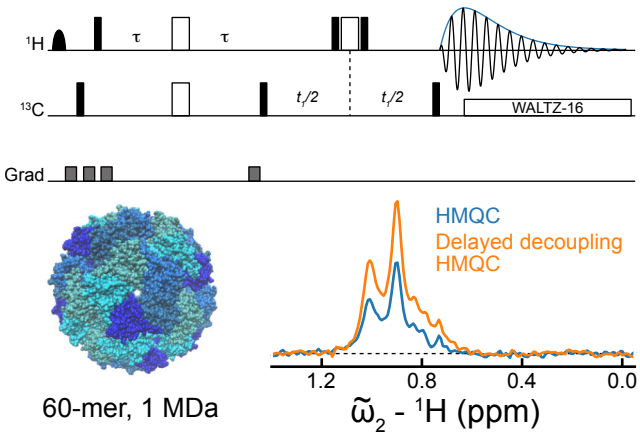
Less is more: A simple methyl-TROSY based pulse scheme offers improved sensitivity in applications to high molecular weight complexes

Nicolas Bolik-Coulon[1-3], Alex I.M. Sever[2], Robert W. Harkness[1-3], James Aramini[1], Yuki Toyama[1-3], D. Flemming Hansen[4], Lewis E Kay[1-3,5]

1) Department of Molecular Genetics, University of Toronto, Toronto, Canada, M5S 1A8
2) Department of Chemistry, University of Toronto, Toronto, ON, Canada, M5S 3H6.
3) Department of Biochemistry, University of Toronto, Toronto, ON, Canada, M5S 1A8.
4) Department of Structural and Molecular Biology, Division of Biosciences, University College London, London, United Kingdom, WC1E 6BT
5) Program in Molecular Medicine, Hospital for Sick Children Research Institute, Toronto, ON, Canada, M5G 0A4.

Keywords: HMQC, Delayed decoupling, Sensitivity gains, MDa protein complexes



Highlights

- Shortened HMQC pulse sequence
- Improved sensitivity for methyl-TROSY based experiments
- Applications to complexes with molecular weights in the MDa range

Abstract

The HMQC pulse sequence and variants thereof have been exploited in studies of high molecular weight protein complexes, taking advantage of the fact that fast and slow relaxing magnetization components are sequestered along two distinct magnetization transfer pathways. Despite the simplicity of the HMQC scheme an even simpler version can be designed, based on elimination of the terminal refocusing period, as a further means of increasing signal. Here we present such an experiment, and show that significant sensitivity gains are realized in studies of proteins varying in molecular masses from 100 kDa to 1 MDa, in some cases by factors of two or more.

Introduction

Despite the tremendous progress in solution NMR spectroscopy over the past several decades (1,2), which has been fueled by advances in hardware, pulse sequences, processing methods, and sample preparation, sensitivity remains a central issue. This is particularly the case for applications to high molecular weight biomolecules where rapid relaxation of signals and spectral overlap are two formidable challenges. The development of TROSY-based approaches, first focusing on backbone amide ^{15}N-$^{1}H^{N}$ spin pairs (3), and subsequently on methyl groups (4), has extended the range of biomolecules that can be studied, offering new insights into the role of structure and dynamics in biomolecular function and misfunction (1,5). Our laboratory has had a longstanding interest in the functional dynamics of large complexes such as oligomeric proteases and unfoldases that regulate protein homeostasis in the cell and in the nucleosome core particle that is the fundamental building block of chromatin. Complexes on the order of 1 MDa have been studied (6) and the often high quality spectra generated allows detailed and quantitative analyses that are normally only reserved for applications to much smaller protein systems. However, even when reasonable spectra are obtained, there is always a distribution of peak intensities, with some correlations much weaker than others. It would be advantageous in these cases, and in applications to cellar complexes that are on the order of or larger than 1 MDa, especially, to develop approaches that increase spectral sensitivity even further.

The HMQC pulse scheme, a simple, yet powerful, experiment for correlating one bond coupled X-^{1}H spins (7,8) has found a prominent role in studies of large protein complexes when applied to methyl groups, as the transfer of magnetization involves a pair of pathways that comprise either fast or slow relaxing components (4). In particular, labeling with $^{13}CH_3$ methyls in an otherwise deuterated protein (9) or nucleic acid (10) background minimizes mixing of the two pathways, with the slowly relaxing pathway dominating in applications to high molecular weight systems. We wondered if it would be possible to improve the sensitivity of the HMQC experiment still further by simply shortening it. In principle, this can be achieved by eliminating the final element that is responsible for refocusing anti-phase magnetization into in-phase ^{1}H magnetization prior to direct detection with X-nucleus decoupling. The advantage is clear. In studies of very

high molecular weight systems, removal of refocusing elements can lead to significant enhancements in signal-to-noise, as illustrated below. A disadvantage is that decoupling during acquisition must be delayed until refocusing is complete, leading to spectra with distorted lineshapes that are not suitable for conventional quantitative analysis and necessitating post-acquisition manipulation to generate spectra with improved (*i.e.*, Lorentzian) peak shapes. Motivated by a recent paper introducing XL-ALSOFAST-HMQC for recording non-TROSY based HMQC spectra in fully protonated proteins that used a shortened refocusing element in combination with delayed decoupling (11) we asked whether suitable spectra could also be obtained by removing the element altogether and what the best approach might be for eliminating the distortions in the resulting lineshapes. Here we have examined spectra of a number of different high molecular weight protein systems that have been processed using a simple *J*-deconvolution procedure to remove peak distortions. These include the HtrA2 protease (equilibrium between 105 kDa trimer and a 210 kDa hexamer) (12), the nucleosome core particle (NCP, 210 kDa) (13), the half-proteasome ($\alpha_7\alpha_7$, 360 kDa) (14), and lumazine synthase (AaLS, 1 MDa) (15); in some cases spectra have been recorded at temperatures as low as 7^{o}C to increase global tumbling correlation times. Significant sensitivity gains are achieved for weak signals.

Materials and Methods

Protein expression and purification

Protein L, HtrA2, NCP, and $\alpha_7\alpha_7$ samples were generated as described previously (10,12,14,16). Sample conditions were: 1.4 mM protein L, 99.9% D_2O, 50 mM sodium phosphate, pH 6.0 (uncorrected); 0.30 mM HtrA2 (protomer concentration), 20 mM HEPES-NaOH (pD 7.4), 1 mM EDTA, 99.9% D2O; 0.15 mM NCP, 20 mM sodium phosphate, pD 6, 0.05% NaN_3, 99.9% D2O; 1 mM $\alpha_7\alpha_7$ (protomer concentration), 25 mM potassium phosphate, pH 6.8 (uncorrected), 50 mM NaCl, 1 mM EDTA, 0.03% NaN_3 and 2 mM DTT, 99.9% D_2O.

In order to produce a sample of lumazine synthase a gBlock™ gene fragment containing the full-length (1-154) *Aquifex aeolicus* lumazine synthase (AaLS) gene was synthesized by IDT (codon-optimized). It was placed into a *Champion* pET vector

(kanamycin resistance) with no N-terminal affinity purification fusion tag using Gibson Assembly® (New England BioLabs Inc.). The coding portion of the plasmid was confirmed via DNA sequencing. The AaLS plasmid was transformed into *E. coli* BL21-CodonPlus(DE3)-RIPL competent cells via heat-shock. Cells were grown in D_2O M9 minimal media, using [^2H,^{12}C]-glucose as the sole carbon source. The sample was labeled as U-^2H, Ile-δ1-$^{13}CH_3$, Leu,Val-$^{13}CH_3$ (only one of the two methyls of Leu,Val was NMR labeled) and Met-Cε-$^{13}CH_3$, as described previously (9,17), achieved with the addition of precursors (60 mg.L^{-1} α-ketobutyric acid, 100 mg.L^{-1} α-ketoisovaleric acid, and 100 mg.L^{-1} ε-$^{13}CH_3$ methionine) 1 hour prior to induction of protein expression. Cells were grown at 37°C until an OD_{600} of ~0.8 was achieved, at which point protein synthesis was induced by adding 0.2 mM IPTG. The cells were then allowed to grow for a further ~20 hours at 25°C. Subsequently, the cells were harvested via centrifugation and resuspended in lysis buffer (20 mM Tris•HCl, pH 8.0, + 1 mg.ml^{-1} lysozyme).

Purification of AaLS consisted of first lysing the cells by ultrasonication, followed by centrifugation at 9000 rpm for 15 minutes. The supernatant was filtered (0.45μm pore size) and loaded onto a HiTrap® Q HP column equilibrated with 20 mM Tris•HCl (pH 8.0). AaLS was eluted using a 0-100% gradient of 20 mM Tris•HCl, 1 M NaCl (pH 8.0) (15). Fractions which contained AaLS were collected and $(NH_4)_2SO_4$ was added to a concentration of 1 M. The solution was loaded onto a HiTrap® Butyl HP column equilibrated with 25 mM NaH_2PO_4, 1 M $(NH_4)_2SO_4$, 1 mM EDTA (pH 7.0). AaLS was eluted using a buffer of 25 mM NaH_2PO_4, 500 mM $(NH_4)_2SO_4$, 1 mM EDTA (pH 7.0). Fractions containing AaLS were collected and diluted 2-fold with 25 mM NaH_2PO_4, 200 mM NaCl, 1 mM EDTA (pH 7.8). The solution was then heated to 80°C for 20 minutes to induce unfolding and aggregation of protein impurities, and subsequently centrifuged at 9000 rpm for 15 minutes to clear these aggregates (15). The resulting supernatant was concentrated, and final purification of AaLS was achieved using a Superose 6 10/300 size-exclusion column equilibrated with 25 mM NaH_2PO_4, 200 mM NaCl, 1 mM EDTA (pH 7.8). Fractions containing AaLS were buffer exchanged into 99.9% D_2O, 25 mM NaH_2PO_4, 100 mM NaCl, 1 mM EDTA, pD 7.5 using an Amicon Ultra-15 concentrator (MWCO = 30 KDa) to produce a 0.2 mM NMR sample (protomer concentration). The concentration

of AaLS was measured using a NanoDrop spectrophotometer and the extinction coefficient (13,980 $M^{-1}.cm^{-1}$) from ExPASy's ProtParam tool (https://web.expasy.org/protparam/).

NMR spectroscopy

All NMR spectra were recorded on a 1 GHz Bruker Ascend Neo spectrometer equipped with a cryogenically cooled *x*,*y*,*z*, pulsed field gradient triple-resonance probe. HMQC spectra were recorded with the number of scans, acquisition times in t_1 and experimental times as listed below. The delay τ for magnetization transfer (Figure 1) was set to 1.8 ms unless otherwise specified. The following are acquisition details for each spectrum: HtrA2, 24 scans (25°C, t_{1max}=21.9 ms, experimental time is 2.4 hrs) and 8 scans (40°C, t_{1max}=21.9 ms, experimental time is 49 min); NCP, 32 scans (25°C, t_{1max}=19.9 ms, experimental time is 2.9 hrs and 7°C, t_{1max}=17.9 ms, experimental time is 2.6 hrs); $\alpha_7\alpha_7$, 32 scans (7°C, t_{1max}=17.9 ms, experimental time is 2.6 hrs) and 8 scans (40°C, t_{1max}=21.9 ms, experimental time is 49 min); AaLS, 32 scans (40°C, t_{1max}=17.3 ms, τ=1.6 ms, experimental time is 2.5 hrs) and 40 scans (25°C, t_{1max}=14.9 ms, τ=1.5 ms, experimental time is 2.7 hrs). Data were processed using the *nmrPipe* suite of programs (18), with peak intensities obtained using *Peakipy* (https://github.com/j-brady/peakipy). ^{1}H $R_{2,S}$ values were recorded as described previously (19).

Results and Discussion

Lineshape of Fourier Transformed Delayed Decoupled FID

Figure 1a shows the pulse scheme that has been used to record delayed decoupled HMQC spectra (ddHMQC) of U-^{2}H, $^{13}CH_3$-labeled proteins. The method is very similar to the 'standard' HMQC except that signal acquisition begins immediately after the final 90° ^{13}C pulse, without first refocusing the anti-phase ^{1}H magnetization. As a result, decoupling is delayed by $1/(2J_{CH})$ from the start of direct detection, where J_{CH} is the one-bond ^{1}H-^{13}C scalar coupling constant, approximately 125 Hz for all methyl groups (124.5±0.8 Hz for Leu, 125.2±0.3 Hz for Ile-δ1, and 126.2±0.3 Hz for Val) (20) with the exception of Met where $J_{CH} \approx$140 Hz. As described previously (11), the delayed decoupling approach leads

to a distorted lineshape in the ^{1}H dimension of ^{13}C-^{1}H spectra that can be calculated as the real part of the Fourier transform of the FID illustrated in Figure 1b (black), where an on-resonance signal has been assumed with a decay constant, R_2 of 100 s^{-1} and J_{CH} = 125 Hz. In order to simplify the calculation $S(t)$ is divided into three time domains, $S_i(t)$, $i \in \{1,2,3\}$, such that $S(t) = S_1(t)-S_2(t)+S_3(t)$, Figure 1b (middle), whose Fourier transforms can be calculated in a straightforward manner, as shown below. It follows that

$$F(\omega) = F_1(\omega) - F_2(\omega) + F_3(\omega) \qquad [1.1]$$

where

$$F_1(\omega) = \int_0^{\infty} S_1(t)\exp(-i\omega t)dt = \int_0^{\infty} \exp(-R_2 t)\exp(-i\omega t)dt$$

$$F_2(\omega) = \int_0^{\frac{1}{2J_{CH}}} S_2(t)\exp(-i\omega t)dt = \int_0^{\frac{1}{2J_{CH}}} \exp(-R_2 t)\exp(-i\omega t)dt \qquad . \qquad [1.2]$$

$$F_3(\omega) = \int_0^{\frac{1}{2J_{CH}}} S_3(t)\exp(-i\omega t)dt = \int_0^{\frac{1}{2J_{CH}}} \sin(\pi J_{CH} t)\exp(-R_2 t)\exp(-i\omega t)dt$$

Evaluation of Eq. [1.2] gives

$$F_1(\omega) = \frac{R_2}{R_2^2 + \omega^2}$$

$$F_2(\omega) = \frac{R_2}{R_2^2 + \omega^2} + \frac{[-R_2\cos(\frac{\omega}{2J_{CH}}) + \omega\sin(\frac{\omega}{2J_{CH}})]\exp(-\frac{R_2}{2J_{CH}})}{R_2^2 + \omega^2}$$

$$F_3(\omega) = \frac{1}{2}\left\{\frac{\omega - \pi J_{CH}}{D_1}[\sin(\frac{\omega}{2J_{CH}})\exp(-\frac{R_2}{2J_{CH}}) - 1] - \frac{R_2\cos(\frac{\omega}{2J_{CH}})\exp(-\frac{R_2}{2J_{CH}})}{D_1}\right\} \qquad [2.1]$$

$$+ \frac{1}{2}\left\{\frac{\omega + \pi J_{CH}}{D_2}[\sin(\frac{\omega}{2J_{CH}})\exp(-\frac{R_2}{2J_{CH}}) + 1] - \frac{R_2\cos(\frac{\omega}{2J_{CH}})\exp(-\frac{R_2}{2J_{CH}})}{D_2}\right\}$$

where

$$D_1 = (\omega - \pi J_{CH})^2 + R_2^2$$
$$D_2 = (\omega + \pi J_{CH})^2 + R_2^2 \quad . \qquad [2.2]$$

Figure 1b (bottom) plots $F_i(\nu)$, along with $F(\nu)$ ($\omega = 2\pi\nu$), with each frequency domain response color coded as for the corresponding time domain signal. $F_1(\nu)$ is a perfect Lorentzian line, since $S_1(t)$ is an exponentially decaying signal that extends to $t = \infty$, but $F_2(\nu)$ and $F_3(\nu)$ are more complex as they contain sinc wiggles, resulting from the truncation of each of the time domains at $t = 1/(2J_{CH}) = 4$ ms. In addition to the truncation artifacts, the $F_3(\nu)$ lineshape is further distorted by the fact that it is comprised of two multiplet components (at frequencies of $\pm J_{CH}/2$ Hz), rather than single lines at 0 frequency for each of $F_1(\nu)$ and $F_2(\nu)$, with each of the two lines antiphase dispersive, as compared to the inphase absorptive components at $\nu = 0$ for $F_1(\nu)$ and $F_2(\nu)$. The multiplet components of $F_3(\nu)$ are broad, as their lifetimes extend to only 4 ms, effectively giving rise to negative lobes on either side of a central peak along with additional wiggles that are due to truncation. Figure 1b (*right*) also shows $F(\nu)$ (black) that clearly illustrates the non-Lorentzian character of the resultant lineshape that is a linear combination of the three contributions, $F_i(\nu)$.

The lineshapes that are produced by the delayed decoupling approach are clearly not optimal as the broad negative tails from each peak will interfere with adjacent resonances (see below). The data can be manipulated in the time domain by dividing the FID by $\sin(\pi J_{CH} t)$ for $0 < t < 1/(2J_{CH})$ so that $S(t)$ is transformed in a post-acquisition manner into $S_1(t)$, resulting in a pure Lorentzian lineshape after Fourier transformation of the time domain. This is an example of *J*-deconvolution, a procedure that dates back to work by Bothner-By and Dadok (21) who selected a doublet from a frequency domain spectrum, inverse transformed the spectral region to the time domain and then divided by $\cos(\pi J t)$ in this case to remove the scalar modulation, as is effectively done here. Simulations that we have done suggest that the deconvolution procedure is quite robust with respect to variability in J_{CH}. For example, a comparison of simulated data with $J_{CH} =$ 140 Hz (*i.e.*, J_{CH} for Met methyl groups), followed by *J*-deconvolution using $J_{CH} = 125$ Hz,

(*i.e.*, J_{CH} for Ile, Leu and Val methyl groups), and Fourier transformation shows that the mismatch leads to a small increase in apparent linewidth, varying from 1%-6% for R_2 values between 10 s^{-1} and 150 s^{-1}.

A similar time domain manipulation has also been used recently in the context of the XL-ALSOFAST HMQC experiment (11), that is optimized for studies of fully protonated proteins with gradient coherence transfer selection, where delayed decoupling was also employed. In this scheme, however, direct detection does not proceed immediately after the final ^{13}C 90° pulse, but awaits the application of a gradient for coherence selection, so that the FID in this case is a mixture of in-phase and anti-phase magnetization at the start of direct acquisition.

A clear limitation of the *J*-deconvolution method, as pointed out in the original paper describing the approach (21), emerges from the near zero values of the signal that result from the sine/cosine modulations introduced by scalar coupled evolution. It, therefore, becomes necessary to divide by numbers close to zero at early positions in the FID, that can result in large noise spikes. In some applications it is possible to judiciously choose acquisition times to avoid zero-crossings of magnetization. In the present case there are no zero-crossings as decoupling commences when the signal becomes in-phase at $t = 1/(2J_{CH})$ from the start of acquisition, however the signal at $t = 0$ is zero and it is, therefore, not possible to restore the first point using this procedure. This results in a DC offset for traces in the ^{1}H dimension that can be 'fixed' via standard baseline correction in the frequency domain, or by backwards linear prediction of the first point of the *J*-deconvoluted spectrum. A processing script that is based on the well-known *nmrPipe* software package (18) combined with an in-house written python based (22,23) *J*-deconvolution program is provided in Supporting Information. Notably, division by $\sin(\pi J_{CH}t)$ for small t values does increase the noise floor, in the applications considered here by 5-15%, but this is significantly less than the sensitivity gains that can be realized in studies of molecular machines where R_2 values are typically very large (see below).

Figure 2 highlights expanded regions from ddHMQC spectra recorded on a highly deuterated nucleosome core particle (NCP) sample comprising two copies each of four histones (H2A, H2B, H3, and H4) along with a 147 basepair DNA fragment (13), where histone 2B (H2B) is labeled as Ile-δ1-$^{13}CH_3$, Leu,Val-$^{13}CH_3$ (referred to as ILV-$^{13}CH_3$-, or

ILVM-$^{13}CH_3$ when Met-Cε-$^{13}CH_3$ is labeled in what follows) and the other histones are NMR silent (fully deuterated; a,b) and on a ILVM-$^{13}CH_3$ $\alpha_7\alpha_7$ half proteasome (c-f) sample at a spectrometer field of 1 GHz. Only one of the two isopropyl methyl groups of Leu and Val is labeled as $^{13}CH_3$ (non-stereospecifically), while the other is $^{12}CD_3$, to minimize proton cross-relaxation pathways (9). In addition, the DNA of the NCP sample was deuterated and methyl groups added at C5 and A6 base positions (10). Spectra were recorded at 7°C (in addition to 40°C for $\alpha_7\alpha_7$) to slow down the overall tumbling of the complexes. On the basis of previous estimates of the assumed isotropic correlation times of NCP and $\alpha_7\alpha_7$ complexes of 115 ns (45°C) (10) and 120 ns (50°C) (14), respectively, correlation times of 340 ns and 390(150) ns are calculated for NCP and $\alpha_7\alpha_7$ at 7°C(40°C). Comparisons of spectra a and b (NCP), or c,d and e,f ($\alpha_7\alpha_7$), show that the *J*-deconvolution approach removes the tails of the lines, improving both resolution and sensitivity.

In order to quantify the sensitivity gains that are obtained by removing the final refocusing period in the ddHMQC pulse scheme, followed by *J*-deconvolution of the time domain data we have first recorded standard HMQC and ddHMQC spectra on a sample of ILV-$^{13}CH_3$ protein L, a 7.5 kDa protein with a correlation time of 5.3 ns at 25°C (24). Given the small size of the protein no sensitivity gain would be expected in the ddHMQC dataset over the HMQC correlation map. Figure 3a compares intensity ratios of crosspeaks recorded using HMQC and ddHMQC schemes (1 GHz) and, as expected, the relative intensities of correlations are similar. Our control studies on protein L pointed out an issue with the efficacy of heteronuclear decoupling in the case of the ddHMQC. While ^{13}C decoupling is efficient when the signal is in-phase at the start of acquisition, we have found that modulation sidebands (on the order of 1% or less of the main peak) are produced with delayed decoupling. These can be observed for small proteins in spectra recorded with high signal to noise (10,000:1, 4 scans in the case of protein L). We have examined a number of different decoupling sequences, including bi-level adiabatic decoupling with different adiabatic shapes (25), however the best performance was obtained using WALTZ-16 (26) (3.3 kHz B_1 field at 1 GHz). Application of a ^{13}C 90° purge pulse or a ^{13}C 180° pulse in alternate scans with no change in the receiver phase immediately prior to turning on the decoupler (*i.e.*, at $t_2 = 1/(2J_{CH})$ from the start of acquisition) so as to ensure only in-phase magnetization, did not improve the situation. Notably, such artifacts were not observed in

any of the applications to high molecular weight systems (see below) with the exception of crosspeaks from I0δ1 and I19δ1 in spectra recorded of the nucleosome particle at 37°C and 25°C (but not at 7°C) that are both associated with the very flexible unfolded N-terminal tail of histone H2B. These, again, were less than 1% of the main correlation.

While no sensitivity gain was observed for protein L, the situation, in contrast, is quite different for larger molecules. Panels b-d of Figure 3 plot the ratios of peak intensities from ddHMQC and HMQC datasets as a function of the transverse relaxation rates of the slow component of methyl ^{1}H magnetization, $R_{2,S}$ (circles) for a number of different protein systems and temperatures. These include, HtrA2, a mitochondrial protease in equilibrium between trimeric (105 kDa) and hexameric conformers (25°C,b; correlation time of the trimer approximately 110 ns (12)), 210 kDa NCP particles (7°C and 25°C,c), and 360 kDa $\alpha_7\alpha_7$, a dimer of disk shaped heptameric 21 kDa protomers (7°C and 40°C,d). Also highlighted are the three-dimensional structures of each complex. The solid line in each plot is the calculated expected intensity ratio after *J*-deconvolution of the ddHMQC data, given by $\exp(R_{2,S}2\tau)$, which takes into account the decay of magnetization during the final refocusing element of the HMQC scheme that is absent in the ddHMQC sequence. In the calculation we have assumed that the fast relaxing component of the magnetization, corresponding to the anti-TROSY component, decays to zero during the first 2τ element in the pulse scheme and the successive t_1 evolution period (see below). It is clear that the sensitivity gain grows as a function of $R_{2,S}$, and hence molecular weight.

As indicated above, both NCP and $\alpha_7\alpha_7$ spectra were recorded at 7°C to increase their tumbling times as a mimic for even bigger structures. In our experience lowering the temperature often deteriorates spectral quality, not necessarily only because of the decreased tumbling rate, but also resulting from μs-ms timescale dynamics that are no longer averaged out at the lower temperatures. With this in mind we sought another system of increased molecular weight that could be studied at physiological temperatures. We have chosen lumazine synthase (AaLS), an enzyme that catalyzes the penultimate step in the biosynthesis of vitamin B2 that is strictly required in the diet of mammals (15). AaLS consists of 60 copies of 16 kDa protomers, giving rise to a virus-like spherical particle of approximately 1 MDa in molecular mass, with an assumed isotropic rotational correlation time that is calculated to be 370 ns and 540 ns at 40°C and 25°C, respectively, in D_2O

solvent from HYDROPRO (27,28). Figures 4a-d illustrate spectral regions of ILVM-$^{13}CH_3$ AaLS (1 GHz, 40°C) from both HMQC (a,c) and ddHMQC (b,d) datasets contoured at the same level. It is clear that the ddHMQC data includes correlations that are absent from the HMQC, and that there are very significant gains in sensitivity for a large number of peaks (compare traces in Figures 4a-g), with an average signal gain of 1.6 (Figure 4h). Larger gains are obtained when the global tumbling is slowed down by lowering the temperature to 25°C (Figure 4i-l).

A comparison of HMQC and ddHMQC datasets of HtrA2 (protomers of 324 residues), recorded at 40°C, 1 GHz and using a sample where only the *proR* methyl groups are $^{13}CH_3$ labeled (29), illustrates an interesting feature of the ddHMQC experiment, Figure 5. In particular, when plotted at similar levels, ddHMQC spectra appear less well resolved than their HMQC counterparts. This reflects the fact that ddHMQC peak intensities are higher than those in HMQC spectra. Assuming that peak heights are, on average, k times larger, for example, the ratio of ddHMQC to HMQC linewidths close to the noise floor (*i.e.*, at much lower intensities than maximum) is approximately $k^{0.5}$ (see Supporting Information), leading to an apparent broadening of the ddHMQC correlations. A second contributing factor is that by shortening the HMQC pulse scheme to generate the ddHMQC it may well be that for "relatively small" proteins on the order of 100 kDa, such as HtrA2, the fast relaxing component is not completely eliminated during the abbreviated pulse scheme. Although the dominant component in spectra would remain the slowly relaxing one, the faster decaying component would contribute to the linewidth. This is illustrated in Figure 5 where HMQC (a), ddHMQC (b), and purged ddHMQC (c) spectra for HtrA2 are highlighted (correlation time of approximately 70 ns at 40°C), plotted at identical levels. In the case of the purged dataset (see Supporting Information) the fast relaxing magnetization is actively eliminated (Figure S1), as described previously (30), giving rise to what appears to be a better resolved spectrum. However, the purge element, of duration $1/(4J_{CH})$, does degrade sensitivity because it both suppresses the fast relaxing component which may not decay to zero in this case, hence eliminating a contribution to the peak intensity, and also because of the additional delay in the sequence for purging (see Supporting Information) which affects the sensitivity of the slow relaxing component as

well. As a result, the ratio of intensities of the purged ddHMQC to the HMQC is less than one in this case (0.94±0.16). In general, therefore, the purge is not recommended.

As described above, significant sensitivity gains can be realized in the ddHMQC experiment, especially in applications to MDa complexes. In the case of AaLS, correlations that could not be observed in HMQC spectra (40°C and 25°C, 1 GHz) were present in the ddHMQC version, as highlighted in Figure 4. It is important to emphasize, however, that the delayed decoupling/*J*-deconvolution approach is not flawless. A comparison of HMQC and ddHMQC spectra shows that the baselines of ddHMQC spectra can be slightly undulated, even after a 4th degree polynomial correction is applied in the final processing step (Figures S2-S4). This reflects the fact that in the *J*-deconvolution scheme the effect of noise in the data is exacerbated by division by $\sin(\pi J_{CH} t)$ for $0 < t < 1/(2J_{CH})$. We have written a simple routine that corrects the baseline further (see Supporting Information) and the "corrected" spectra obtained in this manner are illustrated in Figures S2-S4 as well. It is worth noting that such baseline distortions are also observed in HMQC spectra, but are not as pronounced. Second, as mentioned above, deconvolution does affect the noise floor, which is typically 5-15% higher in ddHMQC datasets (see legend to Figure 4 for values for each spectra). In an effort to eliminate the *J*-deconvolution step we are in the process of developing artificial intelligence-based approaches (31) in which a trained neural network is able to properly reconstruct the initial $1/(2J_{CH})$ of the ddHMQC time domain. This work remains ongoing.

In summary, a simple pulse scheme (even shorter than HMQC!) is described for methyl-TROSY based studies of high molecular weight complexes, using the delayed decoupling approach introduced by Gossert and coworkers (11). The idea of 'simplification through elimination' follows along the lines of a concatenation strategy introduced to complex pulse schemes that eliminates some pulses and delays, allowing evolution from more than one process to occur simultaneously in an effort to decrease relaxation losses (32). The ddHMQC scheme leads to significant gains in applications to MDa complexes, such as AaLS, and it is anticipated that experiments based on this approach will prove extremely beneficial in further studies of even larger molecules.

Acknowledgements

N.B.C, R.W.H., and Y.T. acknowledge post-doctoral support from the Canadian Institutes of Health Research (CIHR), as well as a Japan Society for the Promotion of Science Overseas Research Fellowship (Y.T.) and an Uehara Memorial Foundation postdoctoral fellowship (Y.T.), A.M.S. is grateful for an Ontario Graduate Scholarship. D.F.H. is supported by the Biotechnology and Biological Sciences Research Council UK (BBSRC) (ref: BB/T011831/1). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising. L.E.K. acknowledges support from the CIHR (FND-503573) and the Natural Sciences and Engineering Council of Canada (2015-04347).

Figure Legends

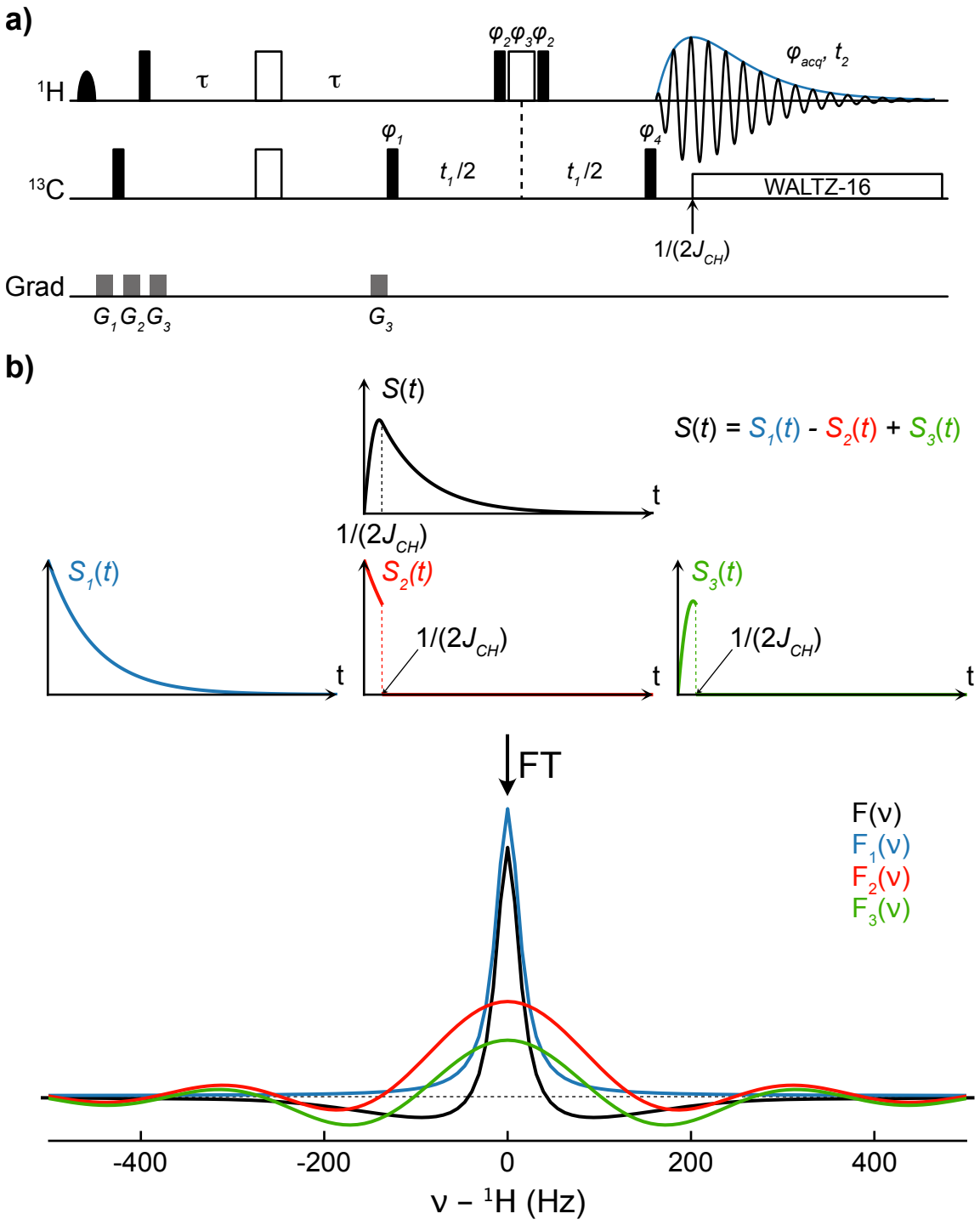


Figure 1. (a) Pulse scheme of the ddHMQC experiment. All 90º (180º) non-selective pulses are denoted by black narrow (white wide) bars, applied with phase x unless indicated otherwise. ^{1}H and ^{13}C pulses are centered at 1 ppm and 20 ppm, respectively, with the exception of the shaped ^{1}H pulse at the start of the sequence which has an EBURP1 profile

(33) (7ms) and is applied on water. WALTZ-16 decoupling (26) (3.3 kHz field at 1 GHz) commences at a time of $1/(2J_{CH})$ after the start of acquisition, once magnetization has refocused to in-phase. The value of τ is set to optimize sensitivity, typically less than $1/(4J_{CH})$. The phase cycle is φ_1 = x,-x; φ_2 = 2(x),2(y),2(-x),2(-y); φ_3 = 2(y),2(-x),2(-y),2(x); φ_4 = 8(x),8(-x); φ_{acq} = 2(x,2(-x),x),2(-x,2(x),-x). A minimum phase cycle of two is required. Gradient strengths (in % maximum) and durations are G_1=(10%,1ms), G_2=(20%,1ms), G_3=(30%,0.5ms). (b) The resulting signal, $S(t)$ (black), can be decomposed into three components $S_i(t)$, $i \in \{1,2,3\}$, with $S(t) = S_1(t)-S_2(t)+S_3(t)$, as indicated. The Fourier transform (FT) of each component, $F_i(\nu)=\mathrm{FT}(S_i(t))$, is shown along with $F(\nu)=\mathrm{FT}(S(t))$, the resulting lineshape in the ^{1}H dimension (R_2 = 100 s^{-1}).

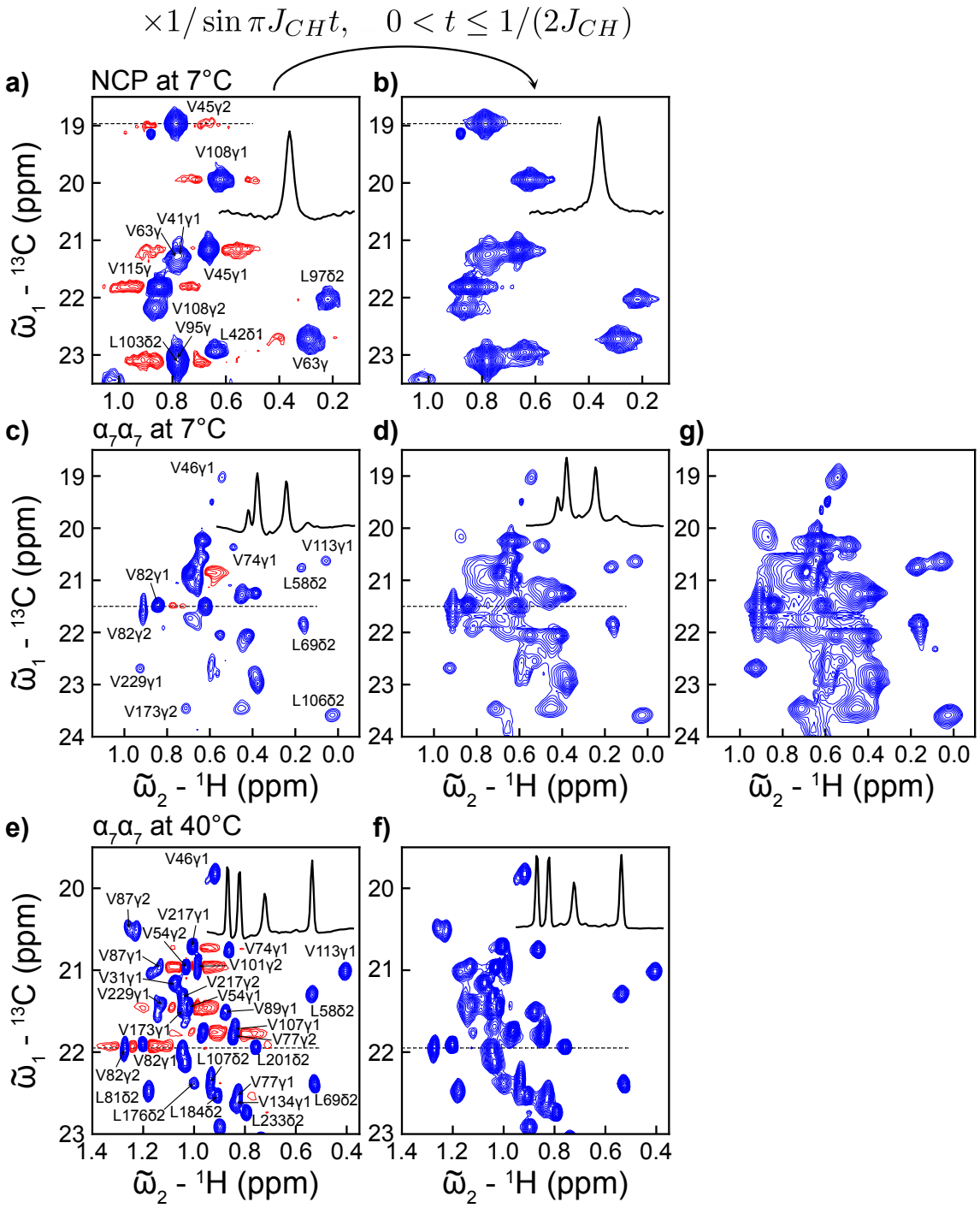


Figure 2. Selected regions of ddHMQC spectra recorded on NCP (a,b) and $\alpha_7\alpha_7$ (c-g) samples (1 GHz) recorded with the scheme of Figure 1, either processed without (a,c,e) or with (b,d,f,g) *J*-deconvolution, contoured starting at the same level for HMQC and ddHMQC datasets for a given sample (with the exception of panel g, same dataset as c and

d, but where the contours start from a level that is twice lower). Traces of regions indicated by dashed lines are shown.

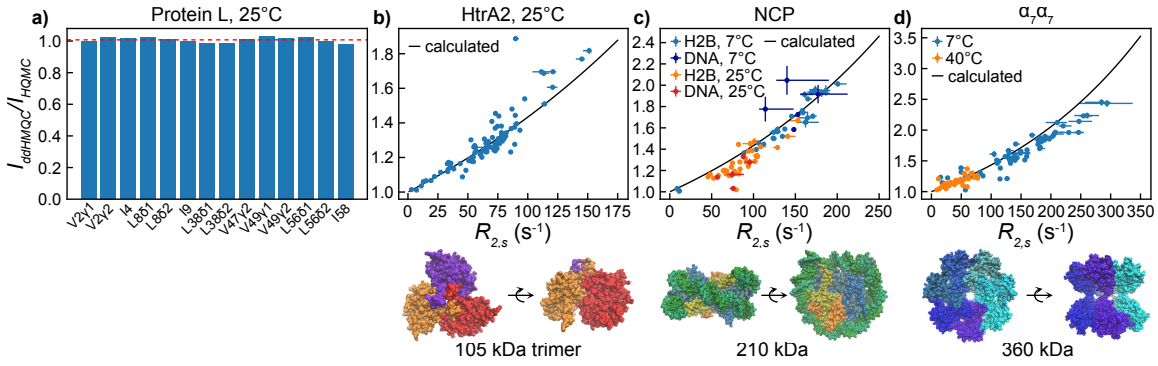


Figure 3. Intensity ratios of correlations measured in ddHMQC and HMQC datasets for molecules ranging in molecular mass from 7.5 kDa to 360 kDa. Although ratios of ≈1 are observed for protein L (a, dashed red line at $I_{ddHMQC}/I_{HMQC} = 1$), ratios increase for larger systems (b-d) and are correlated with ^{1}H transverse relaxation rates of the slowly relaxing components of ^{1}H magnetization, $R_{2,s}$. Cartoon diagrams of the structures of HtrA2 (pdb entry 1lcy (34)), NCP (pdb entry 6esf (35)), and $\alpha_7\alpha_7$ (pdb entry for Thermoplasma acidophilum proteasome 1pma (36)) are shown. In (c) intensity ratios from methyl groups of protein (referred to as H2B) and DNA (referred to as DNA) components are quantified. Noise levels are higher in ddHMQC spectra, with increases compared to the HMQC from 5.9 to 6.5 (protein L, arbitrary units), 1.5 to 1.6 (HtrA2), 1.4 to 1.5 (NCP, 25°C), 1.3 to 1.4 (NCP, 7°C), 3.3 to 3.7 ($\alpha_7\alpha_7$, 40°C) and 1.4 to 1.6 ($\alpha_7\alpha_7$, 7°C). The increased noise floor is not included in the reported signal intensities in the paper.

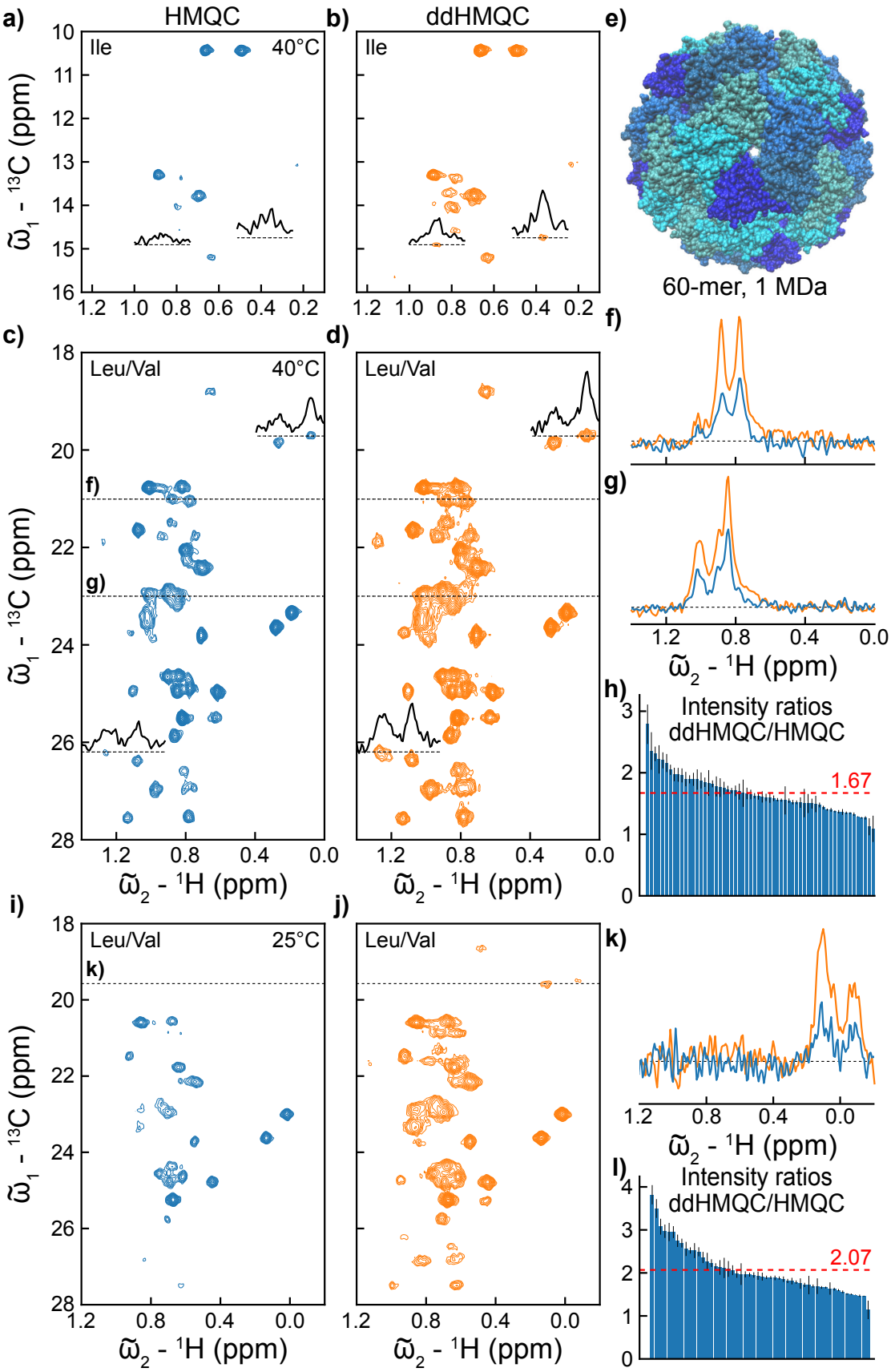

a)
HMQC
Ile
40°C
b)
ddHMQC
Ile
e)
60-mer, 1 MDa
c)
Leu/Val
40°C
d)
Leu/Val
f)
g)
h)
Intensity ratios
ddHMQC/HMQC
1.67
i)
Leu/Val
25°C
k)
j)
Leu/Val
l)
Intensity ratios
ddHMQC/HMQC
2.07
ω̃1 - 13C (ppm)
ω̃2 - 1H (ppm)

Figure 4. (a-d) Comparison of HMQC and ddHMQC spectra recorded on lumazine synthase, 40°C, 1 GHz. (e) Cartoon representation of the structure, pdb entry 1hqk (37), (f,g) Traces at ^{13}C frequencies are highlighted by dashed lines. The spectra are contoured starting at the same level. Several traces are also shown in panels a-d. (h) Intensity ratios of correlations measured in the ddHMQC and HMQC spectra at 40°C. The average ratio is 1.67 and is shown with the dashed red line. (i,j) Comparison of HMQC and ddHMQC spectra (focusing on the Leu/Val region) recorded on lumazine synthase at 25°C, 1 GHz, along with a trace (k) at the ^{13}C frequency highlighted by the dashed lines in the spectra. The spectra are contoured starting at the same level. (l) Intensity ratios of the correlations measured in the ddHMQC and HMQC spectra at 25°C, with the average ratio of 2.07 shown with the dashed red line.

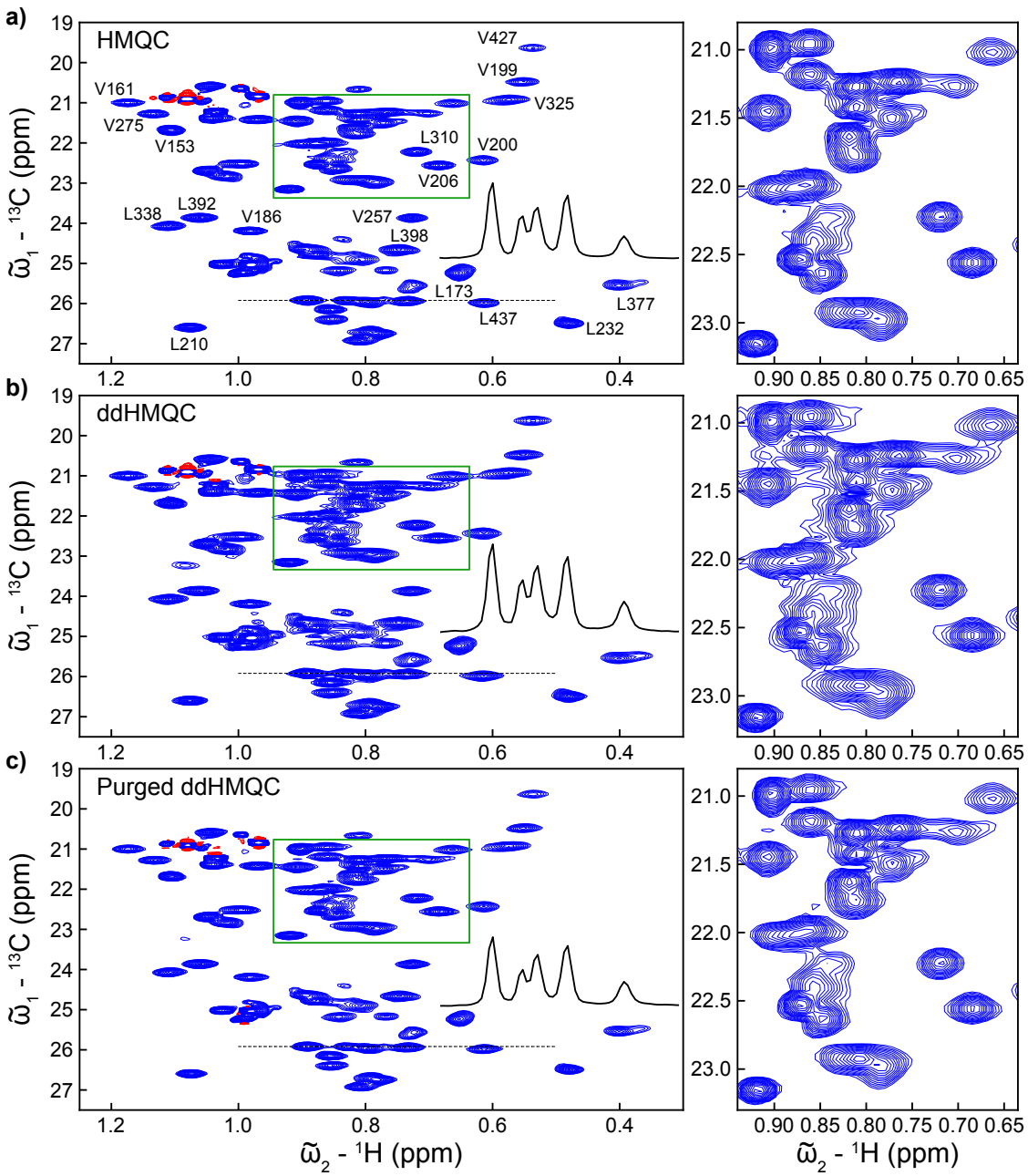


Figure 5. Comparison of selected regions of HMQC (a), ddHMQC (b), and purged ddHMQC (c) spectra for HtrA2, 1 GHz, 40°C, contoured starting at the same level. Regions in green boxes are expanded to the right.

References

1. Alderson TR, Kay LE. NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. Cell. 2021 Feb;184(3):577–95.

2. Biomolecular NMR spectroscopy. Vol. 122. Chemical Reviews; 2022. 9265–10086 p.

3. Pervushin K, Riek R, Wider G, Wuthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci. 1997 Nov 11;94(23):12366–71.

4. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-Correlated Relaxation Enhanced ^{1}H– ^{13}C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. J Am Chem Soc. 2003 Aug;125(34):10420–8.

5. Jiang Y, Kalodimos CG. NMR Studies of Large Proteins. J Mol Biol. 2017 Aug;429(17):2667–76.

6. Ruschak AM, Kay LE. Proteasome allostery as a population shift between interchanging conformers. Proc Natl Acad Sci. 2012 Dec 11;109(50):e3454–62.

7. Mueller L. Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence. J Am Chem Soc. 1979 Aug;101(16):4481–4.

8. Bax A, Griffey RH, Hawkins BL. Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR. J Magn Reson 1969. 1983 Nov;55(2):301–15.

9. Tugarinov V, Kay LE. An Isotope Labeling Strategy for Methyl TROSY Spectroscopy. J Biomol NMR. 2004 Feb;28(2):165–72.

10. Abramov G, Velyvis A, Rennella E, Wong LE, Kay LE. A methyl-TROSY approach for NMR studies of high-molecular-weight DNA with application to the nucleosome core particle. Proc Natl Acad Sci. 2020 Jun 9;117(23):12836–46.

11. Rößler P, Mathieu D, Gossert AD. Enabling NMR Studies of High Molecular Weight Systems Without the Need for Deuteration: The XL-ALSOFAST Experiment with Delayed Decoupling. Angew Chem Int Ed. 2020 Oct 19;59(43):19329–37.

12. Toyama Y, Harkness RW, Kay LE. Dissecting the role of interprotomer cooperativity in the activation of oligomeric high-temperature requirement A2 protein. Proc Natl Acad Sci. 2021 Aug 31;118(35):e2111257118.

13. Luger K, Mäder, Armin W, Richmond, Robin K, Sargent, David F, Richmond, Timothy J. Crystal structure of the nucleosome core particle at 2.8 A˚ resolution. Nature. 1997;389:251–60.

14. Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007 Feb;445(7128):618–22.

15. Sasaki E, Böhringer D, van de Waterbeemd M, Leibundgut M, Zschoche R, Heck AJR, et al. Structure and assembly of scalable porous protein cages. Nat Commun. 2017 Apr;8(1):14663.

16. Mittermaier A, Kay LE. χ_1 Torsion Angle Dynamics in Proteins from Dipolar Couplings. J Am Chem Soc. 2001 Jul 1;123(28):6892–903.

17. Gelis I, Bonvin AMJJ, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, et al. Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. Cell. 2007 Nov;131(4):756–69.

18. Delaglio F, Grzesiek S, Vuister GeertenW, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995 Nov;6(3):257–93.

19. Tugarinov V, Kay LE. Relaxation Rates of Degenerate ^{1}H Transitions in Methyl Groups of Proteins as Reporters of Side-Chain Dynamics. J Am Chem Soc. 2006 Jun 1;128(22):7299–308.

20. Mittermaier A, Kay LE. Effect of deuteration on some structural parameters of methyl groups in proteins as evaluated by residual dipolar couplings. J Biomol NMR. 2002 Mar;23:35–45.

21. Bothner-By AA, Dadok J. Useful manipulations of the free induction decay. J Magn Reson 1969. 1987 May;72(3):540–3.

22. Helmus JJ, Jaroniec CP. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. J Biomol NMR. 2013 Apr;55(4):355–67.

23. van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. Comput Sci Eng. 2011 Mar;13(2):22–30.

24. Sun H, Kay LE, Tugarinov V. An Optimized Relaxation-Based Coherence Transfer NMR Experiment for the Measurement of Side-Chain Order in Methyl-Protonated, Highly Deuterated Proteins. J Phys Chem B. 2011 Dec 15;115(49):14878–84.

25. Kupce Ē, Freeman R, Wider G, Wüthrich K. Suppression of Cycling Sidebands Using Bi-level Adiabatic Decoupling. J Magn Reson A. 1996 Sep;122(1):81–4.

26. Shaka AJ, Keeler J, Frenkiel T, Freeman R. An improved sequence for broadband decoupling: WALTZ-16. J Magn Reson 1969. 1983 Apr;52(2):335–8.

27. Ortega A, Amorós D, García de la Torre J. Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. Biophys J. 2011 Aug;101(4):892–8.

28. García de la Torre J, Huertas ML, Carrasco B. Calculation of Hydrodynamic Properties of Globular Proteins from Their Atomic-Level Structure. Biophys J. 2000 Feb;78(2):719–30.

29. Gans P, Hamelin O, Sounier R, Ayala I, Durá MA, Amero CD, et al. Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. Angew Chem Int Ed. 2010 Mar 8;49(11):1958–62.

30. Korzhnev DM, Kloiber K, Kanelis V, Tugarinov V, Kay LE. Probing Slow Dynamics in High Molecular Weight Proteins by Methyl-TROSY NMR Spectroscopy: Application to a 723-Residue Enzyme. J Am Chem Soc. 2004 Mar;126(12):3964–73.

31. Karunanithy G, Mackenzie HW, Hansen DF. Virtual Homonuclear Decoupling in Direct Detection Nuclear Magnetic Resonance Experiments Using Deep Neural Networks. J Am Chem Soc. 2021 Oct 20;143(41):16935–42.

32. Kay LE, Ikura, Mitsuhiko, Bax, Ad. The Design and Optimization of Complex NMR Experiments. Application to a Triple-Resonance Pulse Scheme Correlating Ha, NH, and 15N Chemical Shifts in 15N-13C-Labeled Proteins. J Magn Reson. 1991 Aug;91:84–92.

33. Geen H, Freeman R. Band-selective radiofrequency pulses. J Magn Reson 1969. 1991 Jun;93(1):93–141.

34. Li W, Srinivasula SM, Chai J, Li P, Wu JW, Zhang Z, et al. Structural insights into the pro-apoptotic function of mitochondrial serine protease HtrA2/Omi. Nat Struct Biol. 2002 Jun 1;9(6):436–41.

35. Bilokapic S, Strauss M, Halic M. Histone octamer rearranges to adapt to DNA unwrapping. Nat Struct Mol Biol. 2018 Jan;25(1):101–8.

36. Löwe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R. Crystal Structure of the 20 *S* Proteasome from the Archaeon *T. acidophilum* at 3.4 Å Resolution. Science. 1995 Apr 28;268(5210):533–9.

37. Zhang X, Meining W, Fischer M, Bacher A, Ladenstein R. X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 Å resolution: determinants of thermostability revealed from structural comparisons. J Mol Biol. 2001 Mar;306(5):1099–114.