This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
NMR spectroscopy is one of the basic tools for molecular structure elucidation. Unfortunately, the resolution of the spectra is often limited by inter-nuclear couplings. The existing workarounds often alleviate the problem by trading it for another deficiency, such as spectral artefacts or difficult sample preparation and, thus, rarely used. We suggest an approach using the coupling deconvolution in the framework of Compressed Sensing (CS) spectra processing that leads to a major increase in resolution, sensitivity, and overall quality of NUS reconstruction. A new mathematical description of the decoupling by deconvolution explains the effects of thermal noise and reveals a relation with the underlying assumption of the CS. The gain in resolution and sensitivity for challenging molecular systems is demonstrated for the key HNCA experiment used for protein backbone assignment applied to two large proteins: intrinsically disordered 441-residue Tau and a 509-residue globular bacteriophytochrome fragment. The approach will be valuable in a multitude of chemistry applications, where NMR experiments are compromised by the homonuclear scalar coupling.

Nuclear magnetic resonance (NMR) is among the main analytical techniques allowing atomic-level studies of proteins. The prerequisite step for most of protein NMR work is a resonance-specific spectral assignment, i.e. association of resonance frequencies with atoms in the protein amino acid chain. The HNCA[24] is by far the most sensitive and thus often the only feasible triple resonance experiment that provides sequential connectivities between neighbouring protein residues. In principle, a sequence-specific resonance assignment could be obtained using the HNCA experiment alone. Unfortunately, low signal resolution relative to the dispersion of the protein $^{13}$C resonances results in massive ambiguity of the assignment even for relatively small proteins. For large systems with many amino acid residues as well as for intrinsically disordered proteins (IDP) characterized by the particularly low resonance dispersion, one has to rely on additional experiments at the expense of sensitivity loss, a significant increase of measurement time, and more complicated and tedious analyses.

Slow transverse relaxation of the $^{13}$C spins, which can be further decreased by deuteration, corresponds to the natural linewidth of 5-8 Hz even for relatively large protein systems. Unfortunately, the practical resolution in the HNCA spectra is usually almost ten times worse. Two main factors limit the resolution in the HNCA: (i) a large number of time-increments in the $^{13}$C dimension needed in the 3D experiment to achieve the high resolution. This leads to too long measurement time that can be unaffordable because of short sample stability and/or limitation on the measurement time at an NMR instrument; (ii) homonuclear one-bond coupling between $^{13}$C and $^{1}$H spins that produces a doublet with separation of approximately 35 Hz for every $^{13}$C signal and thus effectively broadens the spectral line. The former issue is well addressed by using fast pulsing[25] and non-uniform sampling (NUS) techniques. An A large number of methods for handling of the $^{1}$J($^{13}$C$^{a}$-$^{1}$H) coupling had been introduced over the last decades, including biochemical unlabelling of the $^{1}$H carbon atom to $^{13}$C$^{a}$ constant time evolution[19] band-selective homonuclear decoupling, and IPAP decoupling.[21,22] However, broad practical use of these techniques is hindered due to
the inherent compromises in sensitivity, extra demands on sample isotope labelling, inability to deal with serine and threonine residues, and/or significant spectral distortions and artefacts.

A viable alternative to these experimental approaches is the virtual decoupling that is the post-acquisition deconvolution of the J-coupling, i.e. in- and anti-phase peak multiplets, at the signal processing stage.\textsuperscript{24,29} The aim of this communication is to investigate the possibility of effective deconvolution in compressed sensing (CS) algorithms that are among the most powerful for the NUS spectra – to propose a method of selective deconvolution of individual spectral regions; and to demonstrate the relation of the deconvolution to the cornerstone CS concept of sparseness with the resulting benefits for the effectiveness of CS.

The fundamental relation between the NMR signal $f(t)$ detected in the time domain and the spectrum $s$ is

$$f = Fs$$  

(1)

where $F$ is the measurement matrix composed of rows from the inverse Fourier transform matrix for every point in $f$. Thus, reconstruction of a spectrum from $f$ reduces to solving the inverse linear system in Eq. (1). For an undersampled (i.e. NUS) signal, the solution is not unique and additional constraints on the spectrum $s$ are usually imposed. For example, using generalized Tikhonov regularization, the spectrum can be obtained as:

$$s = \arg \min_{x \in \mathbb{C}^n} \|Fx - f\|^2_Q + \|x\|^2_D$$  

(2)

where, for a vector $y$ and matrix $G$, $\|y\|_G^2$ denotes the weighted norm square $y^*Gy$ with $y^*$ denoting the conjugate transpose of $y$; $Q = \sigma^{-2}I$ is the inverse covariance matrix of noise in $f$, which is multiple of the identity matrix $I$ and $\sigma$ is the standard deviation of the noise; $D$ is a diagonal matrix including weightings of the spectrum points and the Tikhonov regularization term. As will be shown below, $Q$ is useful when dealing with the $^1J(C\alpha-C\beta)$ coupling, while matrix is the essential element of the Iterative Reweighted Least Squares (IRLS), one of the most popular algorithms for compressed sensing reconstruction of the NUS spectra (see Supplementary information).\textsuperscript{30}\textsuperscript{31}

Assuming the same value of active $^1J(C\alpha-C\beta)$ coupling for all signals, the measured in experiment $^{13}C\alpha$ signal $f$ and the signal without the J-coupling $\tilde{f}$ are related as:

$$f = CF \tilde{f}$$  

(3)

where $C$ is a diagonal matrix with elements $\cos(\pi Jt)$ for every time point $t$ in $f$.\textsuperscript{23} If points in the measured signal $f$ are corrupted by noise with inverse covariance matrix $Q = \sigma^{-2}I$, the noise in $\tilde{f}$ has the inverse covariance matrix $Q = \sigma^{-2}(CC^t)$. Then, the deconvoluted spectrum is

$$\tilde{s} = \arg \min_{x \in \mathbb{C}^n} \|Fx - f\|^2_Q + \|x\|^2_D$$  

(4)

or equivalently (see Supplementary information),

$$\tilde{s} = \arg \min_{x \in \mathbb{C}^n} \|CFx - f\|^2_Q + \|x\|^2_D.$$  

(5)

The last equation shows that the post-acquisition deconvolution can be achieved in IRLS and any other algorithm based on equation akin to Eq. (2), e.g. Maximum Entropy\textsuperscript{24,27,29} and Multi-Dimensional Decomposition (MDD)\textsuperscript{9} by using measurement matrix $CF$ instead of $F$. Finally, we note that the deconvoluted spectrum contains half of the peaks relative to the undeconvoluted spectrum. Thus, it is \textit{sparser}, and in accordance with the theory of compressed sensing\textsuperscript{33,34} it requires nearly half of the measured data points for successful reconstruction. This means that the virtual decoupling not only enhances spectral resolution but also provides conditions for higher quality CS reconstruction (see theory in SI).

Use of the deconvolution for the HNCA experiment is based on the assumption that $^1J(C\alpha-C\beta)$ coupling constants are nearly the same for all residues in the protein. The variation of the coupling values $\pm 2.5$ Hz\textsuperscript{30} is lower than the line width determined by the transverse relaxation of $^{13}C\alpha$ spins and, thus, does not pose a problem for the reconstruction (see Supplementary information). However, signals (singlets) from Gly residues that do not have $C\beta$ atoms, have no sparse representation in the columns of measurement matrix $CF$. Figure 1 illustrates that this not only corrupts the Gly peaks in the deconvoluted spectrum but also affects other signals and reduces the overall quality of the reconstruction. To tackle this, we suggest a procedure of deconvolution-IRLS (D-IRLS) with the Gly-region selection as outlined in Figure 1 (more details are found in the Supplementary information). We start with reconstructing the full undecoupled spectrum using matrix $F$. Because the $^{13}C\alpha$ atoms usually have distinctly different chemical shifts with values lower than 45 ppm, we can subtract the well-reproduced signals in the Gly region from the original time-
domain signal $f$, which is then used to reconstruct the spectrum with all signals except for Gly using Eq. $(5)$ with the measurement matrix $CF$. Finally, signals of Gly and other residues are combined into the full decoupled spectrum in the frequency domain.

Selection of the NUS acquisition schedule has a profound effect on the reconstruction quality. As $f$ is multiplied by $C^{-1}$ in Eq. $(5)$, the noise is amplified the most for the points in $f$ at times, where $\cos(\pi J t)$ function has small values (i.e. near $t = k/(2J), k = 1, 3, 5$). In the weighted least squares method used to derive Eq. $(4)$, these points are used with low weights and thus contain relatively low information value. In the NUS schedule, it is logical to avoid these points and instead invest spectrometer time into more informative measurements. We used the signal amplitude matched NUS schedule with the sampling density corresponding to $|\cos(\pi J t)|$ and rejecting points with probability less than 0.2\cite{33,34} (Supplementary Figure S1). Additionally, the schedule was in all cases relaxation-matched.

We demonstrate the new D-IRLS procedure using examples of two representative systems: intrinsically disordered human 441-residue Tau protein (the longest hTau40 isoform)\cite{35} and the monomeric variant of the 509-residue globular photosensory module PAS-GAF-PHY of Deinococcus radiodurans phytochrome (DrBphP$_{PSM}$).\cite{36} For each protein, Figure 2 shows the traditional low-resolution 3D HNCA spectrum superimposed with the resolution-enhanced spectrum obtained using D-IRLS with Gly-region selection. For DrBphP$_{PSM}$ the two experiments were reconstructed using nearly the same number of NUS points corresponding to the same measurement time; for Tau, the low resolution experiment was around two times shorter. In the shown examples, the dramatically improved resolution of the D-IRLS spectrum allows us to observe sequential connectivities that are ambiguous in the traditional spectrum.

Figures 2, S2, and S3 demonstrate that, in addition to the enhanced resolution, the D-IRLS spectra show higher or similar sensitivity in comparison to both the traditional low resolution and non-deconvoluted spectra. The peak connecting A87 and A88 in DrBphP$_{PSM}$ spectra (Figure 2B) provides a specific example of this. It is clearly seen in the 1D cross-sections in the D-IRLS spectrum. In the traditional experiment, the weak peak is completely masked by the slope of a stronger peak. In the non-deconvoluted spectrum, only one of the doublet components is present, which gives a completely wrong idea of the peak position.

Figure S2 shows the $^{13}$C/$^{15}$N projections from the spectra of both studied proteins, which confirms the superior quality of the spectra reconstructed with Gly-region selective D-IRLS. While the improved resolution in the spectra is anticipated from the deconvolution, the remarkable sensitivity of the D-IRLS spectrum can be explained by the increased sparsity favourable for the NUS reconstruction.

In order to extensively test the proposed D-IRLS method, we conducted simulations using synthetic peaks added to the 3D HNCA signal of Tau. Adding the simulated components with known positions and intensities to the time domain signal makes it possible to define the precision of the correlation peak parameters derived from the reconstructed spectrum.\cite{39} A detailed description of the simulations can be found in the Supplemen-
Supplementary Information

The Supplementary information contains extended mathematical description of D-IRLS, sample preparation and experimental details, and results of simulations with injected peaks.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 M. Ikura, L. E. Kay and A. Bax, Biochemistry, 1990, 29, 4659–4667.
2 L. E. Kay, M. Ikura, R. Tschudin and A. Bax, Journal of Magnetic Resonance, 1990, 89, 496–514.
3 M. Salzmann, K. Pervushin, G. Wider, H. Senn and K. Wüthrich, Proceedings of the National Academy of Sciences, USA, 1998, 95, 13585–13590.
4 Z. Solyom, M. Schwarten, I. Geist, R. Konrat, D. Willbold and B. Brutsch, Journal of Biomolecular NMR, 2013, 55, 311–312.
5 K. Pervushin, B. Vogeli and A. Eletsky, Journal of the American Chemical Society, 2002, 124, 12898–12902.
6 P. Skop, P. Schanda and B. Brutsch, Journal of Magnetic Resonance, 2007, 187, 163–169.
7 J. C. J. Barna, E. D. Laue, M. R. Mayerg, J. Skilling and S. J. P. Worrall, J. Magn. Reson., 1987, 73, 69–77.
8 J. C. Hoch, A. S. Stern and M. Mobili, Maximum Entropy Reconstruction Update based on original article by Jeffrey C. Hoch, Encyclopedia of Magnetic Resonance, © 1996, John Wiley & Sons Ltd, 2007, [http://doi.wiley.com/10.1002/9780470034590.emrsta0299.pub2](http://doi.wiley.com/10.1002/9780470034590.emrsta0299.pub2)
9 V. Orekhov and V. A. Jaravine, Prog. Nucl. Magn. Reson. Spectrosc., 2011, 59, 271–292.
10 K. Kazimierczuk and V. Y. Orekhov, Ang. Chem. Int. Ed., 2011, 50, 5556–5559.
11 M. Bostock and D. Nietlispach, Concepts in Magnetic Resonance Part A, 2017, 46A, e21438.
12 S. Robson, H. Arhantrah, S. G. Hyberts and G. Wagner, Methods in Enzymology, Elsevier, 2019, vol. 614, pp. 263–291.
13 D. M. LeMaster and D. M. Kushlan, Journal of the American Chemical Society, 1996, 118, 9255–9264.
14 P. E. Coughlin, F. E. Andersen, E. J. Oliver, J. M. Brown, S. W. Homans, S. Pol- lard and J. W. Lustbader, Journal of the American Chemical Society, 1999, 121, 11871–11874.
15 M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. M. Ono and P. Güntert, Nature, 2006, 440, 52–57.
16 P. Lundström, K. Teilum, T. Carstensen, I. Bezsonova, S. Wiesner, D. F. Hansen, T. L. Religa, M. Akke and L. E. Kay, Journal of Biomolecular NMR, 2007, 38, 199–212.
17 K. Takeuchi, Z.-Y. J. Sun and G. Wagner, Journal of the American Chemical Society, 2008, 130, 17210–17211.
18 S. A. Robson, K. Takeuchi, A. Boeszoernenyi, P. W. Coote, A. Dubey, S. Hyberts, G. Wagner and H. Arhantrah, Nature Communications, 2018, 9.
19 R. Powers, A. M. Gronenborn, G. M. Crole and A. Bax, J. Magn. Reson., 1991, 94, 209–213.
20 H. Matsuo, E. Kupče, H. Li and G. Wagner, Journal of Magnetic Resonance-Series B, 1996, 113, 91–96.
21 P. W. Coote, S. A. Robson, A. Dubey, A. Boeszoernenyi, M. Zhao, G. Wagner and H. Arhantrah, Nature Communications, 2018, 1–9.
22 P. Andersson, J. Weigelt and G. Otting, Journal of Biomolecular NMR, 1998, 12, 435–441.
23 M. Ottiger, F. Delaglio and A. Bax, Journal of Magnetic Resonance, 1998, 131, 373–378.
24 A. A. Bothner-By and J. Dadok, J. Magn. Reson., 1987, 72, 540–543.
25 M. A. Delsuc and G. C. Levy, J. Magn. Reson., 1988, 76, 306–315.
26 Z. Serber, C. Richter, D. Moskau, J.-M. Böhlen, T. Gerfin, D. Marek, M. Häberli, L. Baselgia, F. Laukien, A. S. Stern, J. C. Hoch and V. Dötsch, J. Am. Chem. Soc., 2000, 122, 3554–3555.
27 N. Shima, A. S. Stern, C. S. Craik, J. C. Hoch and V. Dötsch, J. Am. Chem. Soc., 2006, 128, 4002–4004.
28 E. J. Candes, M. B. Wakin and S. P. Boyd, Journal of Fourier analysis and applications, 2008, 14, 877–905.
29 N. Shima, H. Kovacs, A. S. Stern, A. M. Nomura, I. Shimada, J. C. Hoch, C. S. Craik and V. Dötsch, J Biomol NMR, 2004, 30, 175–179.
30 R. Kerfah, O. Hamelin, J. Boisbouvier and D. Marion, Journal of Biomolecular NMR, 2015, 63, 389–402.
31 E. J. Candes, M. B. Wakin and S. P. Boyd, Journal of Fourier analysis and applications, 2008, 14, 877–905.
32 K. Kazimierczuk and V. Y. Orekhov, Journal of Magnetic Resonance, 2012, 223, 1–10.
33 Other functional forms are possible for different multiplet types, e.g. sine-modulation for an anti-phase doublet.
34 E. Candes, J. Romberg and T. Tao, IEEE Transactions on Information Theory, 2004, 52, 489–509.
35 J. M. Schmidt, M. J. Howard, M. Maestre-Martínez, C. S. Pérez and F. Löhö, Magnetic Resonance in Chemistry, 2009, 47, 16–30.
36 J. Vavrine, I. Irbragimov and V. Y. Orekhov, Nature Methods, 2006, 3, 605–607.
37 R. L. Narayan, U. H. N. Durr, S. Bibow, J. Biernat, E. Mandelkow and M. Zweckstetter, Journal of the American Chemical Society, 2010, 132, 11906–11907.
38 E. Gustavsson, L. Isaksson, C. Persson, M. Mayzel, U. Brath, L. Vrhovac, J. A. Hällalinen, B. G. Karlsson, V. Orekhov and S. Westenhoff, Biophysical Journal, 2020, 118, 415–421.
39 M. A. Zambrello, A. D. Schuyler, M. W. Maciejewski, F. Delaglio, I. Bezsonova and J. C. Hoch, Methods, 2018, 138–139, 62–68.