$ nuclei during direct detection. When comparing the TROSY–HNCO sequence with the non-TROSY version of the same experiment Salzmann and coworkers obtained an average gain in sensitivity of 2.4 for a fully deuterated sample of the 23 kDa protein gyrase-23B at 20°C [35]. Similar improvements can be expected for the other 3D heteronuclear assignment experiments. Loria and coworkers introduced two improvements to this and other 3D TROSY-based sequences which increased the overall sensitivity by 7–10% and by up to 15% for residues undergoing chemical exchange linebroadening in trioephosphate isomerase (TIM) [36]. These changes involve reducing the constant-time ^{15}N evolution period and minimizing radiation damping by ensuring that water magnetization is never inverted along the $-z$ axis. Fig. 4a shows a 2D slice form a conventional HNCO of TIM compared to the equivalent slice from a TROSY–HNCO (Fig. 4b). The improvement in the TROSY–HNCO is remarkable [36]. TIM is a symmetric dimer of total molecular weight 54 kDa.

Yang and Kay introduced another improvement for the HNC type 3D TROSY sequences [38]. In all of the TROSY-based experiments, as described above, only one of the two possible components of ^{15}N magnetization is transferred to $^{1}H_N$ and subsequently detected while the other fast relaxing component is eliminated by phase cycling. For the 2D ^{1}H–^{15}N correlation experiment this is essential since, in most cases, the faster relaxing component would be detectable resulting in a doubling of the number of resonances. Kay and coworkers noted that in the case of triple resonance TROSY–HNC experiments, active elimination of the second pathway may not be needed since this magnetization relaxes efficiently during the constant time evolution periods. As an example, they calculated that the ratio of the two signals would be 20/1 for a 15 ns correlation time at 800 MHz and 64/1 for a 25 ns correlation time at the same field [38]. Armed with these data a 3D HNC TROSY scheme was developed which allows for the transfer of both pathways, but which is less sensitive to relaxation losses during the transfer than the methods described above. This was achieved by transferring ^{15}N magnetization back to $^{1}H_N$ nuclei using the gradient-enhanced sensitivity-enhanced approach, therefore, maximizing the amount of time that magnetization resides along the $+z$ axis during

the transfer. A gain in sensitivity of 1.19 was measured using the modified TROSY–HNCO sequence for MBP at 5°C ($\tau_c = 46$).

Unfortunately, even these TROSY-based experiments are predicted to fail when the rotational correlation time of the protein or protein complex exceeds 30–40 ns [39]. This is due to that fact that the majority of these experiments are recorded in constant time mode in the ^{13}C dimensions resulting in considerable sensitivity losses. These experiments could be recorded without constant time during ^{13}C evolution periods, but for large monomeric proteins the resolution is not likely to be satisfactory. A series of 4D TROSY-based pulse sequences, namely 4D HNCOCA, HNCACO and $HNCO_{i-1}CA_i$, have been introduced that address these sensitivity and resolution issues [39,40]. In these experiments, the $^{13}C_\alpha$ chemical shift is not recorded in a constant time mode, but the reduced resolution is compensated for by the added dimensionality. As described above, these experiments preserve both the fast and slowly relaxing components of the magnetization, therefore, in order to insure complete elimination of the slowly relaxing component, these sequences should only be used for proteins with correlation times >30 ns at fields at or above 600 MHz. The power of these sequences was demonstrated by the assignment of the ^1H_N, ^{15}N, $^{13}C_\alpha$, ^{13}CO and $^{13}C_\beta$ nuclei in the 67 kDa dimeric form of the protein containing the DNA-binding and oligomerization domains of the p53 tumor suppressor protein [41].

The most recent contribution in this area holds the promise of making possible the study of macromolecular systems of 100 kDa and larger [42,43]. All of the heteronuclear experiments described so far make use of INEPT to transfer magnetization between different nuclei. The efficiency of the INEPT transfer decreases as the molecular rotational correlation time increases; therefore, theory predicts that TROSY will be ineffective for molecules with molecular weights greater than 100 kDa. Cross-correlated relaxation-induced polarization transfer (CRINEPT) combines cross-correlated relaxation induced polarization transfer with INEPT. The rate of polarization transfer by cross-correlated relaxation becomes more efficient as τ_c increases. ^1H_N–^{15}N CRINEPT–TROSY experiments have demonstrated a 3-fold sensitivity enhancement over the conventional INEPT-based TROSY experiment for a 110 kDa protein at 4°C ($\tau_c =$ 70 ns) [42].

4. Residual dipolar couplings

Traditionally, protein structures determined by NMR have relied on distance restraints obtained from ^1H–^1H NOE data, *J*-coupling restraints, hydrogen bond restraints and dihedral angle restraints obtained from ^{13}C chemical shift data. These types of structural information, although strictly local in nature, have been used to successfully determine high-resolution structures for a wide variety of proteins [43]. In cases involving larger proteins that require perdeuteration of all non-exchangeable ^{13}C sites, the number of NOE restraints that can be obtained is limited but tend to extend to greater distances because of decreased spin diffusion and longer mixing times [9]. As described above, selective methyl protonation supplies additional NOE restraints, especially in the protein core [20]. The resultant global folds achieved with perdeuterated/methyl-protonated proteins can be extremely useful, however, additional refinement may be required especially in multi-domain proteins or protein complexes with few restraints connecting the structural elements. Recently, a new class of structural restraint has been introduced that is not strictly local in nature and which represents a major advance in NMR structural studies especially for larger proteins and protein complexes [44]. These restraints are based upon the measurement of residual dipolar couplings between pairs of NMR active nuclei.

Dipolar couplings, a major source of structural data in solid-state NMR, are not normally observed for proteins in solution because of rotational diffusion. It was demonstrated as early as 1987 using small heme-like molecules [45] and later with cyanometmyoglobin [44] at high magnetic field strengths that a small amount of molecular alignment occurs for molecules that possess a nonzero magnetic susceptibility. For cyanometmyoglobin it was shown that this alignment, although small, could result in measurable one-bond ^1H–$^{15}N_H$ residual dipolar couplings.

The dipolar splitting that results from the interaction of two spin 1/2 nuclei, *i* and *j* is given by

$$D_{ij} = -[\gamma_i \gamma_j h/2\pi r^3)]\langle(3\cos^2\theta - 1)/2\rangle \quad (2)$$

where γ are the gyromagnetic ratios, h the Plank's constant, r the distance between the two spins, and θ is the angle that the internuclear vector makes with respect to the magnetic field [44]. The brackets in the equation represent motional averaging so that in the absence of alignment $D_{ij} = 0$. In directly bonded nuclei, like amide $^{1}H_N-^{15}N$ pairs in proteins, these couplings manifest themselves as small field-dependent perturbations to the $^{1}H_N-^{15}N$ scalar coupling as measured in nondecoupled 2D heteronuclear NMR spectra. Dipolar couplings are proportional to the square of the field strength while the scalar coupling is not dependent on the field; therefore, the effects of dipolar and scalar couplings can be separated. This separation is accomplished by measuring the splittings resulting from the combined couplings at several different field strengths [46]. The constant scalar coupling contribution can then be removed. As can be seen in the above equation, the residual dipolar couplings thus measured are angle dependent and can be used as structural restraints that do not rely on nearby nuclei to define their position or orientation; therefore, these represent an entirely new form of structural information that can be extremely important in orienting structural elements that are far removed in the tertiary fold of the protein.

Cynaometmyoglobin, because of the heme moiety, has a very highly anisotropic paramagnetic susceptibility [44]. In fact, although the majority of them are very small, most proteins have a measurable anisotropic magnetic susceptibility tensor. Bax and coworkers have very accurately measured one-bond $^{1}H_N-^{15}N$ dipolar couplings in human ubiquitin [46]. In this diamagnetic protein there is a very small magnetic alignment that results from the sum of the anisotropic magnetic susceptibility contributions of the backbone peptide bonds and the aromatic side chains alone. The $^{1}H_N-^{15}N$ dipolar couplings were determined by fitting resonance intensities from a series of $^{1}J_{NH}$-modulated 2D spectra collected at ^{1}H frequencies of 360, 500 and 600 MHz. The measured residual couplings in ubiquitin are very small indeed, ranging from -0.09 to $+0.10$ Hz.

As was mentioned above for the ubiquitin case, the measurement of these very small residual dipolar couplings with sufficient accuracy proved to be a very difficult task. A major breakthrough in the measurement of dipolar couplings was realized with the introduction of anisotropic media for macromolecular alignment [47]. The first such method involved the use of a dilute aqueous liquid crystalline phase of discotic phospholipid micelles consisting of a 1:2.9 ratio of dihexanoyl phosphatidylcholine (DHPC) and dimyristoyl phosphatidylcholine (DMPC). These lipids form bicelles and a liquid crystalline phase over a narrow concentration and temperature range [47]. Macromolecules dissolved in this media adopt a small degree of alignment due to steric effects. The key to this procedure is the proper tuning of the resultant protein alignment so that only dipolar couplings between closely spaced nuclei give measurable dipolar couplings. If too much alignment is achieved the resulting spectra have too many large dipolar coupling interactions and the spectra become impossible to interpret. Furthermore, the spacing between the bicelles must be much greater than the size of the macromolecule being studied so that T_2 relaxation rates are not adversely affected. Residual dipolar couplings for proteins properly aligned in such media are several orders of magnitude larger than couplings measured based on the alignment caused by the protein's inherent magnetic susceptibility anisotropy alone.

Recently other tunable alignment particles have been introduced [48,49]. Most widely adopted for their ease of use has been the Pf1 filamentous bacteriophage. The Pf1 bacteriophage particle has a diameter of $\sim$60 Å and a length of $\sim$20,000 Å. The coat protein, which packages the DNA, forms an α-helical structure that is parallel to the long axis of the phage forming an alignment of the carbonyl groups. This organization is believed to be the reason for the large magnetic susceptibility anisotropy and the alignment of the particle in the magnetic field. The phage particles have many properties that make them easy to use. They form a stable alignment at a range of concentrations, pH, temperatures and salt concentrations. In addition, they are fully aligned at all magnetic fields used for protein NMR studies. Furthermore, the solutions formed using precious protein samples are stable over long periods of time and both the phage and protein can be separated and retrieved by a simple ultra-centrifugation step [49].

Measurement of dipolar couplings in large proteins is complicated by problems of low spectral resolution

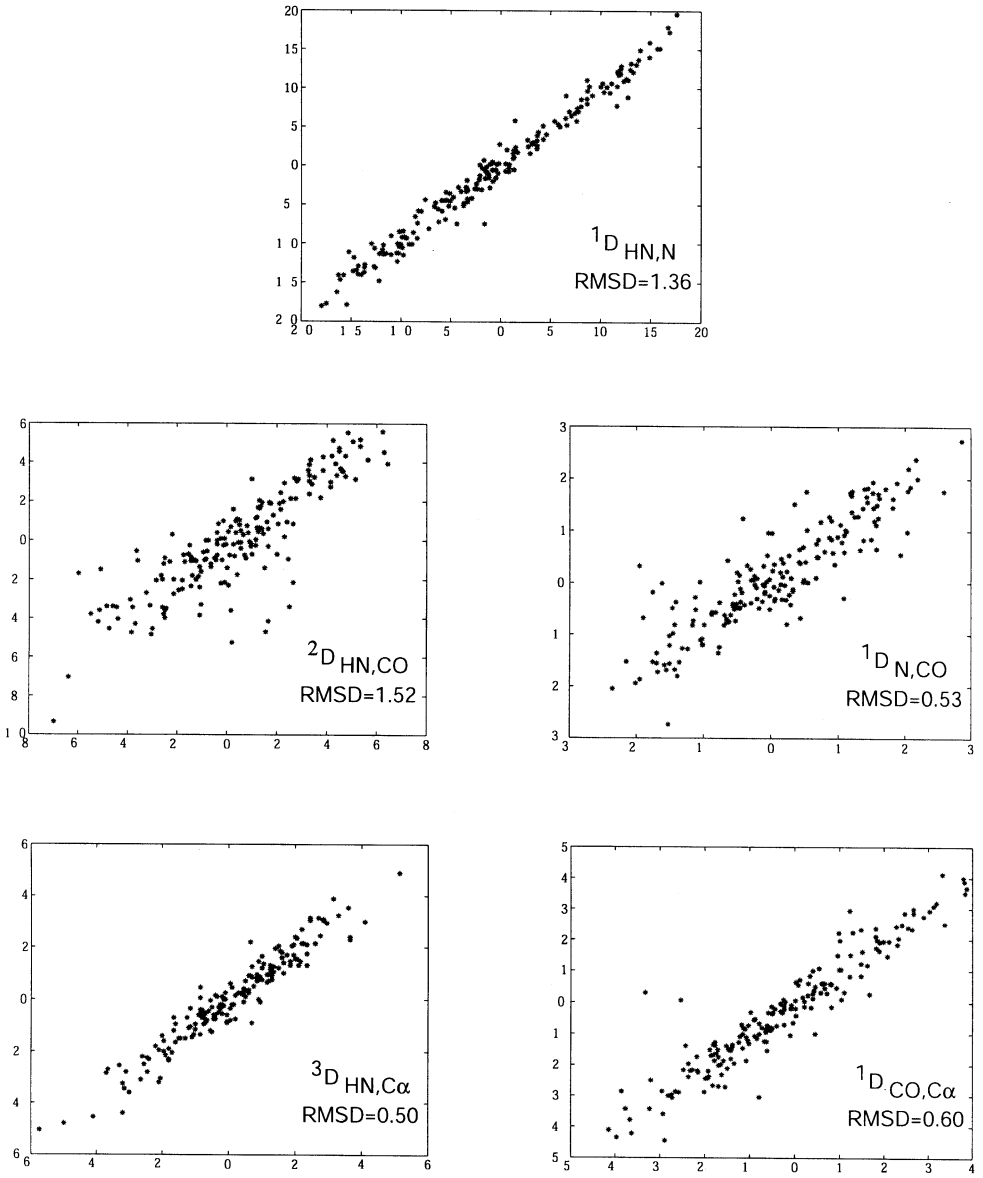

Fig. 5. Correlation between dipolar couplings measured from duplicate experiments on ~1.6 mM HCAII samples (with and without phage). The pair-wise rmsd values obtained are indicated in the lower right hand corner of each plot.

and low sensitivity [50]. In order to accurately measure residual dipolar couplings in large partially aligned $^{2}H/^{13}C/^{15}N$ labeled proteins it becomes necessary to use TROSY-based 3D experiments so that these problems can be overcome. To this end, Yang and coworkers have introduced a suite of TROSY-based triple resonance pulse sequences that are based on the HNCO experiment and which measure the one-bond $^{1}H_N-^{15}N$, $^{15}N-^{13}C'$, $^{13}C'-^{13}C_\alpha$, the two bond $^{1}H_N-^{13}C'$ and the three bond $^{1}H_N-^{13}C_\alpha$ dipolar couplings [51]. Residual dipolar couplings are determined by recording spectra in the presence and absence of phage-induced alignment and by taking the difference of the splitting obtained in each case.

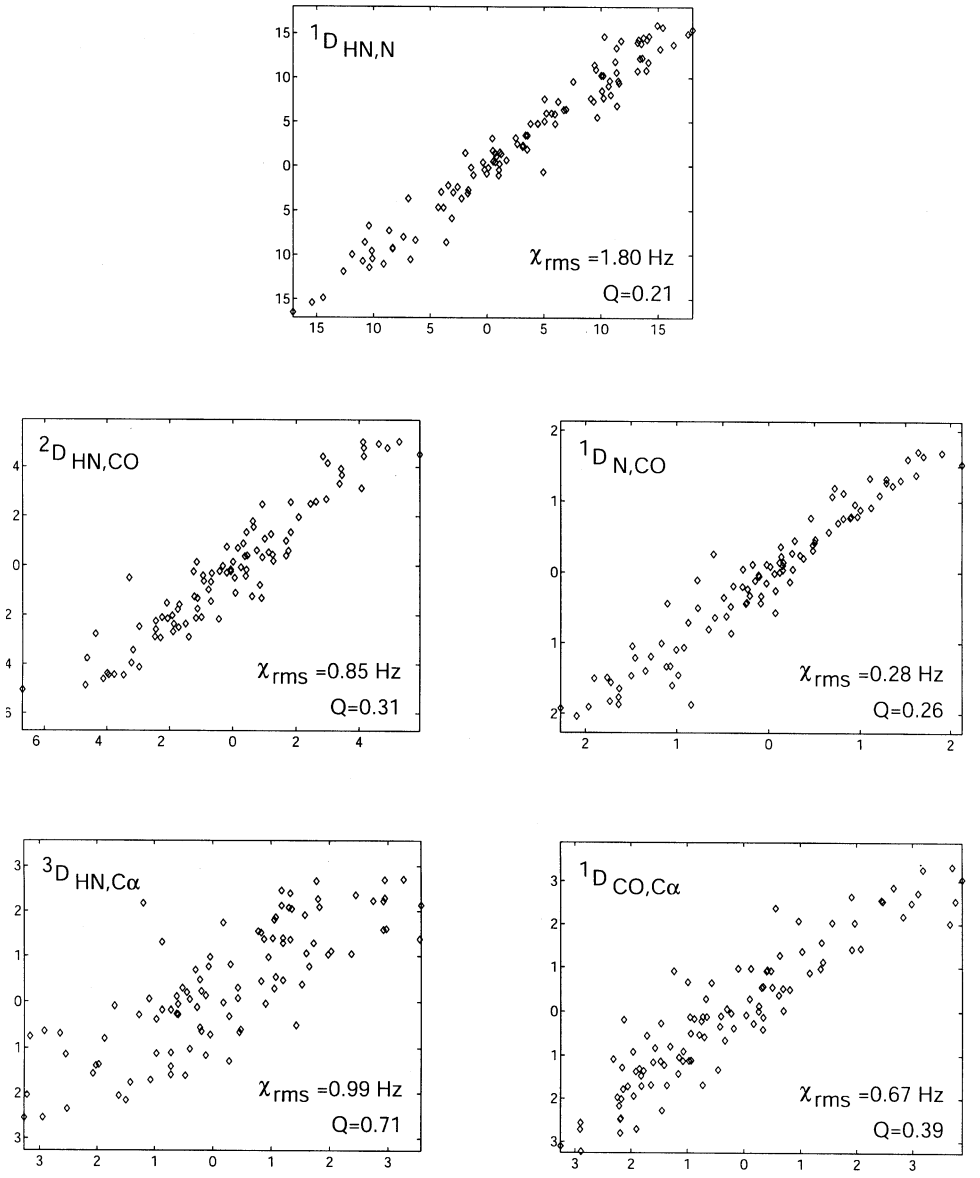

Fig. 6. Comparison of measured and X-ray predicted dipolar couplings from HCAII, including only elements of regular secondary structure. Values of the axial and rhombic components of the alignment tensor of −0.00080 and 0.603 were obtained; Euler angles of $(\alpha, \beta, \gamma) = (126°, 109°, 66°)$ describe the transformation from the X-ray co-ordinate system to the alignment frame. Reprinted with permission from [51].

The precision and accuracy of dipolar coupling data obtained from these experiments was determined. The reproducibility was evaluated by recording each data set twice (with and without phage). Fig. 5 shows the correlation between dipolar couplings measured from duplicate experiments recorded on a 1.6 mM HCA II sample. The agreement between repeat experiments is quite good as indicated by the pairwise rmsd values given in the figure. The accuracy was determined by comparing calculated D_{ij} values based upon the x-ray structure of this protein and the measured D_{ij} coupling for residues found to be in elements of regular

secondary structure. These data for HCA II are presented in Fig. 6 [51].

As described above, the goal of obtaining residual dipolar coupling data is to improve upon NMR solution structures. These can either be high-resolution structures achieved by conventional methods or global folds achieved using perdeuterated or perdeuterated/methyl-protonated samples and a limited set of NOE and dihedral angle restraints. The most straightforward method for incorporating these data into structure calculation protocols is by direct refinement of the orientation of individual bond vectors based upon the measured residual dipolar couplings [52]. In this case, the orientation of each individual bond vector is changed to satisfy dipolar couplings as structures are calculated. As hoped, direct refinement has been shown to improve the accuracy and precision of structures when used in conjunction with nearly complete sets of NOE, coupling constant and chemical shift data [52]. In addition, this procedure worked well for several small proteins with a limited set of NOEs but a well-defined topology [53]. Direct refinement has not been as successful for large perdeuterated proteins with a limited number of NOEs. Several groups have demonstrated that in such cases a large number of violations are observed and that the structures produced have high residual energy and large violations in the standard backbone geometry [21,54]. These groups have speculated that the violations are a result of the fact that for each individual dipolar coupling there are an infinite number of dipole vector orientations which are consistent with the data. All of the couplings taken together lead to an extremely complex energy surface making it difficult to determine a global minimum. Based on these observations a new approach was introduced for the use of dipolar couplings in cases of large proteins where a limited amount of other structural data is available [21]. This approach involves using the peptide plane as a fixed structural unit. For each plane the goal is to use dipolar coupling data to define the Euler angles (α, β, γ) which transform from an initial frame to the peptide alignment frame. Each peptide alignment frame must coincide with the global alignment coordinate system, effectively rotating each peptide into its correct orientation in the overall structure.

A more useful way to express the equation for residual dipolar coupling for proteins dissolved in an anisotropic media is given below where Eq. (3) has been rearranged so that contributions from dynamics are separated from those involving molecular orientation. In addition there is a transformation into the principal alignment frame [21].

$$D_{ij} = -[\gamma_i \gamma_j h/4\pi r^3)]A_a S[(3\cos^2\theta - 1) + (3/2)R\sin^2\theta\cos 2\phi] \quad (3)$$

In Eq. (3) θ and ϕ are now polar angles of the ij vector in the principal alignment frame of the molecule, S is the order parameter and A_a and R are the axial and rhombic components of the alignment tensor. From this equation it can be seen that generating structural restraints from dipolar coupling data for nuclei where r is known requires fitting the Euler angles and the order parameters A_a and R. Clore and coworkers demonstrated that A_a and R could be directly determined from the distribution seen in histograms plotted from dipolar coupling data sets [55]. Mueller and coworkers describe a method for obtaining the Euler angles (α, β, γ) for each peptide plane from a grid search that minimizes the difference between experimental data and dipolar couplings predicted from a trial alignment frame [21]. Based on these findings, utilizing the five dipolar couplings measured with the HNCO-based experiments described above, is more than sufficient to solve for the transformation of each peptide plane into the alignment frame.

The authors addressed two potential problems with this approach [21]. Firstly, was the choice of a suitable trial structure. This problem was solved by using the average structure, calculated from the lowest energy preliminary NMR structures determined solely from NOE and dihedral angle restraints, as the trial structure. Secondly, it is known that dipolar coupling data alone does not define a unique orientation of the peptide within its alignment frame. More specifically, there are eight possible orientations that are consistent with the dipolar data related by 180° rotations of the plane about any of the three alignment axes and by a further rotation of each of these structures by an additional 180° rotation about the plane normal. These eight orientations are depicted in Fig. 7 for the peptide plane connecting residues F149 and N150 in the MBP. The problem lies in choosing the proper orientation

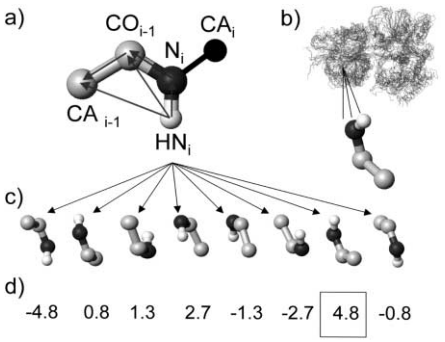

Fig. 7. Summary of the protocol used to choose between the eight possible orientations of peptide planes established on the basis of dipolar coupling data. Dipolar couplings are measured from the five dipole vectors illustrated in (a), resulting in eight possible peptide plane orientations (example shown for the plane bridging residues Phe149 and Asn150 in MBP). The eight orientations (c) are compared with the corresponding plane from the average structure (in the alignment frame) derived on the basis of NOE, dihedral and hydrogen bonding data (b) in order to select a set of restraints. The sum of the dot products of the five dipole vectors from each plane with the corresponding vectors from the average structure shown in (d), illustrating that structure 7 is the best match. Reprinted with permission from Ref. [21].

from the eight possibilities. To do this, the average preliminary structure was used to determine the molecular alignment frame by comparison with all of the measured dipolar couplings. Subsequently, the average structure was rotated into its alignment frame and the orientation of each peptide plane was extracted. Each plane extracted in this manner was then compared with the eight possible peptide orientations obtained from the residual dipolar coupling data. The correct peptide plane orientation chosen by this procedure was then used as an additional restraint in structure calculations. The results shown in Fig. 8, which compare the preliminary structure of MBP based solely on NOE and dihedral angle data with those structures that also include residual dipolar restraints determined as described, are extremely encouraging. This method allowed the global RMSD of the 10 lowest energy structures of MBP to be reduced from 5.5 to 2.2 Å [21]. The successful use of residual dipolar coupling restraints in the structure determination of large proteins and complexes offers exciting possibilities for advancement of the field to even more complex system than those that have been studied to date.

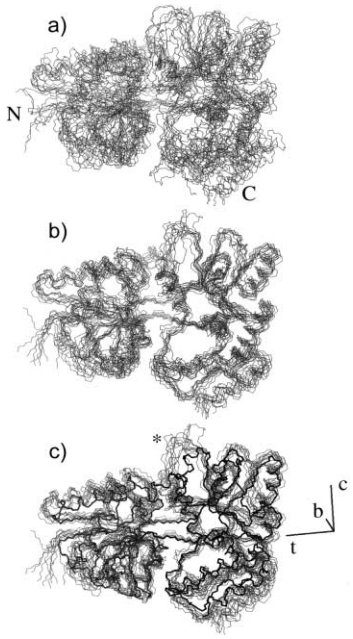

Fig. 8. Comparison of solution structures of the MBP/β-cyclodextrin complex obtained with and without dipolar coupling-based restraints. Best-fit (residues 6-370) superposition of the ten lowest energy structures generated from (a) NOE, dihedral angle, and hydrogen bond restraints (b) NOE, dihedral angle, hydrogen bond and dipolar restraints. In (c) the structures from (b) are aligned with the crystal structure of the MBP/β-cyclodextrin complex, 1dmb (black heavy line). The closure, twist, bend (c,t,b) co-ordinate frame is illustrated, with the bend axis extending out of the plane of the paper. Reprinted with permission from Ref. [21].

5. Proteins in low viscosity fluids

A particularly interesting idea to combat the limitations caused by a reduction in the T_2 times has been offered by Wand and co-workers [56]. Their approach centers on increasing the spin–spin relaxation time by directly reducing the correlation time of the protein. For an isotropically reorienting molecule, the Stokes-Einstein equation relates the tumbling time to the solvent viscosity

$$\tau_m = \frac{1}{6R_{diff}} = \frac{\eta V}{kT}$$

where τ_m is the rotational correlation time, R_{diff}, the spherical rotational diffusion constant, η, the solvent viscosity, V, the sphere volume, k, the Boltzmann constant and T is the temperature in Kelvin. As can be seen, there is a direct correlation between tumbling

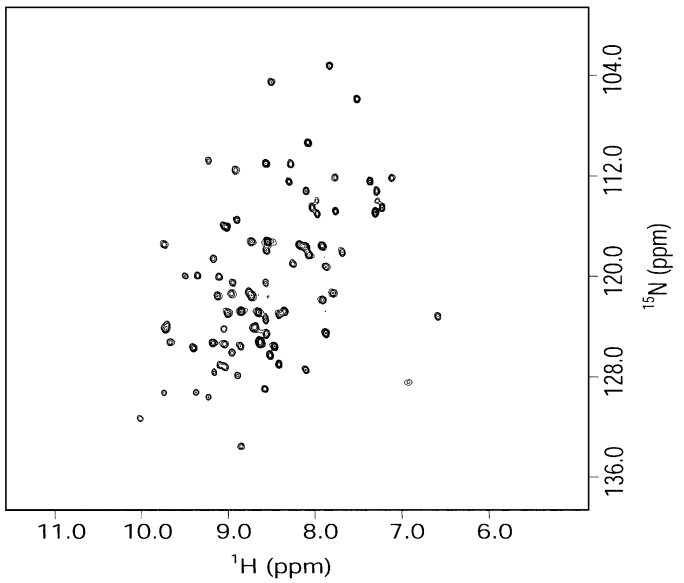

Fig. 9. Contour plot of the ^{1}H–^{15}N HSQC spectrum of recombinant human ubiquitin encapsulated in AOT reverse micelles dissolved in n-pentane. The protein was encapsulated at a concentration of ~0.25 mM in sodium acetate buffer (50 mM) containing 0.25 M NaCl. The spectrum was recorded at 400 MHz (^{1}H) using a cryogenic probe.

time and the solvent viscosity. Since T_2 is inversely proportional to τ_m, the lower the viscosity, the longer T_2 becomes. In order to lower the viscosity of the solvent Wand and his group explored the possibility of using short chain alkanes. In order to liquefy these solvents at room temperature moderate pressures are required. The resulting issues of dissolving a protein in such a hydrophobic solvent are cleverly overcome by encapsulating the protein within the water cavity formed by reverse micelles. Reverse micelles are groupings of surfactant molecules arranged around a water center and can be large enough to contain aqueous solutions of larger proteins.

The reverse micelle of choice in the Wand group is bis(2-ethylhexyl)sulfosuccinate (AOT). Encapsulation is an involved and delicate process due to the handling of flammable substances at elevated pressures. The process involves the preparation of empty micelles followed by loading with protein. Overall there are three distinct phases: liquification of the solvent, solvation of the AOT in the solvent and then transfer of the protein into the AOT-solvent phase by encapsulation. The main technical issue is in maintaining the pressure in order to keep the low viscosity solvent a fluid. Wand and co-workers have designed a custom NMR tube made from zirconium oxide (5 mm inner diameter) that can withstand the necessary pressure and also allows for liquid filling and out-gassing [57]. Fig. 9 shows an ^{1}H–^{15}N HSQC spectrum of the protein ubiquitin to illustrate the validity of the method. The protein is encapsulated in AOT reverse micelles dissolved in liquid propane.

The radius of the reverse micelle dictates the size of the protein that can be encapsulated. This radius is determined by the surfactant concentration and by the relative water content. Wand and co-workers provide a useful rule of thumb for the size of protein, the micelle radius and the effect of different solvents on correlation times:

Protein size	50 kDa	100 kDa
Reverse micelle radius	41 Å	48 Å
τ in water	60 ns	95 ns
τ in butane	11 ns	18 ns
τ in propane	7 ns	11 ns
τ in ethane	2.5 ns	4 ns

The improvement in tumbling characteristic is exceptional and the associated reduction in linewidths reflects this.

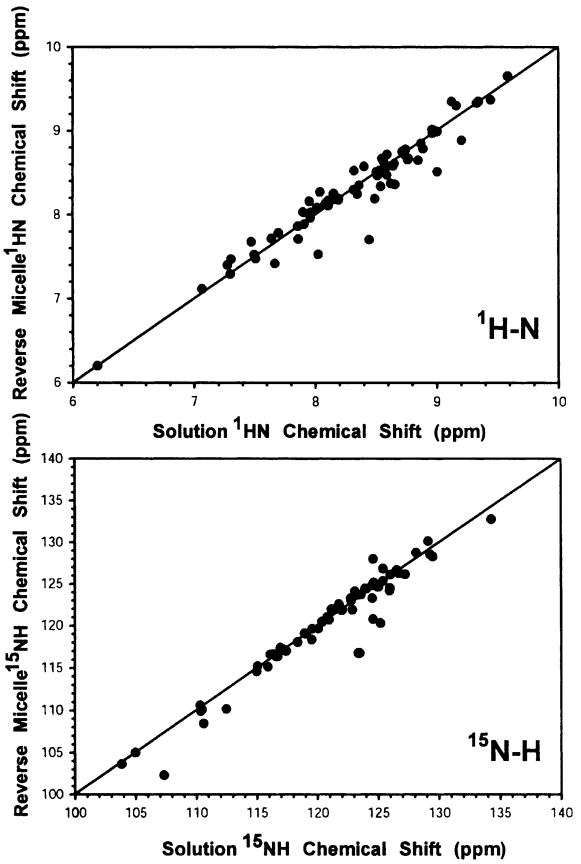

Fig. 10. Correlation of amide ^{1}H and ^{15}N chemical shifts of recombinant human ubiquitin in aqueous buffer with those of the protein encapsulated in AOT reverse micelles dissolved in *n*-pentane, as described in Fig. 9. These data were obtained at 750 MHz with a non-cryogenic probe. Reprinted with permission from Ref. [56].

An important concern to address is whether the encapsulation process and dissolution in a short chain alkane causes unusual artifacts or chemical shifts. Wand et al. compared the ^{1}H–^{15}N chemical shifts of ubiquitin in water and dissolved in AOT reverse micelles. Fig. 10 shows this comparison. Only two regions display minor perturbations, the C-terminus and a tight turn. Additionally, backbone NOE patterns are extraordinarily similar between the two samples. To finally dismiss any concerns in this respect, Wand and co-workers solved the structure of encapsulated ubiquitin and found it to be largely undisturbed when compared to its structure in water [58].

An unexpected bonus of this approach comes in the analysis of encapsulated proteins with cryogenic probe technology [59]. Cryogenic probes cool the receiver coil to cryo-temperatures thereby greatly reducing the noise voltage. The enhanced quality factor of the probe affords a significant increase in sensitivity. Unfortunately, cryogenic probes become somewhat more sensitive to high salt concentrations, often a necessity for acceptable protein spectra. Such sensitivity losses are, of course, due to the bulk solvent characteristics. As noted above, the use of encapsulated proteins in a low conductivity solvent removes the bulk solvent effect. Thus sensitivity losses in cryogenic probes at higher salt concentrations are efficiently overcome. Overall the technique of encapsulating proteins dissolved in low viscosity solvents offers a tremendous advantage by virtue of decreasing molecular tumbling time.

6. Summary

All the techniques described above have the goal of improving resolution and sensitivity for NMR studies of large proteins. The methods are certainly disparate in their approaches. The encapsulation of proteins in low viscosity solvents to directly increase T_2 requires dedicated high pressure hardware to cause a change in a physical property (τ_c); deuteration removes avenues of relaxation thereby increasing T_2; TROSY selects the optimal component of relaxation; segmental labeling, somewhat of an outlier since a physical property is not really altered, produces spectra that are greatly simplified; and dipolar couplings provide new long-range structural restraints that prove vital in those systems where there is a paucity of NOE information. Individually all of these methodologies represent significant advances towards improving the ability to address the complex problems inherent to NMR studies of larger protein systems; however, the ultimate power of these methods lies in their cumulative use. One approach does not preclude the use of another (with the possible exception of being able to extract dipolar couplings from encapsulated proteins), so that combining one or more these methods results in an additive advantage. It is likely that such combinations will allow the size of proteins amenable to study by NMR to reach the 100–200 kDa range.

Acknowledgements

We are indebted to the following people who graciously gave consent to use their figures: Josh Wand (Penn), Tom Muir and David Cowburn (Rockefeller), Pat Loria (Yale), Art Palmer (Columbia), Mark Rance (Cincinnati), Lewis Kay (Toronto), Bennett Farmer II (Bristol Myers Squibb). Portions of this work were supported by NIH RO1 GM55769 (JC), North Carolina State University College of Agriculture and Life Sciences (JC and RT). The Duke University NMR Center was established with grants from the NIH, NSF, and the North Carolina Biotechnology Center, which are gratefully acknowledged.

References

[1] J. Cavanagh, W.J. Fairbrother, A.G. Palmer, N.J. Skelton, Protein NMR Spectroscopy: Principles and Practice, Academic Press, San Diego, 1996.
[2] D.T. Browne, G.L. Kenyon, E.L. Packer, H. Sternlicht, D.M. Wilson, J. Am. Chem. Soc. 95 (1973) 1316.
[3] M.A. Markus, K.T. Dayie, P. Matsudairat, G. Wagner, J. Magn. Reson. Ser. B 105 (1994) 192.
[4] K.L. Constantine, L. Mueller, V. Goldfarb, M. Wittekind, W.J. Metzler, J. Yanchunas Jr., J.G. Robertson, M.F. Malley, M.S. Friedrichs, B.T. Farmer II, J. Mol. Biol. 267 (1997) 1223.
[5] D.S. Garrett, Y.-J. Seok, D.-I. Liao, A. Peterkofsky, A.M. Gronenborn, G.M. Clore, Biochemistry 36 (1997) 2517.
[6] B.T. Farmer II, R.A. Venters, J. Biomol. NMR 7 (1996) 59.
[7] X. Shan, K.H. Gardner, D.R. Muhandiram, N.S. Rao, C.H. Arrowsmith, L.E. Kay, J. Am. Chem. Soc. 118 (1996) 6570.
[8] T. Yamazaki, W. Lee, C.H. Arrowsmith, D.R. Muhandiram, L.E. Kay, J. Am. Chem. Soc. 116 (1994) 11,655.
[9] B.T. Farmer II, R.A. Venters, in: N.R. Krishna, L.J. Berliner (Eds.), Biological Magnetic Resonance, vol. 16:75, Kluwer Academic Publishers/Plenum Press, Dordrecht/New York, 1998.
[10] T. Yamazaki, R. Muhandiram, L.E. Kay, J. Am. Chem. Soc. 116 (1994) 8266.
[11] K.L. Constantine, M.S. Friedrichs, V. Goldfarb, P.D. Jeffrey, S. Sheriff, L. Mueller, Proteins: Struct., Func., Genet. 15 (1993) 290.
[12] D. Nietlispach, R.T. Clowes, R.W. Broadhurst, Y. Ito, J. Keeler, M. Kelly, J. Ashurst, H. Oschkinat, P.J. Domaille, E.D. Laue, J. Am. Chem. Soc. 118 (1996) 407.
[13] R.A. Venters, B.T. Farmer II, C.A. Fierke, L.D. Spicer, J. Mol. Biol. 264 (1996) 1101.
[14] P.E. Hansen, Prog. NMR Spectrosc. 20 (1988) 207.
[15] A. Bax, G.M. Clore, A. Gronenborn, J. Magn. Reson. 88 (1990) 425.
[16] W.J. Metzler, M. Wittekind, V. Goldfarb, L. Mueller, B.T. Farmer II, J. Am. Chem. Soc. 118 (1996) 6800.
[17] M.K. Rosen, K.H. Gardner, R.C. Willis, W.E. Parris, T. Pawson, L.E. Kay, J. Mol. Biol. 263 (1996) 627.
[18] B.O. Smith, Y. Ito, A. Raine, S. Teichmann, L. Ben-Tovim, D. Nietlispach, R.W. Broadhurst, T. Terada, M. Kelly, H. Oschkinat, T. Shibata, S. Yokoyama, E.D. Laue, J. Biomol. NMR 8 (1996) 360.
[19] L.E. Kay, T.E. Bull, L.K. Nicholson, C. Griesinger, H. Schwarbe, A. Bax, D.A. Torchia, J. Magn. Reson. 100 (1992) 538.
[20] N.K. Goto, K.H. Gardner, G.A. Mueller, R.C. Willis, L.E. Kay, J. Biomol. NMR 13 (1999) 369.
[21] G.A. Mueller, W.Y. Choy, D. Yang, J.D. Forman-Kay, R.A. Venters, L.E. Kay, J. Mol. Biol. 300 (2000) 197.
[22] G.W. Vuister, S.J. Kim, C. Wu, A. Bax, J. Am. Chem. Soc. 116 (1994) 9206.
[23] M.J.S. Kelly, C. Krieger, L.J. Ball, Y.H. Yu, G. Richter, P. Schmeider, A. Bacher, H. Oschkinat, J. Biomol. NMR 14 (1999) 79.
[24] F.B. Perler, Cell 92 (1998) 1.
[25] T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino, H. Nakamura, J. Am. Chem. Soc. 120 (1998) 5591.
[26] T. Otomo, K. Teruya, K. Uegaki, T. Yamazaki, Y. Kyogoku, J. Biomol. NMR 14 (1999) 105.
[27] T. Otomo, N. Ito, Y. Kyogoku, T. Yamazaki, Biochemistry 38 (1999) 16040.
[28] R. Xu, B. Ayers, D. Cowburn, T.W. Muir, Proc. Natl. Acad. Sci. USA 96 (1999) 388.
[29] U. Blaschke, G.J. Cotton, T.W. Muir, Tetrahedron 56 (2000) 9461.
[30] K. Pervushin, R. Riek, G. Wider, K. Wuthrich, Proc. Natl. Acad. Sci. USA 94 (1997) 12366.
[31] M. Goldman, J. Magn. Reson. 60 (1984) 437.
[32] K. Pervushin, R. Riek, G. Wider, K. Wuthrich, J. Am. Chem. Soc. 120 (1998) 6394.
[33] K. Pervushin, G. Wider, K. Wuthrich, J. Biomol. NMR 12 (1998) 345.
[34] M. Rance, J.P. Loria, A.G. Palmer III, J. Magn. Reson. 136 (1) (1999) 92.
[35] M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wuthrich, Proc. Natl. Acad. Sci. USA 95 (1998) 13585.
[36] J.P. Loria, M. Rance, A.G. Palmer III, J. Magn. Reson. 141 (1999) 180.
[37] J.P. Loria, M. Rance, A.G. Palmer III, J. Biomol. NMR, 15 (1999) 151.
[38] D. Yang, L.E. Kay, J. Biomol. NMR 13 (1999) 3.
[39] D. Yang, L.E. Kay, J. Am. Chem. Soc. 121 (1999) 2571.
[40] R. Konrat, D. Yang, L.E. Kay, J. Biomol. NMR 15 (1999) 309.
[41] F.A. Mulder, A. Ayed, D. Yang, C.H. Arrowsmith, L.E. Kay, J. Biomol. NMR 18 (2) (2000) 173.
[42] R. Riek, G. Wider, K. Pervushin, K. Wurthrich, Proc. Natl. Acad. Sci. USA 96 (1999) 4918.
[43] G. Wider, K. Wuthrich, Curr. Opin. Struct. Biol. 9 (1999) 594.
[44] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Proc. Natl. Acad. Sci. USA 92 (1995) 9279.
[45] C. Gayathri, A.A. Bothner-By, P.C.M. Van Zijl, C. Maclean, Chem. Phys. Lett. 87 (1982) 192.

[46] N. Tjandra, S. Grzesiek, A. Bax, J. Am. Chem. Soc. 118 (1996) 6264.
[47] N. Tjandra, A. Bax, Science 278 (1997) 1111.
[48] G.M. Clore, M.R. Starich, A.M. Gronenborn, J. Am. Chem. Soc. 120 (1998) 10571.
[49] M.R. Hansen, L. Mueller, A. Pardi, Nat. Struct. Biol. 5 (1998) 1065.
[50] Y-X. Wang, J.L. Marquardt, P. Wingfield, S.J. Stahl, S. Lee-Huang, D. Torchia, A. Bax, J. Am. Chem. Soc. 120 (1998) 7385.
[51] D. Yang, R.A. Venters, G.A. Mueller, W.Y. Choy, L.E. Kay, J. Biomol. NMR 14 (1999) 333.
[52] G.M. Clore, A.M. Gronenborn, N. Tjandra, J. Magn. Reson. 131 (1998) 159.
[53] G.M. Clore, M.R. Starich, C.A. Bewley, M. Cai, J. Kuszewski, J. Am. Chem. Soc. 121 (1999) 6513.
[54] M.W.F. Fischer, J.A. Losonczi, J.L. Weaver, J.H. Prestegard, Biochemistry 38 (1999) 9013.
[55] G.M. Clore, A.M. Gronenborn, A. Bax, J. Magn. Reson. 133 (1998) 216.
[56] A.J. Wand, M.R. Ehrhardt, P.F. Flynn, Proc. Natl. Acad. Sci. USA 95 (1998) 15299.
[57] M.R. Ehrhardt, P.F. Flynn, A.J. Wand, J. Biomol. NMR 14 (1999) 75.
[58] C.R. Babu, P.F. Flynn, A.J. Wand, J. Am. Chem. Soc. JACS (2001) 2691.
[59] P.F. Flynn, D.L. Matiello, H.D.W. Hill, A.J. Wand, J. Am. Chem. Soc. 122 (2000) 4823.