Background Information for Chen et al (V1)
(Propel Literature Review Dec 2022)

Overall Viral Strategy During Early Infection

1) What are some of the first things a virus must do after it infects a cell.

First, the virus needs to redirect the cells macromolecular synthesis capacity from the cell to the virus. This requires that the virus inhibits global macromolecular synthesis by the cell including protein synthesis. However, the virus needs to make sure it's protein synthesis escapes this inhibition, and it usually does this by distinguishing its mRNA with sequence signals that protect it from the inhibition. So this redirection of the protein synthetic machinery has two parts: inhibiting the host cells machinery and allowing viral mRNAs to evade this inhibition.

Second, the virus needs to protect itself from the intracellular innate immune mechanisms that will induce the host cell to commit suicide (e.g. apoptosis) or be killed once the intracellular pathogen is detected. This protection often involves inhibition of signaling pathways that detect pathogen markers. Since some of these pathways involve new protein synthesis (like interferon), inhibition of global host protein synthesis can contribute significantly to suppression of this immune response.

2) What is a + stranded RNA virus?

For double stranded DNA genes, we can distinguish between the two strands based on which strand is actually transcribed into translatable mRNA (which we call the sense strand), and which strand acts as the template for mRNA synthesis during transcription (which we call the antisense strand).

For viruses with single stranded RNA genomes, replication of the genome requires synthesis of the complementary strand using the single-stranded RNA genome as template, followed by use of this complementary strand as the template for making more copies of the single-stranded RNA genome. So both strands will end up being synthesized in infected cells, but usually all the viral proteins come from translation of only one of these two RNA strands, allowing us to functionally distinguish between them. The strand that is translated is called the + strand and the nontranslated complementary strand is called the – strand.

SARS-Cov-2 is a + stranded RNA virus. That means the RNA strand that gets packaged into the virus is the one that is translated into protein. In essence, the genome also doubles as an mRNA ready to be translated once it gets released into the cell.

3) How does SARS-Cov-2 make so many proteins quickly from its genome

Although the SARS-Cov-2 genome can be translated immediately upon release into the cell, there's a challenge it must deal with. Only a single protein is synthesized from each eukaryotic mRNA, and the coding sequence for this protein starts from the AUG translational start codon closest to the 5' end of the mRNA. The ribosome, which travels down the mRNA in a 5' to 3' direction, is instructed by each 3 nucleotides codon after the start codon to add another amino acid to the peptide chain, which is growing from the N-terminus to the C-terminus. Upon reaching a stop codon (UAA, UAG, or UGA), the completed protein is released and the ribosome disassembles, preventing any downstream sequence from being translated. Thus, even if the virus had multiple independent coding sequences in its + strand genome, only the first one closest to the 5' end could initiate translation.

SARS-Cov-2 needs to synthesize multiple proteins quickly from its mRNA-like genome. How does it get around the one mRNA one protein initiation rule? It physically links all the proteins it needs early in infection end to end, so that synthesis of all of them is dependent on initiation of the first protein. What the virus does is synthesize two large polyproteins (Nsp1-11 and Nsp1-16), then use proteases specialized to cleave between the proteins to release individual proteins.

4) Do any of the early proteins synthesized suppress global protein synthesis?

In the two polyproteins synthesized by translation of SARS-Cov-2 mRNA/genome the very first protein in the polyprotein is **Nsp1**. Hence, it is one of the first newly synthesized proteins to accumulate to high levels.

Previous studies on SARS-Cov-1, which shares similar genome structure as SARS-Cov-2, had shown that SARS-Cov-1 Nsp1 inhibited host protein synthesis by binding to the 40S ribosome and by stimulating endonucleolytic cleavage of RNA associated with the ribosome. Hence, the authors knew to focus on SARS-Cov-2 Nsp1 interactions with the ribosome.

Mechanism of Eukaryotic Protein Translation Initiation

(need to understand a mechanisms before you can understand its inhibition)

5) What is Cap-dependent translation initiation?

There are two fundamental types of translation initiation in eukaryotes. The vast majority of mRNAs, including SARS-Cov-2 mRNAs, are translated in a cap-dependent manner. This cap refers to a special chemical structure added to the 5' end of mRNAs soon after their synthesis has commenced. The very first (5' most) nucleotide of transcribed mRNAs has a triphosphate at its 5' end. The cap consists of a modified **(7-methylated) guanosine** that is hooked up "back-to-back" (5' to 5') to this 5' triphosphate.

This **cap structure** is recognized by the translation machinery as the place on the mRNA to start the assembly of the ribosome. As discussed in point 2 above, actual initiation of protein

synthesis usually begins at the **first start codon** (AUG, which encodes the amino acid methionine) in from the cap and ends at the first in frame STOP codon.

6) What is IRES-dependent translation initiation?

This is not relevant to the paper, but is discussed for completeness sake. There is another less frequently used translation initiation strategy used in eukaryotes. This involves a specialized RNA sequence internal to the mRNA that usually folds into a complex structure containing many base-paired hairpins and higher order tertiary interactions. These internal ribosome entry sites (IRES) are usually just upstream of a start codon and they direct the assembly of the ribosome at that start codon to initiate protein synthesis.

With IRESs a eukaryotic mRNA can direct the independent translation initiation of more than one protein, one through the cap-dependent assembly of a ribosome at the 5' end of the mRNA and one through IRES-directed assembly of a ribosome at an internal site on the mRNA.

These IRESs are often used as a molecular biology tool to mark the presence and amount of an mRNA that expresses a protein of interest in a cap-dependent. A reporter gene encoding a fluorescent protein expressed in an IRES-dependent manner can be added 3' to the coding sequence for the protein of interest. The intensity of the fluorescent reporter protein in the cell provides a relative measure of the amount of the marked mRNA in cells.

7) Some basics about the structure and function of ribosomes?

Ribosomes have two subunits, a **large subunit** and a **small subunit**. Each subunit is composed of about 60% ribosomal RNA (rRNA) and 40% proteins. Together the two subunits assemble around the mRNA so that the mRNA is within a **channel** of the ribosome (mainly generated by the small subunit) and the ribosome can slide down the mRNA in a 5' to 3' direction and get instructed by sequential 3 nucleotide codons that specify which amino acid to add next during protein synthesis.

This amino acid covalent linkage activity, which Is called **peptidyl transferase**, is the core catalytic activity of the ribosome. This activity involves the formation of a peptide bond between the 3' C-terminus of a growing peptide chain and the 5' N-terminus of the next amino acid. Much of the huge structure of the ribosome is designed to ensure that this catalytic activity is used with high **fidelity**, i.e. make sure that the correct amino acids are added in the correct order. Both functions of catalysis and fidelity are primarily carried out by the rRNA in the core of each subunit, making the ribosome a **ribozyme** (i.e. RNA based enzyme).

Most of the ribosomal proteins are on the outer surface of the subunits, and in most cases their functions are poorly understood. In fact, in many cases their full structure is poorly understood. The parts that are embedded in the RNA are rigidly positioned enough that cryo-EM can detect specific atomic positions and derive a structure for them. But the parts that might be hanging on the outside of the ribosome don't yield any detectable atomic structure, presumably because these parts flap around too much for a specific structure to be detected.

8) How are the different ribosome structures monitored?

To monitor the different populations of small subunit, large subunit, and full ribosome you need ways to separate these huge macromolecular assemblies and detect them. The individual proteins or rRNA can be separated based on size by gel electrophoresis, but the ribosomal subunits and ribosomes are too large.

Instead, the classic method has been to separate them roughly by size using **centrifugation** through a high density **sucrose gradient**. The gradient is created in a centrifuge tube and the sample containing ribosome components is layered on top. Then the tube is spun in a special rotor so that tubes can swing out, parallel to the horizontal centrifugal force on the tube. This causes the ribosomal components to spin through the dense sucrose gradient straight toward the bottom of the tube.

How fast a ribosome component spins toward the bottom (i.e. its **sedimentation velocity**) depends on the component's **sedimentation coefficient**, which for globular assemblies roughly depends on the size of the assembly. The tube is spun just long enough to maximally separate the components down the length of the tube without having any pellet at the bottom. Sedimentation coefficient standards are added to each sample before they are spun. The standards allow one to correlate distance down the tube with sedimentation coefficient.

After the spin, different layers of the gradient are collected. One way is by dripping the gradient out of a hole punctured in the bottom of the centrifuge tube into a series of small test tubes. The ribosomal components are typically detected by the strong UV absorbance (260 nm wavelength) of the rRNA. The position of the sedimentation standards in each centrifuge tube are usually detected by running a gel of a small portion each fraction. Absorbance peaks at tube positions corresponding to 40S, 60S, and 80S indicate the presence, respectively, of the small ribosomal subunit, the large ribosomal subunit and the whole ribosome. Thus, the small eukaryotic ribosomal subunit is called the **40S subunit** and the large subunit is called the **60S subunit**.

The 80S ribosome is usually assembled on an mRNA to translate it. mRNA experiencing high rates of translation will often have 2, 3, or more ribosomes that had sequentially loaded on and were all traveling down the mRNA in a 5' to 3' direction. These are called polysomes. Each subpopulation of polysomes containing a specific number of ribosomes (> 1) will be represented by an additional A260 absorbance peak past the 80S peak.

9) How does ribosome assembly occur during Cap-dependent translation?

During this early step in the initiation stage of translation, the cell first brings together 3 key components: (1) the 5' cap of the mRNA; (2) methionyl-tRNA (the methionine amino acid covalently linked to a tRNA), which brings in the first amino acid for protein synthesis because its anticodon sequence base pairs with AUG start codons; (3) the 40S subunit. This assembly is achieved by large multiprotein **eukaryotic initiation factor (eIFs**) that recognize and bind these components, and help bring them together.

When the methionyl-tRNA bound to its eIFs is docked into the correct position on a 40S subunit bound to its eIFs, the resulting **pre-initiation complex (PIC)** sediments during sucrose gradient centrifugation slightly faster as a **43S complex**.

This 43S complex is then recruited by cap-binding eIFs to the 5' end of the mRNA. The mRNA is somehow threaded through the RNA channel of this 43S complex to form an even larger **48S complex** containing all three basic components. If you can block the entry of mRNA into the RNA channel you can block translation initiation.

After assembly of the 48S complex at the cap end of the mRNA, the 40S subunit needs to find the fourth component for initiating protein synthesis, the first AUG start codon downstream of the cap. This search is performed by the 40S subunit translocating down the mRNA in a 5' to 3' direction using the methionyl-tRNA anticodon to find the first AUG based on its ability to base pair with the tRNA.

Once the first AUG has been recognized, the fifth and final component can be added, the large (60S) ribosomal subunit. This step, which accompanied by release of many of the eIFs, completes the assembly and positioning of the ribosome at the first start codon downstream of the 5' cap. The translation machinery is now ready to start protein synthesis.

Viral strategies for commandeering host protein synthesis machinery

10) Two big questions regarding how the virus affects the protein translation machinery

As discussed in point 1 in order to wrest control of protein synthesis from the host cell, the virus needs to inhibit host protein synthesis, especially for proteins that can participate in the innate immune response to suppress viral replication or promote host cell death. Hence, one of the major questions in Schubert et al. is how does Nsp1 contribute to this inhibition of the protein translation machinery.

However, if all protein synthesis is universally suppressed, that would hamper the virus's own protein synthesis which is necessary to produce more virus. Hence, a second major question for the paper is does the virus's mRNA escape Nsp1 inhibition of the protein translation machinery, and if so, how?

11) Speculative ideas about how one might design an inhibitory mechanism for protein translation that has the ability to spare viral protein translation

From first principles one can imagine that viral specific escape from any protein translation inhibition might require marking the viral mRNA as being distinct from host mRNA. The simplest way to do this is to have the mark embedded in the sequence of the viral mRNA, presumably outside the protein coding sequence of the mRNA, where the sequence is constrained by the requirement to encode functional proteins under selective pressure. That makes the 5' and 3' untranslated regions of viral mRNA appealing possible locations for such a mark.

One can also imagine that if the distinguishing mark is on the mRNA itself, then a good way to make Nsp1 inhibition be sensitive to that mark is to have the inhibition affect a step where mRNA recognition and incorporation into the translation machinery occurs. You might imagine that if Nsp1 were inhibiting a step for which it is important to be agnostic to the identity of the mRNA, for example, the catalytic peptidyl transferase step, it might be harder to make that inhibition sensitive to any distinguishing marks on the mRNA.