



Partial structure, dampened mobility, and modest impact of a His tag in the SARS-CoV-2 Nsp2 C-terminal region

Miguel Mompeán¹  · Miguel Á. Treviño¹ · Douglas V. Laurents¹ 

Received: 27 July 2021 / Revised: 10 September 2021 / Accepted: 3 October 2021 / Published online: 11 October 2021
© The Author(s) 2021

Abstract

Intrinsically disordered proteins (IDPs) play essential roles in regulating physiological processes in eukaryotic cells. Many viruses use their own IDPs to “hack” these processes to deactivate host defenses and promote viral growth. Thus, viral IDPs are attractive drug targets. While IDPs are hard to study by X-ray crystallography or cryo-EM, atomic level information on their conformational preferences and dynamics can be obtained using NMR spectroscopy. SARS-CoV-2 Nsp2, whose C-terminal region (CtR) is predicted to be disordered, interacts with human proteins that regulate translation initiation and endosome vesicle sorting. Molecules that block these interactions could be valuable leads for drug development. The ¹³Cβ and backbone ¹³CO, ¹HN, ¹³Cα, and ¹⁵N nuclei of Nsp2’s 45-residue CtR were assigned and used to characterize its structure and dynamics in three contexts; namely: (1) retaining an N-terminal His tag, (2) without the His tag and with an adventitious internal cleavage, and (3) lacking both the His tag and the internal cleavage. Two five-residue segments adopting a minor extended population were identified. Overall, the dynamic behavior is midway between a completely rigid and a fully flexible chain. Whereas the presence of an N-terminal His tag and internal cleavage stiffen and loosen, respectively, neighboring residues, they do not affect the tendency of two regions to populate extended conformations.

Keywords SARS-CoV-2 · Nsp2 · NMR spectroscopy · Intrinsically disordered proteins

Abbreviations

CtDR	C-terminal disordered region
IDP	Intrinsically disordered protein
Nsp2	Non-structural protein 2
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2

Introduction

The pandemic provoked by SARS-CoV-2 is sparking unprecedented efforts in the development of vaccines. Nevertheless, drugs which block the activity of viral proteins should also be sought, since SARS-CoV-2 vaccines may well lack availability and complete efficacy, particularly for the immune compromised or against novel strains. Intrinsically disordered proteins (IDPs) play essential roles in

regulating physiological processes in eukaryotic cells by forming weak protein/protein interactions. Many viruses use their own IDPs to trick these processes to neutralize host defenses and promote viral growth (Davey et al. 2011). For example, the small HIV protein Tat (Transactivator of Transcription) is intrinsically disordered (Shojania and O’Neil 2006) and is key to express viral genes; in addition, it manipulates cell signalling networks to downregulate apoptosis and alters cytokine expression (Clark et al. 2016). Because of their key functions, viral IDPs, such as HIV Tat (Hamy et al. 2000), are attractive drug targets. While the flexibility of IDPs generally precludes their study by X-ray crystallography or cryo-EM, atomic level information on their conformational preferences and dynamics can be obtained using NMR spectroscopy.

Using bioinformatics tools, a C-terminal region of SARS-CoV-2 Nsp2 has been predicted to be disordered (Giri et al. 2021). Interestingly, this 45-residue Nsp2 region appears to be more disordered in SARS-CoV-2 than in its close homologs human SARS (responsible for the 2003 outbreak) and bat coronavirus (see Sup. Figure 9 in Giri et al. 2021). A study (Gordon et al. 2020a) of the SARS-CoV-2 interactome reported that Nsp2 interacts with human proteins GIGYF2

✉ Douglas V. Laurents
dlaurents@iqfr.csic.es

¹ “Rocasolano” Institute for Physical Chemistry, Spanish National Research Council, Serrano 119, 28006 Madrid, Spain

and EIF4E2, which regulate translation initiation, as well as WASHC4 and FKBP15, which are involved in endosome vesicle sorting. This investigation has been recently extended by comparative analyses with SARS-CoV-1 and MERS-CoV interactomes and implicated Nsp2 in exosomes, cellular respiration, lipid transport (Gordon et al. 2020b). Nsp2 from human SARS was also reported to interact with Prohibitin1/2, a cell proliferation modulator with tumor suppression activity (Cornilley-Ty et al. 2009). Molecules that block these interactions could be valuable leads for drug development.

Some months after a preprint of this study was made public (Mompeán et al. 2020), a preliminary study was posted on BioRxiv which reported a medium resolution structural model of the complete SARS CoV-2 Nsp2 protein based on CryoEM and AI (Gupta et al. 2021). In their report, electron density for the C-terminal region was either missing (which strongly suggests disorder) or indicated a folded domain rich in β -structure depending on the conditions. Based on analysis of Nsp2 mutants, Gupta et al. (2021) also proposed that E63, E65, G262, G265, K330, and K337 are key for binding host proteins. No Nsp2 C-terminal residues/host protein interactions have yet been identified. The main objective of this study is to characterize the conformational preferences and dynamics of the putatively disordered C-terminal region of SARS CoV-2 Nsp2 (Nsp2 CtDR) using NMR spectroscopy. The assignments obtained could also guide future studies of interactions between Nsp2, human target proteins, and inhibitors.

Finally, His tags are a popular tool for modern protein purification. They are usually removed after the first purification step, but sometimes they are occluded and difficult to remove. His tags left on folded proteins have been reported to influence biochemical activity (Majorek et al. 2014) or increase protein rigidity (Thielges et al. 2011), but their effects on disordered proteins are less well characterized. To address this knowledge gap, we also report the NMR characterization of an Nsp2 CtDR construct with an N-terminal His tag. The secondary structure and dynamics of an additional third sample without the His tag but that had suffered an internal cleavage, possibly due to enterokinase, are also reported.

Results

Bioinformatics analysis

The C-terminal region of Nsp2 was recently predicted to be disordered by Giri et al. (2021), who applied the PONDR[®] and IUPRED algorithms. Nevertheless, this region contains a rather high proportion of hydrophobic residues which generally promote order and folding; in particular its content

of Phe, Ile, Leu, Met, Tyr, Val, and Trp residues is 33.3%, which is close to their average content in folded proteins (32.2%) and much higher than their average content in IDPs (22.0%) (Uversky 2013). To corroborate these findings, we have applied the same tools as Giri et al. (2021) as well as the PrDOS disorder prediction algorithm (Ishida and Kinoshita 2007) to the Nsp2 sequence (Fig. 1). Whereas the

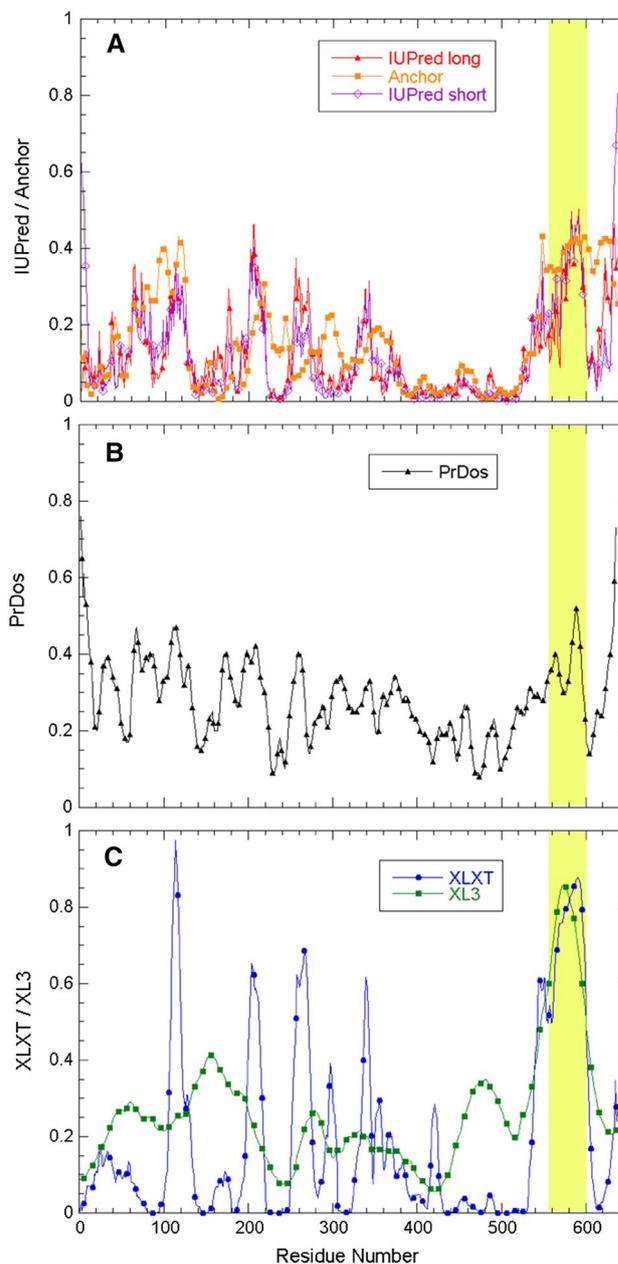


Fig. 1 SARS-CoV-2 Nsp2. Predicted disordered region and primary structure. **A** Sequence-based disorder tendency scores for Nsp2 calculated with IUPred and Anchor, **B** PrDos, and **C** PONDR[®] XLXT and XL3. The region studied here is highlighted in light yellow. **D** Nsp2 primary structure (severe acute respiratory syndrome coronavirus 2). NCBI reference sequence: YP_009742609.1

results show that the C-terminal region spanning approximately residues 555–600 have a high disorder propensity as compared to the rest of the protein, the score falls short of the threshold for some algorithms used by Giri et al. (2021) and is close to the PrDOS threshold for disorder (Fig. 1). Considering the distinct predictions of these bioinformatics tools, and the fact that disorder prediction programs are occasionally wrong (Treviño et al. 2018), we have used NMR spectroscopy to characterize structurally the putatively disordered region of Nsp2.

NMR assignment and assessment of partial structure

The $^{13}\text{C}\beta$ and backbone ^{13}CO , ^1HN , $^{13}\text{C}\alpha$, and ^{15}N nuclei were assigned by analysis of a series of 2D ^1H - ^{15}N HSQC and ^{13}C - ^{15}N CON as well as 3D HNCOC, HNCA, CBCAcoNH, and HncocaNH (Pantoja-Uceda and Santoro 2009) spectra. The degree of completeness of the assignments is reported for the three samples studied and are reported in Table 1 and assignments are deposited in the BMRB database under access code 50,687.

The three assigned 2D ^1H - ^{15}N HSQC spectra of the Nsp2 CtDR in the presence or absence of an N-terminal His tag and with or without an internal cleavage are shown in Fig. 2. The low ^1HN signal dispersion observed in all three spectra suggests that the Nsp2 CtDR is chiefly disordered. The $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, and ^{13}CO conformational chemical shifts ($\Delta\delta$) values

are shown in Fig. 3. Overall, the $\Delta\delta$ values are low, which evinces that the Nsp2 CtDR is mostly unfolded. However, two five-residue stretches spanning residues E_{570} - VLTE_{574} and S_{591} EAVE_{595} tend to adopt β -stands or extended conformations. Their populations within the conformational ensemble are about 10–11% for E_{570} - VLTE_{574} and 12–14% for S_{591} EAVE_{595} . In addition, a few discontinuous residues at the N-terminus of this region also show some tendency to adopt β -strand or extended conformations (Fig. 3). It is important to point out that the sample whose His tag was cleaved contained an additional proteolytic cleavage between residues K_{579} and T_{580} , which was identified on the basis of chemical shift alterations, the absence of an ^1H - ^{15}N crosspeak for T_{580} and the relaxation measurements reported below. This break occurs between the two partly ordered segments mentioned above. Since this break would disrupt hydrogen bonding and stabilizing contacts between the two extended conformations if they were to adopt interacting β -strands, and since populations for E_{570} - E_{574} and S_{591} - E_{595} are similar in both samples, this strongly suggests that the conformational chemical shifts arise from extended conformations instead of β -rich secondary structure.

Dynamics

To assess the extent and contributions of internal motions to the dynamics, we measured relaxation rates in the rotating time frame ($R_{1\rho}$) for the Nsp2 CtDR. This experiment

Table 1 NMR spectral parameters

Experiment	Number of scans	Sweep width (ppm)	Matrix		
Assignment					
1D ^1H	48	12	32 k		
2D ^1H - ^{15}N HSQC	4	11 $^1\text{H}\times 20$ ^{15}N	2 k \times 256		
2D CON ^a	8	11 $^{13}\text{CO}\times 34$ ^{15}N	1 k \times 512		
3D HNCOC ^a	8	11 $^1\text{H}\times 20$ $^{15}\text{N}\times 12$ ^{13}C	2 k \times 48 \times 256		
3D HNCA ^a	8	11 $^1\text{H}\times 20$ $^{15}\text{N}\times 25$ $^{13}\text{C}\alpha$	2 K \times 48 \times 128		
3D CBCAcoNH ^a	8	2 k $^1\text{H}\times 20$ $^{15}\text{N}\times 55$ ^{13}C	2 k \times 48 \times 128		
3D HncocaNH	8	10 $^1\text{H}\times 20$ $^{15}\text{N}\times 10$ ^1H	2 k \times 64 \times 64		
Relaxation					
2D HSQCCTEFT3GPSI ($T_{1\rho}$) ^b	8	10 $^1\text{H}\times 20$ ^{15}N	2 k \times 384		
2D HSQCNOEFGPSI $\{^1\text{H}\}$ - ^{15}N NOE ^c	16	10 $^1\text{H}\times 20$ ^{15}N	2 k \times 384		
Completeness of NMR assignments					
!Sample/Nuclei- >	$^{13}\text{C}\alpha$	$^{13}\text{C}\beta$	^{13}CO	^{15}N	^1HN
+ His tag	82%	82%	82%	92%	100%
No His tag, cut	98%	87%	84%	98%	98%
No His tag, not cut	100%	89%	93%	93%	98%

^aRecorded with non-uniform sampling. The % of sparse sampling was 35% for the 3D HNCOC, HNCA, and CBCAcoNH experiments and 25% for the 2D CON experiment

^bDelays used (in seconds): 0.008, 0.300, 0.036, 0.076, 0.900, 0.150, 0.500, 1.350

^cRecycle delay = 10 s

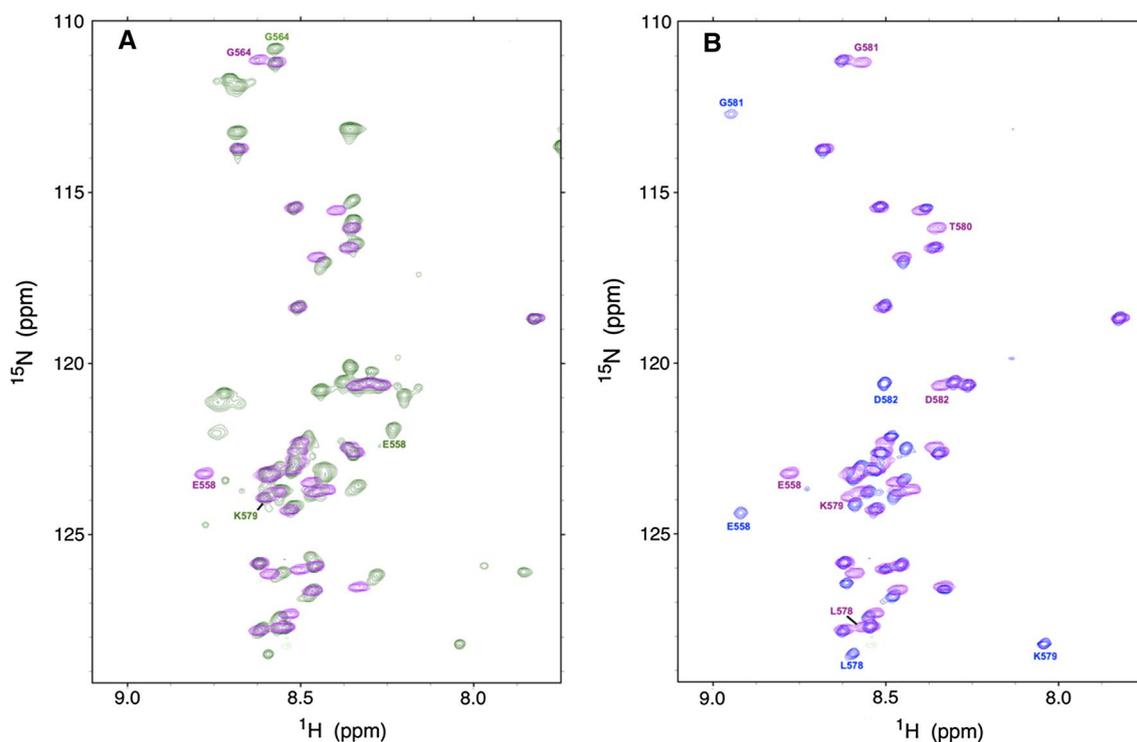


Fig. 2 2D ^1H - ^{15}N HSQC NMR spectra of Nsp2 CtDR in 5 mM KPi, 10 mM NaCl, pH 6.3, 5 °C. **A** Prior to cleavage of the His tag (green) and following removal of the His tag (purple). Peaks with distinct chemical shifts are labeled according to their numbering in the full length Nsp2 protein. Peaks arising from the N-terminal His tag are

not labeled: The His tag sequence is: MAHHHHHHGTGTGSND-
DDD-K. **B** Following His tag cleavage (purple) and with the region containing an additional break between K_{579} and T_{580} (blue). Peaks with the most relevant chemical shift changes are labeled. Each spectrum is shown separately and labeled in Sup. Fig. 1

is sensitive to both fast (ps-ns) and slow (μs -ms) internal motions arising from exchange. Generally low values are found, reflecting flexibility and fast dynamics in the ps-ns timescale. A closer inspection reveals that the N-terminal half of the sequence features increased $R_{1\rho}$ relaxation rates that decrease towards the C-terminus. The higher values in the N-terminus indicate dampened mobility of this region, likely due to conformational exchange in the μs -ms timescale for this segment. In contrast, unrestricted, fast dynamics are observed towards the C-terminus, with gradually decreasing $R_{1\rho}$ relaxation rates (Fig. 4A). The presence of the His tag seems to further stiffen the N-terminal region of the polypeptide. However, beyond the first ten residues, its effect is negligible and the same dynamic behavior is observed, with decreasing $R_{1\rho}$ relaxation rates towards the C-terminus that feature similar relaxation rate values. These results together suggest that the N-terminal part exhibits dampened mobility, more so when extended upstream by the His tag, and that the mobility is progressively less restricted towards the C-terminus. By contrast, the presence of an internal cleavage that yields two fragments has a dramatic effect on this trend, and results in two fragments that show faster internal motions. In particular,

the average $R_{1\rho}$ rates decrease from 4.5 (no His tag) to 3.3 s^{-1} and 2.6 s^{-1} (no His tag, but internally cleaved, N- and C-ter fragments, respectively) (Fig. 4A). These results are point to a length dependence of the potential conformational exchange that underlies restricted mobility, i.e. higher $R_{1\rho}$ rates, in the N-terminus of uncleaved constructs.

On fast ps-ns timescales, the $\{^1\text{H}\}$ - ^{15}N NOE measurements show that the presence of the N-terminal His tag also appears to increase the rigidity of the polypeptide in the first several residues; thereafter, it does not seem to alter the dynamics significantly (Fig. 4B). The presence of the cleavage moderately increases the flexibility in resulting C-terminal fragment, however, on this timescale, the break does not substantially affect the dynamics of residues of the resulting N-terminal fragment (Fig. 4B). Overall the $\{^1\text{H}\}$ - ^{15}N NOE ratios range between 0.2 and 0.4, except for residues near termini which show lower values (Fig. 4B). Whereas these values are lower than typical ratios seen in the rigid elements of folded proteins (0.7–0.8), they are significantly higher than those observed in some other IDPs such as α -synuclein (Masaracchia et al. 2020).

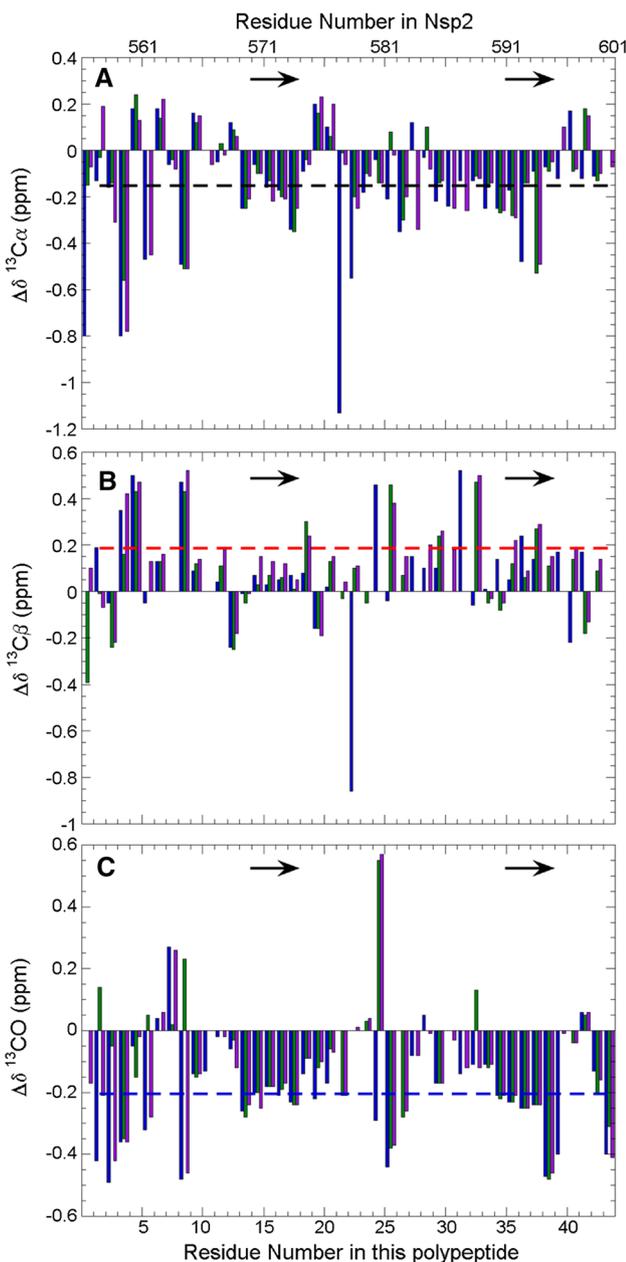


Fig. 3 Conformational shift analysis for Nsp2 CtDR in 5 mM KPi, 10 mM NaCl, pH 6.3, 5 °C. Nsp2 CtDR conformational chemical shifts of **A** $^{13}\text{C}\alpha$ (black), and **B** $^{13}\text{C}\beta$ (red) and **C** ^{13}CO (blue). Values for the Nsp2 CtDR prior to His tag cleavage (green bars), and following His tag cleavage with (blue) or without (purple) an additional cleavage between K_{579} and T_{580} are shown. The dashed black, red, and blue lines mark the values expected for 10% β -strand for $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, and ^{13}CO , respectively. The position of two modestly populated β -strands are marked by arrows

Discussion

Despite recent findings that SARS-CoV-2 Nsp2 is undergoing positive nature selection and is thus important to the virus (Flores-Alanis et al. 2021), this protein is a relatively

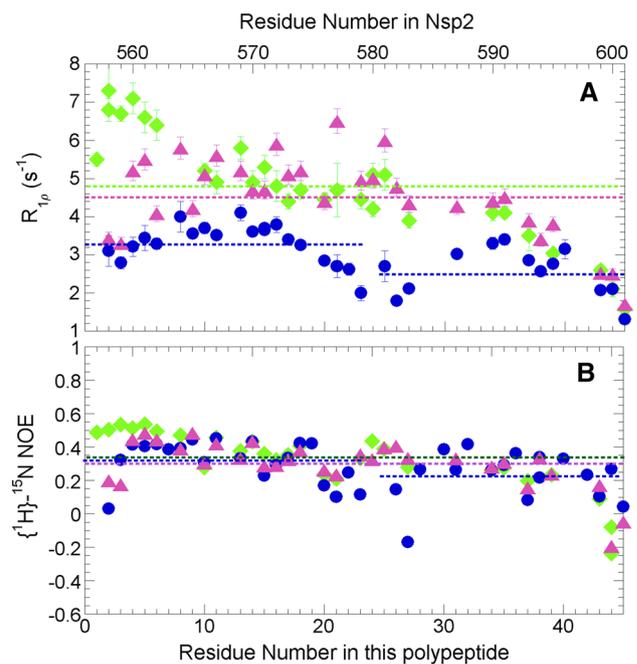


Fig. 4 Residue level dynamics from ^{15}N relaxation. **A** Relaxation rates in the rotating frame ($R_{1\rho}$) for the Nsp2 CtDR with (green) the His tag present, with the His tag and with a cleavage after residue K_{579} (blue) and without the His tag and without the cleavage (purple). In this panel, as well as **B** the mean values for each sample, and both fragments of the cut sample, are represented by the dashed lines with the same green, blue, purple color code. **B** $\{^1\text{H}\}$ - ^{15}N NOE ratio values for the Nsp2 CtDR with (green) the His tag present, with the His tag and with a cleavage after residue K_{579} (blue) and without the His tag and without the cleavage (purple). As the calculated errors are less than 0.02, no error bars are shown as they are smaller than the symbols. The hNOE ratio values are less than values usually seen in folded proteins' rigid regions (0.70–0.80) or flexible zones (0.60–0.65) but higher than values typical of highly flexible segments (≤ 0)

uncharacterized. Here, we show that the C-terminal region of SARS-CoV-2 Nsp2 is intrinsically disordered, which is in general agreement with bioinformatics predictions (Giri et al. 2021). Nevertheless, two five-residue stretches show a small, but significant tendency to adopt extended/ β -strand conformations, and the region is more rigid than a completely disordered protein chain. These findings may be attributed to the high content of β -branched or bulky residues as well as a high density of negatively charged residues which favored extended or β -conformations (Minor and Kim 1994; Zhou and Pang 2018). These β -strands may be present in the folded form of the Nsp2 C-terminal region in a recently advanced structural model (Gupta et al. 2021), although a direct comparison is not possible since no residue level information on this region could be discerned from this medium (5–6 Å) resolution model. These results seem not to depend on the presence of an N-terminal His tag or an internal proteolytic cleavage, which suggests these extended conformations may form independently. Hypothetically, the

C-terminal region, which appears to be rather autonomous with respect to the rest of the protein structure (Gupta et al. 2021) might interact electrostatically with the second and third domains, which are rich in positively charged residues.

Nsp2 has been found to interact with many human proteins and functional complexes (Cornillez-Ty et al 2009; Gordon et al 2020a, b) and some residues in its N-terminal domains have already been implicated in these interactions (Gupta et al. 2021). The Nsp2 CtDR might also participate in interactions with human proteins, as suggested by the relatively high ANCHOR score (Fig. 1) and future experiments will test whether their degree of structure and rigidity increase upon binding. Blocking such an interaction might be an excellent target for future pharmaceuticals. Inhibitor screening of the Nsp2 CtDR in collaboration with the COVID19-NMR consortium is underway. The assigned ^1H - ^{15}N spectrum could serve in the near future to identify where first-generation inhibitors bind and thereby help guide their improvement.

His tags are a common tool for facilitating recombinant protein purification and they are usually removed prior to conformational characterization. The results presented here suggest that this removal may not always be necessary for disordered polypeptides poor in anionic residues since its effect on protein conformation and dynamics is limited to the first neighboring residues. Here, the His tag may serve to dampen mobility of the first residues of the C-terminal region in a manner analogous to what would occur in full length Nsp2. Similar results have been recently reported in a series of eight 100-residue polypeptides corresponding the disordered, prion-like domain of human CPEB3 (Ramírez de Mingo et al. 2020). However, these results may not be general for folded proteins considering that a 2D infrared spectroscopy characterization of His-tagged myoglobin reported an overall decrease in ps dynamics (Thielges et al. 2011) and may affect biological function (Majorek et al. 2014).

An internal proteolytic cleavage seems to impact protein dynamics strongly; in particular on μs -ms timescales its effect is felt throughout the length of the resulting C-terminal fragment of SARS-CoV-2 Nsp2. The study of the structure and dynamics of wild type human frataxin and a C-terminal truncated form also reported more extensive dynamics changes on slower μs -ms timescales (Faraj et al. 2014).

Materials and methods

Bioinformatics

Algorithms IUPred2A and ANCHOR (Mészáros et al. 2018; Erdős and Dosztánzi 2020), which are based on energy estimations, were used to predict the disordered regions in the

Nsp2 sequence. Two versions of IUPred2A, best suited for short and long disordered sequences, were used. The ANCHOR algorithm seeks to detect disordered sequences that undergo folding upon binding. Furthermore, the PrDOS (Ishida and Kinoshita 2007) as well as the XLXT and XL3 algorithms from the PONDR[®] suite (Obradovic et al. 2003) which are neural networks optimized with distinct protein training sets and using different attributes, were also utilized. In all cases, the default threshold values for disordered residues and other program settings were used.

Sample production and isotopic labeling

To enable inhibitor screening and to uncover conformational preferences and dynamics, we have expressed and purified the ^{13}C , ^{15}N -labeled C-terminal region of Nsp2. The full production methodology of this and other SARS-CoV-2 proteins has been recently published (Altincekic et al. 2021). Briefly, the DNA sequence coding for the segment $\text{K}_{557}\text{EIIIFLEGETLPTEVLTEEVVLK-TGDLQPLEQPTSEAVEAPLVGT}_{601}$, from the C-terminus of Nsp2 was synthesized (GenScript, New Jersey, USA) and cloned in plasmid pET45b also coding for an N-terminal hexaHis tag and enterokinase cleavage site or in a pET28a-derived plasmid coding for an N-terminal thioredoxin domain, a hexaHis tag and a TEV-protease cleavage site. These plasmids were transformed in *E. coli* BL21star(DE3) and were expressed in minimal media containing $^{15}\text{NH}_4\text{Cl}$ and ^{13}C -glucose as the sole sources of nitrogen and carbon, respectively, as previously described (Treviño et al. 2018). The domain was purified using the His tag/ Ni^{++} immobilized metal affinity chromatography (IMAC), digested with Enterokinase (NEB Biolabs, MA, USA) or TEV-protease produced in-house using the protocol by Blommel and Fox (2007). Afterwards, the protein was separated from the His tag as the flow-through from IMAC and then subjected to a final ion exchange purification step.

NMR spectroscopy

Samples for NMR spectroscopy contained 1.2–1.5 mM ^{13}C , ^{15}N -labeled Nsp2 CtDR at pH 6.3 in 5.0 mM $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ buffer with 10.0 mM NaCl, 10% D_2O and 50 μM sodium trimethylsilylpropanesulfonate (DSS) as the internal chemical shift reference. These conditions were chosen as an optimal compromise between the low pH and low ionic strength conditions that afford the best quality NMR spectra and physiological conditions. Three series of spectra were recorded at 5 °C: (1) one with an N-terminal His tag whose sequence is MAHHHHHHGTGTGSNDDDD-K, (2) without the His tag and containing an internal cleavage between K_{579} and T_{580} , and (3) without the N-terminal His tag and without the internal cleavage. Based on the sequence and using the “Peptide Cutter Tool” in the ExPASy website (<https://web>.

expasy.org), trypsin may have caused the adventitious cleavage between K₅₇₉ and T₅₈₀. However, it is more likely to have resulted from the intrinsic broad specificity of enterokinase (Chio et al. 2001).

In contrast to cryo-EM which typically employs much colder temperature, i.e. liquid ethane (− 90 °C), NMR spectroscopy can be applied at near physiological temperatures. Here, we chose to use 5 °C instead of the physiological temperature of 37 °C to reduce exchange of the HN groups, allowing more protein residues to be studied. Moreover, previous studies in our laboratory (Ramírez de Mingo et al. 2020) and others (Abyzov et al. (2016) have shown that whereas slightly higher populations and modestly more rigidity is observed at 5 °C, similar conformational and dynamics trends in IDPs are observed at 5 °C as at higher temperatures.

The series of 2D ¹H-¹⁵N HSQC and ¹³C-¹⁵N CON as well as 3D HNCO, HNCA, CBCAcoNH, and HncocaNH (Pantoja-Uceda and Santoro 2009) spectra were recorded on a Bruker Neo 800 MHz (¹H) spectrometer equipped with a cryoprobe and z -gradients. The NMR spectral parameters are listed in Table 1. The spectra were transformed with Topspin 4.0.8 (Bruker Biospin) and were assigned manually with the aid of the program Sparky (Lee et al. 2015).

Calculation of conformational chemical shifts

Following the assignment of the experimental ¹³C α and ¹³CO chemical shifts, the chemical shift values expected for random coil were calculated using the approach of Poulsen and coworkers (Kjaergaard and Poulsen 2011; Kjaergaard et al. 2011) at pH 6.3, 5 °C using the webserver: https://spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS/. These values were subtracted from the experiment chemical shift values to obtain the conformational chemical shift values shown in Fig. 3. To estimate the % population of extended conformations, the conformational chemical shift averaged over five residues was divided by the value of − 1.54 ppm for ¹³C α (the average of − 1.48 (Spera and Bax 1991) and − 1.6 ppm (Santiveri et al 2001) and − 2.21 ppm for ¹³CO (Wishart and Skyes 1994) and multiplied by 100.

Dynamics

A series of 2D ¹H-¹⁵N HSQC-based experiments were recorded to determine the {¹H}-¹⁵N NOE and R₁ ρ relaxation rates to assess dynamics on the ps/ns timescale as well as potential contributions from slow (μ s/ms) exchange. Peak integrals were measured using Topspin 4.0.8. R₁ ρ experiments used a spin lock of 2 kHz, with 8 relaxation delays varying from 8 to 600 ms, and the corresponding relaxation rates were obtained using the program KaleidaGraph (Synergy Software, version 3.6) to fit an exponential equation:

$I(t) = I_0 \cdot \exp(-k \cdot t) + I_\infty$, where $I(t)$ is the peak integral at time t , k is the rate, and I_∞ is the intensity at infinite time to the data. The R₁ ρ uncertainties reported here are those obtained from the least squares fit. In the case of the {¹H}-¹⁵N heteronuclear NOE experiment, two (¹H-saturated and non-saturated) experiments are interleaved with a ten second recycling delay from which the corresponding {¹H}-¹⁵N NOE ratios are obtained. The uncertainties were calculated as the ratio of the noise (estimated as the standard deviation of the integral of several areas lacking peaks) times $\sqrt{2}$ to the peak integral measured without applying the NOE.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00249-021-01575-9>.

Acknowledgements NMR experiments were performed in the “Manuel Rico” NMR Laboratory (LMR) of the Spanish National Research Council (CSIC), a node of the Spanish Large-Scale National Facility for Biomolecular NMR (ICTS R-LRB). We thank Dr. David Pantoja-Uceda for help with NMR spectroscopy.

Author contributions MAT: experimental design, protein expression, and production; MM: experimental design, data acquisition, and analysis; DVL: experimental design, data acquisition and analysis. All authors contributed to writing the MS and approved the final version.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. MM is a Ramón y Cajal Fellow of the Spanish AEI-Ministry of Science and Innovation (RYC2019-026574-I), and a “La Caixa” Foundation (ID 100010434) Junior Leader Fellow (LCR/BQ/PR19/11700003). Funded by project COV20/00764 from the Carlos III Institute of Health and the Spanish Ministry of Science and Innovation to MM and DVL.

Declarations

Conflict of interest The authors declare no conflicts of interest or competing interests.

Availability of data and material The obtained chemical shift data are deposited in the BMRB database under access code 50687. Requests for materials can be addressed to DVL.

Code availability Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication All authors approve this version of the MS for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abyzov A, Salvi N, Schneider R, Maurin D, Ruigrok RWH, Jensen MR, Blackridge M (2016) Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation. *J Am Chem Soc* 138(19):6240–6251. <https://doi.org/10.1021/jacs.6b02424>
- Altincekic N, Korn SM, Qureshi NS, Dujardin M et al (2021) Large-scale recombinant production of the SARS-CoV-2 proteome for high-throughput and structural biology applications. *Front Mol Biosci*. <https://doi.org/10.3389/fmolb.2021.653148>
- Blommel PG, Fox BG (2007) A combined approach to improving large-scale production of tobacco etch virus protease. *Prot Expr Purf* 55(1):53–68
- Chio SI, Song HW, Moon JW, Seong BL (2001) Recombinant enterokinase light chain with affinity tag expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology. *Biotech & Bioeng* 75(6):718–724
- Clark E, Nava B, Caputi M (2016) Tat is a multifunctional viral protein that modulates cellular gene expression and function. *Oncotarget* 8(16):27569–27581
- Cornillez-Ty CT, Liao L, Yates JR, Kuhn P, Buchmeier MJ (2009) Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *J Virol* 83:10314–10318
- Davey NE, Travé G, Gibson TJ (2011) How viruses hijack cell regulation. *Trends Biochem Sci* 36(3):159–169. <https://doi.org/10.1016/j.tibs.2010.10.002>
- Erdős G, Dosztányi Z (2020) Analyzing protein disorder with IUPred2A. *Curr Proto Bioinfo* 70:e99. <https://doi.org/10.1002/cpbi.99>
- Faraj SE, Roman EA, Aran M, Gallo M, Santos J (2014) The alteration of the C-terminal region of human frataxin distorts its structural dynamics and function. *FEBS J* 281:3397–3419. <https://doi.org/10.1111/febs.12869>
- Flores-Alanis A, Cruz-Rangel A, Rodríguez-Gómez A, González J, Torres-Guerrero A, Delgado G, Cravioto R, Morales-Espinosa R (2021) Molecular epidemiology surveillance of SARS-CoV-2 mutations and genetic diversity one year after emerging. *Pathogens* 10:184
- Giri R, Bhardwaj T, Shegane M, Gehi BR, Kumar P, Gadhav K, Oldfield J, Uversky VN (2021) Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. *Cell Mol Life Sci Cell Mol Life Sci* 78:1655–1688. <https://doi.org/10.1007/s00018-020-03603-x>
- Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K et al (2020a) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583(7816):459–468. <https://doi.org/10.1038/s41586-020-2286-9>
- Gordon DE, Hiatt J, Bouhaddou M, Rezelj VV et al (2020b) Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* 370(6521):9403. <https://doi.org/10.1126/science.abe9403>
- Gupta M, Azumaya CM, Moritz M, Pourmal S et al (2021) CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2 a multifunctional protein involved in key host processes. *bioRxiv*. <https://doi.org/10.1101/2021.05.10.443524>
- Hamy F, Gelus N, Zeller M, Lazdins JL, Bailly C, Klimkait T (2000) Blocking HIV replication by targeting Tat protein. *Chem Biol* 7(9):669–676
- Ishida T, Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* 35:W460–464
- Kjaergaard M, Poulsen FM (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR* 50(2):157–165
- Kjaergaard M, Brander S, Poulsen FM (2011) Random coil chemical shifts for intrinsically disordered proteins: Effects of temperature and pH. *J Biomol NMR* 49(2):139–149
- Lee W, Tonelli M, Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinf (oxf Engl)* 31(8):1325–1327. <https://doi.org/10.1093/bioinformatics/btu830>
- Majorek KA, Kuhn ML, Chruszcz M, Anderson WF, Minor W (2014) Double trouble—buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments. *Prot Sci* 23:1359–1368
- Masracchia C, König A, Valiente-Gabioud AA, Peralta P et al (2020) Molecular characterization of an aggregation-prone variant of alpha-synuclein used to model synucleinopathies. *BBA Prot Proteomics* 1868:140298
- Mészáros B, Erdős G, Dosztányi Z (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nuc Acids Res* 46:W329–337
- Minor DL, Kim PS (1994) Measurement of the β -sheet-forming propensities of amino acids. *Nature* 367:660–663
- Mompeán M, Treviño MÁ, Laurents DV (2020) Towards targeting the disordered SARS-CoV-2 Nsp2 C-terminal regions: partial structure and dampened mobility revealed by NMR spectroscopy. *BioRxiv*. <https://doi.org/10.1101/2020.11.09.374173>
- Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK (2003) Predicting intrinsic disorder from amino acid sequences. *Prot Struct Funct Gene* 53:566–572
- Pantoja-Uceda D, Santoro J (2009) Aliasing in reduced dimensionality NMR spectra: (3,2)D HNHA and (4,2)D HN(COCA)NH experiments as examples. *J Biomol NMR* 45(4):351–356
- Ramírez de Mingo D, Pantoja-Uceda D, Hervás R, Carrión-Vázquez M, Laurents DV (2020a) Preferred conformations in the intrinsically disordered region of human CPEB3 explain its role in memory consolidation. *BioRxiv*. <https://doi.org/10.1101/2020.05.12.091587>
- Santiveri CM, Rico M, Jimenez MA (2001) $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts as a tool to delineate β -hairpin structures in peptides. *J Biomol NMR* 19:331–345
- Shojania S, O’Neil D (2006) HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat(1–72) by NMR spectroscopy. *J Biol Chem* 281(13):8347–8356
- Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and $\text{C}\alpha$ and $\text{C}\beta$ ^{13}C nuclear magnetic resonance chemical shifts. *J Am Chem Soc* 113:5490–5492
- Thielges MC, Chung JK, Axup JY, Fayer MD (2011) Influence of histidine tag attachment on picosecond protein dynamics. *Biochemistry* 50:5799–5805. <https://doi.org/10.1021/bi2003923>
- Treviño MÁ, Pantoja-Uceda D, Menéndez M, Gomez MV, Mompeán M, Laurents DV (2018) The singular NMR fingerprint of a polyproline II helical bundle. *J Am Chem Soc* 140(49):16988–17000. <https://doi.org/10.1021/jacs.8b05261>
- Uversky VN (2013) The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord Proteins* 1(1):24684. <https://doi.org/10.4161/idp.24684>

- Wishart DS, Skyes BD (1994) The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical shift data. *J Biomol NMR* 4:171–180
- Zhou H-X, Pang X (2018) Electrostatic interactions in protein structure, folding, binding and condensation. *Chem Rev* 118:1691–1741

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