DYNAMIC MOLECULAR REPROGRAMMING OF HUMAN AND VIRAL PROTEIN SYNTHESIS BY SARS-COV-2 NUCLEOCAPSID PROTEIN

By

Justin Donald Edwards

A capstone project submitted for Graduation with University Honors

May 12, 2023

University Honors
University of California, Riverside

APPROVED

Dr. Seán O'Leary
Department of Biochemistry

Dr. Richard Cardullo, Howard H Hays Jr. Chair
University Honors

ABSTRACT

The COVID-19 pandemic caused by the SARS-CoV-2 virus has had a profound impact on global health and economic stability. Understanding mechanisms of viral interaction with the host cellular machinery is crucial in addressing the challenges of COVID-19. Here, we investigate how the SARS-CoV-2 nucleocapsid protein (N protein) impacts a critical step in the regulation of gene expression, eIF4F-dependent mRNA translation initiation. Previous research suggests that N protein perturbs eIF4F function, suppressing translation, but the molecular mechanisms of this perturbation and its downstream effects remain unclear. Our measurements of N-protein affinity to host and viral mRNAs suggest a hypothesis that eIF4F interactions with N-protein alter eIF4F-mRNA dynamics differently for host and viral mRNAs, aiding the virus in hijacking the host cellular machinery to support its replication and immune system evasion. To test this hypothesis, we directly observe and quantify N protein effects on eIF4F-mRNA dynamics with single-molecule fluorescence microscopy, which enables us to monitor, in real time, changes in kinetics of eIF4F-mRNA interaction induced by external factors such as N-protein. Our results offer new insights into host-virus interplay, and gain a deeper understanding of how, at a molecular level, a key SARS-CoV-2 protein disrupts central biological systems in the host cell during viral pathogenesis. Our findings have the potential to provide insights into the fundamental mechanisms of translation initiation and host-virus interactions, thereby advancing our understanding of the molecular mechanisms underlying SARS-CoV-2 pathogenesis.

ACKNOWLEDGEMENTS

Dr. Seán O'Leary for giving me a chance to work in his lab for the past year and a half and his constant cheerful attitude and support this whole time.

Matthew Guevara for his mentorship, developing the EMSA assay, and countless hours of dedication toward this project. Without him this paper would have never existed.

Hea Jin Hong for use of her protocol for purification of the eIF4F complex alongside her dual luciferase constructs. Without her none of this would have been possible.

Paria Azami for her consultation on RNA constructs and help with troubleshooting various steps of RNA preparation.

Alexandra Huang for her help with preparation of multiple protein preps and sharing various protocols.

Melanie Randall for her help with gathering data for the entire project. Without her this paper would not be completed on time.

Special thanks to Dr. Song for his consultation on the nucleocapsid protein.

TABLE OF CONTENTS

Introduction……..…….4

- Translation Initiation Through Eukaryotic Initiation Factor 4F (eIF4F)
- Insights into Early Stage Viral Infection of SARS-CoV-2
- The Nucleocapsid Protein
- Poly(A) Binding Protein

Materials & Methods……..8

- Purifications Protocols of eIF4F complex
- Purification of N protein and PABPC1
- Preparation of RNA
- Electro Mobility Shift Assay
- Single-Molecule Assay

Results & Discussion……..18

- N protein and PABP bind mRNA poly(A) tails comparably
- N protein advantages SARS-CoV-2 in binding to eIF4F
- N protein selectively promotes the SARS-CoV-2 RNA
- Poly(A) Binding Protein increases eIF4F mRNA lifetimes

Conclusion……..……26

- Future Research
- Limitations

References……...…..…….……..…....28

INTRODUCTION

Given the threat posed by the COVID-19 pandemic, it is imperative to deepen our understanding of the intricate interplay between the virus and its host, in order to inform effective interventions and combat this global health crisis. A comprehensive understanding of the major proliferative factors of SARS-CoV-2 has become an urgent and imperative research priority. While much attention has been given to the virus's spike protein and its role in viral entry, the importance of the nucleocapsid protein cannot be left nescient.

Translation Initiation Through Eukaryotic Initiation Factor 4F (eIF4F)

eIF4F plays a critical role in the early stages of translation initiation, as it recognizes the $m^7G(5')ppp(5')N$ "cap" structure at the 5′ end of host and viral RNAs, where N is any nucleotide. The eIF4F complex consists of three subunits: eIF4E, which recognizes the 5′ cap dinucleotide; eIF4G, which scaffolds the complex and interacts with other translation factors; and eIF4A, which acts as an RNA helicase to unwind the mRNA. [Gingras, *et al.*] It is essential for viral and host translation. eIF4F is required for recruitment of the ribosomal 43S pre-initiation complex and this recruitment leads to a multistep process that culminates in assembly of the translationally-active 80S ribosome on the mRNA allowing entry into the subsequent steps of translation of elongation and termination. [Poulin, *et al.*]

Insights into Early Stage Viral Infection of SARS-CoV-2

To propagate, the SARS-CoV-2 virus must translate its genomic RNA into proteins that facilitate a multi-pronged strategy to take over host cell systems. Upon infecting the host cell, the virus injects two components, the viral genomic RNA and the nucleocapsid (N) protein, into the cytosol. The viral RNA is a positive-sense RNA that is able to directly translate proteins without

transcription (DNA to RNA replication). The RNA possesses a 5′ cap and a 3′ polyadenylate tail that interact with the eukaryotic initiation factor 4F (eIF4F) complex essential for translation initiation, in addition to other translation factors. It has been conventionally assumed that immediately following infection, the viral RNA collaborates with eIF4F to initiate translation of the first two open reading frames (ORFs) of the viral genome, which in turn encode 16 nonstructural proteins (NSPs). [Breyne, *et al.*] The 5′ untranslated region of the RNA harbors several stable secondary structural elements, including a four-way junction located near the AUG start codon. According to Zhichao, Et al's paper titled "Secondary structure of the SARS-CoV-2 5′-UTR," this structural element is similar to the internal ribosome entry site (IRES) of the Hepatitis C Virus. As the AUG start codon is situated downstream and in the vicinity of the four-way junction structure, the SARS-CoV-2 virus may have the ability to undergo cap-independent translation, bypassing eIF4F-mediated translation initiation. The mechanism by which SARS-CoV-2 begins this initial first translation of its genomic RNA is not well understood; in this work we sought to gain detailed insights into how the eIF4F and the genomic RNA interact and whether or not there is a host or viral preference in this interaction.

The Nucleocapsid Protein

The nucleocapsid (N) protein that enters the host cell bound to the viral RNA is another crucial component of the SARS-CoV-2 virus, and is thought to have multiple roles in viral replication and packaging of the genome. *[Surjit , Et al]* When the viral genome enters the cell, many N proteins dissociate from the RNA and both molecules begin a multipronged strategy for host cell takeover. *[Duarte, Et al]* An important pathway involving N protein is its contribution to the formation of double membrane vesicles used to replicate the viral genome. N protein has shown to have different kinds of inhibitory effects on host antiviral immune defenses. *[Setten, Et*

al Mu, Et al] That being said N protein has shown to induce a strong immune response when present in the cell even though it is the most abundant protein in SARS-CoV-2. *[Yoshimoto, Et al Wu, Et al]* N protein localizes mostly in the cytosol, but the nuclear fraction potentially inhibits nuclear export of host mRNA, thus selectively promoting viral RNA replication through depletion of host mRNAs in the cytosol. *[Reed, Et al]* These studies showcase how N protein is an incredibly dynamic molecule and has many roles throughout the viral life cycle.

Previous research has suggested that the N protein may also interfere with translation initiation by interacting with the eIF4F complex and inhibiting its activity, which is necessary for cap-dependent translation of the first two open reading frames during early infection *[Tsai, Et al]*. However, this presents a contradiction because if the N protein inhibits eIF4F formation, as suggested in studies on bovine coronavirus N protein, how can the virus propagate? One possible explanation is that N protein could mediate an interaction that allows its own viral RNA to translate while only inhibiting host mRNA translation. In this work we aimed to determine whether N protein interacts differently with viral and host mRNAs, how N protein alters eIF4F–mRNA interaction dynamics, and if it does, whether this activity differs between host and viral mRNAs.

Poly(A) Binding Protein

The involvement of Poly(A) Binding Protein (PABP) in translation initiation discussion has been widely studied, leading to contrasting mechanistic proposals. PABP serves as one of the critical structural components of the “closed loop” model of translation. In this model, PABP bound to the mRNA’s 3′ poly(A) tail associates with the 4F complex by binding the 4G subunit. This network of interactions brings the mRNA 5′ and 3′ ends into close spatial proximity, and thus a loop is formed. *[Wells, et al. Gallie, et al.]* PABP binding to the initiation complex greatly

enhances the efficiency of translation, including by promoting the recruitment of the ribosome 40S subunit to the mRNA•eIF4F•PABP complex. *[Gallie, Et al]*

Prior research has shown that N protein plays a role in binding to host mRNA in the cytosol with a preference for the 3′ UTR decreasing RNA stability. *[Syed, Et al]* Notably, the Syed study brings up a limitation which is that N protein tends to bind its genomic RNA in the cell, so its ability to target host mRNA is hindered through competitive binding. Further research is needed to fully understand this mechanism upon which N protein interacts with host viral RNA and the downstream effects of this interaction. In this work, we therefore also sought to delineate the dynamics of the interplay between N protein and PABP and how these dynamics impact the function of eIF4F complexes with viral and host mRNA.

MATERIALS & METHODS

Purifications Protocols of eIF4F complex

The purification of each subunit of the eIF4F complex was undergone by our lab before I arrived, the protocols were developed by HeaJin Hong and we graciously thank her for sharing these protocols with us.

Preparation of Recombinant Human eIF4E

The pET-28a(+) vectors containing the sequences of eIF4E or eIF4E(S209D), each fused with an N-terminal Protein G tag, were designed similarly to Feokistova et. al. The constructs contained a sequence encoding the peptide MA(pAzF) between the Protein G tag and the N-terminus of eIF4E, where pAzF is p-azidophenylalanine (pAzF). This is done in order to incorporate the cy5 dye into the protein via cuaac click reaction mechanism. The pULTRA expression vector, was co-transformed into an E. coli BL95 strain with the pET28a(+)-eIF4E plasmid, and transformants were selected on LB agar containing Kanamycin (50 μg/mL) and Spectinomycin (100 μg/mL). A single colony from this selection was used to inoculate a 10 mL LB broth starter culture containing both antibiotics, which was grown overnight at 37 °C. The following day, the starter culture was used to inoculate 1 L of LB broth containing the antibiotics. The culture was grown at 37 °C to OD600 ~0.6, and then 1 mM pAzF was added along with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) to induce recombinant protein overexpression. Induction was carried out for 5 h at 30 °C in darkness in order to avoid photobleaching of fluorescent cy5 dye. The cells were spun down and suspended in lysis buffer (20 mM HEPES-KOH pH 7.5, 400 mM KCl, 5 mM imidazole, 1×EDTA- free protease inhibitor cocktail (0.24 mg/mL benzamidine 8 hydrochloride hydrate, 2 μM pepstatin, and 0.6 μM

leupeptin hemisulfate), 1× phenylmethylsulfonyl fluoride (PMSF). Proteins were purified immediately after induction, without freezing of the cell pellet as freezing kills the protein. The cells were lysed using a sonicator 80 % amplitude, setting 3, for 45 seconds at 5 min intervals and the lysate was immediately clarified by centrifugation (22,000 rpm for 30 min). The filtered lysate was passed through 5 mL Ni-NTA agarose, followed by 3 washes with 5 column volumes of wash buffer (20 mM HEPES-KOH pH 7.5, 400 mM KCl, 10 mM imidazole) until no further detectable protein eluted from the column, as assessed by Bradford reagent. Ni-NTA-bound protein was then eluted with elution buffer (20 mM HEPES-KOH pH 7.5, 400 mM KCl, 200 mM imidazole), and eluate fractions were analyzed by 10 % SDS-PAGE. Fractions that contained eIF4E were combined, and labeled overnight at 4 °C with 50 μM Cy5-DBCO. The labeled protein was desalted to remove unreacted dye and the sample was buffer exchanged into TEV cleavage buffer (20 mM HEPES-KOH, 50 mM KCl, 10 % (v/v) glycerol). The desalted sample was then incubated with TEV protease (50 units) overnight at 4 °C for complete cleavage of protein G from the 4E. Once the cleavage was complete, the sample was loaded onto an SP-column maintained at ~4 °C. The column was washed with 20 column volumes of low-salt buffer (20 mM HEPES-KOH, 50 mM KCl, 10 % Glycerol), then eluted with a linear gradient from 0% to 100% high-salt buffer (20 mM HEPES-KOH, 500 mM KCl, 10 % Glycerol). Fractions containing Cy5-eIF4E typically eluted at ~250 mM KCl. Based on inspection of an SDS-PAGE gel of the eluate fractions, imaged for Cy5 fluorescence by a Typhoon imager, Fractions containing labeled protein were combined, concentrated to < 1 mL by centrifugal ultrafiltration and loaded onto a Superdex 75 Increase column equilibrated in storage buffer (20 mM HEPES-KOH pH 7.5, 200 mM KCl, 10 % Glycerol, 1 mM TCEP) for further purification. Fractions were assessed for purity by SDS-PAGE, and for labeling efficiency by measuring the

ratio of 640 nm to 280 nm absorbance. The 1,2,3-triazole formed in the azide-alkyne cycloaddition reaction that conjugates the fluorophore to the protein is expected to absorb at 280 nm. Considering the contribution of this additional absorbance, the typical labeling efficiency was ~40%; this assessment was supported by SDS-PAGE analysis of the labeled protein, which contained two bands, only one of which was conjugated to Cy5. Purified Cy5-eIF4E and Cy5-eIF4E were stored at 4 °C in darkness and used for a maximum of one week after each purification.

Preparation of Recombinant Human eIF4A

The plasmid was transformed into BL21 cells and transformants were selected on LB–agar plates containing ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL). A single colony from this selection was used to grow a 10 mL overnight culture at 37 °C. This starter culture was then used to inoculate 1 L LB containing the selective antibiotics. The cells were grown to OD600 of ~0.6, then protein overexpression was induced with 1 mM IPTG for 3 h at 37 °C. After induction, the cells were harvested by centrifugation at 4,000 rpm for 15 min and were frozen and stored at -80 °C until purification. The frozen cell pellet was resuspended in lysis buffer (20 mM HEPES-KOH, pH 7.5, 500 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, 0.5% (v/v) NP-40). The cells were lysed by sonication with a microtip attachment (80 % amplitude, setting 3, for 45 seconds at 5 min intervals) and the lysate was immediately clarified by centrifugation (22,000 rpm for 30 min). During centrifugation, a gravity-flow column containing 1 mL Ni-NTA agarose was washed and pre-equilibrated with the lysis buffer. The cell-free supernatant was diluted twofold with this lysis buffer, and filtered through a 0.22 μm syringe filter. The filtrate was then applied to the equilibrated Ni-NTA

column, which was washed with a further ~20 mL lysis buffer. The column was further washed with 40 mL high-salt buffer (20 mM HEPES-KOH, pH 7.5, 1 M NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol), followed by low-salt buffer (20 mM HEPES-KOH, pH 7.5, 500 mM NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol). The protein was then eluted with elution buffer (50 mM Na-phosphate buffer, pH 7.5, 500 mM NaCl, 100 mM Na2SO4, 250 mM imidazole, 2 mM DTT). After elution, the eluate was buffer-exchanged into low-salt buffer (20 mM Tris HCl pH 7.5, 100 mM KCl, 5% glycerol, 2 mM DTT, 0.1 mM EDTA) using a 10 DG column, and was TEV-protease (50 units) cleaved overnight at 4°C to remove the APEX tag. The TEV-protease cleaved sample was loaded onto a 5 mL Q-Sepharose HP column, equilibrated in low-salt buffer and maintained at ~4 °C. The column was washed with 20 column volumes of low-salt buffer, then eluted with a linear gradient from 0% to 100% high-salt buffer (20 mM Tris HCl pH 7.5, 500 mM KCl, 5% glycerol, 2 mM DTT, 0.1 mM EDTA) over 50 column volumes. eIF4A typically eluted at ~265 mM KCl. Eluate fractions were analyzed by SDS-PAGE, and fractions containing pure eIF4A were then buffer exchanged into a storage buffer (20 mM Tris HCl, pH 7.5, 2 mM DTT, 0.1 mM EDTA, 10% glycerol, 100 mM KCl) using Superdex-75, flash frozen under liquid nitrogen and was stored at -80 °C.

Preparation of Recombinant Human eIF4G

The eIF4G cell pellets were a gift from Dr. Rong Hai and we thank him for offering to grow the cells for us. The eIF4G cells were frozen at -80ºC until purification. To extract the protein, the cells were lysed in lysis buffer (20 mM HEPES pH 7.5, 400 mM KCl, 5 mM Imidazole, 10 % Glycerol, 10 mM b-mercaptoethanol, 1 x protease inhibitor, 1 x PMSF, 0.1 % NP-40), by sonication in short bursts (80 % amplitude, setting 3, for 45 seconds at 5 min

intervals), then centrifuged to clarify the lysate.. The lysate was then filtered using a 0.45µm syringe filter. The protein was loaded onto a Pre-equilibrated 5 mL chelating column and washed with lysis and wash buffer (20 mM HEPES pH 7.5, 400 mM KCl, 10 mM Imidazole, 10 % glycerol, 10 mM b-mercaptoethanol, 1 x protease inhibitor, 1 x PMSF, 0.1 % NP-40) then washed again with with a low imidazole (100mM) wash buffer (20 mM HEPES pH 7.5, 400mM KCl, 100 mM Imidazole, 10 % glycerol, 10 mM b-mercaptoethanol, 0.1 % NP-40). The sample was then eluted with elution buffer (20 mM HEPES pH 7.5, 400mM KCl, 500 mM Imidazole, 10 % glycerol, 10 mM b-mercaptoethanol, 0.1 % NP-40). The sample was incubated overnight with 40 uL of the equilibrated FLAG resin at 4ºC. After overnight incubation, the sample was washed three times with the equilibration buffer (20 mM HEPES pH 7.5, 400 mM KCl, 10 % glycerol, 10 mM b-mercaptoethanol, 0.1 % NP-40). Finally, sample was eluted off the flag resin with elution buffer (20 mM HEPES pH 7.5, 400 mM KCl, 150 ug/mL FLAG peptide, 10 % glycerol, 10 mM b-mercaptoethanol, 0.1 % NP-40). After elution the 4G sample was flash frozen and stored at -80ºC until use. Sample was run on SDS PAGE every month to check for degradation as large constructs like this protein tend to degrade over time.

Purification of N protein and PABPC1

The following proteins were purified by us for the purpose of this paper. Purification protocols were developed over the course of many months as these purifications are a very long and tedious process typically taking up to one week to perform.

Preparation of Recombinant Nucleocapsid Protein

To purify the SARS-CoV-2 Nucleocapsid Protein, a pET-28a(+) vector was first constructed with an insert containing a sequence encoding a hexahistidine tag followed by the

sequence encoding the protein, which was codon-optimized for E. coli. The plasmid was then transformed into BL21 (DE3) CodonPlus cells, and transformants were selected on LB-agar plates containing kanamycin (50 μg/mL) and chloramphenicol (25 μg/mL). A single colony of transformants was used to inoculate a 10 mL LB starter culture, which was grown overnight at 37 ˚C. The next day, the starter culture was used to inoculate 1L of LB cultures containing the selective antibiotics. The cells were grown at 37°C until the OD600 reached ~0.6, at which point protein overexpression was induced by the addition of 0.25mM IPTG, and allowed to proceed for 3 hours at 30°C. The cells were then pelleted by centrifugation at (2,200 rpm for 30 minutes) and stored at -80°C until purification. To extract the protein, the cells were lysed in lysis buffer (25 mM Tris-pH 7.5, 500 mM NaCl, 5 mM β-mercapto, 10% glycerol, 5 mM imidazole, 5 mM MgCl2, 1 mM NaN3) by sonication in short bursts (80 % amplitude, setting 3, for 45 seconds at 5 min intervals), then centrifuged to clarify the lysate.. The lysate was then filtered using 0.8μm and then 0.2μm syringe filter. The protein was loaded onto a nickel affinity column and washed with wash buffer (25 mM Tris-pH 7.5, 300 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, 20 mM imidazole , 5 mM MgCl2, 1 mM NaN3) then eluted with elution buffer (25 mM Tris-pH 7.5, 300 mM NaCl, 5 mM β-mercapto, 10% glycerol, 400 mM imidazole , 5 mM MgCl2, 1 mM NaN3). In order to remove the His-Tag, the sample was incubated overnight with TEV protease. To prepare the sample for further purification by FPLC, the protein was exchanged into TEV buffer (25 mM Tris-pH 7.5, 300 mM NaCl, 5 mM β-mercapto, 10% glycerol, 5 mM MgCl2, 1 mM NaN3, 20mM Imidazole). The sample was then further purified by FPLC with a heparin column to remove nucleic acid contamination. Finally, the sample was concentrated and the buffer exchanged to a storage buffer (25 mM Tris-pH 7.5, 300 mM NaCl, 10% glycerol) for future use.

Preparation of Recombinant Human Poly(A) Binding Protein (PABPC)

A pGST vector was constructed with an insert containing a sequence encoding a Glutathione Glutathione-S-transferase (GST) fusion protein (~26 kDa) followed by the sequence encoding Human Poly(A)-Binding Protein (PABPC1). The plasmid was transformed into BL21 CodonPlus cells. Transformants were selected on LB-agar plates containing ampicillin (100 µg/mL) and chloramphenicol (25 µg/mL). A single transformant colony was used to inoculate a 10 mL LB starter culture, which was grown overnight at 37 ˚C. The next day, the starter culture was used to inoculate 1L of LB cultures containing the selective antibiotics. Cells were grown at 37°C to OD600 of ~0.6, then protein overexpression was induced by addition of 0.5mM IPTG, and allowed to proceed for 3hrs at 37°C. The cells were pelleted by centrifugation at (2,200 rpm for 30 minutes) and stored at -80°C until purification. To extract the protein, the cells were lysed in lysis buffer (10 mM Na2HPO4 (pH 7.4), 1.8 mM KH2PO4 (pH 7.4), 637 mM NaCl, 2.7 mM KCl, 10% glycerol, 1x PMSF, 1X Protease Inhibitor Cocktail) by sonication in short bursts (80 % amplitude, setting 3, for 45 seconds at 5 min intervals), then centrifuged to clarify the lysate. Clarified Supernatants containing GST-tagged PABPC1 were incubated for 30 mins-1hr on ice with 50 µg/ml DNase (50µL for every 50mL Lysate). The lysate was then filtered using 0.8µm and then 0.2µm syringe filter. The protein was loaded onto a GSTrap column and washed with wash buffer (10 mM Na2HPO4 (pH 7.4), 1.8 mM KH2PO4 (pH 7.4), 637 mM NaCl, 2.7 mM KCl and 1 mM dithiothreitol (DTT)) then eluted with a elution buffer (20mM Na2HPO4 (pH7.4), 100 mM NaCl and 20 mM glutathione). In order to remove the GST-Tag, the sample was incubated overnight with TEV protease. To remove reduced glutathione from sample, the protein was exchanged into heparin buffer (20 mM Na2HPO4 (pH 7.5), 100 mM NaCl, 10% glycerol, 2 mM dithiothreitol (DTT)). The sample was then further purified by FPLC with a

heparin column to remove nucleic acid contamination. Finally, the sample was concentrated and the buffer exchanged to a storage buffer for future use.

Preparation of RNA

Purified plasmids were digested at the 3' end of the insert sequence with EcoRI-HF to linearize the template with a 5' overhang for in vitro transcription. For 60 µL transcription reactions, about 11 µg of linearized DNA template was incubated for 4 h at 37 °C with 12.5 mM of each NTP, 3% (v/v) DMSO, 25 mM MgCl2, 17 mM DTT, and T7 RNA polymerase (12,500 units/mL) in transcription buffer composed of 0.4 M Tris-HCl (pH 8.1), 10 mM spermidine, and 0.01% (v/v) Triton X-100. The transcription product was extracted using acidic phenol chloroform (pH 4.5) and precipitated overnight in ethanol at –20 °C. RNA was redissolved in water. RNAs were capped using the Vaccinia capping system and poly(A)-tailed at their 3' ends with poly(A) polymerase, following the manufacturer's protocol. RNA concentration was quantified by UV absorbance spectrophotometry using a NanoDrop.

Electrophoretic Mobility Shift Assay

Various concentrations of viral or host RNA were added to varying concentrations of PABP or N protein in binding buffer (3xTBE, 500mM NaCl, 10mM DTT, 5 mg/mL BSA). Additional salt was added using a protein specific storage buffer, to bring the NaCl concentration to desired concentration, and glycerol concentration to 10 % (v/v) in all reactions. The binding reaction was incubated at room temp for 30 mins. The samples were then loaded onto a thin (~0.7cm tall) 0.5% TBE agarose gel. The gel was pre-run at 80V for 30 minutes, at 4 °C, then the binding reaction mixtures were loaded and the gel was electrophoresed at 80 V for 1 h at 4°C. The gel was post-stained with 0.5x Gel-Red for 15 minutes, then destained in H_2O twice for 15

minutes intervals. The gel was then visualized on a Bio-Rad ChemiDoc gel imager, using the Gel Red filter.

Single-Molecule Assay

Single-Molecule Experiments

A customized Pacific Biosciences RS II instrument was used to monitor real-time eIF4E interaction with the 5' UTR of the host and viral RNA. For the experiment, a zero-mode waveguide (ZMW) chip (Pacific Biosciences SMRT Cell; part number 100-171-800) was first treated with NeutrAvidin mixture (16.6 μM neutravidin, 0.67 mg/mL BSA) in smBuffer 30 mM HEPES-KOH pH 7.5, 3 mM Mg(OAc)2 and 100 mM KOAc) at room temperature for 5 min, which allowed the NeutrAvidin to bind the biotin-PEG layer on base of the chip. Polyadenylated viral RNA was prepared for immobilization through annealing to biotin-5'-(dT)-3'-Cy3, by heating a mixture containing 1 μM RNA and 100 nM oligo in 0.1 M BisTris pH 7.0, 0.3 M KCl to 98 °C for 2 minutes in a thermocycler, then slow-cooling to 4 °C over 20 minutes. The resulting hybrid duplex was immobilized by incubating 20 μL of the annealing mixture (10 nM RNA mixture final) on the chip surface for ~10 min at room temperature. The immobilization mixture was then removed, and the chip was washed three times with smBuffer. The chip was then pre-blocked with 1 μM unlabeled eIF4E, 5% (v/v) each Biolipidure 203 and 206, and 5 mg/mL purified BSA, to prevent non- specific interaction of Cy5-eIF4E with the chip surface. After pre- blocking, the chip was washed a further three times with smBuffer. After the third wash, 20 μL was added to the chip surface of smBuffer supplemented with BSA, with an oxygen scavenging system (2.5 mM PCA (protocatechuic acid), 1× PCD, and with triplet-state quencher. After loading the chip onto the instrument, movie files were acquired for 10 minutes at 10 frames

per second, under illumination with a 532 nm laser at a power of 0.70 μW/μm2, to excite Cy3 fluorophores.

Single-Molecule Data Analysis

Single-molecule fluorescence traces were extracted from raw movie files with custom MATLAB scripts. Traces containing FRET signals for eIF4E–RNA interaction were selected by visual inspection. Selection criteria based on multiple factors such as: a stable Cy3 signal at the very beginning of the movie or a single excitation signal, a single Cy3 photobleaching event at the end of the trace, and visually apparent FRET to Cy5 where the Cy3 signal moves down and Cy5 signal moves up. Assignment of event timings in the traces was carried out using a hidden Markov model approach. Event timings identified from analysis were converted to empirical cumulative probability distributions for the times between events, and the event durations, then fit with single- or double-exponential models to determine k on or k off, according to the equations:

$$P(t) = 1 - e^{-kt} \quad (1)$$

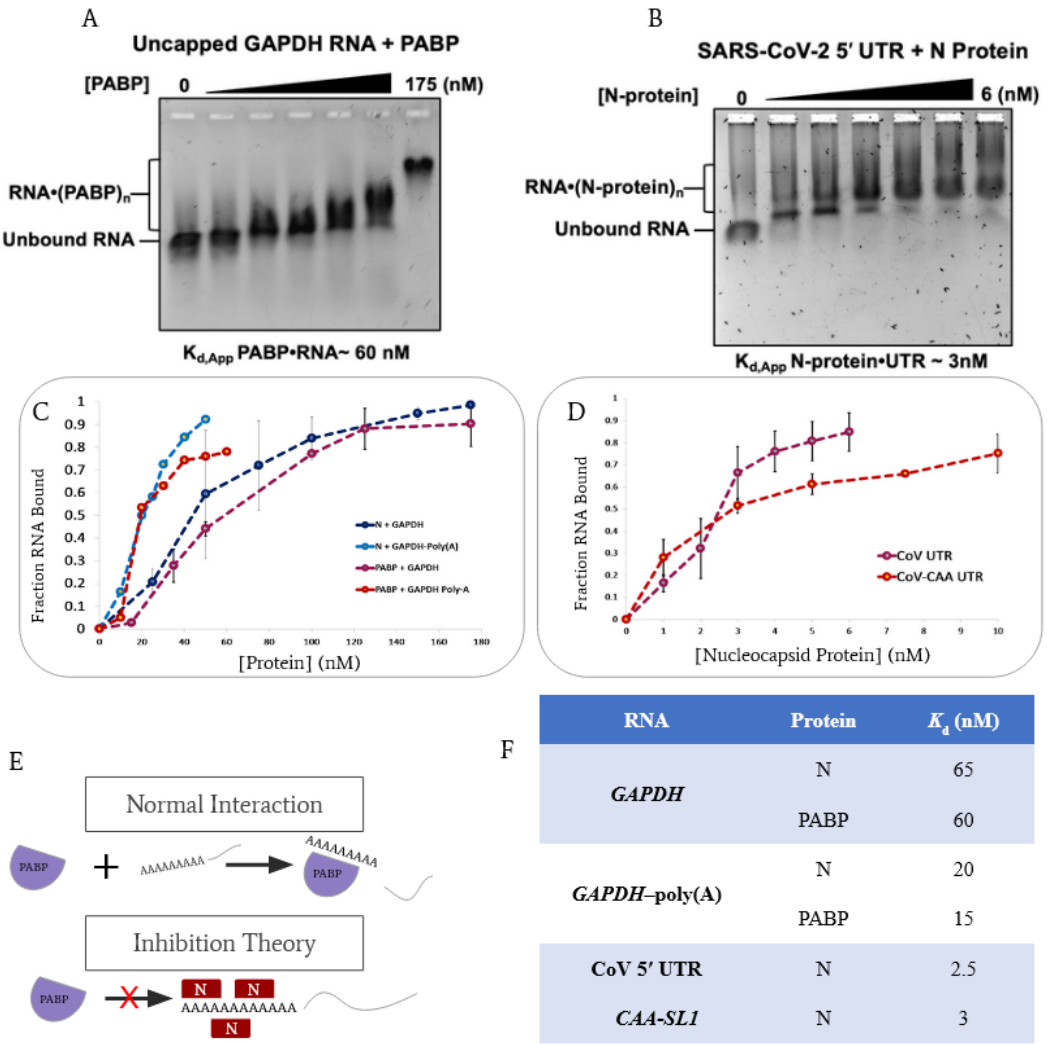
$$P(t) = A(1 - e^{-k1t}) + (1 - A)(1 - e^{-k2t}) \quad (2)$$

Fitting was carried out in MATLAB by non-linear least-squares regression using the Trust-Region algorithm. Goodness-of-fit was confirmed by inspection of the root-mean-squared errors for the fits.

Results & Discussion

N protein and PABP bind mRNA poly(A) tails comparably

We evaluated the binding affinity of N protein to RNA and the poly(A) tail by conducting Electrophoretic Mobility Shift Assays (EMSAs) on both human and viral RNAs, specifically the well translated host mRNA, *GAPDH*, and the CoV 5′ UTR, respectively. These mRNAs were titrated with either N protein or PABP, in order to compare the thermodynamics of binding between both proteins to mRNAs with and without a Poly(A) tail. (Figure 1 A and B)



F

RNA	Protein	K_d (nM)
GAPDH	N	65
	PABP	60
GAPDH–poly(A)	N	20
	PABP	15
CoV 5′ UTR	N	2.5
CAA-SL1	N	3

Figure 1: Thermodynamics of Protein RNA Interactions. A) Sample image of gel shift assay used to calculate the K_d of GAPDH RNA with PABP. **B)** Sample image of gel shift assay used to calculate the K_d of CoV 5′UTR RNA with N protein. **C)** Graph of N protein and GAPDH concentrations in nanomolar (x-axis) with fraction of RNA bound of both RNA with and without the poly(A) tail (y-axis) measured via fluorescence intensity of the bands from gel image and corrected for background. **D)** Graph of the concentration of N protein (x-axis) with fraction of RNA bound for the CoV 5′UTR with and without the first stem loop (y-axis). **E)** A model for the

potential inhibition by N protein via inhibition of PABP association to the Poly(A) tail. **F)** Chart of each measured K_d listed by RNA and Protein.

The N protein exhibited a high degree of specificity in its binding to the viral 5′ UTR of SARS-CoV-2, as evidenced by an equilibrium dissociation constant (K_d) of 5 nM (Figure 1D,F). To investigate the potential dependence of the N protein on the first stem loop of the CoV 5′ UTR, which previous studies have suggested to have implications for translation [Vora, et al.], we introduced CAA sequence repeats into the SARS-CoV-2 5′ UTR, resulting in a mutant strain that lacked the first stem loop (Figure 1D). Although the K_d values remained relatively unchanged, it became apparent that nearly twice the amount of N protein was required to saturate the solution, indicating a loss of cooperative binding. This diminished cooperativity suggests that in the absence of the first stem loop, the N protein is unable to effectively recruit other proteins, as it tends to form homomultimers [McBride, *et al*.]

We also assessed the binding affinity of the N protein to the human GAPDH mRNA (Glyceraldehyde 3-phosphate dehydrogenase), a gene that is known for its high translational activity within the human genome (Figure 1C). In the absence of a poly(A) tail, the equilibrium dissociation constant (K_d) was approximately 70 nM; however, in the presence of the poly(A) tail, the affinity notably increased (K_d ~20 nM). This enhanced affinity observed in the presence of the poly(A) tail implies that the N protein is specifically localized to the 3’ end poly(A) tail region of host RNA.

Upon evaluating Human PABPC1, we observed a greater affinity for polyadenylated GAPDH mRNA compared to its affinity for the untailed mRNA (K_d ~15 nM vs. ~60 nM in the

absence of the tail). The similarity in the affinities of the N protein and PABPC1 for the poly(A) tail suggests the existence of competitive binding between these proteins.

The binding of the N protein to the poly(A) tail carries two potential implications. Firstly, as the N protein is recognized as a virulent protein capable of triggering a host immune response [Yoshimoto *et al*, Wu *et al*], its binding to the poly(A) tails of host mRNA molecules may render them susceptible to deadenylase activity, thereby reducing the pool of host mRNA. Previous studies have reported such reduction in the host mRNA pool during late infection [Reed *et al*]. Secondly, the competitive binding between the N protein and PABP may lead to a decrease in the translational efficiency of host mRNA, while viral mRNA remains relatively unaffected due to the notably higher affinity of the N protein for the viral 5′ UTR.

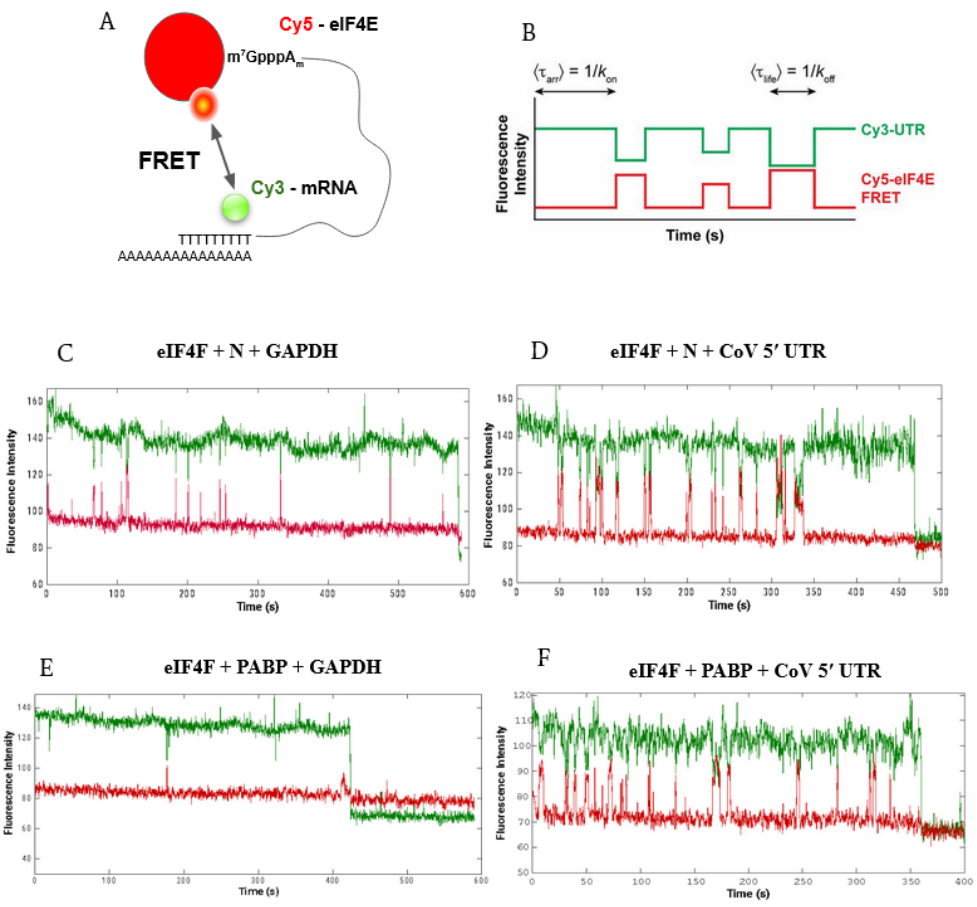


Figure 2: Sample traces of changes in eIF4F dynamics. A) Model of Cy5 - Cy3 interaction in single-molecule assay **B)** Sample assignment of trace **C)** eIF4F - N protein with GAPDH mRNA trace. **D)** eIF4F - N protein with CoV 5′UTR mRNA trace. **E)** eIF4F - PABP with GAPDH mRNA trace. **F)** eIF4F - PABP with CoV 5′UTR mRNA trace.

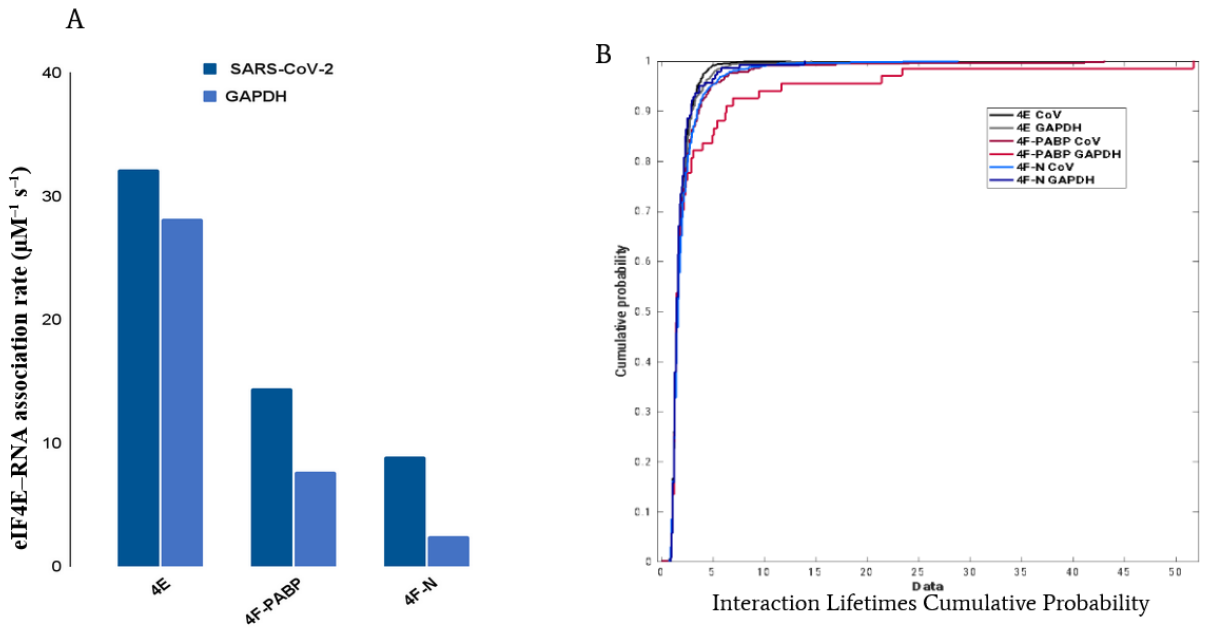


Figure 3: Single-Molecule Dynamics of N protein and PABP on Translation Initiation. A) eIF4E-mRNA association rates for eIF4E alone (15 nM), or in the eIF4F complex, for *GAPDH* and SARS-CoV-2 RNAs, and with PABP or N protein. **B)** Cumulative probability distribution of lifetimes of association rates for each molecule previously listed.

N protein advantages SARS-CoV-2 in binding to eIF4F

In order to elucidate and compare the impact of the N protein and PABP on the dynamics of eIF4F-mRNA recognition with both the CoV 5′ UTR and GAPDH, we employed a single-molecule fluorescence assay to determine the association and dissociation rates of eIF4F. To assess the association rate between the mRNA and eIF4E, we utilized fluorescent labeling with Cy5 (eIF4E) and Cy3 (mRNA) dyes (Figure 2A). When subjected to illumination with a 532 nm laser, the Cy3 label undergoes excitation and in close proximity to the Cy5 label, fluorescence resonance energy transfer (FRET) takes place. Since this interaction is distance-dependent, we can infer "cap" binding when FRET events occur.

The eIF4E subunit alone bound the CoV 5′ UTR with an association rate of 32 μM^{-1} s^{-1} similar to the association rate for *GAPDH* 28 μM^{-1} s^{-1} (Figure 2 C). However, when the trimeric eIF4F complex, comprising eIF4E along with eIF4G and eIF4A, was formed in the presence of N protein (20 nM) and delivered to the CoV 5′ UTR, the association rate of eIF4E markedly decreased to 9 μM^{-1} s^{-1}. This value closely resembles the association rate of eIF4F in the absence of N protein, approximately ~6 μM^{-1} s^{-1} [HeaJin Hong, unpublished data]. In contrast, when eIF4F and N protein were present, the association rate of eIF4E with GAPDH decreased even further, reaching 2 μM^{-1} s^{-1}. Remarkably, this value was significantly lower than the association rate of eIF4F in the absence of N protein, approximately 5 μM^{-1} s^{-1} [HeaJin Hong, unpublished data]. These findings indicate that N protein is responsible for suppression of binding in the eIF4F complex and its interaction with host mRNA.

The off rate of the CoV 5′ UTR in the presence of the eIF4F complex demonstrates that the presence of N protein leads to decreased dissociation rates for viral RNA. Conversely, the off rates for GAPDH remain unchanged compared to the rate observed with eIF4E alone (Figure 3B). Efficient translation of mRNA can be enhanced by two mechanisms that increase the probability of binding the 43S ribosomal subunit within the eIF4F complex. The first mechanism involves rapid association, where repeated interactions between eIF4F and mRNA eventually result in contact with the 43S subunit. The second mechanism involves decreasing the dissociation rate of the eIF4F-mRNA complex, prolonging the interaction time and facilitating contact with the 43S subunit. Due to the decreased off rates and normal on rates (in terms of eIF4F mRNA rates without N protein), viral mRNA gains a significant advantage in recruiting the 43S ribosomal subunit and initiating translation. On the other hand, the presence of N protein puts host mRNA at a disadvantage as it decreases association rates and increases dissociation

rates, thereby hindering both potential aspects of recruitment. The conclusions drawn from our previous data section support these findings, as they demonstrate that the N protein exhibits specificity towards the viral 5′ UTR, thereby enabling the mitigation of translational inhibition specifically for the virus, while negatively regulating host mRNA interaction with eIF4F.

Poly(A) Binding Protein increases eIF4F mRNA lifetimes

Upon the addition of PABP at a concentration of 20 nM, the association rate of eIF4F with the CoV 5′ UTR was measured to be 16 $\mu M^{-1}\ s^{-1}$ (Figure 3A). This value represents a nearly 3.1-fold increase compared to the rate observed in the absence of PABP (Hea Jin Hong, unpublished data), which aligns with the well-established role of PABP in enhancing cap-dependent translation [Gallie et al.]. Conversely, the association rate for GAPDH remained relatively unchanged, measuring 8 $\mu M^{-1}\ s^{-1}$, even with the presence of PABP.

Interestingly, the lifetimes of eIF4F cap binding events for both the CoV and GAPDH complexes were significantly prolonged when PABP was present (Figure 3B), indicating a stabilizing effect on the overall interaction. This observation is consistent with the concept of the "closed loop" model of translation. Significantly, the eIF4F + PABP complex with GAPDH demonstrated a noteworthy decrease in dissociation rates, emphasizing the capacity of PABP to enhance translation of host mRNA. This effect is achieved through the reduction of k_{off} values and the extension of complex lifetimes, ultimately increasing the probability of association with the 43S ribosomal subunit.

Conclusion

In this study, our objective was to gain a comprehensive understanding of the interactions between eIF4F and the viral genomic RNA, as well as to determine any preferential binding towards host or viral mRNA. Our findings reveal that the N protein plays a crucial role in conferring a translation initiation advantage for viral mRNA by negatively regulating the interactions between host mRNA and the eIF4F complex. Importantly, the viral mRNA avoids this negative regulation due to the specific binding preference of the N protein for the CoV 5′ UTR. By competitively binding to the poly(A) tail of host mRNA, the N protein potentially inhibits the closed loop model of translation and renders host mRNA susceptible to degradation by deadenylases, thus reducing the pool of available host mRNA for translation. Further investigations are required to fully elucidate the mechanisms underlying the N protein's interaction with host mRNA and to explore the downstream consequences of this interaction.

Furthermore, our study demonstrates that PABP follows a previously established mechanism for enhancing translation initiation. This involves decreasing the k_{off} rates and prolonging the lifetimes of translation initiation events, thereby facilitating the recruitment of the 43S ribosomal subunit. Future research endeavors should focus on investigating the dynamic interplay between the N protein and eIF4F to gain a comprehensive understanding of the factors contributing to the viral-preferred model of translation initiation.

Limitations

- The activity of eIF4E differs each prep due to concentration uncertainty and overall enzyme activity. This can lead to varying rates between each set of single-molecule experiments as the same concentration of eIF4E can lead to dispersed averages amongst

eIF4F association rates. To counteract this, we will test the activity of eIF4E before experimental use using a static quenching reaction similar to Slepenkov, *et al*.

- We only tested the 5′ untranslated region of the CoV mRNA, and not the full sequence, as our lab is not equipped to work with the full genomic RNA (~30 kb), or constructs to prepare it, which are biohazardous. Since the full mRNA construct is significantly larger, it thus may have different interactions with eIF4F than the isolated UTR. However, we are confident our results provide physiologically-relevant information on translation, as eIF4F interacts only with the viral RNA 5′ end during translation initiation.

Future Research

- We will test the effects of N protein on translation downstream of eIF4F–cap recognition through an *in-vitro* translation assay. With this we can test if N protein's interactions with eIF4F impacts translational efficiency or just initiation.
- We will quantify the dynamics of eIF4F–mRNA recognition in the presence of both PABP and N protein, in our single molecule assay. These experiments will probe whether the closed loop model is disrupted in the presence of N protein. They will also begin to answer many long asked questions on the closed loop model of translation.

References

Bartas, Martin, Adriana Volná, Christopher A Beaudoin, Ebbe Toftgaard Poulsen, Jiří Červeň, Václav Brázda, Vladimír Špunda, Tom L Blundell, and Petr Pečinka. "UNHEEDED SARS-COV-2 Proteins? A Deep Look into Negative-Sense RNA." *Briefings in Bioinformatics* 23, no. 3 (2022). https://doi.org/10.1093/bib/bbac045.

Bernstein, Philip, Stuart W. Peltz, and Jeffrey Ross. "The Poly(A)-Poly(A)-Binding Protein Complex Is a Major Determinant of Mrna Stability in Vitro." *Molecular and Cellular Biology* 9, no. 2 (1989): 659–70. https://doi.org/10.1128/mcb.9.2.659-670.1989.

De Breyne, Sylvain, Caroline Vindry, Olivia Guillin, Lionel Condé, Fabrice Mure, Henri Gruffat, Laurent Chavatte, and Théophile Ohlmann. "Translational Control of Coronaviruses." *Nucleic Acids Research* 48, no. 22 (2020): 12502–22. https://doi.org/10.1093/nar/gkaa1116.

Gallie, D R. "The Cap and Poly(a) Tail Function Synergistically to Regulate Mrna Translational Efficiency." *Genes & Development* 5, no. 11 (1991): 2108–16. https://doi.org/10.1101/gad.5.11.2108.

Gallie, D R. "The Cap and Poly(a) Tail Function Synergistically to Regulate Mrna Translational Efficiency." *Genes & Development* 5, no. 11 (1991): 2108–16. https://doi.org/10.1101/gad.5.11.2108.

Gingras, Anne-Claude, Brian Raught, and Nahum Sonenberg. "EIF4 Initiation Factors: Effectors of Mrna Recruitment to Ribosomes and Regulators of Translation." *Annual Review of Biochemistry* 68, no. 1 (1999): 913–63. https://doi.org/10.1146/annurev.biochem.68.1.913.

Gouveia, Duarte, Guylaine Miotello, Fabrice Gallais, Jean-Charles Gaillard, Stéphanie Debroas, Laurent Bellanger, Jean-Philippe Lavigne, et al. “Proteotyping SARS-COV-2 Virus from NASOPHARYNGEAL SWABS: A Proof-of-Concept Focused on a 3 Min Mass Spectrometry Window.” *Journal of Proteome Research* 19, no. 11 (2020): 4407–16. https://doi.org/10.1021/acs.jproteome.0c00535.

Harapan, Harapan, Naoya Itoh, Amanda Yufika, Wira Winardi, Synat Keam, Haypheng Te, Dewi Megawati, Zinatul Hayati, Abram L. Wagner, and Mudatsir Mudatsir. “Coronavirus Disease 2019 (COVID-19): A Literature Review.” *Journal of Infection and Public Health* 13, no. 5 (2020): 667–73. https://doi.org/10.1016/j.jiph.2020.03.019.

Hong, Hea Jin, Matthew G. Guevara, Eric Lin, and Seán E. O’Leary. “Single-Molecule Dynamics of SARS-COV-2 5’ Cap Recognition by Human EIF4F.” *BioRxiv*, 2021. https://doi.org/10.1101/2021.05.26.445185.

Kamitani, Wataru, Cheng Huang, Krishna Narayanan, Kumari G Lokugamage, and Shinji Makino. “A Two-Pronged Strategy to Suppress Host Protein Synthesis by SARS Coronavirus NSP1 Protein.” *Nature Structural & Molecular Biology* 16, no. 11 (2009): 1134–40. https://doi.org/10.1038/nsmb.1680.

Lo, Chen‐Yu, Tsung‐Lin Tsai, Chao‐Nan Lin, Ching‐Hung Lin, and Hung‐Yi Wu. “Interaction of Coronavirus Nucleocapsid Protein with the 5′‐ and 3′‐ends of the Coronavirus Genome Is Involved in Genome Circularization and Negative‐strand RNA Synthesis.” *The FEBS Journal* 286, no. 16 (2019): 3222–39. https://doi.org/10.1111/febs.14863.

Lokugamage, Kumari G., Krishna Narayanan, Cheng Huang, and Shinji Makino. “Severe Acute Respiratory Syndrome Coronavirus Protein NSP1 Is a Novel Eukaryotic Translation Inhibitor That Represses Multiple Steps of Translation Initiation.” *Journal of Virology* 86, no. 24 (2012): 13598–608. https://doi.org/10.1128/jvi.01958-12.

McBride, Ruth, Marjorie van Zyl, and Burtram Fielding. “The Coronavirus Nucleocapsid Is a Multifunctional Protein.” *Viruses* 6, no. 8 (2014): 2991–3018. https://doi.org/10.3390/v6082991.

Miao, Zhichao, Antonin Tidu, Gilbert Eriani, and Franck Martin. “Secondary Structure of the SARS-COV-2 5’-UTR.” *RNA Biology* 18, no. 4 (2020): 447–56. https://doi.org/10.1080/15476286.2020.1814556.

Mu, Jingfang, Jiuyue Xu, Leike Zhang, Ting Shu, Di Wu, Muhan Huang, Yujie Ren, et al. “SARS-COV-2-Encoded Nucleocapsid Protein Acts as a Viral Suppressor of RNA Interference in Cells.” *Science China Life Sciences* 63, no. 9 (2020): 1413–16. https://doi.org/10.1007/s11427-020-1692-1.

Nabeel-Shah, Syed, Hyunmin Lee, Nujhat Ahmed, Giovanni L. Burke, Shaghayegh Farhangmehr, Kanwal Ashraf, Shuye Pu, et al. “SARS-COV-2 Nucleocapsid Protein Binds Host Mrnas and Attenuates Stress Granules to Impair Host Stress Response.” *iScience* 25, no. 1 (2022): 103562. https://doi.org/10.1016/j.isci.2021.103562.

Passmore, Lori A., and Jeff Coller. “Roles of Mrna Poly(a) Tails in Regulation of Eukaryotic Gene Expression.” *Nature Reviews Molecular Cell Biology* 23, no. 2 (2021): 93–106. https://doi.org/10.1038/s41580-021-00417-y.

Reed, Mark L., Brian K. Dove, Richard M. Jackson, Rebecca Collins, Gavin Brooks, and Julian A. Hiscox. “Delineation and Modelling of a Nucleolar Retention Signal in the Coronavirus Nucleocapsid Protein.” *Traffic* 7, no. 7 (2006): 833–48. https://doi.org/10.1111/j.1600-0854.2006.00424.x.

Setten, Ryan L., John J. Rossi, and Si-ping Han. “The Current State and Future Directions of Rnai-Based Therapeutics.” *Nature Reviews Drug Discovery* 18, no. 6 (2019): 421–46. https://doi.org/10.1038/s41573-019-0017-4.

Slobodin, Boris, Urmila Sehrawat, Anastasia Lev, Ariel Ogran, Davide Fraticelli, Daniel Hayat, Binyamin Zuckerman, et al. *Cap-independent translation and a precisely localized RNA sequence enable SARS-COV-2 to control host translation and escape anti-viral response*, 2021. https://doi.org/10.1101/2021.08.18.456855.

Sonenberg, Nahum. “Protein Synthesis in Eukaryotes: Internal Initiation.” *eLS*, 2001. https://doi.org/10.1038/npg.els.0000544.

Surjit, Milan, and Sunil K. Lal. “The Nucleocapsid Protein of the SARS Coronavirus: Structure, Function and Therapeutic Potential.” *Molecular Biology of the SARS-Coronavirus*, 2009, 129–51. https://doi.org/10.1007/978-3-642-03683-5_9.

Surjit, Milan, Ravinder Kumar, Rabi N. Mishra, Malireddy K. Reddy, Vincent T. Chow, and Sunil K. Lal. “The Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein Is Phosphorylated and Localizes in the Cytoplasm by 14-3-3-Mediated Translocation.” *Journal of Virology* 79, no. 17 (2005): 11476–86. https://doi.org/10.1128/jvi.79.17.11476-11486.2005.

Tsai, Tsung-Lin, Ching-Houng Lin, Chao-Nan Lin, Chen-Yu Lo, and Hung-Yi Wu. “Interplay between the Poly(a) Tail, Poly(a)-Binding Protein, and Coronavirus Nucleocapsid Protein Regulates Gene Expression of Coronavirus and the Host Cell.” *Journal of Virology* 92, no. 23 (2018). https://doi.org/10.1128/jvi.01162-18.

Vora, Setu M., Pietro Fontana, Tianyang Mao, Valerie Leger, Ying Zhang, Tian-Min Fu, Judy Lieberman, et al. “Targeting Stem-Loop 1 of the SARS-COV-2 5′ UTR to Suppress Viral Translation and NSP1 Evasion.” *Proceedings of the National Academy of Sciences* 119, no. 9 (2022). https://doi.org/10.1073/pnas.2117198119.

Wells, Sandra E., Paul E. Hillner, Ronald D. Vale, and Alan B. Sachs. “Circularization of Mrna by Eukaryotic Translation Initiation Factors.” *Molecular Cell* 2, no. 1 (1998): 135–40. https://doi.org/10.1016/s1097-2765(00)80122-7.

Wu, Wenbing, Ying Cheng, Hong Zhou, Changzhen Sun, and Shujun Zhang. “The SARS-COV-2 Nucleocapsid Protein: Its Role in the Viral Life Cycle, Structure and Functions, and Use as a Potential Target in the Development of Vaccines and Diagnostics.” *Virology Journal* 20, no. 1 (2023). https://doi.org/10.1186/s12985-023-01968-6.

Ye, Qiaozhen, Alan M.V. West, Steve Silletti, and Kevin D. Corbett. “Architecture and Self-Assembly of the SARS-COV-2 Nucleocapsid Protein.” *bioRxiv*, 2020. https://doi.org/10.1101/2020.05.17.100685.

Yoshimoto, Francis K. “The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS COV-2 or N-COV19), the Cause of Covid-19.” *The Protein Journal* 39, no. 3 (2020): 198–216. https://doi.org/10.1007/s10930-020-09901-4.