

CTD of SARS-CoV-2 N protein is a cryptic domain for binding ATP and nucleic acid that interplay in modulating phase separation

Mei Dang | Jianxing Song 

Department of Biological Sciences,
Faculty of Science, National University of
Singapore, Singapore

Correspondence

Jianxing Song, Department of Biological
Sciences, Faculty of Science; National
University of Singapore, 10 Kent Ridge
Crescent, Singapore 119260.
Email: dbssjx@nus.edu.sg

Funding information

Ministry of Education of Singapore,
Grant/Award Number: R-154-000-B92-114

Abstract

SARS-CoV-2 nucleocapsid (N) protein plays essential roles in many steps of the viral life cycle, thus representing a key drug target. N protein contains the folded N-/C-terminal domains (NTD/CTD) and three intrinsically disordered regions, while its functions including liquid–liquid phase separation (LLPS) depend on the capacity in binding various viral/host-cell RNA/DNA of diverse sequences. Previously NTD was established to bind various RNA/DNA while CTD to dimerize/oligomerize for forming high-order structures. By NMR, here for the first time we decrypt that CTD is not only capable of binding S2m, a specific probe derived from SARS-CoV-2 gRNA but with the affinity even higher than that of NTD. Very unexpectedly, ATP, the universal energy currency for all living cells with high cellular concentrations (2–16 mM), specifically binds CTD with K_d of 1.49 ± 0.28 mM. Strikingly, the ATP-binding residues of NTD/CTD are identical in the SARS-CoV-2 variants while ATP and S2m interplay in binding NTD/CTD, as well as in modulating LLPS critical for the viral life cycle. Results together not only define CTD as a novel binding domain for ATP and nucleic acid, but enforce our previous proposal that ATP has been evolutionarily exploited by SARS-CoV-2 to complete its life cycle in the host cell. Most importantly, the unique ATP-binding pockets on NTD/CTD may offer promising targets for design of specific anti-SARS-CoV-2 molecules to fight the pandemic. Fundamentally, ATP emerges to act at mM as a cellular factor to control the interface between the host cell and virus lacking the ability to generate ATP.

KEYWORDS

ATP; Nucleic acid, liquid–liquid phase separation (LLPS), NMR spectroscopy, nucleocapsid (N) protein, SARS-CoV-2

1 | INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused the ongoing pandemic,¹ which already led to >243 millions of infections and > 4.94 millions of deaths. It belongs to a large family of positive-stranded

RNA coronaviruses with ~30 kb genomic RNA (gRNA) packaged with nucleocapsid (N) protein into a membrane-enveloped virion. SARS-CoV-2 has four structural proteins: namely the spike (S) protein that recognizes the host-cell receptors angiotensin converting enzyme-2 (ACE2), membrane-associated envelope (E),

membrane (M) proteins, and N protein. SARS-CoV-2 N protein is a 419-residue multifunctional protein (Figure S1a), which is composed of the folded N-terminal domain (NTD) over residues 44–173 and C-terminal domain (CTD) over 248–365 (Figure S1b), as well as three intrinsically-disordered regions (IDRs) respectively over 1–43, 174–247, and 366–419. Previous studies have established that its NTD is an RNA-binding domain (RBD) functioning to bind various RNA and DNA of specific and non-specific sequences,² while CTD acts to dimerize/oligomerize to form high-order structures.³ Coronavirus N proteins appear to have two major categories of functions: while their primary role is to assemble gRNA and N protein to form the viral gRNA–N protein (vRNP) complex into the new virions at the final stage of the infection, they also act to suppress the immune system of the host cell and to hijack cellular machineries to achieve the replication of the virus, such as to interfere in the formation of stress granules (SGs) and to localize gRNA onto the replicase–transcriptase complexes (RTCs).^{4–9}

Very recently, liquid–liquid phase separation (LLPS), the emerging principle for commonly organizing the membrane-less organelles (MLOs) or compartments critical for cellular physiology and pathology,^{10–12} has been identified as the key mechanism underlying the diverse functions of SARS-CoV-2 N protein.^{4–9} Nevertheless, all known functions of N proteins including LLPS appear to be dependent on its capacity in binding various viral/host-cell nucleic acids including single- and double-stranded RNA/DNA of diverse sequences. Indeed, LLPS of SARS-CoV-2 N protein appears to be mainly driven by its dynamic and multivalent interactions with nucleic acids of specific and non-specific sequences, as N protein with nucleic acids completely removed lacks the capacity in LLPS.^{4,9} In particular, although the detailed mechanism still remains poorly understood, the final package of the RNA genome into new virions requires the complex but precise interaction between gRNA and N protein, which should be extremely challenging for SARS-CoV-2 with such a large RNA genome (~30 kb). In this context, any small molecules capable of intervening in the interaction of N protein with nucleic acids are anticipated to critically modulate key steps of the viral life cycle, some of which may manifest the anti-SARS-CoV-2 activity.

Mysteriously, ATP, the universal “biological fuel” for all living cells, has cellular concentrations from 2 to 16 mM depending on the types of cells, which are much higher than required for its classic functions.^{12–19} On the other hand, viruses lack the ability to generate ATP.¹³ Recently emerging results indicate that ATP appears to have a novel category of energy-independently functions

at mM, which include the biphasic modulation of LLPS and specific binding to a list of the folded nucleic-acid-binding domains.^{12,14–19} Very unexpectedly, we recently found that ATP biphasically modulates LLPS of SARS-CoV-2 N protein as well as to specifically bind a pocket on its NTD with the dissociation constant (K_d) of 3.3 ± 0.4 mM.⁴ Therefore, ATP at mM not only has novel functions to control protein hemostasis in living cells, but also appears to be evolutionarily hijacked by SARS-CoV-2 to promote its life cycle in the host cells.^{4,16}

So far, the exact mechanisms for ATP and nucleic acids to interact with SARS-CoV-2 N protein as well as to modulate its LLPS still remains not well understood although its understanding is critical for developing therapeutic strategies/molecules to fight the pandemic by targeting N protein. Here by use of S2m as a specific probe, which is a 32-mer nucleic acid stem-loop II motif (Figure S1c) derived from SARS-CoV-2 gRNA,^{20,21} we utilized NMR spectroscopy and DIC microscopy to address three key questions: (a) can CTD of SARS-CoV-2 N protein functioning to dimerize also bind ATP and S2m? (b) if yes, how do ATP and S2m interplay in binding NTD and CTD? (c) can ATP and S2m also interplay in modulating LLPS of SARS-CoV-2 N protein?

The obtained results here decode for the first time that CTD is not only capable of binding S2m, but with the affinity even higher than NTD, thus defining CTD as a novel nucleic-acid-binding domain. Furthermore, ATP also specifically binds CTD uniquely at two pockets located over the dimerization interface with the average K_d of 1.5 ± 0.2 mM, which is the highest affinity identified so far for the folded domains binding ATP at mM. Strikingly, ATP and S2m interplay in binding both NTD and CTD as well as in modulating LLPS of SARS-CoV-2 N protein. Intriguingly, the ATP-binding residues of NTD/CTD are identical in the SARS-CoV-2 variants (Figure S2), thus strongly implying that ATP is essential for SARS-CoV-2 to complete its life cycle in the host cell. Therefore, the present study further validates the capacity of N protein in binding nucleic acids as a key drug target. Most importantly, the identification of the unique ATP-binding pockets on NTD/CTD offers the promising foundation for design of specific anti-SARS-CoV-2 molecules to fight the pandemic.

2 | RESULTS

2.1 | ATP specifically binds two pockets on the dimerization interface of CTD

CTD of SARS-CoV-2 N protein was recently determined to exist with a well-folded structure only as a dimer as

demonstrated by its crystal structure³ and NMR assignment of SARS-CoV-2 CTD (247–364) in solution.²² So far, it still remains completely unexplored whether ATP is able to bind CTD. Here we first address this question by NMR-based titrations of ATP into CTD because NMR spectroscopy is a very powerful tool for gaining the residue-specific knowledge even for the very weak binding events associated with unstable and aggregation-prone protein samples, which are usually not amenable to investigations by most other biophysical methods such as ITC.^{22–25}

As shown in Figure 1a, the HSQC spectrum of CTD is well-dispersed with peaks highly superimposable to those of CTD in a slightly different solution condition,²² thus indicating that CTD is well-folded in solution. Furthermore, in the very up-field region of one-dimensional proton NMR spectra (III of Figure 1a), the dimeric CTD has two large peaks respectively at -0.25 and -0.58 ppm as well as some small peaks, which are from the methyl groups with the close contact to the aromatic rings only manifested in the well-folded proteins.

Addition of ATP at concentrations below 1 mM only led to minor shift of HSQC peaks, suggesting that ATP has no binding to CTD at concentrations below mM. Nevertheless, upon increasing ATP concentrations, a small set of HSQC peaks was found to undergo the large shift, which was largely saturated at 8 mM (II of Figure 1a). Noticeably, ATP only induced a slight shift of the peak at -0.25 ppm but not the other at -0.58 ppm. As only slight changes were observed on these characteristic up-field peaks, CTD appears to remain as a dimer even upon binding ATP. On the other hand, as in the ATP solution used for titrations, $MgCl_2$ was added for stabilizing the ATP structure, we thus titrated CTD with $MgCl_2$ in the same buffer. However, even with $MgCl_2$ concentrations up to 20 mM, no major change of HSQC spectra of CTD has been observed (Figure S3), unambiguously indicating that $MgCl_2$ has no detectable binding to CTD at these concentrations.

Detailed analysis of HSQC spectra in the presence of ATP at different concentrations revealed that only 12 HSQC peaks were significantly shifted which are from

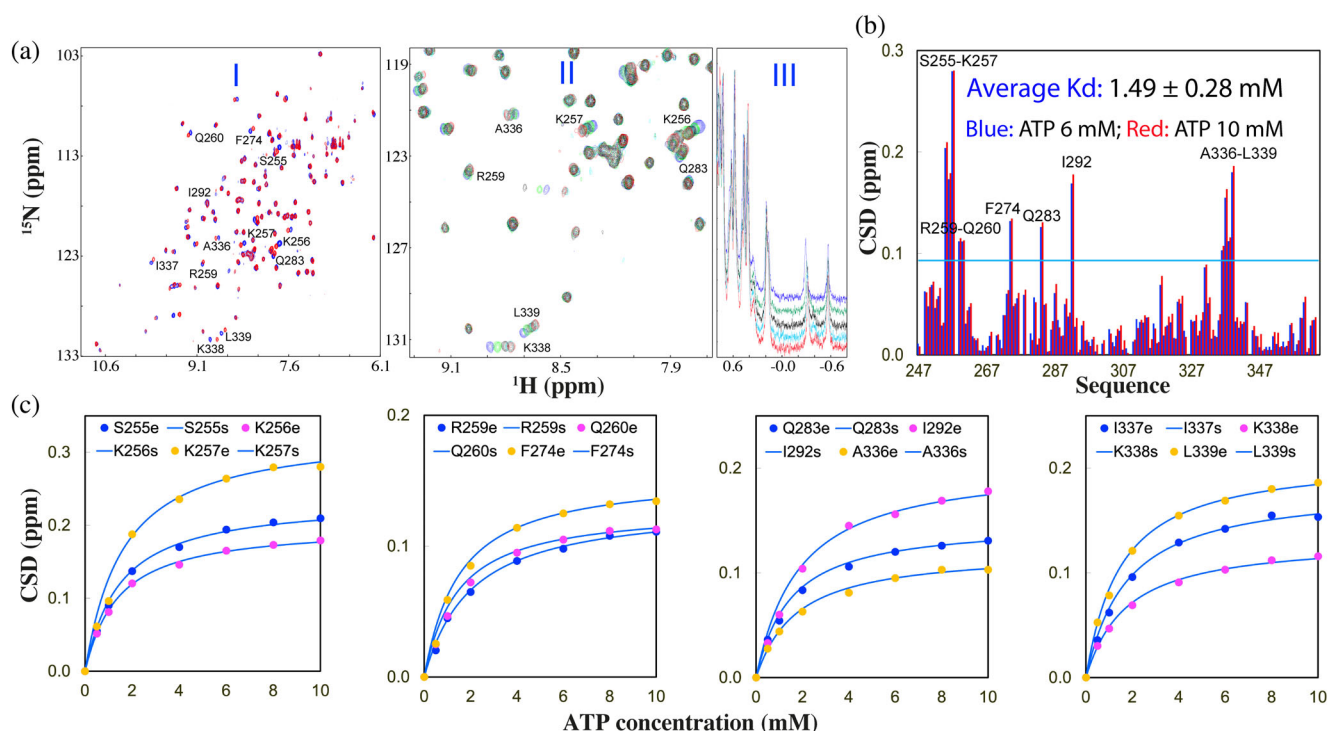


FIGURE 1 ATP specifically binds CTD of SARS-CoV-2 N protein. (a) (I) Superimposition of HSQC spectra of CTD in the free state (blue) and in the presence of ATP at 10 mM (red). (II) Zoom of HSQC spectra of CTD in the absence (blue) and in presence of ATP at different concentrations. For clarity, only spectra in the presence of ATP at four concentrations were shown here: 0 mM (blue); 1 mM (green); 4 mM (black); 6 mM (cyan); and 10 mM (red). The assignments of the significantly perturbed residues are labeled. (III) One dimensional NMR proton spectra of CTD showing the very up-field NMR peaks of methyl groups in the presence of ATP at 0 mM (blue); 1 mM (green); 4 mM (black); 6 mM (cyan) and 10 mM (red). (b) Residue-specific chemical shift difference (CSD) of CTD in the presence of ATP at 6 mM (blue) and 10 mM (red). The significantly perturbed residues are labeled, which are defined as those with the CSD values at 10 mM > 0.095 (average value + one standard deviation) (cyan line). (c) Fitting of the residue-specific dissociation constants (K_d) of the 12 residues: experimental (dots) and fitted (lines) values for the CSDs induced by addition of ATP at different concentrations

residues Ser255-Lys257, Arg259-Gln260, Phe274, Gln283, Ile292 and Ala336-Leu339 (Figure 1a, b). Subsequently, we conducted the fitting of the shift tracings to obtain the residue-specific K_d values of 12 residues (Figure 1c, Table S1), from which the average K_d was calculated to be 1.49 ± 0.28 mM.

Although the 12 residues are distributed over the whole CTD sequence (Figure 1b), they are clustered together to form two pockets on the dimeric CTD structure (Figure 2a). Therefore, with NMR-derived constraints, the structure of the ATP-CTD complex was constructed (Figure 2b) by the well-established HADDOCK program.^{15,25,26} As illustrated by the lowest energy structure, ATP has two binding pockets on the dimeric CTD (Figure 2b), which are located over the dimerization interface and constituted by the residues from both monomers (Figure 2b). Interestingly, the two ATP-binding pockets appears to be highly positively charged (Figure 2c). In the complex, the purine rings of ATP have close contacts with the highly-positive surface constituted by Lys256, Arg259, and Arg262 of one monomer as well as Lys338 of another monomer, thus establishing π -cation and π - π interactions between the ATP purine ring and side chains of Arg/Lys residues as previously observed on other ATP-protein complexes.¹⁵⁻¹⁷ On the

other hand, the triphosphate groups have electrostatic interactions with the positive surface formed mainly by Lys342. Compared to the ATP-binding pockets previously identified in other domains capable of binding ATP at mM,^{4,16,18} the pocket on CTD is highly positive constituted by a cluster of Arg/Lys residues, which may contribute to its high binding affinity for ATP. Intriguingly, the 12 residues were found to be identical in the variants (Figure S2), strongly implying that the binding of ATP to CTD might be essential for SARS-CoV-2 to complete its life cycle in the host cell.

2.2 | ATP and S2m interplay in binding NTD

Very recently, NTD of SARS-CoV-2 N protein was determined by NMR and HADDOCK docking to utilize a conserved surface to bind various nucleic acids with K_d of $\sim \mu\text{M}$.² Furthermore, previously we have also determined that ATP specifically binds a pocket within the conserved nucleic-acid-binding surface of NTD of SARS-CoV-2 N protein.⁴ Briefly, 11 residues distributed over the whole NTD sequence have been significantly perturbed upon binding ATP, which include Asn48, Ser51, Leu56, Thr57,

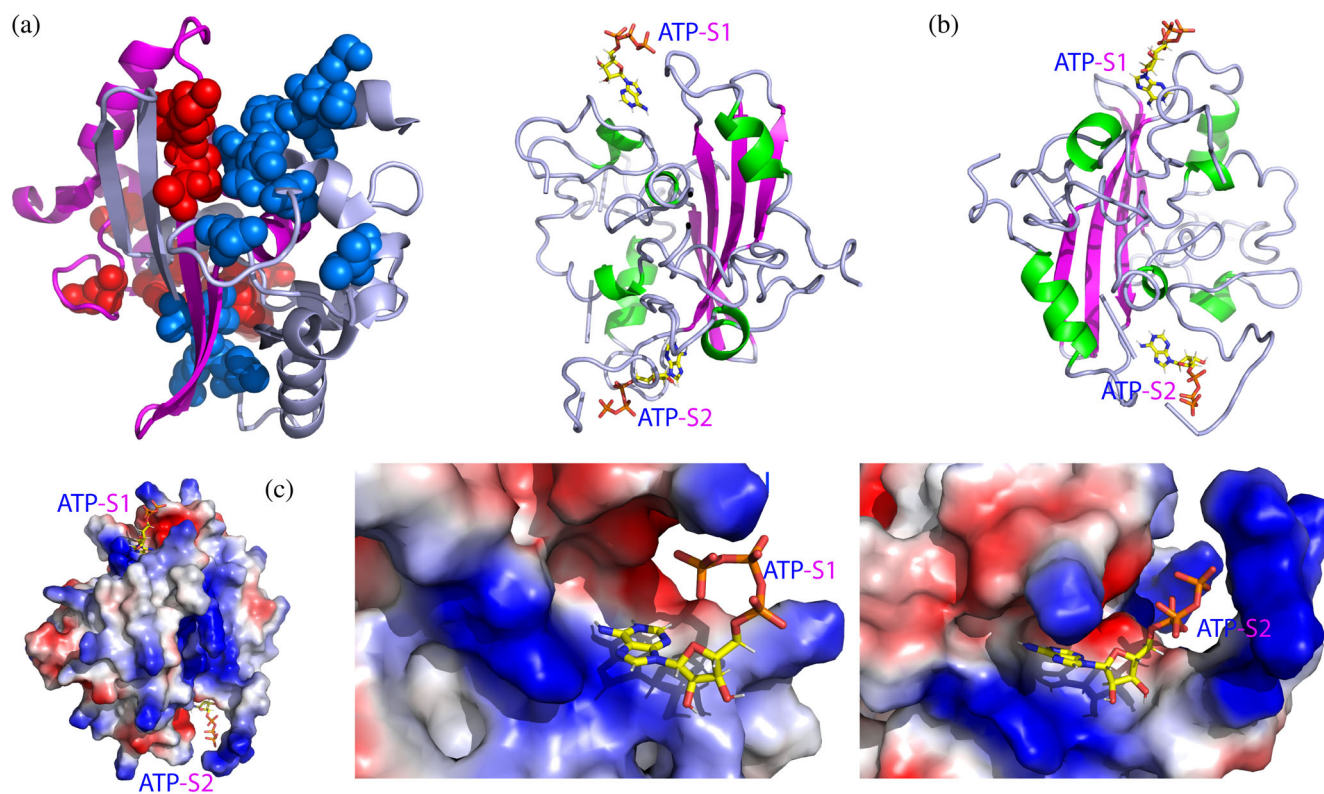


FIGURE 2 The docking structure of the ATP-CTD complex. (a) 12 significantly perturbed residues displayed in sphere on the dimeric CTD structure upon addition of ATP at 10 mM. The residues in one monomer are colored in blue while those in another monomer in red. Structure of the ATP-CTD complex with ATP in sticks and CTD in ribbon (b) and in electrostatic potential surface (c)

Arg89, Ala90, Arg92, Ser105, Arg107, Ala155, and Tyr172. The average K_d of 3.3 ± 0.4 mM was obtained by fitting of the shift tracings of the 11 residues.⁴ In fact, NTD represents the first viral domain capable of binding ATP at mM^{4,16} and unexpectedly all 11 residues are identical in the SARS-CoV-2 variants (Figure S2), thus implying that the binding of ATP to NTD might be also essential for SARS-CoV-2 to complete its life cycle in the host cells.

Here, we first aimed to establish a NTD-nucleic acid binding assay by titrating the ¹⁵N-labeled NTD sample with S2m (Figure S1c), which is a highly conserved sequence among coronaviruses and has been previously utilized to identify the nucleic-acid-binding domains of SARS-CoV-1 proteins.²⁰ Indeed, as monitored by NMR, S2m could bind NTD characteristic of broadening of HSQC peaks upon a stepwise addition of S2m (Figure 3a), similar to the previous observations on other nucleic acid sequences.² Briefly, at 1:0.5, (NTD:S2m), many HSQC peaks became largely broadened (I of

Figure 3a) while at 1:1, some peaks already became too broad to be detected (II of Figure 3a). At 1:2.5, a large portion of HSQC peaks are too broad to be detectable (III of Figure 3a). Further addition of S2m led to no major change of HSQC spectra, indicating that the binding of NTD with S2m is largely saturated at 1:2.5.

On the other hand, over the very up-field region of one-dimensional proton NMR spectra (IV of Figure 3a), NTD has two large peaks respectively at -0.78 and -1.37 ppm together with several small peaks, which are from the methyl groups with the close contact to the aromatic rings. Interestingly, upon stepwise addition of S2m, the two large peaks mainly underwent the broadening while some small peaks also showed the shift. The fact that the S2m binding mainly triggered the peak broadening due to the intermediate exchange on NMR time scale implies the K_d value of $\sim \mu\text{M}$,^{2,23–25} completely consistent with the recent report on the binding affinity of NTD with other nucleic acid sequences.²

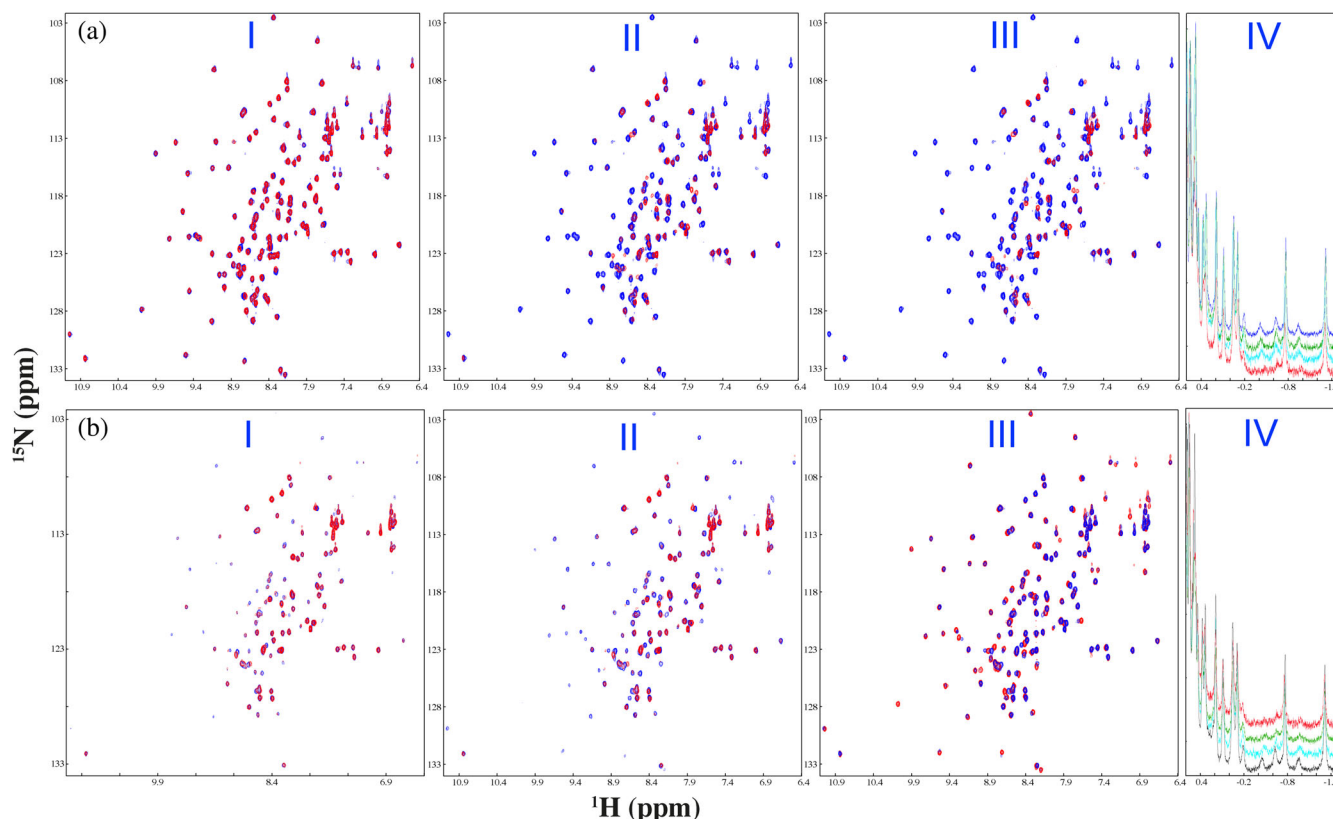


FIGURE 3 NMR characterization of the interplay of ATP and S2m in binding NTD. (A) Superimposition of HSQC spectra of NTD at 100 μM in the free state (blue) and in the presence of S2m (red) at 1:0.5 (I), 1:1.0 (II), and 1:2.5 (III) (NTD:S2m). (IV) One dimensional NMR proton spectra of NTD showing the very up-field NMR peaks of methyl groups in the absence of S2m (blue) and in the presence of S2m at 1:0.5 (green); 1:1 (cyan); and 1:2.5 (red). (B) Superimposition of HSQC spectra of NTD in the presence of S2m at 1:2.5 with additional addition of ATP at 5 mM (I) and 10 mM (II). (III) Superimposition of HSQC spectra of NTD in the presence of S2m at 1:2.5 only (blue) and in the presence of both S2m at 1:2.5 and ATP at 10 mM (red). (IV) One dimensional NMR proton spectra of NTD in the presence of S2m at 1:2.5 (red), as well as with additional addition of ATP at 5 mM (green) and 10 mM (cyan), and in the presence of ATP only at 10 mM (black)

We further assessed whether the ATP binding can affect the binding of S2m with NTD. As shown in Figure 3b, the addition of ATP at 5 mM (I of Figure 3b) and 10 mM (II of Figure 3b) into the NTD sample with the pre-existence of S2m at 1:2.5 (NTD:S2m) could indeed restore some disappeared HSQC peaks but even at 10 mM mimicking the average ATP concentrations in human cells, the HSQC peaks were not completely restored (III of Figure 3b). Consistent with the observation on the amide protons, addition of ATP could also reduce the S2m-induced broadening of the very up-field peaks from the methyl groups (IV of Figure 3b).

To exclude the possibility that ATP may disrupt the S2m-NTD complex by directly interacting with S2m, we thus collected one-dimensional NMR proton spectra of S2m in the presence of ATP at different concentrations. In particular, as shown in Figure S4, S2m has two NMR peaks respectively at 8.22 and 8.37 ppm from its aromatic ring protons which show no overlap with those of ATP respectively at 8.17 and 8.43. Strikingly the addition of ATP to 5 and 10 mM led to no shift of the two peaks of S2m, indicating that ATP has no detectable interaction with S2m. Therefore, the results together indicate that by binding to NTD but not S2m, ATP is capable of disrupting the S2m-NTD complex. However, even at 10 mM, the interaction between NTD and S2m was not completely blocked. This might be due to the very low binding affinity of ATP to NTD (\sim mM) as compared to that of S2m to NTD (\sim μ M), or/and S2m binds NTD with additional interfaces.²

Moreover, we also stepwise added S2m at 1:1 (I of Figure S5) and 1:2.5 (II of Figure S5) to the NTD sample in the presence of ATP at 10 mM. Interestingly, the HSQC spectrum of NTD in the presence of both ATP at 10 mM and S2m at 1:2.5 still showed some difference from that only in the presence of S2m at 1:2.5 (III of Figure S5), indicating that although the presence of ATP at 10 mM failed to completely block the binding of NTD to S2m, but did intervene in the binding of NTD to S2m to certain extent.

2.3 | ATP and S2m interplay in binding CTD

In parallel, we also titrated the ¹⁵N-labeled CTD with S2m and unexpectedly S2m was found to bind CTD also characterized by the broadening of HSQC peaks (Figure 4a). In fact, many HSQC peaks started to become broad even at 1:0.1 (I of Figure 4a). At 1:0.5, many peaks already became too broad to be detected (II of Figure 4a) while at 1:1 almost all HSQC peaks of the backbone amide protons were disappeared (III of

Figure 4a) due to the binding-triggered intermediate exchange on NMR time scale.^{22–24} Furthermore, although S2m also induced the broadening of two large up-field peaks at -0.25 and -0.58 ppm as well as shift of some small peaks (IV of Figure 4a), the two peaks still retained at 1:1. The result thus suggests that CTD remains folded most likely with the dimeric structure even upon binding S2m. Intriguingly S2m appears to bind CTD with the affinity even higher than that for NTD. Unfortunately, the extensive disappearance of NMR peaks prevents further determination of the three-dimensional structure of the CTD-S2m complex, while the severe aggregation of the CTD-S2m complex at high concentrations make it impossible to solve the crystal structure of the complex too.

We also assessed whether the ATP binding can affect the binding of S2m with CTD. As shown in Figure 4b, the addition of ATP at 5 mM (I of Figure 4b) and 10 mM (II of Figure 4b) to the CTD sample with the pre-existence of S2m at 1:1 (CTD:S2m) could indeed restore some disappeared HSQC peaks but even at 10 mM, the HSQC peaks were not completely restored (III of Figure 4b), indicating that ATP is only able to partly disrupt the binding between CTD and S2m and ATP even at 10 mM is unable to completely displace the binding of S2m from CTD. Similarly, ATP could only partly reduce the broadening of the very up-field peaks from the methyl groups (IV of Figure 4b). This again might be due to the very low binding affinity of ATP to CTD (\sim mM) as compared to that of S2m to CTD (\sim μ M), or/and S2m binds CTD with additional interfaces.

Furthermore, we also stepwise added S2m at 1:0.5 (I of Figure S6) and 1:1 (II of Figure S6) to the CTD sample in the presence of ATP at 10 mM. Interestingly, the HSQC spectrum of CTD in the presence of both ATP at 10 mM and S2m at 1:1 showed some difference from that only in the presence of S2m at 1:1 (III of Figure S6), which clearly indicates that although ATP at 10 mM could not completely block S2m from binding CTD, it did intervene in the binding of S2m to NTD to some degree.

2.4 | ATP and S2m interplay in modulating LLPS of N protein

Very recently, the SARS-CoV-2 N protein was shown to achieve its functionality through LLPS, which was induced by dynamic and multivalent interactions with various nucleic acids including viral and host-cell RNA/DNA as well as non-specific sequences.^{4–9} Indeed, previously we found that both ATP and A24, a 24-mer non-specific nucleic acid, could biphasically modulate

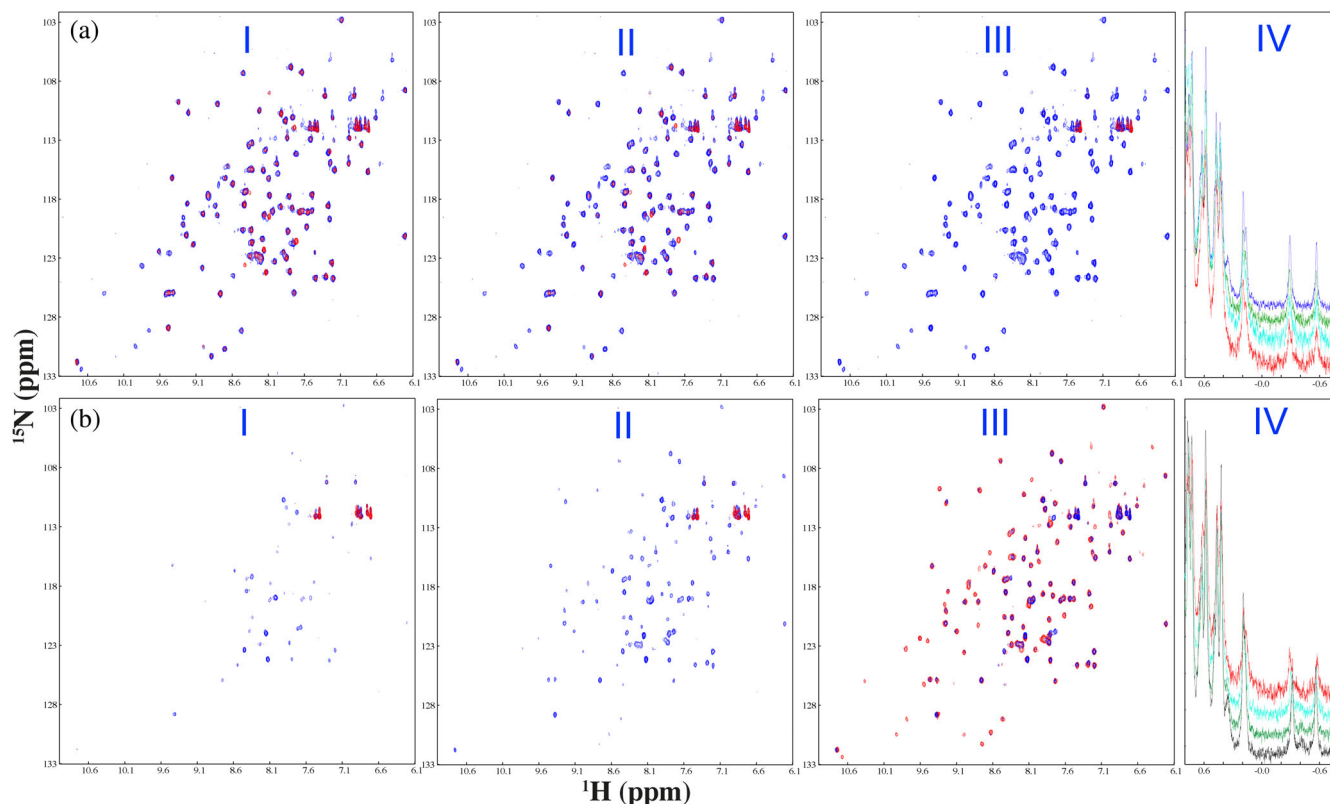


FIGURE 4 NMR characterization of the interplay of ATP and S2m in binding CTD. (a) Superimposition of HSQC spectra of CTD at 200 μM in the free state (blue) and in the presence of S2m (red) at 1:0.1 (I), 1:0.5 (II), and 1:1 (III) (CTD:S2m). (IV) One dimensional NMR proton spectra of CTD showing the very up-field NMR peaks of methyl groups in the absence of S2m (blue) and in the presence of S2m at 1:0.1 (green); 1:0.5 (cyan); and 1:1 (red). (b) Superimposition of HSQC spectra of CTD in the presence of S2m at 1:1 with additional addition of ATP at 5 mM (I) and 10 mM (II). (III) Superimposition of HSQC spectra of CTD in the presence of S2m at 1:1 only (blue) and in the presence of both S2m at 1:1 and ATP at 10 mM (red). (IV) One dimensional NMR proton spectra of CTD in the presence of S2m at 1:1 (red), as well as with additional addition of ATP at 5 mM (cyan) and 10 mM (green), and in the presence of ATP only at 10 mM (black)

LLPS of N protein: induction at low concentrations but dissolution at high concentrations.⁴ Here we asked the questions how the specific S2m motif modulates LLPS of SARS-CoV-2 N protein and whether ATP has any effect on LLPS induced by S2m.

Here we titrated S2m into the full-length N protein with its LLPS monitored both by measuring the turbidity (absorption at 600 nm) and imaging with DIC microscopy, as we previously conducted on SARS-CoV-2 N protein,⁴ FUS,¹⁴ and TDP-43.¹⁹ As shown in Figure S7, S2m could biphasically modulate LLPS of N protein: induction at low ratios and dissolution at high ratios. Briefly, the N protein sample showed no LLPS in the free state, but LLPS could be induced upon addition of S2m as monitored by turbidity and DIC imaging. At 1:0.75 (Nprotein:S2m), the turbidity reached the highest value of 1.92 (Figure S7a) and many liquid droplets with the diameter of $\sim 1 \mu\text{m}$ were formed (Figure S7b). However, further addition of S2m led to the reduction of turbidity and dissolution of the droplets. At 1:1.5 all liquid droplets were dissolved (Figure S7b).

Subsequently we prepared the phase separated N protein sample with the pre-presence of S2m at 1:0.75 and then added ATP into this sample in a stepwise manner, as monitored by turbidity (Figure 5a) and DIC imaging (Figure 5b). At the ratio $<1:50$, ATP only has minor effect on LLPS. However, at ratio of 1:250, ATP started to dissolve liquid droplets as evidenced by the large reduction of turbidity and disappearance of liquid droplets. Strikingly, at 1:750 (Nprotein:ATP), ATP could completely dissolve liquid droplets of N protein induced by S2m. The results demonstrate that ATP and S2m do interplay in modulating LLPS of SARS-CoV-2 N protein.

3 | DISCUSSION

So far, a great success has been achieved for developing the spike-based vaccines to combat the pandemic. Nevertheless, challenges still remain to completely terminate the pandemic, which include the rapidly emerging antibody-resistance variants,²⁷ the adverse effect of the

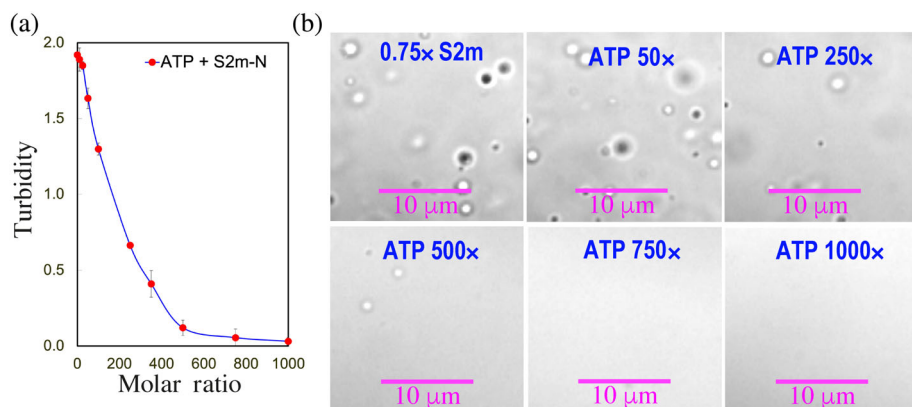


FIGURE 5 ATP disrupts LLPS of SARS-CoV-2 N protein induced by S2m. (a) Turbidity curves of N protein in the presence of S2m at 1:0.75 with additional addition of ATP at different ratios. (b) DIC images of N protein in the presence of S2m at 1:0.75 with additional addition of ATP at different ratios

spike protein.^{28,29} SARS-CoV-2 spike protein has also been identified to provoke the antibody-dependent enhancement (ADE) of infection^{30,31} while its RNA fragments were shown to become integrated into human genome.³² Therefore, any small molecules that directly target SARS-CoV-2 proteins to disrupt the viral life cycle are extremely valuable and urgently demanded to finally terminate the pandemic.

Out of SARS-CoV-2 proteins, N protein is the only one which plays the essential role in almost all key steps of the viral life cycle, thus representing a key target for drug design. Mechanistically, the functions of N protein including LLPS depend on its capacity in interacting with a variety of the viral and host-cell RNA/DNA of diverse sequences. This implies that in evolution only the SARS-CoV-2 variants with their N protein functional in binding nucleic acids can survive and spread. Indeed, the sequences of NTD and CTD critical for binding nucleic acids are highly conserved in the variants of SARS-CoV-2 (Figure S2). In this context, the delineation of the mechanisms for SARS-CoV-2 N protein to interact with nucleic acids as well as its modulation represents a key task not only to gain the biological insight into the mechanism underlying the viral life cycle, but most importantly provides an essential foundation for developing therapeutic strategy/molecules.

However, it is challenging to biophysically characterize the binding mechanisms or/and parameters of SARS-CoV-2 N protein with nucleic acids at least due to: (a) existence of three long intrinsically disordered regions (IDRs) and proneness to oligomerization or/and aggregation; (b) absence of the precise knowledge of the nucleic-acid-binding sites; and (c) induction of LLPS upon adding nucleic acids which thus complicates the interpretation and analysis of the data. In fact, some challenges still exist even for the dissected domains. For example, CTD even with IDRs removed still remains unstable and highly prone to aggregation particularly upon induction by the shearing force provoked such as by the stirring

action required for ITC experiments. As such, while the isolated NTD of SARS-CoV-2 N protein has been well studied and recognized to be a nucleic-acid-binding domain,² it remains unexplored whether CTD of SARS-CoV-2 N protein is able to bind ATP/nucleic acids or not.

In the present study by use of S2m as a specific probe, CTD of SARS-CoV-2 N protein has been decrypted by NMR to be not only capable of binding nucleic acid, but with the affinity even higher than that of NTD. Consequently, the ability of CTD to bind nucleic acids certainly needs to be included not only for establishing the underlying mechanisms for the viral life cycle, but also for design of therapeutic molecules by targeting the nucleic-acid binding of N protein as previously proposed for SARS-CoV-1.^{20,21} In particular, the mechanism of the nucleic-acid binding of CTD has to be elucidated in order to construct the working model of the package of the gRNA–N protein complex for the maturation of SARS-CoV-2 virion.

The present study also decodes an unexpected fact that CTD is also a cryptic domain for binding ATP at mM. Very recently, ATP has been shown to bind a list of the folded nucleic-acid-binding domains with the average K_d values at mM for controlling protein hemostasis in living cells, which include RNA-binding motif (RRM) domains of FUS (K_d of 3.8 mM), TDP-43 (K_d of 2.6 mM), hnRNPA1 (K_d of 4.9 mM)^{15,16} and CIRBP (K_d of 2.9 mM),¹⁷ as well as an all-helical acidic domain of SYN-CRIP (K_d of 3.1 mM).¹⁶ However, it is still very unexpected that NTD of SARS-CoV-2 N protein was recently found to be a non-classic ATP-binding domain with K_d of 3.3 mM⁴ because viruses have no ability to generate ATP. Here CTD of SARS-CoV-2 N protein has been further established to be a novel viral ATP-binding domain with three unique properties: (a) CTD adopts a dimeric fold fundamentally different from all previous ones^{4,16,17}; (b) the dimeric CTD has two ATP-binding pockets located on the dimerization interface; and (c) ATP binds CTP with the average K_d of 1.49 mM, which is the

highest identified so far.^{16,17} The finding that ATP specifically binds both NTD and CTD of SARS-CoV-2 N protein also has a fundamental implication in evolution of the host–pathogen interaction. It appears that in evolution the novel category of functions of ATP at mM is not only extensively incorporated into various biological processes in living cells with high ATP concentrations,¹⁶ but also operates to control the interface between the host cell and SARS-CoV-2 which cannot generate ATP by itself.

As human cells have high ATP concentrations with the lowest (~3 mM) in neurons, so a physiologically relevant question to be addressed is whether the binding of ATP to NTD and CTD has any effect on their interaction with nucleic acids. Here by NMR competition experiments, we showed unambiguously that ATP at 10 mM which mimics the average cellular concentrations can indeed partly disrupt the binding of NTD/CTD with S2m. Mechanistically, this disruption may result from the direct competition of ATP and S2m in binding the overlapped surfaces/pockets or/and from the binding-induced changes of their conformations and dynamics of NTD/CTD.^{24,25} In the future, it is of fundamental interest to biophysically elucidate the underlying structural/dynamic mechanisms as we previously conducted by NMR methodology^{24,25} and computational simulations.³³ Nevertheless, the fact that the NTD/CTD residues involved in binding ATP are identical in the SARS-CoV-2 variants (Figure S2) strongly implies that the ATP-binding might be essential for SARS-CoV-2 to complete its life cycle in the host cell.

Indeed, the interplay of ATP and S2m in binding N protein is sufficient for ATP to dissolve LLPS of N protein induced by S2m, which has been recently shown to be the underlying mechanism for the functionality of N protein.^{4–9} On the other hand, the large difference has been revealed for the capacity of ATP and S2m in inducing and dissolving LLPS of N protein, which is most likely due to the fact previously found on other proteins: although ATP and nucleic acids can bind the folded nucleic-acid-binding domains, the binding affinity of ATP is much lower than those of nucleic acids because ATP can only bind bivalently but nucleic acids may bind multivalently.^{14–19,34} Intriguingly, however, this difference appears to be exploited by SARS-CoV-2 to complete its life cycle in the host cell as we previously proposed.⁴ At the early stage of infection when the concentrations of the gRNA–N protein complex are very low, ATP in the host cell acts to dissolve the phase separated state to achieve the uncoating of the gRNA–N protein condensate in order to release gRNA for initiating the viral life cycle. By contrast, at the later stage of infection when the concentrations of N protein and gRNA become very high, the majority of N protein molecules will become

unbounded with ATP and consequently be able to interact with gRNA to start the package for forming the gRNA–N protein complex essential for the maturation of virions.

In summary, for the first time the present study defines CTD of SARS-CoV-2 N protein to be a novel fold for binding ATP and nucleic acids. This finding not only sheds the light on the previously unknown roles of CTD in the viral life cycle, but also provides the structural basis for design of anti-SARS-CoV-2 molecules by targeting the nucleic-acid binding of N protein. The establishment of the ATP-binding pockets on both NTD and CTD offers the promising potential to discover or/and design small molecules which can bind the pockets with high affinity and specificity in order to disrupt the key functions of N protein by inhibiting its binding to nucleic acids. As the unique ATP-binding pockets on SARS-CoV-2 N protein are most unlikely also presented in the human proteins, these small molecules are expected to have the high specificity to target N protein of SARS-CoV-2 to disrupt its roles in the viral life cycle, thus manifesting the specific anti-SARS-CoV-2 activity. Fundamentally, this study also extends the roles of the ATP-binding to the nucleic-acid binding domains at mM beyond controlling protein hemostasis in living cells to acting as a cellular factor which may operate at the interface between the host cell and viruses which lack the ability to generate ATP.

4 | MATERIALS AND METHODS

4.1 | Preparation of recombinant SARS-CoV-2 N protein as well as its NTD and CTD

The gene encoding 419-residue SARS-CoV-2 N protein was purchased from a local company (Bio Basic Asia Pacific Pte Ltd), which was cloned into an expression vector pET-28a with a TEV protease cleavage site between N protein and N-terminal 6xHis-SUMO tag used to enhance the solubility.⁴ The DNA fragments encoding its NTD (44–180) and CTD (247–364) were subsequently generated by PCR reaction and subcloned into the same vector.⁴

The recombinant N protein and its NTD/CTD were expression in *E. coli* cells BL21 with IPTG induction at 18°C, which were found to be soluble in the supernatant. For NMR studies, the bacteria were grown in M9 medium with addition of (¹⁵NH₄)₂SO₄ for ¹⁵N-labeling. The recombinant proteins were first purified by Ni²⁺-affinity column (Novagen) under native conditions and subsequently in-gel cleavage by TEV protease was conducted. The eluted fractions containing the recombinant

proteins were further purified by FPLC chromatography system with a Superdex-200 column for the full-length and a Superdex-75 column for NTD and CTD.⁴ The purity of the recombinant proteins was checked by SDS-PAGE gels and NMR assignment for both NTD and CTD. ATP was purchased from Merck. Protein concentration was determined by spectroscopic method in the presence of 8 M urea.³⁵ As CTD only exists stably as a dimer, the calculation of the ratio associated with CTD in this study is based on the dimeric form.

4.2 | NMR characterizations of the binding of ATP and S2m to NTD and CTD

NMR samples of NTD and CTD were prepared at 100 μ M and 200 μ M, respectively, in 10 mM sodium phosphate buffer (pH 7.0) in the presence of 150 mM NaCl. ATP was dissolved in the same buffer with the addition of equal molar MgCl_2 to stabilize ATP by forming ATP-Mg complex. The final solution pH was adjusted to 7.0 by use of very diluted HCl or NaOH.

NMR experiments were conducted at 25°C on an 800 MHz Bruker Avance spectrometer equipped with pulse field gradient units and a shielded cryoprobe as described previously.^{4,14,15,19} For NMR titrations to determine residue-specific Kd of CTD residues for binding ATP, two dimensional ^1H - ^{15}N NMR HSQC spectra were collected on the ^{15}N -labeled CTD at 200 μ M in the presence of ATP at 0, 0.5, 1, 2, 4, 6, 8, and 10 mM. For assessing the effect of MgCl_2 , HSQC spectra of CTD were collected under the same conditions with addition of MgCl_2 to reach 5, 10, 15, and 20 mM.

For NMR characterization of the binding of NTD and CTD with S2m, HSQC spectra were collected on the ^{15}N -labeled NTD at 100 μ M or CTD at 200 μ M with addition of S2m at 0; 1:0.05; 1:0.1; 1:0.25; 1:0.5; 1:1; 1:1.5; 1:2; 1:2.5; 1:3.5; 1:5 (NTD or CTD:S2m). For NMR investigation on the interplay of ATP and S2m in binding NTD and CTD, HSQC spectra were collected on the ^{15}N -labeled NTD at 100 μ M or CTD at 200 μ M with addition of ATP and S2m at different combinations. NMR data were processed with NMRPipe³⁶ and analyzed with NMRView.³⁷

4.3 | Calculation of CSD and data fitting

Sequential assignments were achieved based on the deposited NMR chemical shifts for CTD with BMRB ID of 50,518.²² To calculate chemical shift difference (CSD), HSQC spectra collected without and with ATP at

different concentrations were superimposed. Subsequently, the shifted HSQC peaks were identified and further assigned to the corresponding CTD residues. The CSD was calculated by an integrated index with the following formula^{4,14,15,23}:

$$\text{CSD} = ((\Delta^1\text{H})^2 + (\Delta^{15}\text{N})^2/4)^{1/2}.$$

In order to obtain residue-specific dissociation constant (Kd), we fitted the shift tracings of the CTD residues with significant shifts (CSD > average + STD) by using the following formula^{4,15,23}:

$$\text{CSD}_{\text{obs}} = \text{CSD}_{\text{max}} \left\{ \frac{([P] + [L] + \text{Kd}) - \left[([P] + [L] + \text{Kd})^2 - 4[P][L] \right]^{1/2}}{2[P]} \right\}.$$

here [P] and [L] are molar concentrations of CTD and ligands (ATP), respectively.

4.4 | Molecular docking

The structures of the ATP-CTD complex were constructed by use of the well-established HADDOCK software^{4,14,15,25,26} in combination with crystallography and NMR system (CNS),³⁸ which makes use of CSD data to derive the docking that allows various degrees of flexibility. Briefly, HADDOCK docking was performed in three stages: (a) randomization and rigid body docking; (b) semi-flexible simulated annealing; and (c) flexible explicit solvent refinement.

The crystal structure³ of CTD (PDB ID of 6YUN) were used for docking to ATP. Only one cluster of structures with slight differences of ATP orientations were observed for the 10 lowest-energy ATP-CTD complex. Out of them, the ATP-CTD structure with the lowest energy score was selected for the detailed analysis and display by Pymol (The PyMOL Molecular Graphics System).

4.5 | LLPS imaged by differential interference contrast (DIC) microscopy

The formation of liquid droplets was imaged on 50 μ l of the N protein samples by DIC microscopy (OLYMPUS IX73 Inverted Microscope System with OLYMPUS DP74 Color Camera) as previously described.^{4,14,19} The full-length N protein samples were prepared at 10 μ M in 25 mM HEPES buffer (pH 7.0) with 70 mM KCl. The turbidity (absorption at 600 nm) were measured for all DIC samples with three repeats.

ACKNOWLEDGEMENT

This study is supported by Ministry of Education of Singapore (MOE) Tier 1 Grants R-154-000-B92-114 to Jianxing Song.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Mei Dang: Data curation (equal); formal analysis (equal); validation (equal); writing – original draft (equal). **Jianxing Song:** Conceptualization (equal); formal analysis (equal); funding acquisition (equal); project administration (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal).

ORCID

Jianxing Song  <https://orcid.org/0000-0003-0224-6322>

REFERENCES

- Wu F, Zhao S, Yu B, et al. New coronavirus associated with human respiratory disease in China. *Nature*. 2020;579:265–269.
- Dinesh DC, Chalupska D, Silhan J, et al. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog*. 2020;16:e1009100.
- Zinzula L, Basquin J, Bohn S, et al. High-resolution structure and biophysical characterization of the nucleocapsid phosphoprotein dimerization domain from the Covid-19 severe acute respiratory syndrome coronavirus 2. *Biochem Biophys Res Commun*. 2021;538:54–62.
- Dang M, Li Y, Song J. ATP biphasically modulates LLPS of SARS-CoV-2 nucleocapsid protein and specifically binds its RNA-binding domain. *Biochem Biophys Res Commun*. 2021;19:50–55.
- Lu S, Ye Q, Singh D, et al. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat Commun*. 2021;12:502.
- 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;11:6041.
- Iserman C, Roden CA, Boerneke MA, et al. Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Mol Cell*. 2020;80:1078–1091.
- Perdikari TM, Murthy AC, Ryan VH, Watters S, Naik MT, Fawzi NL. SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J*. 2020;17:e106478.
- Carlson CR, Asfaha JB, Ghent CM, et al. Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Mol Cell*. 2020;17:1092–1103.
- Hyman AA, Weber CA, Jülicher F. Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol*. 2014;30:39–58.
- Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. *Science*. 2017;357:6357.
- Patel A, Malinowska L, Saha S, et al. ATP as a biological hydrotrope. *Science*. 2017;356:753–756.
- Lehninger's principles of biochemistry. 5th ed. New York: W.H. Freeman and Company, 2005 pp. 502e503.
- Kang J, Lim L, Lu Y, Song J. A unified mechanism for LLPS of ALS/FTLD-causing FUS as well as its modulation by ATP and oligonucleic acids. *PLoS Biol*. 2019;17:e3000327.
- Kang J, Lim L, Song J. ATP binds and inhibits the neurodegeneration-associated fibrillization of the FUS RRM domain. *Commun Biol*. 2019;20:223.
- Song J. ATP energy-independently controls protein homeostasis with unique structure and diverse mechanisms. *Protein Sci*. 2021;30:1277–1293.
- Zhou Q, Usluer S, Zhang F, Lenard AJ, Bourgeois BMR, Madl T. ATP regulates RNA-driven cold inducible RNA binding protein phase separation. *Protein Sci*. 2021;30:1438–1453.
- Kim TH, Payliss BJ, Nosella ML, et al. Interaction hot spots for phase separation revealed by NMR studies of a CAPRIN1 condensed phase. *Proc Natl Acad Sci U S A*. 2021;118:e2104897118.
- Dang M, Li Y, Song J. ATP biphasically modulates LLPS of TDP-43 PLD by specifically binding arginine residues. *Commun Biol*. 2021;4:714.
- Chen CY, Chang CK, Chang YW, et al. 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.
- Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral Res*. 2014;103:39–50.
- Korn SM, Lambertz R, Fürtig B, et al. (1)H, (13)C, and (15)N backbone chemical shift assignments of the C-terminal dimerization domain of SARS-CoV-2 nucleocapsid protein. *Biomol NMR Assign*. 2021;15:129–135.
- Williamson MP. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc*. 2013;73:1–16.
- Qin H, Lim LZ, Song J. Dynamic principle for designing antagonistic/agonistic molecules for EphA4 receptor, the only known ALS modifier. *ACS Chem Biol*. 2015;10:372–378.
- Qin H, Shi J, Noberini R, Pasquale EB, Song J. Crystal structure and NMR binding reveal that two small molecule antagonists target the high affinity ephrin-binding channel of the EphA4 receptor. *J Biol Chem*. 2008;283:29473–29484.
- Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*. 2003;125:1731–1737.
- Wang P, Nair MS, Liu L, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*. 2021;593:130–135.
- Lei Y, Zhang J, Schiavon CR, et al. SARS-CoV-2 spike protein impairs endothelial function via downregulation of ACE 2. *Circ Res*. 2021;128:1323–1326.
- Hoepel W, Chen HJ, Geyer CE, Allahverdiyeva S, Manz XD, de Taeye SW, Aman J, Mes L, Steenhuis M, Griffith GR, Bonta PI, Brouwer PJM, Caniels TG, van der Straten K, Golebski K, Jonkers RE, Larsen MD, Linty F, Nouta J, van Roomen CPAA, van Baarle FEHP, van Drunen CM, Wolbink G, Vlaar APJ, de Bree GJ, Sanders RW, Willemsen L, Neele AE, van de Beek D,

- Rispens T, Wuhrer M, Bogaard HJ, van Gils MJ, Vidarsson G, de Winther M, den Dunnen J. High titers and low fucosylation of early human anti-SARS-CoV-2 IgG promote inflammation by alveolar macrophages. *Sci Transl Med* 2021;13:eabf8654.
30. Lee WS, Wheatley AK, Kent SJ, DeKosky BJ. Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies. *Nat Microbiol.* 2021;5:1185–1191.
 31. Liu Y, Soh WT, Kishikawa JI, et al. An infectivity-enhancing site on the SARS-CoV-2 spike protein targeted by antibodies. *Cell.* 2021;184:3452–3466.
 32. Zhang L, Richards A, Barrasa MI, et al. Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. *Proc Natl Acad Sci USA.* 2021;118:e2105968118.
 33. Shi J, Han N, Lim L, et al. Dynamically-driven inactivation of the catalytic machinery of the SARS 3C-like protease by the N214A mutation on the extra domain. *PLoS Comput Biol.* 2011;7:e1001084.
 34. Ukmar-Godec T, Hutten S, Grieshop MP, et al. Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nat Commun.* 2019;10:2909.
 35. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 1995;4:2411–2423.
 36. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* 1995;6:277–293.
 37. Johnson BA, Blevins RA. NMRView: A computer program for the visualization and analysis of NMR data. *J Biomol NMR.* 1994;4:603–614.
 38. Brunger AT, Adams PD, Clore GM, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst D.* 1998;54:905–921.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Dang M, Song J. CTD of SARS-CoV-2 N protein is a cryptic domain for binding ATP and nucleic acid that interplay in modulating phase separation. *Protein Science.* 2022;31:345–56. <https://doi.org/10.1002/pro.4221>