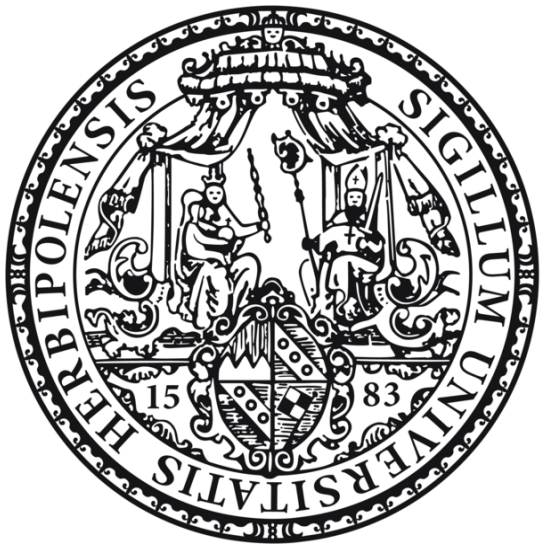

Single-Molecule Approaches To Study Frameshifting Mechanisms

Einzelmolekülansätze zur Untersuchung von Frameshifting-Mechanismen

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Infection and Immunity

submitted by

Lukáš Pekárek
from
Kutná Hora, Czech Republic

Würzburg 2023

Submitted on: ……..

Office stamp

Members of the Thesis Committee

Chairperson:	Prof. Dr. Jörg Schultz
Primary Supervisor:	Prof. Dr. Neva Caliskan
Supervisor (Second):	Prof. Dr. Markus Sauer
Supervisor (Third):	Prof. Dr. Utz Fischer

“Never confuse education with intelligence,
you can have a PhD and still be an idiot.”

Richard P. Feynman

Abstract

The RNAs of many viruses contain a frameshift stimulatory element (FSE) that grants access to an alternate reading frame via −1 programmed ribosomal frameshifting (PRF). This −1PRF is essential for effective viral replication. The −1PRF efficiency relies on the presence of conserved RNA elements within the FSE, such as a slippery sequence, spacer, and a downstream secondary structure – often a hairpin or a pseudoknot. The PRF efficiency is also affected by *trans*-acting factors such as proteins, miRNAs and metabolites. The interactions of these factors with the RNA and the translation machinery have not yet been completely understood. Traditional ensemble methods used previously to study these events focus on the whole population of molecular species. This results in innate averaging of the molecular behavior and a loss of heterogeneity information.

Here, we first established the experimental workflow to study the RNA structures and the effect of potential *trans*-acting factors using single-molecule force spectroscopy technique, optical tweezers. Additionally, to streamline the data analysis, we developed an algorithm for automatized data processing.

Next, we harnessed this knowledge to study viral RNA elements responsible for stimulation of PRF and how the presence of *trans*-acting factors affects the RNA behavior. We further complemented these single-molecule structural data with ensemble functional assays to gain a complex view on the dynamics behind the programmed ribosomal frameshifting.

Specifically, two different viral RNA elements have been studied in the presented work. First, the dynamics of SARS-CoV-2 FSE and the role of extended sequences have been explored. Then, the mode of action of the host-encoded *trans*-acting factor ZAP-S inhibition of SARS-CoV-2 PRF has been examined. Finally, the mechanism of the *trans*-acting viral factor induced PRF in Encephalomyocarditis virus (EMCV) has been uncovered.

Zusammenfassung

Die RNAs vieler Viren enthalten ein Lese-Rasterverschiebung-stimulierendes Element (FSE), das über die −1 programmierte ribosomale Rasterverschiebung (PRF) Zugriff auf einen alternativen Leserahmen gewährt. Dieser −1PRF ist für eine effektive Virusreplikation unerlässlich. Die −1PRF-Effizienz beruht auf dem Vorhandensein konservierter RNA-Elemente innerhalb des FSE, wie z.B. einer Slippery-Sequenz, einem Platzhalter und einer nachgelagerten Sekundärstruktur – oft eine Haarnadel oder ein Pseudoknoten. Die −1PRF-Effizienz wird auch durch *trans*-aktive Faktoren wie Proteine, miRNAs und Metaboliten beeinflusst. Die Wechselwirkungen dieser Faktoren mit der RNA und der Translationsmaschinerie sind noch nicht vollständig verstanden. Traditionelle Ensemble-Methoden, die früher zur Untersuchung dieser Ereignisse verwendet wurden, konzentrieren sich auf die gesamte Population molekularer Spezies. Dies führt zu einer inhärenten Durchschnittsbildung des molekularen Verhaltens und einem Verlust von Heterogenitätsinformationen.

Hier haben wir zunächst den experimentellen Arbeitsablauf zur Untersuchung der RNA-Strukturen und der Wirkung potenzieller *trans*-aktiver Faktoren mithilfe der Einzelmolekül-Kraftspektroskopietechnik Optischer Pinzetten etabliert. Um die Datenanalyse zu optimieren, haben wir außerdem einen Algorithmus zur automatisierten Datenverarbeitung entwickelt.

Als nächstes nutzten wir dieses Wissen, um virale RNA-Elemente zu untersuchen, die für die Stimulierung von −1PRF verantwortlich sind, und wie sich das Vorhandensein *trans*-aktiver Faktoren auf das Verhalten der RNA auswirkt. Wir haben diese Einzelmolekülstrukturdaten weiter durch Ensemble-Funktionsassays ergänzt, um einen komplexen Überblick über die Dynamik hinter der programmierten ribosomalen Rasterverschiebung zu erhalten.

Konkret wurden in der vorgestellten Arbeit zwei verschiedene virale RNA-Elemente untersucht. Zunächst wurden die Dynamik des SARS-CoV-2-FSE und die Rolle erweiterter Sequenzen untersucht. Anschließend wurde die hemmende Wirkungsweise des vom Wirt kodierten *trans*-wirkenden Faktors ZAP-S auf SARS-CoV-2-PRF untersucht. Schließlich wurde der Mechanismus der, durch den *trans*-aktiven Virusfaktor induzierten PRF beim Enzephalomyokarditis-Virus (EMCV), entschlüsselt

Contents

Chapter 1

"All models are wrong, but some are useful."

George E.P. Box

1. Introduction

1.1. Central role of RNA in biology

RNA was originally regarded as a mere "messenger" of the information stored in DNA to form a protein. However, it is now known that besides its "coding" function, RNA encompasses a whole myriad of other functions (1-4). Starting with translation-related non-coding RNAs – transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), the first supplying the amino acids to the ribosomes, whose huge fraction is formed by the latter. RNAs not only facilitate the translation but also regulate the faith of the mRNA (miRNA, siRNA, etc.). Moreover, RNAs also have catalytical functions and can even catalyze synthesis of other RNAs, thus shedding light on the possible explanation of the early evolution of life (1, 2). Among these RNA enzymes (ribozymes), the most prominent one, without any doubt, is the ribosome itself (5-7).

RNA differs chemically from DNA by the presence of the 2' hydroxyl (-OH) group on the ribose moiety of the nucleotide. This 2' hydroxyl group makes the RNA more reactive and labile to hydrolysis, especially under alkaline conditions. RNA and DNA also differ in the bases they consist of. While adenine (A), guanine (G), and cytosine (C) are present in both, uracil (U) is present in RNA, and thymine (T) is characteristic of DNA. U and T differ only in the presence of methyl group at position 5 of the base heterocycle. As C can spontaneously undergo hydrolytic deamination and form U, the employment of T in DNA allows the cellular repair machinery to distinguish the degraded C from the intentional T. Additionally, unlike DNA, RNA does not form long double helices, but neither is it simply single-stranded, it rather forms many short hairpins or helices (discussed further below).

One might easily find themselves thinking that this versatility does not, at first glance, align with the simple nature of RNA, consisting of only four nucleotides. This apparent paradox comes from the ability of RNA nucleotides (**i**) to interact beyond the Watson-Crick-Franklin base pairing space and (**ii**) to form intra- as well as inter-molecular bonds. This way, RNA molecules can explore an extremely wide folding landscape (8-10). The structural flexibility of RNA grants its functional versatility, thus making it a central molecule in many biological processes. Among all those processes, translation is the hallmark of RNA functionality as the three RNA species (rRNA, tRNA, and mRNA) work together to synthesize new proteins.

1.2. Translation

1.2.1. Translation is universal

Translation is an ancient process utilized by all living forms that translates the genetic information from the language of nucleic acids into a protein based on the "universal" rules known as genetic code. Therefore, despite the vast evolutionary distance between prokaryotes and eukaryotes, the process of translation is conserved and thus still bears a lot of similarities between both domains (**Figure 1.1**) (11). Due to its key role in gene expression as well as high energy requirements, translation is a tightly regulated process (12). There are several quality control mechanisms that ensure the translation is efficient and any sort of translation discontinuation is resolved (13-17).

Translation process is facilitated by the ribosome, which is a huge biomolecular complex consisting of rRNAs and proteins. Despite huge differences in size and mass, the core of ribosomes is well conserved among all the domains of life (11, 18, 19). The ribosome consists of two parts, small and large subunits (SSU and LSU), each of them carrying different roles. The SSU binds the mRNA and facilitates the base-pairing between the mRNA codons and cognate tRNAs, also known as the decoding function of SSU. There are three different sites accommodating the tRNAs – A-, P-, and E-site. The first one is the A-site, which accommodates the newly arrived aa-tRNAs. Next is the P-site which contains the tRNA with the growing peptidyl chain. Finally, the E-site contains the translocated tRNA, which transferred its peptidyl chain, and now has to dissociate from the ribosome (19). The peptide bond formation is orchestrated by the peptidyl transfer center (PTC) of the LSU, which consists mainly of the rRNA, thus being responsible for the ribozyme character of the ribosome. As the ribosome goes through the mRNA in the 5'→3' direction in 3 nucleotide steps, it accommodates the aminoacylated tRNAs complementary to the mRNA codons and catalyzes the formation of the peptide bonds. The reading frame is ensured due to the base-pairing between the tRNAs and mRNA. As the peptide chain is growing, it is pushed out of the ribosome through the exit tunnel. The ribosomal movement along the mRNA is facilitated by changes in relative positions between the subunits and the tRNAs: mRNA base paired complex.

Translation process consists of four different phases: (**i**) initiation, (**ii**) elongation, (**iii**) termination, (**iv**) recycling (15, 19). During initiation (**i**), the small ribosomal subunit is recruited to the mRNA by initiation factors (IFs or eIFs, for prokaryotes and eukaryotes, respectively), recognizes the start AUG codon, and finally, the ribosome is assembled by the recruitment of the large subunit. Once the ribosome is successfully positioned at the start codon with the (f)Met-tRNA at the P-site, the next phase - elongation (**ii**) takes place (19, 20).

Elongation (**ii**) is the most conserved process between the eukaryotic and prokaryotic translation, catalyzed by elongation factors (EFs and eEFs for prokaryotes and eukaryotes, respectively), which show high homology between the phylogenetic domains (11). During elongation, the ribosome moves along the mRNA in the 5'-3' direction in three-nucleotide steps. At each step, a new aminoacylated tRNA (aa-tRNA) is brought to the ribosome via the formation of a tertiary complex containing the aa-tRNA, EF-Tu or eEF-1A (depending on the phylogenetic domain), and GTP. Once the aa-tRNA is positioned at the A-site of the ribosome, GTP is hydrolyzed and the tertiary complex disassembles and EF-Tu or eEF-1A:GDP dissociates (21, 22). Then peptide bond is formed between the aa-tRNA at the A-site and the peptidyl residue of the peptidyl-tRNA at the P-site. The tRNAs are then translocated from the A- and P-sites to P- and E-sites, respectively (19). This step is catalyzed by another GTPase protein called EF-G or eEF-2, respectively. Once translocated, the tRNA at the E-site dissociates from the ribosome, and the whole cycle can be repeated until the ribosome reaches the STOP codon and translation is terminated (19).

Termination (**iii**) and ribosome recycling (**iv**) are two processes inherently coupled with each other. Initially, the release factors (RFs or eRFS) are recruited to the ribosome to facilitate the release of the peptide chain from the tRNA. This is followed by the disassembly of the ribosome and the release of the large subunit. Finally, the small subunit also dissociates from the mRNA. The released subunits and the mRNA can be then reactivated, initiated and next round of translation can take place (19, 23, 24).

Each step of translation undergoes tight regulation, which is crucial for the survival and fitness of the cell. Errors in gene expression can lead to dis- or mal-functional proteins as well as a waste of cellular resources (12). On the other hand, changes in translation offer a quick way how to react to changes inside the cell or in the outer environment (25, 26). The circumvention of some of these tight rules governing the translation poses an attractive approach to fine-tuning gene expression at the translation level. Thus, these non-canonical events represent yet another layer of gene expression regulation. On top of that, these recoding translation events are especially captivating to the cellular parasites, namely viruses, that harness them to exploit and restrict the cellular translation machinery. Therefore, we can find recoding strategies being employed at literary every step of the translation (25-28) (**Figure 1.2**).

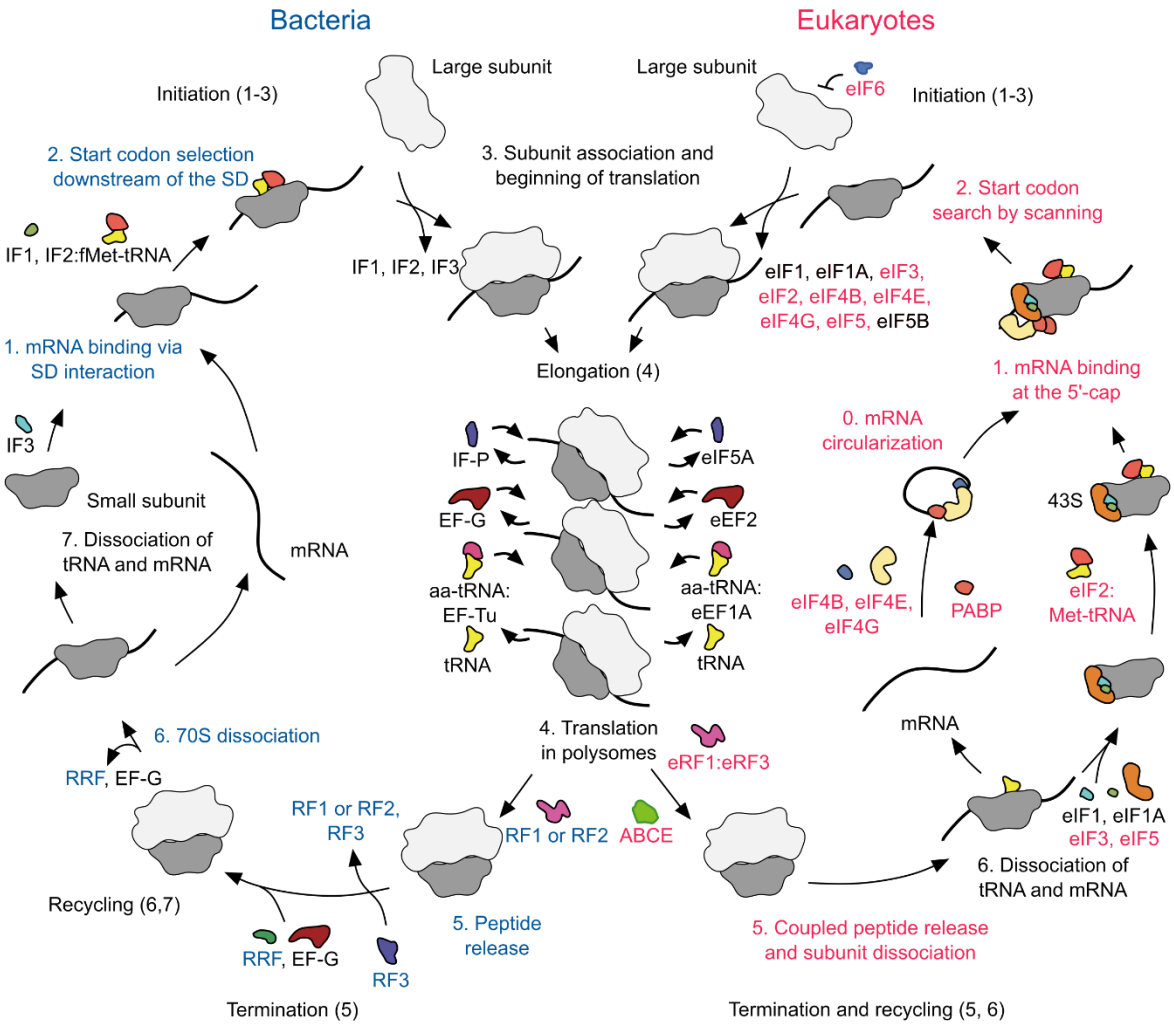

Figure 1.1 Comparison of bacterial and eukaryotic translation. The translation consists of four stages – initiation, elongation, termination, and ribosome recycling. Each of these steps is assisted by protein factors – initiation factors (IFs, or eIFs, for bacteria and eukaryotes, respectively), elongation factors (EF, or eEF), release factors (RFs or eRFs), and recycling factors. Elongation is the most conserved between bacteria and eukaryotes and is assisted by homologous elongation factors (all homologous factors and common steps of translation are labeled black in the figure). The other steps of the translation have diverged and include several stages (indicated by numbers) that differ between bacteria (blue) and eukaryotes (red). The initiation, termination, and release factors catalyzing these steps include many nonhomologous proteins specific to either bacteria (green) or eukaryotes (red). aa-tRNA, aminoacyl-tRNA. Adapted from (11).

1.2.2. Non-canonical initiation, elongation, and termination

The exploitation of each of the translation phases provides additional and different effects on the translation machinery. Alternative ways of initiation, such as the internal ribosomal entry through the internal ribosomal entry site (IRES) allow translation initiation without one or more initiation factors. This way, the canonical translation initiation can be

arrested by viral factors, and the virus can take full advantage of the cellular resources (27, 29, 30). Leaky scanning or non-AUG initiations further expand the proteome diversity and, moreover, allow fine-tuning of the extent of translation. However, the efficiency of non-AUG initiation is relatively low (27, 31, 32). Finally, translation re-initiation takes advantage of the close proximity of one open reading frame to the downstream one and allows to translation of two (or more) open reading frames from a single molecule of mRNA (15, 31).

Once the ribosome is initiated at the mRNA and starts translating, other recoding strategies has been described to interfere with the elongation or termination of translation (25, 28). In general, there are four main categories of these recording events: Stop-codon redefinition or readthrough (**i**), during which some ribosomes do not terminate but instead insert a standard amino acid (readthrough) or a non-standard amino acid such as pyrrolysine or selenocysteine (redefinition); Stop-go (**ii**), during which the peptide bond at a specific site between two amino acids is not formed resulting in two peptide chains being produced during the translation; Bypassing (**iii**), during which (a fraction of) ribosomes skip over a part of the RNA sequence and continues translating; and ribosomal frameshifting (**iv**), in which a fraction of ribosomes shifts the reading frame by 1 or 2 nt either forward or backward resulting in translation of different peptide chain. Thus, there is a broad repertoire of non-canonical mechanisms that can be used to extend the proteome and exploit the translation machinery (25-28, 33).

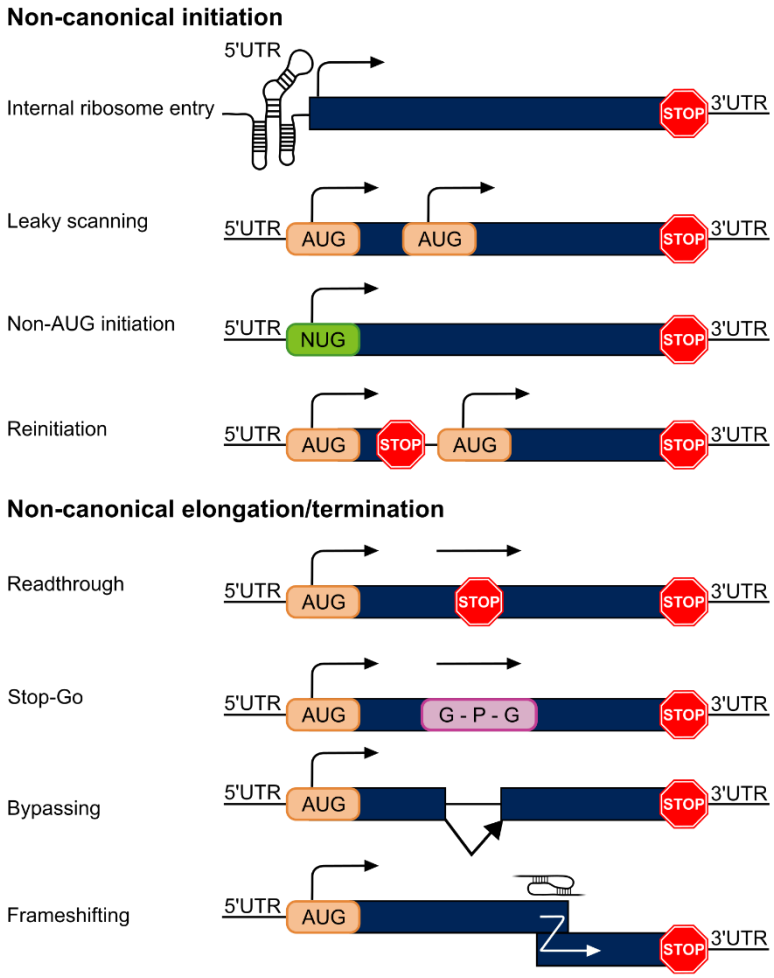

Figure 1.2 Non-canonical translation events. The translation can be manipulated at each phase. Most translated mRNAs contain a 5'UTR followed by a single open reading frame marked by an AUG start codon and ending with one of the stop codons followed by a 3'UTR (In the case of the eukaryotic mRNAs they also contain the poly(A)-tail. Black arrows indicate translation initiation (at the start of an ORF) or continuation of translation. Non-canonical initiation events include internal ribosomal entry, leaky scanning, non-AUG initiation, and reinitiation. Non-canonical elongation/termination events are called recoding and include readthrough, stop-go, bypassing, and frameshifting. The white arrow in frameshifting represents the shift of the ribosome to a different reading frame. Adapted from (27).

1.3. −1 Programmed ribosomal frameshifting (−1PRF)

One of the most studied recoding mechanisms is ribosomal frameshifting, specifically −1 Programmed ribosomal frameshifting (−1 frame, which results in slippage of the ribosome into −1 reading frame by shifting one nt upstream (backwards) (26, 27). −1PRF is a universal translation recoding event which allows effective expansion of the information capacity of an mRNA by encoding two different proteins in a single mRNA molecule (and controls their relative ratio). This expansion of information capacity is

particularly attractive for organisms and entities with constrained genome size, such as viruses. Over the years, evolution came up with a canonical motif of the RNA sequence which stimulates the ribosomal frameshifting (34-37). The RNA sequence of a −1PRF site consists of three parts acting in *cis* (38). On top of that, there has also been evidence for the existence of factors acting in a *trans* manner (25, 26). Since its first discovery, this recoding mechanism has been described in many clinically important viruses, such as Coronaviruses (SARS-CoV-1 and 2) or retroviruses (HIV-1 and 2) (26, 27, 35, 39).

1.3.1. *Cis*-acting elements of −1PRF

The canonical −1PRF site, or frameshifting stimulatory element (FSE) consists of three regions (**Figure 1.3A**): slippery sequence (**i**), spacer (**ii**), and stimulatory structure (**iii**). **The slippery sequence** (**i**) is the site of the actual slippage of the ribosome into a different reading frame. It is a heptanucleotide stretch of RNA with a consensus sequence of X_XXY_YYZ (XXX represents any three identical nucleotides, YYY represents AAA or UUU and Z represents A, C or U). The energy difference between the tRNA: mRNA codon-anticodon base pairs in 0- and −1-frames determines the maximal frameshifting efficiency (37, 40-43).

However, the slippery sequence on its own is not enough to induce frameshifting efficiently (37, 42-45). To increase the chances of ribosomal slippage, a **stimulatory structure** (**ii**) downstream of the slippery sequence is employed to halt the progress of the ribosome and thus provide a time window for the ribosome to explore the alternative frames. The ability to pause ribosome depends (partially) on the structure stability (Gibbs free energy of the RNA structure, or at least the base of the stem) as well as on the overall size of the structure (46-52). Thus, the stimulatory structures are often pseudoknots rather than stem-loops, although these cases were also reported (34, 35, 43, 47, 49-55).

Finally, since the ribosome is a huge ribonucleoprotein complex, the stimulatory structure cannot be positioned immediately downstream of the slipper sequence. Therefore, there is a short stretch of 5-9 nucleotides **spacer** (**iii**), that ensures correct distance between the other two elements (36, 37).

1.3.2. *Trans*-acting elements of −1PRF

On top of these *cis* elements encoded in the RNA sequence, some studies recently also reported the effect of different factors acting in *trans* (56-67). In addition to a rather static effect of *cis*-acting elements, trans-acting factors pose a dynamic regulation of frameshifting efficiency in time and space (25, 26, 67). This additional layer of translation fine-tuning allows more efficient use of cellular resources. Examples of the *trans*-factors

include proteins, miRNAs, cellular metabolites or small molecules (26, 56, 58, 62-64, 66-69). In principle, the *trans*-factors can act either directly on the mRNA or indirectly on other proteins and affect the translation control mechanisms.

Trans-acting factors are utilized by both – host as well as the pathogen (viruses). Among the viral factors, *Cardioviruses*, Encephalomyocarditis virus (EMCV) and Theiler's Murine Encephalomyelitis virus (TMEV), are a well-documented example of fine-tuning the −1PRF efficiency through a virus-encoded 2A protein (65-67, 70). *Cardioviruses* employ several examples for non-canonical translation, they employ IRES, Stop-Go, and −1PRF during the expression of their genome (57, 71, 72). The −1PRF in cardioviruses regulates the expression of the replication machinery enzymes encoded in the 0-frame, the frameshifting event results in the expression of a short 2B* protein with no known function (57, 67, 70, 71). The gene encoding the 2A protein is placed just before the −1PRF element so that it is always expressed regardless of the frameshifting (EMCV and 2A protein is discussed more in detail in **Chapter 6**). Another example of a virus-employed *trans*-factor is the Porcine reproductive and respiratory syndrome virus, a member of the *Arteriviridae* family. There the viral protein nsp1β forms a complex with host-encoded poly(C)-binding protein. This complex then interacts with a C-rich region downstream of the slippery sequence and stimulates the frameshifting (73, 74).

Host cells also encode proteins that are involved in translation regulations and cellular defense. Frameshifting represents an attractive target for cellular defense since the FSE is generally well-conserved throughout evolution (25, 75). Therefore, disrupting the frameshifting efficiency poses a strategy that would be hard to evolve resistance against. Moreover, the stable stimulatory structure is crucial for efficient frameshifting so the ribosome can be stalled. Resolving the stalled ribosomes and unwinding the stable RNA structures is one of the translation quality control strategies, such as ribosome-associated protein quality control, so it is no surprise that in the last few years, studies have shown the involvement of cellular proteins in frameshifting regulation (15, 64).

One of the first such cellular proteins described was Annexin 2 (ANXA2) (58), which was shown to reduce the frameshifting of Infectious Bronchitis Virus (IBV). ANXA2 is a cytoskeletal protein that has recently been shown to play a role in the calcium-dependent IRES-mediated translation initiation of MYC (76). However, how ANXA2 affects these processes is unknown. Another reported cellular trans-factor is the Shiftless (SFL) protein, which has been shown to act as a wide broad recoding inhibitor (64, 77).

Finally, the recent Covid-19 pandemic, caused by the SARS-CoV-2, prompted many efforts towards identifying the trans-acting agents targeting the −1PRF site RNA (discussed in detail in **Chapter 5**) (63, 69).

1.3.3. −1PRF thermodynamics, kinetics and mechanisms

There are several aspects affecting the kinetics as well as the thermodynamics of the frameshifting. Many previous studies attempted to characterize these aspects from various points of view – What is the role of slippery sequence? (36, 40, 42, 43) How does the stimulatory structure affect the frameshifting? (34, 46-48, 50, 51, 53, 55, 61) When during the translation does the frameshifting actually occur? (43, 54, 78-81) Each of those studies provides a piece into the puzzle of programmed ribosomal frameshifting.

Briefly, frameshifting most likely occurs during the late phase of translocation (**Figure 1.3B** and **C**) when the tRNAs are moved from the A- and P- site into P- and E-sites, respectively. During the translocation step, the base-pairings between tRNAs and mRNAs are broken and reformed, therefore this step is prone to exploration of alternative frames (26, 53, 80). The evidence further suggests that the slowing down the translocation increases the chances of alternative frames being explored (43, 82, 83). The EF-G (eEF2) binding can then facilitate rescue of the ribosome from being trapped in the rotated/chimeric state (43, 80, 83).

In a general view, both slippery sequence and the stimulatory structure together determine the final frameshifting efficiency. However, their respective roles are completely different. The slippery sequence determines the thermodynamics of the frameshifting – based on the energy differences between tRNA:mRNA base-pairings in the 0- and −1-frames – the maximum frameshifting efficiency is determined (37, 40, 42, 84).

However, the slippery sequence on its own is not enough to induce frameshifting efficiently. This is likely because the translation, specifically the translocation step, is too fast for the thermodynamic equilibrium to be established (26, 43). Therefore, the ribosome has to be slowed down or even stalled on the slippery sequence to allow establishing of the thermodynamic equilibrium (43). This can be achieved through several routes, which can be combined. One of the most well-documented ways of stalling the ribosome is a stable RNA structure downstream of the slippery sequence (as described above). The translocation is catalyzed by the EF-G (eEF2) binding to the A-site and hydrolysis of the GTP, therefore, occluding the A-site is another strategy. This can be achieved either through a downstream RNA structure or by *trans*-acting factors (46, 67, 83, 85). Another proposed strategy to stall or at least slow down the ribosome is the stalling of the nascent peptide in the exit tunnel (69, 86-88). Finally, in bacteria it has been shown that a Shine-Dalgarno-like sequence upstream of the slippery sequence can interfere with the ribosomal progress through interactions with the rRNA (54, 82). Taken together, while slippery sequence determines the maximal frameshifting efficiency through the thermodynamic differences between base-pairings in the respective frames, the thermodynamic equilibrium can be achieved only by slowing down the ribosome.

Additionally, some studies suggested that amino acid starvation can also induce frameshifting (89-91). This alternative mechanism has been described by Caliskan et al., (92). tRNAs are present in the cells at the different levels (concentrations), thus reflecting the differences in codon usage. Moreover, if the supply of an amino acid is limited, the amino acylated tRNA levels can be even further altered. This mechanism is slower and independent of the stimulatory structure necessary for the canonical −1PRF site (**Figure 1.3D**).

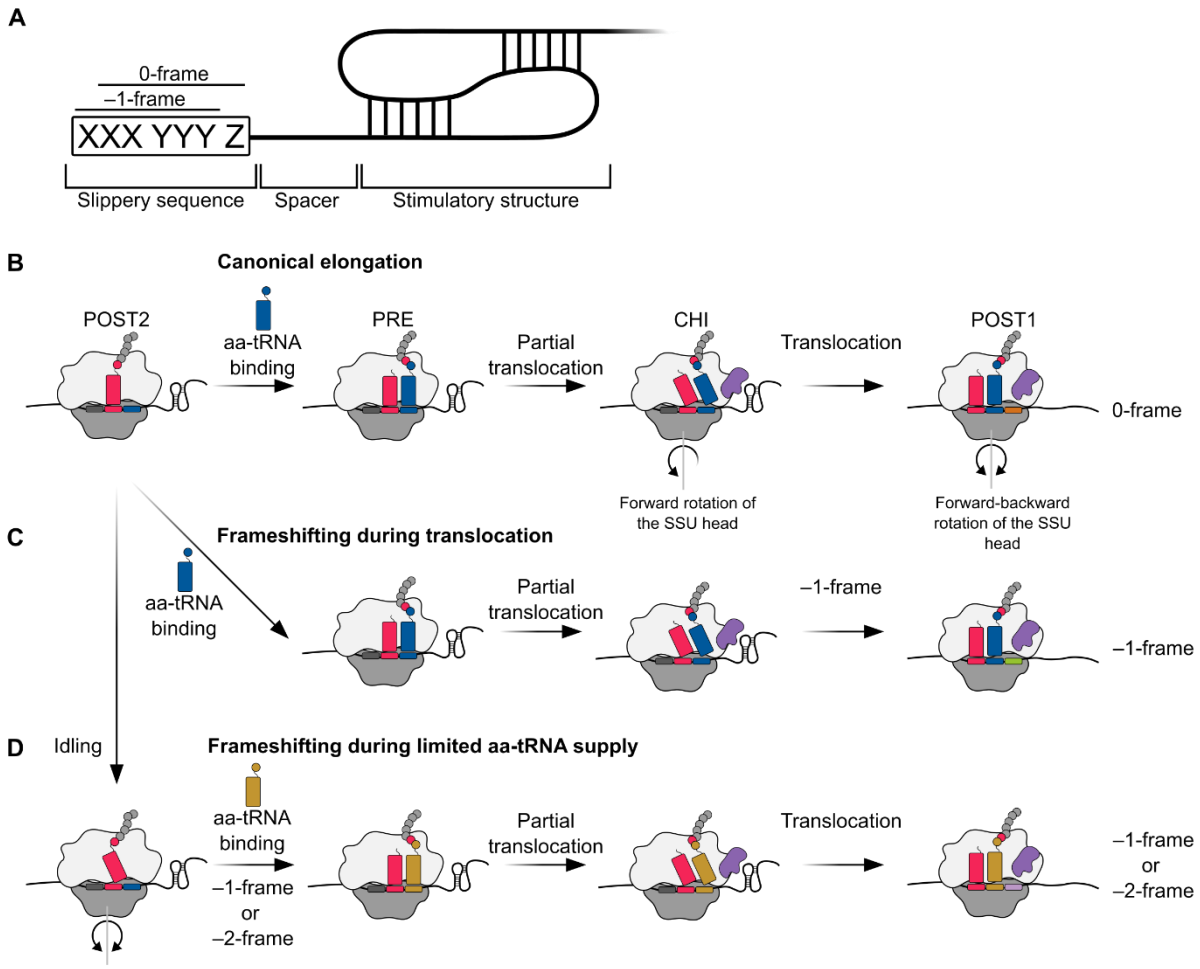

Figure 1.3 Programmed ribosomal frameshifting. (**A**) a scheme of canonical RNA motif of −1 programmed ribosomal frameshifting. (**B**) In the canonical elongation pathway, movement of the ribosome along the mRNA and the peptide bond formation is facilitated by specific forward and backward movements of the small subunit (SSU) head and the hydrolysis of GTP by the elongation factor EF-G (eEF2 in eukaryotes). POST: post-translocation state, PRE: pre-translocation state, CHI: chimeric state. (**C**) In case of canonical −1 frameshifting, ribosomal stalling is caused by a secondary structure of the mRNA leading to frameshifting during translocation. (**D**) An alternative frameshifting pathway is mediated by the limitation of the aminoacyl-site codon respective aminoacylated tRNA (aa-tRNA), here an idling step gives the time to overcome the limitation by shifting into the −1- or −2-frame. Adapted from (26)

1.4. RNA structure and function

As described above, RNA is rarely a passive passenger but rather can be seen as an active player, actively altering or even governing some of the most fundamental processes in the cell. The key aspect of this versatility is the relationship between RNA function and structure. Unlike proteins, RNAs are much more promiscuous in exploring or alternating between different conformations and thus also different functions.

While the primary sequence of RNA is restrained to only four "letters" (A, U, G, C), the variety of base pairing options is essential for the function of different RNAs (9). The primary structure of RNA refers to the linear sequence of nucleotides that make up the RNA molecule (**Figure 1.4A**). It is crucial for the spatial organization of the RNA molecules (secondary, tertiary structures, etc.) as it determines the specific order of nucleotides that are available for base pairing and, therefore, the potential for the formation of different secondary structures. This potential is then reflected in the free-energy landscape (FEL) of each RNA molecule that, on top of the possible conformations, also provides information about the probabilities of each of these conformations (9, 93-95). A change in a single nucleotide in the primary structure could disrupt a structure or create a new structure that was not present before (96). Therefore, the primary and secondary structures of RNA are intimately connected, and changes in one can have important functional consequences for the other. Understanding the relationship between these two structures is essential for explaining the complex roles that RNA plays in biological processes.

1.4.1. RNA secondary structure

The secondary structure refers to the local folding pattern of the RNA molecules. These folding patterns are formed through base pairing between complementary nucleotides, which create double-stranded regions (stems) and single-stranded regions (loops). The various options of arrangement of these individual stems and loops give rise to a multitude of secondary structure motifs (**Figure 1.4B**). The specific arrangement of stems and loops gives rise to the unique secondary structure of an RNA molecule, which can play a critical role in determining its function (10, 95, 97).

Among the structural motifs, we can recognize the nested structures (**Figure 1.4B**), in which, in the case of multiple stems, any stem-loop is entirely contained within another, larger stem-loop (98-101). These include simple hairpins, bulges, internal loops, or three/four-way junctions. In contrast, we also recognize so-called non-nested structures (**Figure 1.4C**). These are characterized by the presence of overlapping stems where (part of) a loop of one stem-loop is forming another stem. These motifs often involve short-to-long-distance interactions making the final structure more spatially complex. Examples of

these structural motifs include pseudoknots, kissing loops, and G-quadruplexes. The non-nested nature of these structures makes their prediction quite challenging (98, 102, 103).

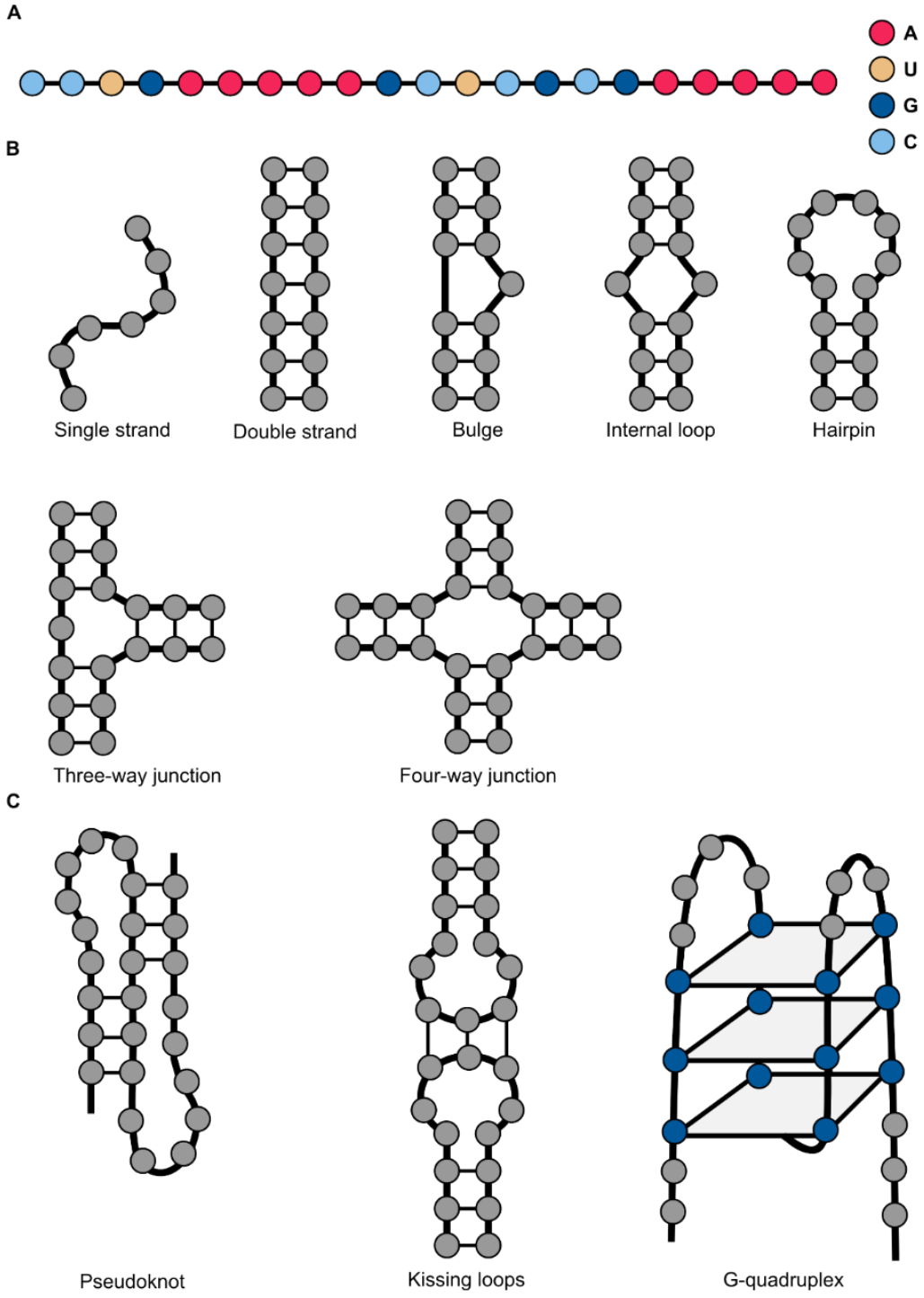

Figure 1.4 RNA primary and secondary structure motifs. (**A**) The primary structure of an RNA molecule consists of four different nucleotides and is determined by the order of these nucleotides in the 5' to 3' direction. (**B**) Examples of nested secondary structure motifs. (**C**) Examples of non-nested secondary structure motifs.

1.4.2. Tertiary structure in RNA

RNA tertiary structure refers to the three-dimensional shape adopted by an RNA molecule, which arises from the folding of its secondary structure into a compact spatial conformation (**Figure 1.5A** and **B**). Unlike secondary structure, which is primarily determined by base pairing interactions between complementary nucleotides, tertiary structure is determined by a combination of base pairing, base stacking, and other non-canonical interactions, including non-Watson-Crick-Franklin base pairings, hydrogen bonds, electrostatic interactions, and hydrophobic interactions (9, 95, 97) (**Figure 1.5C**).

The complexity of RNA tertiary structure arises from the large number of potential non-canonical interactions between the nucleotides in the molecule, as well as the influence of the surrounding environment, such as the presence of (multivalent) ions or other small molecules (55, 104, 105). The specific tertiary structure adopted by an RNA molecule is essential to its biological function, as it can determine its interactions with other molecules, such as proteins or other RNAs (61, 106-108). For example, ribozymes, which are RNA molecules with catalytic activity, rely on specific tertiary structures to position key functional groups for efficient catalysis (8, 109).

Determining the tertiary structure of an RNA molecule can be challenging, as it requires advanced experimental techniques such as X-ray crystallography, NMR spectroscopy, or cryo-electron microscopy (69, 110-112). Additionally, predicting the tertiary structure of an RNA molecule computationally is still an active area of research, as current methods are often limited in their accuracy and ability to model large and complex RNA molecules (113-115).

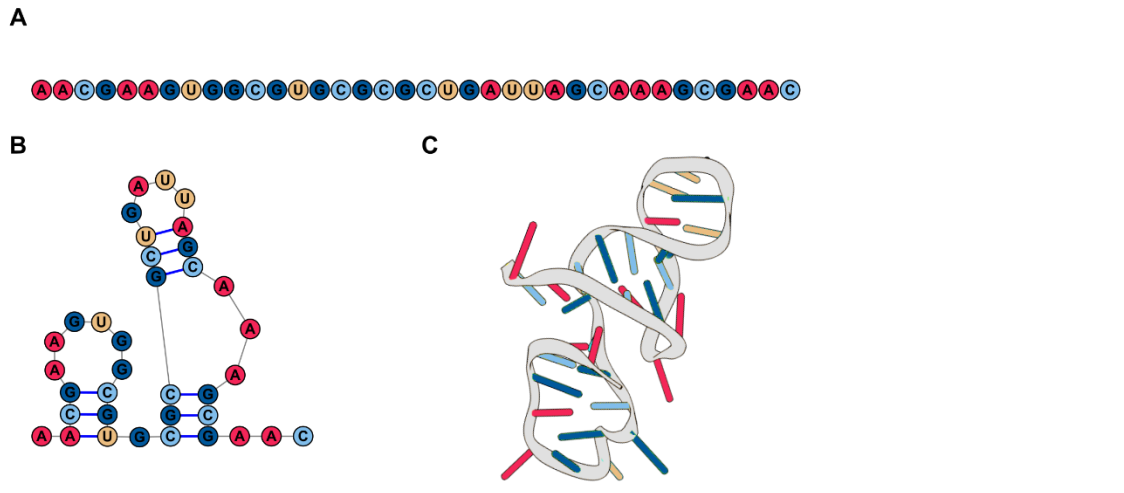

Figure 1.5 From primary to tertiary structure of an RNA. (**A**) a primary structure of an RNA consisting of specifically ordered nucleotides. (**B**) The secondary structure of the same RNA molecule (as predicted by an RNA prediction software (citation)) captures the base pairing interactions between different bases. (**C**) The tertiary structure represents the spatial organization of the same RNA molecule.

1.5. Single-molecule methods are a powerful tool to unveil the structural diversity of RNA

Traditional ensemble methods used previously to study these events focus on the whole population of molecular species. Consequently, the inherent heterogeneity within the system was inevitably lost, as the experimental approaches resulted in an averaged representation of molecular behavior.

To overcome the limitations of ensemble techniques, single-molecule methods such as force spectroscopy (optical tweezers or atomic-force microscopy) and single-molecule fluorescence offer the exciting possibility of studying individual molecules. These approaches have been successfully applied to probe RNA structures and investigate cellular processes like transcription and translation, providing novel insights into translation dynamics (51, 61, 83, 88, 116-124).

Moreover, several groups have successfully employed the optical tweezers to study frameshift RNAs. Halma et al. 2019 performed a detail dissection of the folding landscape of the West Nile virus frameshift RNA and account the high frameshifting efficiency of this PRF site to its folding versatility (52). Recently, also the pseudoknot RNA of the SARS-CoV-2 PRF site has been studied by the optical tweezers (125).

1.5.1. Optical tweezers – principle

Optical tweezers is a popular single molecule method to measure dynamics of RNA molecules. Recently, commercial devices have been introduced to the market, which allow scientists without a strong background in physics to employ this powerful technique to answer their biological questions. Nevertheless, understanding the fundamental principles of this method is beneficial for proper experimental planning and successful troubleshooting.

The ability of light beam to drag and move small dielectric objects has been first described by Arthur Ashkin in 1980s (126), and he was awarded with the Nobel prize in physics for this discovery in 2018. The principle of optical tweezers is based on the use of a focused laser beam with a gaussian distribution of light intensity (**Figure 1.6A**). When a dielectric bead happens to be in this beam, the light at the bead surface is partially reflected and refracted. As the light is refracted or reflected, the momentum of the photons changes and the momentum is transferred to the bead (in agreement with the momentum conservation law). Consequently, the two different event (light reflection or refraction) result in the momentum changes in opposite directions and two opposite-direction forces acting on the bead. Force resulting from the light reflection (scattering) acts on the bead in the direction of the light ray, while the force from the light refraction acts in the opposite

direction, dragging the bead back to the focus point. Similarly, the bead is dragged towards the beam focus (highest light intensity) in the directions perpendicular to the light ray direction (**Figure 1.6A**). Thus, the focused beam creates an optical trap (127-129).

For small displacements of the bead from the trap center, the trap behaves as a Hookean spring. Therefore, the force applied on the bead can be determined as it is approximately linearly dependent on the distance of the bead center from the trap center. The spring constant (trap stiffness) is dependent on the bead size, laser light intensity and the difference of optical density between the beads and the measurement environment (127-131). The trap stiffness is determined empirically during the experiments (131).

1.5.2. Optical tweezers setup

The optical tweezers device is composed of several optical devices on its own (**Figure 1.6B**). In the simplest setup, the OT consists of a bright-field microscope combined with the trapping beam part (129). The bright-field microscope serves for orientation in the flow cell and overall quality control of the examined beads, aggregates, bubbles, etc. The trapping beam moiety guides the trapping laser through an arrangement of several telescope lenses, mirrors, other optical utensils to the objective and flow cell and then further to the condenser and detectors. The trap focus can be adjusted or moved through the telescope (z-axis) and stirring mirrors (x and y-axis). In the dual-trap setup, the laser beam is also split to create two traps that can be moved or adjusted independently (129).

This can be further complemented with other parts, such as total internal reflection fluorescence or confocal microscope to allow the direct comparison of the force-distance readout with fluorescence information (129, 131, 132). In the standard force-ramp experimental set up (**Figure 1.6C**) a single strand molecule of RNA is hybridized with single strands of DNA handles complementary to RNA regions to yield a complex consisting of an ssRNA region in the middle flanked by two RNA/DNA handle regions. The handles are labelled at the 3’ and 5’ ends with digoxigenin and biotin, respectively. This complex is then immobilized at the surface of small polystyrene beads via interaction with anti-digoxigenin antibodies or streptavidin coupled to the bead surface, respectively. The beads are caught in the optical trap and by moving the traps apart the single-stranded part of the RNA/DNA complex is unfolded. Throughout the experiment, two parameters are detected – the distance between the two beads and the force applied by the optical trap on the construct. The experimental design and set-up are further discussed in **Chapter 2**.

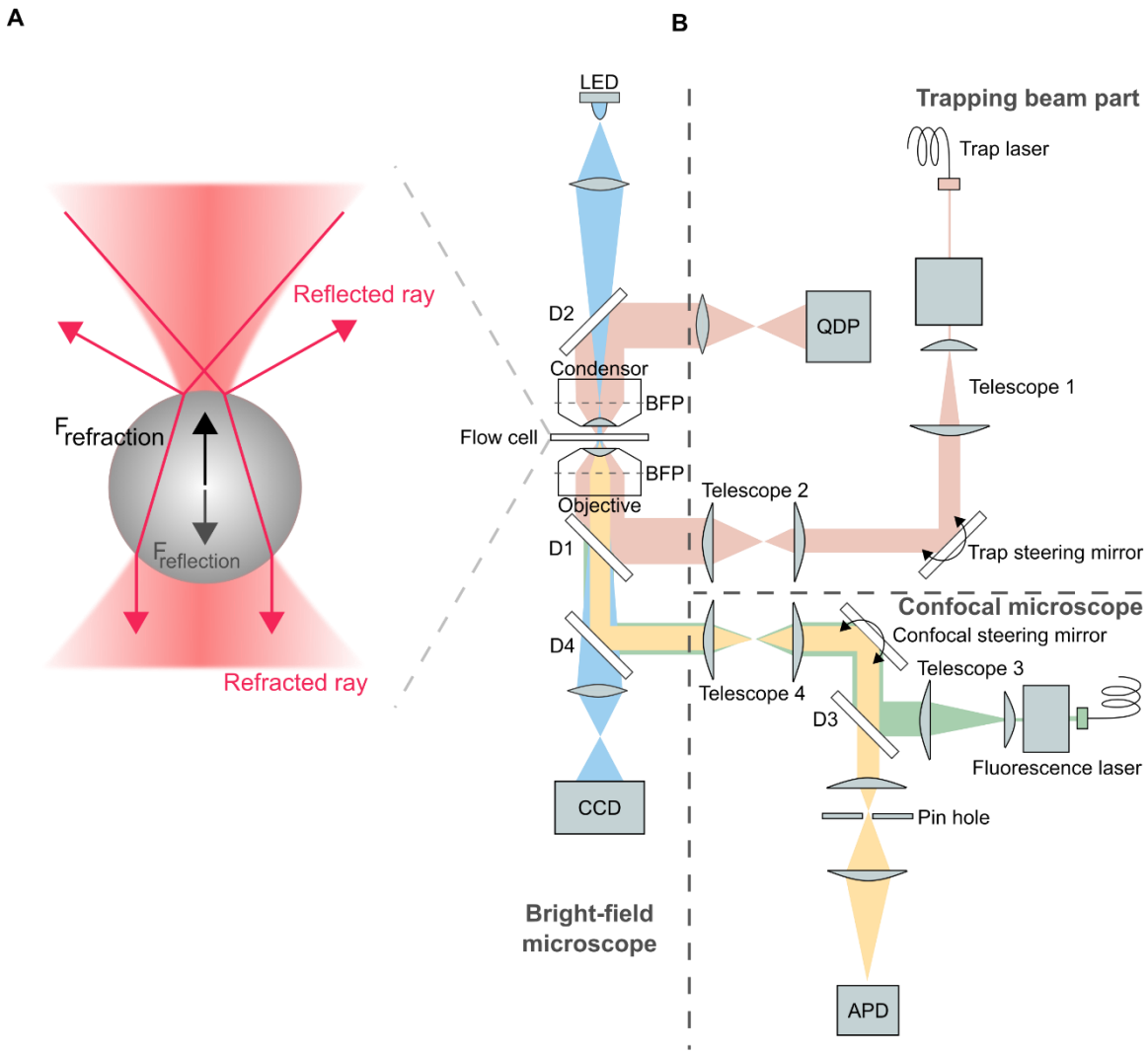

Figure 1.6 Scheme of optical tweezers coupled with a confocal microscope. (**A**) A dielectric sphere larger than the wavelength of light either reflects or refracts light (pink arrows) focused by a high-NA lens. The change in the direction of each ray corresponds to a change in the momentum of the light and an equal and opposite change in bead momentum. Reflected rays of light lose forward momentum that is gained by the bead, leading to a net force ($F_{reflection}$, gray arrow) pushing the bead along the direction of propagation of the light. Refracted rays are deflected forward because of the high incidence angle of the light, which generates momentum change and reactive force ($F_{refraction}$, black arrow) that pulls the bead towards the focus. (**B**) A high-power laser generates the trapping beam (pink), which is expanded by telescope 1. Beam-steering optics (steerable mirror (SM)) control the tilt in the beam axis. A high-numerical aperture objective focuses the trapping beam into the sample. Telescope 2 images the steering plane (at SM) onto the objective back focal plane (BFPO), so that tilting the beam displaces the trap in the sample plane. A condenser collects the light scattered by the trapped particle. A lens images the light at the condenser back focal plane (BFPC) onto a position-sensitive quadrant photodetector (QPD) for position/force detection. Two dichroic mirrors (D1 and D2) reflect the trapping beam and transmit visible light (blue) for bright-field illumination (light-emitting diode (LED)) and imaging (charge-coupled device (CCD)) of the

sample plane. A fluorescence excitation laser (green) is expanded and directed into the objective. The objective focuses the beam to a diffraction-limited spot on the sample plane and collects light emitted within the spot. The excitation spot is displaced in the sample plane by a steering mirror. The emitted light (yellow) travels back along the emission path, passing through a dichroic mirror and into a pin hole aperture to reject out-of-focus light. The emission light is detected by an avalanche photodiode (APD). The trapping and fluorescence excitation beams are interlaced by two out-of-phase acousto-optic modulators (AOM). Adapted from (129).

1.5.3. Things under tension – Applying external force on nucleic acids

The folding of biopolymers can be intuitively described by the free energy landscape (FEL). FEL shows the energy of the biopolymer as the function of the folding (133-136). The local or global minimal represent the stable conformations with the saddles representing the intermediate states (**Figure 1.7A**). In case of RNA, the FEL is mainly determined by the base-pairing interactions (but also other non-Watson-Crick-Franklin interactions) together determining the RNA conformation (9, 133, 137-139).

For the practical purposes, it is often convenient to depict the FEL as a 2D plot describing the dependency of biopolymer free energy on a chosen parameter, a reaction coordinate (**Figure 1.7A**). In case of optical tweezers, the ultimate readout of the experiment is the extension of the studied molecule, therefore extension of the studied molecule is usually chosen as the reaction coordinate. The FEL profile can be altered by the external environment changes, which include changes in the temperature, pH or salt concentration. Force represents an additional way how to probe the biopolymer folding and tilt the FEL profile (**Figure 1.7B**). By applying the force exclusively on the single molecule examined in between the beads, the force effect can be documented without affecting the surround environment (experimental conditions) (140).

Essentially, the FEL shows the differences between two (or more) states – folded vs. unfolded (local minima), with a transition state (local maximum) in between. The energy difference between the starting state (folded) and transition state is the activation energy ($\Delta G^{\ddagger}$) and $X_U^{\ddagger}$ is the distance to the transition state. The external force favors the extended (unfolded) state, therefore tilting the FEL towards the unfolded state. Depending on the amount of force applied, the probability of the two states is altered (**Figure 1.7C**).

The effect of force on the (bio)polymer can be described through mathematical models. In the case of biopolymers, especially nucleic acids, the two most commonly used models are the worm-like chain model (WLC) and freely-jointed chain model (FJC) (137, 138, 140-143). FJC is often used to describe the behavior of single-stranded nucleic acids, and it treats the polymer as a simply connected stiff monomers, with monomer length *l* (sometimes described as Kuhn length), which do not affect each other. The WLC, and its

extensible version (exWLC), describes a polymer that is semi-flexible – the individual monomers are not moving independently, but rather they are pointing in somewhat similar direction (143). Therefore, the overall curve profile is smoother. The polymer is characterized by the persistence length, a parameter describing the bending stiffness of the polymer. For the WLC to be applicable, it is crucial that the persistence length of a polymer is within a few orders of magnitude of the polymer. WLC can be used to describe the behavior of double stranded as well as single stranded nucleic acids, given that the correct value (estimate) of persistence length is used.

The chain models allow to link the applied external force with the polymer extension (X), which is defined as the end-to-end distance. For most of the experimental cases, the persistence length (L_P), or Kuhn length, is much smaller than the contour length (L_C, $L_P << L_C$), which characterizes the total length of the polymer. Given these aspects, the polymer behaves according to the central limit theorem, therefor without external force applied (F=0 pN) the actual end-to-end distance of the polymer is smaller than the contour length ($X < L_C$) (141, 143). When the external force is applied (F > 0 pN), the loss of entropy due to the limitation of the freedom of movement must be compensated by the work performed by the optical traps. The analysis of the experimental optical tweezers data is further discussed in **Chapter 3**.

However, single stranded RNA is a peculiar kind of biopolymer. It can fold on itself and form various conformation through the base-pairings (and other interactions). These secondary or tertiary structures are then unfolded during the experiments by applying the external force (47, 142, 144). The stability of the RNA structure depends on several things, among others the number of base-pairs (and the GC content), loop size, pH, salt concentration and ion type (monovalent, bivalent) (104, 105, 137, 142). The stability of the structure to the pulling is also dependent on the direction of the applied force(142). The different geometries of hairpins and pseudoknots (or other structures) relative to the applied force result in their different mechanical response, which can be utilized to distinguish various secondary and tertiary folding (**Figure 1.7F** and **G**) (52, 133, 139, 142, 145). In case of a simple hairpin, the force intensity is highest at the bottom of the stem (**Figure 1.7F**), while in case of a pseudoknot, the force exertion is distributed among more base-pairs (**Figure 1.7G**). Therefore, the pseudoknots show often higher mechanical stability comparing to the stem-like structure (34, 47, 49, 50, 52, 55, 146).

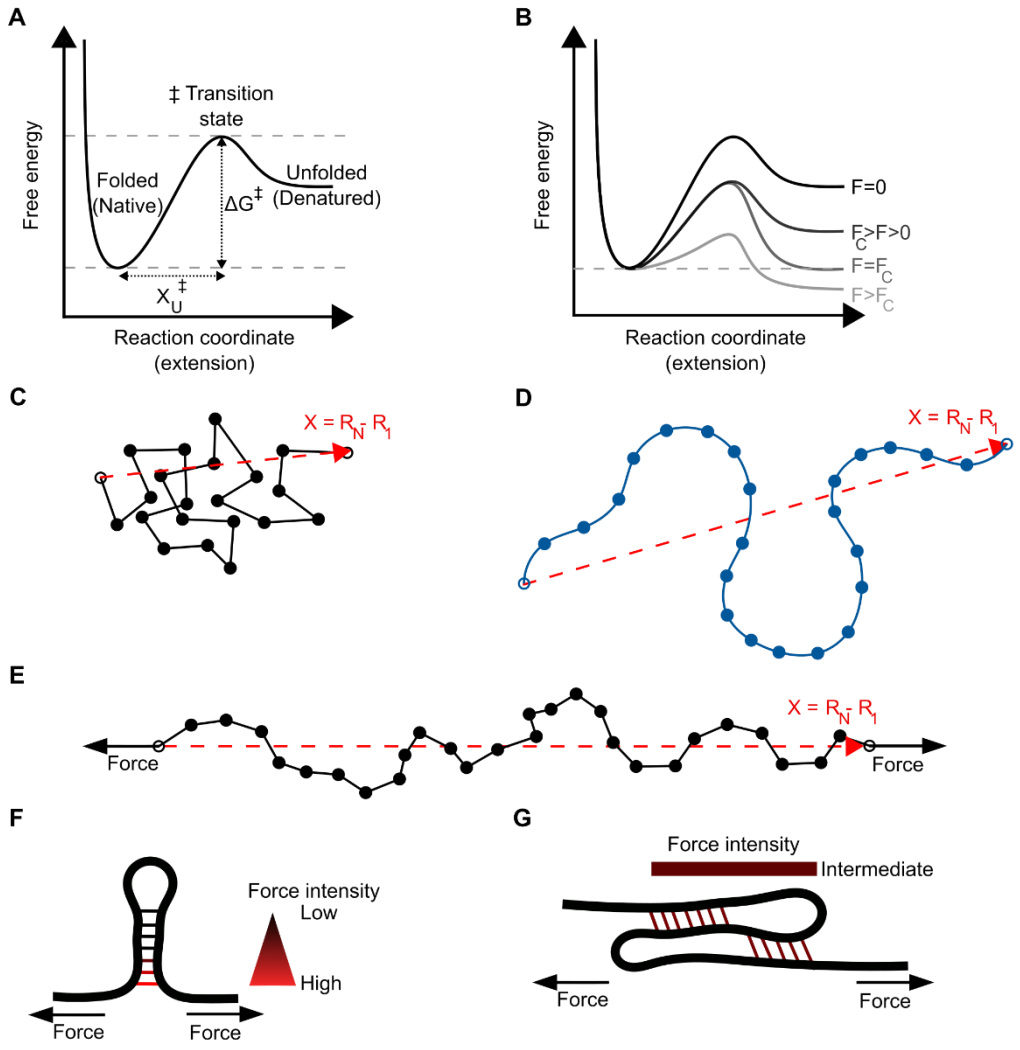

Figure 1.7 RNA as a biopolymer under external force. (**A**) Schematic representation of a free-energy landscape. $X_U^{\ddagger}$ is depicted as the distance from the folded (native) state to the transition state. $\Delta G^{\ddagger}$ represents the activation energy necessary to reach the transition state. (**B**) Free-energy landscape tilted by the constant external force. As the constant external force increases, the free-energy landscape is tilted to favor the unfolded state. At $F=F_C$, the free energies of folded and unfolded states are equal. (**C**) Freely jointed chain. (**D**) Worm-like chain. (**E**) External force applied on a polymer. The work done due to the external force compensates the reduced freedom of movement (entropy). (**F**) Force intensity distribution during applying an external force on an RNA hairpin. (**G**) Force intensity distribution during applying an external force on an RNA pseudoknot.

1.6. Scope of the thesis

Based on recent findings of –1PRF *trans*-acting factors existence, it is tempting to speculate that these factors may play a crucial role in both – precise regulation of infection progress in the case of pathogens as well as cellular defense pathways employed by the immune system. In my PhD, I aim to dissect the mechanical basis of –1PRF and to unveil the mechanisms of how *trans*-acting factors direct frameshifting efficiency rates. To this end, I will employ single-molecule optical tweezers coupled with the confocal microscope.

The aim of this thesis was to understand the mechanistic processes governing the programmed ribosomal frameshifting. Some of the major aims include:

1. **To establish the single-molecule optical tweezers technique in the Caliskan lab**. It is critical to have an appropriate experimental design and accurate analysis techniques to study RNAs due to the tendency of forming multiple conformations. For this reason, careful experimental design, execution as well as data analysis to study the molecules of interest was established (**Chapter 2** and **3**).
2. **To unveil the mechanical aspects of frameshifting.** It is becoming clear that to understand mechanisms of translation recoding events, such as frameshifting, one needs to go beyond the limits of ensemble techniques (**Chapter 4**).
3. **To study the role of *trans*-acting factors in −1PRF.** Properties of *cis*-acting factors, such as slippery sequence, spacer length, or RNA structure stability, have a rather static effect on the frameshifting efficiency. The engagement of *trans*-acting factors brings time-space dynamics to the frameshifting control. This is attractive not only to pathogens for precise timing during infection but also to host cells to fine-tune the complex regulation of cellular pathways (**Chapter 5** and **6**).

1.7. References:

1. Higgs, P.G. and Lehman, N. (2015) The RNA World: molecular cooperation at the origins of life. *Nature Reviews Genetics*, **16**, 7-17.
2. Bernhardt, H.S. (2012) The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)a. *Biology Direct*, **7**, 23.
3. Eddy, S.R. (2001) Non–coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, **2**, 919-929.
4. Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology*, **19**, 327-341.
5. Cech, T.R. (2000) Structural biology. The ribosome is a ribozyme. *Science*, **289**, 878-879.
6. Steitz, T.A. and Moore, P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends in Biochemical Sciences*, **28**, 411-418.
7. Lilley, D.M.J. (2001) The Ribosome Functions as a Ribozyme. *ChemBioChem*, **2**, 31-35.
8. Assmann, S.M., Chou, H.-L. and Bevilacqua, P.C. (2023) Rock, scissors, paper: How RNA structure informs function. *The Plant Cell*.
9. Westhof, E. and Fritsch, V. (2000) RNA folding: beyond Watson–Crick pairs. *Structure*, **8**, R55-R65.
10. Bevilacqua, P.C. and Blose, J.M. (2008) Structures, Kinetics, Thermodynamics, and Biological Functions of RNA Hairpins. *Annual Review of Physical Chemistry*, **59**, 79-103.
11. Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G. and Yusupov, M. (2012) One core, two shells: bacterial and eukaryotic ribosomes. *Nature Structural & Molecular Biology*, **19**, 560-567.
12. Lindqvist, L.M., Tandoc, K., Topisirovic, I. and Furic, L. (2018) Cross-talk between protein synthesis, energy metabolism and autophagy in cancer. *Curr Opin Genet Dev*, **48**, 104-111.
13. Holcik, M. and Pestova, T.V. (2007) Translation mechanism and regulation: old players, new concepts. Meeting on translational control and non-coding RNA. *EMBO reports*, **8**, 639-643.
14. Roy, B. and Jacobson, A. (2013) The intimate relationships of mRNA decay and translation. *Trends in Genetics*, **29**, 691-699.
15. Schuller, A.P. and Green, R. (2018) Roadblocks and resolutions in eukaryotic translation. *Nature Reviews Molecular Cell Biology*, **19**, 526-541.
16. Gebauer, F. and Hentze, M.W. (2004) Molecular mechanisms of translational control. *Nature Reviews Molecular Cell Biology*, **5**, 827-835.
17. Hershey, J.W., Sonenberg, N. and Mathews, M.B. (2012) Principles of translational control: an overview. *Cold Spring Harb Perspect Biol*, **4**.
18. Williamson, J.R. (2009) The Ribosome at Atomic Resolution. *Cell*, **139**, 1041-1043.
19. Schmeing, T.M. and Ramakrishnan, V. (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature*, **461**, 1234-1242.
20. Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene*, **234**, 187-208.
21. Loveland, A.B., Demo, G. and Korostelev, A.A. (2020) Cryo-EM of elongating ribosome with EF-Tu*GTP elucidates tRNA proofreading. *Nature*.
22. Loveland, A.B., Demo, G., Grigorieff, N. and Korostelev, A.A. (2017) Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature*, **546**, 113-117.
23. Dunkle, J.A. and Cate, J.H.D. (2010) Ribosome Structure and Dynamics During Translocation and Termination. *Annual Review of Biophysics*, **39**, 227-244.
24. Matheisl, S., Berninghausen, O., Becker, T. and Beckmann, R. (2015) Structure of a human translation termination complex. *Nucleic acids research*, **43**, 8615-8626.

25. Caliskan, N., Peske, F. and Rodnina, M.V. (2015) Changed in translation: mRNA recoding by −1 programmed ribosomal frameshifting. *Trends in Biochemical Sciences*, **40**, 265-274.
26. Riegger, R.J. and Caliskan, N. (2022) Thinking Outside the Frame: Impacting Genomes Capacity by Programmed Ribosomal Frameshifting. *Frontiers in Molecular Biosciences*, **9**.
27. Firth, A.E. and Brierley, I. (2012) Non-canonical translation in RNA viruses. *J. Gen Virol*, **93**, 1385-1409.
28. Rodnina, M.V., Korniy, N., Klimova, M., Karki, P., Peng, B.-Z., Senyushkina, T., Belardinelli, R., Maracci, C., Wohlgemuth, I., Samatova, E. *et al.* (2020) Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Research*, **48**, 1056-1067.
29. Pestova, T.V. and Hellen, C.U.T. (2003) Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes Dev*, **17**, 181-186.
30. Fernández, I.S., Bai, X.-C., Murshudov, G., Scheres, S.H.W. and Ramakrishnan, V. (2014) Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell*, **157**, 823-831.
31. Ryabova, L.A., Pooggin, M.M. and Hohn, T. (2006) Translation reinitiation and leaky scanning in plant viruses. *Virus Res*, **119**, 52-62.
32. Hinnebusch, A.G. (2011) Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiol Mol Biol Rev*, **75**, 434-467, first page of table of contents.
33. Jaafar, Z.A. and Kieft, J.S. (2019) Viral RNA structure-based strategies to manipulate translation. *Nature Reviews Microbiology*, **17**, 110-123.
34. Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J. and Brierley, I. (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, **441**, 244-247.
35. Brierley, I. and Dos Ramos, F.J. (2006) Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res*, **119**, 29-42.
36. Lin, Z., Gilbert, R.J.C. and Brierley, I. (2012) Spacer-length dependence of programmed −1 or −2 ribosomal frameshifting on a U 6 A heptamer supports a role for messenger RNA (mRNA) tension in frameshifting. *Nucleic Acids Research*, **40**, 8674-8689.
37. Tsuchihashi, Z. and Brown, P.O. (1992) Sequence requirements for efficient translational frameshifting in the Escherichia coli dnaX gene and the role of an unstable interaction between tRNA(Lys) and an AAG lysine codon. *Genes Dev*, **6**, 511-519.
38. Gesteland, R.F. and Atkins, J.F. (1996) RECODING: Dynamic Reprogramming of Translation. *Annual review of biochemistry*, **65**, 741-768.
39. Horsfield, J.A., Wilson, D.N., Mannering, S.A., Adamski, F.M. and Tate, W.P. (1995) Prokaryotic ribosomes recode the HIV-1 gag-pol-1 frameshift sequence by an E/P site post-translocation simultaneous slippage mechanism. *Nucleic Acids Res*, **23**, 1487-1494.
40. Licznar, P., Mejlhede, N., Prère, M.F., Wills, N., Gesteland, R.F., Atkins, J.F. and Fayet, O. (2003) Programmed translational -1 frameshifting on hexanucleotide motifs and the wobble properties of tRNAs. *EMBO J*, **22**, 4770-4778.
41. Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E. and Baranov, P.V. (2016) Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res*, **44**, 7007-7078.
42. Bock, L.V., Caliskan, N., Korniy, N., Peske, F., Rodnina, M.V. and Grubmüller, H. (2019) Thermodynamic control of −1 programmed ribosomal frameshifting. *Nature communications*, **10**, 4598.
43. Choi, J., O'Loughlin, S., Atkins, J.F. and Puglisi, J.D. (2020) The energy landscape of −1 ribosomal frameshifting. *Science Advances*, **6**, eaax6969.

44. Poulis, P., Patel, A., Rodnina, M.V. and Adio, S. (2022) Altered tRNA dynamics during translocation on slippery mRNA as determinant of spontaneous ribosome frameshifting. *Nature communications*, **13**, 4231.
45. Poulis, P., Peske, F. and Rodnina, M.V. (2023) The many faces of ribosome translocation along the mRNA: reading frame maintenance, ribosome frameshifting and translational bypassing.
46. Bao, C., Zhu, M., Nykonchuk, I., Wakabayashi, H., Mathews, D.H. and Ermolenko, D.N. (2022) Specific length and structure rather than high thermodynamic stability enable regulatory mRNA stem-loops to pause translation. *Nature communications*, **13**, 988.
47. Green, L., Kim, C.-H., Bustamante, C. and Tinoco, I. (2008) Characterization of the Mechanical Unfolding of RNA Pseudoknots. *Journal of Molecular Biology*, **375**, 511-528.
48. Chen, G., Chang, K.Y., Chou, M.Y., Bustamante, C. and Tinoco, I., Jr. (2009) Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc Natl Acad Sci U S A*, **106**, 12706-12711.
49. Zhong, Z., Yang, L., Zhang, H., Shi, J., Vandana, J.J., Lam, D.T., Olsthoorn, R.C., Lu, L. and Chen, G. (2016) Mechanical unfolding kinetics of the SRV-1 gag-pro mRNA pseudoknot: possible implications for -1 ribosomal frameshifting stimulation. *Sci Rep*, **6**, 39549.
50. Hansen, T.M., Reihani, S.N.S., Oddershede, L.B. and Sørensen, M.A. (2007) Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5830-5835.
51. Ritchie, D.B., Foster, D.A. and Woodside, M.T. (2012) Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc Natl Acad Sci U S A*, **109**, 16167-16172.
52. Halma, M.T.J., Ritchie, D.B., Cappellano, T.R., Neupane, K. and Woodside, M.T. (2019) Complex dynamics under tension in a high-efficiency frameshift stimulatory structure. *Proceedings of the National Academy of Sciences*, **116**, 19500.
53. Giedroc, D.P. and Cornish, P.V. (2009) Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res*, **139**, 193-208.
54. Kim, H.-K., Liu, F., Fei, J., Bustamante, C., Gonzalez, R.L. and Tinoco, I. (2014) A frameshifting stimulatory stem loop destabilizes the hybrid state and impedes ribosomal translocation. *Proc Natl Acad Sci U S A*, **111**, 5538-5543.
55. Hori, N., Denesyuk, N.A. and Thirumalai, D. (2016) Salt Effects on the Thermodynamics of a Frameshifting RNA Pseudoknot under Tension. *Journal of Molecular Biology*, **428**, 2847-2859.
56. Howard, M.T., Gesteland, R.F. and Atkins, J.F. (2004) Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides. *RNA*, **10**, 1653-1661.
57. Loughran, G., Firth, A.E. and Atkins, J.F. (2011) Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome. *Proc Natl Acad Sci U S A*, **108**, E1111-1119.
58. Kwak, H., Park, M.W. and Jeong, S. (2011) Annexin A2 binds RNA and reduces the frameshifting efficiency of infectious bronchitis virus. *PLoS One*, **6**, e24067.
59. Petty, R.V., Basta, H.A., Bacot-Davis, V.R., Brown, B.A. and Palmenberg, A.C. (2014) Binding interactions between the encephalomyocarditis virus leader and protein 2A. *J Virol*, **88**, 13503-13509.
60. Li, Y., Treffers, E.E., Napthine, S., Tas, A., Zhu, L., Sun, Z., Bell, S., Mark, B.L., van Veelen, P.A., van Hemert, M.J. *et al.* (2014) Transactivation of programmed ribosomal frameshifting by a viral protein. *Proceedings of the National Academy of Sciences*, **111**, E2172.

61. Chen, Y.-T., Chang, K.-C., Hu, H.-T., Chen, Y.-L., Lin, Y.-H., Hsu, C.-F., Chang, C.-F., Chang, K.-Y. and Wen, J.-D. (2017) Coordination among tertiary base pairs results in an efficient frameshift-stimulating RNA pseudoknot. *Nucleic acids research*, **45**, 6011-6022.
62. Ritchie, D.B., Soong, J., Sikkema, W.K.A. and Woodside, M.T. (2014) Anti-frameshifting Ligand Reduces the Conformational Plasticity of the SARS Virus Pseudoknot. *Journal of the American Chemical Society*, **136**, 2196-2199.
63. Zimmer, M.M., Kibe, A., Rand, U., Pekarek, L., Ye, L., Buck, S., Smyth, R.P., Cicin-Sain, L. and 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**, 7193.
64. Wang, X., Xuan, Y., Han, Y., Ding, X., Ye, K., Yang, F., Gao, P., Goff, S.P. and Gao, G. (2019) Regulation of HIV-1 Gag-Pol Expression by Shiftless, an Inhibitor of Programmed -1 Ribosomal Frameshifting. *Cell*, **176**, 625-635.e614.
65. Napthine, S., Ling, R., Finch, L.K., Jones, J.D., Bell, S., Brierley, I. and Firth, A.E. (2017) Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nature communications*, **8**, 15582.
66. Napthine, S., Bell, S., Hill, C.H., Brierley, I. and Firth, A.E. (2019) Characterization of the stimulators of protein-directed ribosomal frameshifting in Theiler's murine encephalomyelitis virus. *Nucleic Acids Res*, **47**, 8207-8223.
67. Hill, C.H., Pekarek, L., Napthine, S., Kibe, A., Firth, A.E., Graham, S.C., Caliskan, N. and Brierley, I. (2021) Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch. *Nature communications*, **12**, 7166.
68. Henderson, C.M., Anderson, C.B. and Howard, M.T. (2006) Antisense-induced ribosomal frameshifting. *Nucleic Acids Research*, **34**, 4302-4310.
69. Bhatt, P.R., Scaiola, A., Loughran, G., Leibundgut, M., Kratzel, A., Meurs, R., Dreos, R., O'Connor, K.M., McMillan, A., Bode, J.W. *et al.* (2021) Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science*, **372**, 1306-1313.
70. Hill, C.H., Cook, G.M., Napthine, S., Kibe, A., Brown, K., Caliskan, N., Firth, A.E., Graham, S.C. and Brierley, I. (2021) Investigating molecular mechanisms of 2A-stimulated ribosomal pausing and frameshifting in Theilovirus. *Nucleic Acids Res*, **49**, 11938-11958.
71. Carocci, M., Cordonnier, N., Huet, H., Romey, A., Relmy, A., Gorna, K., Blaise-Boisseau, S., Zientara, S. and Kassimi, L.B. (2011) Encephalomyocarditis virus 2A protein is required for viral pathogenesis and inhibition of apoptosis. *J Virol*, **85**, 10741-10754.
72. Caliskan, N. and Hill, C.H. (2022) Insights from structural studies of the cardiovirus 2A protein. *Biosci Rep*, **42**.
73. Beura, L.K., Dinh, P.X., Osorio, F.A. and Pattnaik, A.K. (2011) Cellular poly(c) binding proteins 1 and 2 interact with porcine reproductive and respiratory syndrome virus nonstructural protein 1β and support viral replication. *J Virol*, **85**, 12939-12949.
74. Napthine, S., Treffers, E.E., Bell, S., Goodfellow, I., Fang, Y., Firth, A.E., Snijder, E.J. and Brierley, I. (2016) A novel role for poly(C) binding proteins in programmed ribosomal frameshifting. *Nucleic Acids Res*, **44**, 5491-5503.
75. Kelly, J.A., Woodside, M.T. and Dinman, J.D. (2021) Programmed −1 Ribosomal Frameshifting in coronaviruses: A therapeutic target. *Virology*, **554**, 75-82.
76. Strand, E., Hollås, H., Sakya, S.A., Romanyuk, S., Saraste, M.E.V., Grindheim, A.K., Patil, S.S. and Vedeler, A. (2021) Annexin A2 binds the internal ribosomal entry site of c-myc mRNA and regulates its translation. *RNA Biol*, **18**, 337-354.
77. Suzuki, Y., Chin, W.X., Han, Q., Ichiyama, K., Lee, C.H., Eyo, Z.W., Ebina, H., Takahashi, H., Takahashi, C., Tan, B.H. *et al.* (2016) Characterization of RyDEN (C19orf66) as an Interferon-Stimulated Cellular Inhibitor against Dengue Virus Replication. *PLoS Pathog*, **12**, e1005357.

78. Qu, X., Wen, J.-D., Lancaster, L., Noller, H.F., Bustamante, C. and Tinoco, I., Jr. (2011) The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature*, **475**, 118-121.
79. Dinman, J.D. (2012) Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip Rev RNA*, **3**, 661-673.
80. Caliskan, N., Katunin, V.I., Belardinelli, R., Peske, F. and Rodnina, M.V. (2014) Programmed -1 frameshifting by kinetic partitioning during impeded translocation. *Cell*, **157**, 1619-1631.
81. Kim, H.-K. and Tinoco, I., Jr. (2017) EF-G catalyzed translocation dynamics in the presence of ribosomal frameshifting stimulatory signals. *Nucleic Acids Research*, **45**, 2865-2874.
82. Xie, P. (2016) Model of the pathway of −1 frameshifting: Long pausing. *Biochemistry and Biophysics Reports*, **5**, 408-424.
83. Desai, V.P., Frank, F., Lee, A., Righini, M., Lancaster, L., Noller, H.F., Tinoco, I. and Bustamante, C. (2019) Co-temporal Force and Fluorescence Measurements Reveal a Ribosomal Gear Shift Mechanism of Translation Regulation by Structured mRNAs. *Molecular Cell*, **75**, 1007-1019.e1005.
84. Nguyen, H.A., Hoffer, E.D. and Dunham, C.M. (2019) Importance of a tRNA anticodon loop modification and a conserved, noncanonical anticodon stem pairing in tRNA_{CGG}^{Pro} for decoding. *Journal of Biological Chemistry*, **294**, 5281-5291.
85. Bao, C., Loerch, S., Ling, C., Korostelev, A.A., Grigorieff, N. and Ermolenko, D.N. (2020) mRNA stem-loops can pause the ribosome by hindering A-site tRNA binding. *Elife*, **9**.
86. Kaiser, C.M., Goldman, D.H., Chodera, J.D., Tinoco, I. and Bustamante, C. (2011) The Ribosome Modulates Nascent Protein Folding. *Science*, **334**, 1723-1727.
87. Alexander, L., Goldman, D., Tinoco, I. and Bustamante, C.J. (2017) The Ribosome Alters the Folding of a Multidomain Nascent Protein. *Biophysical Journal*, **112**, 40a-41a.
88. Goldman, D.H., Kaiser, C.M., Milin, A., Righini, M., Tinoco, I. and Bustamante, C. (2015) Mechanical force releases nascent chain–mediated ribosome arrest in vitro and in vivo. *Science*, **348**, 457-460.
89. Temperley, R., Richter, R., Dennerlein, S., Lightowlers, R.N. and Chrzanowska-Lightowlers, Z.M. (2010) Hungry codons promote frameshifting in human mitochondrial ribosomes. *Science*, **327**, 301.
90. Weiss, R., Lindsley, D., Falahee, B. and Gallant, J. (1988) On the mechanism of ribosomal frameshifting at hungry codons. *J Mol Biol*, **203**, 403-410.
91. Gallant, J.A. and Lindsley, D. (1992) Leftward ribosome frameshifting at a hungry codon. *J Mol Biol*, **223**, 31-40.
92. Caliskan, N., Wohlgemuth, I., Korniy, N., Pearson, M., Peske, F. and Rodnina, M.V. (2017) Conditional Switch between Frameshifting Regimes upon Translation of dnaX mRNA. *Molecular Cell*, **66**, 558-567.e554.
93. Woodside, M.T., Behnke-Parks, W.M., Larizadeh, K., Travers, K., Herschlag, D. and Block, S.M. (2006) Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proceedings of the National Academy of Sciences*, **103**, 6190.
94. Hyeon, C. and Thirumalai, D. (2007) Mechanical unfolding of RNA: from hairpins to structures with internal multiloops. *Biophysical journal*, **92**, 731-743.
95. Fallmann, J., Will, S., Engelhardt, J., Grüning, B., Backofen, R. and Stadler, P.F. (2017) Recent advances in RNA folding. *Journal of Biotechnology*, **261**, 97-104.
96. Yang, L., Zhong, Z., Tong, C., Jia, H., Liu, Y. and Chen, G. (2018) Single-Molecule Mechanical Folding and Unfolding of RNA Hairpins: Effects of Single A-U to A·C Pair Substitutions and Single Proton Binding and Implications for mRNA Structure-Induced −1 Ribosomal Frameshifting. *Journal of the American Chemical Society*, **140**, 8172-8184.

97. Vicens, Q. and Kieft, J.S. (2022) Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences*, **119**, e2112677119.
98. Smit, S., Rother, K., Heringa, J. and Knight, R. (2008) From knotted to nested RNA structures: A variety of computational methods for pseudoknot removal. *RNA*, **14**, 410-416.
99. Bellaousov, S., Reuter, J.S., Seetin, M.G. and Mathews, D.H. (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res*, **41**, W471-474.
100. Subpaiboonkit, S., Thammarongtham, C., Cutler, R.W. and Chaijaruwanich, J. (2013) RNA secondary structure prediction using conditional random fields model. *International Journal of Data Mining and Bioinformatics*, **7**, 118-134.
101. Lorenz, R., Wolfinger, M.T., Tanzer, A. and Hofacker, I.L. (2016) Predicting RNA secondary structures from sequence and probing data. *Methods*, **103**, 86-98.
102. Ren, J., Rastegari, B., Condon, A. and Hoos, H.H. (2005) HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA*, **11**, 1494-1504.
103. Moss, W.N. (2013) In Lorsch, J. (ed.), *Methods in Enzymology*. Academic Press, Vol. 530, pp. 3-65.
104. Draper, D.E. (2004) A guide to ions and RNA structure. *Rna*, **10**, 335-343.
105. Tan, Z.J. and Chen, S.J. (2011) Salt contribution to RNA tertiary structure folding stability. *Biophys J*, **101**, 176-187.
106. Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M. and Grzybowska, E.A. RNA–protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biology*, **9**, 190096.
107. Armaos, A., Zacco, E., Sanchez de Groot, N. and Tartaglia, G.G. (2021) RNA-protein interactions: Central players in coordination of regulatory networks. *Bioessays*, **43**, e2000118.
108. Re, A., Joshi, T., Kulberkyte, E., Morris, Q. and Workman, C.T. (2014) RNA-protein interactions: an overview. *Methods Mol Biol*, **1097**, 491-521.
109. Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature Reviews Genetics*, **8**, 776-790.
110. Li, B., Cao, Y., Westhof, E. and Miao, Z. (2020) Advances in RNA 3D Structure Modeling Using Experimental Data. *Frontiers in Genetics*, **11**.
111. Zhang, K., Zheludev, I.N., Hagey, R.J., Haslecker, R., Hou, Y.J., Kretsch, R., Pintilie, G.D., Rangan, R., Kladwang, W., Li, S. *et al.* (2021) Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nat. Struct. Mol. Biol*, **28**, 747-754.
112. Zhang, J., Fei, Y., Sun, L. and Zhang, Q.C. (2022) Advances and opportunities in RNA structure experimental determination and computational modeling. *Nature methods*, **19**, 1193-1207.
113. Magnus, M., Boniecki, M.J., Dawson, W. and Bujnicki, J.M. (2016) SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. *Nucleic Acids Res*, **44**, W315-319.
114. Biesiada, M., Purzycka, K.J., Szachniuk, M., Blazewicz, J. and Adamiak, R.W. (2016) Automated RNA 3D Structure Prediction with RNAComposer. *Methods Mol Biol*, **1490**, 199-215.
115. Li, J. and Chen, S.-J. (2021) RNA 3D Structure Prediction Using Coarse-Grained Models. *Frontiers in Molecular Biosciences*, **8**.
116. Qu, X., Lancaster, L., Noller, H.F., Bustamante, C. and Tinoco, I., Jr. (2012) Ribosomal protein S1 unwinds double-stranded RNA in multiple steps. *Proc Natl Acad Sci U S A*, **109**, 14458-14463.
117. Liu, T., Kaplan, A., Alexander, L., Yan, S., Wen, J.-D., Lancaster, L., Wickersham, C.E., Fredrick, K., Noller, H., Tinoco, I., Jr. *et al.* (2014) Direct measurement of the mechanical work during translocation by the ribosome. *eLife*, **3**, e03406.

118. John, R., Davenport, n., Wuite, G.J.L., Landick, R. and Bustamante, C. (2000) Single-Molecule Study of Transcriptional Pausing and Arrest by E. coli RNA Polymerase. *Science*, **287**, 2497-2500.
119. Svetlov, V. and Nudler, E. (2012) Unfolding the Bridge between Transcription and Translation. *Cell*, **150**, 243-245.
120. Blanchard, S.C. (2009) Single-molecule observations of ribosome function. *Current opinion in structural biology*, **19**, 103-109.
121. Juette, M.F., Terry, D.S., Wasserman, M.R., Zhou, Z., Altman, R.B., Zheng, Q. and Blanchard, S.C. (2014) The bright future of single-molecule fluorescence imaging. *Curr Opin Chem Biol*, **20**, 103-111.
122. Ferguson, A., Wang, L., Altman, R.B., Terry, D.S., Juette, M.F., Burnett, B.J., Alejo, J.L., Dass, R.A., Parks, M.M., Vincent, C.T. *et al.* (2015) Functional Dynamics within the Human Ribosome Regulate the Rate of Active Protein Synthesis. *Molecular cell*, **60**, 475-486.
123. Rundlet, E.J., Holm, M., Schacherl, M., Natchiar, S.K., Altman, R.B., Spahn, C.M.T., Myasnikov, A.G. and Blanchard, S.C. (2021) Structural basis of early translocation events on the ribosome. *Nature*, **595**, 741-745.
124. Holm, M., Natchiar, S.K., Rundlet, E.J., Myasnikov, A.G., Watson, Z.L., Altman, R.B., Wang, H.-Y., Taunton, J. and Blanchard, S.C. (2023) mRNA decoding in human is kinetically and structurally distinct from bacteria. *Nature*, **617**, 200-207.
125. Neupane, K., Zhao, M., Lyons, A., Munshi, S., Ileperuma, S.M., Ritchie, D.B., Hoffer, N.Q., Narayan, A. and Woodside, M.T. (2021) Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers. *Nat. Commun.*, **12**, 4749.
126. Ashkin, A., Dziedzic, J.M., Bjorkholm, J.E. and Chu, S. (1986) Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.*, **11**, 288-290.
127. Moffitt, J.R., Chemla, Y.R., Smith, S.B. and Bustamante, C. (2008) Recent advances in optical tweezers. *Annual review of biochemistry*, **77**, 205-228.
128. Rocha, M.S. (2009) Optical tweezers for undergraduates: Theoretical analysis and experiments. *American Journal of Physics*, **77**, 704-712.
129. Bustamante, C.J., Chemla, Y.R., Liu, S. and Wang, M.D. (2021) Optical tweezers in single-molecule biophysics. *Nature Reviews Methods Primers*, **1**, 25.
130. Hashemi Shabestari, M., Meijering, A.E.C., Roos, W.H., Wuite, G.J.L. and Peterman, E.J.G. (2017) In Spies, M. and Chemla, Y. R. (eds.), *Methods in Enzymology*. Academic Press, Vol. 582, pp. 85-119.
131. Pekarek, L., Buck, S. and Caliskan, N. (2022) Optical Tweezers to Study RNA-Protein Interactions in Translation Regulation. *JoVE*, e62589.
132. Whitley, K.D., Comstock, M.J. and Chemla, Y.R. (2017) High-Resolution "Fleezers": Dual-Trap Optical Tweezers Combined with Single-Molecule Fluorescence Detection. *Methods in molecular biology (Clifton, N.J.)*, **1486**, 183-256.
133. Tinoco, I., Jr., Li, P.T.X. and Bustamante, C. (2006) Determination of thermodynamics and kinetics of RNA reactions by force. *Q Rev Biophys*, **39**, 325-360.
134. Li, P.T.X., Bustamante, C. and Tinoco, I. (2007) Real-time control of the energy landscape by force directs the folding of RNA molecules. *Proceedings of the National Academy of Sciences*, **104**, 7039.
135. Deng, N.-J. and Cieplak, P. (2010) Free energy profile of RNA hairpins: a molecular dynamics simulation study. *Biophysical journal*, **98**, 627-636.
136. Gupta, A.N., Vincent, A., Neupane, K., Yu, H., Wang, F. and Woodside, M.T. (2011) Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nature Physics*, **7**, 631-634.

137. Wen, J.-D., Manosas, M., Li, P.T.X., Smith, S.B., Bustamante, C., Ritort, F. and Tinoco, I. (2007) Force Unfolding Kinetics of RNA Using Optical Tweezers. I. Effects of Experimental Variables on Measured Results. *Biophysical Journal*, **92**, 2996-3009.
138. Manosas, M., Wen, J.D., Li, P.T.X., Smith, S.B., Bustamante, C., Tinoco, I. and Ritort, F. (2007) Force Unfolding Kinetics of RNA using Optical Tweezers. II. Modeling Experiments. *Biophysical Journal*, **92**, 3010-3021.
139. Li, P.T.X., Vieregg, J. and Tinoco, I. (2008) How RNA Unfolds and Refolds. *Annual review of biochemistry*, **77**, 77-100.
140. Tinoco, I., Jr. (2004) Force as a useful variable in reactions: unfolding RNA. *Annual review of biophysics and biomolecular structure*, **33**, 363-385.
141. Wang, M.D., Yin, H., Landick, R., Gelles, J. and Block, S.M. (1997) Stretching DNA with optical tweezers. *Biophys J*, **72**, 1335-1346.
142. Stephenson, W., Wan, G., Tenenbaum, S.A. and Li, P.T. (2014) Nanomanipulation of single RNA molecules by optical tweezers. *J Vis Exp*.
143. Odijk, T. (1995) Stiff Chains and Filaments under Tension. *Macromolecules*, **28**, 7016-7018.
144. Hyeon, C. and Thirumalai, D. (2006) Forced-unfolding and force-quench refolding of RNA hairpins. *Biophysical journal*, **90**, 3410-3427.
145. Li, P.T.X., Bustamante, C. and Tinoco, I. (2006) Unusual mechanical stability of a minimal RNA kissing complex. *Proceedings of the National Academy of Sciences*, **103**, 15847-15852.
146. Chen, G., Wen, J.D. and Tinoco, I., Jr. (2007) Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *RNA*, **13**, 2175-2188.

Chapter 2

"It's a(n optical) trap!"

Gial Ackbar

2. Optical Tweezers to Study RNA-Protein Interactions in Translation Regulation

Lukas Pekarek[1], Stefan Buck[1], Neva Caliskan[1,2]

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

[2]Medical Faculty, Julius-Maximilians University Würzburg, Würzburg, Germany

Corresponding author:

Neva Caliskan (Neva.Caliskan@helmholtz-hiri.de)

Published: Journal of Visualized Experiments, Issue 180, Pages e62589 (2022)

2.1. Summary

This protocol presents a complete experimental workflow for studying RNA-protein interactions using optical tweezers. Several possible experimental setups are outlined including the combination of optical tweezers with confocal microscopy.

2.2. Abstract

RNA adopts diverse structural folds, which are essential for its functions and thereby can impact diverse processes in the cell. In addition, the structure and function of an RNA can be modulated by various *trans*-acting factors, such as proteins, metabolites or other RNAs. Frameshifting RNA molecules, for instance, are regulatory RNAs located in coding regions, which direct translating ribosomes into an alternative open reading frame, and thereby act as gene switches. They may also adopt different folds after binding to proteins or other trans-factors. To dissect the role of RNA-binding proteins in translation and how they modulate RNA structure and stability, it is crucial to study the interplay and mechanical features of these RNA-protein complexes simultaneously. This work illustrates how to employ single-molecule-fluorescence-coupled optical tweezers to explore the conformational and thermodynamic landscape of RNA-protein complexes at a high resolution. As an example, the interaction of the SARS-CoV-2 programmed ribosomal frameshifting element with the *trans*-acting factor zinc-finger antiviral protein (ZAP) is elaborated. In addition, fluorescence-labeled ribosomes were monitored using the confocal unit, which would ultimately enable the study of translation elongation. The

fluorescence coupled OT assay can be widely applied to explore diverse RNA-protein complexes or trans-acting factors regulating translation and could facilitate studies of RNA-based gene regulation.

2.3. Introduction

Transfer of genetic information from DNA to proteins through mRNAs is a complex biochemical process, which is precisely regulated on all levels through macromolecular interactions inside cells. For translational regulation, RNA-protein interactions confer a critical role to rapidly react to various stimuli and signals[1,2]. Some RNA-protein interactions affect mRNA stability and thereby alter the time an RNA is translationally active. Other RNA-protein interactions are associated with recoding mechanisms such as stop-codon readthrough, bypassing, or programmed ribosomal frameshifting (PRF)[3-7]. Recently, a number of RNA-binding proteins (RBPs) have been demonstrated to interact with stimulatory mRNA elements and the translation machinery to dictate when and how much recoding will occur in the cell[7-11]. Thus, to dissect the role of RNA-binding proteins in translation and how they modulate RNA structure and stability, it is pivotal to study the interaction principles and mechanical properties of these RNA-protein complexes in detail.

Decades of work have laid the foundation to study the multi-step and multi-component process of translation, which relies on intricate communication between the RNA and protein components of the translation machinery to achieve speed and accuracy[12-14]. A crucial next step in understanding complex regulatory events is determining the forces, timescales, and structural determinants during translation at high precision[12,15-17]. The study of RNA conformational dynamics and especially how *trans*-acting auxiliary factors act on the RNA structure during translation have been further illuminated by the emergence of single-molecule tools, including optical tweezers or zero-mode waveguides[16-26].

Optical tweezers (OT) represent a highly precise single-molecule technique, which has been applied to study many sorts of RNA-dependent dynamic processes including transcription, and translation[26-32]. The use of optical tweezers has allowed probing of molecular interactions, nucleic acid structures, and thermodynamic properties, kinetics, and energetics of these processes in detail[16,17,22,33-39]. Optical tweezers assay is based on the entrapment of microscopic objects with a focused laser beam. In a typical OT experiment, the molecule of interest is tethered between two transparent (usually polystyrene) beads (**Fig. 2.1A**)[27]. These beads are then caught by optical traps, which behave like springs. Thus, the force applied on the molecule can be calculated based on the bead's displacement from the center of the focused laser beam (trap center). Recently,

optical tweezers have been combined with confocal microscopy (**Fig. 2.1B**), enabling fluorescence or Förster resonance energy transfer (FRET) measurements[40-42]. This opens a whole new field of possible experiments allowing simultaneous measurement and, therefore, precise correlation of force spectroscopy and fluorescence data.

Here, we demonstrate experiments using the optical tweezers combined with confocal microscopy to study protein-RNA interactions regulating translational frameshifting. Between the objective and the condenser, a flow cell with five channels enables continuous sample application with laminar flow. Through the microfluidic channels, various components can be injected directly, which decreases the hands-on time as well as allowing very little sample consumption throughout the experiment.

First, a basic guideline to assist the design of OT experiments is proposed and advantages as well as pitfalls of various setups are discussed. Next, the preparation of samples and experimental workflows are described, and a protocol for the data analysis is provided. To represent an example, we outline the results obtained from RNA stretching experiments to study the SARS-CoV-2 frameshifting RNA element (**Fig. 2.2A**) with the *trans*-acting factor zinc-finger antiviral protein (ZAP), which alters the translation of the viral RNA from an alternative reading frame[43]. Additionally, it is demonstrated that fluorescence-labeled ribosomes can be employed in this OT confocal assay, which would be useful to monitor the processivity and speed of the translation machinery. The method presented here can be used to rapidly test the effect of different buffers, ligands, or other cellular components to study various aspects of translation. Finally, common experimental pitfalls and how to troubleshoot them are discussed. Below, some crucial points in experimental design are outlined.

Construct design

In principle, there are two common approaches to create an OT-compatible RNA construct. The first approach employs a long RNA molecule that is hybridized with complementary DNA handles, thus yielding a construct consisting of two RNA/DNA hybrid regions flanking a single-stranded RNA sequence in the middle (**Fig. 2.2B**). This approach is employed in most OT RNA experiments[33,44,45].

The second approach takes advantage of dsDNA handles with short (around 20 nt) overhangs[15,17]. These overhangs are then hybridized with the RNA molecule. Although more complicated in design, the use of dsDNA handles overcomes some of limitations of the DNA/RNA-hybrid system. In principle, even very long handles (>10kb) can be implemented, which is more convenient for confocal measurements. In addition, the RNA molecule can be ligated to DNA handles to increase tether stability.

End-labeling strategy

The construct must be tethered to beads via a strong molecular interaction. While there are approaches available for covalent bonding of handles to beads[46], strong but non-covalent interactions such as streptavidin-biotin and digoxigenin-antibody are commonly used in OT experiments[15,33,35,45]. In the described protocol, the construct is labeled with biotin or digoxigenin, and the beads are coated with streptavidin or antibodies against digoxigenin, respectively (**Fig. 2.1A**). This approach would be suitable for applying forces up to approximately 60 pN (per tether)[47]. Furthermore, the use of different 5' and 3' labeling strategies allow determining the orientation of the tether formed between the beads[17].

Protein labeling for fluorescence measurements

For the confocal imaging, there are several commonly used approaches for fluorescence labeling. For instance, fluorophores can be covalently attached to amino acid residues that are found natively in proteins or introduced by site-directed mutagenesis through a reactive organic group. Thiol or amine-reactive dyes can be used for labeling of cysteine and lysine residues, respectively. There are several reversible protection methods to increase the specificity of labeling[48,49], however native proteins would typically be labeled at multiple residues. Although the small size of the fluorophore may confer an advantage, non-specific labeling might interfere with the protein activity and thus signal intensity may vary[49]. Also, depending on the labeling efficiency signal intensity may differ between different experiments. Therefore, an activity check should be performed prior to the experiment.

In case the protein of interest contains an N- or C-terminal tag, such as a His-tag or strep-tag, specific labeling of these tags represents another popular approach. Moreover, tag-targeted labeling reduces the chance of the fluorophore interfering with protein activity and can enhance solubility[49]. However, tag-specific labeling usually yields mono-fluorophore labeled proteins, which might be challenging to detect. Another way of specific labeling can be accomplished by employing antibodies.

Microfluidics setup

The combination of OT with a microfluidics system allows a rapid transition between different experimental conditions. Moreover, current systems take advantage of maintaining the laminar flow inside the flow cell, which precludes the mixing of liquids from other channels in the perpendicular direction relative to the flow direction. Therefore,

laminar flow is particularly advantageous for the experimental design. Currently, flow cells with up to 5 channels are commonly employed **(Fig. 2.3**).

2.4. Protocol

2.4.1. Sample preparation

2.4.1.1. Clone the sequence of interest into the vector containing the Lambda DNA fragments, which serves as the handle sequences (**Fig. 2.2**)[43,50].

2.4.1.2. First generate a DNA template for subsequent in vitro transcription via PCR (**Fig. 2.2B**; reaction 1). At this PCR step, the T7 promoter is added in the 5' end of the sense DNA molecule[32,33,43,50]. Set the PCR reaction according to Table 1. Run the PCR in 50 µL aliquots with appropriate cycles in the thermocycler.

2.4.1.3. Prepare the handles by two separate PCR reactions (**Table 2.1, Fig. 2.2B**; reaction 2 and 3). First, generate the 5’ handle by PCR. Then, generate the 3' handle and simultaneously label it with digoxigenin by using a 5' digoxigenin-labeled primer[32,33,43,50].

2.4.1.4. After the PCR, purify the DNA using silica spin columns.

2.4.1.5. Carry out the in vitro transcription reaction using T7 RNA polymerase (**Table 2.2**)[32,33,43,50]. Incubate the reaction at 37 °C for 2-4 h depending on the length of the RNA. Next, add DNase I to the reaction and incubate at 37 °C for 30 min to digest the DNA template. Purify the RNA using silica spin columns.

2.4.1.6. During the labeling reaction of the 5’ handle (**Table 2.3**), add biotin-16-dUTP at the 3' end of the handle by T4 DNA polymerase[38,50]. Perform the reaction at room temperature for 1-2 h. Afterwards, purify the DNA using silica spin columns.

NOTE: Since the 5' handle must be labeled at its 3' end (**Fig. 2.2B**), the labeling cannot be performed during the PCR.

2.4.1.7. Mix the components mentioned above – 5' handle (3' labeled with biotin), 3' handle (5' labeled with digoxigenin), and RNA – in a 1:1:1 molar ratio in annealing buffer (80% formamide, 400 mM NaCl, 40 mM HEPES, pH 7.5, 0.5 mM EDTA, pH 8), to obtain the desired RNA/DNA hybrid (**Table 2.4**). Heat the annealing mixture up to 85 °C for 10 min and then slowly cool down to 4 °C.

2.4.1.8. Mix the annealed sample with 1/10 of volume of 3 M sodium acetate (pH 5), 3 volumes of ice-cold ethanol and incubate at -80 °C for at least 1 h or at -20 °C overnight.

2.4.1.9. Centrifuge the samples at 15,000 × g for 30 min at 4 °C. Discard the supernatant and dry the pellet (usually not visible) under vacuum.

2.4.1.10. Finally, resuspend, the pellet in 50 µL of RNase-free water and make aliquots. Store the aliquots at -80 °C until used. For short term storage, the samples can be also stored at -20 °C.

2.4.2. Instrument setup

NOTE: The following protocol is optimized for the commercial optical tweezers instrument C-Trap from LUMICKS company. Therefore, adjustments to the presented steps might be necessary while using other optical tweezers instruments. If not used, the microfluidics system of the machine is kept in bleach (sodium hypochlorite solution) and must be washed before use.

2.4.2.1. Discard the bleach and fill the syringes with 1 mL of RNase-free water.

2.4.2.2. Add 50 µL of 0.5 M sodium thiosulfate to at least 1 mL of the RNase-free water and thoroughly wash the system (1 bar, at least 0.5 mL) to eliminate the remaining bleach in the system.

2.4.2.3. Discard the sodium thiosulfate solution from the syringes. Replace syringes with fresh ones and wash the system with at least 0.5 mL of RNase-free water.

NOTE: Be careful, that the microfluidics system never runs dry to avoid air bubbles in the system.

2.4.2.4. Put 2 drops of immersion oil (refractive index of 1.33) or approximately 70 µL of water on top of the objective.

2.4.2.5. Place the flow cell inside the holding frame in its position.

2.4.2.6. Put 2 drops of immersion oil (refractive index of 1.51) on top of the flow cell.

2.4.2.7. Turn on the laser device in the tweezers machine. Once it is running, turn on the trapping laser in the software interface at 100%.

2.4.2.8. Using diagnostic cameras (Z finder), adjust the Z-axis to the middle of the chamber between the second and the third reflections (interfaces) where the refraction rings are the biggest, by turning the micro screw.

NOTE: Each time the objective is moved closer to the measuring chamber and the focal plane of the objective crosses the interface between two phases, a reflection can be recognized in the Z-finder mode. There are 4 interfaces possible: (i) water/immersion oil and bottom glass (ii) bottom glass and buffer inside the chamber (iii) buffer inside the chamber and top glass (iv) top glass and immersion oil for condenser.

2.4.2.9. Adjust the condenser position (set trapping laser to approximately 50%) so the condenser touches the immersion oil on top of the measuring chamber.

2.4.2.10. Adjust the focus by moving slowly down/up with the condenser, so approx. 10 light bands are shown in the moon mode (diagnostic cameras).

2.5. Sample measurement

2.5.1. Incubate anti-digoxigenin-coated beads (AD) with the sample constructs (3 µL of 0.1% (w/v) AD bead suspension + 4 µL of sample) and with 1 µL of RNase inhibitors and 8 µL of the assay buffer (300 mM KCl, 5 mM $MgCl_2$, 20 mM HEPES, pH 7.6, 0.05% Tween 20, 5 mM DTT) at RT for 10-20 min. After the incubation, dilute the sample in 500 µL of assay buffer.

NOTE: It is recommended to add oxygen scavengers, particularly during fluorescence measurements to the buffer in order to prevent oxidative damage. Here oxygen scavenger system containing glucose (8.3 mg/mL), glucose oxidase (40 U/mL) and catalase (185 U/mL) was used.

2.5.2. Mix 0.8 µL of 1% (w/v) streptavidin-coated (SA) beads with 1 mL of assay buffer.

2.5.3. Discard water from the syringes and fill the syringes with respective suspensions/solutions. Wash for at least 2 min at approximately 1 bar, and then start catching beads.

NOTE: Depending on the experimental set-up, different channel arrangements may be used (**Fig. 2.3**). Typically, one flow channel is filled with anti-digoxigenin beads carrying the RNA molecule. A second channel is filled with the streptavidin-coated beads. Buffer channel is used to form the tethers (**Fig. 2.3B**). A fourth channel can be employed to load the RNA binding protein, or alternatively RBP can be added directly in the buffer channel (**Fig. 2.3C**).

2.5.4. To capture the beads, move the optical traps apart from each other. First move to the AD channel and catch an AD-bead in trap 1. Next, move the stage to the SA-channel and catch a single SA bead by trap 2.

NOTE: Try to stay at the interface of the buffer and bead channels to avoid losing the already caught bead, or to prevent catching multiple beads by the same trap.

2.5.5. Once the beads of the right size are captured, move to the buffer channel and stop the laminar flow. Next, perform force calibration to check trap stiffness. The respective stiffness values should not differ in the x/y axis by more than 10-15%.

NOTE: Adjust the laser power or the laser split between the traps according to bead size. Force calibration does not have to be done for every bead pair as long as the bead templates match (similarity score > 0.9). However, it should be performed regularly, or at least every time assay conditions are changed.

2.5.6. Start fishing for a tether by moving the beads close to each other, waiting for a few seconds, and then moving them back apart, repeat until a tether is formed. A tether formation results in an increase of measured force upon pulling the two beads away from each other.

NOTE: To avoid formation of multiple tethers, the beads should not be moved too close. Upon catching a tether between the two beads, tether quality can be checked by finding the overstretching plateau. The plateau should be between 50 to 60 pN for a single tether.

2.5.7. Upon obtaining a tether, start the measurement. Depending on the phenomenon studied different measurement setups should be chosen (**Fig. 2.1B-D**).

NOTE: Usually at the beginning of the experiment, a force-ramp experiment is conducted to check the tether quality and probe the behavior. Afterward, one may also start the constant-force or constant-position experiments to study the state transitions further. Once sufficient number of measurements have been performed on an RNA sample to determine its behavior, labeled factors can be added to the system to perform confocal measurements.

2.5.8. To perform fluorescence measurements, turn on the confocal lasers and photon counter unit in the optical tweezers instrument.

2.5.9. Turn on the excitation laser of desired wavelength in the software interface and set the power of the laser to 5% or higher, depending on the fluorophore.

NOTE: While not measuring lower the power setting of the excitation laser to 0% to avoid excessive photodamage to the sample.

2.5.10. Start imaging the sample by using image functions of the software.

NOTE: In order to get well-focused images, the focal plane of the confocal microscope and optical traps have to be aligned. For this purpose, autofluorescence of the polystyrene beads in the blue laser channel can be employed. The focal plane of optical traps is moved up or down in the z-axis until the image of beads reaches its highest diameter. At this position, the fluorescence signal from the molecule tethered between the beads can be measured.

2.5.11. To use the kymograph function, specify the x-y position of the kymograph axis so that it allows detection of the tether between the beads.

2.5.12. Throughout the measurement, buffer composition can be easily changed by either moving the beads to different channels or by changing the buffer supplied in the microfluidics system.

2.6. Data analysis

Raw data pre-processing

2.6.1. By using a simple script, downsample the data (**Fig. 2.4A**) enough to (i) allow faster subsequent data processing but (ii) still contain all the critical information. Usually, 100-5000 Hz is suitable for this purpose.

NOTE: The data gathering frequency in optical tweezers experiments is often higher than it is necessary for the analysis – in the presented experiments, the data gathering frequency is set to 78 125 Hz by default. Since storage space is limited, it is convenient and timesaving to reduce the sampling rate of the data. Here, the raw data were downsampled by a factor of 30.

2.6.2. Next, employ a signal filter to reduce the high frequency measurement noise from the signal (**Fig. 2.4A**). Adjust the filter degree and cut-off frequency parameters accordingly to optimize data output of different experiments (**Fig. 2.5**).

NOTE: Amongst signal filters, Butterworth filter[51] is one of the most widely used. A custom-written python script allowing the pre-processing of raw data is provided in the supplementary data. Downsampling and signal filtering parameters (cut-off frequency, filter degree) need to be optimized for different experiments.

For force-ramp data analysis, use the following steps:

2.6.3. Mark the steps either manually by finding corresponding points on the force trajectory plot or by using custom-written scripts. Unfolding steps are characterized by a sudden drop in force combined with an increase in distance in the force-distance (FD) curve.

2.6.4. Once unfolding events are marked, fit different regions of the FD curve using appropriate models (**Fig. 2.4D**).

NOTE: For the region before the first unfolding step, the tether can be considered "double-stranded" and is commonly fit using an extensible Worm-like-chain model (WLC)[47,52,53]. The parts after the first unfolding event are considered a combination of double-stranded nucleotides (handles) and single-stranded nucleotides (unfolded RNA molecule). Therefore, the fit is more complex – usually a combination of 2 WLC models or WLC and Freely-jointed chain (FJC) models[36,39,52]. The extensible WLC model has two main fit parameters the contour length (L_C) and the persistence length (L_P). Contour length corresponds to the length of the fully stretched molecule and persistence length defines the bending properties of the molecule of interest. The model can be described with the following equation (1). WLC can be used to model the behavior of both folded as well as unfolded regions, although for each of these a separate model with different parameters has to be employed.

$$(1) \qquad x_{WLC} = L_C \left[1 - \frac{1}{2}\left(\frac{k_B T}{F \cdot L_P}\right)^{1/2} + \frac{F}{S}\right]$$

where x is extension, L_C is contour length, F is force, L_P is persistence length, k_B is Boltzmann constant, T is Thermodynamic temperature, and S is stretch modulus.

The second model called Freely-jointed chain (FJC) is commonly used to describe behavior of unfolded single stranded regions. It uses similar parameters of the polymers but treats each unit of the “chain” as a rigid rod, here corresponding to the nucleotides of the unfolded single stranded region. The following equation (2) describes this model:

$$(2) \qquad x_{FJC} = L_C \left[\coth\left(\frac{2F \cdot L_P}{k_B T}\right) - \frac{k_B T}{2F \cdot L_P}\right]\left(1 + \frac{F}{S}\right)$$

NOTE: Our lab has recently developed an algorithm that allows batch processing of the raw force-ramp data called Practical Optical Tweezers Analysis TOol (POTATO)[54]. The algorithm downsamples and filters the data, then it identifies possible unfolding steps and finally performs data fitting. The POTATO is built in a user-friendly graphical user interface (GUI) (https://github.com/REMI-HIRI/POTATO).

Process constant-force data as follows:

NOTE: The following instructions can be analogically applied on constant-position data.

2.6.5. For the constant-force data, plot the distance over time (**Fig. 2.5**). A histogram showing the frequency (counts) of different conformations over the relative change in position is a useful way to characterize various dominant and minor states (**Fig. 2.7**).

2.6.6. Fit the histogram using (multiple) Gaussian functions to estimate the overall percentage of individual conformers at a given force (**Fig. 2.7C**). The Gaussian fits, mean position, and the standard deviation outlines the force-related relationship among different populations.

NOTE: A custom-written python script allowing pre-processing and basic bimodal Gaussian fitting of constant-force data is provided in the supplementary data. Parameters (cut-off frequency, filter degree, expected means, standard deviation values and amplitudes) need to be optimized for different experiments.

2.6.7. Next, employ the Hidden Markov model to further analyze the states, which may uncover additional folding intermediates (conformers)[55]. For further information on the constant-force and Hidden Markov model, one may refer to[55-58].

2.7. Representative Results

In this section, focus is mainly given on measurements of RNA-protein/ligand interactions by the fluorescence optical tweezers. For a description of general RNA optical

tweezers experiments and corresponding representative results, see[32]. For more detailed discussion of the RNA/DNA-protein interactions, also see[1,2,26,59,60].

In principle, binding of an RBP or any other trans-acting factor of interest on the RNA stabilizes, destabilizes, or may alter the conformation of the molecule. Below, a depiction of the mechanical observables for each effect are shown. However, the actual effect observed for a given RNA-protein complex is not limited to these below-mentioned scenarios.

Stabilization

The RNA structure can be specifically recognized and bound by the protein or other ligands[45,61-64]. The formation of the bonds is accompanied by a release of energy. Therefore, an extra energetical barrier must be overcome in order to unfold the given RNA structure. As a result, an increase in the mean unfolding force might be observed[50,65]. The stabilization of the RNA structure by binding of an external agent (protein, small molecule, other trans-acting factors) may also result in a change of the folding kinetics of the structure[45]. For that, further measurements can be performed in the constant-force mode, where less frequent transitions between the folding intermediates as well as force-shift in the equilibrium can be observed.

Destabilization

Some proteins recognize certain sequence motifs rather than specific RNA structures. The binding sites may vary from a highly specific motif to a more general pattern such as GC or AU rich stretches[60,66]. Nevertheless, if the protein preferentially binds to the unfolded single-stranded RNA conformation, the equilibrium between the folded and unfolded state can be shifted towards the unfolded state[36,43,67]. In **Fig. 2.6** and **Fig. 2.7** examples of such behavior are depicted.

Structure alteration

In some instances, RBPs (or other ligands) might combine both mechanisms mentioned above in such a way that the RBP destabilizes the previously dominant conformation and shifts the equilibrium towards an alternative RNA structure[44,68,69]. The switch to an alternative state may result in a change in the observed conformational population frequencies as well as the occurrence or disappearance of individual folding states. These changes can be first observed in force-ramp experiments and can be further investigated by the constant-force (or constant-position) experiments.

Effect of the *trans*-acting factor on RNA folding/unfolding

Here, an RNA sequence corresponding to the -1 programmed ribosomal frameshifting element of SARS-CoV-2 was studied. This RNA element is predicted to form an H-type pseudoknot[70,71]. In the example force-distance trajectories, the RNA unfolds and refolds in two consecutive steps (**Fig. 2.6A**). These two steps likely correspond to the two stem loops that are the prerequisite for the pseudoknot formation. In this case, the pseudoknot was not observed either because the RNA did not fully fold or formed an alternative structure competing with the pseudoknot. Upon addition of the *trans*-acting factor ZAP, a sudden disappearance of the refolding events and a huge hysteresis was observed (**Fig. 2.6B**)[43]. This suggests that the protein binds to the single-stranded state of the RNA, impeding the formation of secondary structures. Furthermore, constant-force experiments confirm the results of force-ramp experiments. Accordingly, while the RNA is fully folded at around 10 pN, the presence of the protein shifts the refolding towards lower forces, and at 10 pN the RNA is still mostly occupying the unfolded state (**Fig. 2.7**).

OT measurements coupled with confocal microscopy

Next, exemplary results are shown for the non-specific as well as specific binding of different fluorophores and labeled ribosomes (**Fig. 2.8**). In the first example, Sytox dye was used to label the tethered DNA/RNA hybrid. With increasing force, the dye binding is more abundant resulting in higher fluorescence signal. Once the force is too high, the tether breaks, and the fluorescence signal is lost (**Fig. 2.8B**). For the experiments with bacterial ribosomes (**Fig. 2.8C**), non-specific labeling of the lysine residues was employed using N-hydroxysuccinimide (NHS) conjugated to a red fluorescent dye. Although there is a risk of decreasing the activity of labeled protein/complex, the big advantage is stronger signal achieved as each ribosome is (on average) labeled by multiple fluorophores. The RNA construct contained a ribosome binding site (RBS) recognized by bacterial ribosomes, which was placed in the 5’ proximity of the studied RNA sequence. Upon binding of the ribosomes, the fluorescence signal is observed on the tether. Fluorescence data can be further analyzed using image analysis tools[72], and the results can be combined with the force data, allowing the study of folding transitions.

2.8. Figure and Table Legends

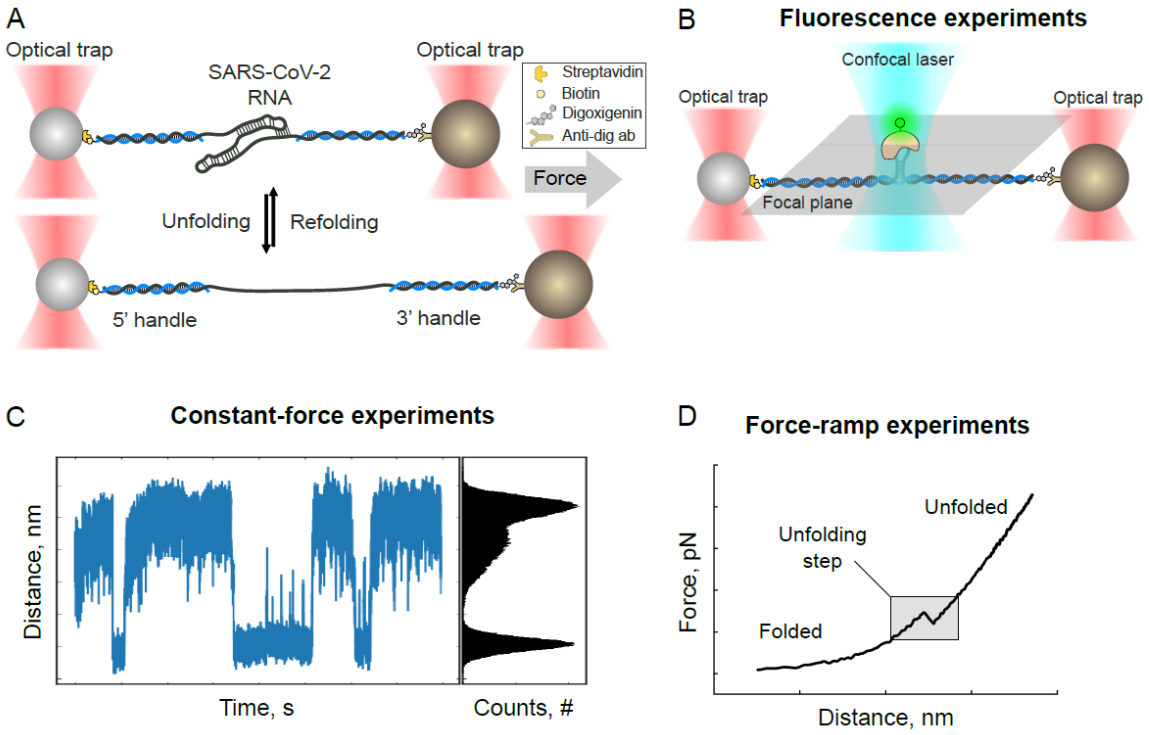

Figure 2.1 Schematic of the OT experiment and possible measurement approaches. (A) Schematic illustrating the optical tweezers experiments with the SARS-CoV-2 frameshifting RNA in the middle. RNA is hybridized to ssDNA handles and immobilized on beads. These are used to exert pulling force on the RNA with a focused laser beam. The force is gradually increased until the RNA is unfolded (bottom). **(B)** Schematic of confocal microscopy combined with optical tweezers to monitor binding of labeled factor to RNA. **(C)** Example constant-force data can be obtained by fixing the force at a constant value over time, which allows to precisely measure dwell time of the conformers. **(D)** Example force-distance (FD) curve obtained from a force-ramp measurement. The unfolding step is observed as a sudden rupture in the FD profile.

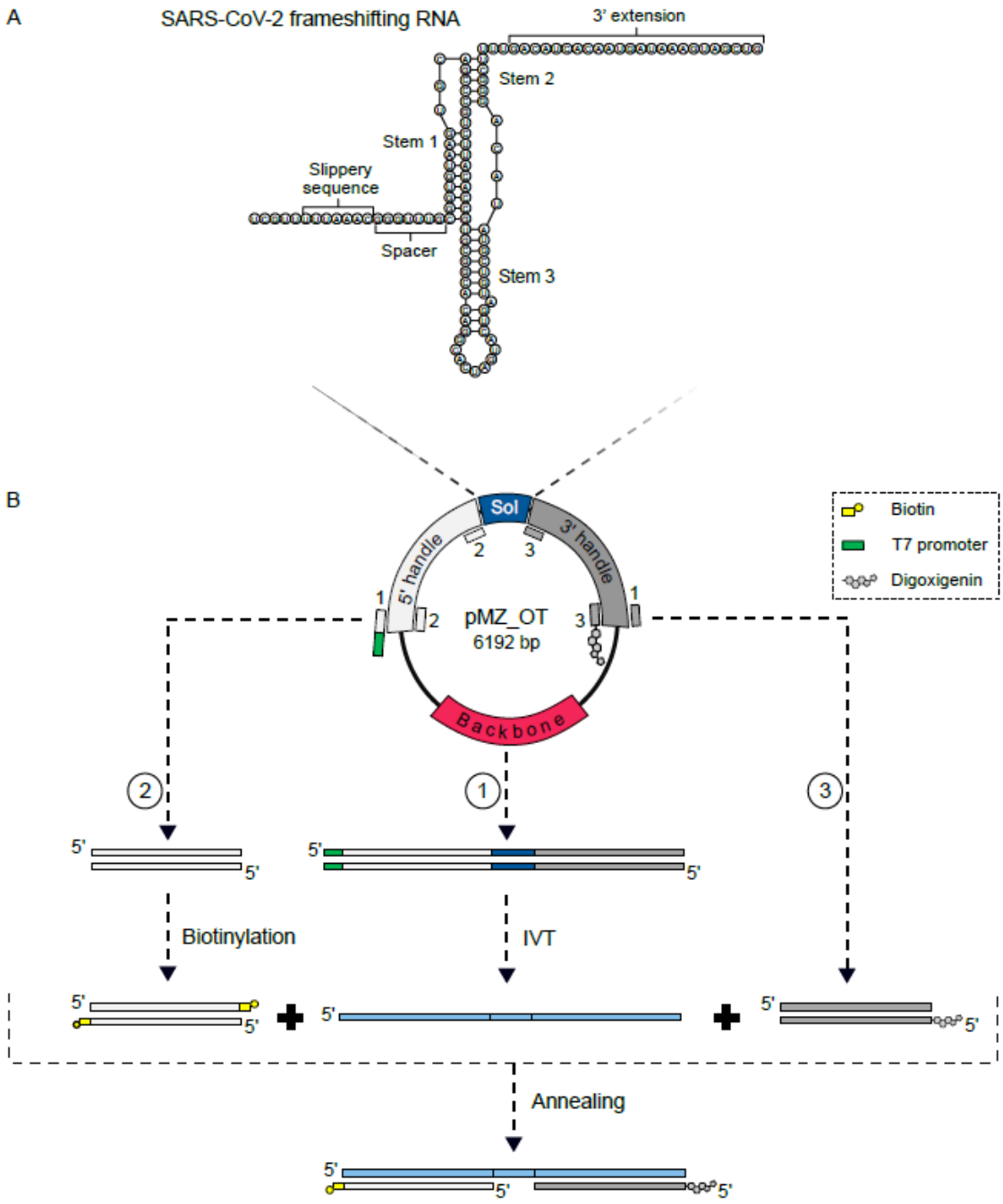

Figure 2.2 A general scheme of OT sample synthesis. (A) Example sequence and predicted secondary structure of the studied SARS-CoV-2 frameshifting RNA employed in the study. **(B)** A vector containing the sequence of interest (SoI) flanked by two handle regions serves as the template for generation of the DNA/RNA construct in 3 PCR reactions. Primers are depicted and numbered in the scheme according to their binding sites in the corresponding PCR. PCR 1 yields the *in vitro* transcription template, which is subsequently used for the *in vitro* transcription (IVT) reaction to generate the long RNA molecule (light blue). PCR 2 yields the 5' handle, which is later 3' labeled with biotin. PCR 3 using the forward primer conjugated to digoxigenin produces the 3' digoxigenin-labeled handle. Finally, the two handles and RNA are annealed to give a DNA/RNA hybrid construct suitable for optical tweezers measurements.

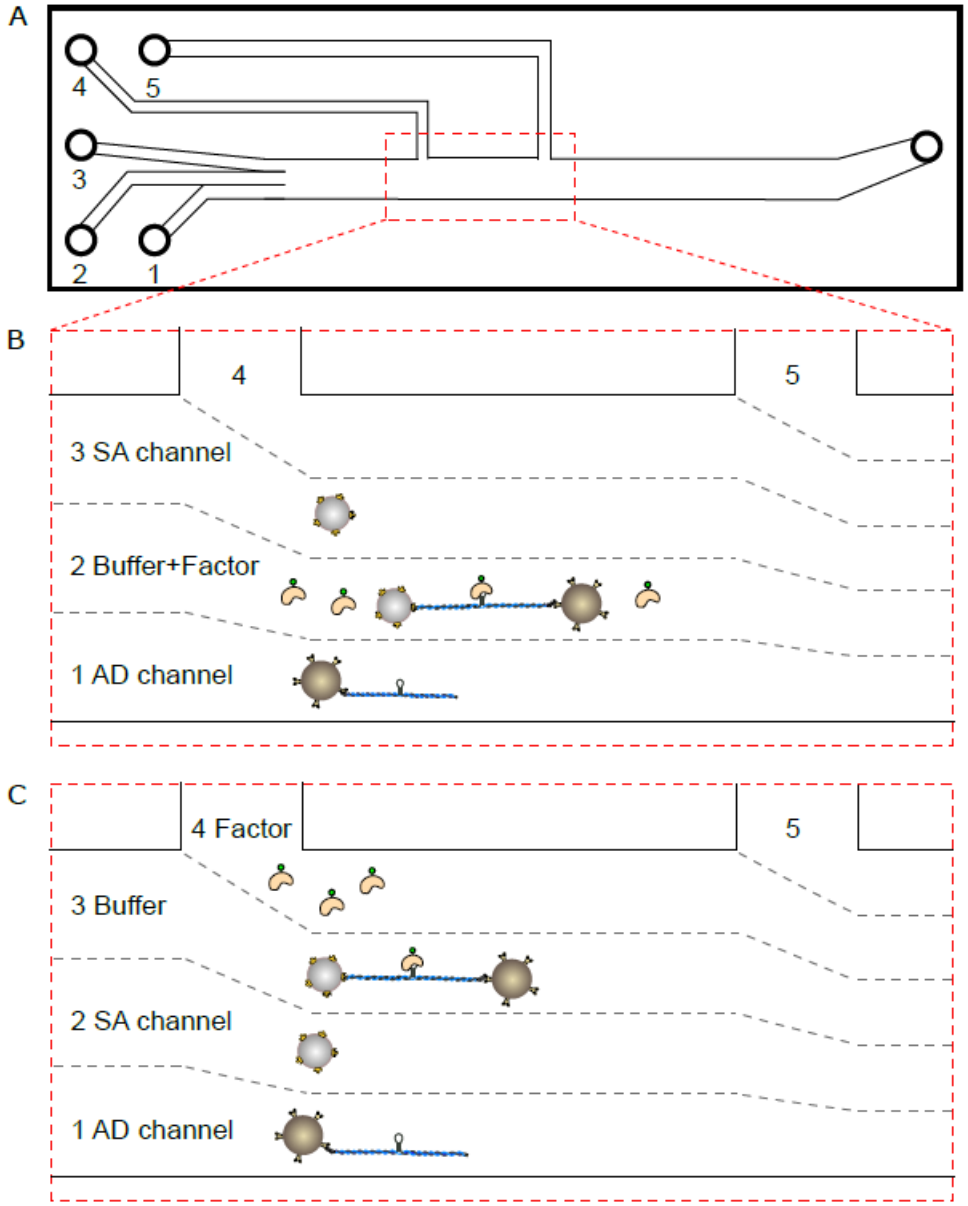

Figure 2.3 Illustration of different microfluidics channel setups. (A) A scheme of the flow cell with 5- microfluidics channels. **(B)** and **(C)** are the zoom-ins of the red-dashed area of **(A)**. **(B)** A simple 3- channel setup with AD beads and SA beads in channels 1 and 3, respectively. Factor is found in channel 2. This setup is suitable for stable proteins with high affinity, thus low concentration is preferred to ensure low fluorescent background. The bead channels on the side allow fixed tether orientation and quick recruitment of new beads if necessary. **(C)** 4-channel setup with Factor in channel 4. Such an arrangement is particularly advantageous for minimal sample consumption. The measurement can be performed directly in channel 4. Alternatively, to avoid background fluorescence signal, the complex can be formed in channel 4 and then the measurement can be performed in channel 3.

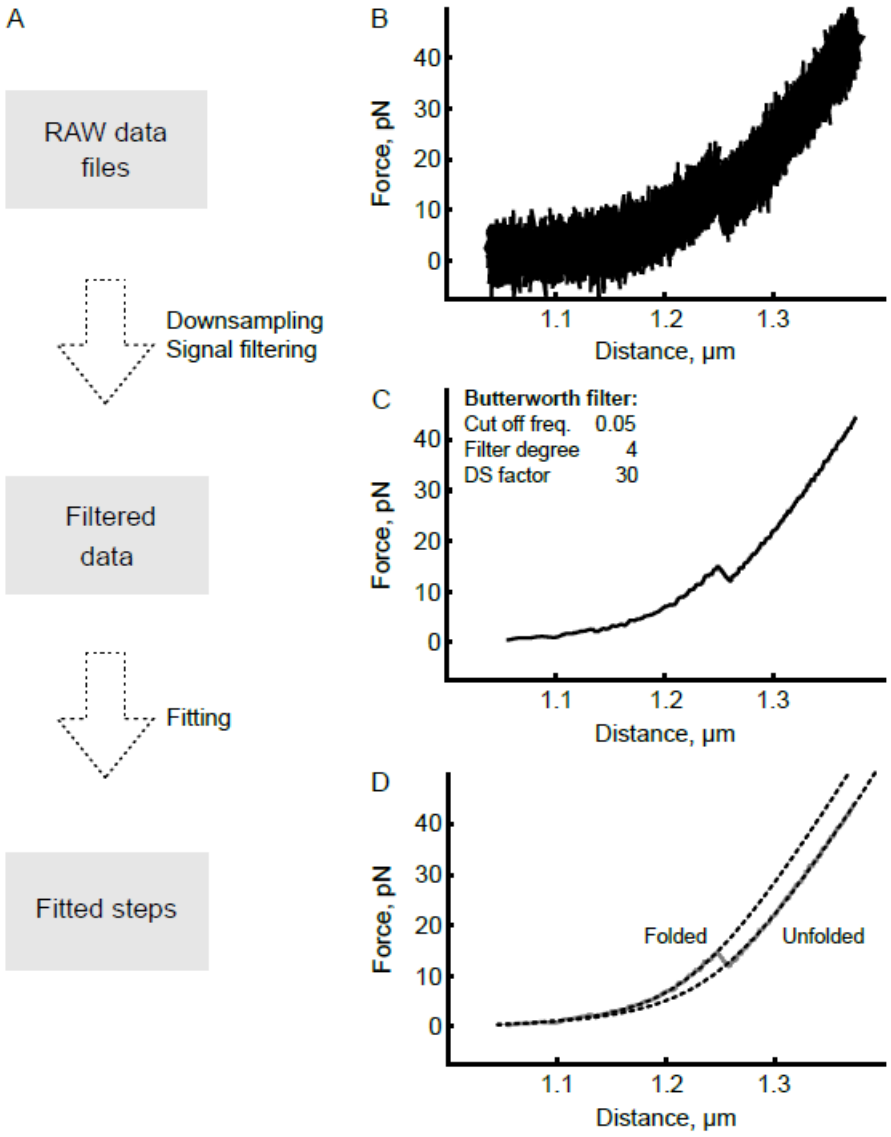

Figure 2.4 Data analysis workflow for force-ramp experiments. (A) Flowchart of the data analysis workflow. The raw data files are first downsampled and filtered, then steps are marked and the individual states are fitted to the corresponding model. **(B)** The raw data contain considerable amount of noise, which obstruct the identification of unfolding/refolding events. Also, in most of the experiments, the frequency of data gathering is higher than necessary. **(C)** Therefore, downsampling and signal filtration are employed to smoothen the data profile. **(D)** The processed curves are finally fitted to the worm-like chain (WLC) model when the molecule is still in the folded state (before the unfolding event), a combination of a WLC model with a freely-jointed chain (FJC) or a second WLC model when the molecule is in an unfolded state (after the unfolding event).

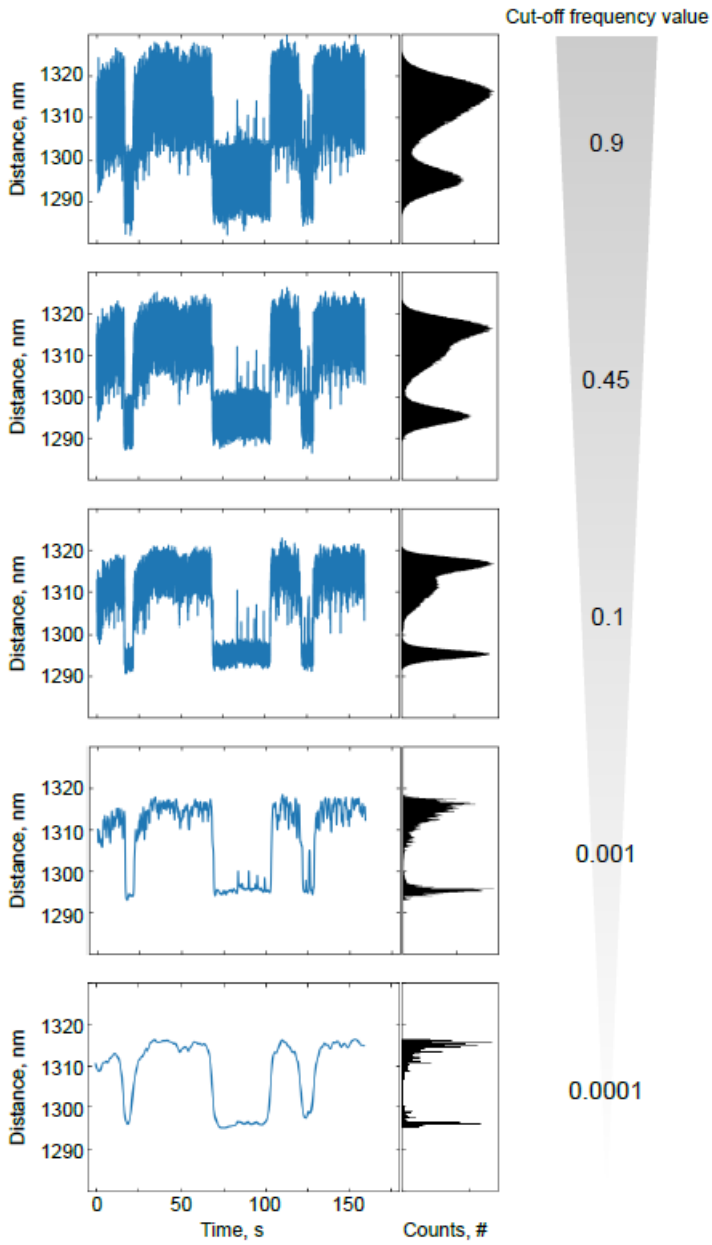

Figure 2.5 The effect of cut-off-frequency on data output. While the raw data output might be burdened with signal noise (top), it is crucial to choose proper signal filtration parameters for data analysis. Although proper filtration would help in the identification of folding intermediates (cut-off frequency 0.1, middle), over filtration (cut-off frequency <0.001, bottom) may result in loss of resolution.

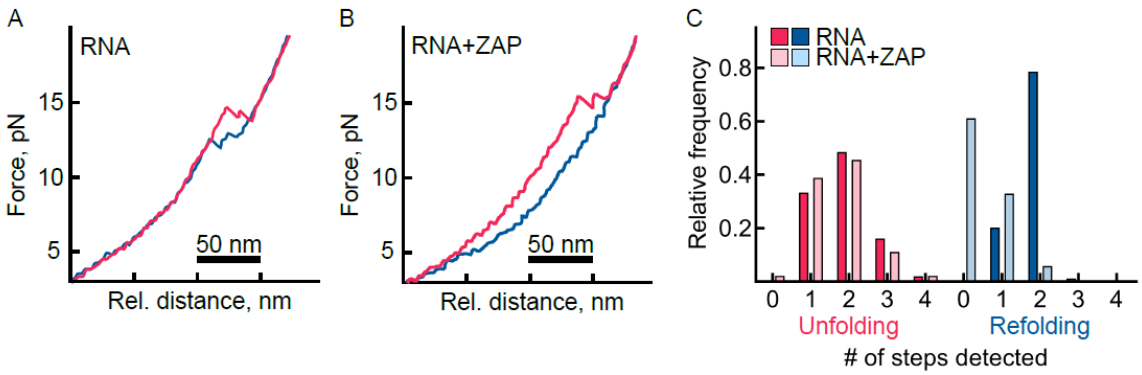

Figure 2.6 Example FD trajectories in the absence and presence of ZAP. (A) Unfolding (pink) and refolding (blue) traces of the SARS-CoV-2 RNA in the absence of ZAP. The sample shows readily refolding with only small hysteresis. **(B)** Unfolding (pink) and refolding (blue) traces of the RNA in the presence of *trans*-factor ZAP (400 nM). The sample shows huge hysteresis, suggesting that the protein binds to the single-stranded RNA and prevents its refolding. **(C)** A bar chart showing the number of unfolding (pink) and refolding (blue) steps in the absence or presence of ZAP. While the distribution of unfolding steps remains almost unaffected by the presence of ZAP, there is a clear drop in the number of refolding steps with ZAP.

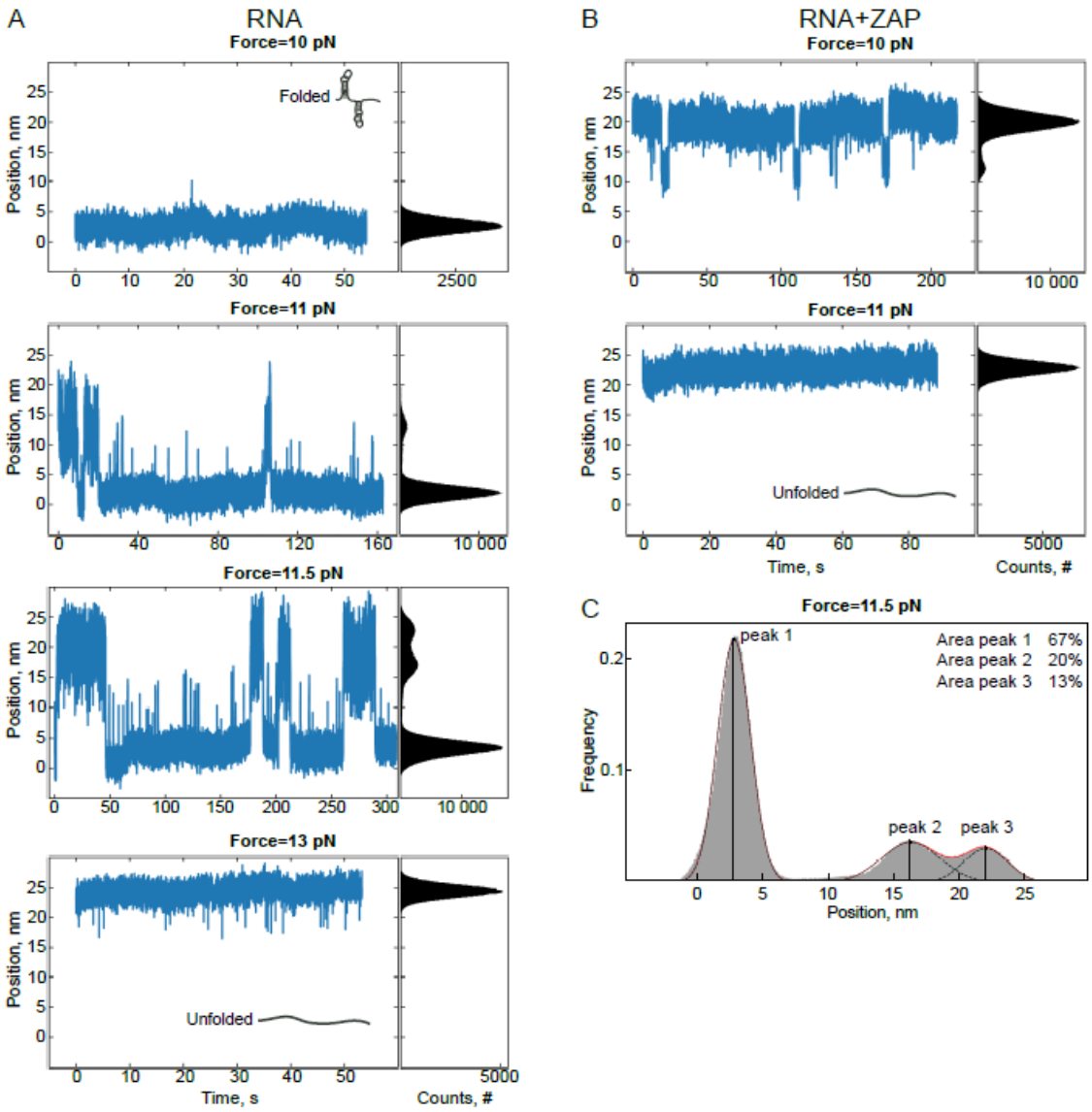

Figure 2.7 Example constant-force data in the absence and presence of ZAP. (A) Constant-force data obtained at forces ranging between 10 (up) to 13 (bottom) pN showing the shift from fully folded state to fully unfolded state of the SARS-CoV-2 frameshifting RNA element. Each graph includes the position vs. time (left) and a histogram plot (right). **(B)** Constant-force data obtained in the presence of ZAP (400 nM). Upon protein binding, the refolding is impaired. At 10 pN, in contrast to RNA alone, in the presence of ZAP RNA mostly exists the unfolded state. Therefore, a shift in the equilibrium force towards lower forces is indicated. **(C)** The histogram of position data can be analyzed by fitting the data to gaussian functions to yield the relative abundance of each state (derived from the area under the curve for each state).

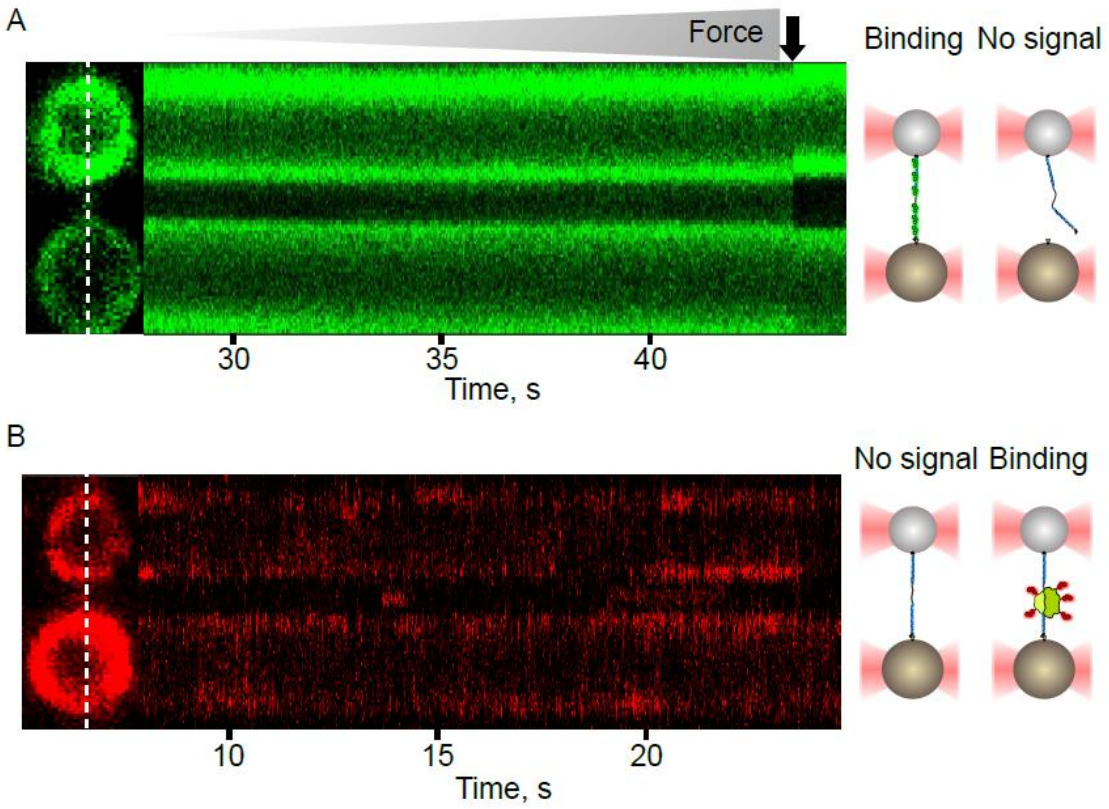

Figure 2.8 OT combined with confocal microscopy. (A) An example kymograph of the SYTOX Green labeled tether (left). Note the increase in signal intensity at increasing forces. The black arrow marks the tether breakage event, which leads to loss of signal. Depiction of the tether with dye bound to it (Binding) and after breakage without dye (No signal) (right). **(B)** Example kymograph of specific binding of the ribosome on the mRNA (left). The binding event can be observed as a fluorescence signal on the tethered between the two beads. Depiction of tether without (No signal) and with fluorescence-labeled ribosomes bound (Binding) (right).

2.9. Discussion

Here, we demonstrate the use of fluorescence-coupled optical tweezers to study interactions and dynamic behavior of RNA molecules with various ligands. Below, critical steps and limitations of the present technique are discussed.

Critical steps in the protocol

As for many other methods, the quality of the sample is pivotal to obtain reliable data. Therefore, to obtain the highest possible quality samples, it is worth it to spend time to optimize the procedure for sample preparation. The optimization steps include proper primer design, annealing temperatures, RNA and protein purification steps.

Throughout the experiment use of filtered tips and solutions is crucial in order to maintain RNase-free conditions. In addition, the microfluidics system is kept in bleach

when not in use. Before starting measurements, it is important to wash the system properly with sodium thiosulfate and RNase-free water to remove the bleach from the system.

In case the same-sized beads are used throughout the experiment, it is not required to perform force calibration each time. Nevertheless, force calibration checks should be done regularly for the reproducibility of experiments.

Modifications and troubleshooting of the method

Fluorophore stability and photobleaching

A complication during fluorescence measurements is photobleaching. Since the time frame to monitor translation can be extended from seconds to minutes depending on the system, photobleaching during the measurements should be also considered and minimized as much as possible[73]. One option is to employ more stable fluorophores, which are less prone to photobleaching, such as recently introduced quantum dots[49,74,75]. Further stability is also achieved by removing oxygen molecules using an "oxygen scavenger" system, such as glucose oxidase coupled with catalase. Glucose oxidase removes oxygen from the environment by turning it into hydrogen peroxide, which is then decomposed by catalase. Alternative oxygen scavenging systems can also be employed[76,77].

Microfluidics

Maintaining a continuous laminar flow is essential for proper measurements. Most importantly, the system should never run dry. Unfortunately, RBPs or other *trans*-acting factors of interest are often available only in small volumes for the experiments, therefore maintaining continuous flow can be challenging and cost intensive. If air bubbles are introduced into the system during the sample application, manual pressure or ethanol wash is usually sufficient for their removal.

Limitations of the method

Combination of OT with confocal microscopy also brings some limitations. First, the focal plane of the confocal unit must be aligned properly with trap centers to allow proper recording of fluorescence signal. Furthermore, for confocal measurements, handles of at least 2 kb at each site are usually needed[17]. Although in principle using longer handles is possible, one should consider the energy contribution of the handles and the change in the persistence length for the accuracy of data analysis[78]. Another crucial point is the oxygen scavengers, which are used to increase the half-life of the fluorophores, also lead to relatively quick changes in pH of the solutions[76]. These changes can be partially compensated by increasing the concentration of the buffering compound; however, during the measurements, samples should be replenished regularly (every 30-60 min) to ensure consistent conditions through the experiment.

2.10. Acknowledgments

We thank Anuja Kibe and Jun. Prof. Redmond Smyth for critically reviewing the manuscript. We thank Tatyana Koch for expert technical assistance. We thank Kristyna Pekarkova for the help with recording experimental videos. The work in our laboratory is supported by the Helmholtz Association and funding from the European Research Council (ERC) Grant Nr. 948636 (to NC).

2.11. Disclosures

The authors have nothing to disclose.

2.12. References

1 Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M., Grzybowska, E. A. RNA–protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biology.* **9** (6), 190096 (2019).

2 Armaos, A., Zacco, E., Sanchez de Groot, N., Tartaglia, G. G. RNA-protein interactions: Central players in coordination of regulatory networks. *BioEssays.* **43** (2), 2000118 (2021).

3 Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. *Journal of General Virology.* **93** (Pt 7), 1385-1409 (2012).

4 Caliskan, N., Peske, F., Rodnina, M. V. Changed in translation: mRNA recoding by −1 programmed ribosomal frameshifting. *Trends in Biochemical Sciences.* **40** (5), 265-274 (2015).

5 Jaafar, Z. A., Kieft, J. S. Viral RNA structure-based strategies to manipulate translation. *Nature Reviews Microbiology.* **17** (2), 110-123 (2019).

6 Eswarappa, S. M. et al. Programmed translational readthrough generates antiangiogenic VEGF-Ax. *Cell.* **157** (7), 1605-1618 (2014).

7 Rodnina, M. V. et al. Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Research.* **48** (3), 1056-1067 (2020).

8 Li, Y. et al. Transactivation of programmed ribosomal frameshifting by a viral protein. *Proceedings of the National Academy of Sciences.* **111** (21), E2172 (2014).

9 Napthine, S. et al. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nature Communications.* **8** (1), 15582 (2017).

10 Patel, A. et al. Molecular characterization of the RNA-protein complex directing -2/-1 programmed ribosomal frameshifting during arterivirus replicase expression. *Journal of Biological Chemistry.* **295** (52), 17904-17921 (2020).

11 Napthine, S., Bell, S., Hill, C. H., Brierley, I., Firth, A. E. Characterization of the stimulators of protein-directed ribosomal frameshifting in Theiler's murine encephalomyelitis virus. *Nucleic Acids Research.* **47** (15), 8207-8223 (2019).

12 Marshall, R. A., Aitken, C. E., Dorywalska, M., Puglisi, J. D. Translation at the Single-Molecule Level. *Annual Review of Biochemistry.* **77** (1), 177-203 (2008).

13 Rodnina, M. V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein science : A publication of the Protein Society.* **25** (8), 1390-1406 (2016).

14 Rodnina, M. V., Fischer, N., Maracci, C., Stark, H. Ribosome dynamics during decoding. *Philosophical Transactions of Royal Society of London B Biological Sciences.* **372** (1716) (2017).

15 Yan, S., Wen, J. D., Bustamante, C., Tinoco, I., Jr. Ribosome excursions during mRNA translocation mediate broad branching of frameshift pathways. *Cell.* **160** (5), 870-881 (2015).

16 Liu, T. et al. Direct measurement of the mechanical work during translocation by the ribosome. *eLife.* **3,** e03406-e03406 (2014).

17 Desai, V. P. et al. Co-temporal force and fluorescence measurements reveal a ribosomal gear shift mechanism of translation regulation by structured mRNAs. *Molecular Cell.* **75** (5), 1007-1019.e1005 (2019).

18 Choi, J., O'Loughlin, S., Atkins, J. F., Puglisi, J. D. The energy landscape of -1 ribosomal frameshifting. *Science Advances.* **6** (1), eaax6969 (2020).

19 Prabhakar, A., Puglisi, E. V. & Puglisi, J. D. Single-molecule fluorescence applied to translation. *Cold Spring Harbor Perspectives in Biology.* **11** (1), a032714, (2019).

20 Bao, C. et al. mRNA stem-loops can pause the ribosome by hindering A-site tRNA binding. *Elife.* **9**, e55799, (2020).

21 Chen, J., Tsai, A., O'Leary, S. E., Petrov, A., Puglisi, J. D. Unraveling the dynamics of ribosome translocation. *Current Opinion in Structural Biology.* **22** (6), 804-814 (2012).

22 Qu, X. et al. The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature.* **475** (7354), 118-121 (2011).
23 Zheng, Q. et al. Ultra-stable organic fluorophores for single-molecule research. *Chemical Society Reviews.* **43** (4), 1044-1056 (2014).
24 Blanchard, S. C. Single-molecule observations of ribosome function. *Current Opinion in Structural Biology.* **19** (1), 103-109 (2009).
25 Juette, M. F. et al. The bright future of single-molecule fluorescence imaging. *Current Opinion in Chemical Biology.* **20,** 103-111(2014).
26 McCauley, M. J., Williams, M. C. Mechanisms of DNA binding determined in optical tweezers experiments. *Biopolymers.* **85** (2), 154-168 (2007).
27 Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E., Chu, S. Observation of a single-beam gradient force optical trap for dielectric particles. *Optics Letters.* **11** (5), 288-290 (1986).
28 Bustamante, C., Smith, S. B., Liphardt, J. & Smith, D. Single-molecule studies of DNA mechanics. *Current Opinion in Structural Biology.* **10** (3), 279-285 (2000).
29 Choudhary, D., Mossa, A., Jadhav, M., Cecconi, C. Bio-molecular applications of recent developments in optical tweezers. *Biomolecules.* **9** (1), 23 (2019).
30 Moffitt, J. R., Chemla, Y. R., Smith, S. B., Bustamante, C. Recent advances in optical tweezers. *Annual Reviews of Biochemistry.* **77,** 205-228 (2008).
31 Li, P. T. X., Vieregg, J., Tinoco, I. How RNA Unfolds and Refolds. *Annual Review of Biochemistry.* **77** (1), 77-100 (2008).
32 Stephenson, W., Wan, G., Tenenbaum, S. A., Li, P. T. Nanomanipulation of single RNA molecules by optical tweezers. *Journal of Visualized Experiments.* (90), e51542, (2014).
33 Halma, M. T. J., Ritchie, D. B., Cappellano, T. R., Neupane, K., Woodside, M. T. Complex dynamics under tension in a high-efficiency frameshift stimulatory structure. *Proceedings of the National Academy of Sciences.* **116** (39), 19500 (2019).
34 Hansen, T. M., Reihani, S. N. S., Oddershede, L. B., Sørensen, M. A. Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proceedings of the National Academy of Sciences of the United States of America.* **104** (14), 5830-5835 (2007).
35 Zhong, Z. et al. Mechanical unfolding kinetics of the SRV-1 gag-pro mRNA pseudoknot: possible implications for -1 ribosomal frameshifting stimulation. *Science Reports.* **6**, 39549 (2016).
36 McCauley, M. J., Rouzina, I., Li, J., Núñez, M. E., Williams, M. C. Significant differences in RNA structure destabilization by HIV-1 GagDp6 and NCp7 proteins. *Viruses.* **12** (5), 484, doi:10.3390/v12050484 (2020).
37 de Messieres, M. et al. Single-molecule measurements of the CCR5 mRNA unfolding pathways. *Biophysics Journal.* **106** (1), 244-252 (2014).
38 Yang, L. et al. Single-molecule mechanical folding and unfolding of RNA hairpins: Effects of single A-U to A·C pair substitutions and single proton binding and implications for mRNA structure-induced -1 ribosomal frameshifting. *Journal of American Chemical Society.* **140** (26), 8172-8184 (2018).
39 McCauley, M. J. et al. Targeted binding of nucleocapsid protein transforms the folding landscape of HIV-1 TAR RNA. *Proceedings of the National Academy of Sciences of the United States of America.* **112** (44), 13555-13560 (2015).
40 Whitley, K. D., Comstock, M. J., Chemla, Y. R. High-resolution "Fleezers": Dual-trap optical tweezers combined with single-molecule fluorescence detection. *Methods in Molecular Biology (Clifton, N.J.).* **1486**, 183-256 (2017).
41 Yerramilli, V. S., Kim, K. H. Labeling RNAs in live cells using malachite green aptamer scaffolds as fluorescent probes. *ACS Synthetic Biology.* **7** (3), 758-766 (2018).

42 Gross, P., Farge, G., Peterman, E. J., Wuite, G. J. Combining optical tweezers, single-molecule fluorescence microscopy, and microfluidics for studies of DNA-protein interactions. *Methods in Enzymology.* **475**, 427-453 (2010).
43 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.* **12** (1), 7193 (2021).
44 Neupane, K., Yu, H., Foster, D. A. N., Wang, F., Woodside, M. T. Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic acids research.* **39** (17), 7677-7687 (2011).
45 Ritchie, D. B., Soong, J., Sikkema, W. K., Woodside, M. T. Anti-frameshifting ligand reduces the conformational plasticity of the SARS virus pseudoknot. *Journal of the American Chemical Society.* **136** (6), 2196-2199 (2014).
46 Janissen, R. et al. Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments. *Nucleic Acids Research.* **42** (18), e137-e137 (2014).
47 Smith, S. B., Cui, Y., Bustamante, C. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science.* **271** (5250), 795 (1996).
48 Puljung, M. C., Zagotta, W. N. Labeling of specific cysteines in proteins using reversible metal protection. *Biophysical Journal.* **100** (10), 2513-2521 (2011).
49 Toseland, C. P. Fluorescent labeling and modification of proteins. *Journal of Chemical Biology.* **6** (3), 85-95 (2013).
50 Hill, C. H. et al. Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch. *Nature Communications.* **12** (1), 7166 (2021).
51 Butterworth, S. On the theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer.* **7,** 536-541 (1930).
52 Wang, M. D., Yin, H., Landick, R., Gelles, J., Block, S. M. Stretching DNA with optical tweezers. *Biophysics Journal.* **72** (3), 1335-1346 (1997).
53 Mukhortava, A. et al. Structural heterogeneity of attC integron recombination sites revealed by optical tweezers. *Nucleic Acids Research.* **47** (4), 1861-1870 (2019).
54 Buck, S., Pekarek, L., Caliskan, N. POTATO: An automated pipeline for batch analysis of optical tweezers data. *bioRxiv.* 2021.2011.2011.468103 (2021).
55 Zhang, Y., Jiao, J. & Rebane, A. A. Hidden Markov modeling with detailed balance and its application to single protein folding. *Biophysical Journal.* **111** (10), 2110-2124 (2016).
56 Sgouralis, I., Pressé, S. An introduction to infinite HMMs for single-molecule data analysis. *Biophysics Journal.* **112** (10), 2021-2029 (2017).
57 Müllner, F. E., Syed, S., Selvin, P. R., Sigworth, F. J. Improved hidden Markov models for molecular motors, part 1: basic theory. *Biophysical Journal.* **99** (11), 3684-3695 (2010).
58 Elms, P. J., Chodera, J. D., Bustamante, C. J., Marqusee, S. Limitations of constant-force-feedback experiments. *Biophysical Journal.* **103** (7), 1490-1499 (2012).
59 Re, A., Joshi, T., Kulberkyte, E., Morris, Q., Workman, C. T. RNA-protein interactions: an overview. *Methods Molecular Biology.* **1097**, 491-521 (2014).
60 Jankowsky, E., Harris, M. E. Specificity and nonspecificity in RNA-protein interactions. *Nature reviews. Molecular Cell Biology.* **16** (9), 533-544 (2015).
61 Lim, F., Peabody, D. S. RNA recognition site of PP7 coat protein. *Nucleic Acids Research.* **30** (19), 4138-4144 (2002).
62 Sunbul, M., Jäschke, A. SRB-2: a promiscuous rainbow aptamer for live-cell RNA imaging. *Nucleic Acids Research.* **46** (18), e110-e110 (2018).
63 Sanchez de Groot, N. et al. RNA structure drives interaction with proteins. *Nature Communications.* **10** (1), 3246 (2019).

64 Zeffman, A., Hassard, S., Varani, G., Lever, A. The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein. *Journal of Molecular Biology.* **297** (4), 877-893 (2000).
65 Mangeol, P. et al. Probing ribosomal protein–RNA interactions with an external force. *Proceedings of the National Academy of Sciences.* **108** (45), 18272 (2011).
66 Luo, X. et al. Molecular mechanism of RNA recognition by Zinc-Finger antiviral protein. *Cell Reports.* **30** (1), 46-52.e44 (2020).
67 Qu, X., Lancaster, L., Noller, H. F., Bustamante, C., Tinoco, I., Jr. Ribosomal protein S1 unwinds double-stranded RNA in multiple steps. *Proceedings of the National Academy of Science U. S. A.* **109** (36), 14458-14463 (2012).
68 Chandra, V., Hannan, Z., Xu, H., Mandal, M. Single-molecule analysis reveals multi-state folding of a guanine riboswitch. *Nature Chemical Biology.* **13** (2), 194-201 (2017).
69 Savinov, A., Perez, C. F., Block, S. M. Single-molecule studies of riboswitch folding. *Biochimica et Biophysica Acta.* **1839** (10), 1030-1045 (2014).
70 Kelly, J. A. et al. Structural and functional conservation of the programmed ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). *Journal of Biological Chemistry.* **295** (31), 10741-10748 (2020).
71 Neupane, K. et al. Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers. *Nature Communications.* **12** (1), 4749 (2021).
72 Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods.* **9** (7), 676-682 (2012).
73 Zheng, Q., Jockusch, S., Zhou, Z., Blanchard, S. C. The contribution of reactive oxygen species to the photobleaching of organic fluorophores. *Photochemistry and Photobiology.* **90** (2), 448-454 (2014).
74 Deerinck, T. J. The application of fluorescent quantum dots to confocal, multiphoton, and electron microscopic imaging. *Toxicologic Pathology.* **36** (1), 112-116 (2008).
75 Rill, N., Mukhortava, A., Lorenz, S., Tessmer, I. Alkyltransferase-like protein clusters scan DNA rapidly over long distances and recruit NER to alkyl-DNA lesions. *Proceedings of the National Academy of Science U. S. A.* **117** (17), 9318-9328 (2020).
76 Swoboda, M. et al. Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments. *ACS Nano.* **6** (7), 6364-6369 (2012).
77 Aitken, C. E., Marshall, R. A., Puglisi, J. D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical Journal.* **94** (5), 1826-1835 (2008).
78 Wen, J.-D. et al. Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophysical journal.* **92** (9), 2996-3009 (2007).

Chapter 3

"The dumbest farmers grow the biggest POTATOes"

Old German proverb

3. POTATO: Automated pipeline for batch analysis of optical tweezers data

Stefan Buck [†1], Lukas Pekarek[†1], Neva Caliskan* [1,2]

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

[2] Medical Faculty, Julius-Maximilians University Würzburg, Würzburg, Germany

[†] Authors contributed equally to this work.

* Corresponding author

Neva Caliskan neva.caliskan@helmholtz-hiri.de

Published: Biophysical Journal volume 121, Issue 15, Pages 2830-2839 (2022)

3.1. Abstract

Optical tweezers is a single-molecule technique that allows probing of intra- and intermolecular interactions that govern complex biological processes involving molecular motors, protein–nucleic acid interactions and protein/RNA folding. Recent developments in instrumentation eased and accelerated optical tweezers data acquisition, but analysis of the data remains challenging. Here, to enable high-throughput data analysis, we developed an automated python-based analysis pipeline called POTATO (Practical Optical Tweezers Analysis TOol). POTATO automatically processes the high-frequency raw data generated by force-ramp experiments and identifies (un)folding events using predefined parameters. After segmentation of the force-distance trajectories at the identified (un)folding events, sections of the curve can be fitted independently to worm-like chain and freely-jointed chain models, and the work applied on the molecule can be calculated by numerical integration. Furthermore, the tool allows plotting of constant force data and fitting of the Gaussian distance distribution over time. All these features are wrapped in a user-friendly graphical interface (https://github.com/REMI-HIRI/POTATO), which allows researchers without programming knowledge to perform sophisticated data analysis.

3.2. Significance

Studying (un)folding of biopolymer structures with optical tweezers under different conditions generates very large datasets for statistical data analysis. Recent technical improvements accelerated data acquisition by coupling modern instruments with

microfluidic systems, at the same time creating the need for a high-throughput, and unbiased data analysis. We developed Practical Optical Tweezers Analysis TOol (POTATO); an open-source python-based tool that can process data gathered by any OT force-ramp experiment in an automated fashion. POTATO is principally designed for data preprocessing, identification of (un)folding events and the fitting of the force-distance curves. In addition, all parameters for preprocessing, statistical analysis and fitting of the curves can be adapted to suit the dataset under analysis in an easy-to-use graphical user interface.

3.3. Introduction

Arthur Ashkin received the Nobel Prize in 2018 for his research on trapping dielectric particles with laser light in optical tweezers (OT) (1). Optical tweezers enable probing of structural dynamics of individual molecules by monitoring internal forces and short-lived intermediate states in real-time (2-5). This technique has been widely used to study structures of nucleic acids and dynamics of RNA/protein folding (6-10). In addition, OT can also be used to probe the molecular interactions between small molecules, proteins, and nucleic acids (11-13). Recently, the combination of optical tweezers with confocal microscopy enabled simultaneous measurements of force and fluorescence that provided unprecedented insights into molecular mechanisms such as timing and order of events during transcription or translation (12,14-16). Basically, in a typical OT experiment, a biopolymer, such as a protein, DNA, or RNA molecule, is tethered between two dielectric beads via labeled handles. The beads are then trapped by focused laser beams, the so-called optical traps. Following this several modes of operation are possible. In force-ramp mode the beads are precisely displaced in a monotonous manner, which applies increasing forces onto the biopolymer (**Fig. 3.1A**). Since trapped beads behave as if they were attached to mechanical springs, the applied force can be calculated from the measured displacement of the beads out of the trap focus according to Hooke's law (**Fig. 3.1B**) (17). This mode is commonly used to determine the elastic properties of the molecule and/or to determine the rupture forces at which transitions in folding and unfolding occur.

On the other hand, a constant-force operation mode allows tracking the molecule of interest in real time as it transitions between different conformational states, yielding kinetic parameters of folding-unfolding of molecules or progressive movements of molecular motors (5). Accordingly, optical tweezers experiments also allow precise calculation of the work done on the system of interest (18,19). Previously, OT instruments were self-built by researchers and thus application required substantial physics and

engineering background. Furthermore, such experiments were highly time demanding and labor intensive because a large amount of data need to be collected for a quantitative analysis. Recently, commercial instruments became available on the market. Another breakthrough was the integration of OT instruments with microfluidic systems, which accelerated both experimental setup and data acquisition (14,15). Nowadays, high-frequency data acquisition allows the generation of large data sets in a relatively short time. Subsequent data analysis, however, still requires custom written scripts to perform data preprocessing, identification of (un)folding events or different folding states, mathematical modeling, and statistical analysis. There are few algorithms developed for the analysis of single-molecule force spectroscopy data, which can perform alignment and pattern recognition functions (20-23). Such algorithms are mostly tailored for atomic force spectroscopy data analysis, thus are not directly applicable for optical tweezers data (20-25). In addition, device manufacturers would provide basic solutions for the analysis of force spectroscopy data, yet processing of the data still require bioinformatics and statistics skills, therefore remain to be a major bottleneck.

Here, we present an automated python-based pipeline for the analysis of optical tweezers force-ramp and constant-force data (POTATO). Using statistical analysis of the time-derivative of force and distance data, both unfolding as well as refolding steps are deduced automatically, and values such as (un)folding force and step length are derived. These values are then directly employed for fitting of force-distance (FD) curves. Additionally, we provide a basic constant-force analysis function. In order to allow the users to modify the analysis parameters to suit their needs, we integrated an easy-to-use graphical user interface (GUI) in POTATO. Since the pipeline allows automated processing of multiple raw data files, our tool reduces the analysis time substantially and the automated analysis ensures reproducibility and eliminates inconsistencies of manual analysis (26). Next, applicability of the tool is demonstrated on an artificially generated dataset, which covers a broad range of possible parameter combinations for force-ramp data, and also on real experimental data (27,28). Finally, we also evaluated the performance of POTATO on a published dataset independently generated using a self-built optical tweezers system (29). Our results indicate that POTATO exhibits a robust performance in identifying (un)folding events with high accuracy, precision, and recall.

3.4. Materials and methods

Algorithm implementation

The algorithm is written in python 3. We designed a graphical user interface and wrapped the code into a windows stand-alone executable with *pyinstaller* to open this tool to a broader audience without a bioinformatics background. The code is freely available on GitHub (https://github.com/REMI-HIRI/POTATO) and the architecture of the python files and GUI is further explained in the Supporting Material.

Artificial data generation

Artificial force spectroscopy data were generated using a custom-written python script (Supporting Material). The fully folded part of FD curves was modeled using an equation for extensible worm-like chain (WLC) models (**Eq. 4**). The partially unfolded region was modeled using a combination of WLC and freely-jointed chain (FJC) models (**Eq. 5 and 6**). For a more detailed description, see the supplementary information.

Optical trapping system

Optical tweezers experiments were performed using a C-Trap® instrument (LUMICKS, NL). This device offers two laser traps combined with a 5-channel laminar-flow microfluidics system and a confocal microscope. Experiments were conducted as described in (27,28,30).

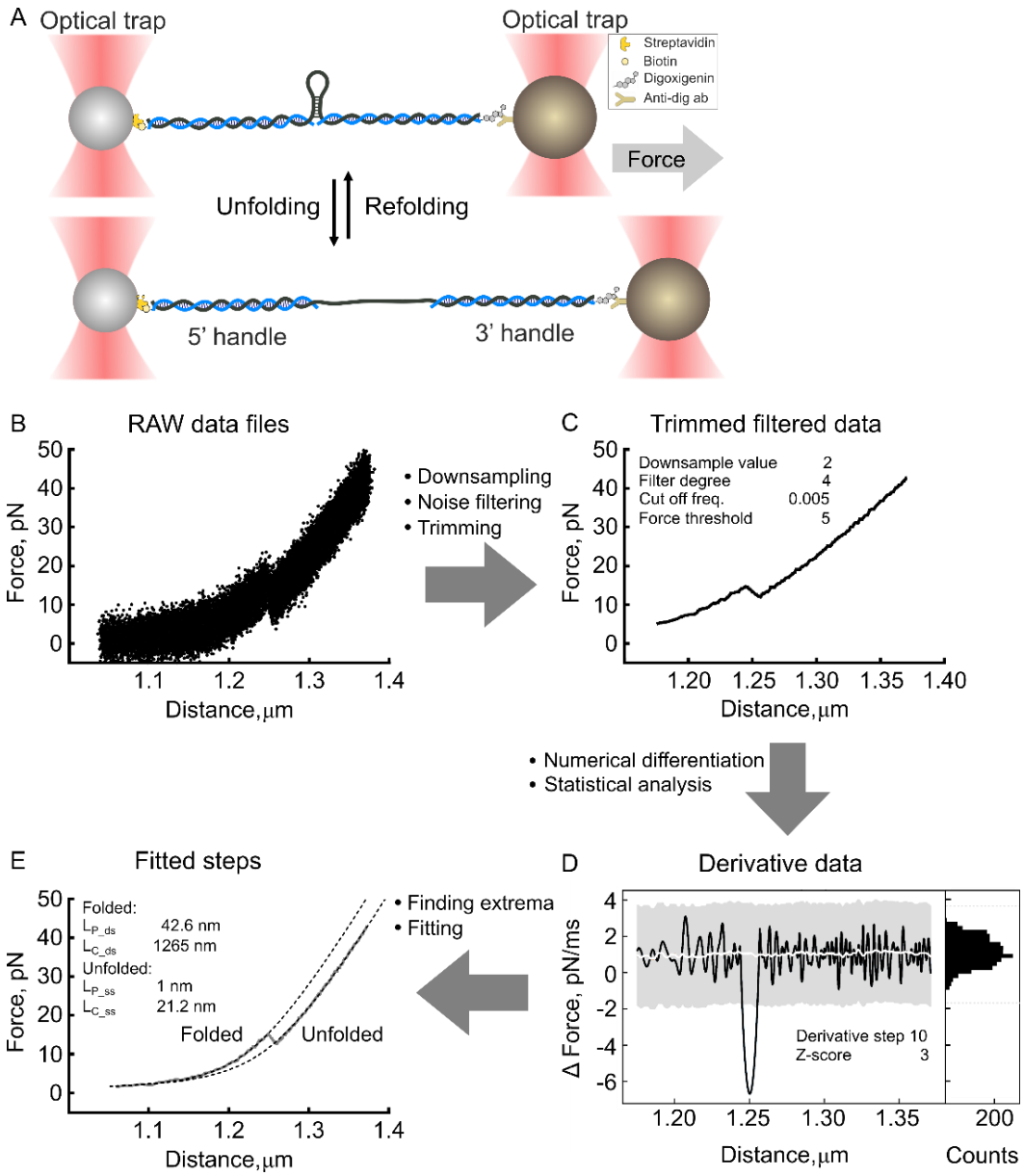

Figure 3.1 Schematic of the pipeline. **(A)** Diagram illustrating the optical tweezers experiments. RNA is hybridized to ssDNA handles and immobilized on beads. These are used to exert a pulling force on the RNA with a focused laser beam. In force-ramp operation mode, the force is gradually increased until the structure in the middle is unfolded (bottom). Release of the force allows the structure to refold (top). *RAW data files* **(B)** are downsampled, the noise is filtered using a Butterworth signal filter, and the data are trimmed at a minimum force threshold to yield the *Trimmed filtered data* **(C)**. Then the time derivative is calculated numerically to yield the *Derivative data* **(D)**; histogram of the derivative value distribution (right) shows two populations - normal-like distribution represents the experimental noise, while the other population of outliers represents the (un)folding steps. The derivative data are then statistically analyzed – the standard deviation and moving median are calculated. Peaks in derivative data that exceed median (white line) ± z-score (grey region) are classified as (un)folding events. The beginning and end of each event are derived. The coordinates of the events are then used to define the region for fitting, yielding the *Fitted steps* **(E)**.

Finally, the output data files are exported according to the selected settings. The FD curve shown here was simulated (see Supplementary Material).

3.5. Results and discussion

Data preprocessing

Raw data (**Fig. 3.1B**) from various input file formats (.h5 or .csv files containing force and distance information) can be loaded and preprocessed before marking the (un)folding events (Supporting Material). Depending on the data collection frequency downsampling can be performed, which accelerates the analysis and saves storage space. Downsampling is especially crucial when data are collected at high frequencies. The instrument we used automatically collects data in the high-frequency mode (78,000 Hz) and the raw data need to be downsampled for ease of analysis. On the other hand, self-built systems allow collecting the data at lower frequencies. In principle, if the data frequency is sufficiently high to detect the molecule while transitioning from folded to unfolded states and vice versa, POTATO can perform the analysis. Therefore, downsampling rate should be defined by the user empirically. We also note that data sets of very low data gathering frequency might not be suitable for the direct analysis by POTATO and therefore may require certain preprocessing steps (see data augmentation in supplementary material). At the next step, a low pass Butterworth filter is employed to reduce the noise out of the signal (**Eq. 1**) (31). This filter allows efficient noise removal while keeping the actual (un)folding events intact and is therefore commonly used (**Fig. 3.1C**). The algorithm then trims the data at a minimum force threshold set by the user (**Table 3.S1**). Similar to downsampling, also the noise filtering can be disabled in the GUI if the loaded data is already preprocessed.

(1) Butterworth filter:

$$G^2(\omega) = \frac{{G_0}^2}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}$$

G is gain, ω is frequency, ω_c is cut-off frequency, and *n* is filter degree.

Force-ramp data analysis

For identification of the (un)folding events, we decided to employ a derivative-based approach, as it has already been shown that this approach allows efficient step recognition (23). There are also few published algorithms based on probabilistic approaches such as the FEATHER pipeline (22). However, it must be noted that these

tools are largely developed for the analysis of atomic force microscopy (AFM) generated data (20-25). Here we aimed to combine step recognition with downstream fitting and work calculations, only based on recorded force and distance values. Furthermore, we tried to keep the pipeline intuitive and adjustable to user needs. Even though we initially employed this tool for LUMICKS FD data in H5 format, POTATO can in principle be used to analyze any dataset independent of the used instrument.

Statistical analysis

In force-ramp trajectories, an unfolding event is characterized by a simultaneous drop in force and a quick increase in distance as the secondary structure of the polymer undergoes a sudden transition from the folded to the unfolded state (**Fig. 3.1C**). Refolding events have opposite characteristics, in which the distance decreases and the force increases upon refolding. When flipped, the refolding data cannot be distinguished from the unfolding data and the processing, therefore step identification can be performed in an identical manner. Ultimately, these (un)folding events can be identified as a local maximum in the derivative of the distance and a local minimum in the derivative of the force (**Eq. 2**). When plotted, the numerical derivative data of both distance and force show two populations of values. The first is a normal-like distribution representing the measurement noise, while outliers from the normal distribution represent the second population – the actual (un)folding events. To distinguish real (un)folding events from background noise, we calculate the moving median and the standard deviation (SD). These are then used to separate the normally distributed data from the extreme values outside a given z-score (i.e. number of standard deviations = 3 by default) (**Fig. 3.1D**). This should include 99.73% of the normally distributed data points. As the initially calculated SD is affected by the outliers, a second SD is calculated from the data points inside the threshold, and the data are sorted again. The cycle is repeated until the difference between initial and secondary SD is < x (with x-default = 5%). After the force- and distance derivatives are sorted, our algorithm finds the local extrema of the derivatives, representing the saddle points of the (un)folding events in the FD curve. Then, it finds the adjacent crossing points of the derivative with the moving median, representing the start or end of the corresponding unfolding events.

(2) Numerical approximation of the derivatives:

$$\frac{dF}{dt} = \frac{F(t+dt) - F(t)}{dt} \approx \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t} = \frac{F(x + step\ d) - F(x)}{step\ d}$$

$$\frac{dD}{dt} = \frac{D(t+dt) - D(t)}{dt} \approx \lim_{\Delta t \to 0} \frac{D(t+\Delta t) - D(t)}{\Delta t} = \frac{D(x + step\ d) - D(x)}{step\ d}$$

F is force, D is distance, *t* is time, *x* is position, and step *d* is a change in position.

Data fitting

Once the respective (un)folding steps are identified, this information is employed for data fitting. Data fitting is performed on the untrimmed data to model the trajectories more precisely. For the characterization of the mechanical properties of the (bio)polymer under tension, the extensible worm-like chain (WLC) model is commonly used relating the applied force and molecular extension (**Eq. 3**) (32). For that, the FD curve is split into multiple parts. The fully folded part (until the first detectable unfolding step) is fitted with an WLC (32) to calculate the persistence length (dsL_P) of the tethered molecule, while the contour length (dsL_C) is fixed. In addition, baseline and offsets in both force and distance are included in the model to compensate for the experimental variability in the FD curves.

The partially and fully unfolded parts of the FD curves are subsequently fitted using a combined model comprising WLC (describing the folded double-stranded handles) and freely jointed chain (FJC) (**Eq. 4, 5**), or another worm-like chain (WLC) model (representing the unfolded single-stranded parts) (**Eq. 6**) **(Fig. 3.1E**) (32,33). To mathematically fit the models, we applied model polymer stretching functions from the free python package *pylake* (LUMICKS).

(3) Extensible worm-like chain model (WLC):

$$x_{WLC} = L_C \left[1 - \frac{1}{2} \left(\frac{k_B T}{(F - F_{off\ set}) \cdot L_P} \right)^{1/2} + \frac{(F - F_{off\ set})}{K_0} \right] - \mathrm{d_{offset}}$$

X is an extension, L_C is contour length, *F* is force, L_P is persistence length, k_B is Boltzmann constant, *T* is thermodynamic temperature, K_0 is stretch modulus, f_{offset} is force offset and d_{offset} is distance offset.

(4) Freely jointed chain (FJC):

$$x_{FJC} = L_C \left[\coth\left(\frac{2F \cdot L_P}{k_B T} \right) - \frac{k_B T}{2F \cdot L_P} \right] \left(1 + \frac{F}{K_0} \right)$$

(5) WLC + FJC:

$$x_{total} = x_{ds} + x_{ss} = x_{WLC} + x_{FJC}$$

(6) WLC + WLC:

$$x_{total} = x_{ds} + x_{ss} = x_{WLC1} + x_{WLC2}$$

Work calculations

Unfolding and refolding force-distance trajectories also yield crucial information on the thermodynamic properties of the molecule under study. Accordingly, the work applied by the optical tweezers instrument onto the system can be calculated from the area under

the FD curve (AUC), here using composite Simpson's rule (**Eq. 7**). First, we determine the work applied to the whole construct, including the handles (**Fig. 3.2A**). The total work on the construct is the sum of the AUC of the folded model until the starting point of the step (W_{ds}) and work performed during the step transition (W_{step}), represented by the rectangular area of the step length times force average ((F_{start} + F_{end}) / 2) (**Fig. 3.2A**). In order to extract the amount of work applied only to the structure of interest ($W_{structure}$, **Fig. 3.2C**), the work applied to the handles, represented by the AUC of the combined model (W_{ss}), is subtracted from the sum of the work on the whole construct (**Eq. 8, Fig 3.2B-C**). It shall be noted that the work derived from these calculations equals the Gibbs free energy of the studied structure provided the system is in thermodynamic equilibrium. However, if the (un)folding trajectories do not coincide, it indicates that the molecule is out of equilibrium. In non-equilibrium scenario, Gibbs free energy can be extracted from the work values (5,18,19,29,34-36) (**Fig. 3.S3**). We would like to note here that while POTATO performs the work calculations, the energy estimations have to be derived by the user separately.

(7) Numerical integration using composite Simpson's rule:

$$\int_a^b f(x)\,dx \approx \frac{h}{3}\sum_{j=1}^{n/2}\left[f\left(x_{2j-2}\right) + 4f\left(x_{2j-1}\right) + f(x_{2j})\right]$$

where $x_j = a + jh$ for j=0, 1, …, n-1 with h=(b-a)/n; $x_0 = a$ and $x_n = b$.

(8) Non-equilibrium work calculation:

$$W_{structure} = W_{ds} + W_{step} - W_{ss}$$

$W_{structure}$ is work needed to unfold the structure of interest. W_{ds} is numerical integration of the fully folded model, W_{ss} is numerical integration of the unfolded model, and W_{step} is numerical integration of the step region between the two models.

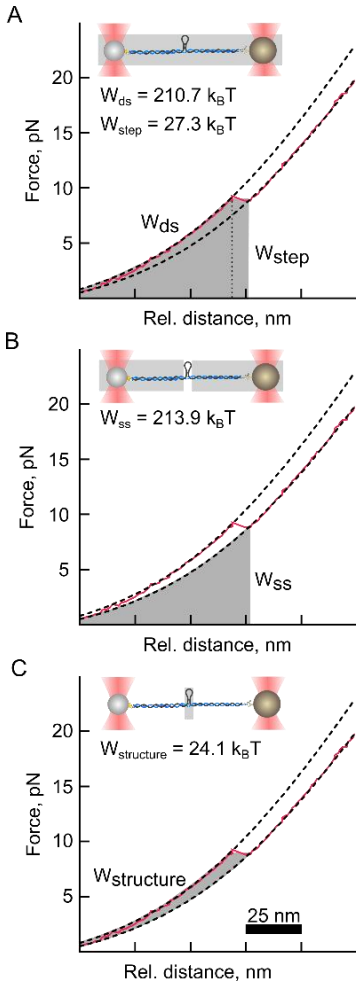

Figure 3.2 Work determination of a simple hairpin. (**A-C**) FD curve obtained during force-ramp experiment of a short stem-loop of 30 nucleotides. Inlets: the optical tweezers construct stretched between the beads with grey regions indicating to what parts of the construct the calculated work relates. (**A**) Marked region (grey) corresponding to the work necessary for stretching of the whole construct including the structure of interest. (**B**) Marked region (grey) corresponding to the work necessary for stretching of the handles and the unfolded single-stranded RNA. (**C**) Marked region (grey) corresponding to the work necessary for stretching of the RNA structure of interest. See the subsequent analysis in Supplementary **Figure 3.S3**.

Constant-force data analysis

In addition to force-ramp experiments, the algorithm we provide can also analyze constant-force data (**Fig. 3.S1** in the Supporting Material). In this way, the dynamics of the structure at a given force can be investigated. This way the equilibrium force at which the chance of the structure to be folded or unfolded are equal can be derived.

The constant-force analysis accepts the same input formats as the force-ramp batch analysis, and data preprocessing is performed similarly by downsampling and filtering of the data without trimming. First, it is necessary to display the constant-force data in order to optimize the preprocessing parameters and the plot's axis (**Fig. 3.S1B)**. At this step, two plots are generated for visualization. In the first plot, distance is plotted against time. Here, the difference in distance corresponds to the change in the contour length of the tethered molecule. The second plot is a histogram of the distance distribution (**Fig. 3.S1C)**. From this histogram, the number of different folding states can be deduced. Afterward, the histogram is fitted with multiple Gaussian functions. According to the position distribution histograms, the user can interactively provide initial estimates for

various parameters including the number, localization, width (standard deviation, z-score), and amplitude of the fits. After the optimization, the model parameters are exported together with the percentage of each folding state as a table in csv format (comma separated values).

Artificial data sets to test the limits of detection

To test the limits of (un)folding events detectable by the POTATO pipeline, an artificial dataset was generated (Supporting Material). In this data set, some curves can show a negative step-length that would not be observed in real unfolding events. We considered these steps as non-identifiable and used them as negative controls. The phenomenon of negative steps can mainly be observed for small contour length changes (ΔL_C) between the models, combined with high force drop (ΔF) values. To test the performance of the algorithm, we defined identifiable steps as events with a drop in force and a simultaneous increase in distance (Supporting Material). To evaluate if a specific parameter combination results in an identifiable curve, **Eq. 9** with x = 0 was solved for all sets of parameters. Each time two parameters were fixed, and the third parameter was optimized.

(9) Minimal step calculation:

$$x = \mathrm{WLC}_{ss}(\mathrm{step}_{\mathrm{end}}) + \mathrm{WLC}_{ds}(\mathrm{step}_{\mathrm{end}}) - \mathrm{WLC}_{ds}(\mathrm{step}_{\mathrm{start}})$$

Where WLC corresponds to expression from **Eq. 3**, "ss" refers to the model corresponding to single-strand values, while "ds" describes the double-stranded region.

A hyperplane showing the interface of theoretically identifiable and non-identifiable steps was generated from these optimized values (**Fig. 3.3A**). This allowed us to classify the generated dataset based on a combination of parameters: One with curves where POTATO is expected to find an unfolding step (x > 0) and the other one where POTATO should not identify the steps (x ≤ 0). After analyzing the artificial dataset (comprising 2520 curves) with different z-scores, the expected results, based on the input parameters when the data were generated, were compared to the steps identified by POTATO. For the default z-score of 3, the expected parameters were then plotted into the 3D plot and colored based on the identification by POTATO (**Fig. 3.3A**). For an unfolding force of 25 pN, the ΔF and ΔL_C values are shown in a 2D plot, making it easier to identify and compare single unfolding events analyzed with different z-scores. It can be seen that all identified steps at this specific unfolding force are above the theoretical threshold and that more unfolding events are identified at z-score 2.5 than at z-score 3 (**Fig. 3.3B**). Accordingly, the effect of the z-score on the derivative of force (**Fig. 3.3C)** and distance (**Fig. 3.3D**) can

be investigated for an individual force-distance trajectory. In the representative trajectory, the local maximum in the derivatives of distance is above the z-score threshold for both cases. In the derivative of force, the local minimum at the same position is only detected for the lower z-score **(Fig. 3.3C-D)**.

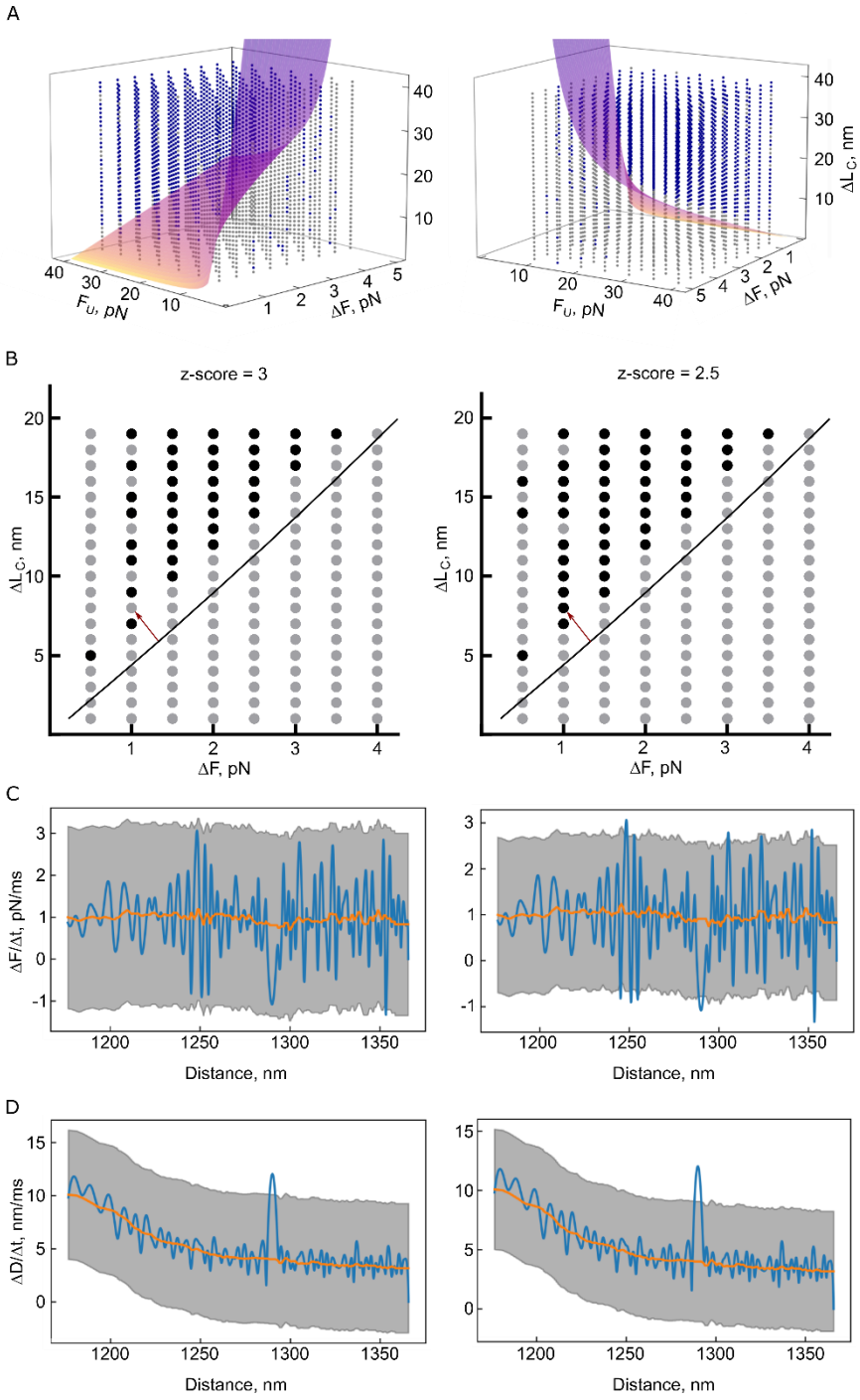

Figure 3.3 Testing the limits of POTATO. For each combination of the parameters unfolding force (F_U), force drop (ΔF), and contour length change (L_C), two parameters were fixed, and the third one was optimized so that the **Eq. 9** (Supporting Material) evaluates to zero. **(A)** A hyperplane was generated from the optimized values that separate the resolvable space above the hyperplane (parameter combinations that result in identifiable steps) from the unresolvable space below the

hyperplane (parameter combinations that result in unidentifiable steps). Each analyzed curve is plotted in blue if its step was identified by POTATO or in grey if it was not recognized. **(B)** Slices of the 3D plot at F_U = 25 pN were analyzed with different z-scores. The black line corresponds to the theoretical limit of resolvable/unresolvable parameter combinations. The black dots represent curves with identified steps, whereas the grey dots represent curves where POTATO could not identify the step. The derivatives of force **(C)** and distance **(D)** of the curve that is marked with a red arrow in **(B)** are displayed at different z-scores.

Next, we calculated performance measures such as accuracy, precision, sensitivity, specificity, and F1-score to validate the performance of POTATO. For a z-score of 3.2, a precision score of 0.974 indicates that most of the positive classified steps were actual steps, and even for a z-score of 2.5, the precision was still above 0.944 (**Table 3.S2**). As expected, higher precision comes with the trade-off to miss certain positive events (recall 0.870 - 0.939), and the optimal z-score has to be chosen depending on the application. For smaller unfolding events that are difficult to detect, lower z-scores should be employed, as for distinct unfolding events the z-score can be set to higher values. This way number of false-positive events detected can be minimized. Since the present dataset was generated using artificial parameter combinations, those might not be found in actual OT measurements. Therefore, it is important to keep in mind that we were exploring the limits of the tool by using these strict parameter constraints. Performance measures would also vary depending on where a specific dataset is located in the parameter space, and which z-scores were employed.

Furthermore, we investigated how accurately POTATO estimates step parameters (F_U, ΔL_C, ΔF). For that, we compared the expected and measured values of these parameters for all curves analyzed (**Fig. 3.4**). We then calculated the linear regression of the true positive values to estimate possible biases of POTATO estimated F_U and ΔL_C values. Our analysis shows that in the case of F_U (**Fig. 3.4A**), the values determined by POTATO are in perfect agreement with the expected values (slope of the linear regression = 0.9912). For ΔL_C (**Fig. 3.4B**), the comparison shows a broader distribution of the measured values with an overall trend suggesting a minor overestimation (slope of the linear regression = 1.0282) of around 3%. Lastly, in the case of ΔF (**Fig. 3.4C**), the trend shows a slight underestimation of the measured values (slope of the linear regression = 0.8517), resulting in a bias of 12-15%. Taken together, our performance measures analysis suggests that the presented tool successfully identifies most (un)folding events correctly with only few false classifications (false positives/false negatives). Accordingly, in most of the cases, performance measures were above 0.9 (**Table 3.S2**). Moreover, we show that POTATO can precisely estimate the parameter values describing the (un)folding

events (F_U, ΔL_C, ΔF, **Fig. 3.4**). Overall, the performance measures and the accuracy of the estimates show that POTATO represents a reliable tool for optical tweezer data analysis.

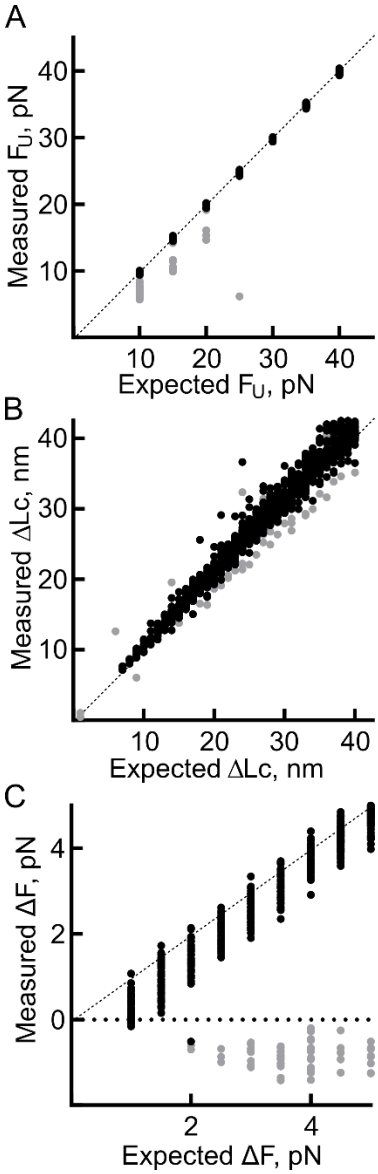

Figure 3.4 Evaluation of the performance of POTATO. The parameters used for the generation of the dataset compared to the parameters identified by POTATO are plotted against each other. All three parameters used for the data generation are evaluated with a z-score of 3. The values of the true positive steps (black) and the values of the false-positive steps (grey) are visualized for **(A)** the unfolding Force (F_U), **(B)** the contour length change (ΔL_C), and **(C)** the force drop (ΔF). A dashed line represents the theoretical perfect correlation between measured and expected value.

Applicability of POTATO on real experimental data

Next, we employed POTATO to test its performance on real experimental data generated from force-distance measurements of the programmed ribosomal frameshifting (PRF) element of the Encephalomyocarditis virus (EMCV) and SARS-CoV-2 (27,28). We compared the POTATO results with manually annotated steps of a subset of our dataset. The results obtained with manual step identification and data fitting were in good agreement with the automated analysis using the pipeline (**Fig. 3.S2A**). Harnessing POTATO in the data processing allowed us to speed up the analysis significantly compared to previous manual analysis. Furthermore, we saw that POTATO is not only

suitable for curves with a single (un)folding event like in the artificial dataset, but we successfully fit force-distance curves with as many as five unfolding steps and we were able to identify even short-lived intermediate states of the unfolding process (**Fig. 3.S2B** and **C**). In addition to the contour length change obtained by curve fitting, also the Gibb's free energy is an important variable to conclude on the nature of the (un)folded structure as the Gibb's free energy is dependent on the base pairing of the RNA. We were able to use the work calculated by the POTATO, to estimate the Gibb's free energy of the structures and thereby distinguish between different secondary structures (27). Here to demonstrate the energy calculation, we used a stem-loop mRNA of 30 nucleotides in length (**Fig. 3.S3**) (28). First, we use mfold (37) to predict the secondary structure and its Gibb's free energy (**Fig. 3.S3A**). Then, we plot the unfolding as well as refolding work distributions calculated by POTATO (**Fig. 3.S3B**). We then use the results of POTATO analysis to estimate the Gibb's free energies by applying (**i**) Crooks fluctuation theorem, and (**ii**) Jarzynski equality with bias correction (**Fig. 3.S3C**) as described in (18,34-36).

To evaluate the performance of POTATO on other published datasets generated using a self-built optical tweezers instrument we analyzed the SARS-CoV-2 pseudoknot RNA force-distance data by Neupane et al (29). Since the dataset provided had a lower data frequency resulting in less than 250 datapoints per FD curve, we first had to artificially augment the datapoints (see Supplementary Material). Despite that, we could still successfully assign the steps and reproduce the unfolding force distribution (**Fig 3.S2**) as well as the contour length estimate (**Table 3.S3**). We were also able to detect the refolding steps force distribution and detected steps as low as 6 pN (**Fig. 3.S2**). In conclusion, regardless of the system used, we demonstrate that the pipeline output matched well with manual data analysis on real-experiment datasets and that POTATO performed analysis of FD trajectories with multiple steps or even short-live intermediates in a reliable way. Therefore, POTATO represents a versatile tool for high-throughput OT data analysis for many upcoming studies.

Limitations

Processing automation comes with trade-offs (38,39). First, the statistical analysis applied in the pipeline might be prone to false-positive event discoveries due to external causes, such as vibration that might induce step-like events in the force-distance profile of gathered data. We split the force-distance data and analyze the derivatives of force and distance separately to minimize this effect. Only the events found by both approaches are considered real (un)folding events. Therefore, the robustness of the analysis is increased.

Second, the pipeline output strongly depends on parameters and threshold values that are applied throughout the analysis. The default values were set empirically to suit our needs. Therefore, it might require optimization to fit specific needs and reach an analysis output consistent with the manual data analysis. User input is required despite the user-friendly GUI environment, and an understanding of the analysis workflow is necessary to adjust the parameters rationally.

The current algorithm does not annotate the repeated folding and unfolding of a structure during force-ramp measurements and identifies this oscillation as independent steps. Nevertheless, this mainly occurs at slow loading rates and does not affect the contour length estimates. To overcome any unexpected issues with the automated analysis, POTATO also includes a tab that allows full manual analysis of the force-ram data files. This should help to eliminate bias caused by omission of certain files from the analysis during the automated analysis.

3.6. Summary

Here we present a publicly available pipeline for batch analysis of optical tweezers data. Our pipeline allows OT raw or preprocessed data processing from force-ramp or equilibrium measurements (constant force/position). These are widely employed experimental approaches in the OT field, applied to nucleic acid structure probing, protein folding, RNA-protein interactions, or even to analyze events as complex as translation. Here, by wrapping our algorithm in a standalone application and designing an intuitive graphical user interface, we aim to open the data analysis to a broader audience without the need for a bioinformatics background. The user can adjust all parameters directly in the GUI without diving into the code to tailor the pipeline to their exact needs. With the parameters optimized for the here presented datasets, POTATO showed high precision and accuracy in the identification of (un)folding events. Moreover, compared to manual data analysis, the pipeline is faster and, most importantly, consistent throughout the analysis, thus yielding reproducible results.

3.7. Supporting material

Supporting Material can be accessed in the GitHub repository (https://github.com/REMI-HIRI/POTATO)

3.8. Author contributions

NC, LP, and SB designed the pipeline. LP and SB wrote the python scripts. LP generated the artificial data. SB analyzed the artificial data. LP and SB performed the optical tweezers experiments. LP analyzed experimental data. LP and SB prepared the figures with input from NC. NC, LP, and SB wrote the manuscript.

3.9. Acknowledgments

We thank Vojtech Vrba for helpful python discussions. We thank Dr. Anke Sparmann for critically reviewing the manuscript. The work in our laboratory is supported by the Helmholtz Association and grants from the European Research Council (ERC) Grant Nr. 948636.

3.10. Disclosures

The authors have nothing to disclose.

3.11. Supplementary information

Script structure

The script is written in Python 3 and split into multiple parts for clarity. The first part, "POTATO_GUI", defines the GUI with all necessary functions and input variables. When the GUI is started, the default values of the input variables are loaded from the "POTATO_config" file. The GUI was created and structured using the standard Tkinter python package. A parallel subprocess initiates from this main process when a folder is selected for force ramp analysis to perform computationally demanding data-processing. This way the GUI remains responsive during computation. All the functions used for data preprocessing and step recognition are defined in the "POTATO_preprocessing" and the "POTATO_find_steps" files respectively. The functions used for curve fitting are defined in another file, "POTATO_fitting". For computation, we mainly use matplotlib and NumPy packages, as well as the lumicks.pylake package for fitting (**Table 3.S4**). The subprocess is a daemon process spawned by the main process and therefore stops as soon as the GUI terminates the main process. The last part, "POTATO_constantF", is executed by the main thread as it only analyzes one constant force file at a time. The results are exported in different CSV files or as PNG images.

Graphical user interface

We designed a graphical user interface (GUI) that allows users to easily adjust the analysis steps and parameters according to their needs and select between three different input data formats. This enables the GUI to load data from every OT instrument. The GUI

is separated into multiple tabs, resulting in easy and intuitive navigation without overloading the individual windows. The "POTATO_config" file, included in the POTATO repository, contains the default parameters, which are loaded into the GUI. The most commonly changed parameters can be found in the first tab, “Folder Analysis”, so a basic analysis can be performed right away (press enter to confirm changed parameters). Alternatively, before each analysis, all parameters can be adjusted in the 'Advanced Settings' tab to suit the data set. In addition, we implemented the possibility to selectively export results. Each analysis creates a new folder with a timestamp directly in the analyzed directory. The used parameters are exported as well so that parameters can be optimized later. The second tab, “Show Single File”, provides a control mechanism for data preprocessing. A single file can be loaded, and the unfiltered data are plotted together with the filtered data, which streamlines troubleshooting. Finally, there is a third tab for “Constant Force Analysis”.

Input data

The presented pipeline accepts three different input data formats. Two of them are based on the default hdf5 output format of Lumicks C-Trap – one is predefined for high-frequency data (using the piezo-tracking function of the instrument), and the second is for low-frequency data (using video recognition). The third data format is a basic CSV file format with force and distance values in the first and second columns. Force data need to be in [pN], whereas the unit of distance data can be specified either as [µm] or [nm] in the GUI. Thus, our pipeline can process force-distance data from virtually any optical tweezers machine. In addition, entire directories containing force-ramp data files can be selected and processed simultaneously.

Data output

Depending on individual analysis requirements, different export settings can be selected. The down sampled and filtered data are exported in CSV format (smooth) for each file by default. The identified (un)folding steps by derivatives of force and distance are exported together with the steps identified by both strategies (common steps) into a single CSV file. All identified steps of all curves in the analyzed folder are gathered into a single results file for quantitative analysis. The respective summary figure containing the plot of preprocessed data, trimmed data, and both derivatives with marked steps is exported. The plots of fitted data, together with the fitting parameters and model data, are exported as PNG and CSV files, respectively.

Artificial data generation

To test the limits of the algorithm, artificial data with a single step per curve were generated. The fully folded part of force-distance curves was modeled using an equation for extensible WLC models (**Eq. 4**). The partially unfolded region was modeled using a combination of WLC and FJC models (**Eq. 5 and 6**). The force value at which the step occurs, the contour length change between the unfolded and folded region, and the drop in force during the step, are the parameters for data generation. The first parameter was set to occur between 10-40 pN with a 5 pN resolution. The curves were generated with a contour length change from 1-40 nm with a 1 nm resolution and a force drop of 1-5 pN with a 0.5 pN resolution. To mimic the (Gaussian) noise affecting the raw data, we employed the NumPy random normal distribution function (1).

Since the (un)folding step is generally defined as a drop in force (one of the parameters) and a sudden increase in distance (not a parameter), the data generated by this script also contained combinations that did not increase distance. We used these curves showing no increase in the distance as negative controls.

Augmentation of low-frequency data

During analysis of the freely available data from Neupane and Zhao et al., 2021, we had to employ the data augmentation approach to increase the precision of the analysis. For the best output, ideally raw data should be directly curated in POTATO and at least >2000 data points are available. The augmentation was performed as follows. For each two consecutive data points in the original data, we divided the linear space between them by factor of 100 to get positions for new data points. Starting from the first original data point, we consecutively added 99 new data values always increasing by the previously calculated increment +/- randomly assigned noise in force and distance dimensions using random gauss function with the parameters mu=0 and sigma=0.5. The newly created files where then analyzed as csv files by POTATO.

Manual data analysis

POTATO GUI also contains a tab that provides the user with the option to manually mark steps, fit models and calculate the work for FD curves. Manual analysis is particularly useful to evaluate the precision of the automated analysis and perform parameter optimization. Furthermore, manual analysis is convenient for analysis of the FD curves that were not analyzed properly by the batch analysis. To speed up the manual analysis we implemented several keyboard shortcuts. Briefly, at the initial step one needs to mark

the region corresponding to the folded state of the molecule and fit it with extensible WLC model. At the next step, (partially) unfolded region is marked and fit of the combined model is performed (WLC+WLC). Afterward, start and end of the unfolding step is marked and the W_{step} is calculated. At the final step, calculated values of W_{ds}, W_{ss} and W_{step} are used to determine the value of $W_{structure}$ (**Eq. 8**). For a detailed description, we suggest the reader to refer to the readme file on Github (https://github.com/REMI-HIRI/POTATO).

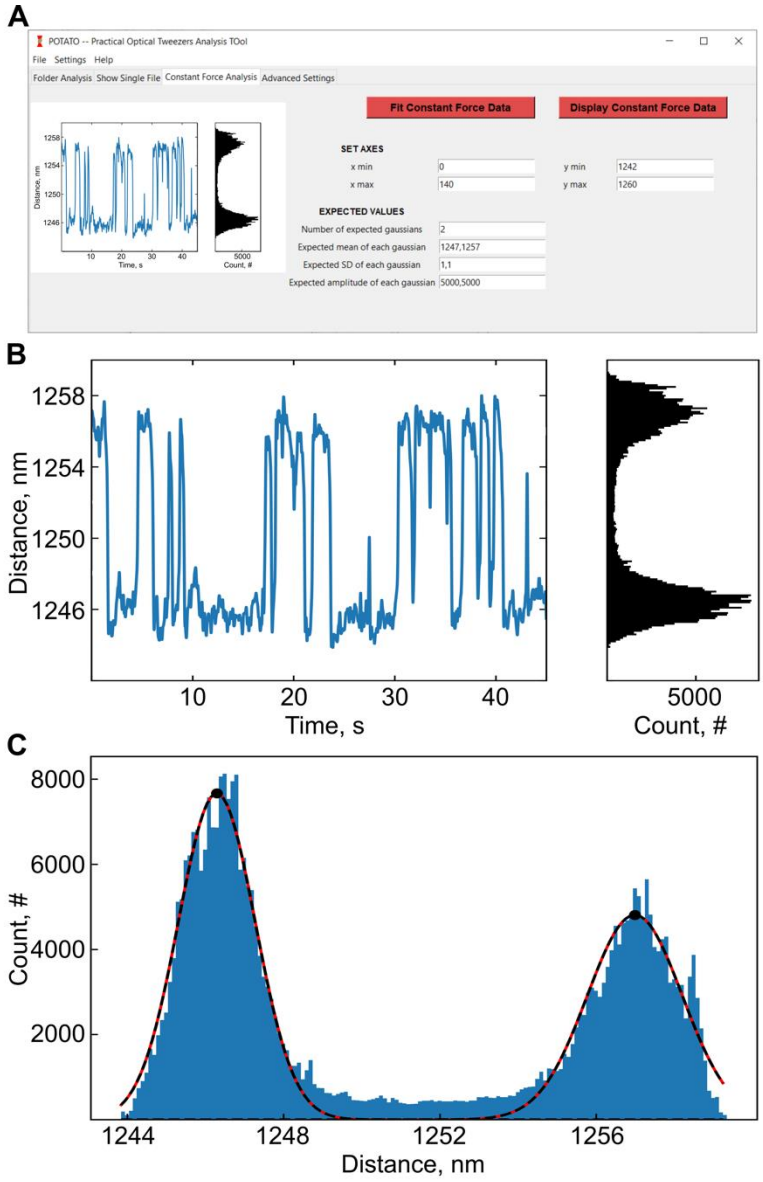

Figure 3.S1 Constant force data analysis in POTATO. (A) GUI tab containing the constant force analysis features, **(B)** Display constant force data output; (left) distance over time plot, (right) histogram of the distance over time values. **(C)** Fit constant force data output showing the histogram of distance values distributions and the two gaussian functions fitted.

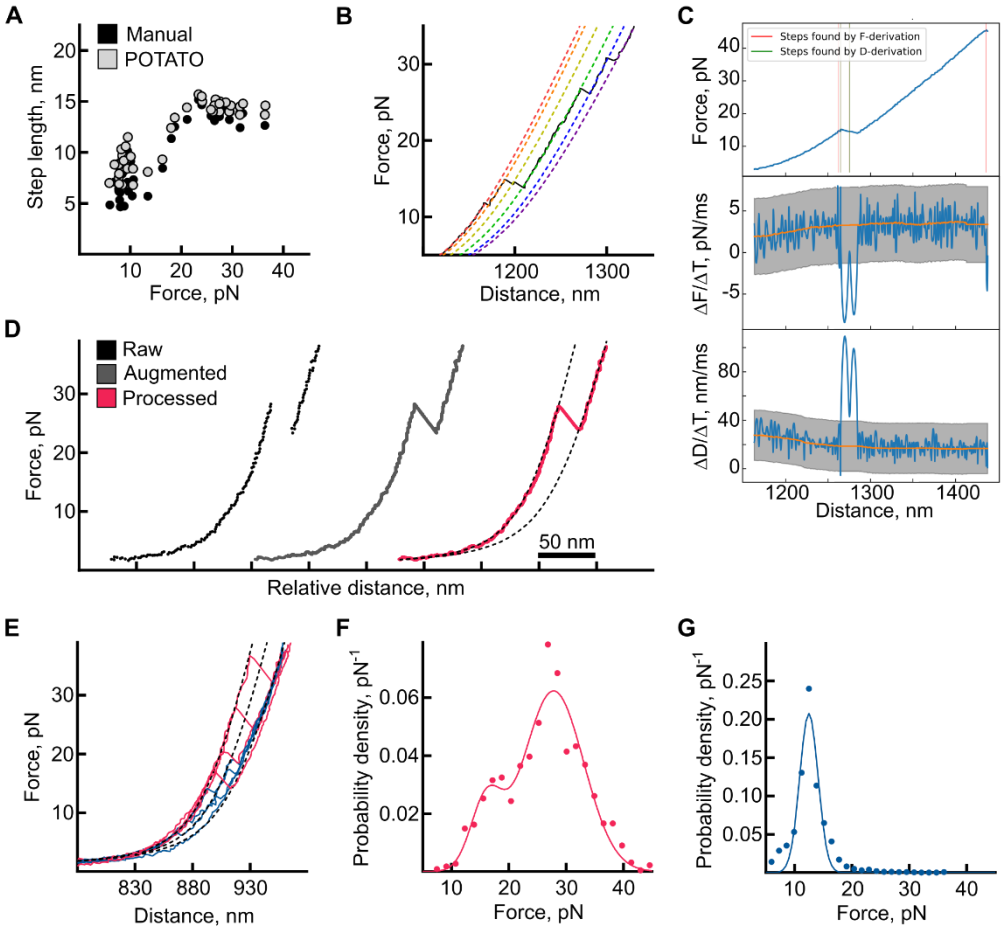

Figure 3.S2 Real data application of POTATO. **(A)** Comparison of unfolding events marked in a subset of the data analysed manually (black) or with POTATO (grey). (**B**) Example FD curve (black, solid) with five unfolding steps fitted by POTATO (colored, dashed). **(C)** Example analysis output from POTATO showing the trimmed FD curve (up), force derivative data (middle), and distance derivative data (bottom). An intermediate conformer is detected by POTATO during the unfolding. Other FD curves confirmed the presence of an even more stable and distinct intermediate step. (**D-G**) Experimental data published in Neupane and Zhao et al., 2021 (subset with 6nt spacer) reanalyzed using POTATO; (**D**) Comparison of different data types and the effect on FD trajectory – raw data as accessible online in Neupane and Zhao et al., 2021 (black), As there were not enough data values, data augmentation was performed (see supplementary methods, grey), and POTATO processed data (pink). **(E)** Four Example unfolding (pink) and refolding (blue) FD curves. (**F**) Unfolding force distribution for unfolding curves showing single unfolding step (N=1378) shows two peaks similarly as in the original analysis. (**G**) Refolding force distribution for all the refolding curves (N=1861) shows a single peak around 12 pN with refolding steps detected at forces as low as 6 pN.

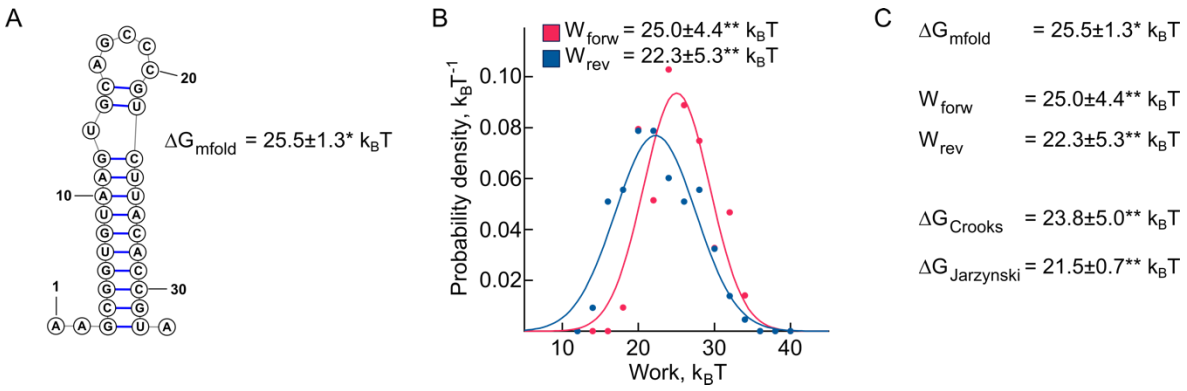

Figure 3.S3 Extracting energy information from the experimental data. (A) Mfold predicted secondary structure of a simple hairpin of 30 nucleotides in lenght. (B) Distributions of measured work values for the unfolding (red) and refolding (blue) FD curves. (C) Energy and work values as predicted by Mfold (ΔG_{mfold}), measured (W_{forw} and W_{rev}) or calculated using Crooks Theorem (ΔG_{Crooks}) and Jarzynski equality ($\Delta G_{Jarzynski}$). * 5% standard error, **standard deviation.

Table 3.S1 Parameters used throughout the pipeline and a short description.

Parameter	Description
Preprocessing	
Downsampling rate	Only every n^{th} value is taken for analysis, speeds up subsequent processing.
Butterworth filter degree	Defines the stringency of the filter.
Cut-off frequency	Signals with a frequency above this threshold are suppressed.
Force threshold, pN	Values lower than the threshold are excluded from the analysis.
Derivative	
Step d	Characterizes the interval between two values used for numerical derivative calculation.
Data frequency, Hz	The frequency at which data is recorded.
Statistics	
z-score	The number of standard deviation used to determine whether a given value is part of a normal distribution.
Moving median window size	The number of values considered for each median calculation.
SD difference threshold	Statistical analysis and data sorting are iterated until the difference between two consecutive SDs is below this value.
Fitting	
dsLp, nm	Persistence length of the double-stranded (folded) part of the tethered construct.
dsLc, nm	Contour length of double-stranded (folded) part of the tethered construct.
dsK0, pN	Stretch modulus of double-stranded (folded) part of the tethered construct.
Force offset, pN	Force offset of a given dataset; compensates for a shift in the dataset.
Distance offset, nm	Distance offset of a given dataset; compensates for a shift in the dataset.
ssLp, nm	Persistence length of the single-stranded (unfolded) part of the tethered construct.
ssLc, nm	Contour length of single-stranded (unfolded) part of the tethered construct.
ssK0, pN	Stretch modulus of single-stranded (unfolded) part of the tethered construct.

Table 3.S2 Dependence of the performance measures on the z-score. Analysis of 2520 simulated data curves with steps occurring between 10-40 pN with different z-score values.

z-score / Parameter	3.2	3	2.7	2.5
True positives	1206	1267	1280	1303
True negatives	1101	1076	1073	1056
False positives	32	57	60	77
False negatives	181	120	107	84
Accuracy	0.915	0.930	0.934	0.936
Precision	0.974	0.957	0.955	0.944
Recall	0.870	0.913	0.923	0.939
Specificity	0.972	0.950	0.947	0.932
F1-Score	0.919	0.935	0.939	0.942

Table 3.S3 Application of POTATO on real experimental data.

	Expected ΔLc, nm	Observed ΔLc, nm	Observed ΔLc, nm (Neupane et al. 2021)
Simple hairpin (30 nt)	17.7	16.4±2.8	-
SARS-CoV-2 frameshift pseudoknot (6 nt spacer)	34.7-36.3	34.8±2.0	35.6±0.4

Table 3.S4 Python packages used in POTATO. Standard packages are not included in the table.

Package name	link
h5py	https://www.h5py.org (2)
Pandas	https://pandas.pydata.org (3)
Scipy	https://www.scipy.org (4)
Matplotlib	https://matplotlib.org (5)
Lumicks.pylake	https://lumicks-pylake.readthedocs.io

Supporting references

1. Harris, C. R., K. J. Millman, . . . T. E. Oliphant. 2020. Array programming with NumPy. *Nature*. 585(7825):357-362, doi: 10.1038/s41586-020-2649-2.
2. Collette, A. 2013. Python and HDF5. O'Reilly.
3. McKinney, W. 2011. pandas: a Foundational Python Library for Data Analysis and Statistics. *Python High Performance Science Computer*.
4. Virtanen, P., R. Gommers, . . . Y. Vázquez-Baeza. 2020. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*. 17(3):261-272, doi: 10.1038/s41592-019-0686-2.
5. Hunter, J. D. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*. 9(3):90-95, doi: 10.1109/MCSE.2007.55.

Supporting citations

References **(40-44)** appear in the Supplementary Material.

3.12. References

1. Ashkin, A., J. M. Dziedzic, . . . S. Chu. 1986. Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.* 11(5):288-290, doi: 10.1364/OL.11.000288.
2. Moffitt, J. R., Y. R. Chemla, . . . C. Bustamante. 2008. Recent advances in optical tweezers. *Annu Rev Biochem*. 77:205-228, doi: 10.1146/annurev.biochem.77.043007.090225.
3. Choudhary, D., A. Mossa, . . . C. Cecconi. 2019. Bio-Molecular Applications of Recent Developments in Optical Tweezers. *Biomolecules*. 9(1):23, doi: 10.3390/biom9010023.
4. Hashemi Shabestari, M., A. E. C. Meijering, . . . E. J. G. Peterman. 2017. Chapter Four - Recent Advances in Biological Single-Molecule Applications of Optical Tweezers and Fluorescence Microscopy. In Methods Enzymol. M. Spies, and Y. R. Chemla, editors. Academic Press, pp. 85-119.
5. Bustamante, C. J., Y. R. Chemla, . . . M. D. Wang. 2021. Optical tweezers in single-molecule biophysics. *Nature Reviews Methods Primers*. 1(1):25, doi: 10.1038/s43586-021-00021-6.
6. Chen, Y.-T., K.-C. Chang, . . . J.-D. Wen. 2017. Coordination among tertiary base pairs results in an efficient frameshift-stimulating RNA pseudoknot. *Nucleic Acids Res*. 45(10):6011-6022, doi: 10.1093/nar/gkx134.
7. Mukhortava, A., M. Pöge, . . . M. Schlierf. 2019. Structural heterogeneity of attC integron recombination sites revealed by optical tweezers. *Nucleic Acids Res*. 47(4):1861-1870, doi: 10.1093/nar/gky1258.
8. Stephenson, W., G. Wan, . . . P. T. Li. 2014. Nanomanipulation of single RNA molecules by optical tweezers. *J Vis Exp*.(90), doi: 10.3791/51542.
9. Zhong, Z., L. Yang, . . . G. Chen. 2016. Mechanical unfolding kinetics of the SRV-1 gag-pro mRNA pseudoknot: possible implications for -1 ribosomal frameshifting stimulation. *Sci Rep*. 6:39549, doi: 10.1038/srep39549.
10. Jiao, J., A. A. Rebane, . . . Y. Zhang. 2017. Single-Molecule Protein Folding Experiments Using High-Precision Optical Tweezers. *Methods Mol Biol*. 1486:357-390, doi: 10.1007/978-1-4939-6421-5_14.
11. Ritchie, D. B., J. Soong, . . . M. T. Woodside. 2014. Anti-frameshifting ligand reduces the conformational plasticity of the SARS virus pseudoknot. *J Am Chem Soc*. 136(6):2196-2199, doi: 10.1021/ja410344b.
12. Desai, V. P., F. Frank, . . . C. Bustamante. 2019. Co-temporal Force and Fluorescence Measurements Reveal a Ribosomal Gear Shift Mechanism of

Translation Regulation by Structured mRNAs. *Molecular Cell*. 75(5):1007-1019.e1005, doi: https://doi.org/10.1016/j.molcel.2019.07.024.
13. Liu, T., A. Kaplan, . . . C. J. Bustamante. 2014. Direct measurement of the mechanical work during translocation by the ribosome. *Elife*. 3:e03406-e03406, doi: 10.7554/eLife.03406.
14. Eriksson, E., J. Enger, . . . D. Hanstorp. 2007. A microfluidic system in combination with optical tweezers for analyzing rapid and reversible cytological alterations in single cells upon environmental changes. *Lab on a Chip*. 7(1):71-76, doi: 10.1039/B613650H, (10.1039/B613650H).
15. Gross, P., G. Farge, . . . G. J. Wuite. 2010. Combining optical tweezers, single-molecule fluorescence microscopy, and microfluidics for studies of DNA-protein interactions. *Methods Enzymol*. 475:427-453, doi: 10.1016/s0076-6879(10)75017-5.
16. Whitley, K. D., M. J. Comstock, . . . Y. R. Chemla. 2017. High-Resolution "Fleezers": Dual-Trap Optical Tweezers Combined with Single-Molecule Fluorescence Detection. *Methods Mol Biol*. 1486:183-256, doi: 10.1007/978-1-4939-6421-5_8.
17. Rocha, M. S. 2009. Optical tweezers for undergraduates: Theoretical analysis and experiments. *American Journal of Physics*. 77(8):704-712, doi: 10.1119/1.3138698.
18. McCauley, M. J., I. Rouzina, . . . M. C. Williams. 2020. Significant Differences in RNA Structure Destabilization by HIV-1 GagDp6 and NCp7 Proteins. *Viruses*. 12(5):484, doi: 10.3390/v12050484.
19. McCauley, M. J., I. Rouzina, . . . M. C. Williams. 2015. Targeted binding of nucleocapsid protein transforms the folding landscape of HIV-1 TAR RNA. *Proc Natl Acad Sci U S A*. 112(44):13555-13560, doi: 10.1073/pnas.1510100112.
20. Kuhn, M., H. Janovjak, . . . D. J. Müller. 2005. Automated alignment and pattern recognition of single-molecule force spectroscopy data. *J Microsc*. 218(Pt 2):125-132, doi: 10.1111/j.1365-2818.2005.01478.x.
21. Bosshart, P. D., P. L. T. M. Frederix, . . . A. Engel. 2012. Reference-free alignment and sorting of single-molecule force spectroscopy data. *Biophysical journal*. 102(9):2202-2211, doi: 10.1016/j.bpj.2012.03.027.
22. Heenan, P. R., and T. T. Perkins. 2018. FEATHER: Automated Analysis of Force Spectroscopy Unbinding and Unfolding Data via a Bayesian Algorithm. *Biophysical Journal*. 115(5):757-762, doi: https://doi.org/10.1016/j.bpj.2018.07.031.
23. Andreopoulos, B., and D. Labudde. 2011. Efficient unfolding pattern recognition in single molecule force spectroscopy data. *Algorithms for Molecular Biology*. 6(1):16, doi: 10.1186/1748-7188-6-16.
24. Gergely, C., B. Senger, . . . J. Hemmerlé. 2001. Semi-automatized processing of AFM force-spectroscopy data. *Ultramicroscopy*. 87(1-2):67-78, doi: 10.1016/s0304-3991(00)00063-2.
25. Roduit, C., B. Saha, . . . S. Kasas. 2012. OpenFovea: open-source AFM data processing software. *Nature Methods*. 9(8):774-775, doi: 10.1038/nmeth.2112.
26. Muhs, K. S., W. Karwowski, . . . D. Kern. 2018. Temporal variability in human performance: A systematic literature review. *International Journal of Industrial Ergonomics*. 64:31-50, doi: https://doi.org/10.1016/j.ergon.2017.10.002.
27. Hill, C. H., L. Pekarek, . . . I. Brierley. 2021. Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch. *Nature Communications*. 12(1):7166, doi: 10.1038/s41467-021-27400-7.
28. Zimmer, M. M., A. Kibe, . . . N. Caliskan. 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, doi: 10.1038/s41467-021-27431-0.

29. Neupane, K., M. Zhao, . . . M. T. Woodside. 2021. Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers. *Nature Communications*. 12(1):4749, doi: 10.1038/s41467-021-25085-6.
30. Pekarek, L., S. Buck, . . . N. Caliskan. 2022. Optical Tweezers to Study RNA-Protein Interactions in Translation Regulation. *JoVE*.(180):e62589, doi: doi:10.3791/62589.
31. Butterworth, S. 1930. On the Theory of Filter Amplifiers. *Experimental Wireless and the Wireless Engineer*. 7:536-541.
32. Odijk, T. 1995. Stiff Chains and Filaments under Tension. *Macromolecules*. 28(20):7016-7018, doi: DOI 10.1021/ma00124a044.
33. Smith, S. B., Y. Cui, . . . C. Bustamante. 1996. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science*. 271(5250):795, doi: 10.1126/science.271.5250.795.
34. Gore, J., F. Ritort, . . . C. Bustamante. 2003. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proceedings of the National Academy of Sciences*. 100(22):12564, doi: 10.1073/pnas.1635159100.
35. Liphardt, J., S. Dumont, . . . C. Bustamante. 2002. Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski's Equality. *Science*. 296(5574):1832, doi: 10.1126/science.1071152.
36. Collin, D., F. Ritort, . . . C. Bustamante. 2005. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*. 437(7056):231-234, doi: 10.1038/nature04061.
37. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 31(13):3406-3415, doi: 10.1093/nar/gkg595.
38. Alberdi, E., L. Strigini, . . . P. Ayton (2009). Why Are People's Decisions Sometimes Worse with Computer Support? In B. Buth, G. Rabe, and T. Seyfarth, eds. Computer Safety, Reliability, and Security. Springer Berlin Heidelberg.
39. Cummings, M. L., F. Gao, . . . K. M. Thornburg. 2016. Boredom in the Workplace: A New Look at an Old Problem. *Hum Factors*. 58(2):279-300, doi: 10.1177/0018720815609503.
40. Harris, C. R., K. J. Millman, . . . T. E. Oliphant. 2020. Array programming with NumPy. *Nature*. 585(7825):357-362, doi: 10.1038/s41586-020-2649-2.
41. Collette, A. 2013. Python and HDF5. O'Reilly.
42. McKinney, W. 2011. pandas: a Foundational Python Library for Data Analysis and Statistics. *Python High Performance Science Computer*.
43. Virtanen, P., R. Gommers, . . . Y. Vázquez-Baeza. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 17(3):261-272, doi: 10.1038/s41592-019-0686-2.
44. Hunter, J. D. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*. 9(3):90-95, doi: 10.1109/MCSE.2007.55.

Chapter 4

"That's one small step for a man,
one giant unfolding for an RNA molecule"
Neil Armstrong, modified

4. Cis-mediated interactions of the SARS-CoV-2 frameshift RNA alter its conformations and affect function

Lukas Pekarek[1], Matthias M. Zimmer[1,§], Anne-Sophie Gribling-Burrer[1,§], Stefan Buck[1,§], Redmond Smyth*,[1,2], Neva Caliskan*,[1,2]

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

[2] Medical Faculty, Julius-Maximilians University Würzburg, Germany

[§] Equal contribution

* Corresponding authors

Neva Caliskan neva.caliskan@helmholtz-hiri.de

Published: Nucleic Acids Research, gkac1184 (2022)

4.1. Abstract

The RNA genome of SARS-CoV-2 contains a frameshift stimulatory element (FSE) that allows access to an alternative reading frame through –1 programmed ribosomal frameshifting (PRF). –1PRF in the *1a/1b* gene is essential for efficient viral replication and transcription of the viral genome. –1PRF efficiency relies on the presence of conserved RNA elements within the FSE. One of these elements is a three-stemmed pseudoknot, although alternative folds of the frameshift site might have functional roles as well. Here, by complementing ensemble and single-molecule structural analysis of SARS-CoV-2 frameshift RNA variants with functional data, we reveal a conformational interplay of the 5' and 3' immediate regions with the FSE and show that the extended FSE exists in multiple conformations. Furthermore, limiting the base pairing of the FSE with neighboring nucleotides can favor or impair the formation of the alternative folds, including the pseudoknot. Our results demonstrate that co-existing RNA structures can function together to fine-tune SARS-CoV-2 gene expression, which will aid efforts to design specific inhibitors of viral frameshifting.

4.2. Introduction

Many viruses, including Coronaviruses, employ several recoding strategies that allow access to an overlapping and functional reading frame, thereby increasing the regulatory potential and the coding repertoire of their genomes (1-7). One such recoding event is –1 programmed ribosomal frameshifting (–1PRF), where ribosomes slip back by one nucleotide in the 3' to 5' direction (–1) into an alternative reading frame. The frameshift stimulatory element (FSE) that promotes frameshifting typically consists of a slippery site (SS, the heptanucleotide UUUAAAC in SARS-CoV-2, **Fig. 4.1A**) and an RNA structure in form of a stem loop or pseudoknot located at a defined distance of 5-9 nucleotides downstream of the SS (1,7-9). The role of the PK structure is to slow down translation elongation by impeding translocation of the tRNAs over the slippery codons and thereby facilitating new codon-anticodon interactions in the alternative reading frame (10,11).

Since the outbreak of the COVID-19 pandemic, several studies explored the structure of the SARS-CoV-2 FSE employing nuclear magnetic resonance (NMR), X-ray crystallography, cryo-electron microscopy (Cryo-EM), optical tweezers and chemical probing techniques (12-20). These structural studies of the PK RNA in solution or engaged with the ribosome show that it folds into an H-type pseudoknot with coaxially stacked SL1 and SL2 stems which form a continuous helix, while SL3 is perpendicularly positioned with respect to this helix. Several groups have proposed that the 5' of the RNA can pass through a ring formed inside the three stems of the PK – this particular PK fold was named the threaded PK (13)(19). On the ribosome, this well-structured RNA directly interacts with the proteins of the small subunit positioned at the mRNA entry channel and thus creates a mechanical hindrance to resist unwinding by the intrinsic helicase activity of the ribosome (12). Recently published SARS-CoV-2 RNA structure probing data *in vitro* and in infected cells indicate that the frameshift element can also be found in alternative conformations (15,16,21,22). *In silico* work by Schlick *et al.* also predicted several alternative folds, including a three-way junction (3WJ) forming with the 3' and 5' immediate ends of the PK and an alternative PK where the SL1 loop base pairs with nucleotides upstream of the FSE (21,23,24). In some cases, sequences upstream or downstream of the core FSE were proposed to affect frameshifting. For example, in SARS-CoV, an attenuator hairpin upstream of the PK was suggested to moderate frameshifting (25-27). In infected cells, several groups identified the 5' upstream attenuator hairpin as part of an extended stem of varying lengths, which they termed the alternative stem or alternative duplex (15,22). Despite the varying length, there is a consensus that the alternative stem would include the slippery sequence and the SL1 of the PK, making it unable to fold into the PK (15,22). Recently, a similar structure was also supported by *in situ* conformation sequencing of RNA inside SARS-CoV-2 particles (28). Lastly, longer-range interactions of the FSE were

predicted in genome-wide studies; these might have functions in structural organization or replication of the genome (15,20). Collectively, these data indicate that, although the isolated FSE contains a pseudoknot as the dominant fold, as the length of the frameshift RNA increases the molecule can be found in multiple alternative conformations, which may have functional roles during frameshifting and viral replication (15,21).

To gain a comprehensive view of the conformational heterogeneity of the SARS-CoV-2 FSE, we used an integrative approach with single-molecule optical tweezers (OT) and chemical probing by dimethyl sulfate mutational profiling (DMS-MaP) (**Fig. 4.1B** and **C**). To determine the functional relevance of the respective conformers, we employed a flow cytometry-based cellular frameshift reporter assay (**Fig. 4.1D**). This allowed us to evaluate the functional effect of different RNA variants on translation from the –1 reading frame (29). We show that the canonical PK is the effective structure to stimulate frameshifting, and that both the standard and the threaded PK folds can induce frameshifting (13,19). Through employing 5' and 3' extended FSE constructs and mutants that interfere with the folding of the PK (**Fig. 4.S1**), we demonstrate that 25-50 nucleotides at the 5' and 3' regions of the FSE promote alternative folds that interfere with the formation of the frameshift stimulatory PK fold and lead to differences in frameshifting efficiencies. Furthermore, we were able to modulate the folding of the FSE by occluding specific base-pair interactions involving SL2 and 3' downstream regions by employing antisense oligonucleotides (ASOs). Using locked nucleic acid (LNA) containing ASOs targeting the SL2 of the pseudoknot we show that we can abolish frameshifting *in vitro* and a in cell-based reporter assay. Taken together, our dynamic- and steady-state analysis of the frameshift RNA suggest that there are alternative conformations of the FSE mediated by *cis*-acting short-range RNA interactions that fine-tune gene expression of SARS-CoV-2, which would provide important clues for future structure-based drug design studies.

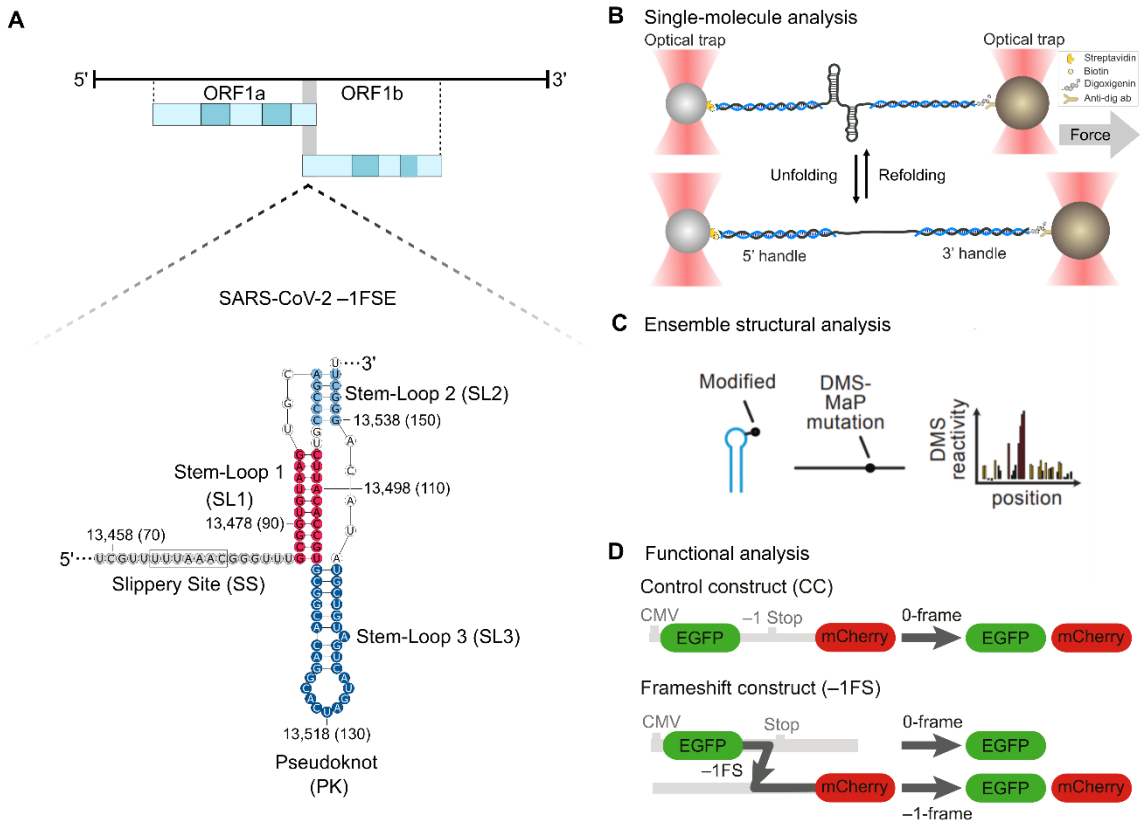

Figure 4.1 An integrated system for probing the structural landscape of the SARS-CoV-2 FSE. (**A**) Schematic representation of the SARS-CoV-2 programmed –1 ribosomal frameshifting element within the SARS-CoV-2 genome. The secondary structure of the FSE is derived from published structural models (12,13). SL1 (in magenta), SL2 (in light blue) and SL3 (in dark blue) with the corresponding genomic nucleotide positions indicated; numbers in brackets refer to the position relative to the longest RNA variant measured in this study (FSE-V4). (**B**) Schematic illustration of optical tweezers experiments. RNA is hybridized to single-stranded DNA handles flanking the SARS-CoV-2 frameshift site and conjugated to functionalized beads. A focused laser beam is used to exert pulling force from one end of the molecule. The force is gradually increased until the RNA fully unfolds (bottom). (**C**) Structural profiles are obtained by DMS-MaP, in which DMS preferably reacts with unpaired A and C residues, which are later read out as mutations and converted into DMS reactivities by comparing DMS-treated samples vs. untreated controls. (**D**) Scheme of the dual-fluorescence frameshift reporter construct. EGFP and mCherry are separated by a StopGo signal as well as by a stop codon in-frame with EGFP. As a result, translation in 0-frame produces only EGFP, whereas translation in –1-frame produces both EGFP and mCherry. The ratio of mCherry to EGFP fluorescence normalized to a control lacking the frameshift signal encoding both EGFP and mCherry is used to quantify frameshift efficiencies (FE). See also Materials and Methods.

4.3. Materials and Methods

Plasmid construction

To generate dual-fluorescence reporter constructs, frameshift sites of SARS-CoV-2 and corresponding mutant/truncated variants were placed between the coding sequences of EGFP and mCherry (parental construct was a gift from Andrea Musacchio (Addgene plasmid #87803 (30)) 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 signals (31) as well as an alpha-helical linker (32). 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. Variants and mutants of the frameshift site of the SARS-CoV-2 in the dual-fluorescence construct, as described in **Fig. 4.2** and **4.4** and **Table 4.S1,** were generated by Golden Gate Assembly.

OT 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 handle sequences of lambda DNA (2 kb each) flanking the RNA structure of interest (29,33). Constructs were generated using Gibson Assembly. The additional RNA variants and mutants were generated by site-directed mutagenesis. Sequences of the variants used in this study are given in **Table 4.S1**.

Optical tweezers constructs

5' and 3' DNA handles, and the template for *in vitro* transcription of the RNA samples 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 in-house purified 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. Following the annealing, samples were concentrated by ethanol precipitation, pellets were resuspended in 50 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) as described before (29,33,34). For the experiments, optical tweezers (OT) constructs were mixed with 4 µl of polystyrene beads coated with antibodies against digoxigenin (AD beads, 0.1% w/v suspension, Ø 2.12 µm, Spherotech), 10 µl of assay buffer (20 mM HEPES, pH 7.6, 300 mM KCl, 5 mM $MgCl_2$, 5 mM DTT and 0.05% Tween 20) and 1 µl of RNase inhibitor (Molox). 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% w/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 streptavidin beads and the complex of OT construct with anti-digoxigenin beads were introduced into the flow cell. During the experiment, single AD and SA beads were trapped in individual traps 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. FD data were recorded at a rate of 78125 Hz.

Raw data files were processed using the Practical Optical Tweezers Analysis TOol (POTATO, https://github.com/REMI-HIRI/POTATO (35)). In brief, raw data were first downsampled by a factor of 30 to speed up subsequent processing, and the noise was filtered using Butterworth filter (0.005 filtering frequency, filter order 4). Folded as well as unfolded regions of the FD curves were then fitted. 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 (34). First, 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. 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 fitting parameters were derived and the results were plotted using Prism 9.2.0 (GraphPad).

Cell culture

HEK293 cells (gift from Prof. Jörg Vogel, HIRI-HZI) were maintained in DMEM (Gibco) supplemented with 10% FBS (Gibco) and 100 µg/ml streptomycin and 100 U/ml penicillin at 37°C with 5% CO_2. Transfections were performed using linear 25 kDa PEI (Polysciences) according to manufacturer's instructions using a 1:12 DNA:PEI ratio. For

co-transfections, plasmids and antisense oligonucleotides were mixed in a 1:40 molar ratio and electroporations were performed in OptiMEM (Gibco) using a Nepa21 Super Electroporator (NEPAGENE) following the manufacturer's instructions. 24 h post-transfection, the cells were analyzed by flow cytometry. Antisense oligonucleotides (IDT) were added to the cells in a final concentration of 100 nM. To increase stability of the oligonucleotides, the phosphate backbone was substituted by phosphothioate and three bases at the 5' and 3' end were locked nucleic acids (LNAs).

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. 4.1D**. Cells were harvested at 24 h post-transfection. After washing with phosphate-buffered saline (PBS), flow cytometry was performed on a NovoCyte Quanteon (ACEA) instrument. FE was calculated according to the following formula:

$$\text{FE}(\%) = \frac{\text{mCherry}_{\text{test}}/\text{EGFP}_{\text{test}}}{\text{mCherry}_{\text{control}}/\text{EGFP}_{\text{control}}} \times 100$$

where mCherry represents the mean –1-frame mCherry intensity, EGFP the mean EGFP intensity, test represents the sample containing the frameshift element and control represents the in-frame control lacking the frameshift element where mCherry and EGFP are produced from the 0-reading frame. To ensure reliability of the present assay, several controls were employed. These included vectors with deleted CMV promoter, deleted slippery sequence, mutated slippery sequence, and a readthrough control (**Fig. 4.S2**). Data represents the results of at least three independent experiments. Data was analyzed and plotted in GraphPad Prism (version 9.2.0). For statistical analysis of FE in the presence of antisense oligonucleotides, an 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 compare test samples to control constructs (non-targeting antisense oligonucleotides).

***In vitro* translation in rabbit reticulocyte lysate (RRL)**

In vitro translations in rabbit reticulocyte lysate (RRL) were performed as described before (29). Briefly, mRNAs were synthesized by *in vitro* transcription with in-house purified T7 polymerase using linearized plasmid DNAs as templates and subsequently translated using rabbit reticulocyte lysate (RRL; Promega). Typical reactions were 0.05 µM template mRNA, 75% v/v RRL, and 20 µM amino acids. Antisense oligonucleotides (IDT) were added to the reactions in a concentration range from 0-0.5 µM. Reactions were

incubated for 1 h at 30°C at which point samples were mixed with three volumes of 1× BOLT™ LDS Sample Buffer (Invitrogen), denatured for 10 min at 70°C, and resolved on NuPAGE™ 4-12% Bis-Tris polyacrylamide gels (Invitrogen). Translation products were detected by western blot. Briefly, after transfer using Trans-Blot (Bio-Rad), nitrocellulose membranes were developed with anti-DYKDDDK-tag antibody (Proteintech 20543-1-AP, 1:3000) and visualized by incubation with a secondary antibody (IRDye® 800CW Goat anti-rabbit, LI-COR, dilution 1:10000)). Bands were detected using an Odyssey Clx infrared imager system (LI-COR) and quantified densitometrically using FIJI software (36). FE was calculated as previously described, by the formula:

$$\text{FE}(\%) = \frac{\text{Intensity}_{-1-\text{frame}}}{\text{Intensity}_{0-\text{frame}} + \text{Intensity}_{-1-\text{frame}}} \times 100$$

The relative FE was calculated as a ratio of FE of each condition to the FE of no-oligonucleotide control in each measurement. Experiments were repeated at least three independent times. Data was plotted and IC50 values were calculated using an inhibitor vs. response model assuming a standard slope (Hill slope= –1) in GraphPad Prism (version 9.2.0) software.

Dimethyl sulfate mutational profiling (DMS-MaP)

A mix of 37 nM of RNA, 37 nM of blocking primer 5' [GTAGCTGTCGAGCTCCTGCGAAG] and 37 nM of blocking primer 3' [GGCGAAGAGCAGGTTGCAGGAT] was first heat denatured at 90°C for 2 min, then annealed by slow cooling (< 1°C/min) to 50°C. RNA was then folded by adding the folding buffer (20 mM HEPES, pH 7.6, 250 mM KCl, 5 mM MgCl2, and 0.3 U/µl of RNasin) and slow cooling (< 1°C/min) to 23°C. DMS was diluted in ethanol to a working concentration of 0.85 M. 1/10 volume of DMS working stock was added to the samples to make a final concentration of 85 mM in a total volume of 33 µl. Samples were incubated at 37°C for 6 min and then quenched with 33 µl of beta-mercaptoethanol (14.2 M stock). For the untreated control, ethanol was used instead of DMS. RNA was then purified by Trizol LS (Sigma) according to the manufacturer's instructions.

Half of the RNA samples were reverse transcribed using 40 U MarathonRT (37,38) in RT Buffer (50 mM Tris-HCl, pH 8.3, 200 mM KCl, 5 mM DTT, 20% glycerol, 1 mM MnCl2), 0.5 mM dNTP mix, 0.3 µM primer [GGCGAAGAGCAGGTTGCAGGAT] and 8 U of RNasin in a final volume of 25 µl. Reverse transcriptions were carried out at 42°C for 4 h. cDNA was diluted 1/8 with nuclease-free water and PCR amplified. Reaction conditions were 8 µl of diluted cDNA, 1× GXL reaction buffer, 0.2 mM dNTPs, 0.25 µM of forward [TCGTCGGCAGCGTCAGCTTCGCAGGAGCTCGACAGCTAC] and reverse primer [GTCTCGTGGGCTCGGAGGGCGAAGAGCAGGTTGCAGGAT], 0.02 U/µl of Q5 High-

Fidelity DNA Polymerase (NEB) in a final volume of 50 µl. Cycling conditions were 2 min at 98°C then 25 cycles of 10 sec at 98°C, 15 sec at 60°C and 15 sec at 72°C then 72°C for 5 min. PCR products were verified on a 2% agarose gel followed by column purification (NucleoSpin Gel and PCR Clean-up kits, Macherey-Nagel) according to the manufacturer's instructions. A final indexing PCR was carried out using Illumina Nextera DNA CD indexes (96 Indexes, 96 Samples, Illumina). Reaction conditions were 40 ng of purified PCR product, 1× Q5 reaction buffer, 0.2 mM dNTPs, 2.5 µl of indexing primer, 0.02 U/µl of Q5 polymerase in a final volume of 14 µl. Indexed PCR products were verified on a 1.5% agarose gel, pooled together in an equimolar ratio, before final purification on a 1.5% agarose gel. The pooled indexed sequencing library was quantified using the NEBNext library Quant Kit for Illumina and paired-end PE150 sequencing was carried out on an Illumina Novaseq instrument (Novogene).

DMS-MaP data was trimmed using cutadapt v 1.18 (39) and aligned to the reference sequence using bowtie2 (40). cutadapt parameters were “–nextseq-trim 20 –max-n 0 -a atcctgcaacctgctcttcgcc -A gtagctgtcgagctcctgcgaag”. bowtie2 parameters were “-D 20 -R 3 -N 1 -L 15 -i S,1,0.50”. Further analysis was carried out using the rf-count and rf-norm modules of RNA Framework package v2.7.2 (41). rf-count parameters were “-m -es”. rf-norm parameters were “-rb AC -sm 3 -nm 1”, meaning that DMS reactivities were calculated by subtracting background mutations in the untreated sample and normalized using 2-8% normalization (42). Alternative structures were detected directly from the DMS data using the clustering algorithm Detection of RNA folding Ensembles using Expectation-Maximization' (DREEM) (43). Briefly, RNA molecules from DMS-MaP experiments often contain multiple modifications that can be used later to distinguish between alternative folds. The DREEM algorithm uses expectation maximization to directly assign individual sequencing reads to distinct structural clusters. The number of clusters and their relative proportions are computed by iteratively maximizing a log-likelihood function based on a multivariate Bernoulli mixture model. The resulting DMS reactivities for each structural class were then used to predict RNA structures using RNAstructure using the default parameters embedded in the DREEM pipeline. Data were plotted using StructureEditor (version 1.0) (44).

4.4. Results

5' and 3' extensions of the FSE can favor alternative conformers

Frameshifting has been shown to be induced by a 68 nucleotide (nt) long pseudoknot structure within the FSE, yet analyses of the SARS-CoV-2 whole genome

structure suggest that the region including the FSE can adopt multiple conformations in cells (15-17,21). In order to determine the translation relevant structural and functional effects proximal and distal nucleotides of the FSE, we designed and tested four variants (**Fig. 4.2A, S1**). The reference RNA, which we call the core FSE of 86 nt, contains the slippery sequence, spacer and the pseudoknot composed of SL1, SL2 and SL3. FSE-V1 contains the FSE and an extended 3' region of 28 nucleotides (total length: 114 nt). FSE-V2 contains extensions immediately 5' and 3' of the FSE, including the 5' attenuator hairpin (length: 141 nt). FSE-V3 contains the 3' extension (28 nt) and a 38 nt extension at the 5' end, including the recently proposed alternative stem 1 (length: 152 nt) (15,22). Lastly, FSE-V4 contains the longest 5' extension of 68 nucleotides and an additional 3' extension of 39 nt (length: 221 nt) (**Fig. 4.S2A**).

We first asked whether upstream and downstream regions of the FSE can alter frameshifting. For that we used a cell-based dual-fluorescence frameshift reporter, in which the SARS-CoV-2 frameshift site is placed between the EGFP gene in the 0-frame and the mCherry gene in the –1-frame (**Fig. 4.1D**) (29). With the core FSE construct, frameshifting efficiency was 41.6±0.3%. The presence of extensions on either 3' (FSE-V1) or 5’ (FSE-V2) of the FSE resulted in a decrease of frameshifting levels to approximately 25% (**Fig. 4.2A**). The FSE-V3 construct containing an additional 5’ extension had a frameshifting efficiency of 30% (**Fig. 4.2A**) The FSE-V4 construct, which has the longest extension among the variants tested had the highest frameshift efficiency of about 37%, close to the stimulation observed with the wild type FSE (**Fig. 4.S3A**).

In order to test whether the effect on frameshifting is related to altered conformation of the FSE variants, we next employed single-molecule optical tweezers (**Fig. 4.2B** and **C**, **Tables 4.S2** and **4.S3**). Here, FSE RNA variants flanked by 2 kb DNA:RNA hybrid handles were gradually stretched at a constant rate and then the applied force was released allowing the RNA molecule to refold. This allows the RNA molecule to shift between folded and unfolded states and sudden changes in measured force-distance (FD) trajectories represent transitions between RNA conformations (29) (**Fig. 4.1B**). The (un)folding force as well as the change in contour length and folding hysteresis (the difference between the observed unfolding and refolding transitions) provides the information on RNA conformational states.

In the FSE RNA, which in addition to the PK contains the slippery sequence and the spacer, we mostly observed a single unfolding step in 72.2% of the FD trajectories (**Fig. 4.2B-C, Table 4.S3**). Most of these unfolding events occurred at forces above 15 pN and exhibited hysteresis, which is typical for structured RNAs, like pseudoknots (19,45-48) (**Fig. 4.2B-C**). In 20.8% of the FD curves, we noted two successive unfolding events (**Table 4.S3**), each with an average of 17.1±3.5 nm change in the contour length, which

would correspond to the opening of ≈32 nucleotides. The total change in contour length of the FSE sample was 35.5±2.5 nm, which was close to the expected length of 37.2 nm for the PK (68 nt) (13,19) (**Fig. 4.S4A** and **Table 4.S2**). The two-step unfolding pattern may represent the sequential unfolding of individual stem-loops, SL2+3 and SL1 or may indicate presence of an alternative fold (**Fig. 4.2B** and **Table 4.S2**). The histogram of the unfolding forces also pointed to the presence of at least two populations in the FSE sample (**Fig. 4.2C** and **Table 4.S2**). First, one unfolding event at the force of 11.9±1.8 pN, and a second event with a mean force of unfolding at 20.1±4.9 pN (**Table 4.S2**). Based on the individual refolding trajectories, the population unfolding at lower forces would correspond to a mixture of stem-loops and a less-stable PK fold (**Fig. 4.2B-C**). The second population unfolding at higher forces corresponds to a highly stable PK fold, which likely represents the threaded PK (13,19) (**Fig. 4.2B-C**).

Previous studies also predicted that the 3' end of the PK can be involved in alternative folds, with unknown functions (15,16,20-22). To test whether the immediate 3' nucleotides lead to the formation of conformations that compete or co-exist with the FSE and affect the dynamics of the canonical FSE, we employed the FSE-V1 variant (**Fig. 4.2A** and **4.S1**). With this RNA variant we noticed a decrease in single-step unfolding events from 72% to 55% as compared to the FSE sample (**Table 4.S3**). In addition, we marked a decrease in the population of a highly stable conformer, which we infer to be the threaded PK, although the less stable PK fold was still pronounced (**Fig. 4.2B-C** and **Table 4.S2**). The total change in contour length of the FSE-V1 sample was slightly increased (37.5±3.6 nm) compared to FSE (**Fig. 4.S4B** and **Table 4.S2**).

Next, we employed the FSE-V2 RNA variant containing the putative attenuator hairpin at the 5' of the FSE, which was previously reported to decrease frameshifting in SARS-CoV-1 and SARS-CoV-2 (25,26) (**Fig. 4.2A** and **4.S1**). The (un)folding behavior of the FSE-V2 variant was similar to the FSE-V1 sample, with no noticeable threaded PK-like unfolding events and a similar change in the contour length (**Fig. 4.2B-C, 4.S4C** and **Table 4.S2**). However, we see an increase in two step unfolding event (**Table 4.S3**). Interestingly, during unfolding of both RNA variants with 5' and/or 3' extensions we also observed a few folding rescue events (**Fig. 4.S5**). These events were characterized by reversible folding behavior during pulling, suggesting that the RNA molecule dynamically explores various intermediate states, while it partially unfolds and immediately refolds into an energetically more favored structure (49).

We then evaluated the effect of the alternative stem (AS) (or the so-called alternative duplex) formed by base pairing between the 5' upstream region of the attenuator hairpin and parts of the FSE including the stem of SL1 (15,22). For that, we designed the FSE-V3 variant containing an 11 nt extension 5' of the attenuator hairpin

(**Fig. 4.2A** and **4.S1**). Similar to the FSE-V1 and FSE-V2 variants, the threaded PK state was absent in the FSE-V3 sample (**Fig. 4.2B-C**). Unlike the other variants, we noted a substantial population of unfolding steps at lower forces (5.4±1.1 pN) and overall higher heterogeneity of FD trajectories, pointing to the presence of less stable short hairpins in the RNA (**Fig. 4.2B-C**, **4.S4D**, **4.S5**, and **Tables 4.S2-3**). Lastly, we tested the effect of further extending the 5' end of the FSE by 68 nucleotides (FSE-V4) (**Fig. 4.S1** and **4.S3**). This sample was found to be highly heterogeneous indicating presence of at least 3 low force unfolding hairpins, together with a higher force unfolding population with a large change in the contour length by 102.0±10 nm (**Fig. 4.S3** and **Table 4.S2**).

Overall, based on our single-molecule analysis we conclude that the PK is the major conformation for the FSE, yet the FSE can adopt pseudoknot structures with varying stabilities. Interestingly, addition of 5’ and 3’ proximal nucleotides to the canonical FSE resulted in the loss of the highly stable RNA structures, suggestive of alternative no-PK folds present together with a low stability PK.

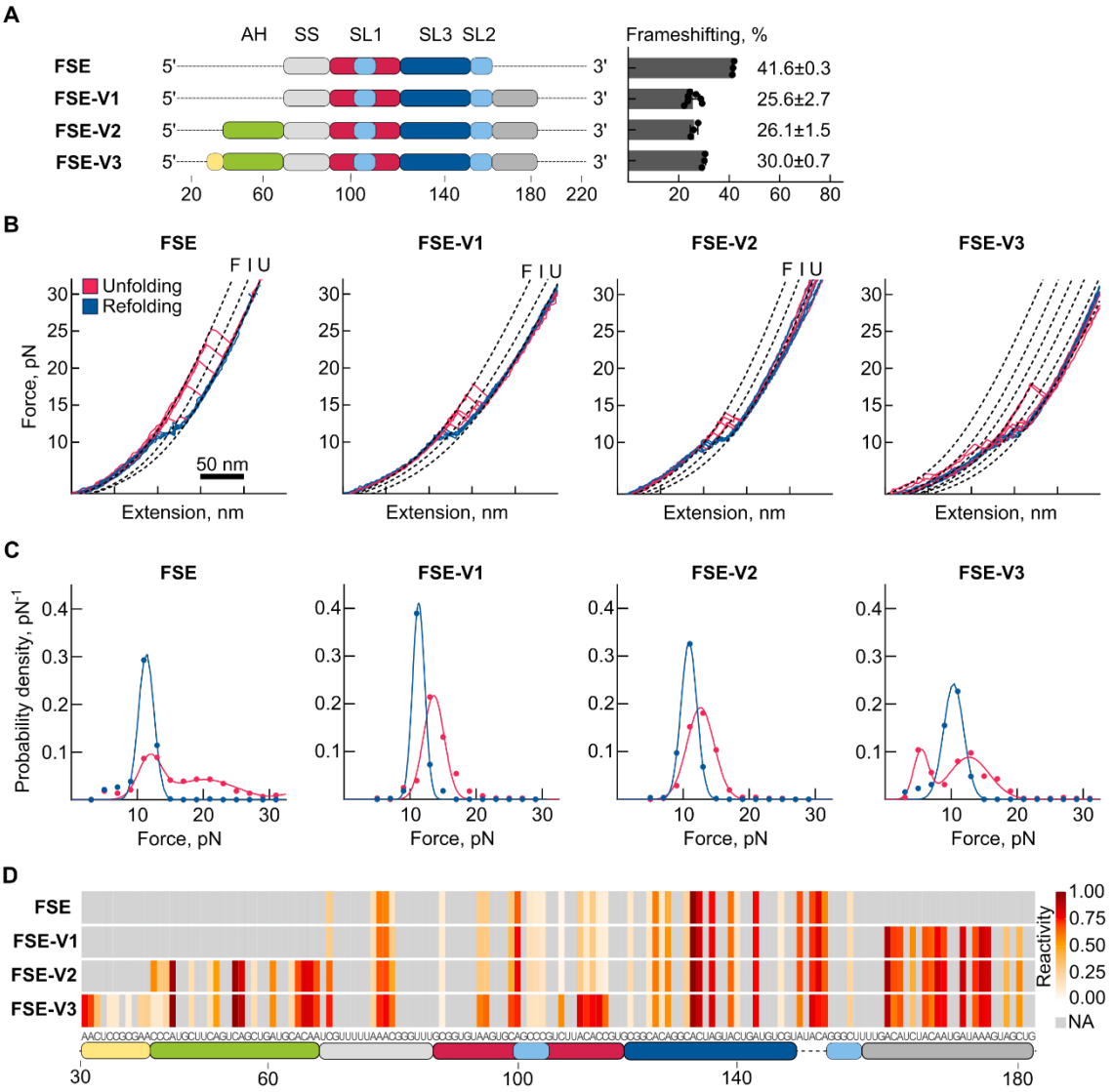

Figure 4.2 Effects of 5' and 3' extensions on frameshift efficiencies and RNA structure. (**A**) Schematic representation of RNA variants with respective parts of the FSE employed in the study. The color scheme matches the SARS-CoV-2 FSE in Figure 1. The yellow and green blocks correspond to the alternative stem 1 forming regions (AS1) and attenuator hairpin (AH), respectively. Numbers represent position relative to the beginning of the FSE-V4. Frameshift efficiencies of the variants are shown on the right. (**B**) Example unfolding (red) and refolding (blue) force-distance (FD) curves. “F” denotes the folded states, “I” the intermediate, and “U” the fully unfolded state. (**C**) Histograms of the force distribution for the unfolding (red) and refolding (blue) steps observed in each sample. (**D**) Reactivity profiles of the RNA variants as determined by DMS-MaP.

Chemical probing of the structural ensembles confirms presence of alternative folds

To investigate the conformational heterogeneity of the FSE and its variants, we next generated structural profiles of the RNAs based on reactivities of RNA to the chemical DMS (50,51) (**Fig. 4.1C**). In the FSE sample, DMS reactivities were consistent with the formation of the pseudoknot, especially the nucleotides within the loop of SL1 involved in the formation of SL2 were unreactive to DMS (**Fig. 4.2D**). Importantly, SL1 loop nucleotides became reactive to DMS in constructs containing deletions in the 3' stem of SL2, together with a deletion of SL3 (FSE-D1) or alone (FSE-D2) (**Fig. 4.S6**). However, in the absence of the SL3 alone (FSE-D3), SL1 loop nucleotides remained unreactive and indicating the formation of a minimal PK composed of SL1 and SL2, which was still able to support frameshifting in a reporter assay, in contrast to the FSE-D1 and FSE-D2 mutants which were unable to frameshift (**Fig. 4.S6**).

In the FSE-V1 and FSE-V2 samples we saw negligible differences compared to the FSE sample (**Fig. 4.2D** and **4.S7**). In the FSE-V3, which contained the putative alternative stem at its 5' end, more prominent changes in DMS reactivities were seen (**Fig. 4.2D** and **4.S7**). The most striking change was the strongly increased reactivity in both strands of the SL1 stem, particularly nucleotides 110-114, which were protected in the FSE sample and other FSE RNA variants (**Fig. 4.2D** and **4.S7**), indicating structural rearrangements of the FSE consistent with the unstable folding states observed in single-molecule pulling experiments (**Fig. 4.2B-C**). FSE-V4 showed increased DMS reactivities within SL1 suggestive of an alternative fold similar to FSE-V3 (**Fig. 4.S3E** and **Fig. 4.S7**).

To probe for alternative folds, we next analyzed DMS-MaP data at the single-molecule level using the Detection of RNA folding Ensembles using Expectation-Maximization' (DREEM) algorithm (15). DREEM clusters single-molecule measurements of RNA structure in DMS-MaP experiments to detect and quantify the relative abundance of alternative structures. In the FSE sample, we detected two alternative conformations (**Fig. 4.3A**). Cluster 1 was detected at 35% abundance and is consistent with the canonical PK structure. Cluster 2 was detected at 65% abundance and contained a shifted SL1 structure, which no longer exposed the loop required for PK formation (**Fig. 4.3A**). In FSE-V1, we identified two conformations, with a decrease in the relative abundance of the putative PK structure from 35% to 25% (**Fig. 4.3B**). The second conformation was predicted to form a 3WJ, consistent with the one proposed by Schlick *et al*. (21,24). Interestingly, formation of this 3WJ conformer is further supported by DMS reactivities at U residues, which were recently demonstrated to be also susceptible to DMS (52) (**Fig. 4.S7**). U bases in the immediate upstream (84-86 nt) and downstream (159-160 nt) of the

PK, which would be base-paired in the 3WJ fold, show substantially lower reactivity compared to the same bases in FSE sample (**Fig. 4.S7**). In FSE-V2, two very similar clusters were detected at relative proportions of 29% and 71% (**Fig. 4.3C**). The most notable difference in DMS reactivities between these two clusters was seen at a single position (C99) in the loop of SL1, suggestive of a subtle change in the PK structure (**Fig. 4.3C**). In FSE-V3, intriguingly, we detected a single structural cluster which maintained the canonical SL3 and SL2 structures. However, SS and SL1 were folded differently as part of the alternative stem structure (AS1) (**Fig. 4.3D**). Finally, FSE-V4 was seen as a single structural cluster, containing a slightly expanded SL3 and a long alternative stem (AS1) similar to the ones reported in chemical probing experiments conducted in cells and virions (15,22,28) (**Fig. 4.3E**).

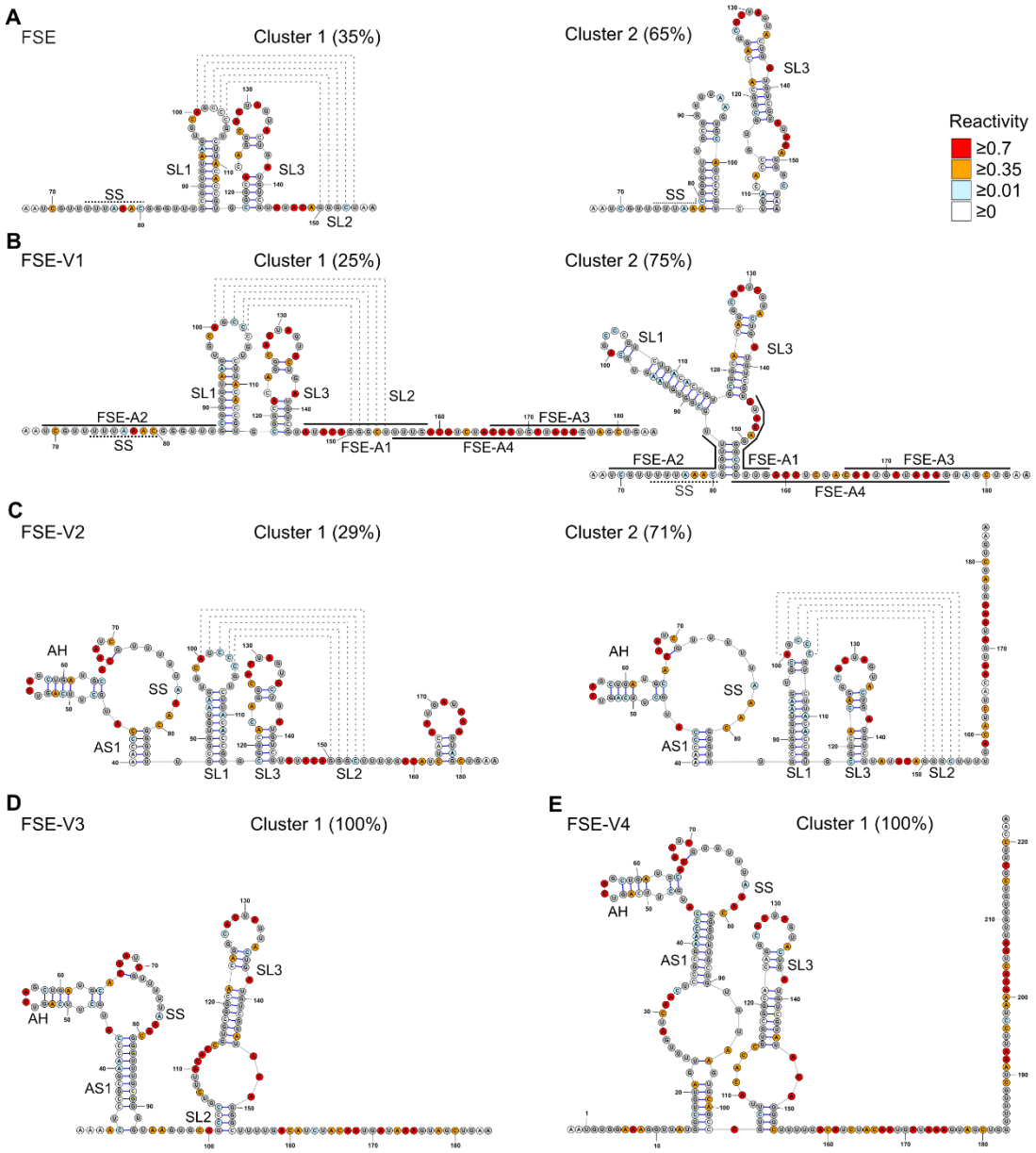

Figure 4.3 DMS-guided cluster analysis of FSE variants. (**A-E**) Secondary structures detected based on clustering of DMS reactivities using DREEM. DREEM does not predict pseudoknots, and dotted lines represent PK interactions supported by DMS-MaP. Structural elements are annotated according to the nomenclature described in the manuscript (**A**) FSE, (**B**) FSE-V1, binding sites of oligonucleotides used in this study are marked, (**C**) FSE-V2, (**D**) FSE-V3, (**E**) FSE-V4.

Mutations of FSE can favor alternative folds

To explore the role of the PK structure and to dissect how alternative conformations of the FSE might impact frameshifting, we designed FSE mutants (FSE-M1-3), based on the FSE-V1 mRNA. These mutants were computationally predicted to either disrupt the PK fold or alter equilibrium between the alternative folds identified above (24) (**Fig. 4.4A** and **4.S1**).

FSE-M1 was designed to interfere with PK formation by disrupting SL2 whilst leaving SL1 and SL3 of the FSE intact (**Fig. 4.4A** and **4.S1**)(24). Most of the FD curves (80.7%) presented a two-step unfolding pattern, first a small step at 9.4±1.3 pN followed by a second step at 12.6±1.1 pN with total contour length change of 26.1±3.4 nm (**Fig. 4.4B**-**C**, **4.S4E**, and **Tables 4.S2**-**3**). Furthermore, the pattern of unfolding in the FSE-M1 mutant showed remarkable similarity to the FSE-D2, which lacks the 3' strand of SL2, indicating that these two steps correspond to opening of the SL3 and SL1, respectively (29). In the remaining 20% of the FD curves, we detected only the second step around 13 pN, denoting that the SL3 either did not form or the transition was below the detection limit (**Table 4.S2**). In addition, DMS-MaP data pointed to increased reactivities in SL2 indicating that this mutant disrupted the pseudoknot structure (**Fig. 4.4D**). Frameshifting efficiencies dropped from 25.6% to 1.3% demonstrating that SL1 and SL3 are not sufficient for frameshifting (**Fig. 4.4A**).

The FSE-M2, contains two point mutations at the 3' end of SL1 (**Fig. 4.4A** and **4.S1**). This mutant is predicted to shift the location of SL1 by forming additional base pairing with the spacer and the C of the slippery sequence (UUUAAAC), forming a structure similar to cluster 2 of FSE (**Fig. 4.3A**)(24). In pulling experiments, FSE-M2 unfolded in two steps in most of the FD curves (69.5%) both at 11.9±1.3 pN with total contour length change of 36.3±1.6 nm (**Fig. 4.4B**-**C, 4.S4F** and **Tables 4.S2-3**). DMS-MaP data indicated an increase in DMS reactivity of two A residues at positions 94-95 found at the 5' strand of SL1, supporting the formation of a shifted SL1 structure (**Fig. 4.3A** cluster 2 and **Fig. 4.4D**). In cell-based reporter assays, frameshifting efficiencies dropped to 1.2% demonstrating that cluster 2 as seen in the FSE sample does not induce frameshifting. This further reinforces the notion that cluster 1 detected in the FSE sample is the frameshift stimulatory fold (**Fig. 4.3A**).

The FSE-M3 RNA variant has two mutations in the 3' strand of SL2 (**Fig. 4.4A** and **4.S1)**. These mutations are designed to stabilize base pairing interactions between the 5' spacer and the 3' strand of SL2, thus promoting the folding of the 3WJ as detected in cluster 2 of FSE-V1 (**Fig. 4.3B**)(24). The majority of unfolding events occurred in three steps (4.8±0.5 pN, 8.0±2.1 pN, and 11.3±1.8 pN) accompanied with similar low force refolding with a total contour length of 69.7±3 nm, corresponding to the full unfolding of all 114 nt (**Fig. 4.4B**-**C, 4.S4G** and **Table 4.S2-3**). The unfolding step at low forces (4.8±0.5 pN) is most likely the opening of the base of the 3WJ (**Table 4.S2**). As expected, DMS-MaP data showed asymmetric changes in DMS reactivity at SL2, with increases in reactivity of the 5' strand of SL2, but the 3' strand of SL2 remaining unreactive (**Fig. 4.4D** and **4.S7**). As seen with the FSE-M1 and FSE-M2 mutants, frameshifting efficiencies dropped to 1.0% demonstrating that the 3WJ is not able to promote frameshifting (**Fig.4 4.4A**).

Another FSE mutant contained two substitutions in SL2 (FSE-M4), which was predicted to form an alternative PK structure (**Fig. 4.S8A** and **4.S1**)(24). However, our single molecule OT analysis did not indicate any PK-like unfolding trajectories (**Fig. 4.S8B-D**). DMS-MaP showed increased reactivities in SL2, but we did not observe any reactivity differences in SL1 and SL3 suggesting that these mutations result in the disruption of the PK fold, leaving SL1 and SL3 intact (**Fig. 4.S8E**). In line with that, FSE-M4 was not able to frameshift efficiently (1%) (**Fig. 4.S8A**).

Thus, we conclude that the correct folding of the pseudoknot is crucial for frameshifting and that alternative folds, such as the 3WJ, can compete with PK formation to decrease frameshifting efficiency.

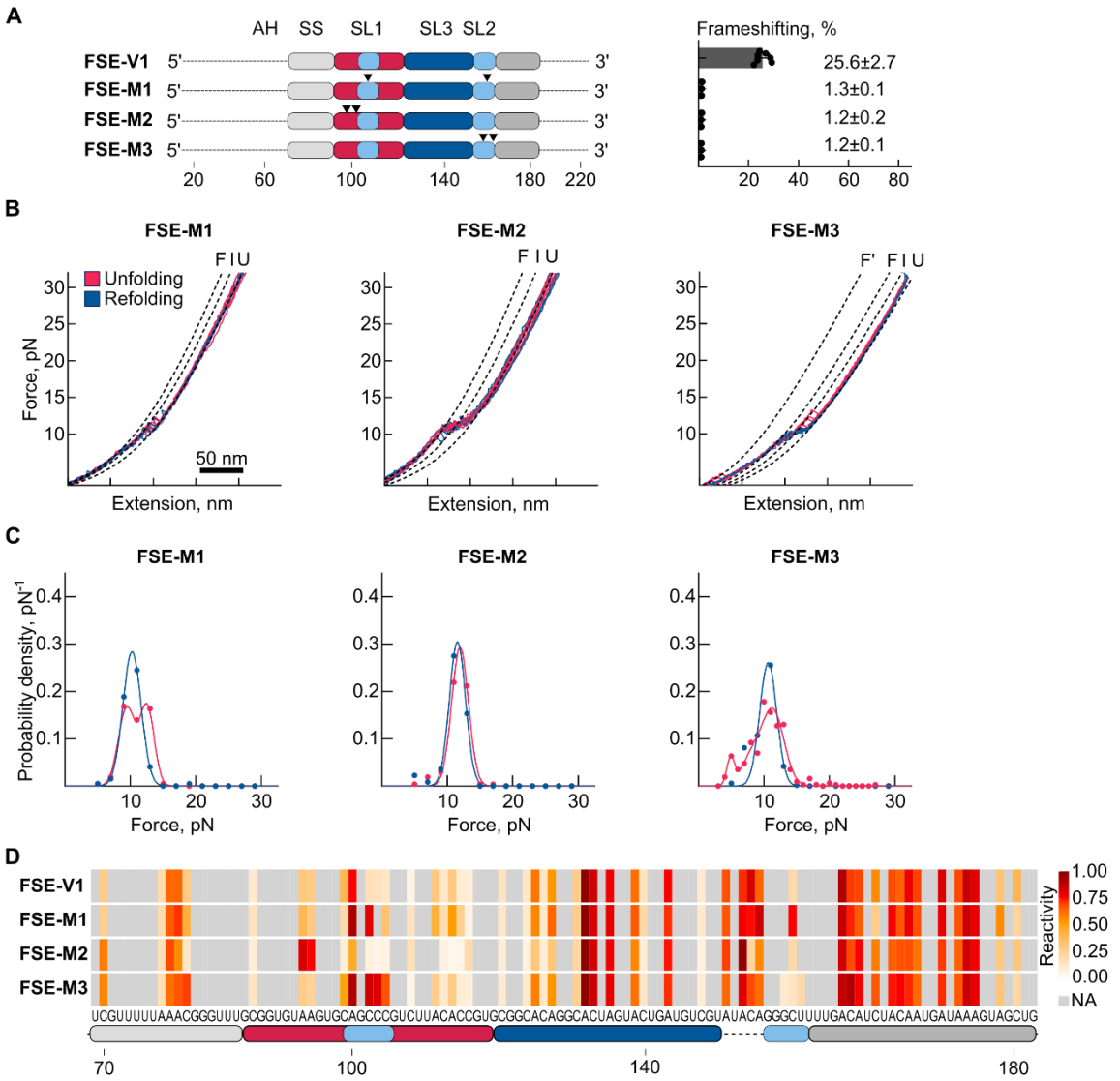

Figure 4.4 Effect of point mutations altering the conformation of the FSE. (**A**) Schematic representation of RNA mutants. Frameshift efficiencies of each RNA as measured by dual-fluorescence assay are plotted on the right. (**B**) Example unfolding (red) and refolding (blue) force-distance (FD) curves. “F” and “F’” denote different folded states, “I” the intermediate state, and “U” the fully unfolded state. (**C**) Histograms of force distribution for the unfolding (red) and refolding (blue) events observed in each RNA sample. (**D**) Reactivity profiles of the RNA mutants as determined by DMS-MaP.

Antisense oligonucleotides can alter FSE conformations

Informed by our analysis of the FSE structural folds, we next tested three antisense oligonucleotides (ASO) designed to alter the balance between PK and alternative folds (**Fig. 4.5A** and **4.S1**).

The first ASO (FSE-A1) hybridizes to the 3' strand of SL2 (position 145-158 relative to the FSE-V4) and thus impairs folding of the canonical PK (**Fig. 4.5A** and **4.3B** cluster 1). Here, we expected to observe (un)folding behavior comparable to the FSE-D2 and FSE-M1 samples, in which the PK forming SL2 was either deleted or mutated (**Fig. 4.4B-C** and **4.S4E**) (29). In agreement, in the presence of the FSE-A1 oligonucleotide about 90% of the (un)folding events occurred in two steps, at forces of 9.4±1.3 pN and 12.6±1.1 pN, showing a shift to low-stability hairpins (**Table 4.S2**). Furthermore, the total change in contour length of 32.7±5.1 nm was similar to the contour length change we observed with the FSE-M1 and FSE-D2 samples which cannot form SL2 (**Fig 4.4, 4.S4E, 4.S4H,** and **Table 4.S2**) (29). Based on these data, we conclude that the unfolding profile of FSE-A1 represents the sequential opening of SL1 and SL3, indicating the successful disruption of the canonical PK (**Fig. 4.5B-C** and **Tables 4.4S2-3**).

The second ASO (FSE-A2) binds to the 5' end of the PK including the SS and the spacer region (position 69-86), which is a few nucleotides downstream of the attenuator stem (**Fig. 4.5A** and **4.S1**). This ASO mimics the ribosome induced unfolding of the RNA 5' to the PK. This means that it would not disrupt folding of the PK directly but may influence alternative folds, for example by preventing base pairing at the base of the 3WJ. In this sample, we noted an increase in single-step unfolding events (from 55.5% in FSE-V1 to 82.7%), and the average force of unfolding was slightly increased to 14.9±2.7 pN, which implies the presence of more PK fold in this sample (**Fig. 4.5B-C**, **4.S4I** and **Tables 4.S2-3**). However, we did not observe the highly stable PK fold, presumably because the 5' immediate region cannot fold into the threaded form once it is hybridized to the ASO (**Fig. 4.5B-C**).

Lastly, we tested FSE-A3, which binds 10 nucleotides downstream of the FSE (position 166-182), a region that was not predicted to be part of any of the major conformers (**Fig. 4.5A**). Accordingly, the unfolding profile and the length of the unfolded region in single-molecule experiments were mostly unchanged compared to the FSE-V1 variant (**Fig. 4.5B-C**, and **Tables 4.S2-3**). In this sample, we observed a small increase in unfolding force (15.0±1.5 pN) and the number of single-step unfolding events as compared to FSE-V1 (65.2%) (**Fig. 4.5B**, **C** and **4.S4J**, and **Tables 4.S2-3**).

Overall, these data show that a preclusion of the base pairing of 5' spacer region and the 3' strand of the SL2 hampers the formation of the alternative folds, further suggesting that FSE can undergo structural rearrangements during translation, when the slippery nucleotides and the spacer region are occluded by the ribosome.

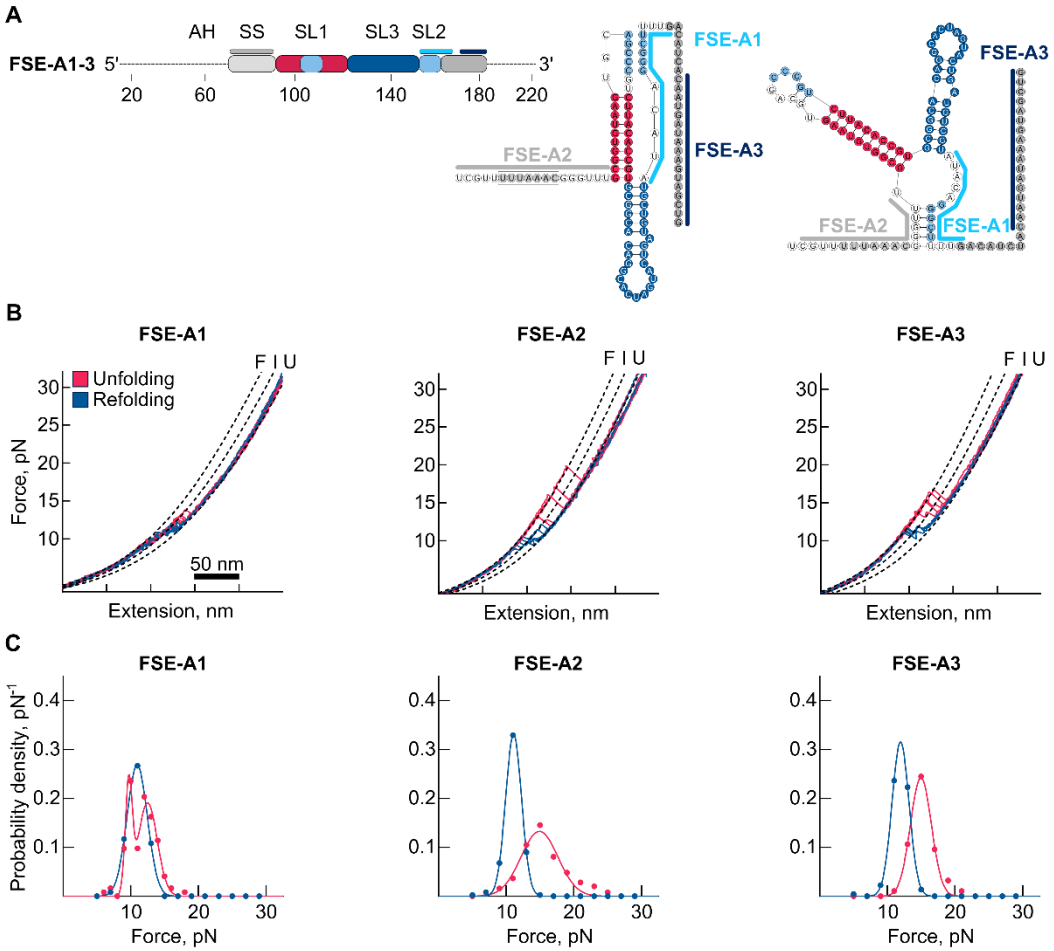

Figure 4.5 Antisense oligonucleotides alter the formation of conformers. (**A**) A schematic representation of the ASO samples. The oligonucleotide binding sites on the RNAs are depicted on the PK and 3WJ conformers. (**B**) Example unfolding (red) and refolding (blue) force-distance (FD) curves. "F" denotes the folded, "I" the intermediate, and "U" the fully unfolded states. (**C**) Histograms of force distribution of the unfolding (red) and refolding (blue) steps observed in each sample.

Antisense oligonucleotides can alter frameshifting efficiencies

To test whether ASOs can affect frameshifting, we employed in cell dual-fluorescence reporter assays and *in vitro* translation reporter assays (53,54). This strategy has been successfully employed to target untranslated regions (UTRs) and FSEs in SARS-CoV-2 and other viruses (13,55-58).

We first tested ASO FSE-A1, which impairs the formation of the PK (**Fig. 4.6A**). Accordingly, we observed a 20% decrease in frameshifting in the cell-based reporter assay (**Fig. 4.6A**). To probe for dose dependency of the effect observed with the FSE-A1 sample, we next used an *in vitro* frameshift reporter assay (**Fig. 4.6B**). In contrast to the cellular reporter assay, where we employed isolated frameshift site fragments, here the frameshift construct was derived from the SARS-CoV-2 genomic RNA corresponding to a 1.5 kb long

region (nucleotides 12686-14190) of the ORF1a/1b (29). Therefore, the *in vitro* SARS-CoV-2 frameshift reporter also enabled testing whether *cis*-RNA interactions within the 1.5 kB long region of SARS-CoV-2 affect frameshifting, albeit not excluding the possibility of other long-range base pairing of the FSE beyond this region. Translation of the *in vitro* frameshift reporter RNA results in a 31.8 kDa long peptide in the 0-reading frame, and a longer product of 59.2 kDa in the –1 frame (29). We observed a strong dose dependent decrease in frameshifting, ending up in complete inhibition of frameshifting at the 10:1 ratio (ASO:RNA) (**Fig. 4.6B** and **C**). On the other hand, with the non-targeting control oligonucleotide, we did not observe a difference in the relative amounts of the –1-, or 0-frame products. The difference in the targeting efficiency of FSE-A1 ASOs in cell-based versus *in vitro* reporter assays can be attributed to low transfection efficiencies, which may limit the cellular levels of ASOs to suboptimal concentrations.

Next, we aimed to test whether we could alter frameshifting efficiencies by targeting alternative conformers of the FSE. Since FSE-A2 hybridizes with the SS, which would interfere with ribosome progression and translation independent of frameshifting, we designed FSE-A4 to interfere with the formation and stability of alternative folds involving 3' downstream regions of the PK. The 3WJ conformer as detected by the DREEM analysis would be only partially targeted with this ASO, nevertheless one can expect some destabilization of the 3WJ (**Fig. 4.6A**). Yet, we detected no significant effect on frameshifting in the presence of the FSE-A4 (**Fig. 4.6B**).

To sum, our *in vitro* and cell-based reporter assays further corroborate that the PK is the critical fold for efficient frameshifting and can be effectively targeted by antisense strategies.

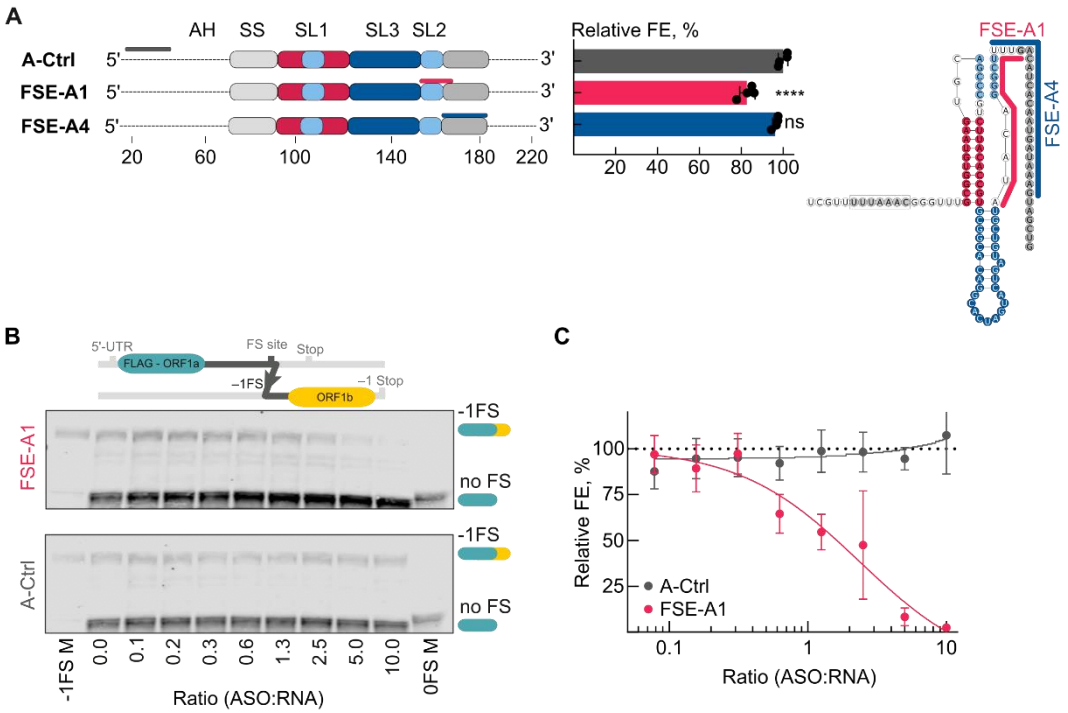

Figure 4.6 Targeting the FSE with antisense oligonucleotides. (**A**) Schematic representation of the binding sites of the ASO employed in the study. *In vivo* dual-fluorescence assay to evaluate the effect of ASOs on PRF in HEK293 cells. The relative frameshifting efficiency (FE) is calculated from the ratio of mCherry to EGFP intensities in the frameshift construct relative to the in-frame control construct lacking the PRF signal. Datapoints represent the mean ± s.d. (n = 4 independent experiments). *P*-values were derived by an ordinary unpaired one-sided ANOVA comparing every ASO to the scrambled FSE-A-Ctrl for each construct. **P* < 0.05, *****P* < 0.0001. (**B**) Upper panel: Scheme of the SARS-CoV-2 mRNA construct used for *in vitro* translation experiments. The N-terminal FLAG-tagged frameshifting reporter construct includes the nucleotides 12686-14190 (≈1.5 kb) of the SARS-CoV-2 genome. Representative western blots ASO titrations in RRL-based *in vitro* translation using the FSE-A1 (upper) or the control (A-Ctrl) oligonucleotide. Frameshift constructs encode for an N-terminal FLAG-tagged version of the 1a-1b polyprotein. 0-frame frameshift product: 32 kDa, –1-frame product: 58 kDa. (**C**) Graph presenting the densitometric quantification of the western blot analysis in (**B**). Datapoints represent the mean ± s.d. (n = 3 independent experiments); IC50 (FSE-A1) = 1.8; R^2 = 0.94. For additional controls, please refer to **Figure 4.S2**.

4.5. Discussion

Due to its crucial role in the viral life cycle and replication, the FSE of SARS-CoV-2 has been extensively studied by functional and structural approaches (12,13,15-19,21,22,28). These studies imply that FSE folding varies substantially depending on whether the FSE RNA is investigated in isolation or in the presence of 5' and 3' extensions that more closely mimic the genomic context. As a result, there is a lack of consensus on

the existence of alternative conformations of the FSE, and their functional relevance to frameshifting.

Here, we used single molecule optical tweezers and DMS-MaP structural probing to investigate the folding and unfolding dynamics and steady-state RNA conformers of SARS-CoV-2 FSE variants. Collectively, our results show that the PK is not the only conformer formed at the frameshift site. Instead, immediate 5' upstream and 3' downstream regions of the FSE promote structural transitions of the FSE. In the presence of these regions, the frameshift stimulatory PK fold co-exists with alternative conformers (**Fig. 4.7A**). Nevertheless, mutagenesis of the FSE to disrupt the PK or to favor formation of the alternative folds convincingly demonstrate that the canonical PK is the only conformation driving efficient frameshifting. Recently, an alternative PK was proposed to form in the presence of 5' extensions, due to the interaction between the SL1 loop and the GGG nucleotides found at the spacer (21,23). Although our chemical probing results cannot definitively exclude the alternative PK fold, our pulling experiments did not detect the steps corresponding to the alternative PK suggesting it is not present in our experimental setup.

Clustering of DMS-MaP reactivities by DREEM analysis detected length-dependent alternative folds at the SARS-CoV-2 frameshift site. In the shorter FSE, FSE-V1 and FSE-V2 variants we detected two distinct conformations. One of the conformations is consistent with the canonical PK fold, whose relative abundance closely matched with the frameshift efficiencies measured in cell-based frameshift assays. The second conformation, which we presumed to be non-frameshift competent one, contained either a shifted SL1 (FSE), a 3WJ (FSE-V1), or a closed SL1 loop that may preclude PK formation (FSE-V2). Strikingly, in the longer FSE-V3 and FSE-V4 variants DREEM analysis detected a single non-PK conformation containing AS1, yet these variants supported higher levels of frameshifting when compared to FSE-V1 and FSE-V2. One possible explanation is that during translation, RNA structures would be unfolded by the helicase activity of the ribosomes allowing the RNA to resample into alternative conformations. Specifically, as the ribosome moves over the 5' portion of the AS1 stem, it would liberate 3' part of the stem allowing the folding of SL1, and eventually the formation of the canonical PK fold (**Fig. 4.7B**). In contrast, structures containing the shifted SL1 cannot frameshift effectively, because helicase disruption of the shifted SL1 would preclude its refolding into the canonical SL1 of the PK. The 3WJ could in principle fold into a PK when base pairing at the slippery sequence is disrupted, as shown by our single molecule pulling data in FSE-A2. On the other hand, in the context of translation, the proximity of the 3WJ to SL1 may reduce the likelihood of the 3WJ refolding into the PK. This can be either due to steric inhibition, insufficient refolding time before the ribosome

hits the SL1, or because the SL1 is kinetically trapped and cannot readily refold into the PK conformation. This model would mechanistically explain how high frameshifting efficiencies can be achieved even though the PK is not identified as the dominant structural conformation in genome based structural probing studies (15,22,28). Whilst the PK fold was observed in Cryo-EM structures of SARS-CoV-2 FSE stalled ribosomes (12), these ribosomes were trapped over the SS using a stop codon in place of the second slippery codon (U_UUA_AAC to U_UUA_UAA). Consequently, the 5' nucleotides occluded by the ribosome would not be base pair into the alternative stem (AS) or three-way junction (3WJ). We therefore predict that ribosomes trapped before the slippery sequence would allow detection of alternative RNA structures, such as the 3WJ conformer, which can be important for fine-tuning SARS-CoV-2 frameshifting for optimal replication.

We also suggest that the presence of less-stable structures can be a strategy employed by RNA viruses to avoid logjams during the replication of their genomes (59,60). In coronaviruses, during the synthesis of the negative strand of the genome, the RNA would be unwound from its 3' end by the RNA-dependent RNA polymerase, which would first unfold the SL2 before reaching the AS1. This would preclude the formation of the PK structure and allow genome replication without hindrance (10,24,61-63). Thus, less stable folds might be preferred to ensure optimal speed of genome replication at different stages of infection.

Finally, we show that ASO targeting of the SL2 prevents the formation of the canonical PK, thus decreasing frameshifting both *in vitro* and in cells. Nevertheless, the conformational landscape that we identified has implications for therapeutic interventions. So far, efforts to target FSE have mainly focused on the isolated PK element. However, if it is formed transiently during translation or at different stages of the infection cycle, the PK may not be the most abundant structure in the cell. This would impact the efficiency of small molecule targeting. In addition, (de)stabilization of the alternative conformers may lead to a shift in the possible RNA folds, which may ultimately impact frameshifting and reduce viral replication. Knowledge of functionally relevant alternative conformers of viral frameshift elements could therefore be targeted to increase the efficiency of viral RNA targeting. Although our work provides mechanistic insights on structural transitions of FSE variants under defined conditions, future investigations will also be needed to understand the potential role of RNA-binding proteins and longer range cis-interactions on the structure and function of frameshift elements in the context of viral infection in cells.

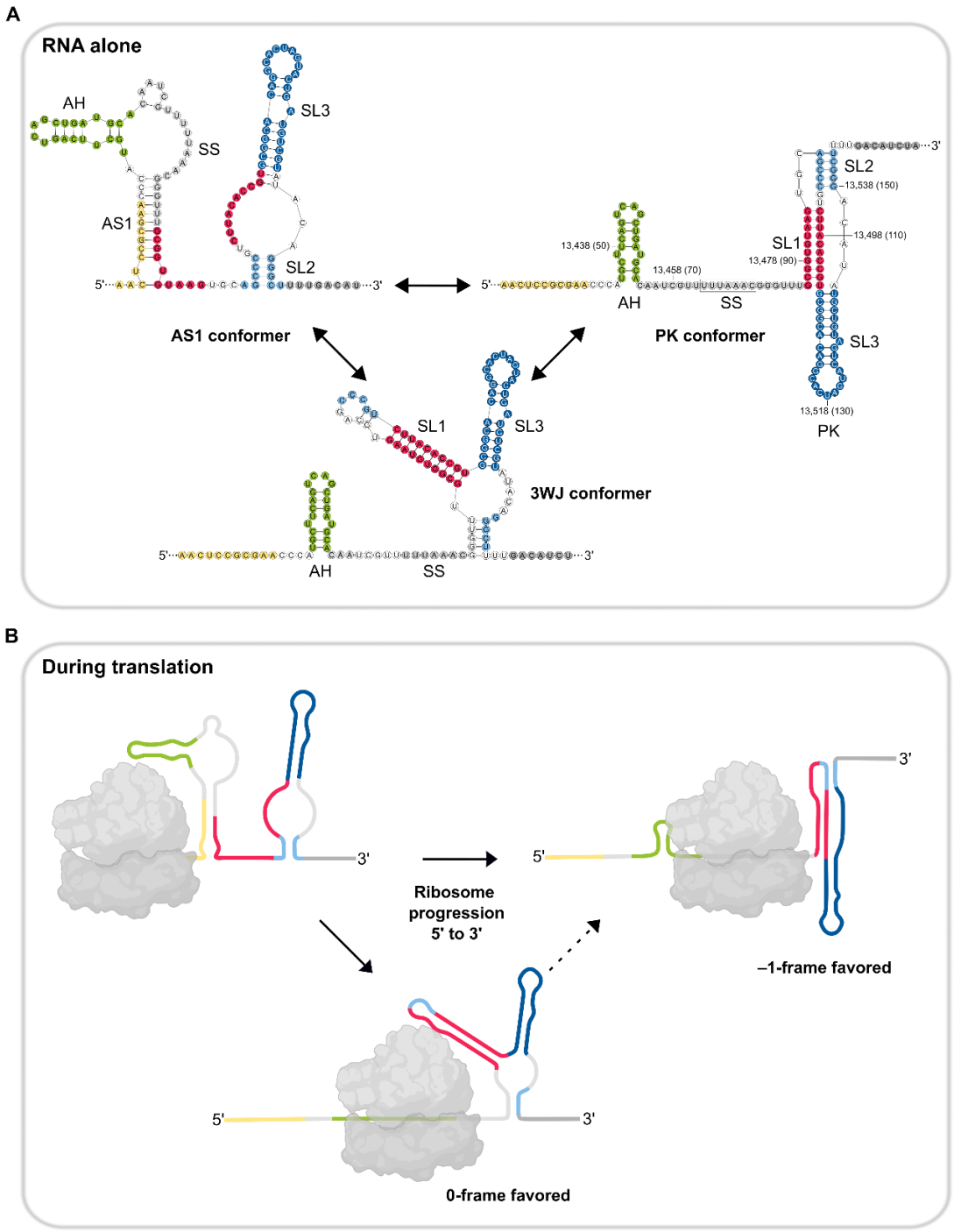

Figure 4.7 Proposed model of PRF site conformational transitions as a translation regulatory switch. (**A**) Transition between the alternative conformation, the canonical pseudoknot and the three-way junction is achieved by a conformational switch between SL1, AS1 and SL2. (**B**) In cells, the viral RNA undergoes translation, replication/transcription or virion packaging depending on its localization and phase of infection. The viral RNA is trapped in an intermediate conformation. During translation, as the ribosome progresses, the AS1 of the intermediate conformer would be unwound by the helicase activity of the ribosome, thus allowing SL1 to form. This could result in formation of either the frameshift stimulatory PK or the three-way junction. As translation progresses, the three-way junction may also fold into the PK.

4.6. Data Availability

Data supporting this study is available at 10.5281/zenodo.6626934.

4.7. Code Availability

Custom scripts were employed to process optical tweezers data. The python algorithm called Practical Optical Tweezers Analysis TOol is available on Github (POTATO, https:// github.com/REMI-HIRI/POTATO.git)(35).

4.8. Author Contributions

LP and SB designed and cloned OT constructs and LP prepared RNA samples and performed most of single-molecule OT experiments and LP and SB analyzed data under supervision of NC. MZ cloned frameshift reporter constructs and performed translation assays. ASGB performed the chemical probing experiments and prepared libraries for DMS-MaP experiments under supervision of RS and RS analyzed the data and performed DREEM analysis. NC, RS and LP drafted, wrote and edited the paper with the contributions from all authors.

4.9. Acknowledgments

We thank Tatyana Koch for expert technical assistance. We thank Dr. Anke Sparmann and members of Caliskan group for critical reading of the manuscript. We thank Patrick Bohn for useful discussions on DMS-MaP experiments.

4.10. Funding

The work in our laboratories is supported by the Helmholtz Association and grants from the European Research Council (ERC) Grant Nr. 948636 (to NC) and Helmholtz Young Investigator Grant VH-NG-1347 (to RS).

4.11. Supplementary material

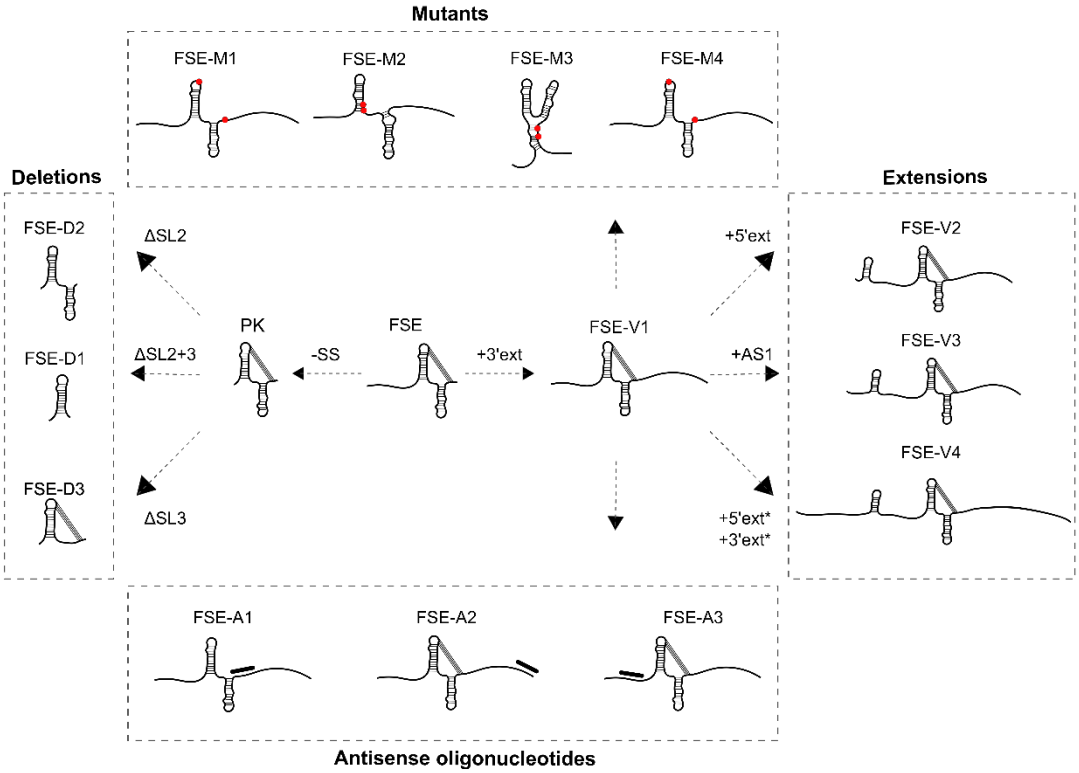

Figure 4.S1 Structural landscape of the SARS-CoV-2 FSE RNA. Schematic depiction of the structural relationship between different RNA variants employed in this study. Related to **Figure 4.1**.

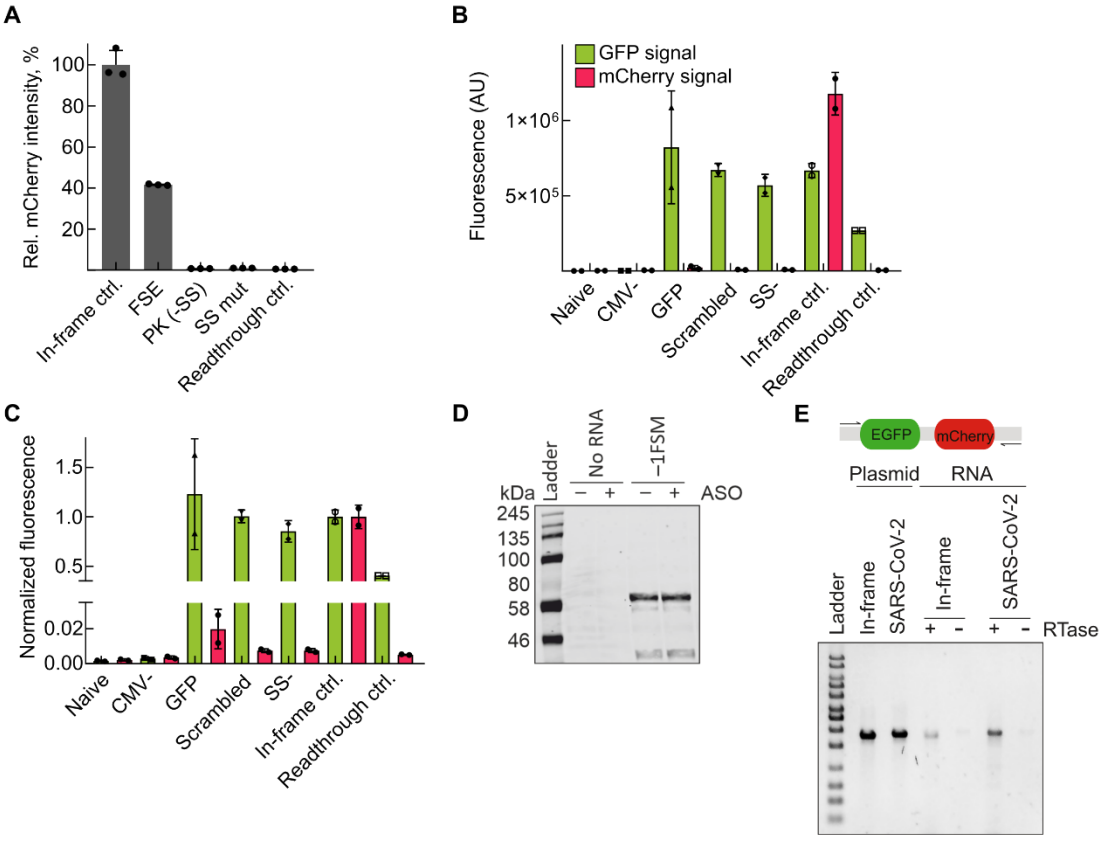

Figure 4.S2 Control experiments of dual-fluorescence assay. (**A**) relative mCherry intensity (analogical to frameshifting efficiency) of selected RNA variants and control samples. FSE contains slippery sequence, spacer and pseudoknot sequence of SARS-CoV-2 RNA. PK(-SS) contains only pseudoknot sequence. SS mut contains mutation in slippery sequence that prevents the frameshifting. Readthrough control contains mCherry gene in 0-frame, thus only upon readthrough event mCherry would be expressed. (**B**) Raw fluorescence values from flow-cytometer for different cell samples. Cells were transfected with no vector (**naïve**), vector lacking CMV promoter (**CMV-**), vector containing only the GFP gene (**GFP**), vector containing GFP in 0-frame, mCherry in –1-frame and the SARS-CoV-2 FSE was scrambled (**scrambled**), vector lacking slippery sequence (**SS-**), vector containing GFP and mCherry genes both in 0-frame without the stop codon in between (**In-frame ctrl.**), vector containing GFP and mCherry genes both in 0-frame with the stop codon in between (**Readthrough ctrl.**). (**C**) Same as B but the values are normalized to the In-frame control. (**D**) In vitro translation experiments in RRL were performed in the presence of –1FS M mRNA, with the mutated frameshift stimulatory element encoding a segment of the in-frame ORF1a-1b. As the control, a non-targeting oligonucleotide was used ("– ASO"). (**E**) Potential splicing of the dual-fluorescence reporter. Schematic depiction of dual-fluorescence reporter construct and position of primer binding sites for PCR amplification of the reporter plasmids and cDNA (**up**). Products of PCR amplification of the in-frame and SARS-CoV-2 frameshift RNA reporter plasmids and oligo(dT)-primed cDNA from cells transfected with the same reporter plasmids are analyzed by using agarose gel electrophoresis (**bottom**).

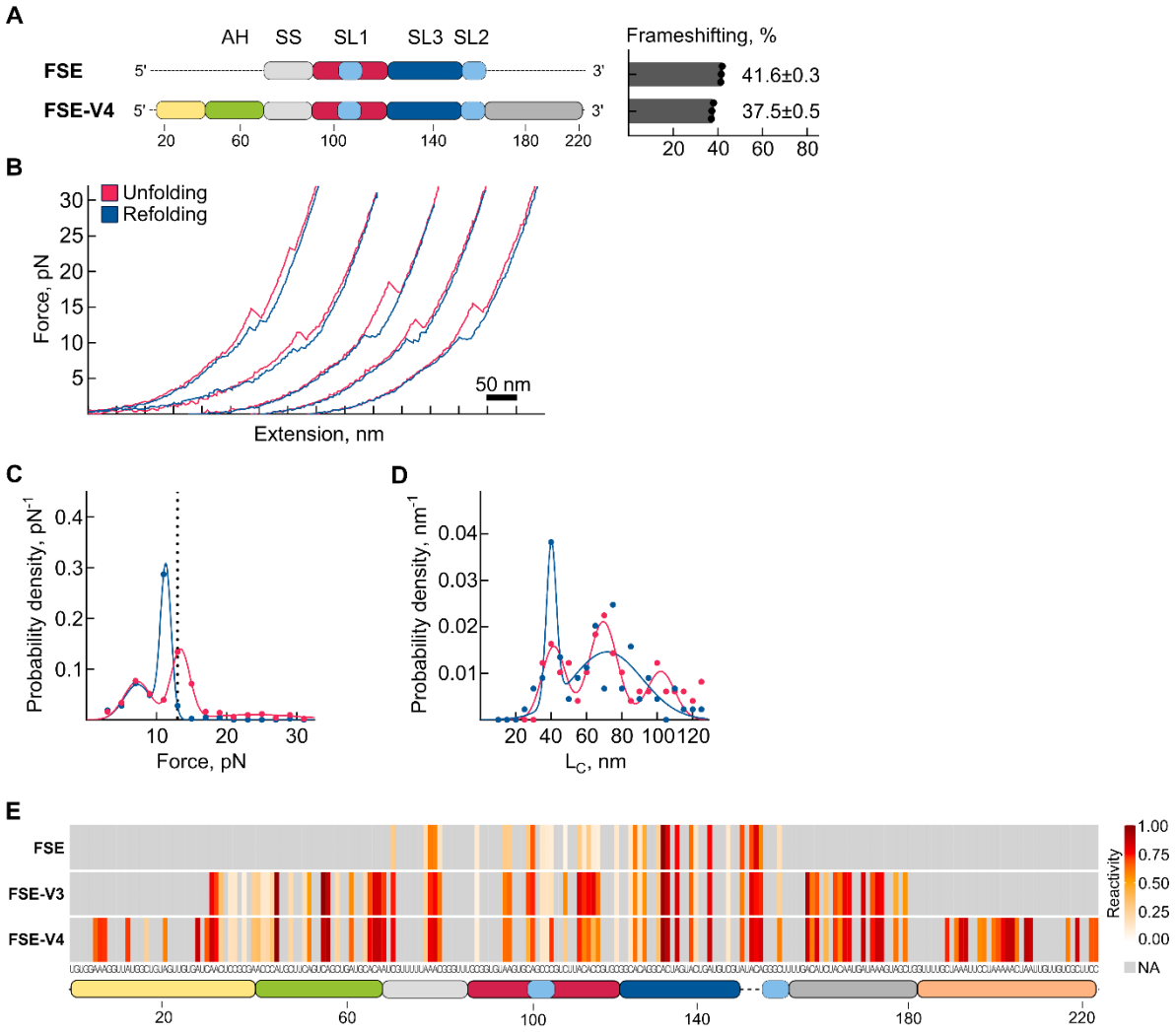

Figure 4.S3 FSE-V4 functional and single-molecule data. (**A**) Schematic representation of RNA variants with respective parts of the FSE employed in the study. Frameshift efficiencies of the variants are shown on the right. (**B**) Example unfolding (red) and refolding (blue) force-distance (FD) curves. The FSE-V4 trajectories exhibit shouldering in the low force regime (<10 pN) resulting in bias of the subsequent fitting. (**C**) Force distribution of the unfolding (red) and refolding (blue) steps observed for the FSE-V4. (**D**) Total contour length distribution of the unfolding (red) and refolding (blue) steps observed for the FSE-V4. (**E**) Reactivity profiles of the RNA variants as determined by DMS-MaP. Related to **Figure 4.2**.

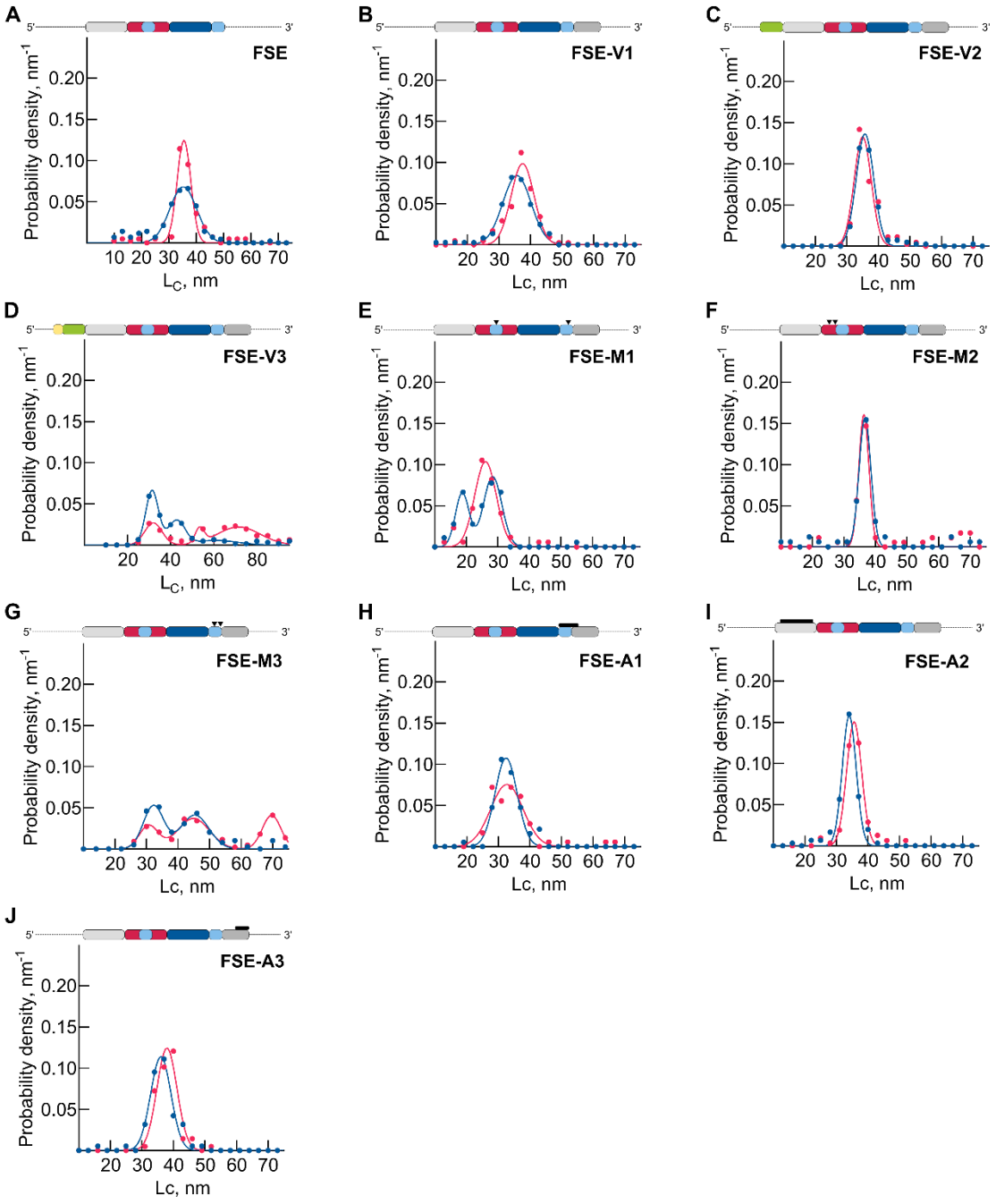

Figure 4.S4 Distributions of (un)folding contour length changes for different RNA variants. (**A**) FSE. (**B**) FSE-V1. (**C**) FSE-V2. (**D**) FSE-V3, (**E**) FSE-M1 (**F**) FSE-M2. (**G**) FSE-M3 (**H**) FSE-A1. (**I**) FSE-A2. (**J**) FSE-A3. Unfolding distributions are in red and refolding are shown in blue. Related to **Figures 4.2-5**.

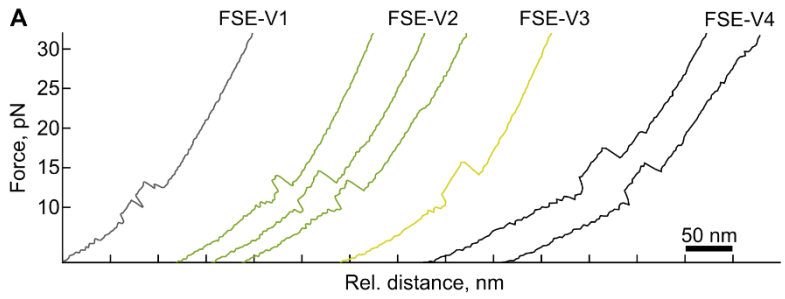

Figure 4.S5 Examples of folding rescue events. (**A**) FD curves showing the folding rescue (indicated by the black arrow) during unfolding for different samples. Related to **Figure 4.2**.

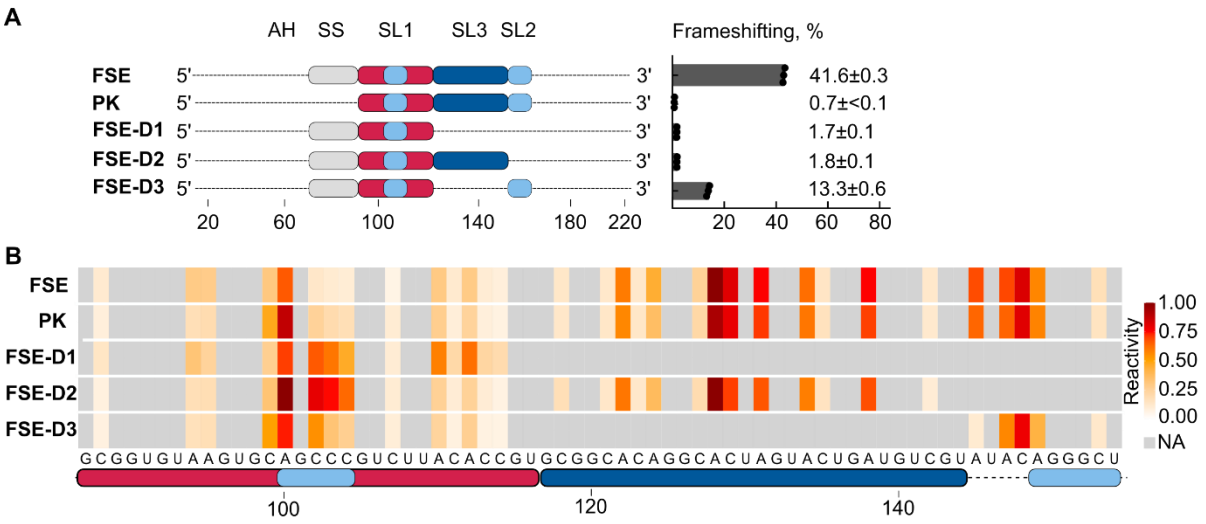

Figure 4.S6 Determination of minimal FS motif. (**A**) A schematic representation of the RNA variants employed with depicted regions of the SARS-CoV-2 FSE and surrounding genomic regions. Frameshifting efficiency (FE) of each RNA as measured by dual-fluorescence assay are indicated at the right. FSE-D1-3 in DF assay contained the slipper sequence. (**B**) Reactivity profiles of the RNA variants as determined by DMS-MaP. Frameshifting values are taken from Zimmer et al., 2021.

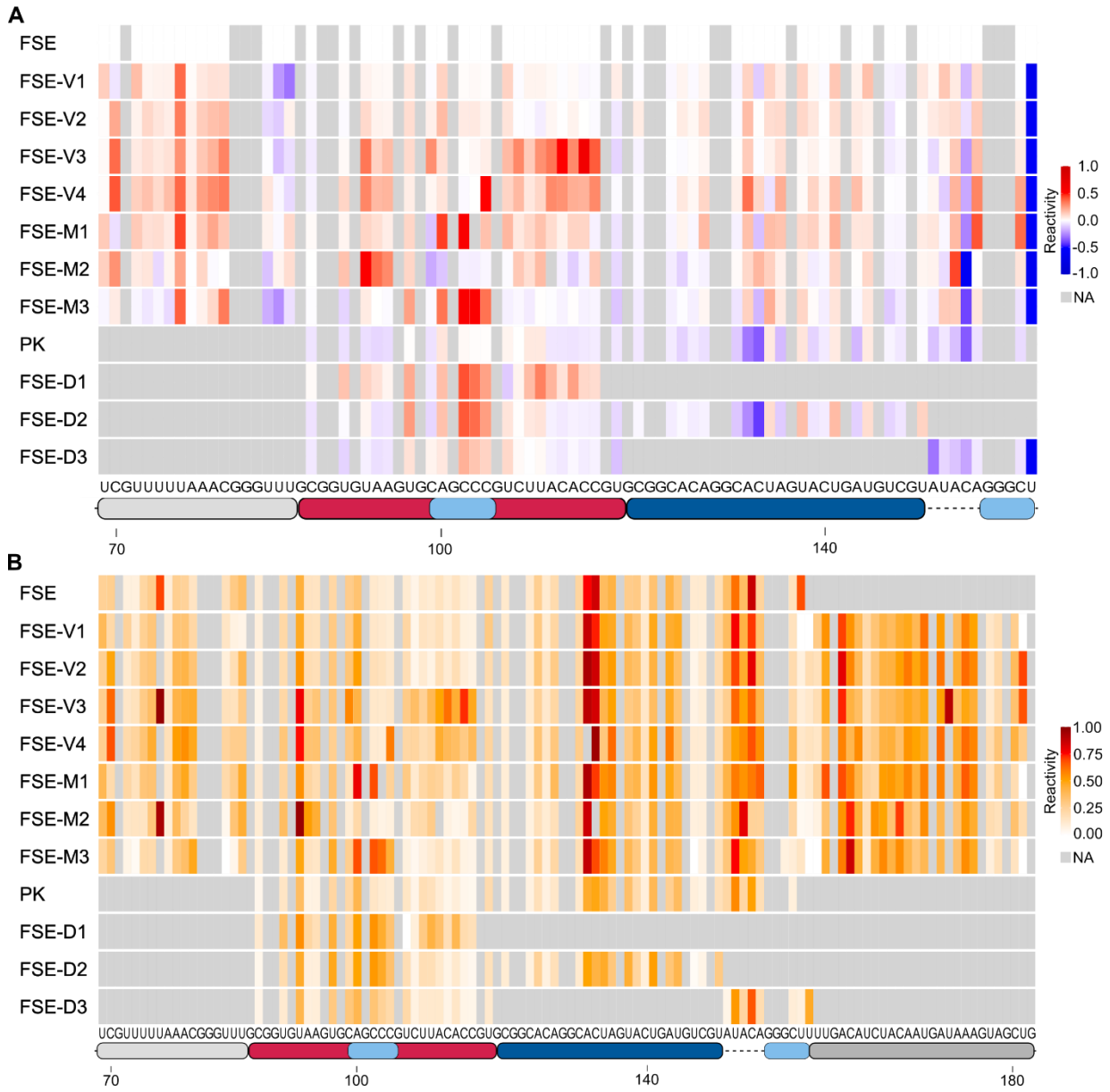

Figure 4.S7 Additional DMS-MaP analysis. (A) Reactivity profiles of the RNA variants as determined by DMS-MaP with U reactivities included. Reactivities of bases were normalized individually within the same base type. **(B)** Differential reactivity profiles of the RNA variants. The values were obtained by subtracting the reactivities of a given base in the FSE sample from the reactivity of a given base in other RNA variants. Related to **Figures 4.2-4**.

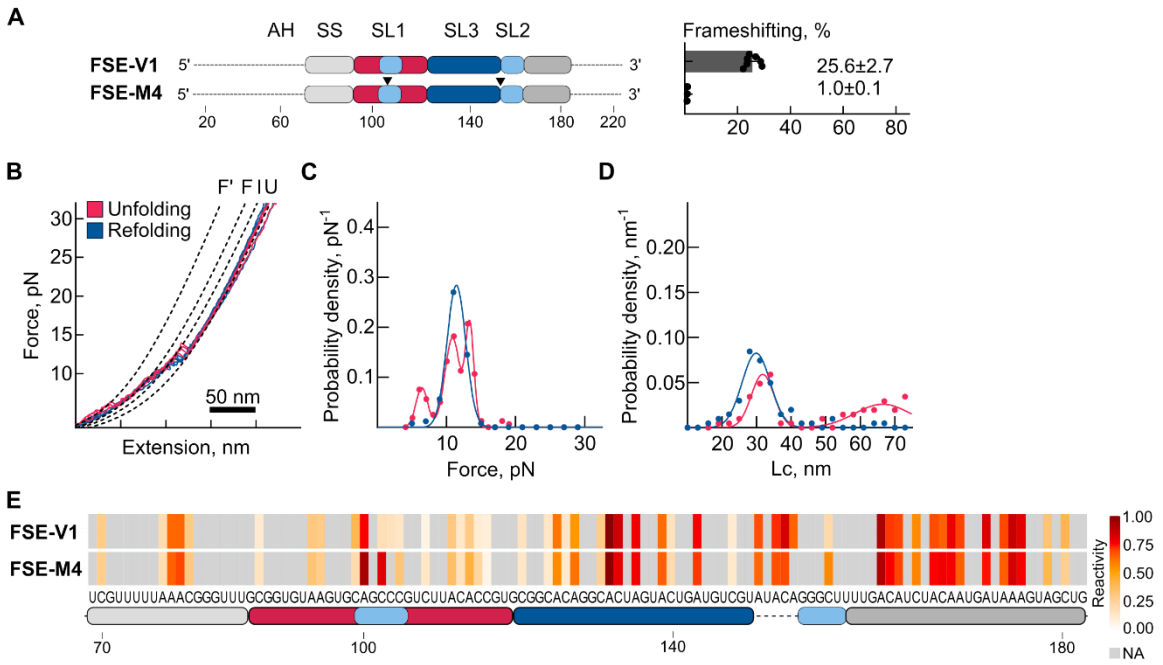

Figure 4.S8 Point mutant stabilizing an alternative pseudoknot. (**A**) Schematic representation of RNA mutants. Frameshift efficiencies of each RNA as measured by the dual-fluorescence assay are plotted at the right. (**B**) Example unfolding (red) and refolding (blue) force-distance (FD) curves. (**C**) Force distribution of the unfolding (red) and refolding (blue) steps observed for the FSE-V4. (**D**) Total contour length distribution of the unfolding (red) and refolding (blue) steps observed for the FSE-V4. (**E**) Reactivity profiles of the RNA variants as determined by DMS-MaP. Related to **Figure 4.2**.

Supplementary Tables

Table 4.S1 Sequences employed to synthesize the RNA variants employed in this study. Nucleotides in small letters indicate the mutations. Slippery site is underlined.

RNA variant	Length	RNA sequence (5'>3')
FSE	86	UCGUU<u>UUUAAAC</u>GGGUUUGCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGCU
FSE-V1	114	UCGUU<u>UUUAAAC</u>GGGUUUGCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGCUUUUGACAUCUACAAUGAUAAAGUAGCUG
FSE-V2	141	CCCAUGCUUCAGUCAGCUGAUGCACAAUCGUU<u>UUUAAAC</u>GGGUUUGCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGCUUUUGACAUCUACAAUGAUAAAGUAGCUG
FSE-V3	152	AACUCCGCGAACCCAUGCUUCAGUCAGCUGAUGCACAAUCGUU<u>UUUAAAC</u>GGGUUUGCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGCUUUUGACAUCUACAAUGAUAAAGUAGCUG
FSE-V4	221	UGUGGAAAGGUUAUGGCUGUAGUUGUGAUCAACUCCGCGAACCCAUGCUUCAGUCAGCUGAUGCACAAUCGUU<u>UUUAAAC</u>GGGUUUGCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGCUUUUGACAUCUACAAUGAUAAAGUAGCUGGUUUUGCUAAAUUCCUAAAAACUAAUUGUUGUCGCUUCC
FSE-M1	114	UCGUU<u>UUUAAAC</u>GGGUUUGCGGUGUAAGUGCAGCgCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGaUUUUGACAUCUACAAUGAUAAAGUAGCUG
FSE-M2	114	UCGUU<u>UUUAAAC</u>GGGUUUGCGGUGUAAGUGCAGCCCGUCUUgCcCCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGCUUUUGACAUCUACAAUGAUAAAGUAGCUG
FSE-M3	114	UCGUU<u>UUUAAAC</u>GGGUUUGCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGcCcUUUGACAUCUACAAUGAUAAAGUAGCUG
FSE-M4	114	TCGTT<u>TTTAAAC</u>GGGTTTGCGGTGTAAGTGCAGCgCGTCTTACACCGTGCGGCACAGGCACTAGTACTGATGTCGTATACtGGGCTTTTGACATCTACAATGATAAAGTAGCTG
PK	67	GCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGC
FSE-D1	30	GCGGUGUAAGUGCAGCCCGUCUUACACCGU
FSE-D2	58	GCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGU
FSE-D3	41	GCGGUGUAAGUGCAGCCCGUCUUACACCGUAUACAGGGCUU

Table 4.S2 Average calculated force values and expected contour length values of different RNA variants employed in this study. The force and contour length peaks do not always correspond. Related to **Figures 4.2-5**.

RNA variant	Direction	Peak	Force, pN	Contour length, nm	Expected contour length, nm	FD curves / molecules
FSE	Unfolding	1	11.9±1.8	35.5±2.5	37.2	287/16
		2	20.1±4.9			
	Refolding	1	11.4±1.6	35.3±4.9		
FSE-V1	Unfolding	1	13.6±1.6	37.5±3.6	37.2	260/17
	Refolding	1	11.3±0.9	35.7±4.5		
FSE-V2	Unfolding	1	12.6±2.1	35.0±2.7	37.2	288/20
	Refolding	1	10.9±1.2	35.7±2.8		
FSE-V3	Unfolding	1	5.4±1.1	32.0±4.1	-	244/6
		2	12.7±3.1	53.5±2.1		
		3	-	71.6±10.0		
	Refolding	1	10.4±1.5	31.4±3.0		
		2	-	42.7±3.9		
		3	-	61.2±9.9		
FSE-V4	Unfolding	1	7.3±1.9	41.5±6.6	-	187/6
		2	13.4±1.3	69.5±8.0		
		3	24.0±6.6	102.0±7.8		
	Refolding	1	7.3±2.0	40.1±2.8		
		2	11.3±0.8	71.7±19.8		
FSE-M1	Unfolding	1	9.4±1.3	26.1±3.4	31.9	118/13
		2	12.6±1.1	-		
	Refolding	1	10.3±1.4	18.7±2.2		
		2	-	28.4±2.9		
FSE-M2	Unfolding	1	11.9±1.3	36.3±1.6	37.2	112/8
	Refolding	1	11.6±1.2	36.7±1.9		
FSE-M3	Unfolding	1	4.8±0.5	30.7±3.0	65.0	208/17
		2	8.0±2.1	44.9±5.1		

		3	11.3±1.8	69.7±3.0		
	Refolding	1	10.6±1.3	32.2±3.2		
		2	-	45.3±4.5		
FSE-A1	Unfolding	1	9.7±0.5	32.7±5.1	31.9	123/10
		2	12.5±1.4	-		
	Refolding	1	11.0±1.5	32.4±3.5		
FSE-A2	Unfolding	1	14.9±2.7	35.6±2.4	37.2	206/16
	Refolding	1	11.1±1.2	34.0±2.2		
FSE-A3	Unfolding	1	15.0±1.5	37.9±3.2	37.2	132/11
	Refolding	1	11.9±1.3	36.0±3.2		
	Refolding	1	11.5±1.0	21.9±2.7		

Table 4.S3 Step assignments and the percentage of (un)folding events observed at each step. Related to **Figures 4.2-5**.

RNA variant		Step #, %				
		0	1	2	3	>3
FSE	Unfolding	2.1	72.2	20.8	4.9	-
	Refolding	1.4	26.1	69.0	3.5	-
FSE-V1	Unfolding	-	55.5	40.1	4.4	-
	Refolding	1.6	17.9	74.8	5.7	-
FSE-V2	Unfolding	-	48.6	49.3	2.0	-
	Refolding	0.7	7.9	85.7	5.7	-
FSE-V3	Unfolding	-	19.7	54.9	21.3	4.1
	Refolding	-	12.3	65.6	20.5	1.6
FSE-V4	Unfolding	-	14.3	33.7	31.6	20.4
	Refolding	-	15.7	36.0	40.4	7.9
FSE-M1	Unfolding	-	19.3	80.7	-	-
	Refolding	1.6	37.7	59.0	1.6	-
FSE-M2	Unfolding	-	5.1	69.5	25.4	-
	Refolding	-	5.7	83.0	7.5	3.8
FSE-M3	Unfolding	-	-	28.2	58.2	13.6
	Refolding	-	1.0	44.9	49.0	5.1
FSE-A1	Unfolding	-	5.0	86.7	6.7	1.7
	Refolding	-	9.5	90.5	-	-

FSE-A2	Unfolding	-	82.7	15.4	1.9	-
	Refolding	2.0	15.7	82.4	-	-
FSE-A3	Unfolding	-	65.2	33.3	1.4	-
	Refolding	-	27.0	71.4	1.6	-

4.12. References

1. Riegger, R.J. and Caliskan, N. (2022) Thinking Outside the Frame: Impacting Genomes Capacity by Programmed Ribosomal Frameshifting. *Front. Mol. Biosci.*, **9**, 842261.
2. Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E. and Baranov, P.V. (2016) Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res*, **44**, 7007-7078.
3. Firth, A.E. and Brierley, I. (2012) Non-canonical translation in RNA viruses. *J. Gen Virol*, **93**, 1385-1409.
4. Schlub, T.E. and Holmes, E.C. (2020) Properties and abundance of overlapping genes in viruses. *Virus Evol*, **6**, veaa009.
5. Dinman, J.D. (2012) Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip. Rev. RNA*, **3**, 661-673.
6. Rodnina, M.V., Korniy, N., Klimova, M., Karki, P., Peng, B.-Z., Senyushkina, T., Belardinelli, R., Maracci, C., Wohlgemuth, I., Samatova, E. *et al.* (2020) Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Res*, **48**, 1056-1067.
7. Caliskan, N., Peske, F. and Rodnina, M.V. (2015) Changed in translation: mRNA recoding by −1 programmed ribosomal frameshifting. *Trends Biochem. Sci.*, **40**, 265-274.
8. Kelly, J.A., Woodside, M.T. and Dinman, J.D. (2021) Programmed −1 Ribosomal Frameshifting in coronaviruses: A therapeutic target. *Virology*, **554**, 75-82.
9. Brierley, I. and Dos Ramos, F.J. (2006) Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res*, **119**, 29-42.
10. Caliskan, N., Katunin, Vladimir I., Belardinelli, R., Peske, F. and Rodnina, Marina V. (2014) Programmed –1 Frameshifting by Kinetic Partitioning during Impeded Translocation. *Cell*, **157**, 1619-1631.
11. Kim, H.-K., Liu, F., Fei, J., Bustamante, C., Gonzalez, R.L. and Tinoco, I. (2014) A frameshifting stimulatory stem loop destabilizes the hybrid state and impedes ribosomal translocation. *Proc Natl Acad Sci U S A*, **111**, 5538-5543.
12. Bhatt, P.R., Scaiola, A., Loughran, G., Leibundgut, M., Kratzel, A., Meurs, R., Dreos, R., O'Connor, K.M., McMillan, A., Bode, J.W. *et al.* (2021) Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science*, **372**, 1306-1313.
13. Zhang, K., Zheludev, I.N., Hagey, R.J., Haslecker, R., Hou, Y.J., Kretsch, R., Pintilie, G.D., Rangan, R., Kladwang, W., Li, S. *et al.* (2021) Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nat. Struct. Mol. Biol*, **28**, 747-754.
14. Wacker, A., Weigand, J.E., Akabayov, S.R., Altincekic, N., Bains, J.K., Banijamali, E., Binas, O., Castillo-Martinez, J., Cetiner, E., Ceylan, B. *et al.* (2020) Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res*, **48**, 12415-12435.
15. Lan, T.C.T., Allan, M.F., Malsick, L.E., Woo, J.Z., Zhu, C., Zhang, F., Khandwala, S., Nyeo, S.S.Y., Sun, Y., Guo, J.U. *et al.* (2022) Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat. Commun.*, **13**, 1128.
16. Manfredonia, I., Nithin, C., Ponce-Salvatierra, A., Ghosh, P., Wirecki, T.K., Marinus, T., Ogando, N.S., Snijder, E.J., van Hemert, M.J., Bujnicki, J.M. *et al.* (2020) Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res*, **48**, 12436-12452.
17. Sun, L., Li, P., Ju, X., Rao, J., Huang, W., Ren, L., Zhang, S., Xiong, T., Xu, K., Zhou, X. *et al.* (2021) In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. *Cell*, **184**, 1865-1883.e1820.

18. Roman, C., Lewicka, A., Koirala, D., Li, N.-S. and Piccirilli, J.A. (2021) The SARS-CoV-2 Programmed −1 Ribosomal Frameshifting Element Crystal Structure Solved to 2.09 Å Using Chaperone-Assisted RNA Crystallography. *ACS Chem. Biol.*, **16**, 1469-1481.
19. Neupane, K., Zhao, M., Lyons, A., Munshi, S., Ileperuma, S.M., Ritchie, D.B., Hoffer, N.Q., Narayan, A. and Woodside, M.T. (2021) Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers. *Nat. Commun.*, **12**, 4749.
20. Ziv, O., Price, J., Shalamova, L., Kamenova, T., Goodfellow, I., Weber, F. and Miska, E.A. (2020) The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol. Cell*, **80**, 1067-1077.e1065.
21. Schlick, T., Zhu, Q., Dey, A., Jain, S., Yan, S. and Laederach, A. (2021) To Knot or Not to Knot: Multiple Conformations of the SARS-CoV-2 Frameshifting RNA Element. *J. Am. Chem. Soc.*, **143**, 11404-11422.
22. Huston, N.C., Wan, H., Strine, M.S., de Cesaris Araujo Tavares, R., Wilen, C.B. and Pyle, A.M. (2021) Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol Cell*, **81**, 584-598.e585.
23. Yan, S., Zhu, Q., Jain, S. and Schlick, T. (2022) Length-dependent motions of SARS-CoV-2 frameshifting RNA pseudoknot and alternative conformations suggest avenues for frameshifting suppression. *Nat. Commun.*, **13**, 4284.
24. Schlick, T., Zhu, Q., Jain, S. and Yan, S. (2021) Structure-altering mutations of the SARS-CoV-2 frameshifting RNA element. *Biophys. J.*, **120**, 1040-1053.
25. Su, M.-C., Chang, C.-T., Chu, C.-H., Tsai, C.-H. and Chang, K.-Y. (2005) An atypical RNA pseudoknot stimulator and an upstream attenuation signal for −1 ribosomal frameshifting of SARS coronavirus. *Nucleic Acids Res*, **33**, 4265-4275.
26. Hu, H.T., Cho, C.P., Lin, Y.H. and Chang, K.Y. (2016) A general strategy to inhibiting viral -1 frameshifting based on upstream attenuation duplex formation. *Nucleic Acids Res*, **44**, 256-266.
27. Kelly, J.A., Olson, A.N., Neupane, K., Munshi, S., San Emeterio, J., Pollack, L., Woodside, M.T. and Dinman, J.D. (2020) Structural and functional conservation of the programmed ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). *J. Biol. Chem.*, **295**, 10741-10748.
28. Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., Yang, M., Xing, X., Wang, Y., Li, M. *et al.* (2021) The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat. Commun.*, **12**, 3917.
29. Zimmer, M.M., Kibe, A., Rand, U., Pekarek, L., Ye, L., Buck, S., Smyth, R.P., Cicin-Sain, L. and Caliskan, N. (2021) The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting. *Nat. Commun.*, **12**, 7193.
30. Pan, D., Klare, K., Petrovic, A., Take, A., Walstein, K., Singh, P., Rondelet, A., Bird, A.W. and Musacchio, A. (2017) CDK-regulated dimerization of M18BP1 on a Mis18 hexamer is necessary for CENP-A loading. *eLife*, **6**, e23352.
31. Loughran, G., Howard, M.T., Firth, A.E. and Atkins, J.F. (2017) Avoidance of reporter assay distortions from fused dual reporters. *RNA*, **23**, 1285-1289.
32. Arai, R., Ueda, H., Kitayama, A., Kamiya, N. and Nagamune, T. (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng. Des. Sel.*, **14**, 529-532.
33. Pekarek, L., Buck, S. and Caliskan, N. (2022) Optical Tweezers to Study RNA-Protein Interactions in Translation Regulation. *J. Vis. Exp. (180),* e62589.
34. Hill, C.H., Pekarek, L., Napthine, S., Kibe, A., Firth, A.E., Graham, S.C., Caliskan, N. and Brierley, I. (2021) Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch. *Nat. Commun.*, **12**, 7166.
35. Buck, S., Pekarek, L. and Caliskan, N. (2022) POTATO: Automated pipeline for batch analysis of optical tweezers data. *Biophys. J.*, **121**, 2830-2839.

36. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nature methods*, **9**, 676-682.
37. Guo, L.T., Adams, R.L., Wan, H., Huston, N.C., Potapova, O., Olson, S., Gallardo, C.M., Graveley, B.R., Torbett, B.E. and Pyle, A.M. (2020) Sequencing and Structure Probing of Long RNAs Using MarathonRT: A Next-Generation Reverse Transcriptase. *J Mol Biol*, **432**, 3338-3352.
38. Zhao, C., Liu, F. and Pyle, A.M. (2018) An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA*, **24**, 183-195.
39. Kechin, A., Boyarskikh, U., Kel, A. and Filipenko, M. (2017) cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J. Comput. Biol.*, **24**, 1138-1143.
40. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**, 357-359.
41. Incarnato, D., Morandi, E., Simon, L.M. and Oliviero, S. (2018) RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications. *Nucleic Acids Res*, **46**, e97-e97.
42. Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A. and Weeks, K.M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nature methods*, **11**, 959-965.
43. Tomezsko, P.J., Corbin, V.D.A., Gupta, P., Swaminathan, H., Glasgow, M., Persad, S., Edwards, M.D., McIntosh, L., Papenfuss, A.T., Emery, A. *et al.* (2020) Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature*, **582**, 438-442.
44. Bellaousov, S., Reuter, J.S., Seetin, M.G. and Mathews, D.H. (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res*, **41**, W471-474.
45. Green, L., Kim, C.-H., Bustamante, C. and Tinoco, I. (2008) Characterization of the Mechanical Unfolding of RNA Pseudoknots. *J Mol Biol*, **375**, 511-528.
46. Zhong, Z., Yang, L., Zhang, H., Shi, J., Vandana, J.J., Lam, D.T.U.H., Olsthoorn, R.C.L., Lu, L. and Chen, G. (2016) Mechanical unfolding kinetics of the SRV-1 gag-pro mRNA pseudoknot: possible implications for −1 ribosomal frameshifting stimulation. *Sci Rep*, **6**, 39549.
47. Ritchie, D.B., Foster, D.A. and Woodside, M.T. (2012) Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc Natl Acad Sci U S A*, **109**, 16167-16172.
48. Halma, M.T.J., Ritchie, D.B., Cappellano, T.R., Neupane, K. and Woodside, M.T. (2019) Complex dynamics under tension in a high-efficiency frameshift stimulatory structure. *Proc. Natl. Acad. Sci.*, **116**, 19500.
49. Li, P.T.X., Bustamante, C. and Tinoco, I. (2007) Real-time control of the energy landscape by force directs the folding of RNA molecules. *Proc. Natl. Acad. Sci.*, **104**, 7039.
50. Cordero, P., Kladwang, W., VanLang, C.C. and Das, R. (2012) Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry*, **51**, 7037-7039.
51. Mailler, E., Paillart, J.C., Marquet, R., Smyth, R.P. and Vivet-Boudou, V. (2019) The evolution of RNA structural probing methods: From gels to next-generation sequencing. *Wiley Interdiscip Rev RNA*, **10**, e1518.
52. Krokhotin, A., Mustoe, A.M., Weeks, K.M. and Dokholyan, N.V. (2017) Direct identification of base-paired RNA nucleotides by correlated chemical probing. *RNA*, **23**, 6-13.
53. Crooke, S.T., Baker, B.F., Crooke, R.M. and Liang, X.-h. (2021) Antisense technology: an overview and prospectus. *Nat. Rev. Drug Discov*, **20**, 427-453.

54. Beane, R.L., Ram, R., Gabillet, S., Arar, K., Monia, B.P. and Corey, D.R. (2007) Inhibiting gene expression with locked nucleic acids (LNAs) that target chromosomal DNA. *Biochemistry*, **46**, 7572-7580.
55. Hillebrand, F., Ostermann, P.N., Müller, L., Degrandi, D., Erkelenz, S., Widera, M., Pfeffer, K. and Schaal, H. (2019) Gymnotic Delivery of LNA Mixmers Targeting Viral SREs Induces HIV-1 mRNA Degradation. *Int J Mol Sci*, **20**, 1088.
56. Komar, A.A. and Hatzoglou, M. (2015) Exploring Internal Ribosome Entry Sites as Therapeutic Targets. *Front Oncol*, **5**, 233.
57. Lulla, V., Wandel Michal, P., Bandyra Katarzyna, J., Ulferts, R., Wu, M., Dendooven, T., Yang, X., Doyle, N., Oerum, S., Beale, R. *et al.* Targeting the Conserved Stem Loop 2 Motif in the SARS-CoV-2 Genome. *J. Virol.*, **95**, e00663-00621.
58. Hagey, R.J., Elazar, M., Pham, E.A., Tian, S., Ben-Avi, L., Bernardin-Souibgui, C., Yee, M.F., Moreira, F.R., Rabinovitch, M.V., Meganck, R.M. *et al.* (2022) Programmable antivirals targeting critical conserved viral RNA secondary structures from influenza A virus and SARS-CoV-2. *Nat Med*, **28**, 1944-1955.
59. Larson, M.H., Landick, R. and Block, S.M. (2011) Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. *Mol Cell*, **41**, 249-262.
60. John, R., Davenport, n., Wuite, G.J.L., Landick, R. and Bustamante, C. (2000) Single-Molecule Study of Transcriptional Pausing and Arrest by E. coli RNA Polymerase. *Science*, **287**, 2497-2500.
61. Lee, T.-H., Lapidus, L.J., Zhao, W., Travers, K.J., Herschlag, D. and Chu, S. (2007) Measuring the Folding Transition Time of Single RNA Molecules. *Biophys. J.*, **92**, 3275-3283.
62. Mahen, E.M., Watson, P.Y., Cottrell, J.W. and Fedor, M.J. (2010) mRNA Secondary Structures Fold Sequentially But Exchange Rapidly In Vivo. *PLOS Biology*, **8**, e1000307.
63. Prabhakar, A., Choi, J., Wang, J., Petrov, A. and Puglisi, J.D. (2017) Dynamic basis of fidelity and speed in translation: Coordinated multistep mechanisms of elongation and termination. *Protein Sci.*, **26**, 1352-1362.

Chapter 5

"Sometimes science is more art than science"

Rick Sanchez

5. The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting

Matthias M. Zimmer[1,†], Anuja N. Kibe[1,†], Ulfert Rand[2], Lukas Pekarek[1], Liqing Ye[1], Stefan Buck[1], Redmond Smyth[1,3], 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.

5.1. Abstract

Programmed ribosomal frameshifting (PRF) is a fundamental gene expression event in many viruses, including SARS-CoV-2. It allows production of essential viral structural and replicative enzymes that are encoded in 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 frameshift elements and thereby impact virulence. This prompted us to comprehensively dissect the interplay between the SARS-CoV-2 frameshift element and the host proteome. We reveal that the short isoform of the zinc-finger antiviral protein (ZAP-S) is a direct regulator of PRF in SARS-CoV-2 infected cells. ZAP-S overexpression strongly impairs frameshifting and inhibits 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 interferes with the folding of the frameshift RNA element. Together, these data identify ZAP-S as a host-

encoded inhibitor of SARS-CoV-2 frameshifting and expand our understanding of RNA-based gene regulation.

5.2. 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 [1]. Analysis of RNA- and protein-interaction networks have rapidly improved our understanding of SARS-CoV-2 viral replication [2,3]. However, detailed mechanistic understanding of the interplay between RNA-protein complexes, which could inform the design of novel antivirals, is still lacking. Here, functionally important RNA elements of the SARS-CoV-2 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 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 genomes and is often used to expand the variability of proteomes or to ensure a defined stoichiometry of protein products [4,5]. In coronaviruses, –1 programmed ribosomal frameshifting on the *1a/1b* gene is essential for efficient viral replication and transcription of the viral genome. In cells, the efficiency of this frameshifting event varies between 20-40% [6,7]. PRF relies on the presence of a slippery heptameric sequence (in coronaviruses U UUA AAC) and a stimulatory RNA secondary structure, such as a pseudoknot (PK) **(Fig. 5.1A)**. Mutations in the slippery sequence and downstream RNA structure drastically impair frameshifting efficiency [8,9].

Traditionally, efforts to understand the mechanism of –1PRF focused on *cis*-acting modulatory elements. Previous work in *in vitro* reconstituted translation systems revealed in detail how ribosome pausing on the slippery codons may lead to a kinetic partitioning between the two reading frames and favor movement of translating ribosomes to the alternative reading frame [6,10]. It has been shown that –1PRF canonically occurs during a late stage of the tRNA translocation step. The stimulatory RNA element causes ribosomes to become trapped in an unusual conformation that is relieved by either the spontaneous unfolding of the blockade or by a –1 slip on the mRNA [6,10]. It is now becoming clear, however, that *cis*-acting elements are not the only determinants of frameshifting in cells and that *trans*-acting viral and cellular factors as well as small molecules or oligonucleotides can alter frameshifting levels [11–14]. Despite these recent insights, fundamental questions remain. It is for example still unclear how important RNA-binding factors are for frameshifting processes in general and how exactly interactions of these

factors with the RNA alter the mechanical properties of RNA or the choice of the reading frame.

At least three plausible routes to modulate frameshifting by *trans*-acting factors can be envisioned. First, the binding of the factor can transform the downstream RNA element to a more stable roadblock. This has been shown for cardiovirus 2A, poly-(C) binding protein and some small molecules such as naphthyridine carbamate tetramer (NCTn) [11,12,15]. In these cases, the specific interaction of the factor with the nucleotides downstream of the slippery codons leads to an increase in frameshifting. Second, the factor can target stalled ribosomes, as was shown for eukaryotic release factors such as eRF1 alone or eRF1/3, which are recruited by Shiftless (SHFL) to the HIV-1 frameshift site[16,17]. Different from the first group of regulators, the interaction of both SHFL and release factors is not dependent on the identity of the frameshift RNA. Therefore, it remains to be solved how the frameshifting ribosome complexes are recognized by these *trans*-acting factors. A possible third route might act via remodeling or destabilization of the frameshift RNA elements through direct interactions with the *trans*-factor. So far, however, no cellular or viral factor has been reported to affect frameshifting efficiency (FE) through that route.

Given the importance of –1PRF for the SARS-CoV-2 life cycle, we set out to comprehensively identify direct protein interactors of the SARS-CoV-2 frameshift RNA element using an *in vitro* RNA-antisense capture and mass spectrometry-based screen [18]. Through this approach, we identified the short isoform of the zinc-finger antiviral protein (ZAP-S, ZC3HAV1) as the most prominent interaction partner amongst over 100 proteins detected. 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 about 20-fold. Apart from SARS-CoV-1 and 2, we were not able to identify other PRF sites that are affected by ZAP-S, which suggests that certain RNA elements are preferentially recognized by ZAP-S. Using a multidisciplinary approach, we further revealed important molecular principles of frameshifting downregulation by ZAP-S. Specifically, we show that ZAP-S can alter the physical properties of the frameshift RNA. Our study highlights that the expression of the SARS-CoV-2 ORF1a/1b can be directly modulated by a host-encoded RNA-binding protein (RBP) during infection. These findings provide new insights on –1PRF regulation and the interplay between SARS-CoV-2 replication and host defense, thereby paving the way for novel RNA-based therapeutic intervention strategies.

5.3. Results

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

To identify potential cellular RNA-binding proteins 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. 5.1A, B)** [18]. Calu-3 cells are lung epithelial cells that are commonly used to study CoV infection [19]. HEK293 cells are routinely used to study RNA-protein interactomes and represent an ideal system to assess possible cell-type specific interactions[20]. To exclude any non-specific binders, an 80 nucleotides long non-structured RNA was employed 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. 5.1B, C)**.

We identified more than 100 proteins that were at least two-fold enriched compared to the control RNA. According to GO term analysis, the majority (80%) of our hits are categorized as RNA-binding proteins **(Fig. 5.S1A)**. For example, we observed enrichment of the viral nucleocapsid protein (N) in infected lysates, which is a well-described RBP [21]. 35% and 30% of the enriched RBPs were involved in splicing and ribosome biogenesis, respectively **(Fig. 5.S1A)**. Among those, 9 proteins were identified in infected and uninfected Calu-3 cells as well as in HEK293 cells, 19 proteins were common to infected and uninfected Calu-3 cells, 18 hits were found only in HEK293 cells, 15 were captured only in uninfected Calu-3 cells, and 40 were present only in infected Calu-3 cells **(Fig. 5.S1B)**. The core interactome of 9 proteins identified in all three cell systems encompasses well-described post-transcriptional regulators, namely HNRNPH1, DHX36, GRSF1, HNRNPH2, HNRNPF, ZAP, MATR3, ELAVL1 and POP1 **(Fig. 5.1C, D, Fig. 5.S1C, Table 5.S1)**. Based on their enrichment and functions in translational regulation, we selected 20 proteins for downstream analysis and functional characterization. These candidates were also hits recently identified in SARS-CoV-2 genome-wide RNA interactome studies **(Fig. 5.S1D)** [3,21–23]. Several of those interactors 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) [24–26]. Translational regulators included ELAVL1, DHX36, SSB, IMP1 and IMP3 [27,28]. Among those, ELAVL1 is a cofactor, which ensures translational fidelity in the context of upstream ORFs [29]. IMP1 and IMP3 were both relatively lowly enriched RBPs in infected and uninfected Calu-3 cells. IMP1 was reported predominantly bound to the 3' untranslated region of genes and IMP3 was mostly bound

to coding regions [28,30]. DHX36 is a multifunctional helicase and is involved in translation and innate immunity [31,32]. G-rich RNA sequence binding protein, GRSF1, is implicated in mitochondrial translation [33]. 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 [26,34–37]. While the longer isoform of ZAP (ZAP-L) was reported to be mainly recruited to membrane-associated sites of viral replication [34,38,39], 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 [34]. Two additional hits were included in the downstream analysis based on their above 4-fold enrichment only in infected Calu-3 lysates. These included the poly-(A) polymerase PAPD4, and GNL2 which has been implied in ribosome biogenesis [40,41].

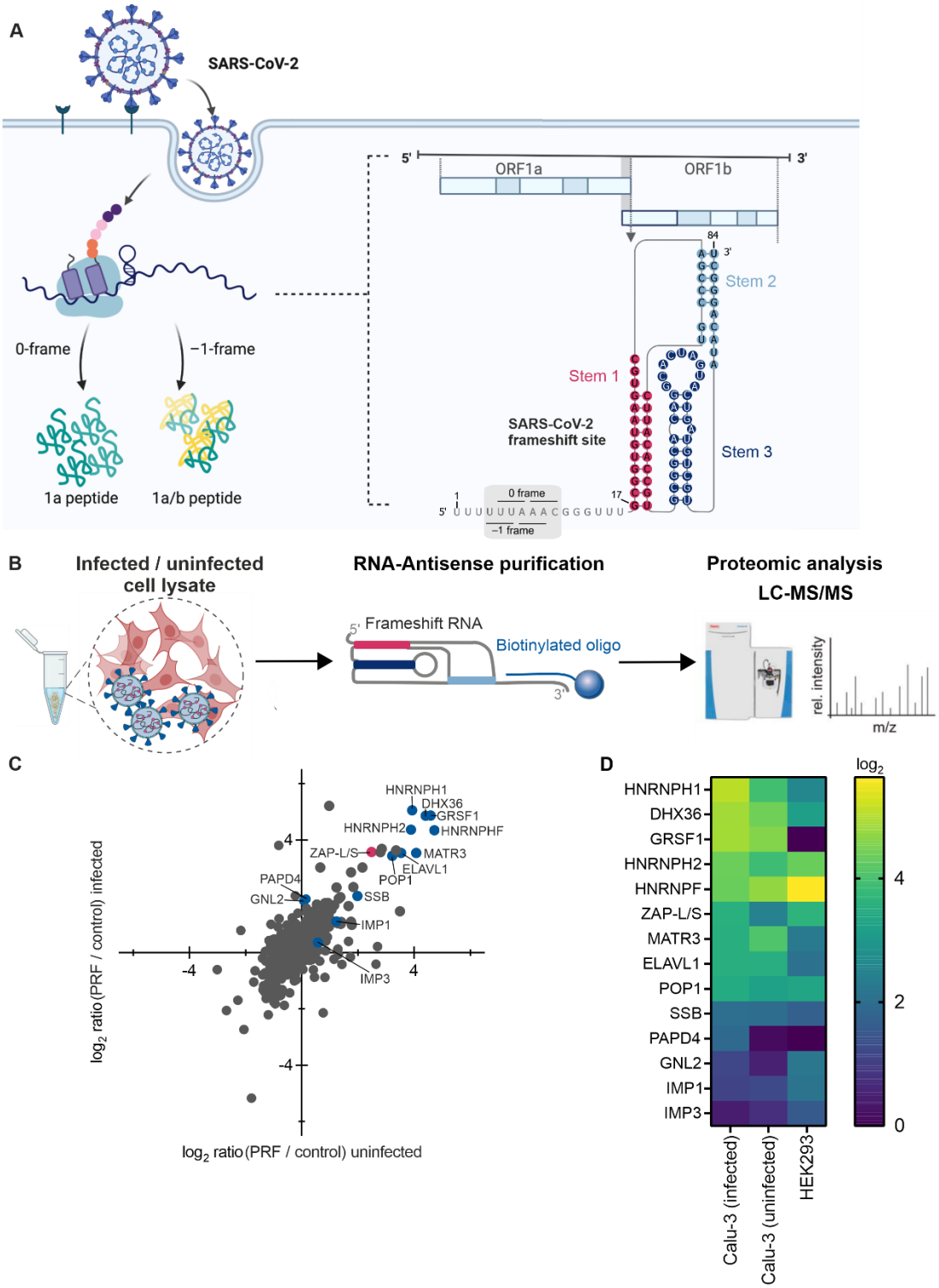

Figure 5.1 *In vitro* RNA-antisense purification-based discovery of protein interactors of the SARS-CoV-2 –1PRF element. (A) Schematic representation of the relevant genomic segment of SARS-CoV-2 as well as the location of the –1PRF element. **(B)** Schematic of *in vitro* interactome capture of protein 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 **Fig. 5.S1D**.

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 a fluorescence-based cellular –1PRF reporter assay. In this assay, the expression of the first ORF, EGFP in the 0-frame, is constitutive, whereas the expression of the following ORF mCherry depends on –1PRF occurring at the preceding SARS-CoV-2 1a/1b frameshift element **(Fig. 5.2A)**. As controls, we used a construct lacking the frameshift element, and the mCherry gene either in –1 or in-frame with respect to EGFP **(Fig. 5.2A, B)**. Frameshifting 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)**. In order 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 (see also **Fig. 5.S2A**). 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 assay, the frameshifting efficiency (FE) of the wild type SARS-CoV-2 1a/1b frameshift site was 35% in HEK293 cells **(Fig. 5.2C, Fig. 5.S2A, B, Table 5.S2)**, in agreement with the published FE for SARS-CoV-1 and SARS-CoV-2 [7,8]. In addition, a vector expressing ECFP-SHFL, a previously described inhibitor of –1PRF in SARS-CoV-2, was employed as a positive control (FE was reduced to 27%) [3]. Among the selected RNA interactors, no change in FE was observed with GNL2, HNRNPF, SSB, IMP1 or IMP3, indicating that binding to the stimulatory RNA element alone is not sufficient for modulating FE. 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. 5.2C)**. 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 **(Table 5.S2)**, 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 **(Fig. 5.S1E)** [42,43]. 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 [42]. 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. 5.1C**). 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 [42,43] **(Fig. 5.S1E, F)**. An increase in ZAP-S protein levels upon SARS-CoV-2 infection was also reported previously [44,45].

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. 5.2D, Fig. 5.S2D)** [21,46]. 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. 5.2D, Fig. 5.S2C, 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 **(Fig. 55.S2C, 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 [46]. 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 [47]. 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. 5.2E)**, 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 [16,48].

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. 5.2F)** [9,49]. 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 [46]. 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 5.2F)**. 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.

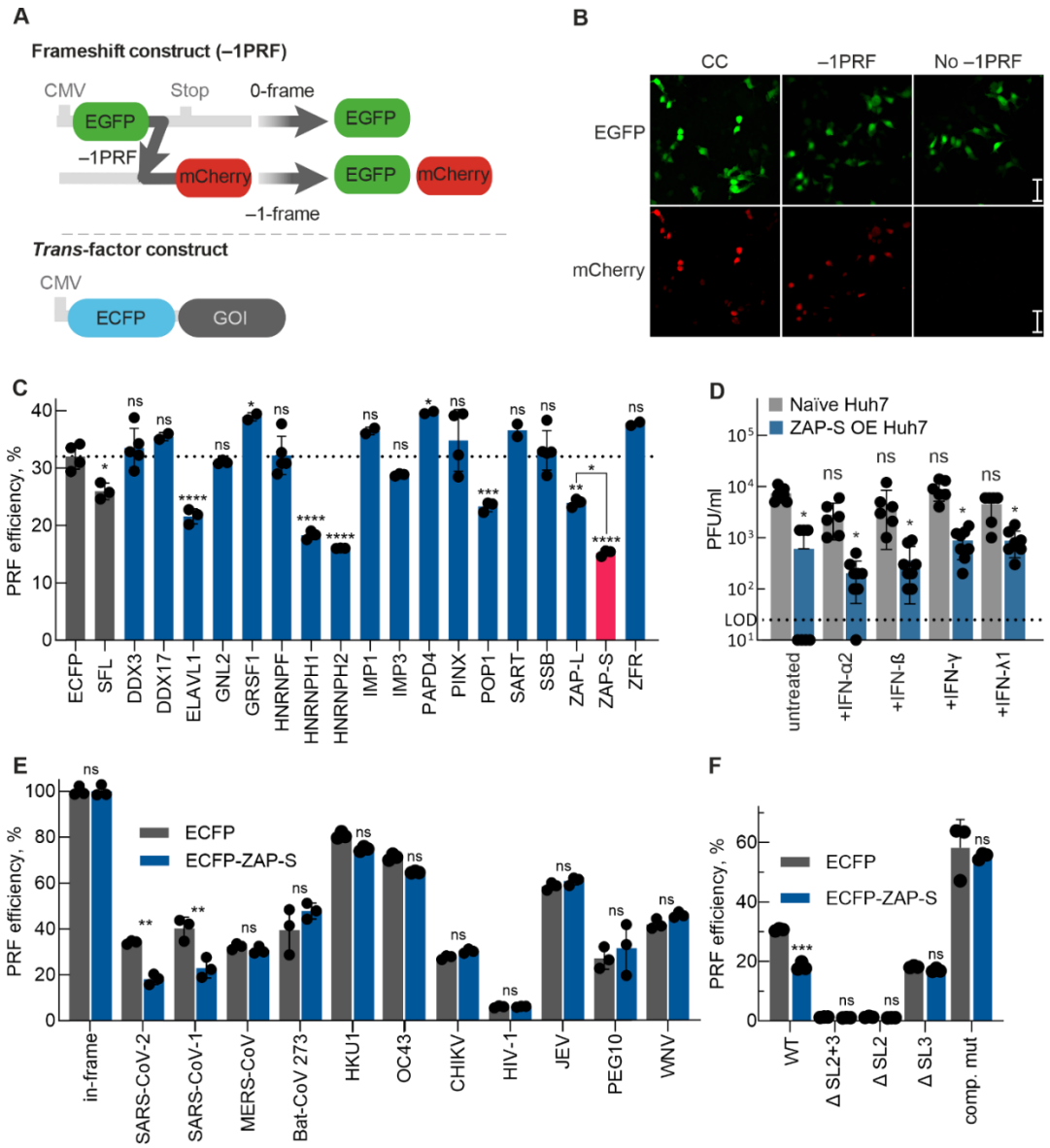

Figure 5.2 A functional screen of SARS-CoV-2 –1PRF element interactors. (A) Schematic representation of the dual-fluorescence frameshift reporter construct. EGFP and mCherry are separated by a self-cleaving 2A peptide as well as by a stop codon in-frame with EGFP. As a result, 0-frame translation would 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 **Table 5.S2** as well as **Fig. 5.4** for schematics of the mutants used here.

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. 5.3A and Fig. 5.S2F)**. 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. 5.3B)**. In accordance with a previous study [7], 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. 5.3B, C, Table 5.S4)**. 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. 5.3C)**.

Several *trans*-acting factors including the cardiovirus 2A and SHFL were shown to bind to ribosomes and as well as frameshifting RNAs [16,50]. 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. 5.3A)**. 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. 5.3D)**. 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. 5.3A)**. 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. 5.3E)**. Similar polysome profiles were obtained with cells overexpressing SHFL, which as a known ribosome interactor acts as positive control[16]. We further confirmed that endogenous ZAP-S also associates with ribosomes in naïve Calu-3 cells via ribosome pelleting **(Fig. 5.3A, F)**. Collectively, these results indicate that ZAP-S associates with ribosomes, either directly, or indirectly through its interactions with the SARS-CoV-2 mRNA.

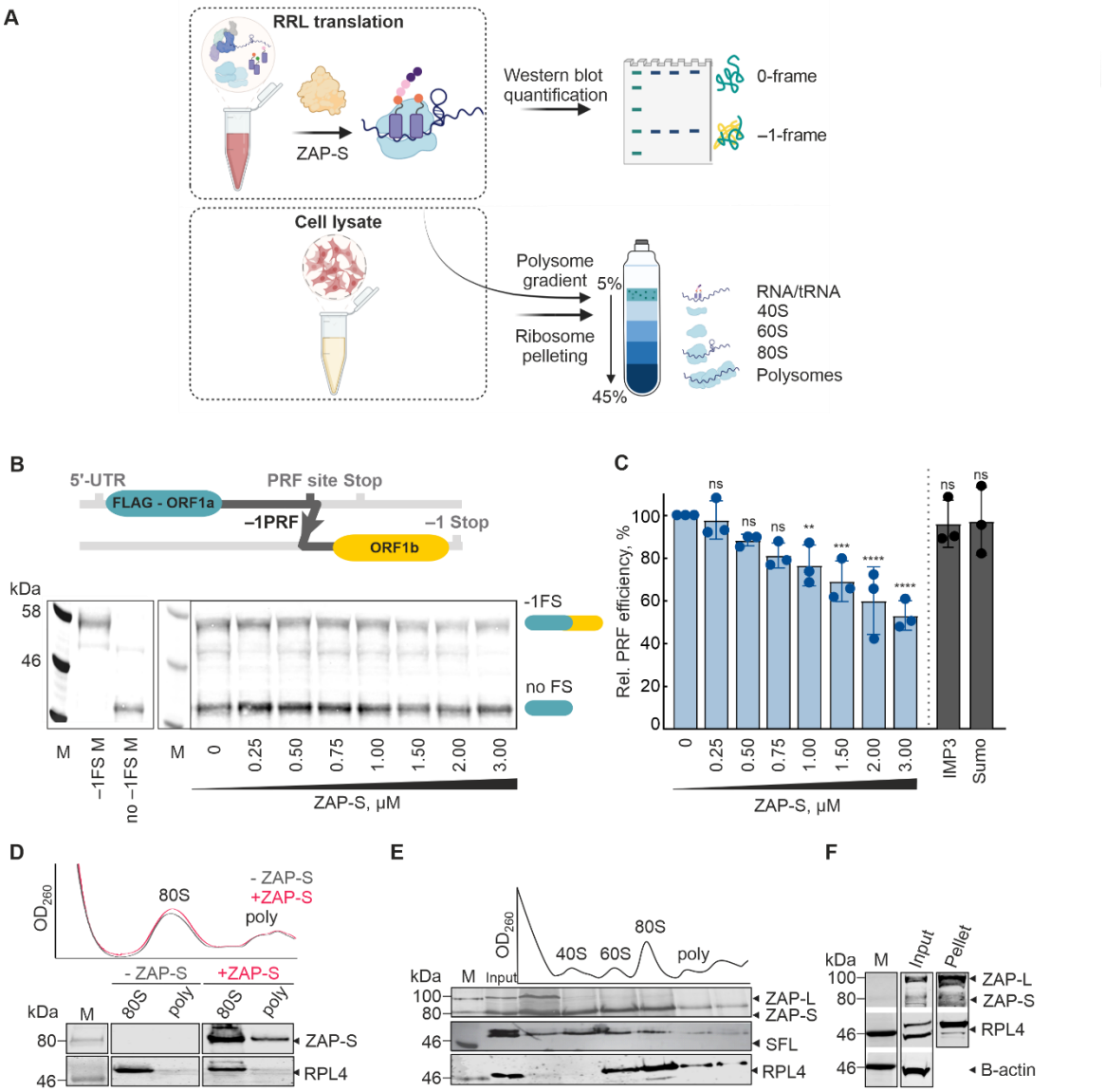

Figure 5.3 Effect of zinc-finger antiviral protein (ZAP) on 1a/1b –1 frameshifting *in vitro*. (A) The strategy of the *in vitro* translation assay using rabbit reticulocyte lysate (RRL) and the experimental workflow to study ribosome association of ZAP-S. **(B)** Schematics of the N-terminal FLAG-tagged 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 **Fig. 5.S2** and **Table 5.S3**. **(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 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. 5.4A, 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. 5.4C)** 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. 5.4D)**. In contrast, deletion of the SL3 region (ΔSL3) only marginally reduced the affinity of ZAP binding (K_D=175±64 nM) **(Fig. 5.4E)**. 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. 5.4F)**. In contrast, ZAP-S binds to the compensatory mutant, with an affinity close to WT RNA (K_D =128±29 nM) **(Fig. 5.4G)**. 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. 5.4H)**. 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 **(Fig. 5.S3G)**. 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) **(Fig. 5.S3H, 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) **(Fig. 5.S4)**. In the absence of ZAP-S, DMS reactivities were consistent with a significant proportion of the RNA folding into a pseudoknot conformation **(Fig. 5.S4)**. 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.

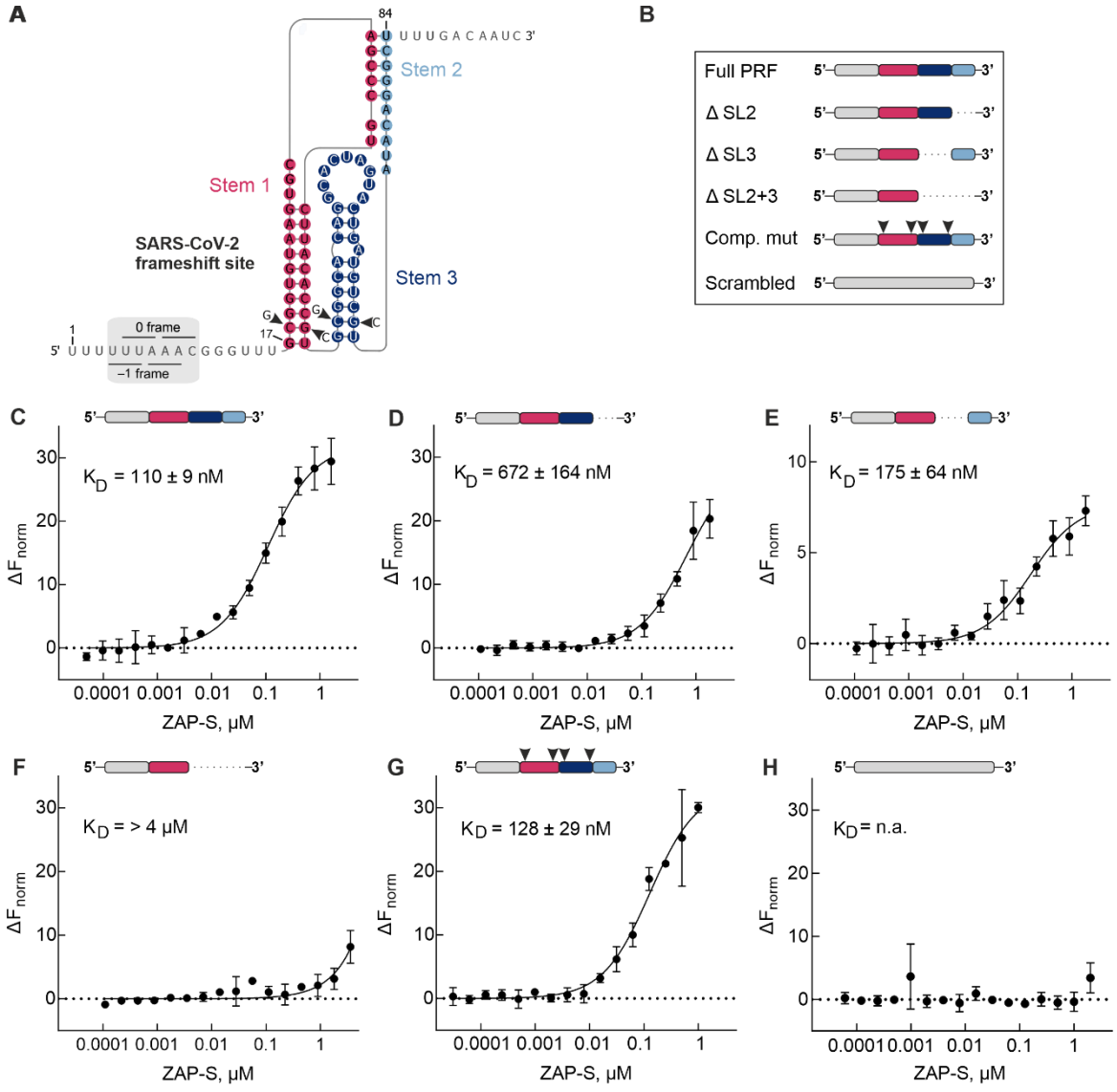

Figure 5.4 *In vitro* characterization of ZAP-S interaction with SARS-CoV-2 –1 PRF RNA. (A) Proposed structure of the PRF element of SARS-CoV-2. Nucleotide substitutions in the

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 **Fig. 5.S4A-F.** For the related DNA sequences of the mutants, see also **Table 5.S2.**

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 [49,51–53]. 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. 5.5A)**. 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 **(Fig. 5.S6, Table 5.S5)**, which agreed with the expected value for the full-length pseudoknot reported previously Fig. **5B)** [51,52]. Moreover, the majority (80%) of RNA molecules unfolded at forces (FU) of 15 -20 pN (**Fig. 5.S5**). For the remaining traces, we observed two consecutive unfolding events with an intermediate contour length change of 17.1±3.5 nm **(Fig. 5.5D)** 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. 5.5D, Fig. 5.S6, Table 5.S5)**. Such a hysteresis during refolding is commonly reported with pseudoknots and other highly structured RNAs [51,52]. 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. 5.5D, Fig. 5.S5** and **5.S6, Table 5.S5**). 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. 5.5D, Fig. 5.S5** and **5.S6, Table 5.S5)**.

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. 5.4B, Fig. 5.5C** 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. 5.5C, 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 (**Fig. 5.S5** and **5.S6, Table 5.S5)**. 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) (**Fig. 5.S6, Table 5.S5**). In the ΔSL3 RNA variant, (**Fig. 5.5 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 (**Fig. 5.S5** and **5.S6, Table 5.S5)**. 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. 5.5F, Table 5.S5)**. The ΔSL2+3 variant is predicted to form the simple hairpin (SL1). Our data also confirmed the presence of a single stem-loop (**Fig. 5.5G, Fig. 5.S5** and **5.S6, Table 5.S5)**, 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. 5.5G)**. Aside from that, F_U was slightly shifted to lower values although our MST results clearly showed no binding of ZAP-S to this RNA variant (**Fig. 5.4F, Fig. 5.S5**). We also tested the effect of non-specific interactions using the control RBP IMP3, and we observed a similar small shift in the F_U. Therefore, we conclude that this effect is due to non-specific interactions and/or molecular crowding (**Fig. 5.5I**). 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. 5.5H, Fig. 5.S5** and **5.S6, Table 5.S5**). 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. 5.5H**).

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. 5.5J, Table 5.S5**). 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. 5.5K**). 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.

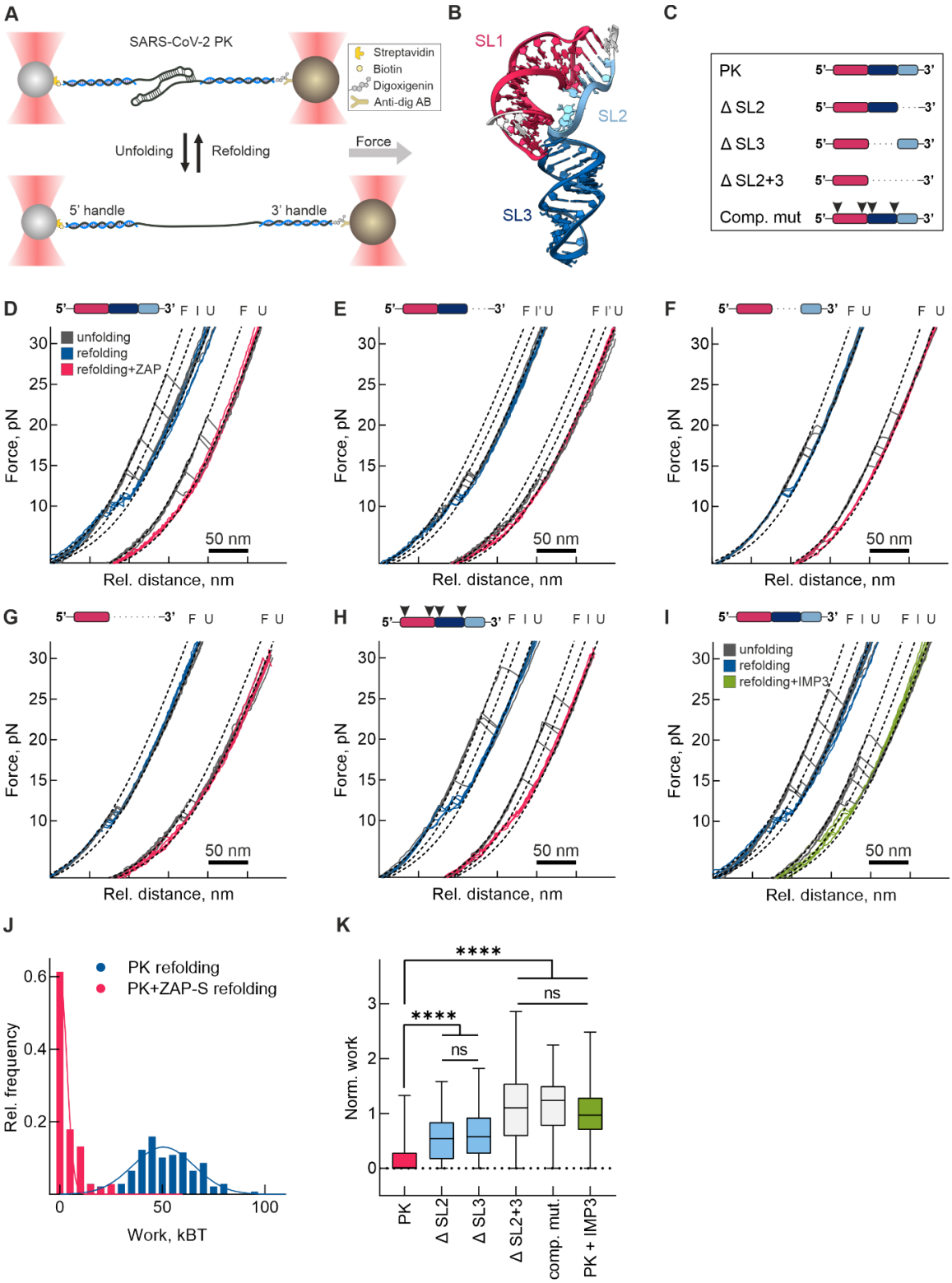

Figure 5.5 Single molecule characterization of mechanical properties of SARS-CoV-2 PRF RNA in the presence of ZAP-S. (A) Schematic illustrating optical tweezers experiments. RNA was hybridized to single-stranded DNA handles flanking the SARS-CoV-2 frameshift site and conjugated to functionalized beads. 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 **Fig. 5.S5, 5.S6 and Table 5.S1**.

5.4. 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 [3,21–23].

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 [12,15]. Rather, ZAP-S inhibits coronaviral frameshifting and shows sequence preference for SARS-CoV-2 the SL2 and SL3. 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 [16]. In contrast, ZAP-S most likely interferes with the refolding of the frameshift RNA an intact SL3 seems to be crucial for this effect. In addition, ZAP-S shows sequence preference for SARS-CoV-2 SL2 and SL3. 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 [54]. 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. 5.6)**. 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. 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 21,46. Further studies are required to deconvolute the multivalent effects of ZAP-S on immunity, viral replication and gene expression 34,55–59. 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.

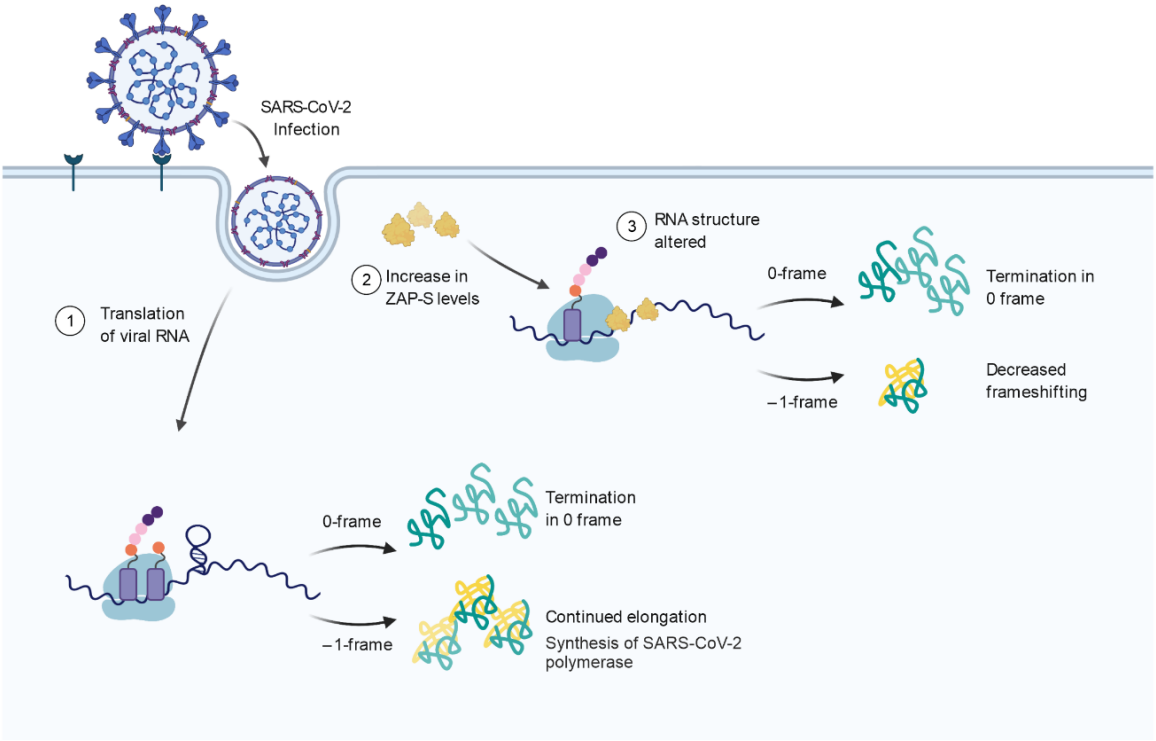

Figure 5.6. Model for ZAP-S mediated inhibition of SARS-CoV-2 frameshifting. (1) Upon infection, the viral RNA is translated by the cellular machinery, and 40% of translation events yield the 1a/1b polyprotein through –1PRF. **(2)** Infection also leads to the induction of antiviral factors including 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.

5.5. 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 [60]. Gene ontology (GO) term analysis was performed with Panther [61]. The list of all identified proteins is given in **Table 5.S1**.

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 [62]) 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 [63] signals as well as an alpha-helical linker [64]. 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. 5.4A** and **Table 5.S3** 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 [65]) 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 [66]. 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 [67]. 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 **Table 5.S3**.

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 [68] and pCMV-VSV-G (gift from Prof. Weinberg, Addgene plasmid # 8454 [69]). 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. 5.2A**. 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 **(Fig. 5.S2F)**. FE was calculated according to the following formula:

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

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 [70]. 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 (**Fig. 5.S2D)**. 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), anti-DDDDK (ab49763), anti-ALFA (FluoTag®-X2 anti-ALFA AlexaFluor 647), anti-ZC3HAV1

(Proteintech 16820-1-AP), anti-RPL4 (Proteintech 67028-1-Ig), anti-RPS6 (Proteintech 14823-1-AP), anti-RYDEN (SHFL) (Proteintech 27865-1-AP). See also Supplementary Table 3. The following secondary antibodies were used: IRDye® 800CW Goat anti-rabbit and IRDye® 680RD Donkey anti-Mouse (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 [71]. FE was calculated as previously described, by the formula intensity (–1-frame)/(intensity (–1-frame) + intensity (0-frame)) 11. 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. 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.

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).

Dimethyl sulfate mutational profiling with sequencing (DMS-MaPseq)

50 ng of RNA was first heat denatured at 90°C for 2 mins followed by chilling on ice for 2 min. RNA was then refolded in 50 mM HEPES pH 7.6, 250 mM KCl, 0.05 mM EDTA, 5% glycerol, 5 mM MgCl2, 0.2 mg/ml yeast tRNA, and 20 U of RNasin for 15 min at 37 °C. Recombinant ZAP-S was buffer exchanged into 250 mM KCl, 50 mM HEPES/KOH pH 7.6, 0.05mM EDTA, 5% glycerol and added to a final concentration of 5 µM and incubated at 25 °C for 10 min. DMS was diluted in EtOH to a working concentration of 1.7 M. 1/10 volume of DMS working stock was added to the samples to make a final concentration of 170 mM in a total volume of 30 µl. Samples were incubated at 37°C for 6 min and then quenched with 30 µl of beta-mercaptoethanol (from a 14.2 M stock). For the untreated control, EtOH was used instead of DMS. RNA was then purified by Trizol LS according to the manufacturer's instructions.

Probed and control RNA was reverse transcribed using 40 U MarathonRT in RT Buffer (50 mM Tris–HCl pH 8.3, 200 mM KCl, 5 mM DTT, 20% glycerol, 1 mM MnCl2), 0.5 mM dNTP mix, 2 µM primer [GGcgaagagcaggttgcaggat] and 8 U of RNasin in a final

volume of 25 µM. Reverse transcriptions were carried out at 42°C for 3 h. cDNA was diluted 1/10 with nuclease free water and PCR amplified using PrimeSTAR GXL DNA polymerase. Reaction conditions were 8 µl of diluted cDNA, 1x GXL reaction buffer, 0.2 mM dNTPs, 0.25 µM of forward [TCGTCGGCAGCGTCAGcttcgcaggagctcgacagctac] and reverse primer [GTCTCGTGGGCTCGGAGGGcgaagagcaggttgcaggat], 0.025 U/µl of polymerase in a final volume of 25 µl. Cycling conditions were 30 sec at 98°C then 25 cycles of 10 sec at 98°C, 15 sec at 60°C and 15 sec at 68°C then 68°C for 5 min. PCR products were verified on a 1.5% agarose gel followed by column purification (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel) according to the manufacturer's instructions. A final indexing PCR was carried out using 40 ng of PCR products using Illumina Nextera DNA CD indexes (96 Indexes, 96 Samples, Illumina). Reaction conditions were 40 ng of purified PCR product, 1x Q5 reaction buffer, 0.2 mM dNTPs, 2.5 µl of indexing primer, 0.02 U/µl of Q5 polymerase in a final volume of 15 µl. Cycling conditions were 30 sec at 98°C then 5 cycles of 10 sec at 98°C, 15 sec at 60°C and 15 sec at 68°C then 68°C for 5 min. Indexed PCR products were verified on a 1.5% agarose gel, pooled together in an equimolar ratio, before final purification on a 1.5% agarose gel. Pooled indexed sequencing library was quantified using the NEBNext library Quant Kit for Illumina and sequenced on an Illumina Miniseq using a 150 cycle High Output reagent kit.

DMS-MaP-seq data was trimmed using cutadapt [72] and aligned to the reference sequence using bowtie2 [73]. Cutadapt parameters were "–nextseq-trim 20 –max-n 0 -a atcctgcaacctgctcttcgcc -A gtagctgtcgagctcctgcgaag". Bowtie2 parameters were "-D 20 -R 3 -N 1 -L 15 -i S,1,0.50 --rdg 5,1 --rfg 5,1 --maxins 600". Further analysis was carried out using the rf-count and rf-norm modules of RNA Framework package [74]. rf-count parameters were "-po -pp -m -ds 75 -q 30 -es -cc". rf-norm parameters were "-rb AC -sm 3 -nm 1". Data were plotted onto PK structures using StructureEditor (version 1.0).

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; cat. no. ABK84-E2-M) and secondary antibody anti-mouse Alexa488 (Cell Signaling Technology (Danvers, MA, 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 (Eching, Germany), #VEC-H-1000).

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 (Beckmann 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.

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. 5.4E)**. 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 (POTATO, https://github.com/lpekarek/POTATO.git, manuscript in preparation). 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/). 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 [50]. 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 [75], 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^-ΔΔCt method relative to the values in uninfected samples, after normalization to the housekeeping gene (geometric mean) GAPDH. Statistical analysis was conducted using an unpaired two-tailed *t*-test with Welch's correction comparing ΔCt values of the respective RNA in uninfected and infected cells. The results were plotted using Prism 8.0.2 (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.

5.6. 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[76].

5.7. 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)[77].

5.8. 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.

5.9. 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.

5.10. Supplementary material

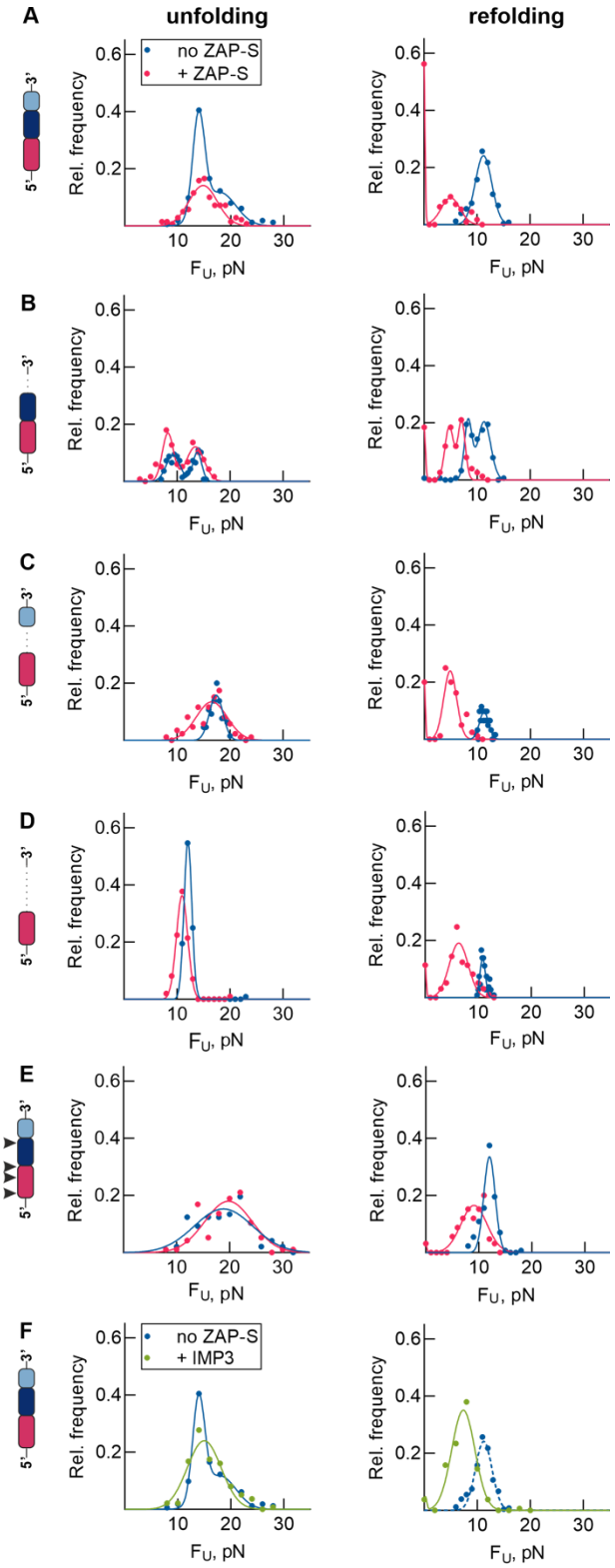

Figure 5.S5 Optical tweezers data related to Fig. 5.5. (A-F) Distributions of unfolding and refolding force (F_U), respectively, in absence (blue) and presence (pink) of ZAP-S protein for different RNA samples measured in OT. (A) wild type PK, (B) ΔSL2 mutant, (C) ΔSL3 mutant, (D) ΔSL2+3 mutant, (E) compensatory mutant, (F) PK in absence (blue) and presence (green) of IMP3.

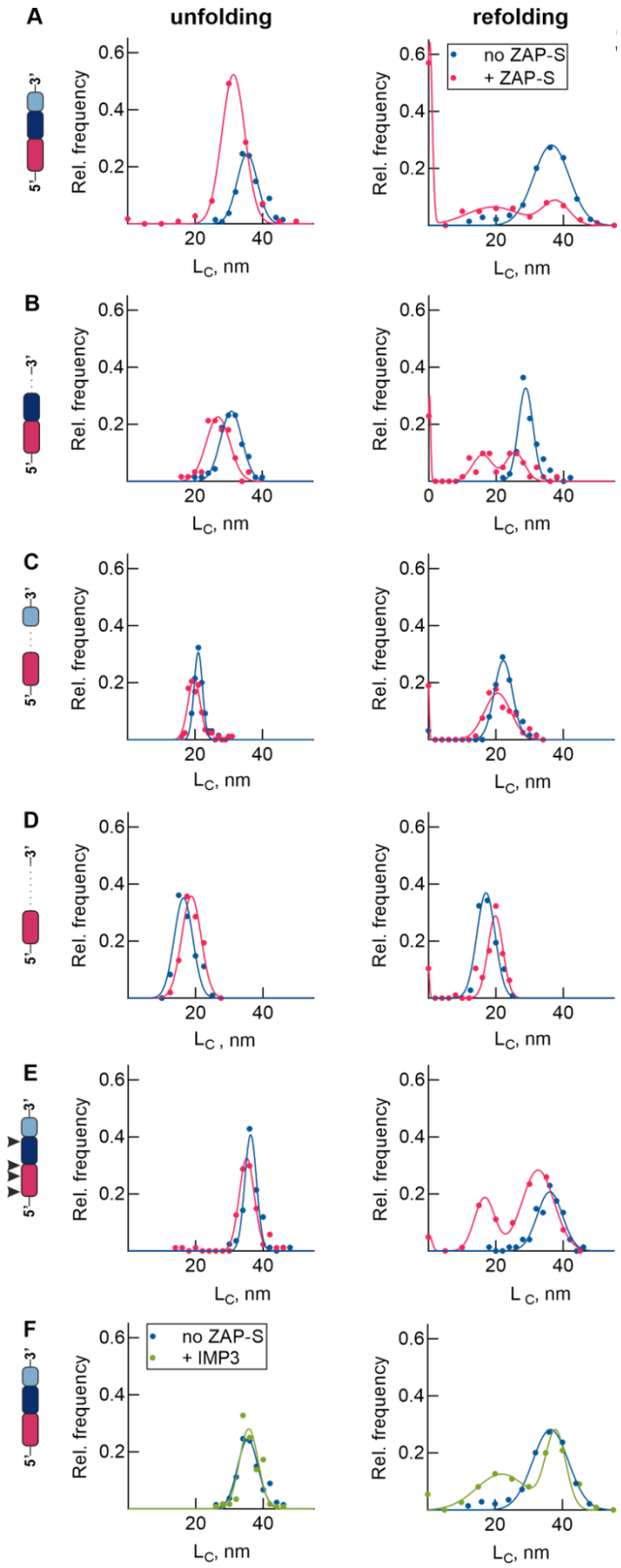

Figure 5.S6 Optical tweezers data related to Fig. 5.5. (A-F) Distributions of unfolding and refolding contour length change (L_C), respectively, in absence (blue) and presence (pink) of ZAP-S protein for different RNA samples measured in OT. (A) wild type PK, (B) ΔSL2 mutant, (C) ΔSL3 mutant, (D) ΔSL2+3 mutant, (E) compensatory mutant, (F) wild type PK in absence (blue) and presence (green) of IMP3.

Table 5.S1: Values of fitted and calculated parameters for individual samples. Errors on the measured values represent standard error on the mean.

	Direction	Peak #	Work, kBT	Contour length, nm	Force, pN	# of FD curves / molecules
PK	unfolding	1	74.1±16.6	35.4±3.0	14.0±1.1	273/24
		2			18.0±3.0	
	refolding	1	50.5±15.0	36.4±5.3	11.2±1.6	
PK+ZAP	unfolding	1	63.3±22.8	31.4±3.4	14.8±2.7	219/11
	refolding	1	0.0±0.9	0.4±0.7	0	
		2		19.2±8.0	4.8±1.8	
		3		37.7±3.9	9.3±0.5	
PK+IMP3	unfolding	1	79.2±24.2	35.7±2.8	15.0±3.2	226/20
	refolding	1	19.8±8.0	22.0±7.5	7.4±2.2	
		2		38.0±3.0		
Δ SL2	unfolding	1	48.9±8.0	30.8±3.1	9.3±1.3	146/8
		2			13.8±0.8	
	refolding	1	39.0±7.9	28.7±2.1	8.3±0.7	
		2			11.3±1.2	
Δ SL2+ZAP	unfolding	1	39.1±4.8	26.9±3.5	8.3±1.1	122/8
		2			13.5±1.7	
	refolding	1	1.4±4.0	0.2±0.3	4.8±0.8	
		2	12.9±2.6	15.9±3.1	7.1±0.7	
		3	23.1±2.7	25.6±2.8		
Δ SL3	unfolding	1	50.4±7.0	21.0±1.2	17.4±1.3	127/12
	refolding	1	30.3±4.4	22.2±2.7	11.3±0.9	
Δ SL3+ZAP	unfolding	1	39.8±13.9	19.7±1.8	16.6±2.9	163/11
	refolding	1	0.8±11.3	20.5±3.9	5.1±1.2	
Δ SL2+3	unfolding	1	24.7±6.0	16.4±2.8	12.1±0.7	216/8
	refolding	1	21.8±6.8	17.1±2.7	10.9±0.5	
ΔSL2+3 +ZAP	unfolding	1	25.1±7.6	18.8±2.8	11.0±1.1	196/11
	refolding	1	9.3±7.2	19.9±2.2	6.3±1.8	
Comp. mut	unfolding	1	114.2±33.9	36.3±1.7	18.9±5.5	158/12
	refolding	1	56.9±12.6	36.0±3.7	12.0±1.1	
Comp. mut+ZAP	unfolding	1	109.3±30.5	35.2±2.2	19.8±4.5	169/16
	refolding	1	29.5±13.6	18.2±5.3	9.2±2.5	
		2		32.8±3.0		

5.11. References

1. Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: Immunity, inflammation and intervention. *Nature Reviews. Immunology* **20**, 363–374 (2020).
2. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
3. Schmidt, N. *et al.* The SARS-CoV-2 RNA-protein interactome in infected human cells. *Nature Microbiology* (2020).
4. Atkins, J. F. & Gesteland, R. F. *Recoding: Expansion of decoding rules enriches gene expression*. vol. 24 (Springer New York, 2010).
5. Caliskan, N., Peske, F. & Rodnina, M. V. Changed in translation: mRNA recoding by -1 programmed ribosomal frameshifting. *Trends in Biochemical Sciences* **40**, 265–274 (2015).
6. Caliskan, N., Katunin, V. I., Belardinelli, R., Peske, F. & Rodnina, M. V. Programmed -1 frameshifting by kinetic partitioning during impeded translocation. *Cell* **157**, 1619–1631 (2014).
7. Neupane, K. *et al.* Anti-frameshifting ligand active against SARS coronavirus-2 is resistant to natural mutations of the frameshift-stimulatory pseudoknot. *Journal of Molecular Biology* **432**, 5843–5847 (2020).
8. Baranov, P. V. *et al.* Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* **332**, 498–510 (2005).
9. Brierley, I., Digard, P. & Inglis, S. C. Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. *Cell* **57**, 537–547 (1989).
10. Chen, J. *et al.* Dynamic pathways of -1 translational frameshifting. *Nature* **512**, 328–332 (2014).
11. Matsumoto, S., Caliskan, N., Rodnina, M. V., Murata, A. & Nakatani, K. Small synthetic molecule-stabilized RNA pseudoknot as an activator for -1 ribosomal frameshifting. *Nucleic Acids Research* **46**, 8079–8089 (2018).
12. Napthine, S. *et al.* Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nature Communications* **8**, 15582 (2017).
13. Belew, A. T. *et al.* Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. *Nature* **512**, 265–269 (2014).
14. Puah, R. Y. *et al.* Selective binding to mRNA duplex regions by chemically modified peptide nucleic acids stimulates ribosomal frameshifting. *Biochemistry* **57**, 149–159 (2018).

15. Napthine, S. *et al.* A novel role for poly(c) binding proteins in programmed ribosomal frameshifting. *Nucleic Acids Research* **44**, 5491–5503 (2016).
16. Wang, X. *et al.* Regulation of HIV-1 gag-pol expression by shiftless, an inhibitor of programmed -1 ribosomal frameshifting. *Cell* **176**, 625-635.e14 (2019).
17. Kobayashi, Y., Zhuang, J., Peltz, S. & Dougherty, J. Identification of a cellular factor that modulates HIV-1 programmed ribosomal frameshifting. *The Journal of Biological Chemistry* **285**, 19776–19784 (2010).
18. Butter, F., Scheibe, M., Mörl, M. & Mann, M. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10626–10631 (2009).
19. Harcourt, J. L., Caidi, H., Anderson, L. J. & Haynes, L. M. Evaluation of the calu-3 cell line as a model of in vitro respiratory syncytial virus infection. *Journal of Virological Methods* **174**, 144–149 (2011).
20. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
21. Lee, S. *et al.* The SARS-CoV-2 RNA interactome. *Molecular Cell* **81**, 2838-2850.e6 (2021).
22. Flynn, R. A. *et al.* Discovery and functional interrogation of SARS-CoV-2 RNA-host protein interactions. *Cell* **184**, 2394-2411.e16 (2021).
23. Kamel, W. *et al.* Global analysis of protein-RNA interactions in SARS-CoV-2-infected cells reveals key regulators of infection. *Molecular Cell* **81**, 2851-2867.e7 (2021).
24. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: Insights into their role in health and disease. *Human Genetics* **135**, 851–867 (2016).
25. Wu, J. *et al.* Cryo-EM structure of the human ribonuclease p holoenzyme. *Cell* **175**, 1393-1404.e11 (2018).
26. Bick, M. J. *et al.* Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *Journal of Virology* **77**, 11555–11562 (2003).
27. Ray, P. S. & Das, S. La autoantigen is required for the internal ribosome entry site-mediated translation of coxsackievirus B3 RNA. *Nucleic Acids Research* **30**, 4500–4508 (2002).
28. Weinlich, S. *et al.* IGF2BP1 enhances HCV IRES-mediated translation initiation via the 3'UTR. *RNA (New York)* **15**, 1528–1542 (2009).
29. Zhang, J. *et al.* hnRNPs and ELAVL1 cooperate with uORFs to inhibit protein translation. *Nucleic Acids Research* (2016) doi:10.1093/nar/gkw991.
30. Degrauwe, N., Suvà, M.-L., Janiszewska, M., Riggi, N. & Stamenkovic, I. IMPs: An RNA-binding protein family that provides a link between stem cell maintenance in normal development and cancer. *Genes & Development* **30**, 2459–2474 (2016).

31. Sauer, M. *et al.* DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nature Communications* **10**, 2421 (2019).
32. Yoo, J.-S. *et al.* DHX36 enhances RIG-i signaling by facilitating PKR-mediated antiviral stress granule formation. *PLoS Pathogens* **10**, e1004012 (2014).
33. Pietras, Z. *et al.* Dedicated surveillance mechanism controls g-quadruplex forming non-coding RNAs in human mitochondria. *Nature Communications* **9**, 2558 (2018).
34. Schwerk, J. *et al.* RNA-binding protein isoforms ZAP-s and ZAP-l have distinct antiviral and immune resolution functions. *Nature Immunology* **20**, 1610–1620 (2019).
35. Guo, X., Carroll, J.-W. N., Macdonald, M. R., Goff, S. P. & Gao, G. The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. *Journal of Virology* **78**, 12781–12787 (2004).
36. Guo, X., Ma, J., Sun, J. & Gao, G. The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 151–156 (2007).
37. Todorova, T., Bock, F. J. & Chang, P. Poly(ADP-ribose) polymerase-13 and RNA regulation in immunity and cancer. *Trends in Molecular Medicine* **21**, 373–384 (2015).
38. Charron, G., Li, M. M. H., MacDonald, M. R. & Hang, H. C. Prenylome profiling reveals S-farnesylation is crucial for membrane targeting and antiviral activity of ZAP long-isoform. *Proceedings of the National Academy of Sciences* **110**, 11085–11090 (2013).
39. Vyas, S., Chesarone-Cataldo, M., Todorova, T., Huang, Y.-H. & Chang, P. A systematic analysis of the PARP protein family identifies new functions critical for cell physiology. *Nat Commun* **4**, (2013).
40. Kwak, J. E., Wang, L., Ballantyne, S., Kimble, J. & Wickens, M. Mammalian GLD-2 homologs are poly(a) polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 4407–4412 (2004).
41. Liang, X. *et al.* Structural snapshots of human pre-60S ribosomal particles before and after nuclear export. *Nature Communications* **11**, 3542 (2020).
42. Blanco-Melo, D. *et al.* Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036-1045.e9 (2020).
43. Sun, L. *et al.* In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. *Cell* **184**, 1865-1883.e20 (2021).

44. 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* **12**, (2021).
45. 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* **16**, e1008845 (2020).
46. Nchioua, R. *et al.* SARS-CoV-2 is restricted by zinc finger antiviral protein despite preadaptation to the low-CpG environment in humans. *mBio* **11**, (2020).
47. Manktelow, E., Shigemoto, K. & Brierley, I. Characterization of the frameshift signal of edr, a mammalian example of programmed -1 ribosomal frameshifting. *Nucleic Acids Research* **33**, 1553–1563 (2005).
48. Napthine, S., Hill, C. H., Nugent, H. C. M. & Brierley, I. Modulation of viral programmed ribosomal frameshifting and stop codon readthrough by the host restriction factor shiftless. *Viruses* **13**, (2021).
49. Bhatt, P. R. *et al.* Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science* **372**, 1306–1313 (2021).
50. Hill, C. H. *et al.* Structural studies of cardiovirus 2A protein reveal the molecular basis for RNA recognition and translational control. *BioRxiv* (2020) doi:10.1101/2020.08.11.245035.
51. 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* **28**, 747–754 (2021).
52. Neupane, K. *et al.* Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers. *Nature Communications* **12**, 4749 (2021).
53. Schlick, T., Zhu, Q., Jain, S. & Yan, S. Structure-altering mutations of the SARS-CoV-2 frame shifting RNA element. *BioRxiv* (2020) doi:10.1101/2020.08.28.271965.
54. Mouzakis, K. D., Lang, A. L., Vander Meulen, K. A., Easterday, P. D. & Butcher, S. E. HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Research* **41**, 1901–1913 (2013).
55. 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. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 24303–24309 (2019).
56. Ficarelli, M. *et al.* KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *eLife* **8**, (2019).

57. Li, M. M. H. *et al.* TRIM25 enhances the antiviral action of zinc-finger antiviral protein (ZAP). *PLoS Pathogens* **13**, e1006145 (2017).
58. 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 Pathogens* **16**, e1008845 (2020).
59. Luo, X. *et al.* Molecular Mechanism of RNA Recognition by Zinc-Finger Antiviral Protein. *Cell Reports* **30**, 46-52.e4 (2020).
60. Turakhiya, A. *et al.* ZFAND1 recruits p97 and the 26S proteasome to promote the clearance of arsenite-induced stress granules. *Molecular Cell* **70**, 906-919.e7 (2018).
61. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research* **41**, D377-86 (2013).
62. Pan, D. *et al.* CDK-regulated dimerization of M18BP1 on a Mis18 hexamer is necessary for CENP-a loading. *eLife* **6**, (2017).
63. Loughran, G., Howard, M. T., Firth, A. E. & Atkins, J. F. Avoidance of reporter assay distortions from fused dual reporters. *RNA (New York)* **23**, 1285–1289 (2017).
64. Arai, R., Ueda, H., Kitayama, A., Kamiya, N. & Nagamune, T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering* **14**, 529–532 (2001).
65. Meyer, S. *et al.* Multi-host expression system for recombinant production of challenging proteins. *Plos One* **8**, e68674 (2013).
66. Andreou, A. I. & Nakayama, N. Mobius assembly: A versatile golden-gate framework towards universal DNA assembly. *Plos One* **13**, e0189892 (2018).
67. Götzke, H. *et al.* The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nature Communications* **10**, 4403 (2019).
68. Brennan, T. V., Lin, L., Huang, X. & Yang, Y. Generation of luciferase-expressing tumor cell lines. *Bio-protocol* **8**, (2018).
69. Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA (New York)* **9**, 493–501 (2003).
70. Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F. & Atkins, J. F. A dual-luciferase reporter system for studying recoding signals. *RNA (New York)* **4**, 479–486 (1998).
71. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).

72. Kechin, A., Boyarskikh, U., Kel, A. & Filipenko, M. cutPrimers: A new tool for accurate cutting of primers from reads of targeted next generation sequencing. *Journal of Computational Biology* **24**, 1138–1143 (2017).
73. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with bowtie 2. *Nature Methods* **9**, 357–359 (2012).
74. Incarnato, D., Morandi, E., Simon, L. M. & Oliviero, S. RNA framework: An all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications. *Nucleic Acids Research* **46**, e97 (2018).
75. Kingston, R. E., Chomczynski, P. & Sacchi, N. Guanidine methods for total RNA preparation. *Current Protocols in Molecular Biology* **Chapter 4**, Unit4.2 (2001).
76. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A, Vizcaíno JA (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res 47(D1):D442-D450
77. Buck, S., Pekarek, L. & Caliskan, N. POTATO: An automated pipeline for batch analysis of optical tweezers data. *BioRxiv* (2021) doi:10.1101/2021.11.11.468103.

Chapter 6

"A virus is a piece of bad news wrapped in a protein"

Sir Peter Medawar

6. 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)

Manuscript modifications: In this thesis, I have only included the supplementary materials I significantly contributed to. The rest of the supplementary data, including the tables, is available online or provided as source data with this thesis.

6.1. Abstract

Programmed –1 ribosomal frameshifting (PRF) in cardioviruses is activated by the 2A protein, a multi-functional virulence factor that also inhibits cap-dependent translational initiation. Here we present the X-ray crystal structure of 2A and show that it selectively binds to a pseudoknot-like conformation of the PRF stimulatory RNA element in the viral genome. Using optical tweezers, we demonstrate that 2A stabilises this RNA element,

likely explaining the increase in PRF efficiency in the presence of 2A. Next, we demonstrate a strong interaction between 2A and the small ribosomal subunit and present a cryo-EM structure of 2A bound to initiated 70S ribosomes. Multiple copies of 2A bind to the 16S rRNA where they may compete for binding with initiation and elongation factors. Together, these results define the structural basis for RNA recognition by 2A, show how 2A-mediated stabilisation of an RNA pseudoknot promotes PRF, and reveal how 2A accumulation may shut down translation during virus infection.

6.2. Introduction

PRF is a translational control strategy employed by many RNA viruses, where it ensures the production of proteins in optimal ratios for efficient virus assembly and enables viruses to expand their coding capacity through the utilisation of overlapping ORFs (reviewed in[1-3]). In canonical PRF, elongating ribosomes pause over a heptanucleotide "slippery sequence" of the form X_XXY_YYZ when they encounter a "stimulatory element" about 5–9 nucleotides downstream in the mRNA. During this time, a –1 frameshift may occur if codon-anticodon re-pairing takes place over the X_XXY_YYZ sequence: the homopolymeric stretches allow the tRNA in the P-site to slip from XXY to XXX, and the tRNA in the A-site to slip from YYZ to YYY[4-7]. A diverse array of stem-loops and pseudoknots are known to induce frameshifting, and the stability, plasticity and unfolding kinetics of these RNA elements are thought to be the primary determinants of PRF efficiency[8-10], along with the thermodynamic stability of the codon-anticodon interactions[6]. Cardioviruses present a highly unusual variation to conventional viral PRF in which the virally encoded 2A protein is required as an essential *trans*-activator[11,12]. 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[12]. 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[12].

2A is a small, basic protein (~17 kDa; 143 amino acids; pI ~9.1) generated by 3C-mediated proteolytic cleavage at the N-terminus[13] and Stop-Go peptide release at the C-terminus[14]. Despite the identical name, it has no homology to any other picornavirus "2A" protein[15], nor any other protein of known structure. The PRF-stimulatory activity of 2A is related to its ability to bind to the RNA stimulatory element[12]. However, 2A also binds to 40S ribosomal subunits[16], inhibits apoptosis[17] 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)[16],

thereby interfering with eIF4F assembly[18]. However, the absence of structural data has precluded a definitive molecular characterisation of this multi-functional protein, and the mechanism by which it recognises RNA elements and stimulates frameshifting remains obscure.

Here we present the crystal structure of 2A from encephalomyocarditis virus (EMCV), revealing a novel RNA-binding fold that we term a “beta-shell”. We show that 2A binds directly to the frameshift-stimulatory element in the viral RNA with nanomolar affinity and equimolar stoichiometry, and we define the minimal RNA element required for binding. Through site-directed mutagenesis and the use of single-molecule optical tweezers, we study the dynamics of this RNA element, both alone and in the presence of 2A. By observing short-lived intermediate states in real-time, we demonstrate that the EMCV stimulatory element exists in at least two conformations and 2A binding stabilises one of these, an RNA pseudoknot, increasing the force required to unwind it. Finally, we report a direct interaction of 2A with both mammalian and bacterial ribosomes. High-resolution cryo-electron microscopy (cryo-EM) characterisation of 2A in complex with initiated 70S ribosomes reveals a multivalent binding mechanism and defines the molecular basis for RNA recognition by the 2A protein. It also reveals a likely mechanism of 2A-associated translational modulation, by competing for ribosome binding with initiation factors and elongation factors. Together, our work provides a new structural framework for understanding protein-mediated frameshifting and 2A-mediated regulation of gene expression.

6.3. Results

Structure of EMCV 2A reveals an RNA-binding fold

Following recombinant expression in *E. coli*, purified 2A was analysed by SEC-MALS, revealing a predominantly monodisperse, monomeric sample (**Fig. 6.1a, b**; observed mass 18032.8 Da vs calculated mass 17930.34 Da), with a small proportion of dimers (observed mass 40836.0 Da). We crystallised the protein and determined the structure by multiple-wavelength anomalous dispersion analysis of a selenomethionyl derivative. The asymmetric unit (ASU) of the $P6_222$ cell contains four copies of 2A related by non-crystallographic symmetry (NCS), and the structure was refined to 2.6 Å resolution (**Table 6.S1**). Unexpectedly, the four molecules are arranged as a pair of covalent ‘dimers’ with an intermolecular disulfide bond forming between surface-exposed cysteine residues (C111). This arrangement is likely an artefact of crystallisation, which took >30 days, possibly due to the gradual oxidation of C111 promoting formation of the crystalline lattice.

The N-terminal 10–12 residues are disordered in all chains except B, in which they make long-range contacts with another chain. Similarly, C-terminal residues beyond 137 are absent or poorly ordered in all chains.

2A adopts a compact, globular fold of the form $\beta_3\alpha\beta_3\alpha\beta$ (**Fig. 6.1c**). Searches of PDBeFOLD[19], DALI[20] and CATHEDRAL[21] databases failed to reveal structural homology to any other protein, so we term this fold a “beta shell”. The most striking feature is a curved, seven-stranded anti-parallel beta sheet (**Fig. 6.1d**). The concave face of the beta sheet is supported by tight packing against the two alpha helices: together, this comprises the hydrophobic core of the fold. In contrast, the solvent-exposed convex face and surrounding loops are enriched with arginine, lysine and histidine residues, conferring a strong positive electrostatic surface potential at physiological pH. Superposition of the four NCS-related chains and an analysis of the atomic displacement factors reveals regions of flexibility (**Fig. 6.1e, f**). In addition to the N- and C- termini, the β2-loop-β3 region (residues 28–37) exists in multiple conformations that deviate by up to 5.8 Å in the position of the C_α backbone. Similarly, the arginine-rich loop between β5 and β6 (“arginine loop”, residues 93–100) is mobile, with backbone deviations of up to 4.5 Å.

Several previous studies have described mutations, truncations or deletions in EMCV 2A that affect its activity[22-24]. Many of the truncations would severely disrupt the fold and the results obtained with these mutants should be interpreted with caution. However, the loop truncation ($2A_{\Delta 94\text{-}100}$) and point mutations made by Groppo *et al.*[23] would have only minor effects (**Fig. 6.S1**). Notably, in 2A, a C-terminal YxxxxLΦ motif predicted to bind eIF4E is within a beta strand, whereas the equivalent motif in 4E-BP1 is alpha-helical[25]. As a result, Y129 is partially buried and distal to both L134 and I135. Overlay of our 2A structure with the structure of the eIF4E:4E-BP1 complex indicates that without a significant conformational change, this motif is unlikely to represent the mechanism by which 2A recognises eIF4E (**Fig. 6.S1**).

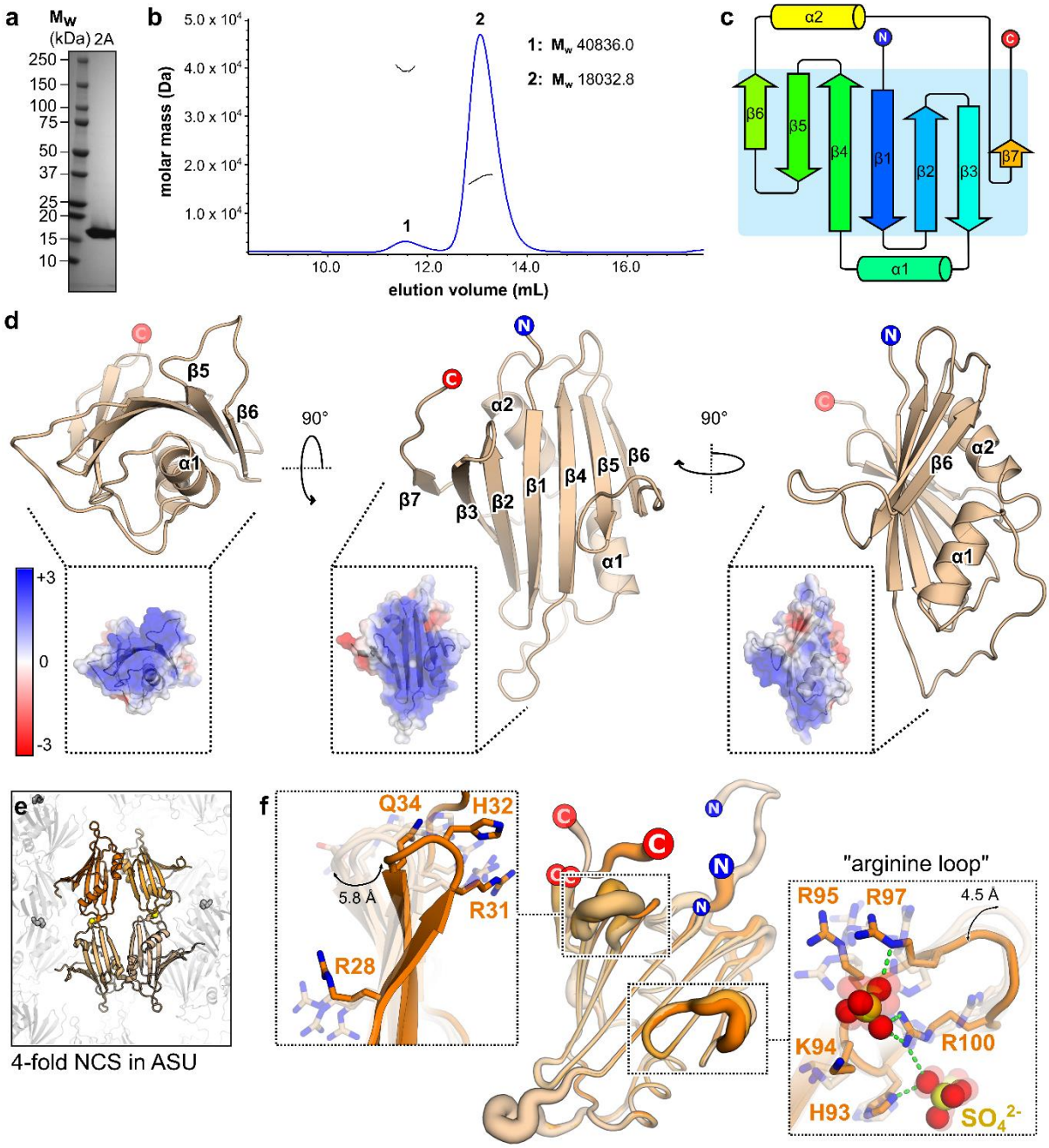

Figure 6.1 2A adopts a highly basic RNA-binding fold with intrinsic flexibility. a, SDS-PAGE analysis of EMCV 2A (Coomassie). Representative gel from five independent purifications. **b,** SEC-MALS analysis of 2A. The differential refractive index is shown across the elution profile (blue) and weight-averaged molar masses of the indicated peaks are listed. **c,** Topological diagram of "beta-shell" fold: a curved central sheet comprising seven antiparallel beta strands, supported by two helices. **d,** Crystal structure of EMCV 2A in three orthogonal views. N- and C- termini are indicated. <*Inset*> Electrostatic surface potential calculated at pH 7.4, coloured between +3 (blue) and -3 (red) kT/e^-. **e,** Four molecules of 2A are present in the asymmetric unit of the crystal, arranged as two pairs of disulfide-linked dimers (spheres). **f,** Superposition of the four NCS-related 2A chains in **e** reveals regions of conformational flexibility. The width of the cartoon is proportional to atomic B-factor. <*Insets*> Close-up view of surface loops exhibiting the greatest variation per molecule.

Flexible sidechains are shown as sticks, and the Cα backbone deviation is indicated in Å. The positions of two sulfate ions from the crystallisation buffer are indicated with spheres. Source data are provided as a Source Data file.

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. 6.2a**). We have previously demonstrated that three conserved cytosines in the loop are essential for 2A binding[12] (**Fig. 6.2a**). 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. 6.2b**). These were fluorescently labelled at the 5′ end, and their binding to 2A was analysed by electrophoretic mobility shift assay (EMSA; **Fig. 6.2c**) and microscale thermophoresis (MST; **Fig. 6.2d, Table 6.S2**).

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. Truncating the disordered N- and C- termini of 2A, or mutating the disulfide-forming C111 residue has no effect on RNA binding (**Fig. 6.S2**). To investigate stoichiometry, we performed an isothermal titration calorimetry (ITC) analysis of the interaction between 2A and EMCV 6 (**Fig. 6.S2**). Equimolar binding was observed, with a measured K_D (246 ± 72 nM), similar to those obtained using MST. The dominance of enthalpy (ΔH, –13.9 ± 0.81 kcal/mol) to the overall free energy of binding (ΔG, –9.02 kcal/mol) indicates an interaction mechanism driven by hydrogen bond or electrostatic contact formation. Finally, reciprocal MST experiments with fluorescently labelled 2A and unlabelled RNA yielded similar K_D values (**Fig. 6.S2**, **Table 6.S2**).

We next asked whether these small RNAs could act as competitors to sequester 2A and reduce the efficiency of PRF in rabbit reticulocyte lysate (RRL) *in vitro* translation reactions programmed with a frameshift reporter mRNA (**Fig. 6.S2**). Indeed, when unlabelled EMCV 1, 3 and 6 were added in excess, they were able to compete with the stimulatory element present in the reporter, thereby reducing the amount of the –1 frame product. In contrast, EMCV 2, 4 and 5 had no such effect, reinforcing the results of direct binding experiments.

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. To test this, we carried out mutagenesis of the 5′ extension and loop C-triplet. Individually, G7C and C37G mutations both reduce 2A-dependent PRF to near-background levels (**Fig. 6.S3**). However, in combination, the G7C+C37G double mutation restores PRF to wild-type levels, and EMSA experiments with these mutants confirm that this is due to inhibition and restoration of 2A binding. Together, this demonstrates the likelihood of a base-pair between positions 7 and 37 that is necessary to form a conformation that 2A selectively recognises. Using this base pair as a restraint, RNA structure prediction[26,27] reveals a pseudoknot-like fold (**Fig. 6.S3**).

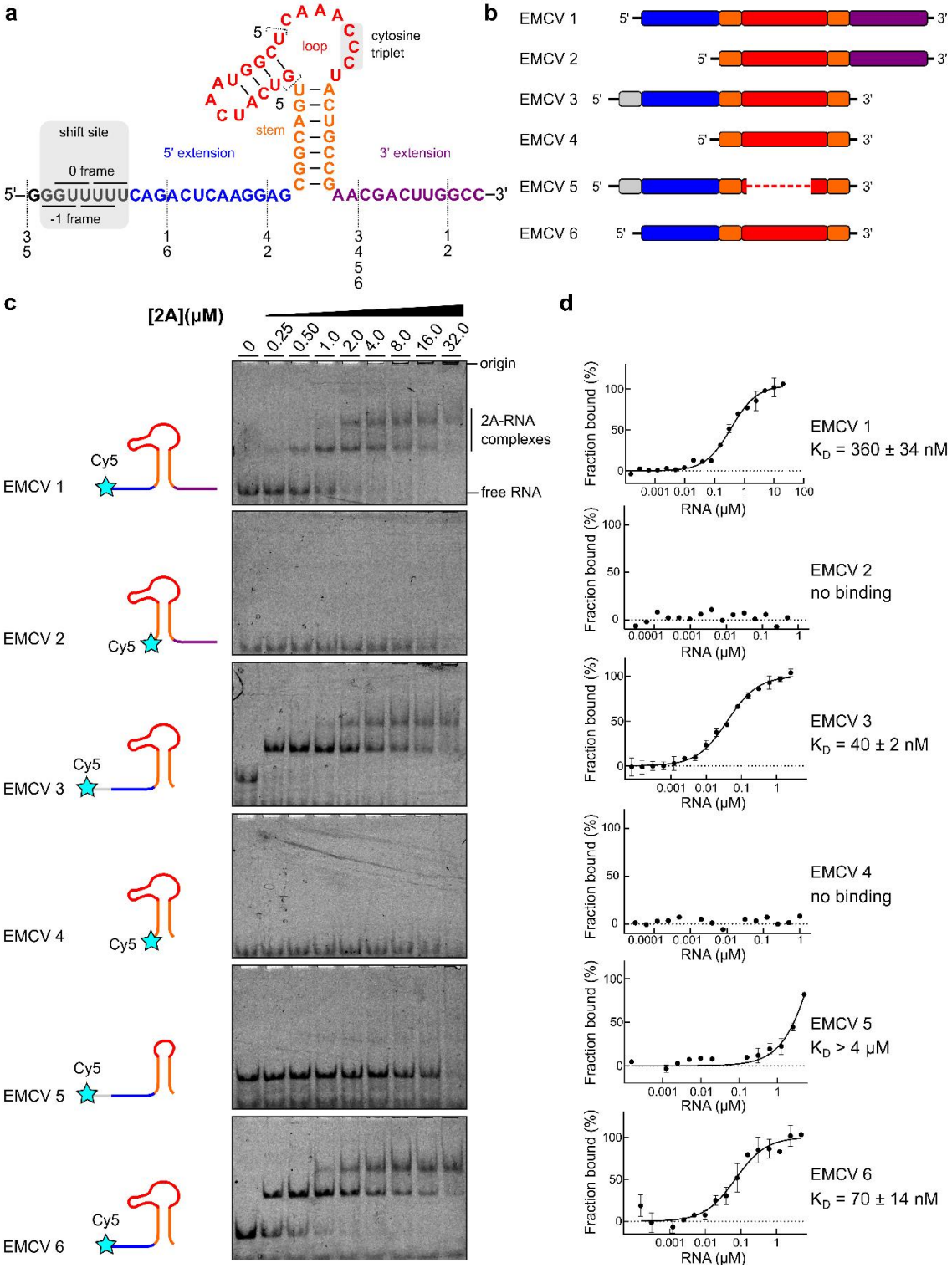

Figure 6.2 2A binds to a minimal 47 nt element in the viral RNA. a-b, Sequences and schematic diagrams of the EMCV 1–6 constructs used to assay 2A binding. **c,** EMSA analyses showing that removal of the 5′ extension (blue) disables 2A binding. **d,** Microscale thermophoresis (MST) was used to quantify 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.

Single-molecule measurements of stimulatory element unwinding reveal several conformations

Information is limited in ensemble measurements of RNA-protein interactions due to molecular averaging. To further explore the effects of 2A on unfolding and refolding of individual EMCV RNA molecules, we used optical tweezers (**Fig. 6.3a**). In force-ramp experiments, a single RNA molecule is gradually stretched and relaxed in several cycles at a constant pulling rate. The applied force allows the RNA molecule to transition between folded and unfolded states, and sudden changes in recorded force-distance trajectories indicate transitions between RNA conformations (**Fig. 6.3c, d**)[28-30]. By mathematically fitting each force-distance curve (**Methods**) we can obtain information on the physical properties of the RNA such as the change in the contour length (maximum possible extension), which indicates whether our data are physically consistent with predicted structures of the EMCV RNA. In addition to the pseudoknot (40 nt, discussed above), *mfold*[31] suggested two other possible conformations for the frameshift stimulatory element: a stem loop (35 nt) and an extended stem-loop with additional interactions between 5' and 3' flanking regions (49 nt) (**Fig. 6.3b, Table 6.S6**). Alongside the wild-type EMCV RNA sequence (WT), we also used a mutant with a substitution in the cytosine triplet (CUC) which is known be crucial for 2A binding and PRF[12] (**Fig. 6.3a**; lower).

We initially monitored the unfolding and refolding of WT and CUC RNAs in the absence of 2A. In WT RNA, the majority of force-distance (FD) trajectories were characterized by a single rip at 9.3 ± 2.3 pN force (**Fig. 6.3c, e**). Upon release of the force, the molecules readily refolded at 6.5 ± 3.0 pN, showing that the process is reversible (**Fig. 6.3c, f, Table 6.S3**). The change in contour length calculated from the fits was approximately 26.3 ± 5.4 nm (**Fig. 6.S4, Table 6.S3**) corresponding to a length of 46 single-stranded nucleotides. This is in close agreement with the predicted 49 nt long extended stem-loop formed by the EMCV PRF RNA (**Fig. 6.3b, Table 6.S6**). Interestingly, we observed similar (un)folding trajectories with the CUC RNA, with a rip occurring at 8.6 ± 4.2 pN and a contour length change of about 27.2 ± 4.3 nm (**Fig. 6.3e, f, Fig. 6.S4, Table 6.S3**), suggesting both RNAs would essentially fold into a stem-loop of similar length.

In a small fraction of WT FD trajectories (~12 %) we observed a single unfolding event at higher forces above 20 pN, while refolding was unchanged (6.5 ± 3.0 pN), suggesting the existence of a WT conformer with resistance to unfolding. Indeed, the putative EMCV pseudoknot would comprise 40 nucleotides, and lead to an expected difference of 23 nm in contour length upon unfolding (**Fig. 6.S4, Table 6.S6**). Since both the predicted pseudoknot and extended-stem loop are of similar length, and the distributions of contour length change are quite broad, this parameter is not precise

enough to unambiguously distinguish between these conformations. On the other hand, the resistance to unfolding and hysteresis during refolding are well-known characteristics of more complex structures such as pseudoknots[9,32], and this is also consistent with our mutational analysis (**Fig. 6.S3;** discussed above). Furthermore, these higher force unfolding events are not observed in the CUC mutant, which is very unlikely to form a pseudoknot (**Fig. 6.S4, Table 6.S3**).

Next, we compared the energetics of folding and unfolding of the conformers. In optical tweezer experiments, the work of unfolding is the work required to extend the folded RNA construct (dsDNA:RNA handles and dsRNA) minus the work required to extend the fully unfolded (dsDNA:RNA handles and ssRNA) construct ($W=W_{ds}-W_{ds+ss}$). Accordingly, free energy values of WT and CUC constructs were calculated as −13.6 ± 4.6 and −14.5 ± 4.7 kcal/mol, respectively, which are close to the *mfold*-predicted Gibbs free energy values for the stem loop (−14 ± 0.7 kcal/mol) and extended stem-loop (−16.2 ± 0.8 kcal/mol) (**Table 6.S3**). This further supports the view that the EMCV WT and CUC mutant RNAs predominantly fold into the predicted stem-loops.

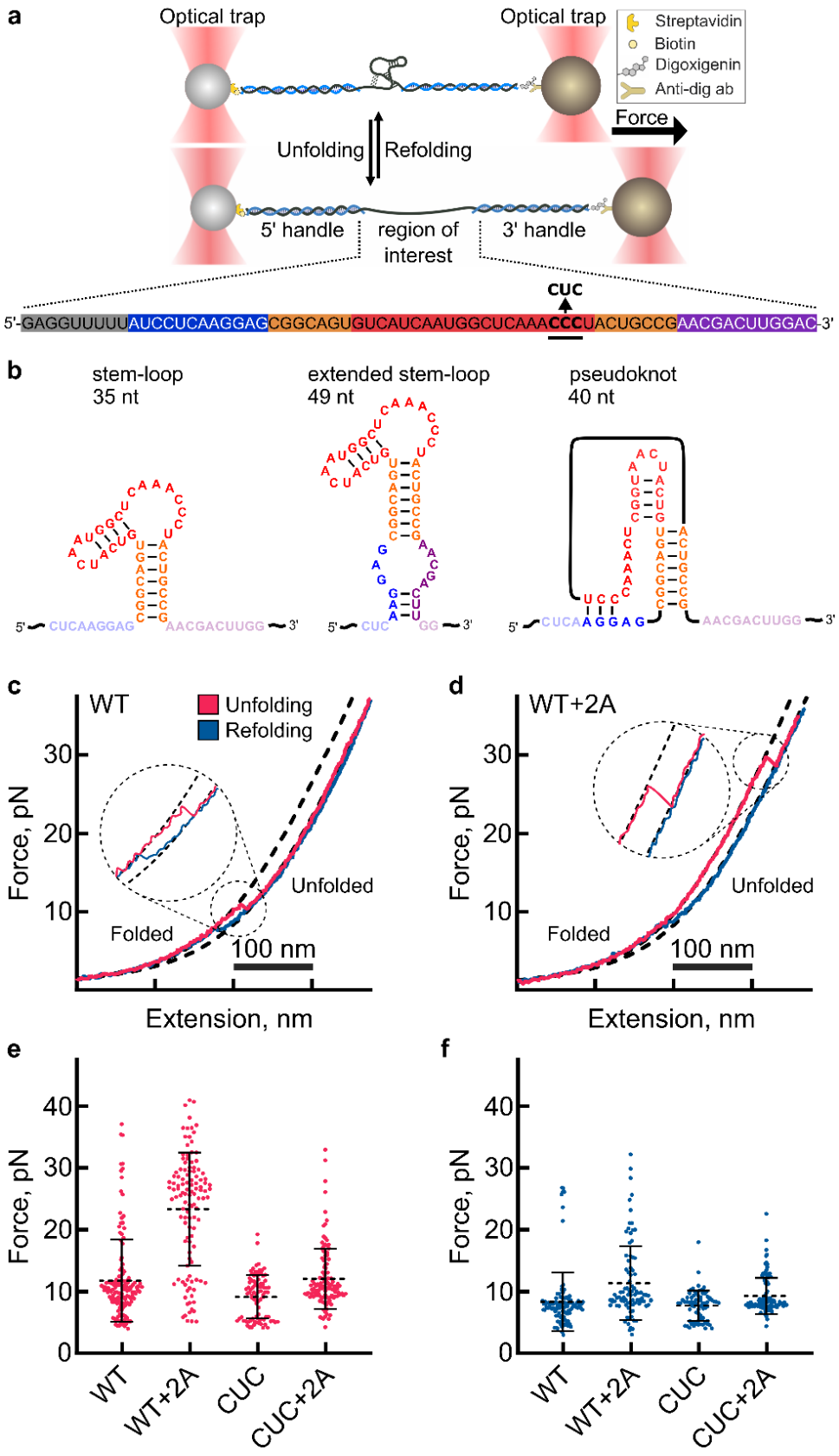

Figure 6.3 Conformations of EMCV frameshifting RNA and effect of 2A on RNA unwinding. **a,** <*Upper*> Schematic diagram illustrating the optical tweezer experiments (right). RNA is hybridized to ssDNA handles and immobilised on beads. These are used to exert pulling force on the RNA with a focused laser beam. <*Lower*> Primary sequence of the construct used in optical

tweezer experiments, colour coded as in **Fig. 6.2**. The location of the cytosine triplet (wild-type, WT) and point mutation (CUC) is indicated. **b,** Predicted conformations of the RNA construct in a. The number of nucleotides involved in each folded structure is indicated. Also see **Table 6.S6**. **c,** Representative force-distance curves of the unfolding (pink) and refolding (blue) transitions of the wild type (WT) CCC RNA element. **d,** Representative force-distance curves of the unfolding (pink) and refolding (blue) transitions of the wild type (WT) CCC RNA element in the presence of 300 nM 2A protein. **e**, Global analysis of all the unfolding force trajectories. Number of individual measurements are WT=117, WT+2A=104, CUC=85, CUC+2A=109. Data (black line) are presented as mean values ± SD error bars. **f**, Global analysis of all refolding force trajectories. Number of individual measurements are WT=111, WT+2A=89, CUC=74, CUC+2A=97. Data (black line) are presented as mean values ± SD error bars. Source data are provided as a Source Data file.

2A favours the formation of an alternative conformation with resistance to mechanical unwinding

We next tested how 2A binding influences RNA stability and resistance of RNA to mechanical unwinding. For the wild-type RNA, global analysis of the unfolding forces reveals a 2A-induced stabilisation, which increased the fraction of unfolding events at higher forces (27.0 ± 4.2 pN) (**Fig. 6.3d, e, Fig. 6.S4, Table 6.S3**). However, refolding of the RNA was mostly unaffected (**Fig. 6.3d, f, Table 6.S3**). Probability distributions of the change in contour length show a peak at around 20.0 ± 3.4 nm (**Fig. 6.S4, Table 6.S3**) which may reflect the unfolding of either the stem-loop (21 nm expected) or the proposed pseudoknot conformation (23 nm expected) (**Fig. 6.3b, Fig. 6.S4**, **Table 6.S6**).

Subsequently, we examined 2A binding to the CUC mutant RNA (**Fig. 6.S4**). In contrast to the wild type RNA, within this population we did not observe any stabilisation in the presence of 2A (**Fig. 6.3e, Fig. 6.S4**). Thus, the unfolding and refolding force distributions overlap with those observed for CUC RNA in absence of 2A (**Fig. 6.3e**, **f**, **Fig. 6.S4**, **Table 6.S3**). We observed a small shift in the distribution of contour length changes towards lower values, which could be either due to non-specific interactions or stochastic noise. Overall, the lack of effect of 2A on the CUC RNA agreed well with the ensemble analysis of 2A:RNA interactions.

To further dissect the effect of 2A on EMCV RNAs, we calculated the work performed on the WT and CUC RNAs during (un)folding in the presence of 2A (**Fig. 6.S4**). For CUC RNA with 2A, the unfolding and refolding work distributions were largely overlapping, so the process can be considered reversible. We obtained a free energy value of −15.5 ± 5.0 kcal/mol, which is within the range of *mfold* predicted free energy values for the CUC stem-loop (−14 ± 0.7 kcal/mol) and the extended stem-loop (–16.2 ± 0.8 kcal/mol). For WT RNA with 2A, the stabilisation effect shifted the calculated free

energy to −26.5 ± 8.7 kcal/mol, thus moving the system away from equilibrium (**Table 6.S3**)[33,34]. The 2A-induced decrease in free energy of the wild-type RNA may be a combination of stabilisation induced by protein binding, and a change in RNA conformation. Taken together, our results support that 2A binding stabilises the EMCV stimulatory RNA element and increases its resistance to mechanical unwinding.

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[16]. 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. 6.4a, 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. 6.4c**), but not 50S (**Fig. 6.4d**). 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. 6.4e, f**).

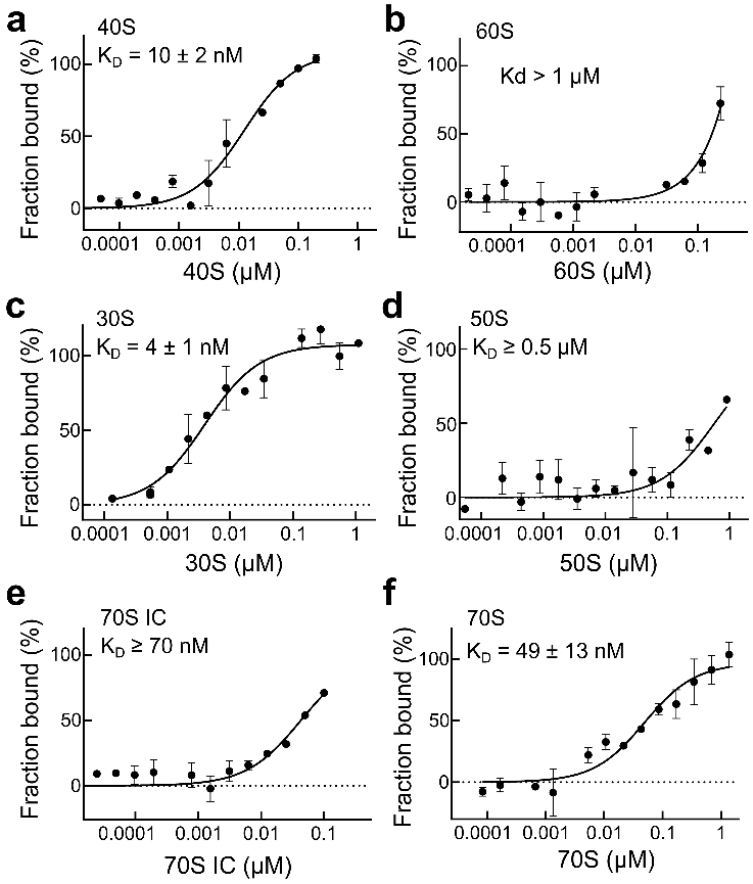

Figure 6.4 2A binds directly to eukaryotic and prokaryotic ribosomes. a, MST binding curves and apparent K_D values using unlabelled 40S subunits at a concentration range of 20 pM – 0.4 μM. All measurements were repeated as two independent experiments and error bars represent the standard deviation from 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. Source data are provided as a Source Data file.

Prokaryotic ribosomes are responsive to 2A-mediated frameshifting

Prokaryotic translation systems are well-established models for studying eukaryotic PRF signals[35,36] but it is unknown whether they can support protein-dependent PRF. To address this, we measured the efficiency of the EMCV signal in a reconstituted prokaryotic translation system and in *E. coli* S30 extracts using frameshift reporter mRNAs (**Fig. 6.S5**). In each case, 2A-dependent PRF was observed, with ~7% of ribosomes changing frame. Mutagenesis of either the shift site or the CCC triplet disabled PRF. Shortening the length of the spacer to one more optimal for prokaryotic ribosomes (from

13 to 12 nt) doubled PRF efficiency to ~15%, comparable to that measured in eukaryotic *in vitro* translation systems (20%)[12]. High concentrations of 2A also had an inhibitory effect on translation, similar to that seen in eukaryotic systems.

Cryo-EM structure of a 2A-ribosome complex reveals the structural basis for RNA recognition and translational pathology

Having validated the use of prokaryotic ribosomes as a model system to study protein-dependent PRF, we prepared complexes between 2A and the initiated 70S ribosomes and imaged them by cryo-EM (**Fig. 6.5a, Table 6.S4**). After processing (**Fig. 6.S6**), the final 3D reconstruction produced a density map of 2.7 Å resolution and revealed three copies of 2A bound directly to 16S rRNA of the 30S subunit in a tripartite cluster (**Fig. 6.5b, c**). The local resolution for 2A was sufficient to allow sidechain modelling and refinement. All three 2A molecules use the same RNA-binding surface (comprising variations of R46, K48, K50, K73, K94, R95 and R97) (**Fig. 6.5d**), to recognise the ribose phosphate backbone through numerous polar and electrostatic contacts (**Fig. 6.6a–c**). We mutated this putative interaction surface (**Fig. 6.S5**) and observed reduced binding to both the stimulatory element RNA and mammalian ribosome subunits and a decreased activity in stimulating PRF *in vitro*. $2A_{R95A/R97A}$ was completely functionally defective, whilst $2A_{K73A}$ and $2A_{R46A/K48A/K50A}$ exhibited moderate and mild effects, respectively.

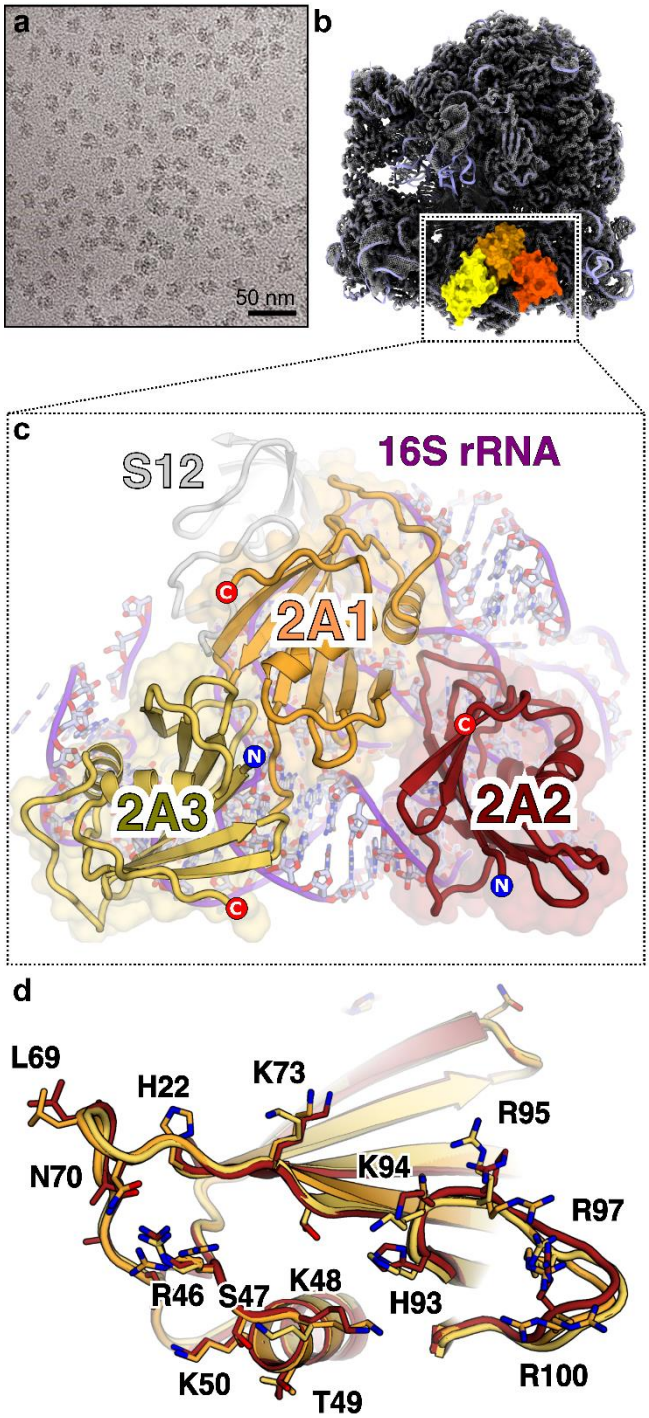

Figure 6.5 2A binds to the 70S ribosome via interactions with the 16S rRNA. a, Cryo-EM analysis of a complex formed between initiated *E. coli* 70S ribosomes and EMCV 2A. Images (× 75,000) were recorded on a Titan Krios microscope. Representative micrograph from dataset of 5730 images. **b,** Cryo-EM map at 2.7 Å resolution after focused classification and refinement. Three copies of 2A (orange, red, yellow) are bound to the 16S rRNA of the small (30S) subunit (blue ribbon). **c**, Close-up view of the 2A binding site. Ribbon diagrams of 2A (coloured as above) and ribosomal RNA (purple) are shown. Protein N- and C- termini are labelled. **d,** Superposition of the three copies of 2A reveals a common RNA-binding surface with conformational flexibility. Residues involved in rRNA binding are labelled and shown as sticks.

By comparing the quality of both the overall density for each 2A molecule, and the side-chain density at the interaction surface, we can rank the three binding sites 2A1 > 2A2 > 2A3 in order of likely affinity. 2A1 is the most well-ordered molecule, and the 2A1 binding site on the rRNA is also the most conserved between prokaryotic and mammalian ribosomes (**Fig. 6.S7**) This is therefore likely the most physiologically relevant site, and it is possible that 2A2 and 2A3 represent lower-affinity sites (a ~40-fold molar excess of 2A was used to prepare grids). 2A1 exemplifies the critical role of the "arginine loop" (**Fig. 6.6d**). R95, R97 and R100 side chains are inserted into a ~90° junction between helices 3 and 4, forming a network of electrostatic interactions that bridge the phosphate groups on both strands. This is further stabilised by the guanidinium groups stacking against each other and exposed bases (G38) (**Fig. 6.6d**). Arginine loop residues also form polar and electrostatic contacts at the 2A2 and 2A3 interfaces (**Fig. 6.6e, f**). Whilst base-specific contacts are rare, 2A2 interacts with U485 which is normally flipped out of helix 17 (**Fig. 6.S8**). Superposition of the rRNA binding sites failed to reveal a common structural motif for RNA recognition (**Fig. 6.S7**), thus conformational plasticity of side chains at the RNA-binding surface (**Fig. 6.5d**) explains how this protein can recognise a several RNA targets. There are also intermolecular contacts between 2A protomers, consistent with our observations of multimers by SEC-MALS (**Fig. 6.1b**) and EMSA (**Fig. 6.2c**). In a subset of the data, a fourth copy of 2A (2A4) was identified to bind helix 33 of the 16S rRNA 'beak' in the 30S head. Although local resolution was only sufficient for docking, 2A4 uses the same RNA-binding surface to recognise the distorted helical backbone (**Fig. 6.S8**).

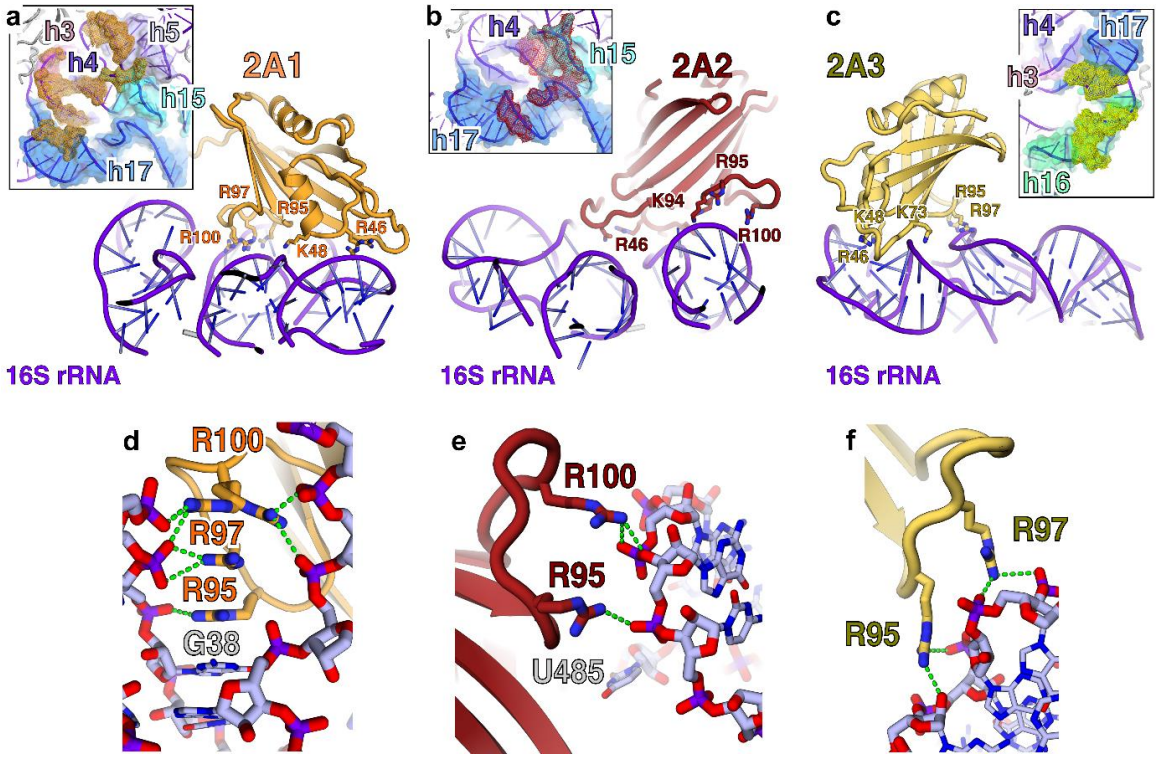

Figure 6.6. The 'arginine loop' plays a central role in RNA recognition. a-c, Details of rRNA recognition by 2A. For each copy of 2A, selected residues involved in interactions are labelled and shown as sticks <*Insets*> View of the rRNA surface bound by each copy of 2A. The rRNA helices are colour-coded and labelled. The 2A contact surface is shown as a coloured mesh (orange, red and yellow, respectively). **d–f,** Close-up view of interactions between the 2A 'arginine loop' residues (R95, R97 and R100) and the rRNA backbone (sticks) for each copy of 2A (orange, red, yellow). Polar or electrostatic contacts are indicated by a green dashed line.

The ribosome is in an unrotated state that would normally be elongation competent, with fMet-tRNA$_i$ base-paired to the initiator codon in the P-site and mRNA available for amino-acyl tRNA delivery to the A-site[37] (**Fig. 6.7a**). There are no 2A-induced rearrangements at the decoding centre (**Fig. 6.S8**) but the presence of 2A on the 30S subunit occludes the binding site for translational GTPases. 2A1 occupies a position that would severely clash with domain II of EF-G in both compact and extended pre- and post-translocation states[38,39] (**Fig. 6.7b**). It also makes direct hydrophobic contacts with the face of S12 that would normally interact with domain III of EF-G. This 2A interaction surface on S12 is directly adjacent to the binding site for antibiotic dityromycin, which inhibits translocation by steric incompatibility with the elongated form of EF-G[40] (**Fig. 6.S8**). 2A1 would also clash significantly with domain II of EF-Tu during delivery of aminoacyl tRNAs to the A-site[41,42] (**Fig. 6.7c**). In a similar way, 2A2 would be detrimental to both EF-G and EF-Tu binding (**Fig. 6.7b, c**). We therefore predict that 2A binding would be inhibitory to

elongation and potentially initiation, via competition with IF2 during pre-initiation complex assembly[43].

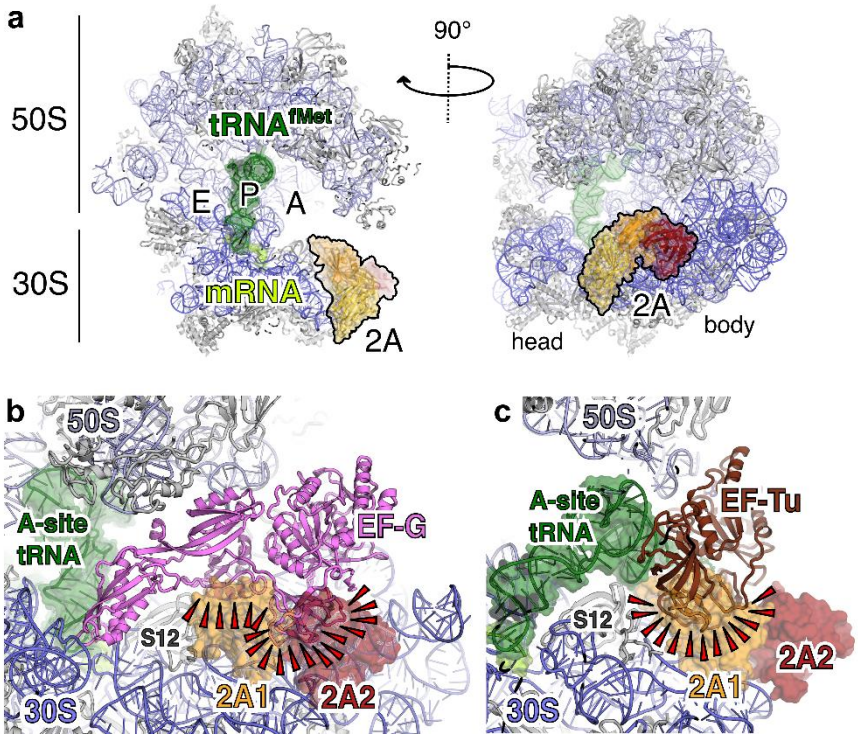

Figure 6.7. 2A binding may clash with translational GTPases. a, Ribbon diagram of initiated 70S-mRNA-tRNA^{fMet}-2A complex. Ribosome sites are labelled A, P and E. The initiator tRNA^{fMet} (dark green), mRNA (light green), and 2A (orange, red, yellow) are shown in two orthogonal views. **b,** Comparison of 70S-2A complex to 70S pre-translocation complex with EF-G (4V7D [http://doi.org/10.2210/pdb4v7d/pdb]). 2A binding would clash (red wedges) with EF-G binding. **c,** Comparison of 70S-2A complex to 70S complex with EF-Tu (5WE6 [http://doi.org/10.2210/pdb5we6/pdb]). 2A binding would clash (red wedges) with EF-Tu binding.

6.4. Discussion

Here we show that 2A adopts an RNA-binding fold, allowing specific recognition and stabilisation of the PRF stimulatory element in the viral RNA and direct binding to host ribosomes. Given this structural framework, we can reinterpret several preceding biochemical and virological observations. Many functions of 2A can be assigned to a single positively charged surface loop ("arginine loop" residues 93–100). Despite the low pairwise sequence identity of 2A proteins amongst Cardioviruses, R95 and R97 are completely conserved. This region was originally described as a nuclear localisation sequence (NLS)[23] and subsequently, we demonstrated that these residues are essential for PRF activity in both EMCV and Theiler's murine encephalomyelitis virus (TMEV), and that their mutation to alanine prevents 2A binding to the stimulatory element in the viral RNA[12,44] (**Fig. 6.S5**). Here we reveal how R95 and R97 also mediate direct 2A binding to the small

ribosomal subunit (**Fig. 6.6d–f**) and therefore also likely confer other 2A-associated translational activities. Importantly, 2A uses the same molecular surface to bind to both the PRF stimulatory element and to ribosomes (**Fig. 6.6d–f, Fig. 6.S5**), so for any given 2A molecule these events are mutually exclusive. This suggests that the primary determinant of −1 PRF is likely to be 2A binding to the stimulatory element, with ribosome binding having a secondary effect. If 2A were to act as a "bridge" between the stimulatory element and the ribosome, this would necessitate two separate interactions surfaces, which we do not observe.

Our cryo-EM structure unexpectedly revealed four distinct 2A:rRNA interfaces (**Fig. 6.6 and Fig. 6.8**). Based on the quality of cryo-EM density and the degree of structural conservation between prokaryotic and mammalian ribosomes, the 2A1 site is likely to be the highest affinity and most physiologically relevant (**Fig. 6.7**). Nevertheless, all sites provide clues as to how RNA-binding specificity is achieved. RNA recognition is driven almost exclusively by electrostatic interactions between arginine or lysine side chains and the ribose phosphate backbone oxygen atoms. The mobility and flexibility of the arginine loop and other residues at the RNA binding surface (**Fig. 6.1f, 6.5d**) illustrates how 2A can recognise a variety of structurally degenerate targets. Whilst superposition of sites failed to reveal a common structural motif (**Fig. 6.S7**), they all include features such as kinks, distortions and junctions between multiple helices. A preference for these features is consistent with our biochemical observations that 2A is unable to bind EMCV 2, 4 and 5 RNAs, which are predicted to form stable, undistorted stem-loops (**Fig. 6.2c, d**). There is a strong likelihood that, in the 2A-bound state, the conformation of the EMCV RNA that stimulates PRF involves additional base-pairs between C-residues in the loop and a GG pair in the 5′ extension (**Fig. 6.S3**). This pseudoknot-like conformation may either pre-exist in equilibrium with other states, or it may be directly induced by 2A binding (**Fig. 6.8**). Whilst we have been unable to capture a snapshot of this molecular recognition event, it likely comprises the structural basis for the molecular "switch" that activates frameshifting during EMCV infection.

Our single-molecule data now also provide a physical explanation for this molecular "switch". It was previously shown that ribosome can exert forces of up to 20 pN during elongation[45]. We show that, in the absence of 2A, both WT and CUC RNAs unfold at forces around ~10 pN, well within the ribosome-achievable force range and hence unlikely to cause a ribosomal pause. However, in the presence of 2A, WT but not CUC RNAs are stabilised to unwind at ~27 pN, presenting a considerable blockade to ribosome progression (**Fig. 6.3d**). This also supports the idea that the failure of the CUC mutant to stimulate PRF is due to its inability to adopt the pseudoknot-like conformation of the "switch" that would normally be selectively recognised and stabilised by 2A.

Our current mechanistic understanding of PRF is largely informed by ensemble kinetic and single-molecule FRET studies of prokaryotic ribosomes[4-6,46-48]. Frameshifting occurs late during the EF-G mediated translocation step, in which the stimulatory element traps ribosomes in a rotated or hyper-rotated state, accompanied by multiple abortive EF-G binding attempts and rounds of GTP hydrolysis. Stability of the RNA stimulatory element structure downstream of the slippery sequence is thought to be an important determinant of the frameshifting efficiency[9,49,50] although the plasticity of this structure, and the ability to adopt alternate conformations, is also a key property[10]. Several recent studies emphasise the importance of the energetics of codon:anticodon base-pairing at the slippery sequence[6,51], suggesting that the primary role of the stimulatory element is to simply pause the ribosome over a permissive slippery sequence in which the tRNA-mRNA base pairing energies in the 0 and −1 frames are similar. Longer pauses at a more stable stimulatory element allow an equilibrium to be established between the 0 and −1 frames, converging on a maximum frameshift efficiency of ~50%. We have demonstrated how 2A-mediated stabilisation of the stimulatory element likely presents a potent elongation blockade allowing this equilibrium to be established (**Fig. 6.3e, 6.8**). However, this mechanism alone cannot explain the very high PRF efficiencies (up to ~70%) observed by ribosome profiling during EMCV infection[12,44].

Based on our structure, it is tempting to speculate that competition between EF-G/eEF2 and 2A1 binding might have a role in prolonging the pause, thereby contributing to the high PRF efficiencies that we observe in 2A-dependent systems[52]. Indeed, direct interactions between the ribosome and PRF stimulatory elements are not unprecedented, with a recent study describing how the HIV-1 stem-loop induces a pause by binding to the 70S A-site and preventing tRNA delivery[48]. The ribosome-bound form of 2A that we observe could therefore be a secondary 'enhancer' of PRF efficiency, acting synergistically with the main stimulatory element. It could also be relevant to the resolution of the elongation blockade: by providing an alternative 2A-binding surface that competes with the viral RNA, the ribosome may help to induce 2A dissociation from the stimulatory element during a pause at the PRF site. Alternatively, it may not be directly relevant to frameshifting *per se*, instead representing a way of interfering with host cell translation as 2A accumulates during infection.

In conclusion, this work defines the structural and molecular basis for the temporally regulated 'switch' behind the reprogramming of viral gene expression in EMCV infection (**Fig. 6.8**). 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.

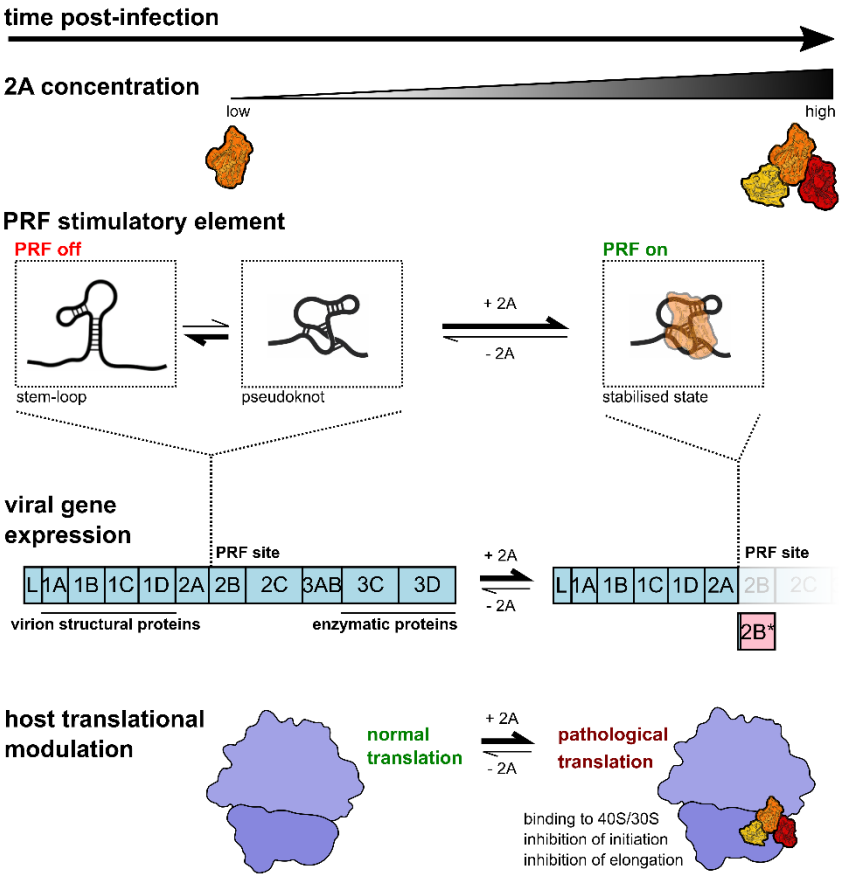

Figure 6.8. Molecular basis for 2A-induced reprogramming of gene expression. The PRF stimulatory RNA element is predicted to form either stem-loop or pseudoknot conformations. As 2A accumulates during EMCV infection, it selectively binds to and stabilises a pseudoknot-like conformation of the PRF stimulatory 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.

6.5. Methods

Materials availability

Further information and requests for resources should be directed to and will be fulfilled by Ian Brierley (ib103@cam.ac.uk). Plasmids generated in this study are available on request. DNA and RNA oligonucleotides are standard synthetic products that are commercially available (see **Table 6.S5**).

Cloning, protein expression and purification

All gene cloning, manipulation and plasmid propagation steps involving pGEX6P1 or pOPT vectors were carried out in *Escherichia coli* DH5α cells grown at 37 °C in 2 × TY or LB media supplemented with appropriate selection antibiotics. EMCV 2A cDNA was amplified by PCR from previously described plasmid 2A_pGEX6P1[12] (primers E2A_F1 and E2A_R1; **Table 6.S5**) and cloned into pOPTnH[53] using NdeI and BamHI sites, thereby introducing a C-terminal GlySerLysHis$_6$ tag. The 2A$_{9\text{-}136}$ truncated construct was cloned in an identical way (primers E2A_F2 and E2A_R2; **Table 6.S 5**). The EMCV 2A R95A/R97A mutant was cloned into pOPTnH after PCR-amplification from a previously described 2A_pGEX6P1 construct containing these mutations[12]. Other EMCV 2A mutants were prepared by PCR mutagenesis, using either the wild-type EMCV 2A_pOPT or 2A$_{9\text{-}136}$_pOPT plasmids as templates, with the following primer pairs (C111S: E2A_mut_F1 and E2A_mut_R1; R46A/K48A/K50A: E2A_mut_F2 and E2A_mut_R2; K74A: E2A_mut_F3 and E2A_mut_R3; **Table 6.S5**). To introduce an N-terminal StrepII-tag ($_{SII\text{-}}$2A), annealed oligonucleotides encoding the StrepII-tag (SII_F and SII_R, **Table 6.S5**) were inserted in-frame at the BamHI site of 2A_pGEX6P1.

Recombinant proteins 2A, 2A$_{9\text{-}136;\ C111S,}$ 2A$_{R95A/R97A}$, 2A$_{R46A/K48A/K50A}$ and 2A$_{K73A}$ were expressed in *E. coli* BL21 (DE3) pLysS cells grown in 2 × TY broth supplemented with 100 µg/mL ampicillin and 12.5 µg/mL chloramphenicol (37 °C, 200 rpm) until an OD$_{600nm}$ of 0.6 – 1.0 was reached. Expression was induced with 0.5 mM IPTG for either 4 h at 37 °C or overnight at 21 °C. For selenomethionyl derivatisation (2A$_{SeMet}$), protein was expressed in *E. coli* B834 cells, grown shaking (210 rpm, 37°C) in SeMet base media (Molecular Dimensions) supplemented with nutrient mix, 40 µg/mL L-selenomethionine and 100 µg/mL ampicillin. Expression was induced as above.

Cells were harvested by centrifugation (4,000 × g, 4°C, 20 min), washed once in ice-cold PBS and stored at -20°C. Pellets from four litres of culture were resuspended in cold lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 30 mM imidazole, supplemented with 50 µg/mL DNase I and EDTA-free protease inhibitors) and lysed by passage through a cell disruptor at 24 kPSI (Constant Systems). Lysate was cleared by centrifugation (39,000 × g, 40 min, 4°C) prior to incubation (1 h, 4°C) with 4.0 mL of Ni-NTA agarose (Qiagen) pre-equilibrated in the same buffer. Beads were washed in batch four times with 200 mL buffer (as above, but without DNase or protease inhibitors) by centrifugation (600 × g, 10 min, 4°C) and re-suspension. Washed beads were pooled to a gravity column prior to elution over 10 column volumes (CV) with 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 300 mM imidazole. Fractions containing 2A were pooled and dialysed (3K molecular weight cut-off (MWCO), 4°C, 16 h) against 1 L buffer A (50 mM Tris-HCl pH 8.0, 400 mM NaCl,

5.0 mM DTT) before heparin-affinity chromatography to remove contaminating nucleic acids. Samples were loaded on a 10 mL HiTrap Heparin column (GE Healthcare) at 2.0 mL/min, washed with two CV of buffer A and eluted with a 40% □ 100% gradient of buffer B (50 mM Tris-HCl pH 8.0, 1.0 M NaCl, 5.0 mM DTT) over 10 CV. Fractions containing 2A were pooled and concentrated using an Amicon® Ultra centrifugal filter unit (10K MWCO, 4,000 × g). Size exclusion chromatography was performed using a Superdex 75 16/600 column pre-equilibrated in 10 mM HEPES pH 7.9, 1.0 M NaCl, 5.0 mM DTT. Purity was judged by 4-20% gradient SDS-PAGE, and protein identity verified by mass spectrometry. Purified protein was used immediately or was concentrated as above (~ 7.0 mg/mL, 390 μM), snap-frozen in liquid nitrogen and stored at -80°C. Variants of 2A, including $2A_{9\text{-}136;C111S}$ and $2A_{SeMet}$ were purified identically to the wild-type protein. The StrepII-tagged variant ($_{SII}$-2A) was expressed and purified using GST-affinity as previously described[12]. Following removal of the GST tag by 3C protease, $_{SII}$-2A was further purified by Heparin affinity and size-exclusion chromatography as above.

Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS)

Per experiment, 100 μL of protein was injected onto a Superdex 75 increase 10/300 GL column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl, 1.0 M NaCl (0.4 mL/min flow, 25°C). Experiments were performed with 5.2 mg/mL 2A (corresponding to a molar concentration of 290 μM). The static light scattering, differential refractive index, and the UV absorbance at 280 nm were measured in-line by DAWN 8+ (Wyatt Technology), Optilab T-rEX (Wyatt Technology), and Agilent 1260 UV (Agilent Technologies) detectors. The corresponding molar mass from each elution peak was calculated using ASTRA 6.1 software (Wyatt Technology).

Protein crystallization

Purified EMCV 2A was concentrated to 5.9 mg/ml in 10 mM HEPES pH 7.9, 1.0 M NaCl, 2.0 mM DTT. Diffraction-quality native 2A crystals were grown at 21°C by sitting-drop vapor diffusion against an 80 μL reservoir of 0.625 M $(NH_4)_2SO_4$, 0.15 M tri-sodium citrate pH 5.7. Notably, crystal growth was only visible after 30 days. Drops were prepared by mixing 200 nL protein and 200 nL crystallization buffer. Selenomethionyl derivative 2A ($2A_{SeMet}$) was concentrated to 5.7 mg/mL in 10 mM HEPES pH 7.9, 1.0 M NaCl, 2.0 mM DTT, and diffraction-quality $2A_{SeMet}$ crystals were grown as above against an 80 μL reservoir of 0.675 M $(NH_4)_2SO_4$, 0.15 M tri-sodium citrate pH 5.7. Crystals were cryo-protected by the addition of 0.5 μL crystallization buffer supplemented with 20% v/v

glycerol, prior to harvesting in nylon loops and flash-cooling by plunging into liquid nitrogen.

X-ray data collection, structure determination, refinement and analysis

Native datasets (**Table 6.S1**) of 900 images were recorded at Diamond Light Source, beamline I03 (λ = 0.9796 Å) on a Pilatus 6M detector (Dectris), using 100% transmission, an oscillation range of 0.2° and an exposure time of 0.04 s per image. Data were collected at a temperature of 100 K. Data were processed with the XIA2[54] automated pipeline, using XDS[55] for indexing and integration, and AIMLESS[56] for scaling and merging. Crystallographic calculations were performed using the default software parameters unless otherwise stated. Processing and refinement statistics are detailed in **Table 6.S1**. Resolution cut-off was decided by a $CC_{1/2}$ value ≥ 0.5 and an I/σ(I) ≥ 1.0 in the highest resolution shell[57]. For multiple-wavelength anomalous dispersion (MAD) phasing experiments, selenomethionyl derivative datasets were ecorded at beamline I03 (peak λ = 0.9796 Å, 12656.0 eV; hrem λ = 0.9763, 12699.4 eV; inflexion λ = 0.9797 Å, 12655.0 eV). Data were processed as above using XIA2, XDS and AIMLESS. The structure was solved by three-wavelength anomalous dispersion analysis of the selenium derivative (space group $P6_222$) performed using the autoSHARP pipeline[58], implementing SHELXD[59] for substructure determination, SHARP for heavy-atom refinement and phasing, SOLOMON[60] for density modification and ARP/wARP[61] for automated model building. This was successful in placing 503/573 (87%) residues in the asymmetric unit, which comprised four copies of the protein related by non-crystallographic symmetry (NCS). This initial model was then used to solve the native dataset by molecular replacement with Phaser [62]. The model was completed manually by iterative cycles of model-building using COOT 0.9.2 [63] and refinement with phenix.refine[64] (Phenix build 1.18.1_3865), using local NCS restraints and one TLS group per chain. Upon completion of model building, ISOLDE 1.1 (Croll, 2018) was used to improve model geometry and resolve clashes prior to a final round of refinement using phenix.refine. MolProbity[65] was used throughout the process to evaluate model geometry. For the electrostatic potential calculations, partial charges were first assigned using PDB2PQR[66], implementing PROPKA to estimate protein pKa values. Electrostatic surfaces were then calculated using APBS[67]. Prior to designation of the “beta shell” as a new fold, structure-based database searches for proteins with similar folds to EMCV 2A were performed using PDBeFOLD[19], DALI[20] and CATHEDRAL[21]. Buried surface areas were calculated using PDBePISA [68].

RNA folding prediction

The simRNAweb server[26] was used for stem-loop and pseudoknot tertiary structure modelling of the EMCV stimulatory element. Experimentally-determined base-pairs were input as secondary structure restraints. Replica exchange Monte Carlo (REMC) simulated-annealing was performed with 10 replicas and 16000000 iterations per cycle. Trajectory files from eight independent simulations were concatenated and clustered, and all-atom PDB files was generated from the lowest energy state in each of the five most populous clusters. The 3D models presented (**Fig. 6.S3**) represent the top cluster for pseudoknots and the top three clusters for stem-loops.

Electrophoretic Mobility Shift Assay (EMSA)

Synthetic RNA oligonucleotides (**Table 6.S5**, IDT) were dissolved in distilled water. RNAs were labelled at the 5′ end with A647-maleimide or Cy5-maleimide conjugates (GE Healthcare) using the 5′ EndTag kit (Vector Labs) as directed by the manufacturer. For each binding experiment, a series of reactions were prepared on ice, each containing 1.0 µL 500 nM RNA, 1.0 µL serially diluted protein at concentrations of 320, 160, 80, 40, 20, 10, 5.0, and 2.5 µM in 10 mM HEPES pH 7.9, 1.0 M NaCl, 5.0 µL 2 × buffer (20 mM Tris-HCl pH 7.4, 80 mM NaCl, 4.0 mM magnesium acetate 2.0 mM DTT, 10% v/v glycerol, 0.02% w/v bromophenol blue, 200 µg/mL porcine liver tRNA, 800 U /mL SUPERase-In [Invitrogen]) and 3.0 µL distilled water. This gave final binding reactions of 10 µL with 50 nM RNA, 1 × buffer, a salt concentration of ~ 140 mM and proteins at concentrations of 32, 16, 8.0, 4.0, 2.0, 1.0, 0.5 and 0.25 µM. Samples were incubated at 37°C for 20 min prior to analysis by native 10% acrylamide/TBE PAGE (25 min, 200 V constant). Gels were scanned with a Typhoon FLA-7000 (GE) using the 635 nm laser / R670 filter. Raw, uncropped image data is available in the **Source Data** file.

Isothermal Titration Calorimetry (ITC)

ITC experiments were performed at 25°C using an automated MicroCal PEAQ-ITC platform (Malvern Panalytical). Proteins and synthetic RNA oligonucleotides (IDT) were dialysed extensively (24 h, 4°C) into buffer (50 mM Tris-HCl pH 7.4, 400 mM NaCl) prior to experiments. RNA (52 µM) was titrated into protein (5 µM) with 1 x 0.4 µL injection followed by 12 × 3.0 µL injections. Control titrations of RNA into buffer, buffer into protein and buffer into buffer were also performed. Data were analysed using the MicroCal PEAQ-ITC analysis software 1.30 (Malvern Panalytical) and fitted using a one-site binding model. Presented traces were representative of two independent titrations.

Microscale Thermophoresis (MST)

For RNA-binding experiments, synthetic EMCV RNA variants (**Table 6.S5**) 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. 6.1e**) 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.

Preparation of constructs for optical tweezer experiments

DNA encoding the frameshifting sequence of EMCV was inserted into plasmid pMZ_lambda_OT using PCR and subsequent Gibson assembly. This plasmid contains the ColE1 origin, ampicillin resistance, ribosome binding site and two 2 kbp handle regions derived from lambda phage DNA (5′ and 3′ handle). For the generation of the mutant plasmid, PCR and blunt-end ligation was used to mutate the CCC triplet in the EMCV stem-loop to CUC. Control constructs (see below) were prepared the same way as mutant constructs. For the control construct without any single-stranded RNA region, a PCR reaction using the EMCV wild-type (CCC) construct as template was conducted with 3' handle forward oligonucleotide and 5' handle reverse oligonucleotide as primers (**Table 6.S5**) After the PCR, the linear products were blunt-end ligated to yield the control constructs. Wild-type and mutant plasmids were subsequently used to generate construct suitable for optical tweezer measurements consisting of the EMCV frameshifting sequence flanked by the 2 kbp long handle regions. Three pairs of primers for PCR were designed allowing the amplification of the *in vitro* transcription template and 5′ and 3′ handles. Subsequently, PCR reactions generated 5′ and 3′ handles and a long template for *in vitro* transcription. The 3′ handle was labelled during PCR using a 5′ digoxigenin-labelled reverse primer. The 5′ handle was labelled with Biotin-16-dUTP at the 3′ end following PCR using T4 DNA polymerase. RNA was transcribed from templates for *in vitro* transcription using T7 RNA polymerase. RNA and both DNA handles (5′ and 3′) were annealed together in a mass ratio 1:1:1 (5 μg each) by incubation at 95 °C for 10 min, 62 °C for 1 hour, 52 °C for 1 hour and slow cooling to 4 °C in a buffer containing 80% formamide, 400 mM NaCl, 40 mM HEPES, pH 7.5, and 1 mM EDTA[69]. Following annealing, the samples were concentrated by ethanol precipitation, the pellets resuspended in 40 μL RNase-free water, split into 4 μL aliquots and stored at –20 °C.

Optical tweezers data collection and analysis

Optical tweezer experiments were performed using a commercial dual-trap instrument equipped with a microfluidics system (C-trap, Lumicks). Optical tweezers (OT)

constructs described above were mixed with 3 µL of polystyrene beads coated with antibodies against digoxigenin (0.1% v/v suspension, Ø 1.76 µm, Lumicks), 8 µL of measurement buffer (20 mM HEPES, pH 7.6, 300 mM KCl, 5 mM $MgCl_2$, 5 mM DTT and 0.05% Tween) and 1 µL of RNase inhibitors. The mixture was incubated for 20 min at room temperature in a final volume of 16 µL, and subsequently diluted by addition of 0.5 mL measurement buffer. Separately, 0.8 µL of streptavidin-coated polystyrene beads (1% v/v suspension, Ø 2 µm, Lumicks) was supplemented with 1 mL of measurement buffer, the flow cell was washed with the measurement buffer and suspensions of both streptavidin beads as well as the complex of OT construct with anti-digoxigenin beads were introduced into the flow cell. Per experiment, an anti-digoxigenin (AD) bead and a streptavidin (SA) bead were optically trapped and brought into close proximity to allow the formation of a tether in between. The beads were moved apart (unfolding) and back together (refolding) at 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 2A protein experiments, protein was diluted to 300 nM in measurement buffer and added to the buffer channel of the optical tweezer flow cell. FD data was recorded at a rate of 78000 Hz. To ensure that the observed effects were indeed a result of interaction with the studied RNA region and not a non-specific binding to handle regions, we also employed constructs containing either no single-stranded RNA sequence (No ssRNA control) [http://dx.doi.org/10.17632/gkpwngy65h.2]. No oxygen scavengers were used during measurements. However, to prevent oxygen damage, all buffers were degassed and contained DTT as reducing agent.

Afterwards, the data were down sampled by a factor of 30 and filtered with a Butterworth filter (0.05 filtering frequency, filter order 4) using a custom-written python algorithm. FD curves were fitted using a custom written Python script, which is based on Pylake package provided by Lumicks (https://lumicks-pylake.readthedocs.io/). Scripts have been deposited to GitHub [https://github.com/REMI-HIRI/EMCV_2A_project]. The fitting procedure was done as described[70]. In brief, first, a fully folded part (until the first detectable unfolding step) was fitted with a worm-like chain model (WLC)[71,72] to determine the persistence length (dsL_P) of the tether while the contour length (dsL_C) parameter was held fixed at 1256 nm (± 1%; 4110 bp*0.305 nm/bp and 4 ss*0.59 nm/ss)[73]. The (partially) unfolded parts of FD curve were then fitted by a model comprising of WLC (describing the folded double stranded handles) and a second WLC model (describing the unfolded single stranded parts)[71,74]. For fitting of the unfolded regions, parameters extracted from the fully folded part fitting (dsL_P, dsL_C, dsK) were used and fixed in the WLC part of the combined model. Persistence length of the single stranded part (ssL_P) was fixed at 1 nm while contour length (ssL_C) of the single stranded part together with the single stranded stretch

modulus (ssK) were optimized. The work performed on the structure while unfolding or refolding was calculated as difference between area under curve (AUC) of the fit for the folded region and AUC of the fit for the unfolded region, counted from the beginning of the FD curve till the unfolding step coordinates[75]. If the unfolding and refolding work distributions were overlapping, Crook's fluctuation theorem was applied to estimate the equilibrium work, which represents free Gibbs energy[75], as intersection between the unfolding and refolding work distributions. Since Crooks fluctuation theorem directly averages work values of unfolding and folding, it is not reliable when system is far from equilibrium or in other cases when folding and unfolding work distributions are very different[33]. In our WT+2A samples, the overlap between folding and unfolding work was not sufficient. Therefore, to more accurately estimate the free energies where large fluctuations exist in work distributions, we applied the Jarzynski's equality as described[34]. We then corrected for the bias in the Jarzynski estimate[76]. Theoretical values of the Gibbs free energies for the predicted RNA structures were obtained using mfold[31]. The FD curves were plotted using Prism 8.0.2 (GraphPad). The RNAstructure software (version 6.2) was also used for prediction of the EMCV RNA element secondary structure[77].

Eukaryotic ribosomal subunit purification

40S and 60S subunits were purified from untreated rabbit reticulocyte lysate (Green Hectares) as previously described[78]. 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 required.

Ribosome binding assays

Assays were conducted in 50 mM Tris-acetate pH 7.5, 150 mM potassium acetate, 5.0 mM magnesium acetate, 0.25 mM spermidine, 10 mM DTT, 0.1 % v/v Triton X-100. Per 60 µL binding reaction, ribosome subunits were diluted to a final concentration of 0.4 µM, and 2A protein was added in excess to a final concentration of 2.4 µM. Twenty microlitres of this mixture was retained for SDS-PAGE analysis of the 'input'. The

remaining 40 µL was incubated at room temperature for 20 min prior to application to a S200-HR size-exclusion microspin column (Cytiva) that had been pre-equilibrated (4 x 500 µL) in the above buffer by resuspension and centrifugation (300 × g, 30 s). Immediately after application, the eluate was collected by centrifugation (300 × g, 60 s).

Western blot

Samples were analysed by 4–20% gradient SDS-PAGE and transferred to a 0.2 µm nitrocellulose membrane. All subsequent steps were carried out at room temperature. Membranes were blocked (5% w/v milk, PBS, 1 h) before incubation (1 h) with primary antibodies in 5% w/v milk, PBS, 0.1% v/v Tween-20. Membranes were washed three times with PBS, 0.1% v/v Tween-20 prior to incubation (1 h) with IRDye fluorescent antibodies in 5% w/v milk, PBS, 0.1% v/v Tween-20. After three washes in PBS, 0.1% v/v Tween-20 and a final rinse in PBS, membranes were imaged using an Odyssey CLx Imaging System (LI-COR). Figures were prepared using ImageStudio Lite 5.2 (LI-COR). Antibodies used were rabbit polyclonal anti-2A[12] (1/1000); mouse monoclonal anti-RPS6 (1/1000, clone A16009C, BioLegend); mouse monoclonal anti-RPL4 (1/1000, clone 4A3, Sigma); goat anti-rabbit IRDye 800 CW (1/10,000, LI-COR) and goat anti-mouse IRDye 680LT (1/10,000, LI-COR). Raw, uncropped blots are available in the **Source Data** file.

***In vitro* transcription**

For *in vitro* frameshifting assays, we cloned a 105 nt DNA fragment (pdluc/EMCV, **Table 6.S5**) containing the EMCV slippery sequence flanked by 12 nt upstream and 86 nt downstream into the dual luciferase plasmid pDluc at the XhoI and BglII sites[79]. This sequence was inserted between the Renilla and firefly luciferase genes such that firefly luciferase expression is dependent on −1 PRF. Wild-type or mutated frameshift reporter plasmids were linearized with FspI and capped run-off transcripts generated using T7 RNA polymerase as described[80]. Messenger RNAs were recovered by phenol/chloroform extraction (1:1 v/v), desalted by centrifugation through a NucAway Spin Column (Ambion) and concentrated by ethanol precipitation. The mRNA was resuspended in water, checked for integrity by agarose gel electrophoresis, and quantified by spectrophotometry.

Messenger RNAs for 70S IC preparation (EMCV_IC, **Table 6.S5**) were produced from a 117 nt long DNA fragment containing the EMCV frameshift site flanked by the bacterial 5′ UTR with Shine-Dalgarno sequence and 18 nt downstream region of the putative structure.

5′GGGAAUUCAAAAAUUGUUAAGAAUUAAGGAGAUAUACAUAUGGA**GGUUUUU**AUCACUCAAGGAGCGGCAGUGUCAUCAAUGGCUCAAACCCUACUGCCGAACGACUUGGCCAGATCT 3′ (slippery sequence in bold, initiation codon underlined)

This sequence was PCR amplified and *in vitro* transcribed using T7 RNA polymerase (produced in-house). Messenger RNAs were purified using the Qiagen RNeasy midiprep kit according to the manufacturer's protocols. The mRNAs were eluted in RNAse-free water, integrity and purity was checked by gel electrophoresis and quantified by spectrophotometry.

70S initiation complex preparation

Ribosomes, translation factors, and tRNAs were of *E. coli* origin. Total *E. coli* tRNA was from Roche, and oligonucleotides were from Microsynth. 70*S* ribosomes from MRE600, EF-Tu, EF-G, IF1, IF2 and IF3 were purified from *E. coli* [81]. fMet-tRNA^{fMet} was prepared and aminoacylated according to published protocols[82,83]. Aminoacylated fMet-tRNA^{fMet} was purified by reversed-phase HPLC on a Wide Pore C5 (10 µM particle size 10 mm x 25 cm) column (Sigma Aldrich). To prepare initiation complexes, 70S ribosomes (1 µM) were incubated with a three-fold excess of an EMCV model mRNA (EMCV_IC, **Table 6.S5**) encoding for 5′…AUGGA**GGUUUUU**AUC…3′ (slippery sequence in bold) and a 1.5-fold excess each of IF1, IF2, IF3, and fMet-tRNA^{fMet} in buffer A (50 mM Tris-HCl pH 7.5, 70 mM NH_4Cl, 30 mM KCl, 7 mM $MgCl_2$) supplemented with GTP (1 mM) for 30 min at 37°C. 70S initiation complexes were purified by centrifugation through a 1.1 M sucrose cushion in buffer A. Before grid preparation, initiation complexes were additionally purified on Sephacryl S-300 gel filtration microspin columns.

Frameshifting assays (*In vitro* translation)

Messenger RNAs were translated in nuclease-treated rabbit reticulocyte lysate (RRL) or wheat germ (WG) extracts (Promega). Typical reactions were composed of 90% v/v RRL, 20 µM amino acids (lacking methionine) and 0.2 MBq [^{35}S]-methionine and programmed with ~50 µg/mL template mRNA. Reactions were incubated for 1 h at 30°C. Samples were mixed with 10 volumes of 2× Laemmli's sample buffer, boiled for 3 min and resolved by SDS-PAGE. Dried gels were exposed to a Storage Phosphor Screen (PerkinElmer) and the screen scanned in a Typhoon FLA7000 using phosphor autoradiography mode. Bands were quantified using ImageQuant™TL 8.1.0 software (GE Healthcare). The calculations of frameshifting efficiency (%FS) took into account the differential methionine content of the various products and %FS was calculated as % −1FS = 100 × (IFS/MetFS) / (IS/MetS + IFS/MetFS). In the formula, the number of methionines

in the stop and frameshift products are denoted by MetS, MetFS respectively; while the densitometry values for the same products are denoted by IS and IFS respectively. All frameshift assays were carried out a minimum of three times.

Ribosomal frameshift assays in *E. coli* employed a coupled T7/S30 *in vitro* translation system (Promega). A ~450 bp fragment containing the EMCV PRF signal (or mutant derivative) was prepared by PCR from plasmid pDluc/EMCV[12] and cloned into the BamHI site of the T7-based, *E. coli* expression vector pET3xc[84]. T7/S30 reaction mixes were prepared according to the manufacturer's instructions (50 µL volumes), including 10 µCi ^{35}S methionine, supplemented with plasmid DNA (4 µg) and incubated at 37 ˚C for 90 mins. Reactions were precipitated by addition of an equal volume of acetone, dissolved in Laemmli's sample buffer and aliquots analysed by SDS-PAGE. PRF efficiencies were calculated as above.

Cryo-EM specimen preparation

Initiated 70S ribosomes in 50 mM Tris-HCl pH 7.5, 70 mM NH_4Cl, 30 mM KCl, 7 mM $MgCl_2$ were diluted tenfold into 20 mM HEPES pH 7.5, 100 mM potassium acetate, 1.5 mM $MgCl_2$, 2.0 mM DTT. 2A protein was dialysed (3K MWCO, 4°C, 16 h) into the same buffer. Crosslinking reactions of 50 µL comprising 75 nM ribosomes, 3.0 µM 2A and 2.0 mM bis(sulfosuccinimidyl)suberate (BS3) were performed on ice (30 min) immediately prior to grid preparation. Quantifoil R 2/2 400-mesh copper supports were coated with an additional ~ 60 Å layer of amorphous, evaporated carbon by flotation[85], and thoroughly dried before use. Grids were made hydrophilic by glow-discharge in air for 30 s. Three microliters of crosslinking reaction was applied to grids which were then blotted for 4.5 s and vitrified by plunging into liquid ethane using a Vitrobot MK IV (FEI) at 4°C, 100% relative humidity.

Cryo-EM data collection and processing

Micrographs were collected at the BiocEM facility (Department of Biochemistry, University of Cambridge) on a Titan Krios microscope (FEI) operating at 300 kV and equipped with a Falcon III detector (**Table 6.S4**). At 75,000 × magnification, the calibrated pixel size was 1.07 Å / pixel. Per 0.6 s acquisition in integration mode, a total exposure of 54.4 e^- / Å^2 was fractionated over 23 frames with applied defocus of –1.5, –1.8, –2.1, –2.4, –2.7 and –3.0 µm. EPU software was used for automated acquisition with five images per hole. After manual inspection, 5730 micrographs were used in subsequent image processing.

Movie frames were aligned and a dose-weighted average calculated with MotionCor 2[86]. The contrast transfer function (CTF) was estimated using CtfFind 4[87]. All subsequent image-processing steps were carried out in RELION 3.1[88] (**Fig. 6.S6**) and all reported estimates of resolution are based on the gold standard Fourier shell correlation (FSC) at 0.143, and the calculated FSC is derived from comparisons between reconstructions from two independently refined half-sets. Reference-free autopicking of 820,475 particles was performed using the Laplacian-of-Gaussian function (200 - 250 Å diameter). Particles were initially downscaled threefold and extracted in a 150-pixel box. Two rounds of 2D classification (into 100 and 200 classes, respectively) were used to clean the dataset to 750,029 'good' particles. An initial reference was generated from a PDB file of a 70S elongation-competent ribosome (PDB ID 5MDZ [http://doi.org/10.2210/pdb5mdz/pdb]) and low-pass filtered to 80 Å resolution. The initial 3D refinement (6.5 Å resolution) showed clear evidence for at least one copy of 2A adjacent to the factor binding site on the 30S subunit. At this stage, two rounds of focussed classification with signal subtraction were performed (6 classes) to separate particles based on additional density near i) the factor binding site and ii) the mRNA entry channel/helicase. The former was successful and 289,741 particles containing three copies of 2A were rescaled to full size and extracted in a 450-pixel box. Following initial 3D refinement, creation of a 15 Å low-pass filtered mask (five-pixel extension and five-pixel soft edge) and post-processing, a reconstruction of 2.93 Å was achieved. After per-particle CTF refinement and polishing, this was increased to 2.50 Å. With the increased angular accuracy provided by the fully rescaled data, focussed classification with signal subtraction and local angular searches was performed again to separate particles based on 2A occupancy at the factor binding site. This final reconstruction (2.66 Å) from 120,749 particles revealed three copies of 2A bound with full occupancy, and clearer details in the vicinity of the 2A binding sites. Calculation of a local resolution map revealed additional low-resolution density adjacent to the beak of the 30S head. Subsequent focussed classification with signal subtraction and refinement confirmed that this was a fourth copy of 2A bound, present in 73,059 particles.

To build the model, the atomic coordinates for a 70S initiation complex (5MDZ [http://doi.org/10.2210/pdb5mdz/pdb]) and three copies of chain A from the 2A crystal structure (above) were docked as rigid bodies into the EM map. Local rebuilding was performed iteratively in COOT[63] and the models refined using phenix real-space refine[64] implementing reference model restraints to preserve geometry.

Visualisation of structural data

All structural figures depicting crystallographic data (cartoon, stick and surface representations) were rendered in PyMOL 2.3.4 (Schrödinger LLC). Structural figures of EM maps with docked components were rendered in ChimeraX 1.1[89].

Data Availability

The atomic coordinates and structure factors for the EMCV 2A X-ray crystal structure have been deposited in the wwPDB database under accession code 7BNY [http://doi.org/10.2210/pdb7bny/pdb]. The 70S IC:2A cryo-EM map has been deposited in the EMDB under accession code EMD-12635 [https://www.ebi.ac.uk/emdb/EMD-12257] and the refined atomic coordinates accompanying this structure have been deposited to the wwPDB under accession code 7NWT [http://doi.org/10.2210/pdb7nwt/pdb]. Previously published structures that were used in this study are also available in the wwPDB: 5WE6 [http://doi.org/10.2210/pdb5we6/pdb], 4V7D [http://doi.org/10.2210/pdb4v7d/pdb] and 5MDZ [http://doi.org/10.2210/pdb5mdz/pdb]. Source data are provided with this paper. All raw data (e.g. uncropped, unannotated gels, western blots, tables of force measurements, MST traces) corresponding to individual figure panels are provided in the Source Data File and have also been deposited in Mendeley Data [http://dx.doi.org/10.17632/gkpwngy65h.2].

Code Availability

The force spectroscopy analysis scripts supporting the current study have been uploaded to GitHub [https://github.com/REMI-HIRI/EMCV_2A_project]. Further information is available on request from Neva Caliskan (neva.caliskan@helmholtz-hiri.de).

6.6. Supplementary material

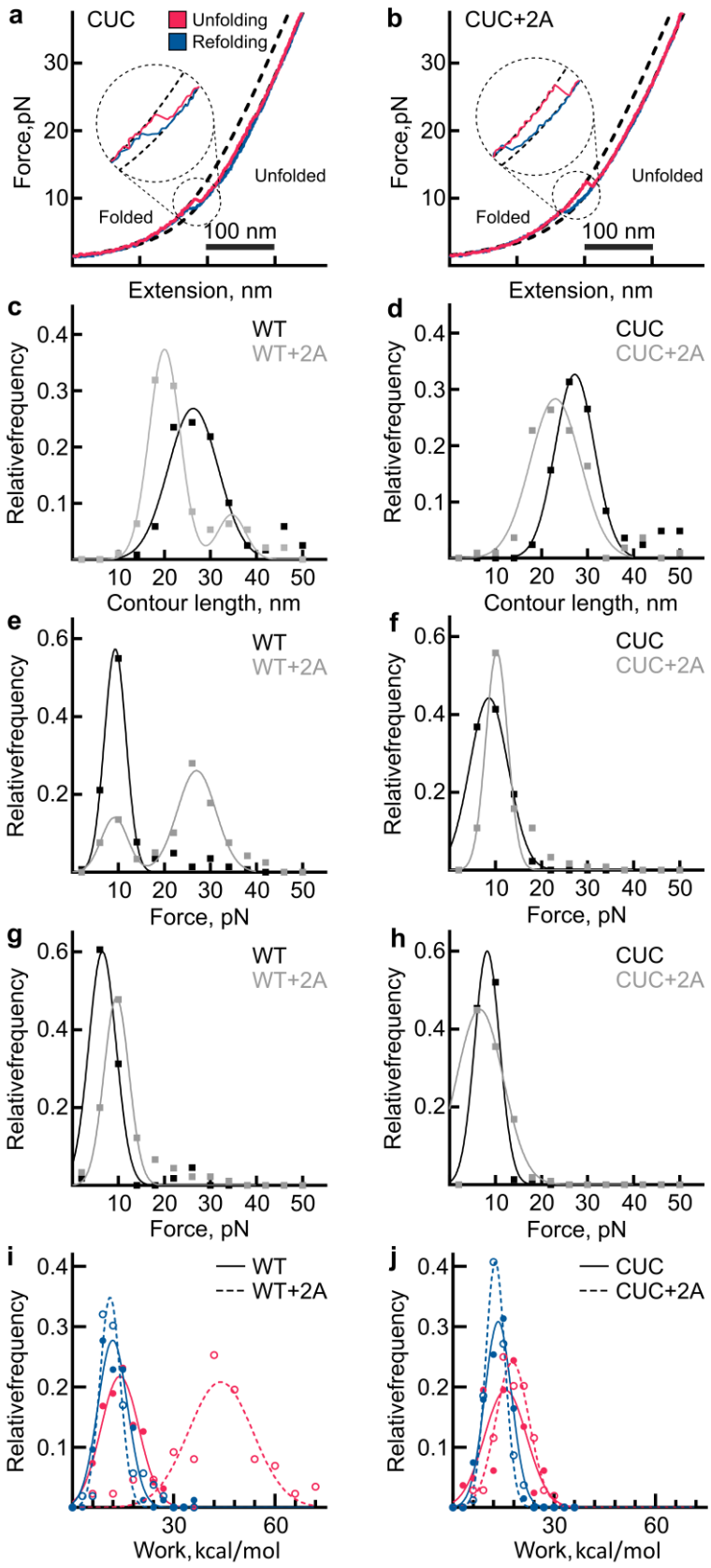

Figure 6.S4 (previous page) – related to Figure 6.3. Details of contour length, force and work distributions observed in optical tweezer experiments. a, Representative force-distance curves of the unfolding (pink) and refolding (blue) transitions of the mutant CUC RNA element (n=85). **b**, Representative force-distance curves of the unfolding (pink) and refolding (blue) transitions of the mutant CUC RNA element in the presence of 300 nM 2A protein (n=109). **c** and **d**, distribution of the contour length changes in the absence (black) and presence (grey) of 2A for WT and CUC RNAs, respectively. Dots represent experimental data points while the line corresponds to the Gaussian fit. **e** and **f**, distribution of the unfolding forces in the absence (black) and presence (grey) of 2A for WT and CUC RNAs, respectively. **g** and **h**, distribution of the refolding forces in the absence (black) and presence (grey) of 2A for WT and CUC RNAs, respectively. **i** and **j**, distribution of unfolding (pink) and refolding (blue) work in the absence (solid) and presence (dashed) of 2A for WT and CUC, respectively.

Table 6.S3. Summary of experimental OT results and predicted mfold data for EMCV WT and CUC RNAs in the absence or presence of 2A. Results were determined by gaussian fitting of the distribution histograms of obtained data (number of tethers used in the experiments was >15, and number of force trajectories range between 85-120). Related to Figure 3 and Supplementary Figure 4. Uncertainties represent standard deviations.

	WT	*WT + 2A*	*CUC*	*CUC + 2A*
Contour length change, nm	26.3±5.4	20.1±3.4	27.2±4.3	23.0±5.5
Unfolding force, pN	9.3±2.3	9.2±2.9 27.0±4.2	8.6±4.2	10.3±2.4
Refolding force, pN	6.5±3.0	9.6±2.7	8.2±2.7	6.6±5.1
Gibbs free energy (mfold), kcal/mol	-14.0±0.7 (SL), -16.2±0.8 (extended SL)		-14.0±0.7 (SL), -16.2±0.8 (extended SL)	
Gibbs free energy (experiment), kcal/mol	-13.6±4.6* -11.6±0.9**	- -26.5±8.7**	-14.5±4.7* -11.0±1.6**	-15.5±5.0* -12.9±1.2**

*Calculated by applying Crook's fluctuation theorem to unfolding/refolding work distributions. Uncertainty represents standard deviation.

**Calculated by applying Jarzynski's equality. Uncertainty represents root of mean square error.

Of note, free energy values of the CUC and WT (without 2A) RNAs were in good agreement regardless of the calculation method employed.

6.7. References

1. Firth, A.E. & Brierley, I. Non-canonical translation in RNA viruses. *J Gen Virol* **93**, 1385-1409 (2012).
2. Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E. & Baranov, P.V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res* **44**, 7007-78 (2016).
3. Korniy, N., Samatova, E., Anokhina, M.M., Peske, F. & Rodnina, M.V. Mechanisms and biomedical implications of -1 programmed ribosome frameshifting on viral and bacterial mRNAs. *FEBS Lett* **593**, 1468-1482 (2019).
4. Chen, J. et al. Dynamic pathways of -1 translational frameshifting. *Nature* **512**, 328-32 (2014).
5. Caliskan, N., Katunin, V.I., Belardinelli, R., Peske, F. & Rodnina, M.V. Programmed -1 frameshifting by kinetic partitioning during impeded translocation. *Cell* **157**, 1619-31 (2014).
6. Choi, J., O'Loughlin, S., Atkins, J.F. & Puglisi, J.D. The energy landscape of -1 ribosomal frameshifting. *Sci Adv* **6**, eaax6969 (2020).
7. Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244-7 (2006).
8. Giedroc, D.P. & Cornish, P.V. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res* **139**, 193-208 (2009).
9. Chen, G., Chang, K.Y., Chou, M.Y., Bustamante, C. & Tinoco, I., Jr. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc Natl Acad Sci U S A* **106**, 12706-11 (2009).
10. Halma, M.T.J., Ritchie, D.B., Cappellano, T.R., Neupane, K. & Woodside, M.T. Complex dynamics under tension in a high-efficiency frameshift stimulatory structure. *Proc Natl Acad Sci U S A* **116**, 19500-19505 (2019).
11. Loughran, G., Firth, A.E. & Atkins, J.F. Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome. *Proc Natl Acad Sci U S A* **108**, E1111-9 (2011).
12. Napthine, S. et al. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat Commun* **8**, 15582 (2017).
13. Jackson, R.J. A detailed kinetic analysis of the in vitro synthesis and processing of encephalomyocarditis virus products. *Virology* **149**, 114-27 (1986).
14. Hahn, H. & Palmenberg, A.C. Deletion mapping of the encephalomyocarditis virus primary cleavage site. *J Virol* **75**, 7215-8 (2001).
15. Yang, X. et al. Structures and Corresponding Functions of Five Types of Picornaviral 2A Proteins. *Front Microbiol* **8**, 1373 (2017).
16. Groppo, R. & Palmenberg, A.C. Cardiovirus 2A protein associates with 40S but not 80S ribosome subunits during infection. *J Virol* **81**, 13067-74 (2007).
17. Carocci, M. et al. Encephalomyocarditis virus 2A protein is required for viral pathogenesis and inhibition of apoptosis. *J Virol* **85**, 10741-54 (2011).
18. Merrick, W.C. eIF4F: a retrospective. *J Biol Chem* **290**, 24091-9 (2015).
19. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-68 (2004).
20. Holm, L. & Laakso, L.M. Dali server update. *Nucleic Acids Res* **44**, W351-5 (2016).
21. Redfern, O.C., Harrison, A., Dallman, T., Pearl, F.M. & Orengo, C.A. CATHEDRAL: a fast and effective algorithm to predict folds and domain boundaries from multidomain protein structures. *PLoS Comput Biol* **3**, e232 (2007).
22. Svitkin, Y.V., Hahn, H., Gingras, A.C., Palmenberg, A.C. & Sonenberg, N. Rapamycin and wortmannin enhance replication of a defective encephalomyocarditis virus. *J Virol* **72**, 5811-9 (1998).

23. Groppo, R., Brown, B.A. & Palmenberg, A.C. Mutational analysis of the EMCV 2A protein identifies a nuclear localization signal and an eIF4E binding site. *Virology* **410**, 257-67 (2011).
24. Petty, R.V., Basta, H.A., Bacot-Davis, V.R., Brown, B.A. & Palmenberg, A.C. Binding interactions between the encephalomyocarditis virus leader and protein 2A. *J Virol* **88**, 13503-9 (2014).
25. Siddiqui, N. et al. Structural insights into the allosteric effects of 4EBP1 on the eukaryotic translation initiation factor eIF4E. *J Mol Biol* **415**, 781-92 (2012).
26. Magnus, M., Boniecki, M.J., Dawson, W. & Bujnicki, J.M. SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. *Nucleic Acids Res* **44**, W315-9 (2016).
27. Ren, J., Rastegari, B., Condon, A. & Hoos, H.H. HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA* **11**, 1494-504 (2005).
28. Chen, G., Wen, J.D. & Tinoco, I., Jr. Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *RNA* **13**, 2175-88 (2007).
29. Li, P.T., Collin, D., Smith, S.B., Bustamante, C. & Tinoco, I., Jr. Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. *Biophys J* **90**, 250-60 (2006).
30. Heller, I., Hoekstra, T.P., King, G.A., Peterman, E.J. & Wuite, G.J. Optical tweezers analysis of DNA-protein complexes. *Chem Rev* **114**, 3087-119 (2014).
31. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-15 (2003).
32. Ritchie, D.B., Foster, D.A. & Woodside, M.T. Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc Natl Acad Sci U S A* **109**, 16167-72 (2012).
33. Mangeol, P. et al. Probing ribosomal protein-RNA interactions with an external force. *Proc Natl Acad Sci U S A* **108**, 18272-6 (2011).
34. Liphardt, J., Dumont, S., Smith, S.B., Tinoco, I., Jr. & Bustamante, C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* **296**, 1832-5 (2002).
35. Leger, M., Sidani, S. & Brakier-Gingras, L. A reassessment of the response of the bacterial ribosome to the frameshift stimulatory signal of the human immunodeficiency virus type 1. *RNA* **10**, 1225-35 (2004).
36. Horsfield, J.A., Wilson, D.N., Mannering, S.A., Adamski, F.M. & Tate, W.P. Prokaryotic ribosomes recode the HIV-1 gag-pol-1 frameshift sequence by an E/P site post-translocation simultaneous slippage mechanism. *Nucleic Acids Res* **23**, 1487-94 (1995).
37. James, N.R., Brown, A., Gordiyenko, Y. & Ramakrishnan, V. Translational termination without a stop codon. *Science* **354**, 1437-1440 (2016).
38. Brilot, A.F., Korostelev, A.A., Ermolenko, D.N. & Grigorieff, N. Structure of the ribosome with elongation factor G trapped in the pretranslocation state. *Proc Natl Acad Sci U S A* **110**, 20994-9 (2013).
39. Lin, J., Gagnon, M.G., Bulkley, D. & Steitz, T.A. Conformational changes of elongation factor G on the ribosome during tRNA translocation. *Cell* **160**, 219-27 (2015).
40. Bulkley, D. et al. The antibiotics dityromycin and GE82832 bind protein S12 and block EF-G-catalyzed translocation. *Cell Rep* **6**, 357-65 (2014).
41. Fislage, M. et al. Cryo-EM shows stages of initial codon selection on the ribosome by aa-tRNA in ternary complex with GTP and the GTPase-deficient EF-TuH84A. *Nucleic Acids Res* **46**, 5861-5874 (2018).
42. Loveland, A.B., Demo, G. & Korostelev, A.A. Cryo-EM of elongating ribosome with EF-Tu*GTP elucidates tRNA proofreading. *Nature* (2020).
43. Hussain, T., Llacer, J.L., Wimberly, B.T., Kieft, J.S. & Ramakrishnan, V. Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation. *Cell* **167**, 133-144 e13 (2016).

44. Napthine, S., Bell, S., Hill, C.H., Brierley, I. & Firth, A.E. Characterization of the stimulators of protein-directed ribosomal frameshifting in Theiler's murine encephalomyelitis virus. *Nucleic Acids Res* **47**, 8207-8223 (2019).
45. Wen, J.D. et al. Following translation by single ribosomes one codon at a time. *Nature* **452**, 598-603 (2008).
46. 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* **111**, 5538-43 (2014).
47. Qin, P., Yu, D., Zuo, X. & Cornish, P.V. Structured mRNA induces the ribosome into a hyper-rotated state. *EMBO Rep* **15**, 185-90 (2014).
48. Bao, C. et al. mRNA stem-loops can pause the ribosome by hindering A-site tRNA binding. *Elife* **9**(2020).
49. Yang, L. et al. Single-Molecule Mechanical Folding and Unfolding of RNA Hairpins: Effects of Single A-U to A.C Pair Substitutions and Single Proton Binding and Implications for mRNA Structure-Induced -1 Ribosomal Frameshifting. *J Am Chem Soc* **140**, 8172-8184 (2018).
50. Zhong, Z. et al. Mechanical unfolding kinetics of the SRV-1 gag-pro mRNA pseudoknot: possible implications for -1 ribosomal frameshifting stimulation. *Sci Rep* **6**, 39549 (2016).
51. Bock, L.V. et al. Thermodynamic control of -1 programmed ribosomal frameshifting. *Nat Commun* **10**, 4598 (2019).
52. Peng, B.Z. et al. Active role of elongation factor G in maintaining the mRNA reading frame during translation. *Sci Adv* **5**, eaax8030 (2019).
53. Neidel, S. et al. Vaccinia virus protein A49 is an unexpected member of the B-cell Lymphoma (Bcl)-2 protein family. *J Biol Chem* **290**, 5991-6002 (2015).
54. Winter, G. *xia2*: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43**, 186-190 (2009).
55. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
56. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-14 (2013).
57. Karplus, P.A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-3 (2012).
58. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol Biol* **364**, 215-30 (2007).
59. Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr A* **64**, 112-22 (2008).
60. Abrahams, J.P. & Leslie, A.G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52**, 30-42 (1996).
61. Perrakis, A., Harkiolaki, M., Wilson, K.S. & Lamzin, V.S. ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr* **57**, 1445-50 (2001).
62. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
63. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
64. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
65. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
66. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. & Baker, N.A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, W665-7 (2004).
67. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-41 (2001).
68. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-97 (2007).

69. Stephenson, W., Wan, G., Tenenbaum, S.A. & Li, P.T. Nanomanipulation of single RNA molecules by optical tweezers. *J Vis Exp* (2014).
70. Mukhortava, A. et al. Structural heterogeneity of attC integron recombination sites revealed by optical tweezers. *Nucleic Acids Res* **47**, 1861-1870 (2019).
71. Wang, M.D., Yin, H., Landick, R., Gelles, J. & Block, S.M. Stretching DNA with optical tweezers. *Biophys J* **72**, 1335-46 (1997).
72. Odijk, T. Stiff Chains and Filaments under Tension. *Macromolecules* **28**, 7016-7018 (1995).
73. Zhang, C. et al. The Mechanical Properties of RNA-DNA Hybrid Duplex Stretched by Magnetic Tweezers. *Biophys J* **116**, 196-204 (2019).
74. Smith, S.B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795-9 (1996).
75. McCauley, M.J., Rouzina, I., Li, J., Nunez, M.E. & Williams, M.C. Significant Differences in RNA Structure Destabilization by HIV-1 GagDp6 and NCp7 Proteins. *Viruses* **12**(2020).
76. Gore, J., Ritort, F. & Bustamante, C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc Natl Acad Sci U S A* **100**, 12564-9 (2003).
77. Bellaousov, S., Reuter, J.S., Seetin, M.G. & Mathews, D.H. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res* **41**, W471-4 (2013).
78. Pisarev, A.V., Unbehaun, A., Hellen, C.U. & Pestova, T.V. Assembly and analysis of eukaryotic translation initiation complexes. *Methods Enzymol* **430**, 147-77 (2007).
79. Fixsen, S.M. & Howard, M.T. Processive selenocysteine incorporation during synthesis of eukaryotic selenoproteins. *J Mol Biol* **399**, 385-96 (2010).
80. Powell, M.L., Brown, T.D. & Brierley, I. Translational termination-re-initiation in viral systems. *Biochem Soc Trans* **36**, 717-22 (2008).
81. Milon, P. et al. Transient kinetics, fluorescence, and FRET in studies of initiation of translation in bacteria. *Methods Enzymol* **430**, 1-30 (2007).
82. Rodnina, M.V., Semenkov, Y.P. & Wintermeyer, W. Purification of fMet-tRNA(fMet) by fast protein liquid chromatography. *Anal Biochem* **219**, 380-1 (1994).
83. Kothe, U., Paleskava, A., Konevega, A.L. & Rodnina, M.V. Single-step purification of specific tRNAs by hydrophobic tagging. *Anal Biochem* **356**, 148-50 (2006).
84. Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**, 60-89 (1990).
85. Passmore, L.A. & Russo, C.J. Specimen Preparation for High-Resolution Cryo-EM. *Methods Enzymol* **579**, 51-86 (2016).
86. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
87. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-21 (2015).
88. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).
89. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).

6.8. Acknowledgements

We thank Dima Chirgadze, Steve Hardwick and Lee Cooper (BiocEM facility) for assistance with CryoEM data acquisition. We thank Ann Mukhortava and Bärbel Lorenz (Lumicks AG) for technical assistance with optical tweezer experiments and Ann Mukhortava for critical reading of the manuscript. We thank Prof. Marina V. Rodnina for providing expression constructs. We thank Matthias Zimmer, Trevor Sweeney, Janet Deane and Tatyana Koch for experimental assistance. We thank Vish Chandrasekaran, Jailson Brito Querido, Sebastian Kraatz and Chris Rae for helpful discussions. A Titan V graphics card used for this research was donated by the NVIDIA Corporation. Remote synchrotron access was supported in part by the EU FP7 infrastructure grant BIOSTRUCT-X (Contract No. 283570). We thank the staff of Diamond Light Source beamline I03 for assistance with crystal screening and data collection. Part of this work was carried out in the laboratory of V. Ramakrishnan, who was funded by the UK Medical Research Council (MC_U105184332), and a Wellcome Trust Senior Investigator award (WT096570). CHH and SN were supported by a Wellcome Trust Investigator Award (202797/Z/16/Z) to IB. CHH is funded by a Sir Henry Dale fellowship (221818/Z/20/Z) from the Wellcome Trust and the Royal Society. AEF is supported by Wellcome Trust (106207/Z/14/Z) and European Research Council (646891) grants to AEF. SCG is funded by a Sir Henry Dale fellowship (098406/Z/12/B) from the Wellcome Trust and the Royal Society. NC, LP and AK are supported by the Helmholtz Association. NC is funded by the European Research Council StG (948636).

6.9. Author Contributions

C.H.H. and S.N. cloned expressed and purified proteins and performed most of the biochemical experiments. L.P. designed OT constructs and performed single-molecule optical tweezers experiments and analyses under the supervision of N.C. C.H.H. and S.C.G performed crystallography experiments. A.K. and N.C. performed MST experiments and analyses. N.C. prepared and purified bacterial initiation complexes for structural analysis. C.H.H. prepared cryo-EM grids and collected and processed cryo-EM data. C.H.H., S.N., A.F., N.C. and I.B. wrote the manuscript with contributions from all authors.

6.10. Competing Interests

The authors declare no competing interests.

Chapter 7

"The best way to have a good idea is to have a lot of ideas"

Linus Pauling

7. Summary and Discussion

Recoding events during translation represent an attractive strategy to fine-tune gene expression. The efficiency of frameshifting is regulated in *cis* by RNA signals embedded in the RNA and in *trans* by host or viral encoded factors.

In this thesis, I summarized our efforts to unveil the underlying regulatory mechanisms of −1PRF using single-molecule methods, specifically optical tweezers, to understand the role of RNA structure in this orchestrated process. Using advanced techniques by itself is not error-proof, therefore in our projects, we always aimed to combine orthogonal methods, functional and structural methods, single-molecule as well as ensemble approaches to acquire knowledge close to the actual truth. Nevertheless, one must be aware of the biases, artifacts, or weaknesses of the methods applied. Therefore, in the following paragraphs, I am going to discuss the critical aspects of our experiments and the findings resulting from them.

7.1. Single-molecule methods offer an unprecedented resolution.

The rise of single-molecule methods by the end of 20th and beginning of the 21st century brought great advances in our understanding of mechanisms underlying the various biological processes (1-7). These techniques offer the detection of changes as small as a few nanometers in real time with millisecond resolution. However, with great power comes great responsibility. These experiments are prone to experimental artifacts and depend strongly on proper experimental design, conduct, and correct data analysis (partially discussed in **chapters 2** and **3**) (8-11). The position of the fluorescent labels can affect the signal intensity as well as the correct folding of the studied molecule (10, 12-14). Similarly, even the choice of the fluorophore can affect the measured values (e.g. apparent K_D values) (15). In the case of optical tweezers, the length of handle regions can affect the measured values and the overall resolution (9, 16). The application of lasers, either for fluorescence or optical trapping purposes, brings the risk of introducing reactive oxygen species into the system, photobleaching of the fluorophore as well as photodamage of the studied biomolecules. Optimizing the laser intensity and the selection of the wavelength are crucial to minimize the photodamage. For the latter, infra-red lasers with wavelengths around 1064 nm are employed, as this wavelength interferes poorly with the biological

(water-based) samples. Finally, oxygen scavenger systems are often used to minimize the photodamage by ROS. While the most commonly employed system consists of the glucose oxidase-catalase pair, it has to be used carefully as it may lead to changes in pH overtime. To overcome this, more advanced oxygen scavenger systems can be employed (17, 18).

In **chapter 3**, we developed a pipeline for automation of optical tweezers data analysis. The data analysis of force-spectroscopy data is laborious and requires thorough understanding of the (polymer) physics. Moreover, the lack of detailed description in the literature makes it hard to reproduce the analysis. Therefore, we decided to streamline as well as standardize this process and make it more accessible to people without formal physics background. Thus, we automatized the data preprocessing, (un)folding event assignment as well as the model fitting steps of the analysis and wrapped this pipeline into a graphical user interface (GUI). Additionally, we added the options to open individual files (for troubleshooting) or perform the (un)folding event assignment and fitting manually.

However, this is just the first step. There are other steps in the data analysis downstream of what has been described in the previous paragraph that still await for the automation and standardization. Statistical analysis of the processed data and especially the quality control of the data processing has to be resolved in order to fully streamline the whole process of optical tweezers data analysis.

As briefly illustrated above, to successfully extract the single-molecule information, there are a lot of pitfalls to avoid, overcome, or at least be aware of. As any other methods, these techniques have their strengths and weaknesses. Thus, we always attempted to complement the single-molecule analysis with ensemble methods and correlate our structural information with functional readout to ensure that the observed results are not a peculiar eccentricity of the experiment but also have biological relevance.

7.2. RNA can act as a regulation switch.

In **chapter 4** (and partially **chapter 5**), we have focused on the SARS-CoV-2 FSE RNA. The stimulatory structure in the SARS-CoV-2 FSE was initially thought to be a three-stemmed pseudoknot. Previous studies have confirmed the existence of the pseudoknot structure, however, they usually focused on short stretches of RNA comprising the expected pseudoknot forming sequence (19-21). Contrary to this evidence, recent work employing other approaches, such as chemical probing, which brings information about whole genome folding in ensemble, they often failed to identify the pseudoknot (22-27). Finally, Bhatt et al. showed ribosomes stalled at the mutated slippery sequence of the SARS-CoV-2 FSE RNA with pseudoknot being present, using cryo-EM (28).

We were able to identify the pseudoknot fold previously proposed by the other studies (19-21), as well as the alternative stem 1 structures proposed by the genome-wide chemical probing studies (22-25). Moreover, we even identified some of the alternative conformations being present and ruled out their ability to induce frameshifting (26, 29, 30).

Our findings suggest that the genomic RNA of SARS-CoV-2 is likely "trapped" in an intermediate state (similar to the alternative stem 1 conformation) and only once this region is unavailable for base-pairing (like during translation by ribosome) the RNA folds into the pseudoknot RNA. This is a very tempting hypothesis. The viral RNA has to be able to undergo several different processes during the life cycle, which are often separated in time and space (cellular localization) – these include translation, replication/transcription or viral packaging. From the evolutionary point of view, it is convenient for the virus to create the RNA roadblock, in the form of pseudoknot, only transiently under certain conditions. While the stable structure is crucial for the induction of –1PRF, it can be fatal for RNA replication. Additionally, the AS1 and pseudoknot conformations differ only by the absence/presence of the SL1. SL3 as well as the pseudoknot forming SL2 are part of this alternative conformer, making the temptation to see this intermediate as "ready-to-switch" even stronger.

In the future, additional work is planned to prove this hypothesis. Computational modeling can answer whether the AS1 alternative conformer can indeed switch to pseudoknot structure without disrupting the SL2, or perhaps if this could be the reason why the "quasi-knot" structure is formed (19, 20). Testing the effect of translating, or stalled, ribosomes on RNA folding is another possible way. This could be tested both in bulk as well as at a single-molecule level. Finally, to prove the relevance of this hypothesis for the RNA replication, testing of the premature termination or drop-off of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of the pseudoknot should also be tested. This is of the highest importance, as the SARS-CoV-2 FSE, specifically the pseudoknot, is an attractive target for therapeutic development. But what if all the efforts so far were aimed towards a structure that is present only transiently?

7.3. *Trans*-factors affecting frameshifting - yet another layer of gene expression regulation.

Cis elements allow the frameshifting to occur without presence of any additional factors. This is particularly convenient as the virus does not need to bring any co-factors along with the RNA. However, this provides rather static control over frameshifting efficiency. As we have shown in **chapter 5** and **6**, an attractive way to further fine-tune the frameshifting is employment of *trans*-factors. Be it a protein, metabolite, small molecule or

perhaps another RNA molecule, engagement of these factors allows conditional regulation of frameshifting, such as spatial or time regulations.

Due to its evolutionary conservation, frameshifting is attractive target not only for us, as humankind, but also for our cells in the never-ending evolutionary arms race. This assumption was confirmed by several studies in the last few years, including the ZAP-S identification described in **chapter 5** (31-33). While in principle the inhibition of viral replication can be achieved by disbalancing the frameshifting efficiency in any direction, most of the documented host factors work through reduction of the frameshifting efficiency, either through impeding the RNA structure formation (ZAP-S) or through interaction with the ribosome (Shiftless). Aiming to inhibit the frameshifting rather than increase it is a reasonable goal from the evolutionary point of view for several reasons: (**i**) translation is highly regulated and energetically demanding process, therefore reduction of non-canonical translation events and tightening of the quality control appears to be more straightforward approach; (**ii**) even innate host mRNAs are often structured to certain degrees and they have to be unwound by the ribosome as they are translated. Therefore, stabilizing the RNA stimulatory structure might result in ribosome stalling and, eventually, a drop-off. (**iii**) the proposed thermodynamic-kinetic mechanism of frameshifting assumed that the maximal frameshifting efficiency is determined by the slippery sequence and the stable RNA structure downstream creates a time window to establish the equilibrium between different reading frames, thus stabilization of the RNA structure might not result in further increase of frameshifting given. Therefore, although more evidence is still required, it is likely that the frameshifting inhibiting *trans*-factors will be more prevalent among the host-encoded proteins.

On the other hand, viruses have completely opposite intentions. In **chapter 6**, we have built on the previously documented phenomenon of the EMCV virus employing its virus-encoded protein 2A to stimulate frameshifting. In our study we have described that this frameshifting stimulation is achieved through RNA structure stabilization in the presence of the 2A protein. Furthermore, the 2A protein binds to the (bacterial) ribosome and potentially competes for binding with elongation factors, thus further slowing down the ribosome progress. The virus can this way not only achieve the time regulation of the frameshifting efficiency but also optimize the use of host-cell resources. Similarly, PPRSV uses host-encoded protein to induce frameshifting (34).

This brings up some intriguing questions. Can the use of *trans*-factors be more common than we have thought? For example, HIV-1 FSE consists of very "slippery" slippery sequence UUUUUUA, yet in general the frameshifting efficiency described in the literature has been reported to range between 5-20%. Why would the virus maintain such slippery sequence if it then does not harness its potential? Or do we simply fail to see the

full picture yet? The stimulatory structure of HIV-1 FSE is generally assumed to be a relatively small stem-loop (35). What if there is no one population of RNAs with frameshifting efficiency of 5-20%, but rather a mixes population of RNAs with background level frameshifting and RNAs with fairly efficient frameshifting due to certain local environment conditions or a specific protein binding to the RNA? I hypothesize that the HIV-1 employs one of its proteins or even abuse a host-encoded protein to regulate its frameshifting efficiency. Identification of such *trans*-factors remains yet to be discovered.

7.4. Next steps – what is lost in (bulk) translation?

The aim of this thesis was to understand the mechanisms of –1PRF. However, so far, we used the single-molecule optical tweezers to acquire the structural information about the –1PRF RNA and linked this stability information with the ensemble functional assay (**chapter 4** and **5**. Fortunately for us, the two information, despite coming from separate experiments, in most of the cases (if not all) agreed together. But is this assumption always valid? Moreover, when studying the effect of *trans*-factors (**chapter 5** and **6**), we determine the mechanism of their effect by extrapolating data obtained in a very artificial and "simple" *in vitro* system onto a complex environment inside cells. This clearly comes with a high risk of artifacts, biases or misinterpretation.

One way to solve this asymmetry, is to introduce not just the RNA structures or RNA:protein interactions, but the whole translation process into the *in vitro* system. This has been achieved due to the hard work of many people previously in bulk for bacterial translations (36-40). Following this success, some teams were also successful in applying the single-molecule methods, smFRET or optical tweezers, onto the reconstituted translation system to study translation in an unprecedent time-resolution to uncover the ordered manner of individual steps, ribosomal reorganization as well as the factor binding (12, 41-52).

Moreover, the combination of smFRET and optical tweezers was used to uncover the time-order of the ribosome translocation and hairpin opening. They also show that the ribosome can "shift" gears between slow and fast paths depending on the RNA stability (53). Recently, also a combination of smFRET and single-molecule imaging CryoEM was employed to compare the kinetics as well as the structures of the prokaryotic and eukaryotic ribosomes (54).

All these achievements clearly prove that the technology as well as the methodology have ripped enough to start uncovering what has been so far lost in bulk translation. This offers an exciting chance to also dissect the mechanisms of recoding events and elucidate the mode of action of different *trans*-acting factors.

7.5. References

1. Moffitt, J.R., Chemla, Y.R., Smith, S.B. and Bustamante, C. (2008) Recent advances in optical tweezers. *Annual review of biochemistry*, **77**, 205-228.
2. Juette, M.F., Terry, D.S., Wasserman, M.R., Zhou, Z., Altman, R.B., Zheng, Q. and Blanchard, S.C. (2014) The bright future of single-molecule fluorescence imaging. *Curr Opin Chem Biol*, **20**, 103-111.
3. Ashkin, A., Dziedzic, J.M., Bjorkholm, J.E. and Chu, S. (1986) Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.*, **11**, 288-290.
4. Binnig, G., Quate, C.F. and Gerber, C. (1986) Atomic Force Microscope. *Physical Review Letters*, **56**, 930-933.
5. Ha, T., Enderle, T., Ogletree, D.F., Chemla, D.S., Selvin, P.R. and Weiss, S. (1996) Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences*, **93**, 6264-6268.
6. Bustamante, C., Cheng, W. and Mejia, Y.X. (2011) Revisiting the Central Dogma One Molecule at a Time. *Cell*, **144**, 480-497.
7. Bustamante, C.J., Chemla, Y.R., Liu, S. and Wang, M.D. (2021) Optical tweezers in single-molecule biophysics. *Nature Reviews Methods Primers*, **1**, 25.
8. Elms, P.J., Chodera, J.D., Bustamante, C.J. and Marqusee, S. (2012) Limitations of constant-force-feedback experiments. *Biophysical journal*, **103**, 1490-1499.
9. Cossio, P., Hummer, G. and Szabo, A. (2015) On artifacts in single-molecule force spectroscopy. *Proceedings of the National Academy of Sciences*, **112**, 14248-14253.
10. König, S.L.B., Kowerko, D. and Sigel, R.K.O. (2013) Kinetic Subpopulations Detected by Single-molecule Spectroscopy: Fundamental Property of Functional Nucleic Acids or Experimental Artefact? *CHIMIA*, **67**, 240.
11. Gore, J., Ritort, F. and Bustamante, C. (2003) Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc Natl Acad Sci U S A*, **100**, 12564-12569.
12. Roy, R., Hohng, S. and Ha, T. (2008) A practical guide to single-molecule FRET. *Nature methods*, **5**, 507-516.
13. Riveline, D. (2013) 'Single molecule': theory and experiments, an introduction. *J Nanobiotechnology*, **11**, S1.
14. Reddington, S.C., Tippmann, E.M. and Dafydd Jones, D. (2012) Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted Click chemistry. *Chemical Communications*, **48**, 8419-8421.
15. Jahnke, K., Grubmüller, H., Igaev, M. and Göpfrich, K. (2021) Choice of fluorophore affects dynamic DNA nanostructures. *Nucleic Acids Research*, **49**, 4186-4195.
16. Wen, J.-D., Manosas, M., Li, P.T.X., Smith, S.B., Bustamante, C., Ritort, F. and Tinoco, I. (2007) Force Unfolding Kinetics of RNA Using Optical Tweezers. I. Effects of Experimental Variables on Measured Results. *Biophysical Journal*, **92**, 2996-3009.
17. Aitken, C.E., Marshall, R.A. and Puglisi, J.D. (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical journal*, **94**, 1826-1835.
18. Swoboda, M., Henig, J., Cheng, H.-M., Brugger, D., Haltrich, D., Plumeré, N. and Schlierf, M. (2012) Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments. *ACS Nano*, **6**, 6364-6369.
19. Neupane, K., Zhao, M., Lyons, A., Munshi, S., Ileperuma, S.M., Ritchie, D.B., Hoffer, N.Q., Narayan, A. and Woodside, M.T. (2021) Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers. *Nat. Commun.*, **12**, 4749.

20. Zhang, K., Zheludev, I.N., Hagey, R.J., Haslecker, R., Hou, Y.J., Kretsch, R., Pintilie, G.D., Rangan, R., Kladwang, W., Li, S. *et al.* (2021) Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nat. Struct. Mol. Biol*, **28**, 747-754.
21. Roman, C., Lewicka, A., Koirala, D., Li, N.-S. and Piccirilli, J.A. (2021) The SARS-CoV-2 Programmed −1 Ribosomal Frameshifting Element Crystal Structure Solved to 2.09 Å Using Chaperone-Assisted RNA Crystallography. *ACS Chem. Biol.*, **16**, 1469-1481.
22. Manfredonia, I., Nithin, C., Ponce-Salvatierra, A., Ghosh, P., Wirecki, T.K., Marinus, T., Ogando, N.S., Snijder, E.J., van Hemert, M.J., Bujnicki, J.M. *et al.* (2020) Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res*, **48**, 12436-12452.
23. Sun, L., Li, P., Ju, X., Rao, J., Huang, W., Ren, L., Zhang, S., Xiong, T., Xu, K., Zhou, X. *et al.* (2021) In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. *Cell*, **184**, 1865-1883.e1820.
24. Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., Yang, M., Xing, X., Wang, Y., Li, M. *et al.* (2021) The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat. Commun.*, **12**, 3917.
25. Lan, T.C.T., Allan, M.F., Malsick, L.E., Woo, J.Z., Zhu, C., Zhang, F., Khandwala, S., Nyeo, S.S.Y., Sun, Y., Guo, J.U. *et al.* (2022) Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat. Commun.*, **13**, 1128.
26. Yan, S., Zhu, Q., Jain, S. and Schlick, T. (2022) Length-dependent motions of SARS-CoV-2 frameshifting RNA pseudoknot and alternative conformations suggest avenues for frameshifting suppression. *Nat. Commun.*, **13**, 4284.
27. Ziv, O., Price, J., Shalamova, L., Kamenova, T., Goodfellow, I., Weber, F. and Miska, E.A. (2020) The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol. Cell*, **80**, 1067-1077.e1065.
28. Bhatt, P.R., Scaiola, A., Loughran, G., Leibundgut, M., Kratzel, A., Meurs, R., Dreos, R., O'Connor, K.M., McMillan, A., Bode, J.W. *et al.* (2021) Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science*, **372**, 1306-1313.
29. Schlick, T., Zhu, Q., Dey, A., Jain, S., Yan, S. and Laederach, A. (2021) To Knot or Not to Knot: Multiple Conformations of the SARS-CoV-2 Frameshifting RNA Element. *J. Am. Chem. Soc.*, **143**, 11404-11422.
30. Schlick, T., Zhu, Q., Jain, S. and Yan, S. (2021) Structure-altering mutations of the SARS-CoV-2 frameshifting RNA element. *Biophys. J.*, **120**, 1040-1053.
31. Rehfeld, F., Eitson, J.L., Ohlson, M.B., Chang, T.-C., Schoggins, J.W. and Mendell, J.T. (2023) CRISPR screening reveals a dependency on ribosome recycling for efficient SARS-CoV-2 programmed ribosomal frameshifting and viral replication. *Cell Reports*, **42**.
32. Zimmer, M.M., Kibe, A., Rand, U., Pekarek, L., Ye, L., Buck, S., Smyth, R.P., Cicin-Sain, L. and 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**, 7193.
33. Schmidt, N., Lareau, C.A., Keshishian, H., Ganskih, S., Schneider, C., Hennig, T., Melanson, R., Werner, S., Wei, Y., Zimmer, M. *et al.* (2021) The SARS-CoV-2 RNA–protein interactome in infected human cells. *Nature Microbiology*, **6**, 339-353.
34. Napthine, S., Treffers, E.E., Bell, S., Goodfellow, I., Fang, Y., Firth, A.E., Snijder, E.J. and Brierley, I. (2016) A novel role for poly(C) binding proteins in programmed ribosomal frameshifting. *Nucleic Acids Res*, **44**, 5491-5503.
35. Low, J.T., Garcia-Miranda, P., Mouzakis, K.D., Gorelick, R.J., Butcher, S.E. and Weeks, K.M. (2014) Structure and Dynamics of the HIV-1 Frameshift Element RNA. *Biochemistry*, **53**, 4282-4291.

36. Michel, E., Duss, O. and Allain, F.H.T. (2018) In Arluison, V. and Valverde, C. (eds.), *Bacterial Regulatory RNA: Methods and Protocols*. Springer New York, New York, NY, pp. 177-195.
37. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nature Biotechnology*, **19**, 751-755.
38. Pestova, T.V. and Hellen, C.U.T. (2005) Reconstitution of eukaryotic translation elongation in vitro following initiation by internal ribosomal entry. *Methods*, **36**, 261-269.
39. Rodnina, M.V., Semenkov, Y.P. and Wintermeyer, W. (1994) Purification of fMet-tRNA(fMet) by fast protein liquid chromatography. *Anal Biochem*, **219**, 380-381.
40. Kothe, U., Paleskava, A., Konevega, A.L. and Rodnina, M.V. (2006) Single-step purification of specific tRNAs by hydrophobic tagging. *Anal Biochem*, **356**, 148-150.
41. Milon, P., Konevega, A.L., Peske, F., Fabbretti, A., Gualerzi, C.O. and Rodnina, M.V. (2007) Transient kinetics, fluorescence, and FRET in studies of initiation of translation in bacteria. *Methods Enzymol*, **430**, 1-30.
42. Kim, H.-K. and Tinoco, I., Jr. (2017) EF-G catalyzed translocation dynamics in the presence of ribosomal frameshifting stimulatory signals. *Nucleic Acids Research*, **45**, 2865-2874.
43. Bao, C., Loerch, S., Ling, C., Korostelev, A.A., Grigorieff, N. and Ermolenko, D.N. (2020) mRNA stem-loops can pause the ribosome by hindering A-site tRNA binding. *Elife*, **9**.
44. Kaiser, C.M., Goldman, D.H., Chodera, J.D., Tinoco, I. and Bustamante, C. (2011) The Ribosome Modulates Nascent Protein Folding. *Science*, **334**, 1723-1727.
45. Goldman, D.H., Kaiser, C.M., Milin, A., Righini, M., Tinoco, I. and Bustamante, C. (2015) Mechanical force releases nascent chain–mediated ribosome arrest in vitro and in vivo. *Science*, **348**, 457-460.
46. Maciuba, K., Zhang, F. and Kaiser, C.M. (2021) Facile tethering of stable and unstable proteins for optical tweezers experiments. *Biophysical Journal*, **120**, 2691-2700.
47. Maciuba, K., Rajasekaran, N., Chen, X. and Kaiser, C.M. (2021) Co-translational folding of nascent polypeptides: Multi-layered mechanisms for the efficient biogenesis of functional proteins. *BioEssays*, **43**, 2100042.
48. Liu, T., Kaplan, A., Alexander, L., Yan, S., Wen, J.-D., Lancaster, L., Wickersham, C.E., Fredrick, K., Noller, H., Tinoco, I., Jr. *et al.* (2014) Direct measurement of the mechanical work during translocation by the ribosome. *eLife*, **3**, e03406.
49. Yan, S., Wen, J.D., Bustamante, C. and Tinoco, I., Jr. (2015) Ribosome excursions during mRNA translocation mediate broad branching of frameshift pathways. *Cell*, **160**, 870-881.
50. Alexander, L., Goldman, D., Tinoco, I. and Bustamante, C.J. (2017) The Ribosome Alters the Folding of a Multidomain Nascent Protein. *Biophysical Journal*, **112**, 40a-41a.
51. Duss, O., Stepanyuk, G.A., Puglisi, J.D. and Williamson, J.R. (2019) Transient Protein-RNA Interactions Guide Nascent Ribosomal RNA Folding. *Cell*, **179**, 1357-1369.e1316.
52. Qureshi, N. and Duss, O. (2023) Real-time tracking of transcription-translation coupling. *Biophysical Journal*, **122**, 488a.
53. Desai, V.P., Frank, F., Lee, A., Righini, M., Lancaster, L., Noller, H.F., Tinoco, I. and Bustamante, C. (2019) Co-temporal Force and Fluorescence Measurements Reveal a Ribosomal Gear Shift Mechanism of Translation Regulation by Structured mRNAs. *Molecular Cell*, **75**, 1007-1019.e1005.
54. Holm, M., Natchiar, S.K., Rundlet, E.J., Myasnikov, A.G., Watson, Z.L., Altman, R.B., Wang, H.-Y., Taunton, J. and Blanchard, S.C. (2023) mRNA decoding in human is kinetically and structurally distinct from bacteria. *Nature*, **617**, 200-207.

Final thoughts – science must be accessible to everyone

Science is evolving. Science is changing (1). We wish to think that the scientific method, although old, is still THE way of verifying new knowledge. At the same time, we, as scientists, are often (unconsciously) finding our way around this concept. Science nowadays is polluted with falsely positive results, which just by chance reached the psychological, yet purely arbitrary, threshold of p-value < 0.05 in the race to publish-or-perish (2-5). The scarce information included in the methods sections (partially due to manuscript size limit from the publishers) often accompanied by the rabbit hole referencing also has its share in the current reproducibility crisis (6, 7). Be it because of poor statistical education among the scientists, random chance, negligence, or perhaps intention, the lack of reproducibility creates a huge burden on current academic research (8), which in turn results in a waste of money (9).

Science is mostly funded by public money. As such, it is our duty to be able to communicate (and defend) our science in front of the public. But if we want to defend or communicate our science in front of the public, we first must be able to defend it in front of ourselves. There are currently efforts to change this trend (10-14). Projects like biorxiv (or other arxiv-type websites), data depositories, or protocols.io (15) aim to remove some of the reproducibility issues. However, to solve this problem, everyone has to act and do their part.

During my PhD journey, I have faced a lot of missing information, gaps in the analysis descriptions, or "data upon reasonable request" barriers. It is outrageous, it is unnecessary, and it is slowing us down. Hiding as an academic strategy has to end (16, 17). I believe in open science, in sharing your data (in database servers), and in sharing the methods and code (on Github or as a pseudocode) (18). As much as it can hurt one's ego if others find mistakes in our experiments or analysis, it is part of the actual scientific method we have been praising. I am glad and very thankful, that during my PhD, besides the experimental manuscripts, I was also part of the two methodological manuscripts (which are sadly behind a paywall, but also available on biorxiv). Our JoVE manuscript (**chapter 2**) aims to open the optical tweezers field to newcomers. While it by no chance comprehends the full complexity of the topic, I hope it can serve as a practical guide and introduction to the field of single-molecule biophysics. As simple as it is, the POTATO is, to my best knowledge, the first truly user-friendly pipeline for optical tweezers data analysis (**chapter 3**). It has flaws and bugs, which we are discovering and correcting as we apply it to more and more data, but it has the potential to not only streamline and speed up the data analysis but also standardize the otherwise often manual analysis in the field, which suffers from the insufficient data analysis description.

I am also happy that we were able to reproduce data on the SARS-CoV-2 pseudoknot RNA published by other groups (19), and further extend the knowledge about its behavior (**chapter 4** and **5**).

Science should be open to everyone, this does not mean just transferring the paywall burden on the authors, it means we as scientists have to share our data, methods and pipelines to everyone.

Science, or academia for that matter, is not perfect. They are just as we make them.

References

1. Hausmann, L., Murphy, S.P. and the Publication Committee of the International Society for, N. (2016) The challenges for scientific publishing, 60 years on. *Journal of Neurochemistry*, **139**, 280-287.
2. Ioannidis, J.P.A. (2005) Why Most Published Research Findings Are False. *PLOS Medicine*, **2**, e124.
3. Rawat, S. and Meena, S. (2014) Publish or perish: Where are we heading? *J Res Med Sci*, **19**, 87-89.
4. Moonesinghe, R., Khoury, M.J. and Janssens, A.C. (2007) Most published research findings are false-but a little replication goes a long way. *PLoS Med*, **4**, e28.
5. Nissen, S.B., Magidson, T., Gross, K. and Bergstrom, C.T. (2016) Publication bias and the canonization of false facts. *Elife*, **5**.
6. Altman, D.G. (2015) Making research articles fit for purpose: structured reporting of key methods and findings. *Trials*, **16**, 53.
7. Brown, A.W., Kaiser, K.A. and Allison, D.B. (2018) Issues with data and analyses: Errors, underlying themes, and potential solutions. *Proceedings of the National Academy of Sciences*, **115**, 2563-2570.
8. Fanelli, D. (2009) How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*, **4**, e5738.
9. Freedman, L.P., Cockburn, I.M. and Simcoe, T.S. (2015) The Economics of Reproducibility in Preclinical Research. *PLOS Biology*, **13**, e1002165.
10. Besançon, L., Bik, E., Heathers, J. and Meyerowitz-Katz, G. (2022) Correction of scientific literature: Too little, too late! *PLOS Biology*, **20**, e3001572.
11. Bergstrom, T.C. (2001) Free Labour for Costly Journals? *Journal of Economic Perspectives*, **15**, 183-198.
12. Lehner, E. and Ziegler, J. (2019) *Free Information, Not Free Labor*.
13. Ioannidis, J.P., Fanelli, D., Dunne, D.D. and Goodman, S.N. (2015) Meta-research: Evaluation and Improvement of Research Methods and Practices. *PLoS Biol*, **13**, e1002264.
14. Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, B.A., Wagenmakers, E.J., Berk, R., Bollen, K.A., Brembs, B., Brown, L., Camerer, C. *et al.* (2018) Redefine statistical significance. *Nature Human Behaviour*, **2**, 6-10.
15. Teytelman, L., Stoliartchouk, A., Kindler, L. and Hurwitz, B.L. (2016) Protocols.io: Virtual Communities for Protocol Development and Discussion. *PLoS Biol*, **14**, e1002538.
16. Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H. *et al.* (2021) Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, **8**, 192.
17. Langille, M.G.I., Ravel, J. and Fricke, W.F. (2018) "Available upon request": not good enough for microbiome data! *Microbiome*, **6**, 8.
18. Allen, C. and Mehler, D.M.A. (2019) Open science challenges, benefits and tips in early career and beyond. *PLoS Biol*, **17**, e3000246.
19. Neupane, K., Zhao, M., Lyons, A., Munshi, S., Ileperuma, S.M., Ritchie, D.B., Hoffer, N.Q., Narayan, A. and Woodside, M.T. (2021) Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers. *Nat. Commun.*, **12**, 4749.

List of Abbreviations

3WJ	three-way junction
AA	amino acid
AD bead	bead coated with antibodies against digoxigenin
AFM	atomic force microscopy
APD	avalanche photodiode
AS	alternative stem
ASO	antisense oligonucleotides
BFP	back focal plane
Bp	base pair
CCD	charge-coupled device
cDNA	coding DNA
CHIKV	Chikungunya virus
COVID-19	coronavirus disease 2019
Cryo-EM	cryogenic electron microscopy
DMS	dimethyl sulfate
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DREEM	Detection of RNA folding Ensembles using Expectation-Maximization
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
DTT	dithiothreitol
EDTA	ethylendiaminetetraacetate
(e)EF	(eukaryotic) elongations factor
EMSA	electrophoretic mobility shift assay
EMCV	Encephalomyocarditis virus
F	force
FD curve	force-distance curve
FE	frameshift efficiency
FEL	free-energy landscape
FJC	freely-jointed chain model
FRET	Förster resonance energy transfer
FSE	frameshift stimulatory element
GDP	guanosine diphosphate
(E)CFP	(enhanced) cyan fluorescent protein

(E)GFP	(enhanced) green fluorescent protein
GTP	guanosine triphosphate
GUI	graphical user interface
HIV	human immunodeficiency virus
IRES	internal ribosome entry site
ITC	isothermal titration calorimetry
IVT	*in vitro* transcription
JEV	Japanese encephalitis virus
L_C	contour length
LNA	locked nucleic acid
L_P	persistence length
MERS-CoV	Middle East respiratory syndrome–related coronavirus
MIME	mutational interference mapping experiment
mRNA	messenger RNA
miRNA	micro RNA
MST	microscale thermophoresis
NMR	nuclear magnetic resonance
Nt	nucleotide
NTP	nucleotide triphosphate
ORF	open reading frame
OT	optical tweezers
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PK	peudoknot
POTATO	Practical Optical Tweezers Analysis TOol
PRF	programmed ribosomal frameshifting
qRT-PCR	quantitative real-time PCR
RBP	RNA-binding protein
RdRP	RNA-dependent RNA polymerases
(e)RF	(eukaryotic) release factor
RNA	ribonucleic acid
RPM	rotations per minute
RRL	rabbit reticulocyte lysate
rRNA	ribosomal RNA
SA bead	streptavidin coated bead
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SD	Shine-Dalgarno sequence

SD	standard devitation
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC-MALS	size exclusion chromatography coupled with multiangle light scattering
SFL	shiftless
SL	stem-loop
SM	steerable mirror
SoI	sequence of interest
SS	slippery sequence
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
TMEV	Theiler's murine encephalomyelitis virus
tRNA	transfer RNA
UTR	untranslated region
W	work
WLC	worm-like chain model
WNV	West Nile virus
WT	wild type
ZAP	Zinc-finger antiviral protein

List of Figures

List of Tables

Statement of individual author contributions to figures

Manuscript 1 (complete reference):
Pekarek, L., Buck, S., Caliskan, N., Optical Tweezers to Study RNA-Protein Interactions in Translation Regulation. 2022, Journal of Visualized Experiments (JoVE), 180, e62589.

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	LP	SB	NC		
2	LP	SB	NC		
3	LP	SB	NC		
4	LP	SB	NC		
5	LP	SB	NC		
6	LP	SB	NC		
7	LP	SB	NC		
8	LP	SB	NC		

Explanations (if applicable):

Manuscript 2 (complete reference):
Buck, S.*, Pekarek, L.*, Caliskan, N., POTATO: Automated pipeline for batch analysis of optical tweezers data. 2022, Biophysical Journal, 15, 121, 2830-2839.
*equal contributions

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

Explanations (if applicable): Supplementary figures were created by LP, supplementary tables were done together by LP and SB.

Manuscript 3 (complete reference):

Pekarek, L., Zimmer, M.M., Gribling, A.S., Buck, S., Smyth, R.P., Caliskan, N., Cis-mediated interactions of the SARS-CoV-2 frameshift RNA alter its conformations and affect function. 2022, Nucleic Acids Research, gkac1184.

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	LP	NC	MZ	RS	
2	LP	NC	RS	MZ	ASG
3	RS	LP	ASG	NC	
4	LP	RS	ASG	NC	SB/MZ
5	LP	NC			
6	MZ	NC	LP		
7	LP	NC	RS	MZ	

Explanations (if applicable):

Manuscript 4 (complete reference):

Zimmer, M.M.*, Kibe, A.*, Rand, U., **Pekarek, L.**, Ye, L., Buck, S., Smyth, P.S., Cicin-Sain, L., Caliskan, N., The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting. 2021, Nature Communication, 12, 7193.

*equal contributions

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

Explanations (if applicable):

Manuscript 5 (complete reference):

Hill, C.H.*, **Pekarek, L.***, Napthine, S.*, Kibe, A., Firth, A.E., Graham, S.C., Caliskan, N., Brierley, I. Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch. 2021, Nature Communication, 12, 7166

* equal contributions

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	CHH	SCG	SN	IB	AF
2	CHH	AK	NC	SN	
3	LP	NC	CHH		
4	AK	NC			
5	CHH	SN	SCG	AF	IB
6	CHH	SN	SCG	AF	IB
7	CHH	SN	SCG	AF	IB
8	CHH	LP	NC	SCG/AF	IB

Explanations (if applicable):

I also confirm my primary supervisor's acceptance.

_____

Doctoral Researcher's Name Date Place Signature

Statement of legal second publication rights

Manuscripts 1-5

Statement of author contributions

Manuscript 1 (complete reference): **Pekarek, L.**, Buck, S., Caliskan, N., Optical Tweezers to Study RNA-Protein Interactions in Translation Regulation. 2022, *Journal of Visualized Experiments (JoVE),* 180, e62589.					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	LP	NC	SB		
Data Collection	LP	SB			
Data Analysis and Interpretation	LP	SB	NC		
Manuscript Writing	LP	SB	NC		
Writing of Introduction	NC	LP	SB		
Writing of Materials & Methods	LP	SB	NC		
Writing of Discussion	NC	LP	SB		
Writing of First Draft	LP	SB			

Manuscript 2 (complete reference): Buck, S.*, **Pekarek, L.***, Caliskan, N., POTATO: Automated pipeline for batch analysis of optical tweezers data. 2022, Biophysical Journal, 15, 121, 2830-2839. *equal contributions					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	LP/SB	NC			
Data Collection	LP/SB				
Data Analysis and Interpretation	LP/SB	NC			
Manuscript Writing	LP/SB	NC			
Writing of Introduction	LP/SB	NC			

Writing of Materials & Methods	LP/SB	NC			
Writing of Discussion	NC	LP/SB			
Writing of First Draft	LP/SB				

Manuscript 3 (complete reference):

Pekarek, L., Zimmer, M.M., Gribling, A.S., Buck, S., Smyth, R.P., Caliskan, N., Cis-mediated interactions of the SARS-CoV-2 frameshift RNA alter its conformations and affect function. 2022, Nucleic Acids Research, gkac1184.

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

Manuscript 4 (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 5 (complete reference):

Hill, C.H.*, **Pekarek, L.***, Napthine, S.*, Kibe, A., Firth, A.E., Graham, S.C., Caliskan, N., Brierley, I. Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch. 2021, Nature Communication, 12, 7166

*equal contribution

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	CHH/NC/ASG	LP/AK/SN	LP/AK	IB/AF	
Data Collection	CHH	LP/AK	SN/AEG		
Data Analysis and Interpretation	CHH	LP/AK	NC/AEG/IB	AF/SN	
Manuscript Writing	CHH/NC	AEG/IB/LP	AF/SN	AK	
Writing of Introduction	CHH/NC	LP/AEG/SN	AF/IB		
Writing of Materials & Methods	CHH	LP/SN/AK	NC/AEG/AF	IB	
Writing of Discussion	CHH/NC/IB	AEG/AF	LP/AK/SN		
Writing of First Draft	CHH	LP/SN	NC/AEG/AF/IB	AK	

The doctoral researcher and the primary supervisor confirm the correctness of the above mentioned assessment.

Doctoral Researcher's Name	Date	Place	Signature

Primary Supervisor's Name	Date	Place	Signature

Appendix A – Publications during candidature

Pekarek, L., Zimmer, M.M., Gribling, A.S., Buck, S., Smyth, R.P., Caliskan, N., Cis-mediated interactions of the SARS-CoV-2 frameshift RNA alter its conformations and affect function. 2022, Nucleic Acids Research, gkac1184.

Buck, S.*, **Pekarek, L.***, Caliskan, N., POTATO: Automated pipeline for batch analysis of optical tweezers data. 2022, Biophysical Journal, 15, 121, 2830-2839.

Pekarek, L., Buck, S., Caliskan, N., Optical Tweezers to Study RNA-Protein Interactions in Translation Regulation. 2022, *Journal of Visualized Experiments (JoVE),* 180, e62589.

Zimmer, M.M., Kibe, A., Rand, U., **Pekarek, L.**, Ye, L., Buck, S., Smyth, P.S., Cicin-Sain, L., Caliskan, N., The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting. 2021, Nature Communication, 12, 7193.

Hill, C.H.*, **Pekarek, L.***, Napthine, S.*, Kibe, A., Firth, A.E., Graham, S.C., Caliskan, N., Brierley, I. Structural and molecular basis for Cardiovirus 2A protein as a viral gene expression switch. 2021, Nature Communication, 12, 7166

*Shared co-first authorship

Appendix B – Acknowledgements

I would like to first thank my primary supervisor **Neva Caliskan** for allowing me to work in her amazing team and supporting me throughout my PhD journey. Thank you for always being open to scientific discussions regardless of how ridiculous the ideas were. The dynamic environment that you created truly stimulated my growth as a scientist in the last four years. I wouldn't reach this far without you always pushing me forward.

I would also like to thank my thesis committee members, **Markus Sauer** and **Utz Fischer**, for their insight and suggestions throughout my PhD.

Throughout my PhD, I had a chance to participate in the guidance of several young scientists. Each of them was unique, and each helped me grow as well. **Mara**, thank you for being my first bachelor student in REMI. We went through many struggles together as we both just started in the lab around the same time. I often did not know the answer to your questions, but we always found out together. **Samuel**, thank you for showing me how obnoxious I must have been to my own bachelor supervisor. You are one of the smartest people I know, and I am sure you will do great despite the hurdles you have to face. **Stefan**, I still cannot believe I was teaching YOU how to write a code at some point. I am very thankful for having the chance to be there throughout your stay in REMI. From colleagues, we became friends, and eventually something even more - we are POTA-Bros – a relationship one of a kind.

I was fortunate to be part of a great team. It would have been much less fun without the REMI people. My huge thanks go to **Tatyana Koch**. I have no idea how you keep things in such a great order, and I am grateful for you pushing us to obey the lab rules. Thank you for always being there whenever I needed to talk about science or my personal life.

Ricarda, you might not seem like the most excited person from the outside, but your dedication to your cause is something I truly admire. Be it long cell culture days, weekends in the lab, or learning R on your own, excitement has many phenotypes. Our morning coffees and a nice chat were often my big motivation to wake up.

Anuja, I have no idea how I will socialize with people once you are not around… You were always there for me, and I hope it also applies back. We went through many things together, starting from the black eye at Oktoberfest and (hopefully not) ending with your wedding. Thank you for making me a part of your life. Like a big sister, you are always the voice of empathy and sensitivity whenever I lack mine.

Matthias, I can hardly describe all the knowledge I gayned thanks to you. You were like my older brother, or uncle, or father-in-law, or whatever you call it in Saarland. Despite REMI being one of the smallest teams at HIRI, one always felt surrounded by many people, thanks to all your personalities. You probably heard that before, but I really like your HUGE… collection of orchids. But jokes ass-ide. Some people are simply unique and irreplaceable, and you are one of them. I am so grateful I got to know you.

Sharing the lab, and for a year also the office, with GARV allowed me to make many great friends. **Redmond**, you were always there for scientific discussions, and I would not join HIRI without you. Thank you & be careful about the melons. **Anne-Sophie**, your choice of music in the lab was always a mood boost for me. **Charlene**, I enjoyed our talks whenever you felt lonely in the lab. **Uddhav**, I am very proud of your growth ever since the start of the PhD. Life was not easy on you, but you did not give up, and I look forward to defending

side-by-side together. **Liquing**, I know "life is hard" but it was nice to go through some of the struggles together. **Shazeb**, thank you for teaching me the difference between hot and spicy, but also for being a great friend. **Marco**, please be careful, you know yourself. My life would have been much less fun without you. "Walls are meant for climbing" but I am glad you chose not to and ended in Europe. **Patrick**, I have never seen a person more excited about what they are doing. I wish I were as enthusiastic as you. Thank you for being a member of our two-member 3D printing club at HIRI. Thank you also for always being there for me, always cheering me up, and showing me the ways not only at the boulder wall but also in life.

Many people from other labs of (not only) HIRI were always open for a beer or just a friendly chat. Amongst others, I would like to thank **Jens,** namely. Thank you for always keeping smiling and pushing for socialization during the times when it was the most needed – the pandemic. You, for sure, have the HIRI spirit in you!

Being an international student, who does not speak the country's language, makes you appreciate all the support from the administrative staff. I don't know how I would manage all the forms filling, reimbursements, and hauspost stuff on my own. **Julia**, thank you for always remaining calm regardless of how stupid my questions were.

Life is a journey, and I could not make it without the support of many people. I am very thankful to my friends and family for supporting us in the crazy idea of moving abroad. Thanks to all the members of "**Pivní Pátek**" for always making sure you have time for a beer whenever we happened to be back in the Czech Republic, and huge thanks to **Pavel** and **Tomáš** for their help while we were moving to Germany.

My parents always supported me in doing what I was passionate about. They helped me in pursuing what they were not allowed to. I will always be thankful for that.

Děkuji, **Mami** a **Tati**.

"Marriage lets you annoy one special person for the rest of your life."

I cannot imagine having a better, more supportive, and loving partner than you, **Kristýna**. Thank you for always being there, supporting me in my crazy ideas, and calming down the really, really crazy ones.

Even during the thesis writing, you were there for me, despite actively "working" on the biggest project of our lives, our son.

Thank you & I love you.

Appendix C – Curriculum vitae

Appendix D - Affidavit

I hereby confirm that my thesis entitled “Single-Molecule Approaches To Study Frameshifting Mechanisms” 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 examination process neither in identical nor in similar form.

Hiermit erkläre ich an Eides statt, die Dissertation ‚Einzelmolekülansätze zur Untersuchung von Frameshifting-Mechanismen‘ 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, 30/06/2023

Place, Date Signature