A novel structural mechanism of ribosomal stop codon readthrough in VEGF-A

Nico O. Wagner¹, Jerome M. Edwards¹, Juliana Abramovich¹, Paromita Gupta², Harish Swaminathan², Silvi Rouskin² and Victoria M. D’Souza*¹

¹Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA. 02138, USA.
²Whitehead Institute for Biomedical Research, Cambridge, MA. 02142, USA.
*Correspondence: dsouza@g.harvard.edu

Abstract

The process of ribosomal recoding is generally regulated by an autonomous mRNA signal downstream of stop-codons. While structural studies have provided mechanistic insights into viral systems, no such studies exist in mammalian systems. Here we define a novel structural mechanism for the VEGF-A readthrough system and show that regulation is multifaceted and complex, requiring a multipartite set of RNA elements located at long distances that interact with each other and with hnRNP A2/B1 to synergistically enhance readthrough levels. The Ax-element downstream of the stop codon adopts a unique multistem (SL-Ax₁₋₃) architecture: SL-Ax₁ interacts with hnRNP A2/B1, while SL-Ax₂ interacts with an RNA element (SL-Au₁) located ~500 nt upstream at the start of the coding sequence. SL-Au₁ also independently binds to hnRNP A2/B1, which manipulates an equilibrium between alternate structures—from a sequestered bulge towards one that allows for the long-range interaction with SL-Ax₂. Overall, our study not only highlights the significance of structural organization of elements within the coding sequence of mRNA, but also provides a functional relevance of the closed-loop mRNA organization in non-canonical translation and suggests complex mechanisms allow for finer integration of many signals for a required output.

Introduction

Recoding during protein translation, either via programmed ribosomal frameshifting (PRF) or stop-codon readthrough (PRT), allows for production of extended proteins or polyproteins by bypassing stop codons. While these mechanisms are widely used by RNA viruses
to densely pack information into their small genomes and to maintain relative protein levels, they have also been discovered in cellular systems, including bacteria, yeast, Drosophilae, and humans, to regulate a wide range of functions from transcriptional regulation to signal transduction and sub-cellular localization. In viruses, ribosomal recoding is essential for maintaining critical relative ratios of proteins, for example structural (Gag) and enzymatic (Pol) in retroviruses and immunomodulatory proteins (in ORF 1a) and replicative proteins (in ORF 1b) in coronaviruses. Similarly, bacterial release factor 2 expression levels are auto-regulated via recoding and in Drosophila, recoding of the headcase protein is critical for tracheal development.

Studies in viruses have provided some insights into structural elements that drive recoding. However, no such insights currently exist for eukaryotic recoding systems, even though over 300 genes in Drosophila and ~100 genes in humans have been identified to potentially undergo functional recoding. Very recently, a new isoform of the mammalian vascular endothelial growth factor A (VEGF-A), produced via readthrough, was implicated in regulating levels of angiogenesis, indicating that recoding may play a role in regulating critical cellular processes, highlighting the need to understand the underlying mechanisms.

Canonical protein translation requires a complex interplay between ribosomes, mRNA and regulatory protein factors, but, in general, is independent of structures formed in the mRNA. On the other hand, recoding mechanisms are heavily reliant upon cis-acting RNA structures, such as stem-loops, kissing loops, or pseudoknots, around stop codons. However, the detailed mechanisms by which such structured elements regulate recoding remain undefined. In the case of readthrough, recoding signals are thought to stall ribosomes by acting as barriers to mRNA unwinding, thus allowing for the recognition of near-cognate tRNAs during PRT. The strained unwinding also potentially acts as a mechanical stressor to force ribosomes into a new reading frame during PRF.
The cis-acting, structured, recoding signals are mostly known to function autonomously. Nevertheless, in some cases their folding can be manipulated by surrounding sequences to either attenuate or enhance activity, and in rare cases they can be influenced by interaction with distal downstream sequences. For example, potato leafroll virus (PLRV), carnation Italian ringspot virus (CIRV), and barley yellow dwarf virus (BYDV) were all found to contain 3’ elements, located 700nt, 3.5kb, and 4kb, respectively, from the frameshifting signals that enhance recoding via long-range base pairing interactions\(^{18-20}\). In even rarer cases, trans-acting protein factors have been implicated in recoding. For example, the encephalomyocarditis virus (EMCV) has been shown to require proteins A2 and porcine reproductive and respiratory syndrome virus (PRRSV) to require nsp1β and poly(C) binding protein for proper recoding to occur\(^{21,22}\). It has been suggested that these proteins, in complex with the RNA, substitute for the barriers usually provided by structured signals\(^{18}\).

VEGF-A mRNA was shown to undergo a stop codon readthrough event to allow for a 22 amino acid extension (Fig. 1a)\(^{11}\). While VEGF-A\(_{164}\), the most common isoform, acts as a key angiogenic agent\(^{23,24}\), the extended VEGF-Ax has been shown to be significantly less angiogenic\(^{11,12}\). In the VEGF system, the 63 nucleotides which code for the Ax-extension immediately downstream of the stop codon, were implicated as the signal required for ribosomal stop-codon readthrough\(^{11}\). Furthermore, it was reported that the event also requires the heterogeneous nuclear ribonucleoprotein (hnRNP) A2/B1 as a trans-acting factor\(^{11}\) in that mutations to the hnRNP A2/B1 binding site on the Ax-element and siRNA knockdowns of the protein itself lead to lower levels of stop-codon readthrough. It is important to note that a subsequent study by Loughran, et al. failed to replicate the same read through capability of the Ax-element\(^{25}\). Here we show that efficient readthrough in VEGF-Ax mRNA requires full-length mRNA, with multiple cis-acting signals in the coding region working synergistically with each other and with the trans-acting hnRNP A2/B1 to confer readthrough activity.
**Results**

**VEGF-Ax readthrough is regulated by distal RNA elements located in the coding region.**

To characterize the VEGF-A system, we used a dual luciferase reporter assay in which various sequences of interest were cloned between renilla and firefly luciferases to determine stop codon readthrough efficiency. First, the natural leakiness of the UGA stop codon was determined to be ~0.05%, which is in agreement with previous studies (Fig. 1b). Second, we tested the full bicistronic coding sequence of the VEGF-A mRNA (VEGF-A+Ax), where we observed ~40-fold increase in readthrough levels to ~2.2%, a level similar to many functionally active recoding systems. To ensure that the observed increase over the natural leakiness of the UGA stop codon was robust and specific, we tested Annexin A2 mRNA (Fig. 1b), a gene with no known readthrough activity. This construct also only allowed for ~0.05% basal leakiness of the stop codon, demonstrating that VEGF-A mRNA indeed has the propensity for stimulating readthrough events. Third, we tested just the 63 nucleotide Ax-element downstream of the stop codon that was previously identified as the signal being responsible for readthrough activity. This region only led to a 4-fold increase over natural leakiness with a 0.2% readthrough activity (Fig. 1b), which is similar to the results obtained by Loughran, et al. Taken together these data indicated that sequences upstream in the coding region influence the process of readthrough in VEGF-Ax, and that the Ax-element alone is only marginally active for readthrough.

In an effort to identify factors responsible for the above observations, we first probed for the presence of potential local enhancers by adding 50, 100 or 200 nucleotides upstream of the Ax-element (Extended Data Fig. 1a). This did not result in increasing the readthrough levels to that of the whole coding sequence, thus implying that there are no local enhancer sequences for the Ax-element, and that sequences very distal and upstream to the stop codon enhance readthrough. We then created truncated constructs in which nucleotides were removed from the 5’ end of the mRNA. Surprisingly, deletion of the first 100 nucleotides dropped the readthrough activity by half (~0.94% readthrough), with a further drop to Ax-only levels at the 250-nucleotide deletion (Fig. 1b). Interestingly, sequence analysis of the upstream region within the first 100 nt presented another potential hnRNP A2/B1 binding sequences (A>GAGG). Altogether, this set
of data demonstrate that the full, bicistronic VEGF-A mRNA is necessary for efficient readthrough to occur and that the Ax-element is incapable of causing significant readthrough without the presence of multiple signals in the coding region: one within the first 100 nucleotides with a potential to bind hnRNP A2/B1 (SL-Au₁), and the second between 200-250 nucleotides (Au₂).

DMS-MaPseq reveals alternate structures in SL-Au₁ and multi-domain organization of the Ax-element.

To gain structural insights into the readthrough process, we chemically probed\textsuperscript{28} the bicistronic native mRNA from the cells used for the luciferase assay by DMS-MaPseq. Overall, the Ax-element folds into three, short stem loops (SL-Ax₁,₂) with five and four nucleotide linker sequences in between them, respectively. This architecture is unusual in that most recoding signals are made up of pseudoknots or single stem loops (Fig. 1c and Extended Data Fig. 1b). In addition, the first seven nucleotides of the Ax-element form an additional stem loop with ten nucleotides upstream of the stop codon. SL-Ax₁ contains the previously identified hnRNP A2/B1 binding site with one of the consensus A\textsuperscript{588}GG motifs predicted to be positioned in the loop\textsuperscript{11}.

DMS mapping of the 5’ end of the mRNA showed that only residues 43-102 give rise to ensembles that converge into a defined fold, which we term SL-Au₁. SL-Au₁ was predicted to form a long stem loop capped with a CCA triloop and has four short helices (1a-1d) interspersed with bulges (Fig. 1c and Extended Data Fig. 1c). In some structures, helix 1c either folded into a stem with a register shift (helix 1c’) or did not form at all and instead configured into a large bulge, suggesting that helix 1c could possibly sample alternate arrangements (Extended Data Fig. 1c). The potential hnRNP A2/B1 binding sequence spans both helix 1a and 1b at the 3’ end of SL-Au₁. Finally, characterization of Au₂ (residues 200-250) also shows the potential to fold into structured elements. However, the configurations of the structures in the ensembles varied significantly, with sequences placed in a stem in one configuration ending up in a loop in the other (Extended Data Fig. 1d); for this study, we thus only focused on the roles of SL-Au₁ and the Ax-element in stop codon readthrough.
Ax and Au₁ elements make long-range interactions.

We next synthesized the individual RNA domains by in-vitro transcription. Given that we observed an effect of the Au₁ region on stop codon readthrough, we tested whether it is able to interact directly with the Ax-element. To do so, we performed Isothermal Titration Calorimetry (ITC) to check for potential long-range interactions between the elements. While titration of SL-Au₁ into SL-Ax₃ did not give rise to any binding interactions, titration of SL-Ax₂ into SL-Au₁ gave rise to specific enthalpically-driven heats of binding between the two domains \( (K_d = 0.95 \pm 0.53 \mu M) \) at a 1:1 stoichiometry \( (n= 0.93 \pm 0.1) \) (Fig. 1d). For comparison, a study by Vander Meulen et al. has characterized a GNRA tetraloop and receptor stem docking interaction — prevalent in maintaining long range structures in RNA molecules — to have an estimated dissociation constant of \( \sim 5.5 \mu M \). It is also important to note that in the Vander Meulen et al. study, the interaction was made bivalent by designing the tetraloop and receptor onto the same construct, which in all likelihood further enhanced the binding. Thus, the relatively strong affinity between SL-Au₁ and SL-Ax₂ suggests a model in which long-range interactions add an additional layer of complexity to the VEGF-A readthrough event.

Structures of the Ax-element.

We then performed structural analysis by Nuclear Magnetic Resonance (NMR) to understand the individual motifs of the Ax-element (Extended Data Table 1). In SL-Ax₁, the 5-bp stem predicted by DMS-MaPseq is extended by an additional non-canonical A₅₈₇ o G₅₉₂ base pair, thus resulting in a four-nucleotide A₅₈₈GGA tetraloop (Fig. 1f and Extended Data Fig. 2a). The first two residues \( (A₅₈₈ and G₅₈₉) \) of the loop are in a stacked configuration with the 5’ nucleotides of the helix, with G₅₈₉ in a syn-conformation. After chain reversal, residue G₅₉₀ also adopts a syn-conformation, stacking on A₅₉₁, which further stacks continuously with the 3’ nucleotides of the helix. Thus, for the two consensus AGG hnRNP A2/B1 binding sequences, A₅₈₄GG is completely sequestered by base pairings in the stem, while the A₅₈₈GG nucleotides are structured in an unusual tetraloop.
As predicted, SL-Ax2 forms a four base pair stem capped with a U₆₀⁷CGGG pentaloop. The first two residues of the pentaloop, U₆₀⁷ and C₆₀₈, continue to stack on the 5’ strand of the stem base pairs, after which the chain turns with the following three residues G₆₀₉GG showing continuously stacking NOES with the 3’ strand of the helix (Fig 1g and Extended Data Fig. 2b). As in the GNRA-type fold, the nucleobase of G₆₁₁ stacks on the ribose of A₆₁₂ leading to an expected upfield shift of the A₆₁₂ H1’ proton to 4.5 ppm (Extended Data Fig. 2c). This positions the Watson-Crick faces of the three contiguous G residues outside the loop. Such a loop configuration where three stacked G residues engage in Watson-Crick interactions with contiguous cytosines present at a long-distance has been observed in the Haloarcula marismortui ribosomal RNA. Thus, our structure of SL-Ax2 indicate that it may be poised for the long-range interactions with structures present in Au₁ as evidenced by ITC analysis.

Finally, the imino proton assignments of SL-Ax3 indicate that it is composed of a five base pair stem-loop with a dinucleotide A₆₃₂A bulge, as predicted by DMS-MaPseq. Continuous stacking of the five base pairs was observed, indicating that the AA bulge is extruded out of the groove (Extended data Fig. 2d). However, dimerization interactions of the loop precluded us from unambiguously assigning and solving the high-resolution structure of SL-Ax3.

**Structure of the Au₁ element.**

To aid in assignments of the relatively large Au₁ construct and to unambiguously corroborate the presence of the equilibrium structure predicted in the SL-Au₁ element, we synthesized smaller segments of Au₁ (1a-1b, 1b-1c, 1b-1c’, 1c’-1d, 1c-1d and 1d; along with the full-length construct (Extended Data Fig. 3-6). Indeed, our data show that SL-Au₁ forms a single stem with a CCA loop and is interspersed with bulges that divide the stem into four short helices, 1a-1d (Extended Data Fig. 3). Furthermore, as predicted, two sets of signals were assigned for the region encompassing residues 56-64 and 81-90, which confirmed an equilibrium between helix 1c and 1c’ (Fig. 1c and Extended Data Fig. 6a).
The structures formed by helices 1a, 1b and 1d do not change in the two configurations adopted by SL-Au₁ (Extended Data Table 1). Separated by a single adenine nucleotide bulge (A₅₀), the distal helices 1a and 1b form an almost continuous stem (Fig. 1e and Extended Data Fig. 4a, 4b). A₅₀ is part of the helix with regular stacking between residues U₄₉ and C₅₁. In this configuration, the two consensus hnRNP A2/B1 binding sequences, A₉₃GG and A₉₆GG, are all involved in base pairing interactions within helix 1b and 1a, respectively (Extended Data Fig. 3).

Furthermore, the proximal helix 1d is rich in GU basepairs (U₆₇-G₇₉, G₆₈-U₇₈ and U₇₀-G₇₆), which make up half of the six base pairs in the helix as evidenced by the three upfield shifted GU cross-peak pairs in the imino region of the NOESY spectrum (Fig. 1e and Extended Data Fig. 5b). The C₇₂CA triloop which caps this helix is well structured, with chain reversal occurring between residues C₇₂ and C₇₃ and stacking with both sides of the helix (Fig. 2a).

The central helix adopts two different configurations (helix 1c and 1c') (Extended Data Fig. 6a). Helix 1c shows the formation of an 8-bp stem (residues 57-64 and 81-88) including two tandem, non-canonical C-C base pairs that are protonation dependent, making the equilibrium between helix 1c and 1c' highly sensitive to pH (Fig. 2b and Extended Data Fig. 5b). In the 1c configuration, the junction between helices 1d and 1c also has an almost continuously stacking architecture with a single adenine nucleotide bulge (A₆₅) stacked inside the helix. On the other end, the junction between 1c and 1b is defined by an internal loop (A₅₆; C₈₉A). While A₅₆ stacks with the helix (Extended Data Fig. 4b), the opposing residue A₉₀ is extruded out toward the minor groove. The H₄' sugar proton of residue C₈₉ gives NOEs to the aromatic H₈ proton of A₉₂, placing it in the major groove and within hydrogen bonding distance of U₅₄-A₉₂ (Fig. 2c). This local triple-base configuration obtained is similar to arginine sandwich motifs. Overall, with continuous stacking of all helices in the helix 1c configuration, the structure of SL-Au₁ has a linear appearance (Fig. 1e).

On the other hand, due to the base pairing register shift in the helix 1c' configuration, residues A₅₆ to A₉₀ and A₈₁ to A₈₅ form two opposing bulges separated by a 5-bp stem (residues 60-65 and 86-90). These four (A₅₆CCA) and five (A₈₁CCCCA) nucleotide bulges flanking helix 1c'
cause bent geometries between junctions with helices 1b and 1d, respectively (Fig. 2d). At the 1b-1c’ junction, residues A56 and C57 stack on helix 1b, whereas residues C58 and A59 stack under helix 1c’ (Fig. 2e). There are no NOEs between C57 and C58, indicating opposing orientations. At the 1c’-1d junction, the A81 in the A81CCCA bulge base pairs with the A65 in the opposite strand (Fig. 2f). The following three cytosine residues show a continuous NOE walk. With only a partial stacking between C84 and A84, the triplet Cs are extruded out in way that seem poised to make long-range interaction with SL-Ax2, as indicated by the ITC data (Fig. 1d, 2d).

**SL-Ax2 and SL-Au1 interact via loop-bulge docking interactions**

To understand the mode of the long-range interaction, we performed small angle X-ray scattering (SAXS) analysis of free SL-Au1 and in complex with SL-Ax2. Furthermore, since the C8CC motif in SL-Au1 and G60GG motif in SL-Ax2 resemble a motif found in *Haloarcula marismortui* ribosomal RNA where the three stacked G residues engage in Watson-Crick interactions with contiguous cytosines¹⁰, we also modeled a similar interaction between SL-Au1 and SL-Ax2 (Fig. 2g). Compared to the free form, the SAXS envelop of the bound form of SL-Au1 showed a clear additional density, which was able to accommodate the structural model of SL-Ax2 in the complex (Fig. 2g).

To unambiguously assign the docking interface, we titrated SL-Ax2 in the 1c’-1d construct, which allowed for identification of the A81-C84 bulge as the one involved in the long-range interaction (Extended Data Fig. 5b). While no global changes in the secondary structure of either 1c’-1d or SL-Ax2 occurred, resonances for three new G-C Watson-Crick base pair formations were observed, supporting complex formation observed by ITC (Fig. 1d). The combination of binding heats observed by ITC, intermolecular base pairing NOEs observed by NMR, and the complex obtained by SAXS all support the formation of the long-range interaction between the SL-Au1 bulge and SL-Ax2 loop.

**hnRNP A2/B1 binds SL-Ax2 and SL-Au1**

We next wanted to test if the previously identified hnRNP A2/B1 binding sites (A584GG and A588GG in SL-Ax1) and the new potential sites (A81GG and A85GG in helix 1a and 1b,
respectively) interacted with hnRNP A2/B1. It was previously reported that mutating the A$_{288}$GG hnRNP A2/B1 binding site led to decreased readthrough levels. Similar to previous hnRNP binding studies, we used an hnRNP construct that consisted of the two RNA-recognition motifs (RRMs), but lacked the aggregation-prone C-terminal domain, previously shown to be involved in oligomerization and nuclear localization of the protein (Fig. 3a). Titration of hnRNP A2/B1 into SL-A$_x$1 and SL-A$_u$1 by ITC gave data that fit well with a single-site binding model with estimated dissociation constants of 200 ± 42 nM and 348.5 ± 22 nM, respectively (n=1.02 ± 0.34 and 1.2 ± 0.33, respectively) (Fig. 3b). These $K_d$'s are in accordance with previously published binding affinity of similar sequences. Nevertheless, a critical difference in the binding mode was observed between the two interactions: while the interaction with SL-A$_x$1 was enthalpically driven ($\Delta H$ -10.4 ± 4.7 kcal/mol and $\Delta S$ -4.4 ± 15.3 cal/mol/K), the interaction with SL-A$_u$1 ($\Delta H$ -3.6 ± 1.7 kcal/mol and $\Delta S$ 17.6 ± 5.7 cal/mol/K) was entropically driven, the latter suggesting rearrangement of RNA structure upon protein binding.

Since hnRNP A2/B1 bound both SL-A$_u$1 and SL-A$_x$1, and since it is able to dimerize through its C-terminal domain, we tested if it is able to mediate the long-range interaction between SL-A$_u$1 and SL-A$_x$1 via its dimerization domain. Using an in-vitro rabbit reticulocyte lysate (RLL) system, we tested the effects of full-length hnRNP A2/B1 and hnRNP A2/B1 lacking a C-terminal domain (ΔCTD) on VEGF-A readthrough efficiencies (Fig. 3c). Recoding was replicated in this assay, with baseline readthrough values of ~4%, which is slightly higher than those in cells. Such differences have previously been reported, and generally arise from tighter regulation in cells. By adding either full-length hnRNP A2/B1 or hnRNP A2/B1-ΔCTD to VEGF-A mRNA, we are able to increase readthrough by a factor of ~2.5 (Fig. 3c) to 9.6%, and further additions of hnRNP A2/B1 reproducibly reduced readthrough. These results not only confirm the stimulatory effect of hnRNP A2/B1 binding, but also point to other hnRNP A2/B1 driven mechanisms to balance required readthrough levels. Importantly, there was no observable difference in readthrough levels between the full-length and dimerization-deficient hnRNP A2/B1 constructs, indicating that hnRNP A2/B1 dimerization is not required for readthrough activity, and that another mechanism must be responsible for mediating the observed long-range
interactions. As a control, using the Murine Leukemia Virus (MLV) readthrough system resulted in constant readthrough levels, showing that the effects of hnRNP A2/B1 are specific to VEGF-A (Fig 3c).

To determine the exact binding site within SL-Au₁, we performed a protein titration on our NMR sample. Addition of even sub-stoichiometric (0.3 equivalents) of RRM1/2 to SL-Au₁ resulted in pushing the equilibrium to the 1c’ configuration, as evidenced by the disappearance of the U₆₀–A₈₅ and A₅₉–U₈₆ base pairing in helix 1c and the appearance of A₆₄–U₈₆ in helix 1c’ (Fig. 3d). Furthermore, addition of one equivalent of protein led to chemical shift changes both in the protein and RNA and gave rise to multiple intermolecular NOEs. Interestingly, our data show that binding occurred specifically to molecules in the 1c’ configuration, as evidenced by the specific disappearance of the U₆₀–A₉₀ base pair, which is representative of helix 1c’. Perturbations of A₉₃ resonances indicated that binding occurs at the consensus hnRNP A2/B1 binding site (Fig. 3d). Overall, these experiments suggest that the role of hnRNP A2/B1 binding to SL-Au₁ may be to modulate the equilibrium toward the 1c’ conformation, which is required for long-range interactions.

Based on recent findings that it is the RRM1 that recognizes the AGG sequence³⁶, we used a shortened hnRNP construct to determine the mode of interaction with SL-Ax₁. Binding between RRM1 and SL-Ax₁ resulted in perturbations of specific residues in the loop, with G₅₉₀ giving clear intermolecular NOEs to an aromatic residue of RRM1, thus providing confirmation of residues A₅₈₈GG sequence as the binding site (Fig. 3e). The second potential A₅₈₄GG binding motif remains sequestered within the stem as unaffected by protein binding; in fact, the majority of the SL-Ax₁ stem structure remaining unchanged upon protein addition. Thus, at both the SL-Au₁ and SL-Ax₁ binding sites, hnRNP A2/B1 recognizes structured loops, as recently described in viral IRES translation regulation³⁵.

The various elements synergistically regulate recoding efficiencies
Given both the unusual multi-stem nature of the Ax-element, and potential unique functions of the individual stems, we wanted to first determine their contributions to recoding. We created constructs in which SL-Ax₂ and SL-Ax₃ elements were removed (Fig. 4a). We also checked DMS-MaPseq reactivities to ensure that in either deletion, the fold of SL-Ax₁ is not perturbed (Extended Data Fig. 7a). Deletions of SL-Ax₃ and both SL-Ax₂ and SL-Ax₃ led to significant reduction in readthrough levels by 40% and 70%, respectively, implying that the entirety of the Ax-element is required to maximize readthrough levels (Fig. 4b). Interestingly, removing the entire Ax-element decreased readthrough levels by almost 80% to a level of 0.5% (Fig. 4b).

While we do not yet understand the structural mechanism by which SL-Ax₃ influences readthrough, our understanding of the long-range interaction between SL-Ax₂ with SL-Au₁ allows us to test for its contribution. We thus created mutants in which the C₆₅CC involved in long range interaction in SL-Au₁ was sequestered by either changing them to Gs along with mutating G₆₀ to C, or by mutating the opposing C₆₅C residues to guanosines. Both of these give rise to three contiguous G-C base pairs, which should preclude inter-domain interaction of this bulge and lead to appreciable decreases in readthrough by ~40% and ~50%, respectively (Fig. 4c). Similarly, disruption of the docking G₆₉GG motif in SL-Ax₂ by replacing the pentaloop with a GAAA tetraloop, led to an equivalent 40% decrease in readthrough activity (Fig. 4c).

Next, we wanted to test the contributions of protein binding motifs to recoding efficiencies. We first tested protein binding efficiency of the previously published dinucleotide A₅₈₇A to U₅₈₇U mutation in SL-Ax₁. ITC experiments on a SL-Ax₁ A₅₈₇A:UU mutant yielded a ~1.5-fold decrease in binding when compared to wild-type SL-Ax₁ (Extended Data Fig. 7b). This correlates with an observed ~50% reduction in readthrough levels (Fig. 4c). Similarly, for experiments on a SL-Au₁ A₉₂A:UU mutant, we observed a ~2-fold increase in Kₐ, or 2-fold decrease in binding affinity (Extended Data Fig. 7c), which correlates to a ~40% decrease in readthrough levels for this mutation (Fig. 4c), suggesting a strong-interplay between hnRNP A2/B1 binding and readthrough of the VEGF-Ax system. Interestingly, a double mutant in which
both hnRNP A2/B1 binding sites were mutated displayed higher readthrough efficiency than either single mutant (Fig. 4c). This epistatic behavior is a strong indication of direct influence of the two binding events on local RNA structure, and consequently their long-range interactions. Furthermore, it suggests that the closed-loop configuration of VEGF mRNA may place SL-Au and SL-Ax in close proximity, thus allowing for the observed epistasis. Overall, these mutational studies show the importance of the various players involved in regulating readthrough levels of the VEGF mRNA.
Discussion

Here we have presented evidence of a complex eukaryotic translational readthrough system employing several novel mechanisms of action. Canonically, readthrough and frameshifting events only require a single *cis*-acting signal in the vicinity of the stop codon. In this study, we performed structural and mutational work to better understand the role of the previously identified readthrough signal, the Ax-element. In addition to the previously identified stop-codon proximal signal, we identified several new components of this readthrough event: two *cis*-acting elements upstream of the stop codon, an additional hnRNP A2/B1 binding site, and long-range RNA-RNA interactions spanning the length of the VEGF-A gene.

The presence of a 5' element conferring ribosomal readthrough is novel. In retroviruses, where ribosomal recoding has been studied more extensively, readthrough signals tend to be localized in the immediately vicinity of the stop codon\textsuperscript{14,16}. While distant downstream elements enhancing frameshifting in Pea enation mosaic virus\textsuperscript{41} and readthrough in Potato leafroll virus\textsuperscript{20} have been studied, to the best of our knowledge, there are no examples of distant upstream signals. Given that these novel *cis*-acting elements, located at ~250 and ~500 nucleotides upstream of the stop codon, are in the coding region, this is an even more unusual setup for ribosomal readthrough. The dispersed, multi-domain architecture of the Ax-element is also unique in that RNA signals that cause recoding in ribosomes are generally made up of a single stem loop or pseudoknot structures. Such an organization, along with the presence of long-range interactions between signals located at the two ends of the coding sequence and multiple hnRNP A2/B1 binding sites, hints towards a complex mechanism for regulation of readthrough in mammalian systems.

Altogether, our study allows us to start putting together a mechanism by which readthrough frequency— and hence relative proportions of VEGF-A (angiogenic) to VEGF-Ax (less angiogenic) isoforms— may be maintained (Fig. 5). Our model suggests that regulation takes place via modulations of alternate mRNA structures by hnRNP A2/B1 levels\textsuperscript{42}. Binding of hnRNP A2/B1 to SL-Au\textsubscript{1} allows for a register shift in the base pairing of helix 1c to the 1c'
configuration, which is capable of making long-range interactions with SL-Ax2 near the vicinity of the stop codon. The two conformations of SL-Au1 were reproducibly observed in cells, adding another example of use of conformational equilibrium in RNA molecules to regulate critical functions in cells. Additionally, while we did not address the roles of SL-Ax1 or the Au2 element ~250 nt upstream of the stop codon in this manuscript, they will further contribute to the complexity of this readthrough system.

Whereas the function of hnRNP A2/B1 binding to SL-Au1 is to modulate the equilibrium, the function of hnRNP A2/B1 binding to SL-Ax1 is unclear. However, given the spacing between the stop codon and the hnRNP A2/B1 binding site at SL-Ax1, the binding of hnRNP A2/B1 may act as a stressor upon the ribosome at the stop codon. This would be similar to EMCV and PRRSV in which binding of A2 protein and nsp1β/poly(C) binding protein, respectively, takes the place of the canonical structured element21,22. Given the apparent lack of an extended, complex secondary structure in VEGF-A mRNA, this hnRNP A2/B1-induced stress, in conjunction with the additive effects of the long-range RNA interaction, may take the place of complex secondary structural elements, such as pseudoknots, found in many other recoding events15.

The long-range cross-talk interaction we have identified is a novel mechanism through which ribosomal recoding events may be regulated. While the broader concept of long-range RNA interactions is not new, it has recently been suggested that more broadly, most if not all mRNAs exhibit 5’ to 3’ communication through a combination of proximity base pairing and RNA-binding proteins43,44. For example, recent studies measuring the spacing between 5’ and 3’ ends of coding mRNA, both in-vitro and in-vivo, have found them to be much closer than would be predicted for a randomly folded RNA43,45-47. While these studies support a previously established closed-loop model of RNA in which the polyA tail and 5’ cap of mRNA are spatially close to regulate canonical translation initiation and termination, our studies show that the closed-loop model can give rise to specific long-range interactions that can have functional consequences on translational elongation.
Methods

In-cellulo translation assay

Sequences of interest for in-cellulo dual luciferase assays were cloned between renilla luciferase (RLuc) and firefly luciferase (FLuc) using BamHI and SacI restriction sites on the p2Luc plasmid\(^2\). To normalize against differential FLuc and RLuc luminescence, mRNA degradation, and ribosome fall-off products, each construct containing a TGA stop codon was cloned in tandem with a control construct containing a GCA codon. This control construct was used to normalize for readthrough activity.

Dual luciferase assays were performed in white 96-well half-area flat bottom plates (Corning). Bovine aortic endothelial cells (BAOECs), grown at 37°C in 5% CO\(_2\), were diluted to 30% confluency in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin. Cells were then plated at 150 µL per well with 5 total replicates per construct. 1 hour post-plating, 11µL transfection mix (480 ng plasmid, 1.2 µL Fugene6 (Promega) and 46.8 µL DMEM) was added to each well. Cells were allowed to express RLuc and FLuc for 2 days prior to taking readings.

Luciferase luminescence readings were collected on a SpectramaxL (Molecular Devices) plate reader at 470 nm with an integration time of 1 second using the reagents in the Dual-Luciferase Reporter Assay System kit (Promega). Prior to collection, cells were lysed in 35 µL of 1x passive lysis buffer (Promega) for 15 minutes. Then 50 µL FLuc substrate was added and allowed to sit for 30 seconds prior to collecting luminescence readings. Similarly, 50µL RLuc substrate was added, allowed to sit, and luminescence was collected again. Readthrough percentages were calculated as previously described\(^2\). P-values were calculated using a 2-tailed T-test.

In-vitro translation assay

Template DNA for rabbit reticulocyte lysate (RRL) assays was prepared by PCR amplifying hnRNP A2/B1 constructs and VEGF-A sequences of interest flanked by RLuc and FLuc. 8 µL PCR product was used as a template for in-vitro transcription using the mMESSAGE
mMACHINE T7 Transcription Kit (Invitrogen). RNA was diluted to 10 µg/µL in nuclease free water.

8 µL RRL (Promega) master mix (100 µL reaction mixture, 17.5 µL nuclease free H₂O, and 2.5 µL methionine) was combined with 1 µL of each RNA (VEGF and hnRNP A2/B1) at the desired concentration. For no protein samples, 1 µL transcription reaction buffer was added instead. Reactions were then allowed to continue for 90 minutes at 30°C prior to placing on ice for 20 minutes. Lysates were then plated at 2.5 µL per well with 3 total replicates per construct in white 96-well half-area flat bottom plates (Corning). Luciferase luminescence readings were collected as described above.

Preparation of hnRNP A2/B1

An hnRNP A2/B1 consisting of RRM1/2 was cloned into a pGEX-6P-1 plasmid using BamHI and NotI restriction sites.

BL-21 cells containing GST-tagged hnRNP A2/B1 constructs in pGEX-6P-1 were inoculated in LB media containing 100 µg/mL ampicillin and grown overnight at 37°C with shaking at 200 rpm. In the morning, 20 mL of overnight culture were diluted into 2 L LB media containing 100 µg/mL ampicillin. Cells were allowed to grow at 37°C with shaking until reaching OD₆₀₀ ~1. Cells were cooled to 20°C and subsequently induced with 100 µM IPTG for 16 hours, after which they were pelleted at 5,000g for 20 minutes. A pellet corresponding to 2 L was resuspended in 50 mL hnRNP lysis buffer (50 mM Phosphate pH 7.5, 200 mM NaCl, 40 mM MgCl₂, 1 mM DTT, 0.1% Triton X-100) and sonicated (Branson Sonifier 450) at setting 7-8 with 50% amplitude for 6 pulses of 60 seconds each. 5 mL of 4% polyethyleneimine (PEI) was added drop wise to the lysed cells and allowed to mix on ice for 30 minutes. The lysed cells were then spun at 27,000g for 20 minutes.
Lysate supernatant was added to Pierce Glutathione Agarose Affinity Purification Media (Thermo Scientific) and allowed to bind for 2 hours on a rocker at 4°C. Following binding, beads were washed twice with 50 mL ice cold hnRNP wash buffer (50 mM Phosphate pH 7.5, 500 mM NaCl, 1 mM DTT). Subsequently, protein-bound beads were loaded onto a column and washed overnight with 2 L hnRNP wash buffer using a pump at a flow rate of 1.8 mL/minute at 4°C. Protein was eluded in 50 mL elution buffer (20 mM Phosphate pH 7.5, 15 mM KCl, 2 mM MgCl₂, 5% glycerol) and 2.4 mg PreScission protease. The column was allowed to rock at 4°C for 8 hours prior to collecting the first elution. hnRNP A2/B1 was concentrated in Ultra Centrifugal Filter Units (Amicon) with a 3 kDa cutoff and washed into the appropriate buffer as needed.

Large-scale synthesis and purification of RNA

RNA constructs were either ordered (IDT or Horizon Discovery) or synthesized in-house using T7 polymerase as previously described⁴⁹. Final RNA was washed into appropriate buffer as needed using Ultra Centrifugal Filter Units (Amicon).

NMR & Structure Calculations

RNA samples were placed in 300 µL NMR buffer (10 mM Tris pH 5.6, 10 mM NaCl or 10 mM HEPES pH 5.6, 50 mM KCl) in D₂O. For assignments of exchangeable proton and nitrogen resonance NOESY and HSQC were collected on samples lyophilized and resuspended in H₂O supplemented with 10-15% D₂O at 280K. For non-exchangeable proton and carbon assignments spectra were collected in 99.96% D₂O. Assignments were done with the use of NMRFx⁵⁰. As previously described, CYANA⁵¹ was used to calculate initial structures of the SL-Au₁, SL-Ax₁ and SL-Ax₂ with manually assigned restraints. Standard torsion-angle (±50° deviations from ideality) and cross-helix phosphate-phosphate restraints were used for α-helical regions. The structure with the lowest target function was used to create a coordinate file for refinement in xplor-nih⁵²,⁵³. Phosphate-phosphate distances were not used at this step. Molecular images were generated with PyMOL (http://www.pymol.org).

ITC
Most ITC experiments were performed in ITC buffer (10 mM HEPES pH 8.0, 50 mM KCl, 1 mM EDTA, 1 mM BME), with the exception of certain SL-Ax constructs (20 mM phosphate pH 7.2, 15 mM KCl, 2 mM MgCl₂, 5% glycerol). RNA constructs were diluted to 5 µM for and hnRNP A2/B1 constructs were diluted to 100 µM. ITC experiments were performed on an ITC-200 microcalorimeter (MicroCal) by loading RNA into the cell and protein into the syringe. Protein was injected in 2 µL increments for a total of 20 injections. Baseline correction, binding curve fitting, and parameter calculation were performed using Origin (OriginLab).

**DMS-MaPseq**

*In-cellulo* DMS-MaPseq samples were prepared by seeding 10-cm cell culture dishes of BAOECs at 30% confluency. 1 hour post-plating, 660µL transfection mix (5 µg plasmid, 26 µL Fugene6 (Promega) and 650 µL DMEM) was added to each well and allowed to grow for 2 days. Sample processing and analysis was performed as previously described²⁸,⁵⁴.

**Small Angle X-Ray Scattering**

SAXS experiments were performed at the SIBYLS beamline (Advanced Light Source, Lawrence Berkeley National Laboratory)⁵⁵. Samples were prepared in NMR buffer (10 mM Tris pH 5.6, 10 mM NaCl) supplemented with 2% glycerol. Data analysis and *ab initio* envelope reconstruction were done using ScÅtter software⁵⁶.
### Extended Data Table 1 - NMR statistics and restraints

|    | AN1 region |    | AN2 region |    |
|----|------------|----|------------|----|
|    | NMR and refinement statistics |    | NMR and refinement statistics |    |
|    | NMR distance and dihedral constraints |    | NMR distance and dihedral constraints |    |
|    |    |    |    |    |
| **Distance restraints** |    |    |    |    |
| Total NOE | 465 | 123 |    |    |
| Intraridge | 297 | 74 |    |    |
| Inter-residue | 366 | 49 |    |    |
| Sequential (i−j) | 144 | 45 |    |    |
| Nonsequential (i−j) | 2 | 4 |    |    |
| Hydrogen bonds | 127 | 37 |    |    |
| Total dihedral angle restraints | 1014 | 272 |    |    |
| Base pair | 0 | 0 |    |    |
| Sugar puckers | 240 | 64 |    |    |
| Backbone | 774 | 208 |    |    |
| Based on A-form geometry | 1014 | 272 |    |    |
| **Structure statistics** |    |    |    |    |
| Violations (mean ± s.d.) | 2.8 ± 0.4 | 0.6 ± 0.5 |    |    |
| Distance constraints (Å) | 41.0 ± 3.5 | 11.8 ± 2.1 |    |    |
| Deviations from idealized geometry |    |    |    |    |
| Bond lengths (Å) | 0.085 ± 0.008 | 0.094 ± 0.001 |    |    |
| Bond angles (°) | 0.628 ± 0.006 | 0.667 ± 0.007 |    |    |
| Improper (°) | 0.352 ± 0.009 | 0.369 ± 0.009 |    |    |
| Average pairwise r.m.s. deviation (Å) |    |    |    |    |
| All RNA heavy | 0.357 | 0.168 | 0.0867 | 0.0486 |
| All nucleotides | 0.377 | 0.178 | 0.0887 | 0.0563 |

---

Footnote: Pairwise r.m.s. deviation was calculated among the refined structures.
Figure 1: Identification and structures of the elements involved in mediating VEGF-A readthrough. (a) Schematic diagram of the VEGF-A mRNA without (top) and with (bottom) the Ax-element. (b) Functional readthrough levels show that only the full-length VEGF-A + Ax mRNA leads to substantial readthrough levels. Error bars indicate standard error (n=3) and statistical tests were performed using a 2-tailed t-test (* p < 0.05, ** p < 0.005). (c) The secondary structure reveals an extended stem-loop for SL-Au, including an alternate conformation (bases shown in orange and blue for simplicity), as well as shorter stem-loops SL-Ax within the Ax-element. hnRNP binding sites are indicated in purple, stop codons are shown in red, and bases involved in the long-range interaction are shown in green. (d) ITC data show that SL-Au is able to specifically interact with SL-Ax, (black squares), but not with SL-Ax, (white squares). Tertiary folds are shown for SL-Au, (PDB 7KUB) (e), SL-Ax, (PDB 7KUC) (f), and SL-Ax, (PDB 7KUD) (g) with the same base coloring as in (c). (h) A model was created showing the docking interaction between SL-Au and SL-Ax.
Figure 2: Structural features of SL-Au₁. (a) Molecular detail of the C₇₂CA triloop that caps SL-Au₁. (b) Helix 1c of SL-Au₁ contains 2 tandem C-C non-canonical basepairs. (c) SL-Au₁ is able to form a triple base-pair interaction between C₈⁹ and the A₉²-U₅₄ base-pair. (d) Helix 1c’ assumes a kinked conformation and is flanked by an A₅₆CCA bulge at the junction with helix 1b (e) and an A₈₁CCCA bulge at the junction with helix 1d (f). (g) The density, as determined by SAXS, of free SL-Au₁ in the 1c’ conformation (blue) shows the bend described above. It also overlays with a model envelope (grey) of the SL-Au₁:SL-Ax, long-range interaction complex. Base coloring is as in Figure 1.
**Figure 3:** hnRNP A2/B1 binds to VEGF-A mRNA to help mediate readthrough. (a) Schematic of the hnRNP A2/B1 domain architecture showing its RNA-recognition motifs (RRM1 and RRM2) and C-terminal domain (CTD), consisting of an Arg-Gly-Gly box (RGG), prion-like domain (PrLD), and M9 transport sequence. (b) ITC binding studies of RRM1/2 to SL-Au (left) or SL-Ax (right) show tight and specific binding. Representative curves of 2 trials are shown. (c) Rabbit reticulocyte lysate assays to which either full length hnRNP A2/B1 (dark grey) mRNA or RRM1/2 (light grey) was added show no difference in readthrough activity for an MLV control construct (top) or a VEGF-A construct (bottom). Error bars indicate standard error (n=3) and statistical tests were performed using a 2-tailed t-test (* p < 0.05). (d) Titration of RRM1/2 (black, 0 molar equivalents; blue, 0.3 molar equivalents; pink, 1 molar equivalent) into SL-Au at 280 K shows the shifting of several $^1$H-$^1$H NMR peaks of base pairs disrupted by protein binding. (e) Titration of RRM1 into SL-Ax at 311K results in $^1$H-$^1$H NOE cross-peaks between G$_{590}$ and aromatic ring hydrogens of the protein.
Figure 4: Each element of the VEGF-mRNA separately contributes toward readthrough activity.

(a) Schematic indicating mutations made to the VEGF-A mRNA. Mutations are labeled and shown in red. For truncated products, the maximal extent of the VEGF-A mRNA is indicated underneath the schematic.

(b) Functional readthrough experiments show that each stem-loop within the Ax-element contributes toward readthrough.

(c) Each tested mutation led to a decrease in readthrough levels, highlighting the importance of each region in mediating ribosomal readthrough. Error bars indicate standard error (n=3) and statistical tests were performed using a 2-tailed t-test (* p < 0.05, ** p < 0.005).
Figure 5: RNA alternate conformations mechanism of VEGF-A readthrough. A mechanistic model of the VEGF-A coding mRNA shows the formation of a linear SL-Au₁ element at the 5’ end of the RNA and three stem loops (SL-Ax₁₋₃) at the 3’ end of the RNA in the absence of hnRNP A2/B1. In the presence of hnRNP A2/B1, SL-Au₁ undergoes a register shift, exposing three cytosine residues. These three cytosine residues are able to partake in long-range Watson-Crick base pairing with SL-Ax₂, thereby promoting translational stop-codon readthrough.
Extended Data Figure 1: VEGF-A readthrough does not rely on local enhancers and DMS reactivities of the readthrough elements (a) Readthrough experiments indicate that the addition of sequences upstream of the Ax-element did not lead to increased levels in readthrough, ruling out the potential effects of local enhancers. Error bars indicate standard error (n=3). DMS-MaPseq data show the predicted secondary structures of the Ax-element (b), the alternate conformations of SL-Au₁ (c), and the alternate conformations of the second upstream signal, Au₂ (d).
Extended Data Figure 2: NMR data confirm the predicted secondary structure of SL-Ax₃.

(a) H-H NMR data at 311K show that SL-Ax₃ assumes the fold predicted by DMS-MaPseq. (b) H-H NMR data (311 K) show that G₆₀₉GG in SL-Ax₂ stack on top of each other and (c) HMQC data (311 K) show the upfield shift of the A₆₁₂H₁ due to stacking on the ribose ring of G₆₁₅. (d) Imino proton NMR assignments at 280 K for SL-Ax₃ were obtained and confirm the secondary structure prediction and show that the A₆₃₂A dinucleotide is extruded out from the stem.
Extended Data Figure 3: NMR imino data confirm the predicted secondary structure of SL-Au₁.

The complete imino proton NMR assignments (280 K) for SL-Au₁ in the 1c conformation were obtained. Imino assignments are colored to match the schematic of the secondary structure, which was subdivided into helices 1a (blue), 1b (red), 1c (grey), and 1d (green).
Extended Data Figure 4: NMR data confirm the predicted structures of SL-Au helices. Shortened RNA constructs were used to ascertain the structure of the full SL-Au element. 1H-1H NMR data (311 K) were collected to assign the peaks for helices (a) 1a-1b and (b) 1b-1c.
Extended Data Figure 5: NMR data reveal the existence of alternate structures of SL-Au helices that allow for long-range interactions. (a) H-H NMR data (311 K) were collected to assign the peaks for the alternate conformation of helix 1c using a shortened helices 1b-1c’ construct. (b) Imino proton NMR assignments (298 K) for helices 1c’-1d (black) along with overlays of free SL-Ax (green) and the SL-Ax:1c’-1d complex (red) shows the formation of long-range basepairing involving G₆₀GG of SL-Ax.
Extended Data Figure 6: NMR data confirm the presence of an equilibrium between helices 1c and 1c' of SL-Au. (a) Multiple sets of signals are present in H-H NMR data (280 K) at 15 mM KCl (black) and 50 mM KCl (blue), showing the salt dependence of the 1c to 1c' equilibrium. Overlays with shortened constructs 1b-1c' (purple) and 1c'-1d (green) show that the 50 mM KCl condition favors the 1c' conformation. (b) The 1c to 1c' equilibrium (280 K) is also pH dependent, showing distinct peaks at pH 5.6 (black) and pH 7.3 (red).
**Extended Data Figure 7: Manipulations to the VEGF-A coding sequence maintain global mRNA structure and reduce hnRNP A2/B1 affinity.** (a) DMS reactivities for SL-Au and the Ax-element were collected for various VEGF-A mutant constructs to show there was no major changes in global fold. Mutations are indicated in red boxes while truncations are indicated by omitted nucleotides. U and G bases (grey) show no DMS reactivity while A and C bases may have low (black), medium (orange), or high (red) DMS reactivity. (b) ITC was used to determine a difference in binding affinity for RRM1/2 to native SL-Ax and SL-Ax with an A₁₅:A:UU mutation. To obtain a monophasic binding curve for the mutant, a phosphate buffer was used for this comparison. (c) An ITC binding curve was collected for RRM1/2 binding to SL-Au with an A₁₅:A:UU mutation in the same HEPES buffer as in Figure 3b. For these ITC experiments, representative curves of 3 trials are shown.
Acknowledgements

This work was supported by HHMI grant 55108516 (to V.M.D’S.) and NSF-GRFP grant 1144152 (to N.O.W.). SAXS data was collected at the Advanced Light Source (ALS), SIBYLS beamline on behalf of US DOE-BER, through the Integrated Diffraction Analysis Technologies (IDAT) program. Additional support comes from the NIGMS project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483.

Accession codes

Atomic coordinates have been deposited in the Protein Data Bank under accession codes PDB 7KUB (SL-Au1), PDB 7KUC (SL-Ax1), and PDB 7KUD (SL-Ax2). Chemical shifts have been deposited in the Biological Magnetic Resonance Data Bank under accession codes 30816 (SL-Au1), 30817 (SL-Ax1), and 30818 (SL-Ax2).

Author contributions

N.O.W., J.M.E., and V.M.D’S. conceived and designed the experiments. N.O.W. performed in-cellulo and in-vitro functional assays. N.O.W. and J.M.E. purified the samples and performed the NMR, ITC, and SAXS experiments. J.A. performed NMR experiments on SL-Ax1. N.O.W., J.M.E., J.A. and V.M.D’S. performed the structural analyses. P.G., H.S., and S.R. performed the DMS-MaPseq experiments and analysis. N.W., J.M.E., and V.M.D’S. interpreted the data and wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.
1. Schueren, F. & Thoms, S. Functional Translational Readthrough: A Systems Biology Perspective. PLoS Genet 12, e1006196 (2016).
2. Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R. & Weissman, J.S. Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. Elife 2, e01179 (2013).
3. Schueren, F. et al. Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. Elife 3, e03640 (2014).
4. Stiebler, A.C. et al. Ribosomal readthrough at a short UGA stop codon context triggers dual localization of metabolic enzymes in Fungi and animals. PLoS Genet 10, e1004685 (2014).
5. Loughran, G. et al. 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 (2018).
6. Brierley, I. et al. An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J 6, 3779-85 (1987).
7. Yoshinaka, Y., Katoh, I., Copeland, T.D. & Oroszlan, S. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc Natl Acad Sci U S A 82, 1618-22 (1985).
8. Steneberg, P. & Samakovlis, C. A novel stop codon readthrough mechanism produces functional Headcase protein in Drosophila trachea. EMBO Rep 2, 593-7 (2001).
9. Curran, J.F. & Yarus, M. Use of tRNA suppressors to probe regulation of Escherichia coli release factor 2. J Mol Biol 203, 75-83 (1988).
10. Jungreis, I. et al. Evolutionary Dynamics of Abundant Stop Codon Readthrough. Mol Biol Evol 33, 3108-3132 (2016).
11. Eswarappa, S.M. et al. Programmed translational readthrough generates antiangiogenic VEGF-Ax. Cell 157, 1605-18 (2014).
12. Xin, H., Zhong, C., Nudleman, E. & Ferrara, N. Evidence for Pro-angiogenic Functions of VEGF-Ax. Cell 167, 275-284 e6 (2016).
13. Dever, T.E., Dinman, J.D. & Green, R. Translation Elongation and Recoding in Eukaryotes. Cold Spring Harb Perspect Biol 10(2018).
14. Gesteland, R.F. & Atkins, J.F. Recoding: dynamic reprogramming of translation. Annu Rev Biochem 65, 741-68 (1996).
15. Bidou, L., Rousset, J.P. & Namy, O. Translational errors: from yeast to new therapeutic targets. FEMS Yeast Res 10, 1070-82 (2010).
16. Firth, A.E. & Brierley, I. Non-canonical translation in RNA viruses. J Gen Virol 93, 1385-409 (2012).
17. 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).
18. Barry, J.K. & Miller, W.A. A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. Proc Natl Acad Sci U S A 99, 11133-8 (2002).
19. Cimino, P.A., Nicholson, B.L., Wu, B., Xu, W. & White, K.A. Multifaceted regulation of translational readthrough by RNA replication elements in a tombusvirus. *PLoS Pathog* **7**, e1002423 (2011).

20. Xu, Y. et al. A Stem-Loop Structure in Potato Leafroll Virus Open Reading Frame 5 (ORF5) Is Essential for Readthrough Translation of the Coat Protein ORF Stop Codon 700 Bases Upstream. *J Virol* **92**(2018).

21. Napthine, S. et al. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat Commun* **8**, 15582 (2017).

22. Napthine, S. et al. A novel role for poly(C) binding proteins in programmed ribosomal frameshifting. *Nucleic Acids Res* **44**, 5491-5503 (2016).

23. Leung, D.W., Cachianes, G., Kuang, W.J., Goeddel, D.V. & Ferrara, N. Vascular endothelial growth factor is a secreted angiogenic mitogen. *Science* **246**, 1306-9 (1989).

24. Melincovici, C.S. et al. Vascular endothelial growth factor (VEGF) - key factor in normal and pathological angiogenesis. *Rom J Morphol Embryol* **59**, 455-467 (2018).

25. Loughran, G., Howard, M.T., Firth, A.E. & Atkins, J.F. Avoidance of reporter assay distortions from fused dual reporters. *RNA* **23**, 1285-1289 (2017).

26. Greentzmann, G., Ingram, J.A., Kelly, P.J., Gesteland, R.F. & Atkins, J.F. A dual-luciferase reporter system for studying recoding signals. *RNA* **4**, 479-86 (1998).

27. Cridge, A.G., Crowe-McAuliffe, C., Mathew, S.F. & Tate, W.P. Eukaryotic translational termination efficiency is influenced by the 3' nucleotides within the ribosomal mRNA channel. *Nucleic Acids Res* **46**, 1927-1944 (2018).

28. Zembradt, M. et al. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* **14**, 75-82 (2017).

29. Vander Meulen, K.A., Davis, J.H., Foster, T.R., Record, M.T., Jr. & Butcher, S.E. Thermodynamics and folding pathway of tetraloop receptor-mediated RNA helical packing. *J Mol Biol* **384**, 702-17 (2008).

30. Hansen, J.L., Schmeing, T.M., Moore, P.B. & Steitz, T.A. Structural insights into peptide bond formation. *Proc Natl Acad Sci U S A* **99**, 11670-5 (2002).

31. Brown, J.A. Unraveling the structure and biological functions of RNA triple helices. *Wiley Interdiscip Rev RNA* **11**, e1598 (2020).

32. Jain, N., Morgan, C.E., Rife, B.D., Salemi, M. & Tolbert, B.S. Solution Structure of the HIV-1 Intron Splicing Silencer and Its Interactions with the UP1 Domain of Heterogeneous Nuclear Ribonucleoprotein (hnRNP) A1. *J Biol Chem* **291**, 2331-44 (2016).

33. Kooshapur, H. et al. Structural basis for terminal loop recognition and stimulation of pri-miRNA-18a processing by hnRNP A1. *Nat Commun* **9**, 2479 (2018).

34. Morgan, C.E. et al. The First Crystal Structure of the UP1 Domain of hnRNP A1 Bound to RNA Reveals a New Look for an Old RNA Binding Protein. *J Mol Biol* **427**, 3241-3257 (2015).

35. Tolbert, M. et al. HnRNP A1 Alters the Structure of a Conserved Enterovirus IRES Domain to Stimulate Viral Translation. *J Mol Biol* **429**, 2841-2858 (2017).

36. Wu, B. et al. Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1. *Nat Commun* **9**, 420 (2018).

37. Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. & Burd, C.G. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **62**, 289-321 (1993).
38. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum Genet* **135**, 851-67 (2016).

39. Han, S.P., Tang, Y.H. & Smith, R. Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J* **430**, 379-92 (2010).

40. Kwak, H., Park, M.W. & Jeong, S. Annexin A2 binds RNA and reduces the frameshifting efficiency of infectious bronchitis virus. *PLoS One* **6**, e24067 (2011).

41. Gao, F. & Simon, A.E. Multiple Cis-acting elements modulate programmed -1 ribosomal frameshifting in Pea enation mosaic virus. *Nucleic Acids Res* **44**, 878-95 (2016).

42. Wu, M.T. & D’Souza, V. Alternate RNA Structures. *Cold Spring Harb Perspect Biol* **12**(2020).

43. Vicens, Q., Kieft, J.S. & Rissland, O.S. Revisiting the Closed-Loop Model and the Nature of mRNA 5'-3' Communication. *Mol Cell* **72**, 805-812 (2018).

44. Yoffe, A.M., Prinsen, P., Gelbart, W.M. & Ben-Shaul, A. The ends of a large RNA molecule are necessarily close. *Nucleic Acids Res* **39**, 292-9 (2011).

45. Lai, W.C. et al. mRNAs and IncRNAs intrinsically form secondary structures with short end-to-end distances. *Nat Commun* **9**, 4328 (2018).

46. Adivarahan, S. et al. Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. *Mol Cell* **72**, 727-738 e5 (2018).

47. Khong, A. & Parker, R. mRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. *J Cell Biol* **217**, 4124-4140 (2018).

48. Harper, S. & Speicher, D.W. Purification of proteins fused to glutathione S-transferase. *Methods Mol Biol* **681**, 259-80 (2011).

49. D’Souza, V., Dey, A., Habib, D. & Summers, M.F. NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol* **337**, 427-42 (2004).

50. Norris, M., Fetler, B., Marchant, J. & Johnson, B.A. NMRfx Processor: a cross-platform NMR data processing program. *J Biomol NMR* **65**, 205-216 (2016).

51. Guntert, P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-78 (2004).

52. Schwieters, C.D., Kuszewski, J.J., Tjandra, N. & Clore, G.M. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**, 65-73 (2003).

53. Schwieters, C.D. & Clore, G.M. Using small angle solution scattering data in Xplor-NIH structure calculations. *Prog Nucl Magn Reson Spectrosc* **80**, 1-11 (2014).

54. Rouskin, S., Zuberadt, M., Washietl, S., Kellis, M. & Weissman, J.S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701-5 (2014).

55. Dyer, K.N. et al. High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol Biol* **1091**, 245-58 (2014).

56. Forster, S., Apostol, L. & Bras, W. Scatter: software for the analysis of nano- and mesoscale small-angle scattering. *Journal of Applied Crystallography* **43**, 639-646 (2010).