ARTICLE

A novel high-dimensional NMR experiment for resolving protein backbone dihedral angle ambiguities

Clemens Kauffmann[1] · Krzysztof Kazimierczuk[3] · Thomas C. Schwarz[1] · Robert Konrat[1] ·
Anna Zawadzka-Kazimierczuk[1,2]

Abstract

Intrinsically disordered proteins (IDPs) are challenging established structural biology perception and urge a reassessment of the conventional understanding of the subtle interplay between protein structure and dynamics. Due to their importance in eukaryotic life and central role in protein interaction networks, IDP research is a fascinating and highly relevant research area in which NMR spectroscopy is destined to be a key player. The flexible nature of IDPs, as a result of the sampling of a vast conformational space, however, poses a tremendous scientific challenge, both technically and theoretically. Pronounced signal averaging results in narrow signal dispersion and requires higher dimensionality NMR techniques. Moreover, a fundamental problem in the structural characterization of IDPs is the definition of the conformational ensemble sampled by the polypeptide chain in solution, where often the interpretation relies on the concept of 'residual structure' or 'conformational preference'. An important source of structural information is information-rich NMR experiments that probe protein backbone dihedral angles in a unique manner. Cross-correlated relaxation experiments have proven to fulfil this task as they provide unique information about protein backbones, particularly in IDPs. Here we present a novel cross-correlation experiment that utilizes non-uniform sampling detection schemes to resolve protein backbone dihedral ambiguities in IDPs. The sensitivity of this novel technique is illustrated with an application to the prototypical IDP α-Synuclein for which unexpected deviations from random-coil-like behaviour could be observed.

Introduction

Cross-correlated relaxation (CCR) experiments have long been established as a unique tool to study protein structure and dynamics (Brutscher 2000; Kumar et al. 2000). CCR

✉ Robert Konrat
robert.konrat@univie.ac.at

✉ Anna Zawadzka-Kazimierczuk
anzaw@chem.uw.edu.pl

[1] Max Perutz Laboratories, Department of Structural and Computational Biology, University of Vienna, Vienna Biocenter Campus 5, 1030 Vienna, Austria

[2] Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

[3] Centre of New Technologies, University of Warsaw, Banacha 2C, 02-097 Warsaw, Poland

experiments quantify correlated interferences of dipolar (dd) and/or chemical shift anisotropy (CSA) interactions. The possibility to use these effects to study protein backbone geometry was first demonstrated in 1997 (Reif et al. 1997) by investigating the backbone angle ψ from the dipole–dipole (dd–dd) interferences of $H_i^\alpha C_i^\alpha - H_i^N N_i$. Other CCR rates probing dihedral angles along the protein backbone were soon proposed, combining different dd–dd (Yang et al. 1998; Chiarparin et al. 1999; Pelupessy et al. 1999a, b; Chiarparin et al. 2000; Crowley et al. 2000; Pelupessy et al. 2003), dd-CSA (Yang et al. 1997, 1998; Chiarparin et al. 1999; Kloiber and Konrat 2000a, b) and CSA-CSA (Skrynnikov et al. 2000; Pelupessy et al. 2003) interactions, see (Schwalbe et al. 2001; Vögeli and Vugmeyster 2019) for more extensive overviews.

Every such combination of backbone interactions comes with its unique angular dependency, ranging from simple Karplus relations to more complex expressions. Due to its

non-bijective nature, an isolated CCR rate does not allow to determine the underlying angle(s) unambiguously. However, by combining and analyzing multiple CCR experiments at once, these ambiguities were shown to be resolvable (Kloiber et al. 2002) if a singular average structure can be assumed. Intrinsically disordered proteins (IDPs) prove to be more challenging in this regard. Their conformational flexibility leads to highly averaged and thus ambiguous observables, i.e. the underlying structural ensemble is underdetermined. This poses major challenges: Conventional point estimate approaches are ill-equipped for modelling heterogeneous ensembles. Not only does the high dimensionality of the problem lead to a steep increase of computational effort but also the experimental underdetermination bears the inherent risk of overfitting. While both difficulties appear to be manageable by Bayesian and Maximum Entropy methods, implementation details and subtleties are still being investigated and improved (Mantsyzov et al. 2014, 2015; Olsson et al. 2013; Bonomi et al. 2016; Hummer and Köfinger 2015; Köfinger et al. 2019; Cesari et al. 2016, 2018; Rangan et al. 2018).

The second challenge lies in the limited number of observables. Despite the conceptual and technical advances made, the underdetermination of structural ensembles still can only be amended by combining an increasing amount of experiments. The above-mentioned studies resorted to well-established NMR experiments such as scalar couplings, chemical shifts, RDCs and/or NOEs.

We argue that CCR rates provide a unique and valuable source of information that has been largely overlooked in the past, with few exceptions (Stanek et al. 2013), mostly due to the experimental challenges involved. The chemical shifts of IDPs are distributed over a very narrow spectral range, reflecting the high mobility of the polypeptide chain; a nucleus experiences (at the fast time-scale) various chemical environments which leads to observation of an averaged chemical shift. Therefore, to achieve peak separation, high-dimensional experiments need to be employed. The above-mentioned CCR experiments were designed as two-dimensional (2D) or three-dimensional (3D), which do not necessarily provide sufficient resolution for IDPs. Taking into account that CCR measurements require quantitative analysis of peak intensities, which can be disturbed by even slight peak overlap, higher dimensionality is indispensable, as demonstrated by the four-dimensional (4D) C'_i CSA— $H^N_i N_i$dd CCR experiment of Stanek et al. (2013).

However, not only preexisting CCR experiments are worthy of consideration, the complex conformational averaging of IDPs demands the development of novel orthogonal experiments that complement existing ones. Surprisingly, most reported CCR experiments neglect the potential of dipolar interactions between not-covalently bound nuclei, Crowley et al. (2000) being the only exception. Thus, we aim to extend the palette of existing CCR experiments by an entirely new interaction: Interfering $H^N_i H^\alpha_{i-1}$dd and C'_{i-1} CSA relaxation reveals a highly informative ψ-dependent observable with reduced ambiguity compared to more commonly used CCR rates and scalar couplings.

We verify the technique with Ubiquitin and demonstrate convincing agreement between structural parameters derived from PDB structures and experimental cross-correlation rates. As a first application to intrinsically disordered proteins, we show that the prototypical IDP α-Synuclein displays surprising deviations from random-coil-like behaviour which is undetectable by conventional chemical-shift-based methods.

Methods

NMR experiment

The pulse sequence of the experiment for $H^N_i H^\alpha_{i-1}$dd – C'_{i-1} CSA CCR rate measurement is shown in Fig. 1. It includes three indirectly-detected dimensions (C^α_{i-1}, C'_{i-1} and N_i) and one directly-detected H^N_i dimension. The employed method of CCR rate quantification is called quantitative spectroscopy, which means that two independent data sets (later referred to as *reference* and *cross*) are acquired. The coherence entering the CCR block of the pulse sequence is partially preserved and partially converted to another one. The conversion is CCR-mediated, thus the measurement of the intensities of signals originating from both components allows for quantification of the CCR effect. In the *cross* spectrum the observable magnetization originates from the converted coherence, while in the *reference* spectrum it originates from the preserved coherence. Crucial is the fact that the peak intensities in the *reference* spectrum are proportional to $\cosh(\Gamma T_c)$ where Γ is the corresponding CCR rate and T_c is the time of the CCR evolution, while in the *cross* spectrum they are proportional to $\sinh(\Gamma T_c)$ which is why their values (I_{ref} and I_{cross}) are not identical (see Fig. 2). Therefore the CCR rates are calculated (separately for each residue of the protein under investigation) using the following formula:

$$\Gamma = \frac{1}{T_c} \text{arctanh}\left(\frac{I_{cross}}{I_{ref}}\right) \tag{1}$$

Importantly, the formula should be modified if the *reference* and *cross* experiments were acquired with different number of scans, which is a good practice due to significantly lower sensitivity of the *cross* experiment. In such a case, the intensity ratio should be divided by the ratio of number of scans acquired in *cross* and *reference* experiments.
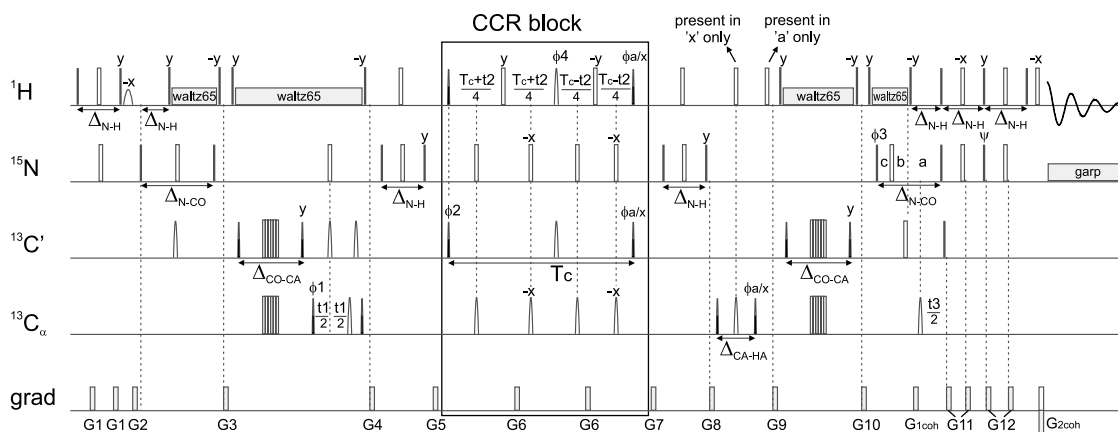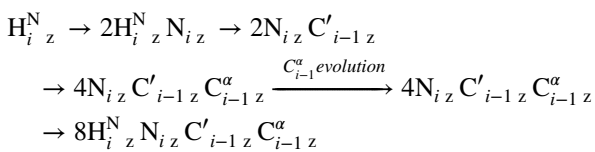
Fig. 1 Pulse sequence of the experiment for the $H^N_i H^\alpha_{i-1} dd - C'_{i-1}$ CSA CCR rate measurement. C^α evolution is in real-time mode, C' evolution is in constant-time mode and N evolution is in semi-constant time mode: $a = (t3 + \Delta_{N-CO})/2$, $b = t3(1 - \Delta_{N-CO}/t3^{max})/2$, $c = \Delta_{N-CO}(1 - t3/t3^{max})$. The delays were set as follows: $\Delta_{N-H} = 5.4$ ms, $\Delta_{N-CO} = 28$ ms, $\Delta_{CO-CA} = 9.1$ ms, $\Delta_{CA-HA} = 3.2$ ms, $T_c = 28$ ms. Unless noted explicitly, pulse phases are set to x. Phase $\phi a/x$ depends on the version of experiment; for reference experiment x, for cross experiment y. 16-step phase cycle was used: $\phi 1 =$ x, -x, $\phi 2 = 2(x)$, $2(-x)$, $\phi 4 = 4(x)$, $4(y)$, $4(-x)$, $4(-y)$. Receiver phase $\phi rec = \phi 1 + \phi 2 + 2 \cdot \phi 4$. Selective pulses affecting only H^N nuclei were pc9 (Kupce and Freeman 1994) for $\pi/2$ pulses and reburp (Geen and Freeman 1991) for π pulses. Simultaneous inversion of C^α and C' spins was achieved using 6-element composite pulse (Shaka 1985). Selective pulses for C' and C^α nuclei were q5 (for $\pi/2$ pulses) and q3 (for π pulses) (Emsley 1969). The pulses labelled "present in 'a'/'x' only" were executed only in reference/cross version of the experiment, respectively
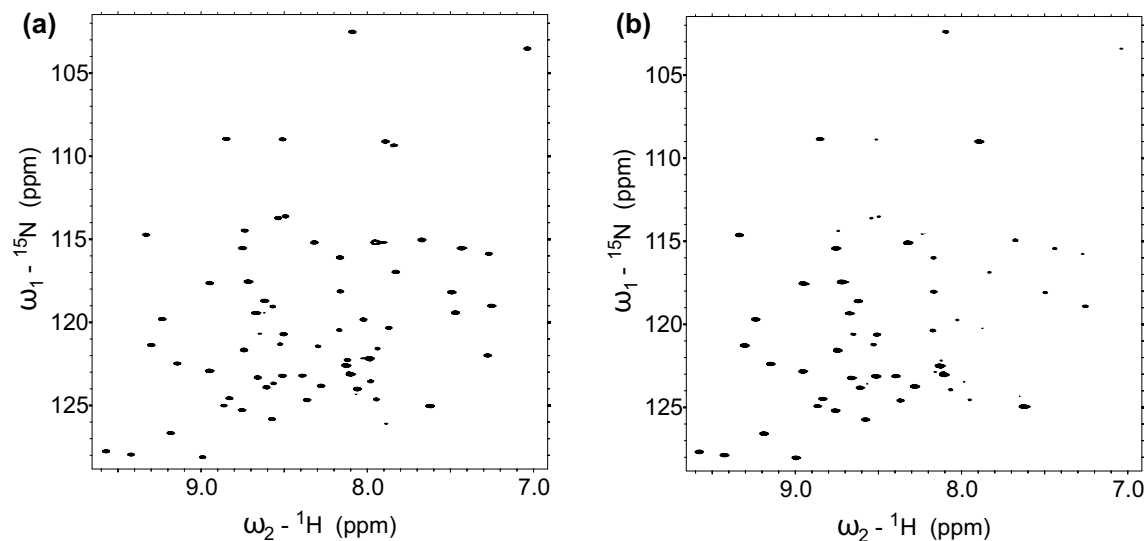
Fig. 2 Comparison of 2D versions of the experiment (with no C^α and C' evolutions) for the *reference* (**a**) and *cross* (**b**) experiments, showing differences in relative peak intensities of some residues

The coherence transfer pathway is shown below. The beginnings of *reference* and *cross* experiments are identical, up to the CCR block:

$$H^N_{i\ z} \rightarrow 2H^N_{i\ z} N_{i\ z} \rightarrow 2N_{i\ z} C'_{i-1\ z}$$

$$\rightarrow 4N_{i\ z} C'_{i-1\ z} C^\alpha_{i-1\ z} \xrightarrow{C^\alpha_{i-1} evolution} 4N_{i\ z} C'_{i-1\ z} C^\alpha_{i-1\ z}$$

$$\rightarrow 8H^N_{i\ z} N_{i\ z} C'_{i-1\ z} C^\alpha_{i-1\ z}$$

Then, during the CCR block, the product operator $8H^N_{i\ z} N_{i\ z} C'_{i-1\ z} C^\alpha_{i-1\ z}$ is partially preserved and partially converted into $16H^N_{i\ z} N_{i\ z} C'_{i-1\ z} C^\alpha_{i-1\ z} H^\alpha_{i-1\ z}$. This conversion occurs for total time T_c. Also, the evolution of C'_{i-1} nuclei occurs here. Importantly, coherence of no other CCR rate occurs during the CCR block, thus providing a clean result. After the CCR block, the coherence ($8H^N_{i\ z} N_{i\ z} C'_{i-1\ z} C^\alpha_{i-1\ z}$ for the *reference* experiment and $16H^N_{i\ z} N_{i\ z} C'_{i-1\ z} C^\alpha_{i-1\ z} H^\alpha_{i-1\ z}$ for

the *cross* experiment) is gradually back-converted into an observable H_i^N transverse magnetization. In particular, in the second INEPT after the CCR block, $8N_{i\,z}C'_{i-1\,z}C^\alpha_{i-1\,z}H^\alpha_{i-1\,z}$ operator in *cross* version is converted into $4N_{i\,z}C'_{i-1\,z}C^\alpha_{i-1\,z}$ one using the J_{CA-HA} scalar coupling. Notably, the operator involving a glycine residue contains two alpha protons and thus will not be converted into an observable magnetization using this pulse scheme. This is the reason why the presented experiment does not allow to determine CCR rates for glycine residues. The last indirect evolution (of N_i nuclei) occurs during the $2N_{i\,z}C'_{i-1\,z} \rightarrow 2H_i^N N_{i\,z}$ INEPT block.

The proposed pulse sequence was designed in a way which precludes the evolution of any other CCR rate. Nonetheless, the results still may be perturbed by another factor, namely the dispersion of the scalar coupling constants between alpha carbon and alpha proton J_{CA-HA} throughout the protein. After the CCR block, in the *cross* version of the experiment this coupling is evolved. It is however not evolved in the *reference* experiment. Therefore deviation from this assumed coupling constant will cause perturbation of the obtained Γ value. For deviations of $\pm 5\%$ of the 146 Hz value of the J-coupling, the Γ perturbation is not big and typically does not exceed 1.5 %.

An important issue is parameters of amide-proton selective pulses in the CCR block of the pulse sequence. The excitation range, defined using 'offset' and 'bandwidth' parameters of the pulse, should cover the whole amide-proton region, but not overlap with the alpha-proton region. For residues with H^N outside or H^α inside the excitation range, the measured ratio of peak intensities from reference and cross spectra will not provide a correct value of the CCR rate. As the resonance assignment is already known at the stage of CCR rates measurements, it is possible to adjust the offset and/or bandwidth to match the particular protein. In general, the problem is less pronounced for IDPs, which feature narrower chemical shift ranges, than for folded proteins. The values used in the experiments shown in the present study (offset of 8.3 ppm and bandwidth of 3.5 ppm) matched well both proteins used, Ubiquitin and α-Synuclein. The H^N of one of the Ubiquitin residues (Ile36) was outside of the excitation range, but the peak involving this nucleus provides information on the CCR rate of the preceding Gly35 residue, which—being a glycine—is not useable anyway.

The pulse sequence of the presented experiment can be obtained from the authors upon request.

Data analysis

The expected angular dependence was modelled in accordance with Yang et al. (1997) assuming model-free dynamics (Lipari and Szabo 1982),

$$\Gamma^{DD,CSA}_{AB,C}(\psi, \theta(\psi)) = \frac{4}{15}\frac{\mu_0\hbar}{4\pi}\frac{\gamma_A\gamma_B}{r^3_{AB}(\psi)}B_0\gamma_C \times f_C \times \tau_c S^2, \quad (2)$$

where

$$f_C = \frac{1}{2}[\sigma_{xx}(3\cos^2\theta_{AB,X} - 1) + \sigma_{yy}(3\cos^2\theta_{AB,Y} - 1) \\ + \sigma_{zz}(3\cos^2\theta_{AB,Z} - 1)], \quad (3)$$

with $A = H_i^N$, $B = H_{i-1}^\alpha$ and $C = C'_{i-1}$. γ is the gyromagnetic ratio, μ_0 is the vacuum permeability, $\hbar$ is the reduced Planck constant, B_0 is the magnetic field strength, τ_c is the global correlation time, S^2 is the local order parameter, $\sigma_{xx,yy,zz}$ are the tensor components of the diagonal CSA tensor (in ppm), r_{AB} is the internuclear distance between A and B, θ denotes the projection angles between the dipolar unit vector AB and the principal axes X, Y, Z of the CSA tensor coordinate system.

Γ as a function of ψ was calculated numerically using an Avogadro-generated (Hanwell et al. 2012) backbone geometry with $\psi = -180°$ (Table 1) and rotating around the C^α–C' bond in 1° steps.

Parameters were adapted primarily from Engh and Huber (2006), angles involving hydrogens were taken from Momany et al. (1975). The principal axes of the carbonyl CSA tensor were set in accordance with Teng et al. (1992): The Z-axis was defined as the cross product of the C'–O and the C'–C^α bond unit vectors, the X- and Y-axis as clockwise rotations of the C'–O bond unit vector around the Z-axis by 82° and −8°, approximating the O–C'–N angle with 120°.

The tensor components of Ubiquitin were adapted from Cisnetti et al. (2004). σ_{xx} and σ_{zz} were set according to the reported averages as 249.4 ppm and 87.9 ppm. Using the suggested calibration, the average σ_{yy} was derived from the chemical shifts (BMRB ID 17769, Cornilescu et al. (1998)) as 191.1 ppm. Uncertainties were estimated by allowing the tensor components to vary within their reported standard deviations (x, y, z = 6.1, 6.1, 5.4 ppm) while still matching the chemical shift range, yielding a lower limit of $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}$ = 243.3, 211.7, 87.4 ppm and an upper limit of $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}$ = 255.5, 172.9, 93.3 ppm.

Table 1 The model protein backbone in x,y,z-coordinates with $\psi = \omega = -180°$

Atom	x (Å)	y (Å)	z (Å)
C^α_0	0.00000	0.00000	0.00000
H^α_0	− 0.33381	0.81771	− 0.46895
C'_0	− 0.52182	− 0.01279	1.43821
O_0	0.26150	− 0.02925	2.38639
N_1	− 1.84661	− 0.00472	1.55557
H^N_1	− 2.47453	0.06788	0.78069

A correlation time τ_c of 4.1 ns was assumed (Schneider et al. 1992), order parameters were taken from Tjandra et al. (1995) with an average S^2 of 0.84, a lower limit of 0.70 and an upper limit of 0.91 (excluding the reported outlier of 0.565 at L73).

Neighbour-corrected Structural Propensity Calulator (ncSPC) values for Ubiquitin and α-Synuclein (BMRB ID 17769 and 6968, respectively) were calculated using the tool of Tamiola and Mulder (2012) with default settings and the Tamiola et al. (2010) library.

CCR rates expected for random-coil-like residues were estimated using the random coil library of Mantsyzov et al. (2015) with a total of 152870 ψ-angles (excluding glycine and proline residues). Rates were calculated according to Equation 2 and averaged. The effective correlation time was estimated from the experimentally observed values: The range of Ubiquitin (0.67 to 10.39 s^{-1}) was normalized to the mean correlation time (3.44 ns $= \tau_c S^2$). Dividing the observed range of α-Synculein (1.38 to 8.16 s^{-1}) by this factor yields an estimate for the effective correlation time of 2.40 ns. Thus, the effective correlation time was estimated to lie within 2 and 3 ns, equating to an expected Γ for random-coil-like residues between 3.10 and 4.66 s^{-1}.

Experimental

The sample of 1 mM ^{13}C,^{15}N-uniformly labeled Ubiquitin dissolved in 10 mM potassium phosphate buffer of pH= 6.5 was purchased from ASLA Biotech. The sample of 1.35 mM ^{13}C,^{15}N-uniformly labeled α-Synuclein was produced using the protocol of Wrasidlo et al. (2016). The protein was equilibrated in 20mM sodium phosphate buffer at pH 6.5 containing 200mM NaCl, 0.5mM EDTA, 0.02% NaN$_3$ and 1x cOmplete^{TM} Protease Inhibitor Cocktail (Roche) with addition of 10% D$_2$O for lock.

All the experiments were performed on a Bruker AVANCE III HD 800 MHz spectrometer equipped with a 5 mm TCI-HCN cryo-probe. The experiments for Ubiquitin were performed at 298 K and for α-Synuclein—at 284.5 K. The experimental parameters are gathered in Table 2.

The amide-proton selective pulses employed in the CCR block of the pulse sequence were defined within the pulse program using WaveMaker. Offset was set equal to 8.3 ppm and bandwidth to 3.5 ppm.

Data from conventional experiments were processed using the fast Fourier transform algorithm implemented in mddnmr software (Orekhov et al. 2004–2019). Data from NUS experiments were processed with compressed sensing (Kazimierczuk and Orekhov 2011), using the iterative soft thresholding algorithm implemented in mddnmr (Orekhov et al. 2004–2019) with parameters set to obtain good quantitativeness, i.e. 500 iterations and parameter $\lambda = 0.01$ (resulting in a relative threshold change from 1.00 in the first iteration to 0.995 in the final one). The data were displayed and analyzed using Sparky (Goddard and Kneller 2002).

Results and discussion

In order to establish the method as a reliable tool for ψ angle determination, the experiment was first tested on a protein of known structure, Ubiquitin. Using a 3D version of the experiment (no C$^\alpha_{i-1}$ evolution) with conventional sampling of the evolution time space, the consistency of the obtained Γ rates with the solution structures (10 in total) of Ubiquitin (Cornilescu et al. 1998), PDB code 1d3z, was checked. In Fig. 3 the experimentally obtained Γ rates (for each residue) are plotted against the average ψ angles of the protein structures, horizontal bars indicate the reported range of ψ values. The experimental uncertainties, shown on the plot as vertical error bars, were estimated based on the spectral noise, calculated in Sparky as a median of absolute values of 10000 randomly chosen spectral points, using the uncertainty propagation method. The potential errors originating from variation of the J_{CA-HA} scalar coupling values throughout the protein are not shown. Assuming J_{CA-HA} deviated by 5 % from 146 Hz, the resulting uncertainty would be around 1/5 of the noise-originating uncertainties. The theoretically

Table 2 Experimental parameters for all data sets (dim—dimensionality, ni—number of hypercomplex increments, ns—number of scans, conv—conventional sampling scheme, NUS—non-uniform sampling scheme)

Sample	dim	ni	C$^\alpha$		C′		N		Version	ns	Time, h
			sw,Hz	t_{max},s	sw, Hz	t_{max},s	sw,Hz	t_{max},s			
Ubiq.	2D	335 (conv)	–	–	–	–	2550	0.1314	ref	4	1 h 10 min
									cross	80	23 h
Ubiq.	3D	3536 (conv)	–	–	3000	0.0173	2550	0.0267	ref	4	23 h 20 min
									cross	8	46 h 45 min
Ubiq.	4D	680 (NUS)	6000	0.01	3000	0.0173	2550	0.0267	ref	4	9 h
									cross	28	63 h
α-syn.	4D	600 (NUS)	5600	0.0098	2200	0.0172	2700	0.0278	ref	4	8 h
									cross	28	55 h 45 min
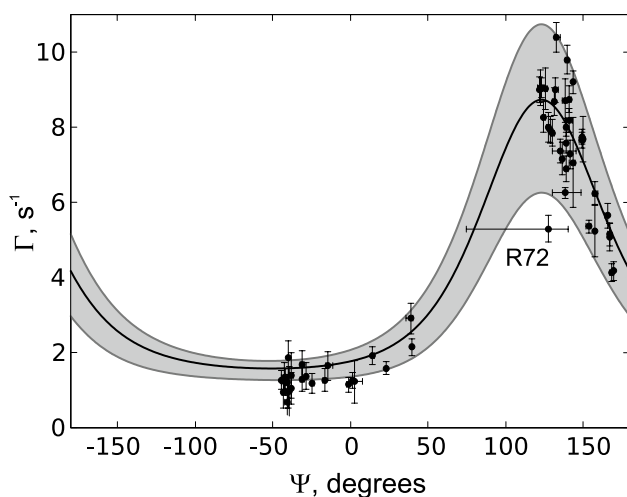
Fig. 3 Comparison of the measured and expected CCR rates $\Gamma_{H_i^N H_{i-1}^\alpha, C'_{i-1}}$ of Ubiquitin, disregarding glycine residues. Measured rates are plotted as black dots against their reported average angle ψ (PDB code 1d3z). The expected average rate with tensor components xx, yy, zz = 249.4, 191.1, 87.9ppm, τ_c = 4.1 ns and S^2 = 0.84, is represented by the black line. Expected variations are depicted in grey, with a lower limit of xx, yy, zz = 243.3, 211.7, 87.4ppm and S^2 = 0.70 and an upper limit of xx, yy, zz = 255.5, 172.9, 93.3ppm and S^2 = 0.91. Horizontal error bars indicate the reported range of angles, vertical error bars are estimated from spectral noise

expected ψ dependence is shown in black, the grey area depicts the uncertainty estimate due to variations of the CSA tensor and the local order parameter (see "Methods" section). Note that the rates could not be quantified for all residues: Besides glycines (see "Methods" section), CCR rates could not be determined for three residues preceding prolines as well as seven other residues (Thr7, Leu8, Ile23, Phe45, Asp52, Leu73, Arg74) due to a very low peak intensity in the reference spectrum.

Clearly, the measured rates agree well with the rates calculated from the PDB structures. The apparent outlier of

the highlighted residue R72 is to be expected. As part of the flexible C-terminal tail, its dynamics and structural averaging are reflected both in a low Γ and the horizontal uncertainty range reported in the PDB structures.

To further assess the influence of ψ uncertainties, a comparison of different PDB entries is shown in Fig. 4: The aforementioned set of NMR derived solution structures 1d3z, the crystal structure 1ubq (Vijay-Kumar et al. 1987) and the ensemble 2k39 (Lange et al. 2008). Average CCR rates were calculated for all structures and compared to the experimentally obtained rates. While the overall agreement between them is apparent, it is worth noting that the ensemble of Lange et al. (2008) (116 structures) gives the highest Pearson R of 0.975. We conclude that the observed deviations are mostly due to the aforementioned experimental and theoretical uncertainties as indicated in Fig. 3.

As Fig. 3 illustrates, the Γ rate can be consistent with two different ψ angles. To resolve these ambiguities and determine the correct conformation, additional experiments need to be analyzed in parallel. Still, a qualitative interpretation is straightforward: Low rates correspond to helical motifs, high rates to β-strands. Note that the low range between ca. $-140°$ and $50°$ is less steep and pronounced. Thus, ψ angles corresponding to β-like elements can be determined with higher precision than angles of helical residues.

Recording the experiment in its 4D version, suitable to crowded spectra of IDPs, requires using of non-uniform sampling (NUS). We have chosen the compressed-sensing IST algorithm for NUS data processing. To check the reliability of the CCR rates calculated using spectra reconstructed from NUS data, we compared the CCR rates of Ubiquitin from two sparse data sets with the rates obtained from the 3D conventional spectra. One sparse data set was obtained by randomly choosing a subset of points from a conventional 3D data set, while another was a 4D dataset recorded in a NUS manner. The comparison proving the methods reliability is shown in Fig. 5.
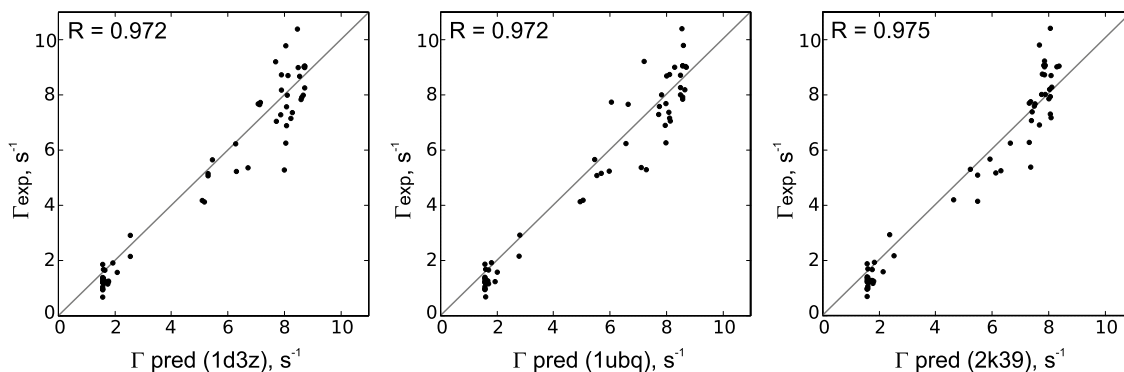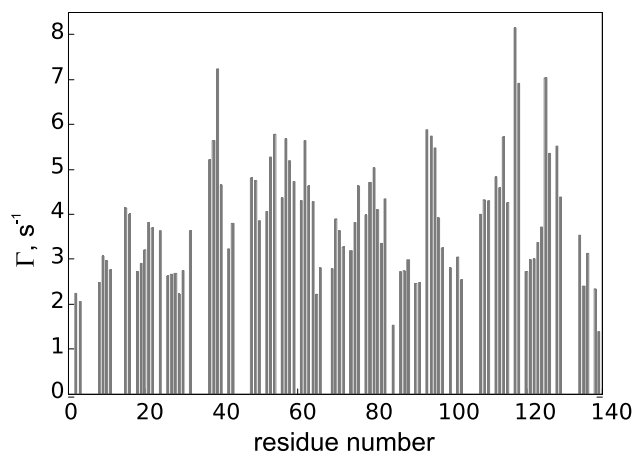

Fig. 4 Comparison of calculated CCR rates of three different PDB structures of Ubiquitin (x axes) with the experimentally obtained CCR rates (y axes). R is the Pearson correlation coefficient

(a)

(b)
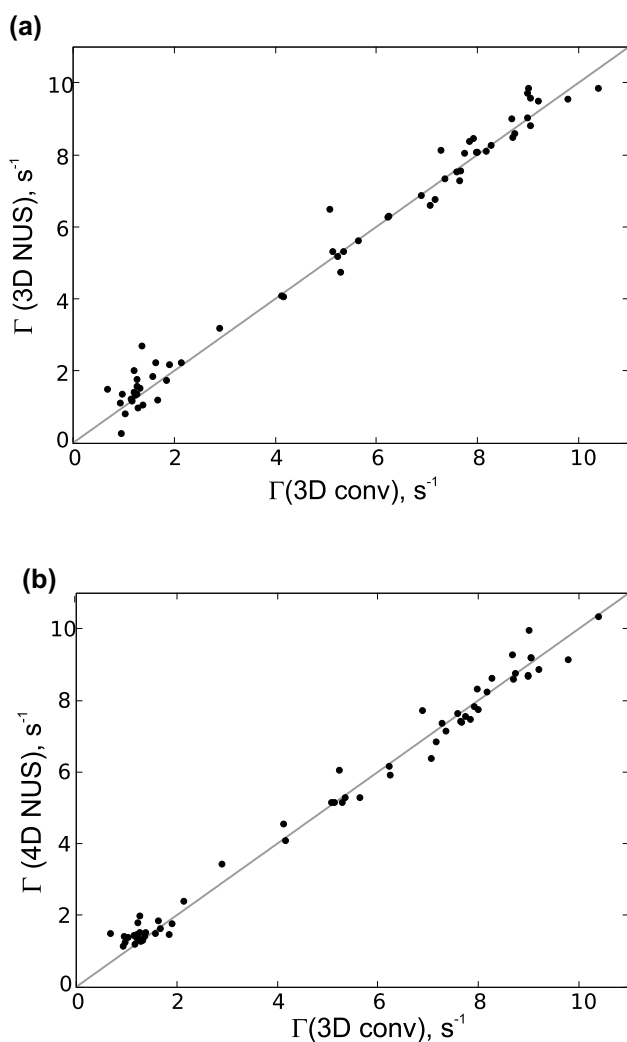
Fig. 5 Comparison of CCR rates values for Ubiquitin sample obtained using 3D conventional (fully-sampled) data with the values obtained using: **a** a subset of 680 random points chosen from the 3D conventional data set, **b** 4D NUS dataset
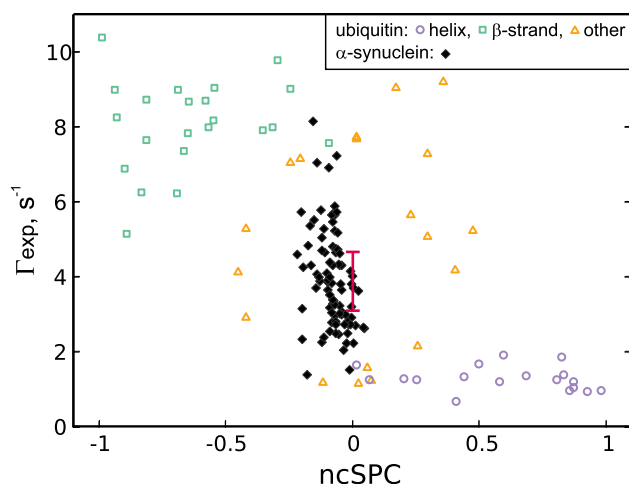

Fig. 6 $\Gamma_{H_i^N H_{i-1}^\alpha, C'_{i-1}}$ CCR rates of α-Synculein obtained using the 4D version of the proposed experiment

(ncSPC) (Tamiola and Mulder 2012) for both Ubiquitin and α-Synuclein, see Fig. 7. Referring to the globular protein Ubiquitin first, we can see that both ncSPC and $\Gamma_{H_i^N H_{i-1}^\alpha, C'_{i-1}}$ allow to distinguish between helical and β-strand structures. For helical residues, the ncSPC values are positive and the corresponding CCR rates are below $2\ s^{-1}$. In contrast, for β-strand residues ncSPCs are negative and CCR rates exceed $5\ s^{-1}$. The observed vertical spread for residues located in the β region is much higher than in helices which is expected

Finally, the experiment was acquired for the model IDP α-Synuclein as a prototypical showcase. CCR rates could be determined for 92 residues (see Fig. 6), which constitutes 80 % of non-glycine residues for which the $C'_{i-1}-N_i-H_i^N$ chemical shifts were known from the BMRB entry 6968. The remaining 20 % of peaks (i.e. 23 peaks) were overlapped. Importantly, this still represents a noticeable decrease compared to the 3D H^N–N–C' projection in which a total of 30 residues are overlapping.

It should be noted that due to conformational averaging these rates cannot be related to a single dihedral angle ψ as in Fig. 3, however, the same underlying functional form is applicable. To assess the potential structural information content of the $H_i^N H_{i-1}^\alpha$dd—C'_{i-1} CSA CCR rate, we use the chemical shift derived structural propensity score provided by the neighbour-corrected Structural Propensity Calculator

Fig. 7 $\Gamma_{H_i^N H_{i-1}^\alpha, C'_{i-1}}$ rates of Ubiquitin and α-Synuclein plotted against their corresponding ncSPC scores. The data corresponding to Ubiquitin residues is plotted with empty markers corresponding to their secondary structure (as in PDB entry 1d3z): helix—circle, β-strand—square, other—triangle. The data for α-Synuclein is plotted using filled markers. Expected values for random-coil-like residues are indicated by the error bar: Centered at 0 ncSPC, the bar indicates the expected range for Γ (as described in the Methods—"Data analysis" section)

given the underlying functional form of Γ (Fig. 3). Interestingly, secondary structure motifs other than helix and strand show a noticeable vertical and horizontal spread, illustrating the shortcomings of chemical shift derived secondary structure assessments. Nevertheless, the observed increase in resolution is worth noting and indicates an interesting (and potentially valuable) complementarity between CCR rates and ncSPCs for elucidating conformational averaging in mobile parts of (folded) proteins.

The IDP α-Synuclein shows a strikingly different pattern. As expected, the ncSPC range is substantially narrower (-0.22 up to 0.05) compared to a globular protein (due to structural averaging), which shows that no secondary structure motifs are constantly adapted. However, the CCR rates are significantly more diverse, ranging from 1.38 to 8.16 s^{-1}, thus exceeding our estimates for random-coil-like residues of 3.10 to 4.66 s^{-1}). This surprising heterogeneity indicates that residues in α-Synuclein are far from identical, but rather showing differential local structural propensities which are not displayed by the chemical shifts. We thus conclude that the analysis of CCR rates offers unique possibilities to investigate structural dynamics of IDPs in solution.

Conclusion

A novel pulse sequence was presented that allows quantifying cross-correlated relaxation between the $H_i^N H_{i-1}^\alpha$ dipole–dipole and the C'_{i-1} chemical shift anisotropy. Validated against Ubiquitin, this ψ-dependent interference proves to be highly sensitive and capable to distinguish between helical and β-strand structural elements. This is due to its non-trivial angular distance dependency, which results in a strongly decreased ambiguity compared to other commonly used scalar couplings and CCR rates.

We argue that this unique feature is of particular importance in the context of intrinsically disordered proteins (IDPs), where quantitative analysis is challenging due to the effects of ensemble averaging. Using α-Synuclein as an example, we show that the proposed CCR rate is not only highly specific to structural motifs of folded proteins but maintains its functional range even in highly disordered proteins. These findings suggest surprising deviations from the random-coil-like behaviour chemical-shift-based methods would suggest. This might not only be due to the unique geometrical dependency of the CCR rate but also due to its sensitivity to local dynamics. While chemical shifts report on simple population averages, relaxation rates are relatively weighted by their effective correlation time. Thus, the contribution of compact substates, even if sparsely populated, is amplified by their dynamics. We conclude that CCR rates are uniquely suited to characterize IDP ensembles and especially their compact substates. A systematic approach combining

various CCR rates to characterize backbone dihedral angle distributions of IDPs as well as folded proteins is currently under investigation in our lab.

References

Bonomi M, Camilloni C, Cavalli A, Vendruscolo M (2016) Metainference: a bayesian inference method for heterogeneous systems. Sci Adv 2(1):e1501177

Brutscher B (2000) Principles and applications of cross-correlated relaxation in biomolecules. Concepts Magn Reson 12(4):207–229

Cesari A, Gil-Ley A, Bussi G (2016) Combining simulations and solution experiments as a paradigm for rna force field refinement. J Chem Theory Comput 12(12):6192–6200

Cesari A, Reißer S, Bussi G (2018) Using the maximum entropy principle to combine simulations and solution experiments. Computation 6(1):15

Chiarparin E, Pelupessy P, Ghose R, Bodenhausen G (1999) Relaxation of two-spin coherence due to cross-correlated fluctuations of dipole-dipole couplings and anisotropic shifts in NMR of 15N,13C-labeled biomolecules. J Am Chem Soc 121(29):6876–6883

Chiarparin E, Pelupessy P, Ghose R, Bodenhausen G (2000) Relative orientation of C(α)H(α)-bond vectors of successive residues in proteins through cross-correlated relaxation in NMR. J Am Chem Soc 122(8):1758–1761

Cisnetti F, Loth K, Pelupessy P, Bodenhausen G (2004) Determination of chemical shift anisotropy tensors of carbonyl nuclei in proteins through cross-correlated relaxation in NMR. ChemPhysChem 5(6):807–814

Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120(27):6836–6837

Crowley P, Ubbink M, Otting G (2000) ϕ angle restraints in protein backbones from dipole-dipole cross-correlation between ^{1}H^N-^{15}N and ^{1}H^N-^{1}H$^\alpha$ vectors. J Am Chem Soc 122(12):2968–2969

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for nmr using a quaternion description of their overall propagators. J Magn Reson (1969) 97(1):135–148

Engh RA, Huber R (2006) Structure quality and target parameters. American Cancer Society, chap 18.3, pp 382–392

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

Goddard TD, Kneller DG (2002) Sparky 3, University of California, San Francisco. http://www.cgl.ucsf.edu/home/sparky/

Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4(1):17

Hummer G, Köfinger J (2015) Bayesian ensemble refinement by replica simulations and reweighting. J Chem Phys 143(24):243150

Kazimierczuk K, Orekhov VY (2011) Accelerated NMR spectroscopy by using compressed sensing. Angewandte Chemie - International Edition 50(24):5556–5559

Kloiber K, Konrat R (2000a) Measurement of the protein backbone dihedral angle σ based on quantification of remote CSA/DD interference in inter-residue $13C'(i-1)$-$13C(\alpha)(i)$ multiple-quantum coherences. J Biomol NMR 17(3):265–268

Kloiber K, Konrat R (2000b) Peptide plane torsion angles in proteins through interference: facile discrimination between Type-I and Type-II—turns. J Am Chem Soc 122(17):12033–12034

Kloiber K, Schüler W, Konrat R (2002) Automated NMR determination of protein backbone dihedral angles from cross-correlated spin relaxation. J Biomol NMR 22(4):349–363

Kumar A, Christy Rani Grace R, Madhu PK (2000) Cross-correlations in NMR. Prog Nucl Magn Reson Spectrosc 37(3):191–319

Kupce E, Freeman R (1994) Wideband excitation with polychromatic pulses. J Magn Reson A 108(2):268–273

Köfinger J, Stelzl LS, Reuter K, Allande C, Reichel K, Hummer G (2019) Efficient ensemble refinement by reweighting. J Chem Theory Comput 15(5):3390–3401

Lange OF, Lakomek NA, Farès C, Schröder GF, Walter KFA, Becker S, Meiler J, Grubmüller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution. Science 320(5882):1471–1475. https://doi.org/10.1126/science.1157092

Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104(17):4546–4559

Mantsyzov AB, Maltsev AS, Ying J, Shen Y, Hummer G, Bax A (2014) A maximum entropy approach to the study of residue-specific backbone angle distributions in α-synuclein, an intrinsically disordered protein. Protein Sci 23(9):1275–1290

Mantsyzov AB, Shen Y, Lee JH, Hummer G, Bax A (2015) Mera: a webserver for evaluating backbone torsion angle distributions in dynamic and disordered proteins from nmr data. J Biomol NMR 63(1):85–95

Momany FA, McGuire RF, Burgess AW, Scheraga HA (1975) Energy parameters in polypeptides. vii. geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J Phys Chem 79(22):2361–2381

Olsson S, Frellsen J, Boomsma W, Mardia KV, Hamelryck T (2013) Inference of structure ensembles of flexible biomolecules from sparse, averaged data. PLoS ONE 8(11):1–7

Orekhov VY, Jaravine V, Mayzel M, Kazimierczuk K (2004–2019) MddNMR—reconstruction of NMR spectra from NUS signal using MDD and CS. http://mddnmr.spektrino.com

Pelupessy P, Chiarparin E, Ghose R, Bodenhausen G (1999a) Efficient determination of angles subtended by $C(\alpha)$-$H(\alpha)$ and N-H(N) vectors in proteins via dipole-dipole cross-correlation. J Biomol NMR 13(4):375–380

Pelupessy P, Chiarparin E, Ghose R, Bodenhausen G (1999b) Simultaneous determination of Ψ and Φ angles in proteins from measurements of cross-correlated relaxation effects. J Biomol NMR 14(3):277–280

Pelupessy P, Ravindranathan S, Bodenhausen G (2003) Correlated motions of successive amide N-H bonds in proteins. J Biomol NMR 25(4):265–280

Rangan R, Bonomi M, Heller GT, Cesari A, Bussi G, Vendruscolo M (2018) Determination of structural ensembles of proteins: restraining vs reweighting. J Chem Theory Comput 14(12):6632–6641

Reif B, Hennig M, Griesinger C (1997) Direct measurement of angles between bond. Science 276(May):1230–1234

Schneider DM, Dellwo MJ, Wand AJ (1992) Fast internal main-chain dynamics of human ubiquitin. Biochemistry 31(14):3645–3652

Schwalbe H, Carlomagno T, Hennig M, Junker J, Reif B, Richter C, Griesinger C (2001) Cross-correlated relaxation for measurement of angles between tensorial interactions. Methods Enzymol 338:35–81

Shaka A (1985) Composite pulses for ultra-broadband spin inversion. Chem Phys Lett 120(2):201–205

Skrynnikov NR, Konrat R, Muhandiram DR, Kay LE (2000) Relative orientation of peptide planes in proteins is reflected in carbonyl—carbonyl chemical shift anisotropy cross-correlated spin relaxation. J Am Chem Soc 122(29):7059–7071

Stanek J, Saxena S, Geist L, Konrat R, Koźmiński W (2013) Probing local backbone geometries in intrinsically disordered proteins by cross-correlated NMR relaxation. Angewandte Chemie - International Edition 52(17):4604–4606. https://doi.org/10.1002/anie.201210005

Tamiola K, Mulder F (2012) Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. Biochem Soc Trans 40(5):1014–1020

Tamiola K, Acar B, Mulder FAA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. J Am Chem Soc 132(51):18000–18003

Teng Q, Iqbal M, Cross TA (1992) Determination of the carbon-13 chemical shift and nitrogen-14 electric field gradient tensor orientations with respect to the molecular frame in a polypeptide. J Am Chem Soc 114(13):5312–5321

Tjandra N, Feller SE, Pastor RW, Bax A (1995) Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation. J Am Chem Soc 117(50):12562–12566

Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8Åresolution. J Mol Biol 194(3):531–544. https://doi.org/10.1016/0022-2836(87)90679-6

Vögeli B, Vugmeyster L (2019) Distance-independent cross-correlated relaxation and isotropic chemical shift modulation in protein dynamics studies. ChemPhysChem 20(2):178–196. https://doi.org/10.1002/cphc.201800602

Wrasidlo W, Tsigelny IF, Price DL, Dutta G, Rockenstein E, Schwarz TC, Ledolter K, Bonhaus D, Paulino A, Eleuteri S, Skjevik ÅA, Kouznetsova VL, Spencer B, Desplats P, Gonzalez-Ruelas T, Trejo-Morales M, Overk CR, Winter S, Zhu C, Chesselet MF, Meier D, Moessler H, Konrat R, Masliah E (2016) A de novo compound targeting α-synuclein improves deficits in models of Parkinson's disease. Brain 139(12):3217–3236

Yang D, Konrat R, Kay LE (1997) A multidimensional NMR experiment for measurement of the protein dihedral angle ψ based on cross-correlated relaxation between $1H(\alpha)$-$13C(\alpha)$ dipolar and 13C' (carbonyl) chemical shift anisotropy mechanisms. J Am Chem Soc 119(49):11938–11940

Yang D, Gardner KH, Kay LE (1998) A sensitive pulse scheme for measuring the backbone dihedral angle ψ based on cross-correlation between 13 C. J Biomol NMR 11(2):213–220

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.