JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JACS

Article

Atomic-Resolution Structure of SARS-CoV-2 Nucleocapsid Protein N-Terminal Domain

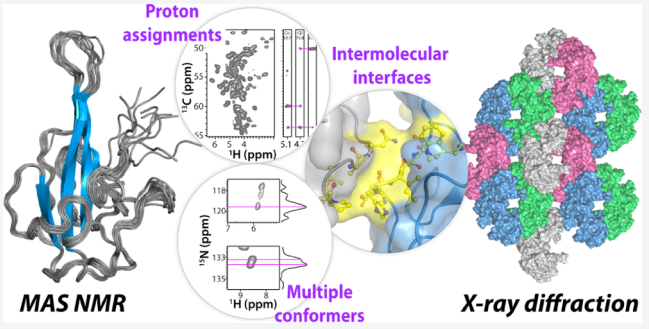
Sucharita Sarkar,[‖] Brent Runge,[‖] Ryan W. Russell, Kumar Tekwani Movellan, Daniel Calero, Somayeh Zeinalilathori, Caitlin M. Quinn, Manman Lu, Guillermo Calero, Angela M. Gronenborn,* and Tatyana Polenova*

Cite This: https://doi.org/10.1021/jacs.2c03320

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The nucleocapsid (N) protein is one of the four structural proteins of the SARS-CoV-2 virus and plays a crucial role in viral genome organization and, hence, replication and pathogenicity. The N-terminal domain (N^{NTD}) binds to the genomic RNA and thus comprises a potential target for inhibitor and vaccine development. We determined the atomic-resolution structure of crystalline N^{NTD} by integrating solid-state magic angle spinning (MAS) NMR and X-ray diffraction. Our combined approach provides atomic details of protein packing interfaces as well as information about flexible regions as the N- and C-termini and the functionally important RNA binding, β-hairpin loop. In addition, ultrafast (100 kHz) MAS ^{1}H-detected experiments permitted the assignment of side-chain proton chemical shifts not available by other means. The present structure offers guidance for designing therapeutic interventions against the SARS-CoV-2 infection.



INTRODUCTION

SARS-CoV-2, a positive-sense single-stranded RNA virus from the β-coronavirus family,[1] is the causative agent of the COVID-19 pandemic that killed millions of people and brought the world economy to a grinding halt.[2,3] The SARS-CoV-2 genome encodes four structural proteins: spike (S) glycoprotein, envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein.[4,5] All play crucial roles in the viral life cycle and pathogenicity, including host immunity evasion.[6] Due to its important role in genome packaging and ribonucleoprotein (RNP) formation, the N protein represents a potential target for therapeutic interventions.[7−9]

The N protein comprises two folded domains, the N-terminal (N^{NTD}, residues 40−174) and C-terminal (N^{CTD}, residues 246−365) domains, connected by an ∼70 amino acid linker region that contains a 13-residue serine/arginine motif, as well as extensive intrinsically disordered regions (IDRs) at the N- and C-termini[10−15] (Figure 1a). All domains including the N^{NTD} play an important role in the RNA genome interaction.[16−18]

Several structures of β-coronavirus N^{NTD} domains have been reported, all of which possess the same architecture, resembling a right hand.[19−24] The core structure is made of a four-stranded antiparallel β-sheet, the palm, from which the β2, β3 hairpin prominently protrudes. It contains several basic residues, and this basic finger and the palm have been implicated in RNA binding.[25] The loop connecting β2 and β3 is flexible, in agreement with the missing density in this region of most X-ray structures (see below).[19,26] The N-terminal disordered tail projects outward and may contribute to RNA binding.[19]

Here, we report the atomic-resolution structure of crystalline N^{NTD}, determined by combining X-ray crystallography and solid-state magic angle spinning (MAS) NMR spectroscopy. The protein crystallized in the $P2_12_12_1$ space group with four chains in the asymmetric unit, and the X-ray structure was solved at a 1.7 Å resolution. The MAS NMR structure of an individual N^{NTD} chain, at 0.7 Å rmsd resolution, was determined using a single crystalline U−^{13}C,^{15}N−N^{NTD} sample, based on 2968 nonredundant ^{13}C−^{13}C, ^{15}N−^{13}C, and ^{15}N−^{1}H distance restraints. Several inter-chain contacts were identified in ^{13}C−^{13}C correlation experiments, both for chains in the asymmetric unit as well as across asymmetric units. Side-chain proton chemical shifts were assigned from high-frequency (100 kHz) MAS NMR correlation experiments and provided important structural information, such as the tautomeric state of the H107 residue. Our results illustrate the

Received: March 28, 2022

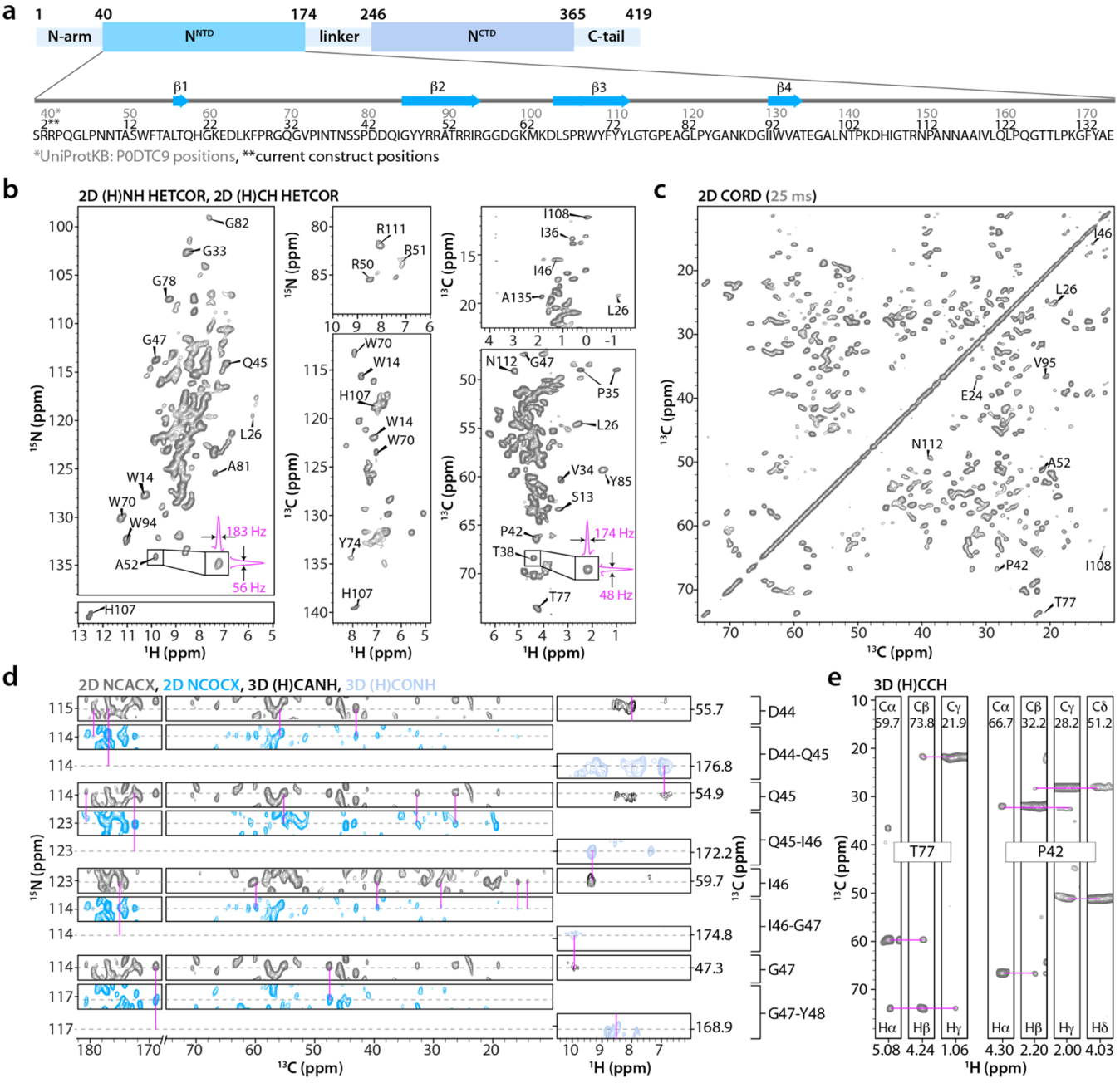


Figure 1. Domain delineation, amino acid sequence, and magic angle spinning (MAS) NMR spectra used for resonance assignment of SARS-CoV-2 N^{NTD}. (a) Top: domain organization of SARS-CoV-2 nucleocapsid (N) protein; N-terminal domain (N^{NTD}) and C-terminal domain (N^{CTD}). Bottom: N^{NTD} primary sequence and β-strands (blue arrows); residues 2−136 (black) in the current N^{NTD} construct (this work) correspond to residues 40−174 (gray) in the full-length N protein. (b) Selected regions of ^{1}H-detected two-dimensional (2D) (H)NH HETCOR and (H)CH HETCOR spectra of U−^{13}C,^{15}N−N^{NTD}. The expansions around the A52 and T38 cross-peaks in (H)NH and (H)CH spectra depict one-dimensional (1D) slices to illustrate the line widths in the two frequency dimensions. (c) Aliphatic region of the 2D CORD spectrum (25 ms mixing time). (d) Sequential assignments for the D44−G47 stretch of residues are illustrated with representative strips of 2D NCACX (gray), NCOCX (blue), ^{1}H-detected three-dimensional (3D) (H)CANH (black), and (H)CONH (pale blue) spectra. (e) Selected strips from the ^{1}H-detected 3D (H)CCH spectrum for residues T77 and P42 illustrating ^{13}C and ^{1}H side-chain resonance assignments. CORD, NCACX, and NCOCX spectra were recorded at an MAS frequency of 14 kHz; (H)CANH and (H)CONH spectra were acquired at an MAS frequency of 60 kHz; and HETCOR and (H)CCH spectra were acquired at an MAS frequency of 100 kHz. The number of scans and the number points in the direct and indirect dimensions are as follows: 2D (H)NH HETCOR - 32 scans, 1024 t_2 points, 1034 t_1 points; 2D (H)CH HETCOR - 64 scans, 1024 t_2 points, 2310 t_1 points; 2D CORD - 192 scans, 2048 t_2 points, 840 t_1 points; 2D NCACX - 2048 scans, 2048 t_2 points, 96 t_1 points; 2D NCOCX - 1536 scans, 3072 t_2 points, 96 t_1 points; 3D (H)CANH - 48 scans, 2048 t_3 points, 112(^{15}N) t_2 points, 32(^{13}C) t_1 points; 3D (H)CONH - 32 scans, 2048 t_3 points, 112(^{15}N) t_2 points, 32(^{13}C) t_1 points; and 3D (H)CCH - 8 scans, 1024 t_3 points, 264 (^{13}C) t_2 points, 264 (^{13}C) t_1 points.

power of integrating orthogonal structural techniques, here MAS NMR and X-ray diffraction, for assessing details of protein conformations. The atomic-resolution structure of crystalline N^{NTD} reported here will guide the development of small-molecule inhibitors and biologics for treatment as well as biosensors for the detection of SARS-CoV-2 infection.[27]

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

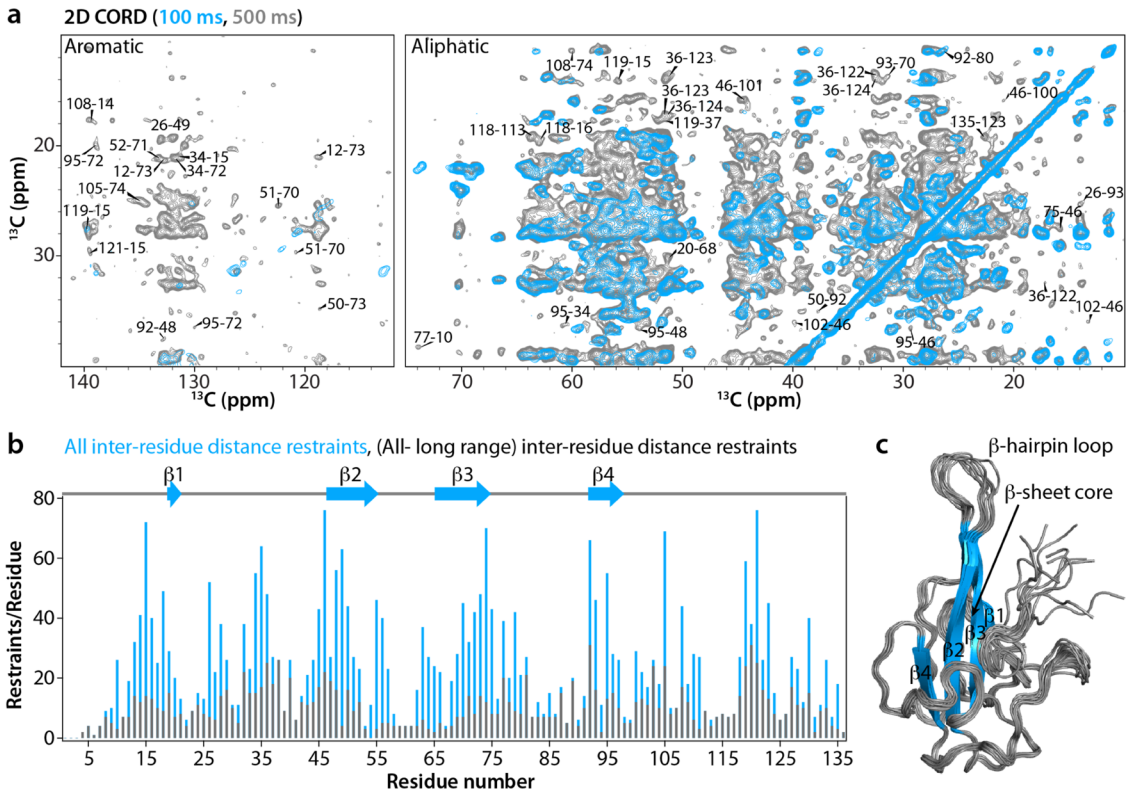


Figure 2. Correlation spectra, inter-residue distance restraints, and MAS NMR structure of a single N^{NTD} chain. (a) Superposition of representative regions of 2D CORD spectra of U−^{13}C,^{15}N−N^{NTD} acquired with the mixing times of 100 ms (blue) and 500 ms (gray). Aromatic and aliphatic regions are shown in the left and right panels, respectively. Representative cross-peaks between amino acids are labeled by residue numbers. (b) The number of all inter-residue distance restraints and long-range inter-residue distance restraints are plotted for each residue along the polypeptide chain. (c) Superposition of the 10 lowest-energy MAS NMR structures of a single chain of SARS-CoV-2 N^{NTD}. β-strands are colored in blue and labeled. The number of scans and the number points in the direct and indirect dimensions are as follows: 2D CORD (100 ms mixing time) - 96 scans, 3072 t_2 points, 840 t_1 points; 2D CORD (500 ms mixing time) - 192 scans, 3072 t_2 points, 667 t_1 points.

RESULTS

Resonance Assignments. Chemical shift assignments and distance restraints for N^{NTD} structure calculation were obtained using a single sample of fully protonated crystalline U−^{13}C,^{15}N−N^{NTD} comprising residues 40−174 (current construct residues 2−136; Figure 1a; see the Materials and Methods section for experimental details). A total of eleven 2D and three 3D ^{1}H- and ^{13}C-detected high-frequency (100 and 60 kHz) MAS NMR experiments were recorded (Figure 1b−e and Supporting Information Table S1). The spectra are of remarkably high resolution, with line widths as narrow as 35 Hz for ^{15}N, 48 Hz for ^{13}C, and 174 Hz for $^1H^N$ (Figure 1b).

2D CORD,[28] NCACX, and NCOCX at a 25 ms mixing time, as well as ^{1}H-detected 2D (H)NH HETCOR, 3D (H)CANH, and (H)CONH spectra (Figure 1c,d), were used for sequential backbone assignments, and ^{13}C and ^{15}N chemical shifts are complete for 128 of 136 residues. For five residues, F28, P84, P113, P124, and E136, partial backbone chemical shift assignments were obtained, and, for 119 residues, backbone amide proton (H^N) chemical shifts were assigned. The resonances of the first two residues, R2 and R3, are missing in the spectra, likely due to disorder. Overall, good agreement is observed between ^{1}H and ^{15}N chemical shifts determined in this work and those reported previously from solution NMR.[25,29] MAS NMR assignments for a representative stretch of residues D44−G47 are illustrated in Figure 1d.

Side-chain ^{13}C chemical shifts and inter-residue correlations were obtained from 2D NCACX, NCOCX, and CORD spectra, the latter acquired with 25, 100, and 500 ms mixing times (Figures 1c and 2a). High spectral resolution permitted the unambiguous assignment of numerous cross-peaks, including those corresponding to aliphatic-to-aromatic (left panel) and aliphatic-to-aliphatic (right panel) side-chain correlations (Figure 2a). To determine side-chain and backbone ^{1}H chemical shifts, a 3D (H)CCH correlation experiment was recorded at an MAS frequency of 100 kHz (Figure 1e). In conjunction with spectra acquired at an MAS frequency of 60 kHz, 84 side-chain proton resonances for 71 residues and H^α resonances for 65 residues were assigned. For 11 Ala, 3 Val, 4 Ser, and 1 His residues, complete ^{13}C, ^{15}N, and ^{1}H backbone and side-chain chemical shifts were obtained. Overall, assignments for 132 residues were attained (Figure S1) on the basis of 3728 cross-peaks in various spectra (Table 1). All chemical shifts are summarized in Table S2 of the Supporting Information.

Gratifyingly, many side-chain protons of aromatic residues could be unambiguously assigned from the ^{1}H-detected 100 kHz MAS NMR spectra (Figure 1b). For example, for W70 and W94, located in the β-sheet core and assumed to be involved in RNA binding, side-chain protons were assigned fully (W70) or partially (W94). Moreover, the tautomeric state of H107 was determined (see below).

Structure of a Single N^{NTD} Chain Determined by MAS NMR. The structure of an N^{NTD} single chain was calculated using 2968 nonredundant distance and 101 ϕ/ψ torsion angle restraints. Of these, 2197 are unambiguous ^{13}C−^{13}C, 763 are

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Table 1. Summary of Samples and the Number of Assigned Peaks

	no. of assigned peaks[a]
U−^{13}C,^{15}N−N^{NTD} (MAS NMR)	
intraresidue	1943
sequential ($\|i - j\| = 1$)	495
medium range ($1 < \|i - j\| < 5$)	306
long range ($\|i - j\| \geq 5$)	972
long range ($\|i - j\| \geq 5$) (interchain)	12
total assigned peaks (MAS NMR)	**3728**
U−^{15}N−N^{NTD} (solution NMR)	
intraresidue	159
total assigned peaks	**3887**

[a]Cross-peaks present in different experiments are counted only once.

^{15}N−^{13}C, and 4 are ^{1}H−^{15}N distance restraints, including 968 long-range ($|i - j| \geq 5$) restraints (Table 2 and Figure S2 of the Supporting Information). The number of restraints per residue is plotted in Figure 2b. The 10 lowest-energy MAS NMR structures in the structural ensemble and an average structure of a single chain of N^{NTD} are shown in Figures 2c and S3 of the Supporting Information, respectively. All MAS NMR distance restraints are summarized in Table 2. With nearly 22 restraints/residue on average, the N^{NTD} structure determined in this study represents a notable technical advance being one of only two MAS NMR structures of proteins larger than 100 residues per chain determined using more than 20 restraints/residue and reaching the maximum accuracy and precision attained for MAS NMR protein structures[30] (see below).

Table 2. Summary of MAS NMR Restraints and Structure Statistics

MAS NMR distance restraints	^{13}C−^{13}C	^{15}N−^{13}C	^{1}H−^{15}N
unambiguous	2197	763	4
intraresidue	807	505	0
sequential ($\|i - j\| = 1$)	119	258	4
medium range ($1 < \|i - j\| < 5$)	303	0	0
long range ($\|i - j\| \geq 5$)	968	0	0
ambiguous	4		
total number of restraints assigned	**2968** (21.8 restraints per residue)		
MAS NMR Dihedral Angle Restraints			
Φ	101		
Ψ	101		
Structure Statistics from 10 Lowest-Energy Subunits			
violations (mean ± sd)			
distance restraints ≥ 7.2 Å (Å)	0.144 ± 0.001		
dihedral angle restraints ≥ 5° (deg)	1.528 ± 0.137		
max. distance restraint violation (Å)	1.254		
max. dihedral angle restraint violation (deg)	17.267		
deviations from idealized geometry			
bond lengths (Å)	0.008 ± 0.000		
bond angles (deg)	0.774 ± 0.012		
impropers (deg)	0.516 ± 0.016		
average pairwise rmsd (Å)[a]			
backbone (N, C$^{\alpha}$, C′)	**0.7 ± 0.2**		
heavy	**1.2 ± 0.1**		

[a]Disordered N-terminus (residues 1−9) excluded.

Like all coronavirus N^{NTD} structures,[16,19,24−26,31,32] the MAS NMR-derived structure exhibits the overall shape of a right hand, made up of a four-stranded β-sheet, comprising β1 (L18−T19), β2 (I46−R55), β3 (D65−Y74), and β4 (I92−A96). At its center, a long β-hairpin protrudes out from the palm (Figure 2c). The irregular regions at the N- and C-termini exhibit well-defined backbone and side-chain orientations in the MAS NMR structure, except for the first eight amino acids (R2−N9) and the last residue (E136) (see Figures S2 and S3 of the Supporting Information). The lack of long-range inter-residue distance restraints for the N-terminal tail residues (P4−N9) and β-hairpin loop (I56−K64) suggests that they are dynamic (Figure 2b). The precision of the single-chain MAS NMR structure is 0.7 ± 0.2 Å, as measured by the pairwise atomic backbone rmsd for the 10 lowest-energy structures (excluding the disordered N-terminal tail, residues R2−N9) (Table 2 and Figure S3 of the Supporting Information).

X-ray Crystal Structure of the N^{NTD}. The protein crystallized in the $P2_12_12_1$ orthorhombic space group with four monomers (chains A−D) in the asymmetric unit (Figures 3a and S4a of the Supporting Information). Two views of the four chains are provided in Figure 3a, and chain A is depicted in Figure 3b. Details of the β2, β3 hairpin and loop region, as well as the difference electron density map, are shown in Figure 3c. Complete statistics for X-ray data collection, phasing, and refinement are provided in Table 3. The average pairwise rmsd value between the four chains in the asymmetric unit is 0.5 ± 0.1 Å for the backbone atoms (excluding common missing residues in all four chains, R2−N9, Q20−D25, R57−P68, and P124−E136) (Figure S5b and Table S3 of the Supporting Information). A positively charged region, comprising arginine residues in the β-sheet (R50, R51, R54, R55, R69) and at the tip of the β-hairpin finger (R57, K62), may possibly contribute to RNA binding (Figure S4c of the Supporting Information).[19,25]

Interesting details about the intratetramer interfaces in the crystal can be noted in the structure, with five unique types of contacts formed by the residues within the tetramer (Figure 4a). Specifically, (i) the A−B interface comprises several residues (T18, H21, I36, R54, R56−G59, K64, L66, S67, V120, Q122, Y134, and A135) from both chain A and chain B; (ii) the A−C interface is very small and involves residues G59 and D60 of chain A contacting K131 of chain C; (iii) the B−C interface comprises several residues, such as R30−Q32, P42, D43, E98−L101, and P124−T127 of the chain B palm region, which are in contact with the C-terminal tail (residues L121−L129, K131, G132, Y134, A135), as well as H21, G22, and K23 of chain C; (iv) the B−D interface packs the palm regions of chains B and D against each other; and (v) the D−C interface comprises residues I56 and P113−A117 of chain D and T16, H21, A117, A118, V120, and Q122 of chain C.

The X-ray and the MAS NMR structures of the individual chains are in good agreement, with a backbone rmsd of 1.1 Å between the X-ray structure (averaged over the four chains in the asymmetric unit) and the MAS NMR structure (averaged over the ensemble of 10 lowest-energy structures) (Figure S5 and Table S3 of the Supporting Information). Upon exclusion of chain D, which possesses the highest degree of disorder in the X-ray structure, the corresponding value becomes 0.7 Å. The average pairwise backbone rmsd between the four chains in the X-ray structure and within the ensemble of 10 lowest-energy MAS NMR structures are, both, 0.5 ± 0.1 Å (Table S3

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

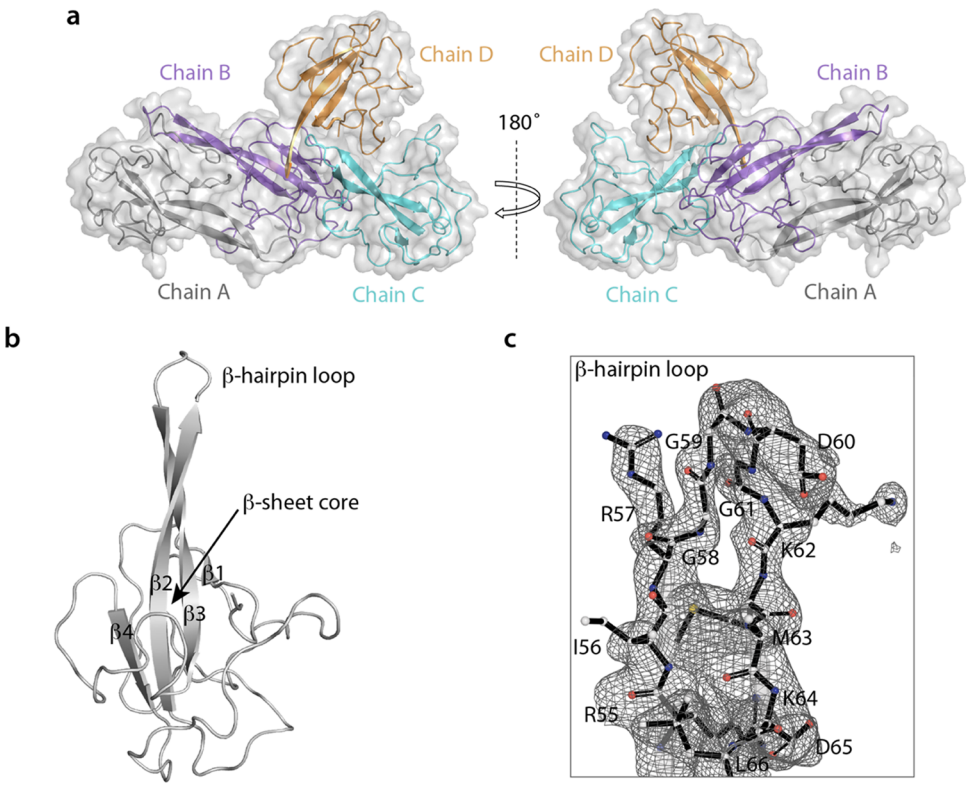


Figure 3. X-ray crystal structure of SARS-CoV-2 N^{NTD}. (a) Ribbon and surface representation of the four N^{NTD} chains in the asymmetric unit shown in two different orientations (PDB: 7UW3); chain A (gray), chain B (purple), chain C (cyan), and chain D (orange). (b) Structure of chain A (ribbon representation) with the strands in the β-sheet core labeled $\beta1-\beta4$. (c) Electron density map for the β-hairpin loop of chain A superimposed on the atomic model in stick representation.

of the Supporting Information). Side-chain conformations for most residues in all four chains of the X-ray structure vary little, except for residues N10, R30, R55, and I56, which exhibit major differences (Figure S6 of the Supporting Information). Unlike the MAS NMR structure, which was determined using distance restraints or/and chemical shifts for most residues, except R2, R3, and E136, density is either missing or very weak for residues R2−N9 and E136 in all chains and for residues Q20−D25, R57−P68 and P124−E136 in chain D.

In addition to the five intratetramer interfaces, each tetramer (arbitrarily designated as "tetramer 0") contacts 10 neighboring tetramers (tetramers 1−10) resulting in four distinct intertetramer interfaces, classified by symmetry operators. The nomenclature for a specific chain (A) in a tetramer (0) is denoted as 0_A. The packing of tetramers in the crystallographic lattice is depicted in Figure 4b. Intertetramer interface 1 is formed by two tetramers, 1 and 2, adjacent to tetramer 0. This interface comprises residues G82−P84 and A87−K89 of chains C (0_C) and D (0_D) of tetramer 0 and equivalent residues G82−P84 and A87−K89 in 2_D and 1_C, respectively (colored light gray in Figure 4b(i)). Intertetramer interface 2 is formed by four tetramers, 3, 4, 5, and 6 (colored green in Figure 4b(ii)) around tetramer 0. Several residues in 0_A (N10−W14, T16, T38−S41, D44, R50−R57, K62, M63, L66, R69, Y71, Y73, P79, D106−V120), 0_B (N10−S13, T16, T38−S40, R50−R57, G58−D60, M63, L66, R69, Y71, Y73, P79, G82−K89, R111−V120, Q122), 0_C (N10, T11, W14, R30, N37−I46, Y48, L75−A87, V95−I119), and 0_D (N10, T11, W14, N39, S41−D44, I46, Y48, L75−A87, A96−N116) are involved in crystallographically inequivalent interfaces (see Table S4). Intertetramer interface 3 is formed by two tetramers, 7 and 8 (colored pink in Figure 4b(iii)), adjacent to tetramer 0. Residues H21−D25, K27, P29−Q32, A96−G99, and Q122−A135 in 0_A, Q20−K23, R57, K62, K64, L66−P68, and Y134 in 0_B, residues R50, R51, T53−R57, G61−M63, and D65−L66 in 0_C, and I36 and V120−L123 of 0_D have contacts with residues comprising tetramers 7 and 8. Intertetramer interface 4 comprises two tetramers, 9 and 10 (colored blue in Figure 4b(iv)), adjacent to tetramer 0. The residues involved in this interface are E80, G82−P84, and A87−K89 of 0_A and 9_A and W14, D106−R112 in 10_B and 0_B, respectively. All intra- and intertetramer contacts for each N^{NTD} residue are summarized in Table S4.

These unique intra- and intertetramer interfaces are reflected in distinct correlations in the MAS NMR spectra. In the (H)NH spectrum recorded at an MAS frequency of 100 kHz, multiple resolved peaks or broad (unresolved) peaks are observed for residues that are found in several different local environments, with ^{15}N peak widths of ∼85−110 Hz (Figure 5b), whereas those for amino acids in single conformations are ∼40−60 Hz (Figure 5a,c). Examples of correlations corresponding to intratetramers A−B, A−C, and B−D interfaces, as well as to the intertetramer interfaces, are shown in Figure 4a,b, bottom panels.

One notable example of unique intratetramer contacts, evidenced by multiple cross-peaks in the MAS NMR spectra, is seen for the H21−Y134 pair of residues (Figure 4a, bottom panel). It is evident from the X-ray structure that the interchain distances are much shorter than the intrachain ones (3.5−4.6 and 7.0−8.0 Å, respectively); hence, only interchain correlations are expected in the spectra (Figure 6). Another interesting example involves residues D60 and M63 in the $\beta2$,

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Table 3. X-ray Data Collection and Refinement Statistics[a]

SARS-CoV-2 N^{NTD}	
Data Collection	
wavelength (Å)	0.9794
space group	$P2_12_12_1$
cell dimensions	
a, *b*, *c* (Å)	58.76, 92.76, 95.59
α, *β*, *γ* (deg)	90, 90, 90
resolution range (Å)	37−1.70
R_{sym} or R_{merge}	0.024(0.65)[b]
$I/\sigma I$-CC1/2	27.5(1.1)-99(42)
completeness (%)	98.9(98.2)
redundancy	11(7)
Refinement	
refinement program	COOT
resolution range (Å)	37−1.70
no. reflections	59,316
R_{work}/R_{free}	24.4/28.5
no. of nonhydrogen atoms	
protein	3642
solvent (water)	650
B-factors	
protein	48.5
solvent (water)	42
Rms deviations	
bond lengths (Å)	0.003
bond angles (deg)	0.68
PDB ID	7UW3

[a]Molecular replacement. [b]Values in parentheses are for the highest-resolution shell.

*β*3-hairpin loop, for which intra- and intertetramer correlations were identified based on the following considerations: (i) D60 from chain A with no intertetramer contacts has a unique intratetramer correlation with K131 at the A−C interface. In contrast, there are no intratetramer correlations involving D60 from other chains with K131, and only intertetramer correlations are present; (ii) M63 has no intratetramer interactions; (iii) M63 from chains A and B is in close proximity to N39 in the symmetry-related tetramer interface 2; and (iv) M63 from chain C forms contacts with A135 across the tetramer interface 3 (Figure 4b, bottom panel, and Figure 7).

In addition to the intra- and intertetramer contacts discussed above, we also observed multiple conformers for many residues, as is evident from their distinct backbone and side-chain chemical shifts. For example, T77 exhibits several resolved resonances with unique chemical shifts each (Figure 4c, left panel). A unique T77C^{β}−A117C′ cross-peak is found for one of the two conformers (designated as conformer b, $T77^{b}$) in the 100 ms mixing time CORD spectrum. This correlation is missing for the second conformer, $T77^{a}$. Both conformers exhibit significant chemical shift differences (ΔN = 1.4 ppm, ΔC^{β} = 0.3 ppm, ΔC^{γ} = 0.3 ppm, $\Delta C'$ = 0.6 ppm), consistent with distinct local environments for T77. Indeed, in the X-ray structure, T77 in 0_A and 0_B has no intra- or intertetramer contacts within 7 Å, consistent with this conformer being $T77^{a}$. In contrast, T77 in chains 0_C and 0_D exhibits intertetramer contacts with A117 from chains 4_B and 3_A, respectively, therefore suggesting that these resonances correspond to the $T77^{b}$ conformer (Figure 4c, right panels). Likewise, at least two distinct conformations are seen for A52, whose backbone chemical shifts are different (ΔN = 2.0 ppm, ΔC^{α} = 0.4 ppm), suggesting that they exist in unique local environments (Figure S7, top panel, Supporting Information). This finding is fully consistent with the X-ray structure (Figure S7, bottom panel, Supporting Information), where A52 in 0_A and 0_B forms intertetramer contacts with D43 and L101 from chains 6_C and 5_D, respectively, while A52 in chains 0_C and 0_D forms contacts with E24 and K27 from chains 8_A and 0_B, respectively.

Tautomeric State of Histidine-107 in the Crystal. In the ^{1}H-detected 2D (H)NH spectrum acquired with a CP contact time of 0.3 ms, H107 gives rise to two distinct $^{15}N^{\varepsilon 2}$−$H^{\varepsilon 2}$ cross-peaks at $\delta(^{15}N)$ = 170.2 ppm/$\delta(^1H)$ = 12.4 ppm and $\delta(^{15}N)$ = 170.4 ppm/$\delta(^1H)$ = 12.6 ppm (Figures 1b and 8a). In 2D (H)CH spectra, multiple $C^{\varepsilon 1}$−$H^{\varepsilon 1}$ and $C^{\delta 2}$−$H^{\delta 2}$ correlations are also observed (Figure 8a). Taken together, these results suggest the presence of local heterogeneity around H107, consistent with two distinct local environments seen in the X-ray crystal structure (Table S4 of the Supporting Information).

The tautomeric state of H107 was ascertained by a ^{1}H-detected 2D (H)NH experiment acquired with a CP contact time of 4 ms (Figure 8a). The cross-peaks at $\delta(^{15}N) \sim 170$ ppm/$\delta(^1H)$ = 7.9 ppm, $\delta(^{15}N) \sim 170$ ppm/$\delta(^1H)$ = 7.7 ppm, and $\delta(^{15}N) \sim 249$ ppm/$\delta(^1H)$ = 7.9 ppm were unambiguously assigned as $N^{\varepsilon 2}$−$H^{\varepsilon 1}$, $N^{\varepsilon 2}$−$H^{\delta 2}$, and $N^{\delta 1}$−$H^{\varepsilon 1}$ correlations, respectively, based on the $H^{\varepsilon 1}$ and $H^{\delta 2}$ chemical shift assignments from 2D (H)CH and 3D (H)CCH experiments (Figure 8a). These results suggest that H107 is the $N^{\varepsilon 2}$−H tautomer.[33−36] Further evidence comes from a solution HMBC spectrum recorded at the pH of the crystallization (pH 6.3) (Figure S8 of the Supporting Information), where $N^{\varepsilon 2}$−$H^{\varepsilon 1}$ and $N^{\varepsilon 2}$−$H^{\delta 2}$ correlations were observed at $\delta(^{15}N)$ = 170 ppm/$\delta(^1H)$ = 7.7 ppm and $\delta(^{15}N)$ = 170 ppm/$\delta(^1H)$ = 6.9 ppm, respectively. The $N^{\delta 1}$−$H^{\varepsilon 1}$ correlation at $\delta(^{15}N)$ = 243.4 ppm/$\delta(^1H)$ = 7.7 ppm is consistent with the H107 being the $N^{\varepsilon 2}$−H tautomer.[35,36]

The deshielded $H^{\varepsilon 2}$ resonance at $\delta(^1H) \sim 12.5$ ppm suggests that the H107 imidazole may be involved in hydrogen bonding or close to a negatively charged group.[34,37] The presence of a $^{15}N^{\varepsilon 2}$−$H^{\varepsilon 2}$ correlation in the (H)NH spectrum acquired with a 0.3 ms CP contact time also suggests a short nitrogen−hydrogen distance, possibly a directly bonded $N^{\varepsilon 2}$−H. Indeed, in the X-ray crystal structure, H107 is in close proximity to D44 and the interatomic H107$N^{\varepsilon 2}$−D44$O^{\delta 1,\delta 2}$ distances are 2.9, 2.9, 3.2, and 3.4 Å in chains A, B, C, and D, respectively (Figure 8b). Taken together, these data indicate that the (H107)$N^{\varepsilon 2}$−H is close to the carboxylate side chain of D44.

DISCUSSION

The NTD domain of the SARS-CoV-2 virus N protein has been previously structurally characterized by X-ray crystallography and solution NMR.[16,19,25] Here, we present a structure that was determined by integrating MAS NMR and X-ray diffraction, providing important novel findings of distinct conformers, made possible by the remarkably high resolution of the MAS NMR spectra. The structural heterogeneity of N^{NTD} is an outcome of crystallization, as seen in other N^{NTD} crystal structures,[19,26,31] and underscores the ability of the protein to form multiple types of contacts involving distinct conformers with unique local environments.

From the technical standpoint, the current study represents a notable advance for determining protein structures by MAS

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

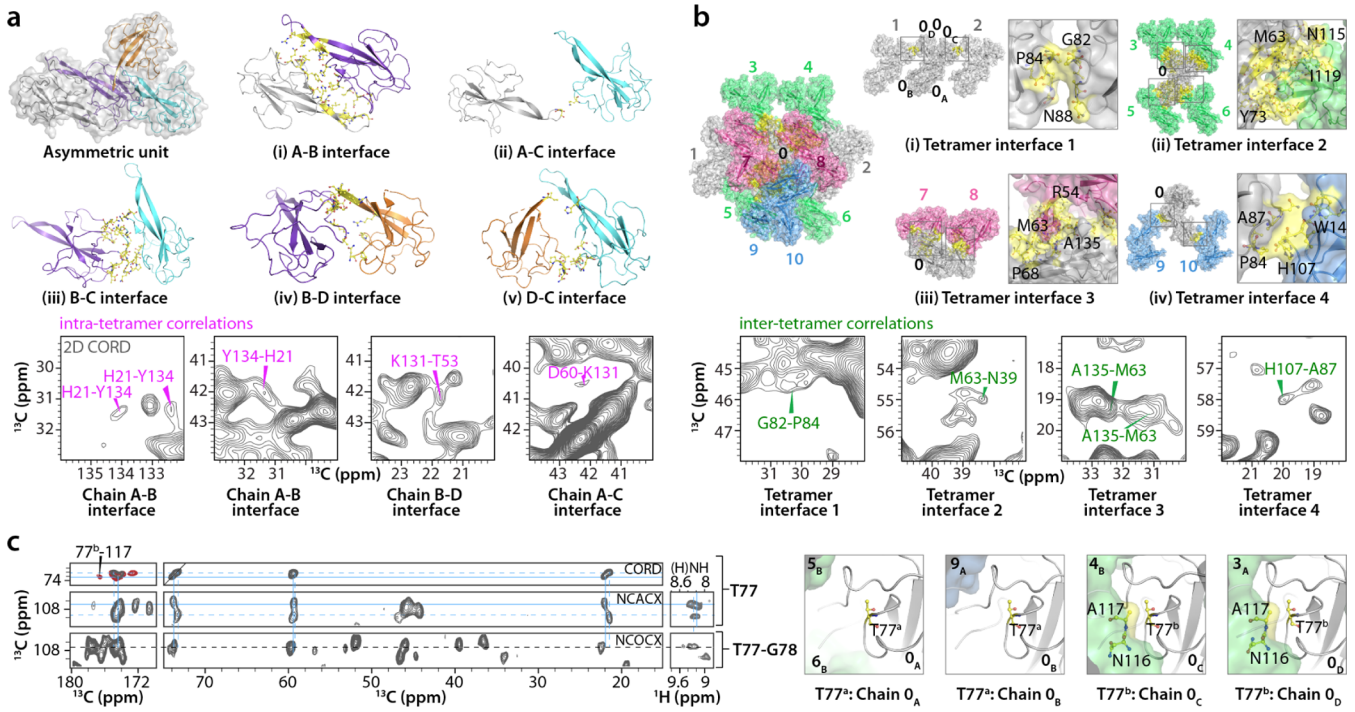


Figure 4. Interchain interfaces and crystal packing in the N^{NTD} structure. (a) Intratetramer interfaces. Top and middle: five unique interchain interfaces in the asymmetric unit of N^{NTD} crystal; chain A (gray), chain B (purple), chain C (cyan), and chain D (orange). Interface residues are in yellow stick representation. (b) Intertetramer interfaces. Top-left: each single tetramer (numbered 0) forms intertetramer interfaces with 10 neighboring tetramers (1−10). Intertetramer interface residues are colored yellow. Top-right and middle-right: four unique intertetramer interfaces are formed based on symmetry operation. The nomenclature for a specific chain (A) in a tetramer (0) is 0_A. Symmetry-related interfaces are boxed and expanded, with individual residues labeled and depicted in stick representation. Selected regions of a 2D CORD spectrum (100 ms mixing time) showing intratetramer correlations (magenta) and intertetramer correlations (green) (a and b, bottom panels). (c) Left: representative strips of the 2D CORD (top strips, 25 ms mixing time, gray; 100 ms mixing time, red), 2D NCACX (middle strip), 2D NCOCX (bottom strip), and 2D (H)NH HETCOR (right strips) spectra illustrating the sequential assignment for T77−G78. Resonances for two conformers, a and b, of T77 are indicated by dotted and solid lines, respectively. Right: interchain contacts of T77 for each chain are colored yellow. The number of scans and the number points in the direct and indirect dimensions are as follows: 2D CORD (25 ms mixing time) - 192 scans, 2048 t_2 points, 840 t_1 points; 2D CORD (100 ms mixing time) - 96 scans, 3072 t_2 points, 840 t_1 points; 2D NCACX - 2048 scans, 2048 t_2 points, 96 t_1 points; 2D NCOCX - 1536 scans, 3072 t_2 points, 96 t_1 points; and 2D (H)NH HETCOR - 80 scans, 3072 t_2 points; 400 t_1 points.

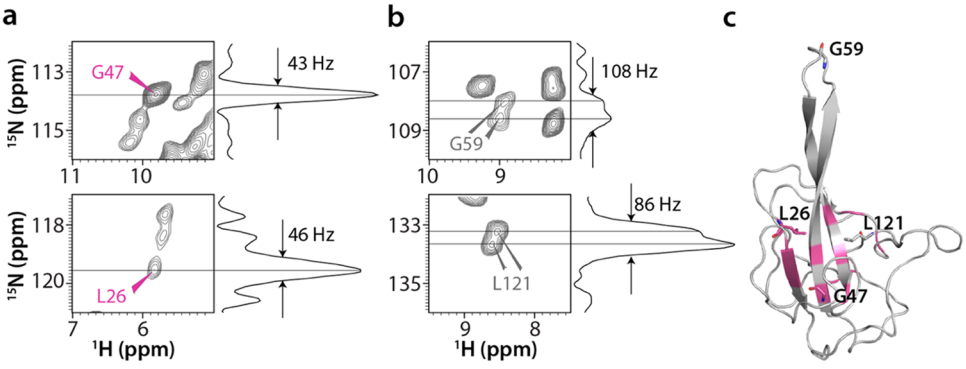


Figure 5. Selected amino acids that exhibit multiple backbone amide resonances in the 2D CORD spectra of crystalline SARS-CoV-2 N^{NTD}. (a, b) Individual single cross-peaks reporting on a unique environment (a, pink labels) and doubled cross-peaks reporting on different environments for the different chains (b, gray labels) with their corresponding 1D ^{15}N slices. (c) The location of amino acids (pink) possessing single amide backbone cross-peaks mapped onto the structure of chain A in the X-ray structure.

NMR since a single sample of only 3.6 mg of U−^{13}C,^{15}N-labeled N^{NTD} packed in a 1.3 mm MAS rotor was sufficient to obtain all necessary spectra. The same sample (∼0.5 mg) was subsequently packed in a 0.7 mm MAS rotor for ^1H-detected experiments at 100 kHz MAS and 20.0 T, and these additional experiments yielded unique information on the side-chain protons. Resonances for 98% of all amino acids were assigned, and a large number of correlations corresponding to 2968 distance restraints, including 968 long-range restraints, were obtained from 14 2D and 3D data sets. As a result, no preparations of isotopically diluted samples were necessary. At 21.8 restraints per residue, the N^{NTD} single-chain structure reported here is one of the highest precision and accuracy MAS NMR structures determined to date.

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

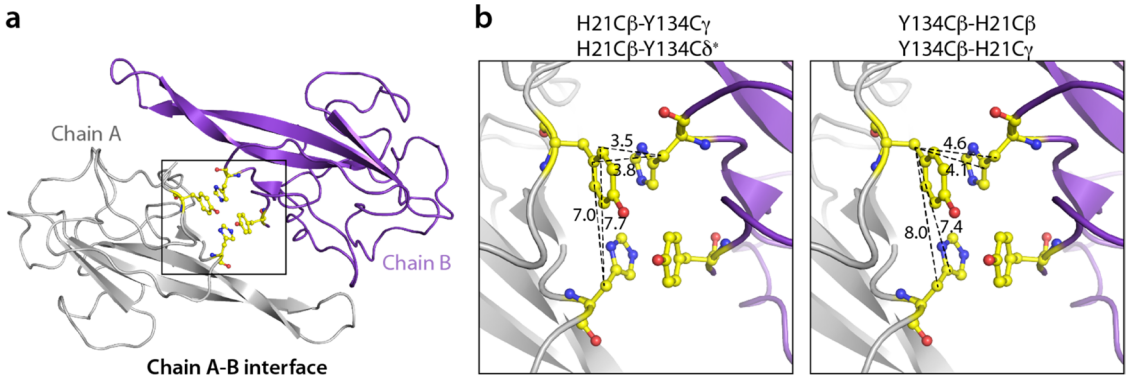


Figure 6. H21−Y134 intra- and interchain contacts at the A−B interface in the X-ray structure. (a) The A−B dimer in the asymmetric unit. (b) Close-up view of the interchain contacts formed between H21 and Y134 from chains A and B. Interchain contacts (dotted lines) are shorter than the intrachain contacts.

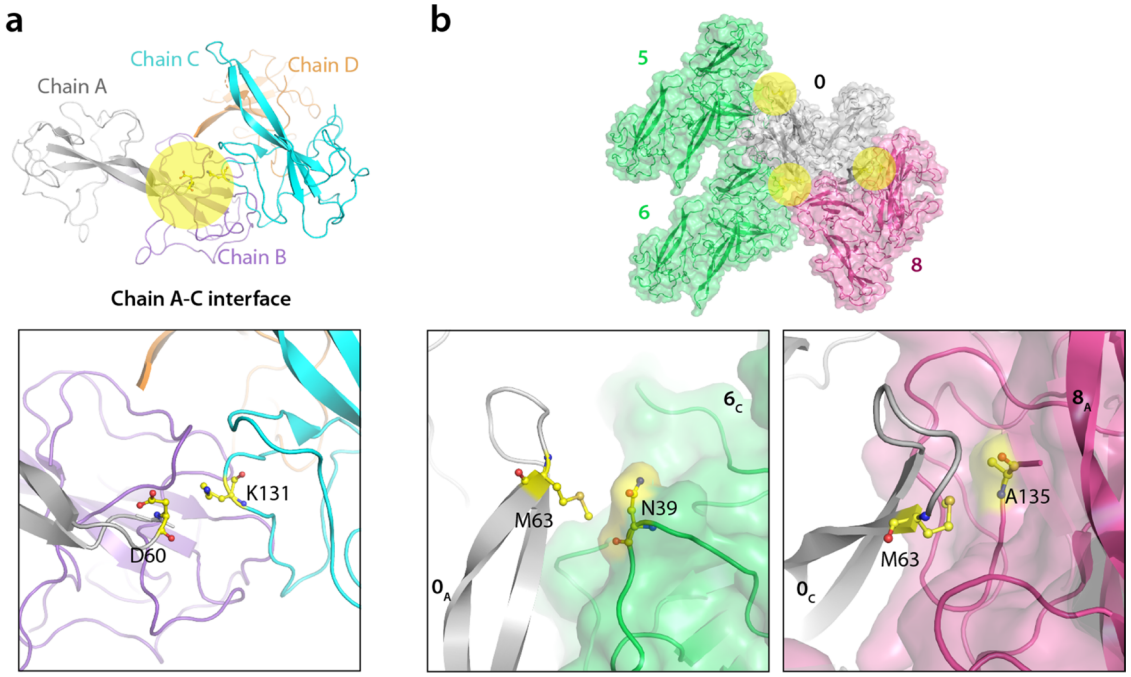


Figure 7. Intra- and intertetramer contacts involving the β-hairpin loop. (a) The A−C interface in the asymmetric unit of X-ray structure is colored yellow (top), and the D60−K131 contact for which correlations are observed in MAS NMR spectra is shown at the bottom. (b) The three intertetramer interfaces around M63 in the crystal lattice are colored yellow (top) with a detailed view provided in the bottom two panels (bottom). The numbering of the tetramers and colors is as in Figure 4b.

Notable is the complementarity of information obtained by MAS NMR and X-ray diffraction. In the X-ray structure, the atomic level details for individual chains and information on the quaternary arrangement in the crystal are obtained, while the strength of MAS NMR lies in providing information about dynamically disordered regions, proton positions, protonation and tautomeric states, and contacts with water molecules. To understand the dynamics of the different regions of N^{NTD} in solution and in the crystalline forms and their role in RNA binding, it will be interesting in future studies to perform measurements of relaxation rates, chemical shifts, and dipolar anisotropy tensors.

CONCLUSIONS

We have determined the structure of SARS-CoV-2 N^{NTD} by integrating MAS NMR and X-ray diffraction. Our combined approach provided atomic details of packing interfaces as well as information about disordered residues at the N- and C-termini and the functionally important RNA binding, β-hairpin loop. In addition, ^{1}H-detected experiments at an MAS frequency of 100 kHz permitted the assignment of side-chain proton chemical shifts not available by other means. The present structure offers guidance for designing therapeutic interventions against the SARS-CoV-2 infection.

MATERIALS AND METHODS

Expression and Purification of N^{NTD}. The recombinant plasmid for expressing SARS-CoV-2 N^{NTD} (residues 40−174, current construct residue numbering 2−136) was prepared from GenScript based on the sequence previously reported for N^{NTD}[19] and *E.coli* codon-optimized. The template coding for SARS-CoV-2 N^{NTD} was subcloned into a pET28a(+) vector fused with an N-terminal hexahistidine tag (His_6), followed by a tobacco etch virus (TEV) protease cleavage site, His_6-TEV-N^{NTD}. For the expression of U−^{13}C,^{15}N−N^{NTD} and U−^{15}N−N^{NTD}, transformed *E.coli* BL21

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

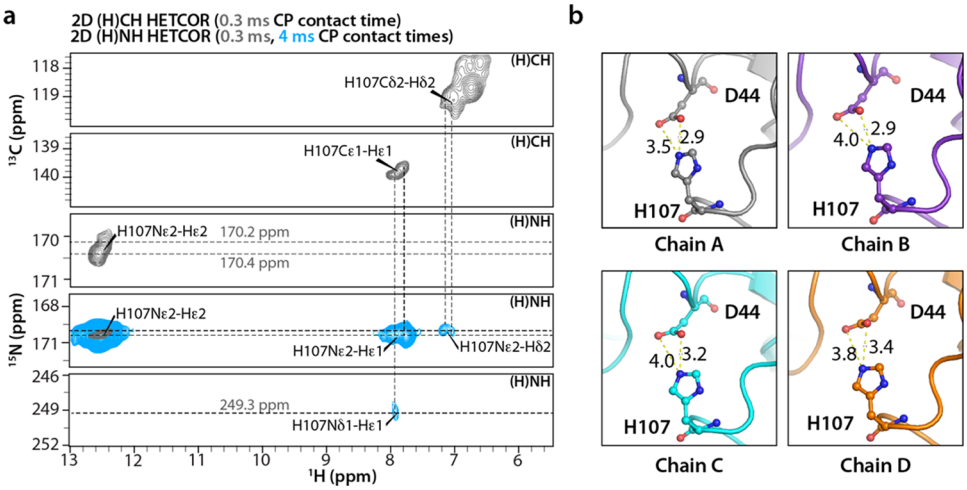


Figure 8. Side-chain imidazole state of H107 in the N^{NTD} crystal. (a) Strips extracted from 2D (H)CH and (H)NH HETCOR spectra acquired with CP contact times of at 0.3 ms (gray) and 4 ms (blue) (2D (H)NH HETCOR spectra only). H107 side-chain ^{13}C, ^{15}N, and ^{1}H resonances are labeled. (b) Conformations of H107 and the neighboring D44 residues in chains A–D of the crystal. The interatomic H107$N^{\varepsilon 2}$–D44$O^{\delta 1,\delta 2}$ distances vary between 2.9 and 3.4 Å. The number of scans and the number points in the direct and indirect dimensions are as follows: 2D (H)CH HETCOR - 64 scans, 1024 t_2 points, 2310 t_1 points; 2D (H)NH HETCOR - 16 scans, 1024 t_2 points, 512 t_1 points.

(DE3) cells were cultured in 5 mL of Luria-Bertani (LB) medium containing 100 μg/mL of kanamycin. LB preculture was incubated at 37 °C with agitation until the OD_{600} reached 1.0–1.2. Fifty milliliters of M9 medium, supplemented with 1 g/L $^{15}NH_4Cl$ (U–^{15}N–N^{NTD}) or 1 g/L $^{15}NH_4Cl$ and 2 g/L U–$^{13}C_6$–glucose (U–^{13}C,^{15}N–N^{NTD}), was inoculated with 1 mL of the LB preculture and incubated overnight at 37 °C. Following the overnight growth, 50 mL of M9 medium was transferred to 1 L of the isotopically labeled M9 medium and incubated at 37 °C. Cells were grown to an OD_{600} of 1.0 and induced with a final isopropylthio-β-galactoside (IPTG) concentration of 400 μM. The protein was expressed at 20 °C for 16–18 h, and cells were harvested by centrifugation at 4000*g* for 10 min at 4 °C. The cell pellet was resuspended in lysis buffer (20 mM tris–HCl, 500 mM NaCl, 20 mM imidazole, 0.02% NaN_3, pH 8.0) and flash-frozen (−80 °C) for short-term storage.

Cells were opened after treatment with 1 mM phenylmethylsulfonyl fluoride (PMSF) by sonication at 40% power for 5 min (15 s pulse on and 45 s pulse off) on ice. The cellular lysate was clarified by centrifugation at 14,000*g* for 1 h at 4 °C. His_6-tagged N^{NTD} was purified by affinity chromatography over a 5 mL HisTrap HP column (GE Healthcare). For elution, a gradient of 20–500 mM imidazole in 20 mM tris–HCl (pH 8.0), 500 mM NaCl, and 0.02% NaN_3 was employed. The His_6-tag was cleaved with TEV protease (1:25 ratio of TEV protease to His_6–U–^{13}C,^{15}N–N^{NTD}) for 12–16 h at 4 °C and again fractionated over a 5 mL HisTrap HP column (GE Healthcare). Fractions were eluted in 20 mM tris–HCl, 500 mM NaCl, 20 mM imidazole, 0.02% NaN_3, pH 8.0, and the pure protein was buffer-exchanged into crystallization buffer (20 mM tris–HCl, 50 mM NaCl, pH 6.0) and NMR buffer (20 mM tris–HCl, 150 mM NaCl, 90:10 H_2O/D_2O, pH 8.0). The buffer-exchanged N^{NTD} was concentrated to 12 mg/mL for solution NMR and 30 mg/mL for crystallization.

Crystallization of N^{NTD}. Small-scale crystallization was carried out at room temperature (~20 °C) using the sitting-drop vapor diffusion. Two microliters of U–^{13}C,^{15}N–N^{NTD} (20 mM tris–HCl, 50 mM NaCl, pH 6.0) was mixed with 2 μL of crystallization buffer (100 mM MES, 30% PEG4000, pH 6.5), modified from a previously published crystallization condition (PDB: 6WKP).[26]

U–^{13}C,^{15}N–N^{NTD} for MAS NMR experiments was crystallized using a large-scale sitting-drop method based on the volumetric proportions of 500 μL sitting-drop crystallization wells (Figure S9 of the Supporting Information). A series of presterilized Petri dishes were used to form a concentric sitting-drop vessel with a reservoir (volume capacity of 25–75 mL) and three droplet wells (optimal volume of 300–1000 μL). Similar to the small-scale crystallization, the crystallization droplet mixture comprised 250 μL of N^{NTD} and 250 μL of crystallization buffer. The large-scale sitting-drop vessel was sealed using vacuum grease and left undisturbed at 20 °C for 5 days. Once crystallization was complete, the protein crystals were harvested and packed into a Bruker 1.3 mm rotor by ultracentrifugation at 10,000*g* for 15 min at 4 °C. The fully packed 1.3 mm rotor contained 3.6 mg of hydrated protein crystals. The same sample (~0.5 mg) was subsequently packed in a 0.7 mm rotor for experiments at 100 kHz MAS and 20.0 T.

Diffraction Data Collection and Structure Determination. X-ray diffraction data of protein crystals were collected at beamline 12-2 at the Stanford Synchrotron Radiation Lightsource (SRRL). All diffraction data used for analysis were collected from crystals grown in 100 mM MES, 30% PEG4000, pH 6.0 at 100 K. All diffraction data were indexed and integrated using the program XDS[38] and scaled using the program AIMLESS from the CCP4 suite.[39] The structure was solved by molecular replacement (MOLREP, CCP4 suite) using one monomer of PDB:ID 6M3M. Structure refinement was carried out in Phenix[40] with manual building in COOT[41] (PDB:ID 7UW3, this work).

Solution NMR Spectroscopy. A 2D ^{1}H–^{15}N HSQC spectrum of 850 mM U–^{15}N–N^{NTD} in 20 mM tris–HCl, 150 mM NaCl, 90:10 H_2O/D_2O buffer (pH 8.0) was recorded at 25 °C on a 14.1 T Bruker Neo spectrometer equipped with a triple-resonance inverse detection (TXI) probe. The Larmor frequencies were 600.13 MHz (^{1}H), 150.9 MHz (^{13}C), and 60.8 MHz (^{15}N). Backbone and side-chain ^{1}H and ^{15}N chemical shift assignments (Figure S10 and Table S5 of the Supporting Information) were obtained by comparison with SARS-CoV-2 N^{NTD} (BMRB:34511) and SARS-CoV-1 N^{NTD} (BMRB:6372) chemical shifts in the BMRB.[25,29] ^{1}H–^{15}N HMBC spectra were recorded at pH 6.3 to match the crystallization pH, with delays set to 5.4, 25, and 50 ms, corresponding to 1/2 of ^{1}J and $^{2,3}J$ coupling constants of 92, and 10 and 20 Hz, respectively.

MAS NMR Spectroscopy. MAS NMR spectra of U–^{13}C,^{15}N–N^{NTD} protein crystals were recorded on a 14.1 T Bruker AVIII spectrometer outfitted with a 1.3 mm HCN probe. The Larmor frequencies were 599.8 MHz (^{1}H), 150.8 MHz (^{13}C), and 60.7 MHz (^{15}N). The MAS frequencies were controlled to within ±10 Hz by a Bruker MAS controller. The actual sample temperature was maintained at ~25 °C throughout the experiments.

Typical 90° pulse lengths were 1.3–1.5 μs for ^{1}H, 2.6–2.9 μs for ^{13}C, and 3.2–3.5 μs for ^{15}N. ^{1}H–^{13}C and ^{1}H–^{15}N cross-polarizations were performed with an 80–100% linear amplitude ramp on ^{1}H, with contact times of 1–1.5 and 2–2.5 ms, respectively. The center of the

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX–XXX

ramp was matched to the Hartmann−Hahn condition at the first spinning sideband. 2D $^{13}C-^{13}C$ CORD,[28] 2D NCACX, and 2D NCOCX spectra were recorded at an MAS frequency of 14 kHz. CORD mixing times were 25, 100, 250, and 500 ms, and the ^{1}H radio frequency (rf) field strength during CORD mixing was 14 kHz. Band-selective $^{15}N-^{13}C$ spectrally induced filtering in combination with cross-polarization (SPECIFIC-CP)[42] with a contact time of 6.0−7.5 ms. SPINAL-64[43] decoupling (90−100 kHz) was used during the evolution and acquisition periods.

2D $^{13}C-^{13}C$ RFDR, ^{1}H-detected 2D (H)NH, and (H)CH HETCOR, as well as 3D (H)CANH and (H)CONH spectra, were recorded at an MAS frequency of 60 kHz with a 2.4 ms RFDR mixing time. Swept-low power TPPM (15 kHz, slTPPM)[44] was used for ^{1}H-heteronuclear decoupling during acquisition. WALTZ-16[45] broad-band decoupling was used for ^{13}C and ^{15}N decoupling during ^{1}H acquisition. For 3D ^{1}H-detected (H)CANH and (H)CONH spectra, CA-N and CO-N CP contact times were 6−7.5 ms with a contact-amplitude spin lock of about 25 kHz on ^{13}C and a tangent-modulated amplitude spin lock of the mean rf field amplitude of about 35 kHz on ^{15}N.[46]

Additional MAS NMR spectra were recorded on a 20.0 T Bruker AVIII spectrometer outfitted with a 0.7 mm HCND and a 1.3 mm HCN probe. The Larmor frequencies were 850.4 MHz (^{1}H), 213.9 MHz (^{13}C), and 86.2 MHz (^{15}N). The MAS frequency was 100 kHz, controlled to within ±50 Hz by a Bruker MAS controller. The sample temperature was maintained at ∼25 °C throughout the experiments. Pulse lengths (90°) were 1.3 μs for ^{1}H, 3.15 μs for ^{13}C, and 3 μs for ^{15}N. The (H)NH spectrum was recorded using a back CP (HN) of an 800 μs contact time with an 80−100% linear amplitude ramp on ^{1}H; the rf field strengths were 145 kHz for ^{1}H and 48 kHz for ^{15}N. The forward CP (NH) used a 200 μs contact time, with an 80−100% linear amplitude ramp on ^{1}H; the rf field strengths were 134 kHz for ^{1}H and 48 kHz for ^{15}N. For (H)CH and (H)CCH experiments, the ^{13}C CP rf field strength was set to 30 kHz; for forward and back CP, linear amplitude ramps on ^{1}H were 80−100 and 100−80%; the ^{1}H rf field strengths were set at 138 and 132 kHz; the contact times were 600 and 175 μs, respectively. The CC RFDR mixing time was 0.56 ms. For all spectra, the ^{1}H rf field strengths for water suppression and proton decoupling were set at 1/4 ω_r, and a WALTZ sequence at 10 kHz was used for heteronuclear decoupling of both ^{13}C and ^{15}N. An additional 2D (H)NH spectrum was recorded at 60 kHz MAS, using a 1.3 mm HCN probe. The CP contact time was 4 ms, and the remainder of all conditions were identical to those at 14.1 T (see above).

Data Processing. All MAS NMR data were processed using Bruker TopSpin and NMRPipe.[47] ^{1}H resonances are referenced with respect to water at 4.7 ppm and ^{13}C and ^{15}N to the external standards adamantane and ammonium chloride, respectively. All 2D and 3D data sets were processed by applying 30, 45, 60, and 90° shifted sine bell apodization, followed by a Lorentzian-to-Gaussian transformation in all dimensions. Forward linear prediction to twice the number of original data points was applied in the indirect dimension for some data sets, followed by zero filling. 2D and 3D ^{1}H-detected data sets were processed with Gaussian and/or square sine window apodization and quadrature baseline correction.

Resonance Assignments. Spectra were analyzed using CCPN[48] and Sparky,[49,50] and MAS NMR backbone and side-chain $^{1}H-^{15}N$ resonance assignments were initially carried out by comparison with solution NMR chemical shifts[25,29] and verified by *de novo* backbone assignment based on 2D $^{13}C-^{13}C$ CORD (25 ms mixing time) and RFDR spectra, combined with 2D NCACX (25 ms mixing time), 2D NCOCX (25 ms mixing time), ^{1}H-detected 2D (H)NH HETCOR, 3D (H)CANH, and 3D (H)CONH spectra. Side-chain carbon and nitrogen resonances were assigned using 2D CORD, 2D NCACX, 2D NCOCX, and 2D (H)NH spectra, and side-chain and backbone protons were assigned using ^{1}H-detected 2D (H)NH, (H)CH HETCOR, and 3D (H)CCH experiments.

Structure Calculation of SARS-Cov2 N^{NTD}. The MAS NMR structure of a single N^{NTD} chain was calculated in Xplor-NIH version 2.53[51−53] using $^{13}C-^{13}C$, $^{15}N-^{13}C$, and $^{1}H-^{15}N$ distance restraints, extracted from 2D CORD (100, 250, and 500 ms mixing times), NCACX, NCOCX, and (H)NH HETCOR spectra and backbone dihedral angles predicted by TALOS−N[54] from the experimental ^{1}H, ^{13}C, and ^{15}N chemical shifts. The bounds for the distance restraints were set to 1.5−6.5 Å (4.0 ± 2.5 Å) and 2.0−7.2 Å (4.6 ± 2.6 Å) for intra- and inter-residue restraints, respectively, consistent with our previous studies.[30,55]

Calculations were seeded using the primary sequence as extended strands. One thousand structures were generated with molecular dynamics simulated annealing in torsion angle space with two successive annealing schedules and a final gradient minimization in Cartesian space, essentially as described previously[30,55] and detailed below.

Two successive annealing schedules were used, the first in vacuum with the REPEL module and the second with an implicit solvent refinement using the EEFx module.[56] The 10 lowest-energy structures were selected and served as input for the second schedule, and the 10 lowest-energy structures were selected and served as input for the final ensemble (PDB: 7SD4). Standard terms for bond lengths, bond angles, and impropers were applied to enforce the correct covalent geometry.

The first annealing calculation was essentially identical to that reported previously,[30,55] with initial random velocities at 3500 K constant-temperature molecular dynamics run for the shorter of 800 ps or 8000 steps, with the time step size allowed to float to maintain constant energy. Subsequently, simulated annealing calculations at reduced temperatures in steps of 25−100 K were carried out for the shorter of 0.4 ps or 200 steps. Force constants for distance restraints were ramped from 10 to 50 kcal/(mol·Å)^2. Dihedral angle restraints were disabled for high-temperature dynamics at 3500 K and subsequently applied with a force constant of 200 kcal/(mol·rad^2). The force constant for the radius of gyration was geometrically scaled from 0.002 to 1, and a hydrogen bond term, HBPot, was used to improve hydrogen bond geometries.[57] After simulated annealing, structures were minimized using a Powell energy minimization scheme.

For the second schedule performed in implicit solvent, all parameters were set as in the example EEFx of Xplor-NIH. Annealing was performed at 3500 K for 15 ps or 15,000 steps, whichever was completed first. The starting time step was 1 fs and was self-adjusted in subsequent steps to ensure conservation of energy. Random initial velocities were assigned about a Maxwell distribution at the starting temperature of 3500 K. Subsequently, the temperatures were reduced to 25 K in steps of 12.5 K. At each temperature, 0.4 ps dynamics were run with an initial time step of 1 fs. Force constants for distance restraints were ramped from 2 to 30 kcal/(mol·Å^2). The dihedral restraint force constants were set to 10 kcal/(mol·rad^2) for high-temperature dynamics at 3,000 K and 200 kcal/(mol·rad^2) during cooling. After the EEFx module, structures were minimized using a Powell energy minimization scheme.

Structure Analysis and Visualization. Atomic rmsd values were calculated using routines in Xplor-NIH (version 2.53).[51−53] The visualization of structural ensembles was rendered in PyMOL,[58] using in-house shell/bash scripts. Secondary structure elements were classified according to STRIDE[59] and manual inspection.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.2c03320.

Summary of resonance assignment of SARS-CoV-2 N^{NTD}; X-ray crystal structure; 10 lowest-energy conformers of MAS NMR structures; all and selected side-chain conformations of MAS NMR and X-ray structures; multiple resonances for A52 in the MAS NMR spectra; histidine protonation state in solution NMR; mass spectrum, crystallization, and purification of N^{NTD}; 2D

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

$^{1}H-^{15}N$ HSQC NMR spectrum of U−^{15}N−N^{NTD}; summary of all NMR experiments; MAS NMR and solution NMR chemical shifts of N^{NTD}; rmsd values between X-ray and MAS NMR structures; intra- and intertetramer contacts in N^{NTD} crystals; and author contributions (PDF)

AUTHOR INFORMATION

Corresponding Authors

Angela M. Gronenborn − *Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States; Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0001-9072-3525; Email: amg100@pitt.edu

Tatyana Polenova − *Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States; Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0002-0346-1131; Email: tpolenov@udel.edu

Authors

Sucharita Sarkar − *Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0002-0996-1647

Brent Runge − *Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States*

Ryan W. Russell − *Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States*

Kumar Tekwani Movellan − *Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States*

Daniel Calero − *Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States*

Somayeh Zeinalilathori − *Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States*

Caitlin M. Quinn − *Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States;* orcid.org/0000-0001-6172-7324

Manman Lu − *Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States; Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0002-4156-4975

Guillermo Calero − *Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States; Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0003-3730-4676

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.2c03320

Author Contributions

[‖]S.S. and B.R. contributed equally to this work.

Notes

The authors declare no competing financial interest.

The MAS NMR atomic structure coordinates and X-ray structure coordinates of SARS-CoV-2 N^{NTD} have been deposited in the Protein Data Bank under accession codes 7SD4 and 7UW3, respectively. MAS NMR chemical shifts of SARS-CoV-2 N^{NTD} have been deposited in the Biological Magnetic Resonance Data Bank under accession code 30955.

ACKNOWLEDGMENTS

The authors thank Dr. Shi Bai for assistance with acquiring solution HMBC spectra and Roman Zadorozhnyi for helpful discussions. This work was supported by the National Institutes of Health (NIH Grant P50AI1504817, Technology Development Project 2). The authors acknowledge the National Institutes of Health (NIGMS Grant P30GM110758) for the support of core instrumentation infrastructure at the University of Delaware. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health (NIH Grant P30GM133894).

REFERENCES

(1) Gorbalenya, A. E.; Baker, S. C.; Baric, R. S.; de Groot, R. J.; Drosten, C.; Gulyaeva, A. A.; Haagmans, B. L.; Lauber, C.; Leontovich, A. M.; Neuman, B. W.; Penzar, D.; Perlman, S.; Poon, L. L. M.; Samborskiy, D. V.; Sidorov, I. A.; Sola, I.; Ziebuhr, J. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **2020**, *5*, 536−544.

(2) Ibn-Mohammed, T.; Mustapha, K. B.; Godsell, J.; Adamu, Z.; Babatunde, K. A.; Akintade, D. D.; Acquaye, A.; Fujii, H.; Ndiaye, M. M.; Yamoah, F. A.; Koh, S. C. L. A critical analysis of the impacts of COVID-19 on the global economy and ecosystems and opportunities for circular economy strategies. *Resour., Conserv. Recycl.* **2021**, *164*, No. 105169.

(3) Nicola, M.; Alsafi, Z.; Sohrabi, C.; Kerwan, A.; Al-Jabir, A.; Iosifidis, C.; Agha, M.; Agha, R. The socio-economic implications of the coronavirus pandemic (COVID-19): a review. *Int. J. Surg.* **2020**, *78*, 185−193.

(4) Brian, D. A.; Baric, R. S. Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* **2005**, *287*, 1−30.

(5) Cui, J.; Li, F.; Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17*, 181−192.

(6) Beyer, D. K.; Forero, A. Mechanisms of antiviral immune evasion of SARS-CoV-2. *J. Mol. Biol.* **2021**, No. 167265.

(7) Masters, P. S.; Sturman, L. S. Background paper functions of the coronavirus nucleocapsid protein. *Adv. Exp. Med. Biol.* **1990**, *276*, 235−238.

(8) McBride, R.; van Zyl, M.; Fielding, B. C. The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **2014**, *6*, 2991−3018.
(9) Savastano, A.; Ibáñez de Opakua, A.; Rankovic, M.; Zweckstetter, M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat. Commun.* **2020**, *11*, No. 6041.
(10) Chang, C. K.; Sue, S. C.; Yu, T. H.; Hsieh, C. M.; Tsai, C. K.; Chiang, Y. C.; Lee, S. J.; Hsiao, H. H.; Wu, W. J.; Chang, W. L.; Lin, C. H.; Huang, T. H. Modular organization of SARS coronavirus nucleocapsid protein. *J. Biomed. Sci.* **2006**, *13*, 59−72.
(11) Lo, Y. S.; Lin, S. Y.; Wang, S. M.; Wang, C. T.; Chiu, Y. L.; Huang, T. H.; Hou, M. H. Oligomerization of the carboxyl terminal domain of the human coronavirus 229E nucleocapsid protein. *FEBS Lett.* **2013**, *587*, 120−127.
(12) Chen, I. J.; Yuann, J. M. P.; Chang, Y. M.; Lin, S. Y.; Zhao, J. C.; Perlman, S.; Shen, Y. Y.; Huang, T. H.; Hou, M. H. Crystal structure-based exploration of the important role of Arg106 in the RNA-binding domain of human coronavirus OC43 nucleocapsid protein. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 1054−1062.
(13) Chang, C. K.; Chen, C. M. M.; Chiang, M. H.; Hsu, Y. L.; Huang, T. H. Transient oligomerization of the SARS-CoV N protein—implication for virus ribonucleoprotein packaging. *PLoS One* **2013**, *8*, No. e65045.
(14) Cubuk, J.; Alston, J. J.; Incicco, J. J.; Singh, S.; Stuchell-Brereton, M. D.; Ward, M. D.; Zimmerman, M. I.; Vithani, N.; Griffith, D.; Wagoner, J. A.; Bowman, G. R.; Hall, K. B.; Soranno, A.; Holehouse, A. S. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **2021**, *12*, No. 1936.
(15) Chang, C. K.; Hou, M. H.; Chang, C. F.; Hsiao, C. D.; Huang, T. H. The SARS coronavirus nucleocapsid protein—Forms and functions. *Antivir. Res.* **2014**, *103*, 39−50.
(16) Peng, Y.; Du, N.; Lei, Y. Q.; Dorje, S.; Qi, J. X.; Luo, T. R.; Gao, G. F.; Song, H. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *EMBO J.* **2020**, *39*, No. e105938.
(17) Yang, M.; He, S. H.; Chen, X. X.; Huang, Z. X.; Zhou, Z. L.; Zhou, Z. C.; Chen, Q. Y.; Chen, S. D.; Kang, S. S. Structural insight Into the SARS-CoV-2 nucleocapsid protein C-terminal domain reveals a novel recognition mechanism for viral transcriptional regulatory sequences. *Front. Chem.* **2021**, *8*, No. 624765.
(18) Schiavina, M.; Pontoriero, L.; Uversky, V. N.; Felli, I. C.; Pierattelli, R. 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. *Biomol. NMR Assign.* **2021**, *15*, 219−227.
(19) Kang, S.; Yang, M.; Hong, Z. S.; Zhang, L. P.; Huang, Z. X.; Chen, X. X.; He, S. H.; Zhou, Z. L.; Zhou, Z. C.; Chen, Q. Y.; Yan, Y.; Zhang, C. S.; Shan, H.; Chen, S. D. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* **2020**, *10*, 1228−1238.
(20) Grossoehme, N. E.; Li, L. C.; Keane, S. C.; Liu, P. H.; Dann, C. E.; Leibowitz, J. L.; Giedroc, D. P. Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. *J. Mol. Biol.* **2009**, *394*, 544−557.
(21) Jayaram, H.; Fan, H.; Bowman, B. R.; Ooi, A.; Jaaram, J.; Collisson, E. W.; Lescar, M.; Prasady, B. V. V. X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: Implications for nucleocapsid formation. *J. Virol.* **2006**, *80*, 6612−6620.
(22) Chen, C. Y.; Chang, C. K.; Chang, Y. W.; Sue, S. C.; Bai, H. I.; Riang, L.; Hsiao, C. D.; Huang, T. H. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* **2007**, *368*, 1075−1086.
(23) Huang, Q. L.; Yu, L. P.; Petros, A. M.; Gunasekera, A.; Liu, Z. H.; Xu, N.; Hajduk, P.; Mack, J.; Fesik, S. W.; Olejniczak, E. T. Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* **2004**, *43*, 6059−6063.
(24) Saikatendu, K. S.; Joseph, J. S.; Subramanian, V.; Neuman, B. W.; Buchmeier, M. J.; Stevens, R. C.; Kuhn, P. Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *J. Virol.* **2007**, *81*, 3913−3921.
(25) Dinesh, D. C.; Chalupska, D.; Silhan, J.; Koutna, E.; Nencka, R.; Veverka, V.; Boura, E. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.* **2020**, *16*, No. e1009100.
(26) Chang, C.; Michalska, K.; Jedrzejczak, R.; Maltseva, N.; Endres, M.; Godzik, A.; Kim, Y.; Joachimiak, A. *Crystal Structure of RNA-Binding Domain of Nucleocapsid Phosphoprotein from SARS CoV-2, Monoclinic Crystal Form*; Worldwide Protein Data Bank, 2020.
(27) Elledge, S. K.; Zhou, X. X.; Byrnes, J. R.; Martinko, A. J.; Lui, I.; Pance, K.; Lim, S. A.; Glasgow, J. E.; Glasgow, A. A.; Turcios, K.; Iyer, N. S.; Torres, L.; Peluso, M. J.; Henrich, T. J.; Wang, T. T.; Tato, C. M.; Leung, K. K.; Greenhouse, B.; Wells, J. A. Engineering luminescent biosensors for point-of-care SARS-CoV-2 antibody detection. *Nat. Biotechnol.* **2021**, *39*, 928−935.
(28) Hou, G.; Yan, S.; Trebosc, J.; Amoureux, J.-P.; Polenova, T. Broadband homonuclear correlation spectroscopy driven by combined R2(n)(v) sequences under fast magic angle spinning for NMR structural analysis of organic and biological solids. *J. Magn. Reson.* **2013**, *232*, 18−30.
(29) Zhong, N.; Huang, Q.; Jin, C.; Xia, B. ^{1}H, ^{13}C, and ^{15}N resonance assignments of the N-terminal domain of the SARS CoV nucleocapsid protein. *J. Biomol. NMR* **2005**, *31*, 79−80.
(30) Russell, R. W.; Fritz, M. P.; Kraus, J.; Quinn, C. M.; Polenova, T.; Gronenborn, A. M. Accuracy and precision of protein structures determined by magic angle spinning NMR spectroscopy: for some 'with a little help from a friend'. *J. Biomol. NMR* **2019**, *73*, 333−346.
(31) Kang, S.; Yang, M.; He, S. H.; Wang, Y. M.; Chen, X. X.; Chen, Y. Q.; Hong, Z. S.; Liu, J.; Jiang, G. M.; Chen, Q. Y.; Zhou, Z. L.; Zhou, Z. C.; Huang, Z. X.; Huang, X.; He, H. H.; Zheng, W. H.; Liao, H. X.; Xiao, F.; Shan, H.; Chen, S. D. A SARS-CoV-2 antibody curbs viral nucleocapsid protein-induced complement hyperactivation. *Nat. Commun.* **2021**, *12*, No. 2697.
(32) Ye, Q.; Lu, S.; Corbett, K. D. Structural basis for SARS-CoV-2 nucleocapsid protein recognition by single-domain antibodies. *Front. Immunol.* **2021**, *12*, No. 719037.
(33) Wei, Y.; de Dios, A. C.; McDermott, A. E. Solid-state 15N NMR chemical shift anisotropy of histidines: experimental and theoretical studies of hydrogen bonding. *J. Am. Chem. Soc.* **1999**, *121*, 10389−10394.
(34) Li, S. H.; Hong, M. Protonation, tautomerization, and rotameric structure of histidine: a comprehensive study by magic-angle-spinning solid-state NMR. *J. Am. Chem. Soc.* **2011**, *133*, 1534−1544.
(35) Pelton, J. G.; Torchia, D. A.; Meadow, N. D.; Roseman, S. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. *Protein Sci.* **1993**, *2*, 543−558.
(36) Jung, J.; Byeon, I.-J. L.; Wang, Y.; King, J.; Gronenborn, A. M. The structure of the cataract-causing P23T mutant of human γD-crystallin exhibits distinctive local conformational and dynamic changes. *Biochemistry* **2009**, *48*, 2597−2609.
(37) Hong, M.; Fritzsching, K. J.; Williams, J. K. Hydrogen-bonding partner of the proton-conducting histidine in the influenza M2 proton channel revealed from ^{1}H chemical shifts. *J. Am. Chem. Soc.* **2012**, *134*, 14753−14755.
(38) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−132.
(39) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S.

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 235−242.
(40) Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 1948−1954.
(41) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.
(42) Baldus, M.; Petkova, A. T.; Herzfeld, J.; Griffin, R. G. Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. *Mol. Phys.* **1998**, *95*, 1197−1207.
(43) Bräuniger, T.; Wormald, P.; Hodgkinson, P. Improved proton decoupling in NMR spectroscopy of crystalline solids using the SPINAL-64 sequence. *Monatsh. Chem.* **2002**, *133*, 1549−1554.
(44) Lewandowski, J. R.; Sein, J.; Blackledge, M.; Emsley, L. Anisotropic collective motion contributes to nuclear spin relaxation in crystalline proteins. *J. Am. Chem. Soc.* **2010**, *132*, 1246−1248.
(45) Shaka, A. J.; Keeler, J.; Frenkiel, T.; Freeman, R. An improved sequence for broad-band decoupling—WALTZ-16. *J. Magn. Reson.* **1983**, *52*, 335−338.
(46) Barbet-Massin, E.; Pell, A. J.; Retel, J. S.; Andreas, L. B.; Jaudzems, K.; Franks, W. T.; Nieuwkoop, A. J.; Hiller, M.; Higman, V.; Guerry, P.; Bertarello, A.; Knight, M. J.; Felletti, M.; Le Marchand, T.; Kotelovica, S.; Akopjana, I.; Tars, K.; Stoppini, M.; Bellotti, V.; Bolognesi, M.; Ricagno, S.; Chou, J. J.; Griffin, R. G.; Oschkinat, H.; Lesage, A.; Emsley, L.; Herrmann, T.; Pintacuda, G. Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J. Am. Chem. Soc.* **2014**, *136*, 12489−12497.
(47) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277−293.
(48) Stevens, T. J.; Fogh, R. H.; Boucher, W.; Higman, V. A.; Eisenmenger, F.; Bardiaux, B.; van Rossum, B. J.; Oschkinat, H.; Laue, E. D. A software framework for analysing solid-state MAS NMR data. *J. Biomol. NMR* **2011**, *51*, 437−447.
(49) Goddard, T. D.; Kneller, D. G. *SPARKY 3*; University of California: San Francisco, 2004.
(50) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015**, *31*, 1325−1327.
(51) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **2003**, *160*, 65−73.
(52) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47−62.
(53) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. Xplor-NIH for molecular structure determination from NMR and other data sources. *Protein Sci.* **2018**, *27*, 26−40.
(54) Shen, Y.; Bax, A. Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. *Methods Mol. Biol.* **2015**, *1260*, 17−32.
(55) Lu, M.; Russell, R. W.; Bryer, A. J.; Quinn, C. M.; Hou, G. J.; Zhang, H. L.; Schwieters, C. D.; Perilla, J. R.; Gronenborn, A. M.; Polenova, T. Atomic-resolution structure of HIV-1 capsid tubes by magic-angle spinning NMR. *Nat. Struct. Mol. Biol.* **2020**, *27*, 863−869.
(56) Tian, Y.; Schwieters, C. D.; Opella, S. J.; Marassi, F. M. A practical implicit solvent potential for NMR structure calculation. *J. Magn. Reson.* **2014**, *243*, 54−64.
(57) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. A three-dimensional potential of mean force to improve backbone and sidechain hydrogen bond geometry in Xplor-NIH protein structure determination. *Protein Sci.* **2020**, *29*, 100−110.
(58) *The PyMOL Molecular Graphics System*, version 2.0; Schrödinger, LLC, 2000.
(59) Heinig, M.; Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **2004**, *32*, W500−W502.

https://doi.org/10.1021/jacs.2c03320
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX