Biomolecular NMR Assignments
https://doi.org/10.1007/s12104-021-10009-8

ARTICLE

The highly flexible disordered regions of the SARS-CoV-2 nucleocapsid N protein within the 1–248 residue construct: sequence-specific resonance assignments through NMR

Marco Schiavina[1,2] · Letizia Pontoriero[1,2] · Vladimir N. Uversky[3] · Isabella C. Felli[1,2] · Roberta Pierattelli[1,2]

Received: 10 December 2020 / Accepted: 15 February 2021

Abstract

The nucleocapsid protein N from SARS-CoV-2 is one of the most highly expressed proteins by the virus and plays a number of important roles in the transcription and assembly of the virion within the infected host cell. It is expected to be characterized by a highly dynamic and heterogeneous structure as can be inferred by bioinformatics analyses as well as from the data available for the homologous protein from SARS-CoV. The two globular domains of the protein (NTD and CTD) have been investigated while no high-resolution information is available yet for the flexible regions of the protein. We focus here on the 1–248 construct which comprises two disordered fragments (IDR1 and IDR2) in addition to the N-terminal globular domain (NTD) and report the sequence-specific assignment of the two disordered regions, a step forward towards the complete characterization of the whole protein.

Keywords SARS-CoV-2 · Covid-19 · Nucleocapsid protein · NMR spectroscopy · ^{13}C detection · IDPs

Biological context

Coronaviruses (CoVs) are relatively large viruses containing a single-stranded positive-sense RNA genome encapsulated within a membrane envelope (Cui et al. 2019). There are four classes of CoVs, called α, β, γ, and δ, with the class β-coronavirus including CoVs that can infect humans, such as the severe acute respiratory syndrome virus (SARS-CoV), the Middle East respiratory syndrome virus (MERS-CoV), and the COVID-19 causative agent SARS-CoV-2 (Masters 2006; Surjit and Lal 2008). Similar to SARS-CoV and MERS-CoV, SARS-CoV-2 attacks the lower respiratory system causing viral pneumonia, but it may also affect the gastrointestinal system, heart, kidney, liver, and central nervous system leading to multiple organ failure (Huang et al. 2020; Wang et al. 2020). The severe rate of this virus spread, based on its unexpectedly high infectivity, demands rapid action towards both the development of a vaccine and potent viral inhibitors to weaken or eliminate major life-threat symptoms.

The SARS-CoV-2 nucleocapsid protein N is a structurally heterogeneous, 419 amino-acid-long, multidomain RNA-binding protein that is found inside the viral envelope (Fig. 1). This protein, as already established for its SARS-CoV homologue, stabilizes viral RNA by forming a ribonucleoprotein complex (RNP) and plays a fundamental role in the transcription and assembly of the virion once the host cell is infected (Chang et al. 2009, 2014). The self-association of the N protein is also responsible for the formation of a shell, the capsid, which protects the genetic material from external agents. The N protein includes two functional domains known as N- and C-terminal domains, or NTD and CTD respectively, that are

Marco Schiavina and Letizia Pontoriero have contributed equally.

Isabella C. Felli
felli@cerm.unifi.it

Roberta Pierattelli
roberta.pierattelli@unifi.it

[1] Magnetic Resonance Center – CERM, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, FI, Italy

[2] Department of Chemistry “Ugo Schiff”, University of Florence, Via della Lastruccia 3-13, 50019 Sesto Fiorentino, FI, Italy

[3] Department of Molecular Medicine and USF Health Byrd Alzheimer’s Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

Published online: 03 March 2021

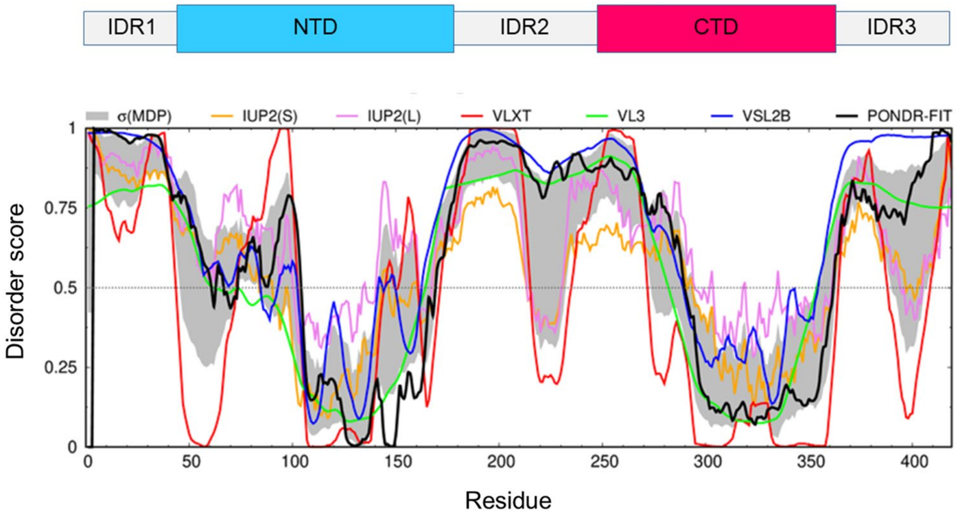


Fig. 1 Bioinformatics analysis of the intrinsic disorder predisposition of the SARS-CoV-2 nucleocapsid N protein obtained using IUPred short (golden line), IUPred long (purple line), PONDR® VLXT (red line), PONDR® VL3 (green line), PONDR® VSL2B (blue line), PONDR® FIT (black line). The gray shadow region signifies the error distribution σ(MDP) around the mean disorder profile calculated by averaging of the disorder profiles of individual predictors. Protein regions with a disordered score consistently larger than 0.5 are considered disordered, whereas regions with disorder scores between 0.2 and 0.5 are considered as flexible. Over the plot, the domain organization used in the text is reported

responsible for RNA binding (NTD) and homo-dimerization (CTD) (Chang et al. 2006). Bioinformatics analysis predicts the presence of three long intrinsically disordered regions in the polypeptide chain as reported in Fig. 1 (Giri et al. 2020). These regions are believed to be responsible for an intricate mechanism that leads to the regulation of the formation of the RNP complex. They are also engaged in many interactions with other viral proteins or host proteins, as was already demonstrated for the homologous nucleocapsid protein of the CoV that causes SARS (Chang et al. 2014; Giri et al. 2020). To date there is no structural and dynamic information with atomic resolution for the entire N protein due to its highly disordered nature. The structures of the globular NTD and CTD domains have been determined (Kang et al. 2020; Peng et al. 2020; Dinesh et al. 2020). However, there is no atomic resolution information on the disordered parts of this protein. On the other hand, the role of disorder is not accidental and is very relevant for the modulation of the mechanisms leading to the infection (Goh et al. 2012, 2013). In addition, the N proteins of the different variants of CoVs seem to be genetically stable (Giri et al. 2020), which makes them excellent candidates for developing antiviral therapies that have not been explored to date.

In this frame, we provide here the backbone assignment of the two disordered regions flanking the NTD, the N-terminal IDR1 and the serine-rich disordered region IDR2, in the 1–248 residue construct (IDR1-NTD-IDR2). These data will contribute to the efforts of the research consortium covid19-nmr (www.covid19-nmr.de) enabling follow-up applications, such as residue-resolved drug screening and interaction mapping.

Methods and experiments

Construct design

This study uses the SARS-CoV-2 NCBI reference genome entry NC_045512.2, identical to GenBank entry MN908947.3. The definition of domain boundaries for the IDR1-NTD-IDR2 fragment (1–248) was guided by the SARS-CoV homologue (Chang et al. 2014).

A codon-optimized expression construct of SARS-CoV-2 IDR1-NTD-IDR2 inserted into the pET29b(+) plasmid was obtained from Twist Bioscience.

Sample preparation

Uniformly ^{13}C, ^{15}N-labelled IDR1-NTD-IDR2 protein was expressed in *E. coli* strain BL21 (DE3). The culture was grown in 1 L LB medium at 37 °C until OD_{600} reached 0.8, then transferred in 250 mL of labelled minimal medium (4x) containing 0.25 g/L $^{15}NH_4Cl$ (Cambridge Isotope Laboratories), 0.75 g/L [U]$^{13}C_6$-D-glucose (Eurisotop). After 1 h of metabolite clearance, the culture was induced with 0.2 mM isopropyl-beta-thiogalactopyranoside (IPTG) at 18 °C for 16/18 h.

The cell pellet was resuspended in 25 mM 2-Amino-2-(hydroxymethyl)-1,3-propanediol (TRIS), 1.0 M sodium chloride, 5% glycerol, DNAse, RNAse and 500 µL of 100 × stock of protease inhibitor cocktail (SIGMA) at pH 8.

Cells were disrupted by sonication. The supernatant was cleared by centrifugation (50′, 30,000×*g*, 4 °C), then the cleared supernatant was dialyzed overnight at 4 °C into 25 mM TRIS pH 7.2 (binding buffer).

The protein was purified with ion-exchange chromatography using an HiTrap SP FF 5 mL column and a 70% gradient of 25 mM TRIS, 1 M NaCl pH 7.2. Fractions containing pure protein were pooled and concentrated using 15 mL and 0.5 mL Centricon centrifugal concentrators (MW cut-off 10 kDa).

Final NMR samples were 280 µM IDR1-NTD-IDR2, 25 mM TRIS pH 6.5, 450 mM sodium chloride, 0.02% NaN_3, 5% (v/v) D_2O in water.

NMR experiments

All the NMR experiments were acquired at 298 K. Carbon-13 direct detected NMR experiments were acquired on a 16.4 T Bruker AVANCE NEO spectrometer operating at 700.06 MHz ^{1}H, 176.05 MHz ^{13}C, and 70.97 MHz ^{15}N frequencies, equipped with a 5 mm cryogenically cooled probehead optimized for ^{13}C direct detection (TXO). Proton direct detected NMR experiments were acquired on a 28.3 T Bruker AVANCE NEO spectrometer operating at 1200.85 MHz ^{1}H, 301.97 MHz ^{13}C, and 121.70 MHz ^{15}N equipped with a 3 mm cryogenically cooled triple-resonance probehead (TCI).

Backbone assignment was performed by analyzing 2D and 3D ^{1}H and ^{13}C direct detected experiments. In particular, 2D-[^{1}H, ^{15}N]-HSQC, 2D-[^{1}H, ^{15}N]-BEST-TROSY (BT), 2D-CON, 2D-(H)CACO and 2D-(H)CBCACO experiments were performed. Moreover, a series of 3D experiments were acquired: 3D-(H)CBCACON, 3D-(H)CBCANCO, 3D-BT-HNCACB, and 3D-BT-HN(CO)CACB. To compare the resonance values obtained through the carbon detected spectra with the ones obtained with the proton detected ones, 3D-HNCO and 3D-HN(CA)CO were also collected.

All the 2D-^{13}C detected experiments were acquired in a version optimized for the detection of the highly flexible regions of the protein (Felli and Pierattelli 2012). Carbon-13 homonuclear decoupling was achieved through the IPAP virtual decoupling approach (Bermel et al. 2006a). 2D-(H)CACO and 2D-(H)CBCACO exploit constant-time evolution in the indirect dimension (Pontoriero et al. 2020). The 2D-CON was acquired both with the ^{13}C start variant (Bermel et al. 2006b) as well as with the 2D-(HCA)CON variant (Bermel et al. 2009) to ensure direct detection of proline ^{15}N resonances. 3D-(H)CBCACON and 3D-(H)CBCANCO (Bermel et al. 2009) were acquired with high resolution in all detected dimensions. Most relevant acquisition parameters are reported in Table 1.

Pulse lengths and carrier frequencies generally used for triple resonance experiments were used for the ^{13}C detected experiments and are summarized hereafter. The ^{1}H carrier was placed at 4.7 ppm for non-selective hard pulses. ^{13}C pulses were given at 176.7 ppm, 55.9 ppm, and 45.7 ppm for C′, C^{α} and C^{ali} regions, respectively.

Table 1 Experimental parameters used to collect the NMR experiments

Experiments	Dimension of acquired data			Spectral width (ppm)			NS[a]	d1 + aq (s)[b]	Spectrometer frequency (^{1}H) (MHz)
	t1	t2	t3	F1	F2	F3			
^{1}H detected									
^{1}H-^{15}N BEST-TROSY	512 (^{15}N)	9676 (^{1}H)		41	15		16	0.5	1200
BT-HNCACB	96 (^{13}C)	90 (^{15}N)	6144 (^{1}H)	75	41	14	96	0.2	1200
BT-HN(CO)CACB	96 (^{13}C)	80 (^{15}N)	6144 (^{1}H)	75	41	14	96	0.2	1200
HN(CA)CO	128 (^{13}C)	128 (^{15}N)	4096 (^{1}H)	7	28	18	8	1.0	1200
HNCO	128 (^{13}C)	220 (^{15}N)	4096 (^{1}H)	7	28	18	4	1.0	1200
^{13}C detected									
CON	512 (^{15}N)	1024 (^{13}C)		34	31		32	1.6	700
(HCA)CON	220 (^{15}N)	1024 (^{13}C)		40	31		16	0.9	700
(H)CACO	330 (^{13}C)	1024 (^{13}C)		34	30		32	1.0	700
(H)CBCACO	476 (^{13}C)	1024 (^{13}C)		59	30		32	1.0	700
(H)CBCACON	128 (^{13}C)	96 (^{15}N)	1024 (^{13}C)	58	34	30	4	1.0	700
(H)CBCANCO	96 (^{13}C)	96 (^{15}N)	1024 (^{13}C)	58	34	30	16	1.0	700

[a] Number of acquired scans

[b] Relaxation delay (acquisition time plus recovery delay d1)

^{15}N pulses were given at 124.0 ppm. Q5 and Q3 shapes (Emsley and Bodenhausen 1990) of durations of 300 and 231 µs, respectively, were used for ^{13}C band-selective π/2 and π flip angle pulses except for the π pulses that should be band-selective on the C$^{\alpha}$ region (Q3, 900 µs), and for the adiabatic π pulse (Böhlen and Bodenhausen 1993) to invert both C′ and C$^{\alpha}$ (smoothed Chirp 500 µs, 20% smoothing, 80 kHz sweep width, 11.3 kHz RF field strength). Composite pulse decoupling was applied on ^{1}H (Waltz-16) (Shaka et al. 1983) and ^{15}N (Garp-4) (Shaka et al. 1985) with an RF field strength of 3 kHz and 1 kHz respectively.

^{1}H detected experiments, acquired at 1.2 GHz, exploited the BEST-TROSY approach (3D-BT-HNCACB and 3D-BT-HN(CO)CACB) or the sensitivity enhanced approach (3D-HNCO and 3D-HN(CA)CO) for the 3D experiments. The 2D-[^{1}H, ^{15}N]-BEST-TROSY used sensitivity-enhanced gradient echo/antiecho coherence selection (Czisch and Boelens 1998; Schulte-Herbrüggen and Sørensen 2000) and Band-Selective Excitation Short-Transient (BEST) (Schanda et al. 2006; Lescop et al. 2007; Solyom et al. 2013) approach using exclusively shaped proton pulses. The inter-scan delay was set to 0.2 s. A 2D-[^{1}H, ^{15}N]-HSQC was also acquired in its fast version which exploits Watergate 3-9-19 pulses for water suppression (Mori et al. 1995). 3D-BT-HNCACB, and 3D-BT-HN(CO)CACB used echo/antiecho gradient selection and semi-constant time in the ^{15}N dimension (Schulte-Herbrüggen and Sørensen 2000; Solyom et al. 2013). 3D-HNCO and 3D-HN(CA)CO used sensitivity enhanced approach and selective pulse on the solvent for the water suppression (Kay et al. 1994). C′ and C$^{\alpha}$/C$^{\beta}$ selective excitation was exploited through band selective pulses.

Carrier frequencies used for triple resonance experiments in ^{1}H detected experiments were the same as for ^{13}C detected experiments except for the ^{15}N carrier placed at 118.0 ppm. Pulse shapes and lengths for ^{13}C band-selective pulses were G4 (Emsley and Bodenhausen 1992) and Q3 (Emsley and Bodenhausen 1990) shapes of durations of 205 and 128 µs, respectively, used for ^{13}C band-selective π/2 and π flip angle pulses except for the π pulses that should be band-selective on the C$^{\alpha}$ region (Q3, 525 µs). The ^{1}H band-selective pulses on the amide region were Pc9 (Kupce and Freeman 1994) or Eburp2 (Geen and Freeman 1991) for the π/2 and Reburp (Geen and Freeman 1991) or Bip (Smith et al. 2001) for π pulses.

All the spectra were acquired, processed, and analysed by using Bruker TopSpin 4.0.8 software. Chemical shifts were referenced using the ^{1}H and ^{13}C shifts of DSS. Nitrogen chemical shifts were referenced indirectly using the conversion factor derived from the ratio of NMR frequencies (Markley et al. 1998).

The sequence-specific assignment was performed with the aid of CARA (Keller 2004) and its tool NEASY (Bartels et al. 1995).

Bioinformatics tools

Several commonly utilized bioinformatics tools were used to predict or evaluate some of the protein features. Peculiarities of the distribution of intrinsic disorder predisposition along the amino acid sequence of the SARS-CoV-2 nucleocapsid protein N were evaluated by several members of the PONDR family (PONDR® VLXT (Romero et al. 2001), PONDR® VL3 (Obradovic et al. 2003), PONDR® VSL2 (Obradovic et al. 2005), and PONDR® FIT (Xue et al. 2010), together with the two versions of IUPred2A designed to predict short and long disordered regions (Mészáros et al. 2018).

The online tool ncSPC available at https://st-protein02.chem.au.dk/ncSPC/ was used to calculate the secondary structure propensity with the obtained assignment (Tamiola and Mulder 2012).

Assignments and data deposition

The 2D HN spectrum recorded on the IDR1-NTD-IDR2 (1–248) construct of the SARS-CoV-2 nucleocapsid protein N is shown in Fig. 2. The 2D HN spectrum clearly shows a set of well-resolved NMR signals deriving from the globular NTD domain, as one can verify by superimposing the available sequence-specific assignment (BMRB 34511, Dinesh et al. 2020). In addition, a set of signals, with smaller dispersion and higher intensity, are observed. These are expected to originate from the flexible and disordered fragments of the protein (black contours in Fig. 2).

The 2D CON spectrum (Fig. 3) provides information regarding the highly flexible and disordered protein regions. Due to the very different structural and dynamic properties of the globular NTD domain, with the chosen set-up the NMR signals of this region are very weak or absent in the 2D CON. This is exploited to selectively detect the resonances deriving from the two disordered protein regions. Proline residues can be directly monitored through the observation of the C′$_{i-1}$-N$_{i}$ correlations that fall in a very clean region of the CON spectrum ($132 < \delta(^{15}N) < 140$ ppm). The observation of only 7 well-resolved cross-peaks in this region (out of 17 expected for this construct) indeed confirms that C′ direct detection selectively picks up the signals of the disordered regions (5 proline residues present in the IDR1 region and 2 in the IDR2 one, Fig. 3 bottom squared region).

Sequence-specific assignment of the resonances can be performed by combining the information available in the 2D ^{13}C-detected spectra with that provided by two 3D experiments, the (H)CBCACON and the (H)CBCANCO (Bermel et al. 2009).

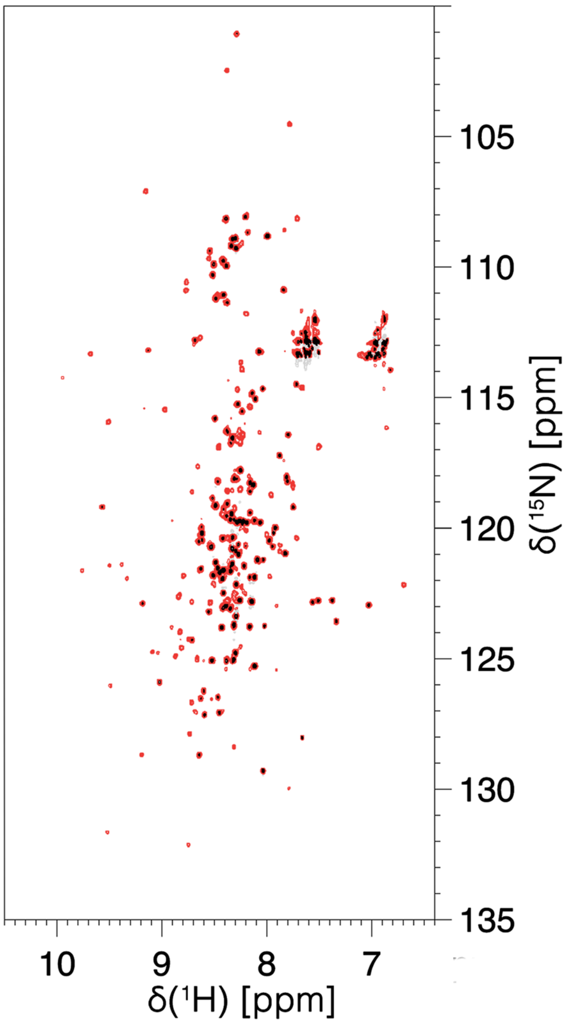


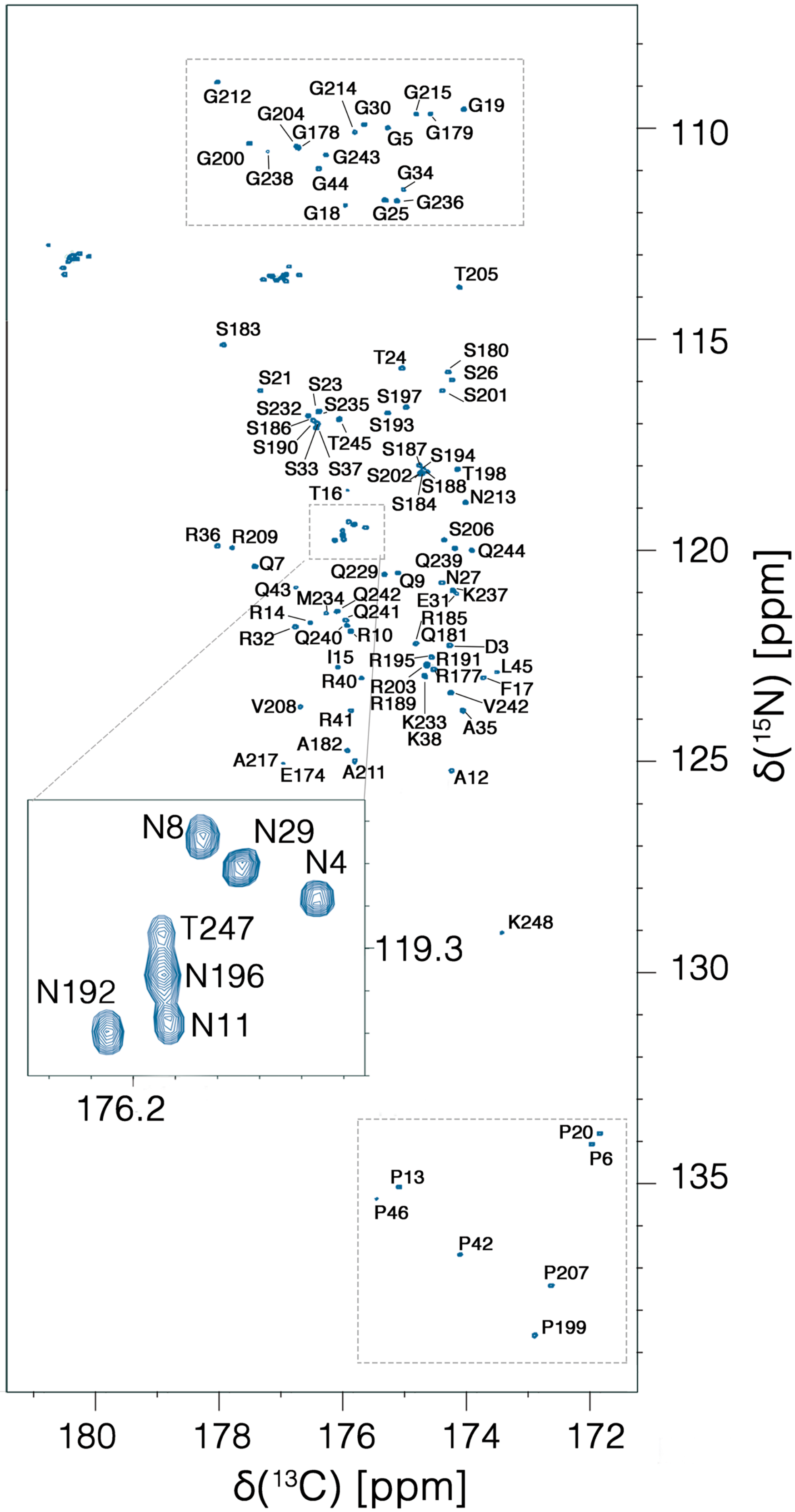
Fig. 2 The 2D HN BEST-TROSY of IDR1-NTD-IDR2 construct of the SARS-CoV-2 nucleocapsid protein. The figure shows the superimposition of two different processing of the same spectrum: the black one is optimized for the resolution and the red one is optimized for the signal to noise ratio. The spectrum was collected on a 28.3 T Bruker AVANCE NEO spectrometer operating at 1200.85 MHz ^{1}H, 301.97 MHz ^{13}C, and 121.70 MHz ^{15}N equipped with a 3 mm cryogenically cooled triple-resonance probehead (TCI)

It is worth noting that proline resonances provide a useful starting point for sequence-specific assignment. The particular ^{15}N chemical shift range expected for proline nitrogen signals (N_i) and the fact that this is correlated to resonances of the preceding amino acid (C'_{i-1}, C^{α}_{i-1}, C^{β}_{i-1}) through the 2D CON and 3D (H)CBCACON spectra constitute two features that allow us to unambiguously identify the type of dipeptide (X_{i-1}-Pro_i pair) that gives rise to specific signals as highlighted in Fig. 4. Indeed, the characteristic chemical shifts of C^{α} and C^{β} resonances enable us to recognize glycine, alanine, serine, and threonine residues; the remaining X-Pro pairs can then be easily identified as deriving from leucine and arginine residues by comparison with the primary sequence of the protein. Therefore, already at this very early stage of the sequence-specific assignment process, most of the observed resonances in this region could be assigned to specific amino acids uniquely considering the type of X-Pro pairs present in the intrinsically disordered regions (all resonances could be unambiguously assigned except for the two Gly-Pro pairs). Similarly, inspecting the opposite region of the CON spectrum at low ^{15}N chemical shifts (Fig. 3, top squared region) allows us to identify correlations involving ^{15}N nuclear spins of glycine residues; correlation to the carbonyl carbon of the previous amino acid (C'_{i-1}-N_i) contributes to an excellent resolution allowing us to count 16 resolved cross peaks in this region in the simple 2D mode. This is in line with the number of glycine residues present in the flexible disordered fragments. The classification of these resonances in X_{i-1}-Gly_i pairs achieved through inspection of the (H)CBCACON provides further input for their identification, as described above for the case of X_{i-1}-Pro_i pairs. Complete comparative analysis of the 3D (H)CBCACON and 3D (H)CBCANCO spectra enables the identification of the vast majority of the expected resonances of disordered regions. The excellent resolution obtained in the 2D reference spectra, the CON as well as the (H)CACO and (H)CBCACO, provides valuable support for the analysis of crowded regions of the spectra and to the discrimination between different residue types (Pontoriero et al. 2020).

The information retrieved for the intrinsically disordered regions of the spectra can then be used as a starting point to identify the spin systems also in $^{1}H^{N}$ detected 3D spectra. The latter are much more crowded due to more extensive cross-peak overlap, as well as because the signals of the globular region are also observed. In addition, cross peak intensities are highly heterogeneous due to the different structural and dynamic properties of the globular and disordered domains as well as due to the effects of solvent exchange processes. Therefore, the combined analysis of the two datasets greatly simplifies the identification of the signals deriving from the intrinsically disordered regions. As a further aid to discriminate the different sets of signals, spectra can be processed to enhance resolution, at the expense of signal-to-noise, taking advantage of the long-lived ^{15}N coherences of highly flexible regions of the protein as well as exploiting the long FID acquisition times that are possible through the BEST-TROSY approach (Schanda et al. 2006; Lescop et al. 2007; Solyom et al. 2013).

As a result, 98% of the disordered fragment IDR1 (only the first methionine is missing) (BMRB 50619) and 91% of the fragment IDR2 (BMRB 50618) could be assigned in a sequence-specific manner (C′, C^{α}, C^{β}, N, H^{N}) (vide infra). It is interesting to note how the combined use of these complementary datasets (^{13}C′- and $^{1}H^{N}$-detected 3D experiments) provides information that is particularly useful to achieve sequence-specific assignment of intrinsically disordered

Fig. 3 The 2D-CON of IDR1-NTD-IDR2 construct of the SARS-CoV-2 nucleocapsid protein. The high resolution provided by this experiment allows us to easily resolve resonances in the usually very crowded Gly-region (upper squared region) and to directly observe correlations involving proline residues (lower squared region). In the expansion shown in the center of the map the resolution of several repeating fragments comprising asparagine residues can be appreciated (the assignment reported is referred to the amide nitrogen of the mentioned amino acid). The spectrum was acquired on a 16.4 T Bruker AVANCE NEO spectrometer operating at 700.06 MHz ^{1}H, 176.05 MHz ^{13}C, and 70.97 MHz ^{15}N frequencies, equipped with a 5 mm cryogenically cooled probehead optimized for ^{13}C direct detection (TXO)



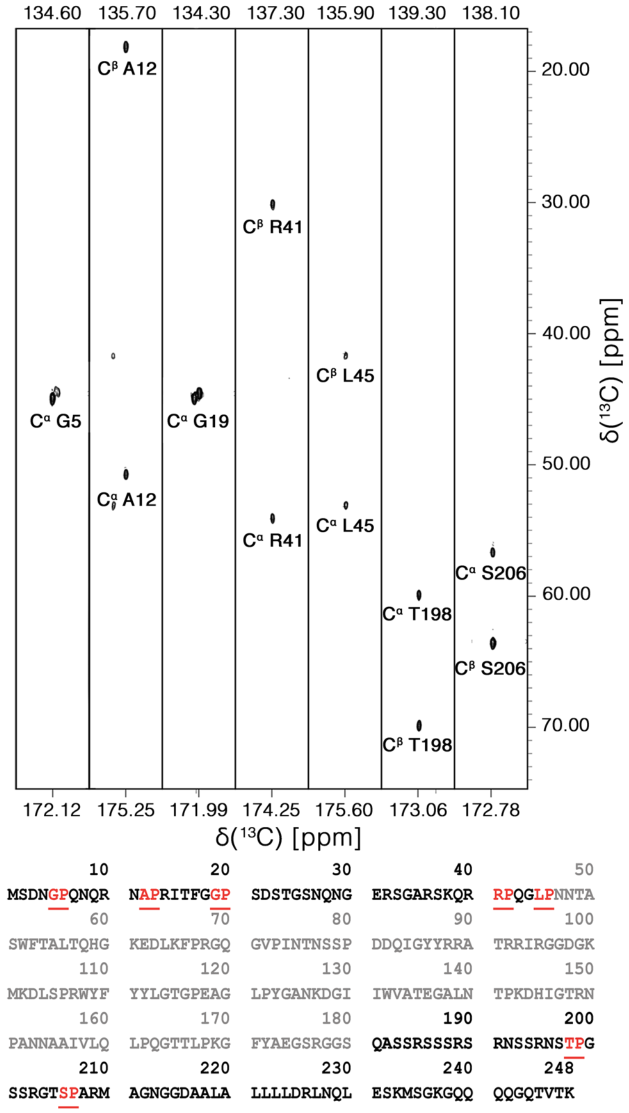


Fig. 4 Seven strips derived from the 3D-(H)CBCACON experiment extracted at the ^{15}N chemical shift of proline residues. The C′, C$^{\alpha}$ and C$^{\beta}$ frequencies belong to the preceding amino acid leading to the X-Pro assignment. The lower part of the figure reports the IDR1-NTD-IDR2 primary sequence in which X-Pro pairs are highlighted. Five proline residues are found in the IDR1 and two in IDR2 domain. The primary sequence of NTD domain is reported in grey. The 3D spectrum was acquired on a 16.4 T Bruker AVANCE NEO spectrometer operating at 700.06 MHz ^{1}H, 176.05 MHz ^{13}C, and 70.97 MHz ^{15}N frequencies, equipped with a 5 mm cryogenically cooled probehead optimized for ^{13}C direct detection (TXO)

regions also within highly heterogeneous proteins. The set of 2D spectra (HN, CON, (H)CACO, (H)CBCACO), provided they are acquired with high resolution, then becomes a very useful tool to achieve atomic resolution for the vast majority of the amino acids in the highly flexible disordered regions of complex, heterogeneous proteins.

The first two disordered regions of the N protein from SARS-CoV-2 (IDR1 and IDR2) can now be investigated at atomic resolution providing experimental information regarding the many interaction sites that can be predicted through different approaches (Kumar et al. 2008; Giri et al. 2020). The resonances of characteristic amino acids involved in interactions with RNA, such as arginine, serine, glutamine, and glycine residues, which are very abundant in the IDR1 and IDR2 disordered domains, can be detected and most of them can be resolved already in the 2D mode also at physiological pH and temperature conditions. Several signals in low complexity regions, such as the polyQ (238–242) or some repeats located in different positions in the primary sequence (for example the Asn-Arg region reported in the expanded panel in the middle of Fig. 3) can be resolved allowing their high-resolution investigation.

Chemical shifts were then used to determine secondary structural propensities as shown in Fig. 5. The data confirm the disordered nature of these fragments, with a moderate propensity to sample a helical conformation in the leucine-rich region (218–232), where few residues (Leu 221, Leu 222, Leu 223, Leu 224, Asp 225, Arg 226, and Leu 230) escaped detection likely because of the signal broadening due to conformational exchange. These experimental results are in agreement with the bioinformatics analysis reported in Fig. 1, which predicts a high extent of disorder for the two IDR regions as well as the presence of some structure in the region 215–232.

The NMR resonance assignments of the IDR1 and IDR2 domains of the N protein from SARS-CoV-2 open the way to understanding the role of these flexible parts of the nucleocapsid protein in modulating its function. The suite of ^{13}C detected 2D experiments (CON, (H)CACO, (H)CBCACO) in conjunction with 2D HN correlation experiments provide an excellent tool to monitor at atomic resolution their role in the interactions with RNA, with viral proteins or with proteins of the host, as well as with small molecules as potential drugs, opening the way to radically novel, unexplored approaches in drug discovery.

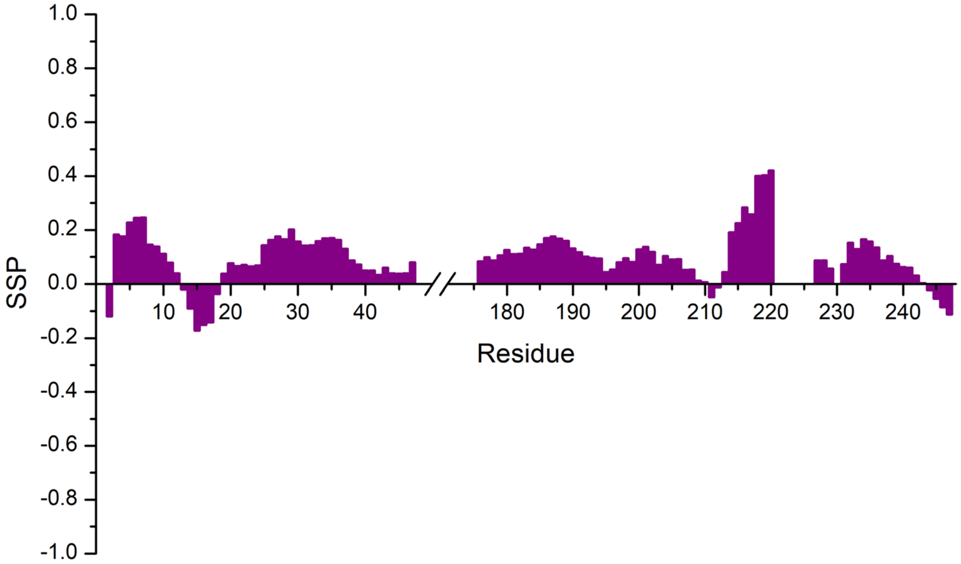


Fig. 5 Secondary Structure Propensity (SSP) plot obtained with the assignment reported on the BMRB (50619 and 50618) for the two assigned regions 1–47 and 176–248. Chemical shift values for H^N, N, C′, C^α, and C^β nuclei were used. The two regions result to be highly disordered with a slight tendency to be in an α-helix conformation for the residues 216–220

Acknowledgements The support and the use of resources of the CERM/CIRMMP centre of Instruct-ERIC is gratefully acknowledged. This work was supported in part by the Fondazione CR Firenze and by the Italian Ministry for University and Research (FOE funding). Fabio Almeida, Andreas Schlundt and Leonardo Gonnelli are acknowledged for the stimulating discussions.

Author contributions ICF and RP conceived the project and planned the experiments. LP purified and labelled the protein for NMR experiments. MS and LP acquired and analysed the data under the supervision of ICF and RP. VNU performed the bioinformatics analysis. All the authors contributed to writing the paper.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement. Fondazione CR Firenze; Italian Ministry for University and Research (FOE funding).

Availability of data and materials The chemical shift values for the ^{1}H, ^{13}C and ^{15}N resonances of the first two flexible linkers of the SARS-CoV-2 nucleoprotein have been deposited in the BioMagResBank (https://www.bmrb.wisc.edu) under accession number 50619 (IDR1, residues 1–47) and 50618 (IDR2, residues 176–248). Spectral raw data (upon request) and assignments are also accessible through https://covid19-nmr.de.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

Bartels C, Xia TH, Billeter M, Güntert P, Wütrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 6:1–10. https://doi.org/10.1007/BF00417486

Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R (2006a) Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. J Magn Reson 178:56–64. https://doi.org/10.1016/j.jmr.2005.08.011

Bermel W, Bertini I, Felli IC, Lee YM, Luchinat C, Pierattelli R (2006b) Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J Am Chem Soc 128:3918–3919. https://doi.org/10.1021/ja0582206

Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P (2009) H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. J Magn Reson 198:275–281. https://doi.org/10.1016/j.jmr.2009.02.012

Böhlen JM, Bodenhausen G (1993) Experimental aspects of chirp NMR spectroscopy. J Magn Reson Ser A 102:293–301. https://doi.org/10.1006/jmra.1993.1107

Chang CK, Sue SC, Yu TH, Hsien CM, Tsai CK, Chiang YC, Lee SJ, Hsiao HH, Wu WJ, Chang WL, Lin CH, Huang TH (2006) Modular organization of SARS coronavirus nucleocapsid protein. J Biomed Sci 13:59–72. https://doi.org/10.1007/s11373-005-9035-9

Chang CK, Hsu YL, Chang YH, Chao FA, Wu MC, Huang YS, Hu CK, Huang TH (2009) Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. J Virol 83:2255–2264. https://doi.org/10.1128/jvi.02001-08

Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH (2014) The SARS coronavirus nucleocapsid protein—forms and functions. Antiviral Res 103:39–50. https://doi.org/10.1016/j.antiviral.2013.12.009

Cui J, Li F, Shi ZL (2019) Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 17:181–192. https://doi.org/10.1038/s41579-018-0118-9

Czisch M, Boelens R (1998) Sensitivity enhancement in the TROSY Experiment. J Magn Reson 134:158–160. https://doi.org/10.1006/jmre.1998.1483

Dinesh DC, Chalupska D, Silhan J, Veverka V, Boura E (2020) Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. PLoS Pathog 16:e1009100. https://doi.org/10.1371/journal.ppat.1009100

Emsley L, Bodenhausen G (1990) Gaussian pulse cascades: new analytical functions for rectangular selective inversion and inphase excitation in NMR. Chem Phys Lett 165:469–476. https://doi.org/10.1016/0009-2614(90)87025-M

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. J Magn Reson 97:135–148. https://doi.org/10.1016/0022-2364(92)90242-Y

Felli IC, Pierattelli R (2012) Recent progress in NMR spectroscopy: toward the study of intrinsically disordered proteins of increasing size and complexity. IUBMB Life 64:473–481. https://doi.org/10.1002/iub.1045

Geen H, Freeman R (1991) Band-selective radiofrequency pulses. J Magn Reson 93:93–141. https://doi.org/10.1016/0022-2364(91)90034-Q

Giri R, Bhardwaj T, Shegane M, Gehi BR, Kumar P, Godhave K, Oldfield CJ, Uversky VN (2020) Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. Cell Mol Life Sci 25:1–34. https://doi.org/10.1007/s00018-020-03603-x

Goh GKM, Dunker AK, Uversky VN (2012) Understanding viral transmission behavior via protein intrinsic disorder prediction: coronaviruses. J Pathog 2012:738590. https://doi.org/10.1155/2012/738590

Goh GKM, Dunker AK, Uversky V (2013) Prediction of intrinsic disorder in MERS-CoV/HCoV-EMC supports a high oral-fecal transmission. PLoS Curr 5:1–25. https://doi.org/10.1371/currents.outbreaks.22254b58675cdebc256dbe3c5aa6498b

Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Kang S, Yang M, Hong Z, Zhang L, Huang Z, Chen X, He S, Zhou ZZ, Chen Q, Yan Y, Zhang C, Shan H, Chen S (2020) Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. Acta Pharm Sin B 10:1228–1238. https://doi.org/10.1016/j.apsb.2020.04.009

Kay LE, Xu GY, Yamazaki T (1994) Enhanced-Sensitivity triple-resonance spectroscopy with minimal H_2O saturation. J Magn Reson Ser A 109:129–133. https://doi.org/10.1006/jmra.1994.1145

Keller R (2004) The computer aided resonance assignment tutorial. Goldau, Switz. Cantina Verlag 1–81

Kumar M, Gromiha MM, Raghava GPS (2008) Prediction of RNA binding sites in a protein using SVM and PSSM profile. Proteins Struct Funct Genet 71:189–194. https://doi.org/10.1002/prot.21677

Kupce E, Freeman R (1994) Wideband excitation with polychromatic pulses. J Magn Reson Ser A 108:268–273. https://doi.org/10.1006/jmra.1994.1123

Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. J Magn Reson 187:163–169. https://doi.org/10.1016/j.jmr.2007.04.002

Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sukes BD, Wrigth PE, Wütrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. Pure Appl Chem 70:117–142. https://doi.org/10.1023/A:1008290618449

Masters PS (2006) The molecular biology of coronaviruses. Adv Virus Res 65:193–292. https://doi.org/10.1016/S0065-3527(06)66005-3

Mészáros B, Erdös G, Dosztányi Z (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res 46:W329–W337. https://doi.org/10.1093/nar/gky384

Mori S, Abeygunawardana C, Johnson MO, Vanzijl PCM (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new Fast HSQC (FHSQC) detection scheme that avoids water saturation. J Magn Reson Ser B 108:94–98. https://doi.org/10.1006/jmrb.1995.1109

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK (2003) Predicting intrinsic disorder from amino acid sequence. Proteins Struct Funct Genet 53:566–572. https://doi.org/10.1002/prot.10532

Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins Struct Funct Genet 61:176–182. https://doi.org/10.1002/prot.20735

Peng Y, Du N, Lei Y, Dorje S, Qi J, Luo T, Gao GF, Sonh H (2020) Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. EMBO J 39:1–12. https://doi.org/10.15252/embj.2020105938

Pontoriero L, Schiavina M, Murrali MG, Pierattelli R, Felli IC (2020) Monitoring the interaction of α-synuclein with calcium ions through exclusively heteronuclear nuclear magnetic resonance experiments. Angew Chem Int Ed 59:18537–18545. https://doi.org/10.1002/anie.202008079

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Ak D (2001) Sequence complexity of disordered protein. Proteins 42:38–48. https://doi.org/10.1002/1097-0134(20010101)42:1<3c38::aid-prot50>3e3.0.co;2-3

Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043. https://doi.org/10.1021/ja062025p

Schulte-Herbrüggen T, Sørensen OW (2000) Clean TROSY: compensation for relaxation-induced artifacts. J Magn Reson 144:123–128. https://doi.org/10.1006/jmre.2000.2020

Shaka AJ, Keeler J, Freeman R (1983) Evaluation of a new broadband decoupling sequence: WALTZ-16. J Magn Reson 53:313–340. https://doi.org/10.1016/0022-2364(83)90035-5

Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552. https://doi.org/10.1016/0022-2364(85)90122-2

Smith MA, Hu H, Shaka AJ (2001) Improved broadband inversion performance for NMR in liquids. J Magn Reson 151:269–283. https://doi.org/10.1006/jmre.2001.2364

Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR 55:311–321. https://doi.org/10.1007/s10858-013-9715-0

Surjit M, Lal SK (2008) The SARS-CoV nucleocapsid protein: a protein with multifarious activities. Infect Genet Evol 8:397–405. https://doi.org/10.1016/j.meegid.2007.07.004

Tamiola K, Mulder FAA (2012) Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. Biochem Soc Trans 40:1014–1020. https://doi.org/10.1042/BST20120171

Wang D, Hu B, Hu C et al (2020) Clinical characteristics of 138 hospitalized patients with 2019 novel cvoronavirus–infected pneumonia in Wuhan. China JAMA 323:1061. https://doi.org/10.1001/jama.2020.1585

Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 1804:996–1010. https://doi.org/10.1016/j.bbapap.2010.01.011

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