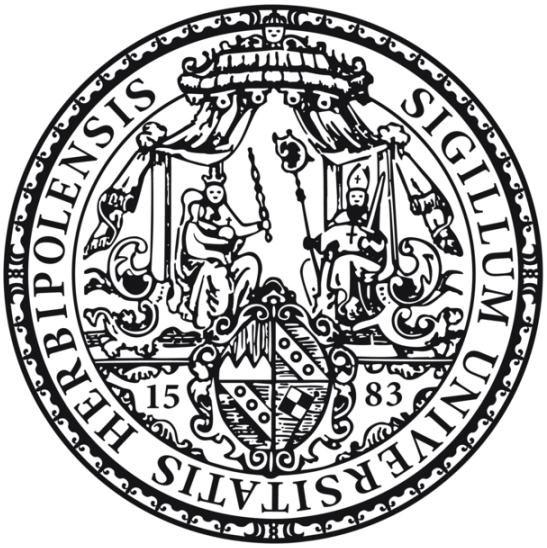


Translational landscape and regulation of recoding in virus-infected cells

Translationslandschaft und Regulierung der Rekodierung in virusinfizierten Zellen

Doctoral thesis for a doctoral degree

at the Graduate School of Life Sciences,

Julius-Maximilians-Universität Würzburg,

Section Infection and Immunity

submitted by

Anuja Kibe

from

Pune, India

Würzburg 2022

Submitted on: ……..

Office stamp

Members of the Thesis Committee

Chairperson: Prof. Dr. Christian Janzen

Primary Supervisor: Prof. Dr. Neva Caliskan

Supervisor (Second): Prof. Dr. Redmond Smyth

Supervisor (Third): Prof. Dr. Florian Erhard

Supervisor (Fourth): Prof. Dr. Luka Cicin-Sain

Abstract

RNA viruses rely entirely on the host machinery for their protein synthesis and harbor non-canonical translation mechanisms, such as alternative initiation and programmed –1 ribosomal frameshifting (–1PRF), to suit their specific needs. On the other hand, host cells have developed a variety of defensive strategies to safeguard their translational apparatus and at times transiently shut down global translation. An infection can lead to substantial translational remodeling in cells and translational control is critical during antiviral response. Due to their sheer diversity, this control is likely unique to each RNA virus and the intricacies of post-transcriptional regulation are unclear in certain viral species.

Here, we explored different aspects of translational regulation in virus-infected cells in detail. Using ribosome profiling, we extensively characterized the translational landscape in HIV-1 infected T cells, uncovering novel features of gene regulation in both host and virus. Additionally, we show that substantial pausing occurs prior to the frameshift site indicating complex regulatory mechanisms involving upstream viral RNA elements that can act as *cis*-regulators of frameshifting.

We also characterized the mechanistic details of *trans*- modulation of frameshifting by host- and virus-encoded proteins. Host antiviral protein ZAP-S binds to the SARS-CoV-2 frameshift site and destabilizes the stimulatory structure, leading to frameshift inhibition. On the other hand, EMCV 2A protein stabilizes the viral frameshift site, thereby, activating EMCV frameshifting. While both proteins were shown to be antagonistic in their mechanism, they interact with the host translational machinery. Furthermore, we showed that frameshifting can be regulated not just by proteins, but also by small molecules. High-throughput screening of natural and synthetic compounds identified two potent frameshift inhibitors that also impeded viral replication, namely trichangion and compound **25**. Together, this work largely enhances our understanding of gene regulation mechanisms in virus-infected cells and further validates the druggability of viral –1 PRF site.

Zussamenfassung

RNA-Viren sind bei der Proteinsynthese vollständig auf die Maschinerie des Wirts angewiesen und verfügen über nicht-kanonische Translationsmechanismen wie alternative Initiation und –1 programmiertes ribosomales Frameshifting (–1PRF), um ihre spezifischen Bedürfnisse zu erfüllen. Auf der anderen Seite haben die Wirtszellen eine Vielzahl von Abwehrstrategien entwickelt, um ihren Translationsapparat zu schützen und die globale Translation gegebenenfalls vorübergehend abzuschalten. Eine Infektion kann zu einer erheblichen Umgestaltung der Translation in den Zellen führen und die Kontrolle der Translation ist für die antivirale Reaktion von entscheidender Bedeutung. Aufgrund ihrer großen Vielfalt ist diese Kontrolle wahrscheinlich für jedes RNA-Virus einzigartig, und die Feinheiten der posttranskriptionellen Regulierung sind bei bestimmten Virusarten noch unklar. Hier haben wir verschiedene Aspekte der Translationsregulation in virusinfizierten Zellen im Detail untersucht. Mithilfe von Ribosomen-Profiling haben wir die Translationslandschaft in HIV-1-infizierten T-Zellen umfassend charakterisiert und dabei neue Merkmale der Genregulation sowohl im Wirt als auch im Virus aufgedeckt. Darüber hinaus konnten wir zeigen, dass Ribosomen vor der Frameshift-Stelle zu einem erheblichen Maße pausieren, was auf komplexe Regulationsmechanismen hinweist, an denen vorgelagerte virale RNA-Elemente beteiligt sind, die als *cis*-Regulatoren des Frameshifting wirken können. Darüber hinaus haben wir die mechanistischen Details der trans-Modulation des Frameshifting durch vom Wirt und vom Virus kodierte Proteine charakterisiert. Das antivirale Wirtsprotein ZAP-S bindet an die SARS-CoV-2 Frameshift-Stelle und destabilisiert die stimulierende Struktur, was zu einer Hemmung des Frameshifting führt. Auf der anderen Seite stabilisiert das EMCV-2A-Protein die virale Frameshift-Stelle und aktiviert dadurch das EMCV-Frameshifting. Obwohl sich beide Proteine in ihrem Mechanismus als antagonistisch erwiesen haben, interagieren sie mit der Translationsmaschinerie des Wirts. Darüber hinaus haben wir gezeigt, dass das Frameshifting nicht nur durch Proteine, sondern auch durch kleine Moleküle reguliert werden kann. Durch ein Hochdurchsatz-Screening natürlicher und synthetischer Verbindungen wurden zwei potente Frameshift-Inhibitoren identifiziert, die auch die virale Replikation behinderten, nämlich Trichangion und Compound 25. Zusammengenommen verbessert diese Arbeit unser Verständnis der Mechanismen der Genregulierung in virusinfizierten Zellen und bestätigt die Medikamentenfähigkeit der viralen -1 PRF-Seite.

Contents

Chapter 1

1. General introduction

Translation, a central process in gene expression, converts genomic information provided by the messenger RNA (mRNA) into a functional protein. Eukaryotic translation is complex and cyclical, consisting of four main phases: initiation, elongation, termination and ribosome recycling [1-6]. Each step is orchestrated by a myriad of cellular factors and regulatory mechanisms. Controlling gene expression by mRNA translation enables rapid responses to environmental and physiological stresses. Such control is especially critical during viral infection, where substantial translational remodeling may occur due to the absolute dependence of RNA- and DNA- viruses on host cellular machinery. Optimal viral protein synthesis occurs at the expense of cellular proteins, and many viruses have evolved non-canonical mechanisms to customize translation to meet their specific needs. Such alternative viral strategies lead to deviations from the rules that seem relatively consistent across the translation of cellular genes. However, to appreciate the diversity of non-canonical translation mechanisms and their regulation, it is essential to understand the standard canonical translation of cellular mRNAs.

1.1 Canonical mechanism of eukaryotic translation

Canonically, the majority of mRNAs in eukaryotes possess a specialized 'label' – namely the 7-methylguanosine cap (m^7G) on the 5' end, which, when 'activated', is bound to eIF4F – a multifactorial complex consisting of cap-binding eIF4E, an RNA-helicase eIF4A and a large scaffold eIF4G [7, 8]. Similarly, the 3' end of eukaryotic mRNA is a poly (A) tail that is associated with the poly(A)-binding protein (PABP), which can then interact with eIF4G to form a closed loop [2, 6, 9]. Eukaryotic translation begins with the formation of a ternary complex comprising of Met-$tRNA_i$ and eIF2+GTP [2, 6, 10, 11] (**Figure 1**; Step 1). This complex is then loaded onto the 40S subunit of the ribosome and, along with eIF1, eIF1A, eIF3, and eIF5 forms a 43S pre-initiation complex (PIC) (**Figure 1**; Step 2). Then, in a process facilitated by eIF4F, the 5' capped end of the mRNA recruits the 43S PIC, which starts scanning for an AUG initiation codon (**Figure 1**; Step 3). Initiation factors eIF1 and eIF1A are vital in recognizing the AUG codon and surrounding nucleotide context. Upon encountering the start codon, base-pairing between the mRNA AUG codon and the P-site-

bound Met-tRNA$_i$ anticodon halts the scanning process. eIF2-bound GTP can be hydrolyzed during scanning in an eIF5-dependent reaction and dephosphorylation of eIF2 occurs on AUG recognition [12, 13]. Together, these changes trigger a conformational change in eIF2 and stimulated by the binding of eIF5B, eIF2.GDP dissociates in complex with eIF5. The PIC now bound by eIF5B.GTP accelerates the recruitment of the 60S subunit, which is thought to be activated by the release of eIF6 [14, 15]. Hydrolysis of GTP associated with eIF5B triggers the release of eIF5B and eIF1A [16]. Initiation ends with the formation of the 80S initiation complex with the ribosomal A-site ready to accept an amino-acyl transfer RNA (tRNA) and enter the elongation stage (**Figure 1**; Steps 4,5).

During the elongation phase, cognate amino-acylated tRNAs which match the mRNA codon at the A-site of the ribosome are delivered to the ribosomal A-site by elongation factor eEF1A (**Figure 1**; Step 6). Next, a peptide bond is formed between the incoming amino acid and the peptidyl-tRNA in the P-site, and the growing nascent polypeptide chain is transferred from the P-site tRNA to the A-site tRNA (**Figure 1**; Step 7). Then, translocation, mediated by elongation factor eEF2, transfers the deacetylated P-site tRNA to the E-site, and the peptidyl A-site tRNA to the P-site (**Figure 1**; Step 8). This process also moves the mRNA through the ribosome, and the A-site is now open for the next elongation cycle. Translational fidelity is highest during the translocation step, where spontaneous frameshifts occur only at a rate of 10^{-5} per codon [17]. The conserved eEF2 post-translational diphthamide modification is critical for maintaining the translational reading frame [18, 19]. The elongation cycle is repeated until every codon is translated and the ribosome reaches a stop codon, where termination occurs (**Figure 1**; Step 9).

Translation termination is mediated by release factors eRF1 and eRF3 when the stop codon (UAA, UAG and UGA) enters the ribosomal A-site (**Figure** 1; Steps 9, 10). eRF1, which is similar in size and shape to a tRNA, explicitly recognizes the stop codons and hydrolyzes peptidyl tRNA [20-22]. eRF3, a translational GTPase, promotes peptide release and increases termination efficiency in a GTP hydrolysis-dependent manner [23]. The eRF3-eRF1 complex interacts with the ribosome, following which eRF3 dissociates upon GTP hydrolysis. Then, eRF1 transfers a water molecule at the peptidyl-transferase center that hydrolyzes the nascent peptide, thus releasing the peptide and completing the termination process. Next, facilitated by the conserved ATP-binding cassette sub-family E member 1 (ABCE1) and the release factors, the post-termination 80S ribosome complex is recycled and disassembled into subunits, making it ready for the next round of translation.

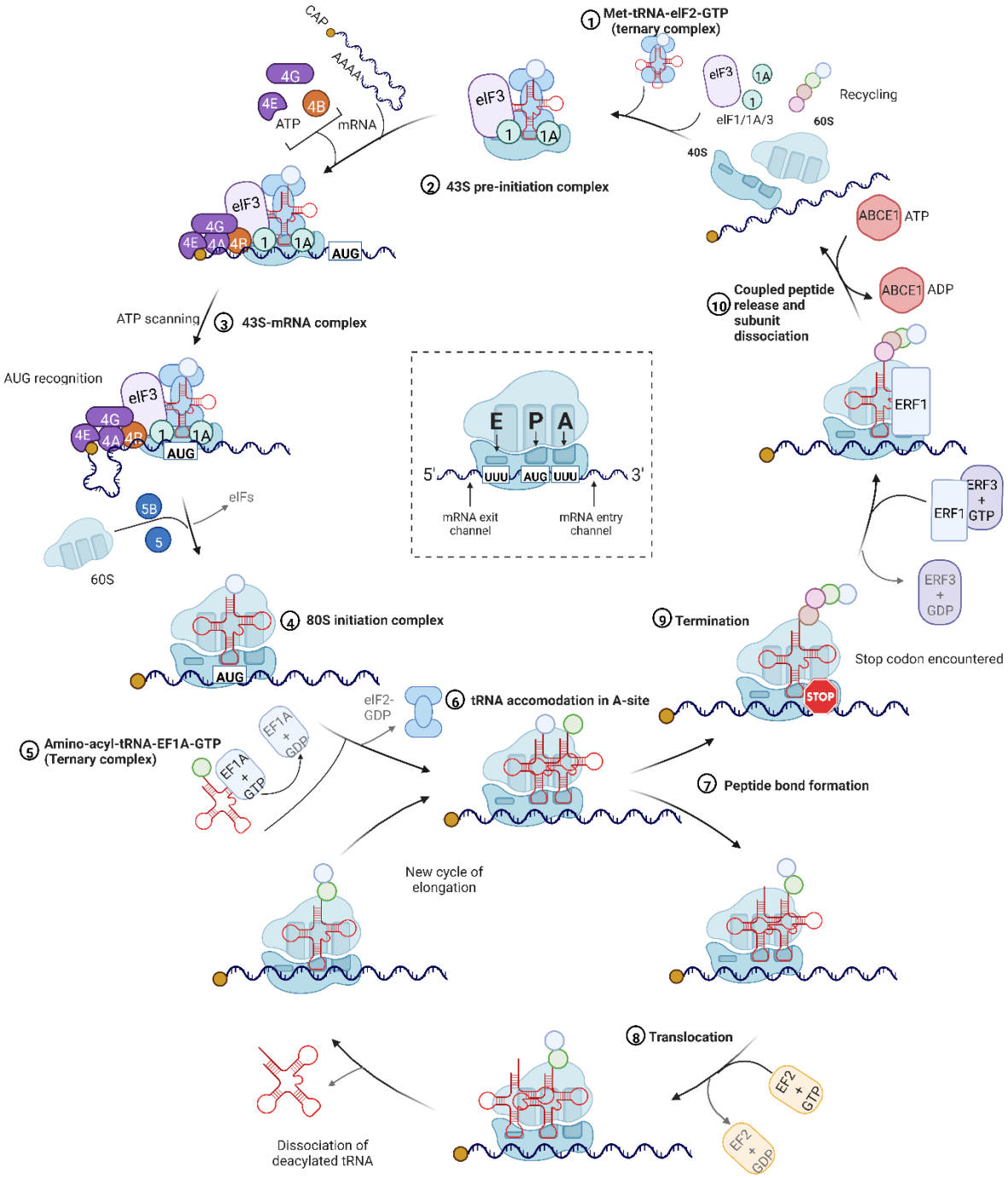


Figure 1: Schematic overview of the canonical eukaryotic translation process. (1) Formation of the ternary complex eIF2-GTP-Met-tRNA$_i$Met, (2) formation of the 43S pre-initiation complex upon interaction with the 40S subunit and initiation factors, (3) Activation of the mRNA by binding of the eIF4F complex to the 5' cap structure is followed by binding of the 43S PIC to the eIF4F complex and the mRNA scanning along the 5' UTR in the 5'-3'direction. Upon recognition of the AUG start codon and formation of the 48S initiation complex, the 60S subunit binds to the 48S initiation complex catalyzed by eIF5B, releasing eIF1A and eIF5B to (4) assemble an elongation competent 80S ribosome. (5) Elongating amino-acyl-tRNAs are recruited by a ternary complex of eEF1A, amino-acyl-tRNA and GTP (6) and accommodated in the ribosomal

A-site. Here, the cognate codon-anticodon interaction triggers the hydrolysis of GTP and tRNA accommodation.(7) Peptide bond formation takes place before (8) eEF2-catalyzed translocation by one codon triplet shifting the A-site tRNA to the P-site, the P-site tRNA to the E-site while the E-site tRNA dissociates allowing the cycle to start over again.(9) Once a stop codon is encountered, a ternary complex of eERF1, eERF3 and GTP is recruited to the A-site. Accommodation of the tRNA-mimetic eERF1 recruits ABCE1, which splits the ribosomal subunits in an ATP-dependent manner leading to the release of the polypeptide and ribosome recycling enabling another round of translation initiation. Figure prepared in Biorender.

1.2 Non-canonical translation events are abundant in RNA viruses

Generally, the canonical eukaryotic translation process described above, relies on the recognition of the 5' cap, which permits the synthesis of only a single protein from a given mRNA [7, 24]. RNA viruses are required to express many proteins, both structural and enzymatic, at times from a single transcriptional unit (e.g. in HIV-1) for their replication and thus need to bypass this limitation. Additionally, many RNA viruses lack the machinery to add the specialized 'labels' unique to eukaryotic mRNAs, namely the 5' m^7G cap and poly (A) tail, which have been described above. Thus, viruses have alternative mechanisms for ribosome recruitment and the formation of a closed loop during the translation initiation process. Furthermore, with the largest viral genome size being only around 30 kb (e.g. some members of the *Coronaviridae* family), there is a distinctive selective pressure on viruses to expand their coding capacity. The compact genome size forces the virus to employ different translational strategies for accessing multiple open reading frames (ORFs) within the same mRNA. Due to all these reasons, non-canonical translation mechanisms, including internal ribosome entry, leaky scanning, ribosome shunting, reinitiation, stop codon readthrough, translational bypassing, and programmed ribosomal frameshifting (**Figure 2**), are abundant amongst RNA viruses. This translational reprogramming is often regulated by viral RNA structural elements and host or viral proteins and occurs at every stage of eukaryotic translation [25-29].

1.2.1 Non-canonical initiation

Internal ribosome entry sites (IRESs) are complex structural motifs usually found in the 5' untranslated region (UTR) or the intergenic regions of polycistronic RNAs [30-32]. These highly organized structures recruit ribosomes to mRNAs independent of initiation signals such as a 5' cap. Following the ribosome recruitment, IRESs ensure remodeling of the small subunit to accommodate an internal region of the template into the RNA-binding channel, thus reducing the need for initiation factors. Having said that, depending on the structure

and minimum required initiation factors, IRESs are divided into four classes (1-4). Although they have varying uses, IRESs allow the virus to bypass cap-dependent translation, which could be suppressed due to viral activity or innate cellular immunity. IRES elements also give the viruses a unique opportunity to increase their coding capacity. This has been demonstrated in several members of *Dicistroviridae* (e.g. Cricket paralysis virus (CrPV), *Picornaviridae* (e.g. Theiler's murine encephalomyelitis virus (TMEV)) and *Flaviviridae* (e.g. Hepatitis C virus (HCV)) [33-36]. Here, a small number of ribosomes recruited by the IRES initiate in an alternative overlapping +1 open reading frame (ORF), producing the normal 0-frame polyprotein and a +1 product.

While IRESs allow cap-independent internal entry of ribosomes, ribosomal shunting allows ribosomes cap-dependent access to downstream ORFs. However, unlike canonical cap-dependent initiation, ribosomal shunting, does not depend on the 40S PIC scanning process and occurs when the ribosome completely bypasses parts of the 5' UTR to reach a start codon. [37] This mechanism begins with the ribosomes being loaded in a cap-dependent manner on the mRNA, however the scanning process is limited, and the ribosome reaches an mRNA sequence called as the 'donor site'. When the 40S subunit encounters such a site it jumps over parts of the 5' UTR sequence landing on an 'acceptor' start codon. This process may or may not lead to the formation of a short ORF. At times when the short ORF is translated, such as in the case of *Adenoviruse*s or cauliflower mosaic virus, the 40S subunit remains attached to the mRNA and reinitiates through interactions with the rRNA or the release factors [38, 39].

Deviant cap-dependent initiation can also occur through leaky scanning or non-AUG initiation, events. During translation initiation, the 40S PIC scans the mRNA linearly until it encounters the first AUG codon [40]. However, in certain cases, ribosomes fail to initiate at the first AUG codon and continue scanning until they reach an alternative AUG codon downstream, termed as leaky scanning. They may also initiate at near cognate non-AUG codons, termed as non-AUG initiation. Both processes allow production of multiple N-terminally different protein isoforms of a single protein or production of distinct proteins through a shift in the reading frame. Now the efficiency of translation initiation is largely defined by the sequence context of the potential initiation codon, primarily on the nucleotides at positions –3 and +4 [40, 41]. The optimal context is purines in both of these positions, often called as the Kozak sequence. Pyrimidines in these positions reduce the recognition efficiency, which can lead to leaky scanning. This mechanism is seen in

members of the *Orthomyxoviridae, Luteovirus* and *Tungrovirus*, and in some cases combined with non-AUG initiation can leads to production of upto three or four distinct proteins (e.g. *Caulimoviridae*) [42-44].

Non-canonical initiation can be coupled with alternative termination leading to a termination-reinitiation event that occurs when the stop codon of an upstream ORF is in close proximity to the start codon of the downstream ORF [45, 46]. Here, the ribosomes translate the upstream ORF up until termination where a fraction of 40S subunits remain attached to the mRNA and directly reinitiate at the start codon of the downstream ORF. In viruses such as the *Caliciviruses* and influenza B, this process often occurs with an interplay of 18S rRNA and release factors [47-50]. This allows the production of essential proteins from a downstream ORF (such as BM2 ORF in influenza encoding for proton channel), which otherwise is unable to recruit the 40S subunits directly [49, 51]. Thus, reinitiation is an elegant mechanism allowing coupled translation of adjacent ORFs to produce protein products in defined ratios.

1.2.2 Non-canonical elongation and termination

Besides alternative initiation mechanisms, non-canonical elongation or alteration in termination signal recognition, loosely termed as 'recoding' mechanisms are also common gene regulation processes used by RNA viruses. Translational recoding falls into three main categories: stop-codon redefinition or readthrough, bypassing and programmed ribosomal frameshifting [52-54].

Stop-codon readthrough (SCR) or programmed readthrough occurs when a near-cognate tRNA decodes a stop codon as a sense codon and translation continues until the next termination codon [55-57]. Usually, translation termination is highly efficient, however, the sequence context of the stop codon, local mRNA topology as well as *trans*-acting factors that interact with the region downstream of the stop codon are important for SCR. Programmed readthrough can lead to production of proteins with a C-terminal extension, which can have effects on its localization, function and stability and thereby influence several cellular processes. In *Luteoviruses*, SCR motif in the end of the coat protein gene generates a polypeptide at a defined frequency, which is essential for aphid transmission [58]. SCR may also redefine the meaning of the codons, in which the stop codons UAG or UGA code for selenocysteine and pyrrolysine giving rise to distinct proteins [59-62].

Translational bypassing occurs when the translating ribosome skips a portion of the mRNA and forms a polypeptide from a discontinuous frame [52, 63]. This process represents an exquisite coordination between the mRNA elements, the nascent peptide and the ribosome. Although bypassing is not known to be utilized by eukaryotic viruses, a well-documented example is *gene60* of bacteriophage T4 [64, 65]. Here, the ribosome translates the first 46 codons of *gene60* mRNA up to the glycine codon GGA followed by a stop codon UAG. Instead of termination at UAG, peptidyl-tRNA^{Gly} disengages from pairing with the GGA codon called as the 'take-off' codon, is retained within the ribosome as it bypasses 50nt in the mRNA and re-pairs at the GGA codon downstream in the mRNA, called as the 'landing site' codon. Bypassing in the gene60 mRNA is dependent on the identity of peptidyl tRNA, nascent peptide as well as RNA secondary structures at the decoding site [66, 67].

Ribosome skipping/'Stop-go'/'Stop-Carry On' is mediated by the amino acid motif D(V/I)ExNPGP, which, together with the nascent peptide in the ribosome exit tunnel, skips the formation of a peptide bond between glycine and the final proline [68]. Nonetheless, translation can continue with proline as the N-terminal amino acid of the downstream product. It was first discovered at the junction between the 2A and 2B proteins of the *Picornaviruses* [69]. Although multiple models have been suggested for this process, it has been proposed, the structure of nascent peptide, prevents tRNA from binding at the ribosomal A-site and instead allows eRF1 to bind which along with eRF3 releases the nascent peptide and an unconventional reinitiation event occurs at the proline of a downstream peptide. This leads to production of multiple protein products (2A and 2B) from a single ORF [70, 71].

Most of the aforementioned non-canonical events can increase the coding capacity of the viral genomes and allow access to alternative reading frames. Nonetheless, for this particular purpose, recoding through programmed ribosomal frameshifting (PRF) is the main route that is utilized by all domains of life [72-79]. During PRF a proportion of ribosomes are directed into a different reading frame by shifting in the + or – direction relative to the 5' UTR by one or up to six nucleotides. Influenza A utilizes a low level of +1PRF, where ribosomes shift one nucleotide downstream, to produce a protein called as PA-X, which is involved in host cell shut-off and suppressing immune response. Here, the frameshift site is UCC UUU CGU and the presence of the rare CGU codon is said to stimulate a +1 frameshift [80, 81]. The sparseness of cognate amino-acyl tRNAs, would lead to a potential ribosomal pause, providing the time window for frameshifting to take place. Besides this,

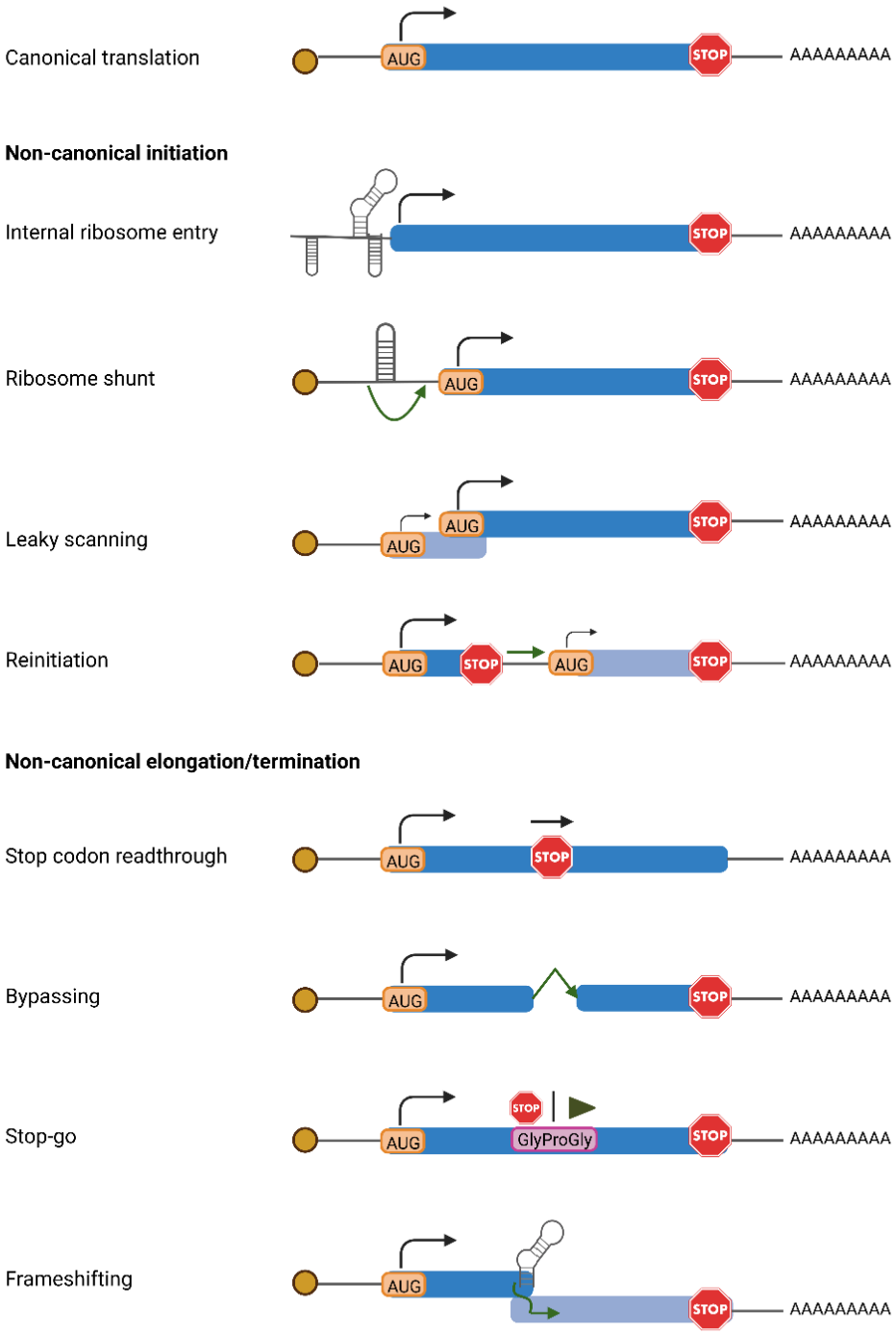


Figure 2: Examples of non-canonical translational mechanisms. Canonical eukaryotic mRNA translation is shown in the top panel. Black arrows indicate translation initiation (at the start of an ORF) or continuation of translation. Green arrows indicate the probable movement of ribosomes/ 40S subunit in a non-canonical manner. The ORFs are shown in different shades of blue, when two different polypeptides are synthesized. In the stop–carry on mechanism, both termination and initiation steps are non-canonical, as indicated by the red square and green circle. Figure made in Biorender. (Adapted from [25, 77]).

+1 frameshifting is also said to occur in members of plant-infecting *Closteroviridae* family, and appears to be utilized for the expression of the viral polymerase [82]. However, in viruses, +1PRF is poorly understood and several questions about its mechanisms remain unresolved. In comparison to +1PRF, –1PRF, where ribosomes shift one nucleotide upstream, is a widely used variety of PRF in RNA viruses and its mechanisms are well-studied.

1.3 –1PRF and its regulation

–1PRF was first reported in Rous sarcoma alpharetrovirus where it led to the production of the Gag-Pol polyprotein from overlapping *gag* and *pol* reading frames [83, 84]. Since then, related signals have been widely reported in several clinically important viruses, *e.g.* Human immunodeficiency viruses (HIV) - 1 and 2, Coronaviruses, *etc*. **–**1PRF is employed to control the stoichiometry of proteins or increase the coding capacity of genomes.

1.3.1 *Cis-acting elements of –1PRF*

The ability of the ribosome to undergo –1PRF depends on crucial *cis*-acting elements present in the mRNA. These include a heptanucleotide slippery sequence where the ribosomes undergo a translational shift into the **–**1 alternative reading frame and a downstream stimulatory RNA structure which slows down ribosome progression [84-88]. The two *cis*-acting elements are separated by a spacer region (about 5-9 nt) that ensures the correct positioning of the ribosome on the slippery site [89-91]. The slippery sequence is typically composed of X_XXY_YYZ, where X are any identical nucleotides, Y is A or U and Z can be A, C or U. On the other hand, the RNA structural element is frameshift-site specific, e.g. single stem-loop (in case of *1a/1b* site of astroviruses [92]) or complex H-type pseudoknot (in coronaviral mRNAs) (**Figure 3**) [85]. Furthermore, additional RNA elements are also said to play a role in frameshift regulation. For SARS-CoV-2 and SARS-CoV, a conserved upstream hairpin structure termed as ‘attenuator’ is said to decrease frameshifting efficiency (**Figure 3**) [93, 94].

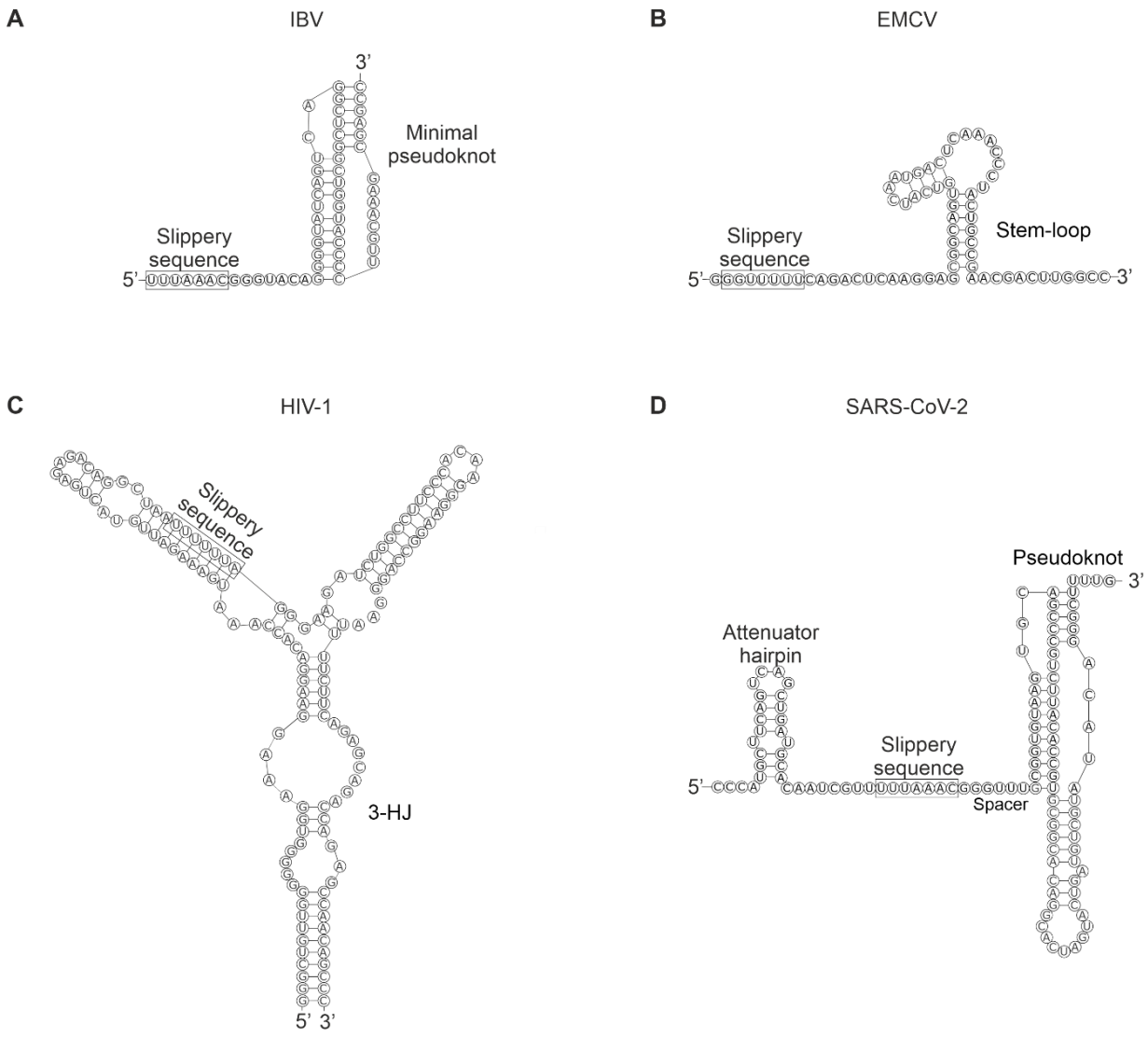


Figure 3: Diversity of the *cis*-acting elements of –1PRF signals. (A) Minimal pseudoknot – IBV mRNA [95] (B) Stem-loop – EMCV [96] (C) Extended 3-helix junction (3HJ) – HIV-1 [97] (D) Complex pseudoknot and attenuator – SARS-CoV-2 [93]

1.3.2 *Trans-acting elements of –1PRF*

Besides, *cis*-acting elements present in the mRNA, *trans*-acting factors of –1PRF can be pathogen- or host- encoded proteins, or molecules such as non-coding small RNAs and natural or synthetic compounds. The *trans*-factors either directly interact with the RNA structural element or ribosomes, or indirectly affect frameshifting by binding to other proteins influencing overall translation dynamics [75]. An elegant example of pathogen-encoded protein-mediated frameshift regulation is the *trans*-activation of −1PRF of Encephalomyocarditis virus (EMCV) and Theiler's Murine Encephalomyelitis virus (TMEV) by their respective 2A proteins, which are discussed in detail in **Chapter 3** (for EMCV) [96, 98-101]. Briefly, −1PRF in these cardioviruses occurs at a conserved GGUUUUU motif followed by slightly differing 3' stem-loop structures at the 2A-2B ORF junction producing a *trans*-frame protein 2B* with unknown functions. Here, binding of the 2A protein stabilizes the frameshift stimulatory structure and activates EMCV −1PRF (**Figure 4A**). In arteriviruses such as the Porcine reproductive and respiratory syndrome virus (PRRSV), the viral protein nsp1β forms a complex with cellular poly(C)-binding proteins and interacts with the conserved C-rich region downstream of the slippery sequence, stimulating −1 as well as −2 frameshifting [102, 103]. Several interferon induced host proteins such as the zinc-finger antiviral protein (ZAP-S) (discussed in detail in **Chapter 3**) as well as Shiftless (SHFL) were shown to inhibit frameshifting in SARS-CoV-2 and HIV-1 [104, 105]. While ZAP-S interacts with the frameshift site and destabilizes the RNA stimulatory structure, SHFL recruits eRF1-eRF3 to the stalled complexes resulting in premature translation termination (**Figure 4B, C**). Similarly, the host proteins Annexin A2 (ANXA2) and eRF1 were shown to decrease frameshift efficiencies of IBV and HIV-1 mRNAs, respectively, although the mechanisms of these proteins remain to be studied [106, 107]. Besides proteins, co-translational folding of the nascent polypeptide chain is shown to stimulate frameshifting in alphaviruses [108, 109]. Similar regulation was observed in case of SARS-CoV-2, where the C-terminus nascent peptide co-translationally folds and interacts with the ribosome exit tunnel, enhancing frameshifting [110].

−1PRF is regarded as an attractive antiviral target due to its conserved nature and importance during viral replication. Both antisense oligonucleotides and synthetic compounds were used for targeting −1PRF structural regions and some have been shown to affect frameshifting efficiencies and may decrease viral replication in several viruses including SARS-CoV-2, HIV-1, and Mouse mammary tumor virus (MMTV) [90, 111-113]. [114-

[118]. However, the application of frameshift-targeting small molecules was largely used for the recent SARS-CoV-2 virus where large-scale screening of FDA-approved drugs and commercial small-molecule libraries was implemented in hope of new antivirals. [119-121]. Recently, the first screen of natural compounds targeting the frameshift site was conducted leading to the discovery of two molecules that impede SARS-CoV-2 frameshifting and thereby affect viral replication [122] (discussed in detail in **Chapter 4**).

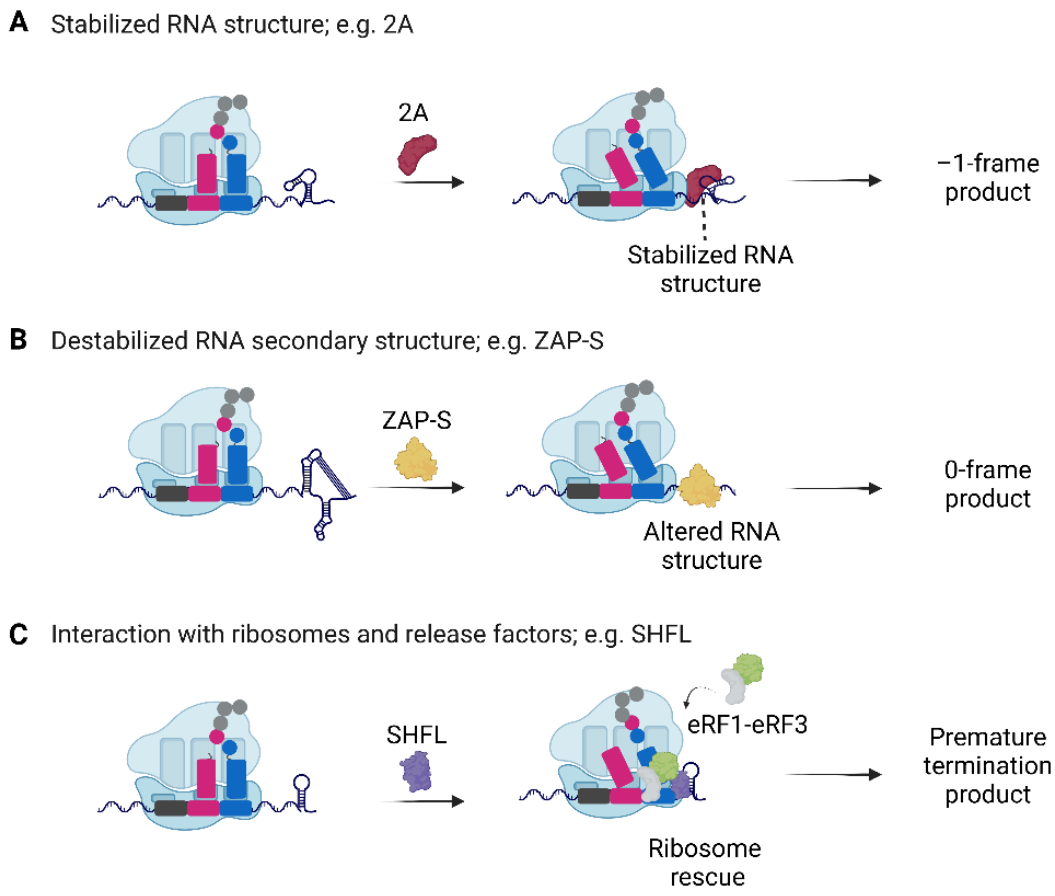


Figure 4: Differential mechanisms of protein-mediated −1PRF. Three examples, namely (A) the cardioviral 2A protein, (B) the zinc-finger antiviral protein ZAP-S and (C) the SHFL protein are shown as representatives of how host- and pathogen-encoded proteins can alter the secondary structure of a frameshift mRNA or recruit release factors resulting in enhanced or decreased frameshifting. 2A and ZAP-S specifically bind to the frameshift RNA and stabilize or destabilize the structure, respectively. SHFL is not known to interact with frameshift RNAs but rather recruits release factors to form a premature termination codon. Figure prepared in Biorender.

1.3.3 Parameters of –1PRF

Cis- and if present, *trans*- acting factors act synergistically to regulate –1PRF [72]. Canonically, frameshifting is stimulated by *cis*-acting elements embedded in the mRNA. *In vitro*, codon-resolved analysis of –1PRF site of IBV frameshift motif, revealed that frameshifting occurs in the late-stage of translocation when the two tRNAs both bound on the slippery codons, move from A- and P-sites to P- and E-sites, respectively (**Figure 5B**) [87]. Several molecular determinants including the thermodynamic stability, mechanistic features, conformational heterogeneity and structural plasticity of the RNA stimulatory structure are said to influence –1PRF. These structural elements can mediate slowing down or stalling of ribosomes at the slippery codons or can sterically hinder A-site tRNA binding [87, 88, 97, 123-131]. However, the propensity of frameshifting on a given mRNA as well as the efficiency is largely determined by the thermodynamics of base-paring in the heptanucleotide slippery sequence, *i.e.* the free-energy difference between the 0- and alternative frame [132]. This occurs in the case of longer stalling times, provided by a stable secondary structure. However, during shorter stall times, the pause becomes too short for the ribosome to overcome the energy barrier and the frameshifting would decrease due to increased translocation [132]. Shorter pauses can occur in case of possible *trans*-acting factors, such as ZAP-S, that alter the frameshift stimulatory structure. In these cases, mechanistic and kinetic contributions of *cis*- and *trans*- acting elements would also affect the frameshift efficiencies.

Non-canonically, –1PRF can also occur via the hungry codon pathway/ aa-tRNA depletion-stimulated frameshifting (ADF) pathway, which does not require a secondary structure and the elongation pause is provided by either a rare codon or limited aa-tRNA (**Figure 5C**) [133-138]. The slow decoding of the A-site causes the slippage of the P-site tRNA into the –1-frame changing the identity of the A-site codon. In some mRNAs, availability of aa-tRNAs can also influence the choice of frameshifting pathways, seen in the case of HIV-1 mRNAs. In HIV-1 mRNA, the frameshifting efficiency is modulated by the amount of Leu-tRNALeu that reads UAA codon. This particular isoacceptor is limited in T-lymphocyte cells, which are the usual HIV-1 targets. When UAA decoding occurs normally, the canonical –1PRF pathway is translocation-dependent. However, when this decoding is delayed, frameshifting is mediated by the ADF pathway [134].

Thus, in the ongoing evolutionary arms race between the virus and the host, it is clear that RNA viruses use a fascinating plethora of non-canonical translation events to evade host restriction mechanisms imposed by cellular translational machinery, host immune response or availability of aa-tRNAs. Due to the variety of unusual translational phenomena, annotating translated sequences is particularly challenging. It is crucial to identify the complete repertoire of viral-translated proteins for understanding viral replication, defining the epitopes of immune surveillance as well as identifying potential novel antiviral targets. A well-established technique for understanding the complex viral as well as cellular translational landscape during infection is ribosome profiling.

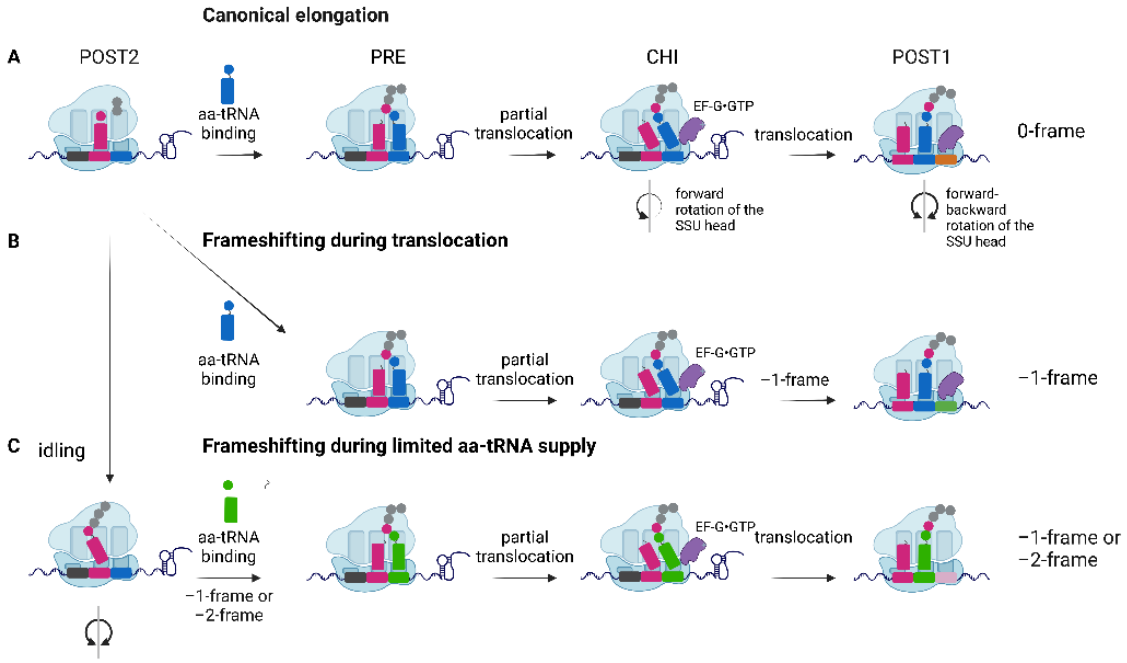


Figure 5: Schematics of –1PRF mechanisms. (A) During canonical elongation, movement of the ribosome along the mRNA and the peptide bond formation is facilitated by specific forward and backward movements of the small subunit (SSU) head and the hydrolysis of GTP by the elongation factor EF-G (eEF-2 in eukaryotes). POST: post-translocation state, PRE: pre-translocation state, CHI: chimeric state. (B) During canonical –1PRF, ribosomal stalling is caused by a secondary structure of the mRNA leading to frameshifting during translocation by tandem slippage of two 0-frame t-RNAs. (C) During aa-tRNA depletion-stimulated frameshifting (ADF) pathway, limitation of the A-site codon leads to idling, which can be overcome by shifting into the –1- or –2-frame. Figure made in Biorender. Adapted from [72] .

1.4 Ribosome profiling is a powerful tool to decipher viral complexity

Ribosome profiling is a genome-scale method that maps the precise position of translating ribosomes by deep sequencing of the mRNA footprints that are occupied by the ribosomes and are thus physically shielded from nuclease digestion (**Figure 6A**) [139-142]. It essentially provides a snapshot of the translatome at a single nucleotide sub-codon resolution [139, 143]. With this high resolution information, precise boundaries of translated regions and the specific translational reading frame can be identified. Furthermore, it would allow us to pinpoint regions whether translation of overlapping reading frames occurs due to presence in different mRNA isoforms or due to non-canonical translation events such as alternative initiation or programmed ribosomal frameshifting.

Translational start sites can be mapped directly by performing ribosome profiling under special conditions, such as antibiotic pretreatment of cells with harringtonine or lactidomycin, which preferentially captures initiating ribosome (**Figure 6C**) [140, 144-148]. So far, ribosome profiling has been applied to a number of RNA viruses including SARS-CoV-2, TMEV, IBV, PRRSV, murine leukemia virus and Hepatitis B virus (HBV) [98, 103, 149-154]. Indeed, high resolution analysis of murine coronavirus and Epstein Barr virus revealed the presence of upstream ORFs (uORFs) in the 5' UTR, which could influence translational regulation or cellular immune response [155, 156]. Additionally, intricacies of non-canonical elongation such as PRF can be resolved by careful analysis of the frameshift sites in the viral genome. A notable example is the PRRSV genome where the density of ribosome footprints mapped to ORF1A is higher than the footprint density at ORF1B. As a result of –1PRF, only a small proportion of the ribosomes translating ORF1A shift reading frame and proceed to ORF1B. Thus, it is possible to estimate frameshifting efficiency by calculating the ratio of footprint densities between the two open reading frames [103]. One of the important features involved in the –1PRF mechanism is the occurrence of the ribosomal pause which allows the ribosome enough time to sample alternative reading frames. Indeed, this ribosomal pause was shown to occur *in vivo* in several viruses, e.g. SARS-CoV-2 and EMCV, through ribosome profiling [96, 157]. In fact, ribosome stacking or collisions were observed at the frameshift site and stop-go motif of TMEV using a variation of this technique, called as disome profiling, which captures ribosome protected fragments of collided ribosomes (**Figure 6D**) [98]. Furthermore, specialized protocols which involved tagging host ribosomal proteins with different compositions and conducting ribosome profiling of these ribosomes has led to the discovery that such ribosomal heterogeneity can

lead to preferential translation of a subset of mRNAs. For example, interaction of RPL10A/uL1with the IRES is important for the translation of IRES containing mRNAs (Chen & Dickman, 2017). Although this study was not conducted in virus-infected cells, it suggests that perhaps viral translation can occur through heterogenous ribosomes, pointing to the sheer importance of translational control during infection.

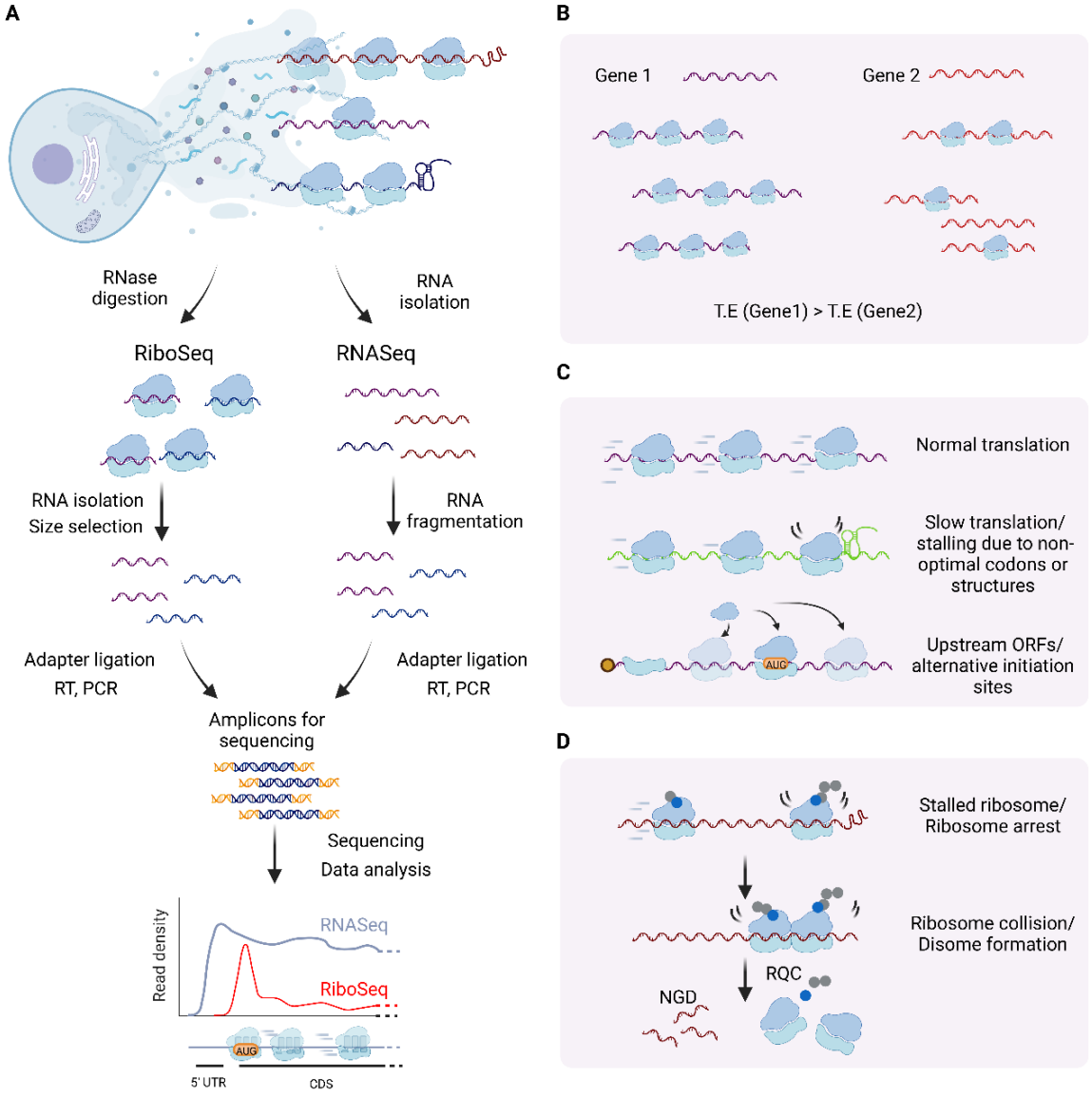


Figure 6: The ribosome profiling technique and its applications. (A) Schematics of the conventional ribosome profiling workflow. Ribosome profiling technique can be used to study (B) differential translational efficiency of genes, (C) speed of elongation or stalling and non-canonical translation events and (D) ribosome arrest or disome formation.

1.5 Translational control is at the forefront in infection

During the virus-host battle for cellular translational machinery, one of the important host defense response is to induce translational shut-off and activate rapid synthesis of antiviral proteins to limit collateral damage [158]. This leads to the following fundamental questions, what impact do various viral infections have on the range of host mRNAs being translated, and how or to what extent does each virus have the ability to translate its own mRNAs more effectively at the expense of those of the host? This important aspect of understanding viral infection can be revealed through ribosome profiling by measuring the global translational changes that occur in cellular and viral genes during infection. Control of protein production reflects both the regulation of transcript abundance and the efficiency with which these mRNAs are translated into proteins. Historically, the analysis of global gene expression focused on RNA sequencing, micro-arrays and mass spectroscopy techniques which identified the transcriptomic and proteomic changes giving an indirect evidence of genome-wide translational control. Conducting parallel RNA sequencing and ribosome-profiling measurements allows direct assessment of differential translation of any cellular and viral genes during infection, providing opportunities to answer fundamental questions relating to translational control during viral infection (**Figure 6B**). Recent ribosome profiling study in SARS-CoV-2 revealed that viral infection increases translational efficiency of mRNAs associated with unfolded protein response (UPR) [159]. This can be a defense strategy of the host, as massive production of viral proteins stresses the endoplasmic reticulum and the cell compensates by reducing ER protein levels. Interestingly, it was shown that coronaviruses take advantage of this response to enhance their viral replication, pointing to co-evolution of the host and virus [159].

Host immune response also leads to the upregulation of interferon-induced antiviral proteins which can have a drastic effect on all aspects of viral translation. This can be clearly observed in case of ZAP-S and SHFL which inhibit viral programmed ribosomal frameshifting [104, 105]. Modulation of −1PRF efficiencies – both an increase and decrease – can impede viral replication. However, viruses can, to some extent, tolerate an increase in efficiency [160]. It is, thus, interesting to note that cellular proteins (ZAP-S, SHFL) tend to decrease frameshift efficiencies while viral proteins (at least in case of 2A) induce −1PRF, warranting extensive characterization of the mechanisms with which they target viral translational regulation.

Although non-canonical translation events are abundant in viruses, current studies have revealed that several cellular mRNAs also rely on non-canonical initiation events to inhibit viral replication. Upstream ORFs play an important role in dynamic translation control triggered by viral infection and other stresses [161-165]. A powerful antiviral defense is the activation of protein kinase R (PKR), which phosphorylates eIF2α thereby inhibiting translation initiation and bulk protein synthesis [166-168]. Despite the general translation inhibition, certain mRNAs involved in stress-response showed persistent translation. These mRNAs were seen to possess at least one efficiently translated uORF that translationally represses the main ORF under normal conditions [169]. Examples of such uORF-mediated regulation include mRNAs of transcription factors such as ATF4 involved in global stress response pathways and proapoptotic factors such as DDIT3. Perhaps, translation of these genes provides cellular resistance to infection or ultimately apoptosis. Additionally, several recent studies corroborate the importance of not just uORFs but also small ORFs as translational regulators during cellular stress conditions [165, 170]. Thus, stress-induced alternative mechanisms such as reinitiation, non-AUG initiation, leaky scanning add to the translational and proteomic diversity of not just viral but also cellular genome. Do eukaryotic genomes also regulate their stress response through other non-canonical translation mechanisms? Indeed, bioinformatic analyses suggest that approximately 10% of cellular mRNAs contain putative frameshift motifs and in fact, –1PRF events were confirmed experimentally on mammalian genes *PEG-10*, *ma3* and *ma5* in humans [115, 171-174]. Thus, it is likely that eukaryotes also contain potential frameshift motifs and like alternative initiation, these may contribute towards antiviral response. Overall it is clear that viral infection can lead to substantial translational reprogramming in cells, possibly unique to each virus, and that translational control is critical for modulation of gene expression in both host and virus during infection.

2. Scope of the thesis

Several studies have shown that RNA viruses employ non-canonical mechanisms, such as –1PRF, to meet their specific translational needs and viral protein synthesis often occurs at the expense of cellular proteins during infection. Furthermore, translational control in host cells is at the forefront of antiviral response, with cells activating their stress response genes and employing alternative mechanisms to ensure the translation of cellular mRNAs. Several proteins- both host and viral- are involved in this post-transcriptional gene regulation. The aim of this thesis was to understand both viral and host translational regulation during infection. This is essential for the design of new therapeutics to fight viral infection or alter the cellular innate immune response. Some of the questions we aim to answer are;

1) How does the translational landscape of both host and pathogen dynamically change upon viral infection? (**Chapter 1** – using HIV-1 as a model system)

2) What are the mechanisms with which host or viral proteins regulate viral translation, more specifically the non-canonical –1PRF? (**Chapter 3 and supplement**– using SARS-CoV-2 and EMCV as model systems).

3) How can we target conserved viral –1PRF elements to impede viral replication in the host cell? (**Chapter 4** – using SARS-CoV-2 as a model system)

Chapter 2

The translational landscape of HIV-1 infected cells

Anuja Kibe[1], Stefan Buck [1], Redmond P. Smyth [1,3], Florian Erhard [2], Neva Caliskan [1,3*]

[1]Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Zentrum für Infektionsforschung (Helmholtz Centre for Infection Research), Josef-Schneider-Strasse 2, 97080, Würzburg, Germany

[2]Institute for Virology and Immunobiology, Julius-Maximilians-Universität, 97080, Würzburg, Würzburg, Germany

[3]Medical Faculty, Julius-Maximilians University Würzburg, 97074, Würzburg, Germany

[*]Corresponding author: Neva Caliskan neva.caliskan@helmholtz-hiri.de

Manuscript in preparation

Abstract

HIV-1 is a complex retrovirus which produces multiple proteins through a combination of alternative splicing, translational and post-translational mechanisms, which are tightly regulated throughout infection. Here, we have conducted in-depth ribosome profiling to gain insights into the role of translational regulation of host and viral transcripts during HIV-1 infection of human T-cells. Our results indicate that initial host response is translationally regulated, which is followed by alterations in the host transcriptome at later stages of infection. Analysis of ribosome profiling reads from host and viral transcripts reveals extensive translation of short open reading frames (sORFs), which are temporally expressed throughout the viral replication cycle. Additionally, careful analysis of disome profiling reads revealed extensive ribosome stacking upstream of the HIV-1 frameshift site. Together, these data reveal the key role of post-transcriptional regulation in HIV-1 replication and highlight unique mechanisms that can be targeted for anti-viral therapies.

Introduction

The type 1 human immunodeficiency virus (HIV-1) is a member of the Lentivirus genus from the Retroviridae family and the etiologic agent of the acquired immunodeficiency syndrome. During infection, the positive-sense single-stranded viral RNA is reverse transcribed into a double-stranded DNA molecule that integrates into our genome, forming a provirus. The host RNA polymerase II (Pol II) transcribes the provirus DNA into a capped and polyadenylated RNA that may undergo complex alternative splicing [175-177]. This extensive splicing process leads to three classes of HIV mRNAs: the unspliced RNA genome, which encodes Gag/Gag-Pol; partially spliced transcripts encoding Vif, Vpr, and Env/Vpu; and completely spliced mRNAs encoding Tat, Rev and Nef [178, 179]. Alternative RNA splicing in HIV-1 is temporally regulated, with completely spliced mRNA species being exported to the cytoplasm early in the infection to form regulatory proteins, while the incompletely spliced and unspliced RNAs are exported at later infection stages [178-180].

Similar to the alternative splicing process, the translation of HIV-1 transcripts is also tightly regulated. HIV-1 protein synthesis depends entirely on host translation machinery and viral proteins are mainly synthesized by a cap-dependent mechanism similar to canonical host translation. Additionally, multiple mechanisms of non-canonical initiation such as leaky scanning, ribosome shunting, and internal ribosome entry-sites (IRESs) are used for the production and regulation of different viral proteins [181-183]. IRESs have been described in

the 5' UTR of HIV-1 RNAs as well as the coding region of Gag for unspliced RNAs [182, 184-192]. The activity of the 5' UTR IRES was proposed to be stimulated during conditions of cellular stress and is said to be up-regulated as the HIV-1 replication cycle proceeds [193, 194]. HIV-1 protease leads to the proteolysis of eIF4GI, eIF4G2 and PABP which leads to a profound inhibition of cap-dependent translation and 43S scanning [184, 195-198]. Furthermore, HIV downregulates genes involved in ribosome biogenesis such as rRNA transcription, pre-rRNA processing and ribosome maturation in primary CD4+ T cells [199]. Despite this global translation decrease, it was recently shown that viral proliferation is ensured by hypermethylation of the m7G cap, which allows HIV-1 RNAs to enter separate specialized protein translation pathways [200]. Besides initiation, the translation of full-length HIV-1 RNA is also regulated at the elongation step. The Gag-Pol polyprotein which contains viral enzymes including protease, reverse transcriptase and integrase is translated through programmed −1 ribosomal frameshifting (−PRF), which occurs with an efficiency of 5-10% [201]. HIV-1 frameshifting requires two key elements- a highly conserved UUUUUUA slippery sequence and a downstream stable RNA secondary structure, predicted to be a stem-loop, a conserved three-helix junction (3HJ) or a pseudoknot [201-206]. This frameshift event maintains the correct ratio of the structural Gag protein and the fusion Gag-Pol polyprotein which is critical for viral replication.

Although a large number of proteome and transcriptome analyses have been conducted in HIV-1 infected cells, little is known regarding the translational landscape of host and viral proteins [207-212]. Additionally, most available studies of HIV-1 translation have been conducted mainly using *in vitro* translation extracts or reporter constructs. Although, these have been crucial in revealing the myriad of mechanisms controlling HIV-1 expression, a global view of translational control during a productive replication cycle is still lacking. Here, we have conducted in-depth ribosome profiling (RiboSeq) and RNA-sequencing (RNASeq) from cytoplasmic extracts of HIV-1 infected T-cells to gain insights into the role of translational regulation of host and viral transcripts. Our results indicate a modest impact on cellular translation rates and reveal that initial response to infection is at a translational level followed by transcriptomic changes in later infection stages. We discovered novel short ORFs (sORFs) in the human and HIV genome that were temporally expressed during the course of infection. Additionally, extensive non-AUG translation initiation was observed in the 5' untranslated regions of HIV-1, pointing to the presence of upstream ORFs (uORFs) that are translated during infection. Careful analysis of RiboSeq

data revealed a major stalling event upstream of the HIV-1 frameshift site. We confirmed ribosome stacking at this position through disome profiling. Stalling at this position may alter the kinetics of translation near the frameshift site, thereby modulating HIV-1 frameshifting. Thus, our work adds more layers to post-transcriptional regulation in HIV-1 replication and highlights unique mechanisms that can be targeted for anti-viral therapies.

Results

Changes in transcript abundance and translation of host and viral mRNAs

To study the global transcriptional and translational regulation during HIV-1 infection, SupT1 cells were infected with HIV-1-iGFP (NL4-3 strain), such that more than 85% cells express GFP 24 hours post infections (hpi), as assessed by flow cytometry analysis (**Supp. Fig. 1A**). Cytoplasmic lysates from mock and HIV-1 infected cells were collected at 8, 16 and 24 hpi to monitor changes via transcriptome sequencing (RNASeq) and ribosome profiling (RiboSeq) (**Fig. 1A**). Quality control analysis was done as described previously which confirmed the overall good quality of the datasets [156]. First, the length distribution of mapped RiboSeq reads were within the expected range of RPFs peaking at 30nt (**Supp. Fig 1B**) [139, 213]. Next, PRICE analysis was used for determination of frame distribution. Majority of CDS-mapping RPFs are in the correct frame *i.e.* 0-frame which is translated in cellular mRNAs, indicating appropriate nuclease digestion (**Supp. Fig 1C**). Lastly, the majority of our reads start within coding transcripts at the canonical start codon reflective of the length of the coding sequence (CDS) and a small percentage of the RPFs map to the 5' UTR (**Supp. Fig 1D**). Only a small percentage (~1%) of the cellular mRNA pool was derived from HIV-1 mRNA at 8 hpi, which increased to 6% at 24 hpi. However, ribosome associated viral RNA was only detected from 16 hpi (~0.5% of total ribosome-bound reads), and by 24 hpi, ~2.5% of the ribosome-protected fragments contained HIV-1 sequences (**Fig. 1B)**.

Comparison of differentially expressed host genes in the mock and infected timepoints indicates that at 8hpi there were no genes significantly downregulated in RNASeq and a small subset of genes significantly upregulated as compared to the uninfected mock. On the other hand, ~200 genes were differentially regulated in the RiboSeq dataset (**Supp. Fig. 2A**). At 16hpi, while the number of translationally regulated genes remained constant, we found ~600 genes with changes in their transcript abundance (**Supp. Fig. 2B**). At 24hpi, overall changes in the transcriptome and translatome correlate well (**Supp. Fig. 2C**). In all

time points, however, there was only a modest change in the translational efficiencies (TE) of cellular genes, with a minor subset of 80 genes showing differential TE in each time point.

In order to obtain a global assessment of cellular and viral translation we directly compared RNASeq and RiboSeq reads. Based on the significance of the fold changes, the differentially expressed genes were classified in four groups: (1) genes classified as *exclusive* are only translationally regulated and the total mRNA amount stays the same; (2) *forwarded* genes are those where there is change in the mRNA levels, while the translation efficiency stays constant; (3) genes named as *intensified* change both mRNA levels and ribosome per mRNA ratio in the same direction; and (4) *buffered* genes are those where the mRNA- and ribosome protected fragment fold changes are changing at opposite rates leaving the overall gene expression at similar levels (**Fig. 1C,** gene lists provided in **Supp. Table 1**). Accordingly, at 8hpi we saw ~800 genes classified as *exclusive*, indicating that the change in RiboSeq counts of these genes as seen in **Supp. Fig. 2A** was solely due to their translational efficiency. At later timepoints (16 and 24 hpi) this pattern changed showing more *forwarded* classified genes. Overall, these results indicate that perhaps the initial stress response to infection is translationally regulated, and the longer the infection persists, the cellular changes are driven at the level of cytoplasmic transcript abundance, which may result from changes in transcription, nuclear export, and transcript stability or mRNA degradation.

Our gene ontology analysis on RiboSeq data indicates that the upregulated genes at 8hpi, when most translational regulation occurs, are those involved in cell cycle and DNA replication, whereas downregulated genes are involved in regulation of apoptosis and electron transport chain. This could suggest that the host cells try to undergo apoptosis, which is actively inhibited by the virus. At later stages of infection (16-24 hpi), we saw an upregulation of genes involved in immune processes and cholesterol metabolism. The modification of cholesterol pathway agrees with the studies suggesting HIV-protein Nef reduces the efficiency of reverse cholesterol transport causing accumulation of cholesterol in macrophages [214, 215]. Interestingly, there is a translational downregulation in protein transport and all ribosome related processes at 24hpi in the RiboSeq data. Gene ontology analysis on the RNASeq data shows an upregulation of cellular differentiation, cholesterol metabolism, cell migration and communication, as well as immune response genes, and a downregulation in translational processes, tRNA transport and RNA processing. The

downregulation in both transcriptional and translational efficiency of the ribosome-related genes, including those involved in the ribosome assembly and rRNA processing, suggests a modest global decrease in translation levels caused due to HIV-1 infection.

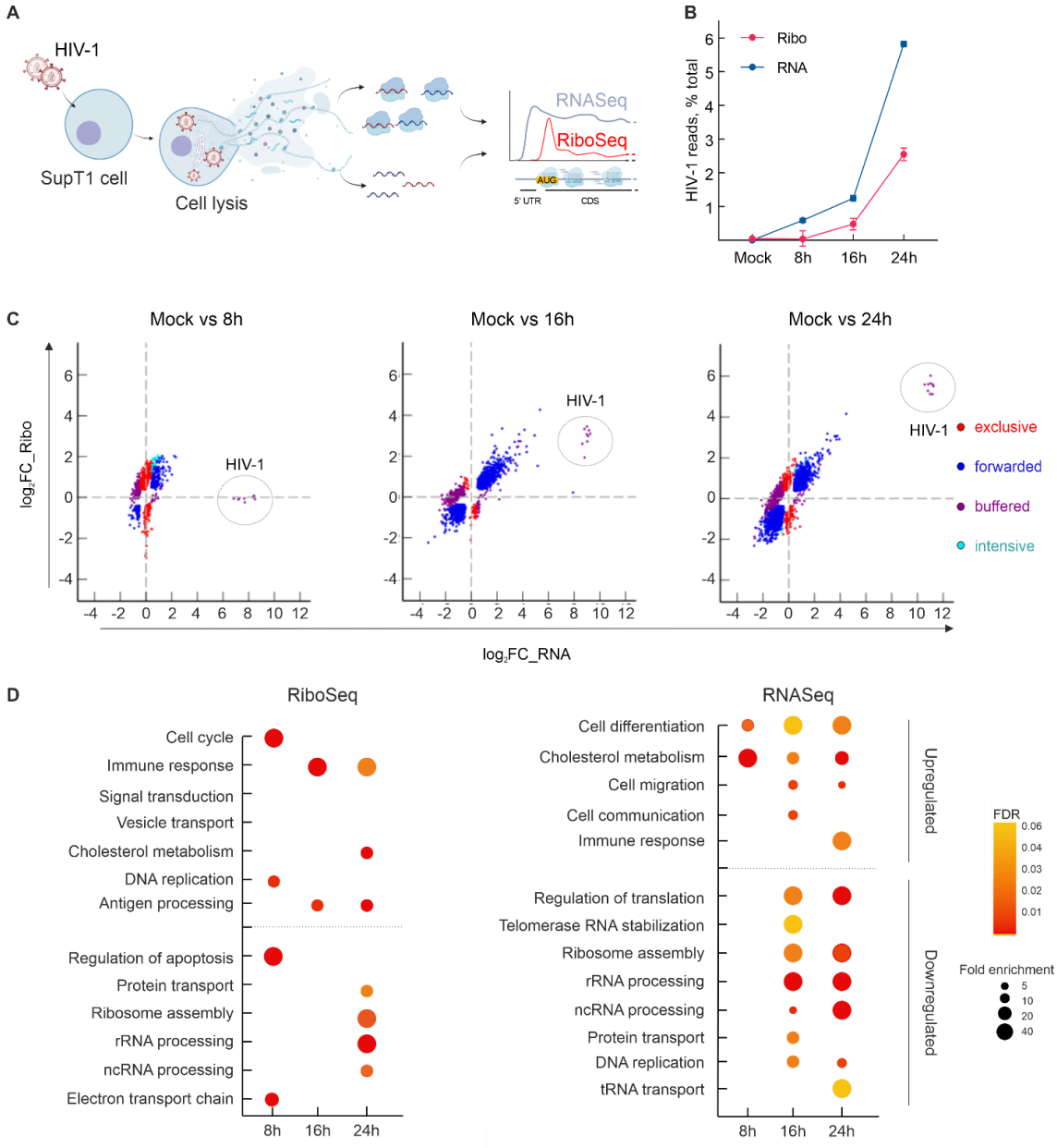


Figure 1: Global transcriptional and translational changes in HIV-1 infected cells. A) Schematic representation of the procedure to monitor transcript abundance and translation in HIV-1 infected cells. Briefly, SupT1 cells were infected or not (Mock) with HIV-1-iGFP (NL4.3 strain). At 0, 8, 16 and 24 hours post infection (hpi), cells were lysed to recover the cytoplasmic fraction and prepare ribosome profiling and RNA-seq libraries subjected to high-throughput sequencing. B) Percentage of total RNAseq and RiboSeq reads mapping to the HIV-1 genome at each time point. C) Scatter-plot of the fold-change (log_2) in cytoplasmic RNASeq and RiboSeq of the Mock-infected and HIV-1 infected cells compared to each time

point of infection. Red dots are exclusive genes - only translationally regulated and the total mRNA amount stays the same; blue dots are forwarded genes - change in the mRNA levels and the translation efficiency stays constant; light blue dots are intensified genes - change both mRNA levels and ribosome per mRNA ratio in the same direction; and purple dots are buffered genes - mRNA- and ribosome protected fragment fold changes are changing at opposite rates. D) Gene ontology analysis of differentially expressed genes in RiboSeq and RNASeq at each time point.

HIV-1 infection causes a global decrease in host translation

To observe the effect of HIV-1 on global translation levels, we conducted polysome profile analysis on cellular lysates of 8 and 24 hours infected time-points with their respective mock-infected cells. At 8 hpi, there was no significant change between the profiles of mock and infected cells (**Fig. 2A**). However, at 24 hpi, there was a reduction in the amplitude of the polyribosome peaks in the infected cells compared to the mock, with a subtle increase in the monosome (80S) peak (**Fig. 2B**). The results suggest that the translation initiation process is targeted by HIV-1 infection. Previous studies with ribosomal RNA profiles and sucrose gradient analysis, have confirmed this change in polysome levels upon HIV-1 infection [216]. We then compared the differential translational efficiency of host house-keeping genes to the HIV-1 genes at later time-points of infection (**Fig. 2C**). We observed a substantial increase in TEs of all viral genes at 16 hpi, which remained constant for *tat*, *rev* and *vpr* at 24 hpi and showed minor increase (fold change of 1.5 atleast) for the rest of the viral genes at this timepoint. This indicates that ribosome occupancy on viral RNAs is temporally regulated during the course of infection. Compared to viral RNAs, there was a subtle decrease in the translational efficiencies of host housekeeping genes such as *GAPDH, ACTB* and *RPL13A* at later stages of infection. Overall, HIV-1 infection does not seem to completely shut down host translation, but rather decreases it at the timepoints we monitored. However, despite this global translational suppression, the HIV-1 mRNAs continue to be preferentially expressed, possibly through specialized initiation mechanisms.

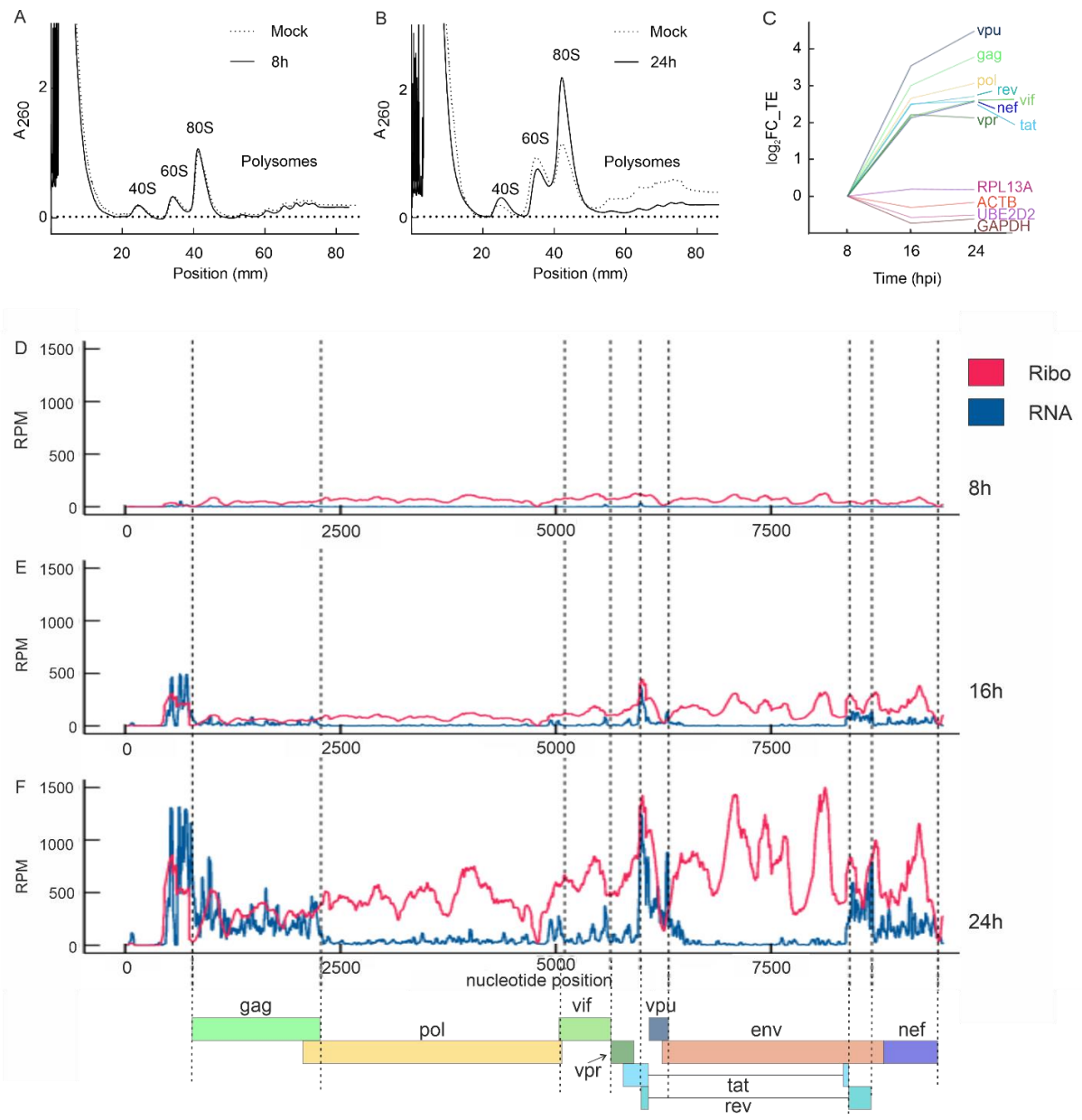


Figure 2: Translational efficiencies and landscape of host and viral transcripts. Polysome profile analysis of HIV-1 infected cells at A) 8 hpi and B) 24 hpi. Briefly, mock and HIV-1 infected cells were lysed and lysates were subjected to 5–45% sucrose gradient ultracentrifugation, subsequently fractionated and absorbance of fractions was checked at 260 nm. C) Fold change (log_2) of translational efficiencies of HIV-1 and host housekeeping gene transcripts at 16 and 24 hpi normalized to 8 hpi. Distribution of RNA-seq (pink) and ribosome profiling (blue) reads across the HIV-1 genome in SupT1 cells at D) 8 hpi, E) 16 hpi and F) 24 hpi. Positions of viral genes are depicted.

Translational landscape of host and HIV-1 transcripts

Having characterized the global effect of HIV-1 on host translation, we proceeded to investigate in detail the translational landscape on HIV-1 transcripts. As described earlier, we could detect RNASeq reads of viral origin at all time-points of infection (**Fig. 1B, 2D-F**). We could also detect some ribosome footprints at 8 hpi, suggesting that some of the incoming viral RNAs are translated early on in the infection cycle (**Fig. 2D**). At later time points, translation was detected on all canonical viral coding sequences (**Fig. 2E, F**). In the HIV 5' UTR, density of RPFs was more than that of the known coding sequences. Here, it is important to note that HIV-1 mRNA is spliced and since 5' UTR is common among all splice isoforms, the increased RPF density may reflect cumulative ribosome density. Furthermore, the 5' UTR region is highly structured and multiple RNA-binding proteins (including) Gag are tightly bound to the UTR. This could possibly restrict nuclease access to all regions and some proteins may co-purify with the monosomes. Although, it is unlikely these would produce fragments which are similar in length to RPFs, we cannot exclude the possibility. In contrast to RiboSeq, RNASeq density is essentially constant across the entire genome, and at 24h, it increases steadily 5' to 3' reflecting cumulative density summed over the genomic RNA and subgenomic transcripts (**Fig. 2D-F**). The ribosome occupancy on Pol was much lower than in all other viral CDSs, due to the presence of the HIV-1 PRF site responsible for Pol production, which is discussed in detail in the later results.

Translation initiation across the HIV-1 and host genome

To map the canonical as well as novel initiation sites, we performed RiboSeq experiments in the presence of harringtonine, which results in the accumulation of ribosomes at translation initiation sites. This strategy led to a clear enrichment of ribosomes at the canonical start codons of the viral transcripts (**Fig. 3A**). Additionally, harringtonine treatment resulted in accumulation of ribosomes at alternative translation initiation sites. We took a closer look at the 5' UTR of HIV-1, as majority of initiating ribosomes within viral transcripts in the harringtonine dataset originated from the 5' UTR. We observed that several of the peaks were located at near-cognate AUG codons (**Fig. 3B**). However, it remains to be investigated whether these reads from the 5' UTR are bonafide RPFs and play any significant role in regulation of HIV-1 translation or in infection.

Next, we proceeded to characterize and identify non-canonical short open reading frames (sORFs) in the host as well as the viral genome. For this purpose, we employed PRICE analysis and identified ~1,000,000 sORFs in the human genome (gene list provided in **Supp. Table 2**). We increased the constraints applied to the dataset (see Materials and Methods) and eventually identified ~2400 upstream ORFs (uORFS) and ~700 each of downstream ORFs (dORFs) and internal ORFs (iORFs) in the human genome. uORFs are known to play a role in regulation of downstream ORF translation, especially during cellular stress conditions [165, 217]. Interestingly, we found atleast 200 genes where the expression of the uORF appeared to be temporally regulated. One such example is the two uORFs in the 5' region of the MDM2 gene, located 228 nt and 98 nt upstream of the canonical AUG codon. Previously, these uORFs have been identified as translational repressors of downstream MDM2 [218]. Interestingly, we see that the RPF densities at these uORFs increase from 0 to 24 hpi. Due to limited number of RiboSeq reads mapped to the CDS of MDM2, we could not quantify the effect on protein expression. We also identified a novel uORF located in the Importin 7 (IPO7) gene. Here, we saw decreased RPF densities on the uORF as well as the canonical ORF as infection proceeds, possibly due to a decrease in RNA levels (**Fig. 4A, B, Supp. Table 2**). Both MDM2 and IPO7 proteins have been known to play a role in viral infection. MDM2 enhances Tat-mediated HIV-1 replication through inhibition of p53 [219, 220]. IPO7 has been implicated in the regulation of ribosomal biogenesis and acts as an import factor for several proteins, including ribosomal proteins and HIV-1 integrase protein [221, 222]. On the HIV genome, we only identified 2 novel HIV-1 iORFs, within the vif and pol region. The gradual increase in RPF densities at 16 and 24h corresponding to these iORFs correlated well with the increase in RNA levels (**Fig. 4C, D, Supp. Table 2**).

Novel pause sites at the HIV-1 frameshift site

In addition to non-canonical initiation, HIV-1 utilizes a unique and well-studied mode of gene regulation, −1PRF. Canonically, the frameshift stimulatory site of HIV-1 is the heptanucleotide slippery sequence UUUUUUA and a downstream stem loop structure (**Fig. 5A**). We calculated the frameshifting efficiency of HIV-1 in infected T-cells using the ratio of ribosome footprints in the Pol and Gag coding sequences. The percentage of frameshifting was ~15-20% and generally remained constant throughout the course of infection, although slightly higher at 8h (**Fig. 5B**).

Looking specifically at the canonical frameshift site, we observed an accumulation of ribosomes near the site, more specifically a pause at the UUA leucine codon of the slippery site (**Fig. 5C upper panel;** peak 2). We saw several peaks (peak 3 and 4) approximately 12 and 24 nt upstream of the slippery site paused codon, which could represent ribosomes slowing down near the frameshift site. Previously, ribosome collisions at the frameshift site of Theiler's murine encephalomyelitis virus (TMEV) were shown to be crucial for viral frameshifting [98]. To observe the existence of collided ribosomes near the frameshift site, we modified the ribosome profiling protocol to include collided ribosomes or disomes [223, 224]. Disomes are routinely excluded during preparation of ribosome profiling libraries by the inclusion of a size-selection step (in this study, 25–34 nt) which selects for monosome-protected fragments. For mock as well as infected samples, we carried out DisomeSeq where fragments of 50-80 nt were size-selected. The read lengths were broader with local peaks of around 54 and 60–63 nt, consistent with expected lengths of RNA protected by disomes (**Supp. Fig. 1B**) [98, 223]. The densities of these disome protected fragments were plotted on the viral genome at the inferred P site position of the upstream stalling ribosome (**Fig. 5C lower panel**). Due to the variation in read-lengths the P-site position is not codon-resolved but rather shows the potential P-site distribution. Nevertheless, we did not observe ribosomes colliding at the UUA Leucine codon. However, a minor enrichment of reads corresponding to such ribosomes was observed upstream of the slippery sequence. This is a relatively proline rich region (~30% Pro), which is known to stall ribosomes due to exceptionally slow peptide bond formation [225].

Another pause was seen in the RiboSeq data at Arginine, 23 amino acids upstream of the paused Leucine codon (**Fig. 5C upper panel**). This paused codon lies at the base of the proposed extended frameshift site, where the upstream RNA sequence forms a 3- helix junction [97, 203]. This peak was several orders of magnitude higher than the pause at the slippery site. We also observed major ribosome collisions at this site in the disome data (**Fig. 5C lower panel**). This pause may be due to the ribosome resolving the extended frameshift structure or due to the nature and charge of the amino acids located at the P-site as well as the nascent peptide chain.

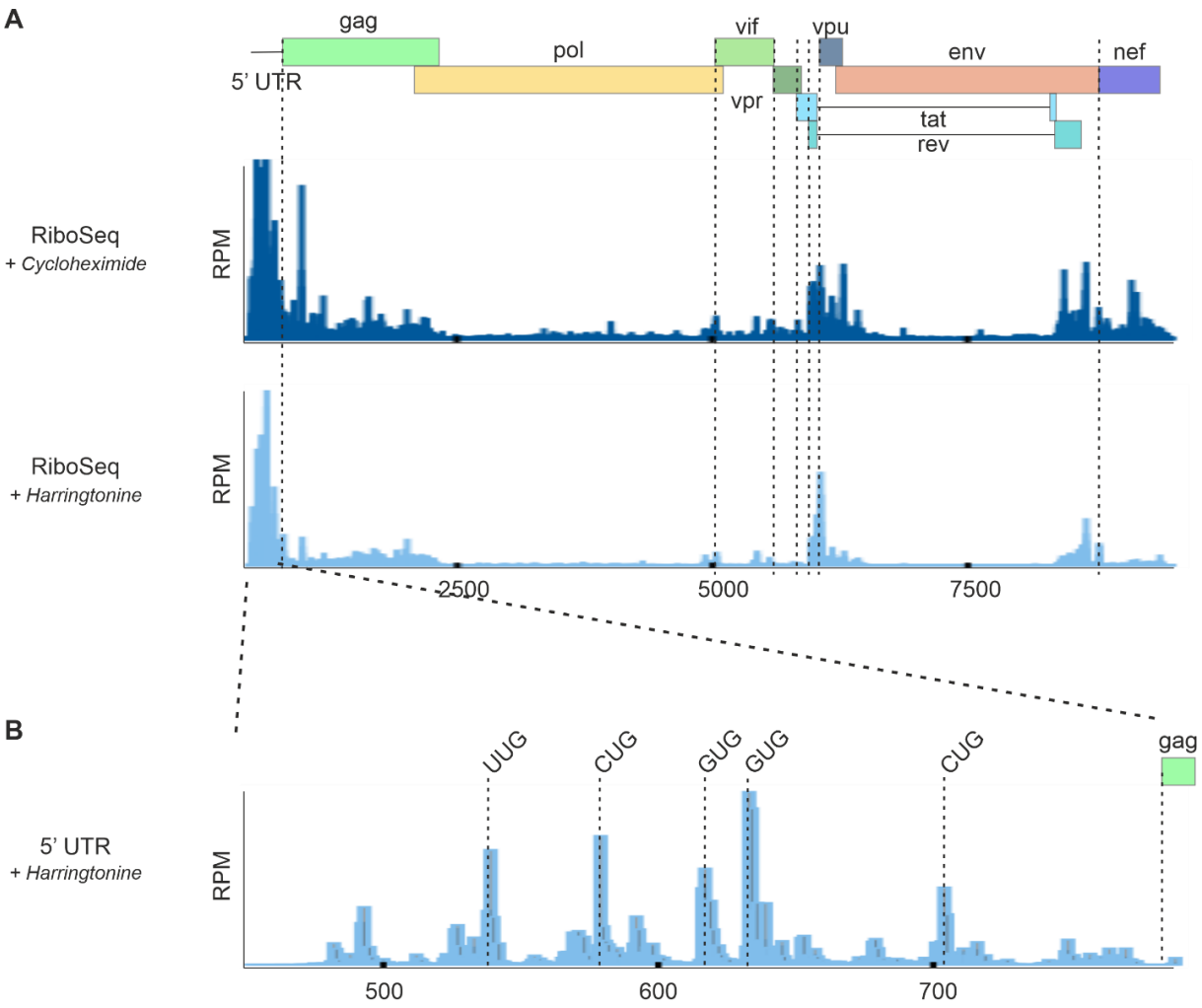


Figure 3: Translation initiation sites in viral transcripts. A) Distribution of ribosome P-sites around annotated start and stop codons in all viral transcripts in cycloheximide (dark blue) and harringtonine (light blue) libraries. Canonical start codons of viral transcripts are depicted. B) Close-up view of the HIV-1 5'UTR showing the distribution of ribosome P-sites and the position of putative non-AUG start-codons.

Experimental analysis of pause sites at the HIV-1 −1PRF site

To confirm these pause sites *in vitro*, we performed the ribosome pausing assay using rabbit reticulocyte lysates treated with harringtonine. We employed reporter mRNAs containing a flag tag followed by nucleotides 64 – 2687 (Δ1870 - 1881) of the HIV-1 genome to best mimic the native genomic context of viral frameshifting. To accurately mark the position of the predicted pause products, control mRNAs in which a stop codon has been introduced after the UAA and AGG respectively were also prepared (**Fig. 5D**). These mRNAs were

translated in RRL for 3 min prior to addition of the translation initiation inhibitor harringtonine. Translation was continued at 25°C, samples withdrawn at the indicated times post-harringtonine addition and translation products separated on a 15% SDS-PAGE gel. As shown in **Fig. 5E**, we observe a pause at the AGG indicating ribosome stacking occurs *in vitro*. However, the appearance of the pause product is transitory (albeit spread over several minutes), indicating this is an intermediate rather than a dead-end product. A distinct translational pause was also observed at the slippery site. However, unlike the pause at the AGG site, the slippery site pause was more pronounced and persists till the translation end-point, indicating ribosomal drop-off at the slippery site. To confirm the presence during 'normal' translation conditions, we conducted the same assay in absence of harringtonine and observed a product migrating at the same position as the respective control (**Fig. 5F, Supp. Fig. 3B**). These results indicate that a percentage of ribosomes drop off at the base of the HIV-1 frameshift stem-loop to form an alternative product. In contrast to the *in vivo* scenario, *in vitro* the slippery site pause is more distinct. This may be a bias of the *in vitro* experimental conditions or perhaps the presence of regulatory proteins which strengthen the pause in cells. In contrast, we observed no product at the AGG pause site for the WT RNA, confirming that the ribosome stacking at this position is transient and can be resolved (**Fig. 5F, Supp. Fig. 3B)**. Furthermore, to understand if this stacking is due to the nature of the amino acid at the P-site or presence of secondary structure, we mutated the arginine codon. We also mutated the following lysine codon from AAA to AAG (also a lysine) to remove the potential slippery sequence context downstream of AGG. When the arginine was mutated to uncharged Glycine, we observed no product at AGG, as seen in the WT RNA. When we mutate Arginine to charged Lysine, we observe ribosomal drop-off at the AGG pause site (**Fig. 5F, Supp. Fig. 3B)**. However, this can be attributed to the presence of multiple lysines (KKK) in the nascent peptide chain as translation proceeds, in line with previous studies [226-229].

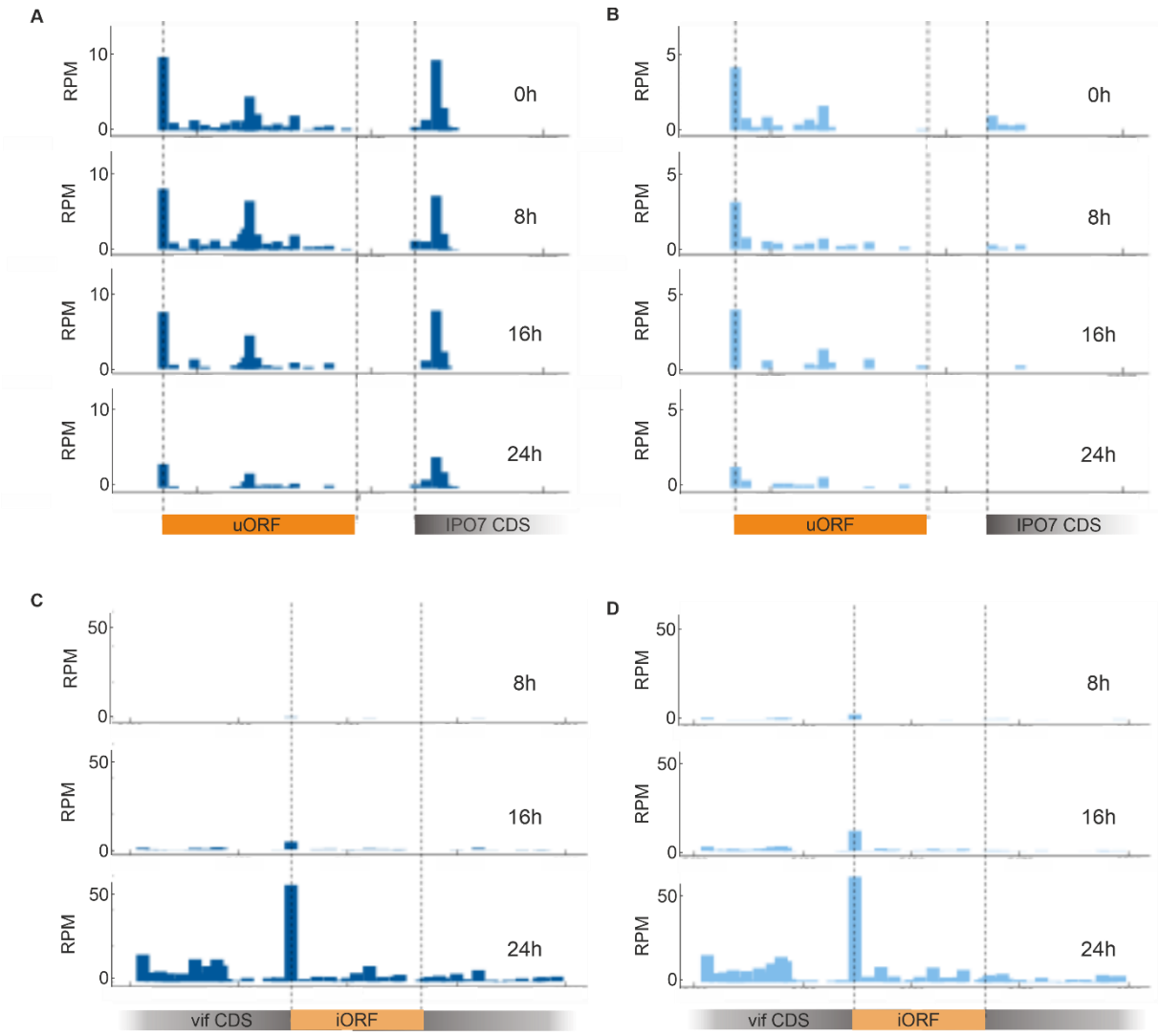


Figure 4: Novel sORFs in host and viral transcripts. Distribution of ribosome P-sites in the uORF detected in the 5' UTR of IPO7 gene in A) cycloheximide (dark blue) and B) harringtonine (light blue) libraries at each time point. Distribution of ribosome P-sites in the iORF detected in the vif coding region in C) cycloheximide (dark blue) and D) harringtonine (light blue) libraries at each time point.

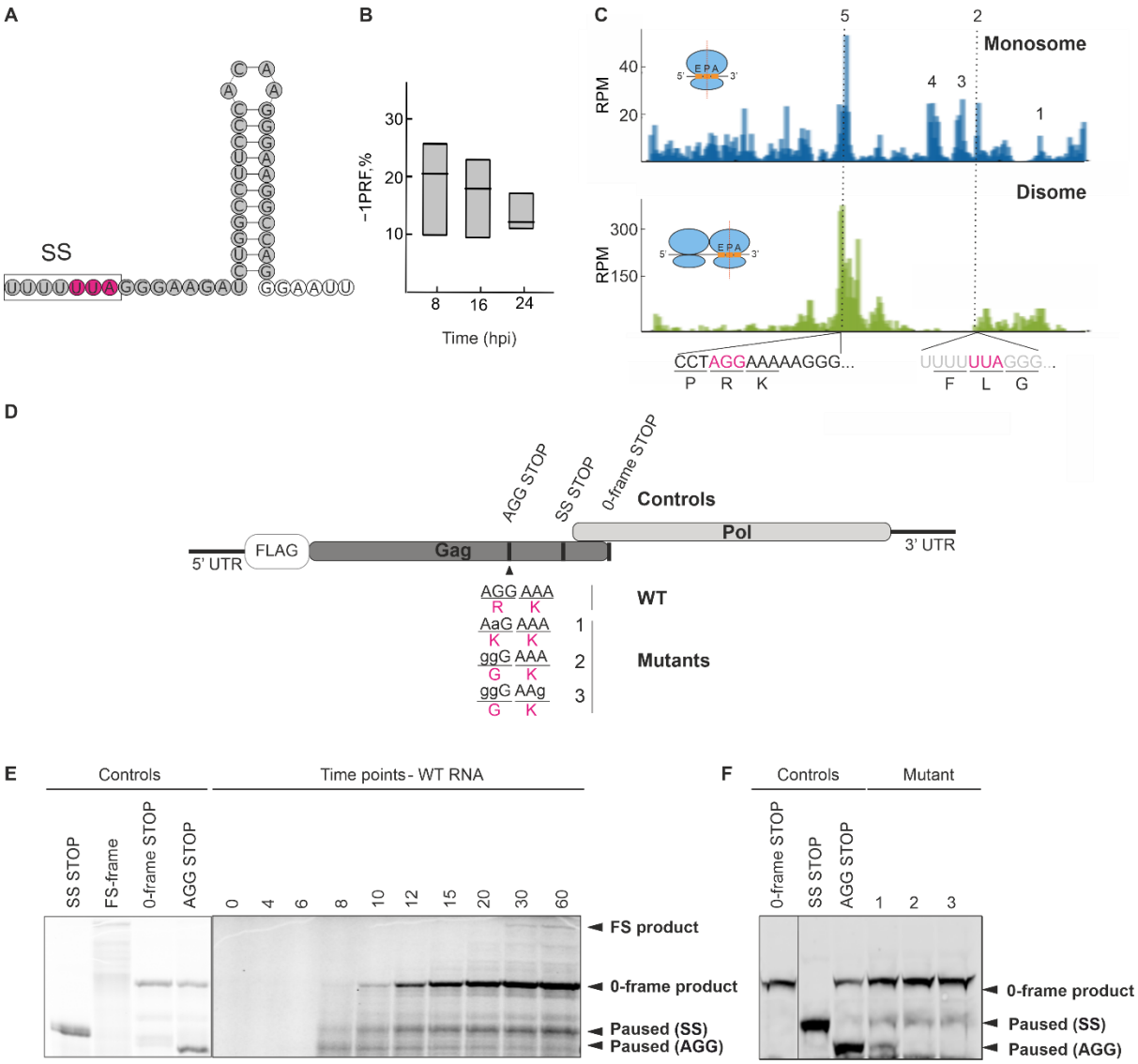


Figure 5: Investigation of pause sites near the HIV-1 −1PRF site by ribosome and disome profiling. A) Predicted frameshift stimulatory stem-loop structure of HIV-1. Slippery sequence (SS) is boxed and *in vitro* pause site at UUA is depicted in pink. B) Percentage of Gag-Pol ribosome frameshifting efficiency at each time point of infection. C) Distribution of stalling ribosome P-sites of the 200nt region up to and including the −1PRF site in ribosome profiling libraries (dark blue) and disome profiling libraries (green). Peaks are denoted by numbers and dotted lines represent the peak at the stalling codon of the slippery sequence and upstream AGG codon respectively. D) Schematics of the N-terminal FLAG-tagged frameshifting reporter consisting of the nucleotides 64 – 2687 (Δ1870 - 1881) of the HIV-1 genome, indicating the controls and mutation sites. E) Time course of translation of reporter mRNA and controls in RRL. Briefly, translation was allowed to proceed at 26°C in the presence of [^{35}S] methionine for 3 min prior to addition of harringtonine to a final concentration of 200 µM. Samples were withdrawn at the indicated times after harringtonine addition, and translation products separated on a 15% SDS-PAGE gel and detected by autoradiography. The predicted position of the pause products was determined from the controls (see text). F) RNAs were translated in RRL at 30°C in absence of harringtonine.

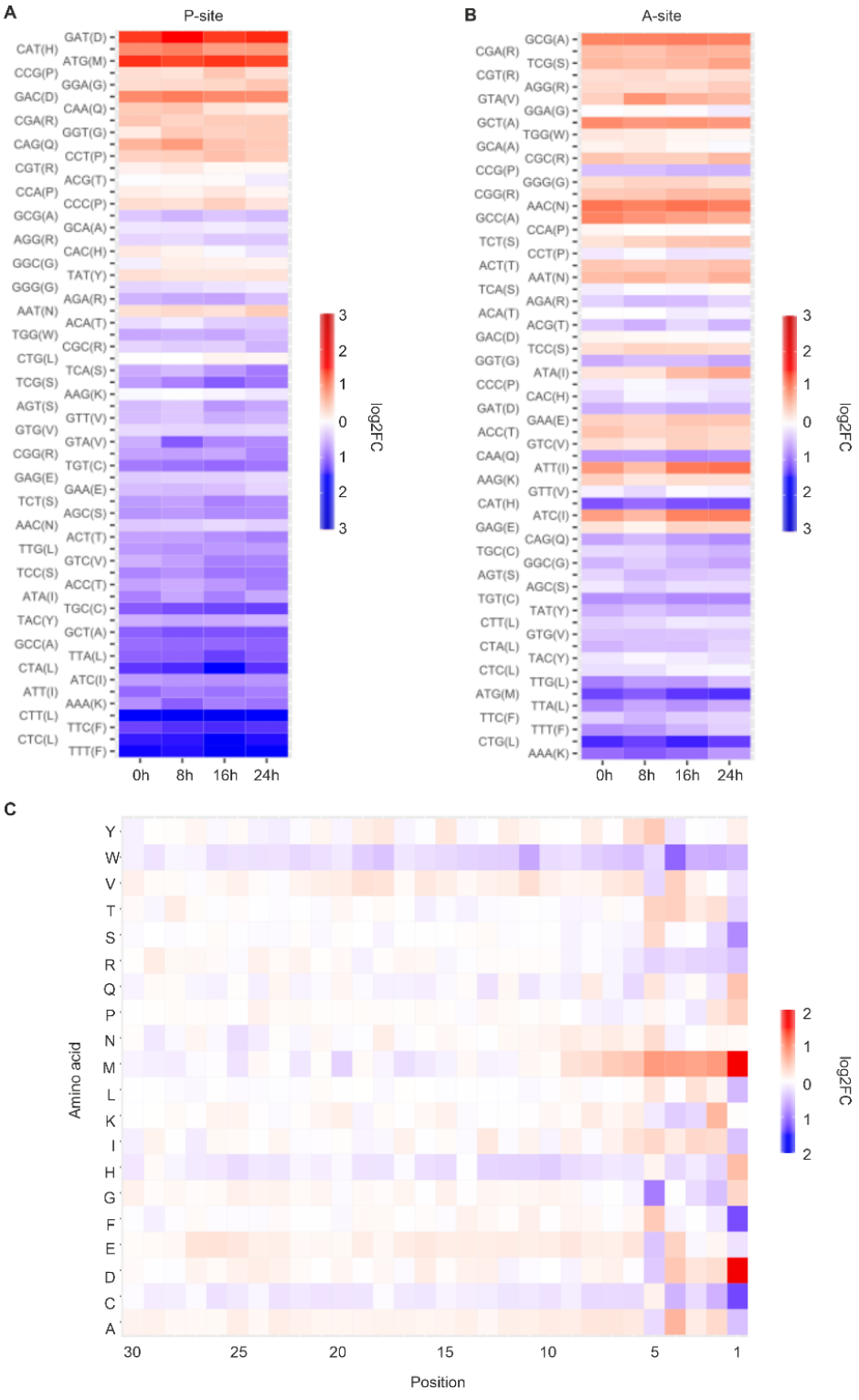


Figure 6: Stalling landscape of human T-cell genome during HIV-1 infection. Codon specific stalling in the A) P-site and B) A-site on a genome wide scale. Fold change (log_2) between the expected human codon usage frequencies and observed stalling frequencies of the codons, normalized to the randomized control, are plotted at each timepoint. C) Distribution of amino acids along the exit tunnel of stalled ribosomes at 24 hpi. Fold change (log_2) between the stalling frequencies of the codon and the codon frequencies in the randomized control are plotted. (See Materials and methods)

Stalling landscape of the human genome during infection

Like HIV-1, we observed multiple stalling events during elongation in our human RiboSeq data. Viral infections are known to induce host integrated stress response, leading to preferential translation of stress response genes by altering codon usage and cellular tRNA pool [230, 231]. Furthermore, HIV-1 has been shown to manipulate the host tRNA abundance and take advantage of this conserved stress response to translate its A/U-rich genome [232-234]. We checked if HIV-1 mediated translational remodeling leads to a change in the global stalling landscape of the host during early- and late stages of infection. To obtain high-confidence stall sites, we devised a method to detect stalls above the noise level (See Materials & Methods). As a control, we generated a similar profile using a collection of random sites of the same RiboSeq data. As expected, in the P-site, we see stalling at the canonical AUG start-codon, which is not seen in the random dataset (**Fig. 6A**). Furthermore, we see stalling events at Proline (P), Glycine (G) and Aspartic acid (D), which were shown to be conserved stall sites in most organisms, including humans [227]. Similarly, we checked the codon enrichment at the A-site of stalled ribosomes. We saw an enrichment of certain codons of Alanine (A), Serine (S), Asparagine (N) and Isoleucine (I) at the A-site (**Fig. 6B**). This A-site stalling can be contributed to the presence of rare codons, *e.g* we see stalling at the TCG Serine codon which is very rare (4%) as compared to the commonly used AGC Serine codon where we see no stalling events. A-site stalling can also occur when a certain amino acid is depleted or in limiting amounts, such as in cases of starvation. T-cell proliferation and activation as well as immune response pathways depends on the availability of alanine, asparagine and isoleucine [235-238]. It is likely that these amino acids are present in limiting amounts in SupT1 cells which may explain their enrichment in the A-sites of stalled ribosomes. In fact, HIV-1 restricts alanine uptake in T-cells through Vpu-mediated downregulation of amino acid transporter SNAT-1 [237]. Therefore, this may be a viable viral strategy to regulate immune cell activation. Although, enrichment of Alanine did not change temporally during infection, we found that stalling at the A-site due to Isoleucine codons mildly increases during the course of infection. Although Alanine and Aspartic acid show the highest propensity to stall a ribosome at A- and P-site respectively, it is interesting to note that no codon, except perhaps Arginine, has the potential to stall ribosomes at both of these sites (**Supp. Fig. 3A**). Overall, there is not a massive change in the stalling landscape of SupT1 cells, upon HIV-1 infection.

During protein synthesis, the nascent peptide chain that emerges from the ribosome through the ribosomal exit tunnel also plays a role in ribosomal stalling [239-243]. Newly synthesized, positively charged amino acids, *e.g.* multiple lysines in a row, were shown to induce stalling by interacting with the negatively charged exit tunnel [226-228, 239, 242]. Another study found the presence of negatively charged amino acids in the exit tunnel of the stalled ribosome [243]. We therefore analyzed the 30 amino acids upstream of stall sites that would span the ribosome exit tunnel, using random sites in the same dataset as a control (**Fig. 6C**). We found a large contribution of charged amino acids throughout the exit tunnel, but particularly strong lysine and glutamate enrichment at positions 2 and 4, respectively (**Fig. 6C**). Additionally, it has been shown that aromatic amino acids, Phenylalanine, Tryptophan, Tyrosine, in nascent peptides enhance ribosome arrest in bacteria [244, 245]. While Tyrosine appeared to be enriched in positions 2-5 and phenylalanine in position 5, we found no significant enrichments of tryptophan (**Fig. 6C**).

Discussion

HIV-1 is a complex retrovirus which produces multiple proteins through a combination of alternative splicing, translational and post-translational mechanisms, which are temporally regulated throughout infection [179, 180]. Additionally, it utilizes a myriad of non-canonical initiation and elongation translation mechanisms including scanning, leaky scanning, internal ribosome entry sites, adenosine methylation and programmed ribosomal frameshifting to suit its specific translational needs [182, 183, 201]. Here, we describe in detail the translational landscape of host and viral transcripts during HIV-1 replication through ribosome profiling and RNA sequencing.

Global analysis of the host response to early HIV-1 infection (8hpi) revealed a minor perturbance in the cytoplasmic transcriptome and remodeling of the translatome, indicating that changes in TE of certain host mRNAs may play a dominant role in cellular response to infection. Amongst the translationally regulated genes we identified several defense response genes known to severely abrogate HIV-1 replication such as Schlafen family member 11 (SLFN11) [246], G3BP stress granule assembly factor 1 (G3BP1) [247], and Dead box helicases (DDXs) [248] which were upregulated and a translational downregulation of interferon induced transmembrane proteins (IFITMs) [249]. We also observed translational upregulation of signal transducer and activator of transcription 1 (STAT1) which is shown to be activated by HIV-1 and plays an integral role in HIV-1-induced inflammatory

responses [250]. Late HIV-1 infection, in our case 16 and 24 hpi, induced a strong perturbation of the cellular cytoplasmic transcriptome and a significant change in the translational status of a restricted group of transcripts. Cellular transcripts where expression is down-regulated are mainly related to the regulation of gene expression and include factors involved in ribosome biogenesis, translation, tRNA biogenesis, mRNA splicing and processing, perhaps resulting from a cellular response to the stress. Amongst the commonly differentially regulated genes between all the time points, Synaptotagmin-like protein 3 (SYTL3) gene showed extensive changes in transcript abundance as well as ribosome occupancy. SYTL3 protein plays a role in vesicular trafficking and is said to regulate the release of viral particles [251]. We also saw transcriptional upregulation in GTPase of the immunity-associated protein family (GIMAP) genes in all datasets which are profoundly expressed in immune cells, playing an important role in lymphocyte development and are commonly associated with autoimmune and inflammatory diseases [252, 253].

It has been proposed that HIV-1 induces host translation shut-off to favor the translation of its own transcripts [184, 195-198, 216]. Indeed, HIV-1 infection is known to decrease cap- and poly(A)-dependent initiation of translation through viral protease-mediated degradation of eIF4G and PABP [181, 196]. One of the mechanisms through which HIV-1 mRNA translation is sustained is the distinct composition of the HIV-1 ribonucleoprotein complexes, specifically the association of HIV-1 unspliced transcripts with nuclear cap binding protein 80 (CBP80) [216]. Interestingly, we observed a consistent increase in translational efficiency of CBP80 throughout the course of infection. Although HIV-1 infection affected the translational status of certain mRNAs, we did not observe a drastic translational shut-off, as seen in RNA viruses such as picornaviruses, at the time-points tested. However, HIV-1 directly relies on the host machinery for synthesis of its capped and polyadenylated viral mRNAs which can be translated in a cap-dependent manner, hence the lack of translational shut-off is not surprising [182, 185].

One of the prevalent mechanisms of regulation of translation initiation is the expression of uORFs, which are extensively found in viruses as well as in eukaryotes [254-256]. uORFs generally inhibit translation of the main ORF by restricting the fraction of initiating/scanning 40S ribosomal subunits that can reach the canonical start codon [162, 217, 255]. Careful analysis of the translational landscape of viral transcripts uncovered several translation initiation sites in the HIV 5' UTR, pointing to the presence of uORFs. However, further studies would be required to demonstrate if the translation from these HIV-1 uORFs

is productive and whether it plays a role in HIV-1 translation regulation or immune response. It was recently observed that the cellular amount of uORF-encoded peptides increases during stress [165]. Through PRICE analysis, we identified novel uORFs in the human genome where expression of the uORFs was temporally regulated. Further analysis of the uORFs discovered in our study revealed that a large number of these genes are involved in T-cell activation, differentiation and receptor signaling pathways. Immune responses, in particular, require precise, dynamic gene regulation at the post-transcriptional level that must activate rapidly as threats rise and resolve efficiently as they disappear. Thus, temporal expression of uORFs would offer an elegant strategy for the cell to actively respond to environmental stresses.

Besides translation initiation, elongation of HIV-1 can also occur through non-canonical means, specifically through programmed ribosomal frameshifting. Studies examining the secondary structure of the HIV-1 genomic RNA within capsids have revealed that the frameshift site is part of a conserved three-helix junction (3HJ) [203, 257]. Despite the extensive *in vitro* analysis of the frameshift mechanisms, not much is known about the role of stalling in HIV-1 −1PRF. Our disome and monosome analysis revealed a significant pausing event upstream of the slippery sequence. Here, the ribosome would be positioned at the base of the three-helix RNA structure. It is likely that the RNA structure promotes ribosome stacking which can decrease the rate of translation, thereby facilitating pausing at the slippery sequence. It has been known that factors such as increased translation initiation rates lead to increased polysome density, which can cause ribosomes to stack at the frameshift site. This in turn affects the rate of mRNA refolding during translation and leads to a decrease in overall frameshift efficiency [160, 258]. It is likely that the extended structure leading to the pause functions in a similar fashion, decreasing the number of ribosomes that can stack at the frameshift site, allowing frameshifting to occur at the appropriate level. Extensive ribosome stalling can also occur due to the presence of charged amino acids in the nascent peptides. In this case, the nascent peptide contains nearly 30% charged amino acids which could contribute towards the pausing. Furthermore, the pause occurs at the RKK motif and several studies in yeast have shown that consecutive Lysine or Arginine codons can induce translation arrest, leading to the formation of disomes [223, 259-261]. Additional studies would be required to determine the exact mechanism of this pause and its function in HIV-1 frameshifting. However, we hypothesize that ribosome stacking

upstream of the frameshift site is an additional layer of frameshift regulation to maintain the critical ratio of structural and enzymatic proteins during HIV-1 replication.

Taken together our work largely enhances the understanding of post-transcriptional regulation of both host and viral mRNAs during HIV-1 infection. It presents new paradigms of non-canonical mechanisms of HIV-1 translation, thereby paving the way for novel antiviral targeting.

Materials and methods

Cell culture

SupT1 cells (NIH-AIDS reagent database) and Jurkat cells (NIH-AIDS reagent database) were maintained in RPMI supplemented with 10% FBS (Gibco), 100 µg/ml streptomycin, 100 U/ml penicillin and 2 mM Glutamine. HEK293 cells (gift from Prof. Jörg Vogel, HIRI-HZI) and Huh7 cells (gift from Dr. Mathias Munschauer, HIRI-HZI), were maintained in DMEM (Gibco) supplemented with 10% FBS (Gibco) and 100 µg/ml streptomycin and 100 U/ml penicillin.

HIV-1 virus purification and infection

18 million HEK293T cells were seeded in a 15 cm^2 plate with 15 ml media 24h prior to transfection. Cells were then co-transfected with 10 µg pNL4-3 ΔEnv-iGFP (NIH AIDS Research and Reference Reagent program, Cat. #11100) and 2 µg pCMV VsVG (gift from Dr. Redmond Smyth, HIRI-HZI) using PEI (DNA:PEI =1:12). 48h post transfection, supernatant was collected and centrifuged at 2000g for 20 min to remove cell debris. The supernatant was then filtered through 45 micron filters, loaded onto the 20% sucrose cushion (Tris-HCl 50mM pH 7.4, NaCl 100 mM, EDTA 0.5mM) with a ratio 4:1 and centrifuged at 4500 rpm for 20h in a Beckmann rotor JLA16.250. The virus pellet was resuspended in PBS, treated with DNase I for 1h at 37°C, distributed into microfuge tubes and kept at -80°C until further use.

For ribosome, disome and polysome profiling experiments, cells were infected with 50 µl of virus suspension per 40 million cells by spinoculation at 1500 g for 30 min at 37 °C, in presence of 8 µg/ml polybrene (Sigma). Cells were pelleted at 1000 RPM for 3 min at 37 °C, washed once with warm media, resuspended in 40 ml RPMI medium and incubated at 37 °C.

Ribosome profiling and RNA sequencing

Ribosome profiling samples were prepared as previously described (McGlincy and Ingolia., 2017) with modifications for suspension cell lines. Briefly, at each time point of infection, cells were treated with 100 μg/ml cycloheximide in DMSO at 37 °C for 2 min. For harringtonine treated samples, cells were treated with 2 μg/ml cycloheximide in DMSO at 37 °C for 5 min, followed by cycloheximide treatment. Cells were immediately pelleted at 1000 RPM for 3 min at 4 °C, washed once with cold cycloheximide containing PBS and the cell pellet was snap-frozen in liquid nitrogen and stored at -80 °C, till further use. Cell pellet was thawed in the presence of ice-cold lysis buffer containing 1% Triton-X and triturated 10 times through a 26G needle. The lysate was centrifuged at 17000g for 10 min at 4 °C and the supernatant was recovered. Cell lysates were subjected to RiboSeq and RNASeq.

The methodologies employed for ribosome profiling were based on the original protocols of Ingolia and colleagues [213, 262], except rRNA contamination was removed using riboPOOLs rRNA depletion kit (siTOOLs Biotech) using the manufacturer's instructions. For disome profiling, a broad range of RPFs (50-80nt) were size selected. Quality of the library was assessed by using a BioAnalyzer via the High Sensitivity DNA Kit (Agilent; 5067-4626). Sequencing experiments were performed by either Novogene or the Core Unit Systems Medicine (University of Würzburg): Sequencing of Riboseq – For cycloheximide treated samples- first replicate was sequenced by Novogene (Novaseq SE50- 50 million reads per sample), the next 2 replicates were sequenced at the Core Unit Systems Medicine (Illumina Nextseq 2000 P3 SE100 - ~150 million reads/sample). Harringtonine treated samples were sequenced at the Core Unit Systems Medicine (Illumina Nextseq 2000 P3 SE100 - ~75 million reads/sample). Disome profiling samples were sequenced at the Core Unit Systems Medicine (Illumina Nextseq 2000 P3 SE100 - ~75 million reads/sample).

For RNASeq, total RNA was isolated from the cell-lysate using Trizol (Invitrogen), following manufacturer's instructions. RNA-seq libraries from these samples were constructed using CORALL Total RNA-Seq Library Prep Kit (Lexogen), according to manufacturer's instructions. Quality of the library was assessed by using a BioAnalyzer via the High Sensitivity DNA Kit (Agilent). Sequencing experiments were performed by either Novogene or the Core Unit Systems Medicine (University of Würzburg): For one replicate, sequencing was done at Novogene (Novaseq PE150- 20 million reads per sample) and the

other two replicates were sequenced at the Core Unit Systems Medicine (Illumina Nextseq 500 Mid-output SE150 - ~16 million reads/sample).

Bioinformatic analysis

For RiboSeq

For the computational processing of the ribosome profiling reads, PRICE pipeline was utilized [263]. In addition to performing all pre-processing and pre-filtering steps, it determines the P-site codon for each individual Ribo-Seq read with a probabilistic model. Here, the demultiplexed FASTQ files were run on the pipeline using a JSON file specifying the path to the raw sequencing data and mapping parameters (**Source Data**). With the correctly prepared JSON file, BASH scripts streamlining the pipeline were created as instructed (https://github.com/erhard-lab/price). Finally, the BASH scripts were executed and the PRICE coverage data, a report folder with mapping statistics and BAM files with the mapped reads were generated by the pipeline. Visualization of PRICE coverage information was performed with the built-in *RiboView* functionality of *gedi*. The .bam files generated by PRICE were assigned to genomic features with featureCounts tool [264]. For analysis of differential gene expression, the deltaTE pipeline was utilized [265]. The classification system of the pipeline was adjusted to also distinguish between RiboSeq up- and downregulated genes. Since deltaTE always compares two timepoints to each other, the RScript was adjusted for each comparison. Furthermore, the option to consider batch effects was selected. The pipeline was ultimately executed as instructed (https://github.com/SGDDNB/translational_regulation). The RScript, count matrices, gene lists, fold change tables, as well as some interactive plots are included in the supplemental material. For disome profiling, the preprocessing steps were similar to ribosome profiling, except P-site determination was performed as described below..

For gene ontology analysis the online tool of the *Panther* classification system was employed. GO analysis was performed on the human host genome and enrichment for biological processes was selected. Uploaded genes were provided with Ensembl identifiers.

For RNASeq reads

In general, the preprocessing steps of the RiboSeq and RNASeq analysis were the same. As the preprocessing is already integrated in the PRICE pipeline for the Ribo-Seq analysis, the pre-processing steps of the pipeline were manually implemented. The linker sequences

were trimmed and the data was filtered using *fastp*. rRNA and Mycoplasma was removed by aligning the reads to the specific genomes using Bowtie2. The remaining reads were mapped to the human genome using STAR tool and reads were deduplicated using DedupUMI. For the visualization of mapped RNA-Seq and Ribo-Seq reads in .bam format, the Integrative Genomics Viewer (IGV) was used.

Ribosomal stalling patterns in host translation

To investigate ribosomal stalling, P-site codon information was utilized. For monosomes, P-sites were determined by PRICE, however, for disomes P-site determination was performed manually. Sequencing reads with a length between 53-67 nucleotides were processed. Here, fixed P-site offsets were applied. The offset for the stalling ribosome was presumed to be 16 nucleotides from the 3'-end of the read. A self-written python script is available in the supplemental material.

For the ribosomal stalling patterns, cycloheximide treated RiboSeq datasets were used. All codon coverage peaks were sorted into bins depending on the genomic distance to the next peak so that individual translational active regions end up in a single bin. If the distance between peaks exceeded 100 nucleotides, a new bin was created. Bins with less than four entries were discarded. From the remaining bins, the highest peaks were considered potential stalling sites when the read count was at least two times higher than the mean of all peaks in the bin. Depending on the size of each bin a different number of maximum peaks were selected, specifically the number of considered peaks was determined by floor division for bins with more than 10 entries (number of max values = binsize // 10). For comparison, a random dataset was created from the same bin distribution, but the peaks of each bin were selected with a random function and there was no criterium of minimal coverage. The A-site codon downstream and the 30 codons upstream (coding for nascent peptide) were also identified in a similar manner. For sequence extraction and translation into an amino acid sequence, the Biopython library was employed [266]. Python scripts guiding through the workflow are available in the supplemental material. The frequencies of the P-site and A-site potential stalling events were compared to the human codon usage frequencies obtained from CoCoPUTs [267, 268]. The fold change of expected to obtained frequencies was calculated and finally the fold changes (log_2) were normalized with the fold change (log_2) values obtained from the random subsample. For the nascent peptide analysis, the frequency of every amino acid was individually counted for each position of

the ribosomal exit tunnel and the fold change (log_2) in frequency was directly calculated in comparison to the random subsample.

Plasmid construction

To generate reporter constructs for *in vitro* translation and ribosome pausing, golden gate compatible vectors were generated and protein-coding sequences were introduced by Golden Gate Assembly using AarI cut sites [269]. The reporter vector contained ß-globin 5′ and 3′ UTRs as well as a 30 nt long poly-(A) tail. The insert was derived from nucleotides 64 – 2687 (Δ1870 - 1881) of the HIV-1 genome; a 3×FLAG-tag was introduced at the N-terminus to facilitate detection. To generate 0% and 100% –1PRF controls, the –1PRF site was mutated by disrupting the PK structure as well as the slippery sequence. For ribosome pausing, controls were generated by adding two stop codons (UAA) downstream of the UUA codon of slippery site as well as pause site (AGG codon) discovered in this study.

Ribosome pausing

mRNAs were in vitro transcribed using T7 polymerase purified in-house using linearized plasmid DNA as the template. RNAs were translated in nuclease-treated rabbit reticulocyte lysate (RRL) (Promega) programmed with ~50 μg/ml template mRNA. A typical reaction mixture was composed of 90% (vol/vol) RRL, 20 μM amino acids (lacking methionine), and 0.2 MBq [^{35}S]-methionine. The translational inhibitor harringtonine was added 3 min after the start of the reaction in order to obtain synchronous initiation (final concentration, 5 μg/ml). Aliquots of 2.5 μl were withdrawn from the translation reaction mixture at specified intervals and stopped by the addition an equal volume of 10 mM EDTA, 100 μg/ml RNase A (Invitrogen). Following incubation at room temperature for 20 min, 3 Vol of 2× Bolt™ LDS Sample Buffer (Invitrogen) was added, and the samples were denatured at 70 °C for 10 min before loading on a 20x20 cm 15% SDS-PAGE. Samples were resolved at 300 V for 3 h. For autoradiography, the gels were dried at 80 °C for2 h under vacuum, exposed to phosphor screens (Fujifilm) for at least 5 days. The phosphor screens were then scanned with a Typhoon™ FLA 7000 (GE Healthcare).

***In vitro* translation**

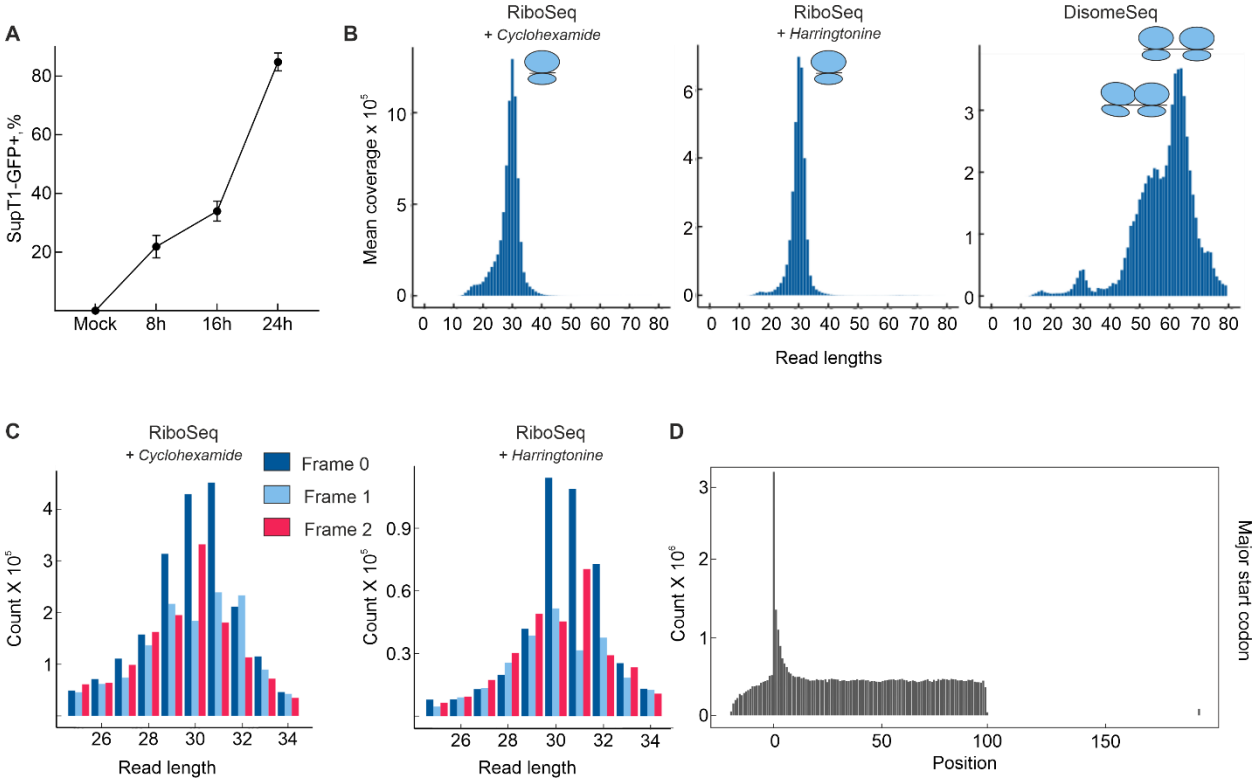
For in vitro translation assays, conditions were kept the same except reactions were prepared without [^{35}S]-methionine, instead 20 μM amino acids (lacking cysteine) was added. Reactions were incubated for 1 h at 30 °C. Samples were mixed with 3 Vol of 1X

Bolt™ LDS Sample Buffer (Invitrogen), denatured at 70 °C for 10 min, before loading on a NuPAGE™ 4–12% Bis-Tris polyacrylamide gel (Invitrogen). The products were detected using western blot and transfer was done using Trans-Blot (Bio-Rad). The nitrocellulose membranes were developed using anti-DDDDK primary (Abcam ab49763) and IRDye® 680RD donkey anti-rabbit secondary antibody (LI-COR). Bands were visualized using an Odyssey Clx infrared imager system (LI-COR).

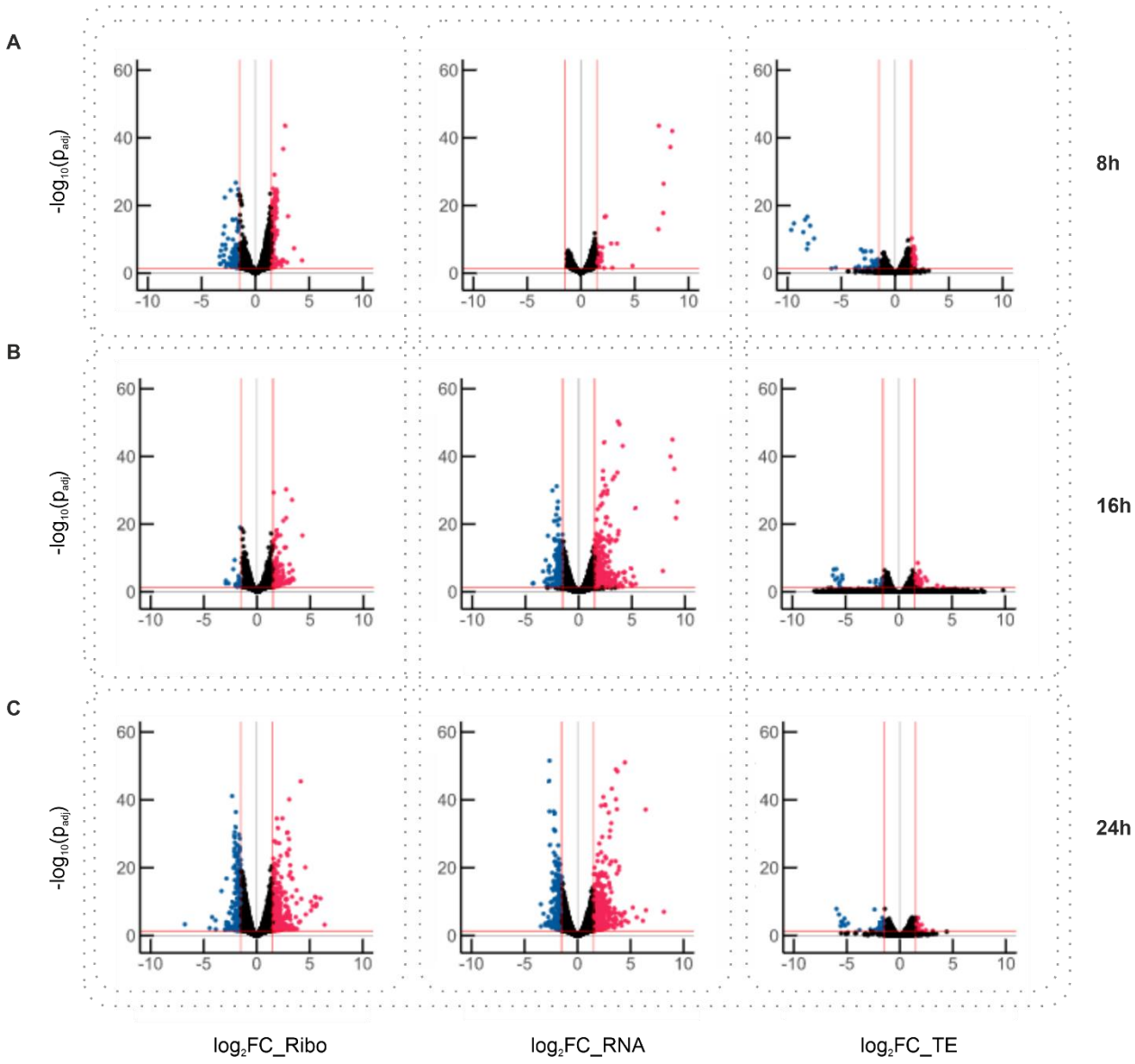
Data availability

Supplementary information indicated in the manuscript is included in the **Source data** provided with this thesis.

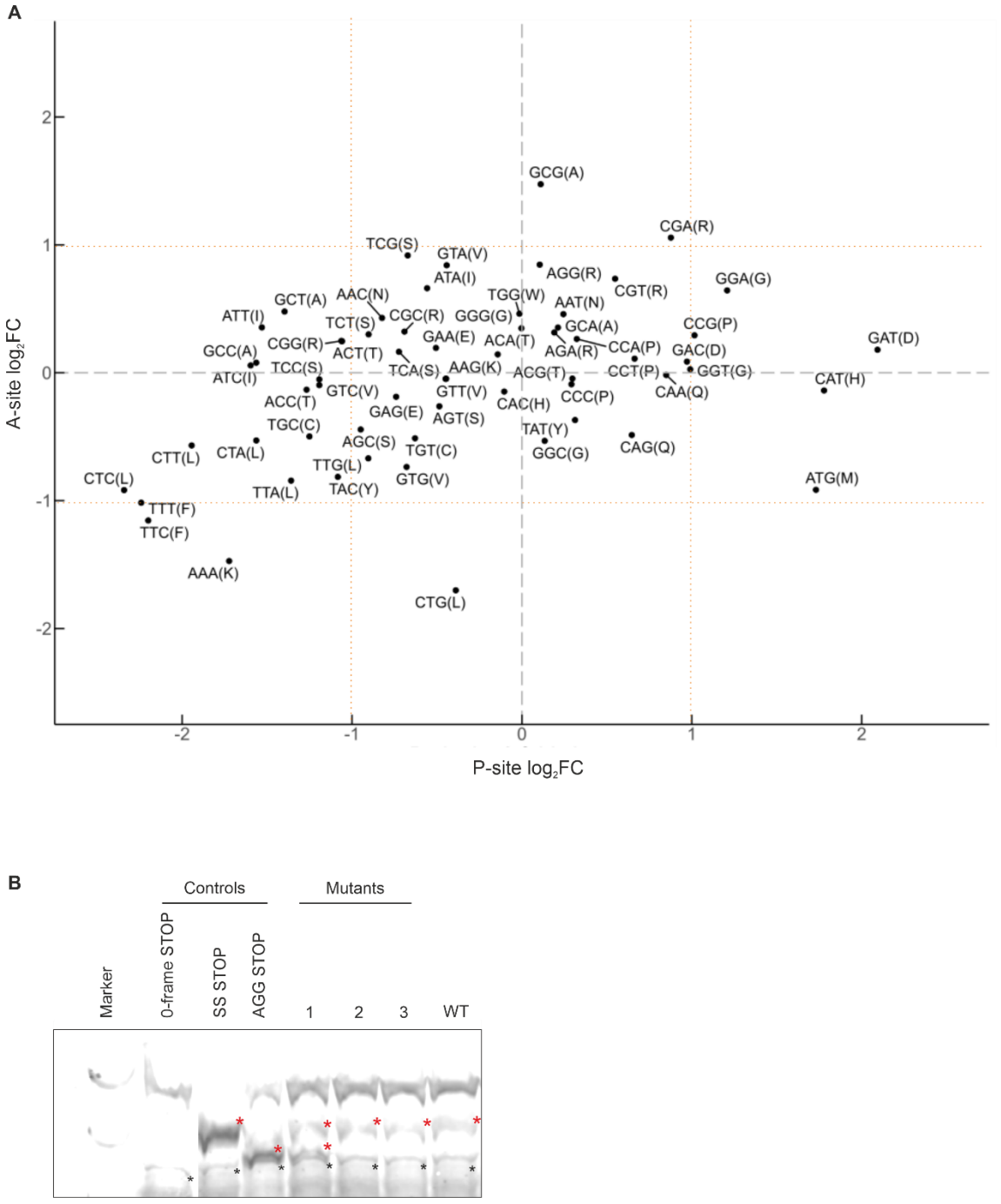
Supplementary Information



Supp Fig 1: HIV-1 infection and quality control of the ribosome and disome profiling libraries, related to Fig. 1. A) Percentage of GFP-positive SupT1 cells at each time point. B) Read length distribution of mapped reads of the ribosome profiling and disome profiling libraries. C) The frame distribution per read length of reads overlapping a CDS in the ribosome profiling libraries. D) Number of reads starting within coding transcripts. The x axis is the relative position of the read start within the three regions (5'-UTR: <0; CDS: 0-100; 3'-UTR: >100). The major isoform is used as reference for each gene (the longest coding transcript).



Supp Fig 2: Host differential gene expression in HIV-1 infected cells, related to Fig.1. Differences in ribosome occupancy (left), transcription (middle) and translation efficiency (right) were determined using deltaTE at A) 8 hpi, B) 16 hpi and C) 24 hpi. Volcano plots show relative fold changes (log_2) between the mock and infected samples at each time point. The y-axis shows the $-1log_{10}$ value of the false discovery rate (FDR)-corrected p-values. The red lines mark the thresholds of a log_2 fold change +/- 1.5 (vertical) and p-value of 0.05 (horizontal). The significantly down regulated genes are colored in blue, whereas the significantly up regulated genes are colored in red. Interactive versions of the plots are included in the Source data. (A) 8 hpi compared to mock. (B) 16 hpi compared to mock. (C) 24 hpi compared to mock.



Supp. Fig. 3: Stalling sites of host and viral genome, related to Fig. 5 and 6. A) Comparison of codon-specific stalling at the A- and P-sites at 24 hpi. Fold change (log_2) between the expected human codon usage frequencies and observed stalling frequencies of the codons, normalized to the randomized control, were determined for A- and P-sites and plotted against each other. B) RNAs described in Fig. 5E, including the WT RNA, were translated in RRL at 30°C in absence of harringtonine. Translation products separated on a 12% SDS-PAGE gel and FLAG-tagged peptides were identified via western blotting using anti-DDDDK antibody. The red star indicates genuine pause sites and the black star indicates a non-specific product detected in the RRL by the anti-rabbit secondary antibody.

Chapter 3

Revealing the host antiviral protein ZAP-S as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting

Matthias M. Zimmer[1,†], Anuja Kibe[1,†], Ulfert Rand[2], Lukas Pekarek[1], Luka Cicin-Sain[2], Neva Caliskan[1,3,*]

[1] Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Zentrum für Infektionsforschung (Helmholtz Centre for Infection Research), Josef-Schneider-Strasse 2, 97080, Würzburg, Germany

[2] Helmholtz Zentrum für Infektionsforschung, Inhoffenstrasse 7, 38124, Braunschweig, Germany

[3] Medical Faculty, Julius-Maximilians University Würzburg, 97074, Würzburg, Germany

[†] ***These authors contributed equally to this work.***

[*]Corresponding author:.

Neva Caliskan neva.caliskan@helmholtz-hiri.de

Published: Nature Communications volume 12, Article number: 7193 (2021)

Manuscript modifications: In this thesis, I have only included the supplementary figures I significantly contributed to. The rest of the supplementary data, including the tables, is available online or provided as source data with this thesis. Supplementary figure 3 contains additional data studying interactions of eukaryotic ribosomal subunits with ZAP-S *in vitro* **(Supp. Fig. 3J)**. This was part of the follow-up work I performed and the 'Results' and 'Materials and methods' sections have been updated accordingly.

Abstract

Programmed ribosomal frameshifting (PRF) is a fundamental gene expression event in many viruses including SARS-CoV-2, which allows production of essential structural and replicative enzymes from an alternative reading frame. Despite the importance of PRF for the viral life cycle, it is still largely unknown how and to what extent cellular factors alter mechanical properties of frameshifting RNA molecules and thereby impact virulence. This prompted us to comprehensively dissect the interplay between the host proteome and the SARS-CoV-2 frameshift element. Here, we reveal that zinc-finger antiviral protein (ZAP-S) is a direct and specific regulator of PRF in SARS-CoV-2 infected cells. ZAP-S overexpression strongly impairs frameshifting and viral replication. Using *in vitro* ensemble and single-molecule techniques, we further demonstrate that ZAP-S directly interacts with the SARS-CoV-2 RNA and ribosomes and interferes with the folding of the frameshift RNA. Together these data illuminate ZAP-S as *de novo* host-encoded specific inhibitor of SARS-CoV-2 frameshifting and expand our understanding of RNA-based gene regulation.

Introduction

The novel severe acute respiratory syndrome-related coronavirus (SARS-CoV-2), the causal agent of Coronavirus Disease 2019 (COVID-19), emerged rapidly to become a global threat to human health [270]. Global analyses of RNA- and protein-interaction networks have increased our understanding of SARS-CoV-2 viral replication in a short time [271, 272]. However, there is a lack of detailed mechanistic understanding of the interplay between RNA-protein complexes, which could inform the design of novel antivirals. Here, functionally important RNA elements of the viral genome represent ideal targets due to their evolutionary conservation. One of those well-conserved RNA elements is the programmed ribosomal frameshift site.

A hallmark of infections by the SARS-CoV-2 and many other viruses is the –1 programmed ribosomal frameshifting (–1PRF) event which allows translation of multiple proteins from the same transcript. Frameshifting increases the coding potential of the genomes and is often used to expand the variability of proteomes, adapt to changing environments, or ensure a defined stoichiometry of protein products [54, 77]. In coronaviruses, –1 frameshifting on the *1a/1b* gene is fundamental for efficient viral replication and transcription of the viral genome. In cells, efficiency of this frameshifting event varies between 20-40% [87, 273].

Programmed ribosomal frameshifting relies on the presence of a slippery heptameric sequence (in coronaviruses U UUA AAC) and an RNA secondary structure such as a pseudoknot **(Fig. 1A)**. Mutations in the slippery sequence and downstream RNA structure drastically impair frameshifting efficiency [85, 274].

Traditionally, efforts to understand the mechanism of –1PRF focused on *cis*–acting modulatory elements. Previous work in purified translation systems explained in unprecedented detail how ribosome pausing on the slippery codons may lead to a kinetic partitioning and favor movement of translating ribosomes to an alternative reading frame [87, 124]. It has been shown that –1PRF may occur during a late stage of the tRNA translocation step with the stimulatory element causing ribosomes to become trapped in an unusual conformation that is relieved by either the spontaneous unfolding of the blockade or a –1 slip on the mRNA [87, 124]. Recently, it is becoming clear that *cis*-acting elements are not the only determinants of frameshifting in cells and *trans*-acting viral and cellular factors as well as small molecules or oligonucleotides can alter frameshifting levels [96, 114, 275]. Despite this momentum, fundamental questions such as how pertinent RNA-binding factors are for frameshifting processes in general and how exactly these interactions alter the mechanical properties of RNA as well as the choice of the reading frame remain to be exploited.

Based on current knowledge, there would be at least three potential routes to modulate frameshifting by *trans*-acting factors. First, the binding of the factor can transform the downstream RNA element to a more stable roadblock, which was shown for cardiovirus 2A, poly-(C) binding protein and some small molecules such as the NCT-8 [96, 102, 114]. In these cases, the specific interaction of the factor with the nucleotides downstream of the slippery codons leads to an increase in frameshifting. Alternatively, eukaryotic release factors such as eRF1 alone or eRF1/3 recruited by Shiftless (SFL) to the HIV-1 frameshift site were shown to target stalled ribosomes [105, 107]. In this case, different from the first group of regulators the interaction of both SFL and release factors was not dependent on the identity of the frameshift RNA. Therefore, it remains to be solved how the frameshifting ribosome complexes would be recognized by these *trans*-acting factors. A third route could potentially work through remodeling or destabilization of the frameshifting RNA elements through direct interactions between the RNA and the *trans*-factor. However, so far there has been no cellular or viral factor reported to affect frameshifting efficiency (FE) through this route.

These prompted us to comprehensively identify and study direct interactions between the host cell proteome and the SARS-CoV-2 frameshifting RNA element. Firstly, to decipher interactors of the frameshifting RNA element, we employed an *in vitro* RNA-antisense capture and mass spectrometry-based screen [276]. Through this approach, we identified the short isoform of zinc-finger antiviral protein (ZAP-S, ZC3HAV1), as a prominent RNA interaction partner. We demonstrated that ZAP-S acts as a host-encoded inhibitor of SARS-CoV-2 1a/1b frameshifting *in vivo* and *in vitro*. Intriguingly, ZAP-S overexpression reduced the replication of SARS-CoV-2 by more than 90%, highlighting the importance of the protein in the viral life cycle. The effect of ZAP-S on SARS-CoV-2 frameshifting was specific, because barring the closely related SARS-CoV-1, other viral and cellular PRF levels were not affected by ZAP-S *in vivo*. Using a multidisciplinary approach, we further probed this effect and revealed important clues on molecular principles of frameshifting downregulation by ZAP-S. Amongst them, we show that ZAP-S can alter the physical properties of the PRF RNA, which brings a unique dimension to frameshift mechanisms. Our study highlights for the first time that the expression of the SARS coronavirus ORF1a/1b, can be directly and specifically modulated by a host-encoded RNA-binding protein during infection. These findings provide substantial new insights on PRF regulation and the interplay between SARS-CoV-2 replication and host defense, thereby paving the way for novel RNA-based therapeutic intervention strategies.

Results

SARS-CoV-2 PRF RNA capture identifies novel host interactors

To identify potential cellular RNA-binding proteins (RBPs) that interact with the –1PRF element of SARS-CoV-2, an *in vitro* synthesized RNA fragment corresponding to nucleotides 13456-13570 of the SARS-CoV-2 genome was incubated with lysates of SARS-CoV-2-infected and uninfected Calu-3 cells and naïve HEK293 cells **(Fig. 1B)** [276]. Calu-3 cells are lung epithelial cells that are commonly used to study CoV infection [277]. HEK293 cells are routinely used to study RNA-protein interactomes, therefore they represented an ideal system to assess possible cell-based variations [278]. To exclude any non-specific binders, we used an 80 nucleotides long non-structured RNA as a control. RNAs were captured by a biotinylated antisense DNA-oligo, and interacting proteins were identified by LC-MS/MS (liquid chromatography tandem mass spectrometry) analysis **(Fig. 1B, C)**.

In our SARS-CoV-2 frameshift RNA capture, more than 100 proteins were at least two-fold enriched. According to our GO term analysis, the majority (80%) of identified hits have been described as RNA-binding proteins **(Supplementary Fig. 1A)**. As for viral proteins, we saw an enrichment of the viral nucleocapsid protein (N) in infected lysates, which is a well-described RNA-binding protein [279]. In addition, 35% and 30% of the enriched RBPs were involved in splicing and ribosome biogenesis, respectively **(Supplementary Fig. 1A)**. Among those, 19 proteins were common to infected and uninfected Calu-3 cells, 18 hits were identified only in HEK293 cells, 15 were captured only in uninfected Calu-3 cells, and 40 were present only in infected Calu-3 cells **(Supplementary Fig. 1B)**. The core interactome of 9 proteins identified in all three cell systems encompasses well-described post-transcriptional regulators **(Fig. 1C, D, Supplementary Fig. 1C, Supplementary Table 1)**. Proteins recently identified in genome-wide interactome studies as direct RNA interaction partners for SARS-CoV-2 were selected for downstream functional characterization [272, 279-281]. Several of these have been shown to play a role in RNA processing, including splicing (such as HNRNPs F, H1, and H2), RNA trimming (POP1) and RNA surveillance (ZAP) [282-284]. Translational regulators included IGF2BP1, ELAVL1, DHX36, and SSB [285, 286]. ELAVL1 is a cofactor which ensures translational fidelity in the context of uORFs [287]. DHX36 is a multifunctional helicase and is involved in translation and innate immunity [288, 289]. G-rich RNA sequence binding protein, GRSF1, is implicated in mitochondrial translation [290]. Another

multifunctional protein that was identified in our screen was ZAP, which is an interferon-induced antiviral factor with two isoforms (ZAP-S and ZAP-L). Both isoforms of ZAP are implied in various RNA-related mechanisms, including RNA decay and translation [284, 291-294]. While the longer isoform of ZAP (ZAP-L) was reported to be mainly recruited to membrane-associated sites of viral replication [291, 295, 296], the shorter cytoplasmic form of ZAP (ZAP-S) has been identified as an immune-regulatory protein through its interaction with the 3′ untranslated region of interferon mRNAs [291]. Two additional hits were included in the downstream analysis based on their above fourfold enrichment only in infected Calu-3 lysates. These included the poly-(A) polymerase PAPD4, and GNL2 which has been implied in ribosome biogenesis [297, 298].

RNA interactors specifically inhibit SARS-CoV-2 frameshifting in cells

To explore the potential role of the RNA binders in SARS-CoV-2 frameshifting, we designed an *in vivo* fluorescence-based –1PRF assay. In this assay, the expression of the first ORF EGFP in the 0-frame would be constitutive, whereas the expression of the following ORF mCherry would depend on –1PRF occurring at the preceding SARS-CoV-2 1a/1b frameshifting fragment **(Fig. 2A)**. As controls, we used a construct lacking the –1PRF stimulatory sequence, and the mCherry gene is placed either in –1 or in-frame with respect to EGFP **(Fig. 2A, B)**. Frameshift efficiencies were calculated as the ratio of mCherry to EGFP in the test construct normalized to the in-frame control **(see also Materials and Methods)**. To study the effect of the *trans*-acting factors on SARS-CoV-2 frameshifting, cells were co-transfected with both the dual-fluorescence reporter plasmid and the plasmid encoding the putative *trans*-factor as an N-terminal ECFP fusion. This allowed gating of ECFP + cells, which express the trans-acting protein of interest **(Supplementary Fig. 2A)**. To benchmark the assay, a vector expressing only ECFP was used as a control to compensate for the spectral overlap between ECFP and EGFP. Using this fluorescence reporter system, the frameshifting efficiency (FE) of SARS-CoV-2 was measured to be ca. 35% in HEK293 **(Fig. 2C, D, Supplementary Fig. 2A, B, Supplementary Table 2)**, in agreement with the published FE for SARS-CoV-1 as well as the SARS-CoV-2 [273, 274]. In addition, vector expressing ECFP-SFL, a previously described inhibitor of –1PRF in SARS-CoV-2, was used as a positive control [272]. Among the selected RNA interactors, no change in FE was observed with GNL2, HNRNPF, IGF2BP1 or SSB which points to the fact that binding to the stimulatory RNA element is not sufficient for modulating PRF. Furthermore, control proteins that were not significantly enriched in the

interactome capture, such as SART, DDX3, PINX and ZFR, did not lead to significant changes in FE, corroborating the specificity of the flow-cytometry-based frameshifting assay **(Fig. 2C)**. Two hits, namely GRSF1 and PAPD4, led to a small but statistically significant increase in FE. Proteins with the strongest effect on FE were HNRNPH1, HNRNPH2 and ZAP-S, where frameshifting was substantially reduced by up to 50%. Despite equal expression levels **(Supplementary Table 2)**, the large isoform of ZAP (ZAP-L) reduced frameshifting levels to a much lower degree compared to HNRNPH1, HNRNPH2 and ZAP-S.

We also compared the relative mRNA expression levels of the selected RBPs in published RNA-seq datasets from infected Calu-3, Huh7.5.1 cells and COVID-19 patients **(Supplementary Fig. 1E)** [299, 300]. HNRNPH1 and HNRNPH2 expression levels did not change upon infection, whereas IMP3 and ZAP transcripts were enriched by more than 6-fold in patient samples [299]. We therefore decided to include IMP3 as a control RNA-binding protein for the downstream analysis due to its relatively low enrichment in the screen (log_2 enrichment 0.4-0.7) (**Fig. 1C**). Notably, among all the hits we analyzed, ZAP was the only factor that was also induced in infected Calu-3 and Huh7.5.1 cells. We also analyzed expression levels of these candidates by quantitative RT-PCR in SARS-CoV-2 infected Calu-3 cells compared to uninfected controls at 72 hours post-infection. As seen in the RNA-seq data, only ZAP showed a significant (ca. 20-fold) increase in mRNA levels upon infections [299, 300] **(Supplementary Fig. 1E, F)**. An increase in ZAP-S protein levels upon SARS-CoV-2 infection was also reported previously [301, 302].

Next, to test whether ZAP-S is functionally relevant during SARS-CoV-2 infection, Huh7 cells stably overexpressing ALFA-tagged ZAP-S were infected with SARS-CoV-2. In line with previous reports using RNAi, ZAP-S overexpression reduced viral replication after 24 hours by approximately 20-fold **(Fig. 2D, Supplementary Fig. 2D)** [279, 303]. We further tested whether the addition of interferons had a synergistic effect but observed no further enhancement of the effect of ZAP-S upon treatment with IFN-α2, INF-ß, IFN-ɣ, and IFN-ƛ1 **(Fig. 2D, Supplementary Fig. 2C, E).** In addition, we also measured the viral N protein levels via immunofluorescence, which is one of the early markers of SARS-CoV-2 infection. Levels of the N protein were also decreased upon ZAP overexpression **(Supplementary Fig. 2C, D and E)**. Taken together, our results showed that ZAP-S has the potential to restrict SARS CoV-2 replication in our cellular system, similar to published

results in Calu-3 cells [303]. Based on its strong induction upon infection, inhibition of viral frameshifting and antiviral function, we decided to focus on ZAP-S for further experiments.

To investigate the specificity of ZAP-S for the SARS-CoV-2 frameshift element, we tested whether the overexpression of ZAP-S affects –1PRF of other RNAs, e.g., different Coronaviruses (SARS-CoV-1, MERS-CoV, Bat-CoV-273, two additional human coronavirus HKU1 and OC43), Arboviruses (West Nile Virus (WNV), Japanese Encephalitis Virus (JEV), Chikungunya Virus (CHIKV)), and Human Immunodeficiency Virus-1 (HIV-1). Our analysis also included the embryonic gene PEG10, which represents an established example for –1PRF in humans [173]. Among the frameshift sites investigated, only the FE of SARS-CoV-1 was reduced significantly in the presence of ZAP-S (decrease by ca. 50%) **(Fig. 2E)**, likely due to the high degree of similarity between the SARS-CoV-1 and CoV-2 frameshift sites. This specificity is unlike the SHFL protein, which affects several PRF genes, including the cellular PEG10 [105, 304].

In order to understand if the inhibitory effect of ZAP-S on viral frameshifting is dependent on specific interactions with the SARS-CoV-2 frameshift element, we introduced sequential truncations within the predicted stem loops (SL) of the SARS-CoV-2 frameshift stimulatory pseudoknot (PK). We prepared a series of mutants – namely ΔSL2, ΔSL3 and ΔSL2+3 – which were deletions of the predicted SL2 region (nucleotides 13535-13542), SL3 region (nucleotides 13505-13532) and both SL2 and SL3 (nucleotides 13505-13542), respectively. Frameshifting was completely abolished in the ΔSL2 and ΔSL2+3 mutants, which is in line with minimal sequence requirements for frameshifting in other coronaviruses **(Fig. 2F)** [85, 110]. Due to the absence of PRF in ΔSL2 and ΔSL2+3, we were not able to evaluate the effect of ZAP-S with these mutants. With the ΔSL3 mutant, FE was severely reduced (to ~20%) and remained unaffected by the presence of ZAP-S. ZAP has been shown to bind CG dinucleotides [303]. Therefore, we tried to address four of these by compensatory mutants which would maintain the predicted base pairing. These compensatory C<-->G mutations led to an increase of the FE up to 60%, which might be due to stabilization of the pseudoknot or alternatively due to effects on alternative folds. Notably, the PRF-inhibitory effect of ZAP-S was no longer observed in this compensatory mutant **(Fig 2F)**. Taken together, ZAP-S seems to require an intact PK sequence or a particular RNA fold for its effect, since mutations or truncations in the RNA either decreased or completely abolished its effect.

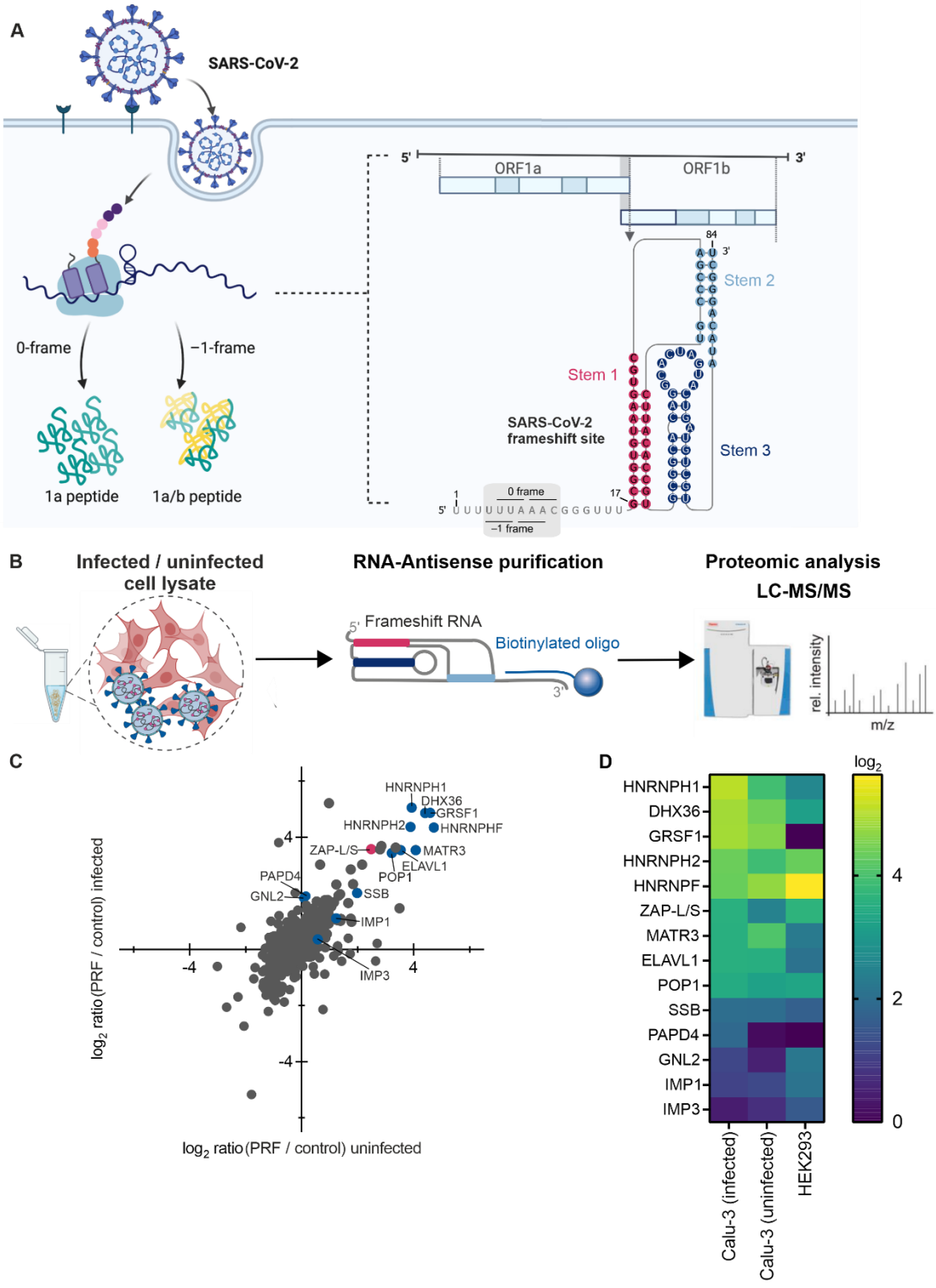


Fig. 1. ***In vitro*** **RNA-antisense purification-based discovery of protein interactors of the SARS-CoV-2 –1PRF element. (A)** Schematic representation of the relevant genomic segment of SARS-CoV-2 as well as the location of the –1PRF element. **(B)** Schematic of *in vitro* interactome capture of protein interactors of the SARS-CoV-2 –1PRF element. *In vitro* synthesized RNA fragment numbered 1-84 corresponding to nucleotides 13456 – 13570 of the SARS-CoV-2 genome, was incubated with lysates of naïve HEK293 cells as well as SARS-CoV-2-infected and uninfected Calu-3 cells. The –1PRF RNA was captured by a biotinylated antisense DNA oligo and isolated proteins were subjected to LC-MS/MS. **(C)** Representative

scatter plot of log_2-ratios comparing proteins captured in uninfected vs. SARS-CoV-2-infected Calu-3 cells. Core interactors common between uninfected and SARS-CoV-2-infected Calu-3 cells as well as uninfected HEK293 cells are highlighted in blue, ZAP is highlighted in pink. **(D)** Heatmap representing the enrichment (log_2) of core interactors. See also **Supplementary Fig. 1D**.

ZAP-S decreases SARS-CoV-2 frameshifting efficiency *in vitro*

We next focused on characterizing ZAP-S mediated regulation of frameshifting *in vitro* using the rabbit reticulocyte lysate (RRL) translation system and recombinant ZAP-S. **(Fig. 3A and Supplementary Fig. 2F)**. We employed reporter mRNAs containing nucleotides 12686-14190 of the SARS-CoV-2 genome to best mimic the native genomic context of viral frameshifting. Control RNAs exclusively producing either the 0-frame (nsp9-11) or –1-frame products (nsp9-11 + partial nsp12) were employed as size markers for the western blot **(Fig. 3B)**. In accordance with a previous study [273], SARS-CoV-2 FE was about 46% in the absence of ZAP-S. Upon titration of increasing amounts of ZAP-S, we observed a corresponding decrease in FE. At the highest concentration of ZAP-S (3 µM), FE was reduced from 46% to ~26% **(Fig. 3B, C)**. These results establish that ZAP-S acts on the native SARS-CoV-2 mRNA directly and that no cofactors are required for its action. To ensure that the observed effect was specific to ZAP-S and not mediated by non-specific RNA-protein interactions, we also tested IMP3, an RBP that we identified as a weak interactor with the RNA frameshifting element in our screen, and the SUMO-tag alone. Neither the addition of IMP3, nor the addition of SUMO alone led to a change in frameshifting levels **(Fig. 3C)**.

Several *trans*-acting factors including the cardiovirus 2A and SHFL were shown to bind to ribosomes and as well as frameshifting RNAs [99, 105]. Thus, to explore whether ZAP-S interacts with the translation machinery, we performed polysome profiling of the RRL translating the SARS-CoV-2 frameshift reporter mRNA in presence and absence of ZAP-S **(Fig. 3A)**. Both polysome profiles were similar, suggesting that ZAP-S does not significantly change bulk translation in RRL. In addition, ZAP-S was detected in the monosome (80S) as well as the polysome fractions; the latter represent the actively translating pool of ribosomes (**Fig. 3D)**. To confirm that the interactions of ZAP-S with ribosomal subunits and polysomes also occurs within cells, we conducted polysome profiling of HEK293 cells overexpressing ZAP-S **(Fig. 3A)**. Also in that case, ZAP-S was detected in ribosomal fractions, including polysomes. -In this experimental set-up, we could also detect endogenous ZAP-L in free RNA fractions and to a small extent in

ribosome fractions **(Fig. 3E)**. Similar polysome profiles were obtained with cells overexpressing SHFL, which as a known ribosome interactor acts as positive control [105]. We further confirmed that endogenous ZAP-S also associates with ribosomes in naïve Calu-3 cells via ribosome pelleting **(Fig. 3A, F)**. Next, we purified eukaryotic ribosomal sub-units to test whether ZAP-S directly interacts with the ribosomes *in vitro*. Accordingly, we observed that ZAP-S interacts with both ribosomal subunits with nanomolar affinity (40S - K_D=95±30 nM; 60S - K_D=691±390 nM) (**Supplementary Fig. 3J**). Collectively, these results indicate that ZAP-S associates with ribosomes directly. However, in the cells, it may also bind indirectly through its interactions with the SARS-CoV-2 mRNA.

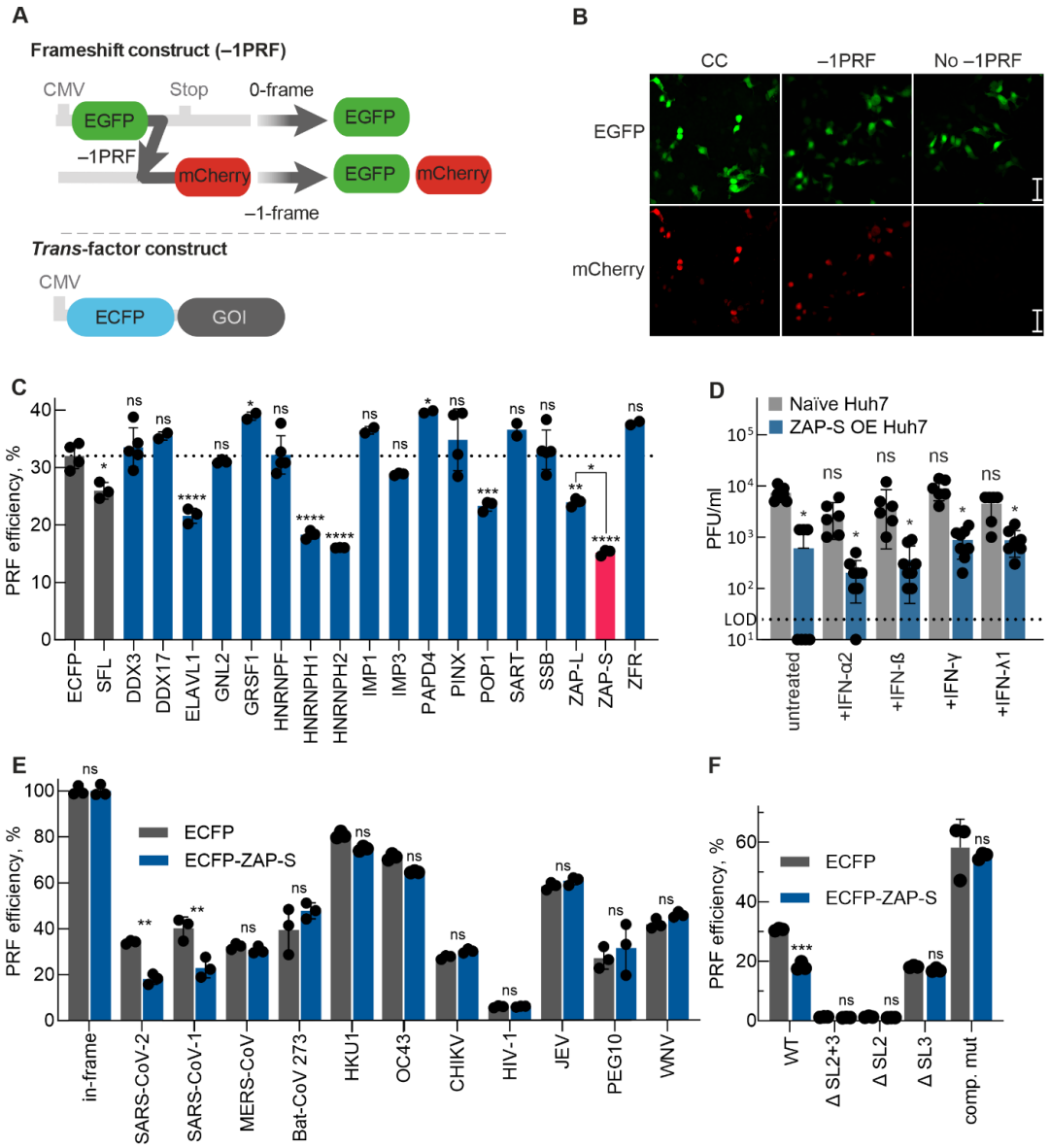


Fig. 2. A functional screen of SARS-CoV-2 –1PRF element interactors. (A) Schematic representation of the dual-fluorescence frameshift reporter construct. EGFP and mCherry are separated by a self-cleaving 2A

peptide as well as by a stop codon in-frame with EGFP. As a result, 0-frame translation would produce only EGFP, whereas –1PRF would produce both EGFP and mCherry. The ratio of mCherry to EGFP fluorescence is used to quantify the FE. The *trans*-factor construct is an N-terminal fusion of ECFP with the protein of interest to be analyzed. The control construct consists of ECFP alone. **(B)** Confocal microscopy images of cells transfected with the EGFP-mCherry control (CC- no –1PRF site included after EGFP and mCherry in-frame with EGFP), –1PRF, and no PRF (no –1PRF site and stop codon after EGFP) constructs. The size bar represents 50 µm. n = 1 independent experiment. **(C)** Comparison of relative FE of cells overexpressing *trans*-factors as ECFP fusion proteins. Data points represent the mean ± s.d. (n = 3 independent experiments). P values were calculated using an ordinary unpaired one-sided ANOVA comparing every condition to the ECFP control. ZAP-L and ZAP-S were separately compared to each other. * $P < 0.05$ – ** $P < 0.01$ – *** $P < 0.001$ – **** $P < 0.0001$. Exact P values: SFL – 0.03, DDX3 – 0.99, DDX17 – 0.72, ELAVL1 – < 0.0001, GNL2 – 0.99, GRSF – 0.03, HNRNPF – 0.99, HNRNPH1 – < 0.0001, HNRNPH2 – < 0.0001, IMP1 – 0.39, IMP3 – 0.68, PAPD4 – 0.01, PINX – 0.74, POP1 – 0.0005, SART – 0.36, SSB – 0.99, ZAP-L – 0.001, ZAP-S – < 0.0001, ZFR – 0.12 **(D)** Virus titers in the supernatant of infected naïve Huh7 or ZAP-S overexpressing Huh7 cells (ZAP-S OE) at 24 hours post infection. Treatment with IFN-ɣ (500 U/ml), IFN-β (500 U/ml), or IFN-ƛ1 (5 ng/ml) was done one hour before infection. Boxes show mean values ± s.d. (n = 4 independent experiments). The dotted line represents the limit of detection (LOD). P values were calculated using an ordinary unpaired one-sided ANOVA comparing every condition to untreated naïve infected Huh7 cells. Exact P values: untreated + ZAP-S – 0.01, INF-α2+ ZAP-S – 0.04, INF-ß + ZAP-S – 0.49, INF-γ + ZAP-S – 0.049, INF-λ + ZAP-S – 0.049 **(E)** *In vivo* dual-fluorescence of additional –1PRF RNAs in HEK293 cells in the presence and absence of ZAP-S. SARS-CoV-1 – severe acute respiratory syndrome-related coronavirus 1, MERS-CoV – Middle East respiratory syndrome-related coronavirus, Bat-CoV-273 – Bat Coronavirus 273, HKU1 – Human coronavirus *HKU1*, OC43 – Human Coronavirus *OC43*, CHIKV – Chikungunya Virus, HIV-1 – Human Immunodeficiency Virus 1, JEV – Japanese Encephalitis Virus, PEG10 – paternally expressed 10, WNV – West Nile Virus. Data points represent the mean ± s.d. (n = 3 independent experiments). P values were calculated using an ordinary unpaired one-sided ANOVA comparing every condition to the ECFP control. * $P < 0.05$ – ** $P < 0.01$. Exact P values: SARS-CoV-2 – 0.001, SARS-CoV-1 – 0.001. **(F)** *In vivo* dual-fluorescence of mutants of SARS-CoV-2 –1PRF RNA in HEK293 cells in the presence and absence of ZAP-S. Datapoints represent the mean ± s.d. (n = 3 independent experiments). P values were calculated using an unpaired one-sided ANOVA comparing values of the ECFP control. * $P < 0.05$. Exact P values: WT – 0.0003. See also **Supplementary Table 2** as well as **Fig. 4** for schematics of the mutants used here.

ZAP-S directly interacts with the SARS-CoV-2 frameshift motif

In order to further dissect the interplay between the SARS-CoV-2 frameshifting RNA and ZAP-S, we performed RNA-protein binding assays using the highly sensitive microscale thermophoresis assay (MST) **(Fig. 4A, B)**. The wild type (WT) PK, derived from nucleotides 13456-13570 of the SARS-CoV-2 genome, was *in vitro* transcribed and

Cy5-labeled at the 3' end. We also tested the stem-loop truncation variants we designed earlier and stem-loop mutants of the stimulatory pseudoknot.

For the wild type SARS-CoV-2 PK, we observed that ZAP-S interaction occurs with a high affinity (K_D = 110±9 nM) **(Fig. 4C)** indicating that ZAP-S is a direct interaction partner of the frameshift signal. Next, with the ΔSL2 mutant, we detected a weak interaction with ZAP-S which was characterized by a K_D of 672±164 nM **(Fig. 4D)**. In contrast, deletion of the SL3 region (ΔSL3) only marginally reduced the affinity of ZAP binding (K_D=175±64 nM) **(Fig. 4E)**. On the other hand, deletion of both SL2 and SL3 (ΔSL2+3), which is predicted to fold into a short stem-loop (SL1) completely abolished ZAP-S binding **(Fig. 4F)**. In contrast, ZAP-S binds to the compensatory mutant, with an affinity close to WT RNA (K_D =128±29 nM) **(Fig. 4G)**. A negative control RNA with the same nucleotide composition as the WT PRF site but a disrupted pseudoknot RNA fold did not bind ZAP-S **(Fig. 4H)**. Furthermore, we tested the binding of two control proteins, IMP3 and SUMO, to the SARS-CoV-2 frameshift motif. Compared to ZAP-S, IMP3 showed an almost 7-fold lower affinity to the RNA (K_D= 806±252 nM). No interaction between SUMO and the frameshift element was detected **(Supplementary Fig. 3G)**. Based on these data, we hypothesized that ZAP-S has multiple binding sites in the putative SL2 and SL3 regions of the pseudoknot. We then carried out electrophoretic mobility shift assays (EMSAs), which confirmed multiple binding events on the WT pseudoknot RNA, but none with the RNA variant lacking the SL2 and SL3 regions (ΔSL2+3) **(Supplementary Fig. 3H, I)**. To further analyze potential changes in the SARS-CoV-2 RNA structure in the presence of ZAP-S we carried out dimethyl sulfate (DMS) mutational profiling with sequencing (DMS-MaPseq)**(Supplementary Fig. 4)**. In the absence of ZAP-S, DMS reactivities were consistent with a significant proportion of the RNA folding into a pseudoknot conformation **(Supplementary Fig. 4)**. In the presence of ZAP-S, we witnessed decreases in DMS reactivities in both the loop regions of SL2 and SL3, as well as increases in reactivities in the stems of SL1 and SL2. Overall, our MST and DMS-MaPseq analysis suggest SL2 and SL3 as the main binding sites for ZAP-S.

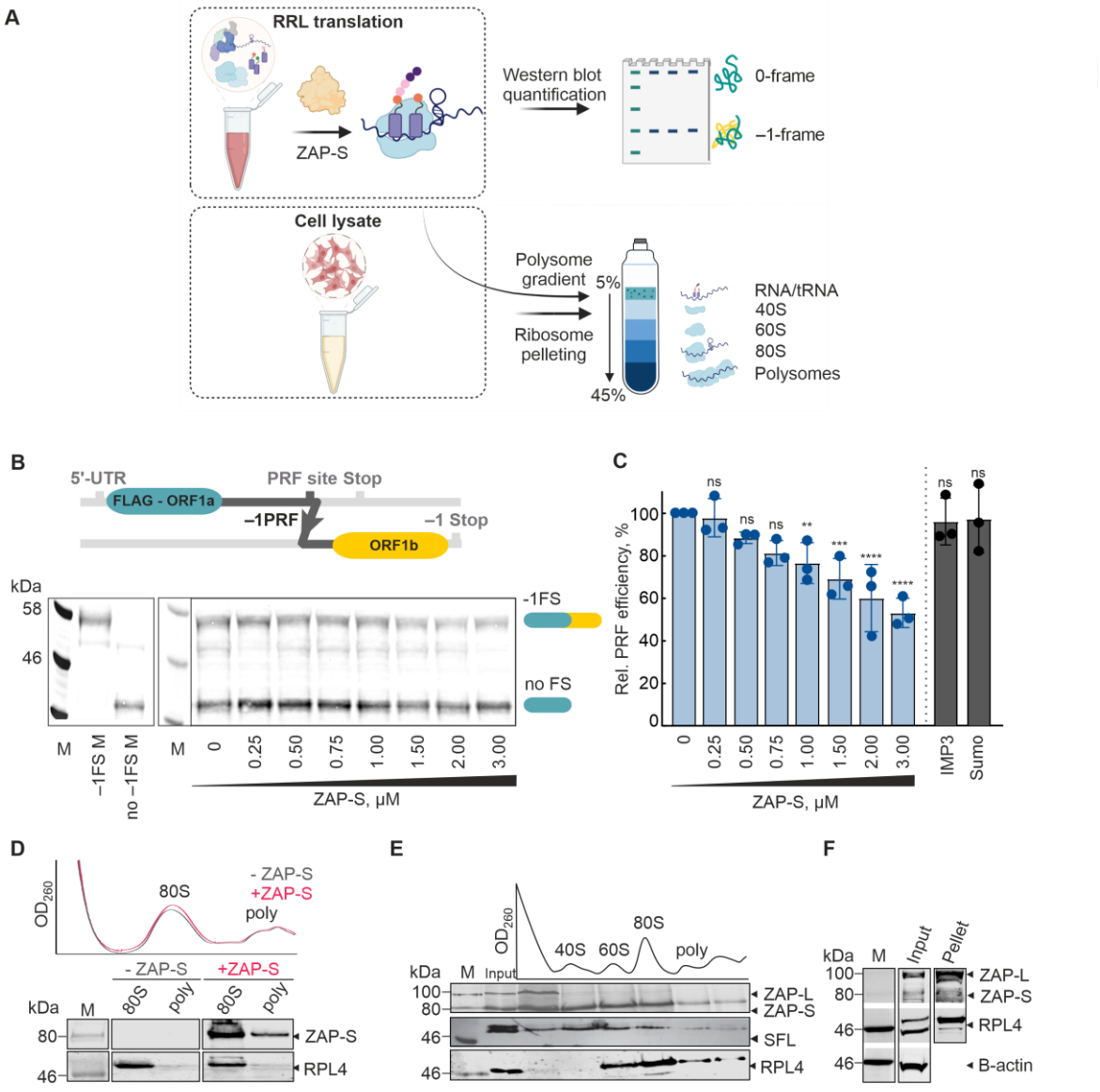


Fig. 3. Effect of zinc-finger antiviral protein (ZAP) on 1a/1b –1 frameshifting *in vitro*. (A) The strategy of the *in vitro* translation assay using rabbit reticulocyte lysate (RRL) and the experimental workflow to study ribosome association of ZAP-S. **(B)** Schematics of the N-terminal FLAG-tagged frameshifting reporter consisting of the nucleotides 12686-14190 (~1.5 kb) of the SARS-CoV-2 genome. RNAs were translated in RRL in the presence of increasing concentrations of ZAP-S ranging from 0 to 3 µM. FLAG-tagged peptides generated by ribosomes that do not frameshift (no –1PRF) or that enter the −1 reading frame (−1PRF) were identified via western blotting using anti-DDDDK antibody. FE was calculated as previously described [11], by the formula: Intensity (–1-frame)/ (Intensity (–1-frame) + Intensity (0-frame)). Size markers - M (Marker), –1PRF M (–1-frame marker), and no –1PRF M (0-frame marker). n = 3 independent experiments. **(C)** Changes in FE observed in the presence of ZAP-S from (B) (normalized to 0 µM ZAP as shown in B). P values were calculated using an ordinary unpaired one-sided ANOVA comparing every concentration to the no ZAP control. * P < 0.05 – ** P < 0.01 – *** P < 0.001 – **** P < 0.0001. Exact P values: 0.25 µM – 0.82, 0.50 µM – 0.26, 0.75 µM – 0.06, 1.00 µM – 0.009, 1.50 µM – 0.0002, 2.00 µM – < 0.0001, 3.00 µM – < 0.0001. See also **Supplementary Fig. 2 and Supplementary Table 3**. **(D)** Polysome profiling analysis of ZAP-S in

RRL. RRL translating the FLAG-tagged SARS-CoV-2 frameshifting reporter was subjected to 5-45% sucrose gradient ultracentrifugation, and subsequently fractionated. Levels of RPL4, as well as ZAP in each fraction, were analyzed by western blotting using anti-RPL4 and anti-ZC3HAV1 (ZAP) antibodies. n = 2 independent experiments. **(E)** Ribosome pelleting of untreated Calu-3 cells. Naïve Calu-3 cells were lysed and loaded onto sucrose cushions. Levels of RPL4, ZAP, and β-actin in the pellets were analyzed by western blotting using anti-RPL4, anti-ZC3HAV1 (ZAP) and anti-β-actin antibodies. n = 3 independent experiments. **(F)** Polysome profiling analysis of ZAP-S in cells. HEK293 cells transiently expressing ZAP-S were lysed, subjected to 5-45% sucrose gradient ultracentrifugation, and subsequently fractionated. Levels of ribosomal proteins, ZAP as well as SHFL in each fraction, were analyzed by western blotting using anti-RPL4, anti-ZC3HAV1 (ZAP) and anti-RYDEN (SHFL) antibodies. n = 3 independent experiments.

ZAP-S prevents the refolding of the stimulatory RNA

Since ZAP-S directly interacted with the frameshift element, we next tested whether this binding alters the RNA structure and/or mechanical stability of the RNA using single-molecule optical-tweezers assays. To this end, an RNA containing the 68 nucleotides long wild type SARS-CoV-2 pseudoknot (nucleotides 13475 – 13542 of SARS-CoV-2 genome) was hybridized to DNA handles and immobilized on polystyrene beads. We employed exclusively the sequence corresponding to the putative pseudoknot to preclude the formation of alternative conformers [110, 112, 305, 306]. We used the force-ramp method to probe the forces required for (un)folding of the RNA in the presence and absence of ZAP-S. Briefly, the frameshift RNA was gradually stretched at a constant rate, and then the applied force was released while recording the molecular end-to-end extension distances. This allows the RNA molecule to transition between folded and unfolded states, and sudden changes in measured force-distance trajectories indicate transitions between various RNA conformations **(Fig. 5A)**. By mathematically fitting each force-distance trajectory, we can obtain information on the physical properties of the RNA such as the change in the contour length (number of nucleotides unfolded) or the force required for (un)folding (**Methods**). With the SARS-CoV-2 putative pseudoknot, in the absence of ZAP-S, we mainly observed a single-step unfolding event leading to a contour length (L_C) change of 35.4±3.0 nm **(Supplementary Fig. 6, Supplementary Table 1)**, which agreed with the expected value for the full-length pseudoknot reported previously Fig. **5B)** [112, 306]. Moreover, the majority (80%) of RNA molecules unfolded at forces (F_U) of 15 -20 pN (**Supplementary Fig. 5**). For the remaining traces, we observed two consecutive unfolding events with an intermediate contour length change of 17.1±3.5 nm **(Fig. 5D)** likely corresponding to the sequential unfolding of the pseudoknot

structure. By decreasing the force, the RNA refolded in two steps, both at about 11 pN **(Fig. 5D, Supplementary Fig. 6, Supplementary Table 1)**. Such a hysteresis during refolding is commonly reported with pseudoknots and other highly structured RNAs [112, 305].

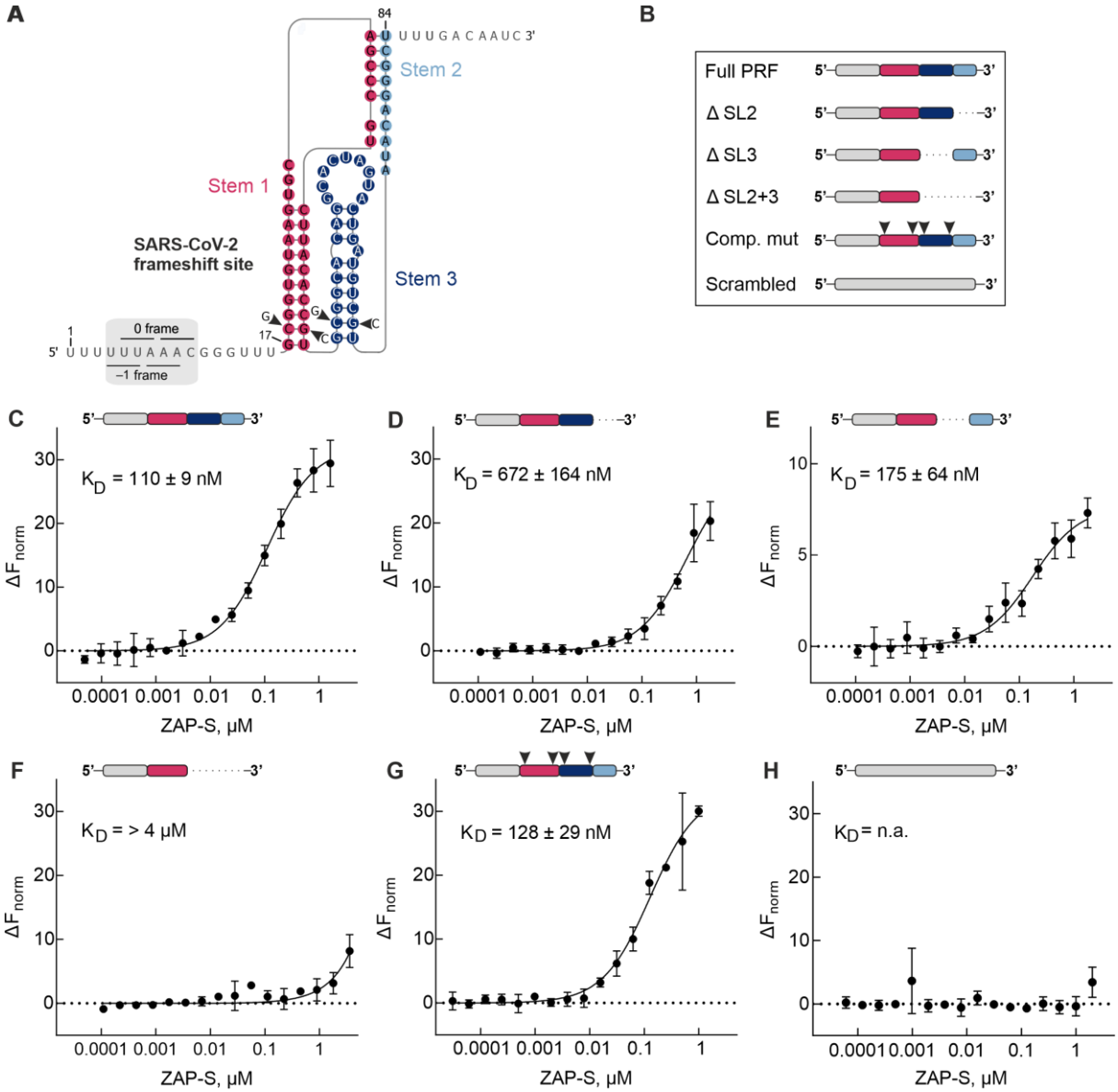


Fig. 4. ***In vitro*** **characterization of ZAP-S interaction with SARS-CoV-2 –1 PRF RNA. (A)** Proposed structure of the PRF element of SARS-CoV-2. Nucleotide substitutions in the compensatory mutant are indicated (arrowheads). **(B)** Schematic representations of the RNAs studied. **(C-H)** Microscale thermophoresis assay to monitor ZAP-S binding to **(C)** Full PRF, **(D)** ΔSL2 mutant, **(E)** ΔSL3 mutant, **(F)** ΔSL2+3 mutant, **(G)** compensatory mutant, **(H)** scrambled mutant. Unlabeled protein (40 pM to 2 μM) was

titrated against 3' pCp-Cy5 labeled RNA (5 nM) and thermophoresis was recorded at 25°C with 5% LED intensity and medium MST power. Change in fluorescence (ΔF_{norm}) was measured at MST on-time of 2.5 s. Data were analyzed for ΔF_{norm} using standard functions of MO. Affinity Analysis software and data was plotted and K_D was determined using Graphpad Prism 9.2.0. Data represent mean ± s.d. of three measurements (n = 3). For the related thermophoretic traces, see also **Supplementary Fig. 4A-F.** For the related DNA sequences of the mutants, see also **Supplementary Table 2.**

When we performed the measurements in the presence of ZAP-S RNA unfolding trajectories remained almost/mostly unaffected, suggesting that the interaction neither stabilizes nor destabilizes the RNA structure (**Fig. 5D, Supplementary Fig. 5 and 6, Supplementary Table 1**). On the other hand, strikingly, refolding of the RNA into its native fold was impaired with less or no detectable transitions into the folded state **(Fig. 5D, Supplementary Fig. 5 and 6, Supplementary Table 1)**.

To better characterize the sequence or structural constraints that are important for the ZAP-S mediated effect, we also employed the same set of truncation mutants of the wild type SARS-CoV-2 PK used earlier **(Fig. 4B, Fig. 5C and E-H)**. Truncation of SL2 region (ΔSL2) is expected to prevent the formation of the PK, and instead RNA would fold into two consecutive stem loops **(Fig. 5C, E).** With ΔSL2 both the change in L_C (30.8±3.1 nm) and F_U (peak 1 - 9.3±1.3 pN, peak 2 - 13.8±0.8 pN) were lower compared to the wild type PK, and RNA was able to refold back readily, which was in line with the formation of predicted stem-loops. In the presence of ZAP-S with the ΔSL2 variant, force of unfolding was unchanged, but three distinct populations of refolding were observed based on the change in the contour length (**Supplementary Fig. 5 and 6, Supplementary Table 1)**. In one population no refolding was seen (0.2±0.3 nm), the second one showed similar step sizes during (un)folding (25.6±2.8 nm), and the third one represented a partially refolded state, which was likely a simple hairpin based on the lower contour length change (15.9±3.1 nm) (**Supplementary Fig. 6, Supplementary Table 1**). In the ΔSL3 RNA variant, (**Fig. 5 C, F**), the RNA was predicted to fold into a shorter PK. In agreement with this prediction, we measured higher forces of unfolding (17.4±1.3 pN) and hysteresis during refolding, yet the change in contour length (21.0±1.2 nm) was lower than the wild type PK (**Supplementary Fig. 5 and 6, Supplementary Table 1)**. In the presence of ZAP-S, no refolding was observed in about 20% of ΔSL3 curves, and we observed a significant decrease in the refolding work **(Fig. 5F, Supplementary Table 1)**. The ΔSL2+3 variant is predicted to form the simple hairpin (SL1). Our data also confirmed the presence of a single

stem-loop (**Fig. 5G, Supplementary Fig. 5 and 6, Supplementary Table 1)**, with the contour length value of (16.4±2.8 nm). Here, only about 10% of traces did not refold in the presence of ZAP-S **(Fig. 5G)**. Aside from that, FU was slightly shifted to lower values although our MST results clearly showed no binding of ZAP-S to this RNA variant (**Fig. 4F, Supplementary Fig. 5**). We also tested the effect of non-specific interactions using the control RBP IMP3, and we observed a similar small shift in the FU. Therefore, we conclude that this effect is due to non-specific interactions and/or molecular crowding (**Fig. 5I**). Finally, with the compensatory mutant (comp. mut.), which has a stack of 4Gs at the SL1 and SL3, unfolding forces were slightly higher than with the WT pseudoknot (18.9±5.5 pN). Nevertheless, the contour length change matched with the expected PK structure (36.3±1.7 nm) (**Fig. 5H, Supplementary Fig. 5 and 6, Supplementary Table 1**). While we cannot exclude that the compensatory mutant forms an alternative structure to the wild type PK, we hypothesize that this stabilization might be caused by the stacking interactions between G stretches at the base of the stems. Interestingly, force-extension behavior of this alternative PK was only minimally affected by ZAP-S binding (**Fig. 5H**).

To further compare the effect of ZAP-S on SARS-CoV-2 RNA variants, we calculated the work performed during refolding of the RNAs in the presence and absence of ZAP-S (**Fig. 5J, Supplemental Table 1**). Since work is calculated as a numerical integration of FD curves (**Methods**), employing of the refolding work enabled us to account for the ZAP-S effect on both refolding force as well as the total contour length change in a single value, thus allowing a better comparison among different samples. In the wild type PK work performed during refolding in the presence of ZAP-S was negligible, and the majority of traces (more than 60%) do not show any detectable refolding. Since the other RNAs differed in their lengths and other physical properties, we normalized the refolding work performed on each RNA in the presence of ZAP-S to work performed in the absence of ZAP-S. This allowed a non-biased comparison of the effect of the *trans*-acting factor. No significant difference in work was detected with PK in the presence of IMP3 control, or ΔSL2+3 and comp. mut. in the presence of ZAP-S (**Fig. 5K**). Conversely, in ΔSL2 and ΔSL3 RNA variants the refolding work was still affected by ZAP-S, albeit to a lesser degree when compared to wild type PK. Overall, we were able to quantify the effect of ZAP-S on refolding of the pseudoknot RNA and we suggest that SL2 and SL3 are crucial for the function of ZAP-S.

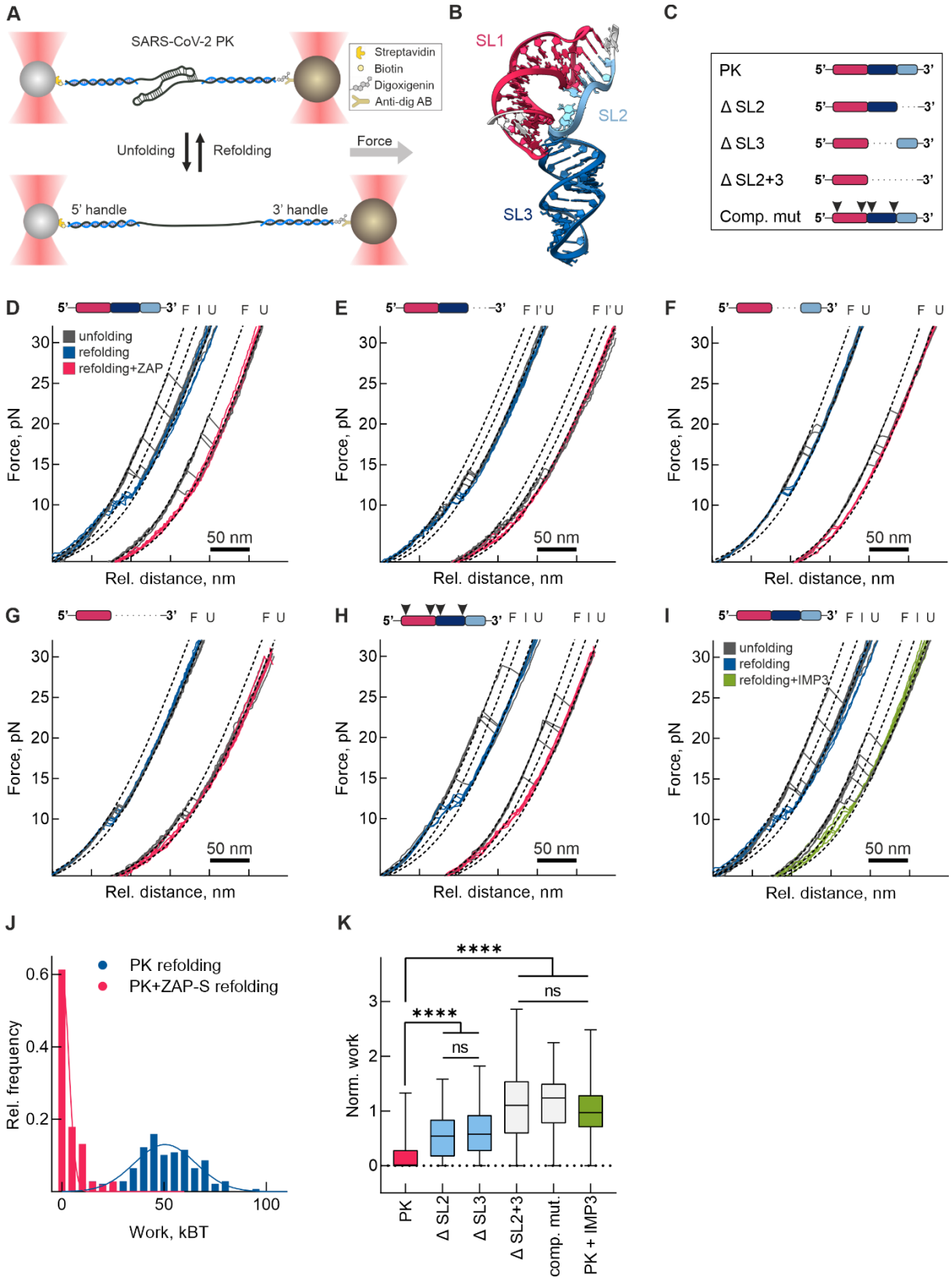


Fig. 5. Single molecule characterization of mechanical properties of SARS-CoV-2 PRF RNA in the presence of ZAP-S. (A) Schematic illustrating optical tweezers experiments. RNA was hybridized to single-

stranded DNA handles flanking the SARS-CoV-2 frameshift site and conjugated to functionalized beads. A focused laser beam was used to exert pulling force from one end of the molecule. The force was gradually increased until the RNA was fully unfolded (bottom). **(B)** 3D structure of SARS-CoV-2 pseudoknot RNA (PK) derived from Zhang *et al.* 2021 and colored according to the scheme used in **Fig. 4**. **(C)** Schematic representations of the RNAs studied. **(D - I)** Example unfolding and refolding traces of PK in the presence or absence of ZAP-S, "F" denotes the folded state, "I" the intermediate, and "U" the fully unfolded state, (**D**) PK (N=273 FD curves from 24 molecules no ZAP-S, N=219 FD curves from 24 molecules +ZAP-S samples), **(E)** ΔSL2 mutant (N=146 FD curves from 8 molecules no ZAP-S, N=122 FD curves from 8 molecules +ZAP-S samples), **(F)** ΔSL3 mutant (N=127 FD curves from 12 molecules no ZAP-S, N=163 FD curves from 11 molecules +ZAP-S samples), **(G)** ΔSL2+3 mutant (N=216 FD curves from 8 molecules no ZAP-S, N=196 FD curves from 11 molecules +ZAP-S samples), **(H)** compensatory mutant (N=158 FD curves from 12 molecules no ZAP-S, N=169 FD curves from 16 molecules +ZAP-S samples), **(I)** PK in absence (blue) and presence (green) of IMP3 (N=273 FD curves from 24 molecules no ZAP-S, N=226 FD curves from 20 molecules +ZAP-S samples). **(J)** Distribution of refolding work in presence (pink) and absence (blue) of ZAP-S. **(K)** Normalized refolding work in the presence of ZAP-S or IMP3. Data points represent the mean ± s.d. (box) and min and max values (whiskers). P values were calculated using an ordinary unpaired one-sided ANOVA followed by Dunnett's multiple comparisons test. * $P < 0.05$ – **** $P < 0.00001$. See also **Supplementary Fig. 5, 6 and Supplementary Table 1**.

Discussion

Programmed ribosomal frameshifting (–1PRF) is essential for coronavirus replication. In this study, we explored whether *trans*-acting host or viral factors can modulate SARS-CoV-2 –1PRF. We discovered that the short isoform of the interferon-induced zinc-finger antiviral protein ZAP-S can strongly impair SARS-CoV-2 frameshifting and decrease viral replication. ZAP-S was also one of the prominent common hits in genome-wide screens for proteins that interacted with SARS-CoV-2 RNA [272, 279-281].

Similar to previously reported *trans*-acting protein regulators of frameshifting (such as cardiovirus 2A and SHFL), we show that ZAP-S interacts with the translation machinery, suggesting that this might be a common feature of PRF modulators. Yet, unlike the cardiovirus 2A or cellular poly(C)-binding protein, ZAP-S does not mediate formation of a more stable mRNA roadblock to induce frameshifting [96, 102]. Rather, ZAP-S inhibits coronaviral frameshifting through recognition of a specific RNA motif. This sequence preference is not common. For example, SHFL interacts with stalled ribosomes and recruits release factors to terminate translation irrespective of the type of frameshift RNA [105]. In contrast, ZAP-S most likely interferes with the refolding of the frameshift RNA and an intact SL3 seems to be crucial for this effect. In addition, ZAP-S shows sequence preference

for SARS-CoV-2 SL2 and SL3. ZAP-S seems to have multiple binding modes and may bind to folding intermediate structures, which includes SL3, which we showed was essential for PRF inhibition. ZAP-S binding may thus slow down the folding rate of the native pseudoknot structure. Interestingly, SL3 is identical in SARS-CoV-1 and –2 frameshift motifs but shows a higher degree of sequence diversity in other coronaviruses. Furthermore, the compensatory mutant with a stretch of four Gs at the base of SL1 and SL3 was the most effective in stimulating PRF. The high FE of this mutant can be explained by the thermodynamic stability of the first 3-4 base pairs of the stem loops, near the mRNA entry channel of the ribosome [97]. Strikingly, the *trans*-acting factor ZAP-S showed no strong effect on this RNA variant, even though it interacts with the RNA element in the steady state. Here the effect of ZAP-S might not be prominent, either because binding site or structure is somehow altered due to the mutations or due to faster refolding kinetics of the PRF stimulatory element. This supports the notion that binding of ZAP-S is a prerequisite but not sufficient for its modulatory effect. Furthermore, it may explain why not all binders identified in our screen or in other studies are affecting frameshifting levels.

Ultimately, based on our findings, we propose the following model for the inhibition of –1PRF by ZAP-S **(Fig. 6)**. ZAP-S binding to the frameshift RNA alters the stimulatory RNA structure and reduces the chance of elongating ribosomes to encounter the stimulatory pseudoknot. Without this stimulatory structure, the elongation pause during the next round of translation would be too short for codon-anti-codon interactions to be established in the –1-frame. Thus, ZAP-S would likely allow translation to proceed and terminate at the 0-frame UAA stop codon found immediately downstream of the slippery sequence. The resulting decrease in the amounts of the 1a/1b polypeptides may lead to a reduction in the levels of the viral RNA-dependent RNA polymerase (RdRP) from the –1-frame.

In addition to the direct interaction with the frameshifting RNA element, ZAP-S also associates with the ribosomes, although how it interacts, direct or indirectly, or whether this interaction is functionally relevant awaits further investigation. We envision ZAP-S binding to the native PK and inhibiting its interactions with the ribosome as one possibility. Overall, these findings establish ZAP-S as unique cellular factor, which has a direct role in modulating SARS-coronavirus frameshifting. In accordance with previously published results, we demonstrate that overexpression of ZAP-S reduces the replication of SARS-CoV-2 [279, 303]. Further studies are required to deconvolute the multivalent effects of ZAP-S on immunity, viral replication and gene expression [291, 302, 307-310]. Given the plethora of

mechanisms by which *trans*-regulators of PRF can act, it is conceivable that viral- and host-encoded *trans*-factors follow a multitude of routes to impact frameshift paradigms. Taken together, our study establishes ZAP-S as a novel regulator of SARS-CoV-2 frameshifting and determines one (potential) mechanism by which ZAP-S mediates a SARS-CoV-2 antiviral response.

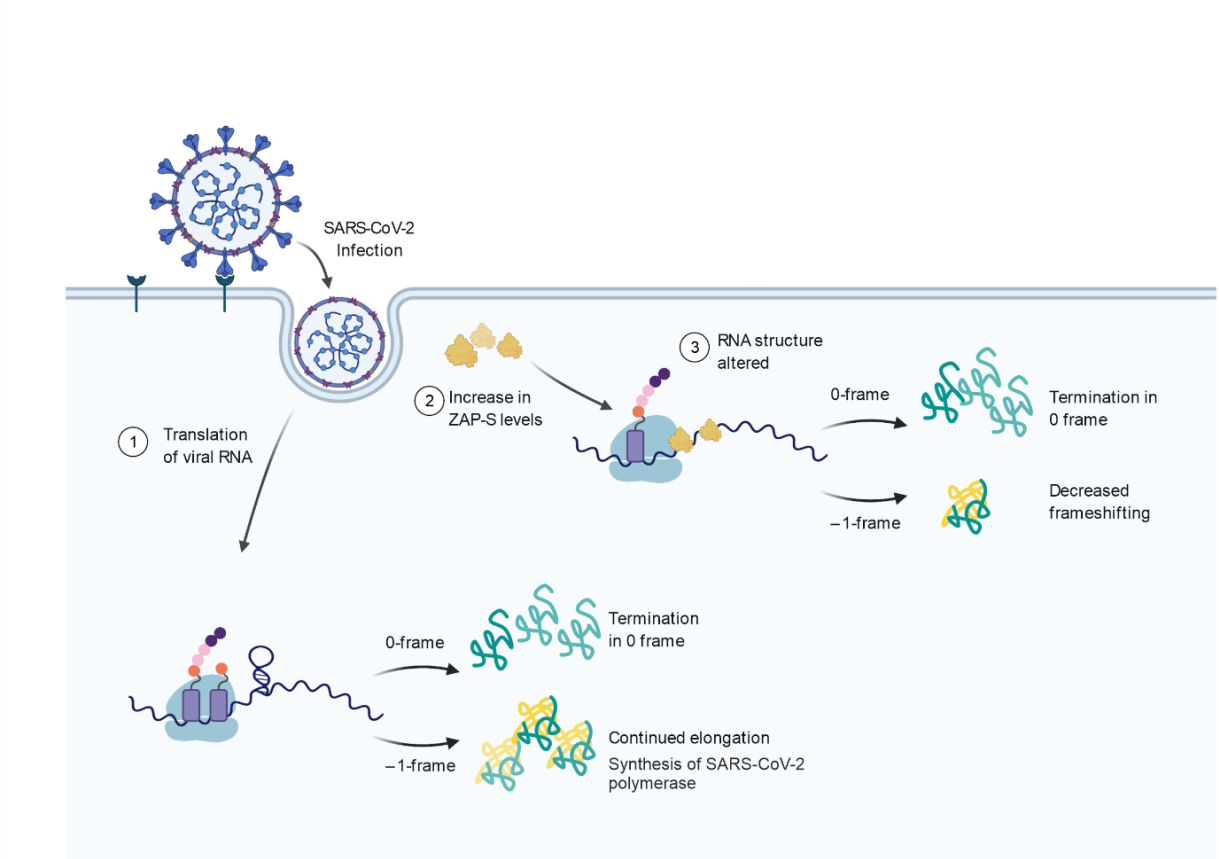


Fig. 6. Model for ZAP-S mediated inhibition of SARS-CoV-2 frameshifting. (1) Upon infection, the viral RNA is translated by the cellular machinery, and 40% of translation events yield the 1a/1b polyprotein through –1PRF. **(2)** Infection also leads to the induction of antiviral factors including ZAP-S. **(3)** ZAP-S binding to the frameshift RNA alters RNA refolding and thereby reduces the chance of elongating ribosomes to encounter the stimulatory structure. Thus, the elongation pause is too short for codon-anti-codon interactions to be established in the –1-frame and ZAP-S allows translation to proceed without a strong roadblock effect. This leads to termination at the canonical 0-frame UAA stop codon found just downstream of the slippery sequence. The resulting decrease in the amounts of the 1a/1b polypeptides reduces the levels of the viral RNA-dependent RNA polymerase (RdRP) from the –1-frame.

Materials and methods

RNA affinity pulldown mass spectrometry

RNA antisense purification was performed according to a protocol based on [18]. Briefly, $6*10^7$ HEK293 cells per condition were lysed in a buffer containing 20 mM Tris/HCl pH 7.5, 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 0.5 % Igepal CA630 (Sigma-Aldrich), 1× cOmplete™ Protease Inhibitor Cocktail (Roche), 40 U/ml RNase inhibitor (Molox). The cleared lysate was incubated with *in vitro* transcribed RNA corresponding to the SARS-CoV-2 –1PRF site, which was immobilized on streptavidin hydrophilic magnetic beads (NEB) by biotin-streptavidin interaction. After three washes with binding buffer (50 mM HEPES/KOH pH 7.5, 100 mM NaCl, 10 mM MgCl2) and two washes with wash buffer (50 mM HEPES/KOH pH 7.5, 250 mM NaCl, 10 mM MgCl2), bound proteins were eluted by boiling the sample in 1× NuPAGE LDS sample buffer (Thermo Fisher Scientific) supplemented with 40 mM DTT. For infected as well as uninfected Calu-3 cells the procedure was performed similarly. In order to inactivate the virus, the lysis buffer contained Triton X-100 and inactivation was confirmed by plaque assays.

For LC-MS/MS, the eluted proteins were alkylated using iodoacetamide followed by acetone precipitation. In solution digests were performed in 100 mM ammonium bicarbonate and 6 M urea using Lys-C and after reducing the urea concentration to 4 M with trypsin. Peptides were desalted using C18 stage tips and lyophilized. LC-MS/MS was performed at the RVZ Proteomics Facility (Würzburg, Germany) and analyzed as described previously [311]. Gene ontology (GO) term analysis was performed with Panther [312]. The list of all identified proteins is given in the **Source Data File**.

Plasmid construction

To generate dual-fluorescence reporter constructs frameshift sites of SARS-CoV-1, SARS-CoV-2, MERS-CoV, BtCoV 273, Human coronavirus *HKU1*, Human Coronavirus *OC43*, HIV-1, JEV, PEG10, WNV were placed between the coding sequence of EGFP and mCherry (parental construct was a gift from Andrea Musacchio (Addgene plasmid # 87803 [313]) by site-directed mutagenesis or golden gate assembly in a way that EGFP would be produced in 0-frame and mCherry in –1-frame. EGFP and mCherry were separated by StopGo [314] signals as well as an alpha-helical linker [315]. A construct with no PRF insert and mCherry in-frame with EGFP served as a 100% translation control and was used to

normalize EGFP and mCherry intensities. Mutants of the frameshift site of SARS-CoV-2 in the dual fluorescence as described in **Fig. 4A** and **Supplementary Table 2** were generated by golden gate assembly.

To generate screening vectors, protein-coding sequences of DD3X (NM_001193416.3), DDX17 (NM_001098504.2), DDX36 (NM_020865.3), ELAVL1 (NM_001419.3), GNL2 (NM_013285.3), GRSF1 (NM_001098477.2), HNRNPF (NM_001098204.2), HNRNPH1 (NM_001364255.2), HNRNPH2 (NM_001032393.2), IGF2BP1 (IMP1) (NM_006546.4), IGF2BP3 (IMP3) (amplified from a vector kindly provided by Dr. Andreas Schlundt), MATR3 iso 2 (NM_018834.6), MMTAG2 (NM_024319.4), NAF1 (NM_138386.3), NHP2 (NM_017838.4), PAPD4 (NM_001114393.3), PINX1 (NM_001284356.2), POP1 (NM_001145860.2), RAP11B (NM_004218.4), RSL1D1 (NM_015659.3), SART1 (NM_005146.5), SHFL (NM_018381.4), SSB (NM_001294145.2), SURF6 (NM_001278942.2), TFRC (NM_003234.4), ZC3HAV1 (ZAP) (NM_024625.4), ZFR (NM_016107.5), and ZNF346 (NM_012279.4) were placed in frame with the coding sequence for ECFP in pFlp-Bac-to-Mam (gift from Dr. Joop van den Heuvel, HZI, Braunschweig, Germany [316]) via Gibson Assembly.

Golden Gate compatible vectors for heterologous overexpression in *E. coli*, *in vitro* translation in RRL, and lentivirus production, were generated by Golden Gate or Gibson Assembly. A dropout cassette was included to facilitate the screening of positive colonies. Protein-coding sequences were introduced by Golden Gate Assembly using AarI cut sites [269]. pET-SUMO-GFP (gift from Prof. Utz Fischer, Julius-Maximilians-University, Würzburg, Germany) was used as the parental vectors for protein overexpression in *E. coli*. The lentivirus plasmid was a gift from Prof. Chase Beisel (HIRI-HZI, Würzburg, Germany). An ALFA-tag was included to facilitate the detection of the expressed protein [317]. The frameshift reporter vector for the *in vitro* translation contained ß-globin 5' and 3' UTRs as well as a 30 nt long poly-(A) tail. The insert was derived from nucleotides 12686–14190 of SARS-CoV-2 (NC_045512.2); a 3×FLAG-tag was introduced at the N-terminus to facilitate detection. To generate 0% and 100% –1PRF controls, the –1PRF site was mutated by disrupting the pseudoknot structure as well as the slippery sequence.

Optical tweezers constructs were based on the wild type SARS-CoV-2 frameshift site (nucleotides 13475-13541) cloned into the plasmid pMZ_lambda_OT, which encodes for

the optical tweezer handle sequences (2 kb each) flanking the RNA structure (130 nt). Constructs were generated using Gibson Assembly. Sequences of all plasmids and oligos used in this study are given in **Source Data File**.

Cell culture, transfections, generation of polyclonal stable cell lines

HEK293 cells (gift from Prof. Jörg Vogel, HIRI-HZI) and Huh7 cells (gift from Dr. Mathias Munschauer, HIRI-HZI), were maintained in DMEM (Gibco) supplemented with 10% FBS (Gibco) and 100 µg/ml streptomycin and 100 U/ml penicillin. Calu-3 cells (ATCC HTB-55) were cultured in MEM (Sigma) supplemented with 10% FBS. Cell lines were kept at 37 °C with 5% CO2. Transfections were performed using PEI (Polysciences) according to manufacturer's instructions. For co-transfections, plasmids were mixed at a 1:1 molar ratio.

VSV-G envelope pseudo-typed lentivirus for the generation of stable cell lines was produced by co-transfection of each transfer plasmid with pCMVdR 8.91 [318] and pCMV-VSV-G (gift from Prof. Weinberg, Addgene plasmid # 8454 [319]). 72 h post-transfection, the supernatant was cleared by centrifugation and filtration. The supernatant was used to transduce naïve Huh7 cells in the presence of 10 µg/ml polybrene (Merck Millipore). After 48 h, the cells were selected with 10 µg/ml blasticidin (Cayman Chemical) for 10 days to generate polyclonal cell lines.

SARS-CoV-2 infection

For infection with SARS-CoV-2, we used the strain hCoV-19/Croatia/ZG-297-20/2020, a kind gift of Prof. Alemka Markotic (University Hospital for Infectious Diseases, Zagreb, Croatia). The virus was raised for two passages on Caco-2 cells (HZI Braunschweig). Calu-3 cells (ATCC HTB-55) were infected with 2000 PFU/ml corresponding to an MOI of 0.03 at 24 h post-infection, cells were collected and lysed for proteomic and ribosome-interaction experiments. To study the effect of ZAP-S on SARS-CoV-2 infection, Huh-7 cells were employed. One hour before infection, Huh-7 cells both naïve or ZAP-S-overexpressing cells were either pre-stimulated with IFN-β (500 U/ml), IFN-ɣ (500 U/ml), IFN-ƛ1 (5 ng/ml), or left untreated. Cells were infected with 20,000 PFU/ml, corresponding to an MOI of 0.03 at 24 h post-infection, cell culture supernatants were collected and titrated by plaque assay on Vero E6 cells (ATCC CRL-1586). Briefly, confluent Vero E6 cells in 96-well plates were inoculated with dilutions of the virus-containing supernatants

for one hour at 37 °C, the inoculum was removed and cells were overlaid with MEM containing 1.75% methyl-cellulose. At three days post-infection, whole wells of the plates were imaged using an IncuCyte S3 (Sartorius) at 4x magnification, and plaques were counted visually.

Flow cytometry

HEK293 cells were transiently transfected with either the control construct or the –1PRF construct encoding for the dual-fluorescence EGFP-mCherry translation reporter as outlined in **Fig. 2A**. Cells were harvested at 24 h post-transfection and fixed with 0.4% formaldehyde in PBS. After washing with PBS, flow cytometry was performed on a FACSAria III (BD Biosciences) or a NovoCyte Quanteon (ACEA) instrument. Flow cytometry data were analyzed with FlowJo software (BD Biosciences). ECFP-positive cells were analyzed for the ratio between mCherry and EGFP **(Supplementary Fig. 2F)**. FE was calculated according to the following formula:

$$(1)\ \mathrm{FE}(\%) = \frac{mCherry_{test}/_{EGFP_{test}}}{mCherry_{control}/_{EGFP_{test}}}$$

where mCherry represents the mean mCherry intensity, EGFP the mean EGFP intensity, test represent the tested sample and control represents the in-frame control where mCherry and EGFP are produced in an equimolar ratio [320]. Data represent the results of at least three independent experiments.

Purification of recombinant proteins

Recombinant ZAP-S N-terminally tagged with 6×His-SUMO was purified from *E. coli* Rosetta 2 cells (Merck) by induction with 0.2 mM isopropyl β-d-1-thiogalactopyranoside for 18 h at 18 °C. Cells were collected, resuspended in lysis buffer (50 mM HEPES/KOH pH 7.6, 1 M NaCl, 1 mM DTT, 1 mM PMSF) and lysed in a pressure cell. The lysate was cleared by centrifugation and ZAP-S was captured using Ni-NTA resin (Macherey-Nagel). After elution with 500 mM imidazole, ZAP-S was further purified and the bound nucleic acids removed by size exclusion chromatography (HiLoad® 16/600 Superdex® 200) in 20 mM HEPES/KOH pH 7.6, 1 M KCl, 1 mM DTT, 20% glycerol. Protein identity was verified by SDS-PAGE as well as western blotting (**Supplementary Fig. 2D)**. Purified ZAP-S was rapidly frozen and

stored in aliquots at -80 °C. His-SUMO IGF2BP3 as well as His-SUMO were kind gifts from Dr. Andreas Schlundt (Goethe University, Frankfurt, Germany).

Western blots

Protein samples were denatured at 95 °C and resolved by 12% SDS-PAGE at 30 mA for 2 h. After transfer using Trans-Blot (Bio-Rad), nitrocellulose membranes were developed using the following primary antibodies: anti-His-tag (ab18184, dilution 1:1000), anti-DDDDK (ab49763, dilution 1:3000), anti-ALFA (FluoTag®-X2 anti-ALFA AlexaFluor 647, dilution 1:1000), anti-ZC3HAV1 (Proteintech 16820-1-AP, dilution 1:3000), anti-RPL4 (Proteintech 67028-1-Ig, dilution 1:10000), anti-RPS6 (Proteintech 14823-1-AP, dilution 1:500), anti-RYDEN (SHFL; Proteintech 27865-1-AP, dilution 1:1000). The following secondary antibodies were used: IRDye® 800CW Goat anti-rabbit (dilution 1:25000) and IRDye® 680RD Donkey anti-Mouse (dilution 1:15000; both LI-COR). Bands were visualized using an Odyssey Clx infrared imager system (LI-COR) or a Typhoon7000 (GE Healthcare).

***In vitro* translation assays**

mRNAs were *in vitro* transcribed using T7 polymerase purified in-house using linearized plasmid DNA as the template. These mRNAs were capped (Vaccinia Capping System, NEB) and translated using the nuclease-treated rabbit reticulocyte lysate (RRL; Promega). Typical reactions were comprised of 75% v/v RRL, 20 μM amino acids, and were programmed with ~50 μg/ml template mRNA. ZAP-S was buffer exchanged into 250 mM KCl, 50 mM HEPES/KOH pH 7.6, 0.05mM EDTA, 5% glycerol, 1 mM DTT, Rnasin and titrated in the range of 0-3 μM. Reactions were incubated for 1 h at 30 °C. Samples were mixed with 3x volumes of 1X NuPAGE™ LDS Sample Buffer (Invitrogen), boiled for 3 min, and resolved on a NuPAGE™ 4 to 12% Bis-Tris polyacrylamide gel (Invitrogen). The products were detected using western blot (method as described above). The nitrocellulose membranes were developed using anti-DDDDK primary (Abcam ab49763) and IRDye® 680RD donkey anti-mouse secondary antibody (LI-COR). Bands were visualized using an Odyssey Clx infrared imager system (LI-COR). Bands corresponding to the –1 or 0-frame products, 58 kDa and 33 kDa respectively, on western blots of *in vitro* translations were quantified densitometrically using ImageJ software [321]. FE was calculated as previously described, by the formula

(2) intensity (–1-frame)/(intensity (–1-frame) + intensity (0-frame)) [114].

The change in FE was calculated as a ratio of FE of each condition to the FE of no-protein control in each measurement. Experiments were repeated at least 3 independent times.

Microscale thermophoresis

Short frameshifting RNA constructs were *in vitro* transcribed using T7 polymerase as described above. RNAs were labeled at the 3' end using pCp-Cy5 (Cytidine-5'-phosphate-3'-(6-aminohexyl) phosphate) (Jena Biosciences). For each binding experiment, RNA was diluted to 10 nM in Buffer A (50 mM Tris-HCl pH 7.6, 250 mM KCl, 5 mM MgCl2, 1 mM DTT, 5% glycerol supplemented with 0.05% Tween 20 and 0.2 mg/ml yeast tRNA). A series of 16 tubes with ZAP-S dilutions were prepared in Buffer A on ice, producing ZAP-S ligand concentrations ranging from 40 pM to 2 μM. For measurements, each ligand dilution was mixed with one volume of labeled RNA, which led to a final concentration of 5.0 nM labeled RNA and 20 pM to 1 μM. For ribosomes, The reaction was mixed by pipetting, incubated for 10 min at room temperature, followed by centrifugation at 10,000 × g for 5 min. Capillary forces were used to load the samples into Monolith NT.115 Premium Capillaries (NanoTemper Technologies). Measurements were performed using a Monolith Pico instrument (NanoTemper Technologies) at an ambient temperature of 25 °C. Instrument parameters were adjusted to 5% LED power, medium MST power, and MST on-time of 2.5 seconds. An initial fluorescence scan was performed across the capillaries to determine the sample quality and afterward, 16 subsequent thermophoresis measurements were performed. Data of three independently pipetted measurements were analyzed for the ΔFnorm values determined by the MO. Affinity Analysis software (NanoTemper Technologies). Graphs were plotted and binding affinities were calculated using GraphPad Prism 9.2.0 software. Similar experiments were conducted with ribosomes in buffer containing 50 mM HEPES-KOH pH 7.6, 250 mM KCl, 5% Glycerol, 1 mM DTT, 0.05% tween, 0.1mg/ml yeast tRNA (Invitrogen) with ligand concentrations ranging between 0.00003 to 2 μM for both subunits. The measurements were performed as described above.

Electrophoretic Mobility Shift Assay (EMSA)

EMSAs to visualize the stoichiometry of ZAP-S binding to SARS-CoV-2 PRF RNA variants were performed as described previously with some modifications [48]. Briefly, 100 nM RNA labelled with Cy5 at the 3’end was incubated with serial dilutions of

ZAP-S in Buffer A supplemented with 5% glycerol. Reactions were separated by 0.5% agarose electrophoresis in 1x TBE prior to visualization using a Typhoon7000 imager (GE Healthcare).

Microscopy

HEK293 cells were cultured on glass slides and transfected as described above. The cells were fixed with 4% paraformaldehyde in 1x PBS for 15 min at room temperature. After washing with 1x PBS, cells were mounted in ProLong Antifade Diamond without DAPI (Invitrogen). Microscopy was performed using a Thunder Imaging System (Leica) using 40% LED power and the 40x objective. EGFP was excited at 460-500 nm and detected at 512-542 nm. mCherry was excited at 540-580 nm and detected at 592-668 nm. The images were processed with the LasX software (Leica). For immunofluorescence, Huh-7 cells naïve or overexpressing ZAP-S were prestimulated or infected as mentioned above. Cells were fixed with 6% paraformaldehyde in PBS for 1 h at room temperature, followed by washing with PBS. Cells were permeabilized with 0.1% Triton X-100 in PBS for 10 min at room temperature, washed with PBS, and blocked with 2% BSA in PBS for 1 h. Antibody labelling was performed with recombinant anti-nucleocapsid protein SARS-CoV-2 (Abcalis, Germany; #ABK84-E2-M) and secondary antibody anti-mouse Alexa488 (Cell Signaling Technology, USA; #4408), each step was followed by three washing steps with PBS containing 0.05% Tween-20. Finally, cells were overlaid with Vectashield Mounting Medium (Biozol, Germany).

Polysome profiling analysis

A plasmid expressing ZAP-S N-terminally tagged with a His-tag was transfected into HEK293 cells using PEI, as described above. To check endogenous ZAP-S expression, HEK cells were transfected with a plasmid containing the same backbone and His-tag. At 24 h post-transfection, cycloheximide (VWR) was added to the medium at a final concentration of 100 µg/ml to stop translation. Approximately 107 HEK cells were lysed with 500 µl lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM MgCl2, 1 mM DTT, 100 µg/ml Cycloheximide, 1% Triton X-100), and the lysate was clarified by centrifugation at 17,0000 × g for 10 min at 4 °C. Polysome buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM MgCl2, 1 mM DTT, 100 µg/ml Cycloheximide) was used to prepare all sucrose solutions. Sucrose density gradients (5%–45% w/v) were freshly made in SW41 ultracentrifuge tubes (Beckman) using a Gradient Master (BioComp Instruments)

according to manufacturer's instructions. The lysate was then applied to a 5%–45% sucrose continuous gradient and centrifuged at 35,000 rpm (Beckman Coulter Optima XPN) for 3 h, at 4 °C. The absorbance at 254 nm was monitored and recorded and 500 µl fractions were collected using a gradient collector (BioComp instruments). The protein in each fraction was pelleted with trichloroacetic acid, washed with acetone, and subjected to western blotting, as described above. For polysome profiling analysis of RRL a similar procedure was followed except SARS-CoV-2 mRNA was *in vitro* transcribed and translated in RRL as described above for 20 min at 30 °C and 300 µl of this lysate was applied to a sucrose gradient.

Ribosome pelleting assay

Calu-3 lysates were prepared as described above. 300 µl of the lysate was loaded onto a 900 µl 1 M sucrose cushion in polysome buffer (described above) in Beckman centrifugation tubes. Ribosomes were pelleted by centrifugation at 75,000 rpm for 2 h, at 4 °C, using a Beckman MLA-130 rotor (Beckman Coulter Optima MAX-XP). After removing the supernatant, ribosome pellets were resuspended in polysome buffer and were used for western blotting, as described above.

Eukaryotic ribosomal subunit purification

40S and 60S subunits were purified from untreated rabbit reticulocyte lysate (Green Hectares) as previously described[322]. Briefly, ribosomes were pelleted by centrifugation (4°C, 270,000 × g, 4.5 h) and resuspended in 20 mM Tris-HCl pH 7.5, 4.0 mM $MgCl_2$, 50 mM KCl, 2.0 mM DTT. Following treatment with 1.0 mM puromycin and addition of KCl to 0.5 M, 40S and 60S subunits were separated by centrifugation (4°C, 87,000 × g, 16 h) through a sucrose density gradient (10 □ 30% sucrose in 20 mM Tris-HCl pH 7.5, 2.0 mM DTT, 4.0 mM $MgCl_2$, 0.5 M KCl). After analysis by SDS-PAGE, uncontaminated fractions were pooled, and exchanged into 20 mM Tris-HCl pH 7.5, 100 mM KCl, 2.0 mM $MgCl_2$, 2.0 mM DTT, 250 mM sucrose using Amicon centrifugal concentrators (4°C,100K MWCO). Ribosome subunits were snap-frozen in liquid nitrogen and stored at -80°C until further use.

Optical tweezers constructs

5' and 3' DNA handles, and the template for *in vitro* transcription of the SARS-CoV-2 putative pseudoknot RNA were generated by PCR using the pMZ_lambda_OT vector. The

3′ handle was labeled during the PCR using a 5′ digoxigenin-labeled reverse primer. The 5′ handle was labeled with Biotin-16-dUTP at the 3′ end following PCR using T4 DNA polymerase. The RNA was *in vitro* transcribed using T7 RNA polymerase. Next, DNA handles (5′ and 3′) and *in vitro* transcribed RNA were annealed in a mass ratio 1:1:1 (5 µg each) by incubation at 95 °C for 10 min, 62 °C for 2 h, 52 °C for 2 h and slow cooling to 4 °C in annealing buffer (80% formamide, 400 mM NaCl, 40 mM HEPES, pH 7.5, and 1 mM EDTA, pH 8) to yield the optical tweezer suitable construct **(Fig. 4E)**. Following the annealing, samples were concentrated by ethanol precipitation, pellets were resuspended in 40 µl RNase-free water, and 4 µl aliquots were stored at –80°C until use.

Optical tweezers data collection and analysis

Optical tweezers measurements were performed using a commercial dual-trap platform coupled with a microfluidics system (C-trap, Lumicks). For the experiments, optical tweezers (OT) constructs were mixed with 4 µl of polystyrene beads coated with antibodies against digoxigenin (AD beads, 0.1% v/v suspension, Ø 2.12 µm, Spherotech), 10 µl of assay buffer (20 mM HEPES, pH 7.6, 300 mM KCl, 5 mM MgCl2, 5 mM DTT and 0.05% Tween 20) and 1 µl of RNase inhibitor. The mixture was incubated for 20 min at room temperature in a final volume of 19 µl and subsequently diluted by the addition of 0.5 ml assay buffer. Separately, 0.8 µl of streptavidin-coated polystyrene beads (SA beads, 1% v/v suspension, Ø 1.76 µm, Spherotech) were mixed with 1 ml of assay buffer. The flow cell was washed with the assay buffer, and suspensions of both streptavidin beads and the complex of OT construct with anti-digoxigenin beads were introduced into the flow cell. During the experiment, an anti-digoxigenin (AD) bead and a streptavidin (SA) bead were trapped and brought into proximity to allow the formation of a tether. The beads were moved apart (unfolding) and back together (refolding) at a constant speed (0.05 µm/s) to yield the force-distance (FD) curves. The stiffness was maintained at 0.31 and 0.24 pN/nm for trap 1 (AD bead) and trap 2 (SA bead), respectively. For experiments with ZAP-S protein, recombinantly expressed ZAP-S was diluted to 400 nM in assay buffer and introduced to the flow cell. FD data were recorded at a rate of 78125 Hz.

Raw data files were processed using our custom-written python algorithm called Practical Optical Tweezers Analysis TOol. In brief, raw data were first down sampled by a factor of 20 to speed up subsequent processing, and the noise was filtered using Butterworth filter (0.05 filtering frequency, filter order 2). FD curves were fitted using a custom written

Python script, which is based on Pylake package provided by Lumicks (https://lumicks-pylake.readthedocs.io/en/stable/). For data fitting, we employed a combination of two worm-like chain models (WLC1 for the fully folded double-stranded parts and WLC2 for the unfolded single-stranded parts) as described previously [99]. Firstly, the initial contour length of the folded RNA was set to 1256 ± 5 nm, and the persistence length of the double-stranded part was fitted [50]. Then, the persistence length of the unfolded RNA was set to 1 nm, and the contour length of the single-stranded part was fitted. The work performed on the structure while unfolding or refolding was calculated as difference between area under curve (AUC) of the fit for folded region and AUC of the fit for unfolded region, counted from the beginning of the FD curve till the unfolding step coordinates. To be able to compare the effect of protein presence on different structures we decided to normalize the refolding work in each pair (protein-/protein+) to the protein- sample. We used the PK+IMP3 value as molecular crowding control and further normalized all the ZAP+ values to it. This allowed us to quantitatively compare the effect of ZAP on different RNA molecules. Data were statistically analyzed, and the results were plotted using Prism 9.2.0 (GraphPad).

qRT-PCR

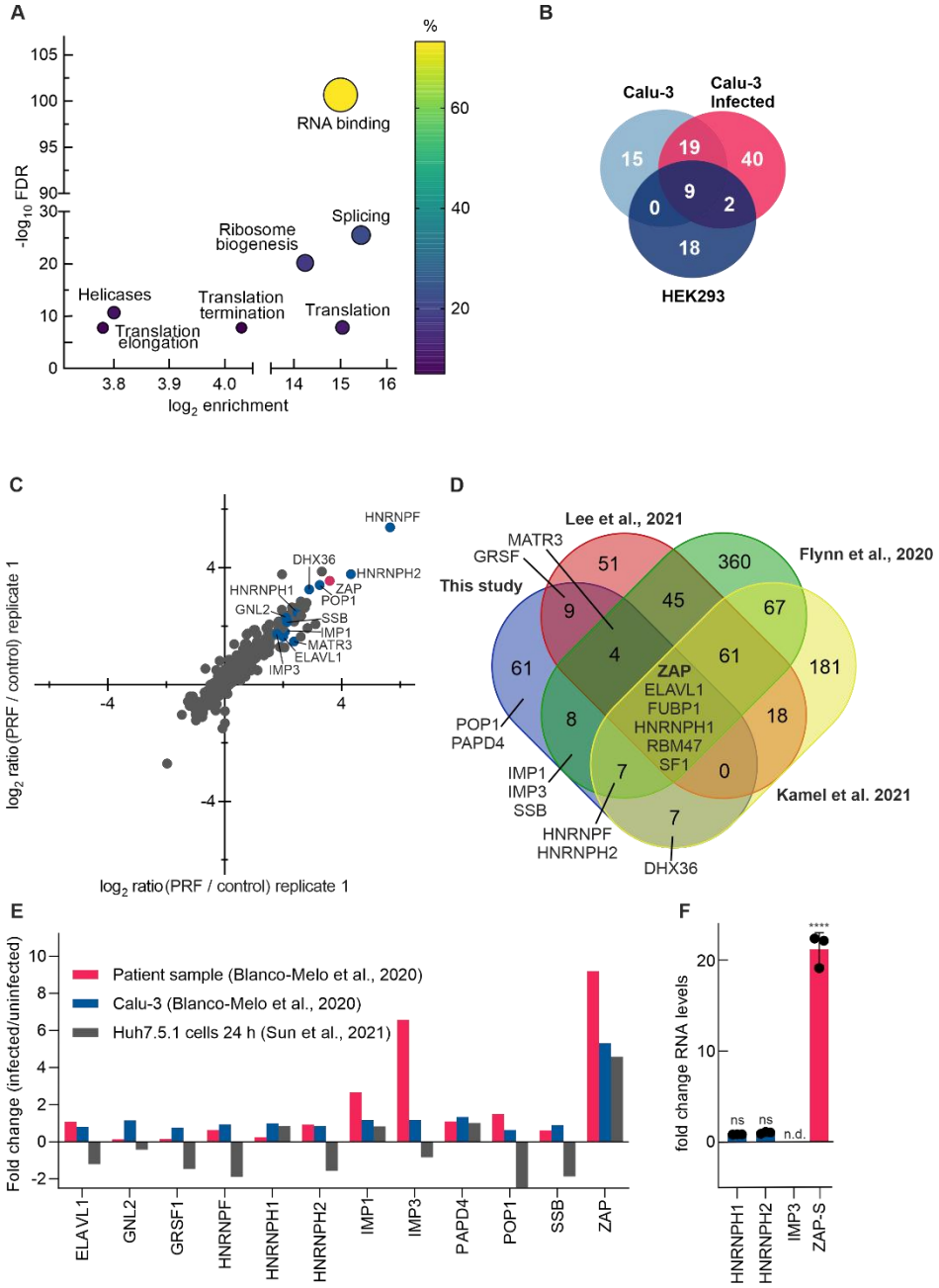
Total RNA was isolated as described previously [323], and the reverse transcription using RevertAid (Invitrogen) was primed by oligo(dT). Reactions of quantitative real-time PCR (qRT-PCR) were set up using POWER SYBR green Master-mix (Invitrogen) according to manufacturer's instructions and analyzed on the CFX96 Touch Real-Time PCR Detection System (Bio-Rad) under the following cycling condition: 50°C for 2 min, 95°C for 2 min, followed by 40 cycles of 95°C for 15 s and 60°C for 30 s, and ending with a melt profile analysis. The fold change in mRNA expression was determined using the 2^-$\Delta\Delta C_t$ method relative to the values in uninfected samples, after normalization to the housekeeping gene (geometric mean) GAPDH. Statistical analysis was conducted comparing ΔC_t values of the respective RNA in uninfected and infected cells and results were plotted using Prism 9.2.0 (GraphPad).

Quantification and statistical analysis

All statistical analysis and software used have been mentioned in the Figure Legends and Materials & Methods. Ordinary one-sided ANOVA was followed by a Brown-Forsythe test to ensure equal variance among the samples. Finally, a Dunnett's multiple comparisons test

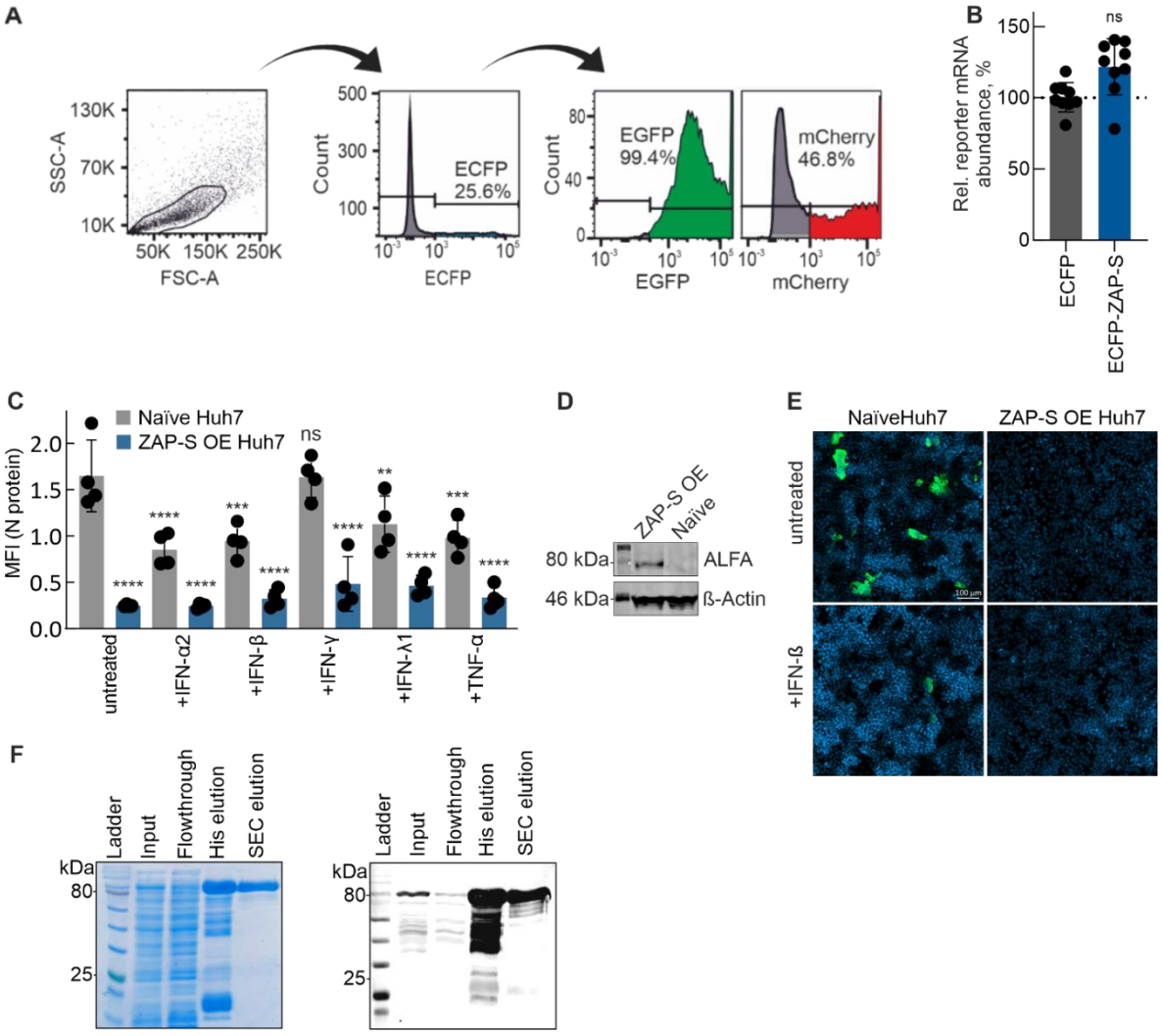
was employed to identify the differentially regulated conditions compared to our control constructs. Statistical analysis was performed using GraphPad Prism version 9.2.0. Measurements from the *in vitro* western blot assay and *in vivo* dual fluorescence assay resulted from 3 technical replicates. Measurements from single-molecule experiments resulted from a specified number (n) of traces from a single experiment. For the ensemble MST analysis, all analysis for ΔFnorm from 3 individual replicates was performed in Nanotemper MO. Affinity software. Data was plotted and KD was determined using Graphpad Prism version 9.2.0 nonlinear regression, binding-saturation function.

Supplementary Information



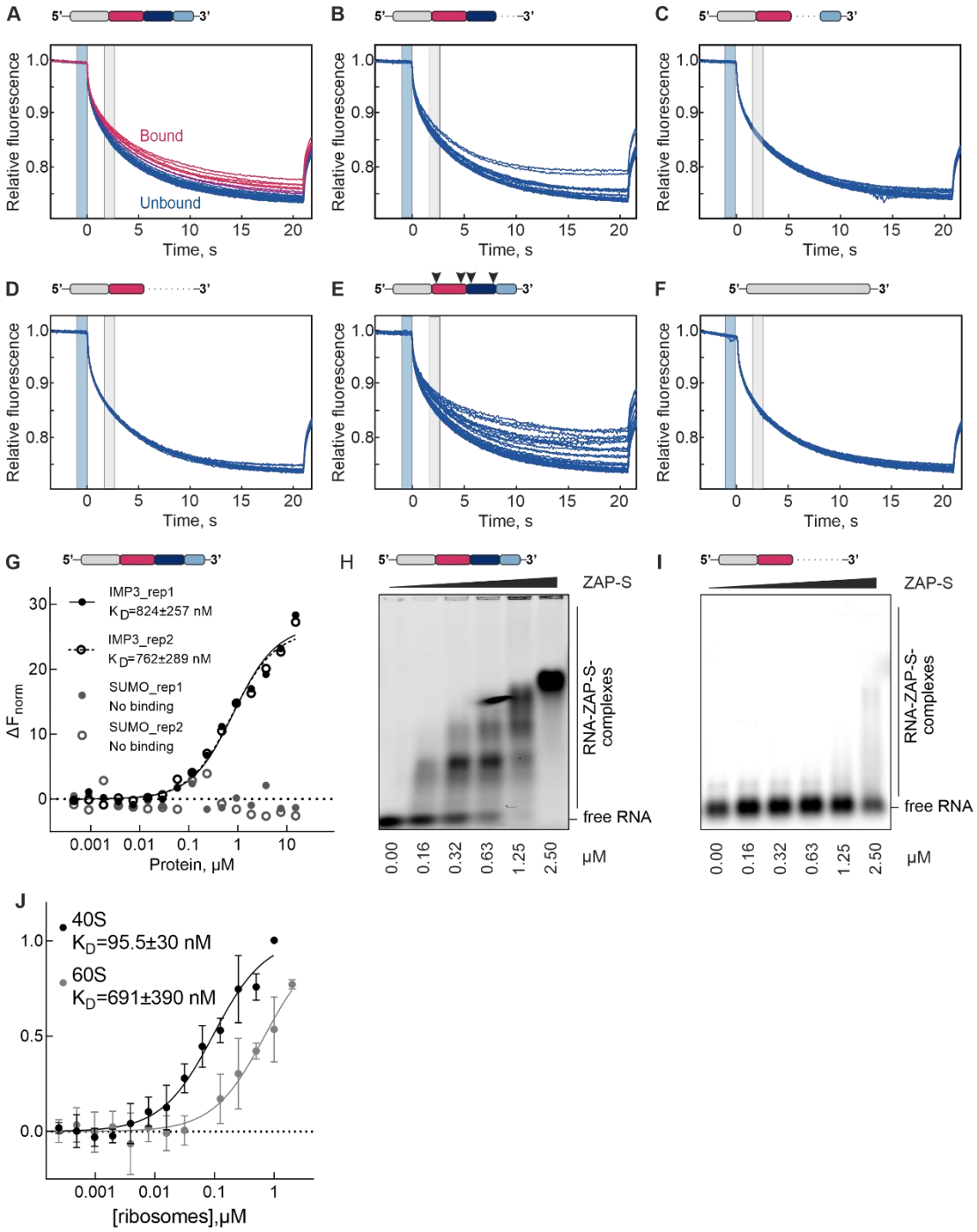
Supplementary Fig. 1. Capture and analysis of frameshift RNA interactors, related to Fig. 1. (A) Gene ontology (GO) term analysis of SARS-CoV-2 frameshift RNA interactions. FDR – false discovery rate. **(B)** Venn Diagram comparing the hits of the *in vitro* RNA antisense purification in HEK293, uninfected Calu-3 as well as SARS-CoV-2 infected Calu-3 cells. **(C)** Scatter plot of log2-transformed ratio of RNA-antisense purification in HEK293 cells. **(D)** Venn Diagram comparing the hits of the vitro RNA antisense purification of the SARS-CoV-2 frameshift site from this study with the hits of genome-wide interactome captures from the literature [20,21,22]. **(E)** Fold change of SARS-CoV-2 PRF element interactors in published RNAseq datasets. **(F)** Expression profiles of selected genes in Calu-3 infected cells with SARS-CoV-2 at 72 hours. RNA levels

were quantified by qRT-PCR and normalized to the respective RNA abundance in uninfected cells (shown as $2^{-\Delta\Delta CT}$). Data points represent the mean ± s.d. (n = 3 independent experiments). P values were calculated using an ordinary unpaired one-sided ANOVA comparing ΔC_t values of the respective RNA in uninfected and infected cells. Exact p values: ZAP-S – <0.0001. n.d. - not detected. See also **Supplementary Table 2**.



Supplementary Fig. 2. Effect of ZAP-S on FE of various PRF sites, related to Fig. 2 and 3. (A) Gating strategy for flow cytometry determining FE in HEK293 cells. Cell populations were determined based on SSC and FSC. ECFP-positive cells were analyzed for the mean intensities of EGFP and mCherry. **(B)** Abundance of reporter mRNA upon overexpression of ZAP-S determined by qRT-PCR. Datapoints represent the mean ± s.d. (n = 3 independent experiments with three technical replicates each). P values were calculated using an ordinary unpaired one-sided ANOVA comparing raw Ct values to the ECFP control. Exact P values: **** P < 0.0001, Naïve Huh7 IFN-β – 0.0003, Naïve Huh7 IFN-λ1 – 0.0086, Naïve Huh7 TNF-α – 0.0005. **(C)** Quantification of immunostaining of N-protein in infected naïve Huh7 or ZAP-S overexpressing Huh7 cells (ZAP-S OE) at 24 hours post-infection. Treatment with IFN-ɣ (500 U/ml), IFN-β (500 U/ml), or IFN-ƛ1 (5 ng/ml) was done one hour before infection. Boxes show mean values ± s.d. (n = 4 technical replicates). **(D)** Western blot of naïve as well as ALFA-tagged ZAP-S-overexpressing Huh7 cells. ALFA-ZAP-S was

detected using anti-ALFA antibody, ß-actin serves as a loading control. n = 3 independent experiments. **(E)** Confocal micrograph of immunostaining of N-protein in infected naïve Huh7 or ZAP-S overexpressing Huh7 cells (ZAP-S OE) at 24 hours post-infection. **(F)** Coomassie-stained SDS-PAGE and western blot of heterologous expression of ZAP-S in *E. coli* as well as the purification steps. ZAP-S was detected using an anti-ZC3HAV1 (ZAP) antibody. n = 1 experiment



Supplementary Fig. 3. Thermophoresis raw data and interaction of IMP3 and SUMO with SARS-CoV-2 FS RNA, related to Fig. 4. Thermophoretic time-traces of microscale thermophoresis (MST) measurements of binding between ZAP-S and **(A)** Full PRF, **(B)** ΔSL2 mutant, **(C)** ΔSL3 mutant, **(D)**

ΔSL2+3 mutant, **(E)** compensatory mutant and **(F)** scrambled mutant. Blue and grey boxes in the time-course traces represent the temperature jump and MST-on time (2.5 s), respectively. In all cases, there is no adsorption of the labeled protein to the capillaries. See Fig. 4 for resulting binding curves. **(G)** Microscale thermophoresis to monitor binding of IMP3 and SUMO to SARS-CoV-2 FS PK. Unlabeled protein (0.4 nM to 15 μM) was titrated against 3' pCp-Cy5 labeled RNA (5 nM) and thermophoresis was recorded at 25°C with 5% LED intensity and medium MST power. Change in fluorescence (ΔF_{norm}) was measured at MST on-time of 2.5 s. Data were analyzed using standard functions in the MO. Affinity Analysis software and K_D was determined using Graphpad Prism 9.2.0. Data represent mean +/- s.d. of each independent measurement (n=2 independent experiments). Electrophoretic mobility Assay (EMSA) of **(H)** SARS-CoV-2 pseudoknot and **(I)** ΔSL2+3 mutant. 100 nM RNA labelled with Cy5 at the 3'end was incubated with serial dilutions of ZAP-S in Buffer A supplemented with 5% glycerol. Reactions were separated by 0.5% agarose electrophoresis in 1x TBE prior to visualization. n = 2 independent experiments.

Data and materials availability

Supplementary tables and source data are provided with this thesis and available online. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD029656 [324].

Code availability

Custom scripts were employed to process optical tweezers data. Python algorithm called Practical Optical Tweezers Analysis TOol is available on Github (POTATO, https://github.com/lpekarek/POTATO) [325].

Acknowledgements

U.R. and L.P. contributed equally to this work. We thank Dr. Zeljka Macak-Safranko and Prof. Alemka Markotic (University of Zagreb) for providing the SARS-CoV-2 virus isolate prior to publication. We thank Dr. Andreas Schlundt for kind gifts of IGF2BP3 and SUMO proteins (Goethe University, Frankfurt, Germany). We thank Dr. Joop van den Heuvel (HZI) for his helpful suggestions on protein purification. We thank Dr. Anke Sparmann, Prof. Jörg Vogel, Prof. Lars Dölken, Prof. Utz Fischer and Prof. Thomas Pietschmann for critical reading of the manuscript. We thank expert technical assistance by Tatyana Koch (HIRI-HZI). We thank Ayse Barut for cell maintenance for infection studies (HZI). We thank Dr. Andreas Schlosser and Stephanie Lamer from the Rudolf Virchow Center for the LC-MS/MS analysis. Figures were partially generated using BioRender.com (licensed for commercial printing to A.K.). This project is funded fully or in part by the Helmholtz Association. L.C.S. funded through MWK Niedersachsen Grant Nr. 14-76103-184

CORONA-2/20. N.C. received funding from the European Research Council (ERC) Grant Nr. 948636.

Contributions

M.Z. and A.K. designed and cloned the constructs, purified proteins, and performed most of the biochemical experiments. L.P., S.B. and N.C. designed the OT constructs, L.P. performed most of the single molecule experiments and processed the data with the help of S.B. S.B. and L.P. have written the scripts for automatized analysis of the single molecule data. U.R. performed the SARS-CoV-2 infection assays and collected lysates for downstream biochemical analysis. L.Y. and R.S. performed DMS-MaPseq experiments and analyzed the data. L.C.-S., R.S., and N.C. supervised the study. M.Z., A.K., L.P. and N.C. wrote the paper. All authors contributed to the review and editing of the final paper.

Supplement to Chapter 3

In this supplement, I will discuss the contributions made by me to Hill, Pekarek, Napthine *et al.*, 2021 study[99]. The discussed results shed light on the mechanisms of EMCV frameshift modulation by its 2A protein and also characterize the interactions of the 2A protein with prokaryotic and eukaryotic ribosomes. This enabled us to get structural insights on the mechanisms of 2A protein and contributes to the overall knowledge of virus-encoded –1PRF protein modulators.

Supplementary tables are available online https://doi.org/10.1038/s41467-021-27400-7

Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch

Chris H. Hill*[†1,2,3], Lukas Pekarek[†4], Sawsan Napthine[†1], Anuja Kibe[4], Andrew E. Firth[1], Stephen C. Graham*[1], Neva Caliskan*[4,5] and Ian Brierley*[1]

[†] These authors contributed equally

*Corresponding authors:
Chris H. Hill chris.hill@york.ac.uk
Stephen C. Graham scg34@cam.ac.uk
Neva Caliskan neva.caliskan@helmholtz-hiri.de
Ian Brierley ib103@cam.ac.uk

[1] Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, UK. CB2 1QP.

[2] MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Francis Crick Ave, Cambridge, UK. CB2 0QH

[3] **Present address:** Department of Biology, University of York, Wentworth Way, York, YO10 5DD.

[4] Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany

[5] Medical Faculty, Julius-Maximilians University Würzburg, Josef-Schneider-Straße 2/D15, 97080 Würzburg, Germany

Published: Nature Communications volume 12, Article number: 7166 (2021)

Introduction

Cardioviruses present a highly unusual variation to conventional viral PRF in which the virally encoded 2A protein is required as an essential *trans*-activator[326, 327]. Here, the spacing between the slippery sequence and stem-loop is 13 nt, significantly longer than typically seen, and 2A protein has been proposed to bridge this gap through interaction with the stem-loop[327]. This allows for temporal control of gene expression as the efficiency of –1 frameshifting is linked to 2A concentration, which increases with time throughout the infection cycle[327].

2A is a small, basic protein (~17 kDa; 143 amino acids; pI ~9.1) generated by 3C-mediated proteolytic cleavage at the N-terminus[328] and Stop-Go peptide release at the C-terminus[329]. The PRF-stimulatory activity of 2A is related to its ability to bind to the RNA stimulatory element[327]. However, 2A also binds to 40S ribosomal subunits[330], inhibits apoptosis[331] and contributes to host cell shut-off by inhibiting cap-dependent translation. A C-terminal YxxxxLΦ motif has been proposed to bind to and sequester eIF4E in a manner analogous to eIF4E binding protein 1 (4E-BP1)[330], thereby interfering with eIF4F assembly[332]. However, the mechanism by which it recognises RNA elements and stimulates frameshifting remains obscure.

Here, we show that 2A binds directly to the frameshift-stimulatory element in the viral RNA with nanomolar affinity and we define the minimal RNA element required for binding. We report a direct interaction of 2A with both mammalian and bacterial ribosomes. Together, this work enables us to better understand protein-mediated frameshifting and 2A-mediated regulation of gene expression.

Results

2A binds to a minimal 47 nt pseudoknot in the viral RNA

The RNA sequence that directs PRF in EMCV consists of a G_GUU_UUU slippery sequence and a stimulatory stem-loop element downstream (**Fig. 1a**). We have previously demonstrated that three conserved cytosines in the loop are essential for 2A binding[327] (**Fig. 2a**). To map the interaction between 2A and the stimulatory element in more detail, we prepared a series of synthetic RNAs with truncations in the shift site, loop, and 5′ and 3′ extensions on either side of the stem (EMCV 1–6; **Fig. 1b**). These were fluorescently

labelled at the 5′ end, and their binding to 2A was analysed by microscale thermophoresis (MST; **Fig. 1C, Supplementary Table 2**).

Binding of 2A to EMCV 1 RNA is high affinity (K_D = 360 ± 34 nM). Removal of the 3′ extension, as in EMCV 3 and EMCV 6, further increases the affinity (K_D values of 40 ± 2 and 70 ± 14 nM, respectively), perhaps by removing competing base-pairing interactions. There is no substantial difference between affinities of EMCV 3 and 6, which differ only by the presence of the shift site. Removal of the 5′ extension, as in EMCV 2 and EMCV 4, completely abolishes 2A binding, and truncation of the loop, including a putative second stem (EMCV 5) reduces binding to micromolar levels. Finally, reciprocal MST experiments with fluorescently labelled 2A and unlabelled RNA yielded similar K_D values (**Supplementary Fig. 2**, **Supplementary Table 2**).

The failure of 2A to bind to EMCV 2, 4 and 5 was unexpected as these RNAs retain the main stem and the conserved cytosine triplet in the putative loop region. A possible explanation is that the frameshift-relevant state may include an interaction between the loop and the 5′ extension, forming a different conformation that 2A selectively recognises

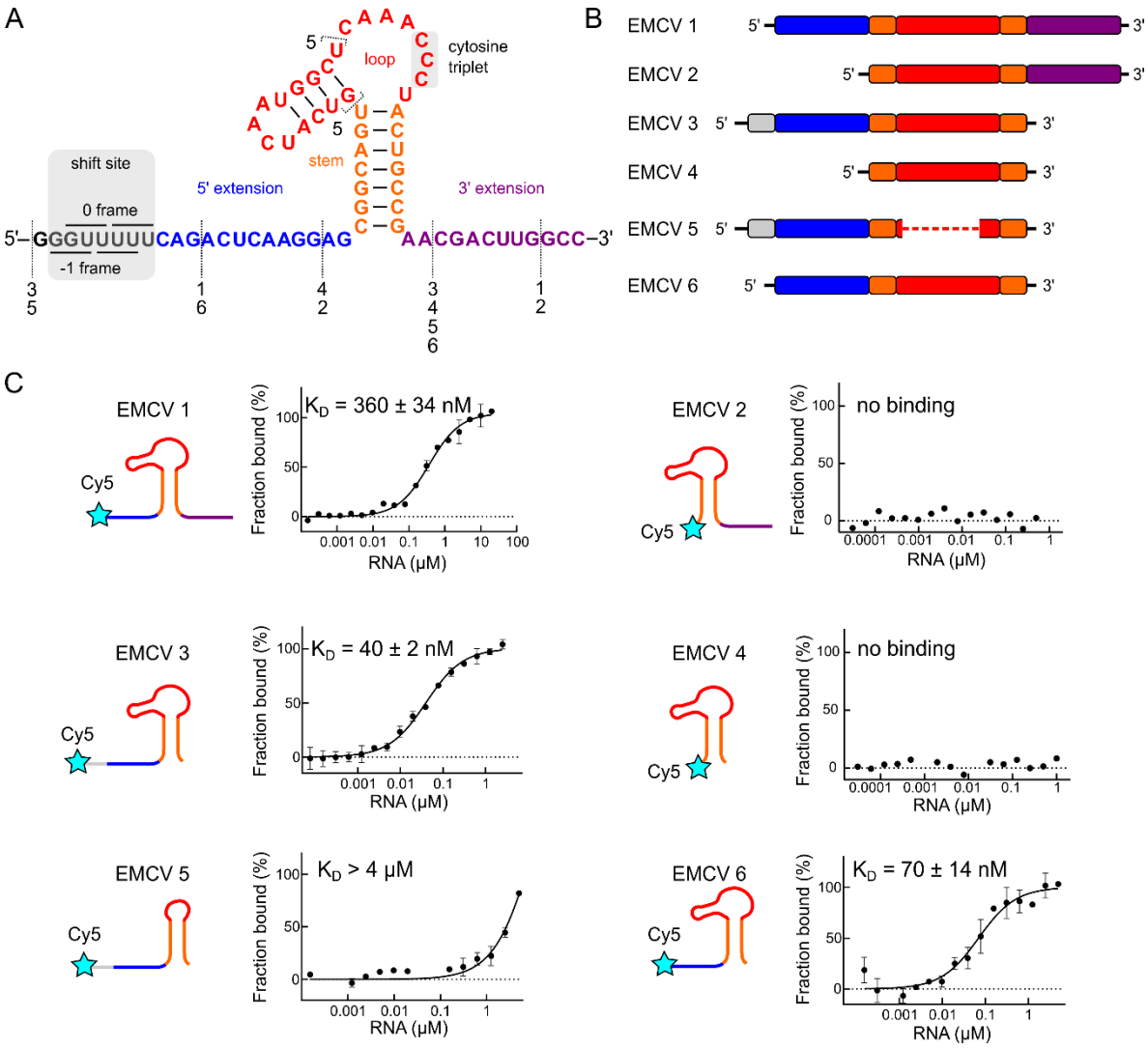


Figure 1. 2A binds to a minimal 47 nt element in the viral RNA. a-b, Sequences and schematic diagrams of the EMCV 1–6 constructs used to assay 2A binding. **c,** EMSA analyses showing that removal of the 5′ extension (blue) disables 2A binding. **d,** Microscale thermophoresis (MST) was used to quantify the interactions observed in **c**. All measurements were repeated as two independent experiments and error bars represent the standard deviation from the mean. RNA concentration ranges between 60 pM – 20 µM (for EMCV 1) and 150 pM – 5 µM (for EMCV 2–6). Source data are provided as a Source Data file.

2A interacts with the small ribosomal subunit in both eukaryotes and prokaryotes

In addition to its role as a component of the stimulatory element, 2A has been reported to bind to 40S subunits in EMCV-infected cells[330]. To determine if the interaction of 2A with the 40S subunit can be reproduced *ex vivo*, we purified ribosomal subunits from native RRL and analysed 2A-subunit interactions by MST (**Fig. 4a, b**). Consistent with previous data, 2A forms a tight complex with 40S (apparent K_D = 10 ± 2 nM) but not 60S. This apparent

selectivity for the small subunit was also observed with purified prokaryotic ribosome subunits. 2A binds with very high affinity to 30S (apparent K_D = 4 ± 1 nM; **Fig. 4c**), but not 50S (**Fig. 4d**). We next examined binding of 2A to intact 70S ribosomes and to reconstituted, mRNA-bound 70S ribosomes at the initiation stage (70S IC; initiator $tRNA^{Met}$ in the P-site and an empty A-site). We were able to detect high affinity interactions with both uninitiated and initiated 70S ribosomes (**Fig. 4e, f**).

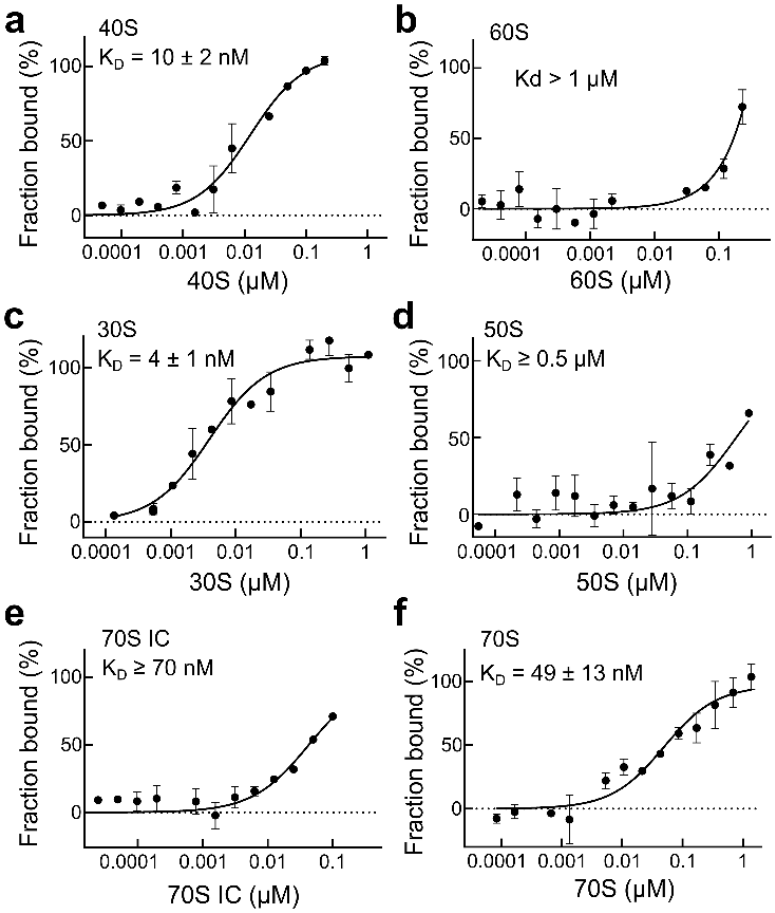
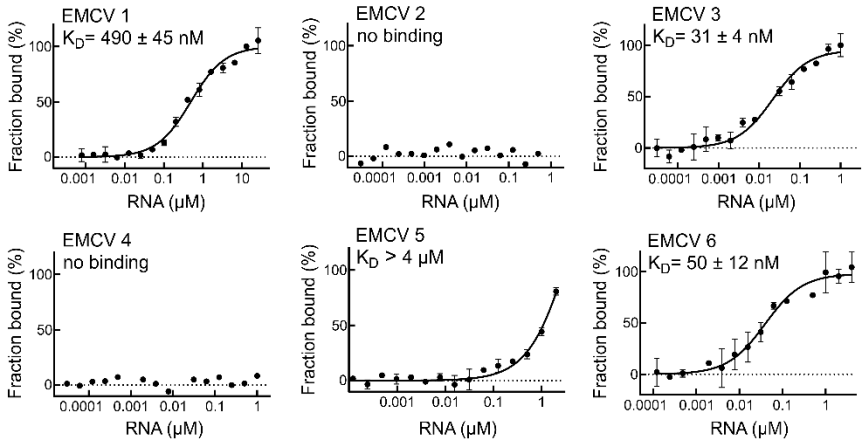


Figure 2. 2A binds directly to eukaryotic and prokaryotic ribosomes. a, MST binding curves and apparent K_D values using unlabelled 40S subunits at a concentration range of 20 pM – 0.4 μM. All measurements were repeated as two independent experiments and error bars represent the standard deviation from the mean. 2A binds with high affinity to the small ribosomal subunit. **b,** As in **a** with 60S subunits. Error bars as above. **c,** Binding curve and apparent K_D values using unlabelled 30S subunits at a concentration range of 30 pM – 1 μM. Error bars as above. 2A shows a strong interaction with the prokaryotic small subunit. **d,** As in **c** with 50S subunits at a concentration range of 27 pM – 0.9 μM. **e,** Binding curves and reported K_D values for 2A-70S IC interactions. Error bars as above. **f,** Same as **e**, with 2A and vacant 70S.



Supplementary Figure 2 – related to **Figure 2. The 2A-RNA interaction occurs with nanomolar affinity.** MST binding curves and reported K_D values of fluorescently labelled 2A protein (5 nM) and short unlabelled RNAs (as in **Fig. 1a, b**) at concentrations between 800 pM – 26 µM for EMCV 1 and 120 pM – 4 µM for EMCV 2–6. All measurements were repeated as two independent experiments and error bars represent the standard deviation from the mean.

Conclusion

Through the contributions to the study discussed in this thesis, it is clear that 2A binds directly to the frameshift RNA and host ribosomes. We also define the minimal RNA element required for this binding.

Overall, this entire study defines the structural and molecular basis for the temporally regulated 'switch' behind the reprogramming of viral gene expression in EMCV infection (**Fig. 3**). At the heart of this is 2A: an RNA-binding protein with the remarkable ability to discriminate between stem-loop and pseudoknot conformers of the PRF stimulatory element. We also reveal how 2A interferes with host translation by specifically recognising distinct conformations within the ribosomal RNA. Together, this illustrates how the conformational plasticity of one RNA-binding surface can contribute to multiple functions through finely tuned relative affinities for different cellular targets.

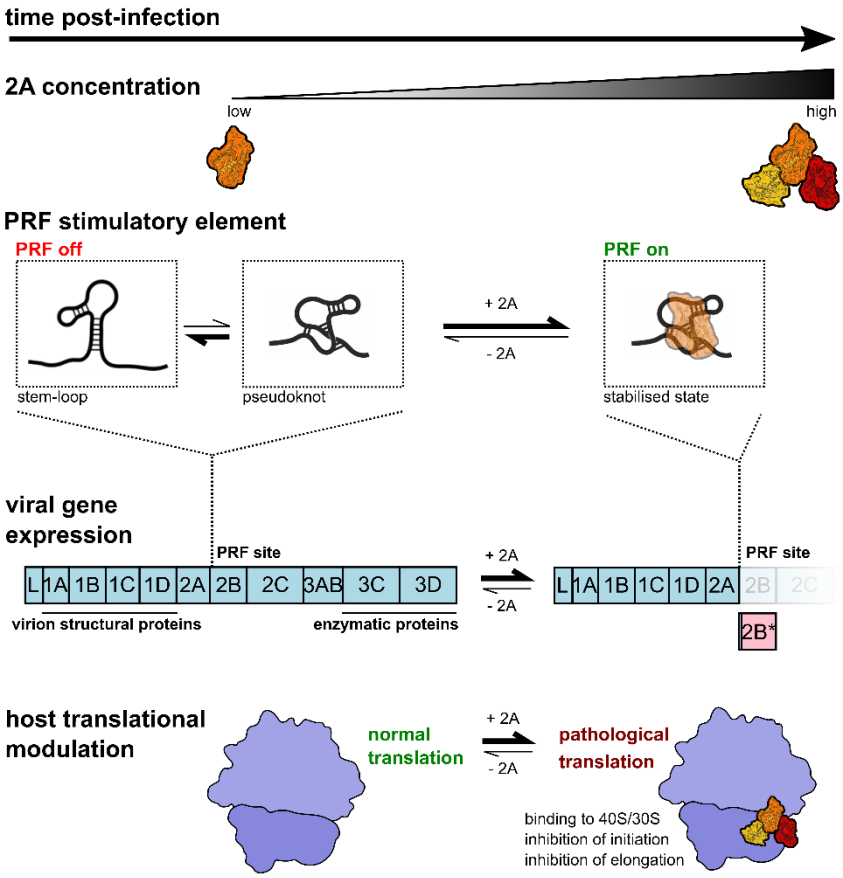


Figure 3. Molecular basis for 2A-induced reprogramming of gene expression. The PRF stimulatory RNA element is predicted to form either stem-loop or pseudoknot conformations. As 2A accumulates during EMCV infection, it selectively binds to and stabilises a pseudoknot-like conformation of the PRF stimulatory element, thereby enabling PRF, producing *trans*-frame product 2B* and downregulating the expression of enzymatic viral proteins later in infection. 2A also binds directly to the small ribosomal subunit at the translational GTPase factor binding site, progressively inhibiting both initiation and elongation as it accumulates. This may contribute to the shutdown of host cell translation during lytic infection.

Materials and methods

Microscale Thermophoresis (MST)

For RNA-binding experiments, synthetic EMCV RNA variants (**Supplementary Table 5**) were dissolved in distilled water and labelled at the 5' end with Dylight 650 maleimide conjugates (Thermo Scientific) using the 5′ EndTag kit (Vector Labs) as directed by the manufacturer. For each binding experiment, RNA was diluted to 10 nM in MST buffer (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT supplemented with 0.05% Tween 20) and a series of 16 tubes with 2A dilutions were prepared on ice in MST buffer, producing 2A ligand concentrations ranging from 0.00015 to 5 µM for EMCV RNA 2-6 and 0.00006 to 20 µM for EMCV RNA1. For the measurement, each ligand dilution was

mixed with one volume of labelled RNA, which led to a final concentration of 5.0 nM labelled RNA. The reaction was mixed by pipetting, incubated for 10 min followed by centrifugation at 10,000 × g for 10 min. Capillary forces were used to load the samples into Monolith NT.115 Premium Capillaries (NanoTemper Technologies). Measurements were performed using a Monolith NT.115Pico instrument (NanoTemper Technologies) at an ambient temperature of 25°C. Instrument parameters were adjusted to 5% LED power, medium MST power and MST on-time of 10 seconds. An initial fluorescence scan was performed across the capillaries to determine the sample quality and afterwards 16 subsequent thermophoresis measurements were performed. To determine binding affinities, data of at least two independently pipetted measurements were analysed for the fraction bound (MO.Affinity Analysis software, NanoTemper Technologies). For the non-binders, since the maximum amplitude would numerically be zero, deltaFnorm values were divided by the average maximum amplitude of the dataset to plot fraction bound. Data were fitted to the Kd model using MO.Affinity Analysis software (NanoTemper) and were plotted using Prism 8.0.2 (GraphPad).

Conjugation of a fluorescent label to the surface-exposed cysteine residue (C111) observed in the 2A crystal structure (**Fig. 1e**) provided a convenient way of studying binding to multiple unlabelled targets by MST, in such a way that the observed affinities would be directly comparable. For this experiment, EMCV 2A protein was labelled using the Protein Labelling Kit RED-Maleimide (NanoTemper Technologies) according to the manufacturer's instructions. In brief, 2A protein was diluted in a buffer containing 10 mM HEPES pH 7.9, 1.0 M NaCl and dye was mixed at a 1:3 molar ratio at room temperature for 30 min in the dark. Unreacted dye was removed on a spin gel filtration column equilibrated with 10 mM HEPES pH 7.9, 1.0 M NaCl. The labelled 2A protein was diluted to 10 nM in MST buffer. Synthetic EMCV RNA variants were used in dilutions ranging from 0.0008 to 26 μM for RNA 1 and 0.00003 to 1 μM for RNA 2-6. For the measurement, each RNA ligand dilution was mixed with one volume of labelled protein 2A, which led to a final concentration of protein 2A of 5.0 nM. Similar experiments were conducted with ribosomes in MST buffer, with ligand concentrations ranging between 0.00002 to 0.4 μM for 40S and 60S, 0.00003 to 1 μM for 30S, 0.000027 to 0.9 μM for 50S, 0.0008 to 1.375 μM for empty 70S and 0.000003 to 0.1 μM for 70S IC. The measurements were performed as described above.

Eukaryotic ribosomal subunit purification

40S and 60S subunits were purified from untreated rabbit reticulocyte lysate (Green Hectares) as previously described[322]. Briefly, ribosomes were pelleted by centrifugation (4°C, 270,000 × g, 4.5 h) and resuspended in 20 mM Tris-HCl pH 7.5, 4.0 mM $MgCl_2$, 50 mM KCl, 2.0 mM DTT. Following treatment with 1.0 mM puromycin and addition of KCl to 0.5 M, 40S and 60S subunits were separated by centrifugation (4°C, 87,000 × g, 16 h) through a sucrose density gradient (10 □ 30% sucrose in 20 mM Tris-HCl pH 7.5, 2.0 mM DTT, 4.0 mM $MgCl_2$, 0.5 M KCl). After analysis by SDS-PAGE, uncontaminated fractions were pooled, and exchanged into 20 mM Tris-HCl pH 7.5, 100 mM KCl, 2.0 mM $MgCl_2$, 2.0 mM DTT, 250 mM sucrose using Amicon centrifugal concentrators (4°C,100K MWCO). Ribosome subunits were snap-frozen in liquid nitrogen and stored at -80°C until further use.

Chapter 4

Screening of Natural Products and Small Molecules Uncovers Novel Coronavirus 1a/1b Frameshifting Inhibitors with Antiviral Properties

Anuja Kibe[1*], Walid A. M. Elgaher[2*], Ulfert Rand[3], Matthias M. Zimmer[1], Andreas M. Kany[2], Jennifer Hermann[2,4], Rolf Müller[2,4,5,6], Luka Cicin-Sain[3], Anna K. H. Hirsch[2,4,5,6†], Neva Caliskan[1,7†]

[1] Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Josef-Schneider-Strasse 2, 97080, Würzburg, Germany

[2] Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) - HZI, Campus Building E8.1, 66123 Saarbrücken, Germany

[3] Helmholtz Centre for Infection Research (HZI), Inhoffenstrasse 7, 38124, Braunschweig, Germany

[4] German Center for Infection Research (DZIF), Hannover-Braunschweig, Germany

[5] Helmholtz International Lab for Anti-infectives, Campus Building E8.1, 66123 Saarbrücken, Germany

[6] Department of Pharmacy, Saarland University, Campus Building E8.1, 66123 Saarbrücken, Germany

[7] Medical Faculty, Julius-Maximilians University Würzburg, 97080, Würzburg, Germany

** Authors contributed equally to this work*

[†]Corresponding authors:

Neva Caliskan neva.caliskan@helmholtz-hiri.de

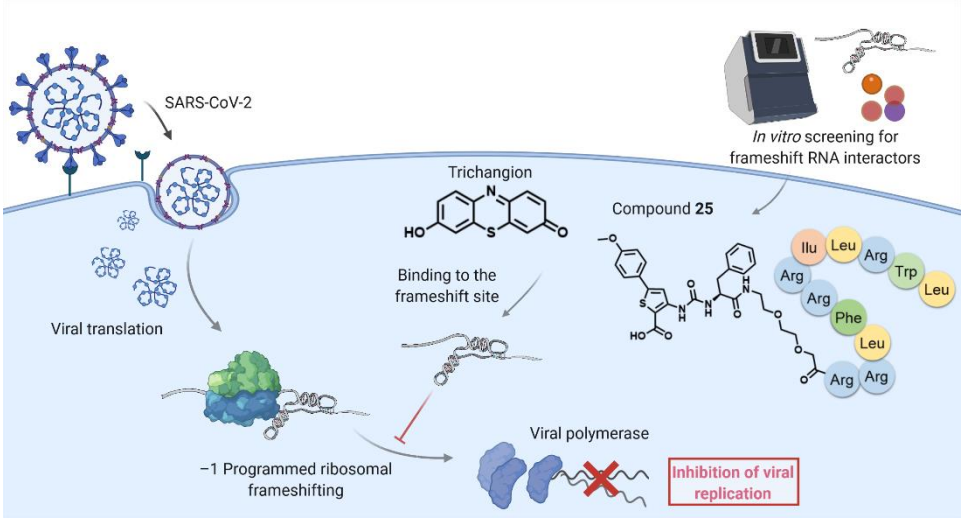
Anna K. H. Hirsch anna.hirsch@helmholtz-hips.de

Under review; Available as preprint (http://dx.doi.org/10.2139/ssrn.4157446)

Abstract

The novel severe acute respiratory syndrome-related coronavirus (SARS-CoV-2) and its emerging variants continue to pose a global health emergency. Like most coronaviruses, SARS-CoV-2 replication depends on programmed ribosomal frameshifting (PRF), which, due to its conserved nature, represents a promising antiviral target. Here, we employ a microscale thermophoresis (MST) assay to screen two diverse in-house libraries of natural products and synthetic small molecules for their frameshift RNA-binding properties. We identify trichangion and the ureidothiophene-coupled antimicrobial peptide **25** as potent binders. Using MST and surface plasmon resonance (SPR), we determine the binding affinity and interaction kinetics of the hit compounds. Moreover, we model the binding of the new PRF inhibitors into the structure of the SARS-CoV-2 RNA frameshift element, thus suggesting potential binding sites for RNA-targeting molecules. Furthermore, we demonstrate that these compounds substantially reduce frameshift levels *in vitro* and suppress viral propagation in SARS-CoV-2-infected cells. This work highlights the potential of natural products, especially the phenothiazine class, and synthetic amphiphilic cationic peptides to target frameshifting RNA elements, which sets the stage for the development of novel antivirals.

Graphical abstract



Introduction

The novel coronavirus SARS-CoV-2, the causal agent of coronavirus disease 2019 (COVID-19), and its emerging variants, have ravaged the global health and economy. Concerted efforts towards vaccine development and an amassed knowledge on viral RNAs, proteins and life cycle have been instrumental to the success of the current mRNA-based vaccines. However, the global spread of recent variants -Delta and Omicron-, in which acquired mutations enhance the viral transmission rates and ability to evade the hosts' immune system, attests to the reality that emerging novel viruses will continue to pose a challenge [333, 334]. Moreover, coronaviruses are known to persist within several wildlife hosts [335]. Population growth, globalization and changes to the ecosystems enhance the number of interspecies contacts creating an imminent threat of zoonotic transfer and novel viral epidemics [336]. Thus, there is an urgent need for new antivirals that target conserved elements of the coronavirus genome, rendering them effective against recurrent outbreaks as well as emerging virus variants.

Coronaviruses are single-stranded, positive-sense RNA viruses. The ~30 kb RNA genome has at least nine different open reading frames, of which ORF1a/b is the most prominent occupying nearly two-thirds of the genome [337]. A hallmark of many RNA viruses, including SARS-CoV-2, is the highly conserved gene expression event- –1 programmed ribosomal frameshifting (–1PRF). In coronaviruses, –1PRF on the 1a/1b gene is essential for the expression of ORF1b, which encodes proteins responsible for replication and transcription of the viral genome. Here, frameshifting is directed by two crucial elements separated by a spacer- a slippery heptameric sequence UUUAAAC and a downstream stable RNA secondary structure, which forms a pseudoknot **(Figure 1B)** [338]. Importantly, this so-called frameshift stimulatory element (FSE) is mostly conserved among different betacoronaviruses and several emerging SARS-CoV-2 variants [339, 340]. Mutations and drugs that affect –1PRF levels in SARS-CoV-2 and several other viruses such as HIV-1 and SARS-CoV were shown to impede viral propagation [121, 341-345]. Due to the functional importance and high conservation of this RNA element, the frameshift site of coronaviruses is regarded as a novel target for interventions by proteins [104] or small molecules [114, 116, 119, 346-350].

Given the high relevance of the FSE, it is necessary to search for further novel ligands targeting FSE that can act as future therapeutic templates. Previous work focused on using commercial libraries and on repurposing of FDA-approved drugs to target the conserved

frameshift element [119-121, 346, 351]. One such molecule, MTDB, was shown to be resistant to mutations in the novel coronavirus [120, 273], however another study did not observe any MTDB induced specific frameshift inhibition [344]. This molecule is proposed to interact with the frameshift site, albeit with very low affinity (210±20 μM) and at high concentrations affects the conformational plasticity of the –1PRF site [352]. On the other hand, molecules like merafloxacin, nafamostat and KCB261770 are proposed to be effective against several coronaviruses [119-121]. However, the exact mechanism of these molecules and whether or not they directly interact with the FSE has not yet been studied.

Natural products from microorganisms, fungi, plants and insects possess large structural and chemical diversity, and they continue to inspire new drug discovery [353]. Currently, approximately 53% of anti-infective compounds and 79% of anticancer pharmaceutical agents in clinical use are natural products or their derivatives [354]. Additionally, several studies reported plant products and secondary metabolites that target various aspects of SARS-CoV-2 replication and inhibit infection [355, 356]. Nevertheless, potential of natural compounds to target RNA molecules is vastly unexploited. So far, natural products such as macrolides, aminoglycosides, tetracyclines and chloramphenicol have been identified to target bacterial rRNA [357, 358]. However, there has not been a systematic screen of natural product libraries to assess their effect on viral RNA molecules.

This prompted us to screen diverse classes of compounds ranging from natural products to synthetic small molecule libraries for interactors of the highly conserved SARS-CoV-2 frameshifting RNA element in order to identify modulators of viral replication. Our microscale thermophoresis (MST)-based screening revealed two hits (trichangion and compound **25**) as the first natural product and peptide, respectively, interacting with the SARS-CoV-2 FSE. Furthermore, our studies show that these compounds inhibit SARS-CoV-2 frameshifting by approximately 50%, while not altering overall translation *in vitro*. Treatment of SARS-CoV-2-infected cells with these molecules significantly reduced viral propagation. These findings pinpoint phenothiazines and amphiphilic-cationic peptides as potential frameshifting inhibitors, which may assist the development of novel, highly effective antiviral drugs in the future.

Results and Discussion

MST-Based Screening Identifies SARS-CoV-2 Frameshift RNA Interactors

In order to identify new compounds that interact with the SARS-CoV-2 frameshift RNA, we established a high-throughput MST-based screen (**Figure 1A)**. The MST technique measures the change in fluorescence intensity of the macromolecule (F_{norm}) due to migration along a heat gradient. Upon binding of a ligand, the thermophoretic movement of the macromolecule will change, resulting in a difference in the fluorescence intensity, which we refer to as ΔF_{norm}. We employed a 3′Cy5-labeled minimal mRNA pseudoknot motif derived from nucleotides 13403–13516 of the SARS-CoV-2 genome as the bait for this *in vitro* screen (**Figure 1B**). We tested a collection of 481 compounds comprising 385 natural products with diverse chemical structures and biological activities isolated from myxobacteria and fungi and 96 synthetic small molecules from our in-house library of anti-infective agents. We placed an arbitrary ΔF_{norm} cut-off of 5 to define hit compounds with a significant interaction with the FSE. Using this filter, 458 molecules showed no or a minor change in thermophoresis (ΔF_{norm} <5) (**Figure 1C**). Out of 23 hits, two molecules were strong interaction partners of the SARS-CoV-2 pseudoknot and were chosen for further characterization, namely the natural product trichangion (ΔF_{norm} = 34 ± 7) and peptide compound **25** (ΔF_{norm} = 129 ± 2).

Trichangion is a 7-hydroxyphenothiazin-3-one, which was isolated from the Gram-negative myxobacterium *Sorangium cellulosum* **(Figure S1A)**. It was the only molecule with an intrinsic fluorescence property, hence the ΔF_{norm} value was corrected for background thermophoresis of the molecule **(Figure 1C)**. Intriguingly, trichangion and other myxobacterial secondary metabolites were shown to reduce the formation of processing bodies (P-bodies), cytoplasmic mRNA–protein domains, but their mode of action remains unknown [359]. The second hit, compound **25**, belongs to a series of *Pseudomonas aeruginosa* quorum sensing inhibitors [360]. Compound **25** consists of a hydrophobic ureidothiophene-2-carboxylic acid linked to a cationic undecapeptide (RRLFRRILRWL) **(Figure S1B)**. The latter is a known antimicrobial peptide (AMP) [361].

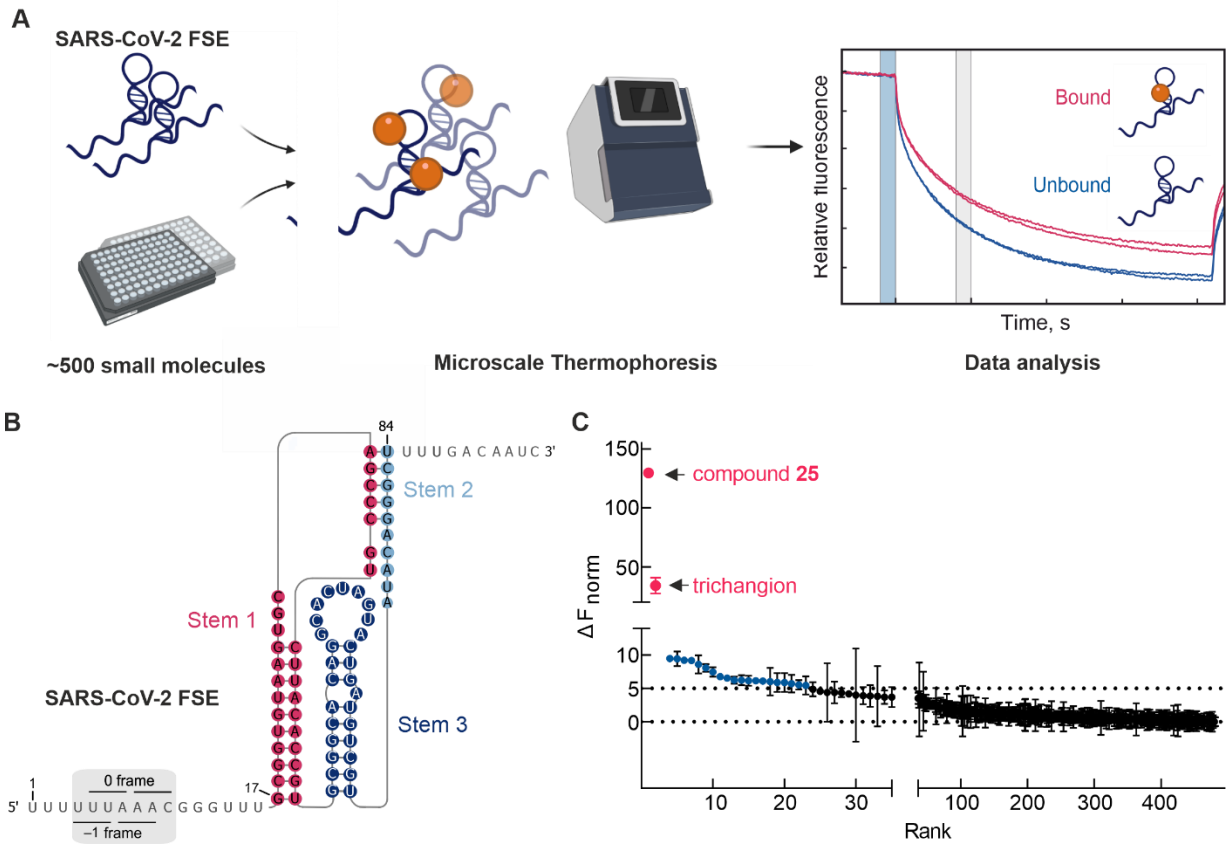


Figure 1. Identification of SARS-CoV-2 frameshift interactors using *in vitro* MST screening. (A) Schematic of the MST screening to identify potential small-molecule interactors. Compounds (100 µM) were incubated with 3' pCp-Cy5-labeled RNA (5 nM) and thermophoresis was recorded at 25 °C with 5% LED intensity and medium MST power. Fluorescence (F_{norm}) was measured at MST on-time of 5 s. Blue and gray boxes in the time-course traces represent the temperature jump and MST-on time (5 s), respectively. (B) Schematic representation of the relevant genomic segment of SARS-CoV-2, corresponding to nucleotides 13456–13570 of the SARS-CoV-2 RNA genome numbered as 1–84 as well as the proposed structure of FSE. (C) Change in fluorescence (ΔF_{norm}) values of all the tested compounds, the strongest interaction partners trichangion and compound **25** are highlighted in pink, molecules with $\Delta F_{norm} > 5$ are highlighted in blue. ΔF_{norm} was calculated as the absolute value of ($F_{norm(RNA)}$- $F_{norm(RNA+molecule)}$), where $F_{norm(RNA)}$ denotes F_{norm} of the labeled RNA in presence of DMSO and $F_{norm(RNA+molecule)}$ denotes fluorescence of labeled RNA in presence of molecule. Data represent the average of two independent experiments and mean values ± s.d. are shown (see also Methods). Dotted lines represent $\Delta F_{norm} = 0$ and 5. See also **Figure S1**.

Binding Affinity, Kinetics, and Selectivity of the Hit Molecules

Having identified trichangion and compound **25** as ligands of SARS-CoV-2 FSE, we determined the binding affinities, kinetic parameters and selectivity of the hit compounds using a four-channel SPR system as an orthogonal label-free approach to MST. For this purpose, we prepared the minimal SARS-CoV-2 FSE RNA (**Figures 2A and 3A**) with a biotin-tag at the 3′ end. In addition, we also tested biotinylated random RNA (GAGA array) as well as a biotinylated ssDNA oligonucleotide as negative controls. The three constructs were immobilized onto a streptavidin-coated sensor chip via affinity-based capture on channels 1–3, respectively, resulting in immoblization levels of about 1000 resonance unit (RU) (**Figure S2**) [345]. Both trichangion and compound **25** displayed a concentration-dependent binding response upon injection over SARS-CoV-2 frameshift RNA with an equilibrium dissociation constant (K_D) of 16 μM and 1.4 μM and binding response of 20 and 4 RU, respectively (**Figures 2B, 3B** and **Table S1**). Markedly, compound **25** showed more than two orders of magnitude faster association rate (k_a) compared to trichangion, which to some extent could be explained by the multicationic nature of the compound (**Table S1**). On the other hand, trichangion exhibited a 20-fold slower dissociation rate (k_d) and longer residence time (τ) on SARS-CoV-2 FSE compared to compound **25** (**Table S1**). Owing to the fast kinetics of compound **25**, we calculated the equilibrium affinity of both hit compounds by plotting the steady-state response against concentration. The resulting affinity (K_D) values were concordant with those estimated from the association and dissociation kinetics **(Figure S3A,B)**. Additionally, trichangion and compound **25** bound the random RNA, but with lower binding response (RU) (13 and 3 RU respectively) and affinity (K_D of 25 and 2.2 μM respectively) (**Figures S3C, D, S4A, C** and **Table S1**). Intriguingly, both compounds show no significant binding to the ssDNA oligonucleotide (**Figures S3E, F S4B, D and Table S1**).

For the binding affinity measurements of compounds using MST, the affinity curves were conducted in presence of 0.2 mg/mL yeast tRNA to avoid the liability of non-specific interactions to the compounds. The K_D values of trichangion and compound **25** were in the low micromolar range of 26.8 ± 11 μM and 6.6 ± 1 μM , respectively (**Figures 2C, 3C and Table S2**). We also tested the binding specificity of the molecules to the structured pseudoknot by analyzing their interactions with a stem-loop truncation variant as well as the random RNA. Trichangion binding was negligible (K_D >200 μM) upon deletion of both SL2 and SL3 (ΔSL2 + 3), a 48 nt sequence that is predicted to fold into a short double-

stranded stem-loop (SL1) (**Figure 2D and Table S2**). Additionally, trichangion did not interact with the random RNA (**Figure 2E and Table S2)**. Compound **25** interacts with both the ΔSL2 + 3 mutant as well as the random RNA, albeit with much lower affinity (K_D = 37.6 ± 2 µM and K_D = 60.9 ± 8 µM respectively) **(Figure 3D, E and Table S2).** Overall, our MST and SPR results suggest that trichangion and compound **25** preferentially interact with the structured FSE RNA in the low micromolar range.

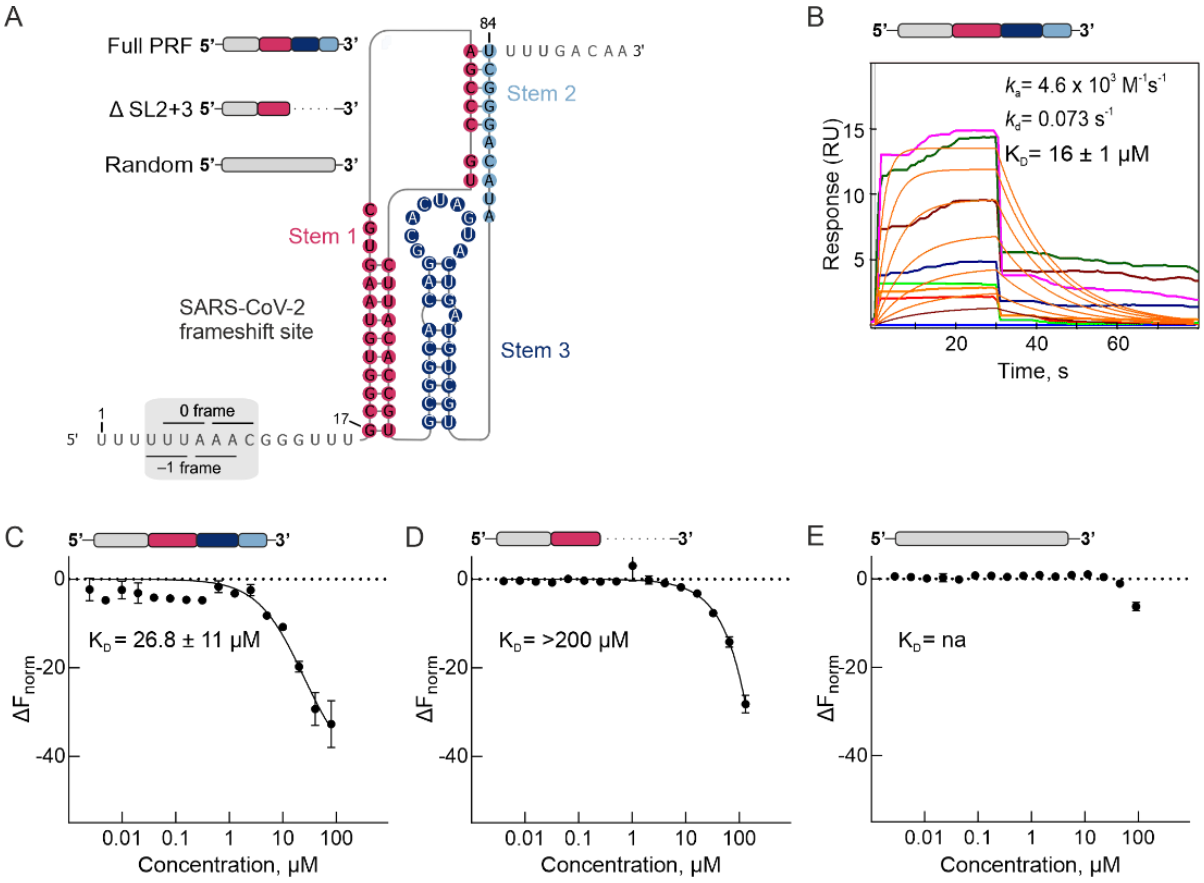


Figure 2. Determination of binding affinities of trichangion using SPR and MST. (A) Proposed structure of the PRF element of SARS-CoV-2 and schematic representations of the RNAs studied. (B) Interaction of trichangion with SARS-CoV-2 FSE detected via SPR. Interactions of trichangion detected via MST with (C) SARS-CoV-2 FSE (D) ΔSL2 + 3 mutant and (E) Random RNA. For SPR, sensorgrams overlay of trichangion or compound **25** injected at concentrations of 1.56–100 µM over a streptavidin–biotin-captured FSE are shown. For MST of trichangion, SARS-CoV-2 FSE (2 nM–80 µM), ΔSL2 + 3 mutant (4 nM–130 µM) and random RNA (2 nM – 90 µM) was titrated against 5 nM of the compound. In all measurements, thermophoresis was measured at 5% LED intensity, 25 °C, medium MST power and on-time of 1.5s. Data represent the average of two independent experiments and mean values ± s.d. are shown. For both MST and SPR, binding kinetics and affinity values were calculated by global fitting of the association and dissociation curves to a 1:1 binding model. See also Figures S2, S3 and Tables S1, S2.

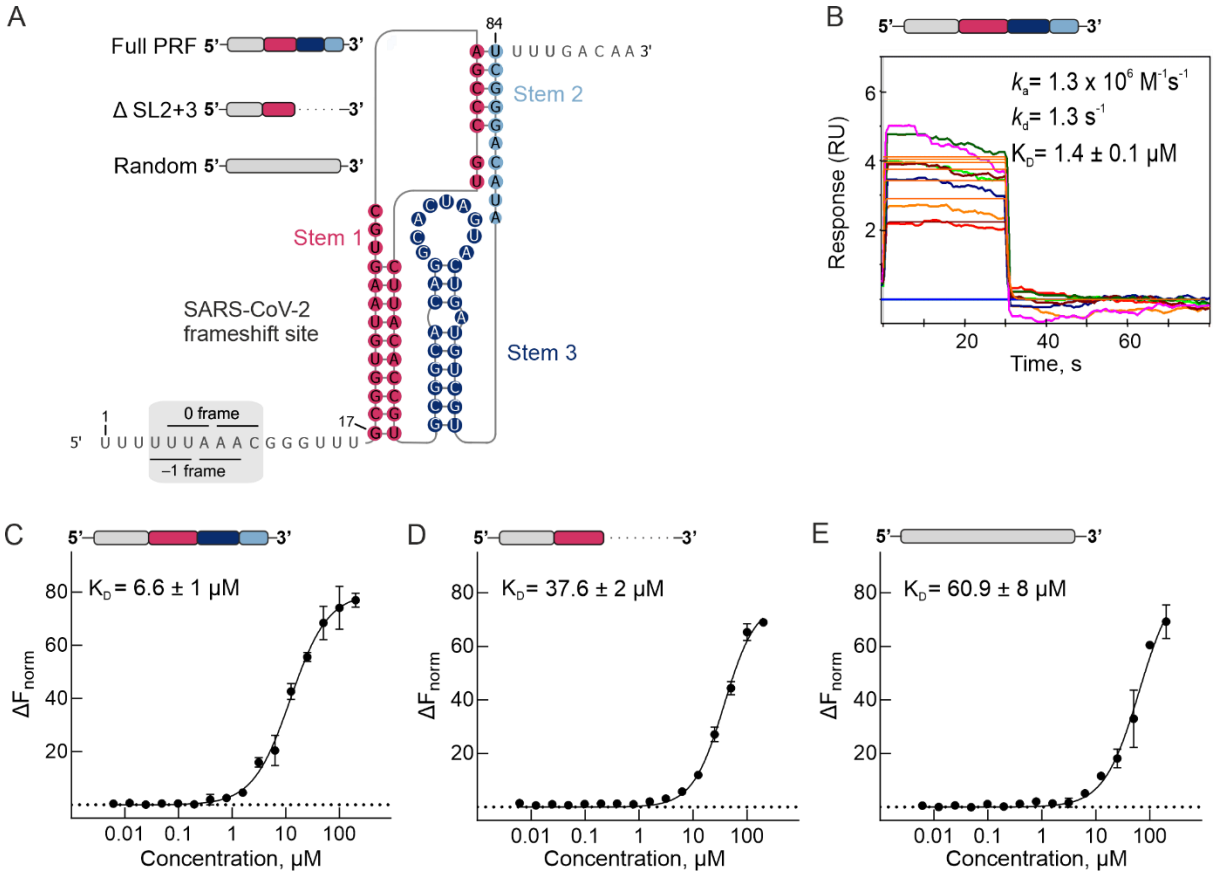


Figure 3. Determination of binding affinities of compound **25** using SPR and MST. (A) Proposed structure of the PRF element of SARS-CoV-2 and schematic representations of the RNAs studied. (B) Interaction of compound 25 with SARS-CoV-2 FSE detected via SPR. Interactions of compound 25 detected via MST with (C) SARS-CoV-2 FSE (D) ΔSL2 + 3 mutant and (E) Random GAGA-repeat RNA. For SPR, sensorgrams overlay of trichangion or compound **25** injected at concentrations of 1.56–100 µM over a streptavidin–biotin-captured FSE are shown. For MST of compound **25**, molecule (3 nM–200 µM) was titrated against 5 nM of 3' pCp-Cy5 labeled RNAs (5 nM). In all measurements, thermophoresis was measured at 5% LED intensity, 25 °C, medium MST power and on-time of 1.5s. Data represent the average of two independent experiments and mean values ± s.d. are shown. For both MST and SPR, binding kinetics and affinity values were calculated by global fitting of the association and dissociation curves to a 1:1 binding model (see Methods). See also Figures S2, S3 and Tables S1, S2.

Binding Site Detection for SARS-CoV-2 Frameshift RNA and Molecular Docking of the Hit Molecules

To get an insight into the mode of interaction of trichangion and compound **25** with the frameshift element, we performed molecular docking using two resolved conformations of the SARS-CoV-2 frameshift RNA, namely the fork-shaped structure (PDB ID: 6XRZ)[362]

and the linear topology (PDB ID: 7MLX) [363]. Both RNA structures were captured in the apo state, therefore we first explored putative druggable binding sites within the two conformations.

The fork-shaped tertiary structure determined by cryo-EM comprises the slippery sequence and spacer at the 5′ end followed by a three-stem pseudoknot (**Figure 4A**). It is characterized by an inclination of stem 3 (magenta) at stem 1 (gold) by about 50°. In addition, the 5′ end is threaded through a ring formed by the second strand of stem 1, stem 2 (green), stem 3, and the unpaired segment (J3/2) (**Figure 4A**). The linear conformation determined by X-ray crystallography represents a minimal construct comprised of a three-stem pseudoknot, lacking the slippery site and spacer. It also shows threading of the 5′ end of stem 1 through a ring formed between the second strand of stem 1 (gold) and loop 3 (**Figure 4B**).

Using the alpha site finder method [364] implemented in the molecular operating environment (MOE) software, we identified four potential ligand-binding sites (I–IV) in the fork-shaped structure (**Table S3** and **Figure S5**). Site I is defined mainly by the nucleotide residues in the major groove of stem 1 and extends to residues from the slippery site and the AUACA residues of the J3/2 site (**Table S3** and **Figure S5A**). This region possesses the largest size and the highest number of hydrophobic contacts (**Table S3**). Site II is delineated by the major groove residues of stem 3 and has a slightly smaller size and less hydrophobicity as compared to site I (**Table S3** and **Figure S5B**). Site III is located in the major groove of stem 2; it is about one third of the size and hydrophobicity of site I and has a more solvent-exposed surface (**Table S3** and **Figure S5C**). Site IV is formed by the converging nucleotides of stem 1, stems 2, stem 3, and the J3/2 segment, shaping a ring structure (**Table S3** and **Figure S5D**). It represents a narrow and short tunnel in the pseudoknot, which may accommodate hydrophobic planar small molecules such as trichangion.

For the linear three-stem pseudoknot topology, two putative binding sites (I and II) were predicted (**Table S3** and **Figure S6**). Site I is located in the major groove, spanning stem 1, stem 2 and loop 3 (**Table S3** and **Figure S6A**). It represents a merged version corresponding to sites I, III, and IV in the forked conformation. Site II is composed of the major groove residues of stem 3, similar to site II in the forked structure (**Table S3** and **Figure S6B**). Interestingly, alignment of the two FSE conformations revealed similar architecture and good superposition at stem 1 (site I) and stem 3 (site II) (**Figure 4C**).

Given the large concavity and hydrophobicity of site I, which includes stem 1 and the J3/2 site (loop 3 in the linear conformation), and due to its high sequence conservation among coronaviruses and its crucial role for –1PRF, this region is most likely to be the primary binding site for FSE-targeting inhibitors [93, 362]. Consequently, we docked trichangion and compound **25** into site I of the forked frameshift RNA structure. The compounds displayed binding energies of –3 and –15 kcal/mol, respectively, in line with the experimental affinity values determined by MST and SPR.

At pH 7.4, trichangion predominates as a zwitterion with a delocalized negative charge on the oxygen atoms at C3 and C7 and a positive charge on N10. Binding is triggered by electrostatic attractions (hydrogen bonding and ion interaction) between the heteroatoms of trichangion and the phosphate groups of U35, C36, and U37 at stem 1 and C4-NH_2 of C77 residue (C13536 in the full-length genome) at J3/2 site **(Figure 4D** and **S7)**. It is noteworthy that the phenothiazine core of trichangion is structurally related to the acridines, which are known nucleic acids binders [365, 366]. Accordingly, it is plausible that trichangion may bind the FSE at other smaller binding pockets, *e.g.*, site IV (ring site) or through intercalation between the base pairs. Docking into site IV indicated that trichangion optimally fits into the ring hole with a binding energy of –4 kcal/mol. The tricyclic molecule establishes hydrophobic and electrostatic interactions with G18 and G44 of stem 1, and A76 of J3/2 region (**Figure S8**).

Docking of compound **25** into site I showed that the conjugated peptide entirely occupies the major groove of stem 1 owing to its flexibility and its electrostatic and hydrophobic complementarity to the RNA surface (**Figure 4E**). The polycationic peptide establishes multiple hydrogen bonds between its arginine residues and the phosphate backbone of stem 1 as well as the C=O of C77 and N3 of G79 at the J3/2 site. Additionally, compound **25** is engaged in hydrophobic interactions through the tryptophan residue and ureidothiophene moiety with A23 and C43, respectively **(Figure S9)**. The predicted contribution of the ureidothiophene fragment to binding and its structural similarity to the –1PRF inhibitor MTDB indicate a beneficial role of small lipophilic motifs for FSE binding besides the electrostatic interactions [93, 117].

Altogether, these data suggest that binding of trichangion and compound **25** to site I might stabilize the threaded frameshift element conformation, leading to an interference with unfolding and refolding of the structure. Alternatively, intercalation (trichangion) or

binding of a bulky ligand (compound **25**) would significantly affect the overall topography of the frameshift structure, which may result in formation of a –1PRF-incompetent unthreaded pseudoknot conformation, or even an inhibition of the interaction with the ribosome [362].

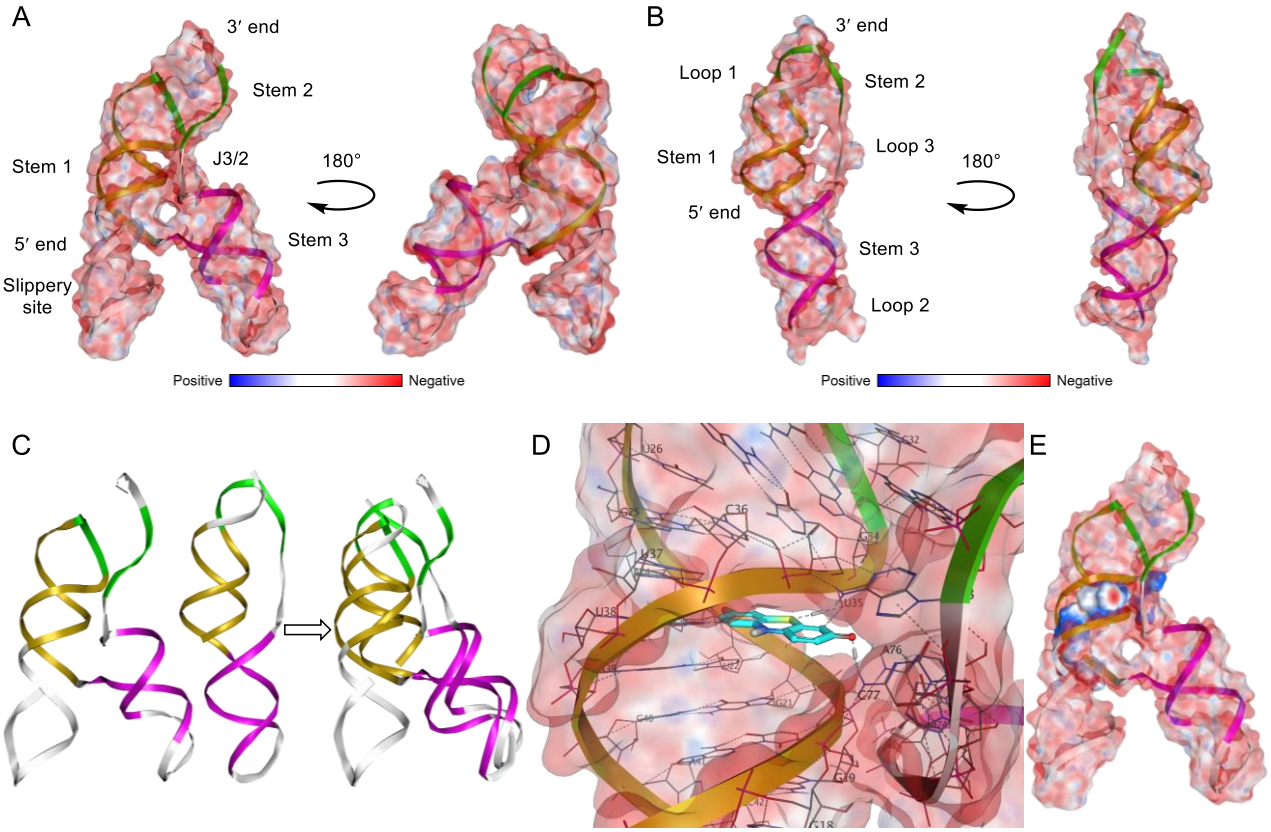


Figure 4. Electrostatic surface and ribbon representations of SARS-CoV-2 FSE in (A) fork-shaped (PDB ID: 6XRZ) and (B) linear (PDB ID: 7MLX) conformations. The three-stem pseudoknot is colored in gold (stem 1), green (stem 2), and magenta (stem 3). (C) Structure alignment of the CEM structure (fork-shaped) and the X-ray structure (linear) showing marked superposition at stem 1 and stem 3. (D) Predicted binding mode of trichangion (cyan) into site I through electrostatic interactions (dashed lines) with nucleotides of stem 1 and J3/2 region. (E) Electrostatic surface presentation for the putative binding mode of compound **25** into the major-groove of stem 1 and J3/2 site, showing marked shape, hydrophobicity and electrostatic complementarity (see also **Figure S9**).

Trichangion and Compound 25 Reduce Frameshifting Efficiency of SARS-CoV-2 *In vitro* and Impede Viral Replication

For functional effects, it is important that the interaction of the compounds with the SARS-CoV-2 FSE leads to a change in the frameshifting efficiency of the viral RNA genome. To test this, we used the *in vitro* rabbit reticulocyte lysate (RRL)-based translation assay and monitored translation of the SARS-CoV-2 FSE in the presence of the two hit molecules. We employed a reporter mRNA containing nucleotides 12686–14190 of the SARS-CoV-2 genome to best mimic the native genomic context (**Figures 5A**). SARS-CoV-2 frameshifting efficiency (FE) was about 25% in the presence of DMSO, as described previously [121]. At increasing concentrations of both molecules, we observed a corresponding decrease in frameshifting (**Figures 5B-E and Table S4**). For trichangion, the FE inhibitory effect starts to plateau at 111 μM and at the highest concentration shows frameshift inhibition of ~50% (**Figure 5D**). Similarly, for compound **25**, a 50% frameshift inhibition is observed at the highest concentration of 167 μM with the inhibitory effect plateauing at 74 μM (**Figures 5E**). These results indicate that trichangion (IC_{50}=30 μM) and compound **25** (IC_{50}=11 μM) can reduce ribosomal FE *in vitro*, with compound **25** displaying more potency (**Figure S10C**), which is consistent with their RNA binding affinities. Importantly, reduction in viral replication does not require complete inhibition of frameshifting – any substantial change in the stoichiometric ratios of structural and enzymatic proteins (the latter are produced via –1PRF) can interfere with the replication process [117, 119, 121, 341, 367, 368]. Notably, our hit compounds did not substantially decrease the product translated by the 0-frame of the reporter mRNA, indicating that these molecules are specific to the –1PRF, and do not influence general translation, initiation or termination processes.

Encouraged by these results, we investigated the effect of our hit compounds on SARS-CoV-2 replication in infected Huh-7 cells, a human hepatocyte cell line which has been extensively used for SARS-CoV-2 research [369]. We tested a concentration of 50 μM for both molecules, which showed ~40% reduction in frameshifting in the *in vitro* assays and was above the IC_{50} of frameshift inhibition. Trichangion reduced the viral load by 10 fold at 50 μM concentration, similar to the interferons IFN-β and IFN-γ (**Figure S10A**). Compound **25** reduced the viral load by ~7 fold (**Figure S10B**). The smaller effect on viral titer as compared to the prominent effect on FE of compound **25** might be caused by inadequate cellular permeation, amongst other factors.We also evaluated the compounds

for cytotoxicity and both compounds did not show a significant reduction of cell viability at concentration of 50 µM (**Figure S10B**). Encouraged by the stronger viral inhibition by trichangion, we proceeded to determine the IC_{50} by titrating the compound in A549 lung cells constitutively expressing ACE2 and TMPRSS2 infected with a mNeonGreen SARS-CoV-2 reporter virus [370, 371] . Trichangion significantly impeded viral replication at 39 µM and completely inhibited the virus at higher concentrations (**Figure 5F**). We observed an IC_{50} of ~47 µM (**Figure S10D**). Moreover, trichangion did not cause a dose-dependent cell death (as estimated by cellular confluence) and at the highest concentration was comparable to that of untreated cells (**Figure 5G**).

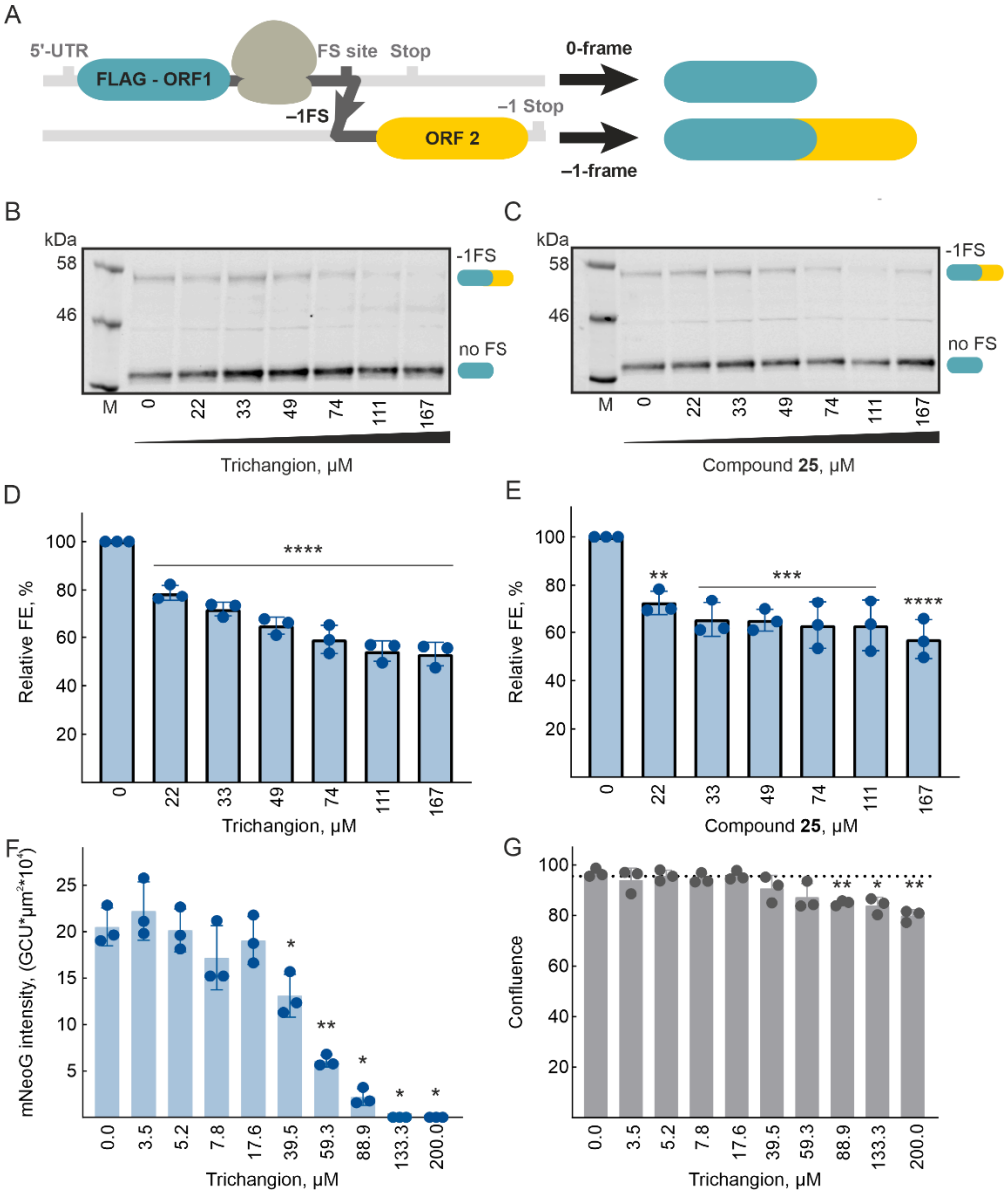


Figure 5. Effect of trichangion and compound 25 on SARS-CoV-2 –1 frameshifting *in vitro* and viral replication *in vivo.* (A) Schematics of the N-terminal FLAG-tagged frameshifting reporter consisting of the

nucleotides 12686–14190 (~1.5 kb) of the SARS-CoV-2 genome. RNAs were translated in RRL in the presence of increasing concentrations of the molecule ranging from 0 to 167 μM. An equal amount of DMSO was added to the no compound (0 μM) control. FLAG-tagged peptides generated by ribosomes that do not frameshift (no –1PRF) or that enter the −1 reading frame (−1PRF) were identified via Western blotting using the anti-DDDDK antibody. FE was calculated as previously described [104], by the formula: FE=Intensity (–1-frame)/ (Intensity (–1-frame) + Intensity (0-frame)). 0-frame and –1PRF FLAG-tagged products identified via Western blotting in the presence of increasing concentrations of (B) trichangion and (C) compound **25**. Changes in FE observed in the presence of (D) trichangion and (E) compound **25** calculated from (B) and (C), respectively, and normalized to responses at 0 μM of the compounds. (F) Changes in viral replication upon trichangion titration in A549-AT cells infected with icSARS-CoV-2-mNeonGreen. Integrated intensity of green fluorescence originating from the reporter virus was measured. (G) Cell confluency measured in parallel in infected A549-AT cells in presence of trichangion and normalized to the DMSO (no-compound) control. For all assays, boxes show mean values ± s.d. (*n* = three independent experiments). *P* values were calculated using an ordinary unpaired one-sided ANOVA comparing every concentration to the no compound control (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$). See also **Figure S10** and **Table S4**.

Conclusions

Programmed ribosomal frameshifting plays an indispensable role in the expression of the RNA-dependent RNA polymerase (RdRp) and other vital proteins of the coronaviruses. In this study, we discovered the first naturally occurring compound (trichangion) and the first peptide (compound **25**) that can selectively bind to the frameshift element of SARS-CoV-2 and impede its function in cells in the low micromolar range. Binding site exploration and structural docking studies indicate that the hit compounds most probably bind in the major groove of stem 1 and J3/2 site (loop 3) as the primary binding site (I). Moreover, the driving force for binding might be a combination of electrostatic and hydrophobic interactions, in addition to an intercalating property for trichangion. Furthermore, we show that these compounds reduce propagation of live SARS-CoV-2 virus with no signs of cytotoxicity under the conditions tested. Although, the potency range of these hit molecules is moderate for potential drug candidates, a major advantage of a small and simple natural product such as trichangion is its structural malleability for optimization of activity and drug-like properties. Another strategy for optimization could be the hybridization of trichangion and compound **25** by replacing the ureidothiophene motif for trichangion in the Arg-rich peptide, which would benefit from both modes of interaction and was shown to improve affinity and selectivity for RNA binding [372].

Overall, our study pioneers the use of natural compounds as RNA binders affecting frameshifting and illustrates that the viral frameshift site is a valid target not only for small

molecules but also for natural compounds. It highlights the phenothiazine-based trichangion and the ureidothiophene-linked cationic peptide (compound **25**) as promising scaffolds for developing –1PRF inhibitors. While much more effort is required for drug development, we hope that this work serves as a valuable starting point and demonstrates the use of high-throughput screening of (natural) compound libraries not only against SARS-CoV-2 but also future coronavirus variants.

Materials and methods

Microscale Thermophoresis

Short frameshifting RNA constructs was *in vitro* transcribed using T7 polymerase purified in-house using PCR fragments as the template. RNAs were labeled at the 3' end using pCp-Cy5 (Cytidine-5'-phosphate-3'-(6-aminohexyl) phosphate) (Jena Biosciences). For the initial screening of molecules, RNA was folded by heating at 70^{o} C for 5 min, followed by snap-cooling on ice and diluted to 10 nM in the MST buffer (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$ supplemented with 0.05% Tween-20, RNAse inhibitor (Molox) and 0.2 mg/mL yeast tRNA). The small molecules were diluted to 100 μM in the above buffer albeit with 10% DMSO. 10 μL of the RNA was mixed with 10 μL of the diluted compound to yield final concentration of 5 nM and 50 μM of the RNA and compound, respectively in MST buffer with 5% DMSO. The reaction was mixed by pipetting, incubated for 10 min at room temperature, followed by centrifugation at 10,000 × *g* for 5 min. Capillary forces were used to load the samples into Monolith NT.115 Premium Capillaries (NanoTemper Technologies). Measurements were performed using a Monolith Pico instrument (NanoTemper Technologies) at an ambient temperature of 25 °C. Instrument parameters were adjusted to 5% LED power, medium MST power, and MST on-time of 5 seconds. Data of two capillaries was used to determine the F_{norm} and error using the MO. Affinity Analysis software (NanoTemper Technologies) and ΔF_{norm} was calculated using the formula ($F_{norm(RNA)}$- $F_{norm(RNA+molecule)}$ ± SQRT (($Error_{(RNA)}$)^2 +($Error_{(RNA+molecule)}$)^2). For determination of binding affinity, all RNAs were folded and diluted as described above. A series of 16 tubes with compound **25** were prepared in MST buffer, producing ligand concentrations ranging from 6 nM to 200 μM. For measurements, each ligand dilution was mixed with one volume of labeled RNA, which led to a final concentration of 5 nM labeled RNA and 3 nM to 100 μM compound **25**. The reactions were incubated and loaded as described above. With trichangion, experiment was similarly

performed albeit unlabeled SARS-CoV-2 RNA was titrated, which led to a final concentration of 5 nM trichangion and 2 nM–80 µM of SARS-CoV-2 FSE, 4 nM- 130 µM of ΔSL2 + 3 mutant and RNA 2 nM – 90 µM of random RNA. An initial fluorescence scan was performed across the capillaries to determine the sample quality and afterwards, 16 subsequent thermophoresis measurements were performed. All measurements were performed using a Monolith Pico instrument (NanoTemper Technologies) at an ambient temperature of 25 °C, instrument parameters were adjusted to 5% LED power, medium MST power and MST on-time of 1.5 seconds. Data of two independently pipetted measurements were analyzed for the ΔF_{norm} and binding affinity values determined by the MO. Affinity Analysis software (NanoTemper Technologies). Graphs were plotted using GraphPad Prism 9.2.0 software.

Surface Plasmon Resonance (SPR)

The SPR binding studies were performed using a Reichert 4SPR surface plasmon resonance spectrometer (Reichert Technologies, Buffalo, NY, USA), and medium density streptavidin-derivatized carboxymethyldextran hydrogel SAD500M sensor chips (XanTec Bioanalytics, Düsseldorf, Germany). Tris-buffered saline with Tween 20 (TBST) buffer (10 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, 0.05% v/v Tween-20, pH 7.8), containing 1% v/v DMSO and 0.001% v/v RNase inhibitor (Molox), was used as the running buffer for polynucleotides capturing and binding studies. All running buffers were filtered and degassed prior to use. The 3′-biotin-labeled SARS-CoV-2 RNA FSE, random RNA, and ssDNA oligonucleotide were captured onto flow cells 1–3, respectively, according to the standard protocol provided by Xantec bioanalytics. The flow cell 4 was left blank to serve as a reference. The system was initially conditioned with activation buffer I (50 mM NaOH and 1 M NaCl), TBST buffer, activation buffer II (6 M guanidine-HCl, pH 1.5), and TBST buffer (10 min each), respectively. The 3′-biotinylated polynucleotides at a concentration of 167 nM in the running buffer were injected at a flow rate of 10 µL/min for 7 min. Unoccupied streptavidin surface was blocked by 10 µg/mL biotin solution in the running buffer at a flow rate of 30 µL/min for 2 min. Nine buffer blanks were injected over both reference and active surfaces to attain stable immobilization levels of approximately 1000 micro refractive index unit (µRIU). Binding experiments were performed at 20 °C. Compounds dissolved in DMSO were diluted with TBST buffer to a final concentration of 100 µM with 1% v/v DMSO. Seven concentrations of the compounds (100–1.56 µM) were prepared by two-fold serial dilutions and were injected at a flow rate of 50 µL/min. The

association time was set to 30 s, and the dissociation phase was recorded for 60 s. Single-cycle kinetics were applied for K_D determination. 1 M NaCl in the running buffer was used for surface regeneration. Differences in the bulk refractive index due to DMSO were corrected by a calibration curve (seven concentrations: 0.25–2% v/v DMSO in TBST buffer). Experiments were performed in duplicate. Data processing and analysis were performed by Scrubber software (Version 2.0c, 2008, BioLogic Software). Sensorgrams were double referenced by sequential subtractions of the corresponding curves obtained from the reference flow cell 4 and the running buffer (blank). SPR responses are expressed in resonance unit (RU). The K_D values were calculated by global fitting of the kinetic curves as well as by curve fitting of the steady state responses against concentration according to a 1:1 Langmuir interaction model.

In Vitro Translation Assays

mRNAs were *in vitro* transcribed using T7 polymerase as described above using linearized DNA fragments as templates. These mRNAs were capped (Vaccinia Capping System, NEB) and translated using the nuclease-treated rabbit reticulocyte lysate (RRL; Promega). Typical reactions were comprised of 75% v/v RRL, 20 μM amino acids, and were programmed with ∼50 μg/mL template mRNA. Small molecules were serially diluted ranging from 22 μM - 167 μM final concentrations. Samples were mixed with 3x volumes of 1X NuPAGE™ LDS Sample Buffer (Invitrogen), boiled for 3 min, and resolved on a NuPAGE™ 4 to 12% Bis-Tris polyacrylamide gel (Invitrogen). The products were detected using western blot (method as described above). The nitrocellulose membranes were developed using anti-DDDDK primary (Abcam ab49763) and IRDye® 680RD donkey anti-mouse secondary antibody (LI-COR). Bands were visualized using an Odyssey Clx infrared imager system (LI-COR). Bands corresponding to the –1 or 0-frame products, 58 kDa and 33 kDa respectively, on western blots of *in vitro* translations were quantified densitometrically using ImageJ software [373]. FE was calculated as previously described, by the formula intensity (–1-frame)/(intensity (–1-frame) + intensity (0-frame)) [104]. The change in FE was calculated as a ratio of FE of each condition to the FE of no-protein control in each measurement. Experiments were repeated at least 3 independent times and times and compared by an ordinary one-sided ANOVA, which was followed by a Brown-Forsythe test to ensure equal variance among the samples. Finally, a Dunnett's multiple comparisons test was employed to identify the differentially regulated conditions compared

to the cells only exposed to DMSO. Statistical analysis was performed using GraphPad Prism version 9.2.0.

Computational Chemistry

All computational work was performed using Molecular Operating Environment (MOE), version 2020.09, Chemical Computing Group ULC, 910–1010 Sherbrooke St. W. Montreal, Quebec, H3A 2R7, Canada. Computational procedure was adapted from reported protocol with slight modifications [374].

Preparation of Ligands and RNA Structures

The 2D structures of the MST hits were sketched using ChemDraw professional 20.0 and were imported into the MOE window. The compounds were subjected to an energy minimization up to a gradient of 0.001 kcal mol^{-1} $Å^2$ using the MMFF94x force field and R-field solvation model, then they were saved as mdb file. The predominant protonation status of the compounds in aqueous medium at pH 7 was calculated via the compute | molecule | wash command in the database viewer window. Cryo-electron microscopy structure of the SARS CoV-2 RNA FSE (PDB ID: 6XRZ) [362] and the crystal structure of the antibody fragment-bound SARS CoV-2 FSE (PDB ID: 7MLX) [363] were used for the molecular docking studies. Potential was set up to Amber10:EHT as a force field and R-field for solvation. Addition of hydrogen atoms and correction of library errors were performed via the QuickPrep module.

Binding Site Identification and Ligand–RNA Docking

Binding sites were defined for the prepared RNA structures using the site finder program implemented in MOE. The calculated binding sites were prioritized according to their propensity for ligand binding (PLB), size, and hydrophobicity indices. For docking experiments, dummy atoms were created for each binding site separately, and then they were selected as the docking site in the Dock window. Docking placement was set to a triangle matcher with an induced fit refinement option. The first scoring function was alpha HB with 100 poses, followed by a refinement score affinity dG with 10 poses.

SARS-CoV-2 Infection

For initial infection with SARS-CoV-2, we used the strain hCoV-19/Croatia/ZG-297-20/2020, a kind gift of Prof. Alemka Markotic (University Hospital for Infectious Diseases,

Zagreb, Croatia). The virus was raised for two passages on Caco-2 cells (HZI Braunschweig). For infection experiments, Huh-7 cells were employed. One h before infection, cells were either pre-stimulated with IFN-β (500 U/mL), IFN-ɣ (500 U/mL), respective compounds at 50 µM or left untreated. Cells were infected with 20000 pfu/mL, corresponding to an MOI of 0.03 at 24 h post-infection, cell culture supernatants were collected and titrated by plaque assay on Vero E6 cells (ATCC CRL-1586). Briefly, confluent Vero E6 cells in 96-well plates were inoculated with dilutions of the virus-containing supernatants for one h at 37 °C, the inoculum was removed and cells were overlaid with MEM containing 1.75% methylcellulose. At three days post-infection, whole wells of the plates were imaged using an IncuCyte S3 (Sartorius) at 4x magnification, and plaques were counted visually. For titration analysis and IC_{50} determination in infected live cells, the integrated intensity of green fluorescence was measured in parallel originating from the recombinant icSARS-CoV-2-mNeonGreen [370] used for infection. Cell confluence was measured using an IncuCyte S3 (Sartorius) microscope and its integrated phase contrast-based segmentation analysis. For cell confluence, data was normalized to the no-compound control. Experiments were repeated at least 3 independent times and compared by an ordinary one-sided ANOVA, which was followed by a Brown-Forsythe test to ensure equal variance among the samples. Finally, a Dunnett's multiple comparisons test was employed to identify the differentially regulated conditions compared to the cells only exposed to DMSO. Statistical analysis was performed using GraphPad Prism version 9.2.0.

Cytotoxicity Assay

For cell viability assays, 5,000 Huh-7 cells were seeded into 96-well plates in 100 µL DMEM and cultured for 24 h. After that, 1 µL test compound in DMSO was added to reach a concentration of 50 µM. As controls, DMSO as well as blasticidin were used. 48 h after addition of the compounds, media was removed and cells were washed once with PBS. 100 µL 0.5 mg/mL 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) in DMEM were added and incubated for 2 h. MTT was replaced by 100 µL acidified isopropanol. To dissolve formazan crystals, the suspension was triturated, incubated 30 min in the dark and triturated again. OD was measured at 540 nm to quantify the formazan as a measure for healthy cellular metabolism. The OD 540 of five independent measurements was plotted and compared by an ordinary one-sided ANOVA, which was followed by a Brown-Forsythe test to ensure equal variance among the samples. Finally, a Dunnett's multiple comparisons test was employed to identify the differentially regulated conditions

compared to the cells only exposed to DMSO. Statistical analysis was performed using GraphPad Prism version 9.2.0.

Acknowledgements

We thank Dr. Frank Surup for help with the supply of the small molecule Simone 402A (HZI, Braunschweig, Germany). We thank Dr. Zeljka Macak-Safranko and Prof. Alemka Markotic (University of Zagreb) for providing the SARS-CoV-2 virus isolate prior to publication. We thank Ayse Barut for support in tissue culture assays and Susanne Talay for the organizational support in the BSL-3 facility of HZI. We thank expert technical assistance by Tatyana Koch (HIRI-HZI). Figures were partially or fully generated using BioRender.com (licensed for commercial printing to A.K. (VO23JOZOPV)). L.C.S and U.R were funded through the "AVASON" project of the "GO-Bio initial" program of the German Ministry of Education and Research (BMBF) and the Project "Virological and immunological determinants of COVID-19 pathogenesis – lessons to get prepared for future pandemics (KA1-Co-02 "COVIPA")", a grant from the Helmholtz Association's Initiative and Networking Fund. This project is funded fully or in part by the Helmholtz Association. N.C. and A. K. H. H. received funding from the European Research Council (ERC) Grant Nr. 948636 and Nr. 757913, respectively.

Data availability: Supplementary tables are provided with this thesis

Supplementary information

Chemical Structures of Identified Hits

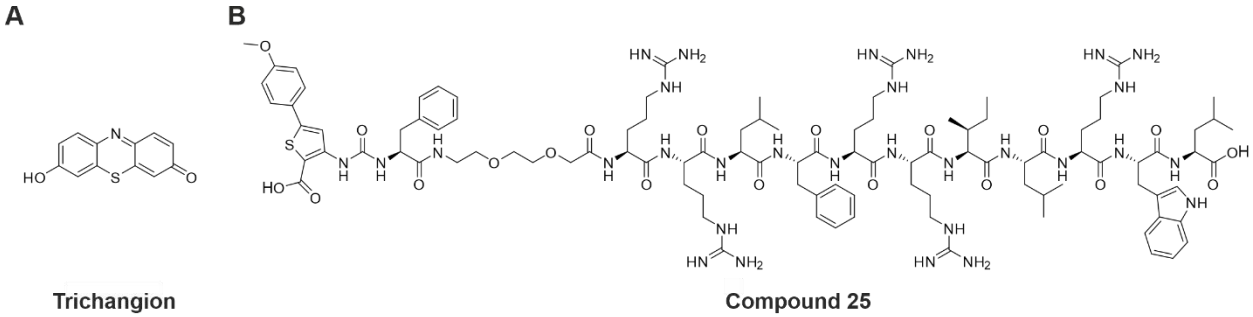


Figure S1. Chemical structure of (A) trichangion and (B) compound **25** (Related to Figure 1).

SPR Capturing Levels, Binding Affinities and Kinetic Parameters of the Hit Compounds

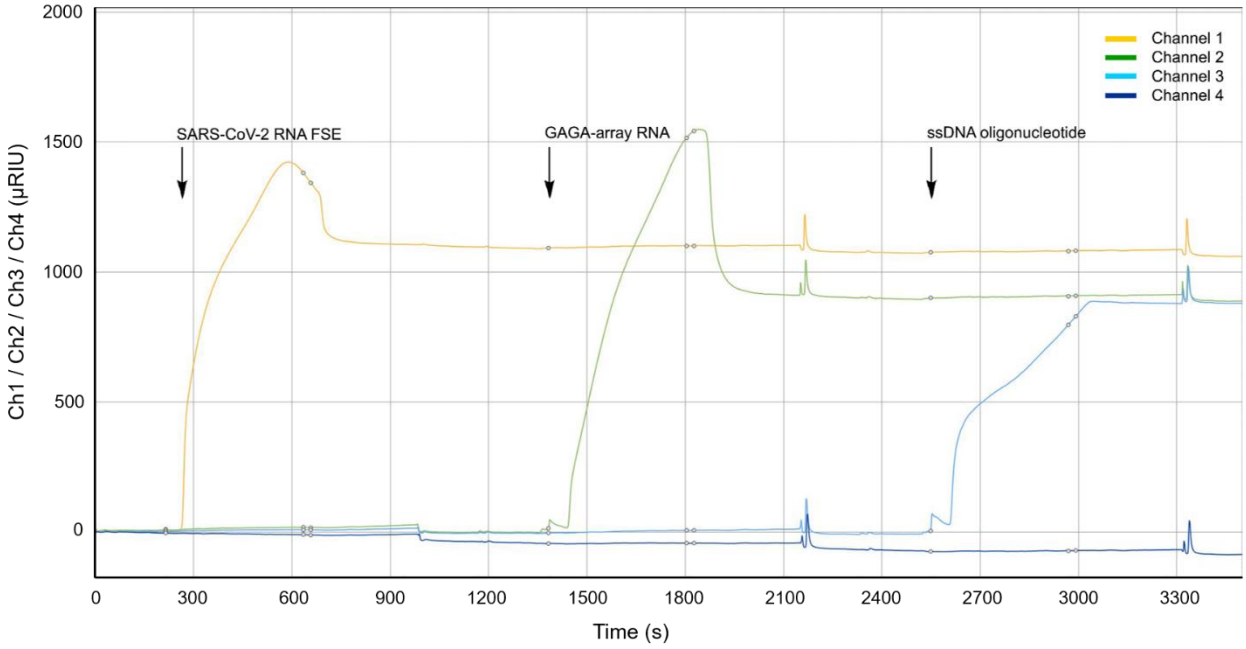


Figure S2. Sensorgrams of capturing biotinylated SARS-CoV-2 RNA FSE (channel 1), random RNA (channel 2), and ssDNA oligonucleotide (channel 3) on SAD500M sensor chip: The orange, green, cyan, and blue curves represent channel 1, channel 2, channel 3, and channel 4 (reference), respectively (Related to Figures 2B, 3B).

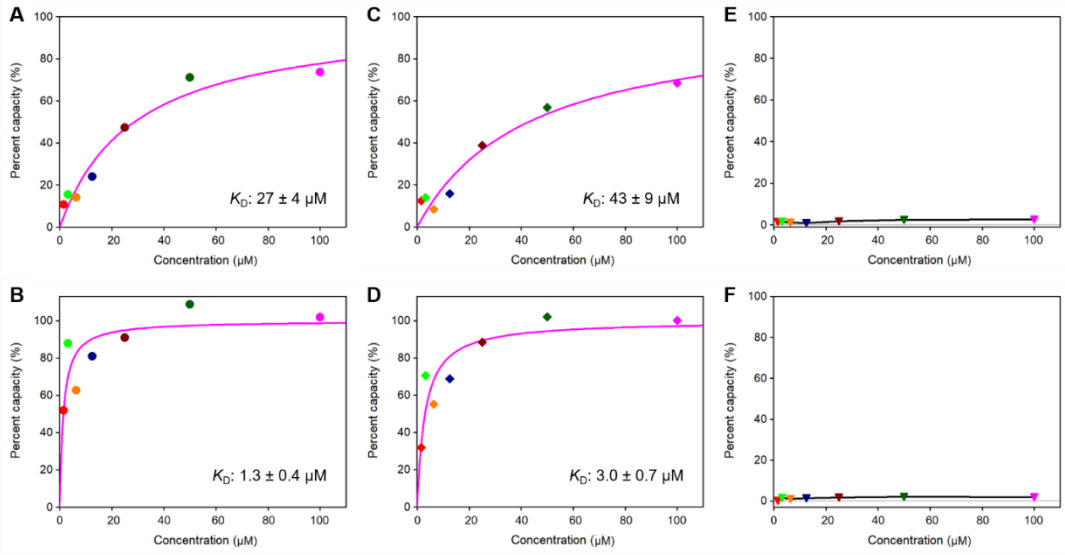


Figure S3. Langmuir binding isotherms of trichangion (A, C, and E) and compound **25** (B, D, and F) flowing over captured SARS-CoV-2 FSE RNA (A and B), random RNA (C and D), and ssDNA oligo (E and F), respectively. Steady state responses are expressed as percent of maximum binding capacity (Related to Figures 2 and 3).

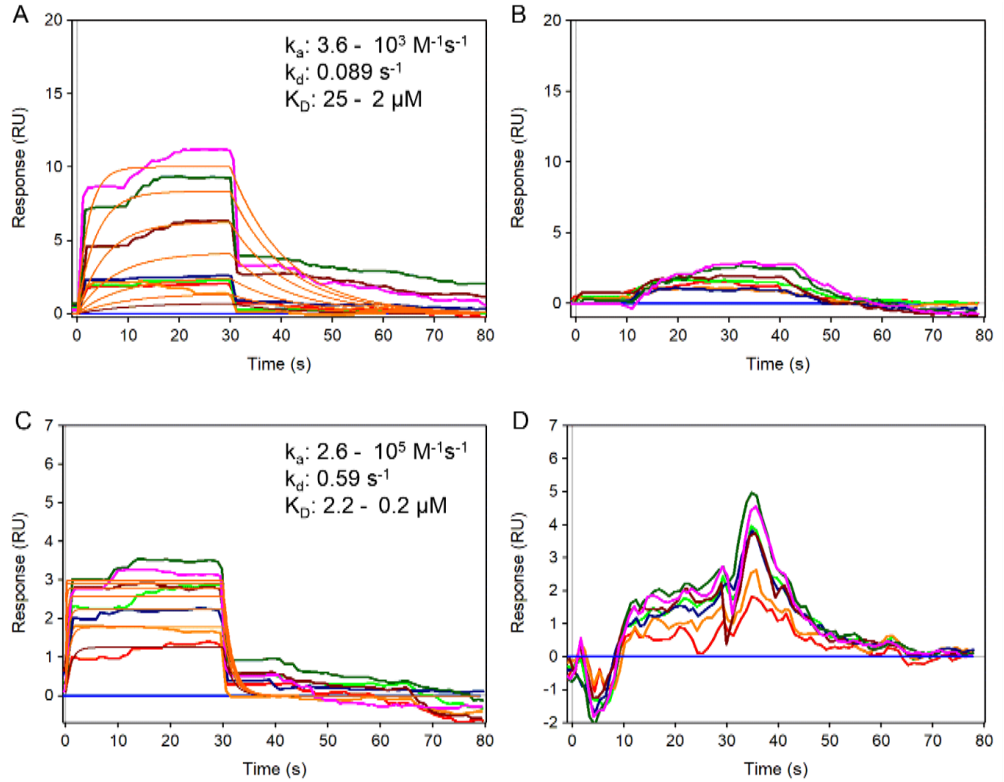


Figure S4. Sensorgrams overlay of trichangion injected at concentrations of 1.56–100 µM over a streptavidin–biotin-captured random RNA (A) and ssDNA oligo (B) and compound **25** injected at concentrations of 1.56–100 µM over a streptavidin–biotin-captured random RNA (C) and ssDNA oligo (D). Kinetic parameters were calculated by global fitting of the association and dissociation curves to a 1:1 binding model (red traces) (Related to Figures 2 and 3).

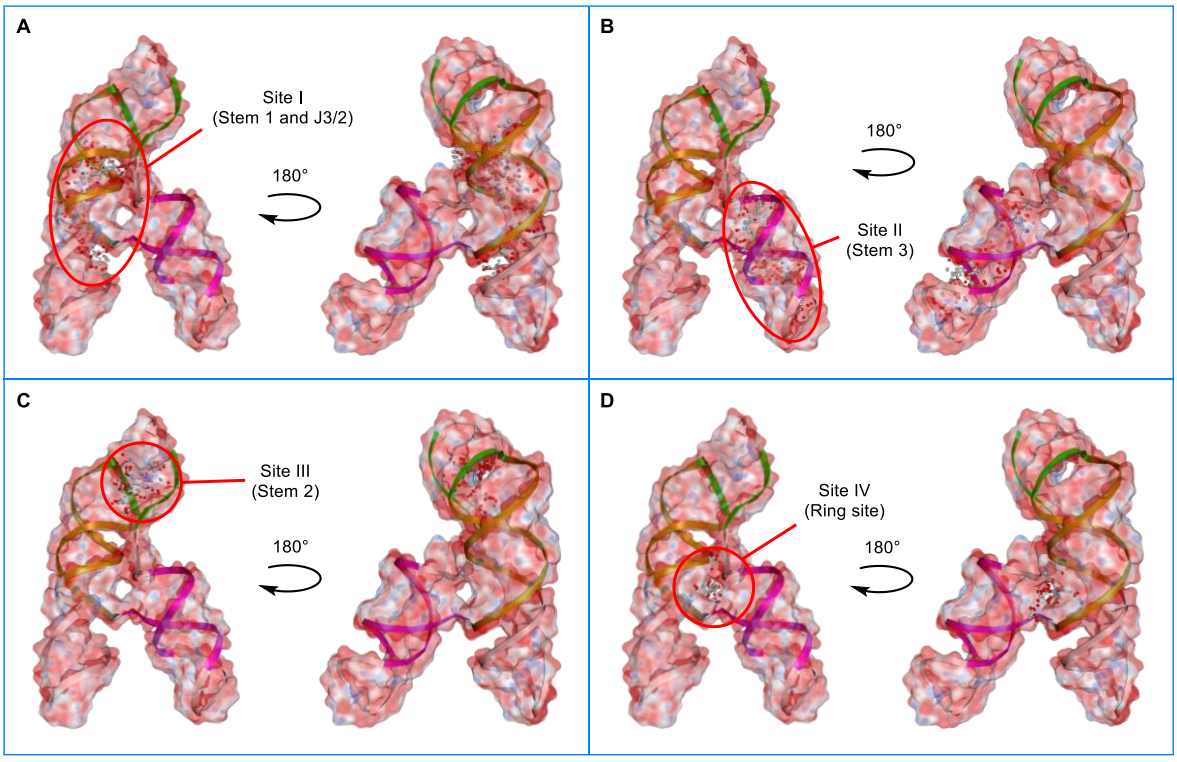


Figure S5. Top four identified ligand-binding sites (A-D represents sites I–IV respectively) in the FSE of SARS-CoV-2 RNA (PDB ID: 6XRZ)[1] represented as clusters of α-spheres. Hydrophilic and hydrophobic cavities are indicted as red and white α-spheres, respectively. Electrostatic surface and ribbon representations of FSE of SARS-CoV-2 RNA are in the same color code as in Figure 4. The three-stem pseudoknot is colored in gold (stem 1), green (stem 2), and magenta (stem 3).

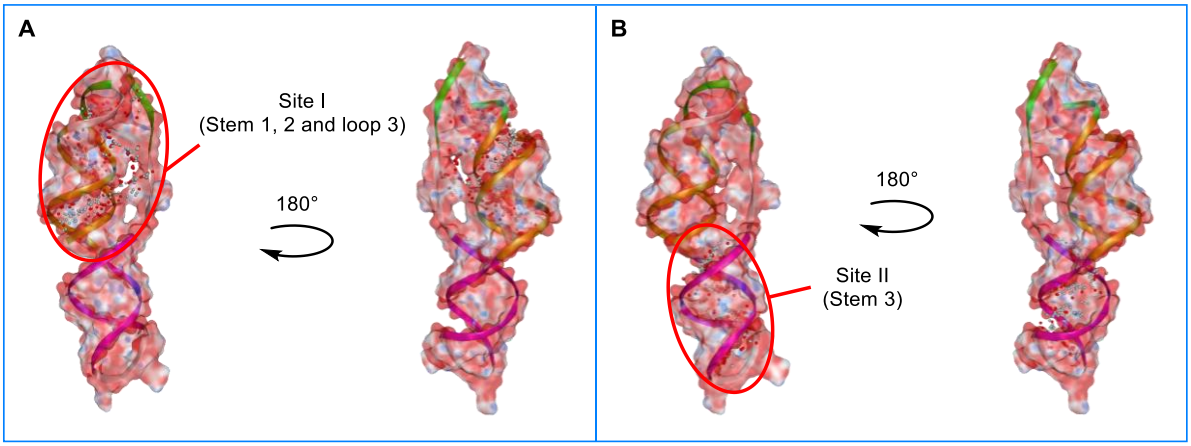


Figure S6. Top two calculated ligand-binding sites (A and B represents sites I and II respectively) in the pseudoknot structure of SARS-CoV-2 RNA (PDB ID: 7MLX)[2] represented as clusters of α-spheres. Hydrophilic and hydrophobic cavities are indicted as red and white α-spheres, respectively. Electrostatic surface and ribbon representations of FSE of SARS-CoV-2 RNA are in the same color code as in Figure 4. The three-stem pseudoknot is colored in gold (stem 1), green (stem 2), and magenta (stem 3).

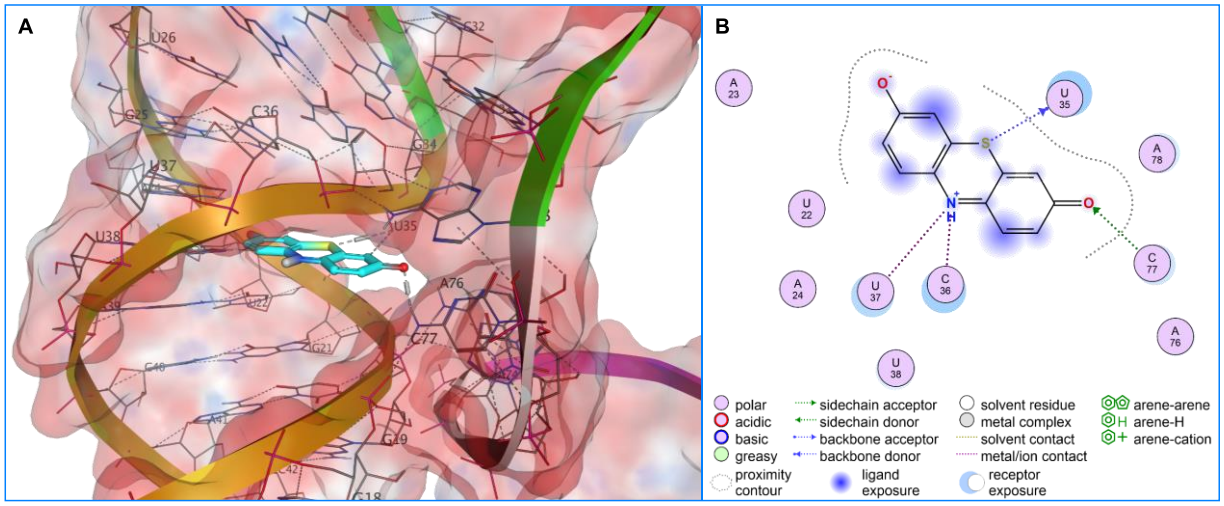


Figure S7. (A) Potential binding mode of trichangion (cyan) into site I (stem 1 and J3/2 site) of SARS-CoV-2 RNA FSE (PDB ID: 6XRZ).[1] Hydrogen bonds and ionic interactions (dashed lines). Electrostatic surface and ribbon representations of the three-stem pseudoknot are in the same color code as in Figure 4. (B) 2D Ligand–RNA interactions.

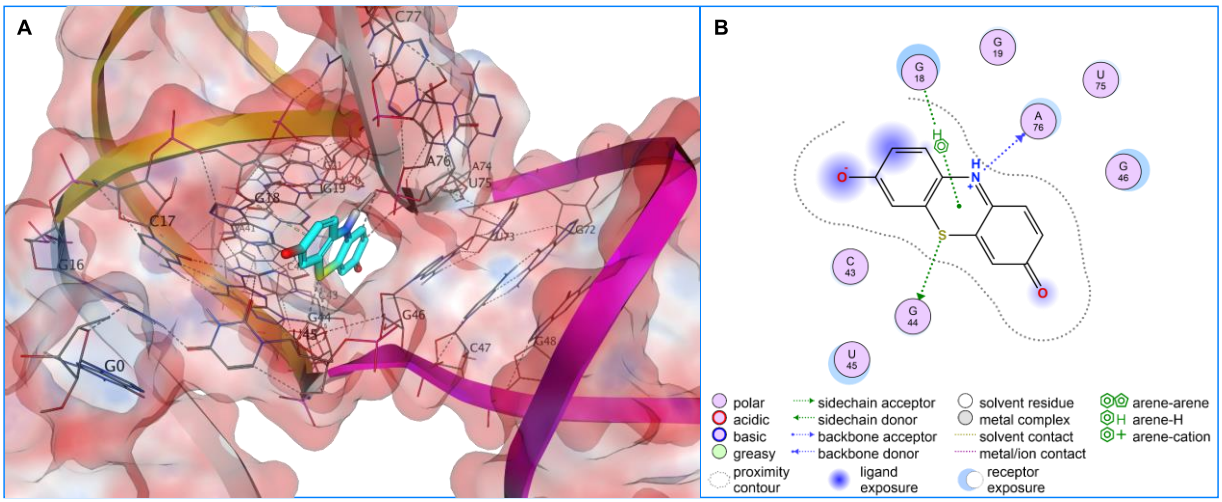


Figure S8. (A) Potential binding mode of trichangion (cyan) into site IV (ring site) of SARS-CoV-2 RNA FSE (PDB ID: 6XRZ).[1] Hydrogen bonds, ionic interactions, and hydrophobic contacts (dashed lines). Electrostatic surface and ribbon representations of the three-stem pseudoknot are in the same color code as in Figure 4. (B) 2D Ligand–RNA interactions.

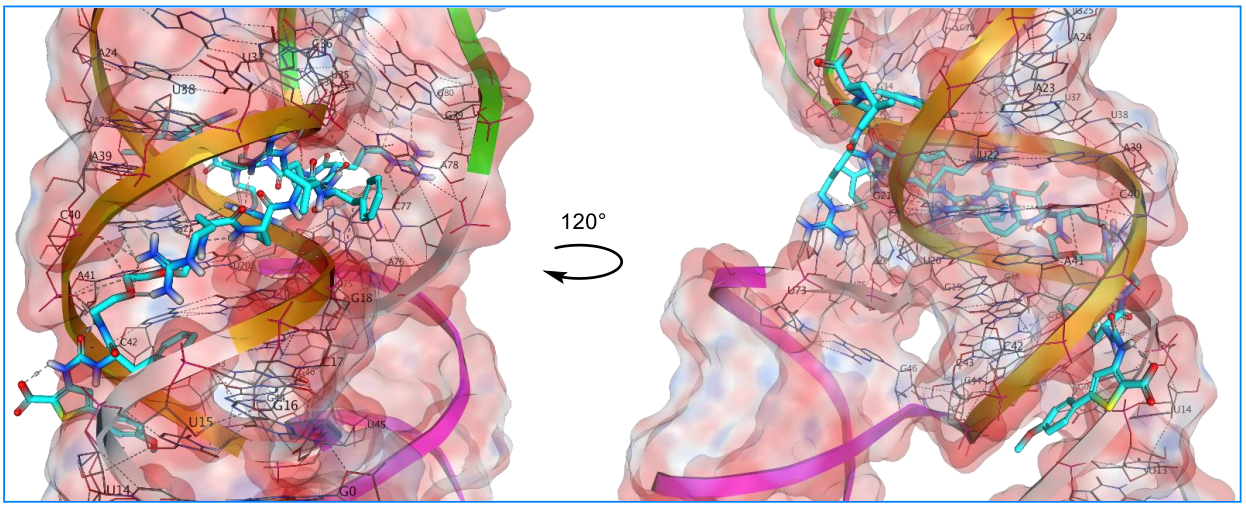


Figure S9. (A) Potential binding mode of compound **25** (cyan) into site I (stem 1 and J3/2 site) of SARS-CoV-2 RNA FSE (PDB ID: 6XRZ).[1] Hydrogen bonds, ionic interactions, and hydrophobic contacts (dashed lines). Electrostatic surface and ribbon representations of the three-stem pseudoknot are in the same color code as in Figure 4.

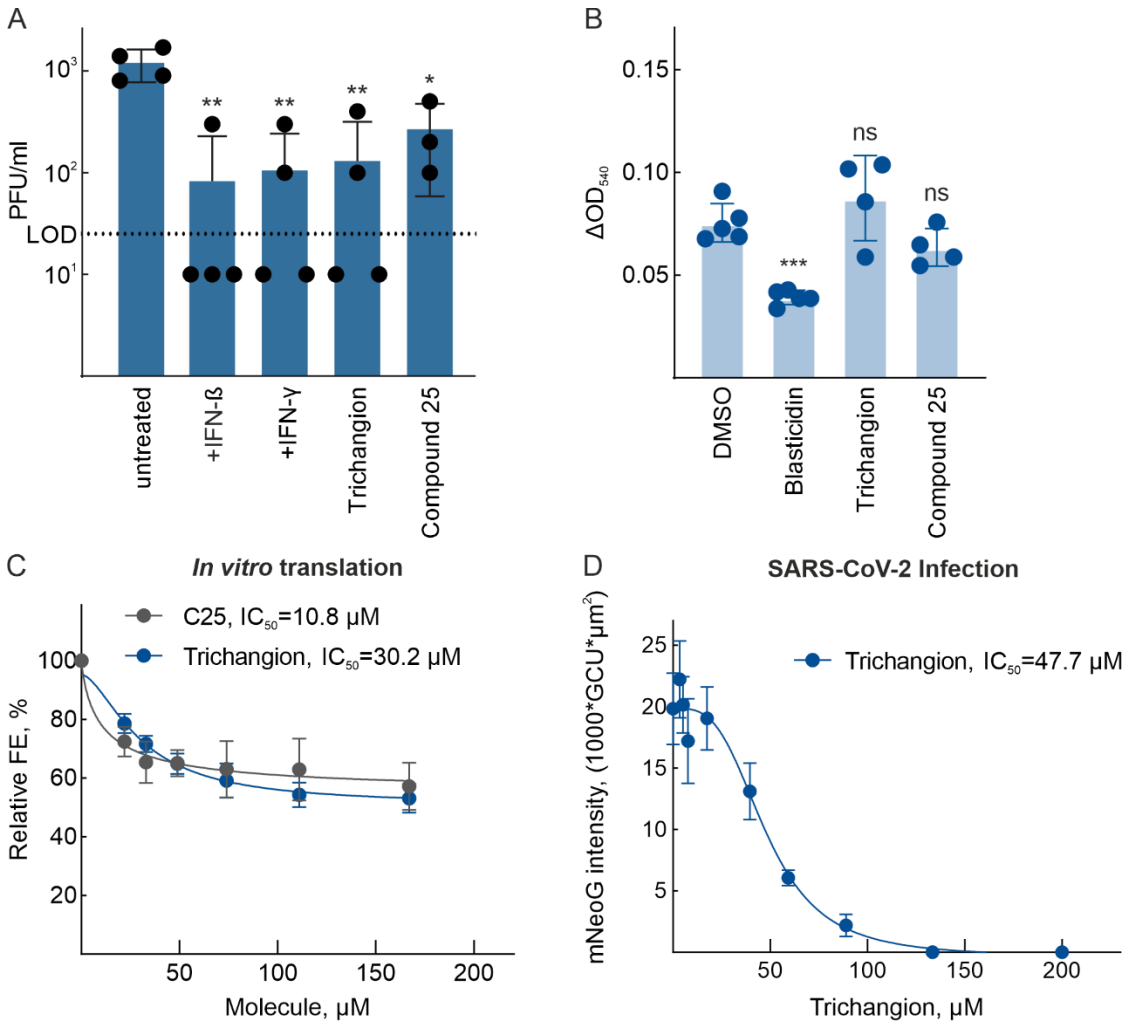


Figure S10. Inhibitory effects on SARS-CoV-2 replication and cytotoxicity evaluation of trichangion and compound **25**. (A) Virus titers in the supernatant of infected Huh-7 cells at 24 h post infection. Treatment with IFN-ɣ (500 U/mL), IFN-β (500 U/mL), IFN-λ1 (5 ng/mL) and the compounds was done 1 h before infection. Boxes show mean values ± s.d. ($n = 4$ independent experiments: 3 independent experiments for compound **25**). The dotted line represents the limit of detection (LOD). *P* values were calculated using an ordinary unpaired one-sided ANOVA comparing every condition to untreated cells. $*P < 0.05$, $**P < 0.01$. Exact *P* values: IFN-β – 0.003, IFN-ɣ - 0.003, trichangion – 0.004, compound **25** – 0.02 (B) Cytotoxicity of DMSO (negative control), blasticidin (positive control) and the tested compounds in Huh-7 cells. Cells were treated with DMSO, blasticidin and compounds 48 h before the addition of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT). OD was measured at 540 nm to quantify the formazan as a measure for healthy cellular metabolism. Boxes show mean values ± s.d. ($n = 5$ independent experiments). P values were calculated using an ordinary unpaired one-sided ANOVA comparing every condition to the DMSO control. ***P < 0.001. Exact P value: Blasticidin – 0.0007 (C) IC_{50} determination of the effect of compounds on –1 PRF efficiency of SARS-CoV-2 *in vitro* (data as presented in Figure 5D, E). (D) IC_{50} determination of trichangion-induced inhibition of SARS-CoV-2 viral replication (data as presented in Figure 5F). Related to Figure 5.

Chapter 5

Summary and Discussion

Translational control allows cells to respond rapidly and efficiently during conditions of environmental and physiological stress, including infection. Furthermore, viruses are entirely dependent on host machinery and employ several tactics to limit host antiviral responses and appropriate control over host ribosomes. This battle for ribosomes leads to substantial changes in the translational landscapes of both host and virus. In this thesis, we used an interdisciplinary approach to explore translational control during viral infections. In **Chapter 2**, the translational landscape of HIV-1 infected human T cells was characterized. Novel viral and eukaryotic short ORFs were discovered, the products of which may play a role in infection or cellular stress response. Additionally, it was shown that HIV-1 frameshifting might have complex frameshift regulatory mechanisms involving RNA elements upstream of the canonical frameshift site that can act as *cis*- regulators of frameshifting. Then, **Chapter 3** focuses on the regulation of frameshifting by *trans*-acting factors. Here,we identified several cellular factors that bind to the SARS-CoV-2 –1PRF site and regulate viral frameshifting. Furthermore, host antiviral protein ZAP-S, which was the strongest frameshift inhibitor in our interactome screen, was extensively characterized *in vitro* and *in vivo*. Such frameshift-inhibitory host proteins can perhaps form a part of the global antiviral response in the cells. We also investigated the molecular details of viral 2A protein-mediated EMCV frameshifting, part of which is mentioned in this thesis as supplemental to Chapter 3 [99]. Our results point to mechanistic differences between host- and virus-encoded –1PRF regulators. Both ZAP-S and EMCV 2A interacted with host translation machinery, similar to the previously reported *trans*-acting protein SHFL [105]. Remarkably, although the binding efficiencies of these proteins to their respective frameshift sites were similar, their mode of action was antagonistic. Thus, this work illuminates contrasting features of frameshifting regulation by the host and viruses. Another class of *trans*-acting factors that target the frameshift element and affect its function are small synthetic molecules. Over the years, such synthetic molecules have been extensively studied for their use as novel antivirals. However, till now, there has been no systematic screen of natural compounds that target the –1PRF element. Hence, in **Chapter 4**, we tested diverse classes of compounds ranging from natural products to synthetic molecule libraries for their RNA-binding capacity. This led to the discovery of the first naturally occurring

compound (trichangion) and the first peptide (compound **25**) that can selectively bind to the frameshift element of SARS-CoV-2 and impede its function in cells in the low micromolar range.

Here, I discuss the potential biases of the techniques employed and suggest future work which can be done to further our understanding of translational control in infected cells and –1PRF regulation.

Ribosome profiling uncovers novel features of translational regulation

Ribosome profiling is an indispensable tool for the field of translation, and a reliable experimentally based technique to identify translation events. It bridges the gap between transcriptomic and proteomic studies, providing novel insights into changes in gene expression during particular cellular conditions and stresses as well as mechanisms of translation regulation and protein synthesis. Numerous findings in Chapter 2 are exceedingly unlikely to have been detected by alternative methods. A notable example of this is determining frameshift efficiency of viruses during early stages of infection. The levels of viral proteins are quite low to be able to determine the frameshift efficiency using less sensitive techniques such as western blot. Despite the strengths of this technique, depending in the cell-type and the protocol used, certain biases can be introduced in the RiboSeq data. It was shown that pre-treatment with cycloheximide can bias ribosome density data leading to a heightened initiation peak and a smaller termination peak [140]. Due to this reason, samples are usually harvested by flash-freezing without any pre-treatment. However, in suspension cell lines, like the Sup-T1 cells used in this study, immediate flash-freezing was not possible. Most of the studies related to cycloheximide-mediated biases are performed in yeast and a recent study has shown that human cell lines are resistant to these biases [375]. Moreover, no pretreatment leads to running off of ribosomes towards the 3' UTR leading to a large termination peak and a smaller initiation peak. In cases of such ribosome run-off, whether the distribution of reads within ORFs represents true biology or a technical artifact is difficult to determine. Furthermore, comparison of footprint densities between different regions of the same ORF is difficult, thus having implications for calculation of frameshift efficiencies. Hence, cycloheximide was used for studying the translation in HIV-1 infected Sup-T1 cells where rapid translational changes occur in early stages of infection.

The characterization of novel ORFs is one of the major highlights of ribosome profiling. Previous RiboSeq studies have successfully identified novel uORFs or iORFs in a large

number of viruses such as PRRSV, influenza, SARS-CoV-2, human cytomegalovirus, herpes simplex virus 1 and MuLV [103, 150, 151, 157, 376,154, 263, 377, 378]. Such short ORFs have been implicated in altering the sites of translation initiation, regulating production of downstream proteins and producing novel immune epitopes. We observed ribosome stalling peaks in the 5' UTR of HIV-1, which suggests that HIV-1 also contains such uORFs. HIV-1 *pol* and *vpu* coding regions contain iORFs in alternative reading frames. It is highly likely that translation of these iORFs can occur through non-canonical means, either PRF or alternative initiation mechanisms. Besides viruses, approximately 50% of human protein-coding genes have been identified to contain uORFs and specific uORFs in eukaryotes are known to control expression of downstream proteins, particularly during stressful conditions [161-163, 217, 379]. A proteogenomic approach that combined RNASeq, RiboSeq and MS data identified 2,503 new non-canonical human proteins, of which 72% were cryptic proteins and 28% novel isoforms [380]. In **Chapter 2**, we detected several novel sORFs in the human genome. However, due to imprecise sequence matching with genomic information, RiboSeq-built databases can likely overestimate genuine translation products, especially non-canonical ones [380]. Hence, in future work it will be important to validate the presence of these cryptic, upstream or *trans*-frame peptides through MS or reporter assays.

So far, the presented work in **Chapter 2** was the first RiboSeq study in HIV-1 infected T cells providing valuable information on the impact of translational control during HIV-1 infection. It should be noted, however, during the course of this work, another ribosome profiling study using HIV-1 infected T cells was published and is currently available as a preprint [381]. The ribosome profiling protocol and results of both studies are similar, albeit they included a later timepoint of infection (36h). We observe stalling peaks in the HIV-1 5' UTR as observed similarly in Labaronne *et al.*, indicating that these may be actual ribosome protected fragments, products of which cause immune recognition of HIV-1 particles by the host cell. On the other hand, Labaronne *et al.*did not focus on the HIV-1 frameshift site and it would be interesting to see whether they observe similar pauses upstream of the frameshift site like shown in **Chapter 2**. Furthermore, they did not observe iORFs within *pol* and *vpu* genes. However, it is likely due to differences in bioinformatic analyses these were overlooked in their data, but comparison of raw reads should confirm if these iORFs exist also at later timepoints of infection. Nonetheless, both studies revealed intricate details of HIV-1 translational regulation and non-canonical mechanisms used by this virus to manipulate the host translational machinery.

In case of TMEV frameshifting, ribosome stacking or collisions were seen at the frameshift site, where ribosome translating the StopGo motif collides with the ribosome paused over the slippery sequence [98]. Such collisions are said to contribute to the complex energetic and conformational landscape required for frameshifting. Unlike with TMEV, we did not observe ribosome stacking at the HIV-1 frameshift site. However, we observed ribosome collisions upstream of the canonical frameshift site, which we propose is a regulatory mechanism employed by the virus to ensure appropriate levels of frameshifting. Furthermore, we attempted to identify the enriched codons in each position of the ribosomal exit tunnel in the human disome data. However, a deeper analysis is necessary to correctly define the disome stalling motifs, which would increase our understanding on factors leading to ribosome collisions in the human genome. Such collisions are said to increase during cellular stress conditions such as amino acid starvation or UV-irradiation, however it is yet to be studied whether frequency of ribosome collisions in the human genome increases during viral replication. The cellular response to increased ribosome collisions is usually inhibition of initiation and activation of ribosome quality control (RQC) pathways, which lead to eventual nascent peptide degradation and ribosome recycling [382-384]. This would likely increase the ribosome pool in cells now available for viral translation. Indeed, vaccinia virus was shown to manipulate RQC pathways by co-opting ZNF958 involved in ubiquitination of collided ribosomes [385]. Given the complex interplay between HIV-1 and the host translation system, it raises the question whether HIV-1 interferes with RQC pathways in a similar fashion.

A deeper look into the cellular functions of ZAP-S is required

In **Chapter 3**, we extensively characterized ZAP-S as a novel regulator of SARS-CoV-2 frameshifting, adding another layer to its role as an antiviral protein. Although both isoforms of ZAP were shown to have antiviral activities, they have distinct cellular localization and mechanisms of action [291]. Furthermore, ZAP-S is significantly upregulated during infection, while ZAP-L is constitutively expressed in cells. Despite a large number of studies describing ZAP-S/L interaction with viral RNAs, the regulatory role of ZAP-S/L in host mRNAs is largely unknown. Prior to establishing ZAP-S/L as a viable therapeutic target, it is critical to understand its impact on cellular processes and translational regulation. For this purpose, it would be useful to conduct RiboSeq studies on un-/infected ZAP knockout cell lines. Through these, the impact of ZAP-S/L on the translational control

in cells during normal or stressed conditions could be explored and the key molecular processes regulated by this protein and its isoforms identified.

Further work can reveal intricate details of ZAP-S-mediated –1PRF modulation

Although inhibition of SARS-CoV-2 frameshifting may lead to an overall decrease in the viral titer, we yet do not have direct proof of whether and to what extent the infection is impacted due to inhibition of –1PRF by ZAP. The *in vitro* translation construct containing the viral genomic snippet was designed to represent the wild-type context of *SARS-CoV-2* mRNA. While this resembles viral mRNA during infection, a quantitative analysis of 0- and –1-frame proteins using a western blot would allow us to directly measure the frameshift levels and their correlation with the viral titers. However, this was not possible due to technical reasons. Additionally, ZAP-S performs a plethora of functions during infection, ranging from RNA degradation, translation inhibition to stimulation of immune signalling pathways [291, 386, 387]. Therefore, it is yet unknown whether the significant drop in viral titer was driven solely by –1PRF inhibition or by the cumulative effect of all ZAP-S-led antiviral actions. The RNA mutation studies in **Chapter 3** show that ZAP-S has multiple binding sites on the PRF RNA – specifically in SL2 and SL3 regions. However, the exact domain of the protein that interacts with the –1PRF RNA is still not determined. For this purpose, it would be useful to conduct structural studies such as cryo-EM or X-ray crystallography for the wild-type protein as well as ZAP-S truncation mutants. A novel approach for this aim would be to construct synthetic ZAP-S derived peptide arrays for investigating protein domains specifically binding to –1PRF RNA. Although this approach has been previously solely applied to studying the binding specificity of protein-protein interactions [388], it should be possible to apply the same principle for RNA-binding. While this assay would not replicate the tertiary structure of the protein, it would give an idea of the exact RNA-interacting surface of the protein. The identified RNA-specific peptides can be tested for their frameshift modulatory effects and developed as therapeutics.

Furthermore, our results show that ZAP-S interacts with translating ribosomes in addition to –1PRF RNA. It is yet to be determined whether this interaction is direct or indirect and whether it is functionally relevant to frameshift inhibition. Our preliminarily results using microscale thermophoresis suggest that ZAP-S directly binds to both ribosomal subunits, with higher affinity towards the 40S. This interaction can occur through binding to the rRNA or to the ribosomal proteins. Due to the complex pathways of maintaining

translational fidelity and the role of rRNA in these pathways, it would be interesting to know if this interaction can impact translational accuracy and speed, thereby impacting frameshifting. Here, structural techniques including crosslink-IP-MS as well as cryo-EM could decipher the specificity and mechanism of ZAP-S: ribosome binding.

Overall, this thesis largely enhances the understanding of post-transcriptional regulation of both host and viral mRNAs during infection. It shows that translational regulation is controlled by the complex interplay of the transcriptome with RNA-binding proteins. Characterization of non-canonical translation events and their regulation will likely open doors for viral targeting and immune modulatory intervention strategies to fight current and emerging viral diseases.

Statement of individual author contributions to figures

Manuscript 1 (complete reference):

The translational landscape of HIV-1 infected cells. Anuja Kibe, Stefan Buck, Redmond Smyth, Florian Erhard, Neva Caliskan. *Manuscript in preparation.*

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	AK	SB	NC		
2	AK	SB	NC		
3	AK	SB	NC		
4	AK	SB	NC		
5	AK	SB	NC		

Explanations (if applicable): Supplementary figures and tables were created by AK

Manuscript 2 (complete reference):

The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting. Matthias M. Zimmer*, Anuja Kibe*, Ulfert Rand, Lukas Pekarek, Liqing Ye, Stefan Buck, Redmond P. Smyth, Luka Cicin-Sain & Neva Caliskan. *Nature Communications volume 12, Article number: 7193 (2021)*

*equal contribution

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	MM	AK	NC		
2	MM	AK	UR	NC	LCS
3	AK	MM	NC		
4	AK	MM	NC		
5	LP	AK/MM	SB	NC	
6	AK	MM	NC		

Explanations (if applicable): Supplementary figures 1-3 equal contributions AK and MM, Supplementary tables 2 and 3 – AK

Manuscript 3 (complete reference):

Screening of Natural Products and Small Molecules Uncovers Novel Coronavirus 1a/1b Frameshifting Inhibitors with Antiviral Properties. Anuja Kibe*, Walid A. M. Elgaher*, Ulfert Rand, Matthias M. Zimmer, Andreas M. Kany, Jennifer Hermann, Rolf Müller, Luka Cicin-Sain, Anna K. H. Hirsch, Neva Caliskan. *Under review. JMB-D-22-00620, Available at SSRN: https://ssrn.com/abstract=4157446 or http://dx.doi.org/10.2139/ssrn.4157446*

*equal contribution

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	AK	NC	WE	AKH	
2	AK	NC	WE	AKH	
3	AK	NC	WE	AKH	
4	WE	AKH	AK	NC	
5	AK	NC	UR	WE	AKH

Explanations (if applicable): Supplementary figure S10B created by AK

I also confirm my primary supervisor's acceptance.

_____

Doctoral Researcher's Name Date Place Signature

Statement of legal second publication rights

Manuscript 2

Statement of author contributions

Manuscript 1 (complete reference):

The translational landscape of HIV-1 infected cells. Anuja Kibe, Stefan Buck, Redmond Smyth, Florian Erhard, Neva Caliskan. *Manuscript in preparation.*

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AK/NC	FE	RP		
Data Collection	AK				
Data Analysis and Interpretation	AK	SB	NC		
Manuscript Writing	AK/NC				
Writing of Introduction	AK/NC				
Writing of Materials & Methods	AK				
Writing of Discussion	AK/NC				
Writing of First Draft	AK				

Manuscript 2 (complete reference):

The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting. Matthias M. Zimmer*, Anuja Kibe*, Ulfert Rand, Lukas Pekarek, Liqing Ye, Stefan Buck, Redmond P. Smyth, Luka Cicin-Sain & Neva Caliskan. *Nature Communications volume 12, Article number: 7193 (2021)*

*equal contribution

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	AK/MZ/NC				
Methods Development	AK/MZ/ NC		LP/LY/UR/RPS		
Data Collection	AK/MZ	LP	UR	LY	

Data Analysis and Interpretation	AK/MZ	NC	LP/UR/RPS		
Manuscript Writing	AK/MZ/NC	LP	UR/LY/RPS/LCS		
Writing of Introduction	AK/ MZ/NC				
Writing of Materials & Methods	AK/MZ	LP	UR/LY/RPS		
Writing of Discussion	AK/MZ/NC	LP			
Writing of First Draft	AK/MZ	LP			

Manuscript 3 (complete reference):

Screening of Natural Products and Small Molecules Uncovers Novel Coronavirus 1a/1b Frameshifting Inhibitors with Antiviral Properties. Anuja Kibe*, Walid A. M. Elgaher*, Ulfert Rand, Matthias M. Zimmer, Andreas M. Kany, Jennifer Hermann, Rolf Müller, Luka Cicin-Sain, Anna K. H. Hirsch, Neva Caliskan. *Under review. JMB-D-22-00620, Available at SSRN: https://ssrn.com/abstract=4157446 or http://dx.doi.org/10.2139/ssrn.4157446*

*equal contribution

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AK/NC/WE/AKH				
Data Collection	AK/WM	MM/UR			
Data Analysis and Interpretation	AK/WM	NC/AKH	MM/UR	LCS	
Manuscript Writing	AK/NC/WE/AKH	MM/UR	LCS/RM		
Writing of Introduction	AK/NC/WE/AKH	MM/UR	LCS/RM		
Writing of Materials & Methods	AK//WE	MM/UR			
Writing of Discussion	AK/NC/WE/AKH	MM/UR	LCS/RM		
Writing of First Draft	AK				

Appendix A – Publications during candidature

Hill, C. H.*, Cook, G. M.*, Napthine, S.*, **Kibe, A**., Brown, K., Caliskan, N., Firth, A. E., Graham, S. C., & Brierley, I. (2021). Investigating molecular mechanisms of 2A-stimulated ribosomal pausing and frameshifting in Theilovirus. Nucleic acids research, 49(20), 11938–11958. *https://doi.org/10.1093/nar/gkab969*

Hill, C. H.*, Pekarek, L.*, Napthine, S.*, **Kibe, A**., Firth, A. E., Graham, S. C., Caliskan, N., & Brierley, I. (2021). Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch. Nature communications, 12(1), 7166. *https://doi.org/10.1038/s41467-021-27400-7*

Zimmer, M. M.*, **Kibe, A.*,** Rand, U., Pekarek, L., Ye, L., Buck, S., Smyth, R. P., Cicin-Sain, L., & Caliskan, N. (2021). The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting. Nature communications, 12(1), 7193. *https://doi.org/10.1038/s41467-021-27431-0*

Liao, C., Sharma, S.*, Svensson, S. L.*, **Kibe, A**.*, Weinberg, Z., Alkhnbashi, O. S., Bischler, T., Backofen, R., Caliskan, N., Sharma, C. M., & Beisel, C. L. (2022). Spacer prioritization in CRISPR-Cas9 immunity is enabled by the leader RNA. Nature microbiology, 7(4), 530–541. *https://doi.org/10.1038/s41564-022-01074-3*

Ye, L., Gribling-Burrer, A. S., Bohn, P., **Kibe, A**., Börtlein, C., Ambi, U. B., Ahmad, S., Olguin-Nava, M., Smith, M., Caliskan, N., von Kleist, M., & Smyth, R. P. (2022). Short- and long-range interactions in the HIV-1 5' UTR regulate genome dimerization and packaging. Nature structural & molecular biology, 29(4), 306–319. *https://doi.org/10.1038/s41594-022-00746-2*

Kibe, Anuja* and Elgaher, Walid A. M*. and Rand, Ulfert and Zimmer, Matthias M. and Kany, Andreas M. and Hermann, Jennifer and Müller, Rolf and Cicin-Sain, Luka and Hirsch, Anna K. H. and Caliskan, Neva, Screening of Natural Products and Small Molecules Uncovers Novel Coronavirus 1a/1b Frameshifting Inhibitors with Antiviral Properties. JMB-D-22-00620, Available at SSRN: https://ssrn.com/abstract=4157446 or http://dx.doi.org/10.2139/ssrn.4157446

* Equal contribution

Appendix B – Acknowledgements

I would like to start by thanking my primary supervisor, Prof. Neva Caliskan. Thank you for taking a chance on me and letting me be one of your first PhD students. For always offering advice, motivation and support during the toughest of times. For all the scientific discussions, constructive criticism and opportunities to explore our scientific ideas, all of which have made me a better scientist. Importantly, thank you for being a fine example of what a boss 'woman in science' looks like.

A big thanks to my thesis committee for their helpful insights and readiness to provide guidance. Prof. Redmond Smyth, I cannot thank you enough for your support on all projects. You are and always will be one of the coolest scientists I know. Prof. Florian Erhard, I would start by apologizing for the incessant emails asking for help during my early PhD days, but thank you for your incredible patience and help. Prof. Luka Cicin-Sain, thank you for all the constructive feedback and insights on the project and for co-organizing seminars, allowing us to present our naïve results without any judgement.

Thanks to the Helmholtz institute for RNA-based infection research (HIRI) for the supportive environment and plentiful resources. I am honoured to be a part of this institute right from the beginning and watch it grow with time. Thanks to the graduate school of life sciences (GSLS) for providing a nurturing and supportive atmosphere.

I would like to thank Dr. Anne-Sophie Gribling-Smyth and Dr. Lydia Hadjeras for their advice with ribosome profiling, virus preparation and general PhD pep-talks.

Someone once tweeted that you can't and shouldn't be friends with your colleagues and I would like to say that this advice, like most advice on Twitter, is absolutely wrong. Clearly, this person has never worked with the REMIs, who aren't just some of my closest friends but also my family. Matthias, you have been my support since the beginning of my PhD journey, and all I want to say is Thanks, Gorge! Lukas, here's to more SciComm discussions and laughing at each other's jokes when no one else does. Ricarda, I would not swap you for anyone else. Stefan, you are a literal calm in the stormy office, a great cat-dad and the best person to work with on the HIV project. Sophia, you are a better student than I ever was and I know you're going to kill it wherever you go next. Tatyana, the clean and tidy part of our messy lab lives, thank you for always being there for us. I can't possibly forget our neighbours- GARVs- who we shared the lab, reagents and basically, all our

problems with. Uddhav, Shazeb, Marco, Patrick, Liqing, Melanie, Charlene – thank you for all the fun and help. Working at HIRI, gave me the opportunity to meet some of the best people. Shuba, thank you for being my home away from home. Thank you to each and every one of my friends here for the football, parties, holidays or even just hanging out. I cannot name every one of you in this space, but you all are awesome, each of you get a beer, and I look forward to having you in India.

I could not have been here without the support of my family back home. *Aai,* the strongest person I know, who got rid of every obstacle in our path, thank you for everything. *Baba*, I will never forget the tiny microscope you got home from work one day, thank you for being my personal encyclopedia and encouraging my love for science. *Tai*, I spent my childhood trying to be exactly like you, thank you for being the only one at that time who encouraged me to embrace my weird self. Parth, thank you for always patiently listening to my rants about everything, you are the best partner anyone could ask for and you have my whole heart. I would also like to thank my extended family, especially my grandmother, my aunts and my sisters – some of the fiercest women I have seen, who handle everything life offers with ease and grace. I have been fortunate enough to gain a second set of parents through Parth, thank you *Kaka*, *Kaku* for all the care packages, phone calls and never-ending support. A big hug to my chosen family- Shriya, Tanvi, Anushka and Shounak- I hope I am as helpful in your life journeys as you are in mine.

I would like to thank my therapist, Sus, for being incredible at their job and making me capable of handling my anxiety and imposter syndrome. Half of my PhD was during the coronavirus pandemic, thank you to all the frontline workers who made it possible for me to come to work and continue my projects. I guess vaccinations don't just save lives, but also help you complete your academic degrees.

Lastly, I would like to dedicate this thesis to my grandfather, Dr. Balwant Joshi, professor in organic chemistry to the world but to me, he was just my dear *Ajoba*. A strict man where academics were concerned but at his core just a huge soft mush with undying love for sweets and the old Indian adage 'Biotechnology *madhe khup* scope *ahe*'. Unfortunately, he was not here to see me follow his footsteps and set out on this journey of science. *Ajoba,* if there is indeed a happier place, I know you are there. I raise a *pedha* to you and hope I have made you proud.

Appendix D - Affidavit

I hereby confirm that my thesis entitled “Translational landscape and regulation of recoding in virus-infected cells” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another rexamination process neither in identical nor in similar form.

Hiermit erkläre ich an Eides statt, die Dissertation ‚Translationslandschaft und Regulierung der Rekodierung in virusinfizierten Zellen‘ d.h. ins-besondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, 12/08/22

Place, Date Signature

Bibliography

1. Schuller, A.P. and R. Green, *Roadblocks and resolutions in eukaryotic translation.* Nature Reviews Molecular Cell Biology, 2018. **19**(8): p. 526-541.
2. Sonenberg, N. and A.G. Hinnebusch, *Regulation of translation initiation in eukaryotes: mechanisms and biological targets.* Cell, 2009. **136**(4): p. 731-45.
3. Dever, T.E. and R. Green, *The elongation, termination, and recycling phases of translation in eukaryotes.* Cold Spring Harb Perspect Biol, 2012. **4**(7): p. a013706.
4. Dever, T.E., J.D. Dinman, and R. Green, *Translation Elongation and Recoding in Eukaryotes.* Cold Spring Harb Perspect Biol, 2018. **10**(8).
5. Sokabe, M. and C.S. Fraser, *Toward a Kinetic Understanding of Eukaryotic Translation.* Cold Spring Harb Perspect Biol, 2019. **11**(2).
6. Jackson, R.J., C.U. Hellen, and T.V. Pestova, *The mechanism of eukaryotic translation initiation and principles of its regulation.* Nat Rev Mol Cell Biol, 2010. **11**(2): p. 113-27.
7. Furuichi, Y., *Discovery of m(7)G-cap in eukaryotic mRNAs.* Proc Jpn Acad Ser B Phys Biol Sci, 2015. **91**(8): p. 394-409.
8. Hinnebusch, A.G., *The scanning mechanism of eukaryotic translation initiation.* Annu Rev Biochem, 2014. **83**: p. 779-812.
9. Amrani, N., et al., *Translation factors promote the formation of two states of the closed-loop mRNP.* Nature, 2008. **453**(7199): p. 1276-80.
10. Nakamoto, T., *Mechanisms of the initiation of protein synthesis: in reading frame binding of ribosomes to mRNA.* Mol Biol Rep, 2011. **38**(2): p. 847-55.
11. Conte, M.R. and S. Pennell, *Translation Initiation: A Eukaryotic Perspective*, in *Encyclopedia of Biophysics*, G.C.K. Roberts, Editor. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 2644-2650.
12. Algire, M.A., D. Maag, and J.R. Lorsch, *Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation.* Mol Cell, 2005. **20**(2): p. 251-62.
13. Majumdar, R. and U. Maitra, *Regulation of GTP hydrolysis prior to ribosomal AUG selection during eukaryotic translation initiation.* Embo j, 2005. **24**(21): p. 3737-46.
14. Ceci, M., et al., *Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly.* Nature, 2003. **426**(6966): p. 579-84.
15. Si, K., et al., *Molecular cloning and functional expression of a human cDNA encoding translation initiation factor 6.* Proceedings of the National Academy of Sciences, 1997. **94**(26): p. 14285-14290.
16. Fringer, J.M., et al., *Coupled release of eukaryotic translation initiation factors 5B and 1A from 80S ribosomes following subunit joining.* Mol Cell Biol, 2007. **27**(6): p. 2384-97.
17. Kurland, C.G., *TRANSLATIONAL ACCURACY AND THE FITNESS OF BACTERIA.* Annual Review of Genetics, 1992. **26**(1): p. 29-50.
18. Ortiz, P.A., et al., *Translation elongation factor 2 anticodon mimicry domain mutants affect fidelity and diphtheria toxin resistance.* J Biol Chem, 2006. **281**(43): p. 32639-48.
19. Djumagulov, M., et al., *Accuracy mechanism of eukaryotic ribosome translocation.* Nature, 2021. **600**(7889): p. 543-546.
20. Nakamura, Y. and K. Ito, *tRNA mimicry in translation termination and beyond.* WIREs RNA, 2011. **2**(5): p. 647-668.
21. Song, H., et al., *The Crystal Structure of Human Eukaryotic Release Factor eRF1—Mechanism of Stop Codon Recognition and Peptidyl-tRNA Hydrolysis.* Cell, 2000. **100**(3): p. 311-321.
22. Lawson, M.R., et al., *Mechanisms that ensure speed and fidelity in eukaryotic translation termination.* Science, 2021. **373**(6557): p. 876-882.
23. Eyler, D.E. and R. Green, *Distinct response of yeast ribosomes to a miscoding event during translation.* Rna, 2011. **17**(5): p. 925-32.

24. Mitchell, S.F., et al., *The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway.* Mol Cell, 2010. **39**(6): p. 950-62.
25. Firth, A.E. and I. Brierley, *Non-canonical translation in RNA viruses.* J Gen Virol, 2012. **93**(Pt 7): p. 1385-1409.
26. Miras, M., et al., *Non-canonical Translation in Plant RNA Viruses.* Frontiers in Plant Science, 2017. **8**.
27. Balvay, L., et al., *Translational control of retroviruses.* Nat Rev Microbiol, 2007. **5**(2): p. 128-40.
28. Sorokin, II, et al., *Non-Canonical Translation Initiation Mechanisms Employed by Eukaryotic Viral mRNAs.* Biochemistry (Mosc), 2021. **86**(9): p. 1060-1094.
29. Jaafar, Z.A. and J.S. Kieft, *Viral RNA structure-based strategies to manipulate translation.* Nature Reviews Microbiology, 2019. **17**(2): p. 110-123.
30. Kieft, J.S., *Viral IRES RNA structures and ribosome interactions.* Trends Biochem Sci, 2008. **33**(6): p. 274-83.
31. Balvay, L., et al., *Structural and functional diversity of viral IRESes.* Biochim Biophys Acta, 2009. **1789**(9-10): p. 542-57.
32. Kwan, T. and S.R. Thompson, *Noncanonical Translation Initiation in Eukaryotes.* Cold Spring Harb Perspect Biol, 2019. **11**(4).
33. Kerr, C.H., et al., *IRES-dependent ribosome repositioning directs translation of a +1 overlapping ORF that enhances viral infection.* Nucleic Acids Res, 2018. **46**(22): p. 11952-11967.
34. Walewski, J.L., et al., *Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame.* Rna, 2001. **7**(5): p. 710-21.
35. van Eyll, O. and T. Michiels, *Non-AUG-initiated internal translation of the L* protein of Theiler's virus and importance of this protein for viral persistence.* J Virol, 2002. **76**(21): p. 10665-73.
36. Kong, W.P. and R.P. Roos, *Alternative translation initiation site in the DA strain of Theiler's murine encephalomyelitis virus.* J Virol, 1991. **65**(6): p. 3395-9.
37. Pooggin, M.M. and L.A. Ryabova, *Ribosome Shunting, Polycistronic Translation, and Evasion of Antiviral Defenses in Plant Pararetroviruses and Beyond.* Frontiers in Microbiology, 2018. **9**.
38. Ryabova, L.A. and T. Hohn, *Ribosome shunting in the cauliflower mosaic virus 35S RNA leader is a special case of reinitiation of translation functioning in plant and animal systems.* Genes Dev, 2000. **14**(7): p. 817-29.
39. Yueh, A. and R.J. Schneider, *Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA.* Genes Dev, 2000. **14**(4): p. 414-21.
40. Kozak, M., *Pushing the limits of the scanning mechanism for initiation of translation.* Gene, 2002. **299**(1-2): p. 1-34.
41. Kozak, M., *Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes.* Cell, 1986. **44**(2): p. 283-92.
42. Dinesh-Kumar, S.P. and W.A. Miller, *Control of start codon choice on a plant viral RNA encoding overlapping genes.* Plant Cell, 1993. **5**(6): p. 679-92.
43. Fütterer, J., et al., *Rice tungro bacilliform virus open reading frames II and III are translated from polycistronic pregenomic RNA by leaky scanning.* J Virol, 1997. **71**(10): p. 7984-9.
44. Williams, M.A. and R.A. Lamb, *Effect of mutations and deletions in a bicistronic mRNA on the synthesis of influenza B virus NB and NA glycoproteins.* J Virol, 1989. **63**(1): p. 28-35.
45. Powell, Michael L., T.David K. Brown, and I. Brierley, *Translational termination–re-initiation in viral systems.* Biochemical Society Transactions, 2008. **36**(4): p. 717-722.
46. Jackson, R.J., C.U. Hellen, and T.V. Pestova, *Termination and post-termination events in eukaryotic translation.* Adv Protein Chem Struct Biol, 2012. **86**: p. 45-93.
47. Luttermann, C. and G. Meyers, *A bipartite sequence motif induces translation reinitiation in feline calicivirus RNA.* J Biol Chem, 2007. **282**(10): p. 7056-65.

48. Meyers, G., *Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependent reinitiation mechanism.* J Biol Chem, 2003. **278**(36): p. 34051-60.
49. Horvath, C.M., M.A. Williams, and R.A. Lamb, *Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide.* Embo j, 1990. **9**(8): p. 2639-47.
50. Napthine, S., et al., *Expression of the VP2 protein of murine norovirus by a translation termination-reinitiation strategy.* PLoS One, 2009. **4**(12): p. e8390.
51. Powell, M.L., et al., *Further characterisation of the translational termination-reinitiation signal of the influenza B virus segment 7 RNA.* PLoS One, 2011. **6**(2): p. e16822.
52. Rodnina, M.V., et al., *Translational recoding: canonical translation mechanisms reinterpreted.* Nucleic Acids Res, 2020. **48**(3): p. 1056-1067.
53. Namy, O., et al., *Translational Recoding and RNA Modifications*, in *Fine-Tuning of RNA Functions by Modification and Editing*, H. Grosjean, Editor. 2005, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 309-340.
54. Atkins, J.F. and R.F. Gesteland, *Recoding: expansion of decoding rules enriches gene expression*. Vol. 24. 2010: Springer Science & Business Media.
55. Manjunath, L.E., et al., *Mammalian proteome expansion by stop codon readthrough.* Wiley Interdiscip Rev RNA, 2022: p. e1739.
56. Beier, H. and M. Grimm, *Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs.* Nucleic Acids Res, 2001. **29**(23): p. 4767-82.
57. Namy, O. and J.-P. Rousset, *Specification of Standard Amino Acids by Stop Codons*, in *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, J.F. Atkins and R.F. Gesteland, Editors. 2010, Springer New York: New York, NY. p. 79-100.
58. Brault, V., et al., *Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74.* Embo j, 1995. **14**(4): p. 650-9.
59. Greiss, S. and J.W. Chin, *Expanding the genetic code of an animal.* J Am Chem Soc, 2011. **133**(36): p. 14196-9.
60. Brugère, J.F., et al., *Pyrrolysine in archaea: a 22nd amino acid encoded through a genetic code expansion.* Emerg Top Life Sci, 2018. **2**(4): p. 607-618.
61. De Lise, F., et al., *Programmed Deviations of Ribosomes From Standard Decoding in Archaea.* Front Microbiol, 2021. **12**: p. 688061.
62. de la Torre, D. and J.W. Chin, *Reprogramming the genetic code.* Nature Reviews Genetics, 2021. **22**(3): p. 169-184.
63. Baranov, P.V., J.F. Atkins, and M.M. Yordanova, *Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning.* Nat Rev Genet, 2015. **16**(9): p. 517-29.
64. Huang, W.M., et al., *A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60.* Science, 1988. **239**(4843): p. 1005-12.
65. Weiss, R.B., W.M. Huang, and D.M. Dunn, *A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60.* Cell, 1990. **62**(1): p. 117-26.
66. Agirrezabala, X., et al., *Ribosome rearrangements at the onset of translational bypassing.* Science Advances, 2017. **3**(6): p. e1700147.
67. Samatova, E., et al., *High-efficiency translational bypassing of non-coding nucleotides specified by mRNA structure and nascent peptide.* Nature Communications, 2014. **5**(1): p. 4459.
68. Brown, J. and M. Ryan, *Ribosome "Skipping": "Stop-Carry On" or "StopGo" Translation*. 2010. p. 101-121.
69. Atkins, J.F., et al., *A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go).* Rna, 2007. **13**(6): p. 803-10.
70. Doronina, V.A., et al., *Site-specific release of nascent chains from ribosomes at a sense codon.* Mol Cell Biol, 2008. **28**(13): p. 4227-39.
71. Donnelly, M.L.L., et al., *Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'.* J Gen Virol, 2001. **82**(Pt 5): p. 1013-1025.

72. Riegger, R.J. and N. Caliskan, *Thinking Outside the Frame: Impacting Genomes Capacity by Programmed Ribosomal Frameshifting.* Front Mol Biosci, 2022. **9**: p. 842261.
73. Penn, W.D. and S. Mukhopadhyay, *Abracadabra, One Becomes Two: The Importance of Context in Viral -1 Programmed Ribosomal Frameshifting.* mBio, 2022: p. e0246821.
74. Champagne, J., et al., *Slippy-Sloppy translation: a tale of programmed and induced-ribosomal frameshifting.* Trends Genet, 2022.
75. Penn, W.D., et al., *Regulators of Viral Frameshifting: More Than RNA Influences Translation Events.* Annu Rev Virol, 2020. **7**(1): p. 219-238.
76. Advani, V.M. and J.D. Dinman, *Reprogramming the genetic code: The emerging role of ribosomal frameshifting in regulating cellular gene expression.* BioEssays, 2016. **38**(1): p. 21-26.
77. Caliskan, N., F. Peske, and M.V. Rodnina, *Changed in translation: mRNA recoding by -1 programmed ribosomal frameshifting.* Trends Biochem Sci, 2015. **40**(5): p. 265-74.
78. Dinman, J.D., *Mechanisms and implications of programmed translational frameshifting.* WIREs RNA, 2012. **3**(5): p. 661-673.
79. Ketteler, R., *On programmed ribosomal frameshifting: the alternative proteomes.* Front Genet, 2012. **3**: p. 242.
80. Firth, A.E., et al., *Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the +1 direction.* Open Biology, 2012. **2**(10): p. 120109.
81. Jagger, B.W., et al., *An overlapping protein-coding region in influenza A virus segment 3 modulates the host response.* Science, 2012. **337**(6091): p. 199-204.
82. Agranovsky, A.A., et al., *Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease.* Virology, 1994. **198**(1): p. 311-24.
83. Jacks, T. and H.E. Varmus, *Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting.* Science, 1985. **230**(4731): p. 1237-42.
84. Jacks, T., et al., *Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region.* Cell, 1988. **55**(3): p. 447-58.
85. Brierley, I., P. Digard, and S.C. Inglis, *Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot.* Cell, 1989. **57**(4): p. 537-47.
86. Brierley, I., R.J.C. Gilbert, and S. Pennell, *Pseudoknot-Dependent Programmed —1 Ribosomal Frameshifting: Structures, Mechanisms and Models*, in *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, J.F. Atkins and R.F. Gesteland, Editors. 2010, Springer New York: New York, NY. p. 149-174.
87. Caliskan, N., et al., *Programmed -1 frameshifting by kinetic partitioning during impeded translocation.* Cell, 2014. **157**(7): p. 1619-31.
88. Kim, H.K., et al., *A frameshifting stimulatory stem loop destabilizes the hybrid state and impedes ribosomal translocation.* Proc Natl Acad Sci U S A, 2014. **111**(15): p. 5538-43.
89. Kontos, H., S. Napthine, and I. Brierley, *Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency.* Mol Cell Biol, 2001. **21**(24): p. 8657-70.
90. Howard, M.T., R.F. Gesteland, and J.F. Atkins, *Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides.* Rna, 2004. **10**(10): p. 1653-61.
91. Lin, Z., R.J. Gilbert, and I. Brierley, *Spacer-length dependence of programmed -1 or -2 ribosomal frameshifting on a U6A heptamer supports a role for messenger RNA (mRNA) tension in frameshifting.* Nucleic Acids Res, 2012. **40**(17): p. 8674-89.
92. Lewis, T.L. and S.M. Matsui, *An astrovirus frameshift signal induces ribosomal frameshifting in vitro.* Archives of Virology, 1995. **140**(6): p. 1127-1135.
93. Kelly, J.A., et al., *Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2).* The Journal of Biological Chemistry, 2020. **295**(31): p. 10741-10748.
94. Cho, C.-P., et al., *Regulation of Programmed Ribosomal Frameshifting by Co-Translational Refolding RNA Hairpins.* PLOS ONE, 2013. **8**(4): p. e62283.

95. Napthine, S., et al., *The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting.* J Mol Biol, 1999. **288**(3): p. 305-20.
96. Napthine, S., et al., *Protein-directed ribosomal frameshifting temporally regulates gene expression.* Nature Communications, 2017. **8**(1): p. 15582.
97. Mouzakis, K.D., et al., *HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome.* Nucleic Acids Res, 2013. **41**(3): p. 1901-13.
98. Hill, C.H., et al., *Investigating molecular mechanisms of 2A-stimulated ribosomal pausing and frameshifting in Theilovirus.* Nucleic Acids Research, 2021. **49**(20): p. 11938-11958.
99. Hill, C.H., et al., *Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch.* Nature Communications, 2021. **12**(1): p. 7166.
100. Loughran, G., E. Firth Andrew, and F. Atkins John, *Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome.* Proceedings of the National Academy of Sciences, 2011. **108**(46): p. E1111-E1119.
101. Napthine, S., et al., *Characterization of the stimulators of protein-directed ribosomal frameshifting in Theiler's murine encephalomyelitis virus.* Nucleic Acids Res, 2019. **47**(15): p. 8207-8223.
102. Napthine, S., et al., *A novel role for poly(C) binding proteins in programmed ribosomal frameshifting.* Nucleic Acids Res, 2016. **44**(12): p. 5491-503.
103. Cook, G.M., et al., *Ribosome profiling of porcine reproductive and respiratory syndrome virus reveals novel features of viral gene expression.* eLife, 2022. **11**: p. e75668.
104. Zimmer, M.M., et al., *The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting.* Nature Communications, 2021. **12**(1): p. 7193.
105. Wang, X., et al., *Regulation of HIV-1 Gag-Pol Expression by Shiftless, an Inhibitor of Programmed -1 Ribosomal Frameshifting.* Cell, 2019. **176**(3): p. 625-635.e14.
106. Kwak, H., M.W. Park, and S. Jeong, *Annexin A2 binds RNA and reduces the frameshifting efficiency of infectious bronchitis virus.* PLoS One, 2011. **6**(8): p. e24067.
107. Kobayashi, Y., et al., *Identification of a cellular factor that modulates HIV-1 programmed ribosomal frameshifting.* J Biol Chem, 2010. **285**(26): p. 19776-84.
108. Harrington, H.R., et al., *Cotranslational folding stimulates programmed ribosomal frameshifting in the alphavirus structural polyprotein.* J Biol Chem, 2020. **295**(20): p. 6798-6808.
109. Carmody, P.J., et al., *Coordination of -1 programmed ribosomal frameshifting by transcript and nascent chain features revealed by deep mutational scanning.* Nucleic Acids Research, 2021. **49**(22): p. 12943-12954.
110. Bhatt Pramod, R., et al., *Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome.* Science, 2021. **372**(6548): p. 1306-1313.
111. Yu, C.H., M.H. Noteborn, and R.C. Olsthoorn, *Stimulation of ribosomal frameshifting by antisense LNA.* Nucleic Acids Res, 2010. **38**(22): p. 8277-83.
112. Zhang, K., et al., *Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome.* Nature Structural & Molecular Biology, 2021. **28**(9): p. 747-754.
113. Olsthoorn, R.C., et al., *Novel application of sRNA: stimulation of ribosomal frameshifting.* Rna, 2004. **10**(11): p. 1702-3.
114. Matsumoto, S., et al., *Small synthetic molecule-stabilized RNA pseudoknot as an activator for -1 ribosomal frameshifting.* Nucleic Acids Research, 2018. **46**(16): p. 8079-8089.
115. Cardno, T.S., et al., *HIV-1 and Human PEG10 Frameshift Elements Are Functionally Distinct and Distinguished by Novel Small Molecule Modulators.* Plos One, 2015. **10**(10): p. e0139036.
116. Marcheschi, R.J., K.D. Mouzakis, and S.E. Butcher, *Selection and characterization of small molecules that bind the HIV-1 frameshift site RNA.* ACS Chemical Biology, 2009. **4**(10): p. 844-854.

117. Park, S.-J., Y.-G. Kim, and H.-J. Park, *Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening.* Journal of the American Chemical Society, 2011. **133**(26): p. 10094-10100.
118. Park, S.J., et al., *Identification of novel ligands for the RNA pseudoknot that regulate -1 ribosomal frameshifting.* Bioorg Med Chem, 2008. **16**(8): p. 4676-84.
119. Ahn, D.-G., et al., *A Novel Frameshifting Inhibitor Having Antiviral Activity against Zoonotic Coronaviruses.* Viruses, 2021. **13**(8): p. 1639.
120. Munshi, S., et al., *Identifying Inhibitors of −1 Programmed Ribosomal Frameshifting in a Broad Spectrum of Coronaviruses.* Viruses, 2022. **14**(2).
121. Sun, Y., et al., *Restriction of SARS-CoV-2 replication by targeting programmed -1 ribosomal frameshifting.* Proceedings of the National Academy of Sciences of the United States of America, 2021. **118**(26).
122. Kibe, A., et al., *Screening of Natural Products and Small Molecules Uncovers Novel Coronavirus 1a/1b Frameshifting Inhibitors with Antiviral Properties.* SSRN Electronic Journal, 2022.
123. Choi, J., et al., *The energy landscape of −1 ribosomal frameshifting.* Science Advances, 2020. **6**(1): p. eaax6969.
124. Chen, J., et al., *Dynamic pathways of -1 translational frameshifting.* Nature, 2014. **512**(7514): p. 328-32.
125. Plant, E.P. and J.D. Dinman, *Torsional restraint: a new twist on frameshifting pseudoknots.* Nucleic Acids Research, 2005. **33**(6): p. 1825-1833.
126. Wu, B., et al., *Translocation kinetics and structural dynamics of ribosomes are modulated by the conformational plasticity of downstream pseudoknots.* Nucleic Acids Res, 2018. **46**(18): p. 9736-9748.
127. Nguyen, H.A., E.D. Hoffer, and C.M. Dunham, *Importance of a tRNA anticodon loop modification and a conserved, noncanonical anticodon stem pairing in tRNACGGProfor decoding.* J Biol Chem, 2019. **294**(14): p. 5281-5291.
128. Moomau, C., et al., *Structural and Functional Characterization of Programmed Ribosomal Frameshift Signals in West Nile Virus Strains Reveals High Structural Plasticity Among cis-Acting RNA Elements.* J Biol Chem, 2016. **291**(30): p. 15788-95.
129. Ritchie, D.B., D.A. Foster, and M.T. Woodside, *Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding.* Proc Natl Acad Sci U S A, 2012. **109**(40): p. 16167-72.
130. Halma, M.T.J., et al., *Complex dynamics under tension in a high-efficiency frameshift stimulatory structure.* Proc Natl Acad Sci U S A, 2019. **116**(39): p. 19500-19505.
131. Bao, C., et al., *mRNA stem-loops can pause the ribosome by hindering A-site tRNA binding.* Elife, 2020. **9**.
132. Bock, L.V., et al., *Thermodynamic control of −1 programmed ribosomal frameshifting.* Nature Communications, 2019. **10**(1): p. 4598.
133. Temperley, R., et al., *Hungry codons promote frameshifting in human mitochondrial ribosomes.* Science, 2010. **327**(5963): p. 301.
134. Korniy, N., et al., *Modulation of HIV-1 Gag/Gag-Pol frameshifting by tRNA abundance.* Nucleic Acids Research, 2019. **47**(10): p. 5210-5222.
135. Gallant, J.A. and D. Lindsley, *Leftward ribosome frameshifting at a hungry codon.* J Mol Biol, 1992. **223**(1): p. 31-40.
136. Olubajo, B. and E.W. Taylor, *A -1 frameshift in the HIV-1 env gene is enhanced by arginine deficiency via a hungry codon mechanism.* Mutat Res, 2005. **579**(1-2): p. 125-32.
137. Caliskan, N., et al., *Conditional Switch between Frameshifting Regimes upon Translation of dnaX mRNA.* Mol Cell, 2017. **66**(4): p. 558-567.e4.
138. Barak, Z., D. Lindsley, and J. Gallant, *On the Mechanism of Leftward Frameshifting at Several Hungry Codons.* Journal of Molecular Biology, 1996. **256**(4): p. 676-684.
139. Ingolia, N.T., et al., *Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling.* Science, 2009. **324**(5924): p. 218-23.

140. Ingolia, N.T., L.F. Lareau, and J.S. Weissman, *Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes.* Cell, 2011. **147**(4): p. 789-802.
141. Steitz, J.A., *Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA.* Nature, 1969. **224**(5223): p. 957-64.
142. Wolin, S.L. and P. Walter, *Ribosome pausing and stacking during translation of a eukaryotic mRNA.* Embo j, 1988. **7**(11): p. 3559-69.
143. Brar, G.A. and J.S. Weissman, *Ribosome profiling reveals the what, when, where and how of protein synthesis.* Nat Rev Mol Cell Biol, 2015. **16**(11): p. 651-64.
144. Fresno, M., A. Jiménez, and D. Vázquez, *Inhibition of translation in eukaryotic systems by harringtonine.* Eur J Biochem, 1977. **72**(2): p. 323-30.
145. Gao, X., et al., *Quantitative profiling of initiating ribosomes in vivo.* Nat Methods, 2015. **12**(2): p. 147-53.
146. Lee, S., et al., *Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution.* Proceedings of the National Academy of Sciences, 2012. **109**(37): p. E2424-E2432.
147. Schneider-Poetsch, T., et al., *Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin.* Nature Chemical Biology, 2010. **6**(3): p. 209-217.
148. Zhang, H., et al., *Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development.* PLoS Biol, 2018. **16**(7): p. e2003903.
149. Alexander, M.R., et al., *Ribosome-Profiling Reveals Restricted Post Transcriptional Expression of Antiviral Cytokines and Transcription Factors during SARS-CoV-2 Infection.* Int J Mol Sci, 2021. **22**(7).
150. Finkel, Y., et al., *SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis.* Nature, 2021. **594**(7862): p. 240-245.
151. Finkel, Y., et al., *The coding capacity of SARS-CoV-2.* Nature, 2021. **589**(7840): p. 125-130.
152. Yuan, S., et al., *Multiomics interrogation into HBV (Hepatitis B virus)-host interaction reveals novel coding potential in human genome, and identifies canonical and non-canonical proteins as host restriction factors against HBV.* Cell Discovery, 2021. **7**(1): p. 105.
153. Dinan, A.M., et al., *Comparative Analysis of Gene Expression in Virulent and Attenuated Strains of Infectious Bronchitis Virus at Subcodon Resolution.* J Virol, 2019. **93**(18).
154. Irigoyen, N., et al., *Ribosome profiling of the retrovirus murine leukemia virus.* Retrovirology, 2018. **15**(1): p. 10.
155. Bencun, M., et al., *Translational profiling of B cells infected with the Epstein-Barr virus reveals 5' leader ribosome recruitment through upstream open reading frames.* Nucleic Acids Research, 2018. **46**(6): p. 2802-2819.
156. Irigoyen, N., et al., *High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling.* PLOS Pathogens, 2016. **12**(2): p. e1005473.
157. Puray-Chavez, M., et al., *The Translational Landscape of SARS-CoV-2-infected Cells Reveals Suppression of Innate Immune Genes.* mBio, 2022. **13**(3): p. e00815-22.
158. Hoang, H.-D., T.E. Graber, and T. Alain, *Battling for Ribosomes: Translational Control at the Forefront of the Antiviral Response.* Journal of Molecular Biology, 2018. **430**(14): p. 1965-1992.
159. Echavarría-Consuegra, L., et al., *Manipulation of the unfolded protein response: A pharmacological strategy against coronavirus infection.* PLOS Pathogens, 2021. **17**(6): p. e1009644.
160. Charbonneau, J., et al., *The 5' UTR of HIV-1 full-length mRNA and the Tat viral protein modulate the programmed -1 ribosomal frameshift that generates HIV-1 enzymes.* Rna, 2012. **18**(3): p. 519-29.
161. Chen, H.H. and W.Y. Tarn, *uORF-mediated translational control: recently elucidated mechanisms and implications in cancer.* RNA Biol, 2019. **16**(10): p. 1327-1338.

162. Young, S.K. and R.C. Wek, *Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response.* J Biol Chem, 2016. **291**(33): p. 16927-35.
163. Lee, D.S.M., et al., *Disrupting upstream translation in mRNAs is associated with human disease.* Nat Commun, 2021. **12**(1): p. 1515.
164. Starck, S.R., et al., *Translation from the 5' untranslated region shapes the integrated stress response.* Science, 2016. **351**(6272): p. aad3867.
165. Moro, S.G., et al., *Impact of uORFs in mediating regulation of translation in stress conditions.* BMC Mol Cell Biol, 2021. **22**(1): p. 29.
166. Sadler, A.J. and B.R. Williams, *Interferon-inducible antiviral effectors.* Nat Rev Immunol, 2008. **8**(7): p. 559-68.
167. Elde, N.C., et al., *Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses.* Cell, 2012. **150**(4): p. 831-41.
168. Rothenburg, S., et al., *Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors.* Nat Struct Mol Biol, 2009. **16**(1): p. 63-70.
169. Andreev, D.E., et al., *Translation of 5' leaders is pervasive in genes resistant to eIF2 repression.* eLife, 2015. **4**: p. e03971.
170. Khitun, A., T.J. Ness, and S.A. Slavoff, *Small open reading frames and cellular stress responses.* Molecular Omics, 2019. **15**(2): p. 108-116.
171. Belew, A.T., et al., *PRFdb: a database of computationally predicted eukaryotic programmed -1 ribosomal frameshift signals.* BMC Genomics, 2008. **9**: p. 339.
172. Clark, M.B., et al., *Mammalian gene PEG10 expresses two reading frames by high efficiency -1 frameshifting in embryonic-associated tissues.* J Biol Chem, 2007. **282**(52): p. 37359-69.
173. Manktelow, E., K. Shigemoto, and I. Brierley, *Characterization of the frameshift signal of Edr, a mammalian example of programmed -1 ribosomal frameshifting.* Nucleic Acids Res, 2005. **33**(5): p. 1553-63.
174. Wills, N.M., et al., *A functional -1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene.* J Biol Chem, 2006. **281**(11): p. 7082-8.
175. Emery, A. and R. Swanstrom, *HIV-1: To Splice or Not to Splice, That Is the Question.* Viruses, 2021. **13**(2).
176. Martin Stoltzfus, C., *Chapter 1 Regulation of HIV-1 Alternative RNA Splicing and Its Role in Virus Replication*, in *Advances in Virus Research*. 2009, Academic Press. p. 1-40.
177. Karn, J. and C.M. Stoltzfus, *Transcriptional and posttranscriptional regulation of HIV-1 gene expression.* Cold Spring Harb Perspect Med, 2012. **2**(2): p. a006916.
178. Emery, A., et al., *Characterizing HIV-1 Splicing by Using Next-Generation Sequencing.* J Virol, 2017. **91**(6).
179. Ocwieja, K.E., et al., *Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing.* Nucleic Acids Res, 2012. **40**(20): p. 10345-55.
180. Nguyen Quang, N., et al., *Dynamic nanopore long-read sequencing analysis of HIV-1 splicing events during the early steps of infection.* Retrovirology, 2020. **17**(1): p. 25.
181. Guerrero, S., et al., *HIV-1 replication and the cellular eukaryotic translation apparatus.* Viruses, 2015. **7**(1): p. 199-218.
182. de Breyne, S. and T. Ohlmann, *Focus on Translation Initiation of the HIV-1 mRNAs.* Int J Mol Sci, 2018. **20**(1).
183. Ohlmann, T., C. Mengardi, and M. López-Lastra, *Translation initiation of the HIV-1 mRNA.* Translation (Austin), 2014. **2**(2): p. e960242.
184. Ventoso, I., et al., *HIV-1 protease cleaves eukaryotic initiation factor 4G and inhibits cap-dependent translation.* Proc Natl Acad Sci U S A, 2001. **98**(23): p. 12966-71.
185. Ricci, E.P., et al., *Lentiviral RNAs can use different mechanisms for translation initiation.* Biochem Soc Trans, 2008. **36**(Pt 4): p. 690-3.
186. Yilmaz, A., C. Bolinger, and K. Boris-Lawrie, *Retrovirus translation initiation: Issues and hypotheses derived from study of HIV-1.* Curr HIV Res, 2006. **4**(2): p. 131-9.

187. Daudé, C., et al., *HIV-1 sequences isolated from patients promote expression of shorter isoforms of the Gag polyprotein.* Archives of Virology, 2016. **161**(12): p. 3495-3507.
188. Herbreteau, C.H., et al., *HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon.* Nature Structural & Molecular Biology, 2005. **12**(11): p. 1001-1007.
189. Locker, N., N. Chamond, and B. Sargueil, *A conserved structure within the HIV gag open reading frame that controls translation initiation directly recruits the 40S subunit and eIF3.* Nucleic Acids Research, 2011. **39**(6): p. 2367-2377.
190. Plank, T.-D.M., J.T. Whitehurst, and J.S. Kieft, *Cell type specificity and structural determinants of IRES activity from the 5' leaders of different HIV-1 transcripts.* Nucleic Acids Research, 2013. **41**(13): p. 6698-6714.
191. Vallejos, M., et al., *Functional and Structural Analysis of the Internal Ribosome Entry Site Present in the mRNA of Natural Variants of the HIV-1.* PLOS ONE, 2012. **7**(4): p. e35031.
192. Deforges, J., et al., *Two ribosome recruitment sites direct multiple translation events within HIV1 Gag open reading frame.* 2017. **45**(12): p. 7382-7400.
193. Gendron, K., et al., *The activity of the HIV-1 IRES is stimulated by oxidative stress and controlled by a negative regulatory element.* Nucleic Acids Res, 2011. **39**(3): p. 902-12.
194. Amorim, R., et al., *HIV-1 transcripts use IRES-initiation under conditions where Cap-dependent translation is restricted by poliovirus 2A protease.* PLoS One, 2014. **9**(2): p. e88619.
195. Álvarez, E., L. Menéndez-Arias, and L. Carrasco, *The Eukaryotic Translation Initiation Factor 4GI Is Cleaved by Different Retroviral Proteases.* Journal of Virology, 2003. **77**(23): p. 12392-12400.
196. Castelló, A., et al., *HIV- 1 Protease Inhibits Cap- and Poly(A)-Dependent Translation upon eIF4GI and PABP Cleavage.* PLOS ONE, 2009. **4**(11): p. e7997.
197. Prévôt, D., et al., *Characterization of a novel RNA-binding region of eIF4GI critical for ribosomal scanning.* The EMBO Journal, 2003. **22**(8): p. 1909-1921.
198. Ohlmann, T., et al., *In Vitro Cleavage of eIF4GI but not eIF4GII by HIV-1 Protease and its Effects on Translation in the Rabbit Reticulocyte Lysate System.* Journal of Molecular Biology, 2002. **318**(1): p. 9-20.
199. Kleinman, C.L., et al., *HIV-1 infection causes a down-regulation of genes involved in ribosome biogenesis.* PLoS One, 2014. **9**(12): p. e113908.
200. Singh, G., et al., *HIV-1 hypermethylated guanosine cap licenses specialized translation unaffected by mTOR.* Proceedings of the National Academy of Sciences, 2022. **119**(1): p. e2105153118.
201. Jacks, T., et al., *Characterization of ribosomal frameshifting in HIV-1 gag-pol expression.* Nature, 1988. **331**(6153): p. 280-3.
202. Low, J.T., et al., *Structure and Dynamics of the HIV-1 Frameshift Element RNA.* Biochemistry, 2014. **53**(26): p. 4282-4291.
203. Watts, J.M., et al., *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-6.
204. Staple, D.W. and S.E. Butcher, *Solution structure of the HIV-1 frameshift inducing stem-loop RNA.* Nucleic Acids Res, 2003. **31**(15): p. 4326-31.
205. Dinman, J.D., et al., *The frameshift signal of HIV-1 involves a potential intramolecular triplex RNA structure.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5331-6.
206. Huang, X., et al., *Highly conserved RNA pseudoknots at the Gag-Pol junction of HIV-1 suggest a novel mechanism of -1 ribosomal frameshifting.* Rna, 2014. **20**(5): p. 587-93.
207. Bauby, H., et al., *HIV-1 Vpr Induces Widespread Transcriptomic Changes in CD4+ T Cells Early Postinfection.* mBio, 2021. **12**(3): p. e01369-21.
208. Golumbeanu, M., et al., *Proteo-Transcriptomic Dynamics of Cellular Response to HIV-1 Infection.* Scientific Reports, 2019. **9**(1): p. 213.
209. Zhang, W., et al., *Transcriptomics and Targeted Proteomics Analysis to Gain Insights Into the Immune-control Mechanisms of HIV-1 Infected Elite Controllers.* EBioMedicine, 2018. **27**: p. 40-50.

210. Greenwood, E.J.D., et al., *Temporal proteomic analysis of HIV infection reveals remodelling of the host phosphoproteome by lentiviral Vif variants.* eLife, 2016. **5**: p. e18296.
211. Carlin, E., et al., *Extensive proteomic and transcriptomic changes quench the TCR/CD3 activation signal of latently HIV-1 infected T cells.* PLOS Pathogens, 2021. **17**(1): p. e1008748.
212. Hyrcza Martin, D., et al., *Distinct Transcriptional Profiles in Ex Vivo CD4+ and CD8+ T Cells Are Established Early in Human Immunodeficiency Virus Type 1 Infection and Are Characterized by a Chronic Interferon Response as Well as Extensive Transcriptional Changes in CD8+ T Cells.* Journal of Virology, 2007. **81**(7): p. 3477-3486.
213. Ingolia, N.T., et al., *The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments.* Nature Protocols, 2012. **7**(8): p. 1534-1550.
214. Mujawar, Z., et al., *Human immunodeficiency virus impairs reverse cholesterol transport from macrophages.* PLoS Biol, 2006. **4**(11): p. e365.
215. Rose, H., et al., *HIV infection and high density lipoprotein metabolism.* Atherosclerosis, 2008. **199**(1): p. 79-86.
216. Sharma, A., et al., *Thriving under stress: selective translation of HIV-1 structural protein mRNA during Vpr-mediated impairment of eIF4E translation activity.* PLoS Pathog, 2012. **8**(3): p. e1002612.
217. Calvo, S.E., D.J. Pagliarini, and V.K. Mootha, *Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans.* Proc Natl Acad Sci U S A, 2009. **106**(18): p. 7507-12.
218. Jin, X., et al., *The two upstream open reading frames of oncogene mdm2 have different translational regulatory properties.* J Biol Chem, 2003. **278**(28): p. 25716-21.
219. Raja, R., et al., *HIV-1 Tat potently stabilises Mdm2 and enhances viral replication.* Biochem J, 2017. **474**(14): p. 2449-2464.
220. Breton, Y., et al., *Expression of MDM2 in Macrophages Promotes the Early Postentry Steps of HIV-1 Infection through Inhibition of p53.* J Virol, 2019. **93**(7).
221. Golomb, L., et al., *Importin 7 and exportin 1 link c-Myc and p53 to regulation of ribosomal biogenesis.* Mol Cell, 2012. **45**(2): p. 222-32.
222. Yang, Y.-C. and B. Sugden, *Epstein-Barr Virus Limits the Accumulation of IPO7, an Essential Gene Product.* Frontiers in Microbiology, 2021. **12**.
223. Meydan, S. and N.R. Guydosh, *Disome and Trisome Profiling Reveal Genome-wide Targets of Ribosome Quality Control.* Molecular Cell, 2020. **79**(4): p. 588-602.e6.
224. Mito, M., Y. Mishima, and S. Iwasaki, *Protocol for Disome Profiling to Survey Ribosome Collision in Humans and Zebrafish.* STAR Protocols, 2020. **1**(3): p. 100168.
225. Krafczyk, R., et al., *Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria.* Communications Biology, 2021. **4**(1): p. 589.
226. Charneski, C.A. and L.D. Hurst, *Positively charged residues are the major determinants of ribosomal velocity.* PLoS Biol, 2013. **11**(3): p. e1001508.
227. Chyżyńska, K., et al., *Deep conservation of ribosome stall sites across RNA processing genes.* NAR Genomics and Bioinformatics, 2021. **3**(2): p. lqab038.
228. Koutmou, K.S., et al., *Ribosomes slide on lysine-encoding homopolymeric A stretches.* Elife, 2015. **4**.
229. Tesina, P., et al., *Molecular mechanism of translational stalling by inhibitory codon combinations and poly(A) tracts.* The EMBO Journal, 2020. **39**(3): p. e103365.
230. Torrent, M., et al., *Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions.* Science Signaling, 2018. **11**(546): p. eaat6409.
231. Nunes, A., et al., *Emerging Roles of tRNAs in RNA Virus Infections.* Trends Biochem Sci, 2020. **45**(9): p. 794-805.
232. van der Kuyl, A.C. and B. Berkhout, *The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus.* Retrovirology, 2012. **9**: p. 92.
233. van Weringh, A., et al., *HIV-1 Modulates the tRNA Pool to Improve Translation Efficiency.* Molecular Biology and Evolution, 2011. **28**(6): p. 1827-1834.

234. Kim, S.Y., et al., *Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression.* Journal of Virology, 1989. **63**(9): p. 3708-3713.
235. Wu, J., et al., *Asparagine enhances LCK signalling to potentiate CD8+ T-cell activation and anti-tumour responses.* Nature Cell Biology, 2021. **23**(1): p. 75-86.
236. Ron-Harel, N., et al., *T Cell Activation Depends on Extracellular Alanine.* Cell Reports, 2019. **28**(12): p. 3011-3021.e4.
237. Matheson, N.J., et al., *Cell Surface Proteomic Map of HIV Infection Reveals Antagonism of Amino Acid Metabolism by Vpu and Nef.* Cell Host Microbe, 2015. **18**(4): p. 409-23.
238. Gu, C., et al., *Isoleucine Plays an Important Role for Maintaining Immune Function.* Current Protein & Peptide Science, 2019. **20**(7): p. 644-651.
239. Lu, J. and C. Deutsch, *Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates.* Journal of Molecular Biology, 2008. **384**(1): p. 73-86.
240. Yanagitani, K., et al., *Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA.* Science (New York, N.Y.), 2011. **331**(6017): p. 586-589.
241. Tenson, T. and M. Ehrenberg, *Regulatory Nascent Peptides in the Ribosomal Tunnel.* Cell, 2002. **108**(5): p. 591-594.
242. Woolstenhulme, C.J., et al., *Nascent peptides that block protein synthesis in bacteria.* Proc Natl Acad Sci U S A, 2013. **110**(10): p. E878-87.
243. Sabi, R. and T. Tuller, *A comparative genomics study on the effect of individual amino acids on ribosome stalling.* BMC Genomics, 2015. **16**(10): p. S5.
244. Cymer, F., et al., *Exploration of the arrest peptide sequence space reveals arrest-enhanced variants.* J Biol Chem, 2015. **290**(16): p. 10208-15.
245. Ito, K. and S. Chiba, *Arrest Peptides: Cis-Acting Modulators of Translation.* Annual Review of Biochemistry, 2013. **82**(1): p. 171-202.
246. Li, M., et al., *Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11.* Nature, 2012. **491**(7422): p. 125-128.
247. Cobos Jiménez, V., et al., *G3BP1 restricts HIV-1 replication in macrophages and T-cells by sequestering viral RNA.* Virology, 2015. **486**: p. 94-104.
248. Sithole, N., et al., *The roles of DEAD box helicases in the life cycle of HIV-1.* Lancet, 2015. **385 Suppl 1**: p. S89.
249. Lee, W.J., et al., *IFITM proteins inhibit HIV-1 protein synthesis.* Sci Rep, 2018. **8**(1): p. 14551.
250. Chaudhuri, A., et al., *STAT1 signaling modulates HIV-1-induced inflammatory responses and leukocyte transmigration across the blood-brain barrier.* Blood, 2008. **111**(4): p. 2062-72.
251. Gerber, P.P., et al., *Rab27a controls HIV-1 assembly by regulating plasma membrane levels of phosphatidylinositol 4,5-bisphosphate.* Journal of Cell Biology, 2015. **209**(3): p. 435-452.
252. Ciucci, T. and R. Bosselut, *Gimap and T cells: a matter of life or death.* Eur J Immunol, 2014. **44**(2): p. 348-51.
253. Filén, S. and R. Lahesmaa, *GIMAP Proteins in T-Lymphocytes.* J Signal Transduct, 2010. **2010**: p. 268589.
254. Balistreri, G., C. Bognanni, and O. Mühlemann, *Virus Escape and Manipulation of Cellular Nonsense-Mediated mRNA Decay.* Viruses, 2017. **9**(1).
255. Zhang, H., Y. Wang, and J. Lu, *Function and Evolution of Upstream ORFs in Eukaryotes.* Trends in Biochemical Sciences, 2019. **44**(9): p. 782-794.
256. Hinnebusch, A.G., et al., *Translational control by 5'-untranslated regions of eukaryotic mRNAs.* Science (New York, N.Y.), 2016. **352**(6292): p. 1413-1416.
257. Wilkinson, K.A., et al., *High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states.* PLoS Biol, 2008. **6**(4): p. e96.
258. Brakier-Gingras, L., J. Charbonneau, and S.E. Butcher, *Targeting frameshifting in the human immunodeficiency virus.* Expert Opin Ther Targets, 2012. **16**(3): p. 249-58.

259. Guydosh, N.R. and R. Green, *Translation of poly(A) tails leads to precise mRNA cleavage.* Rna, 2017. **23**(5): p. 749-761.
260. Dimitrova, L.N., et al., *Nascent Peptide-dependent Translation Arrest Leads to Not4p-mediated Protein Degradation by the Proteasome*.* Journal of Biological Chemistry, 2009. **284**(16): p. 10343-10352.
261. Letzring, D.P., et al., *Translation of CGA codon repeats in yeast involves quality control components and ribosomal protein L1.* Rna, 2013. **19**(9): p. 1208-17.
262. McGlincy, N.J. and N.T. Ingolia, *Transcriptome-wide measurement of translation by ribosome profiling.* Methods, 2017. **126**: p. 112-129.
263. Erhard, F., et al., *Improved Ribo-seq enables identification of cryptic translation events.* Nature Methods, 2018. **15**(5): p. 363-366.
264. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-930.
265. Chothani, S., et al., *deltaTE: Detection of Translationally Regulated Genes by Integrative Analysis of Ribo-seq and RNA-seq Data.* Current Protocols in Molecular Biology, 2019. **129**(1): p. e108.
266. Cock, P.J.A., et al., *Biopython: freely available Python tools for computational molecular biology and bioinformatics.* Bioinformatics, 2009. **25**(11): p. 1422-1423.
267. Alexaki, A., et al., *Codon and Codon-Pair Usage Tables (CoCoPUTs): Facilitating Genetic Variation Analyses and Recombinant Gene Design.* Journal of Molecular Biology, 2019. **431**(13): p. 2434-2441.
268. Athey, J., et al., *A new and updated resource for codon usage tables.* BMC Bioinformatics, 2017. **18**(1): p. 391.
269. Andreou, A.I. and N. Nakayama, *Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly.* PLoS One, 2018. **13**(1): p. e0189892.
270. Tay, M.Z., et al., *The trinity of COVID-19: immunity, inflammation and intervention.* Nat Rev Immunol, 2020. **20**(6): p. 363-374.
271. Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.* Nature, 2020. **583**(7816): p. 459-468.
272. Schmidt, N., et al., *The SARS-CoV-2 RNA–protein interactome in infected human cells.* Nature Microbiology, 2021. **6**(3): p. 339-353.
273. Neupane, K., et al., *Anti-Frameshifting Ligand Active against SARS Coronavirus-2 Is Resistant to Natural Mutations of the Frameshift-Stimulatory Pseudoknot.* J Mol Biol, 2020. **432**(21): p. 5843-5847.
274. Baranov, P.V., et al., *Programmed ribosomal frameshifting in decoding the SARS-CoV genome.* Virology, 2005. **332**(2): p. 498-510.
275. Belew, A.T., et al., *Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway.* Nature, 2014. **512**(7514): p. 265-9.
276. Butter, F., et al., *Unbiased RNA-protein interaction screen by quantitative proteomics.* Proc Natl Acad Sci U S A, 2009. **106**(26): p. 10626-31.
277. Harcourt, J.L., et al., *Evaluation of the Calu-3 cell line as a model of in vitro respiratory syncytial virus infection.* 2011. **174**(1-2): p. 144-149.
278. Castello, A., et al., *Insights into RNA biology from an atlas of mammalian mRNA-binding proteins.* Cell, 2012. **149**(6): p. 1393-406.
279. Lee, S., et al., *The SARS-CoV-2 RNA interactome.* Mol Cell, 2021. **81**(13): p. 2838-2850.e6.
280. Kamel, W., et al., *Global analysis of protein-RNA interactions in SARS-CoV-2-infected cells reveals key regulators of infection.* Mol Cell, 2021. **81**(13): p. 2851-2867.e7.
281. Flynn, R.A., et al., *Discovery and functional interrogation of SARS-CoV-2 RNA-host protein interactions.* Cell, 2021. **184**(9): p. 2394-2411.e16.
282. Geuens, T., D. Bouhy, and V. Timmerman, *The hnRNP family: insights into their role in health and disease.* Hum Genet, 2016. **135**(8): p. 851-67.
283. Wu, J., et al., *Cryo-EM Structure of the Human Ribonuclease P Holoenzyme.* Cell, 2018. **175**(5): p. 1393-1404.e11.
284. Bick, M.J., et al., *Expression of the zinc-finger antiviral protein inhibits alphavirus replication.* J Virol, 2003. **77**(21): p. 11555-62.

285. Ray, P.S. and S. Das, *La autoantigen is required for the internal ribosome entry site-mediated translation of Coxsackievirus B3 RNA.* Nucleic Acids Res, 2002. **30**(20): p. 4500-8.
286. Weinlich, S., et al., *IGF2BP1 enhances HCV IRES-mediated translation initiation via the 3′ UTR.* Rna, 2009. **15**(8): p. 1528-1542.
287. Zhang, J., et al., *hnRNPs and ELAVL1 cooperate with uORFs to inhibit protein translation.* Nucleic Acids Res, 2017. **45**(5): p. 2849-2864.
288. Sauer, M., et al., *DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions.* Nat Commun, 2019. **10**(1): p. 2421.
289. Yoo, J.S., et al., *DHX36 enhances RIG-I signaling by facilitating PKR-mediated antiviral stress granule formation.* PLoS Pathog, 2014. **10**(3): p. e1004012.
290. Pietras, Z., et al., *Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria.* Nat Commun, 2018. **9**(1): p. 2558.
291. Schwerk, J., et al., *RNA-binding protein isoforms ZAP-S and ZAP-L have distinct antiviral and immune resolution functions.* Nat Immunol, 2019. **20**(12): p. 1610-1620.
292. Guo, X., et al., *The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs.* J Virol, 2004. **78**(23): p. 12781-7.
293. Guo, X., et al., *The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA.* Proc Natl Acad Sci U S A, 2007. **104**(1): p. 151-6.
294. Todorova, T., F.J. Bock, and P. Chang, *Poly(ADP-ribose) polymerase-13 and RNA regulation in immunity and cancer.* Trends Mol Med, 2015. **21**(6): p. 373-84.
295. Charron, G., et al., *Prenylome profiling reveals S-farnesylation is crucial for membrane targeting and antiviral activity of ZAP long-isoform.* Proc Natl Acad Sci U S A, 2013. **110**(27): p. 11085-90.
296. Vyas, S., et al., *A systematic analysis of the PARP protein family identifies new functions critical for cell physiology.* Nat Commun, 2013. **4**: p. 2240.
297. Kwak, J.E., et al., *Mammalian GLD-2 homologs are poly(A) polymerases.* Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4407-12.
298. Liang, X., et al., *Structural snapshots of human pre-60S ribosomal particles before and after nuclear export.* Nat Commun, 2020. **11**(1): p. 3542.
299. Blanco-Melo, D., et al., *Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19.* Cell, 2020. **181**(5): p. 1036-1045.e9.
300. Sun, L., et al., *In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs.* Cell, 2021. **184**(7): p. 1865-1883.e20.
301. Gonzalez-Perez, A.C., et al., *The Zinc Finger Antiviral Protein ZAP Restricts Human Cytomegalovirus and Selectively Binds and Destabilizes Viral UL4/UL5 Transcripts.* mBio, 2021. **12**(3).
302. Peng, C., et al., *Zinc-finger antiviral protein (ZAP) is a restriction factor for replication of modified vaccinia virus Ankara (MVA) in human cells.* PLoS Pathog, 2020. **16**(8): p. e1008845.
303. Nchioua, R., et al., *SARS-CoV-2 Is Restricted by Zinc Finger Antiviral Protein despite Preadaptation to the Low-CpG Environment in Humans.* mBio, 2020. **11**(5).
304. Napthine, S., et al., *Modulation of Viral Programmed Ribosomal Frameshifting and Stop Codon Readthrough by the Host Restriction Factor Shiftless.* Viruses, 2021. **13**(7).
305. Neupane, K., et al., *Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers.* Nat Commun, 2021. **12**(1): p. 4749.
306. Schlick, T., et al., *Structure-altering mutations of the SARS-CoV-2 frameshifting RNA element.* Biophysical Journal, 2021. **120**(6): p. 1040-1053.
307. Meagher, J.L., et al., *Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences.* Proc Natl Acad Sci U S A, 2019. **116**(48): p. 24303-24309.
308. Ficarelli, M., et al., *KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides.* Elife, 2019. **8**.

309. Li, M.M., et al., *TRIM25 Enhances the Antiviral Action of Zinc-Finger Antiviral Protein (ZAP).* PLoS Pathog, 2017. **13**(1): p. e1006145.
310. Luo, X., et al., *Molecular Mechanism of RNA Recognition by Zinc-Finger Antiviral Protein.* Cell Rep, 2020. **30**(1): p. 46-52.e4.
311. Turakhiya, A., et al., *ZFAND1 Recruits p97 and the 26S Proteasome to Promote the Clearance of Arsenite-Induced Stress Granules.* Mol Cell, 2018. **70**(5): p. 906-919.e7.
312. Mi, H., A. Muruganujan, and P.D. Thomas, *PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees.* Nucleic Acids Res, 2013. **41**(Database issue): p. D377-86.
313. Pan, D., et al., *CDK-regulated dimerization of M18BP1 on a Mis18 hexamer is necessary for CENP-A loading.* Elife, 2017. **6**.
314. Loughran, G., et al., *Avoidance of reporter assay distortions from fused dual reporters.* 2017. **23**(8): p. 1285-1289.
315. Arai, R., et al., *Design of the linkers which effectively separate domains of a bifunctional fusion protein.* Protein Eng, 2001. **14**(8): p. 529-32.
316. Meyer, S., et al., *Multi-host expression system for recombinant production of challenging proteins.* PLoS One, 2013. **8**(7): p. e68674.
317. Götzke, H., et al., *The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications.* Nat Commun, 2019. **10**(1): p. 4403.
318. Brennan, T.V., et al., *Generation of Luciferase-expressing Tumor Cell Lines.* Bio Protoc, 2018. **8**(8).
319. Stewart, S.A., et al., *Lentivirus-delivered stable gene silencing by RNAi in primary cells.* 2003. **9**(4): p. 493-501.
320. Grentzmann, G., et al., *A dual-luciferase reporter system for studying recoding signals.* 1998. **4**(4): p. 479-486.
321. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
322. Pisarev, A.V., et al., *Assembly and analysis of eukaryotic translation initiation complexes.* Methods Enzymol, 2007. **430**: p. 147-77.
323. Kingston, R.E., P. Chomczynski, and N. Sacchi, *Guanidine methods for total RNA preparation.* Curr Protoc Mol Biol, 2001. **Chapter 4**: p. Unit4.2.
324. Perez-Riverol, Y., et al., *The PRIDE database and related tools and resources in 2019: improving support for quantification data.* Nucleic Acids Res, 2019. **47**(D1): p. D442-d450.
325. Buck, S., L. Pekarek, and N. Caliskan, *POTATO: Automated pipeline for batch analysis of optical tweezers data.* Biophysical Journal, 2022. **121**(15): p. 2830-2839.
326. Loughran, G., A.E. Firth, and J.F. Atkins, *Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome.* Proc Natl Acad Sci U S A, 2011. **108**(46): p. E1111-9.
327. Napthine, S., et al., *Protein-directed ribosomal frameshifting temporally regulates gene expression.* Nat Commun, 2017. **8**: p. 15582.
328. Jackson, R.J., *A detailed kinetic analysis of the in vitro synthesis and processing of encephalomyocarditis virus products.* Virology, 1986. **149**(1): p. 114-27.
329. Hahn, H. and A.C. Palmenberg, *Deletion mapping of the encephalomyocarditis virus primary cleavage site.* J Virol, 2001. **75**(15): p. 7215-8.
330. Groppo, R. and A.C. Palmenberg, *Cardiovirus 2A protein associates with 40S but not 80S ribosome subunits during infection.* J Virol, 2007. **81**(23): p. 13067-74.
331. Carocci, M., et al., *Encephalomyocarditis virus 2A protein is required for viral pathogenesis and inhibition of apoptosis.* J Virol, 2011. **85**(20): p. 10741-54.
332. Merrick, W.C., *eIF4F: a retrospective.* J Biol Chem, 2015. **290**(40): p. 24091-9.
333. Lauring, A.S. and P.N. Malani, *Variants of SARS-CoV-2.* The Journal of the American Medical Association, 2021.
334. Jacobsen, H., et al., *Diminished neutralization responses towards SARS-CoV-2 Omicron VoC after mRNA or vector-based COVID-19 vaccinations.* medRxiv, 2021: p. 2021.12.21.21267898.

335. Banerjee, A., et al., *Bats and Coronaviruses.* Viruses, 2019. **11**(1): p. 41.
336. Karesh, W.B., et al., *Ecology of zoonoses: natural and unnatural histories.* (1474-547X (Electronic)).
337. Wu, A., et al., *Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China.* Cell Host & Microbe, 2020. **27**(3): p. 325-328.
338. Brierley, I., P. Digard, and S.C. Inglis, *Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot.* Cell, 1989. **57**(4): p. 537-547.
339. Rangan, R., et al., *RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look.* RNA, 2020. **26**(8): p. 937-959.
340. Ryder, S.P., et al., *Analysis of Emerging Variants in Structured Regions of the SARS-CoV-2 Genome.* Evolutionary Bioinformatics, 2021. **17**: p. 11769343211014167.
341. Plant, E.P., et al., *Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins.* (1098-5514 (Electronic)).
342. Dulude, D., et al., *Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1.* Virology, 2006. **345**(1): p. 127-136.
343. Brierley, I., et al., *Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal.* Journal of Molecular Biology, 1991. **220**(4): p. 889-902.
344. Bhatt, P.A.-O., et al., *Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome.* (1095-9203 (Electronic)).
345. Ofori, L.O., et al., *High-affinity recognition of HIV-1 frameshift-stimulating RNA alters frameshifting in vitro and interferes with HIV-1 infectivity.* Journal of Medicinal Chemistry, 2014. **57**(3): p. 723-732.
346. Haniff, H.S., et al., *Targeting the SARS-CoV-2 RNA Genome with Small Molecule Binders and Ribonuclease Targeting Chimera (RIBOTAC) Degraders.* ACS Central Science, 2020. **6**(10): p. 1713-1721.
347. Kelly, J.A., M.T. Woodside, and J.D. Dinman, *Programmed -1 Ribosomal Frameshifting in coronaviruses: A therapeutic target.* Virology, 2021. **554**: p. 75-82.
348. Manfredonia, I., et al., *Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements.* Nucleic Acids Research, 2020. **48**(22): p. 12436-12452.
349. Anokhina, V.S. and B.L. Miller, *Targeting Ribosomal Frameshifting as an Antiviral Strategy: From HIV-1 to SARS-CoV-2.* Accounts of Chemical Research, 2021. **54**(17): p. 3349-3361.
350. Zafferani, M., et al., *Amilorides inhibit SARS-CoV-2 replication in vitro by targeting RNA structures.* Science Advances. **7**(48): p. eabl6096.
351. Chen, Y., et al., *A drug screening toolkit based on the –1 ribosomal frameshifting of SARS-CoV-2.* Heliyon, 2020. **6**(8): p. e04793.
352. Ritchie, D.B., et al., *Anti-frameshifting Ligand Reduces the Conformational Plasticity of the SARS Virus Pseudoknot.* Journal of the American Chemical Society, 2014. **136**(6): p. 2196-2199.
353. Wang, Z. and L. Yang, *Turning the Tide: Natural Products and Natural-Product-Inspired Chemicals as Potential Counters to SARS-CoV-2 Infection.* Frontiers in Pharmacology, 2020. **11**: p. 1013.
354. Newman, D.J. and G.M. Cragg, *Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019.* Journal of Natural Products, 2020. **83**(3): p. 770-803.
355. Pamuru, R.R., et al., *Targeting Natural Products for the Treatment of COVID-19 - An Updated Review.* Current Pharmaceutical Design, 2020. **26**(41): p. 5278-5285.
356. Verma, S., et al., *Anti-SARS-CoV Natural Products With the Potential to Inhibit SARS-CoV-2 (COVID-19).* Frontiers in pharmacology, 2020. **11**: p. 561334.
357. Aboul-ela, F., *Strategies for the design of RNA-binding small molecules.* Future Med Chem, 2010. **2**(1): p. 93-119.
358. Warner, K.D., C.E. Hajdin, and K.M. Weeks, *Principles for targeting RNA with drug-like small molecules.* Nature Reviews Drug Discovery, 2018. **17**(8): p. 547-558.

359. Martínez, J.P., et al., *Screening of small molecules affecting mammalian P-body assembly uncovers links with diverse intracellular processes and organelle physiology.* RNA Biology, 2013. **10**(11): p. 1661-1669.
360. Sahner, J.H., et al., *Exploring the chemical space of ureidothiophene-2-carboxylic acids as inhibitors of the quorum sensing enzyme PqsD from Pseudomonas aeruginosa.* European Journal of Medicinal Chemistry, 2015. **96**: p. 14-21.
361. Torcato, I.M., et al., *Design and characterization of novel antimicrobial peptides, R-BP100 and RW-BP100, with activity against Gram-negative and Gram-positive bacteria.* Biochimica et Biophysica Acta, 2013. **1828**(3): p. 944-955.
362. Zhang, K., et al., *Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome.* Nature Structural & Molecular Biology, 2021. **28**(9): p. 747-754.
363. Roman, C., et al., *The SARS-CoV-2 Programmed -1 Ribosomal Frameshifting Element Crystal Structure Solved to 2.09 Å Using Chaperone-Assisted RNA Crystallography.* ACS Chemical Biology, 2021. **16**(8): p. 1469-1481.
364. Edelsbrunner, H. and E.P. Mücke, *Three-dimensional alpha shapes.* 1994. **13**(1 %J ACM Trans. Graph.): p. 43–72.
365. Ježek, J., J. Hlaváček, and J. Sebestík, *Biomedical Applications of Acridines: Derivatives, Syntheses, Properties and Biological Activities with a Focus on Neurodegenerative Diseases*. 2017.
366. Yan, Z., et al., *Identification of an aminoacridine derivative that binds to RNA tetraloops.* Journal of Medicinal Chemistry, 2007. **50**(17): p. 4096-4104.
367. Karacostas, V., et al., *Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles.* Virology, 1993. **193**(2): p. 661-71.
368. Dinman, J.D. and R.B. Wickner, *Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation.* J Virol, 1992. **66**(6): p. 3669-76.
369. Mirabelli, C., et al., *Morphological cell profiling of SARS-CoV-2 infection identifies drug repurposing candidates for COVID-19.* Proceedings of the National Academy of Sciences, 2021. **118**(36): p. e2105815118.
370. Xie, X., et al., *An Infectious cDNA Clone of SARS-CoV-2.* Cell Host & Microbe, 2020. **27**(5): p. 841-848.e3.
371. Widera, M., et al., *Generation of a Sleeping Beauty Transposon-Based Cellular System for Rapid and Sensitive Screening for Compounds and Cellular Factors Limiting SARS-CoV-2 Replication.* Frontiers in Microbiology, 2021. **12**.
372. Qi, X., T. Xia, and R.W. Roberts, *Acridine-N peptide conjugates display enhanced affinity and specificity for boxB RNA targets.* Biochemistry, 2010. **49**(27): p. 5782-5789.
373. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nature methods, 2012. **9**(7): p. 676-682.
374. Elgaher, W.A.M., et al., *Discovery and Structure-Based Optimization of 2-Ureidothiophene-3-carboxylic Acids as Dual Bacterial RNA Polymerase and Viral Reverse Transcriptase Inhibitors.* Journal of Medicinal Chemistry, 2016. **59**(15): p. 7212-7222.
375. Sharma, P., et al., *Humans and other commonly used model organisms are resistant to cycloheximide-mediated biases in ribosome profiling experiments.* Nature Communications, 2021. **12**(1): p. 5094.
376. Machkovech, H.M., J.D. Bloom, and A.R. Subramaniam, *Comprehensive profiling of translation initiation in influenza virus infected cells.* PLOS Pathogens, 2019. **15**(1): p. e1007518.
377. Rutkowski, A.J., et al., *Widespread disruption of host transcription termination in HSV-1 infection.* Nature Communications, 2015. **6**(1): p. 7126.
378. Stern-Ginossar, N., et al., *Decoding Human Cytomegalovirus.* Science, 2012. **338**(6110): p. 1088-1093.
379. Iacono, M., F. Mignone, and G. Pesole, *uAUG and uORFs in human and rodent 5'untranslated mRNAs.* Gene, 2005. **349**: p. 97-105.

380. Ruiz Cuevas, M.V., et al., *Most non-canonical proteins uniquely populate the proteome or immunopeptidome.* Cell Rep, 2021. **34**(10): p. 108815.
381. Labaronne, E., et al., *Extensive uORF translation from HIV-1 transcripts conditions DDX3 dependency for expression of main ORFs and elicits specific T cell immune responses in infected individuals.* bioRxiv, 2022: p. 2022.04.29.489990.
382. Juszkiewicz, S., et al., *Ribosome collisions trigger cis-acting feedback inhibition of translation initiation.* Elife, 2020. **9**.
383. Meydan, S. and N.R. Guydosh, *A cellular handbook for collided ribosomes: surveillance pathways and collision types.* Curr Genet, 2021. **67**(1): p. 19-26.
384. Sinha, N.K., et al., *EDF1 coordinates cellular responses to ribosome collisions.* Elife, 2020. **9**.
385. Sundaramoorthy, E., et al., *Ribosome quality control activity potentiates vaccinia virus protein synthesis during infection.* J Cell Sci, 2021. **134**(8).
386. Hayakawa, S., et al., *ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses.* Nat Immunol, 2011. **12**(1): p. 37-44.
387. Kmiec, D., et al., *S-farnesylation is essential for antiviral activity of the long ZAP isoform against RNA viruses with diverse replication strategies.* PLOS Pathogens, 2021. **17**(10): p. e1009726.
388. Volkmer, R., V. Tapia, and C. Landgraf, *Synthetic peptide arrays for investigating protein interaction domains.* FEBS Letters, 2012. **586**(17): p. 2780-2786.