UNIVERSITY OF CALIFORNIA
RIVERSIDE

Structure-Function Investigation of SARS-CoV-2 Viral Proteins

A Dissertation submitted in partial satisfactionof
the requirements for the degree of

Doctor of Philosophy

in

Biochemistry and Molecular Biology

by

Mahamaya Nimain Biswal

June 2022

Dissertation Committee:
Dr. Jikui Song, Chairperson
Dr. Rong Hai
Dr. Sean O'Leary

The Dissertation of Mahamaya Nimain Biswal is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

First and Foremost my sincerest thanks to my graduate advisor, Dr. Jikui Song. I would like to thank him for his continuous support and guidance throughout. His passion towards science always motivated me. I greatly appreciate him for being patient enough to correct me whenever I made a mistake and kind enough to give opportunity to learn new techniques and to make sure my projects followed the right direction.

I am also grateful to my committee members Dr.Rong Hai and Dr. Sean O' Leary for their valuable advice and feedback for my projects during our meetings.

I would like to thank our collaborators Dr. Rong Hai and his Postdoc Dr. Duo Xu and Dr. Gregor Blaha and his graduate student Stephen Diggs for their efforts that contributed to our work.

I would like to thank my past and current lab members for their support and encouragement. I would like to specially thank Dr. Jian Fang, Dr. Jiuwei Lu, and Nelli Khudaverdyan for all their help and being available for valuable discussions throughout this journey of mine.

I would also like to thank the Biochemistry and Molecular Biology Program for their support in the form of Prize money and Fellowships which really supported me and motivated me.

Last but not the least, I would like to thank my Parents and Sister for always believing in me. They are my cheer leaders without whom I wouldn't have reached here and accomplished what I have up till now. Special

thanks to my Husband, words would fall short to express but he has supported me and motivated me through all the ups and downs and has stood by me like a pillar of strength throughout this journey. I would also like to thank the rest of my family members and friends back in India and here in the states to believe in me and support me all along.

ABSTRACT OF THE DISSERTATION

Structure-Function Investigation of SARS-CoV-2 Viral Proteins

by

Mahamaya Nimain Biswal

Doctor of Philosophy, Graduate Program in Biochemistry and Molecular Biology

University of California, Riverside, June 2022
Dr. Jikui Song, Chairperson

March 2022 marks two years since the Covid-19 Pandemic began. Two years on, battle with the virus still continues. Officially the number of lives lost to the virus is reported to be around 6 Million but a recent study highlights that this number could be a huge undercount and the true estimate may account to three times that of the reported value. With the virus still circulating worldwide, it holds the opportunity to continue evolving and emerge as a variant capable of evading the human immune response. Therefore, new drugs will be needed apart from the approved ones to counter the looming threat of resistance. For doing so, deeper understanding of the mechanistic basis of how the different viral proteins function is needed. In this study, we unravel structural basis of NSP 7-8 complex formation and investigate its dynamic assembly to form an active NSP 7-8-12 core replication complex. Structure-guided mutagenesis, combined with biochemical and enzymatic assays, further reveals a structural coupling between the two oligomer interfaces of NSP 7-8, as

well as the importance of these interfaces for the RdRP activity of the NSP7-NSP8-NSP12 complex. In another project, we investigated the pathogen-host interaction based on SARS-CoV-2 Nucleocapsid protein and human G3BP1 protein. This aspect of our study reveals the biochemical and structural basis of the SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$ interaction, revealing a "ϕ-x-F" motif to be the primary and indispensable determinant of the interaction and flanking residues underpins diverse secondary interactions. Mutation of the key interaction residues of the SARS-CoV-2 $N_{1–25}$ -$G3BP1_{NTF2}$ complex leads to disruption of the SARS-CoV-2 N-G3BP1 interaction *in vitro*. In all, our results provide a molecular basis of the strain- specific interaction between SARSCoV-2 N and G3BP1, which has important implications for the development of novel therapeutic strategies against SARS-CoV-2 infection.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER 1

INTRODUCTION

1.1 Covid-19 pandemic caused by SARS-CoV-2 has become global health concern.

The causative agent of the Covid-19 outbreak is Severe Acute Respiratory Syndrome- Coronavirus- 2 (SARS-CoV-2). Coronaviruses consist of a diverse family of single- stranded RNA viruses and are known to cause respiratory and intestinal infections in animals and humans. They were not considered highly pathogenic to humans until the 'SARS-CoV' epidemic broke out in Guangdong Province, China in 2002-2003. Ten years later, in 2012, another highly pathogenic coronavirus called 'MERS-CoV' emerged in Middle Eastern countries[2, 3, 7]. The SARS-CoV-2 pandemic which broke out in the year 2019 has brought unprecedented challenges to the world, in terms of both health and economic spectra. The official global death toll from the pandemic is around 6 million (https://coronavirus.jhu.edu/map.html), and recent articles say the actual number could be as high as three times to this number due to unreported cases[1]. In addition to available vaccines, effective treatments could pave the way for putting a gradual end to the pandemic. For this, deeper understanding and knowledge of viral life cycle and the different proteins that help in the functioning of the virus is essential.

1.2 Coronavirus diversity and origination

Coronaviruses belong to the family of Coronaviridae, which consists of the subfamily Coronavirinae and the order Nidovirales (International Committee on Virus Taxonomy). This subfamily consists of four genera - Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus - based on their phylogenetic relationship and genome structure. Members of the Alphacoronavirus and betacoronavirus are known to cause infection only in mammals. Gammacoronaviruses and deltacoronaviruses have broader host ranges which includes birds[2].

Once the causative agent of 2002-2003 epidemic was identified as SARS-CoV, antibodies to SARS-CoV were found in commercially available civet cats (Paguma larvata) and in commercially available pet breeders. Further investigation revealed that the civet cats only acted as intermediate hosts, and that the SARS-CoV strain found in them had been transmitted to them from other animals. In 2005, two independent teams reported the discovery of a novel coronavirus related to human SARS-CoV, named it ”SARS-CoV related viruses‟ in horshoe bats (genus Rhinolophus). This discovery suggested that bats could be the primary hosts of SARS CoV and civets acts as intermediate host[2].

1.3 CoV Life cycle

CoV’s have one of the largest RNA viral genome ranging from 26 to 32 kilobases inlength. It contains a 5‘ terminal cap structure and its 3‘ end

is polyadenylated. Initial steps of the coronavirus infection involves specific binding of coronavirus spike protein to the cellular entry receptors that have been identified for several coronaviruses which include human amino peptidase N (APN; HCoV-229E), dipeptidyl peptidase 4 (DPP4; MERS-CoV) and angiotensin-converting enzyme 2 (ACE2; HCoV-NL63, SARS-CoV and SARS-CoV-2) [3,4]. CoV-2 virion binds to and makes specific interactions with cellular entry receptor called ACE2 via its Spike protein; this step is also mediated by some host factors such as the cell surface serine protease "Transmembrane Serine Protease 2" (TMPRSS2), which promotes the viral uptake and fusion at the cellular or endosomal membrane. Following entry, the genomic RNA undergoes uncoating and release and is later subjected to immediate translation of the two large open reading frames (ORF's), ORF1a and ORF1b[2]. The resulting polyprotein pp1a and pp1ab are co-translationally and post-translationally processed into the individual non-structural proteins (NSP's) that form the viral replication and transcription complex (RTC). Alongside the expression of NSP's, there is biogenesis of viral replication organelles taking place that consists of characteristic perinuclear double-membrane vesicles (DMV) and convoluted membranes (CM) that helps to create a protective microenvironment for viral genomic RNA replication and transcription of subgenomic mRNA (sg mRNA) made up of the characteristic nested set of coronavirus mRNA's[2]. ORFs that encode structural proteins and interspersed ORFs that encode accessory proteins are transcribed from the 3′ one-third of the genome to form a nested set of

sg mRNAs. Coronavirus accessory proteins are highly variable sets of virus-specific proteins that display limited conservation even within individual species but they are principally thought to contribute to modulating host responses to infection and are determinants of viral pathogenicity[2, 3]. Nevertheless, the molecular functions of many accessory proteins remain largely unknown owing to the lack of homologies to accessory proteins of other coronaviruses or to other known proteins. The translated structural proteins then translocate into endoplasmic reticulum (ER) membranes and transit through the ER-to-Golgi intermediate compartment (ERGIC), where the encapsidated Nucleocapsid (N) protein interacts with the newly produced genomic RNA and results in budding into the lumen of secretory vesicular compartments. Finally, virions are secreted from the infected cell by exocytosis (Fig 1- 1)[3, 4].

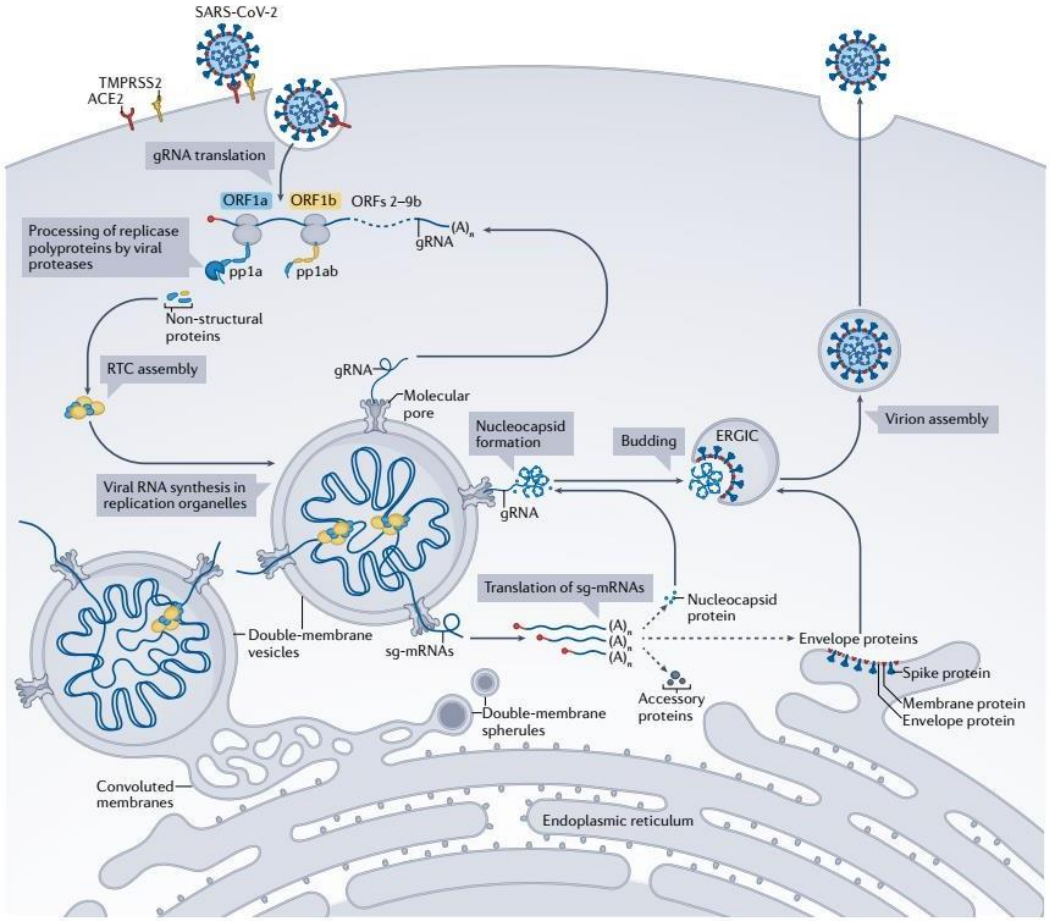

Fig 1-1: SARS CoV-2 infection cycle[3].

1.4 RNA synthesis in Coronaviruses

RNA synthesis in SARS-CoV-2 can be briefly divided into two phases: genome replication and sg mRNA transcription. gRNA translation and proteolytic maturation of replicase proteins is followed by initiation of a complex process of SARS- CoV-2 RNA synthesis and transcription. This depends on interplay between two teams: first, viral RNA and non-structural proteins and second, between host cell proteins and membrane[3]. Specific recognition of RNA template by CoV RTC complex for RNA synthesis

requires the presence of specific RNA sequences and structural elements on the terminal regions of the CoV genome[3, 4]. Replication results in full length viral positive stranded gRNA which can then be translated into replicase polyprotein and could serve as template for synthesis of additional negative stranded RNA or can be packaged into progeny virions. Both Replication and transcription require devoted minus stranded RNA. Full length negative strand RNA can serve as template to make gRNA during replication and nested set of minus strand sg RNA's serve as template for transcription[3,4].

1.5 Coronavirus functions using its arsenal of different proteins: Structural, Non-structural and accessory proteins.

Once the ORF 1a and ORF 1b are translated, it gives rise to two polyprotein called pp1a and 1b. The synthesis of the latter polyprotein requires -1 ribosomal frameshifting near the 3' end of ORF 1a. The two polyproteins are subsequently cleaved by combined effect of host and viral proteases which ultimately leads to the synthesis of sixteen NSP's (NSP 1 to NSP 16)[3,4]. Different NSP's have different dedicated roles in the viral life cycle. NSP1 is used by the virus to inhibit the host gene expression and evade the host immune system and therefore a target for vaccine development. The exact role of NSP2 is not clear but it can interact with NSP3 and cleave ORF 1ab. NSP 3 and NSP 4 are known to interact with other cofactors in order to induce membrane rearrangement for the viral replication and the loss of NSP 3- NSP 4 complex has been shown to

eliminate viral replication. NSP 5 is a cystine like protease, 3CL-PRO which processes 11 cleavage sites between NSP 4 and 16 during replication and also has a conserved domain structure and catalytic residues. NSP 6 generates autophagosomes from the endoplasmic reticulum and is involved in autophagy[4].

NSP7-NSP8-NSP12 (NSP7-8-12) forms the holo RNA-dependent RNA polymerase (RdRp) complex, which is a central component of the replication and transcription machinery. NSP12 belongs to a family of primer-dependent polymerases that catalyze the addition of ribonucleoside triphosphate (rNTP's) to the nascent product RNA as directed by the sequence of the template RNA[11, 14, 15]. NSP12 consists of three domains: a nidovirus RdRp-associated nucleotidyltransferase domain (NiRAN) domain (residues 1-250), an interface region between the NiRAN domain and the RdRp domain (residues 251-398) and the core RdRp domain (residues 399-932)[11, 14]. The NiRAN domain is absent in viruses other than Nidovirales and is an important drug target due to its essential enzymatic activity. Replication and transcription in coronaviruses are thought to be carried out by distinct subcomplexes including holo-RdRP associated with different NSP's as well as accessory proteins. To replicate the genome faithfully, other NSP"s participate in the process. This proofreading activity is performed by NSP14 which is exclusive to Nidovirales'. In NSP 14, this proofreading activity is encoded in the N- terminal exonuclease (ExoN) domain of nsp14, which together with the cofactor nsp10 forms an RNA proofreading complex known to promote faithful replication of the fairly

large genome in the CoV's[38]. The C-terminus of NSP14 has methyltransferase activity. The NSP14-10 complex is responsible for the addition of Cap 0, and the NSP 16-10 complex is responsible for the addition of Cap1 to the newly replicated positive strand RNA as a result of which, the innate immune response to the virus is suppressed.

Components of the core replication machinery, especially NSP 12, and its important cofactor proteins NSP 7 and NSP 8, remain important drug targets for rational drug design. Biochemical experiments by different groups have shown that primer extension ability of SARS-CoV and SARS CoV-2 RdRp (nsp12) is greatly enhanced in the presence of nsp7 and nsp8, corroborating their role as essential cofactors of the RdRp. Recent advancements in cryo-electron microscopy led to the structures of the SARS CoV Apo-holo-RdRp complex and also RNA bound RdRp bound to expanded replication transcription machinery for e.g. NSP 9, NSP13 and NSP14. Crystal structures of the SARS –CoV NSP 7-8[17] and SARS-CoV-2 NSP 7-8[18] complexes have been solved which reports different assembly forms. The structures of SARS-CoV NSP 7-8-12 complex, FCoV NSP 7-8 complex [20] as well as the SARS- CoV-2 NSP 7-8-12 complex along with or without RNA [12-15] have also been solved. Varying arrangement of NSP 7 and NSP 8 proteins are seen in these structures. In the structure representing the hexadecameric form of NSP 7-8 complex, both NSP7 and NSP8 proteins carry a non-native GPLGS tag. The central channel of the hexadecameric structure of SARS CoV NSP 7-8 was proposed to serve an RNA binding site which may mediate

the potential primase activity of this complex[17]. Intriguingly, much recent biochemical and Mass spectrometry evidences pointed out that NSP 7-8 complex exists as a tetramer instead of Hexadecamer[18, 36]. SARS CoV NSP 7-8 fusion protein (NSP7L8) failed to give rise to extended RNA replication product in a denovo RNA synthesis assay done by Subissi etal, posing question on primase activity of NSP 7-8[10]. **In Chapter 2**, we present our investigation done on SARS-CoV-2 NSP 7-8 complex to verify its solution assembly state and further biochemical analysis to understand the dynamic assembly of SARS-CoV-2 NSP 7-8 complex and mechanistic basis of its dynamic transition to form NSP 7-8-12 complex.

1.6. Virus-Host interactions in SARS-CoV-2: Interaction between SARS-CoV-2 Nucleocapsid and human protein G3BP1

Viruses are known to rely on the host cell machinery in order to replicate. Characterization of the factors and the mechanism involved in such pathogen-host interactions will inform our understanding to identify druggable host factors. Host mutation events are rare and aren't selected for viral survival, targeting such host factors limits the ability of the virus to develop 'escape mutations', an event that can appear when viral proteins are targeted by a drug and the viral protein mutates under the drug's selective pressure ultimately developing drug resistance. Also in many cases, a particular host protein is targeted by various different virus families. Therapies designed against such host factors could present as

broad-spectrum anti-virals. Different SARS- CoV-2 proteins are now known to make interactions with different host proteins [31]. The SARS- CoV-2 Nucleocapsid protein (N) is also known to be involved in interacting with different host proteins and exploit the host cell repertoire to provide for its own replication. The N protein has a multifaceted role in aiding the viral replication, virion assembly and also modulating the host cellular activities [40-42]. SARS-CoV-2 N consists of an N-terminal domain (NTD) responsible for RNA binding and a C terminal domain (CTD) that mediates both RNA binding and dimerization[69- 71]. Both the NTD and CTD domains are flanked by Intrinsic disordered regions (IDRs) to reinforce their RNA-binding activities, with a serine-arginine (SR) rich segment located within the second IDR (IDR2) subjected to posttranslational modification for modulation of RNA binding[73, 74].

Recently, multiple studies reported that SARS-CoV-2 Nucleocapsid (N) protein could interact with G3BP1 [44] and play a role in disassembling the stress granules (SG"s) [45]. It does so by interacting with the Stress granule nucleating proteins: ras GTPase activating protein binding protein 1 and 2 (G3BP1 and G3BP2), collectively referred to as G3BP[44].Stress granules (SG's) are dynamic entities and are membraneless organelles which are micron- sized and are transient. Driving factors of stress granule formation are a) Stress granule nucleating proteins- either due to their aggregation because of overexpression or in response to heat shock and osmotic stresses or viral infection b) RNA- due to sudden influx of non-polysomal mRNA as a result of stress induced stalling of pre-initiation complex[58].

Overexpression of some SG proteins leads to nucleation of stress granules in spite of absence of stress or effect due to any drug. Presence of Prion-like domain or intrinsically disordered regions (IDR) or Low complexity regions (LCR) in a protein mediate protein aggregation. Different SG nucleating proteins have different domains that are drivers of nucleation. For e.g. SG protein TIA1 requires both Prion-like domain as well as RNA binding domain which mediate formation of dynamic and reversible SG's. However G3BP1 lacks a Prion-like domain and needs Oligomerization domain: Nuclear transport factor 2 (NTF2) and RNA binding domains (RRM/RGG) to mediate SG assembly[58]. G3BP interacts with many cellular proteins, some of which are SG nucleators: for e.g. Cell cycle-associated protein 1 (Caprin1). G3BP mediated SG assembly is regulated by mutually exclusive binding of Caprin1 and Ubiquitin Specific Peptidase (USP10). Caprin1 binding to G3BP1 promotes SG assembly whereas USP10 binding to G3BP1 inhibits the SG assembly[58, 67].

Another high affinity binding protein to G3BP1 which is also an SG nucleator is Ubiquitin associated Protein 2-like (UBAP2L). UBAP2L acts upstream to G3BP1 during SG assembly. Although UBAP2L forms SG cores along with G3BP1, it does not always co- localize with G3BP1 indicating that UBAP2L can form cores independent of G3BP cores. UBAP2L interacts with G3BP1 through its Domain of Unknown Function (DUF) domain[68]. Deletion of this DUF domain results in breaking the interaction between UBAP2L and G3BP1 and compromising SG formation[88]. Multiple studies have reported that it is this network of these

protein-protein and protein-RNA interactions that regulate the multiphase condensation by fine tuning the strength of interaction between nodes, bridges, caps which allows for substrate dependent “Phase switches" of SG nucleators[48,52,56,57].

Since stress granules pose as an obstacle for viral replication, viruses have evolved to come up with a strategy to attack these stress granules by having specific viral protein that interacts with G3BP1/2. In case of SARS CoV-2, numerous reports have proven the interaction between its Nucleocapsid (N) protein and G3BP1 and that this interaction leads to SG disassembly. The importance of G3BP1 $_{NTF2}$ as an interaction hub is highlighted by studies showing that Non-structural protein 3 (nsP3) from old world alphavirus and Chikungunya virus binds to G3BP1 NTF2L [63, 64]. The old-world alphaviruses target NTF2 in order to recruit G3BP and its associated ribosomal subunits to the viral cytopathic vacuoles (CPV’s). In order to do so, Semliki Forest virus and Chikungunya virus have a dedicated protein called nsP3 which is made up of a duplicated FGDF motif in order to outcompete USP10-G3BP1 $_{NTF2}$ binding existing in cell in the absence of any stress. The nsP3 protein is a protein having multiple domains, the double FGDF motifs is part of a hyper variable domain which is largely disordered and makes multivalent protein interactions [63,64]. The crystal structure of the G3BP1 $_{NTF2}$ domain in complex with nsP3 peptide derived from Semliki Forest virus (SFV) nsP3 or Caprin1 has been reported[64, 67]. Common to these structures is an insertion of a phenylalanine residue from the G3BP1 $_{NTF2}$-interacting peptide into a

hydrophobic pocket formed by the G3BP1 $_{NTF2}$ domain. On the other hand, these structures show large divergence for the Phenylalanine-flanking regions of the peptide sequence

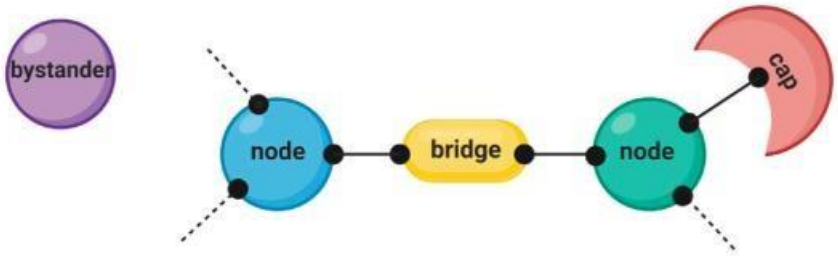

Figure 1-2: Model depicting a SG protein interaction network:

, suggesting a high diversity for the G3BP1 $_{NTF2}$-mediated protein interactions. Recent biochemical and cellular evidence indicated that the interaction between SARS-CoV- 2 N protein and G3BP1 $_{NTF2}$ critically depends on the N-terminal Intrinsic disordered region 1 (IDR1) of SARS-CoV-2 N protein[8,28,37,38]. Deletion of the first 50 residues of SARS-CoV-2 N led to disruption of SG inhibition and reduced viral production, suggesting an important role of the IDR1 of SARS-CoV-2 N in viral infection. Consistently, introduction of an I15A/T16A/ F17A/G18A quadruple mutation to the IDR1 of SARS-CoV-2 N led to abolished G3BP1 binding in an in vitro binding assay, indicating an ITFG motif dependent interaction[76]. However, due to the lack of structural information, the molecular basis for the interaction between SARS-CoV-2 N and G3BP1 remains elusive. In **Chapter 3**, we show our investigation done to determine the molecular basis of the SARS-CoV-2 N - G3BP1 interaction using

crystallography and biochemical analysis. Our results may have important implications in development of therapeutic interventions against SARS-CoV-2 infection.

Finally in **Chapter 4**, we summarize the understanding we hold related to SARS-CoV-2 NSP 7-8 complex and SARS- CoV-2 N-G3BP1 interaction, up till now and discuss the future perspectives related to the same whose further investigation may help unravel new insights into the functioning of these proteins.

CHAPTER 2

Structural and biochemical insights into SARS-CoV-2 NSP 7-8 complex andits dynamic transition to interact with SARS-CoV-2 RdRp

Specific aim: To verify its solution assembly state and further biochemical analysis to understand the dynamic assembly of NSP 7-8 complex and mechanistic basis of dynamic transition into the formation of NSP 7-8-12 complex from NSP 7-8 complex

METHODS

Cloning, expression and purification of SARS-CoV-2 proteins

The DNA fragments encoding SARS-CoV-2 NSP7, NSP8, and NSP12 were chemically synthesized by Integrated DNA Technologies, with codon optimized for bacterial expression. For structural study, the genes for full length NSP7and NSP8 were inserted in tandem into a modified pRSF Duet-1 vector, in which the NSP7 gene was preceded by an N-terminal His6-SUMO tag and ULP1 (ubiquitin like protease) cleavage site and the NSP8 gene was flanked by the NdeI and XhoI cleavage sites. Next, the plasmids were transformed into BL21 (DE3) RIL cell strain (Agilent Technologies). The transformed cells were first grown at 37 ◦C until OD600 reached 0.8. The temperature was then shifted to 16 ◦C, followed by

addition of 0.1 mM isopropyl -D-galactoside for induction. After another 18 h of cell growth, the cells were harvested and the His6- SUMO tagged NSP7 was co-purified with NSP8 using a Ni-NTA affinity column. The NSP7-NSP8 complex was then treated with ULP1 protease to remove the His6-SUMO tag and subjected to further purification by ion-exchange chromatography on a HiTrap Q HP Sepharose column and size exclusion chromatography on a HiLoad 16⁄600 Superdex 75 pg column (GE Healthcare) pre-equilibrated with 25 mM HEPES (pH 7.5), 150 mM NaCl, 5% Glycerol, and 5 mM DTT. The purified NSP7-NSP8 complex was confirmed by SDS-PAGE, concentrated to ∼10 mg⁄ml, and stored at -80 ◦C for further use. For biochemical analysis, the genes for NSP7 and NSP8 were also individually cloned into the pRSF Duet-1 vector, and the gene for NSP12 was cloned into a modified pVP13 vector, N terminally fused to a His6-MBP tag and a TEV cleavage site. The individual WT and mutant NSP7 and NSP8 proteins were purified in the same manneras described for the NSP7-NSP8 complex. For the RdRP assay, NSP12 protein was purified sequentially through Ni-NTA chromatography, ion-exchange chromatography on a Q HP column (GE Healthcare), tag removal via TEV cleavage, and size exclusion chromatography on a Hi Load 16⁄600 Superdex 200 pg column (GE Healthcare) pre- equilibrated with 25 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol, and 5 mM DTT. For analytical gel filtration analysis of the NSP7–NSP8–NSP12 complex, His6-SUMO-NSP7, His6-SUMO-NSP8 and His6-MBP–NSP12 were co-expressed in BL21 (DE3) RIL cells, and co-purified using a Ni-NTA column, followed by

size-exclusion chromatography on a Superdex200 increase 10/300 gl column (GE Healthcare). The mutations of NSP7 and of NSP8 were introduced through site directed mutagenesis and purified in the same manner as the wild-type proteins. For enzymatic comparison of NSP12 derived from codon-optimized and non-codon optimized gene sequences, the native NSP12-encoding DNA sequence (SARS-CoV-2 isolate: Wuhan-Hu-1/2020, NC 045512) was also inserted into the in-house His6-MBP vector. Expression and purification of the NSP12 protein derived from the native gene sequence followed the same procedure as that for the NSP12 protein derived from the codon-optimized gene sequence, as described above.

Crystallization and X-ray data collection

The crystallization condition for the SARS-CoV-2 NSP7– NSP8 complex was initially identified through sparse matrix screens (Hampton Research Inc.). The crystals were thereafter reproduced by hanging drop vapor diffusion method at 4 ◦C by using 1 l of 5 mg/mL SARS-CoV-2 NSP7-NSP8 complex and 1 L of precipitant solution (0.2 M MgCl2,

0.1 M HEPES pH 7.5, 25% [w/v] Polyethylene glycol 3350). SDS-PAGE analysis of the crystals indicated that the NSP8 protein is dominated by a truncated form in crystals. For crystal harvesting, crystals were soaked in well solution supplemented with 25% glycerol before flash freezing in liquid nitrogen. The X-ray diffraction data for the SARS-CoV-2 NSP7–NSP8

complex were collected on the Beamline 5.0.2 at the Advanced Light Source, Lawrence Berkeley National Laboratory. The diffraction data were indexed, refined, and scaled using the HKL 3000 program[22]. The structure was solved by molecular replacement using the partial structure of SARS-CoV NSP7–NSP8 complex (PDB: 2AHM) as a search model. The resulting electron density revealed two molecules each for NSP7 and NSP8 in the asymmetric unit cell. The structure was further improved by iterative rounds of model building and refinement using COOT[23] and PHENIX[24] software packages. The statistics for data processing and structure refinement are summarized in Table 1.

Analytical size-exclusion chromatography

The solution states of NSP7, NSP8 and NSP7–NSP8 complex were analyzed using size-exclusion chromatography. In essence, 100 uL of protein solution at a concentration of 0.5 mg/ml was loaded onto Superdex 200 increase 10/300 gl column (GE Healthcare) and eluted using 25 mM HEPES (pH 7.5), 100 mM NaCl, 5% glycerol and 5 mM DTT.

Crosslinking assay

To prepare the NSP7-NSP8 mixture, NSP7 and NSP8 proteins were each diluted to 5 mg/mL in 25 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol and 5 mM DTT, and mixed in a 1:1 molar ratio. Next, 50 mM

ethylene glycol bis(succinimidyl succinate) (EGS) dissolved in DMSO was added into 15uL of the NSP7-NSP8 mixture to reach a final concentration of 5 μM. The reaction mixtures were then incubated on ice for 2 h before being quenched by 50 mM Tris–HCl (pH 7.5). Subsequently, the samples were subjected to SDS-PAGE analysis.

Thermal shift assay

Thermal shift assay for NSP7 WT and mutants were conducted using a BioRad CFX Connect Real-Time PCR detection system. For each measurement, 20 uL of sample

mixture contains 5.5 uM WT or mutant NSP7 dissolved in buffer containing 20 mM HEPES (pH 7.5), 10% glycerol, 150 mM NaCl, and 1× GloMelt Dye. The sample plates were heated from 25 to 95◦C with heating increments of 0.5◦C. Fluorescence intensity was recorded within the excitation/emission ranges of 470/510 nm. Each sample was prepared in triplicate for the measurement.

RdRP polymerase assay

A minimal hairpin RNA substrate was used as previously reported[13]. In essence, the RNA (/5'-FAM/rUrUrUrUrCrArUrGrCrUrArCrGrCrGrUrArGrUrUrUrUrCrUrArCrGrCrG-

3') was purchased from Integrated DNA Technologies. The RNA was annealed by heating the solution to 75◦C and gradually cooling to 4◦C in the buffer containing 50 mM NaCl and 10 mM Na-HEPES (pH 7.5). The polymerase assay mixture contained 5 M RNA dissolved in 100 mM NaCl, 20 mM Na-HEPES (pH 7.5), 5% (v/v) glycerol, 10 mM MgCl2 and 5 mM -mercaptoethanol, in the presence of the indicated NSP12 (5 µM), NSP8 (15 µM) and/or NSP7 (15 µM) proteins. The reaction was initiated by addition of NTPs(150 µM UTP, GTP, and CTP and 300 µM ATP), followed by incubation at 37◦C for 20, 40 and 60 min. 2× loading dye (7 M urea, 50 mM EDTA pH 8.0, 89 mM Tris-base and 28 mM Taurine) were added to stop the reaction. The reaction samples were then separated on 7 M urea, 20% acrylamide gels (8 cm × 8 cm × 1 mm) in 45 mM Tris-base, 14 mM Taurine, and 0.3 mM EDTA. 6-FAM-labeled RNA products were visualized by ChemiDoc Imager (Bio-Rad Laboratories, Inc.)

RESULTS

Biochemical characterization of the SARS-CoV-2 NSP7– NSP8 assembly

To examine the assembly state of NSP7 and NSP8 proteins in solution, we performed size-exclusion chromatography with NSP7, NSP8, or a mixture of NSP7–NSP8 (1:1 molar ratio). Notably, the NSP7 protein eluted at a volume close to what is expected for its monomeric form, whereas NSP8 eluted at a volume corresponding to its dimeric form (Figure 2-1 B), in line with previous reports that NSP7 and NSP8 exist as a monomerand a dimer in solution, respectively[20, 25]. On the other hand, the NSP7-NSP8 mixture eluted at a volume corresponding to what is expected for a 2:2 tetrameric form (~62 kDa) (Figure 2-1 B), far lower than what is expected for a hexadecameric NSP7-NSP8 complex (~240 kDa).The caveat of protein size estimation by size-exclusion chromatography is that the elution volume of a protein can also be affected by its shape. Therefore, we also performed *in vitro* cross-linking assays on NSP7 and NSP8 using ethylene glycol bis(succinimidyl succinate) (EGS) to evaluate the assembly states of NSP7 and NSP8. SDS-PAGE analysis of EGS-treated NSP7 and NSP8 products revealed dominant monomeric and dimeric forms, respectively (Figure 2-1 C, left and middle).Under the same reaction condition, crosslinking of the NSP7-NSP8 complex resulted in a strong SDS-PAGE band corresponding to the NSP7-NSP8 heterotetramer, but no appreciable fraction for the hexadecameric form of NSP7-NSP8 (Figure 2-1C, right).

These observations are consistent with the recent mass spectrometry-based observation that NSP7-NSP8 is dominantly present as a tetrameric form in solution[19], suggesting that the SARS-CoV-2 NSP7–NSP8 complex exists as a 2:2 heterotetramer in solution.

Crystal structure of the SARS-CoV-2 NSP7–NSP8 complex

Next, we crystalized the NSP7–NSP8 complex and solved the crystal structure at 2.7 Å resolution (Table 1). The crystal structure of the NSP7-NSP8 complex belongs to the space group *P*21, with each asymmetric unit containing two NSP7 and NSP8 molecules (Figure 2-2 A, B). Despite the fact that full-length NSP7 and NSP8 proteins were prepared for crystallization, we were only able to trace the electron density for the C-terminal domain of NSP8 spanning from residues E77 to S193, with the N-terminal portion most likely proteolytically cleaved during crystallization. On the other hand, we were able to trace the entire NSP7 molecule except for residues L83-Q84 at its C terminus. Analysis of the structure of the NSP7–NSP8 complex reveals a 2:2 NSP7– NSP8 tetrameric complex, formed by two closely-packed NSP7–NSP8 dimers (Figure 2-2 A, B). As previously observed for the NSP7–NSP8–NSP12 complex[11,13–17], NSP8 is comprised of an N-terminal domain, albeit with only two -helices traceable here, followed by a C terminal domain formed by a four-stranded antiparallel-sheet packed against three intervening helices (Figure 2-2 A).NSP7 is comprised of four -helices, with the last one moving apart from the first three to cradle the two N-terminal helices of NSP8, resulting in a mixed six-helix bundle (Figure 2-2 B). Furthermore, helix 1 of NSP8 and helices 1–2 of NSP7 pack against the counterparts of the other NSP7–NSP8 complex to form the tetrameric structure (Figure 2-2 A, B). Analysis of the electrostatic surface of theNSP7–NSP8 complex failed to identify significant

basic patches for potential RNA binding sites, in line with the observation that the N-terminal helices, but not the C- terminal domain, of NSP8 is responsible for RNA binding during replication or transcription of the viral genome[12,14,26].Structural comparison of the NSP8-bound NSP7 with the previously reported solution structure of SARS-CoV NSP7 (PDB 2KYS)[25] revealed that the C-terminal half of 1, along with 2 and 3, is well aligned between the twostructures, with a root-mean-square deviation(RMSD)of 2.2 Å over 51 aligned Cα atoms (Figure 2-7 A). The most pronounced structural deviation lies in 4, which is packed against 1–3 in Free State but breaks away in the NSP8-bound form (Figure 2-7 A). To test how this conformational transition of NSP7 affects its interaction with NSP8, we introduced an alanine mutation to NSP7 L71, located at the interface between 1 and 4 in the NSP7–NSP8 complex, and performed thermal shift assay. In comparison with WT NSP7, L71A-mutated NSP7 (NSP7L71A) shows a reduction of thermal stability by ~9◦C (Supplementary Figure 2-7 B) but slightly increased oligomerization with NSP8 (Supplementary Figure 2-7 C), in line with an effect of the dynamic conformational transition of NSP7 α4 on the complex formation of NSP7-NSP8.

Structural basis for the NSP7-NSP8 interaction

Formation of the NSP7-NSP8 tetramer is mediated by two separate interfaces, with one mediating the NSP7–NSP8 dimerization (denoted as

interface I in the expanded viewsin Figure 2-2 B) and the other mediating the tetramerization (denoted as interface II in the expanded view in Figure 2-2 A, herein). Close inspection of the two interfaces revealed that both the dimeric and heterotetrameric association between NSP7 and NSP8 is dominated by non-polar contacts, involving all the four helices of NSP7 and helices 1 and 2 of NSP8 (Figure 2-2 A, B). At interface I, residues (R80, T84, M87, Q88, T89, L91, F92, R96, L98 and N100) from helix 1 of NSP8 are clustered with residues (K2, D5, V6, T9, L13, S15, V16 and Q19) from helix 1 of NSP7 on one side and with residues (V66, I68, L71, E74 and M75) from helix 2 of NSP7 onthe other side (expanded view in Figure 2-2 B, bottom left). In addition, residues (Q31, F49, K51, V53, S57, L60 and S61) from helices 2 and 3 of NSP7 form another hydrophobic cluster with residues (L103, I106, P116, I119, I120, and L122) from helix 2 and its subsequent linker of NSP8 (expanded view in Figure 2-2 B, bottom right). The formation of NSP7-NSP8 dimer results in a buried surface area of ~1445A˚ 2. Formation of interface II is mediated by helices1 and 2 of NSP7 from one NSP7–NSP8 dimer and helix 1 of NSP8 from the other NSP7–NSP8 dimer, which are orthogonally aligned to each other to create complimentary surfaces for side-chain interactions (Figure 2-2A). Reciprocally, residues (S4, K7, C8, V11, V12, H36, N37 and L40) from NSP7 of one NSP7-NSP8 dimer make van der Waals contacts with residues (V83, M87, M90, T93 and M94) from NSP8 of the other NSP7-NSP8 dimer, resulting in a buried surface areaof ~773 $A^{\circ 2}$ (expanded view in Figure 2-2A). Structure-based sequence analysis of the NSP7 and NSP8

proteins among members of the coronavirus family revealed that SARS-CoV-2 NSP7 and NSP8 are closely related to their counterparts in SARS-CoV, with 99% and 98% sequence identity, respectively, whereas the more distant FCoV NSP7 and NSP8, have only 42% and 41% sequence identity, respectively (Figure 2-8 A,B). The residues located on the two oligomer interfaces of NSP7 and NSP8 fall into highly conserved sites (Figure 2-2 C, D), suggesting a conserved interaction mechanism for the NSP7–NSP8 complex formation across all coronaviruses.

Coupling between heterodimerization and heterotetramerization of NSP7–NSP8

To further understand the structural basis for the NSP7– NSP8 assembly, we selected a number of residues from both oligomer interfaces of NSP7 and NSP8 for mutagenesis (Figure 2-3 A, B), and evaluated their impact on the assembly of the NSP7–NSP8 complex via crosslinking assay. In this assay, we mainly evaluated the formation of NSP8 dimer, which is the dominant form of free NSP8 (Figure 2-1 C), as opposed to NSP7–NSP8 complex formation, given that the NPS7–NSP8 complex is spread over the heterodimeric and heterotetrameric forms of the complex. The relative population of the oligomeric states of the NSP7–NSP8 complex is likely influenced by multiple structural and dynamic factors. Inspection of the crosslinking products of the NSP7– NSP8 mixtures revealed that most of the mutations on the heterodimeric interface I of NSP7 (NSP7F49A and

NSP7L56A in Figure 2-3 C) and of NSP8 (NSP8F92A in Figure 2-3D) lead to an increased NSP8 dimerization, accompanied by reduction of the heterotetramerization of NSP7– NSP8, suggesting the impairment of NSP7–NSP8 association by these mutations. Furthermore, whereas the NSP7 M52A mutation (NSP7M52A) does not appreciably affect the population of NSP8 dimer, it leads to a substantial heterotetramer-to-heterodimer shift of NSP7-NSP8 (Figure 2-3 C), supporting a notion that structural integrity of interface I also affect the heterotetrameric assembly of the NSP7–NSP8 complex. Compared with the interface I mutations of NSP7, the two interface II mutations of NSP7, C8G and V11A, lead to an even more pronounced increase of NSP8 homodimerization at the expense of NSP7-NSP8 heterotetramer (Figure 2-3 E). Note that the homotetramer band of NSP8 also appears more visible for these mutants, further supporting the notion that these interface I mutations severely disrupt the NSP7–NSP8 heterotetramer. Likewise, we observed that the NSP8 interfaceII mutations, M90A (NSP8 M90A) and M94A (NSP8 M94A), lead to an increased NSP8 dimer formation at the expense of the NSP7–NSP8 heterotetramer to an extent that is comparable or even more severe than the interface I mutation NSP8F92A (compare Figure 2-3F with 2-3C). Together, these observations suggest that interface II not only serves to maintain the heterotetrameric assembly of NSP7–NSP8, but also helps to stabilize the heterodimeric assembly of NSP7–NSP8, thereby uncovering a synergistic coupling between the heterodimerization and heterotetramerization of the NSP7–NSP8 complex. Further size-exclusion

chromatography analyses of the NSP7– NSP8 mixtures revealed that mutations on the interface II of NSP7 lead to even more severe loss of the tetrameric NSP7–NSP8 fraction than mutations on interface I (Figure 2-9),which reinforces the notion of a structural coupling between the two oligomer interfaces of NSP7 and NSP8.

Structural comparison of the coronavirus NSP7-NSP8 complexes

Crystal or cryo-electron microscopic (cryo-EM) structures have been reported for coronavirus NSP7–NSP8 complexes under a variety of assembly forms, including the SARS-CoV NSP7–NSP8[17] and NSP7-NSP8-NSP12 complexes[11], the SARS-CoV-2 NSP7–NSP8–NSP12 complexes in the absence or presence of RNA substrates [14–16],as well as the NSP7–NSP8 complex from FCoV[20]. These structures show diverse arrangements of the NSP7 and NSP8 proteins, including the hexadecameric arrangement of the SARS-CoV NSP7–NSP8 complex in which both NSP7 and NSP8 carry an N-terminal non-native GPLGS tag (Figure 2-10 A) [17], 1:2:1 heterotetrameric arrangement of the NSP7–NSP8–NSP12 complexes in which each NSP12 molecule is associated with one NSP8 monomer and one NSP7–NSP8 heterodimer[14-16], and the 1:2 heterotrimeric arrangement of the FCoV NSP7–NSP8 complex[20]. Among these, the central channel of the hexadecameric complex of SARS-CoV NSP7–NSP8 (Figure 2-10 A) has been proposed to serve as an RNA binding site[17], which might mediate the potential primase

activity of this complex [20, 27, 28]. However, this observation was later challenged by the fact that a SARS-CoV NSP7–NSP8 fusion protein (N7L8) had no detectable *de novo* RNA synthesis activity[10] and the biochemical evidence indicating that SARS-CoV NSP7–NSP8 exists as a tetramer in solution [18, 19]. Further analysis of the hexadecameric form of the SARS-CoV NSP7-NSP8 complex revealed that it harbors three alternative repeating units, with each adopting the form of an NSP7–NSP8 heterotetramer (Figure 2-10 B–D, denoted as tetramers I, II and III, respectively). Among these, formation of SARS-CoV

NSP7– NSP8 tetramer I is mediated by the C-terminal domains of NSP7 and NSP8 (Supplementary Figure 2-10 B), as is observed here for the SARS-CoV-2 NSP7–NSP8complex (Figure 2-2 A, B). In contrast, the formation of tetramers II and III in the SARS-CoV hexadecameric NSP7–NSP8 complex is mediated by the N-terminal helices of NSP8 proteins (Figure 2-10 C, D). To test the role of the N-terminal domain of NSP8 in the NSP7-NSP8 assembly, we performed size-exclusion chromatography analysis of SARS-CoV-2 NSP7 mixed with the N-terminally truncated SARS-CoV-2 NSP8 (∆73 NSP8), which showed that NSP7 and ∆73 NSP8 remain co-migrating at an elution volume corresponding to their heterotetrameric form (Supplementary Figure 2-10 E). Consistently, crosslinking analysis of the NSP7–∆73 NSP8 mixture confirmed the predominance of the heterotetrameric form in solution (Supplementary Figure 2-10 F). These data suggest that in solution the complex of SARS-CoV-2 NSP7–NSP8 is mainly mediated by the C-terminal domains of

NSP8, rather than the long helical domain at the N-terminus. Along these lines, structural superposition of the SARS-CoV-2 NSP7–NSP8 complex with that of SARS-CoV (PDB 2AHM) shows that the SARS-CoV-2 NSP7–NSP8 tetramer is well aligned with tetramer I of the SARS-CoV NSP7–NSP8 complex, resulting in an RMSD of 0.64 Å over 458 aligned Cα atoms (Figure 2-4 A), suggesting that the interactions mediating the heterotetrameric assembly of NSP7-NSP8 are shared by SARSCoV-2 and SARS-CoV.

Oligomer interfaces I and II underpin various assembly statesof NSP7, NSP8 and NSP12

Next, we asked how the oligomer interfaces of NSP7 and NSP8 undergo the transition from the heterotetrameric NSP7–NSP8 complex to the 1:2:1 heterotetrameric NSP7– NSP8-NSP12 RdRP complex. Structural superposition of the SARS-CoV-2 NSP7–NSP8 heterotetramer with the free or the RNA-bound SARS-CoV-2 RdRP reveals that the interface I-mediated NSP7–NSP8 heterodimer is preserved in the full RdRP complex (Figure 2-4 B,C and Supplementary Figure 2-11 A–C). Intriguingly, some of the residues of the interface II of the NSP7–NSP8 tetrameric complex (i.e. NSP7 S4, C8, V11, V12, N37 and L40 and NSP8 T84, M87, M90 and M94) engage in intermolecular contacts with NSP12 in a fashion similar to that seen in the NSP7–NSP8 complex, which is dominated by surface complementarity and hydrophobic contacts (Figure 2-2A and Figure 2-11

C). Nevertheless, distinct interaction modalities are observed for the remaining residues of interface II in the RdRP complex (Figure 2-2A and Supplementary Figure 2-10 C). For instance, the side chain of NSP7 N37 donates a hydrogen bond tothe backbone carbonyl group of NSP12 A443 in the RdRP complex (Figure 2-11 C) but interacts with NSP8 V83 side chain through a van der Waals contact in the NSP7-NSP8 heterotetrameric complex (Figure 2-2A). On the other side of the RdRP complex, association of the NSP8 monomer with NSP12 involves both interface I andII of NSP8 (Figure 2-11D), in addition to the -pairing mediated by the C-terminal domain of NSP8 and the polymerase domain of NSP12 (Figure 2-11 E). These observations suggest that interface I and II of NSP7 and NSP8 mediate the assembly of both the NSP7- NSP8 and the NSP7-NSP8-NSP12 complexes. We further compared the structure of the SARS-CoV-2 NSP7-NSP8 heterotetramer with that of the FCoV 2:1 NSP7-NSP8 heterotrimer (Figure 2-4 D). Despite the different stoichiometry of NSP7 and NSP8, the two complexes show high conservation for the interface I and the resulting heterodimeric structure of NSP7-NSP8 (Figure 2-4 D). In fact, the interface II of SARS-CoV-2 NSP7–NSP8 heterotetramer also resembles the heterotrimeric interface of the FCoV complex, but with subtle differences (Figure 2-4 E, F). For instance, NSP7 S4 and L40 interact with NSP8 M90 in the SARS-CoV-2 NSP7- NSP8 heterotetramer. In contrast, the corresponding residues in FCoV NSP7-NSP8 heterotrimer, NSP7 T4 and N40 and NSP8 L90, interact with a different set of residues (i.e. NSP7 L53 and F76) (Figure 2-4 F).

These sequence divergences may explain why the SARS-CoV-2 NSP7–NSP8 is dominated by a heterotetrameric arrangement in solution, while the FCoV NSP7–NSP8 adopts a heterotrimeric arrangement.

Role of the NSP7-NSP8 interface residues in the activity of SARS-CoV-2 RdRP

To understand how the dynamic NSP7–NSP8-NSP12 assembly affects the RNA replication activity of SARS-CoV-2 RdRP, we next performed the primer-dependent RNA replication assay using a 5'-FAM fluorescently labeled, single-stranded RNA substrate that was recently developed[14]. This 29-nt RNA folds into a hairpin structure, containing a 5 base-pair (bp) stem and an 11-nucleotide 5" overhang, which serve as the template and as the primer for efficient detection of nucleotide extension (Figure 2-5 A). First, incubation of WT SARS-CoV-2 NSP7–NSP8–NSP12 with the RNA substrate leads to a time-dependent increase of the extended RNA product (Figure 2-5 B), confirming the primer-dependent replication activity of the recombinant RdRP complex. In contrast, NSP12 alone or any pairwise combination of NSP7, NSP8 and NSP12 fails to generate an appreciable level of RNA product (Figure 2-5 B), consistent with previous observations that the co-presence of NSP7 and NSP8 greatly boosts the RNA replication efficiency of NSP12-mediated RNA replication [10, 14]. Second, introduction of the mutations on the interface I of NSP7 (F49A: NSP7F49A, M52A: NSP7M52A, L56A: NSP7L56A and F49A/M52A/L56A: TM) or NSP8 (F92A) lead to a decrease of RdRP efficiency to various extents, with the

NSP7 F49A/M52A/L56A triple mutation giving rise to a stronger effect than individual mutations (Figure 2-5 C), in line with the impairments of the RdRP assembly by these mutations. Third, introduction of the mutations on the interfaceII of NSP7 (C8G: NSP7 C8G and V11A: NSP7 V11A) or NSP8 (M90A: NSP8 M90A and M94A:NSP8 M94A) lead to an even more severe reduction of RdRP efficiency (Figure 2-5 D). Of particular note, the NSP7C8G and NSP7V11A mutations, which concern the association of both the NSP7-NSP8 and NSP7-NSP8-NSP12 complexes (Figure 2-2 A and Figure 2-11 C), lead to nearly completely abolished activity of the RdRP complex (Figure 2-5 D), thereby confirming the critical role of NSP7 in the RdRP activity. Together, these data reinforce the notion that the two oligomer interfaces of NSP7 and NSP8 critically mediate the assembly and RNA replication activity of the SARS-CoV-2 RdRP. In addition, we investigated the RdRP complex carrying mutations on the potential RNA binding sites of NSP8, including K58A (NSP8K58A), R75A (NSP8R75A) and K82A (NSP8K82A) located on the N-terminal domain. Structural studies of the SARS-CoV-2 NSP7–NSP8–NSP12 complexes with RNA substrate bound [10,12,14,26] revealed that these residues are positioned in close proximity to the backbone of the exiting RNA duplex (Figure 2-12). Indeed, all three mutations lead to a significant reduction of the RdRP efficiency (Figure 2-5 E), confirming the important role of these residues in regulating the RdRP activity.

NSP7 mutation with differential effect on the

NSP7-NSP8 vs NSP7-NSP8-NSP12 assembly

Finally, we seek to identify any NSP7 or NSP8 mutation that perturbs the transition between NSP7–NSP8 and In light of the fact that the side chain of NSP7 N37 serves asa hydrogen bond donor in the NSP7-NSP8-NSP12 complex but not in the NSP7–NSP8 heterotetramer, we mutated this residue into valine and evaluated its effect on the two complexes. Indeed, crosslinking and size-exclusion chromatography analyses revealed that, although the NSP7N37V mutation does not affect the stability of the NSP7–NSP8 heterotetramer appreciably (Figure 2-13 A, B), it leads to a modest, but notable, disruption of the NSP7-NSP8-NSP12 complex (compare Figure 2-13 C and D). Consistently, the NSP7 N37V mutation significantly compromises the replication efficiency of the NSP7–NSP8–NSP12 complex (Figure 2-5 D). The identification of the NSP7 N37V mutation causing differential effect on various assembly states of RdRP provides a new avenue for the development of allosteric inhibitors that specifically inhibit SARS-CoV-2 RdRP activity. It is worth noting that a recent study indicated that bacterial expression of the codon-optimized NSP12 affects its translational rate, thereby compromising its co-translational folding and consequently, the activity of the RdRP complex[29]. To ensure the proper folding of the recombinant NSP12 protein used in this study, which was codon-optimized for protein expression in E. Coli, we expressed the NSP12 protein in the form of an MBP-fusion protein using a modified pVP13 vector witha T5 promoter[21]. Our RdRP assays indicate that, in comparison

with the NSP12 protein sample encoded by a native NSP12-coding sequence (SARS-CoV-2 isolate: Wuhan-Hu- 1/2020, NC 045512), the NSP12 protein sample derived from the codon-optimized construct shows a similar activity, albeit with a slightly higher accumulation of reaction intermediates (Figure 2-14 A,B). This observation therefore validates the NSP12 sample used in our enzymatic assays.

DISCUSSION

The recurrent outbreaks of viruses call for the development of highly efficient inhibitors targeting the fundamental machinery that underpins viral infections, such as RdRP[30]. Uniquely among RNA viruses, NSP7, NSP8, and NSP12 proteins constitute the core components of the coronavirus RdRP machinery that mediates viral replication. Through combined structural, biochemical, and enzymatic analyses, our study uncovers the molecular basis for the dynamic assembly of the NSP7-NSP8 complex and its relationship to RdRP activity, thereby providing insights into the functional regulation of viral replication. This study reveals that the SARS-CoV-2 NSP7–NSP8 complex adopts a heterotetrameric structure in solution. Formation of the NSP7–NSP8 complex involves two related, yet separate, non-polar interfaces, which mediate the heterodimeric and heterotetrameric assembly of NSP7– NSP8 in a synergistic manner (Figure 2-6). The heterotetrameric interface (interface II) is formed by the N-terminal helices of NSP7 and NSP8, which contribute to relatively conserved yet distinct conformations in the different NSP7-NSP8 assembly states (Figure 2-4 and Supplementary Figure 2-10 A–D). The coupling between the heterodimerization and the heterotetramerization of SARS-CoV-2 NSP7– NSP8 likely arises from the fact that the two oligomer interfaces are formed by a distinct, yet overlapping set of structural elements (i.e. NSP7 1 and NSP8 1): Formation of the SARS-CoV-2 NSP7–

NSP8 heterotetramer presumably leads to reduced conformational entropy of NSP71 and NSP8 1, which in turn stabilizes the interface mediated heterodimeric interactions (Figure 2-6). Note that all three components of the SARS-CoV-2 RdRP complex have recently been shown to possess a RNA replication independent function[31]. In this context, this coupled dimerization-tetramerization of the NSP7-NSP8 complex may not only help to shield the NSP7–NSP8 from unwanted protein interactions, but also provide a mechanism for the dynamic transition between different functional states of NSP7 and NSP8. During the assembly of the RdRP complex, the residues on the interface II of NSP7 and NSP8 interact with NSP12 in a similar manner as that in the NSP7- NSP8 heterotetramer, leading to a shift from the NSP7- NSP8 tetramer to the NSP7–NSP8–NSP12 complex (Figure 2-6). The residues on both interfaces are highly conserved across different coronaviruses, in line with their important roles in mediating the assembly of the RdRP and NSP7- NSP8 complexes. This study therefore reveals an unprecedented NSP7–NSP8 interaction mechanism, with important implications in the functional regulation of the RdRP complex during the replication or other stage of the viral infection. SARS-CoV-2 has been mutating since its emergence, resulting in the appearance of several variants. These new SARS-CoV-2 variants also carry mutations in NSP7, NSP8 and NSP12, including the frequently detected NSP12 P323L, NSP7 S25L and NSP8 M129I and I156V mutations [32]. Among these, NSP12 P323L has been identified as one of the hot-spot mutations associated with increased severity of COVID- 19 [33, 34],

suggesting a link to an increased transmission capacity of the SARS-CoV-2 variants carrying this mutation. Structural analyses of the NSP7–NSP8 and NSP7–NSP8–NSP12 complexes revealed that the P323L mutation is located next to one of the NSP8–NSP12 interfaces (Figure 2-15 A). Replacement of NSP12 P323 with a leucine likely results in enhanced van der Waals interaction between NSP8 and NSP12. Likewise, the NSP7 S25L mutation is located near to the NSP7-NSP8 interface in the NSP7–NSP8–NSP12 complex (Figure 2-15 B); replacement of NSP7 S25 with a bulky phenylalanine therefore may lead to enhanced NSP7–NSP8 association within the RdRP complex. In addition, the NSP8 M129I mutation is mapped onto the interface between NSP12 and the second NSP8 molecule (Figure 2-11 E). However, given the fact that this interface is mainly mediated by a -pairing between NSP8 and NSP12, the NSP8 M129I mutation may not generate any significant impact on the NSP8–NSP12 association. On the other hand, none of these frequent NSP7 or NSP8 mutations mentioned above are located at the oligomer interfaces of the NSP7–NSP8 tetramer (Figure 2-15 C,D), supporting the functional relevance of the NSP7–NSP8 tetramer. Our study also demonstrates that the dynamic equilibrium between different assembly states of NSP7-NSP8 can be fine-tuned by the interface mutations. Note that the residues on the oligomer interfaces are highly conserved across coronaviruses, highlighting their functional importance. Introduction of the NSP7N37V mutation affects the stabilities of the NSP7–NSP8 and NSP7–NSP8–NSP12 differently, leading to a population shift from the NSP7–NSP8–NSP12 complex

toward the NSP7–NSP8 complex in solution, and consequent impairment of the RdRP activity. Conceivably, the NSP7 protein carrying the N37V-like mutations could be exogenously introduced into the infected cells to interact with NSP8 to form an NSP12 binding defective complex, thereby interfering with the assembly of an active viral RdRP complex. In this context, exogenous NSP7 with N37V-like mutations may serve to deplete the pool of NSP8 proteins available for RdRP formation in infected cells, leading to allosteric inhibition against SARSCoV-2. Whether this allosteric inhibition scheme can serve as a novel therapeutic strategy to complement existing nucleoside analogue-based treatment (e.g. Remdesivir) [35] awaits future investigation. While this study was in process, two other groups reported the crystal structures of the SARS-CoV-2 NSP7-NSP8 complex (PDB 6YHU, 6M5I and 6WIQ) [35]. All of these structures are consistent with our observed heterotetrameric assembly of NSP7–NSP8.

FIGURES

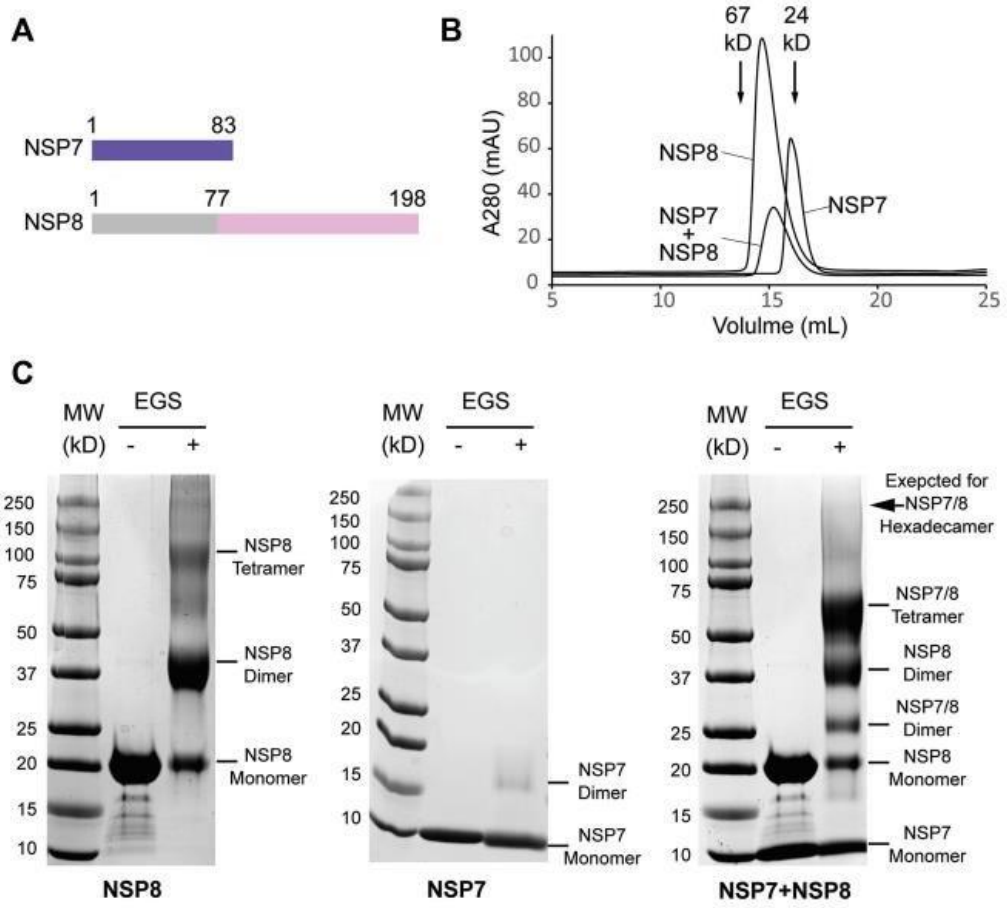

Figure 2-1. Biochemical analysis of the SARS-CoV-2 NSP7-NSP8 complex. **(A)** Domain architecture of SARS-CoV-2 NSP7 and NSP8, with the region of NSP8 missing in the crystal structure colored in grey. (**B**) Size-exclusion chromatography analysis of NSP7, NSP8 and NSP7- NSP8 complex. The elution volumes for standard proteins with known molecular weight are indicated by arrows. (**C)** SDS-PAGE images of NSP8 protein, NSP7 protein, and NSP7-NSP8 mixture, treated with or without cross linker ethylene glycol bis(succinimidyl succinate) (EGS). The individual bands corresponding to distinct assembly states of NSP7 and NSP8 are marked. The position expected for a NSP7–NSP8 hexadecamer on the gel is indicated by arrow.

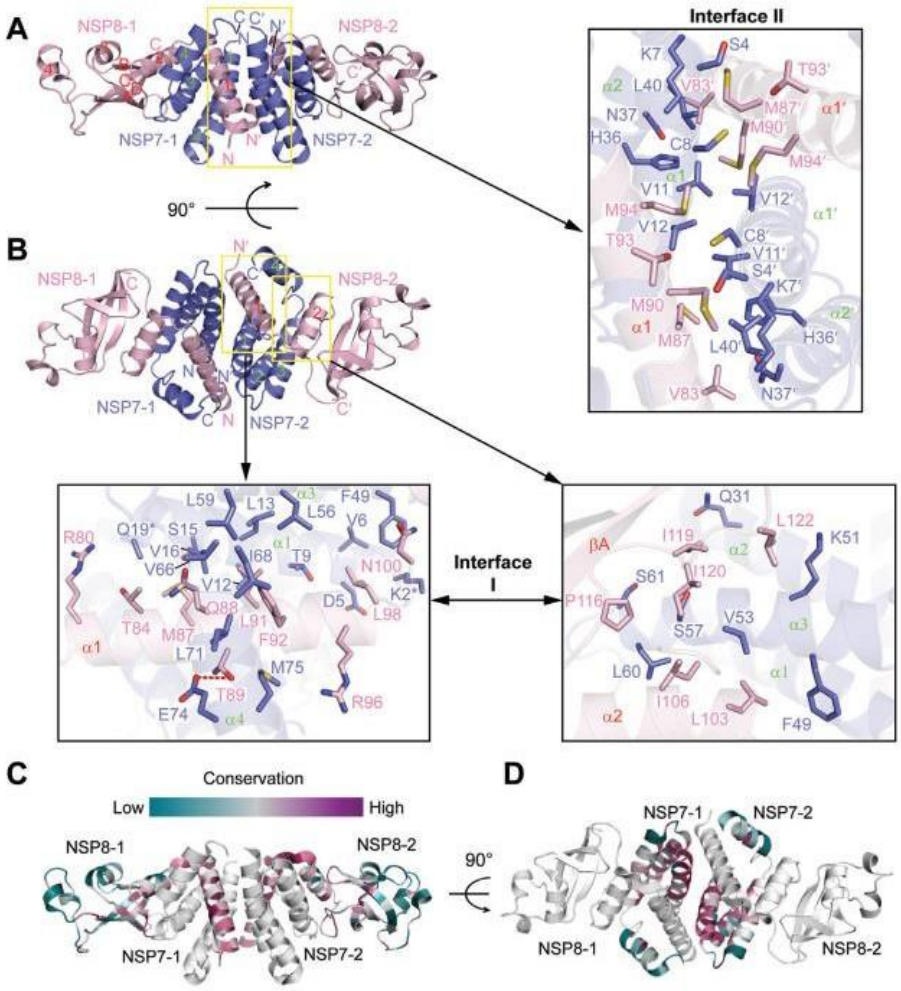

Figure 2-2: Crystal structure of the SARS-CoV-2 NSP7-NSP8 complex. (**A**, **B**) Orthogonal views of the SARS-CoV-2 NSP7-NSP8 complex, with the intermolecular interactions at the two oligomerinterfaces shown in expanded views. The interacting residues are shown in stick representation. The hydrogen bond is shown as a red dashed line. The side chains of NSP7 K2 and Q19, which are untraceable in the crystal structure, are marked by asterisks. (**C**, **D**) Color-coded sequence conservation of SARS-CoV-2 NSP8 (C) and NSP7 (D), analyzed using the ConSurf server.

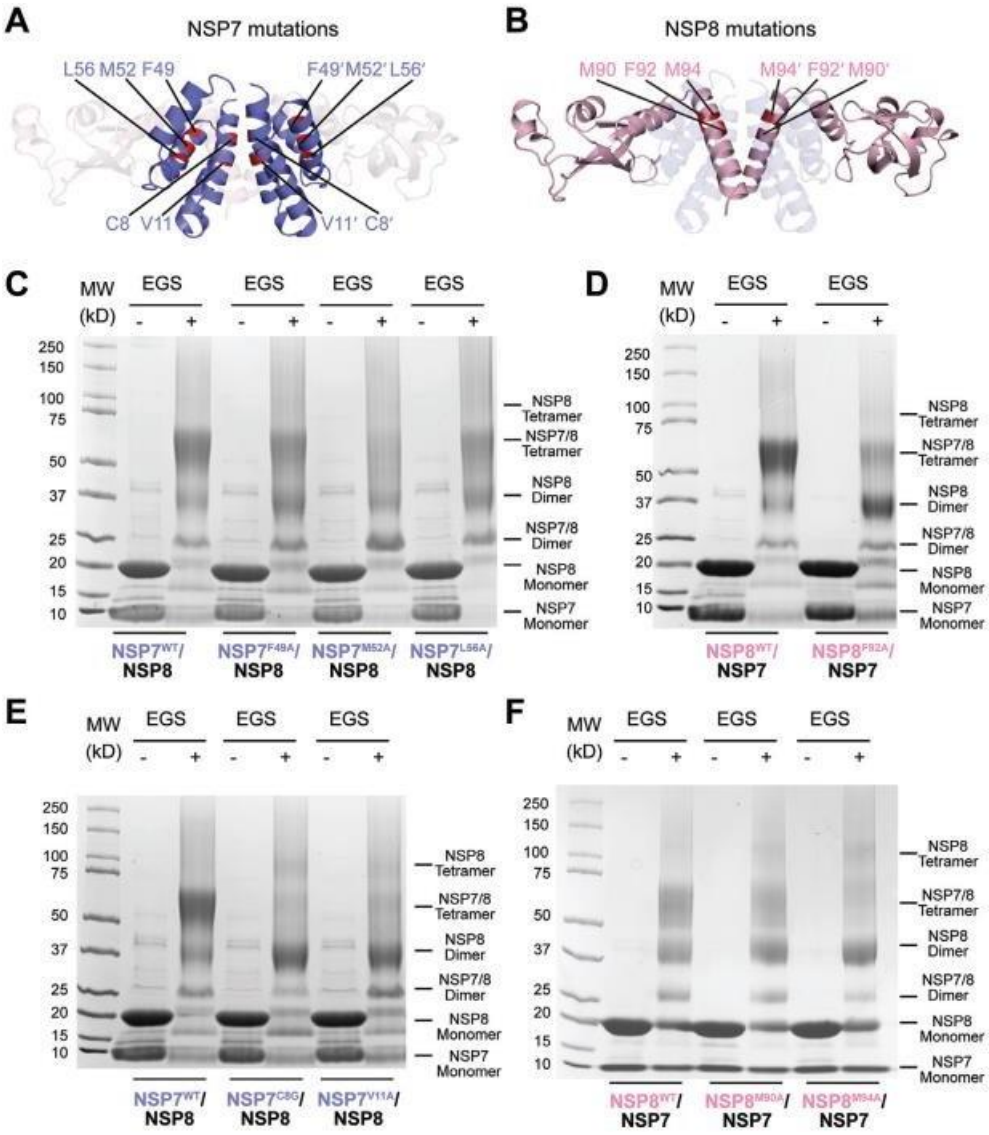

Figure 2-3: Mutational analysis of the SARS-CoV-2 NSP7–NSP8 interaction. (**A**) Residues mutated in NSP7. (**B**) Residues mutated in NSP8. (**C**, **D**) SDSPAGE images of the NSP7-NSP8 complex, WT or mutant on the interface I of NSP7 (C) or NSP8 (D), in the presence and absence of EGS crosslinker. (**E**, **F**) SDS-PAGE images of the NSP7–NSP8 complex, WT or mutant on the interface II of NSP7 (E) or NSP8 (F), in the presence and absence of EGS crosslinker.

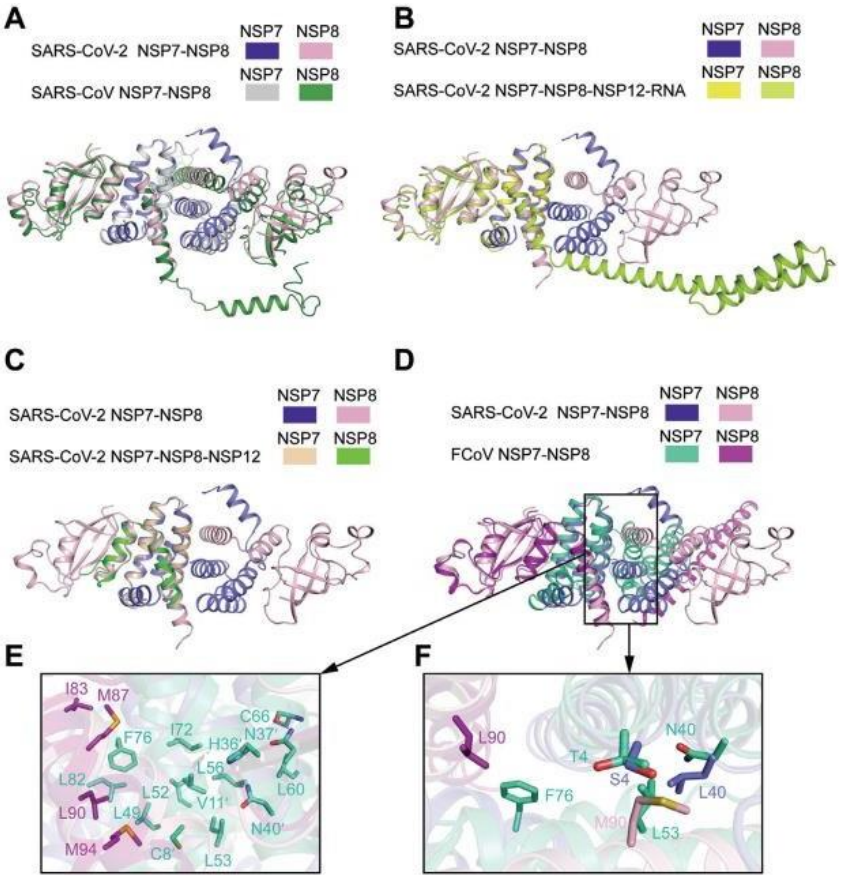

Figure 2-4. Structural analysis of the dynamic assembly of the NSP7-NSP8 complex (**A**) Structural overlay of the SARS-CoV-2 NSP7-NSP8 tetrameric complex with the tetramer I of the SARS-CoV NSP7-NSP8 tetrameric complex (PDB 2AHM). (**B**) Structural overlay of the SARS- CoV-2 NSP7–NSP8 tetrameric complex with the SARS-CoV-2 NSP7–NSP8 dimer in the RNA- bound state of SARS-CoV-2 RdRP (PDB 6YYT). For clarity, NSP12 and the RNA molecule are omitted. (**C**) Structural overlay of the SARS-CoV-2 NSP7-NSP8 tetrameric complex with the SARS-CoV-2 NSP7–NSP8 dimer in the RNA-free state of SARS-CoV-2 RdRP (PDB 6M71). (**D**) Structural overlay of the SARS-CoV-2 NSP7–NSP8 tetrameric complex with the FCoV NSP7– NSP8 dimer in the RNA-free state of FCoV RdRP (PDB 3UB0). (**E**) Close-up view of the intermolecular interactions at the interface II of the FCoV NSP7-NSP8 complex. (**F**) Expanded view of the interface II showing NSP7 and NSP8 residues that are divergent between SARS-COV-2 and FCoV in stick representation.

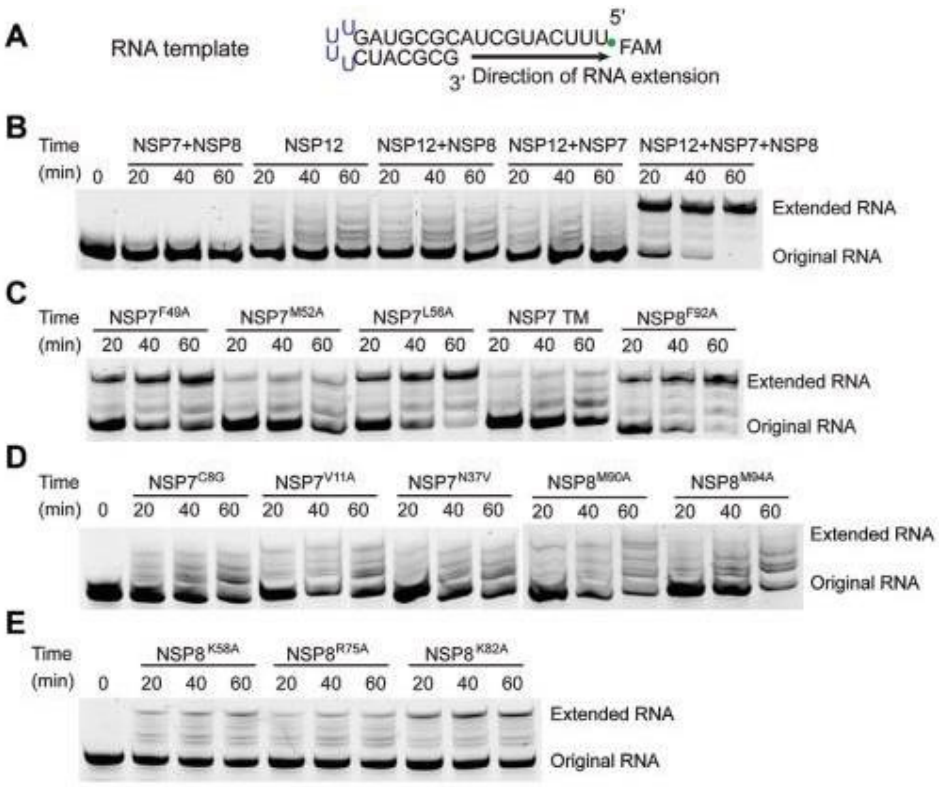

Figure 2-5. RdRP assay using a hairpin RNA substrate. (**A**) Schematic structure of the hairpin RNA. (**B**) Time-dependent incubation of NSP7, NSP8 and/or NSP12 with RNA template. (**C**) Time-dependent incubation of NSP7, NSP8 and NSP12 with RNA template, with NSP7 or NSP8 mutated at interface I. TM refers to NSP7 F49A/M52A/L56A triple mutation. (**D**) Time-dependent incubation of NSP7, NSP8 and NSP12 with RNA template, with NSP7 or NSP8 mutated at interface II. (**E**) Time dependent incubation of NSP7, NSP8 and NSP12 with RNA template, with NSP8 mutated at the N-terminal domain

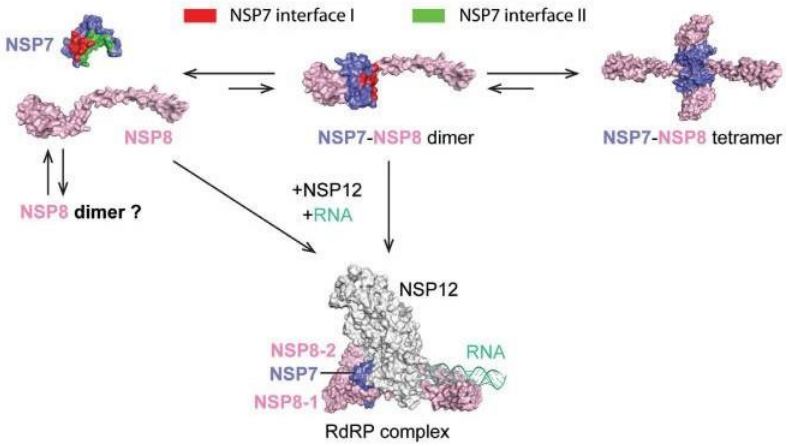

Figure 2-6: A model for the dynamic assembly of the NSP7–NSP8 complex. The mixed populations of NSP7 monomer, NSP8 monomer, NSP8 dimer, NSP7–NSP8 heterodimer, and NSP7–NSP8 heterotetramer co-exist in solution. The population of NSP7–NSP8 heterodimer is transient due to coupled intermolecular interactions between interface I and interface II. In the presence of NSP12 and RNA, NSP7–NSP8 heterodimer and NSP8 monomer associate with NSP12 to form a functional RdRP complex.

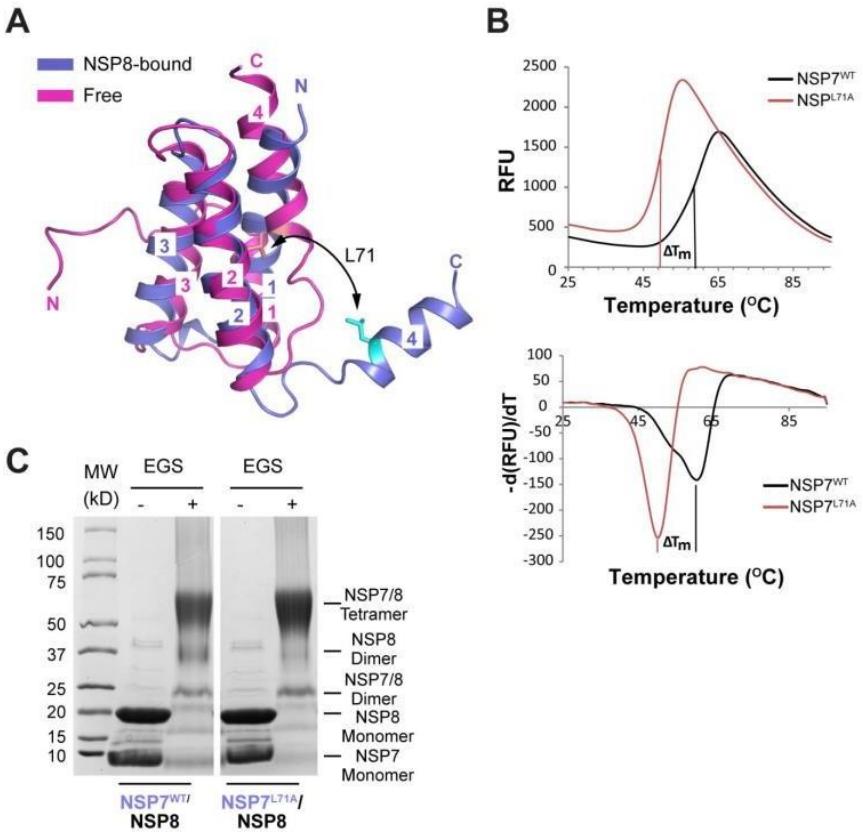

Figure 2-7. Structural and biochemical analysis of the conformational dynamics of NSP7

(A) Structural overlay of NSP8-bound NSP7 of SARS-CoV-2 with SARS-CoV NSP7 in free state (PDB 2KYS). **(B)** Thermal shift assay for the wild type (NSP7WT) and L71A-mutated NSP7 (NSP7L71A), with raw fluorescence data (top) and first derivative of the raw data (bottom) shown. The difference in melting temperature (ΔTm) between NSP7WT and NSP7L71A is indicated. **(C)**SDS-PAGE image showing WT and NSP7 L71A-mutated SARS-CoV-2 NSP7-NSP8 mixture in the presence and absence of ethylene glycol bis(succinimidyl succinate) (EGS) crosslinker.

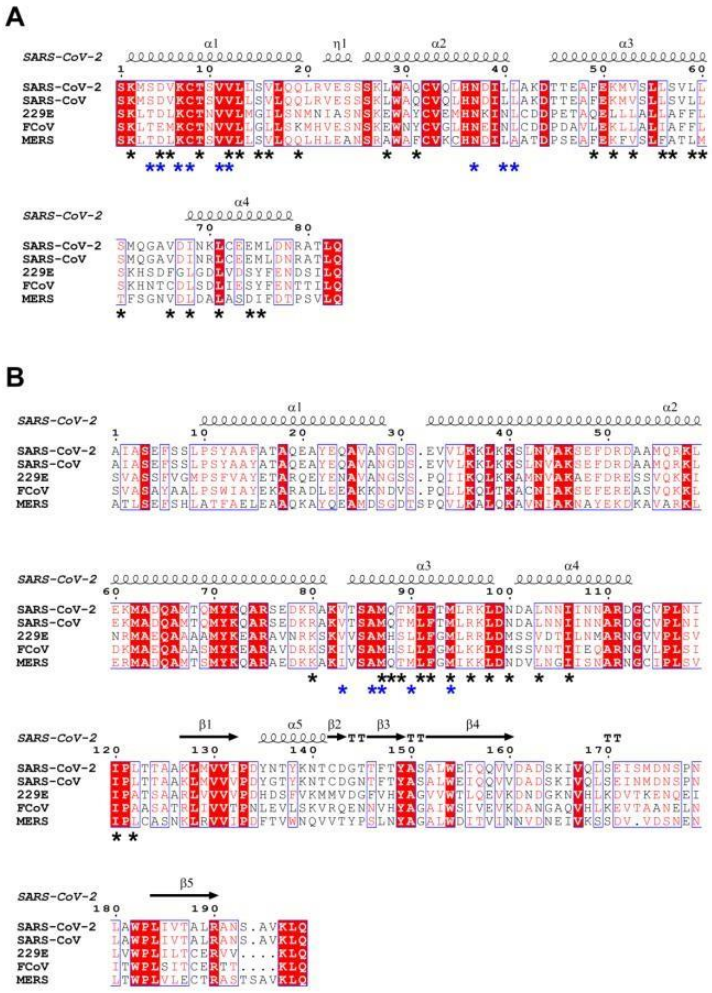

Figure 2-8. Sequence alignment of SARS-CoV-2 NSP7 (A) and NSP8 (B) with homologues from other coronaviruses. Identical or similar residues are boxed and colored in red.

Completely conserved residues are shaded in red. The secondary structures of SARS-CoV-2 NSP7 and NSP8 are marked above the aligned sequences and residues located on interfaces Iand II are marked below the aligned sequences by black and blue asterisks, respectively.

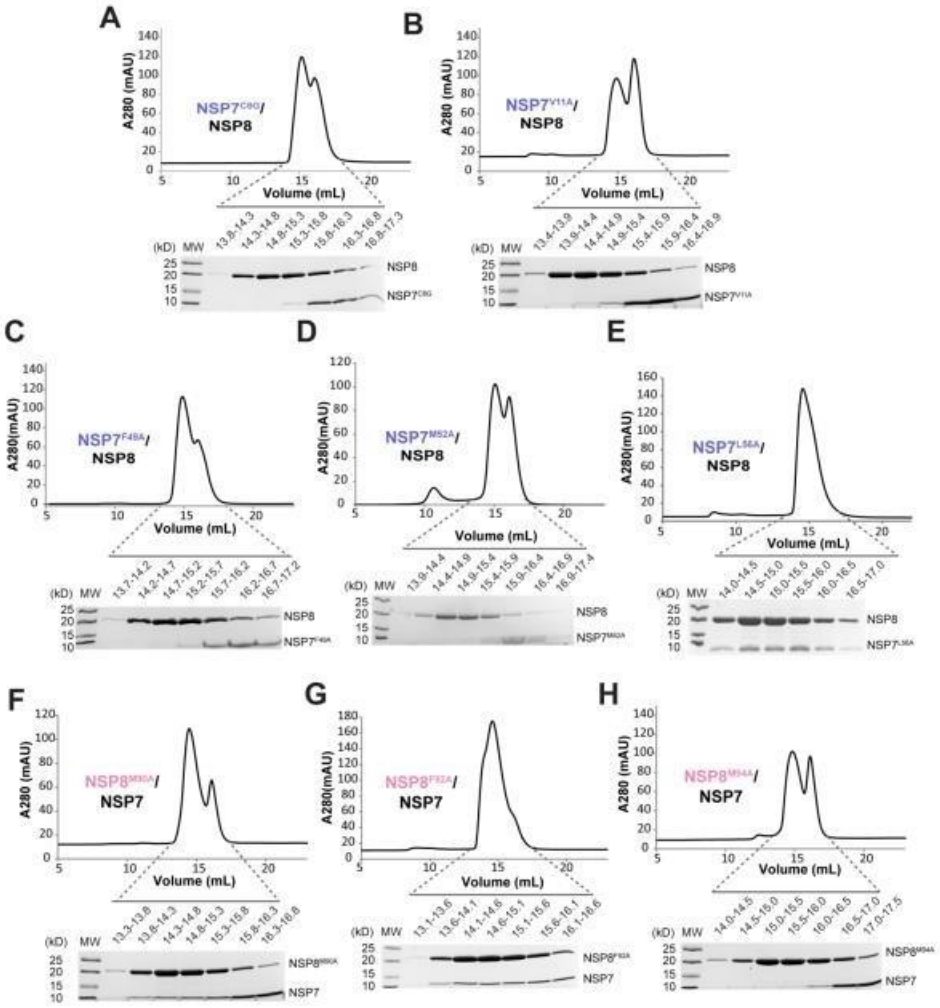

Figure 2-9. Size-exclusion chromatography analyses of NSP7-NSP8 mixtures. (A-E) Gel filtration profile of the NSP7-NSP8 mixture harboring NSP7 C8G (A), NSP7 V11A (B), NSP7 F49A (C)NSP7 M52A (D), or NSP7 L56A (E) mutations. **(F-H)** Gel filtration profile of the NSP7-NSP8 mixture harboring NSP8 M90A (F), NSP8 F92A (G), or NSP8 M94A (H) mutations. The SDS-PAGEimages for selected fractions are shown below.

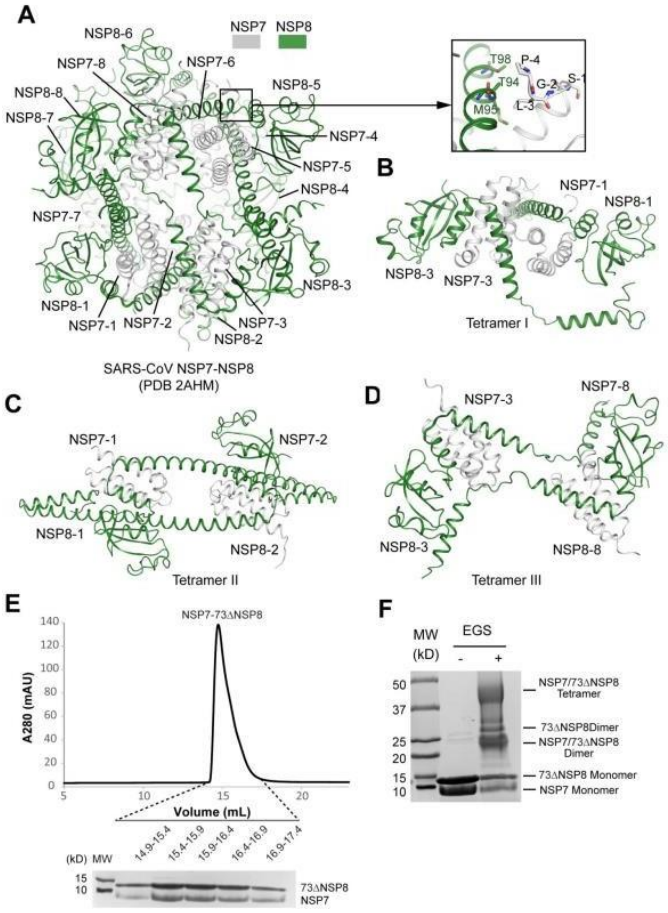

Figure 2-10. Structural and biochemical analysis of the oligomeric state of NSP7-NSP8.(A) Ribbon representation of the crystal structure of the SARS-CoV NSP7-NSP8 hexadecameric complex, with individual NSP7 and NSP8 subunits labeled (NSP7-1 to NSP7-8andNSP8-1to NSP8-8). The N-terminal GPLGS tag of NSP8 participated in the formation of hexadecameric complex, as shown in the expanded view, which may contribute to the stability of the hexadecameric form of the complex in the crystal structure. **(B-D)**Three alternative heterotetrameric assembly of SARS-CoV NSP7-NSP8 derived from the hexadecameric structure in (A). **(E)** Size-exclusion chromatography analysis of the complex formation of NSP8 with N terminal helix deletion (73Δ NSP8) with NSP7. Shown beneath the chromatogram is the SDSPAGE image of the peak fractions. **(F)** SDS-PAGE image shows the NSP7-73ΔNSP8 mixture treated with EGS. The individual bands corresponding to distinct assembly states of NSP7 and NSP8 are marked.

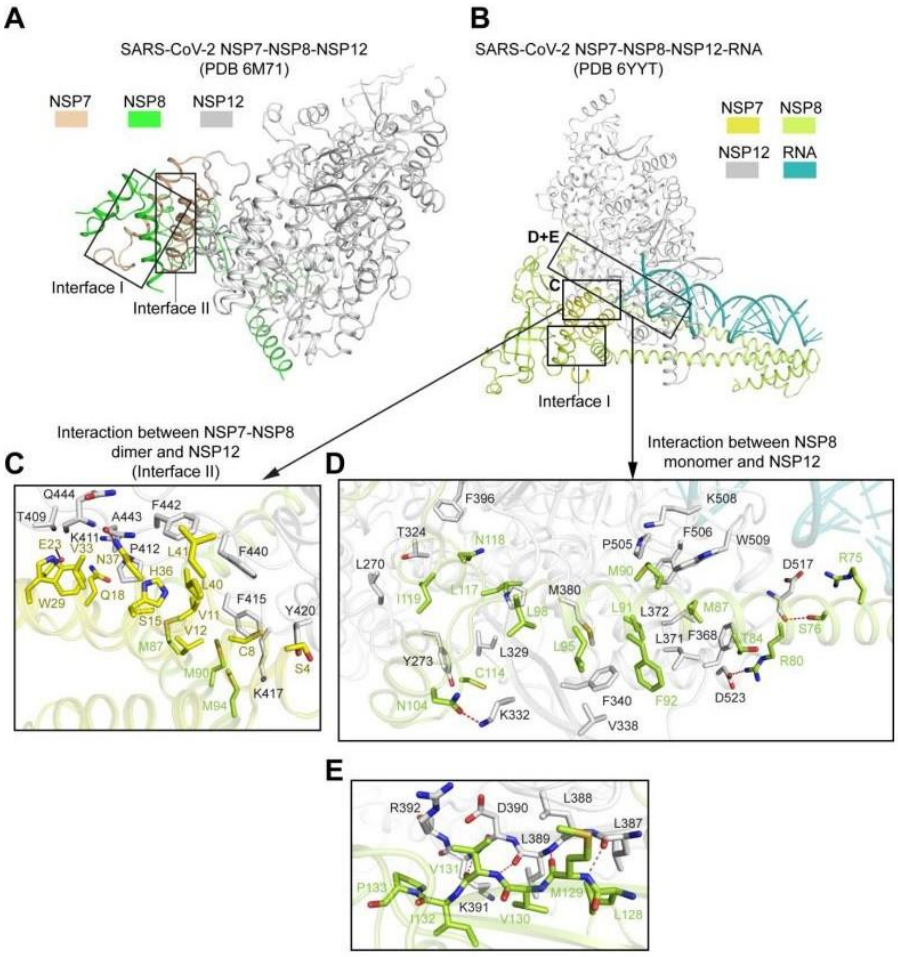

Figure 2-11. Structural analysis of SARS-COV-2 NSP7-NSP8 in the context of the RdRP complex. (A) Cryo-EM structure of NSP7-NSP8-NSP12 (PDB 6M71), with individual subunits color coded. **(B)** Cryo-EM structure of NSP7-NSP8-NSP12-RNA (PDB 6YYT),with individual subunits color coded. **(C)** Expanded view of (B) showing the intermolecular interactions between NSP7-NSP8 and NSP12. **(D-E)** Expanded view of (B) showing the intermolecular interactions between NSP8 monomer and NSP12 involving non-polar contacts (D) and pairing (E).

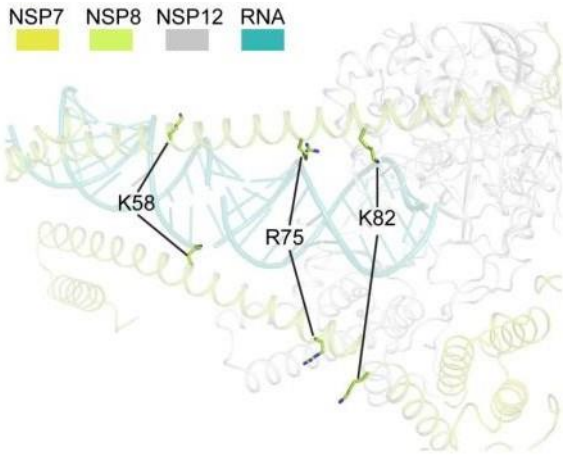

Figure 2-12. Selected RNA-binding sites on NSP8 in the structure of NSP7-NSP8-NSP12- RNA complex (PDB 6YYT). The side chains of NSP8 K58, R75 and K82 are shown in stick representation.

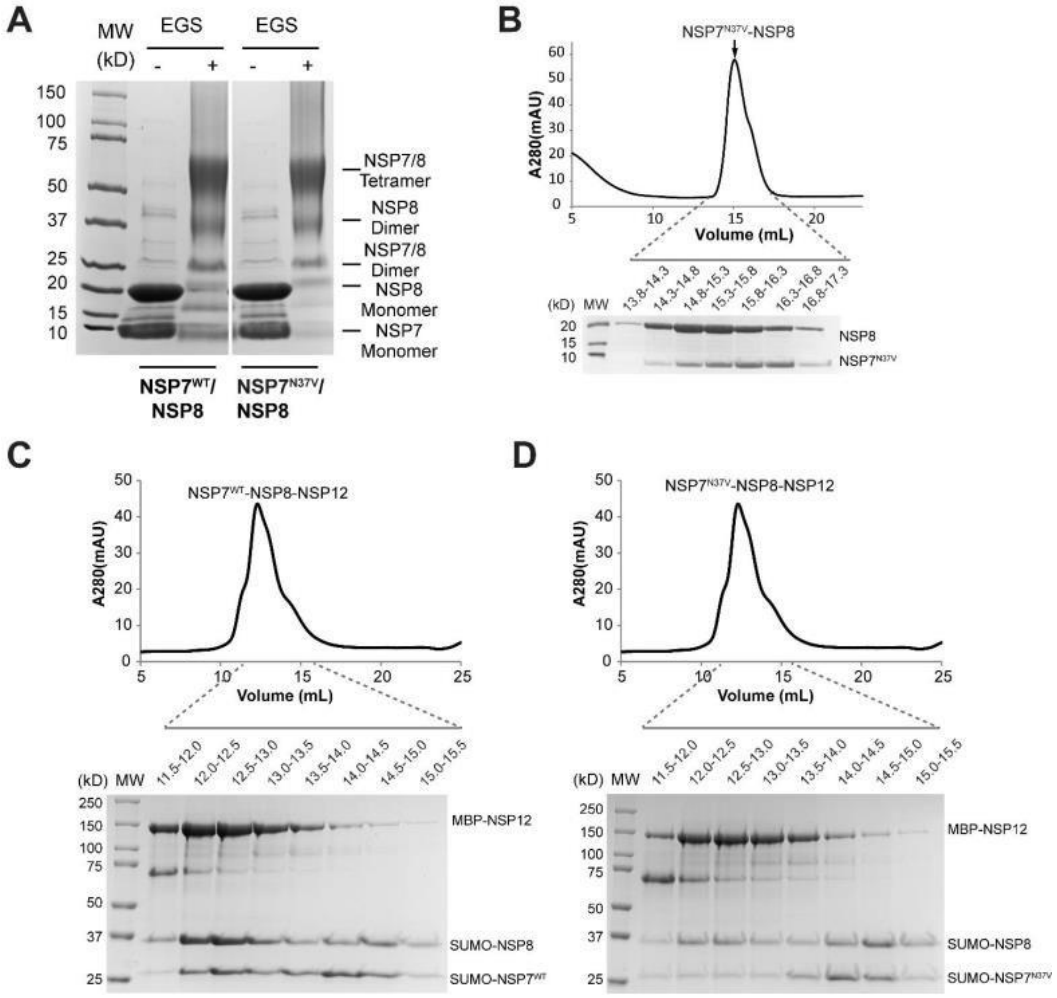

Figure 2-13. Biochemical analysis of NSP7 N37V mutant. (A) SDS-PAGE images of the NSP7- NSP8 mixture harboring WT or N37V mutant, in the presence and absence of EGS crosslinker. **(B)** Gel filtration profile of NSP7N37V-NSP8 mixture, with the SDS-PAGE images of peak fractions shown below. **(C, D)** SDS-PAGE images showing gel filtration fractions of SUMO- NSP7WT mixed with SUMO-NSP8 and MBP-NSP12 (C) and SUMO-NSP7N37V mixed with SUMO-NSP8 and MBPNSP12.

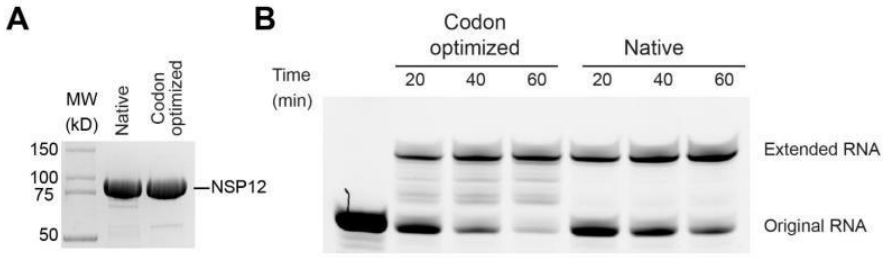

Figure 2-14. Enzymatic comparison of the NSP12 proteins derived from the gene sequence with different codon usage. (A) SDS-PAGE images of the NSP12 proteins derived from the gene sequences either native (Wuhan-Hu-1/2020, NC_045512) or codon-optimized for bacterial expression. **(B)** Time-dependent RdRP assay with the NSP12 protein, derived from either the native or codon-optimized gene sequence, co-incubated with NSP7 and NSP8 proteins.

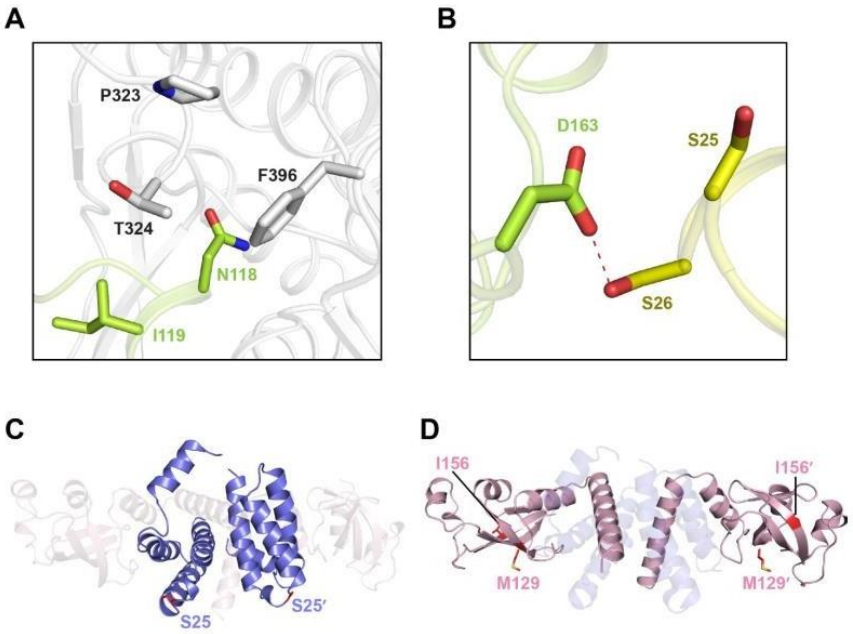

Figure 2-15. Structural analysis of the NSP7 and NSP8 mutation sites from new SARS-CoV- 2 variants. (A) Close-up view of NSP12 P323 and its surrounding protein residues in the NSP7- NSP8-NSP12-RNA complex (PDB 6YYT). NSP12 and NSP8 are colored in grey and lime green, respectively. **(B)** Close-up view of NSP7 S25 and its surrounding protein residues in the NSP7- NSP8-NSP12 complex (PDB 6YYT). NSP7 and NSP8 are colored in yellow and lime green, respectively. Note that the neighboring S26 of NSP7 forms a side-chain hydrogen bond with NSP8 D163. **(C, D)** Mapping of NSP7 S25 (C) and NSP8 M129/I156 (D) onto the structure of the NSP7- NSP8 tetramer in this study (PDB 7JLT).

Table1. Data collection and refinement statistics of SARS-CoV-2 NSP 7-8 complex.

	SARS-CoV-2 NSP7-NSP8 (PDB 7JLT)
Data collection	
Space group	*P* 2_1
Cell dimensions	
a, *b*, *c* (Å)	42.8, 110.6, 42.9
α,β,γ (^{0})	90, 105.7, 90
Resolution (Å)	41.3-2.7 (2.8-2.7)[a]
R_{merge}	0.129 (0.817)
I/σ (*I*)	8.3 (1.1)
$CC_{1/2}$	0.994 (0.628)
Completeness (%)	98.6 (93.1)
Redu dancy	4.0 (3.9)
Refinement	
Resolution (Å)	41.3-2.7 (2.8-2.7)
No. reflections	0405 (956)
R_{work} / R_{free}	23.7/28.3 (30.4/37.5)
No. atoms	
Protein	2935
Water	50
B factors (Å^2)	
Protein	57.93
Water	50.62
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (Å^2)	1.481

[a] Values in parentheses are for highest- resolutionshell.
The data were collected from a single crystal.

CHAPTER 3

Structural basis of SARS-CoV-2 N_{1-25} interaction with human G3BP1 $_{NTF2}$

Specific aim: To determine the molecular basis of the SARS-CoV-2 N - G3BP1interaction using crystallography and biochemical analysis.

METHODS

Cloning, expression and purification of proteins

The DNA fragment encoding SARS-CoV-2 N, codon optimized for bacterial expression, was synthesized from Thermo Fisher Scientific. The cDNAs for full-length human G3BP1 and G3BP2 were purchased from DNASU plasmid repository (https://dnasu.org). The SARS-CoV-2 N_{1-25} peptide was cloned in a modified pRSF vector, in which N1-25 was preceded by an N-terminal His6-SUMO tag and ULP1 (ubiquitin-like protease) cleavage site. Full-length SARS-CoV-2 N, full-length G3BP1, $G3BP1_{NTF2}$ (residues 1-139) and $G3BP2_{NTF2}$ (residues 1-139) were cloned into an in-house His6-MBP vector, preceded by a TEV cleavage site. The plasmids were transformed into BL21 RIL (DE3) cell strain (Agilent technologies) for protein expression. The transformed cells were first grown at 37 °C until cells attained an OD600 of 0.8. The temperature was then lowered to 16 °C,

after which the cells were induced by addition of 0.1 mM isopropyl β-D-galactoside and continued to grow overnight. For His6-SUMO-tagged N1-25, the fusion protein was purified using a Ni-NTA affinity column, followed by removal of the His6-SUMO tag with ULP1 protease and further purification by size-exclusion chromatography on a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) pre equilibrated with 25 mM Hepes (pH 7.5), 0.3 M NaCl and 2 mM DTT. His6-MBP-tagged full-length G3BP1, G3BP1$_{NTF2}$ and G3BP2$_{NTF2}$ proteins were purified sequentially through Ni-NTA chromatography, Q HP column (GE Healthcare) followed by TEV protease treatment. Samples were then subjected to a second round of Ni-NTA chromatography for tag removal. Finally, the proteins were purified by size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) pre equilibrated with 25 mM Hepes (pH 7.5), 300 mM NaCl, and 2 mM DTT. The purified proteins were confirmed by SDS-PAGE, concentrated to ~20 mg/mL and stored at -80 °C for further use.

Crystallization, X-ray data collection and structure determination

For crystallization, ~6 mg/mL G3BP1NTF2 protein dissolved in 25 mM Hepes (pH7.5), 300 mM NaCl and 2 mM DTT was mixed with the N1-25 peptide in a 1:2 molar ratio. The crystallization condition for the SARS-CoV-2 N$_{1-25}$-G3BP1$_{NTF2}$ complex was initially identified through sparse-matrix

screening (Hampton Research Inc.). The crystals were then reproduced by hanging-drop vapor diffusion method at 4 °C, from drops mixed from 1 µL of SARS-CoV-2 N_{1-25}-G3BP1$_{NTF2}$ complex and 1 µL of precipitant solution containing 20% propan-2-ol, 0.1 M MES monohydrate (pH 6.0), 20% PEG MME 20,000. Crystals were soaked for one minute in a cryoprotectant solution, comprised of crystallization buffer and 30% glycerol, before flash frozen in liquid nitrogen. The X-ray diffraction data for the SARS-CoV-2 N_{1-25}-G3BP1$_{NTF2}$ complex were collected on the BL5.0.3 beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory. The diffraction data were indexed, integrated and scaled using the HKL3000 program[22].The structure was solved using the molecular replacement method in PHASER[87] with the structure of human G3BP1NTF2 in complex with SFV nsP3449-471 (PDB ID: 5FW5)as search model. The structure was improved by iterative model building and refinement with Coot[23] and PHENIX[24] software packages. The same R-free test set was used throughout the refinement. The statistics for data collection and structural refinement of the SARS-CoV-2 N_{1-25}-G3BP1$_{NTF2}$ complex is summarized in Table 2.

Isothermal titration calorimetry.

A MicroCal iTC200 system (GE Healthcare) was used to conduct the ITC measurements. All proteins and peptides were dialyzed against buffer

containing 25 mM Hepes (pH7.5) and 300 mM NaCl before titration. For titration of $G3BP1_{NTF2}$ or $G3BP2_{NTF2}$ with SARSCoV-2 N_{1-25}, 0.1 mM human $G3BP1_{NTF2}$ or $G3BP2_{NTF2}$ and 1 mM SARS-CoV-2 N_{1-25} were used. For the rest of titrations, 0.03 mM full-length G3BP1 and 0.3mM SARS-CoV-2 N, full-length or fragments, were used. A total of 20 injections witha spacing of 180s and a reference power of 5 µcal/s were performed at 25 °C. The ITC curves were processed with ORIGIN (MicroCal) software by using a one-site fitting model.

RESULTS

Biochemical characterization of the SARS-CoV-2 N-G3BP1/2 interaction

To identify the interaction elements between SARS-CoV-2 N and G3BP1, we performed isothermal titration calorimetry (ITC) on full-length or truncated fragments of SARS-CoV-1N and G3BP1. Full-length SARS-CoV-2 N and G3BP1 interact strongly, with a dissociation constant (K_d) of 1.0 μM (Fig. 3-1 B and Fig. 3-6 A), and enthalpy and entropy changes of ΔH = -9.9 kcal/mol and ΔS = -5.9 cal/mol/deg, respectively (Table 3). Similar binding thermograms were observed when full-length G3BP1 was titrated withthe C-terminal IDR3-truncated SARS-CoV-2 N (residues 1-364, N_{1-364}), which gave a K_d of 1.7 μM (Fig. 3-1C and Fig. 3-6 B), and enthalpy and entropy changes of ΔH = -9.4 kcal/mol and ΔS = -5.4 cal/mol/deg, respectively (Table 3). Next, we titrated full-length G3BP1 with the SARS-CoV-2 N fragment encompassing the IDR1 and NTD but not the CTD and IDR2 (residues 1-175, N_{1-175}). We observed a K_d of 10.9 μM (Fig. 3-1 C and Fig. 3-6 C), which is ~11- and ~6-fold weaker than those of full-length N or N_{1-364}, respectively. Interestingly, comparison of the titration parameters for full-length SARS- CoV-2 N, N_{1-364} and N_{1-175} reveals that removal of the IDR2 and CTD led to increased reductions for both enthalpy (ΔH = -18.2 kcal/mol) and entropy (ΔS = -38.2 cal/mol/deg) (Table 3), suggesting that, in the context of full-length N or N_{1-364}, the CTD and/or IDR2 contribute to the G3BP1 binding by reducing the entropic cost of the

complex formation. Furthermore, we titrated full-length G3BP1, the $G3BP1_{NTF2}$ domain and the $G3BP2_{NTF2}$ domain with the peptide derived from the first 25 residues of the IDR1 of SARS-CoV-2 N protein (N_{1-25}), which gave K_ds of 7.9 μM, 8.5 μM and 10.9 μM, respectively (Fig. 3-1C and Fig. 3-6 D-F), consistent with the previous observations that the IDR1 of SARS- CoV-2 N protein is primarily responsible for its interaction with G3BP1/2 [44,64,73,74]. Notably, the enthalpy and entropy changes associated with the titrations by the N_{1-25} peptide are equivalent to those of SARS-CoV-2 N_{1-175} within experimental error (Table 3), suggesting a minimal effect of the NTD of SARS-CoV-2 N in G3BP1 binding. Finally, titration of full-length G3BP1 with SARS-CoV-2 N_{26-419} showed no appreciable binding (Fig. 3-1 C and Fig. 3-6 G), indicating that the N_{1-25} is indispensable for the SARS-CoV-2 N-G3BP1 interaction.

Together, these observations identified that the SARS-CoV-2 N_{1-25} and the $G3BP1_{NTF2}$ domain as the major elements for the SARS-CoV-2 N protein-G3BP1 interaction, whereas the CTD and/or IDR2 of SARS-CoV-2 N protein contribute to the interaction entropically.

Crystal structure of the SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$ complex

We then solved the crystal structure of the $G3BP1_{NTF2}$ domain bound to the SARS-CoV-2 N_{1-25} peptide at 2.35 Å resolution (Fig. 3-2 A and Table 2). The SARS-CoV-2 N_{1-25}- $G3BP1_{NTF2}$ complex belongs to the $P2_12_12_1$ space group, with each asymmetric unit containing a $G3BP1_{NTF2}$ homodimer bound to the N_{1-25} peptide in a $G3BP1_{NTF2}$:N_{1-25} ratio of 2:2. We were able to

trace nearly the entire G3BP1$_{NTF2}$ domain, spanning residues V2-F138 (except for S47) (Fig. 3-2A), as well as residues P13-D22 of the SARS-CoV-2 N$_{1-25}$ peptide (Fig. 3-7 A). Note that this segment of SARS-CoV-2 N is strictly conserved in SARS-CoV N, but not in the counterparts of related Middle East Respiratory Syndrome Coronavirus (MERS-CoV) or other coronaviruses (Fig. 3-7 B).

As previously characterized [63, 66, 75], the G3BP1$_{NTF2}$ domain is dominated by a five- stranded antiparallel β-sheet, preceded by three α-helices packed on one side of the β- sheet (Fig. 3-2 A). Two of the G3BP1$_{NTF2}$ domains further undergo a face-to-face β-sheet stacking with each other to form a homodimer. On the outer face of the β-sheet of each G3BP1$_{NTF2}$ monomer, the α1- and α2-helices join with the loop connecting β3 and β4 ($l_{\beta3\beta4}$) to form a long surface groove, cradling the extended SARS-CoV-2 N$_{1-25}$ peptide (Fig. 3-2 B). The association of SARS-CoV-2 N$_{1-25}$ peptide with the G3BP1$_{NTF2}$ is underpinned by their strong surface complementarity (Fig. 3-2 B). Notably, the bulky side chains of N$_{1-25}$ I15 and F17 are accommodated by a ~5.6 Å-wide groove, whereas the downstream residues G18 and G19 snug into a ~3.5 Å-wide groove (Fig. 3-2 B). In fact, our structural modeling analysis indicated that replacement of N$_{1-25}$ G18 with bigger- sized threonine would lead to a steric clash with G3BP1$_{NTF2}$ K123 (Fig. 3-7 C). Together, the association of SARS-CoV-2 N$_{1-25}$ with G3BP1$_{NTF2}$ results in a buried surface area of ~571 Å^2. Structural superposition of the SARS-CoV-2 N$_{1-25}$-bound G3BP1$_{NTF2}$ with the previously reported apo form [75] gives a root-mean-square deviation (RMSD) of 0.43 Å over 225 aligned C*α*

atoms (Fig. 3-2 C), indicative of high similarity. The most divergent regions include the N-terminal tail and the loop connecting α2 and β2, both of which are involved in the interaction with the N_{1-25} peptide (Fig. 3-2 C). These observations highlight that the $G3BP1_{NTF2}$ domain adopts a preconfigured conformation for the interaction with N_{1-25} peptide.

Structural details of the SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$ interaction

The assembly of $G3BP1_{NTF2}$ with SARS-CoV-2 N_{1-25} is mediated by a network of hydrogen-bonding and van der Waals contacts (Fig. 3-3 A,B). Of note, the aromatic ring of SARS-CoV-2 N_{1-25} F17 inserts into the hydrophobic cavity formed by the side chains of $G3BP1_{NTF2}$ V11, F15, Q18, F33 and F124 (Fig. 3-3 A-D). The side chain of N_{1-25} I15 is positioned in parallel with that of N_{1-25} F17, engaging in nonpolar contacts with the side chains of $G3BP1_{NTF2}$ P6, L10 and V11 (Fig. 3-3 B,D). The side chains of N_{1-25} P13, R14 and T16 point away from $G3BP1_{NTF2}$, with N_{1-25} P13 surrounded by the side chains of $G3BP1_{NTF2}$ A121 and N122, the guanidinium of N_{1-25} R14 positioned within a distance for an electrostatic contact with the side chain carboxylate of $G3BP1_{NTF2}$ E14, and N_{1-25} T16 in proximity with the side chain of $G3BP1_{NTF2}$ V120 (Fig. 3-3 A,B). Along one side of the groove, the N_{1-25} I15-G18 segment pairs in parallel with the C-terminal end of $G3BP1_{NTF2}$ β5 (residues A121-Y125), involving both direct and water-mediated main chain hydrogen bonds (Fig. 3-3 A,B). On the other side of the groove, $G3BP1_{NTF2}$ Q18, R32 and K123 form water-mediated or direct hydrogen bonds with the backbone of N_{1-25} F17, G18

and G19, respectively (Fig. 3-3 A,B). Additional intermolecular interactions involve the van der Waals contacts between the backbone of N_{1-25} G18-S21 and the side chains of $G3BP1_{NTF2}$ F33, Q58, E117 and Y125 (Fig. 3-3 A,B).

To test the structural observation, we selected several key SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$ interacting residues for mutagenesis and performed ITC binding assays. Mutation of N_{1-25} I15 or T16 each to alanine reduced the binding by >10-fold and ~3-fold, respectively (Fig. 3-3 E and Fig. 3-8 A, B). Mutation of N_{1-25} F17 to alanine or asparagine abolished the interaction between SARS-CoV-2 N_{1-25} and $G3BP1_{NTF2}$ (Fig. 3-3 E and Fig. 3-8 C,D), supporting the role of these residues in the $G3BP1_{NTF2}$ interaction. Furthermore, we observed that introducing the N_{1-25} G18T mutation led to ~6-fold reduction of the binding affinity (Fig. 3-8 E), supporting the notion that surface complementarity underpins the specific interaction between SARS-CoV-2 N_{1-25} and $G3BP1_{NTF2}$. Conversely, we also observed impairment of the SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$ interaction by several $G3BP1_{NTF2}$ mutations: Introducing $G3BP1_{NTF2}$ V11A and F124A mutations reduced the binding by ~10- and ~20-fold, respectively, while introducing the $G3BP1_{NTF2}$ F15A mutation abolished the binding.

It is worth noting that introduction of the $G3BP1_{NTF2}$ Q18A or F33A mutations led to no appreciable change in binding (for Q18A) or even slightly enhanced the binding affinity (for F33A). The $G3BP1_{NTF2}$ Q18A mutation modestly reduced both the enthalpy (ΔH = -15.1 kcal/mol for

Q18A vs -17.6 kcal/mol for wild-type) and entropy (ΔS = -27.8 cal/mol/degree for Q18A vs -35.0 cal/mol/degree for wild-type) effects , whereas the $G3BP1_{NTF2}$ F33A mutation led to an increase of both enthalpy (ΔH = -21.3 kcal/mol for Q18A vs -17.6 kcal/mol for wild-type) and entropy (ΔS = -51.6 cal/mol/degree for F33Avs -35.0 cal/mol/degree for wild-type) effects. These enthalpy-entropy compensation effects, the origin of which is currently unclear, suggest a certain extent of structural plasticity of the N_{1-25} F17-binding pocket of the $G3BP1_{NTF2}$ domain.

Distinct intermolecular interaction mechanisms among G3BP1 complexes

Structural comparison of the $G3BP1_{NTF2}$ domain bound to SARS-CoV-2 N_{1-25} with that bound to a peptide derived from the SFV nsP3 protein (residues 449-471, $nsP3_{449-471}$) or a peptide derived from Caprin1 (residues 363-382, $Caprin1_{363-382}$) reveals different binding stoichiometry: In the SFV $nsP3_{449-471}$-$G3BP1_{NTF2}$ complex, two FGDF motifs of SFV nsP3 each bind to one $G3BP1_{NTF2}$ homodimer, thereby tethering the G3BP1 molecules into a poly-complex [63]; in contrast, in the SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$ and $Caprin1_{363-382}$-$G3BP1_{NTF2}$ complexes, the $G3BP1_{NTF2}$ domain binds to the N_{1-25} or $Caprin1_{363-382}$ peptide in a 2:2 binding mode (Fig. 3-4 A,B). Nevertheless, residues I15- G18 of N_{1-25}, L449-G452 of the $nsP3_{449-471}$ peptide and Y370-I373 of the $Caprin1_{363-382}$ peptide are anchored to the surface groove of the $G3BP1_{NTF2}$ domain in a similar fashion (Fig. 3-4 A-C and Fig. 3-9 A), with the aromatic rings of N_{1-25} F17, $nsP13_{449-471}$ F451 and $Caprin1_{363-382}$ F372 embraced by the same aromatic cage of the $G3BP1_{NTF2}$ domain (Fig. 3-4

A,B and Fig. 3-9 A,B). In addition, nsP3$_{449-471}$ L449 and Caprin$_{363-382}$ Y370 insert their side chains into the groove similarly as the corresponding N$_{1-25}$ I15 (Fig. 3-4 C,D). Beyond this region, nsP3$_{449-471}$ and Caprin1$_{363-382}$ interact with the G3BP1$_{NTF2}$ domain in a distinct mode than that of N$_{1-25}$ (Fig. 3-4 A, B). Unlike the N$_{1-25}$ peptide that occupies the entire groove, the nsP3$_{449-471}$ peptide exits from the surface groove at the D453 site to engage in a secondary contact with the α1-helix of G3BP1$_{NTF2}$ via helical packing (Fig. 3-4 A vs Fig. 3-2 B). Subsequently, the C-terminal segment of nsP3$_{449-471}$ extends into another G3BP1$_{NTF2}$ homodimer and presents residue F468 for a groove-insertion interaction like that of F452 (Fig. 3-4 A). Similar to nsP3$_{449-471}$, Caprin$_{363-382}$ forms a helical turn at its N-terminal end to interact with the α1-helix of the G3BP1$_{NTF2}$ domain (Fig. 3-4 B), albeit in an opposite direction (Fig. 3-4 D).

Together, our structural and sequence analyses of the G3BP1$_{NTF2}$-binding peptides reveal a ΦxF (Φ denotes a hydrophobic amino acid) consensus motif, in which an invariable phenylalanine (designated as P$_1$ site) and a bulky hydrophobic amino acid at the P$_{-2}$ site dominate the G3BP1$_{NTF2}$ binding via a groove-insertion mechanism (Fig. 3-4 C,D). In addition, the P$_2$ site is populated with a small amino acid (e.g. glycine), except for that in the Caprin1$_{363-382}$ peptide, which contains an isoleucine (I373), coinciding with its structural divergence from the SARS-CoV-2 N$_{1-25}$ and nsP3$_{449-471}$ peptides (Fig. 3-4 B,D). The diverse sequence composition at the ΦxF-flanking regions introduces secondary protein interactions underpinning

various binding outcomes of the G3BP1$_{NTF2}$-interaction partners. Along the line, a recent study reported that residues K36, K50, K59 and K64 of G3BP1 are subjected to ubiquitination in response to heat shock in cells, leading to disassembly and autophagy-independent degradation of SG[76]. Structural inspection of the SARS-CoV-2 N_{1-25}-G3BP1$_{NTF2}$, SFV nsP3$_{449-471}$-G3BP1$_{NTF2}$ and Caprin1$_{363-382}$-G3BP1$_{NTF2}$ complexes indicate that the potential ubiquitination sites of G3BP1 are largely exposed in these complexes (Fig. 3-9 C-E). One exception lies in G3BP1 K59 in the Caprin1$_{363-382}$ complex, which becomes partially shielded from solvent by the Caprin1$_{363-382}$ binding (Fig. 3-9 E). These observations imply that the Caprin1 binding may affect the ubiquitylation of G3BP1 differently than the SARS-CoV-2 N_{1-25} and SFV nsP3$_{449-471}$ bindings.

The SARS-CoV-2 N_{1-25}-binding groove is recurrently present in NTF2 domains

The NTF2 domain is present in a wide array of proteins from diverse species[77]. Sequence analysis of the G3BP1$_{NTF2}$ domain from selected model species (Fig. 3-10 A) or using the ConSurf server[37] reveals that the SARS-CoV-2 N_{1-25}-binding groove evolves more slowly than the surrounding area, with the N_{1-25} F17-binding site falling into the most conserved region (Fig. 3-5 A). Furthermore, structural survey of the 19 NTF2- containing proteins in the protein data bank reveals that the surface groove is evolutionarily preserved in the NTF2 domain of many other proteins. For instance, the NTF2 domain of nuclear transport factor 2-related export protein 1 (NXT1$_{NTF2}$) forms a surface groove in the

corresponding region to harbor the linker sequence connecting the leucine-rich repeat (LRR) and NTF2 domains of nuclear RNA export factor 1 (NXF1), resulting in formation of a heterodimer for mRNA export (Fig. 3-5 B)[78]. Likewise, human NTF2 protein and its counterpart from *C. parvum*, the founding members of the NTF2 family, also form a surface groove in the corresponding region (Fig. 3-5 C,D)[80,81]. As with the $G3BP1_{NTF2}$ domain, the NXT1 and NTF2 proteins form a hydrophobic cavity at the center of the groove, harboring NXF1 F362 and Ran L209, respectively (Fig. 3-5 B,C). These observations suggest groove insertion as a conserved interaction mechanism for the NTF2 domains. A similar surface groove is formed in the NTF2 domain of some other proteins, such as human NXF1, *C. jejuni* PgP2 and *S. cerevisiae* Bre5 (Fig. 5E-G) [65,82,83]. However, the electrostatic and steric properties of these grooves diverge substantially from that of the $G3BP1_{NTF2}$ domain (Fig. 3-5 E-G), implying different interaction mechanisms associated with the groove. Indeed, it was shown that the $NXF1_{NTF2}$ domain binds to the phenylalanine (F5) of an FG peptide derived from the nucleoporins (NUPs) repeat via a hydrophobic pocket distant to the surface groove corresponding to the N_{1-25}-binding site of the $G3BP1_{NTF2}$ domain (Fig. 3-5 E)[65].Taken together, these observations suggest that the SARS-CoV-2 N_{1-25}-binding groove serves as a recurrent structural feature of the NTF2 domain, which is subjected to electrostatic and steric fine tuning for diverse protein interaction behaviors.

DISCUSSION

SARS-CoV-2 N becomes an increasingly attractive drug target not only because it plays a multifaceted role in packing of viral genome, virion assembly and viral transcription[41,84], but also due to the fact that it helps break host defense through inhibition of the G3BP1/2-mediated SG formation in cells[44]. Development of an effective therapeutic strategy against this viral factor necessitates a detailed understanding of its molecular mechanism. Through combined structural and biochemical, this study establishes the structural basis for the interaction between SARS-CoV-2 N and the $G3BP1_{NTF2}$ domain, providing a framework for the development of novel therapeutic strategies against COVID-19 and related viral infection.

Molecular basis for the inhibition of G3BP1-mediated SG formation by SARS-CoV- 2 N protein

The G3BP1-mediated SG formation is underpinned by a heterologous network of interactions involving the protein contacts mediated by the $G3BP1_{NTF2}$ domain and the RNA bindings mediated by the $G3BP1_{RRM}$ domain [47-52]. As proposed in a network-based model[55], the interaction of $G3BP1_{NTF2}$ with the 'bridge' protein Caprin1, which possesses both protein and RNA binding-modules, creates multiple valences for the interaction network, thereby promoting the SG assembly[56]. In contrast, the interaction

with USP10, which is classified as ‘valence cap’ due to lack of multivalent interaction module, serves to halt the propagation of the interaction network inside the SG[57]. Surprisingly, theinteraction of SARS-CoV-2 N with G3BP1 leads to inhibition of G3BP1-mediated liquid-liquid phase separation *in vitro* and in cells [44,74], even though the former contains both protein and RNA interaction modules.

Through fragment-based ITC binding assays, this study identified that the interaction between SARS-CoV-2 N and G3BP1 is primarily mediated by $N_{1\text{-}25}$ and the $G3BP1_{NTF2}$ domain. In addition, the C-terminal regions of SARS-CoV-2 N, including IDR2 and/or CTD, contribute to the interaction by reducing the entropic cost of the complex formation. The fact that both the N- and C-terminal regions of SARS-CoV-2 N contribute to the G3BP1 binding suggests a multivalent engagement between SARS-CoV-2 N and G3BP1, by which the C-terminal region of SARS-CoV-2 N-mediated secondary interaction may interfere with the SG-nucleating activity of G3BP1. In support of this notion, a recent study showed that the SARS-CoV-2 $N_{1\text{-}175}$ fragment exhibits a reduced SG inhibition activity than full-length SARS-CoV-2 N[43]. In addition, given the fact that the SARS-CoV-2 N and Caprin1 interact with G3BP1 in a mutually exclusive manner, it is conceivable that the interaction between SARS-CoV-2 $N_{1\text{-}25}$ and $G3BP1_{NTF2}$ also serves to inhibit the association of G3BP1 with Caprin1, providing another mechanism in attenuating the SG formation. A detailed mechanism by which SARS-CoV-2 N protein inhibits G3BP1-mediated SG formation

remains to be investigated.

A groove-insertion mechanism dictating the SARS-CoV-2 N_{1-25}-G3BP1$_{NTF2}$ interaction

The structure of the SARS-CoV-2 N_{1-25}-G3BP1$_{NTF2}$ complex reveals strong surface complementarity between residues 12-22 of SARS-CoV-2 N_{1-25} and the surface groove of G3BP1$_{NTF2}$. Most notably, the P_1-F17 of SARS-CoV-2 N_{1-25} inserts its aromatic ring into a hydrophobic pocket at the center of the surface groove, which is flanked by a parallel sidechain insertion of P_{-2}-I15. This dual groove-insertion mode is reminiscent of what was previously observed for the complexes of G3BP1$_{NTF2}$ with SFV nsP3 and Caprin1, attesting the "ΦxF" motif as a primary determinant for the G3BP1$_{NTF2}$-mediated protein interaction. In addition, the embedding of N_{1-25} P_2-G18 in a narrow region of the surface groove permits a parallel pairing between N_{1-25} I15-G18 and G3BP1$_{NTF2}$ β5- strand, in line with the previous observation that a P_2-glycine is favored by the G3BP1$_{NTF2}$ domain [58,63].

Unlike the Caprin1 and SFV nsP3 peptides that diverge their ΦxF-flanking regions out of the surface groove to engage helical packing with the α1-helix of G3BP1$_{NTF2}$, the ΦxF- flanking flanking regions of SARS-CoV-2 N_{1-25} peptide remain bound to the groove. Through surface complementarity, residues P13-R14 and G19-S21 of N_{1-25} are anchoredto the two ends of the groove, engaging van der Waals and/or hydrogen-bonding interactions with G3BP1$_{NTF2}$, respectively. These secondary binding sites in the SARS- CoV-2 N_{1-25}-G3BP1$_{NTF2}$ complex may serve as attractive targets for the

development of therapeutic agents that selectively inhibit the SARS-CoV-2 N□G3BP1 association.

NTF2 domain, a platform for diverse protein-protein interactions

The NTF2 domain represents an evolutionarily divergent protein interaction module involved in various cellular activities, such as SG formation[44,47-52,74] and mRNA transport[79]. This study, through structural survey of the NTF2 domains from diverse proteins, uncovers that the surface groove of $G3BP1_{NTF2}$ represents a recurrent feature of the NTF2 domains. Common to many of these NTF2 domains is the formation of a hydrophobic cavity at the center of the groove, which provides a primary docking site for a bulky, hydrophobic side chain from host or viral factors. On the other hand, the NTF2-interacting proteins, such as SARS-COV-2 N, Caprin1 and NXF1, often extend beyond the hydrophobic pocket for secondary contacts, which presumably contributes to their distinct binding affinity and specificity. In this context, the structure of the SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$ complex not only provides a framework for understanding how SARS- CoV N and SARS-CoV-2 N have evolved to target the G3BP1/2, but also provides a basis for identification of other host factors that can potentially be targeted by the SARS- CoV-2 N.

FIGURES

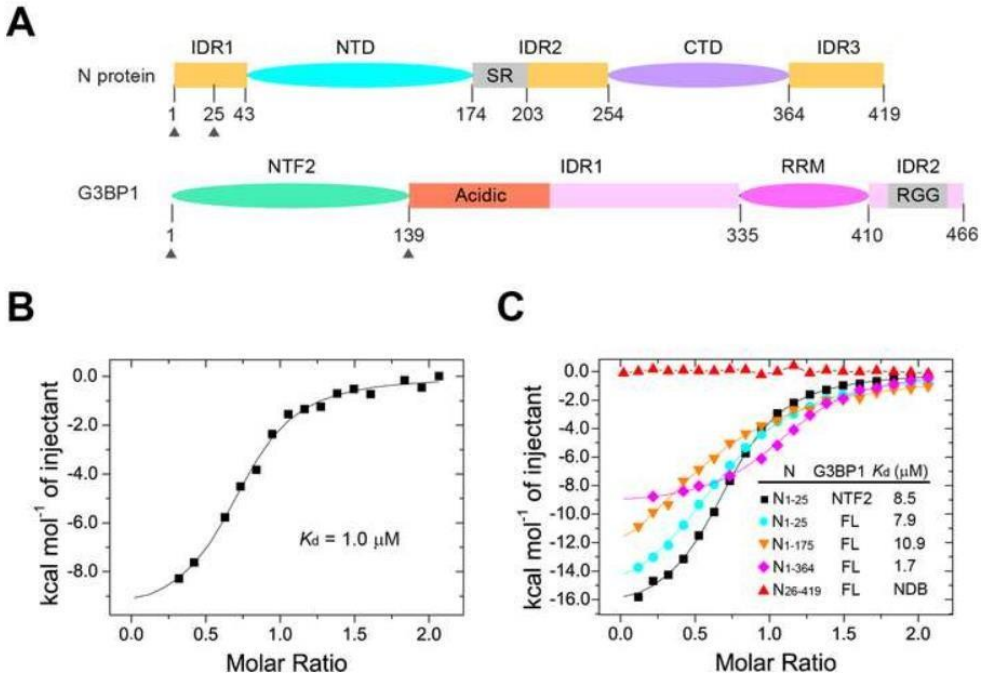

Figure 3-1. Biochemical characterization of the interaction between SARS-CoV-2 N and G3BP1. (A) Domain architecture of SARS-CoV-2 N and G3BP1, with individual domains color coded. The SR-rich region within the IDR2 of SRS-CoV-2 N, the acidic region within the IDR1 of G3BP1 and the RG-rich region (RGG) within the IDR2 of G3BP1 are labeled. The protein fragments (residues 1-25 of SARS-CoV-2 N protein and 1-139 of G3BP1) used for crystallographic study are delimited by arrows. (B) ITC binding assays for full-length SARS-CoV-2 N and G3BP1. (C) ITC binding assays for the truncated fragments of SARS-CoV-2 N and G3BP1.FL, full length; NDB, no detectable binding.

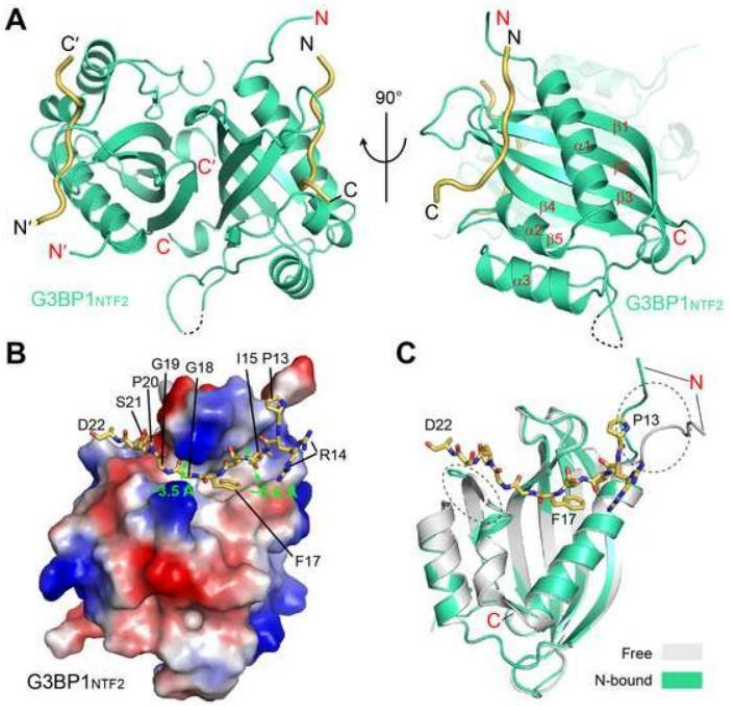

Figure 3-2. Structural overview of the SARS-CoV-2 N_{1-25}□$G3BP1_{NTF2}$ complex. (A) Orthogonal views of the SARS-CoV-2 N_{1-25}□$G3BP1_{NTF2}$ complex, with SARS-CoV-2 N_{1-25} colored in yellow orange and $G3BP1_{NTF2}$ colored in green. The traceable N- and C-termini of SARS-CoV- 2 N_{1-25} or $G3BP1_{NTF2}$ are labeled with 'N' and 'C', respectively. The region (residue S47) with untraceable electron density is shown in dashed line. (B) Electrostatic surface view of the $G3BP1_{NTF2}$ domain bound to the SARS-CoV-2 N_{1-25} peptide (stick representation). The widths of two distinct groove regions of $G3BP1_{NTF2}$ are marked. For clarity, only one monomer of the $G3BP1_{NTF2}$ homodimer is shown. (C) Structural overlay of the $G3BP1_{NTF2}$ domain, free (grey) and in complex with SARS-CoV-2 N_{1-25} (green). The two structurally divergent regions are circled with dotted lines.

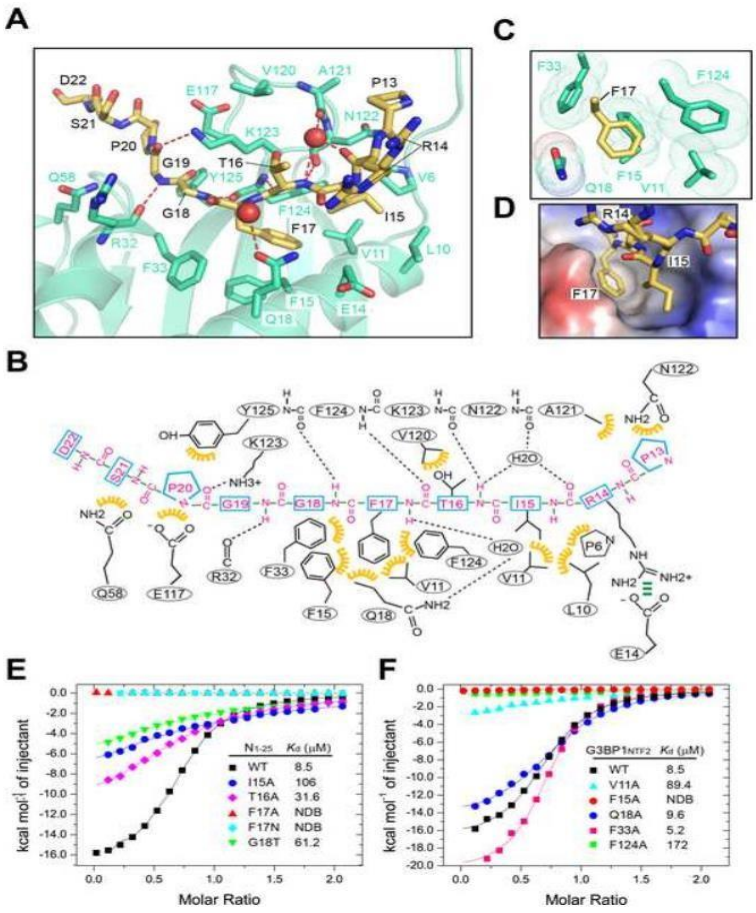

Figure 3-3. Structural and biochemical characterizations of the SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$ interaction. (A) Close-up view of the SARS-CoV-2 N_{1-25} $G3BP1_{NTF2}$ interaction. The interacting residues are shown in stick representation. The hydrogen bonds are shown as dashed lines. The water molecules are shown as red spheres. (B) Schematic view of the SARS- CoV-2 N_{1-25}-$G3BP1_{NTF2}$ interaction. Hydrogen bonds and electrostatic interactions are indicated by black and green dashed lines, respectively, and van der Waals contacts are indicated by yellow gears. (C) Close-up view of the hydrophobic pocket harboring residue F17 of SARS-CoV-2N_{1-25}. The van der Waals radii of the $G3BP1_{NTF2}$ residues are shown in dots. (D) Close-up view of the electrostatic surface of $G3BP1_{NTF2}$ harboring the side chains of I15 and F17 of SARS-CoV-2 N_{1-25}. (E) ITC binding assays for wild-type $G3BP1_{NTF2}$ titrated with wild-type or mutant SARS-CoV-2 N_{1-25} peptide. (F) ITC binding assays for wild-type or mutant $G3BP1_{NTF2}$ titrated with the wild-type SARS-CoV-2 N_{1-25} peptide. NDB, no detectable binding.

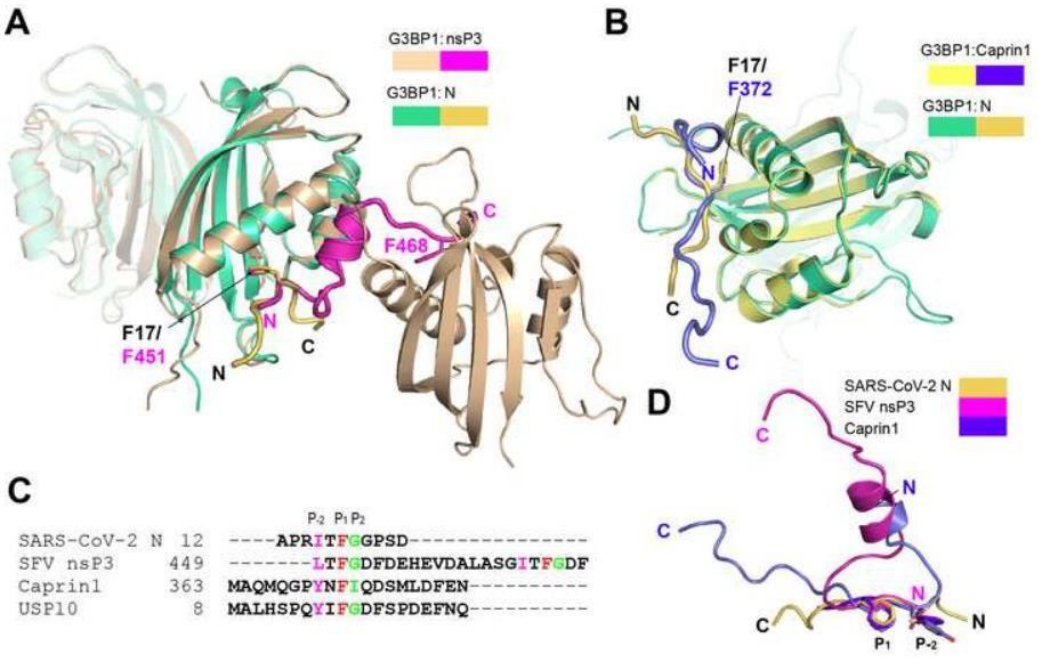

Figure 3-4. Structural comparison of SARS-CoV-2 N_{1-25} G3BP1$_{NTF2}$ with other G3BP1$_{NTF2}$ complexes. (A) Structural overlay between the SARS-CoV-2 N_{1-25} (yellow orange)-G3BP1$_{NTF2}$ (green) and the SFV nsP3$_{449-471}$ (magenta)G3BP1$_{NTF2}$ (wheat) complex. The phenylalanine residues inserting into the hydrophobic pocket of G3BP1$_{NTF2}$ are labeled and shown in stick representation. (B) Structural overlay between the SARS-CoV-2 N_{1-25} (yellow orange)-G3BP1$_{NTF2}$ (green) and the Caprin1$_{363-382}$ (slate) -G3BP1$_{NTF2}$ (yellow) complex. The phenylalanine residues inserting into the hydrophobic pocket of G3BP1$_{NTF2}$ are labeled and shown in stick representation. (C) Sequence alignment of the G3BP1$_{NTF2}$-interacting peptides with the P_{-2}, P_1 and P_2 sites colored in magenta, red and green, respectively. (D) Structural alignment of the G3BP1$_{NTF2}$-interacting peptides, with the P_{-2} and P_1 sites, as well as the N- and C-termini, labeled.

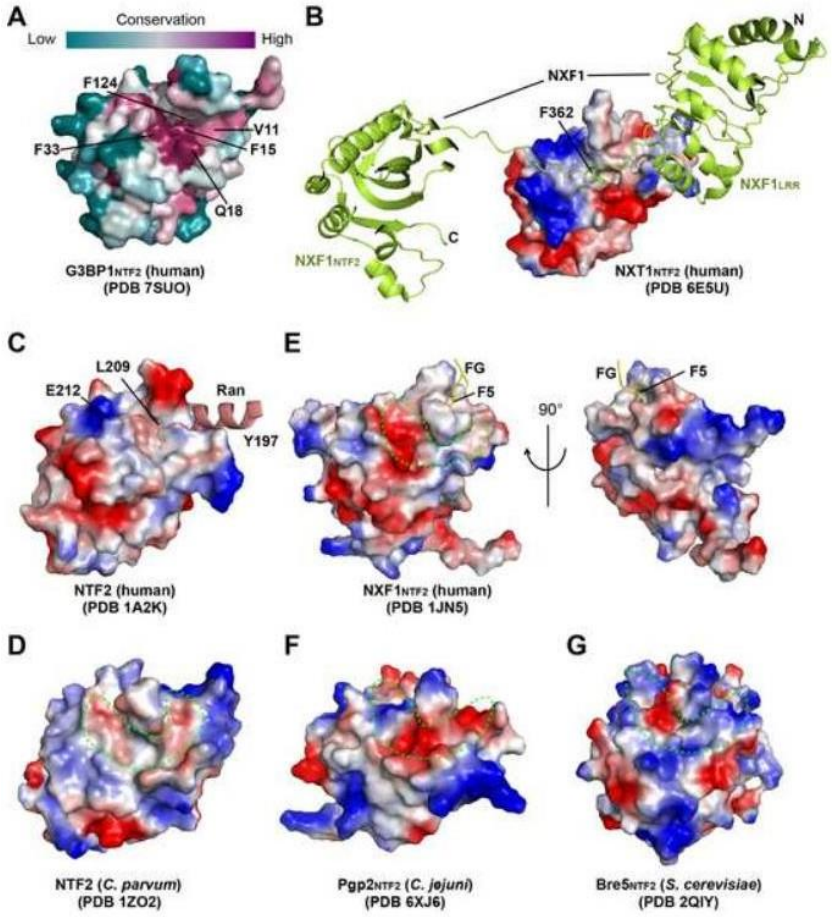

Figure 3-5. Structural evolution of the NTF2 domains. (A) Color-coded sequence conservation of the G3BP1$_{NTF2}$ domain, analyzed using the ConSurf server (https://consurf.tau.ac.il). (B) Electrostatic surface of the NXT1$_{NTF2}$ domain bound to the LRR (NXF1$_{LRR}$)-NTF2 (NXF1$_{NTF2}$) domain linker of the NXF1 protein (limon), with the side chain of NXF1 F362 shown in stick representation. (C) Electrostatic surface of human NTF2 domain bound to residues 197-212 of one Ran molecule (light pink), with the side chain of Ran L209 shown in stick representation. For clarity, the rest of Ran structure is not shown. (D-G) Electrostatic surface of *C. parvum* NTF2 (D), human NXF1$_{NTF2}$ (E), *C. jejuni* Pgp2$_{NTF2}$ (F) and *S. cerevisiae* Bre5$_{NTF2}$ (G), with the individual surface grooves circled by dotting lines. Note that human NXF1$_{NTF2}$ binds to the FG peptide through a hydrophobic pocket positioned separately from the groove.

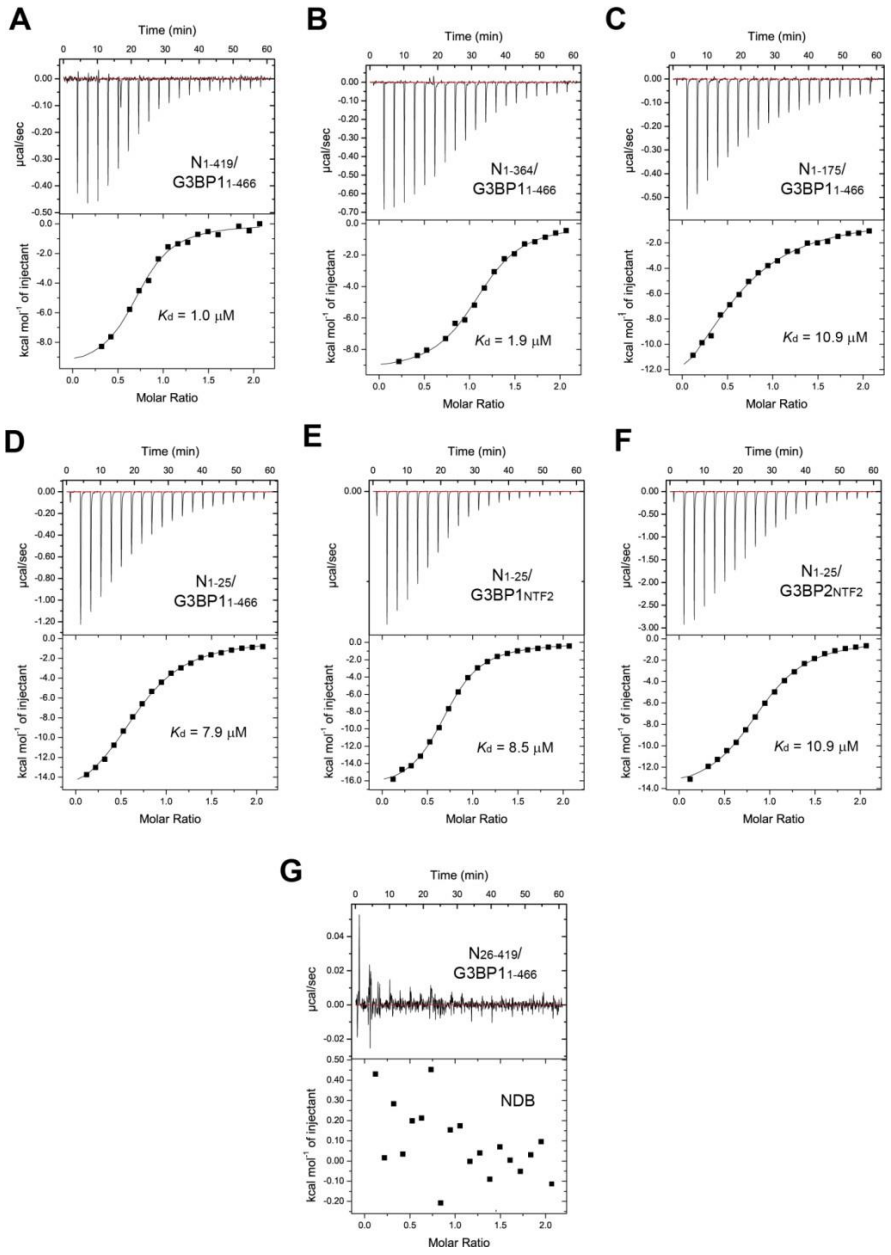

Figure 3-6. ITC binding assays for the interaction between SARS-CoV-2 N and G3BP1. (A) ITC binding assay for full-length G3BP1 titrated with full-length SARS-CoV-2 N. (B) ITC binding assay for full-length G3BP1 titrated with the SARS-CoV-2 N_{1-364} fragment. (C) ITC binding assay for full-length G3BP1 titrated with the SARS-CoV-2 N_{1-175} fragment. (D) ITC binding assay for full- length G3BP1 titrated with the SARS-CoV-2 N_{1-25} peptide. (E) ITC binding assay for the $G3BP1_{NTF2}$ domain titrated with the SARS-CoV-2 N_{1-25} peptide. (F) ITC binding assay for the $G3BP2_{NTF2}$ domain titrated with the SARS-CoV-2 N_{1-25} peptide. (G) ITC binding assay for full- length G3BP1 titrated with the SARS-CoV-2 N_{26-419} fragment. NDB, no detectable binding.

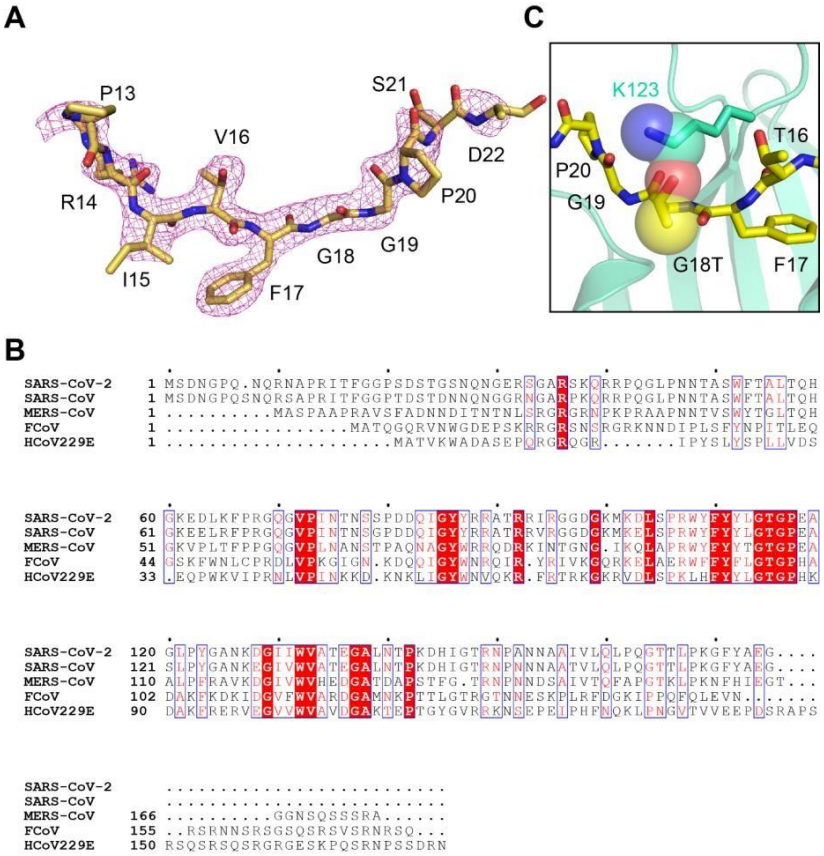

Figure 3-7. Structural and sequence analysis of the SARS-CoV-2 N_{1-25} peptide. (A) Fo-Fc omit map (violet) of the SARS-CoV-2 N_{1-25} peptide bound to the $G3BP1_{NTF2}$ domain at 2.0 contour level. The traceable residues of the SARS-CoV-2 N_{1-25} peptide are labeled and shown in stick representation. (B) Sequence alignment of the N proteins from various coronaviruses. Strictly conserved residues are colored white in red background. Similar residues are colored in red. (C) Structural model of the G18T-substituted SARS-CoV-2 N_{1-25} bound to the $G3BP1_{NTF2}$ domain. The van der Waals radii of N_{1-25} T18 and $G3BP1_{NTF2}$ K123 are shown in spheres, which indicate a steric clash between the two residues.

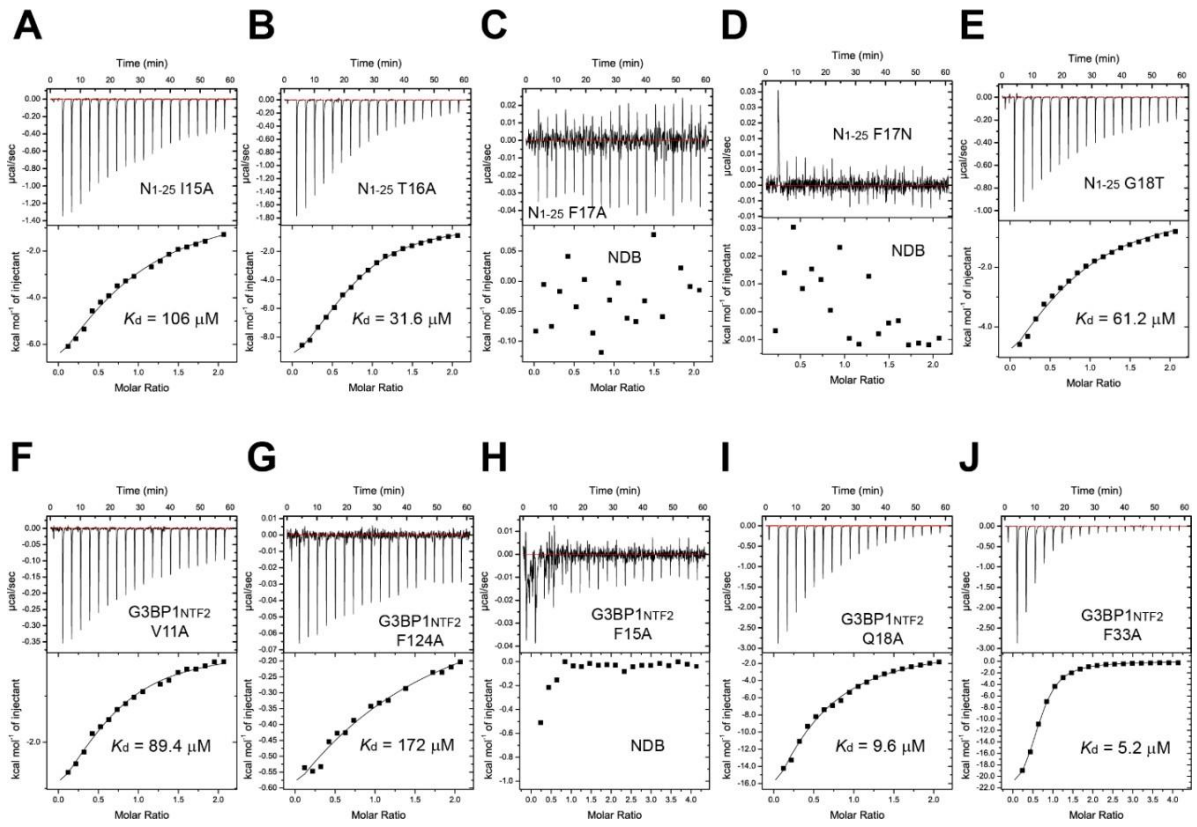

Figure 3-8. ITC binding assays of SARS-CoV-2 N_{1-25} and $G3BP1_{NTF2}$ mutants. (A-E) ITC binding assays of $G3BP1_{NTF2}$ titrated with SARS-CoV-2 N_{1-25} mutants (A) I15A, (B) T16A, (C) F17A, (D) F17N, and (E) G18T. (F-J)) ITC binding assays of $G3BP1_{NTF2}$ mutants (F) V11A, (G) F124A, (H) F15A, (I) Q18A, and (J) F33A, titrated with the SARS-CoV-2 N_{1-25} peptide. NDB, no detectable binding.

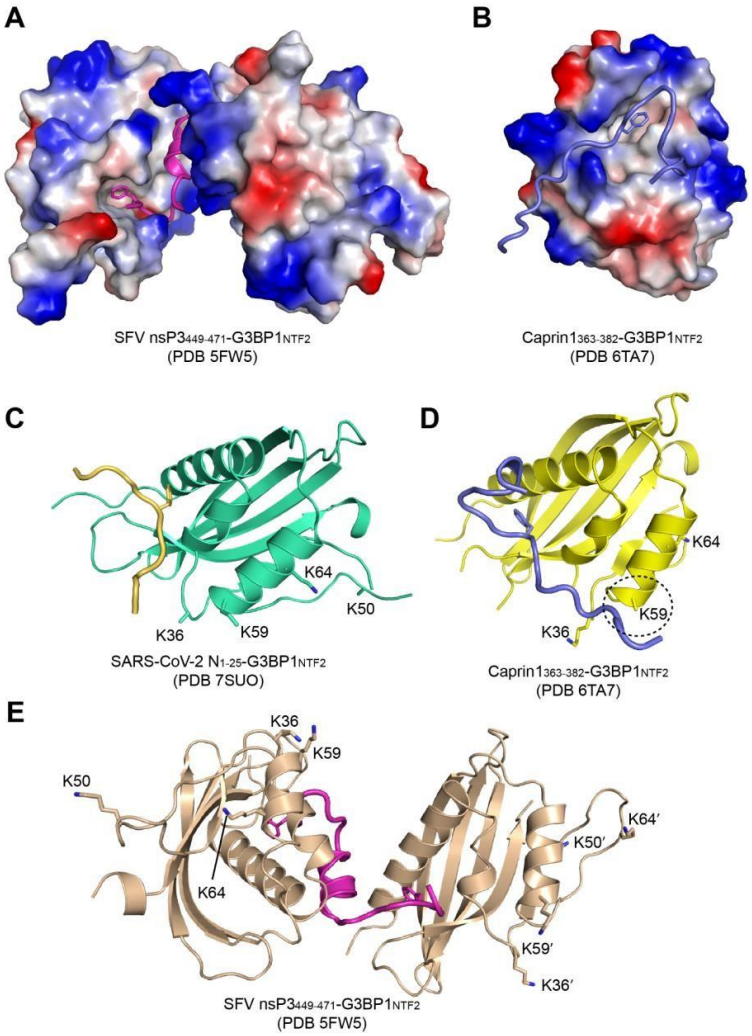

Figure 3-9. Electrostatic surface views of two representative G3BP1$_{NTF2}$ complexes. (A) Electrostatic surface of G3BP1$_{NTF2}$ bound to the nsP3$_{449-471}$ peptide (magenta). (B) Electrostatic surface of G3BP1$_{NTF2}$ bound to the Caprin1$_{363-382}$ peptide (slate). (C-E) Four of the potential ubiquitination sites of G3BP1$_{NTF2}$, K36, K50, K59 and K64, are shown in stick representation in the structure of SARS-CoV-2 N$_{1-25}$-G3BP1$_{NTF2}$ (C), Caprin1$_{363-382}$-G3BP1$_{NTF2}$ (D), and nsP3$_{449-471}$-G3BP1$_{NTF2}$ (E) complexes. Residue K59 of G3BP1 in the Caprin1$_{363-382}$-G3BP1$_{NTF2}$, the solvent access of which is partially blocked by the Caprin1$_{363-382}$ peptide, is circled with dottedline.

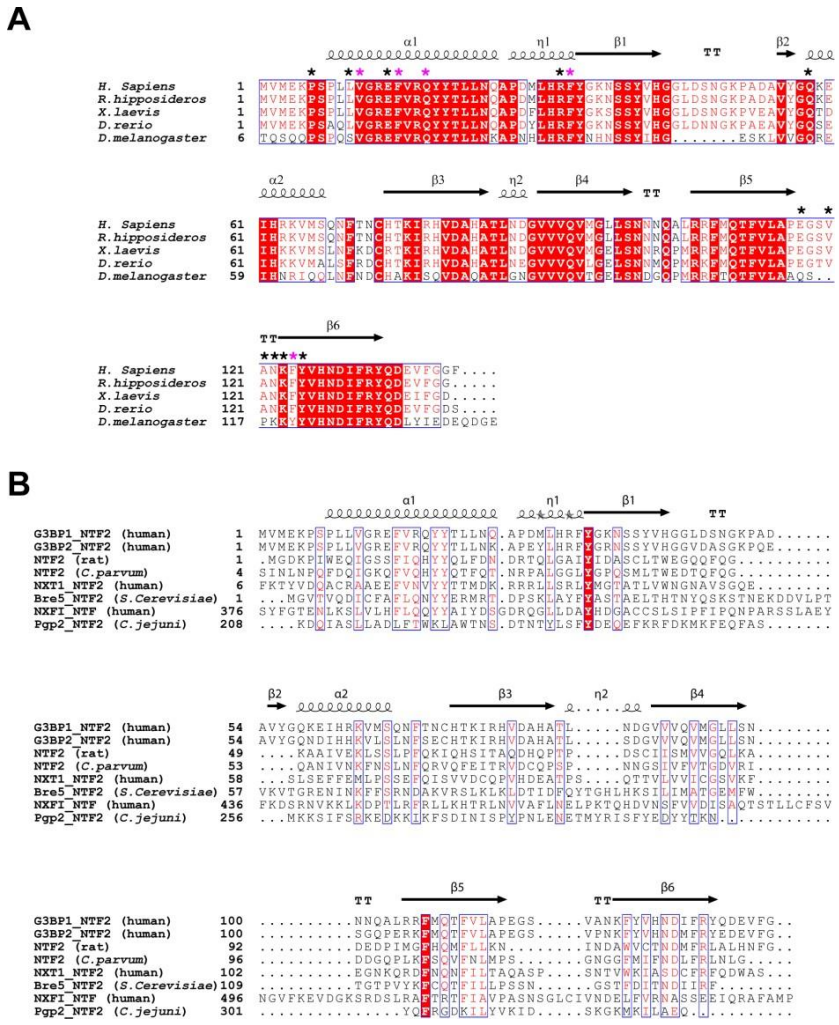

Figure 3-10. Sequence analysis of the NTF2 domains. (A) Sequence alignment of G3BP1$_{NTF2}$ domain from various species, with the corresponding secondary structures of the human G3BP1$_{NTF2}$ domain indicated on top. The SARS-CoV-2 N_{1-25} F17-binding pockets are marked by magenta asterisks, and the rest of the SARS-CoV-2 N_{1-25}-interacting residues are marked by black asterisks. (B) Sequence alignment of the NTF2 domains form various proteins, related to Figure 5. Strictly conserved residues are colored in white in red background. Similar residues are colored in red.

Table 2. Data collection and refinement statistics of the SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$ complex

	SARS-CoV-2 N_{1-25}-$G3BP1_{NTF2}$(PDB 7SUO)
Data collection	
Space group	*P* $2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	42.7, 51.2, 153.2
α,β,γ (^{0})	90, 90, 90
Resolution (Å)	42.53-2.35 (2.43-2.35)[a]
R_{merge}	0.150 (0.660)
$I/\sigma(I)$	10.5 (2.0)
$CC_{1/2}$	0.99 (0.745)
Completeness (%)	99.97 (100.00)
Redundancy	7.9 (6.1)
Refinement	
No. reflections	14,658 (1429)
R_{work} / R_{free}	19.7/24.6 (26.0/32.5)
No. atoms	
G3BP1	2206
N Protein	143
Water	87
B factors (Å^2)	
G3BP1	42.8
N Protein	62.2
Water	44.1
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (^{0})	0.699
Ramachandran (%)	
Favored	95
Allowed	5
Outliers	0

[a] Values in parentheses are for highest-resolution shell.The data were collected from a single crystal.

Table3. Summary of ITC binding parameters for SARS CoV-2 N_{1-25}-G3BP1 $_{NTF2}$ complex

G3BP1	SARS-CoV-2 N protein	K_d (μM)	N value	ΔH (kcal/mol)	ΔS (cal/mol/degree)
G3BP1, FL	N, FL	1.0 ± 0.3	0.74 ± 0.02	-9.9 ± 0.2	-5.9 ± 0.1
G3BP1, FL	N_{1-25}	7.9 ± 0.7	0.70 ± 0.01	-17.6 ± 0.3	-35.4 ± 1.2
G3BP1, FL	N_{26-419}	NDB			
G3BP1, FL	N_{1-175}, WT	10.9 ± 0.1	0.62 ± 0.01	-18.2 ± 0.7	-38.2 ± 2.3
G3BP1, FL	N_{1-364}, WT	1.9 ± 0.2	1.12 ± 0.01	-9.4 ± 0.1	-5.4 ± 0.2
$G3BP1_{NTF2}$, WT	N_{1-25}, WT	8.5 ± 3.3	0.74 ± 0.04	-17.4 ± 0.3	-35.0 ± 1.7
$G3BP2_{NTF2}$, WT	N_{1-25}, WT	10.9 ± 2.4	0.79 ± 0.16	-16.1 ± 2.5	-31.2± 8.6
$G3BP1_{NTF2}$, WT	N_{1-25}, I15A	106 ± 12	0.74 ± 0.01	-15.8 ± 1.5	-34.8 ± 5.3
$G3BP1_{NTF2}$, WT	N_{1-25}, T16A	31.6 ± 1.5	0.73 ± 0.03	-12.7 ± 0.5	-22.2 ± 1.4
$G3BP1_{NTF2}$, WT	N_{1-25}, F17A	NDB			
$G3BP1_{NTF2}$, WT	N_{1-25}, F17N	NDB			
$G3BP1_{NTF2}$, WT	N_{1-25}, G18T	61.2 ± 13.1	0.65 ± 0.08	-9.2 ± 0.4	-11.8 ± 1.7
$G3BP1_{NTF2}$, V11A$^{\#}$	N_{1-25}, WT	89.4 ± 12.3	0.47 ± 0.02	-15.8 ± 2.5	-34.4 ± 8.6
$G3BP1_{NTF2}$, F15A	N_{1-25}, WT	NDB			
$G3BP1_{NTF2}$, Q18A	N_{1-25}, WT	9.6 ± 0.4	0.79 ± 0.01	-15.1 ± 0.2	-27.8 ± 0.7
$G3BP1_{NTF2}$, F33A	N_{1-25}, WT	5.2 ± 1.1	0.69 ± 0.08	-21.3 ± 2.3	-51.6 ± 7.9
$G3BP1_{NTF2}$, F124A	N_{1-25}, WT	172 ± 76	1.00*	-2.2 ± 0.4	9.9 ± 2.1

CHAPTER 4

FUTURE PERSPECTIVES

Although the outbreak of the Covid-19 pandemic has led to an increase in coronavirus- related research, emergence of new variants demands for continuous research in order to develop effective vaccines and novel therapeutic strategies that remain effective against upcoming variants. The global efforts put together to deduce the structures of various components of replication machinery, For e.g. SARS CoV-2 NSP 7-8, SARS- CoV-2 Holo NSP 7-8-12 and SARS-CoV-2 RNA bound NSP 7-8-12 as well as the recent SARS-CoV-2 NSP 7-8-12-13-9-14-10 complex highlighted the roles of these proteins during replication and mechanism of how they may work in conjunction. It is poorly understood as to which different subcomplexes that may be formed during replication and transcription and if they are made up of similar or different set of proteins. The complete repertoire of the host cell factors that may assist in coronavirus replication cycle and participate in formation of RTC remains unknown. Spatial segregation of viral replication machinery in virus induced membranous organelles is an important requisite for successful propagation of virus, yet the exact spatial locations in the cell where the replication takes place during early phase of infection is not known.

In the path downstream to replication organelle formation, key questions that still need answers is determination of possible roles for the intermediates that are released after polyprotein processing. A recent study

pointed out traces of specific polyprotein such as NSP 10-11, NSP 9-10, NSP 7-8-9 by Native Mass spectrometry[18]. The authors point out that even though cleavage sites may be accessible in certain cases yet there is slow processing and prolonged presence of these intermediates which was observed that provides hints that they may have functional implications in regulating the RTC machinery which needs further investigation.

Another key question that needs to be explored is how RNA synthesis is primed on its templates, what is the primase enzyme, and how exactly the primer may be translocated to the RdRp active site to initiate polymerization. In the case of SARS CoV and SARS CoV-2, the NSP 7-8 complexes, appear to assume a tetrameric former, rather than the hexadecamer as previously claimed[17]. This observation argues against the notion that SARS-CoV NSP8 serves as a primase for viral RNA synthesis. In the case of feline intestine peritonitis virus (FIPV) and transmissible-gastroenteritis virus (TGEV), trimeric assembly of NSP 7-8 complex has been reported (FIPV: Heterotrimer:83% and heterodimer: 17%) and a third olimerization pattern is observed for NSP7-8 complex in HCoV- 229E and porcine epidemic diarrhea virus (PEDV), where co-presence of Heterodimer, Heterotrimer and Heterotetramer have been reported (HCoV-229E: 20,12,69%; PEDV :52,6,40%)[18]. The possible reasons for the existence of different NSP 7-8 complexes warrants further investigation, whether they have any functional relevance in coronavirus replication or if they hold any regulatory functions.

Viral-human interactome maps have revealed the different SARS CoV-2

accessory proteins, including both non-structural and structural proteins, which interact with human proteins to either facilitate viral entry into host, hijack proteins that could help them evade innate immune response or could facilitate their replication and survival in the hosts.

Although a large number of host proteins are now known that interact with the different SARS CoV-2 viral proteins, very little structural information is available of such complexes. Knowledge of structural basis of different host and viral protein interaction may help in design of drugs that target multiple or conserved sites of viral and host protein interaction so that it could offer advantage over drugs that target proteins at single site since they are vulnerable to the evolution of drug resistance as a result of viral mutations. So it may be more beneficial to develop drugs which may be crucial for viral lifecycle. It has been shown that co-expression of competitive peptide inhibitors designed specifically to disrupt the ITFG-G3BP1 interaction reduced viral proliferation[75], indicating that the N-G3BP1 interaction is important for viral replication. Our study uncovered the biochemical and structural basis for SARS-CoV-2 N-G3BP1 interaction and revealed that N_{1-25} may serve as a primary determinant of interaction with G3BP1 $_{NTF2}$ and flanking amino acids my serve as secondary determinants for this interaction. Furthermore, our ITC results revealed that N protein CTD or IDR2 may also interact with G3BP1 pointing towards a multivalent interaction mode that might prevail between the full-length N and G3BP1 proteins. Further experiments are needed to uncover the complete mechanism as to how Nucleocapsid protein interaction with G3BP1

disassembles stress granules and if interaction between N protein CTD and or IDR2 with G3BP1 may play a role in lowering the G3BP1's SG-nucleating activity or lowering of its valence. The network theory may have assigned macromolecules their roles in regulating the network of protein-protein or protein-RNA interaction by serving as node, bridge, or cap depending on their valence, yet insights into the relative importance of these proteins and how they relate to one another to govern assembly has been lacking. It is rather intriguing that N protein has properties similar to Caprin-1 but still behaves as cap instead of a node as seen with Caprin-1, in the SG regulation network.

Whether Caprin-1 causes a conformational shift on binding to GB3P that promotes LLPS through IDR's of G3BP and whether USP10 binding to G3BP may have contrasting effects by promoting soluble conformation of G3BP which may induce SG disassembly needs future investigation.

It has been postulated that G3BP1 mediated LLPS is regulated via a molecular switch involving interaction between its three IDR's. Whether N protein mediated interaction with G3BP1 $_{NTF2}$ locks G3BP1 in a closed conformation wherein IDR3 interacts with IDR1 and lowers the RNA binding capacity beyond a threshold concentration in order to keep the SG network intact warrants further investigation.

Given the fact that the SARS-CoV-2 N and Caprin1 interact with G3BP1 in a mutually exclusive manner, it is conceivable that the interaction between SARS-CoV-2 $N_{1–25}$ and G3BP1$_{NTF2}$ also serves to inhibit the association of G3BP1 with Caprin1, which may provide a mechanism in attenuating the

SG formation. On the other hand, other studies point out how G3BP mutant incapable of binding to Caprin-1 or USP 10 remains fully SG competent[57, 58]; in such a case, it is worth to assume that instead of Caprin-1 or USP 10, UBAP2L which is the high affinity binding partner of G3BP may play a critical role, as deletion of the (RNA- binding) RGG motif of UBAP2L abolishes SG formation upon arsenite stress. Notably, UBAP2L contains a self-associative IDR that is likely critical for SG formation. Future studies are needed to investigate if N mediated SG suppression could be modulated because of asymmetric dimethylation of a single arginine residue (R95) in the N-protein RNA binding domain due to G3BP1 binding. It has been shown that N protein interacts with RdRP, whether SG disassembly pathway is linked to the process of N protein recruiting the transcription machinery awaits further studies[65]. Reports also suggest that SARS-CoV-2 Membrane (M) protein interacts with N protein[72]; whether M-N-RdRp complex is tethered to the membrane during replication needs further experimentation.

Indeed, a key gap in knowledge permitting us to understand the molecular basis of SG assembly and dynamics is lack of a conceptual framework built from biophysical principles. As there will be improvements in studying such biophysical principles combined with improved structural techniques, may make it easier to understand the stress granule physics and solve the structures of elusive proteins with their host interaction partners involved in LLPS. This could provide the foundation for development of new antiviral therapies and vaccines.

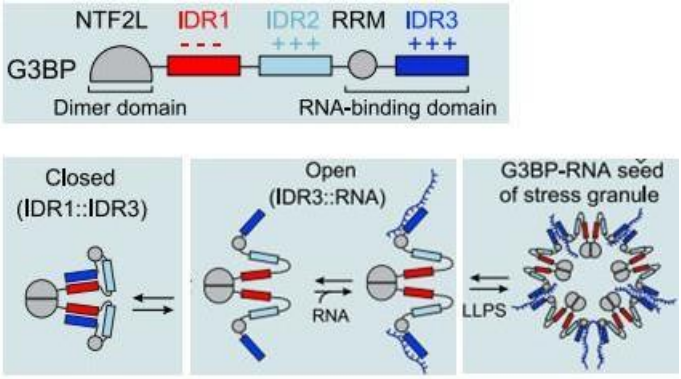

Figure 4-1: The central node G3BP encodes a molecular switch that regulates RNA-dependent LLPS. Interplay between 3 distinct IDRs in G3BP tunes the intrinsic propensity for LLPS [52].

REFERENCES

1. Wang H., Murray C.,Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, Lancet 2022; 399: 1513–36

2. Cui Jie, Li Fang, Shi Zheng-Li, Origin and evolution of pathogenic coronaviruses, Nature Reviews Microbiology, Volume 17, March 2019 ,**181.**

3. V'kovski Philip, Kratzel Annika, Steiner Silvio, Theil Volker, Coronavirus biology and replication: Implications for SARS CoV-2, Nature reviews Microbiology, Volume 19, March 2021

4. Malone Brandon, Urakovya Nadya, Snijder J. Eric, Campbell A Elizabethl; Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design, Nature Reviews, Molecular cell Biology, volume 23 ,January 2022.

5. Wang,M., Cao,R., Zhang,L., Yang,X., Liu,J., Xu,M., Shi,Z., Hu,Z.,Zhong,W. andXiao,G. (2020) Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.*, **30**, 269–271.

6. Zhu,N., Zhang,D., Wang,W., Li,X., Yang,B., Song,J., Zhao,X., Huang,B., Shi,W.,Lu,R. *et al.* (2020) A novel coronavirus from patients with pneumonia in China,2019. *N. Engl. J. Med.*, **382**, 727–733.

7. Wu,A., Peng,Y., Huang,B., Ding,X., Wang,X., Niu,P., Meng,J., Zhu,Z., Zhang,Z., Wang,J. *et al.* (2020) Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell host & microbe*, **27**, 325–328.

8. te Velthuis,A.J., Arnold,J.J., Cameron,C.E., van den Worm,S.H. and Snijder,E.J.(2010) The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Res.*, **38**, 203–214.

9. Posthuma,C.C., Te Velthuis,A.J.W. and Snijder,E.J. (2017) Nidovirus RNA polymerases: complex enzymes handling exceptional RNA genomes. *Virus Res.*,**234**, 58–73.

10. Subissi,L., Posthuma,C.C., Collet,A., Zevenhoven-Dobbe,J.C., Gorbalenya,A.E.,Decroly,E., Snijder,E.J., Canard,B. and Imbert,I. (2014) One severe acute respiratory syndrome coronavirus protein

complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E3900–3909.

11. Kirchdoerfer,R.N. and Ward,A.B. (2019) Structure of the SARS-CoV nsp12polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.*, **10**, 2342.

12. Chen,J., Malone,B., Llewellyn,E., Grasso,M., Shelton,P.M.M., Olinares,P.D.B., Maruthi,K., Eng,E.T., Vatandaslar,H., Chait,B.T. *et al.* (2020) Structural basis forhelicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell*, **182**,1560–1573.

13. Gao,Y., Yan,L., Huang,Y., Liu,F., Zhao,Y., Cao,L., Wang,T., Sun,Q., Ming,Z., Zhang,L. *et al.* (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, **368**, 779–782.

14. Hillen,H.S., Kokic,G., Farnung,L., Dienemann,C., Tegunov,D. and Cramer,P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature*, **584**, 154–156

15. Wang,Q., Wu,J., Wang,H., Gao,Y., Liu,Q., Mu,A., Ji,W., Yan,L., Zhu,Y., Zhu,C. *etal.* (2020) Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell*, **182**, 417–428.

16. Yin,W., Mao,C., Luan,X., Shen,D.D., Shen,Q., Su,H., Wang,X., Zhou,F., Zhao,W., Gao,M. *et al.* (2020) Structural basis for inhibition of the RNA- dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*, **368**,1499–1504.

17. Zhai,Y., Sun,F., Li,X., Pang,H., Xu,X., Bartlam,M. and Rao,Z. (2005) Insights intoSARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat. Struct. Mol. Biol.*, **12**, 980–986.

18. Krichel,B., Bylapudi,G., Schmidt,C., Blanchet,C., Schubert,R., Brings,L., Koehler,M., Zenobi,R., Svergun,D., Lorenzen,K. *et al.* (2021) Hallmarks of alpha-and Betacoronavirus non-structural protein 7+8 complexes. *Sci. Adv.*, **7**

19. Krichel,B., Falke,S., Hilgenfeld,R., Redecke,L. and Uetrecht,C. (2020) Processing of the SARS-CoV pp1a/ab nsp7-10 region. *Biochem. J.*, **477**, 1009–1019.

20. Xiao,Y., Ma,Q., Restle,T., Shang,W., Svergun,D.I., Ponnusamy,R., Sczakiel,G.and Hilgenfeld,R. (2012) Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1

heterotrimer that exhibits primer-independent RNA polymerase activity. *J. Virol.*, **86**, 4444–4454

21. Thao,S., Zhao,Q., Kimball,T., Steffen,E., Blommel,P.G., Riters,M., Newman,C.S.,Fox,B.G. and Wrobel,R.L. (2004) Results from high-throughput DNA cloning of Arabidopsis thaliana target genes using site-specific recombination. *J. Struct. Funct. Genomics*, **5**, 267–276

22. Minor,W., Cymborowski,M., Otwinowski,Z. and Chruszcz,M. (2006) HKL-3000: the integration of data reduction and structure solution–from diffraction images
to an initial model in minutes. *Acta Crystallogr. D. Biol. Crystallogr.*, **62**, 859–866.

23. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.*, **60**, 2126–2132.

24. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D. Biol. Crystallogr.*, **58**, 1948–1954.

25. Johnson,M.A., Jaudzems,K. and Wuthrich,K. (2010) NMR structure of the SARS-CoV nonstructural protein 7 in solution at pH 6.5. *J. Mol. Biol.*, **402**, 619–628.

26. Yan,L., Zhang,Y., Ge,J., Zheng,L., Gao,Y., Wang,T., Jia,Z., Wang,H., Huang,Y.,Li,M. *et al.* (2020) Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat. Commun.*, **11**, 5874.

27. te Velthuis,A.J., van den Worm,S.H. and Snijder,E.J. (2012) The SARS- coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Res.*, **40**, 1737– 1747.

28. Imbert,I., Guillemot,J.C., Bourhis,J.M., Bussetta,C., Coutard,B., Egloff,M.P., Ferron,F., Gorbalenya,A.E. and Canard,B. (2006) A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J.*, **25**, 4933–4942.

29. Wang,B., Svetlov,V., Nudler,E. and Artsimovitch,I. (2021) Lost in translation:codon optimization inactivates SARS-CoV-2 RdRp. BioRxiv, 2021.2001.2024.428004.

30. Wang,B., Thurmond,S., Hai,R. and Song,J. (2018) Structure and

function of Zikavirus NS5 protein: perspectives for drug design. *Cell. Mol. Life Sci.*, **75**, 1723– 1736.

31. Gordon,D.E., Jang,G.M., Bouhaddou,M., Xu,J., Obernier,K., White,K.M.,O"Meara,M.J., Rezelj,V.V., Guo,J.Z., Swaney,D.L. *et al.* (2020) A SARS-CoV-2 protein interaction map reveals targets for drugrepurposing. *Nature*, **583**, 459–468.

32. Reshamwala,S.M.S., Likhite,V., Degani,M.S., Deb,S.S. and Noronha,S.B. (2021) Mutations in SARS-CoV-2 nsp7 and nsp8 proteins and their predictedimpact on replication/transcription complex structure. *J. Med. Virol*.

33. Pachetti,M., Marini,B., Benedetti,F., Giudici,F., Mauro,E., Storici,P., Masciovecchio,C., Angeletti,S., Ciccozzi,M., Gallo,R.C. *et al.* (2020) EmergingSARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.*, **18**, 179.

34. Biswas,S.K. and Mudi,S.R. (2020) Spike protein D614G and RdRp P323L: theSARS-CoV-2 mutations associated with severity of COVID-19. *Genomics & informatics*, **18**, e44.

35. Gordon,C.J., Tchesnokov,E.P., Feng,J.Y., Porter,D.P. and Gotte,M. (2020) The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerasefrom Middle East respiratory syndrome coronavirus. *J. Biol. Chem.*, **295**, 4773– 4779.

36. Konkolova,E., Klima,M., Nencka,R. and Boura,E. (2020) Structural analysis ofthe putative SARS-CoV-2 primase complex. *J. Struct. Biol.*, **211**, 107548.

37. Ashkenazy,H., Abadi,S., Martz,E., Chay,O., Mayrose,I., Pupko,T. and Ben-Tal,N.(2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.*, **44**, W344– 350.

38. Yan L, Yang Y, Li M, Zhang Y, Zheng L, Ge J, et al. . Coupling of N7- Methyltransferase and 3"-5" Exoribonuclease With SARS-CoV-2 Polymerase Reveals Mechanisms for Capping and Proofreading. Cell (2021) 184(13):3474–85 e11. doi: 10.1016/j.cell.2021.05.033

39. Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., et al.,(2020). Genome Composition and Divergence of the Novel Coronavirus (2019- nCoV) Originating in China. Cell Host Microbe 27, 325–328.

40. Chang, C.K., Hou, M.H., Chang, C.F., Hsiao, C.D., Huang, T.H., (2014). The SARS coronavirus nucleocapsid protein– forms and functions. Antiviral Res. 103,39–50.

41. Hsieh, P.K., Chang, S.C., Huang, C.C., Lee, T.T., Hsiao, C. W., Kou, Y.H., Chen,I.Y., Chang, C.K., et al., (2005). Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. J. Virol. 79, 13848–13855.

42. McBride, R., van Zyl, M., Fielding, B.C., (2014). The coronavirus nucleocapsid isa multifunctional protein. Viruses 6, 2991–3018.

43. Liu, X., Verma, A., Garcia Jr., G., Ramage, H., Lucas, A., Myers, R.L., Michaelson, J.J., Coryell, W., et al., (2021). Targeting the coronavirus nucleocapsid protein through GSK-3 inhibition. Proc. Natl. Acad. Sci. U. S. A.118

44. Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V.,et al., (2020). A SARS-CoV-2 protein interaction mapreveals targets for drug repurposing. Nature 583, 459–468.

45. Luo, L., Li, Z., Zhao, T., Ju, X., Ma, P., Jin, B., Zhou, Y., He, S., et al., (2021).SARS-CoV-2 nucleocapsid protein phase separates with G3BPs to disassemblestress granules and facilitate viral production. Science bulletin 66, 1194–1204.

46. Buchan, J.R., Parker, R., (2009). Eukaryotic stress granules: the ins and outs oftranslation. Mol. Cell 36, 932–941.

47. White, J.P., Lloyd, R.E., (2012). Regulation of stress granules in virus systems.Trends Microbiol. 20, 175–183.

48. Guillen-Boixet, J., Kopach, A., Holehouse, A.S., Wittmann, S., Jahnel, M., Schlussler, R., Kim, K., Trussina, I., et al., (2020). RNA-Induced ConformationalSwitching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. Cell 181 346–361 e317.

49. Bley, N., Lederer, M., Pfalz, B., Reinke, C., Fuchs, T., Glass, M., Moller, B., Huttelmaier, S., (2015). Stress granules are dispensable for mRNA stabilizationduring cellular stress. Nucleic Acids Res. 43, e26

50. Kedersha, N., Chen, S., Gilks, N., Li, W., Miller, I.J., Stahl, J., Anderson, P., (2002). Evidence that ternary complex (eIF2-GTP-

tRNA(i)(Met))-deficient preinitiation complexes are core constituents of mammalian stress granules. Mol.Biol. Cell 13, 195–210.

51. Matsuki, H., Takahashi, M., Higuchi, M., Makokha, G.N., Oie, M., Fujii, M., (2013). Both G3BP1 and G3BP2 contribute to stress granule formation. GenesCells: Devoted Mol. Cell. Mech. 18, 135–146.

52. Yang, P., Mathieu, C., Kolaitis, R.M., Zhang, P., Messing, J., Yurtsever, U., Yang, Z., Wu, J., et al., (2020). G3BP1 Is a Tunable Switch that Triggers PhaseSeparation to Assemble Stress Granules. Cell,181, 325–345, e328.

53. Tourriere, H., Chebli, K., Zekri, L., Courselaud, B., Blanchard, J.M., Bertrand, E.,Tazi, J., (2003). The RasGAP-associated endoribonuclease G3BP assembles stress granules. J. Cell Biol. 160, 823–831.

54. McCormick, C., Khaperskyy, D.A., (2017). Translation inhibition and stress granules in the antiviral immune response. Nature Rev. Immunol. 17, 647–660.

55. Parker, F., Maurier, F., Delumeau, I., Duchesne, M., Faucher, D., Debussche, L.,Dugue, A., Schweighoffer, F., et al., (1996). A Ras-GTPase-activating protein SH3-domain-binding protein. Mol. Cell. Biol. 16, 2561–2569.

56. Hofmann, S., Kedersha, N., Anderson, P., Ivanov, P.,(2021). Molecular mechanisms of stress granule assembly and disassembly. Biochim. Biophys.Acta, Mol. Cell. Res.1868, 118876

57. Sanders, D.W., Kedersha, N., Lee, D.S.W., Strom, A.R.,Drake, V., Riback, J.A., Bracha, D., Eeftens, J.M., et al.,(2020). Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. Cell 181 306–324 e328.

58. Kedersha, N., Panas, M.D., Achorn, C.A., Lyons, S., Tisdale, S., Hickman, T.,Thomas, M., Lieberman, J., et al., (2016). G3BP-Caprin1-USP10 complexes mediate stress granule condensation and associate with 40S subunits. J. CellBiol. 212, 845–860.

59. Panas, M.D., Varjak, M., Lulla, A., Eng, K.E., Merits, A., Karlsson Hedestam,G.B., McInerney, G.M., (2012). Sequestration of G3BP coupled with efficient translation inhibits stress granules in Semliki Forest virus infection. Mol. Biol. Cell

23, 4701–4712.

60. White, J.P., Cardenas, A.M., Marissen, W.E., Lloyd, R.E., (2007). Inhibition ofcytoplasmic mRNA stress granule formation by a viral proteinase. Cell Host Microbe 2, 295–305.

61. Reineke, L.C., Lloyd, R.E., (2013). Diversion of stress granules and P-bodiesduring viral infection. Virology 436, 255–267.

62. Pager, C.T., Schutz, S., Abraham, T.M., Luo, G., Sarnow, P., (2013). Modulationof hepatitis C virus RNA abundance and virus release by dispersion of processing bodies and enrichment of stress granules. Virology 435, 472–484.

63. Fros, J.J., Domeradzka, N.E., Baggen, J., Geertsema, C., Flipse, J., Vlak, J.M., Pijlman, G.P., (2012). Chikungunya virus nsP3 blocks stress granule assembly by recruitment of G3BP into cytoplasmic foci. J. Virol. 86, 10873–10879.

64. Schulte, T., Liu, L., Panas, M.D., Thaa, B., Dickson, N., Gotte, B., Achour, A., McInerney, G.M., (2016). Combined structural, biochemical and cellular evidencedemonstrates that both FGDF motifs in alphavirus nsP3 are required for efficient replication. Open Biol. 6

65. Savastano, A., Ibanez de Opakua, A., Rankovic, M., Zweckstetter, M., (2020).Nucleocapsid protein of SARSCoV-2 phase separates into RNA-rich polymerasecontaining condensates. Nature Commun. 11, 6041.

66. Fribourg, S., Braun, I.C., Izaurralde, E., Conti, E., (2001). Structural basis for therecognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor. Mol. Cell 8, 645–656.

67. Schulte, T., Panas, M.D., Williams, L., Kedersha, N., Fleck, J.S., Tan, T.J.C.,Olsson, A., Morro, A.M., et al., (2021). Caprin-1 binding to the critical stress granule protein G3BP1 is regulated by pH. bioRxiv. 2021.2002.2005.429362.

68. Cirillo L, Cieren A, Barbieri S, Khong A, Schwager F, Parker R, Gotta M; UBAP2L Forms Distinct Cores that Act in Nucleating Stress Granules Upstreamof G3BP1., Curr. Biol., 30 (2020), pp. 698-707.e6

69. Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., etal., (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. Acta Pharm. Sin. B 10, 1228–1238.

70. Zhou, R., Zeng, R., von Brunn, A., Lei, J., (2020). Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. Mol. Biomed. 1, 2.

71. Peng, Y., Du, N., Lei, Y., Dorje, S., Qi, J., Luo, T., Gao, G. F., Song, H., (2020).
Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. EMBO J. 39, e105938

72. Chang, C.K., Hsu, Y.L., Chang, Y.H., Chao, F.A., Wu, M. C., Huang, Y.S., Hu, C.K., Huang, T.H., (2009). Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. J. Virol. 83, 2255–2264.

73. Lu, S., Ye, Q., Singh, D., Cao, Y., Diedrich, J.K., Yates, J. R., Villa, E., Cleveland, D.W., Corbett, K.D., (2021). The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and themembrane associated M protein. Nature Commun. 12, 502.

74. Peng, T.Y., Lee, K.R., Tarn, W.Y., (2008). Phosphorylation of the arginine/serinedipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory
activity and cellular localization. FEBS J. 275, 4152–4163

75. Kruse, T., Benz, C., Garvanska, D. H., Lindqvist, R., Mihalic, F., Coscia, F., Inturi,R., Sayadi, A., Simonetti, L., Nilsson, E., Ali, M., Kliche, J., Moliner Morro, A., Mund, A., Andersson, E., et al. (2021) Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities. Nat. Commun. 12, 6761

76. Huang, W., Ju, X., Tian, M., Li, X., Yu, Y., Sun, Q., Ding, Q., Jia, D., (2021).Molecular determinants for regulation of G3BP1/2 phase separation by theSARS-CoV-2 nucleocapsid protein. Cell Discovery 7, 69.

77. Vognsen, T., Moller, I.R., Kristensen, O., (2013). Crystal structures of the humanG3BP1 NTF2-like domain visualize FxFG Nup repeat specificity. PLoS ONE 8, e80947

78. Gwon, Y., Maxwell, B.A., Kolaitis, R.M., Zhang, P., Kim, H. J., Taylor, J.P., (2021). Ubiquitination of G3BP1 mediates stress granule disassembly in a context-specific manner. Science 372, eabf6548.

79. Eberhardt, R.Y., Chang, Y., Bateman, A., Murzin, A.G., Axelrod, H.L., Hwang,W.C., Aravind, L., (2013). Filling out the structural map of the NTF2-like superfamily. BMC Bioinf. 14, 327.

80. Zhang, K., Xie, Y., Mun˜oz-Moreno, R., Wang, J., Zhang, L., Esparza, M., Garcı´a-Sastre, A., Fontoura, B.M.A., et al., (2019). Structural basis for influenza virusNS1 protein block of mRNA nuclear export. Nature Microbiol. 4, 1671– 1679.

81. Stewart, M., Kent, H.M., McCoy, A.J., (1998). Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran. J. Mol. Biol. 277, 635–646.

82. Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dong, A., Wasney, G.A., Gao, M., et al., (2007). Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. Mol. Biochem. Parasitol. 151, 100–110.

83. Li, K., Ossareh-Nazari, B., Liu, X., Dargemont, C., Marmorstein, R., (2007). Molecular basis for bre5 cofactor recognition by the ubp3 deubiquitylating enzyme.
J. Mol. Biol. 372, 194–204.

84. Lin, C.S., Chan, A.C.K., Vermeulen, J., Brockerman, J., Soni, A.S., Tanner, M.E., Gaynor, E.C., McIntosh, L.P., et al., (2021). Peptidoglycan binding by a pocket on the accessory NTF2-domain of Pgp2 directs helical cell shape of Campylobacter jejuni. J. Biol. Chem. 296, 100528

85. Bai, Z., Cao, Y., Liu, W., Li, J., (2021). The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. Viruses 13

86. Mohammad, T., Choudhury, A., Habib, I., Asrani, P., Mathur, Y., Umair, M., Anjum, F., Shafie, A., et al., (2021). Genomic Variations in the Structural Proteins of SARS-CoV-2 and Their Deleterious Impact on Pathogenesis: A Comparative Genomics Approach. Front. Cell. Infect. Microbiol. 11, 765039

87. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., (2007). Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674.

88. C. Huang, Y. Chen, H. Dai, Huan Zhang, M. Xie, Hanbin

Zhang, F. Chen, X.Kang, X. Bai, Z. Chen, UBAP2L arginine methylation by PRMT1 modulates stress granule assembly, Cell Death Differ. 27 (2020) 227–241.