Design of anti-SARS-CoV-2 molecules from hydroxychloroquine (HCQ) for better binding SARS-CoV-2 N protein

Mei Dang, Jianxing Song*
Department of Biological Sciences, Faculty of Science; National University of Singapore; 10 Kent Ridge Crescent, Singapore 119260

Short title: Design of better anti-SARS-CoV-2 molecules from HCQ

*** Corresponding author; Email:** dbssjx@nus.edu.sg.

Abstract

SARS-CoV-2 is the coronavirus causing the ongoing pandemic which already led to >129 millions of infections and >2.8 millions of deaths. Hydroxychloroquine (HCQ) has been shown to own promising potential in clinically combating SARS-CoV-2 but the underlying mechanisms remain highly elusive and its action sites were proposed all on the host cells. Very recently, by NMR spectroscopy, for the first time we have decoded that HCQ specifically binds both N-terminal domain (NTD) and C-terminal domain (CTD) of SARS-CoV-2 nucleocapsid (N) protein to inhibit their interactions with nucleic acids (NAs), as well as to disrupt its NA-induced liquid-liquid phase separation (LLPS) essential for the viral life cycle including the final package of gRNA and N protein into new virions. Here with NMR-derived constraints, we constructed the structure of the HCQ-CTD complex in which HCQ binds two pockets on the dimeric CTD. This unexpected finding thus allowed us to formulate a design strategy by linking two HCQ together to generate a bivalent/multivalent binder for CTD. The linked molecule is anticipated to have significantly enhanced affinity and specificity in binding CTD. We are willing to collaborate with synthetic chemists on this strategy by experimentally charactering the synthetic molecules on their binding and effects on LLPS of SARS-CoV-2 N protein.

Key words: Hydroxychloroquine (HCQ): Bivalent binder; SARS-CoV-2; Nucleocapsid (N) protein; C-terminal Domain (CTD); Liquid-liquid phase separation (LLPS); NMR spectroscopy.

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the etiologic agent of the ongoing catastrophic pandemic (1), which already led to >129 millions of infections and >2.8 millions of deaths. It belongs to a large family of positive-stranded RNA coronaviruses with ~30 kb genomic RNA packaged into a membrane-enveloped virion. SARS-CoV-2 has four structural proteins: the spike (S) protein that recognizes cell receptors angiotensin converting enzyme-2 (ACE2), nucleocapsid (N) protein responsible for packing viral genomic RNA (Fig. 1A), membrane-associated envelope (E) and membrane (M) proteins. Very unusually, SARS-CoV-2 appeared to be highly adapted to human from the very beginning of the outbreak (2) and a large number of mutations have been found for its spike protein (3), thus imposing a great challenge to terminate the pandemic by the currently-available vaccines derived from its spike protein.

N protein of SARS-CoV-2 plays multifunctional roles in the viral life cycle, including to finally assemble genomic RNA (gRNA) into new virions, as well as to localize gRNA to replicase-transcriptase complexes (4-14). SARS-CoV-2 N protein is a 419-residue RNA-binding protein composed of two well-folded domains, namely N-terminal domain (NTD) over residues 44-173 and C-terminal domain (CTD) over residues 248-365, as well as three intrinsically-disordered regions (IDRs) respectively over 1-43, 174-247 and 366-419 with low-complexity sequences (Fig. 1A). Previous studies including those on SARS-CoV-1 N protein revealed that its NTD is an RNA-binding domain (RBD) (13) while CTD exists as a dimer as evidenced by its dimeric crystal structure (Fig. 1B) (14).

Very recently, it was decoded that the N protein functions through liquid-liquid phase separation (LLPS), which could be significantly induced by dynamic and multivalent interactions with various nucleic acids (8-12). In particular, we found that ATP, the universal energy currency mysteriously with very high concentrations in all living cells (15-20), could

specifically bind the NTD as well as biphasically modulate LLPS of SARS-CoV-2 N protein (8). This finding thus raised a therapeutically-critical question as whether other small molecules may bind N protein of SARS-CoV-2 or/and modulate its LLPS?

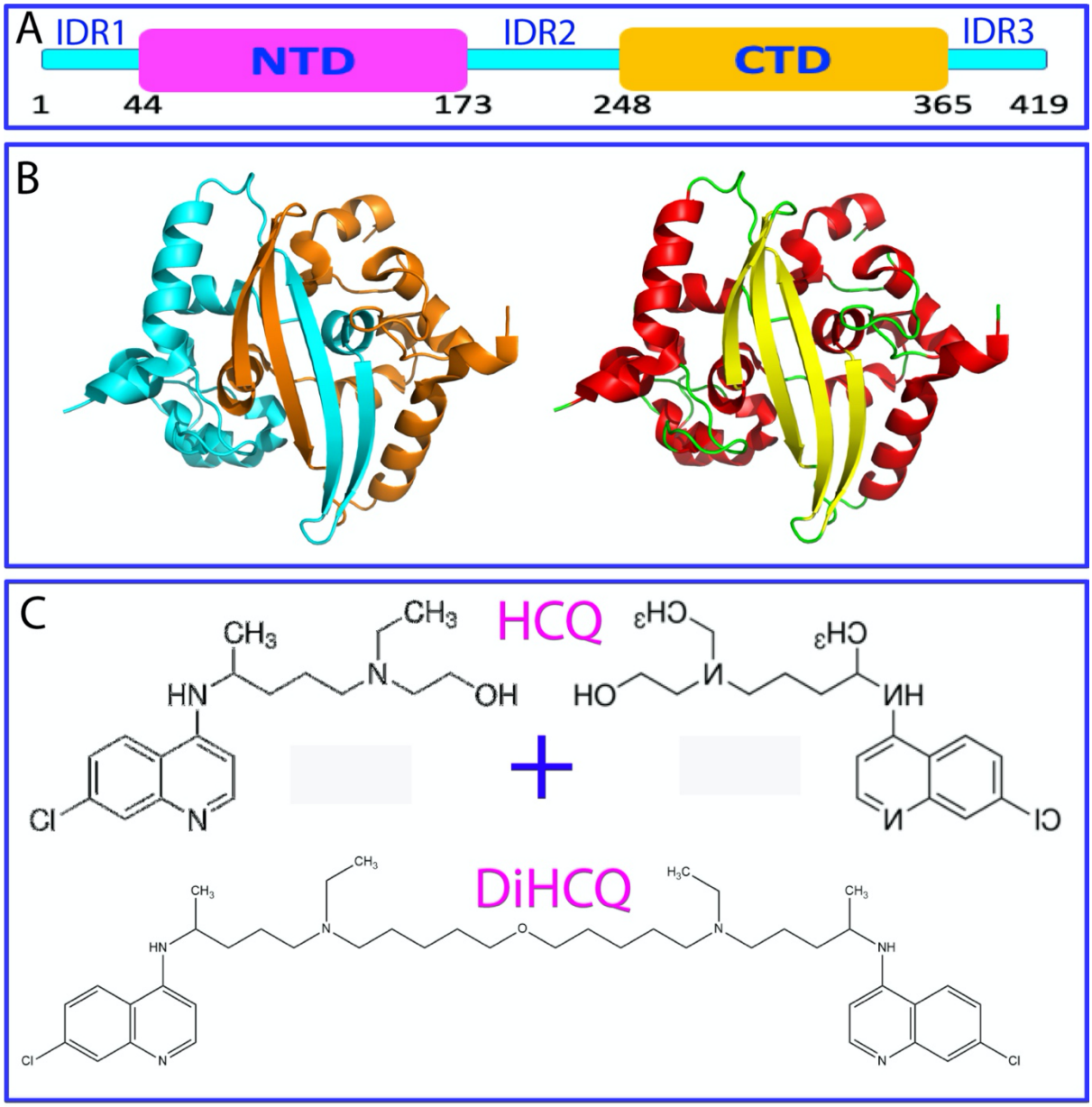

Figure 1. Crystal structure of CTD of SARS-CoV-2 and chemical structures of HCQ abd designed HCQ derivative,
(A) Domain organization of SARS-CoV-2 N protein. (B) Crystal structure of the dimeric CTD of SARS-CoV-2 N protein. (C) Chemical structures of HCQ and linked HCQ derivative (DiHCQ).

Hydroxychloroquine (HCQ) has been extensively reported to have promising potential in clinically combating SARS-CoV-2 (21-23), which is an antimalarial drug with

higher water solubility and lower toxicity (Fig. 1C). However, the mechanisms for its anti-SARS-CoV-2 activity remain almost unknown and so far, almost all proposed action sites for HCQ are on the host cells, which include the interference in the endocytic pathway, blockade of sialic acid receptors, restriction of pH mediated S protein cleavage at the ACE2 binding site and prevention of cytokine storm. In particular, no specific viral target protein has been experimentally identified for HCQ. Only very recently, by DIC microscopy and NMR spectroscopy, for the first time we discovered that HCQ could specifically bind both NTD and CTD of SARS-CoV-2 N protein, as well as dissolve its LLPS (24). This study not only shed the first light on the potential anti-SARS-CoV-2 mechanism of HCQ, but also opens an avenue to optimize HCQ to be the better anti-SAES-CoV-2 drug.

In the present study, by use of NMR-derived constraints, we first successfully constructed the structure of the HCQ-CTD complex in which CTD is bound with two HCQ molecules. As such, we further attempted to improve the binding affinity and specificity of HCQ to CTD by linking two HCQ molecules to obtain bivalent binders designated as DiHCQ. Finally, we also conducted molecular dynamics (MD) simulations for CTD in the free state, in complex with HCQ and DiHCQ. The results indicate the binding of HCQ and DiHCQ significantly increased the dynamics of CTD of SARS-CoV-2 N protein, which may contribute to its anti-SARS-CoV-2 activity.

Results

The structure of the HCQ-CTD complex

Here, by use of NMR HSQC spectroscopy we have monitored the titration of HCQ into the ^{15}N-labeled CTD sample. As shown in Fig. 2A, the ^{15}N-labeled CTD (248-365) has a well-dispersed HSQC spectrum, suggesting the dimeric CTD is well-folded. Subsequently, we characterized the binding of HCQ to CTD by collecting a series of HSQC spectra in the presence of HCQ at different concentrations. Indeed, upon adding HCQ, the shift of a small set of HSQC peaks was observed which is saturatable at high HCQ concentrations as evidenced by Fig. 2A, thus indicating that HCQ does specifically bind CTD.

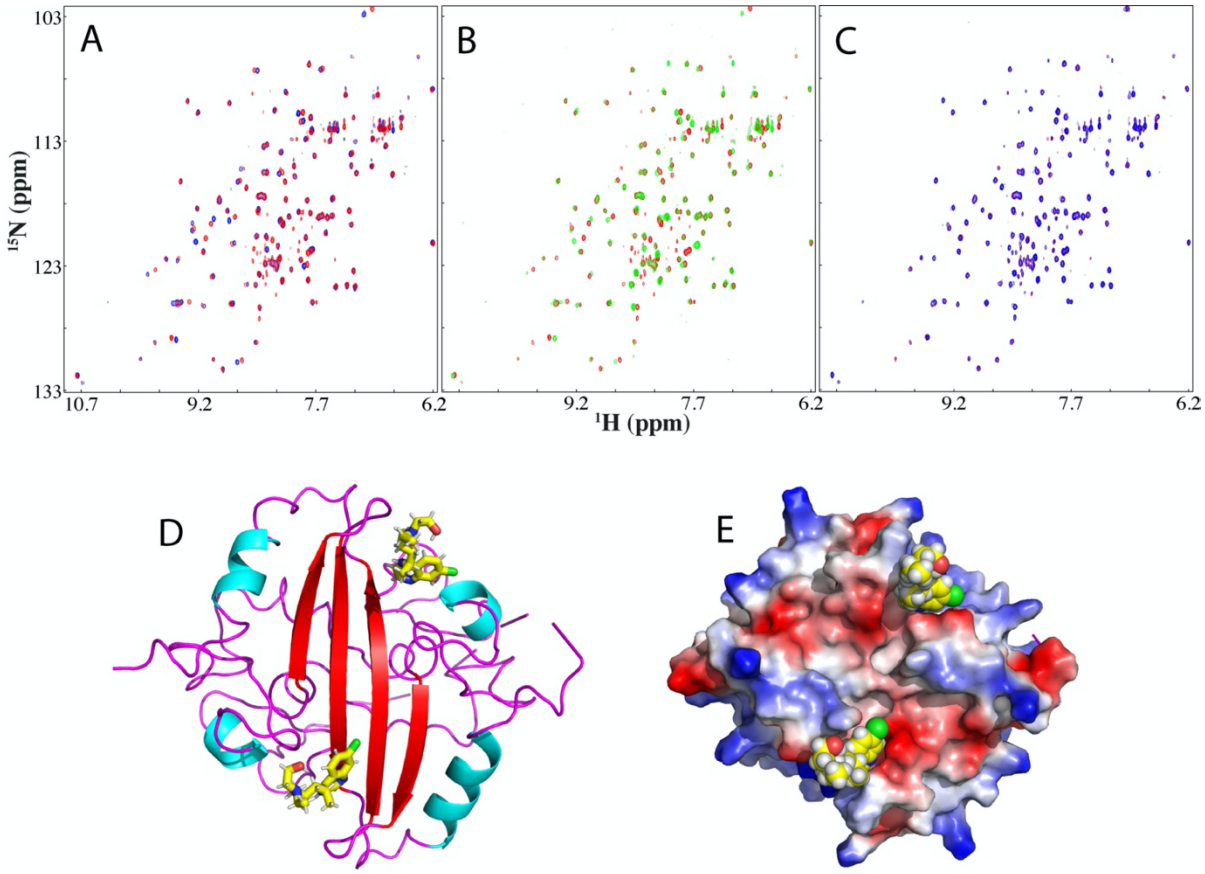

Figure 2. HCQ specifically binds two pockets of the dimeric CTD of SARS-CoV-2 N protein.
(A) Superimposition of HSQC spectra of CTD in the free state (blue) and in the presence of HCQ at the saturated concentrations (red). (B) Superimposition of HSQC spectra of CTD in the presence of HCQ at the saturated concentrations (red) and S2m at the saturated concentrations (green). (C) Superimposition of HSQC spectra of CTD in the presence of HCQ

at the saturated concentrations (red) and with additional addition of S2m at the saturated concentrations (blue). (D) The dimeric CTD structure in complex with two HCQ displayed in stick. (D) The electrostatic potential surface of the dimeric CTD structure in complex with two HCQ displayed in sphere.

We further titrated CTD with S2m, a 32-mer ssDNA, and it induced the disappearance of many HSQC peaks and the shift of a small set of HSQC peaks which is saturatable. The shift patterns induced by HCQ and S2m show a significant difference (Fig. 2B). Strikingly, the addition of S2m at the saturated concentration into the CTD sample saturated with HCQ triggered no significant change of the HSQC spectra (Fig. 2C), implying that the binding of HCQ to CTD blocked its further binding to S2m.

Furthermore, we have assigned HSQC spectra and identified the significantly-perturbed residues. By setting these residues as active residues, we successfully constructed the structure of the HCQ-CTD complex by the well-established program HADDOCK (25) as we previously utilized to build up the small molecule-protein complexes (26,27). As shown by the lowest-energy complex structures, the dimeric CTD has two cavities, both of which are bound with HCQ (Fig. 2D and 2E). Interestingly it appears that the aromatic ring of HCQ plays a key role in binding the CTD residues.

Design of bivalent binder from HCQ

It is well-established that the introduction of bivalent/multivalent binding significantly increases both binding affinity and specificity (28). Therefore, based on the HCQ-CTD structure, we linked two HCQ molecules to obtain a bivalent molecule designated as DiHCQ (Fig. 1C). We then conducted the docking of DiHCQ to CTD with the same setting of the active residues. Interestingly, the lowest-energy structure reveals that DiHCQ could indeed bind to CTD with the two aromatic rings bound to two cavities (Fig. 3). Strikingly, the two cavities are highly overlapped with those for binding HCQ (Fig. 2). This suggests that by properly

introducing the linker, a bivalent binder could indeed be designed. In particular, it is also possible to even engineer the linker to bind the beta-sheet core of CTD to obtain multivalent binders which are expected to have much better binding affinity and specificity (28).

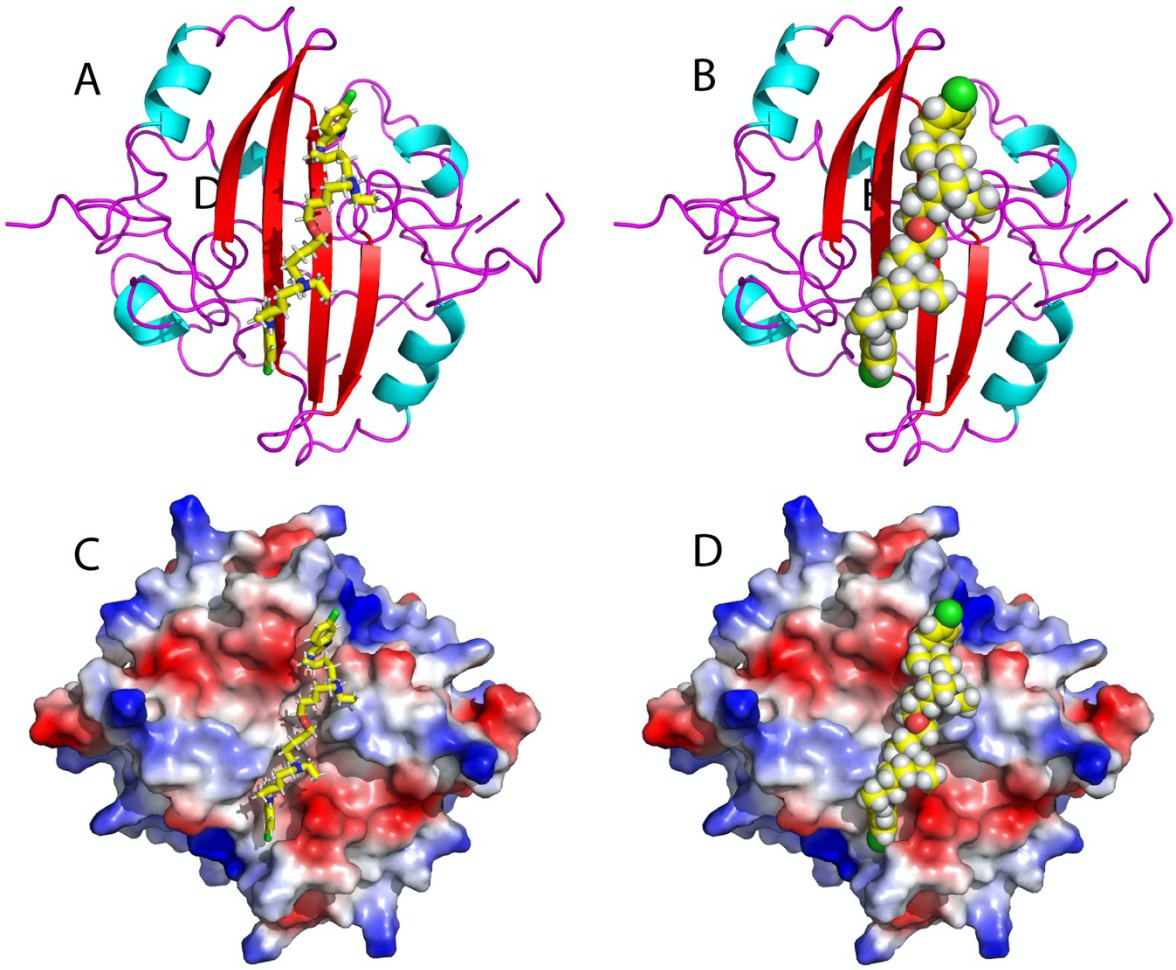

Figure 3. DiHCQ bound to the dimeric CTD of SARS-CoV-2 N protein.
The ribbon representation of the dimeric CTD structure in complex with two HCQ displayed in stick (A) and in sphere (B). The electrostatic potential surface of the dimeric CTD structure in complex with two HCQ displayed in stick (C) and sphere (D).

Molecular dynamics (MD) simulations

To explore the underlying mechanism for HCQ to interfere in the CTD functions as well as to evaluate the design strategy, we conducted 20-ns MD simulations as we previously performed on other protein systems for understanding catalysis, binding and stability (27-33). The obtained structure snapshots are shown for CTD in the free state (Fig. 4A), in complex

with HCQ (Fig. 4B) and with DiHCQ (Fig. 4C). From the results, two major conclusions could be reached: 1) the binding of HCQ appears to increase the dynamics of CTD even on the ns time scale; 2) large structural fluctuations were observed for both HCQ and DiHCQ, implying that further optimization by the integration of combinatorial chemistry, biophysical/NMR characterization and computational evaluation is demanded to enhance their binding affinity.

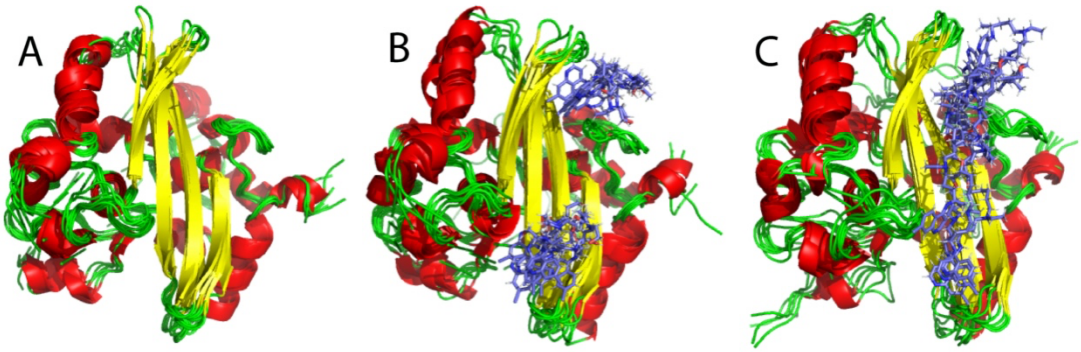

Figure 4. Molecular dynamics (MD) simulations.
Superimposition of 7 structures (one structure for 3 ns interval) of the first set of MD simulation for CTD of SARS-CoV-2 N protein in the unbound state (A), in complex with HCQ (B) and in complex with DiHCQ (C). The HCQ and DiHCQ are displayed in sticks.

Discussion

Although intense efforts have been globally dedicated to developing various vaccines, the rapid emergence of various mutants of the spike protein imposes the extreme challenge to this approach (34). Therefore, effective anti-SARS-CoV-2 drugs might be ultimately essential for terminating the pandemic. In this context, our previous (24) and current efforts provided a promising foundation to develop anti-SARS-CoV-2 drugs from HCQ to target the viral N protein. The high-resolution NMR evidence unambiguously reveals that HCQ could indeed specifically bind both NTD and CTD of SARS-CoV-2 N protein to block their interactions with nucleic acids. Furthermore, HCQ could also dissolve LLPS essential for several key steps of the viral life cycle. Here with NMR-derived constraints we were able to successfully construct the structure of CTD in complex with HCQ, which thus allowed us to propose a strategy to design better anti-SARS-CoV-2 molecules starting from HCQ. By integrating combinatorial chemistry, biophysical/NMR characterization and computational evaluation, bivalent or even multivalent binders for CTD might be designed/optimized to block the viral replication by targeting its N protein.

Materials and methods

Preparation of recombinant SARS-CoV-2 nucleocapsid 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. The DNA fragments encoding its NTD and CTD were subsequently generated by PCR rection and cloned into the same vector.

The recombinant N protein and its NTD/CTD were expression in *E. coli* cells BL21 with IPTG induction at 18 ºC. Both proteins were found to be soluble in the supernatant. For NMR studies, the bacteria were grown in M9 medium with addition of $(^{15}NH_4)_2SO_4$ for ^{15}N-labeling. The recombinant proteins were first purified by Ni^{2+}-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-75 column. The purity of the recombinant proteins was checked by SDS-PAGE gels and NMR assignment for both NTD and CTD. Hydroxychloroquine (HCQ) sulfate was purchased from Merck (HPLC purified, >95%).

LLPS imaged by differential interference contrast (DIC) microscopy

The formation of liquid droplets was imaged on 50 μl of the N protein samples by DIC microscopy (OLYMPUS IX73 Inverted Microscope System with OLYMPUS DP74 Color Camera) as previously described (7,18). The N protein samples were prepared at 20 μM in 25 mM HEPES buffer (pH 7.5) with 70 mM KCl, and HCQ was also dissolved in the same buffer with the final pH adjusted to pH 7.5.

NMR characterizations of the binding of HCQ to NTD and CTD

NMR experiments were conducted at 25 ºC on an 800 MHz Bruker Avance spectrometer equipped with pulse field gradient units and a shielded cryoprobe as described previously (7,18-21,29-31). For NMR HSQC titrations with HCQ, two dimensional ^{1}H-^{15}N NMR HSQC spectra were collected on the ^{15}N-labelled NTD or CTD samples in the absence and in the presence of HCQ at different ratios. The binding-perturbed residues of NTD and CTD structures were displayed by Pymol (The PyMOL Molecular Graphics System).

Calculation of CSD and data fitting

Sequential assignment was achieved and to calculate chemical shift difference (CSD), HSQC spectra collected without and with HCQ at different concentrations were superimposed. Subsequently, the shifted HSQC peaks were identified and further assigned to the corresponding NTD and CTD residues. The chemical shift difference (CSD) was calculated by an integrated index with the following formula (7):

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

In order to obtain residue-specific dissociation constant (Kd), we fitted the shift traces of the 11 residues with significant shifts (CSD > average + STD) by using the one binding site model with the following formula as we previously performed (7, 29-31):

$$CSD_{obs} = CSD_{max}\ \{([P] + [L] + Kd) - [([P] + [L] + Kd)^{2} - 4[P][L]]^{1/2}\}/2[P]$$

Here, [P] and [L] are molar concentrations of RBD and ligands (HCQ) respectively.

Molecular Docking.

NMR-derived constraints were used to guide the docking by HADDOCK software and CNS. The docking was performed in three stages: (1) randomization and rigid body docking; (2) semi-flexible simulated annealing; and (3) flexible explicit solvent refinement.

Acknowledgement

This study is supported by Ministry of Education of Singapore (MOE) Tier 1 Grants R-154-000-B45-114 and R-154-000-B92-114 to Jianxing Song.

Author contributions

Conceived the research: J.S. Performed research and analyzed data: M.D and J.S; Acquired funding: J.S; Wrote manuscript: J.S.

References

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. A new coronavirus associated with human respiratory disease in China. ***Nature***. 2020 Mar;579(7798):265-269.

2. Zhan SH, Deverman BE, Chan YA. SARS-CoV-2 is well adapted for humans. What does this mean for re-emergence? ***BioRxiv***. 2020; doi: https://doi.org/10.1101/2020.05.01.073262.

3. The 'nightmare scenario' for California's coronavirus strain - Los Angeles Times. https://www.latimes.com/science/story/2021-02-24/the-nightmare-scenario-for-californias-coronavirus-strain-here-is-what-we-know

4. R. McBride, M. Van Zyl, B. C. Fielding. (2014) The coronavirus nucleocapsid is a multifunctional protein. ***Viruses,*** 6,2991–3018.

5. M.H. Verheije, M.C. Hagemeijer, M. Ulasli, et al.. (2010) The coronavirus nucleocapsid protein is dynamically associated with the replication-transcription complexes. ***J Virol.*** 84,11575–11579.

6. Chen CY, et al. (2007) Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. ***J Mol Biol.*** 368:1075-86.

7. E.J. Snijder, P.J. Bredenbeek, J.C. Dobbe, et al. (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. ***J Mol Biol.*** 331,991-1004.

8. Dang M , Li Y, Song J. (2021) ATP biphasically modulates LLPS of SARS-CoV-2 nucleocapsid protein and specifically binds its RNA-binding domain ***Biochem Biophys Res Commun.*** 19;541:50-55.

9. S. Lu, Q. Ye, D. Singh, et al. (2020) The SARS-CoV-2 Nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. ***BioRxiv***. doi: 10.1101/2020.07.30.228023

10. A. Savastano, A.I. de Opakua, M. Rankovic, et al. (2020) Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. ***Nat Commun***. 11,6041.

11. T.M. Perdikari, A.C. Murthy, V.H Ryan, et al. (2020) SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. ***EMBO J***. 17;e106478.

12. Carlson CR, Asfaha JB, Ghent CM, Howard CJ, Hartooni N, Safari M, Frankel AD, Morgan DO. (2020) Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. ***Mol Cell.*** 17;80:1092-1103.e4.

13. Dinesh, D.C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., Boura, E. (2020) Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. ***PLoS Pathog*** 16: e1009100-e1009100

14. Zinzula, L., Basquin, J., Bohn, S., Beck, F., Klumpe, S., Pfeifer, G., Nagy, I., Bracher, A., Hartl, F.U., Baumeister, W. (2021) 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*** 538: 54-62

15. Lehninger's Principles of Biochemistry, 5th Edition. W.H. Freeman and Company, NewYork, pp. 502-503 (2005)

16. A. Patel, L. Malinovska, S. Saha, et al. (2017) ATP as a biological hydrotrope, ***Science***, 356,753-756.

17. J. Kang, L. Lim, Y. Lu, et al., (2019) A unified mechanism for LLPS of ALS/FTLD-causing FUS as well as its modulation by ATP and oligonucleic acids, ***PLoS Biol,*** 17,e3000327.

18. J. Kang, L. Lim, J. Song, (2019) ATP binds and inhibits the neurodegeneration-associated fibrillization of the FUS RRM domain, ***Commun Biol,*** 2, 223.

19. M. Dang, J. Kang, L. Lim, et al. (2020) ATP is a cryptic binder of TDP-43 RRM domains to enhance stability and inhibit ALS/AD-associated fibrillation, ***Biochem Biophys Res Commun,*** 522,247-253.

20. Y. He, J. Kang, L. Lim, et al., (2020) ATP binds nucleic-acid-binding domains beyond RRM fold, ***Biochem. Biophys. Res. Commun***. 522, 826-831.

21. Eugenia Quiros Roldan, Giorgio Biasiotto , Paola Magro, Isabella Zanella. (2020) The possible mechanisms of action of 4-aminoquinolines (chloroquine/hydroxychloroquine) against Sars-Cov-2 infection (COVID-19): A role for iron homeostasis? ***Pharmacol Res*** . 158:104904.

22. Sairaj Satarker, Tejas Ahuja, Madhuparna Banerjee, Vignesh Balaji E, Shagun Dogra, Tushar Agarwal, Madhavan Nampoothiri. (2020) Hydroxychloroquine in COVID-19: Potential Mechanism of Action Against SARS-CoV-2 ***Curr Pharmacol Rep***. 24;1-9.

23. Maisonnasse P, Guedj J, Contreras V, et al. Hydroxychloroquine use against SARS-CoV 2 infection in non-human primates. ***Nature***. 2020; 585:584-587.

24. Dang M. Song J. Structural basis of anti-SARS-CoV-2 activity of hydroxychloroquine: specific binding to NTD/CTD and disruption of LLPS of N protein. ***BioRxiv***. 2021. doi: https://doi.org/10.1101/2021.03.16.435741.

25. Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: a protein- protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. 2003, 125, 1731−1737.

26. Qin, H.; Shi, J.; Noberini, R.; Pasquale, E. B.; 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.

27. Lim L, Dang M, Roy A, Kang J, Song J. (2020) Curcumin Allosterically Inhibits the Dengue NS2B-NS3 Protease by Disrupting Its Active Conformation. ***ACS Omega.*** 5:25677.

28. J. Song, F. Ni. (1998) NMR for the design of functional mimetics of protein-protein interactions: one key is in the building of bridges. ***Biochem. Cell Biol.*** 76,177–188.

29. Shi J, Han N, Lim L, Lua S, Sivaraman J. Wang L, Mu Y and Song J. (2011) Dynamically-driven inactivation of the catalytic machinery of the SARS 3C-Like protease by the N214A mutation on the extra domain. ***PLoS Computational Biology***. 7: e1001084.

30. Lim L, Shi J, Mu Y, Song J (2014) Dynamically-driven enhancement of the catalytic machinery of the SARS 3C-like protease by the S284-T285-I286/A mutations on the extra domain.. ***PLoS One***. 9:e101941**.**

31. Wei Z, Song J. (2005) Molecular mechanism underlying the thermal stability and pH-induced unfolding of CHABII. ***J Mol Biol.*** 348:205-18.

32. Lua S, Qin H, Lim L, et al. Structural, stability, dynamic and binding properties of the ALS-causing T46I mutant of the hVAPB MSP domain as revealed by NMR and MD simulations. ***PLoS One.*** 2011;6(11):e27072.

33. Gupta G, Lim L, Song J. NMR and MD Studies Reveal That the Isolated Dengue NS3 Protease Is an Intrinsically Disordered Chymotrypsin Fold Which Absolutely Requests NS2B for Correct Folding and Functional Dynamics. ***PLoS One.*** 2015; 10(8):e0134823.

34. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD, Graham BS, Mascola JR, Chang JY, Yin MT, Sobieszczyk M, Kyratsous CA, Shapiro L, Sheng Z, Huang Y, Ho DD. (2021) Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. ***Nature.*** Mar 8. doi: 10.1038/s41586-021-03398-2.