Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2025 March 19.

Published in final edited form as:

Cell Rep. 2025 February 25; 44(2): 115236. doi:10.1016/j.celrep.2025.115236.

A conserved class of viral RNA structures regulates translation reinitiation through dynamic ribosome interactions

Madeline E. Sherlock[1,2], **Conner J. Langeberg**[1,3], **Katherine E. Segar**[1,4], **Jeffrey S. Kieft**[1,2,4,5,*]

[1]Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

[2]Present address: New York Structural Biology Center, New York, NY, 10027, USA

[3]Present address: Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, USA.

[4]Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA

[5]Lead contact

SUMMARY

Certain viral RNAs encode proteins downstream of their main open reading frame, expressed through “termination-reinitiation” events. In some cases, structures located upstream of the first stop codon within these viral RNAs bind the ribosome, inhibiting ribosome recycling and inducing reinitiation. We used bioinformatics methods to identify new examples of viral reinitiation-stimulating RNAs and experimentally verified their secondary structure and function. We determined the structure of a representative viral RNA-ribosome complex using cryoelectron microscopy (cryo-EM). 3D classification and variability analyses reveal that the viral RNA structure can sample a range of conformations while remaining tethered to the ribosome, enabling the ribosome to find a reinitiation start site within a limited range of mRNA sequence. Evaluating the conserved features and constraints of this entire RNA class within the context of the cryo-EM reconstruction provides insight into mechanisms enabling reinitiation, a translation regulation strategy employed by many other viral and eukaryotic systems.

Graphical Abstract

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

*Correspondence: jkieft@nysbc.org.

AUTHOR CONTRIBUTIONS

M.E.S., conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, visualization, writing – original draft, and writing – review & editing; C.J.L., investigation, formal analysis, and writing – review & editing; K.E.S., investigation and writing – review & editing; J.S.K., funding acquisition, supervision, visualization, and writing – review & editing.

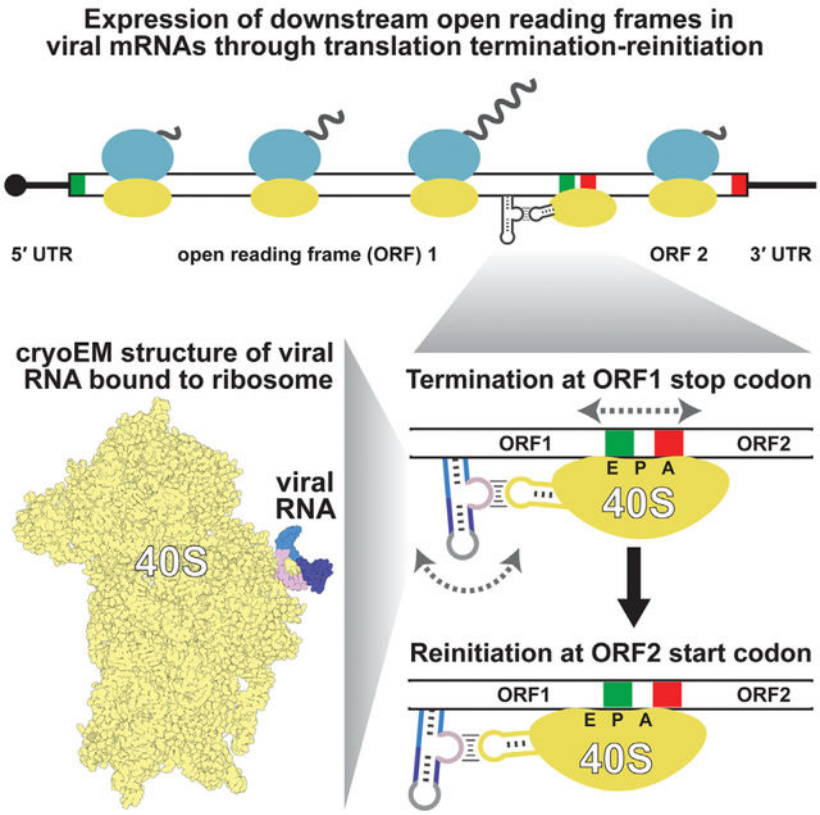
DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.115236.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



In brief

Certain viruses with bicistronic mRNAs use reinitiation to translate the downstream open reading frame. Sherlock et al. use bioinformatics to find new examples of RNA structures that mediate reinitiation, verify their function, and determine the structure of a viral reinitiation-stimulating RNA bound to the ribosome.

INTRODUCTION

Most eukaryotic mRNAs contain one main open reading frame (ORF) encoding a single functional protein. Translation initiation on these mRNAs requires numerous eukaryotic initiation factor (eIF) proteins, and translation starts at a start codon near the mRNA's 5′ end.[1,2] After initiation, the ribosome elongates through the coding region and then terminates on an in-frame stop codon, releasing the protein product.[3,4] Ribosome recycling then involves sequential removal of the large (60S) ribosomal subunit, transfer RNA (tRNA), and the small (40S) ribosomal subunit.[5] Under specific circumstances, termination and peptide release are followed by incomplete ribosome recycling, leading to translation "reinitiation." Specifically, mRNA downstream of the stop codon is translated by the same ribosome as the upstream ORF; thus, a second protein is expressed from a single mRNA by one ribosome.[6–8] Reinitiation is a widespread yet poorly understood translation-level mechanism to control gene expression.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Many viruses use reinitiation to expand the coding and regulatory potential of a single RNA, using ORFs translated exclusively through reinitiation that is promoted by *cis*-acting RNA structures.[9–17] Reinitiation is mechanistically distinct from other viral RNA structure-based strategies to alter translation, such as internal ribosome entry sites (IRESs), ribosomal frameshifting, and stop codon readthrough.[18–22] IRESs can induce translation of a downstream ORF in a bicistronic context but are independent of upstream translation.[23–25] With frameshifting and stop codon readthrough, the ribosome never terminates, creating fusion proteins with C-terminal extensions.[26–28] In contrast, reinitiation requires complete termination, including peptide release, resulting in two separate proteins. There is currently no standard term to broadly describe programmed reinitiation signals. We use "reinitiation-stimulating element" (RSE) to draw a parallel with the nomenclature for programmed ribosomal frameshifting events, which are induced by a frameshift-stimulating element, commonly abbreviated as "FSE."

One type of RSE promotes downstream ORF expression on bicistronic viral mRNAs in both influenza B virus, a segmented negative-sense single-stranded RNA (−ssRNA) virus in the Orthomyxoviridae family, and in Caliciviridae family members, positive-sense ssRNA (+ssRNA) viruses.[11,14,29,30] In these, two ORFs slightly overlap, and the upstream ORF stop codon and downstream ORF start codon can share nucleotides (Figure 1A). Despite distinct phylogenetic lineages, reinitiation in both viral contexts is driven by the same class of RSEs: termination upstream ribosome binding site (TURBS) RNAs.[11]

TURBS RSEs require mRNA sequences 5′ of the stop/start codon, while sequences 3′ of the stop/start are dispensable.[11,14,16] Within TURBS RNAs, motifs 1 and 2 are functionally necessary.[16] Motif 1 comprises conserved nucleotides complementary to 18S ribosomal RNA (rRNA) in the 40S subunit (Figures 1A and 1B), specifically the apex of helix 26 (expansion segment 7 [ES7]). Evidence that the TURBS binds to the ES7 apical loops includes the fact that reinitiation in yeast only occurs with a TURBS RSE mutated in motif 1 to complement this loop in yeast ribosomes.[17,31] Motif 2 comprises nucleotides that flank motif 1, with the 5′ portion termed motif 2 and the 3′ portion termed motif 2* (Figures 1A and 1B). The 2 and 2* sequences are not conserved but can base pair, and mutations that break the putative pairs decrease reinitiation.[31] Interestingly, compensatory mutations that restore base pairing do not fully restore function[17,32]; thus, this secondary structure model does not fully explain the nature of this class of RSE RNAs, and no 3D structural model has been proposed or determined.

TURBS RSE RNAs show positional conservation in that the structure is always fully upstream of the ORF1 stop codon and ORF2 start codon, but the number of intervening nucleotides varies.[13] With some flexibility, there are constraints on the distance between the RSE structure and stop/start codons and on the position of the termination site relative to the reinitiation site.[16] This raises questions as to how the new start codon is found after termination.

TURBS RSE-dependent reinitiation requires fewer eIFs than does canonical initiation.[33] While a near-cognate start codon is tolerated, the N terminus of the resultant protein is a methionine, implicating a role for initiator methionyl-tRNA (Met-tRNA$_i$).[12,16] In an

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in vitro reconstituted system, Met-tRNA$_i$ could be delivered by eIF2, eIF2D, or eIF2D's homolog "multiple copies in T cell lymphoma 1" and "density regulated" protein (MCT1/DENR), which all play roles in ribosome recycling.[33–35] In the same system, eIF3 was not essential,[33] contrary to previous claims.[36] However, eIF3 improves reinitiation efficiency, especially in the context of certain mutations or when other factors are absent.[17,33,36]

Reinitiation promoted by TURBS RNAs' uniqueness compared to other RNA structure-dependent strategies of altering translation motivated this investigation. Previous thorough mutational analyses were limited to a few representative members of the TURBS RNA class, raising questions of how widespread and diverse TURBS elements are. Other questions included how these structured RSEs display their recognition sequence to bind to the ribosome and how they remain tethered to the 40S subunit after termination while allowing a new start site to be found either upstream or downstream of the stop codon. Here, we combine computational, biochemical, and structural biology approaches to investigate the TURBS class of RSEs. We provide new insight into previously observed characteristics of the TURBS RNAs, contextualize shared features across the class, and suggest a model for how the RSE remains bound to the ribosome while a new start codon is located.

RESULTS AND DISCUSSION

Bioinformatic analyses identify new putative viral RSEs

To understand the distribution and diversity of the TURBS class of RSEs, we used comparative sequence analysis and homology searches. We aligned sequences from previously studied RSEs from human norovirus, human sapovirus, feline calicivirus (FCV), rabbit hemorrhagic disease virus (RHDV), and influenza B virus (IBV). These represent four genera within the Caliciviridae family (*Norovirus*, *Sapovirus*, *Vesivirus*, and *Lagovirus*) and the *Betainfluenzavirus* genus of Orthomyxoviridae. We aligned UGGGA (motif 1), hereafter referred to as the ribosome complementary sequence (RCS); the annotated downstream AUG start codon; and motif 2/2* base pairing, hereafter referred to as paired element 1 (P1). For the RNAs that have a second paired element (P2), this was also included (Figure 1B). We used the program Infernal[37,38] with this seed file to query a viral genomic database, identifying potential new TURBS RSE RNAs. Some new examples were expected, since TURBS elements have been proposed to exist in all viruses within the *Norovirus*, *Sapovirus*, *Vesivirus*, and *Lagovirus* genera of Caliciviridae.[13] Other putative TURBSs have not been described previously (Figures 1C and 1D; Table S2).

New representatives in the Caliciviridae family include Newbury 1 agent virus (N1V), the sole member of the *Nebovirus* genus, a larger version predicted to contain P2. Smaller versions were identified in chicken calicivirus (ChCV), the sole member of *Bavovirus*, and turkey calicivirus (TuCV) and goose calicivirus of the genus *Nacovirus*. All putative new TURBs in the Caliciviridae family have consistent genomic context, with the AUG nucleotides matching the annotated start codon for the VP2 minor capsid protein. Interestingly, reinitiation appears to be used by most, but not all, viruses in the Caliciviridae family. Members of the *Recovirus*, *Valovirus*, *Minovirus*, and *Salovirus* genera do not appear to contain a TURBS-class RSE. Notably, *Recovirus* and *Valovirus* members contain

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the stop/start needed for reinitiation (UAAUG) but not the required secondary structural features; these are discussed in later sections.

Outside of the Caliciviridae family, there were examples in many members of the *Hapavirus* genus within the Rhabdoviridae family, which are −ssRNA viruses. All examples in *Hapavirus* are of the smaller version of the TURBS RSE. For most of the *Hapavirus* representatives, the putative RSE RNA is in the G mRNA segment, situated near the 3′ end of the first ORF, just upstream of the downstream, out-of-frame U5 gene (Figure 1A). Interestingly, the U5 protein encodes a viroporin,[39] the same protein class as IBV's BM2 protein also expressed by reinitiation. In some *Hapavirus* genomes, the RSE motif lies within the bicistronic N mRNA segment, poised to reinitiate and express the U4 protein, which encodes a small acidic protein of unknown function.[39]

Functional assays confirm activity of new putative RSE RNAs from diverse viruses

To experimentally validate these new putative TURBS sequences, we used a dual-luciferase reporter. This system quantitatively recapitulates reinitiation events in lysate by testing whether putative RSE sequences can promote translation of a downstream ORF. The bicistronic reporter RNAs contain an RSE sequence inserted between *Renilla* luciferase (Rluc) and firefly luciferase (Fluc) ORFs. The negative control RNA has an in-frame stop codon for the Rluc ORF followed by the Firefly ORF, which is in a different frame and lacks an AUG within the first 50 codons. The positive control RNA lacks a stop codon after the Rluc ORF immediately followed by the in-frame Fluc ORF, leading to uninterrupted translation of both luciferase reporters (Figure 2A). For each putative viral RSE, we tested both the wild-type (WT) and a sequence mutated in the RCS (GGG→CCC). Full-length dual-luciferase RNAs were transcribed *in vitro* and added to lysate to measure translation via the luminescence activity of the two luciferase proteins. The ratio of downstream (Fluc) to upstream (Rluc) luminescence for all viral constructs was normalized to that of the positive control RNA, set to 1 (Figure 2B). The negative control RNA showed near-background Fluc luminescence, as expected.

Using this reporter system, we first validated the previously reported RSEs from RHDV, FCV, and IBV. All resulted in signal consistent with reinitiation above the level of the negative control (Figure 2B), matching previous reports of 3%–20% efficiency for these RSEs.[11,14,41] The GGG→CCC mutation, which should prevent binding to ES7 of the rRNA,[11,14,31] resulted in decreased downstream translation from all tested sequences, indicating that the reporter faithfully reports reinitiation promoted by the RNA structure interacting with the ribosome.

We next tested several putative new RSEs, selecting one representative of each new genus in Caliciviridae and both genomic contexts in the *Hapavirus* genus of Rhabdoviridae (Figure 2B). For each, the WT sequence induces downstream translation consistent with reinitiation. Likewise, in each case, the RCS GGG→CCC mutation decreases downstream translation compared to the WT. This decrease in signal of mutation-containing RNAs is not due to different rates of RNA degradation (Figure S1). We note that the level of relative downstream translation comparing mutant constructs across viruses is low but variable.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

This could be due to some background level of reinitiation that could vary between different stop/start contexts or some internal initiation.

As mentioned previously, some Caliciviridae members contain a stop/start context at the VP1/VP2 junction that could indicate a reinitiation site, but the homology search did not identify a TURBS-matching structure. We tested the sequence corresponding to the region upstream and including the stop/start overlap of St-Valérien virus (*Valovirus* genus of Caliciviridae) for reinitiation activity. Downstream translation for this WT viral sequence was only slightly above the negative control. Although this virus contains a UGGGA sequence >100 nt upstream of the stop codon, a GGG→CCC mutation does not impact the already very low downstream translation level, as expected. This result supports the conclusion that other tested examples are true RSEs, as not all viral sequences drive downstream translation. The expression strategy for VP2 in St-Valérien virus remains unclear.

Newly identified RNA structures fold into similar secondary structures

Although the homology searches strongly suggested a common secondary structure used by these newly identified and diverse RSEs, we confirmed this experimentally using chemical probing. We probed examples from RHDV, FCV, and N1V to represent the larger TURBS RSE version with P2 present and examples from ChCV (*Nacovirus*), TuCV (*Bavovirus*), IBV (*Betainfluenzavirus*), Flanders virus (FLAV; *Hapavirus*), and Wongabel virus (WONV; *Hapavirus*) to represent the smaller version of the structure. Chemical probing revealed that all RNAs folded as expected, with low reactivity for nucleotides predicted to be involved in the base pairing of P1 and P2, when present (Figure 2C). In all cases, the loops had overall higher reactivity, including the nucleotides complementary to ES7 of the 18S rRNA. This is consistent with these complementary sequences being unpaired and, thus, available to pair with ES7. The RSE from RHDV has been proposed previously to contain a third paired element (P3),[32] and this is also supported by the chemical reactivity profile. Overall, these chemical probing data support previously proposed secondary structure models used to create the alignments for homology searches and show that the folding pattern of both previously known and newly identified examples is conserved within this class of viral RSE RNAs.

Conservation and variation among reinitiation-stimulating sequences

After verifying the function and secondary structures of the new TURBS RSEs, we evaluated their conservation and variability. We used the sequence alignment to create consensus sequence and secondary structure models (Figure 3A). The P1 and P2 stems are supported by covariation, indicating that specific nucleotide identities are unimportant if base pairing is maintained, with one exception. When aligned from the bottom of P1, the fourth pair is conserved as C-G in all examples. The P3 stem in RHDV only appears to be present in a few sequences, mostly members of *Lagovirus*, and this helix does not display significant covariation due to this limited distribution.

Due to how the alignment was initially constructed, the reinitiation start codon is fully conserved as an AUG and connected to P1 by a linker 12–23 nt in length with no sequence

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conservation, and the stop codon of the upstream ORF is always present near the start codon. In the larger examples, the nucleotide directly 5′ of the P1 stem is absolutely conserved as a cytidine, whereas this nucleotide can have any identity in the smaller examples of the structure. The largest area of primary sequence conservation is the RCS, which is always UGGGA, but the base pairing possible between the viral RNA and rRNA ES7 spans a larger region in some examples. When allowing for G•U wobble pairs, the RCS is always at least 6 nt in length, as the nucleotide preceding the UGGGA is always a purine that can pair with U1120 of the rRNA. At most, the complementary region comprises 10 nt (Figure 3B), potentially pairing to positions 1,112–1,121 of the 18S rRNA (Figure 3A, highlighted in yellow).

Some reinitiation sequences have extended base pairing with rRNA

The observation that some TURBS might make more base pairs with rRNA has been made previously[13,31] but not experimentally investigated. We therefore performed chemical probing of the RSE RNAs from IBV, RHDV, and FCV in the presence of an *in vitro*-transcribed RNA corresponding to 18S rRNA nucleotides 1,104–1,120, representing the apical portion of the ES7 stem loop. Differential reactivity profiles revealed nucleotides that change flexibility in the presence of the rRNA stem loop (Figure 2C and S2). The IBV RSE can potentially form 8 bp with ES7; 7 nt display a decrease in reactivity in the presence of ES7 RNA. The RHDV RSE can potentially form 7 bp with ES7; six exhibit decreased reactivity. The FCV RSE can potentially form 10 bp with ES7; nine exhibit decreased reactivity. Therefore, some TURBS are capable of forming more than 5 bp to rRNA. The ability to form at 6–10 intermolecular base pairs is intriguing because this requires unwinding of multiple base pairs in ES7. Furthermore, the proposed intermolecular base pairing requires that at least 1 nt for the RSE RNA and at least 1 nt of ES7 rRNA are unpaired, possibly to form linkers to relieve geometric constraints.

Constraints on RNA structure, termination site, and reinitiation site positioning provide insights into mechanism

We examined the genomic context of the newly identified TURBS, including the location of the structured element compared to the stop and start codons and the relative locations of these two codons. For each, we ensured that the annotated stop codon was the first in-frame termination codon for the upstream ORF, and recorded its position within the alignment relative to the reinitiation start codon (Figure 1D). The most common stop/start codon contexts were UA**AUG**, UG**AUG**, **AUG**A, and UGAN-N**AUG**, where N represents any nucleotide identity, the stop codon is underlined, and the start codon is shown in bold (Table S2). All three stop codons are found in this class of RSEs, indicating that any type of termination event is compatible with reinitiation.

We used the distance between the termination reinitiation start sites to calculate the “relative reading frame” for each example, setting the stop codon as the zero position (e.g., UA**AUG** is in the +2 frame, and **AUG**A is in the −1 frame). Allowable relative reading frames are from −8 to +6 for TURBS RSEs (Figure 3C) differing from ribosomal frameshifting due to the relationship between the termination and reinitiation sites. That is, upon termination, the stop codon occupies the A site of the decoding groove, whereas the reinitiation start codon

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

must be placed in the P site to pair with Met-tRNA$_i$. Therefore, reinitiation in the +2 frame (e.g., UA**AUG**) requires the mRNA to move by 5 nt within the decoding groove for the start codon to be properly placed. In contrast, reinitiation in the −5 frame (e.g., **AUG**NNUGA) only requires movement by 2 nt. This likely explains why the reinitiation reading frames skew toward the negative, with nearly all examples accessing reading frames between −8 and +4.

Certain reinitiation frames are incompatible with canonical start and stop codons. For example, the −2 frame cannot be used because the third nucleotide of the start codon must be a guanosine, while the first nucleotide of the stop codon must be a uridine. Codons that are incompatible due to nucleotide identities are −2, 0, and +1, while the −3, −6, and −9 relative reading frames would lead to reinitiation events that immediately terminate due to being in frame with the upstream ORF stop codon. The upstream and downstream ORFs cannot be in the same reading frame if the reinitiation start site is upstream of the termination site. However, reinitiation in the same frame is allowed if the reinitiation start codon is downstream of the termination site, such as +3 or +6. Indeed, both arrangements are represented in both members of the *Nacovirus* genus.

A histogram of the number of nucleotides between the closing of P1 and the start or stop codon shows that the range and distribution of linker lengths are similar for both codons, either 12–23 or 12–24 nt for the start and stop codon, respectively (Figure 3D; Table S2). The lower limit is likely imposed by the footprint of the ribosome, being the minimum distance spanned from the decoding groove to the outside of the mRNA exit channel where the RSE interacts. The upper limit is likely imposed by a requirement for the RNA structure to be near its ES7 binding site when the stop codon is in the A site. The importance of this is supported by data showing that reinitiation cannot be achieved if the start codon is too far downstream of the RSE RNA structure.[16] There are likely complex kinetic and thermodynamic constraints enforcing this restriction on the upper distance limit.

Binding site and architecture of a TURBS RSE RNA structure revealed by cryo-EM

To answer questions regarding the molecular basis of RNA structure-induced reinitiation, we determined a structure of a TURBS RSE bound to the ribosome. We incubated purified 80S ribosomes from rabbit with *in vitro*-transcribed RHDV RNA (6,951–7,039) to assemble a complex that was then used for single-particle cryo-electron microscopy (cryo-EM). Following preliminary analysis and classification (Figure S3), extra map features appeared on the 40S subunit in the ES7 region, which we tentatively assigned to the RHDV RNA (Figure 4A). Most of the usable 80S particles had this added feature, suggesting that most were bound to RHDV RNA. Subtracting the 60S signal then masking for the 40S subunit revealed higher-resolution details of the site of interaction (Figure 4B). The extra map feature on ES7 extends well beyond what is accounted for by the apo ribosome structure (Figures 4C and 4D), with two lobes extending away from the 40S surface. Although the overall resolution of the complex was ~3 Å, the local resolution of the map representing the TURBS RSE was considerably lower, similar to peripheral rRNA expansion segments that extend away from the 40S surface (Figure 5A).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Structural model of the viral RNA in complex with a mammalian ribosome

Guided by the chemical probing data, secondary structure model, and proposed base pairing between the RHDV and ES7 (Figure 2C and S1), we determined the overall architecture of the RNA structure and assigned P1 and P2 to lobes within the map. Once the general topology of the RHDV RNA was clear, we built a model of the rRNA-RSE RNA complex into a locally refined map (Figures 4E–4G). We used an existing 40S structure (PDB: 5FLX) to place the base of ES7. The distal portion of ES7 (nucleotides G1110 to C1123) diverged from existing structures, with differences increasing with distance from the ribosome's core. These residues were built manually (depicted in yellow in Figures 4, 5, and 6). The helix of ES7 is disrupted by a bulged nucleotide, C1109, for which local map resolution is poor (Figure 5A).

The intramolecular base pairing within ES7 is maintained through the U1112-G1121 pair. Beyond this, the helix continues as A-form geometry but consists of intermolecular base pairs between the rRNA and RHDV RNA (depicted in lilac in Figures 4, 5, and 6). The map supports the pairing of 6 nt between the internal loop of RHDV and ES7, leaving U6969 unpaired (Figures 4F and 4G), consistent with chemical probing data (Figure S2). The continuous stacking from ES7 into P2 leads to a structure in which the RSE appears as an "extension" of the ES7, pointing away from the ribosome surface.

The ES7-RHDV RNA base pairing is well defined in the map, and the location and orientation of the P1 and P2 helices within the two lobes is clear due to the strand directionality necessary for base pairing. However, the local resolution of the map in other regions was insufficient to unambiguously build all portions of the structure (Figure 5A). Therefore, we built P1 and P2 by placing A-form helices at an angle best matching the map and refined their location and connectivity to satisfy proper backbone geometry. The precise location of the two unpaired nucleotides of ES7 rRNA (U1119 and A1120) and U6969 of the RHDV RNA are also ambiguous given the local resolution in those regions. These nucleotides (depicted in gray) were thus modeled and refined within the map at favorable backbone angles. Consistent with chemical probing data, the lower local resolution of the map suggests that these are flexible regions of the RNA (Figures 2C and S2).

The maps did not allow us to unambiguously build other parts of the RHDV RNA without overspeculation. Specifically, nucleotides 6,982–7,007, which form a large internal loop, and the P3 stem were not visible, likely due to substantial conformational heterogeneity; this is supported by high chemical reactivity of the loop nucleotides (Figure 2C and S2). Although some additional map features are present in the area adjacent to P1 beyond the 8 nt comprising the helix, it is poorly resolved (>8 Å), and it is unclear which nucleotides might occupy this space. Therefore, we did not model any nucleotides near the 5′ end of P1. Finally, none of the nucleotides upstream or downstream of P1 were built into the model, again due to this ambiguity.

The orientation of the RHDV RSE on the ribosome aligns with the decoding groove

The RHDV RSE on ES7 binds adjacent to the decoding groove where mRNA exits after being translated. Presumably, the RSE RNA passes through the ribosome and then refolds

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and interacts with ES7 during or after termination with a stop codon in the A site. The structure of bound RHDV RSE is consistent with this, as the P1 stem is positioned with its 3′ end pointing at the decoding groove. Notably, there are no map features in the decoding groove corresponding to RNA between the P1 3′ end and the start codon, although this RNA was present. This could be because the sample was generated using purified 80S ribosome and RNA only or for functionally significant reasons (see the discussion in a later section). Nonetheless, the position and orientation of the RHDV RSE on the 40S explains how it binds the subunit and connects to the stop and start codons in the decoding groove.

Proposed identity and interactions for ambiguous regions of the cryo-EM reconstruction

There were features in all cryo-EM maps adjacent to P1 that could not be accounted for by the nucleotides comprising that stem. While we did not build into these features, the flanking nucleotides on the 5′ end of P1 are likely candidates for occupying this space. Among larger TURBS RSE RNAs, the nucleotide immediately preceding P1 (C6963 of RHDV) is always a cytidine (Figure 3A). Systematic truncations of the upstream sequences for both IBV and RHDV indicate that nucleotides 5′ of the P1 stem are necessary for full reinitiation efficiency.[11,14] For IBV, 2 nt upstream of P1 display low chemical probing reactivity in both ES7-bound and unbound forms (Figure 2C and S2), but mutation of the nucleotide immediately upstream of P1 (C727) to an adenosine causes both 5′ nucleotides to display high reactivity, while base pairing of P1 remains intact (Figure S4). These observations suggest that the C immediately upstream of P1 is involved in tertiary contacts, which could contribute to these mysterious map features.

The structural basis for conservation of the fourth base pair of P1 as a C-G pair (Figure 3A) remains unexplained. Mutations of the P1 stem of RHDV so that base pairing is maintained but with different nucleotide identities causes a >80% reduction in reinitiation activity compared to the WT, which was attributed to misfolding into an alternate secondary structure.[32] We propose that this specific C-G is needed for tertiary contacts, perhaps related to conserved C6964. These interactions might position RSE elements correctly in 3D space or could prevent misfolding.

Dynamics in the RSE-ribosome complex allow movement relative to the decoding groove

The position and orientation of the RHDV RSE RNA on the ribosome and analysis of its conformational heterogeneity reveal how it could stay tethered to the ribosome while a new start codon is found. 3D variability analysis and 3D classification show that the viral RNA can occupy a range of positions while bound to the rRNA (Figure 5B). These positions are not fully discrete states, as increasing the number of 3D classes leads to more states populated along the axis of what we infer represents dynamic motions. The small number of particles classified into each minor state significantly limits resolution, resulting in increasingly poor map quality as the number of classes increases (Figure 5A). We speculate that the nucleotide positions that enable these dynamics are the unpaired, flexible residues that have low local resolution and whose positions could not be definitively placed. In particular, the bulged C1109 of ES7 appears to be poised to affect the trajectory of the helix by introducing a kink in the backbone (Figure 5A). Other nucleotides that are more

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

peripheral are U6969 of RHDV, which would specifically allow flexibility in the position of P1, as well as U1119 and A1120 of ES7.

Examining the nature of the dynamics more closely, it appears that the extended helical stack of rRNA, then the viral RNA-rRNA hybrid, then viral RNA (P2) can maintain base pairs but sample different local conformations. These propagate to large differences in the location of peripheral regions of the structure. The two extreme states identified by variability analysis demonstrate a 20° difference in position of the RSE RNA relative to the ribosome structure (Figure 5B). That is, the RHDV RNA remains bound but "rocks" back and forth. This leads to a 13 Å difference in the location of the 3′ end of P1 between the two states. We cannot exclude the possibility that even more widely separated states exist but are populated very rarely.

We propose that the conformational heterogeneity reflects important dynamics for reinitiation function, possibly explaining how the RSE RNA stays bound while a reinitiation start codon is found upstream or downstream of the stop codon. The inferred motion changes the distance between the RSE structure and the mRNA exit channel (Figures 5C and 5D). Specifically, movement of the P1 structure relative to the ribosome could allow movement of the downstream RNA in the decoding groove so that it could reposition relative to the ribosome (Figure 6), allowing different codons to be sampled in the P site after termination.

RSE interactions with proteins are also possible. Specifically, in some positions of the bound RSE, the map is continuous between the RHDV RNA and ribosomal protein regions. "State 1" (Figure 5B) puts nucleotides of RHDV within the 5′ portion of the RCS near residues 199–203 of ribosomal protein S1, while "state 2" puts nucleotides in the 3′ portion of the RCS and rRNA near residues 72–75 of ribosomal protein S27. In the modeled architecture, these proteins would likely interact with the RNA backbone and not rely on specific nucleotide identities, but the details are ambiguous due to the local resolution.

eIF3 binding is compatible with a viral RSE-ribosome complex

While eIF3 is not necessary for reinitiation *in vitro*, it increases efficiency and leads to an increase in fidelity of authentic start codon selection in concert with eIF1.[33] Furthermore, eIF3 complex components crosslink with an RSE RNA.[36] To gain insight into how eIF3 and RSE binding could relate spatially, we modeled an existing structure of eIF3 bound to the 40S subunit[42] onto our map and model of the RHDV RSE-80S complex (Figure S6A). The bound RHDV RNA is positioned so that it should not block any direct interactions between eIF3 and ribosomal proteins, as proposed previously.[33]

In our model, the subunits closest to the RSE structure are eIF3a and eIF3c, which bind ribosomal proteins S1 and S27, respectively, but do not interact directly with ES7.[42] The feature in our cryo-EM reconstruction corresponding to P2 is nestled against the bundle of α helices of eIF3c (Figure S6A), and while they are in proximity, there do not appear to be substantial steric clashes. However, the lobe corresponding to P1 clashes with some of the α helices corresponding to the N-terminal portion of eIF3a. The model therefore predicts that the bound RHDV RSE would sterically hinder binding of eIF3. However,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

several possibilities could alleviate these clashes. First, most of the map features that overlap eIF3a are adjacent to P1; this is the region possibly occupied by flanking 5′ nucleotides, which could interact with eIF3. Second, the steric clashes could also be overcome by structural rearrangements of eIF3a, the RSE RNA structure, or both. If eIF3 can bind the ribosome while the RSE RNA is present, then this differs compared to HCV-like IRESs, which directly bind eIF3[43,44] but reposition many of its subunits (including eIF3a and eIF3c) away from their canonical binding site on the 40S.[45] Importantly, the identity and nature of any RSE-eIF3 interactions and their effects remain speculative, but our model suggests ideas that can be tested by future experiments.

Could eIF3 binding influence reinitiation start site selection?

In our speculative model, the position of eIF3 in relation to RSE RNA and the motion of RSE on the ribosome suggest a mechanism by which this initiation factor could facilitate placement of the new start codon. Specifically, the regions of the RSE map that clash with eIF3 are different when comparing the position of the two extreme states revealed by 3D variability analysis (Figure S6B). This raises the intriguing possibility that binding of eIF3 post termination stabilizes a specific conformation of the RSE, and, in so doing, positions P1 so that downstream mRNA in the decoding groove is in the proper location with the start codon in the P site to enable reinitiation.

Our speculative conformational fit model for eIF3's role in this type of reinitiation contextualizes previously reported data and might alleviate seemingly conflicting reports. It was found that eIF3 is pulled down by the RSE RNA from FCV similar to an HCV-like IRES, proposed to indicate a direct binding event.[36] However, in this same context, a construct carrying a mutation of the RCS was unable to pull down eIF3, strongly suggesting that 40S subunit binding is necessary, in contrast to the proposed direct binding model.[36] If the RSE RNA and eIF3 interact directly, then it would likely be in other regions of the viral RNA, and the affinity of the viral RNA for eIF3 might be too weak to measure via crosslinking unless both are bound to the 40S subunit. In a reconstituted system, eIF3 was not necessary to assemble an initiation complex at the reinitiation start site, but its addition increases the level of reinitiation.[33] These data are consistent with the role proposed by us in stabilizing a reinitiation-competent conformation and influencing start codon placement. Ultimately, additional experiments or structures will be necessary to elucidate molecular interactions, binding, and any influence on RSE RNA position by eIF3.

While this structure provides some insight into the possible role of eIF3, it does not inform on the roles of other translation machinery in inducing reinitiation. While Met-tRNA$_i$ is a necessary component for this type of reinitiation, it can be successfully delivered by multiple factors,[33] and the influence of different conformational states of the RSE RNA on its arrival is unclear (Figure 6).

Model for finding a reinitiation start site within a limited window

We hypothesize that the conformational dynamics of the ribosome-bound RHDV RNA are a shared feature among this entire class of viral RSEs. However, the extent of the conformational heterogeneity and space sampled by each TURBS RSE RNA might differ,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

allowing different mRNA movements within the decoding groove to place the reinitiation start codon. Differential dynamics among members of this class might be caused by variable lengths of flexible linkers connecting paired elements (Figure 3A), different numbers of base pairs between the RCS and ES7 rRNA, and the absence of the P2 stem. These same factors likely play into the overall measured reinitiation efficiency (Figure 2B).[11,14,16] In addition, the thermodynamic stability of the RSE RNA structure, its folding kinetics, and the kinetics of termination at different stop codon identities likely affect observed differences. Future studies that query a more diverse set of structures should enable firm conclusions as to the relative contributions of each of these numerous factors.

The distribution of distances between P1 and the start codon is subject to the same constraints as that between P1 and the stop codon. This suggests that mRNA in the decoding groove can move within a limited range. The upper limit on the start codon location implies that the reinitiating complex does not use 5′-to-3′ scanning, consistent with previous observations.[33] The overall range of distances between P1 and the start or stop codon of 12–24 nt (Figure 3D) likely reflects (1) the footprint of the ribosome on the lower end and (2) a maximal distance beyond which the RCS of the nascently translated and refolded RSE RNA can no longer find its binding site on ES7 on the upper end. It is then noteworthy that the allowable relative reading frames have a stricter upper limit of 9 nt (for reinitiation in the +6 frame), with most examples only requiring a shift spanning the distance of five or fewer nucleotides in the decoding groove (−8 to +2 frame) (Figure 3C). The limit on allowable relative reading frames likely reflects the maximal dynamics of the bound RSE structure that allows sampling of different nucleotides in the P site. Overall, bioinformatic and cryo-EM data support our proposed model for reinitiation start site selection within a limited number of nucleotides in either direction enabled by conformational dynamics (Figure 6). This model warrants future exploration to uncover the mechanistic details underlying this type of reinitiation.

Viral RSEs and IRESs share a binding site but differ in their mechanism

The TURBS RSEs are not the only viral RNA structures that interact with the ES7 portion of rRNA to manipulate translation. The hepatitis C virus (HCV)-like IRESs contain a stem loop that also base pairs with the loop of ES7. Specifically, C1116–C1118 of the ES7 loop base pair with three consecutive guanosines in the loop of HCV IRES domain IIId, while U1115 pairs with a bulged nucleotide in a different portion of IRES domain III[43,46,47] (Figure S5). The other 2 nt of the ES7 loop form hydrogen bonds and stacking interactions.[46–48] Our work shows that binding of the RHDV RSE to rRNA unwinds an additional base pair compared with HCV IRES to rRNA (Figure S5), resulting in two consecutive unpaired ES7 nucleotides. This may lend additional flexibility when bound to a TURBS RSE versus the HCV IRES.

Despite their shared binding location on the ribosome and ability to induce a translation event independent of certain eIFs, there are significant differences between TURBS RSEs and HCV-like IRESs in both structure and function. The HCV-like class of IRESs comprises a larger, more complex multi-domain structure,[49,50] interacting with other regions of the ribosome. In particular, domain II of the HCV IRES reaches into the decoding groove of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the 40S, interacting with the E site to help position the start codon in the P site.[46,48,51] It is thus unsurprising that cryo-EM reconstructions of the HCV IRES bound to the ribosome show features within the decoding groove.[48] In contrast, our maps of the RHDV RSE bound to the ribosome do not show any features in the mRNA channel (Figure S7). While this could be due to construct design, technical limitations, or sample preparation methods, this could reflect an important mechanistic distinction. That is, reinitiation requires a ribosome that first translates the upstream ORF, since the RSE is unable to "*de novo*" place mRNA in the decoding groove on its own.[12,16,30] Therefore, the RSE RNA structure can only initiate translation after termination but before recycling, when the mRNA is already properly positioned within the 40S.

Previously, it has been proposed that the RSEs do not function as IRESs in part because the ribosome must translate through the upstream structure and allow it to refold before proper RSE RNA-rRNA interactions can form.[12,14] While the kinetics of co-translational refolding could certainly impact the efficiency or frequency of reinitiation, we argue that this is likely not the major factor in distinguishing between reinitiation by a TURBS RSEs and HCV-like IRESs. One argument against the post-translation refolding model is that, in both chemical probing experiments and in our cryo-EM maps, the viral RNA (including sequences flanking the structured element) bind the ribosome without a specific type of refolding event. We favor the model of RSEs lacking the ability to efficiently place mRNA in the decoding groove as a distinguishing characteristic compared to an IRES.

Limitations of this study

A limitation of this study is that the cryo-EM reconstructions contain only RSE RNA bound to 80S ribosomes and, thus, are not complexes "captured' during reinitiation or with bound factors. Related to this, the local map resolution corresponding to the RSE was too low to allow unambiguous modeling of all nucleotides, and some parts were not visible. While the placement and dynamics of the RSE were readily observed and enable new hypotheses to be made, elucidation of additional high-resolution details and any interactions with initiation factors await additional structural studies. Our model of the potential role of dynamics and eIF3 remains to be tested. Also, all measurements of reinitiation efficiency were done in cell-free lysate with reporter RNAs. This is a valid way to qualitatively assess function, but viral infection, the cellular context, or different cell types could alter RSE efficiency and are topics for future focused studies on individual RSEs.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Jeffrey Kieft (jkieft@nysbc.org).

Materials availability

Plasmids generated in this study will be made available on request.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Data and code availability

- The cryo-EM map for RHDV reinitiation stimulating TURBS RNA bound to the rabbit ribosome has been deposited in the Electron Microscopy Data Bank under accession code EMD-45307. Coordinates for the resultant model have been deposited in the Protein Data Bank (PDB) under accession number 9C8K.
- This study does not report any custom code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

METHOD DETAILS

Bioinformatics—Sequences of five previously reported examples of reinitiation stimulating RNAs in the TURBS class from were retrieved from the National Center for Biotechnology Information (NCBI) sequence database (genome accession numbers: NC_039476.1, NC_002210.1, NC_001481.2, NC_001543.1, NC_010624.1). The sequences were manually aligned in Stockholm format based on the following four features: motif 1 (‘UGGGA’), motif 2 (5′ portion of the P1 base pairing), motif 2* (3′ portion of the P1 base pairing) and the annotated start codon for the downstream ORF (‘AUG’). This seed alignment file was used to query the RefSeq viral database,[52] downloaded on 01/19/2021, containing one representative sequence for each virus using the program Infernal version 1.1,[38] allowing for an E-value cutoff of 1 (instead of the default 0.05) to find potentially enable identifying sequences with more variation. Iterative searches followed by manual adjustments to the alignment were performed until no new sequences were identified. In total, sequences from 44 total unique viruses were returned that match the TURBS class of RSE RNAs (Table S2, Supplemental File 1). The alignment was subsequently separated into two files based on the presence or absence of the P2 stem, which was added manually to the alignment for the larger versions of the structure. The two consensus models were calculated using R-scape,[53] visualized with R2R,[54] and adjusted and labeled in Adobe Illustrator (Figure 3A). Various features of the structure (e.g., linker lengths, base pairing potential) were calculated manually and histograms were plotted and labeled in Adobe Illustrator.

Plasmid construction and RNA preparation of dual luciferase reporter constructs—Dual luciferase constructs containing intervening viral RNAs of interest were made using the pSGDLucV3.0 vector (Addgene 119760). Insertion sequences were ordered as double-stranded DNA gene fragments (IDT) then amplified via PCR. Vector and inserts were prepared by digesting with BglII (NEB) in NEBuffer r3.1 (NEB) at 37°C for 2 h followed by digestion with PspXI (NEB) in rCutsmart Buffer (NEB) and incubated at 37°C for an additional 2 h. The vector was purified via agarose gel purification while inserts were purified directly from the restriction digest reaction, both using the Wizard SV Gel and PCR Clean-up System (Promega). Inserts were incorporated into the vector via ligation reaction consisting of inserts and vector in a 5:1 M ratio, T4 ligase (NEB), and T4 ligation buffer

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(NEB). The reaction was incubated at 16°C overnight then transformed into DH5α *E. coli* and screened for successful ligation via colony PCR. Sequences were verified via whole plasmid sequencing prior to use. The positive control (in-frame readthrough) and each viral mutant (GGG→CCC) plasmid were constructed using the Q5 site-directed mutagenesis kit (NEB) according to manufacturer protocol and verified via whole plasmid sequencing. Double-stranded DNAs approximately 5 kb in length, corresponding to both luciferase genes separated by control or viral sequence, were produced via PCR as a template and their size verified by 1.5% agarose gel electrophoresis. These PCR products were then used as templates for *in vitro* transcription with the MEGAscript T7 transcription kit (Invitrogen). RNA was purified using the Monarch RNA clean-up kit (NEB) and RNA length and quality was verified via 10% acrylamide gel.

Translation assays using dual luciferase reporters in lysate—Each reaction contained 500 ng of RNA, which was initially heat re-folded by incubating in 33mM HEPES (pH 7.5) at 95°C for 1 min then cooling to room temperature for 5 min at which time $MgCl_2$ was added to a final concentration of 10 mM. The final 50 μL translation reaction included the 5 μL RNA mixture described above and components from the Rabbit Reticulocyte Lysate system (Promega): 35 μL of rabbit reticulocyte lysate, amino acid mixture minus cysteine (final concentration 16.7 μM), amino acid mixture minus methionine (final concentration 16.7 μM), amino acid mixture minus leucine (final concentration 16.7 μM), and KOAc (final concentration 150 mM). Reactions were incubated at 30°C for 4 h. After incubation, reactions were diluted 4X with Passive Lysis Buffer (Promega) and the reaction was split into two technical replicates, each with a final volume of 100 μL. Luciferase activity was measured with a GloMax-Multi Detection system using the Dual-Glo Luciferase Assay System (Promega). Results were analyzed using Excel (Microsoft) software with measurements from at least three independent experiments, using the average of the two technical replicates from each trial. These measured values were used to calculate the ratio of Firefly (downstream) to *Renilla* (upstream) translation (Fluc/Rluc), then these ratios were normalized with the positive control (in-frame readthrough) set to 1. Data were plotted and labeled using Adobe Illustrator.

Reverse transcription-quantitative PCR (RT-qPCR)—Following translation assays in lysate, dual luciferase-containing RNA was isolated by adding Acid-Phenol:Chloroform pH 4.5 (Invitrogen) then removal of the aqueous layer. cDNA was generated by reverse transcription in a 20 μL reaction containing 12 μL of the RNA-containing aqueous layer, 4 μL 5X First strand buffer (Invitrogen), 1 μL 10 mM dNTPs, 1 μL 100mM DTT, 1 μL 2 μM DNA reverse primer, and 1 μL Superscript III enzyme (Invitrogen), which was incubated for 1 h at 50°C followed by heat inactivation for 10 min at 75°C. qPCR was performed in 12 μL reactions containing 5.5 μL of the RT reaction, 0.25 μL of 10 μM forward PCR primer, 0.25 μL of 10 μM reverse PCR primer, 6 μL of 2X PerfeCTa SYBR Green FastMix (Quantabio) and analyzed using a QuantStudio 7 Flex Real-Time PCR System (Thermo Fisher Scientific). Data were initially processed using QuantStudio Design and Analysis v2 (Thermo Fisher Cloud) to calculate cycle thresholds (Ct), which were plotted using GraphPad Prism 10.2.0 then labeled using Adobe Illustrator.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RNA preparation for chemical probing and cryoEM—Template DNA was acquired as gBlock DNA fragments (IDT) and subsequently cloning into the pUC19 vector. Double-stranded DNA (dsDNA) templates, featuring an upstream T7 promoter, were amplified in PCR reactions under the following conditions: 100 ng plasmid DNA, 0.5 μM forward and reverse DNA primers (see Table S1), 500 μM dNTPs, 25 mM TAPS-HCl (pH 9.3), 50 mM KCl, 2 mM $MgCl_2$,1 mM β-mercaptoethanol, and Phusion DNA polymerase (New England BioLabs). Successful amplification of dsDNA was confirmed by 1.5% agarose gel electrophoresis. Subsequently, RNA transcription reactions were carried out using the following conditions: approximately 0.1 μM template DNA, 10 mM NTPs, 75 mM $MgCl_2$, 30 mM Tris-HCl (pH 8.0), 10 mM DTT, 0.1% spermidine, 0.1% Triton X-100, and T7 RNA polymerase. These reactions were incubated at 37°C overnight. Following transcription, insoluble inorganic pyrophosphate was removed through centrifugation at 5000×g for 5 min, after which the RNA-containing supernatant was ethanol precipitated using 3 volumes of 100% ethanol at −80°C for a minimum of 1 h. The mixture was centrifuged at 21000×g for 30 min at 4°C to pellet the RNA, and the ethanolic fraction was carefully decanted. The RNA was resuspended in 9 M urea loading buffer and then purified via denaturing 10% polyacrylamide gel electrophoresis (PAGE). RNA bands were detected via UV shadowing and subsequently excised. These bands were then extracted by crush soaking in diethylpyrocarbonate-treated (DEPC) Milli-Q water at 4°C overnight. The RNA-containing supernatant was concentrated using spin concentrators (Amicon) to achieve the desired RNA concentration in DEPC-treated water. Finally, the RNAs were stored at −80°C, with working stocks stored at −20°C.

***In vitro* chemical probing of RNAs—**RNA structure probing experiments using the selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) reagent NMIA were conducted in accordance with established protocols.[55] Briefly, 240 μM RNA was refolded by heating to 90°C for 5 min, followed by cooling to ambient temperature and incubation with $MgCl_2$ for 20 min. The refolded RNA was incubated with NMIA for 15 min at ambient temperature, with the following modification conditions: 120 nM RNA, 6 mg/mL NMIA or DMSO, 50 mM HEPES-KOH (pH 8.0), 10 mM $MgCl_2$, and 3 nM 6-fluorescein amidite 5′-labeled FAM-RT primer (see Table S1). To quench the modification reaction, NaCl was added to a final concentration of 500 mM, and Na-MES buffer (pH 6.0) was adjusted to 50 mM. Modified RNA was recovered using oligo-dT magnetic beads (Invitrogen Poly(A)Purist MAG Kit). After recovery, the RNAs were washed twice with 70% ethanol and resuspended in water. cDNA was synthesized using SuperScript III (Invitrogen) at 48°C for 1 h, following the manufacturer's instructions. The RNA was subsequently degraded by the addition of NaOH to a final concentration of 200 mM and heating to 90°C for 5 min. Hydrolysis was stopped using an acid-quench solution containing 250 mM NaOAc (pH 5.2), 250 mM HCl, and 500 mM NaCl. DNA was then recovered using magnetic beads, washed twice with 70% ethanol, and eluted in GeneScan 350 ROX Dye Size Standard (ThermoFisher) containing HiDi formamide solution (ThermoFisher). The resulting 5′-FAM-labeled reverse-strand DNA products were analyzed via capillary electrophoresis using an Applied Biosystems 3500 XL instrument. Data analysis to calculate reactivity values at each nucleotide position was performed using HiTrace RiboKit (https://ribokit.github.io/HiTRACE/)[56–59] in MATLAB (MathWorks). For differential chemical

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

probing, normalized reactivities values in the presence of ligand were subtracted from normalized reactivities values in the absence of ligand. Figures were generated using RiboPaint (https://ribokit.github.io/RiboPaint/) in MATLAB then subsequently labeled in Adobe Illustrator. SHAPE reactivity data was mapped onto the predicted secondary structure model based on the alignment.

Ribosome purification from lysate—Nuclease-treated bulk rabbit reticulocyte lysate (Green Hectares) was spun down at 20,000 RPM for 15 min to remove debris, nuclei, and mitochondria.[60] Clarified lysate was filtered with a 0.22 μm filter (Millipore). The supernatant was loaded on to a 30% sucrose cushion (20 mM Tris-HCl pH 7.5, 2 mM $MgOAc_2$, 150 mM KCl, 30% w/v sucrose) and ultracentrifuged for 17.5 h at 36,000 RPM (50.2 Ti rotor) at 4°C to obtain a ribosomal pellet.[61] The pellet was washed and resuspended in a buffer containing 20 mM Tris-HCl pH 7.5, 6 mM $MgOAc_2$, 150 mM KCl, 6.8% w/v sucrose, 1 mM DTT, 1 μL RNasin from Promega (Cat N2618). To remove non-resuspended particles, the resuspended pellet was centrifuged again at 10,000 g for 10 min at 4°C and the supernatant was isolated. 15–30% sucrose gradients were prepared using a buffer containing 20 mM Tris-HCl pH 7.5, 2 mM $MgOAc_2$, 150 mM KCl, and either 15% or 30% w/v sucrose using the Gradient Master (Biocomp). Gradients were cooled to 4°C before use. Pellet supernatant was loaded onto gradients and ultracentrifuged at 19,100 RPM for 17.5 h (SW-28 rotor) at 4°C. Gradients were fractionated using the Piston Gradient Fractionator and Fractionator Software v8.04 (Biocomp), monitoring for absorption at 260 nm. Fractions corresponding to pure 80S ribosomes were pooled and concentrated to an A_{260} of 95 using an Ultra-15 centrifugal filter unit with a nominal molecular weight limit of 100 kDa (Amicon). This concentrated stock was subsequently diluted to 250 nM using buffer containing 20 mM Tris-HCl pH 7.5, 2 mM $MgOAc_2$, 150 mM KCl and 20 μL aliquots were flash frozen in liquid nitrogen and stored at −80°C.

cryoEM sample preparation and image acquisition—RHDV RNA at 20 μM was heat denatured at 95°C and allowed to cool slowly to ambient temperature, then 1 μL of the RNA solution was added to a 20 μL aliquot of 250 nM 80S (described above) and incubated at 37°C for 15 min. Using a FEI Vitrobot mark IV, a 3 μL aliquot of the RHDV-80S solution was applied to grids (holey carbon grids C-flat 1.2/1.3 400 mesh VWR) that had undergone plasma cleaning for 6 s in a mixture of O_2 and H_2 by a Gatan Solarus 950 (Gatan, Inc). Grids were blotted for 3.5–5 s at a force of −5, then plunge frozen in liquid ethane.

Movies were collected using a Thermo Scientific 300 kV Krios G3i microscope equipped with a Falcon 4 Direct Detection Camera and Selectris Imaging Filter. A total of 6,589 micrographs were collected (429 at 30° stage tilt to overcome preferred orientations). Each movie consisted of 267 frames collected over 7.76 s with a total dose of 49.94 $e^-/Å^2$. EPU (Thermo Fischer) was used to monitor data collection and specify defocus values in the range of −0.4 to −1.8 μM. Data were collected at 130,000× magnification with a physical pixel size of 0.97 Å and filter slit width of 10 eV.

cryoEM image processing and map generation—Cryo-EM data were processed using cryoSPARC 4.0.[62] An overview of the processing workflow can be found in Figure S3. Movies were imported then processed (patch motion correction, patch CTF estimation)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

using default parameters. The 80S portion of an existing structure (PDB: 4UJD[63]; was used to generate templates for particle picking. Particles were extracted from micrographs with a box size of 800 × 800 pixels, then downsampled 4X (200 × 200 pixel) to speed up initial classification and refinements. Six initial rounds of 2D classification were performed to remove "junk" particles (e.g., incorrectly picked particles, ice contamination, aggregates). Particles from all good classes were used for *ab initio* 3D reconstruction requesting two classes followed by heterogeneous refinement resulting in one class with good 80S density and another demonstrating good 60S density but poor 40S density and those particles were removed. Four additional rounds of 2D classification were performed to remove particles belonging to classes with poor resolution and/or alignment.

All resulting good particles were subjected to *ab initio* 3D reconstruction and non-uniform refinement,[64] resulting in a map with excellent 60S resolution and relatively poor 40S resolution, as expected due to known dynamics of the small subunit. Nevertheless, density was observed in the 40S in the region of the 40S near the mRNA exit channel beyond what was accounted for by the ribosome and in a similar position compared to IRES structures that bind directly to ES7.[63] These particle locations underwent local motion correction[65] then were re-extracted without downsampling. Masks were created for the 60S and 40S subunit, the 60S signal was subtracted, and the masked 40S underwent non-uniform refinement. A mask was generated for the local region surrounding the RHDV RNA ("RHDV mask") using UCSF Chimera.[66] Focused 3D classification was performed with the RHDV mask using varying numbers of classes. Ultimately six classes resulted in four classes with "good" density in the region corresponding to RHDV RNA and two classes with poor density, which were excluded.

The remaining good 61,643 particles underwent non-uniform refinement, first with the 40S mask then subsequent non-uniform refinement with the RHDV mask. Resolutions for each of these refinements (Figure S3) were estimated using the gold standard Fourier shell correlation (GSFSC) of 0.143. The RHDV mask was also used to perform masked 3D variability analysis[67] for all good particles with a resolution threshold of 8 Å. Initially, multiple modes of variability and resolution thresholds were requested, and one particular mode was always observed, as reported in Figure 5B.

Model building and refinement—The 40S portion of a previously-solved structure (PDB: 5FLX)[47] was fit into the density of the map resulting from the non-uniform "40S mask" refinement using ChimeraX.[68] All regions of our map agreed with the previously solved model except the 18S rRNA nucleotides G1110-C1123 (i.e., the distal portion of ES7). These nucleotides along with the well-resolved portions of RHDV RNA were built into the density of the "RHDV mask" locally refined map. First, base-paired regions were modeled as A-form helices, starting with 18S rRNA base pairing between 1110–1112 and 1121–1123, then the intramolecular pairing between 18S rRNA 1113–1118 and RHDV 6970–6975, then P1 and P2 of RHDV, which were all sequentially fit into the density using ChimeraX. Backbone connectivity and refinement as well as the placement of nucleotides U6969 of RHDV and 1119–1120 of 18S rRNA were initially performed using COOT.[69] The ES7-RHDV RNA was refined against a locally refined map and then this refined structure was added to the rest of the 40S and refined against the 40S map in Phenix.[70] Modeling

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of the structure into the two extreme states of 3D variability analysis (Figure 5B) was accomplished by fitting only this portion of our structural model (RHDV RNA, 18S rRNA 1110–1123) into each state using ChimeraX. The distance separating these two states was calculated in ChimeraX by measuring the distance between the N1 of G6964 of RHDV modeled into each state. The angle measured between the two states was calculated by ChimeraX using the position of G6981 (top of P1) in each state with G1121 of the 18S rRNA serving as an anchor point.

Modeling of eIF3 was accomplished by fitting the 40S portion of a previously determined initiation complex structure into our density and visualizing just the eIF3 complex (PDB: 6ZMW).[42] Modeling of the HCV IRES-ribosome interactions compared to RHSV RSE-ribosome structural model was accomplished by fitting the 40S portion of an existing structure (PDB: 5A2Q)[46] into our map then adding back the IRES RNA. Figures of cryoEM maps and structural models were made using ChimeraX, with the exception of Figure 4 panel F and Figure 5 panel C which were created using PyMol (Schrödinger, LLC) and labeled using Adobe Illustrator.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using Microsoft Excel for translation assays and GraphPad Prism for qPCR. Details on quantification, statistical analyses, and data representation can be found in the legend to Figure 2B and S1 as well as the Method Details section under Translation assays using dual luciferase reporters in lysate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Salimah Thaxton, Nadine Ramirez, and Andrea MacFadden for technical assistance with cloning and additional members of the Kieft laboratory for helpful discussions. We thank Charles Moe at the University of Colorado Boulder Krios Electron Microscopy Center for assistance with cryo-EM data collection. pSGDlucV3.0 was a gift from John Atkins (Addgene plasmid 119760). This work was supported by National Institutes of Health (NIH) grant R35GM118070 (to J.S.K.). M.E.S. was supported by a Jane Coffin Childs post-doctoral fellowship. K.E.S. was supported by NIH grant T32GM136444.

REFERENCES

1. Aitken CE, and Lorsch JR (2012). A mechanistic overview of translation initiation in eukaryotes. Nat. Struct. Mol. Biol. 19, 568–576. [PubMed: 22664984]
2. Hinnebusch AG, and Lorsch JR (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb. Perspect. Biol. 4, a011544. [PubMed: 22815232]
3. Dever TE, and Green R (2012). The elongation, termination, and recycling phases of translation in eukaryotes. Cold Spring Harb. Perspect. Biol. 4, a013706. 10.1101/cshperspect.a013706. [PubMed: 22751155]
4. Hellen CUT (2018). Translation termination and ribosome recycling in eukaryotes. Cold Spring Harb. Perspect. Biol. 10, a032656. [PubMed: 29735640]
5. Young DJ, and Guydosh NR (2022). Rebirth of the translational machinery: The importance of recycling ribosomes. Bioessays 44, 2100269.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

6. Gunisova S, Hronova V, Mohammad MP, Hinnebusch AG, and Valasek LS (2018). Please do not recycle! Translation reinitiation in microbes and higher eukaryotes. FEMS Microbiol. Rev. 42, 165–192. 10.1093/femsre/fux059. [PubMed: 29281028]
7. Sherlock ME, Baquero Galvis L, Vicens Q, Kieft JS, and Jagannathan S (2023). Principles, mechanisms, and biological implications of translation termination–reinitiation. RNA 29, 865–884. [PubMed: 37024263]
8. Skabkin MA, Skabkina OV, Hellen CUT, and Pestova TV (2013). Reinitiation and other unconventional posttermination events during eukaryotic translation. Mol. Cell 51, 249–264. 10.1016/j.molcel.2013.05.026. [PubMed: 23810859]
9. Ahmadian G, Randhawa JS, and Easton AJ (2000). Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism. The EMBO journal 19, 2681–2689. [PubMed: 10835365]
10. Guo L. h., Sun L, Chiba S, Araki H, and Suzuki N (2009). Coupled termination/reinitiation for translation of the downstream open reading frame B of the prototypic hypovirus CHV1-EP713. Nucleic Acids Res. 37, 3645–3659. [PubMed: 19364811]
11. Meyers G (2003). Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependent reinitiation mechanism. J. Biol. Chem. 278, 34051–34060. [PubMed: 12824160]
12. Napthine S, Lever RA, Powell ML, Jackson RJ, Brown TDK, and Brierley I (2009). Expression of the VP2 protein of murine norovirus by a translation termination-reinitiation strategy. PLoS One 4, e8390. [PubMed: 20027307]
13. Powell ML (2010). Translational termination-reinitiation in RNA viruses. Biochem. Soc. Trans. 38, 1558–1564. 10.1042/BST0381558. [PubMed: 21118126]
14. Powell ML, Napthine S, Jackson RJ, Brierley I, and Brown TDK (2008). Characterization of the termination–reinitiation strategy employed in the expression of influenza B virus BM2 protein. Rna 14, 2394–2406. [PubMed: 18824510]
15. Powell ML, Brown TDK, and Brierley I (2008). Translational termination–re-initiation in viral systems. Biochem. Soc. Trans. 36, 717–722. [PubMed: 18631147]
16. Luttermann C, and Meyers G (2007). A bipartite sequence motif induces translation reinitiation in feline calicivirus RNA. J. Biol. Chem. 282, 7056–7065. [PubMed: 17213194]
17. Powell ML, Leigh KE, Pöyry TAA, Jackson RJ, Brown TDK, and Brierley I (2011). Further characterisation of the translational termination-reinitiation signal of the influenza B virus segment 7 RNA. PLoS One 6, e16822. [PubMed: 21347434]
18. Jaafar ZA, and Kieft JS (2019). Viral RNA structure-based strategies to manipulate translation. Nat. Rev. Microbiol. 17, 110–123. 10.1038/s41579-018-0117-x. [PubMed: 30514982]
19. Mailliot J, and Martin F (2018). Viral internal ribosomal entry sites: four classes for one goal. Wiley Interdiscip. Rev. RNA 9, e1458.
20. Firth AE, and Brierley I (2012). Non-canonical translation in RNA viruses. J. Gen. Virol. 93, 1385–1409. 10.1099/vir.0.042499-0. [PubMed: 22535777]
21. Firth AE, Wills NM, Gesteland RF, and Atkins JF (2011). Stimulation of stop codon readthrough: frequent presence of an extended 3′ RNA structural element. Nucleic Acids Res. 39, 6679–6691. [PubMed: 21525127]
22. Giedroc DP, and Cornish PV (2009). Frameshifting RNA pseudoknots: structure and mechanism. Virus Res. 139, 193–208. [PubMed: 18621088]
23. Martinez-Salas E, Francisco-Velilla R, Fernandez-Chamorro J, and Embarek AM (2017). Insights into structural and mechanistic features of viral IRES elements. Front. Microbiol. 8, 2629. [PubMed: 29354113]
24. Filbin ME, and Kieft JS (2009). Toward a structural understanding of IRES RNA function. Curr. Opin. Struct. Biol. 19, 267–276. 10.1016/j.sbi.2009.03.005. [PubMed: 19362464]
25. Wilson JE, Powell MJ, Hoover SE, and Sarnow P (2000). Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. Mol. Cell Biol. 20, 4990–4999. [PubMed: 10866656]
26. Riegger RJ, and Caliskan N (2022). Thinking outside the frame: Impacting genomes capacity by programmed ribosomal frameshifting. Front. Mol. Biosci. 9, 842261. [PubMed: 35281266]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

27. Dinman JD (2012). Mechanisms and implications of programmed translational frameshifting. Wiley Interdiscip. Rev. RNA 3, 661–673. [PubMed: 22715123]
28. Loughran G, Jungreis I, Tzani I, Power M, Dmitriev RI, Ivanov IP, Kellis M, and Atkins JF (2018). Stop codon readthrough generates a C-terminally extended variant of the human vitamin D receptor with reduced calcitriol response. J. Biol. Chem. 293, 4434–4444. [PubMed: 29386352]
29. Herbert TP, Brierley I, and Brown TD (1996). Detection of the ORF3 polypeptide of feline calicivirus in infected cells and evidence for its expression from a single, functionally bicistronic, subgenomic mRNA. J. Gen. Virol. 77, 123–127. [PubMed: 8558120]
30. Horvath CM, Williams MA, and Lamb RA (1990). Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. The EMBO journal 9, 2639–2647. [PubMed: 2114979]
31. Luttermann C, and Meyers G (2009). The importance of inter-and intramolecular base pairing for translation reinitiation on a eukaryotic bicistronic mRNA. Genes Dev. 23, 331–344. [PubMed: 19204118]
32. Wennesz R, Luttermann C, Kreher F, and Meyers G (2019). Structure–function relationship in the 'termination upstream ribosomal binding site'of the calicivirus rabbit hemorrhagic disease virus. Nucleic Acids Res. 47, 1920–1934. [PubMed: 30668745]
33. Zinoviev A, Hellen CUT, and Pestova TV (2015). Multiple mechanisms of reinitiation on bicistronic calicivirus mRNAs. Mol. Cell 57, 1059–1073. 10.1016/j.molcel.2015.01.039. [PubMed: 25794616]
34. Young DJ, Makeeva DS, Zhang F, Anisimova AS, Stolboushkina EA, Ghobakhlou F, Shatsky IN, Dmitriev SE, Hinnebusch AG, and Guydosh NR (2018). Tma64/eIF2D, Tma20/MCT-1, and Tma22/DENR Recycle Post-termination 40S Subunits In Vivo. Mol. Cell 71, 761–774.e5. 10.1016/j.molcel.2018.07.028. [PubMed: 30146315]
35. Young DJ, Meydan S, and Guydosh NR (2021). 40S ribosome profiling reveals distinct roles for Tma20/Tma22 (MCT-1/DENR) and Tma64 (eIF2D) in 40S subunit recycling. Nat. Commun. 12, 2976. [PubMed: 34016977]
36. Pöyry TAA, Kaminski A, Connell EJ, Fraser CS, and Jackson RJ (2007). The mechanism of an exceptional case of reinitiation after translation of a long ORF reveals why such events do not generally occur in mammalian mRNA translation. Genes Dev. 21, 3149–3162. [PubMed: 18056426]
37. Nawrocki EP, Kolbe DL, and Eddy SR (2009). Infernal 1.0: inference of RNA alignments. Bioinformatics 25, 1335–1337. [PubMed: 19307242]
38. Nawrocki EP, and Eddy SR (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29, 2933–2935. [PubMed: 24008419]
39. Walker PJ, Dietzgen RG, Joubert DA, and Blasdell KR (2011). Rhabdovirus accessory genes. Virus Res. 162, 110–125. [PubMed: 21933691]
40. Donnelly MLL, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, and Ryan MD (2001). Analysis of the aphthovirus 2A/2B polyprotein 'cleavage'mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip. J. Gen. Virol. 82, 1013–1025. [PubMed: 11297676]
41. Luttermann C, and Meyers G (2014). Two alternative ways of start site selection in human norovirus reinitiation of translation. J. Biol. Chem. 289, 11739–11754. [PubMed: 24599949]
42. Brito Querido J, Sokabe M, Kraatz S, Gordiyenko Y, Skehel JM, Fraser CS, and Ramakrishnan V (2020). Structure of a human 48 S translational initiation complex. Science 369, 1220–1227. [PubMed: 32883864]
43. Kieft JS, Zhou K, Jubin R, and Doudna JA (2001). Mechanism of ribosome recruitment by hepatitis C IRES RNA. RNA 7, 194–206. [PubMed: 11233977]
44. Sizova DV, Kolupaeva VG, Pestova TV, Shatsky IN, and Hellen CU (1998). Specific interaction of eukaryotic translation initiation factor 3 with the 5′ nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. J. Virol. 72, 4775–4782. [PubMed: 9573242]
45. Hashem Y, Des Georges A, Dhote V, Langlois R, Liao HY, Grassucci RA, Pestova TV, Hellen CUT, and Frank J (2013). Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. Nature 503, 539–543. [PubMed: 24185006]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

46. Quade N, Boehringer D, Leibundgut M, Van Den Heuvel J, and Ban N (2015). Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. Nat. Commun. 6, 7646. [PubMed: 26155016]
47. Yamamoto H, Collier M, Loerke J, Ismer J, Schmidt A, Hilal T, Sprink T, Yamamoto K, Mielke T, Bürger J, et al. (2015). Molecular architecture of the ribosome-bound Hepatitis C Virus internal ribosomal entry site RNA. The EMBO journal 34, 3042–3058. [PubMed: 26604301]
48. Brown ZP, Abaeva IS, De S, Hellen CUT, Pestova TV, and Frank J (2022). Molecular architecture of 40S translation initiation complexes on the hepatitis C virus IRES. The EMBO journal 41, e110581. [PubMed: 35822879]
49. Kieft JS, Zhou K, Jubin R, Murray MG, Lau JY, and Doudna JA (1999). The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. J. Mol. Biol. 292, 513–529. [PubMed: 10497018]
50. Asnani M, Kumar P, and Hellen CUT (2015). Widespread distribution and structural diversity of Type IV IRESs in members of Picornaviridae. Virology 478, 61–74. [PubMed: 25726971]
51. Filbin ME, and Kieft JS (2011). HCV IRES domain IIb affects the configuration of coding RNA in the 40S subunit's decoding groove. Rna 17, 1258–1273. [PubMed: 21606179]
52. Brister JR, Ako-Adjei D, Bao Y, and Blinkova O (2015). NCBI viral genomes resource. Nucleic Acids Res. 43, D571–D577. [PubMed: 25428358]
53. Rivas E, Clements J, and Eddy SR (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat. Methods 14, 45–48. [PubMed: 27819659]
54. Weinberg Z, and Breaker RR (2011). R2R-software to speed the depiction of aesthetic consensus RNA secondary structures. BMC Bioinf. 12, 3–9.
55. Cordero P, Kladwang W, VanLang CC, and Das R (2014). The mutate-and-map protocol for inferring base pairs in structured RNA. Methods Mol. Biol. 1086, 53–77. [PubMed: 24136598]
56. Kim H, Cordero P, Das R, and Yoon S (2013). HiTRACE-Web: an online tool for robust analysis of high-throughput capillary electrophoresis. Nucleic Acids Res. 41, W492–W498. [PubMed: 23761448]
57. Kladwang W, Mann TH, Becka A, Tian S, Kim H, Yoon S, and Das R (2014). Standardization of RNA chemical mapping experiments. Biochemistry 53, 3063–3065. [PubMed: 24766159]
58. Lee S, Kim H, Tian S, Lee T, Yoon S, and Das R (2015). Automated band annotation for RNA structure probing experiments with numerous capillary electrophoresis profiles. Bioinformatics 31, 2808–2815. [PubMed: 25943472]
59. Yoon S, Kim J, Hum J, Kim H, Park S, Kladwang W, and Das R (2011). HiTRACE: high-throughput robust analysis for capillary electrophoresis. Bioinformatics 27, 1798–1805. [PubMed: 21561922]
60. Belin S, Hacot S, Daudignon L, Therizols G, Pourpe S, Mertani HC, Rosa-Calatrava M, and Diaz JJ (2010). Purification of ribosomes from human cell lines. Curr. Protoc. Cell Biol. Chapter 3. Unit 3.40–43.40. 11.
61. Matasova NB, Myltseva SV, Zenkova MA, Graifer DM, Vladimirov SN, and Karpova GG (1991). Isolation of ribosomal subunits containing intact rRNA from human placenta: estimation of functional activity of 80S ribosomes. Anal. Biochem. 198, 219–223. [PubMed: 1799204]
62. Punjani A, Rubinstein JL, Fleet DJ, and Brubaker MA (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296. [PubMed: 28165473]
63. Yamamoto H, Unbehaun A, Loerke J, Behrmann E, Collier M, Bürger J, Mielke T, and Spahn CMT (2014). Structure of the mammalian 80S initiation complex with initiation factor 5B on HCV-IRES RNA. Nat. Struct. Mol. Biol. 21, 721–727. [PubMed: 25064512]
64. Punjani A, Zhang H, and Fleet DJ (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat. Methods 17, 1214–1221. [PubMed: 33257830]
65. Rubinstein JL, and Brubaker MA (2015). Alignment of cryo-EM movies of individual particles by optimization of image translations. J. Struct. Biol. 192, 188–195. [PubMed: 26296328]
66. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. [PubMed: 15264254]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

67. Punjani A, and Fleet DJ (2021). 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. J. Struct. Biol. 213, 107702. [PubMed: 33582281]

68. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, and Ferrin TE (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82. [PubMed: 32881101]

69. Emsley P, and Cowtan K (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132. [PubMed: 15572765]

70. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung L-W, Jain S, McCoy AJ, et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D Struct. Biol. 75, 861–877. [PubMed: 31588918]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- We identify reinitiation-stimulating element (RSE) RNAs in more viruses
- Cryo-EM studies of a viral RSE RNA-ribosome complex reveal conformational dynamics
- Dynamics of ribosome-bound RSE RNA could enable start site selection after termination

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

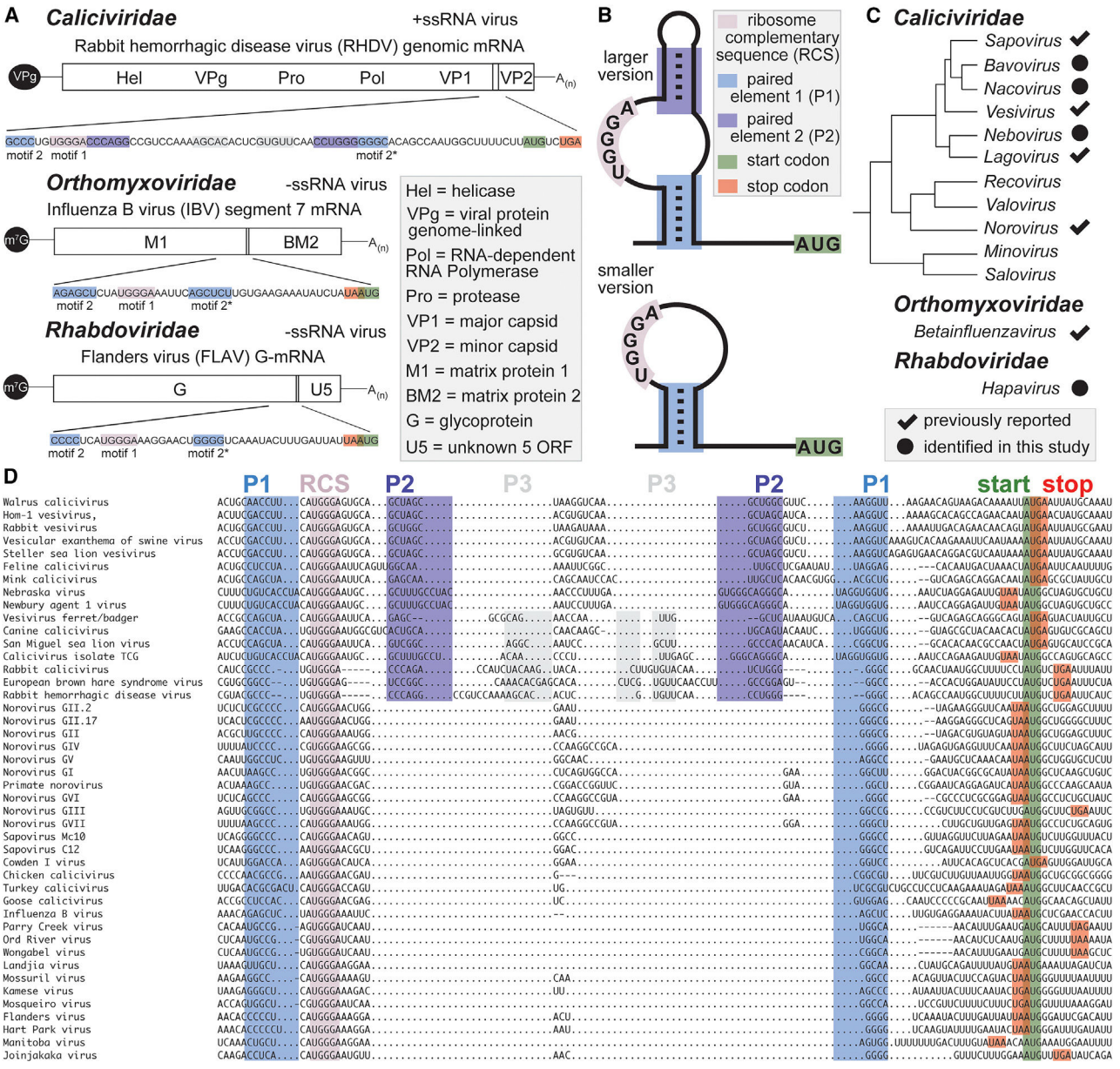


Figure 1. Genomic context and phylogenetic distribution of a class of reinitiation-stimulating RNA elements in viruses.

(A) Genomic locations of reinitiation-stimulating elements in viruses in different families.

(B) Secondary structure cartoons of the two major types of reinitiation-stimulating elements, located upstream (5′) of the downstream reinitiation start codon. Structures contain either two helical elements (P1 and P2) with the ribosome complementary sequence (RCS) embedded in an asymmetric internal loop (top) or a single helical element (P1) with the RCS located in the apical loop (bottom).

(C) Phylogenetic distribution of previously reported and newly identified reinitiation-stimulating RNAs in viruses.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(D) Alignment of the sequence, secondary structure (P1–P3), and downstream start codon of all RNAs belonging to the termination upstream ribosome binding site (TURBS) class of reinitiation-stimulating element RNAs.

See Table S2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

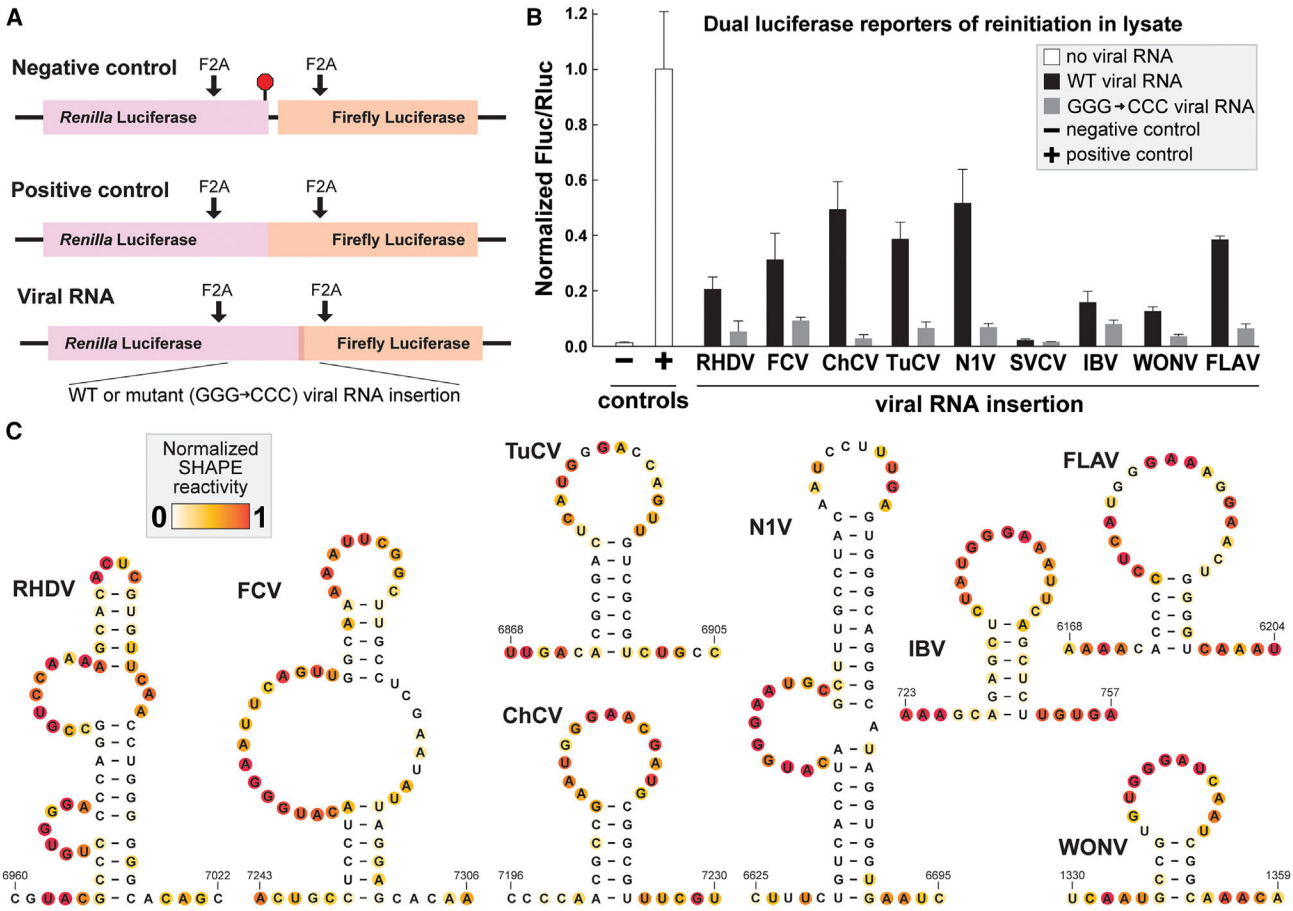


Figure 2. Structure and function of reinitiation-stimulating RNA representatives from diverse viruses.

(A) Design of reinitiation reporter constructs, which encode Rluc in the upstream ORF and Fluc in the downstream ORF. Positive and negative controls and mutant design are described in the main text. The F2A sites (indicated with arrows) cause peptide bond skipping, liberating the Rluc and Fluc proteins from peptides resulting from the inserted viral coding sequences.[40]

(B) Translation assays performed in rabbit reticulocyte lysate containing dual luciferase reporters carrying no viral RNA (controls) or putative WT or mutant reinitiation-stimulating sequences from rabbit hemorrhagic disease virus (RHDV), feline calicivirus (FCV), chicken calicivirus (ChCV), turkey calicivirus (TuCV), Newbury agent 1 virus (N1V), St-Valérien calicivirus (SVCV), influenza B virus (IBV), Flanders virus (FLAV), or Wongabel virus (WONV). The ratio of luminescence from the downstream (Fluc) to upstream (Rluc) proteins was normalized to the positive readthrough control. Error bars represent standard deviation.

(C) Reactivity to the chemical probing reagent *N*-methylisatoic anhydride, assayed *in vitro* by selective 2′ hydroxyl acylation analyzed by primer extension (SHAPE) of putative reinitiation-stimulating RNAs mapped onto the predicted secondary structure model. The full sequence that was probed, including flanking hairpins added to normalize reactivity, can be found in Table S1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

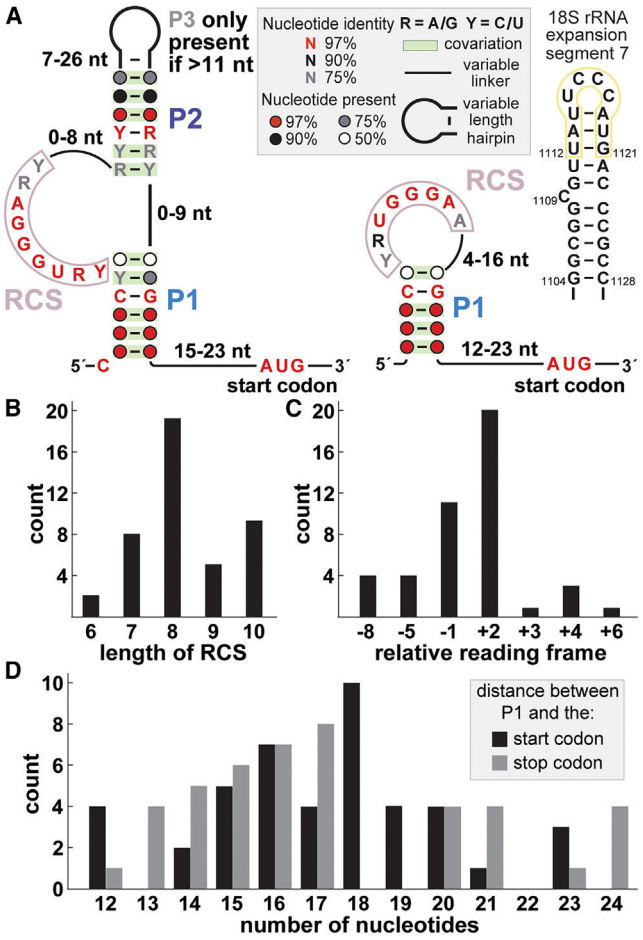


Figure 3. Conserved features among all viral TURBS-like RSE RNAs.

(A) Consensus sequence and secondary structure models of all large (left) and small (right) versions of the TURBS class of reinitiation-stimulating elements derived from 43 unique viruses. The sequence and secondary structure of the 18S rRNA from G1104–C1126, a portion of helix 26 also known as ES7, is depicted, with the nucleotides complementary to the RCS within the viral RNA outlined in yellow. See also Table S2.

(B) Histogram of the number of base pairs that could form between the viral RNA and rRNA, as determined by the length of consecutive complementary nucleotides, allowing for G•U wobble pairs.

(C) Histogram of all possible reinitiation reading frames, comparing the position of the downstream start codon to the upstream stop codon.

(D) Histogram of all possible lengths of the linker between the last nucleotide within the P1 stem and the first nucleotide in either the start (black) or stop (gray) codon among all examples of TURBS RNAs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

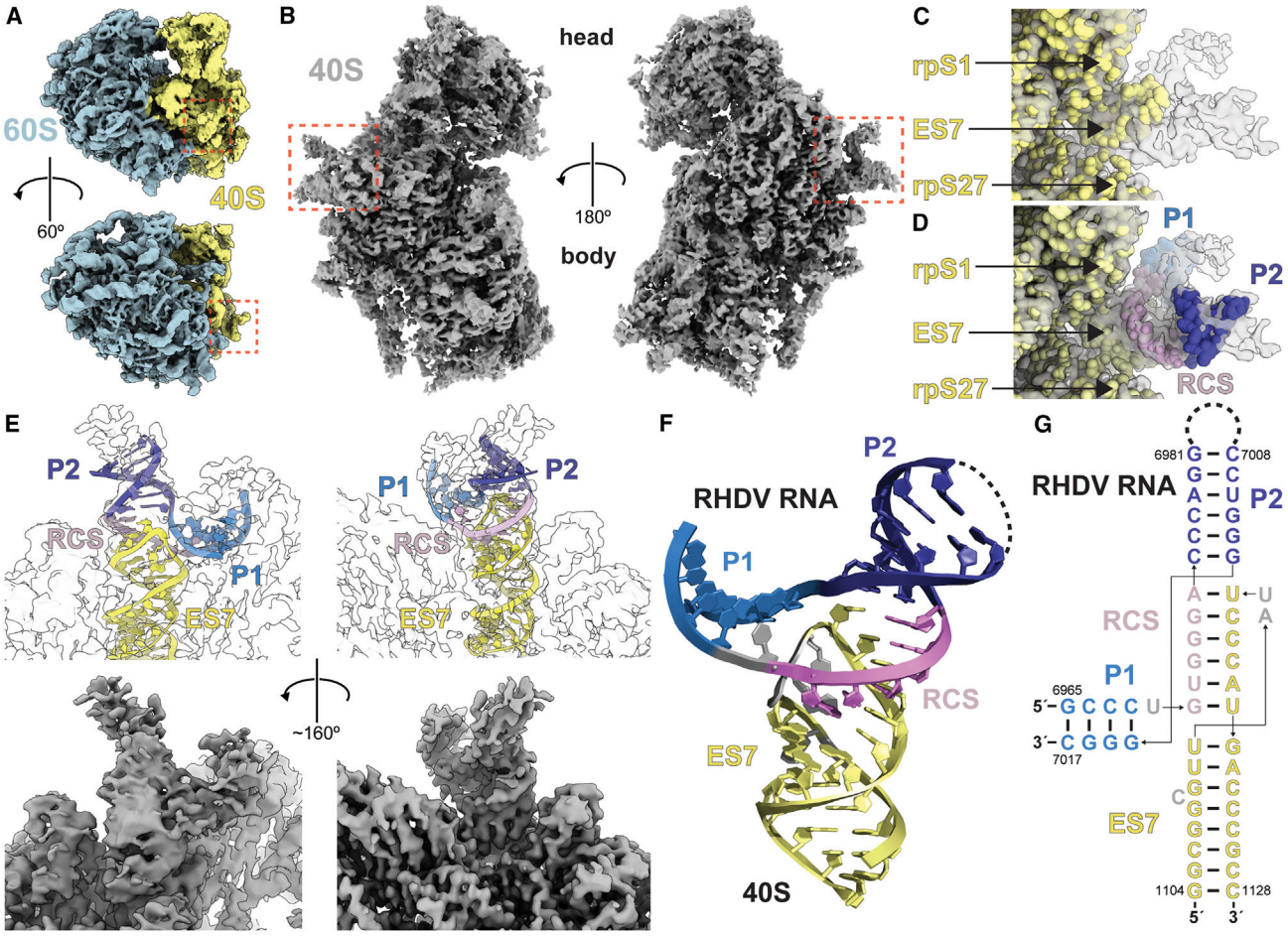


Figure 4. Cryo-EM reconstruction of an RSE-ribosome complex.

(A) Refined map of the full RHDV RNA-80S complex prior to masking. The feature corresponding to RHDV RNA is boxed in red.

(B) Refined map of the masked 40S subunit of the RHDV RNA-ribosome complex. The feature corresponding to RHDV RNA is boxed in red.

(C and D) Cryo-EM map of the RHDV RNA-ribosome complex, zoomed in on the region boxed in red in (B). An apo-40S structure (C; PDB: 6ZMW) and our model (D; PDB: 9C8K) were each fit into the map, revealing extra density corresponding to the RHDV RNA.

(E) A model of the RHDV RNA (top) built into the density of a locally refined map (bottom), including intermolecular base pairs between the RCS of RHDV and the apical loop of ES7.

(F) Model of the RHDV RNA with ribosomal ES7. Nucleotides shown in gray (RHDV U6969 and 18S rRNA U1119, A1120, and C1109) could not be unambiguously built into the density.

(G) Secondary structure of the RHDV RNA bound to the apical portion of ES7, as determined by the model in (E). Colors match (F).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

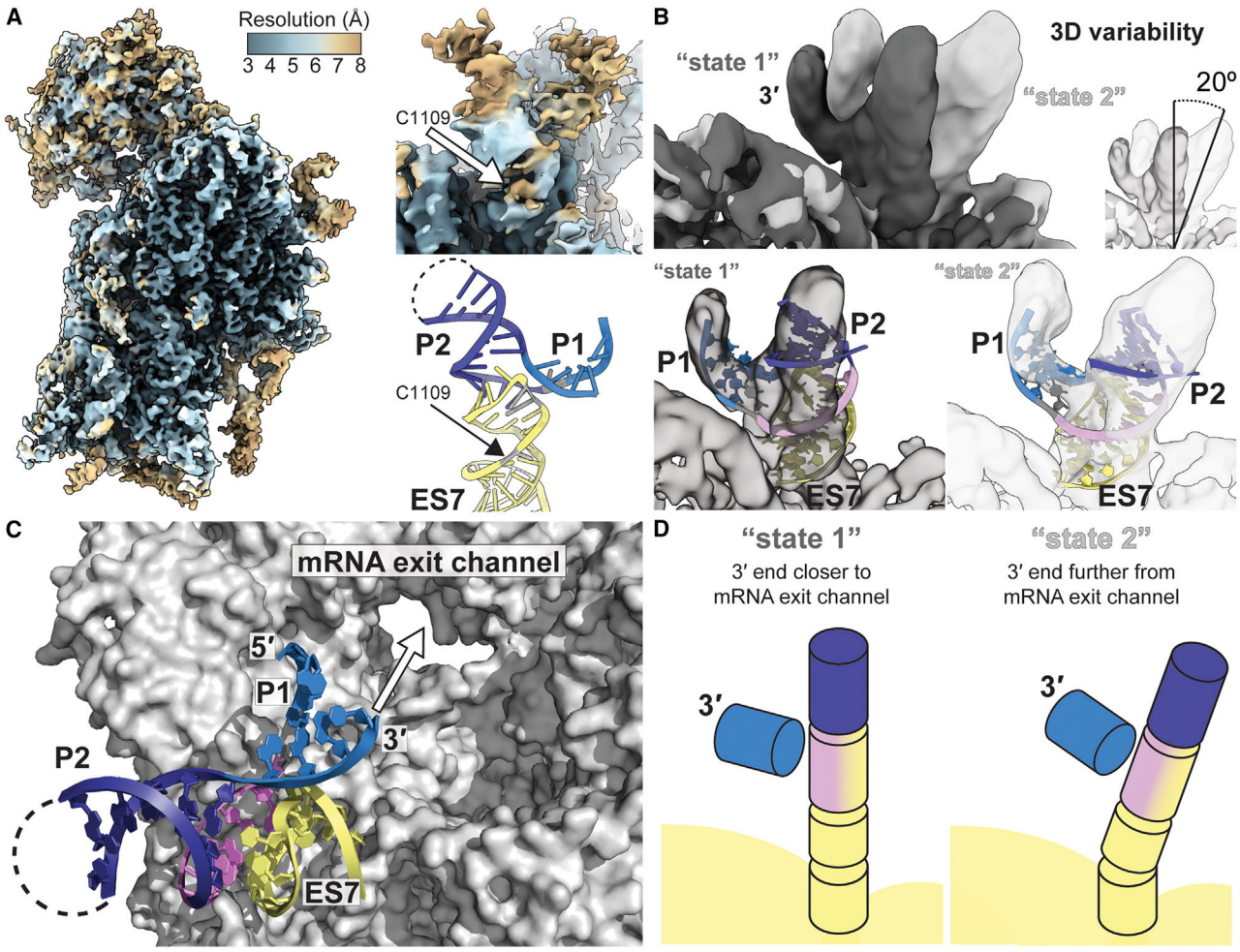


Figure 5. Conformational heterogeneity reveals dynamics related to reinitiation start site selection.

(A) Local resolution of the masked, refined 40S-RHDV RNA map. The position of 18S rRNA C1109 is depicted within the local resolution map and the model with the same view.

(B) Maps representing the two extreme states of 3D variability among the particles, which differ in their orientation by 20°. The model of RHDV RNA bound to the apical portion of ES7 rRNA (G1110–C1123) was fit into the two extreme states.

(C) Surface view of the 40S with the model of RHDV RNA bound to ES7, with the view looking down into the mRNA exit channel from the decoding groove.

(D) Simplified model of the orientation of helices in the two extreme states of 3D variability.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

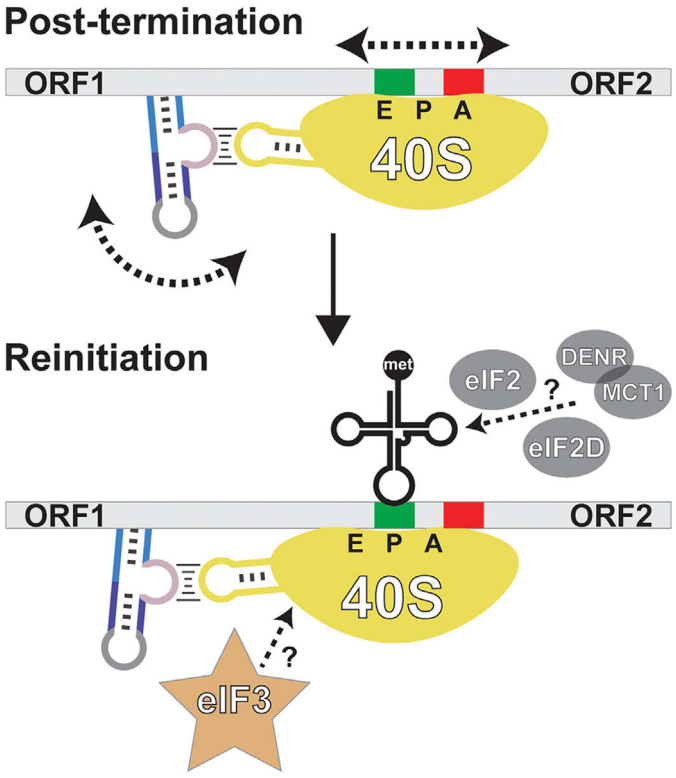


Figure 6. Model of the structural bases for reinitiation start site selection within a limited window.

During termination, the stop codon is in the A site. We propose that conformational flexibility and dynamics allow the RHDV RNA to remain bound to ES7 of the 40S while sampling different states. This motion moves the 3′ end of the structure, which could propagate to the downstream mRNA in the decoding groove and allow sampling of nearby RNA, including the nearby reinitiation start codon in the P site. Other factors, including eIF3, could select for certain conformations of RHDV-ES7 and influence start site selection. Met-tRNA$_i$ is used in this type of reinitiation, which can be delivered by eIF2, eIF2D, or DENR/MCT1 *in vitro*,[33] but the mechanisms of delivery during infection have not yet been defined.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
BglII restriction enzyme	New England Biolabs	R0144S
PspXI restriction enzyme	New England Biolabs	R0656S
T4 DNA ligase enzyme	New England Biolabs	M0202S
Q5 site-directed mutagenesis kit	New England Biolabs	E0554S
MEGAscript™ T7 transcription kit	Thermo Fisher Scientific	A57622
N-Methylisatoic anhydride (NMIA)	Sigma Aldrich	129887
Invitrogen Poly(A)Purist MAG Kit	Thermo Fisher Scientific	AM1922
Superscript III Reverse Transcriptase	Thermo Fisher Scientific	18080093
GeneScan™ 350 ROX™ dye Size Standard	Thermo Fisher Scientific	401735
Critical commercial assays		
Rabbit Reticulocyte Lysate, Nuclease-Treated	Promega	L4960
2X PerfeCTa SYBR Green FastMix	Quantabio	95072
Deposited data		
CryoEM Map of ribosome-RNA complex	This study	EMDB: 45307
Model of ribosome-RNA complex	This study	PDB: 9C8K
Oligonucleotides		
See Table S1		
Software and algorithms		
RiboKit	https://ribokit.github.io/	HiTRACE, RiboPaint
Infernal		version 1.1
MATLAB	MathWorks	version 9.9
Prism	GraphPad	version 10.2.0
Illustrator	Adobe Creative Cloud	version 27.9
ChimeraX	UCSF	version 1.6.1
PHENIX		version 1.20.1
COOT		version 0.9.8.92
PyMol	Schrodinger	version 2.5.5
cryoSPARC		version 4.4.0
QuantStudio Design and Analysis	Thermo Fisher Cloud	version 2
Other		
C-Flat™ Holey Carbon Grids for Cryo TEM	VWR	102100-136