Structural basis of anti-SARS-CoV-2 activity of hydroxychloroquine: specific binding to NTD/CTD and disruption of LLPS of 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: Structural basis of anti-SARS-CoV-2 activity of HCQ

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

Abstract

SARS-CoV-2 is the coronavirus causing the catastrophic pandemic which already led to >120 millions of infections and >2.6 millions of deaths. Hydroxychloroquine (HCQ) has been shown to own promising potential in clinically combating SARS-CoV-2 but the underlying mechanisms still remain almost unknown. So far, all action sites are proposed on the host cells, and in particular no specific viral target protein has been experimentally identified. In this study, by use of DIC microscopy and NMR spectroscopy, for the first time we have decoded that HCQ specifically binds to 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 package of gRNA and N protein into new virions. These results suggest that HCQ may achieve its anti-SARS-CoV-2 activity by interfering in several key steps of the viral life cycle. The study not only provides a structural basis for the anti-SARS-CoV-2 activity of HCQ, but also indicates that SARS-CoV-2 N protein and its LLPS represent key targets for further optimization and development of anti-SARS-CoV-2 drugs.

Key words: Hydroxychloroquine (HCQ): SARS-CoV-2; Nucleocapsid (N) protein; N-terminal domain (NTD); 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 >120 millions of infections and >2.6 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 and a large number of mutations have been found for its spike protein (2), thus imposing a great challenge to terminate the pandemic by the currently-available vaccines derived from its spike protein.
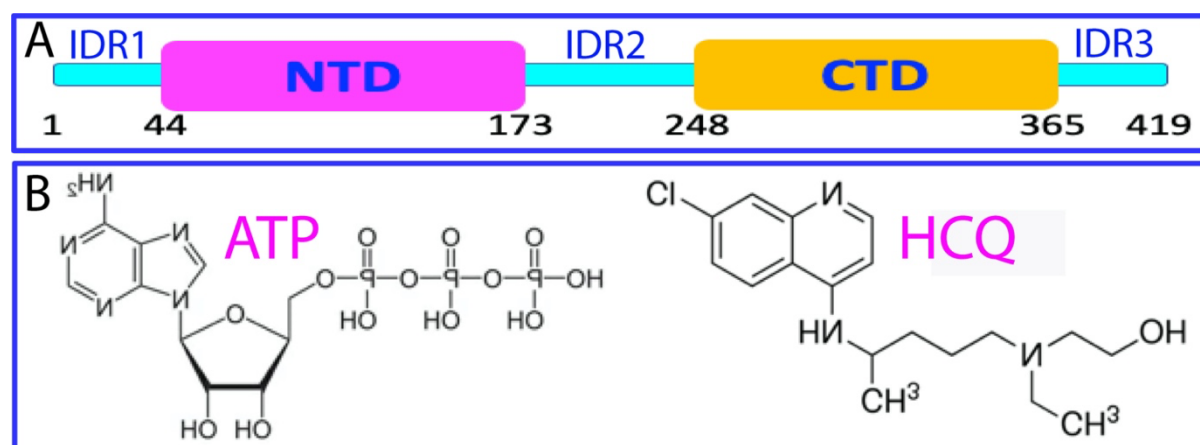
Figure 1. (A) Domain organization of SARS-CoV-2 N protein. (B) Chemical structures of ATP and HCQ.

Intriguingly, hydroxychloroquine (HCQ) has been reported to have promising potential in clinically combating SARS-CoV-2 (3,4), which is an antimalarial drug with higher water solubility and lower toxicity (Fig. 1B). However, the mechanisms for its anti-SARS-CoV-2 activity remain almost unknown and so far, the proposed action sites for HCQ are all 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, to date no specific viral target protein has been experimentally identified for HCQ.

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 (5-11). 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) while CTD functions to dimerize/oligomerize to form high-order structures (5-15). 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 (7-11). In particular, we found that ATP (Fig. 1B), the universal energy currency mysteriously with very high concentrations in all living cells (16-21), could specifically bind the NTD as well as biphasically modulate LLPS of SARS-CoV-2 N protein (7). Inspection of the chemical structures of ATP and HCQ revealed their modular similarity: relatively hydrophobic aromatic rings and polar tails (Fig. 1B). Therefore, by DIC microscopy and NMR spectroscopy, here we characterized the binding of HCQ to both NTD and CTD, as well as the effect on LLPS of SARS-CoV-2 N protein. Very unexpectedly, we decrypted for the first time that HCQ not only specifically binds both NTD and CTD, but also disrupts LLPS of SARS-CoV-2 N protein. Therefore, our 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.

Our results also suggest that SARS-CoV-2 N protein and its LLPS represent critical targets for

future discovery/design of anti-SARS-CoV-2 drugs.

Results

HCQ specifically binds NTD of N protein

Here we first assessed whether HCQ is able to bind NTD of SARS-CoV-2 N protein

which is an RNA-binding domain (14). As shown in Fig. 2A, the ^{15}N-labeled NTD (44-180)

has a well-dispersed HSQC spectrum, indicating that it is well-folded. Subsequently, we

characterized the binding of HCQ to NTD by collecting 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 could be saturated at high HCQ concentrations as evidenced by Fig. 2A,

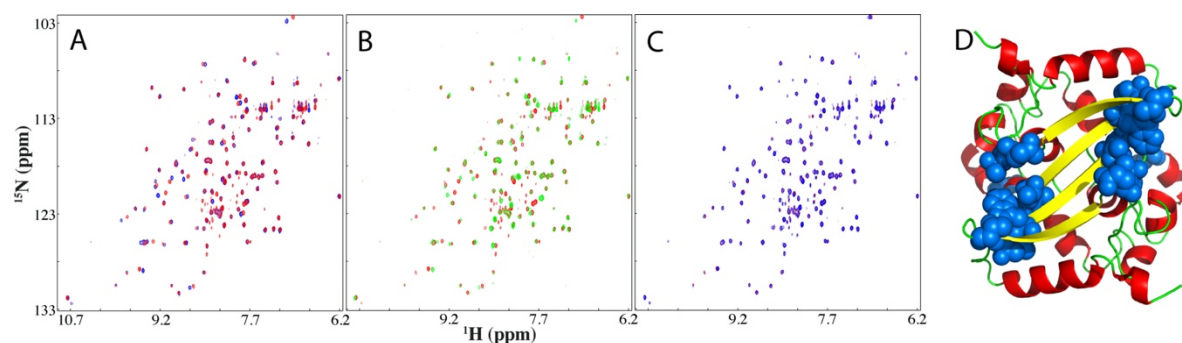thus indicating that HCQ indeed specifically binds NTD.
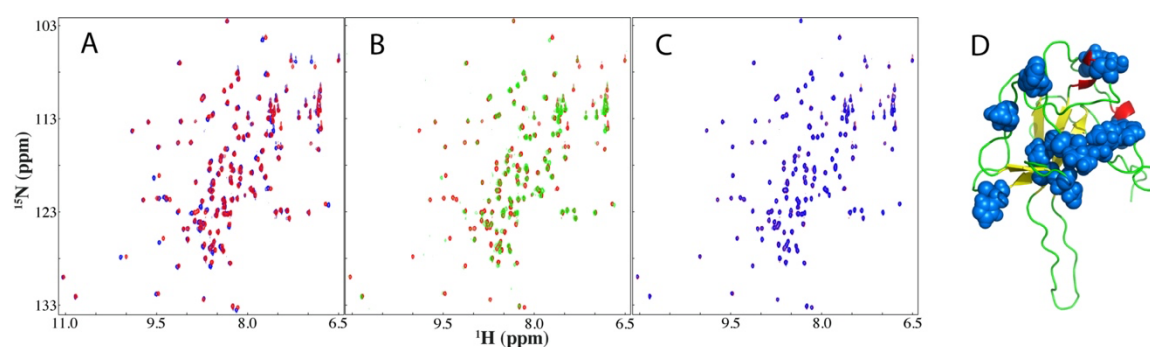
Figure 2. HCQ specifically binds NTD of SARS-CoV-2 N protein. (A) Superimposition of HSQC spectra of NTD in the free state (blue) and in the presence of HCQ at the saturated concentrations (red). (B) Superimposition of HSQC spectra of NTD in the presence of HCQ at the saturated concentrations (red) and S2m-ssDNA at the saturated concentrations (green). (C) Superimposition of HSQC spectra of NTD in the presence of HCQ at the saturated concentrations (red) and with additional addition of S2m at the saturated concentrations (blue). (D) NTD structure with the perturbed residues displayed in spheres upon adding HCQ.

Here we also titrated NTD with S2m, a SARS-CoV-2 ssDNA oligonucleic acid which

was derived from that of SARS-CoV-1 (12). S2m can induce the shift of a small set of HSQC

peaks which is also saturatable. Nevertheless, the shift patterns induced by HCQ and S2m show

a significant difference (Fig. 2B). Very interestingly, additional addition of S2m at the saturated

concentration to the NTD sample with HCQ at its saturated concentration triggered no

significant change of the HSQC spectra (Fig. 2C), implying that the binding of HCQ to NTD

largely blocks its further binding to S2m. Furthermore, we have also achieved the assignments of HSQC spectra and identified the significantly-perturbed residues (Fig. 2D).

HCQ specifically binds CTD of N protein

Subsequently we further assessed whether HCQ can bind CTD of SARS-CoV-2 N protein which is responsible for dimerization/oligomerization (15). As shown in Fig. 3A, the ^{15}N-labeled CTD (248-365) also has a well-dispersed HSQC spectrum, suggesting the dimeric CTD is also well-folded. Subsequently, we characterized the binding of HCQ to NTD 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 also observed which is saturatable at high HCQ concentrations as evidenced by Fig. 3A, thus indicating that HCQ does specifically bind CTD.


Figure 3. HCQ specifically binds 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) CTD structure with the perturbed residues displayed in sphere upon adding HCQ.

We further titrated CTD with S2m and it can also induce 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. 3B). Strikingly, the addition of S2m at the saturated concentration to the CTD sample saturated with HCQ triggered no significant change of the HSQC spectra

(Fig. 3C), implying that the binding of HCQ to CTD also blocks its further binding to S2m. Furthermore, we have successfully assigned HSQC spectra and identified the significantly-perturbed residues as mapped back to the dimeric CTD structure (Fig. 2D).

HCQ dissolves LLPS of N protein

N protein of SARS-CoV-2 at 20 μM has a turbidity value of 0.03 (Fig. 4A) and showed no droplets as imaged by DIC. Addition of HCQ up to 1:750 failed to induce LLPS of N protein as monitored by turbidity (Fig. 4A) and imaged by DIC. Nevertheless, upon adding ssDNA at 1:0.75, many liquid droplets were formed (Fig. 4B) with a large turbidity value (Fig. 4A). Interestingly, additional addition of HCQ into this sample led to dissolution of the droplets, as reflected by the decrease of turbidity (Fig. 4A) and reduction of the droplet numbers and sizes (Fig. 4B). At 1:250 (N-protein:HCQ), the turbidity was reduced to be only 0.125 (Fig. 4A) while most droplets were dissolved (Fig. 4B). Intriguingly, further addition of HCQ appeared to trigger aggregation of the sample.
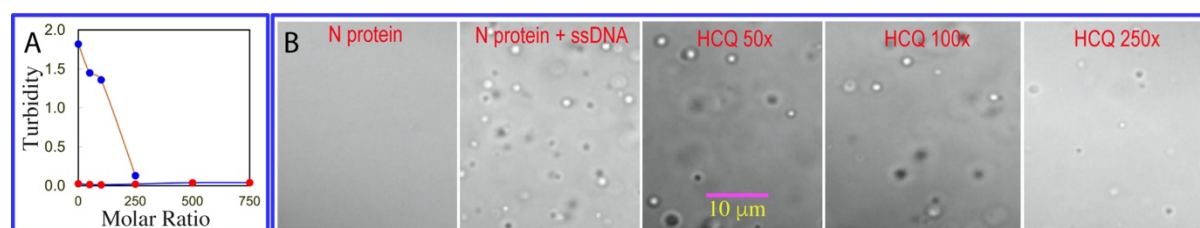
Figure 4. HCQ dissolves NA-induced LLPS of SARS-CoV-2 N protein. (A) Turbidity curves of N protein without (blue) and with the pre-existence of ssDNA at 1:0.75 (red) upon addition of HCQ at different ratios. (B) DIC images of N protein with the pre-existence of ssDNA at 1:0.75 upon additional addition of HCQ at different ratios.

The results here indicate that fundamentally different from ATP that can biphasically modulate LLPS of N protein: namely induction of LLPS at low concentrations but dissolution at high concentrations (7), HCQ has no capacity in inducing LLPS likely because its polar tail is unable to bind Arg residues as the triphosphate chain of ATP does (7,18-21). Nevertheless, HCQ has a strong capacity in dissolving LLPS of N protein induced by oligonucleic acid, most

likely by blocking the interactions of its NTD and CTD with oligonucleic acid which may be

essential for driving LLPS of SARS-CoV-2 N protein (7-11).

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 (22). On the other hand, in principle any step of the viral cycle can be utilized for developing therapeutic interference. For example, previously for SARS-CoV-1, we have been focused on the structurally- and dynamically-driven allosteric mechanisms of the 3C-like protease (23-28). Recently, several groups including us discovered that LLPS of SARS-CoV-2 N protein and its modulation appear to play key roles in the viral life cycle (Fig. 5). Briefly, the dissolution of phase separated condensates of gRNA and N protein appears to be essential for uncoating gRNA, while LLPS is required at least for forming the replicase-transcriptase complexes, as well as the final package of gRNA and N protein into new virions. In particular, our results imply that ATP appears to be hijacked by SARS-CoV-2 to promote its life cycle (Fig. 5) (7).
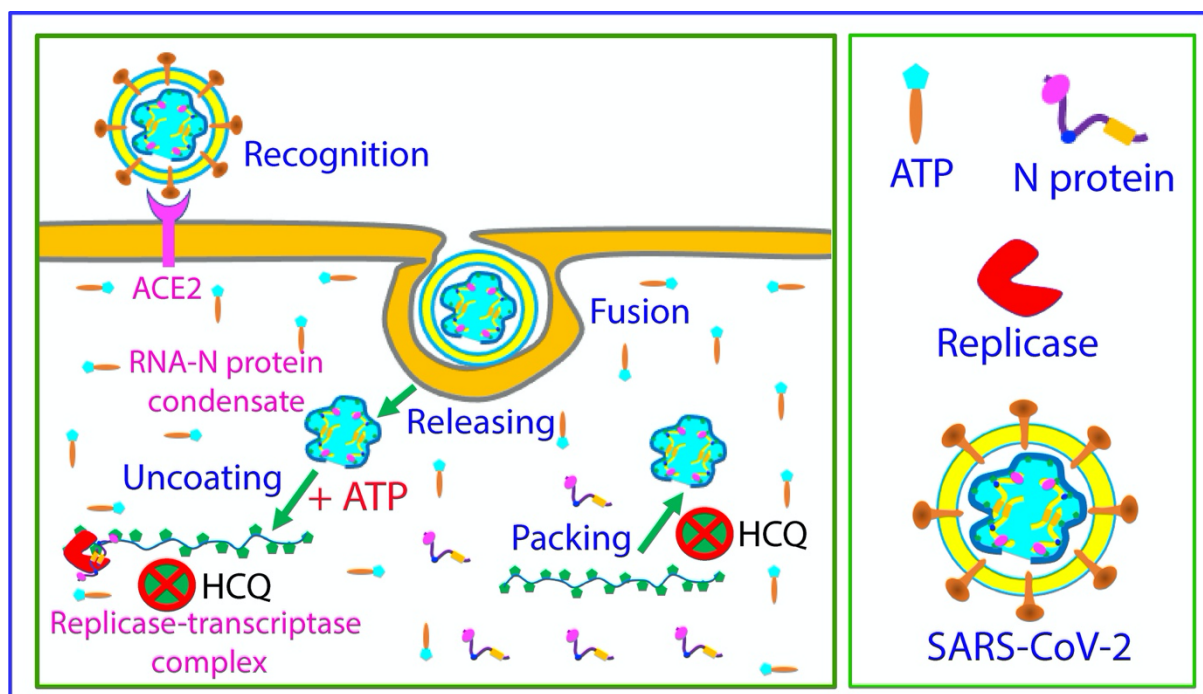
Figure 5. Key roles of LLPS of SARS-CoV-2 N protein and its modulation by ATP.
A proposed scheme to illustrate that ATP appears to be hijacked by SARS-CoV-2 to prompt its life cycle including the initial uncoating of the gRNA-N-protein condensate, subsequent localizing to forming replicase-transcriptase complex, and final package of gRNA and N

protein. In this context, HCQ may manifest the anti-SARS-CoV-2 activity by disrupting required LLPS at least for two key steps: the formation of replicase-transcriptase complexes, as well as the final package of gRNA and N protein into new virions.

Most importantly, our results here provide the first high-resolution NMR evidence that HCQ can in fact specifically bind both NTD and CTD of SARS-CoV-2 N protein to block their interactions with nucleic acids, thus providing a potential mechanism for the anti-SARS-CoV-2 activity of HCQ by targeting the viral protein. Indeed, previously it was found that HCQ was able to block the maturation of SARS-CoV-2 virions but the underlying mechanism was proposed to result from the HCQ-induced change of host cell structures/conditions (3,4). Our current results suggest that the HCQ-induced disruption of LLPS of N protein and gRNA may at least partly account for the inhibition of the maturation. Our study also indicates that the essential LLPS of SARS-CoV-2 N protein is druggable as well as NTD and CTD of its N protein represent key drug targets. Future efforts with the integrated computational and experimental approaches may lead to establishing the promising strategy for optimization or/and design of anti-SARS-CoV-2 drugs (32,33).

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 °C. 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):

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

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 (ATP) respectively.

Acknowledgement

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. 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

3. 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.

4. 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.

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

6. 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.

7. 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.

8. 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

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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

15. 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

16

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

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

18. 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.

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

20. 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.

21. 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.

22. 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.

23. Shi J, Wei Z and Song J. (2005) Dissection Study on the Severe Acute Respiratory Syndrome 3C- like Protease Reveals the Critical Role of the Extra Domain in Dimerization of the Enzyme: Defining the extra domain as a new target for design of highly specific protease inhibitors. *J. Biol. Chem.* 279: 24765-24773.

24. Shi J, Song J. (2008) The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *FEBS J.* 2006 Mar;273(5):1035-45.

25. Shi J, Sivaraman J, Song J. (2008) Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. *J Virol.* 82:4620-9.

26. 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.

27. 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.

28. Lim L, Gupta G, Roy A, Kang J, Srivastava S, Shi J, Song J. (2019) Structurally- and dynamically-driven allostery of the chymotrypsin-like proteases of SARS, Dengue and Zika viruses. *Prog Biophys Mol Biol.* 143:52-66.

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

30. Qin H, Lim LZ, Wei Y, Song J. (2014) TDP-43 N terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA. *Proc Natl Acad Sci U S A.* 111:18619-24.

31. Qin H, Shi J, Noberini R, Pasquale EB, Song J. (2008) 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.* 283: 29473

32. 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.

33. 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