

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

**Nico O. Wagner**

Harvard University

**Jerome M. Edwards**

Harvard University

**Juliana Abramovich**

Harvard University

**Paromita Gupta**

Whitehead Institute for Biomedical Research

**Harish Swaminathan**

Whitehead Institute for Biomedical Research

**Silvi Rouskin**

Whitehead Institute for Biomedical Research

**Victoria M. D'Souza** (✉ [dsouza@g.harvard.edu](mailto:dsouza@g.harvard.edu))

Harvard University

---

## Research Article

**Keywords:** VEGF, recoding, translation, NMR, structural biology, RNA

**Posted Date:** February 9th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-186157/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# 1 A novel structural mechanism of ribosomal stop codon readthrough in 2 VEGF-A

3 Nico O. Wagner<sup>1</sup>, Jerome M. Edwards<sup>1</sup>, Juliana Abramovich<sup>1</sup>, Paromita Gupta<sup>2</sup>, Harish  
4 Swaminathan<sup>2</sup>, Silvi Rouskin<sup>2</sup> and Victoria M. D'Souza<sup>1\*</sup>

5 <sup>1</sup>Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA. 02138,  
6 USA.

7 <sup>2</sup>Whitehead Institute for Biomedical Research, Cambridge, MA. 02142, USA.

8 \*Correspondence: dsouza@g.harvard.edu

9

## 10 Abstract

11 The process of ribosomal recoding is generally regulated by an autonomous mRNA signal  
12 downstream of stop-codons. While structural studies have provided mechanistic insights into  
13 viral systems, no such studies exist in mammalian systems. Here we define a novel structural  
14 mechanism for the VEGF-A readthrough system and show that regulation is multifaceted and  
15 complex, requiring a multipartite set of RNA elements located at long distances that interact  
16 with each other and with hnRNP A2/B1 to synergistically enhance readthrough levels. The Ax-  
17 element downstream of the stop codon adopts a unique multistem (SL-Ax<sub>1-3</sub>) architecture: SL-  
18 Ax<sub>1</sub> interacts with hnRNP A2/B1, while SL-Ax<sub>2</sub> interacts with an RNA element (SL-Au<sub>1</sub>) located  
19 ~500 nt upstream at the start of the coding sequence. SL-Au<sub>1</sub> also independently binds to  
20 hnRNP A2/B1, which manipulates an equilibrium between alternate structures— from a  
21 sequestered bulge towards one that allows for the long-range interaction with SL-Ax<sub>2</sub>. Overall,  
22 our study not only highlights the significance of structural organization of elements within  
23 the coding sequence of mRNA, but also provides a functional relevance of the closed-loop  
24 mRNA organization in non-canonical translation and suggests complex mechanisms allow for  
25 finer integration of many signals for a required output.

26

## 27 Introduction

28 Recoding during protein translation, either via programmed ribosomal frameshifting  
29 (PRF) or stop-codon readthrough (PRT), allows for production of extended proteins or  
30 polyproteins by bypassing stop codons. While these mechanisms are widely used by RNA viruses

31 to densely pack information into their small genomes and to maintain relative protein levels<sup>1</sup>,  
32 they have also been discovered in cellular systems, including bacteria, yeast, *Drosophila*, and  
33 humans<sup>1,2</sup>, to regulate a wide range of functions from transcriptional regulation to signal  
34 transduction and sub-cellular localization<sup>3-5</sup>. In viruses, ribosomal recoding is essential for  
35 maintaining critical relative ratios of proteins, for example structural (Gag) and enzymatic (Pol)  
36 in retroviruses and immunomodulatory proteins (in ORF 1a) and replicative proteins (in ORF 1b)  
37 in coronaviruses<sup>6,7</sup>. Similarly, bacterial release factor 2 expression levels are auto-regulated via  
38 recoding and in *Drosophila*, recoding of the headcase protein is critical for tracheal  
39 development<sup>8,9</sup>.

40

41 Studies in viruses have provided some insights into structural elements that drive  
42 recoding. However, no such insights currently exist for eukaryotic recoding systems, even though  
43 over 300 genes in *Drosophila* and ~100 genes in humans have been identified to potentially  
44 undergo functional recoding<sup>1,10</sup>. Very recently, a new isoform of the mammalian vascular  
45 endothelial growth factor A (VEGF-A), produced via readthrough, was implicated in regulating  
46 levels of angiogenesis<sup>11,12</sup>, indicating that recoding may play a role in regulating critical cellular  
47 processes, highlighting the need to understand the underlying mechanisms.

48

49 Canonical protein translation requires a complex interplay between ribosomes, mRNA  
50 and regulatory protein factors<sup>13</sup> but, in general, is independent of structures formed in the mRNA.  
51 On the other hand, recoding mechanisms are heavily reliant upon *cis*-acting RNA structures, such  
52 as stem-loops, kissing loops, or pseudoknots, around stop codons<sup>14-16</sup>. However, the detailed  
53 mechanisms by which such structured elements regulate recoding remain undefined. In the case  
54 of readthrough, recoding signals are thought to stall ribosomes by acting as barriers to mRNA  
55 unwinding, thus allowing for the recognition of near-cognate tRNAs during PRT<sup>17</sup>. The strained  
56 unwinding also potentially acts as a mechanical stressor to force ribosomes into a new reading  
57 frame during PRF<sup>15,17</sup>.

58

59           The *cis*-acting, structured, recoding signals are mostly known to function autonomously.  
60 Nevertheless, in some cases their folding can be manipulated by surrounding sequences to either  
61 attenuate or enhance activity, and in rare cases they can be influenced by interaction with distal  
62 downstream sequences. For example, potato leafroll virus (PLRV), carnation Italian ringspot  
63 virus (CIRV), and barley yellow dwarf virus (BYDV) were all found to contain 3' elements,  
64 located 700nt, 3.5kb, and 4kb, respectively, from the frameshifting signals that enhance recoding  
65 via long-range base pairing interactions<sup>18-20</sup>. In even rarer cases, *trans*-acting protein factors have  
66 been implicated in recoding. For example, the encephalomyocarditis virus (EMCV) has been  
67 shown to require proteins A2 and porcine reproductive and respiratory syndrome virus (PRRSV)  
68 to require nsp1 $\beta$  and poly(C) binding protein for proper recoding to occur<sup>21,22</sup>. It has been  
69 suggested that these proteins, in complex with the RNA, substitute for the barriers usually  
70 provided by structured signals<sup>18</sup>.

71  
72           VEGF-A mRNA was shown to undergo a stop codon readthrough event to allow for a 22  
73 amino acid extension (Fig. 1a)<sup>11</sup>. While VEGF-A<sub>164</sub>, the most common isoform, acts as a key  
74 angiogenic agent<sup>23,24</sup>, the extended VEGF-Ax has been shown to be significantly less  
75 angiogenic<sup>11,12</sup>. In the VEGF system, the 63 nucleotides which code for the Ax-extension  
76 immediately downstream of the stop codon, were implicated as the signal required for ribosomal  
77 stop-codon readthrough<sup>11</sup>. Furthermore, it was reported that the event also requires the  
78 heterogeneous nuclear ribonucleoprotein (hnRNP) A2/B1 as a *trans*-acting factor<sup>11</sup> in that  
79 mutations to the hnRNP A2/B1 binding site on the Ax-element and siRNA knockdowns of the  
80 protein itself lead to lower levels of stop-codon readthrough. It is important to note that a  
81 subsequent study by Loughran, *et al.* failed to replicate the same read through capability of the  
82 Ax-element<sup>25</sup>. Here we show that efficient readthrough in VEGF-Ax mRNA requires full-length  
83 mRNA, with multiple *cis*-acting signals in the coding region working synergistically with each  
84 other and with the *trans*-acting hnRNP A2/B1 to confer readthrough activity.

## 85 **Results**

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

87 To characterize the VEGF-A system, we used a dual luciferase reporter assay in which  
88 various sequences of interest were cloned between renilla and firefly luciferases to determine stop  
89 codon readthrough efficiency<sup>26</sup>. First, the natural leakiness of the UGA stop codon was  
90 determined to be ~0.05%, which is in agreement with previous studies<sup>27</sup> (Fig. 1b). Second, we  
91 tested the full bicistronic coding sequence of the VEGF-A mRNA (VEGF-A+Ax), where we  
92 observed ~40-fold increase in readthrough levels to ~2.2%, a level similar to many functionally  
93 active recoding systems<sup>1</sup>. To ensure that the observed increase over the natural leakiness of the  
94 UGA stop codon was robust and specific, we tested Annexin A2 mRNA (Fig. 1b), a gene with no  
95 known readthrough activity. This construct also only allowed for ~0.05% basal leakiness of the  
96 stop codon, demonstrating that VEGF-A mRNA indeed has the propensity for stimulating  
97 readthrough events. Third, we tested just the 63 nucleotide Ax-element downstream of the stop  
98 codon that was previously identified as the signal being responsible for readthrough activity<sup>11</sup>.  
99 This region only led to a 4-fold increase over natural leakiness with a 0.2% readthrough activity  
100 (Fig. 1b), which is similar to the results obtained by Loughran, *et al.*<sup>25</sup>. Taken together these data  
101 indicated that sequences upstream in the coding region influence the process of readthrough in  
102 VEGF-Ax, and that the Ax-element alone is only marginally active for readthrough.

103

104 In an effort to identify factors responsible for the above observations, we first probed for  
105 the presence of potential local enhancers by adding 50, 100 or 200 nucleotides upstream of the  
106 Ax-element (Extended Data Fig. 1a). This did not result in increasing the readthrough levels to  
107 that of the whole coding sequence, thus implying that there are no local enhancer sequences for  
108 the Ax-element, and that sequences very distal and upstream to the stop codon enhance  
109 readthrough. We then created truncated constructs in which nucleotides were removed from the  
110 5' end of the mRNA. Surprisingly, deletion of the first 100 nucleotides dropped the readthrough  
111 activity by half (~0.94% readthrough), with a further drop to Ax-only levels at the 250-nucleotide  
112 deletion (Fig. 1b). Interestingly, sequence analysis of the upstream region within the first 100 nt  
113 presented another potential hnRNP A2/B1 binding sequences (A<sub>93</sub>GGAGG). Altogether, this set

114 of data demonstrate that the full, bicistronic VEGF-A mRNA is necessary for efficient  
115 readthrough to occur and that the Ax-element is incapable of causing significant readthrough  
116 without the presence of multiple signals in the coding region: one within the first 100 nucleotides  
117 with a potential to bind hnRNP A2/B1 (SL-Au<sub>1</sub>), and the second between 200-250 nucleotides  
118 (Au<sub>2</sub>).

119

### 120 **DMS-MaPseq reveals alternate structures in SL-Au<sub>1</sub> and multi-domain organization of the Ax-** 121 **element.**

122 To gain structural insights into the readthrough process, we chemically probed<sup>28</sup> the  
123 bicistronic native mRNA from the cells used for the luciferase assay by DMS-MaPseq. Overall,  
124 the Ax-element folds into three, short stem loops (SL-Ax<sub>1-3</sub>) with five and four nucleotide linker  
125 sequences in between them, respectively. This architecture is unusual in that most recoding  
126 signals are made up of pseudoknots or single stem loops (Fig. 1c and Extended Data Fig. 1b). In  
127 addition, the first seven nucleotides of the Ax-element form an additional stem loop with ten  
128 nucleotides upstream of the stop codon. SL-Ax<sub>1</sub> contains the previously identified hnRNP A2/B1  
129 binding site with one of the consensus A<sub>588</sub>GG motifs predicted to be positioned in the loop<sup>11</sup>.

130

131 DMS mapping of the 5' end of the mRNA showed that only residues 43-102 give rise to  
132 ensembles that converge into a defined fold, which we term SL-Au<sub>1</sub>. SL-Au<sub>1</sub> was predicted to  
133 form a long stem loop capped with a CCA triloop and has four short helices (1a-1d) interspersed  
134 with bulges (Fig. 1c and Extended Data Fig. 1c). In some structures, helix 1c either folded into a  
135 stem with a register shift (helix 1c') or did not form at all and instead configured into a large  
136 bulge, suggesting that helix 1c could possibly sample alternate arrangements (Extended Data Fig.  
137 1c). The potential hnRNP A2/B1 binding sequence spans both helix 1a and 1b at the 3' end of SL-  
138 Au<sub>1</sub>. Finally, characterization of Au<sub>2</sub> (residues 200-250) also shows the potential to fold into  
139 structured elements. However, the configurations of the structures in the ensembles varied  
140 significantly, with sequences placed in a stem in one configuration ending up in a loop in the  
141 other (Extended Data Fig. 1d); for this study, we thus only focused on the roles of SL-Au<sub>1</sub> and the  
142 Ax-element in stop codon readthrough.

143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171

#### **Ax and Au<sub>1</sub> elements make long-range interactions.**

We next synthesized the individual RNA domains by *in-vitro* transcription. Given that we observed an effect of the Au<sub>1</sub> 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-Au<sub>1</sub> into SL-Ax<sub>3</sub> did not give rise to any binding interactions, titration of SL-Ax<sub>2</sub> into SL-Au<sub>1</sub> gave rise to specific enthalpically-driven heats of binding between the two domains ( $K_d = 0.95 \pm 0.53 \mu\text{M}$ ) 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  $\sim 5.5 \mu\text{M}$ <sup>29</sup>. 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-Au<sub>1</sub> and SL-Ax<sub>2</sub> 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-Ax<sub>1</sub>, the 5-bp stem predicted by DMS-MaPseq is extended by an additional non-canonical A<sub>587</sub> o G<sub>592</sub> base pair, thus resulting in a four-nucleotide A<sub>588</sub>GGA tetraloop (Fig. 1f and Extended Data Fig. 2a). The first two residues (A<sub>588</sub> and G<sub>589</sub>) of the loop are in a stacked configuration with the 5' nucleotides of the helix, with G<sub>589</sub> in a *syn*-conformation. After chain reversal, residue G<sub>590</sub> also adopts a *syn*-conformation, stacking on A<sub>591</sub>, which further stacks continuously with the 3' nucleotides of the helix. Thus, for the two consensus AGG hnRNP A2/B1 binding sequences, A<sub>584</sub>GG is completely sequestered by base pairings in the stem, while the A<sub>588</sub>GG nucleotides are structured in an unusual tetraloop.

172 As predicted, SL-Ax<sub>2</sub> forms a four base pair stem capped with a U<sub>607</sub>CGGG pentaloop. The  
173 first two residues of the pentaloop, U<sub>607</sub> and C<sub>608</sub>, continue to stack on the 5' strand of the stem  
174 base pairs, after which the chain turns with the following three residues G<sub>609</sub>GG showing  
175 continuously stacking NOES with the 3' strand of the helix (Fig 1g and Extended Data Fig. 2b).  
176 As in the GNRA-type fold, the nucleobase of G<sub>611</sub> stacks on the ribose of A<sub>612</sub> leading to an  
177 expected upfield shift of the A<sub>612</sub> H1' proton to 4.5 ppm (Extended Data Fig. 2c). This positions  
178 the Watson-Crick faces of the three contiguous G residues outside the loop. Such a loop  
179 configuration where three stacked G residues engage in Watson-Crick interactions with  
180 contiguous cytosines present at a long-distance has been observed in the *Haloarcula marismortui*  
181 ribosomal RNA<sup>30</sup>. Thus, our structure of SL-Ax<sub>2</sub> indicate that it may be poised for the long-range  
182 interactions with structures present in Au<sub>1</sub> as evidenced by ITC analysis.

183

184 Finally, the imino proton assignments of SL-Ax<sub>3</sub> indicate that it is composed of a five base  
185 pair stem-loop with a dinucleotide A<sub>632</sub>A bulge, as predicted by DMS-MaPseq. Continuous  
186 stacking of the five base pairs was observed, indicating that the AA bulge is extruded out of the  
187 groove (Extended data Fig. 2d). However, dimerization interactions of the loop precluded us  
188 from unambiguously assigning and solving the high-resolution structure of SL-Ax<sub>3</sub>.

189

### 190 **Structure of the Au<sub>1</sub> element.**

191 To aid in assignments of the relatively large Au<sub>1</sub> construct and to unambiguously  
192 corroborate the presence of the equilibrium structure predicted in the SL-Au<sub>1</sub> element, we  
193 synthesized smaller segments of Au<sub>1</sub> (1a-1b, 1b-1c, 1b-1c', 1c'-1d, 1c-1d and 1d; along with the  
194 full-length construct (Extended Data Fig. 3-6). Indeed, our data show that SL-Au<sub>1</sub> forms a single  
195 stem with a CCA loop and is interspersed with bulges that divide the stem into four short helices,  
196 1a-1d (Extended Data Fig. 3). Furthermore, as predicted, two sets of signals were assigned for the  
197 region encompassing residues 56-64 and 81-90, which confirmed an equilibrium between helix 1c  
198 and 1c' (Fig. 1c and Extended Data Fig. 6a).

199

200 The structures formed by helices 1a, 1b and 1d do not change in the two configurations  
201 adopted by SL-Au<sub>1</sub> (Extended Data Table 1). Separated by a single adenine nucleotide bulge (A<sub>50</sub>),  
202 the distal helices 1a and 1b form an almost continuous stem (Fig. 1e and Extended Data Fig. 4a,  
203 4b). A<sub>50</sub> is part of the helix with regular stacking between residues U<sub>49</sub> and C<sub>51</sub>. In this  
204 configuration, the two consensus hnRNP A2/B1 binding sequences, A<sub>93</sub>GG and A<sub>96</sub>GG, are all  
205 involved in base pairing interactions within helix 1b and 1a, respectively (Extended Data Fig. 3).  
206 Furthermore, the proximal helix 1d is rich in GU basepairs (U<sub>67</sub>-G<sub>79</sub>, G<sub>68</sub>-U<sub>78</sub> and U<sub>70</sub>-G<sub>76</sub>), which  
207 make up half of the six base pairs in the helix as evidenced by the three upfield shifted GU cross-  
208 peak pairs in the imino region of the NOESY spectrum (Fig. 1e and Extended Data Fig. 5b). The  
209 C<sub>72</sub>CA triloop which caps this helix is well structured, with chain reversal occurring between  
210 residues C<sub>72</sub> and C<sub>73</sub> and stacking with both sides of the helix (Fig. 2a).

211  
212 The central helix adopts two different configurations (helix 1c and 1c') (Extended Data  
213 Fig. 6a). Helix 1c shows the formation of an 8-bp stem (residues 57-64 and 81-88) including two  
214 tandem, non-canonical C-C base pairs that are protonation dependent, making the equilibrium  
215 between helix 1c and 1c' highly sensitive to pH (Fig. 2b and Extended Data Fig. 5b). In the 1c  
216 configuration, the junction between helices 1d and 1c also has an almost continuously stacking  
217 architecture with a single adenine nucleotide bulge (A<sub>65</sub>) stacked inside the helix. On the other  
218 end, the junction between 1c and 1b is defined by an internal loop (A<sub>56</sub>; C<sub>89</sub>A). While A<sub>56</sub> stacks  
219 with the helix (Extended Data Fig. 4b), the opposing residue A<sub>90</sub> is extruded out toward the minor  
220 groove. The H4' sugar proton of residue C<sub>89</sub> gives NOEs to the aromatic H8 proton of A<sub>92</sub>, placing  
221 it in the major groove and within hydrogen bonding distance of U<sub>54</sub>-A<sub>92</sub> (Fig. 2c). This local triple-  
222 base configuration obtained is similar to arginine sandwich motifs<sup>31</sup>. Overall, with continuous  
223 stacking of all helices in the helix 1c configuration, the structure of SL-Au<sub>1</sub> has a linear appearance  
224 (Fig. 1e).

225  
226 On the other hand, due to the base pairing register shift in the helix 1c' configuration,  
227 residues A<sub>56</sub> to A<sub>59</sub> and A<sub>81</sub> to A<sub>85</sub> form two opposing bulges separated by a 5-bp stem (residues  
228 60-65 and 86-90). These four (A<sub>56</sub>CCA) and five (A<sub>81</sub>CCCA) nucleotide bulges flanking helix 1c'

229 cause bent geometries between junctions with helices 1b and 1d, respectively (Fig. 2d). At the 1b-  
230 1c' junction, residues A<sub>56</sub> and C<sub>57</sub> stack on helix 1b, whereas residues C<sub>58</sub> and A<sub>59</sub> stack under helix  
231 1c' (Fig. 2e). There are no NOEs between C<sub>57</sub> and C<sub>58</sub>, indicating opposing orientations. At the 1c'-  
232 1d junction, the A<sub>81</sub> in the A<sub>81</sub>CCCA bulge base pairs with the A<sub>65</sub> in the opposite strand (Fig. 2f).  
233 The following three cytosine residues show a continuous NOE walk. With only a partial stacking  
234 between C<sub>84</sub> and A<sub>84</sub>, the triplet Cs are extruded out in way that seem poised to make long-range  
235 interaction with SL-Ax<sub>2</sub>, as indicated by the ITC data (Fig. 1d, 2d).

236

### 237 **SL-Ax<sub>2</sub> and SL-Au<sub>1</sub> interact via loop-bulge docking interactions**

238 To understand the mode of the long-range interaction, we performed small angle X-ray  
239 scattering (SAXS) analysis of free SL-Au<sub>1</sub> and in complex with SL-Ax<sub>2</sub>. Furthermore, since the  
240 C<sub>82</sub>CC motif in SL-Au<sub>1</sub> and G<sub>609</sub>GG motif in SL-Ax<sub>2</sub> resemble a motif found in *Haloarcula*  
241 *marismortui* ribosomal RNA where the three stacked G residues engage in Watson-Crick  
242 interactions with contiguous cytosines<sup>30</sup>, we also modeled a similar interaction between SL-Au<sub>1</sub>  
243 and SL-Ax<sub>2</sub> (Fig. 2g). Compared to the free form, the SAXS envelop of the bound form of SL-Au<sub>1</sub>  
244 showed a clear additional density, which was able to accommodate the structural model of SL-  
245 Ax<sub>2</sub> in the complex (Fig. 2g).

246 To unambiguously assign the docking interface, we titrated SL-Ax<sub>2</sub> in the 1c'-1d construct,  
247 which allowed for identification of the A<sub>81</sub>-C<sub>84</sub> bulge as the one involved in the long-range  
248 interaction (Extended Data Fig. 5b). While no global changes in the secondary structure of either  
249 1c'-1d or SL-Ax<sub>2</sub> occurred, resonances for three new G-C Watson-Crick base pair formations were  
250 observed, supporting complex formation observed by ITC (Fig. 1d). The combination of binding  
251 heats observed by ITC, intermolecular base pairing NOEs observed by NMR, and the complex  
252 obtained by SAXS all support the formation of the long-range interaction between the SL-Au<sub>1</sub>  
253 bulge and SL-Ax<sub>2</sub> loop.

254

### 255 **hnRNP A2/B1 binds SL-Ax<sub>1</sub> and SL-Au<sub>1</sub>**

256 We next wanted to test if the previously identified hnRNP A2/B1 binding sites (A<sub>584</sub>GG  
257 and A<sub>588</sub>GG in SL-Ax<sub>1</sub>) and the new potential sites (A<sub>93</sub>GG and A<sub>96</sub>GG in helix 1a and 1b,

258 respectively)<sup>11</sup> interacted with hnRNP A2/B1. It was previously reported that mutating the  
259 A<sub>588</sub>GG hnRNP A2/B1 binding site led to decreased readthrough levels<sup>11</sup>. Similar to previous  
260 hnRNP binding studies<sup>32-36</sup>, we used an hnRNP construct that consisted of the two RNA-  
261 recognition motifs (RRMs), but lacked the aggregation-prone C-terminal domain, previously  
262 shown to be involved in oligomerization and nuclear localization of the protein<sup>37-39</sup> (Fig. 3a).  
263 Titration of hnRNP A2/B1 into SL-Ax<sub>1</sub> and SL-Au<sub>1</sub> by ITC gave data that fit well with a single-site  
264 binding model with estimated dissociation constants of  $200 \pm 42$  nM and  $348.5 \pm 22$  nM,  
265 respectively ( $n=1.02 \pm 0.34$  and  $1.2 \pm 0.33$ , respectively) (Fig. 3b). These  $K_d$ 's are in accordance with  
266 previously published binding affinity of similar sequences<sup>36</sup>. Nevertheless, a critical difference in  
267 the binding mode was observed between the two interactions: while the interaction with SL-Ax<sub>1</sub>  
268 was enthalpically driven ( $\Delta H -10.4 \pm 4.7$  kcal/mol and  $\Delta S -4.4 \pm 15.3$  cal/mol/K), the interaction  
269 with SL-Au<sub>1</sub> ( $\Delta H -3.6 \pm 1.7$  kcal/mol and  $\Delta S 17.6 \pm 5.7$  cal/mol/K) was entropically driven, the latter  
270 suggesting rearrangement of RNA structure upon protein binding.

271  
272 Since hnRNP A2/B1 bound both SL-Au<sub>1</sub> and SL-Ax<sub>1</sub>, and since it is able to dimerize  
273 through its C-terminal domain<sup>38</sup>, we tested if it is able to mediate the long-range interaction  
274 between SL-Au<sub>1</sub> and SL-Ax<sub>1</sub> via its dimerization domain. Using an *in-vitro* rabbit reticulocyte  
275 lysate (RLL) system, we tested the effects of full-length hnRNP A2/B1 and hnRNP A2/B1 lacking  
276 a C-terminal domain ( $\Delta$ CTD) on VEGF-A readthrough efficiencies (Fig. 3c). Recoding was  
277 replicated in this assay, with baseline readthrough values of ~4%, which is slightly higher than  
278 those in cells. Such differences have previously been reported<sup>40</sup>, and generally arise from tighter  
279 regulation in cells. By adding either full-length hnRNP A2/B1 or hnRNP A2/B1- $\Delta$ CTD to VEGF-  
280 A mRNA, we are able to increase readthrough by a factor of ~2.5 (Fig. 3c) to 9.6%, and further  
281 additions of hnRNP A2/B1 reproducibly reduced readthrough. These results not only confirm the  
282 stimulatory effect of hnRNP A2/B1 binding, but also point to other hnRNP A2/B1 driven  
283 mechanisms to balance required readthrough levels. Importantly, there was no observable  
284 difference in readthrough levels between the full-length and dimerization-deficient hnRNP  
285 A2/B1 constructs, indicating that hnRNP A2/B1 dimerization is not required for readthrough  
286 activity, and that another mechanism must be responsible for mediating the observed long-range

287 interactions. As a control, using the Murine Leukemia Virus (MLV) readthrough system resulted  
288 in constant readthrough levels, showing that the effects of hnRNP A2/B1 are specific to VEGF-A  
289 (Fig 3c).

290

291 To determine the exact binding site within SL-Au<sub>1</sub>, we performed a protein titration on  
292 our NMR sample. Addition of even sub-stoichiometric (0.3 equivalents) of RRM1/2 to SL-Au<sub>1</sub>  
293 resulted in pushing the equilibrium to the 1c' configuration, as evidenced by the disappearance  
294 of the U<sub>60</sub>-A<sub>85</sub> and A<sub>59</sub>-U<sub>86</sub> base pairing in helix 1c and the appearance of A<sub>64</sub>-U<sub>86</sub> in helix 1c' (Fig.  
295 3d). Furthermore, addition of one equivalent of protein led to chemical shift changes both in the  
296 protein and RNA and gave rise to multiple intermolecular NOEs. Interestingly, our data show  
297 that binding occurred specifically to molecules in the 1c' configuration, as evidenced by the  
298 specific disappearance of the U<sub>60</sub>-A<sub>90</sub> base pair, which is representative of helix 1c'. Perturbations  
299 of A<sub>93</sub> resonances indicated that binding occurs at the consensus hnRNP A2/B1 binding site (Fig.  
300 3d). Overall, these experiments suggest that the role of hnRNP A2/B1 binding to SL-Au<sub>1</sub> may be  
301 to modulate the equilibrium toward the 1c' conformation, which is required for long-range  
302 interactions.

303

304 Based on recent findings that it is the RRM1 that recognizes the AGG sequence<sup>36</sup>, we used  
305 a shortened hnRNP construct to determine the mode of interaction with SL-Ax<sub>1</sub>. Binding between  
306 RRM1 and SL-Ax<sub>1</sub> resulted in perturbations of specific residues in the loop, with G<sub>590</sub> giving clear  
307 intermolecular NOEs to an aromatic residue of RRM1, thus providing confirmation of residues  
308 A<sub>588</sub>GG sequence as the binding site (Fig. 3e). The second potential A<sub>584</sub>GG binding motif remains  
309 sequestered within the stem as unaffected by protein binding; in fact, the majority of the SL-Ax<sub>1</sub>  
310 stem structure remaining unchanged upon protein addition. Thus, at both the SL-Au<sub>1</sub> and SL-Ax<sub>1</sub>  
311 binding sites, hnRNP A2/B1 recognizes structured loops, as recently described in viral IRES  
312 translation regulation<sup>35</sup>.

313

314 **The various elements synergistically regulate recoding efficiencies**

315           Given both the unusual multi-stem nature of the Ax-element, and potential unique  
316 functions of the individual stems, we wanted to first determine their contributions to recoding.  
317 We created constructs in which SL-Ax<sub>2</sub> and SL-Ax<sub>3</sub> elements were removed (Fig. 4a). We also  
318 checked DMS-MaPseq reactivities to ensure that in either deletion, the fold of SL-Ax<sub>1</sub> is not  
319 perturbed (Extended Data Fig. 7a). Deletions of SL-Ax<sub>3</sub> and both SL-Ax<sub>2</sub> and SL-Ax<sub>3</sub> led to  
320 significant reduction in readthrough levels by 40% and 70%, respectively, implying that the  
321 entirety of the Ax-element is required to maximize readthrough levels (Fig. 4b). Interestingly,  
322 removing the entire Ax-element decreased readthrough levels by almost 80% to a level of 0.5%  
323 (Fig. 4b).

324  
325           While we do not yet understand the structural mechanism by which SL-Ax<sub>3</sub> influences  
326 readthrough, our understanding of the long-range interaction between SL-Ax<sub>2</sub> with SL-Au<sub>1</sub>  
327 allows us to test for its contribution. We thus created mutants in which the C<sub>82</sub>CC involved in  
328 long range interaction in SL-Au<sub>1</sub> was sequestered by either changing them to Gs along with  
329 mutating G<sub>61</sub> to C, or by mutating the opposing C<sub>62</sub>C residues to guanosines. Both of these give  
330 rise to three contiguous G-C base pairs, which should preclude inter-domain interaction of this  
331 bulge and lead to appreciable decreases in readthrough by ~40% and ~50%, respectively (Fig. 4c).  
332 Similarly, disruption of the docking G<sub>609</sub>GG motif in SL-Ax<sub>2</sub> by replacing the pentaloop with a  
333 GAAA tetraloop, led to an equivalent 40% decrease in readthrough activity (Fig. 4c).

334  
335           Next, we wanted to test the contributions of protein binding motifs to recoding  
336 efficiencies. We first tested protein binding efficiency of the previously published dinucleotide  
337 A<sub>587</sub>A to U<sub>587</sub>U mutation in SL-Ax<sub>1</sub><sup>11</sup>. ITC experiments on a SL-Ax<sub>1</sub> A<sub>587</sub>A:UU mutant yielded a  
338 ~1.5-fold decrease in binding when compared to wild-type SL-Ax<sub>1</sub> (Extended Data Fig. 7b). This  
339 correlates with an observed ~50% reduction in readthrough levels (Fig. 4c). Similarly, for  
340 experiments on a SL-Au<sub>1</sub> A<sub>92</sub>A:UU mutant, we observed a ~2-fold increase in K<sub>d</sub>, or 2-fold  
341 decrease in binding affinity (Extended Data Fig. 7c), which correlates to a ~40% decrease in  
342 readthrough levels for this mutation (Fig. 4c), suggesting a strong-interplay between hnRNP  
343 A2/B1 binding and readthrough of the VEGF-Ax system. Interestingly, a double mutant in which

344 both hnRNP A2/B1 binding sites were mutated displayed higher readthrough efficiency than  
345 either single mutant (Fig. 4c). This epistatic behavior is a strong indication of direct influence of  
346 the two binding events on local RNA structure, and consequently their long-range interactions.  
347 Furthermore, it suggests that the closed-loop configuration of VEGF mRNA may place SL-Au1  
348 and SL-Ax2 in close proximity, thus allowing for the observed epistasis. Overall, these mutational  
349 studies show the importance of the various players involved in regulating readthrough levels of  
350 the VEGF mRNA.

## 351 Discussion

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

360  
361 The presence of a 5' element conferring ribosomal readthrough is novel. In retroviruses,  
362 where ribosomal recoding has been studied more extensively, readthrough signals tend to be  
363 localized in the immediately vicinity of the stop codon<sup>14,16</sup>. While distant downstream elements  
364 enhancing frameshifting in Pea enation mosaic virus<sup>41</sup> and readthrough in Potato leafroll virus<sup>20</sup>  
365 have been studied, to the best of our knowledge, there are no examples of distant upstream  
366 signals. Given that these novel *cis*-acting elements, located at ~250 and ~500 nucleotides upstream  
367 of the stop codon, are in the coding region, this is an even more unusual setup for ribosomal  
368 readthrough. The dispersed, multi-domain architecture of the Ax-element is also unique in that  
369 RNA signals that cause recoding in ribosomes are generally made up of a single stem loop or  
370 pseudoknot structures. Such an organization, along with the presence of long-range interactions  
371 between signals located at the two ends of the coding sequence and multiple hnRNP A2/B1  
372 binding sites, hints towards a complex mechanism for regulation of readthrough in mammalian  
373 systems.

374  
375 Altogether, our study allows us to start putting together a mechanism by which  
376 readthrough frequency— and hence relative proportions of VEGF-A (angiogenic) to VEGF-Ax  
377 (less angiogenic) isoforms— may be maintained (Fig. 5). Our model suggests that regulation  
378 takes place via modulations of alternate mRNA structures by hnRNP A2/B1 levels<sup>42</sup>. Binding of  
379 hnRNP A2/B1 to SL-Au<sub>1</sub> allows for a register shift in the base pairing of helix 1c to the 1c'

380 configuration, which is capable of making long-range interactions with SL-Ax<sub>2</sub> near the vicinity  
381 of the stop codon. The two conformations of SL-Au<sub>1</sub> were reproducibly observed in cells, adding  
382 another example of use of conformational equilibrium in RNA molecules to regulate critical  
383 functions in cells. Additionally, while we did not address the roles of SL-Ax<sub>3</sub> or the Au<sub>2</sub> element  
384 ~250 nt upstream of the stop codon in this manuscript, they will further contribute to the  
385 complexity of this readthrough system.

386

387         Whereas the function of hnRNP A2/B1 binding to SL-Au<sub>1</sub> is to modulate the equilibrium,  
388 the function of hnRNP A2/B1 binding to SL-Ax<sub>1</sub> is unclear. However, given the spacing between  
389 the stop codon and the hnRNP A2/B1 binding site at SL-Ax<sub>1</sub>, the binding of hnRNP A2/B1 may  
390 act as a stressor upon the ribosome at the stop codon. This would be similar to EMCV and PRRSV  
391 in which binding of A2 protein and nsp1 $\beta$ /poly(C) binding protein, respectively, takes the place  
392 of the canonical structured element<sup>21,22</sup>. Given the apparent lack of an extended, complex  
393 secondary structure in VEGF-A mRNA, this hnRNP A2/B1-induced stress, in conjunction with  
394 the additive effects of the long-range RNA interaction, may take the place of complex secondary  
395 structural elements, such as pseudoknots, found in many other recoding events<sup>15</sup>.

396

397         The long-range cross-talk interaction we have identified is a novel mechanism through  
398 which ribosomal recoding events may be regulated. While the broader concept of long-range  
399 RNA interactions is not new, it has recently been suggested that more broadly, most if not all  
400 mRNAs exhibit 5' to 3' communication through a combination of proximity base pairing and  
401 RNA-binding proteins<sup>43,44</sup>. For example, recent studies measuring the spacing between 5' and 3'  
402 ends of coding mRNA, both *in-vitro* and *in-vivo*, have found them to be much closer than would  
403 be predicted for a randomly folded RNA<sup>43,45-47</sup>. While these studies support a previously  
404 established closed-loop model of RNA in which the polyA tail and 5' cap of mRNA are spatially  
405 close to regulate canonical translation initiation and termination, our studies show that the  
406 closed-loop model can give rise to specific long-range interactions that can have functional  
407 consequences on translational elongation.

## 408 **Methods**

### 409 *In-cellulo* translation assay

410 Sequences of interest for *in-cellulo* dual luciferase assays were cloned between renilla  
411 luciferase (RLuc) and firefly luciferase (FLuc) using BamHI and SacI restriction sites on the p2Luc  
412 plasmid<sup>26</sup>. To normalize against differential FLuc and RLuc luminescence, mRNA degradation,  
413 and ribosome fall-off products, each construct containing a TGA stop codon was cloned in  
414 tandem with a control construct containing a GCA codon. This control construct was used to  
415 normalize for readthrough activity.

416  
417 Dual luciferase assays were performed in white 96-well half-area flat bottom plates  
418 (Corning). Bovine aortic endothelial cells (BAOECs), grown at 37°C in 5% CO<sub>2</sub>, were diluted to  
419 30% confluency in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin. Cells  
420 were then plated at 150 µL per well with 5 total replicates per construct. 1 hour post-plating, 11µL  
421 transfection mix (480 ng plasmid, 1.2 µL Fugene6 (Promega) and 46.8 µL DMEM) was added to  
422 each well. Cells were allowed to express RLuc and FLuc for 2 days prior to taking readings.

423  
424 Luciferase luminescence readings were collected on a SpectramaxL (Molecular Devices)  
425 plate reader at 470 nm with an integration time of 1 second using the reagents in the Dual-  
426 Luciferase Reporter Assay System kit (Promega). Prior to collection, cells were lysed in 35 µL of  
427 1x passive lysis buffer (Promega) for 15 minutes. Then 50 µL FLuc substrate was added and  
428 allowed to sit for 30 seconds prior to collecting luminescence readings. Similarly, 50µL RLuc  
429 substrate was added, allowed to sit, and luminescence was collected again. Readthrough  
430 percentages were calculated as previously described<sup>26</sup>. P-values were calculated using a 2-tailed  
431 T-test.

### 432 433 *In-vitro* translation assay

434 Template DNA for rabbit reticulocyte lysate (RRL) assays was prepared by PCR  
435 amplifying hnRNP A2/B1 constructs and VEGF-A sequences of interest flanked by RLuc and  
436 FLuc. 8 µL PCR product was used as a template for *in-vitro* transcription using the mMMESSAGE

437 mMACHINE T7 Transcription Kit (Invitrogen). RNA was diluted to 10  $\mu\text{g}/\mu\text{L}$  in nuclease free  
438 water.

439

440 8  $\mu\text{L}$  RRL (Promega) master mix (100  $\mu\text{L}$  reaction mixture, 17.5  $\mu\text{L}$  nuclease free  $\text{H}_2\text{O}$ , and  
441 2.5  $\mu\text{L}$  methionine) was combined with 1  $\mu\text{L}$  of each RNA (VEGF and hnRNP A2/B1) at the  
442 desired concentration. For no protein samples, 1  $\mu\text{L}$  transcription reaction buffer was added  
443 instead. Reactions were then allowed to continue for 90 minutes at 30°C prior to placing on ice  
444 for 20 minutes. Lysates were then plated at 2.5  $\mu\text{L}$  per well with 3 total replicates per construct in  
445 white 96-well half-area flat bottom plates (Corning). Luciferase luminescence readings were  
446 collected as described above.

447

448

#### 449 **Preparation of hnRNP A2/B1**

450 An hnRNP A2/B1 consisting of RRM1/2 was cloned into a pGEX-6P-1 plasmid using  
451 BamHI and NotI restriction sites<sup>48</sup>.

452

453 BL-21 cells containing GST-tagged hnRNP A2/B1 constructs in pGEX-6P-1 were  
454 inoculated in LB media containing 100  $\mu\text{g}/\text{mL}$  ampicillin and grown overnight at 37°C with  
455 shaking at 200 rpm. In the morning, 20 mL of overnight culture were diluted into 2 L LB media  
456 containing 100  $\mu\text{g}/\text{mL}$  ampicillin. Cells were allowed to grow at 37°C with shaking until reaching  
457  $\text{OD}_{600} \sim 1$ . Cells were cooled to 20°C and subsequently induced with 100  $\mu\text{M}$  IPTG for 16 hours,  
458 after which they were pelleted at 5,000g for 20 minutes. A pellet corresponding to 2 L was  
459 resuspended in 50 mL hnRNP lysis buffer (50 mM Phosphate pH 7.5, 200 mM NaCl, 40 mM  
460  $\text{MgCl}_2$ , 1 mM DTT, 0.1% Triton X-100) and sonicated (Branson Sonifier 450) at setting 7-8 with  
461 50% amplitude for 6 pulses of 60 seconds each. 5 mL of 4% polyethyleneimine (PEI) was added  
462 drop wise to the lysed cells and allowed to mix on ice for 30 minutes. The lysed cells were then  
463 spun at 27,000g for 20 minutes.

464

465 Lysate supernatant was added to Pierce Glutathione Agarose Affinity Purification Media  
466 (Thermo Scientific) and allowed to bind for 2 hours on a rocker at 4°C. Following binding, beads  
467 were washed twice with 50 mL ice cold hnRNP wash buffer (50 mM Phosphate pH 7.5, 500 mM  
468 NaCl, 1 mM DTT). Subsequently, protein-bound beads were loaded onto a column and washed  
469 overnight with 2 L hnRNP wash buffer using a pump at a flow rate of 1.8 mL/minute at 4°C.  
470 Protein was eluded in 50 mL elution buffer (20 mM Phosphate pH 7.5, 15 mM KCl, 2 mM MgCl<sub>2</sub>,  
471 5% glycerol) and 2.4 mg PreScission protease. The column was allowed to rock at 4°C for 8 hours  
472 prior to collecting the first elution. hnRNP A2/B1 was concentrated in Ultra Centrifugal Filter  
473 Units (Amicon) with a 3 kDa cutoff and washed into the appropriate buffer as needed.

474

#### 475 **Large-scale synthesis and purification of RNA**

476 RNA constructs were either ordered (IDT or Horizon Discovery) or synthesized in-house  
477 using T7 polymerase as previously described<sup>49</sup>. Final RNA was washed into appropriate buffer  
478 as needed using Ultra Centrifugal Filter Units (Amicon).

479

#### 480 **NMR & Structure Calculations**

481 RNA samples were placed in 300 μL NMR buffer (10 mM Tris pH 5.6, 10 mM NaCl or 10  
482 mM HEPES pH 5.6, 50 mM KCl) in D<sub>2</sub>O. For assignments of exchangeable proton and nitrogen  
483 resonance NOESY and HSQC were collected on samples lyophilized and resuspended in H<sub>2</sub>O  
484 supplemented with 10-15% D<sub>2</sub>O at 280K. For non-exchangeable proton and carbon assignments  
485 spectra were collected in 99.96% D<sub>2</sub>O. Assignments were done with the use of NMRFX<sup>50</sup>. As  
486 previously described, CYANA<sup>51</sup> was used to calculate initial structures of the SL-Au<sub>1</sub>, SL-Ax<sub>1</sub> and  
487 SL-Ax<sub>2</sub> with manually assigned restraints. Standard torsion-angle ( $\pm 50^\circ$  deviations from ideality)  
488 and cross-helix phosphate-phosphate restraints were used for  $\alpha$ -helical regions. The structure  
489 with the lowest target function was used to create a coordinate file for refinement in xplor-nih<sup>52,53</sup>.  
490 Phosphate- phosphate distances were not used at this step. Molecular images were generated  
491 with PyMOL ([http:// www.pymol.org](http://www.pymol.org)).

492

#### 493 **ITC**

494 Most ITC experiments were performed in ITC buffer (10 mM HEPES pH 8.0, 50 mM KCl,  
495 1 mM EDTA, 1 mM BME), with the exception of certain SL-Ax1 constructs (20 mM phosphate pH  
496 7.2, 15 mM KCl, 2 mM MgCl<sub>2</sub>, 5% glycerol). RNA constructs were diluted to 5 μM for and hnRNP  
497 A2/B1 constructs were diluted to 100 μM. ITC experiments were performed on an ITC-200  
498 microcalorimeter (MicroCal) by loading RNA into the cell and protein into the syringe. Protein  
499 was injected in 2 μL increments for a total of 20 injections. Baseline correction, binding curve  
500 fitting, and parameter calculation were performed using Origin (OriginLab).

501

### 502 **DMS-MaPseq**

503 *In-cellulo* DMS-MaPseq samples were prepared by seeding 10-cm cell culture dishes of  
504 BAOECs at 30% confluency. 1 hour post-plating, 660 μL transfection mix (5 μg plasmid, 26 μL  
505 Fugene6 (Promega) and 650 μL DMEM) was added to each well and allowed to grow for 2 days.  
506 Sample processing and analysis was performed as previously described<sup>28,54</sup>.

507

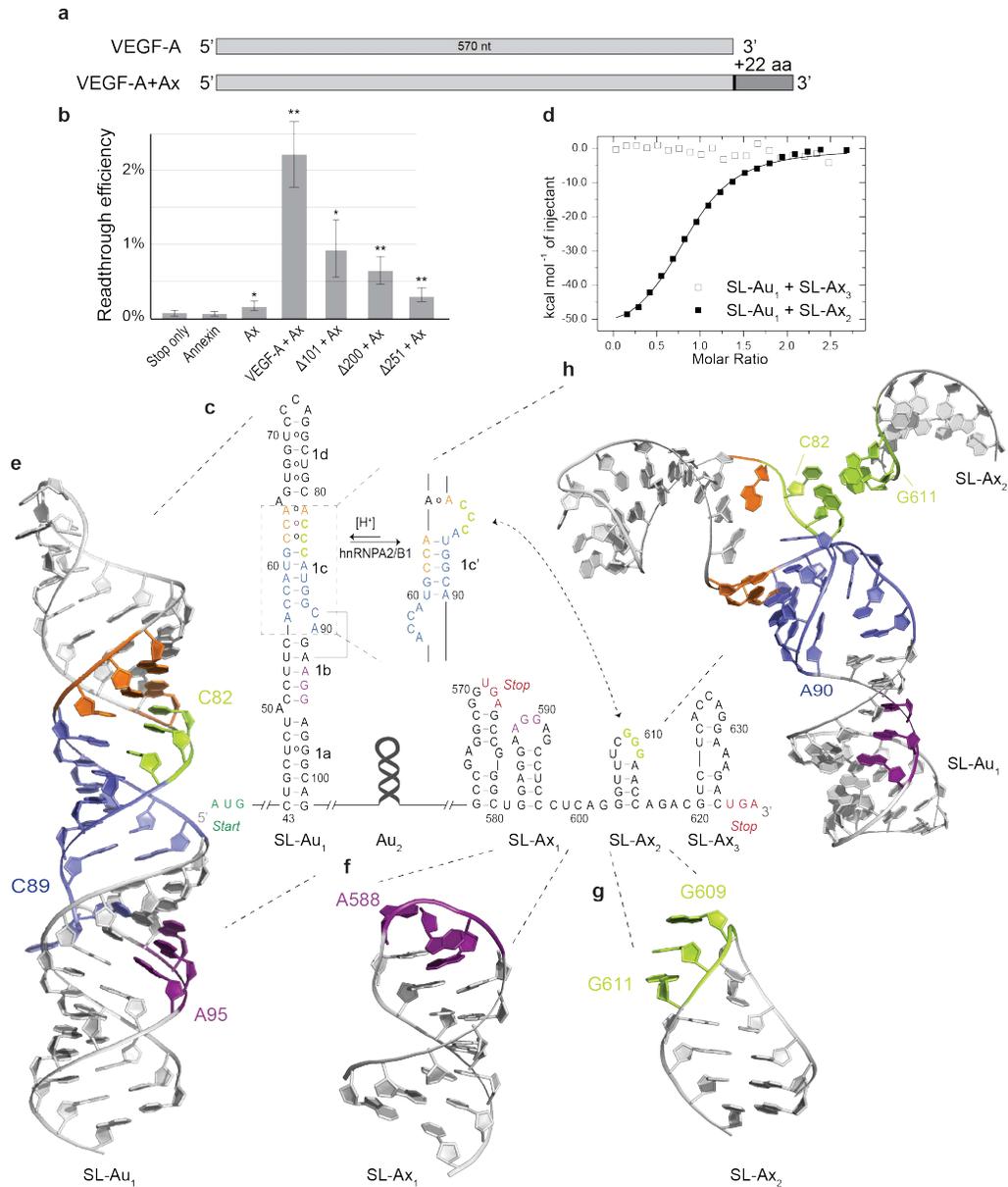
### 508 **Small Angle X-Ray Scattering**

509 SAXS experiments were performed at the SIBYLS beamline (Advanced Light Source,  
510 Lawrence Berkeley National Laboratory)<sup>55</sup>. Samples were prepared in NMR buffer (10 mM Tris  
511 pH 5.6, 10 mM NaCl) supplemented with 2% glycerol. Data analysis and *ab initio* envelope  
512 reconstruction were done using ScÅtter software<sup>56</sup>.

**Extended Data Table 1- NMR statistics and restraints**

NMR and refinement statistics for the Au <sub>1</sub> region		NMR and refinement statistics for the Ax <sub>1</sub> region		NMR and refinement statistics for the Ax <sub>2</sub> region	
	Au <sub>1</sub>		Ax <sub>1</sub>		Ax <sub>2</sub>
NMR distance and dihedral constraints		NMR distance and dihedral constraints		NMR distance and dihedral constraints	
<b>Distance restraints</b>		<b>Distance restraints</b>		<b>Distance restraints</b>	
Total NOE	463	Total NOE	123	Total NOE	105
Intraresidue	297	Intraresidue	74	Intraresidue	65
Inter-residue	166	Inter-residue	49	Inter-residue	40
Sequential ( $ i-j =1$ )	164	Sequential ( $ i-j =1$ )	45	Sequential ( $ i-j =1$ )	38
Nonsequential ( $ i-j >1$ )	2	Nonsequential ( $ i-j >1$ )	4	Nonsequential ( $ i-j >1$ )	2
Hydrogen bonds	127	Hydrogen bonds	37	Hydrogen bonds	19
Total dihedral-angle restraints	1014	Total dihedral-angle restraints	272	Total dihedral-angle restraints	216
Base pair	0	Base pair	0	Base pair	0
Sugar pucker	240	Sugar pucker	64	Sugar pucker	52
Backbone	774	Backbone	208	Backbone	164
Based on A-form geometry	1014	Based on A-form geometry	272	Based on A-form geometry	216
Structure statistics		Structure statistics		Structure statistics	
<b>Violations (mean <math>\pm</math> s.d.)</b>		<b>Violations (mean <math>\pm</math> s.d.)</b>		<b>Violations (mean <math>\pm</math> s.d.)</b>	
Distance constraints (Å)	2.8 $\pm$ 0.4	Distance constraints (Å)	0.6 $\pm$ 0.5	Distance constraints (Å)	0.0 $\pm$ 0.0
Dihedral-angle constraints (°)	41.0 $\pm$ 3.5	Dihedral-angle constraints (°)	11.8 $\pm$ 2.1	Dihedral-angle constraints (°)	5.5 $\pm$ 1.4
<b>Deviations from idealized geometry</b>		<b>Deviations from idealized geometry</b>		<b>Deviations from idealized geometry</b>	
Bond lengths (Å)	0.005 $\pm$ 0.000	Bond lengths (Å)	0.004 $\pm$ 0.000	Bond lengths (Å)	0.004 $\pm$ 0.000
Bond angles (°)	0.628 $\pm$ 0.006	Bond angles (°)	0.667 $\pm$ 0.007	Bond angles (°)	0.616 $\pm$ 0.006
Improper (°)	0.352 $\pm$ 0.009	Improper (°)	0.369 $\pm$ 0.009	Improper (°)	0.342 $\pm$ 0.007
<b>Average pairwise r.m.s. deviation (Å)<sup>a</sup></b>		<b>Average pairwise r.m.s. deviation (Å)<sup>a</sup></b>		<b>Average pairwise r.m.s. deviation (Å)<sup>a</sup></b>	
All RNA heavy	0.357	All RNA heavy	0.0687	All RNA heavy	0.0466
All nucleotides	0.3717	All nucleotides	0.4137	All nucleotides	0.0553

<sup>a</sup> Pairwise r.m.s. deviation was calculated among ten refined structures.



514

515 **Figure 1: Identification and structures of the elements involved in mediating VEGF-A readthrough.**

516 **(a)** Schematic diagram of the VEGF-A mRNA without (top) and with (bottom) the Ax-element. **(b)**

517 Functional readthrough levels show that only the full-length VEGF-A + Ax mRNA leads to substantial  
518 readthrough levels. Error bars indicate standard error (n=3) and statistical tests were performed using a 2-

519 tailed t-test (\* p < 0.05, \*\* p < 0.005). **(c)** The secondary structure reveals an extended stem-loop for SL-

520 Au<sub>1</sub>, including an alternate conformation (bases shown in orange and blue for simplicity), as well as  
521 shorter stem-loops SL-Ax<sub>1-3</sub> within the Ax-element. hnRNP binding sites are indicated in purple, stop

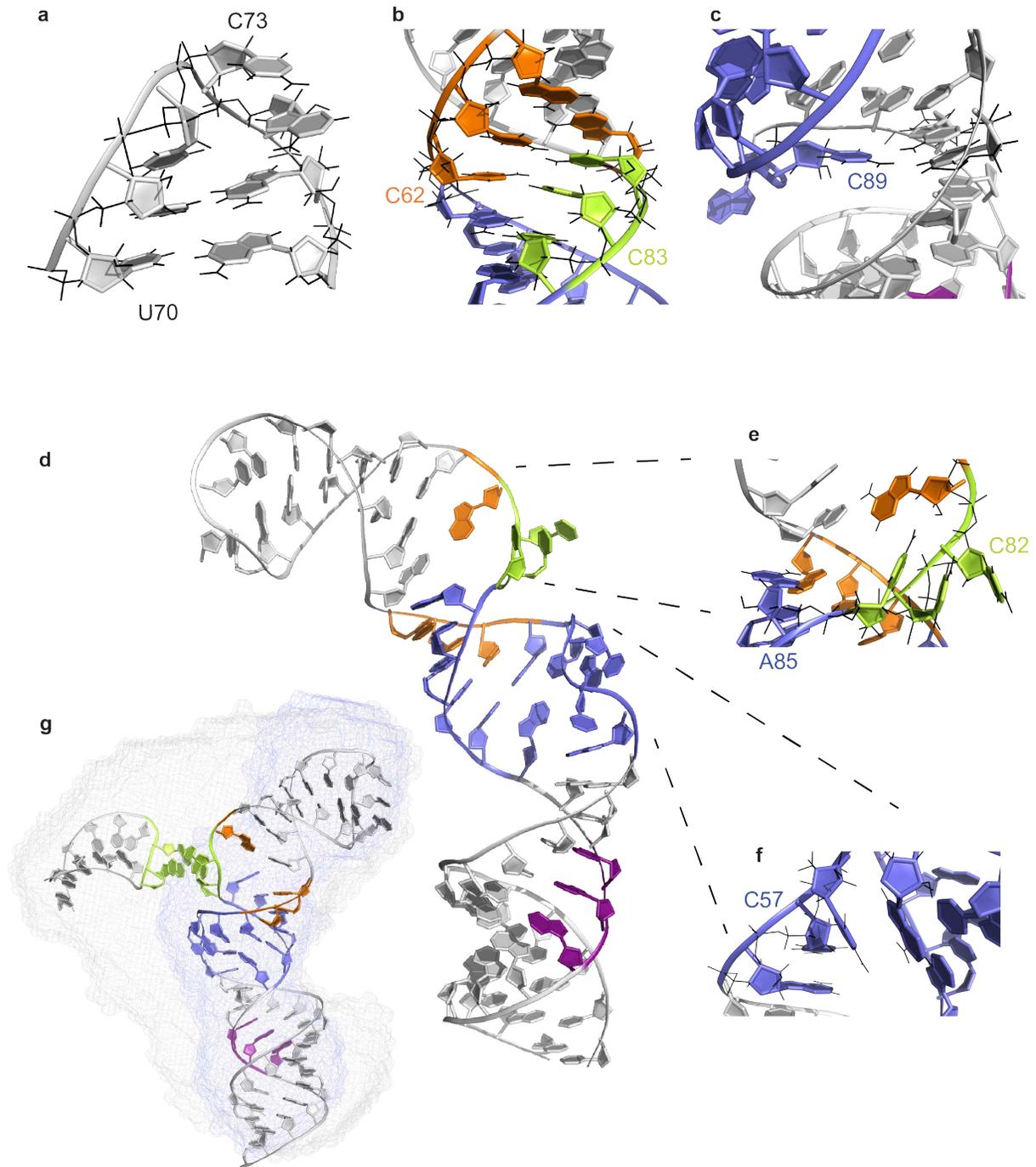
522 codons are shown in red, and bases involved in the long-range interaction are shown in green. **(d)** ITC

523 data show that SL-Au<sub>1</sub> is able to specifically interact with SL-Ax<sub>2</sub> (black squares), but not with SL-Ax<sub>3</sub>

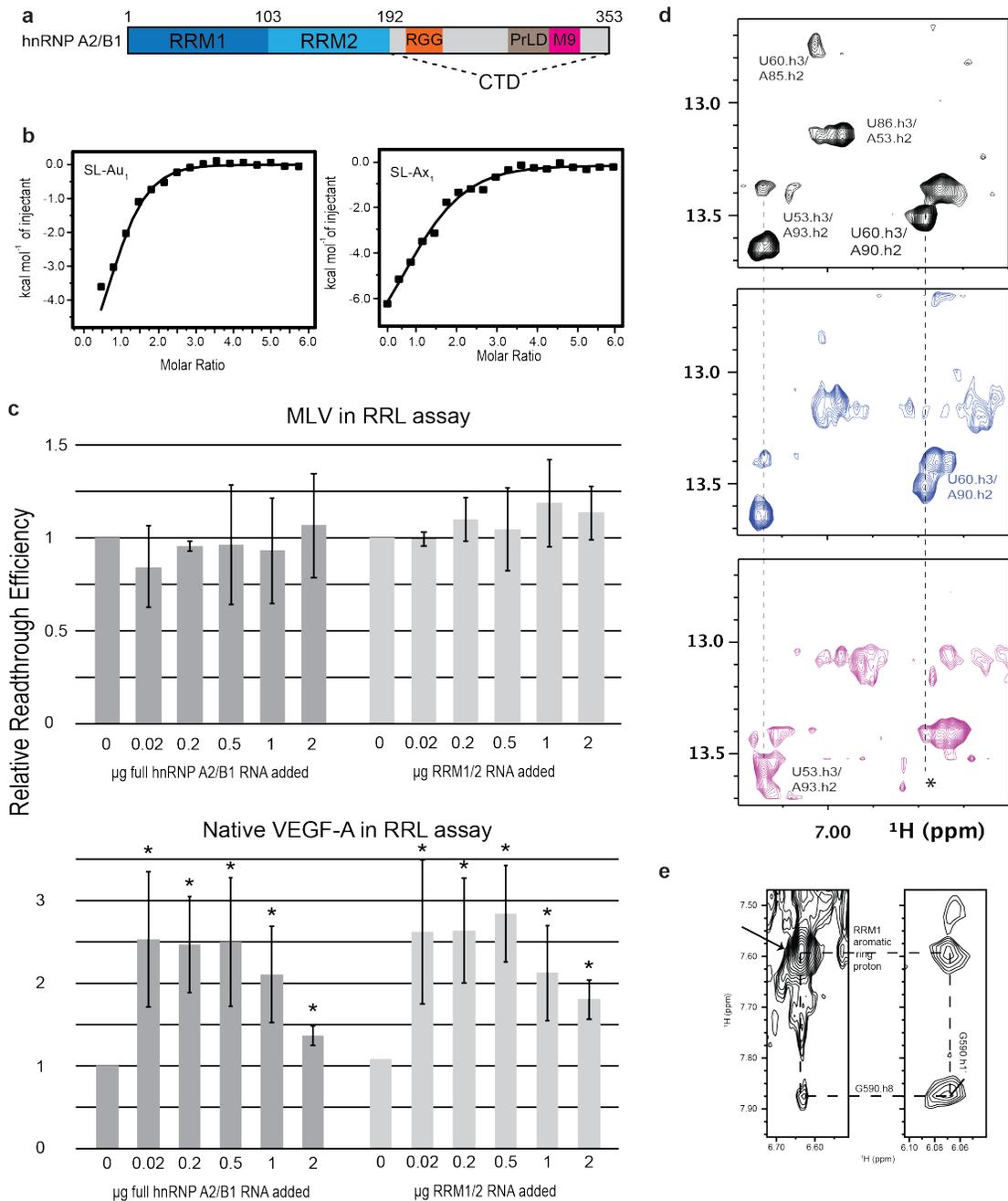
524 (white squares). Tertiary folds are shown for SL-Au<sub>1</sub> (PDB 7KUB) **(e)**, SL-Ax<sub>1</sub> (PDB 7KUC) **(f)**, and SL-

525 Ax<sub>2</sub> (PDB 7KUD) **(g)** with the same base coloring as in **(c)**. **(h)** A model was created showing the docking

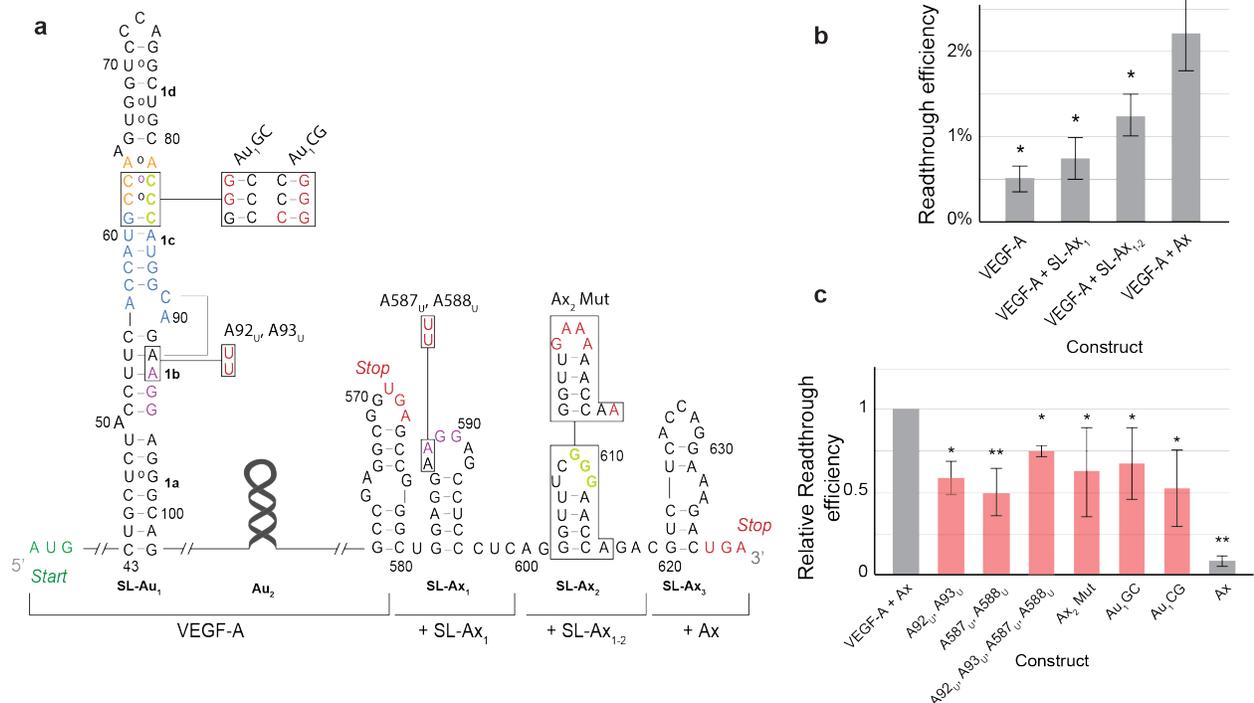
526 interaction between SL-Au<sub>1</sub> and SL-Ax<sub>2</sub>.



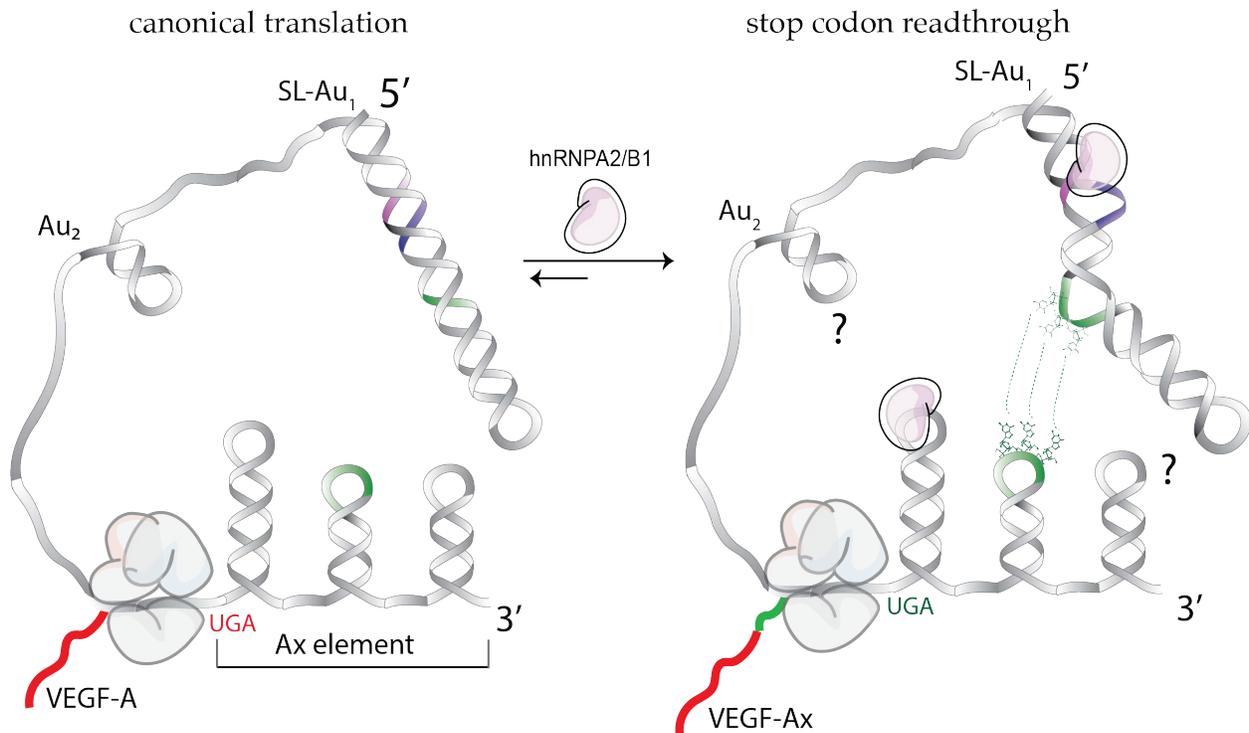
528 **Figure 2: Structural features of SL-Au<sub>1</sub>.** (a) Molecular detail of the C<sub>73</sub>CA triloop that caps SL-Au<sub>1</sub>. (b)  
 529 Helix 1c of SL-Au<sub>1</sub> contains 2 tandem C-C non-canonical basepairs. (c) SL-Au<sub>1</sub> is able to form a triple  
 530 base-pair interaction between C<sub>89</sub> and the A<sub>92</sub>-U<sub>54</sub> base-pair. (d) Helix 1c' assumes a kinked conformation  
 531 and is flanked by an A<sub>56</sub>CCA bulge at the junction with helix 1b (e) and an A<sub>81</sub>CCCA bulge at the  
 532 junction with helix 1d (f). (g) The density, as determined by SAXS, of free SL-Au<sub>1</sub> in the 1c'  
 533 conformation (blue) shows the bend described above. It also overlays with a model envelope (grey) of the  
 534 SL-Au<sub>1</sub>:SL-Ax long-range interaction complex. Base coloring is as in Figure 1.



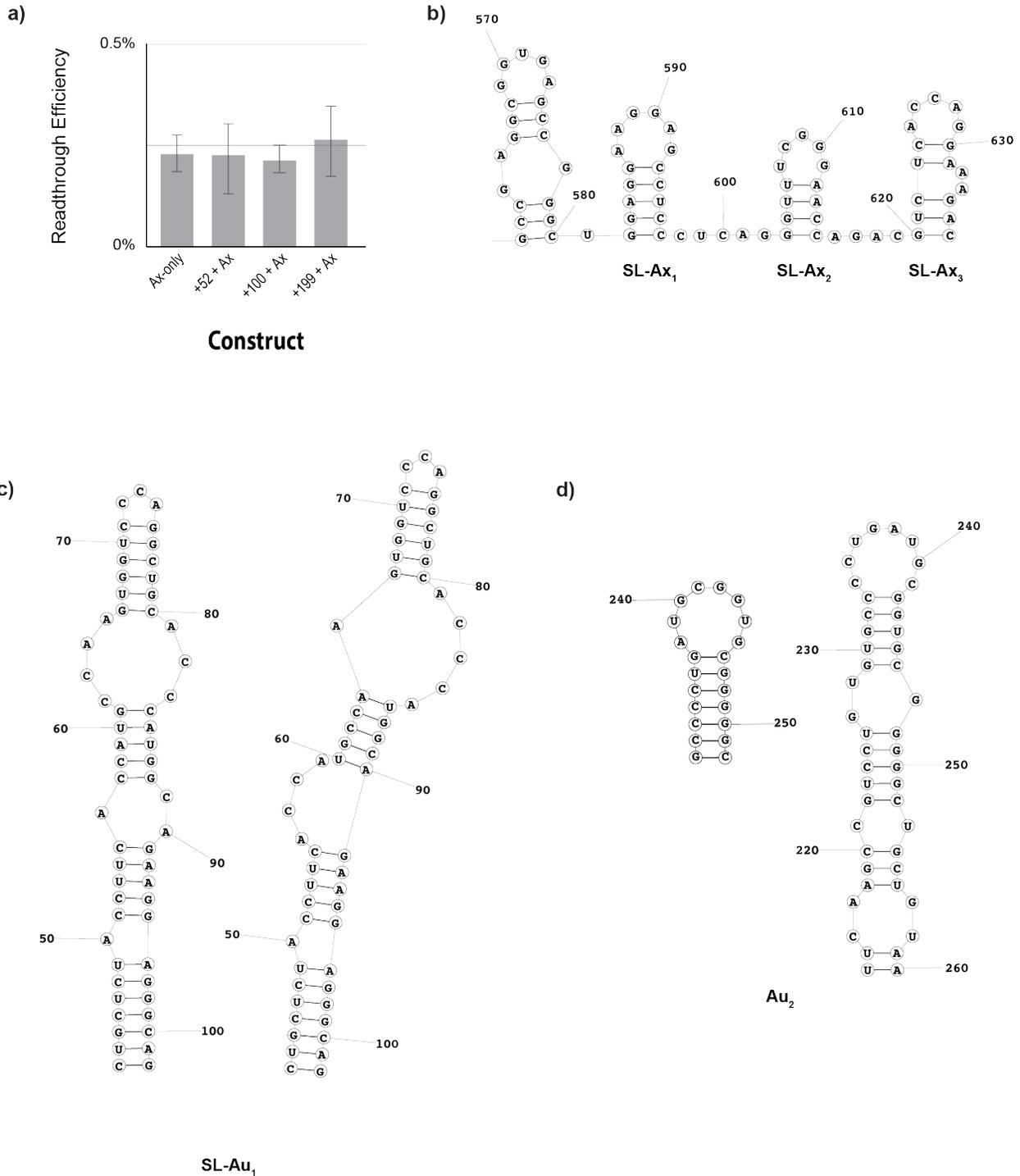
535 **Figure 3: hnRNP A2/B1 binds to VEGF-A mRNA to help mediate readthrough.** (a) Schematic of the  
 536 hnRNP A2/B1 domain architecture showing its RNA-recognition motifs (RRM1 and RRM2) and C-  
 537 terminal domain (CTD), consisting of an Arg-Gly-Gly box (RGG), prion-like domain (PrLD), and M9  
 538 transport sequence. (b) ITC binding studies of RRM1/2 to SL-Au<sub>1</sub> (left) or SL-Ax<sub>1</sub> (right) show tight and  
 539 specific binding. Representative curves of 2 trials are shown. (c) Rabbit reticulocyte lysate assays to  
 540 which either full length hnRNP A2/B1 (dark grey) mRNA or RRM1/2 (light grey) was added show no  
 541 difference in readthrough activity for an MLV control construct (top) or a VEGF-A construct (bottom).  
 542 Error bars indicate standard error (n=3) and statistical tests were performed using a 2-tailed t-test (\* p <  
 543 0.05). (d) Titration of RRM1/2 (black, 0 molar equivalents; blue, 0.3 molar equivalents; pink, 1 molar  
 544 equivalent) into SL-Au<sub>1</sub> at 280 K shows the shifting of several <sup>1</sup>H-<sup>1</sup>H NMR peaks of base pairs disrupted  
 545 by protein binding. (e) Titration of RRM1 into SL-Ax<sub>1</sub> at 311K results in <sup>1</sup>H-<sup>1</sup>H NOE cross-peaks  
 546 between G<sub>590</sub> and aromatic ring hydrogens of the protein.



547  
 548 **Figure 4: Each element of the VEGF-mRNA separately contributes toward readthrough activity.**  
 549 **(a)** Schematic indicating mutations made to the VEGF-A mRNA. Mutations are labeled and shown in red.  
 550 For truncated products, the maximal extent of the VEGF-A mRNA is indicated underneath the schematic.  
 551 **(b)** Functional readthrough experiments show that each stem-loop within the Ax-element contributes  
 552 toward readthrough. **(c)** Each tested mutation led to a decrease in readthrough levels, highlighting the  
 553 importance of each region in mediating ribosomal readthrough. Error bars indicate standard error (n=3)  
 554 and statistical tests were performed using a 2-tailed t-test (\* p < 0.05, \*\* p < 0.005).



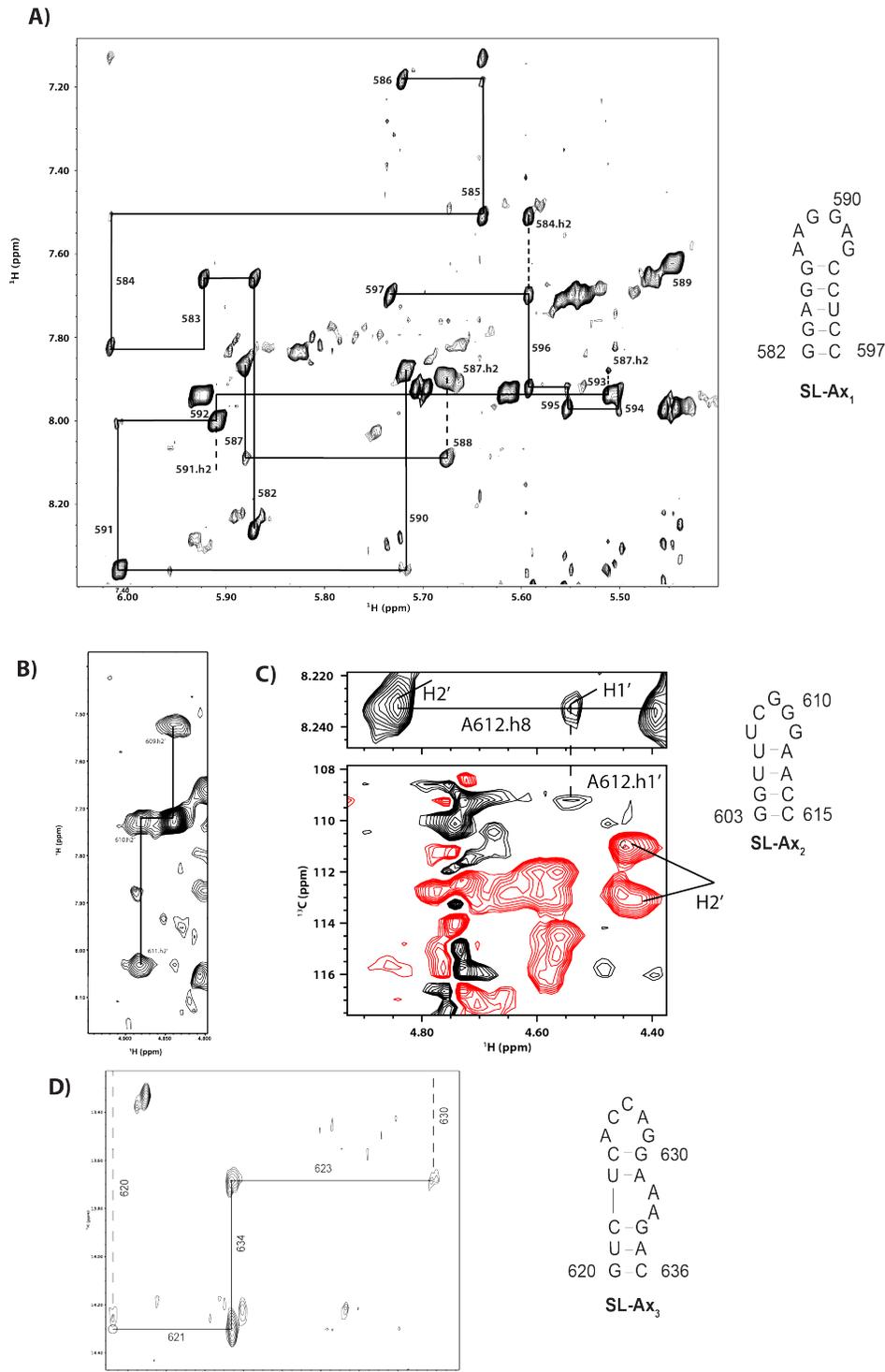
555  
 556 **Figure 5: RNA alternate conformations mechanism of VEGF-A readthrough.** A mechanistic model  
 557 of the VEGF-A coding mRNA shows the formation of a linear SL-Au<sub>1</sub> element at the 5' end of the RNA  
 558 and three stem loops (SL-Ax<sub>1-3</sub>) at the 3' end of the RNA in the absence of hnRNP A2/B1. In the presence  
 559 of hnRNP A2/B1, SL-Au<sub>1</sub> undergoes a register shift, exposing three cytosine residues. These three  
 560 cytosine residues are able to partake in long-range Watson-Crick base pairing with SL-Ax<sub>2</sub>, thereby  
 561 promoting translational stop-codon readthrough.



562

563 **Extended Data Figure 1: VEGF-A readthrough does not rely on local enhancers and DMS**

564 **reactivities of the readthrough elements (a)** Readthrough experiments indicate that the addition of  
 565 sequences upstream of the Ax-element did not lead to increased levels in readthrough, ruling out the  
 566 potential effects of local enhancers. Error bars indicate standard error (n=3). DMS-MaPseq data show the  
 567 predicted secondary structures of the Ax-element (**b**), the alternate conformations of SL-Au<sub>1</sub> (**c**), and the  
 568 alternate conformations of the second upstream signal, Au<sub>2</sub> (**d**).



569

570 **Extended Data Figure 2: NMR data confirm the predicted secondary structure of SL-Ax<sub>1-3</sub>.**

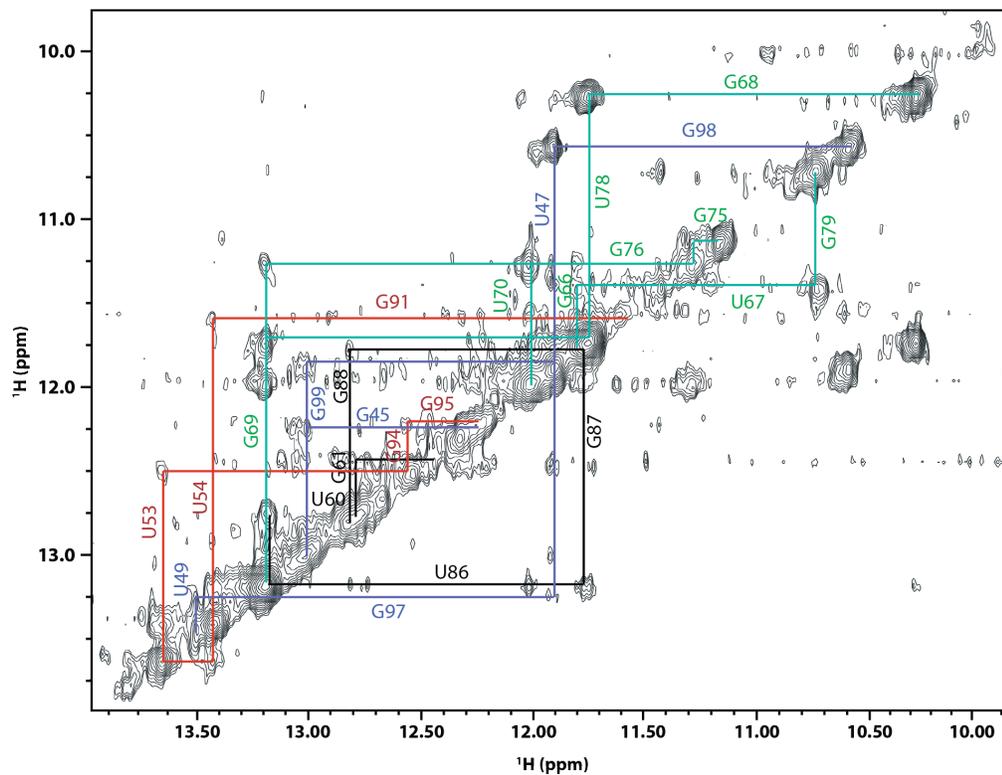
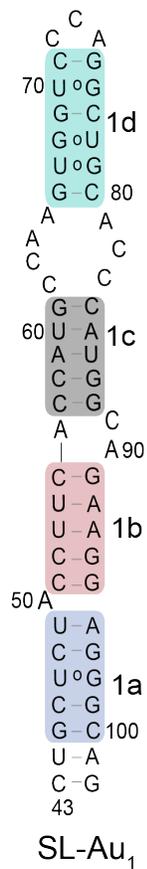
571 **(a)** <sup>1</sup>H-<sup>1</sup>H NMR data at 311K show that SL-Ax<sub>1</sub> assumes the fold predicted by DMS-MaPseq. **(b)** <sup>1</sup>H-<sup>1</sup>H

572 NMR data (311 K) show that G<sub>609</sub>GG in SL-Ax<sub>2</sub> stack on top of each other and **(c)** HMQC data (311 K)

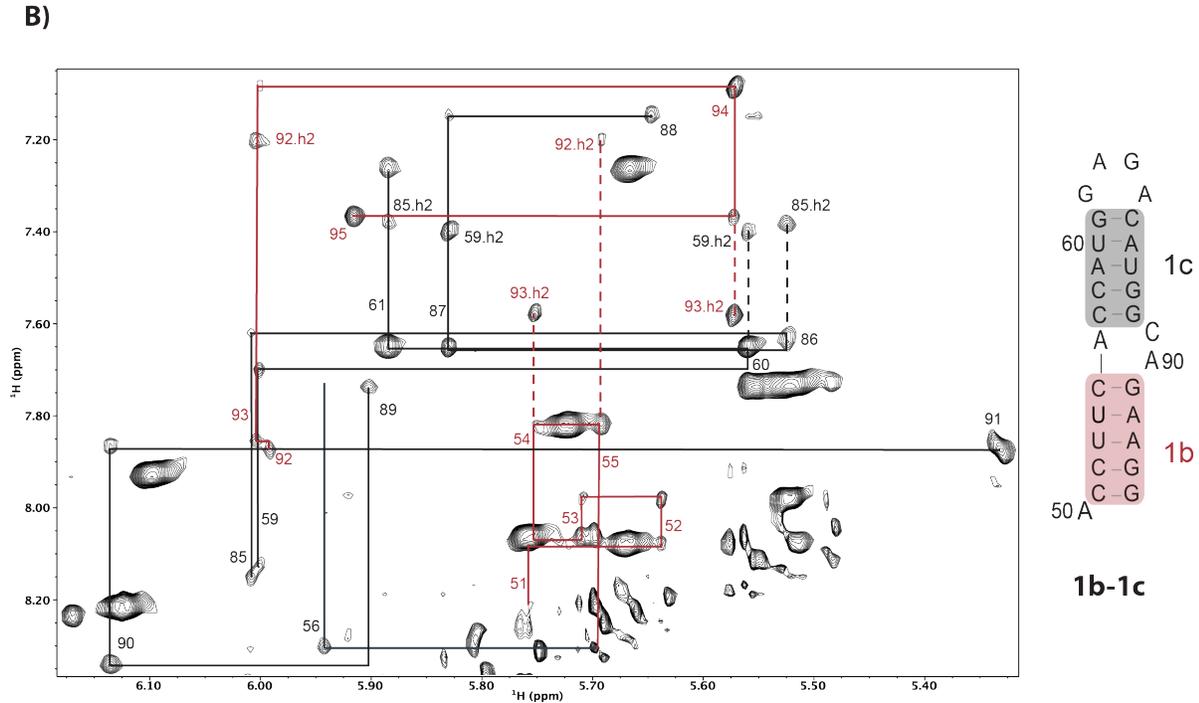
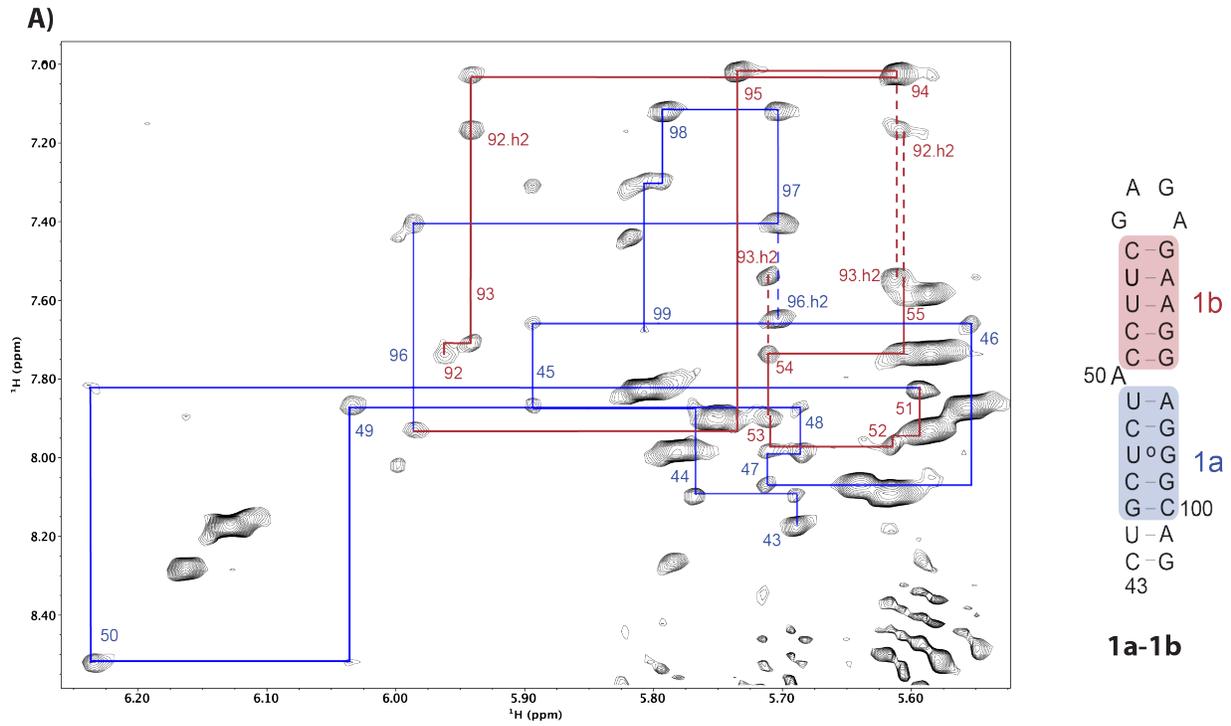
573 show the upfield shift of the A<sub>612</sub> H1' due to stacking on the ribose ring of G<sub>611</sub>. **(d)** Imino proton NMR

574 assignments at 280 K for SL-Ax<sub>3</sub> were obtained and confirm the secondary structure prediction and show

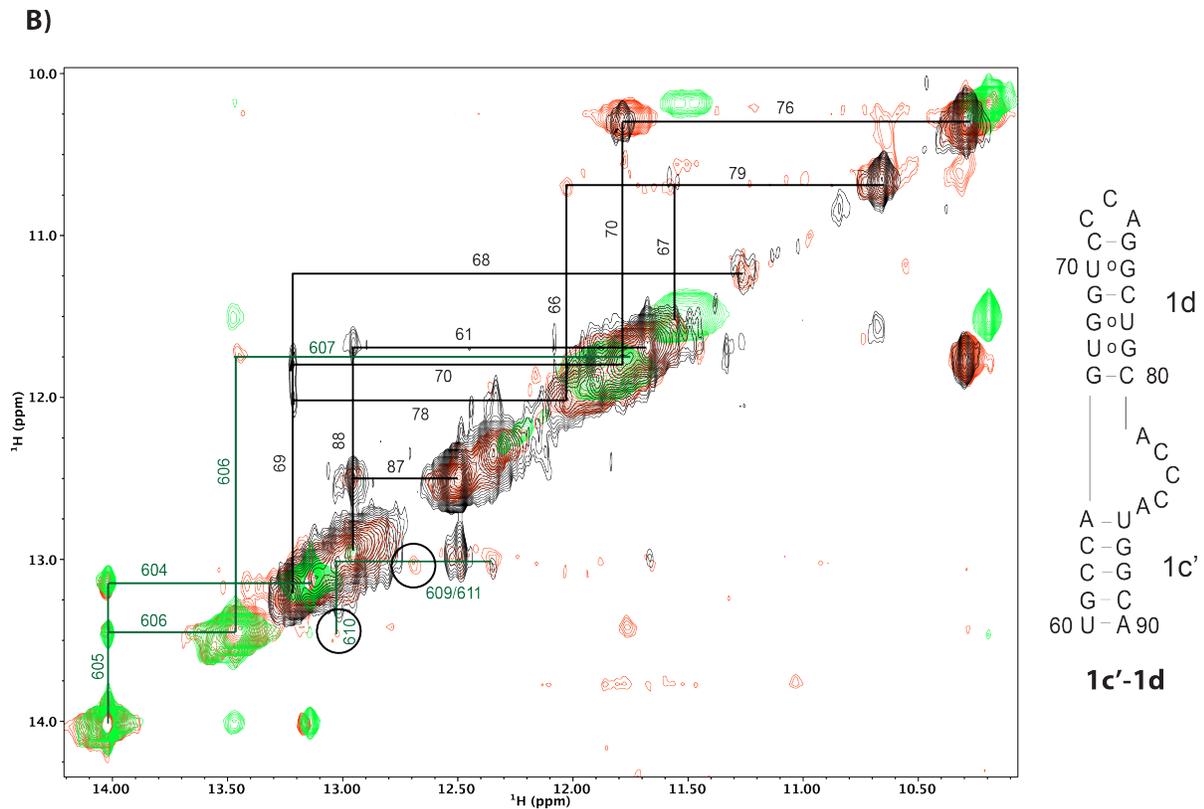
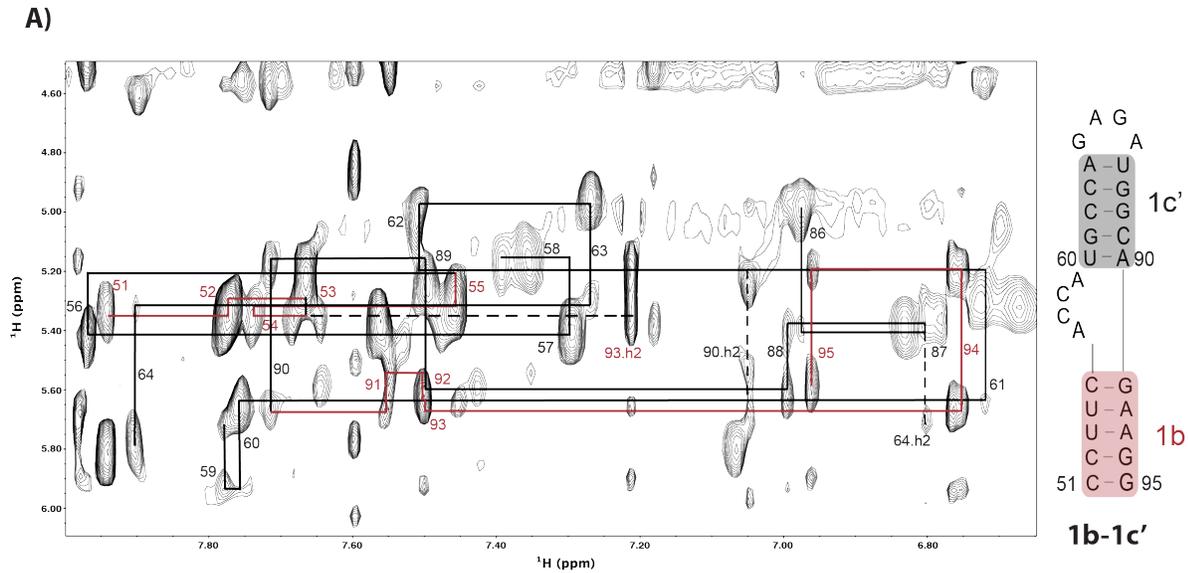
575 that the A<sub>632</sub>A dinucleotide is extruded out from the stem.



576  
 577 **Extended Data Figure 3: NMR imino data confirm the predicted secondary structure of SL-Au<sub>1</sub>.**  
 578 The complete imino proton NMR assignments (280 K) for SL-Au<sub>1</sub> in the 1c conformation were obtained.  
 579 Imino assignments are colored to match the schematic of the secondary structure, which was subdivided  
 580 into helices 1a (blue), 1b (red), 1c (grey), and 1d (green).



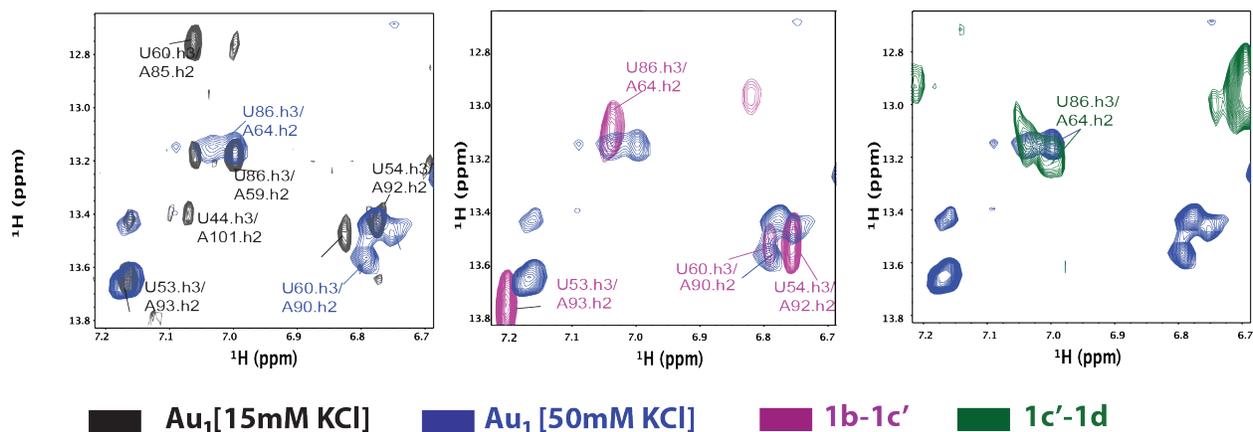
581  
 582 **Extended Data Figure 4: NMR data confirm the predicted structures of SL-Au<sub>1</sub> helices.** Shortened  
 583 RNA constructs were used to ascertain the structure of the full SL-Au<sub>1</sub> element. <sup>1</sup>H-<sup>1</sup>H NMR data (311 K)  
 584 were collected to assign the peaks for helices **(a)** 1a-1b and **(b)** 1b-1c.



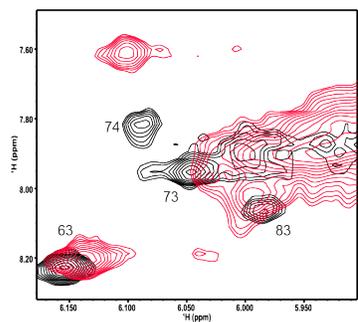
585

586 **Extended Data Figure 5: NMR data reveal the existence of alternate structures of SL-Au<sub>1</sub> helices**  
 587 **that allow for long-range interactions. (a)** <sup>1</sup>H-<sup>1</sup>H NMR data (311 K) were collected to assign the peaks  
 588 for the alternate conformation of helix 1c using a shortened helices 1b-1c' construct. **(b)** Imino proton  
 589 NMR assignments (298 K) for helices 1c'-1d (black) along with overlays of free SL-Ax<sub>1</sub> (green) and the  
 590 SL-Ax<sub>1</sub>:1c'-1d complex (red) shows the formation of long-range basepairing involving G<sub>60</sub>GG of SL-Ax<sub>1</sub>.

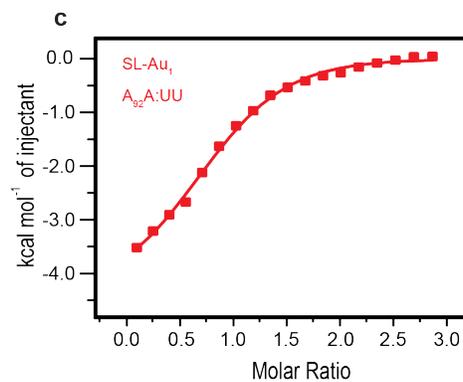
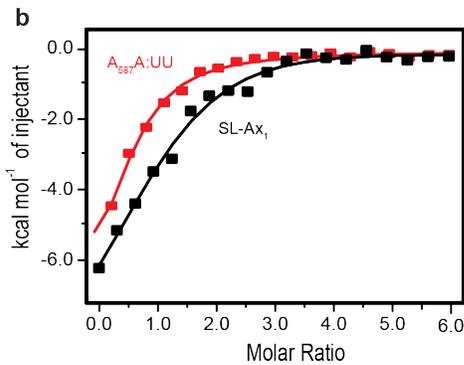
A)



B)



591  
 592 **Extended Data Figure 6: NMR data confirm the presence of an equilibrium between helices 1c and**  
 593 **1c' of SL-Au.** (a) Multiple sets of signals are present in <sup>1</sup>H-<sup>1</sup>H NMR data (280 K) at 15 mM KCl (black)  
 594 and 50 mM KCl (blue), showing the salt dependence of the 1c to 1c' equilibrium. Overlays with  
 595 shortened constructs 1b-1c' (purple) and 1c'-1d (green) show that the 50 mM KCl condition favors the  
 596 1c' conformation. (b) The 1c to 1c' equilibrium (280 K) is also pH dependent, showing distinct peaks at  
 597 pH 5.6 (black) and pH 7.3 (red).



598

599 **Extended Data Figure 7: Manipulations to the VEGF-A coding sequence maintain global mRNA**  
 600 **structure and reduce hnRNP A2/B1 affinity. (a)** DMS reactivities for SL-Au<sub>1</sub> and the Ax-element were  
 601 collected for various VEGF-A mutant constructs to show there was no major changes in global fold.  
 602 Mutations are indicated in red boxes while truncations are indicated by omitted nucleotides. U and G  
 603 bases (grey) show no DMS reactivity while A and C bases may have low (black), medium (orange), or  
 604 high (red) DMS reactivity. **(b)** ITC was used to determine a difference in binding affinity for RRM1/2 to  
 605 native SL-Ax<sub>1</sub> and SL-Ax<sub>1</sub> with an A<sub>587</sub>A:UU mutation. To obtain a monophasic binding curve for the  
 606 mutant, a phosphate buffer was used for this comparison. **(c)** An ITC binding curve was collected for  
 607 RRM1/2 binding to SL-Au<sub>1</sub> with an A<sub>92</sub>A:UU mutation in the same HEPES buffer as in Figure 3b. For  
 608 these ITC experiments, representative curves of 3 trials are shown.

609 **Acknowledgements**

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

615

616 **Accession codes**

617 Atomic coordinates have been deposited in the Protein Data Bank under accession codes PDB  
618 7KUB (SL-Au<sub>1</sub>), PDB 7KUC (SL-Ax<sub>1</sub>), and PDB 7KUD (SL-Ax<sub>2</sub>). Chemical shifts have been  
619 deposited in the Biological Magnetic Resonance Data Bank under accession codes 30816 (SL-Au<sub>1</sub>),  
620 30817 (SL-Ax<sub>1</sub>), and 30818 (SL-Ax<sub>2</sub>).

621

622 **Author contributions**

623 N.O.W., J.M.E., and V.M.D'S. conceived and designed the experiments. N.O.W. performed *in-*  
624 *cellulo* and *in-vitro* functional assays. N.O.W. and J.M.E. purified the samples and performed the  
625 NMR, ITC, and SAXS experiments. J.A. performed NMR experiments on SL-Ax<sub>1</sub>. N.O.W., J.M.E.,  
626 J.A. and V.M.D'S. performed the structural analyses. P.G., H.S., and S.R. performed the DMS-  
627 MaPseq experiments and analysis. N.W., J.M.E., and V.M.D'S. interpreted the data and wrote the  
628 manuscript.

629

630 **Competing financial interests**

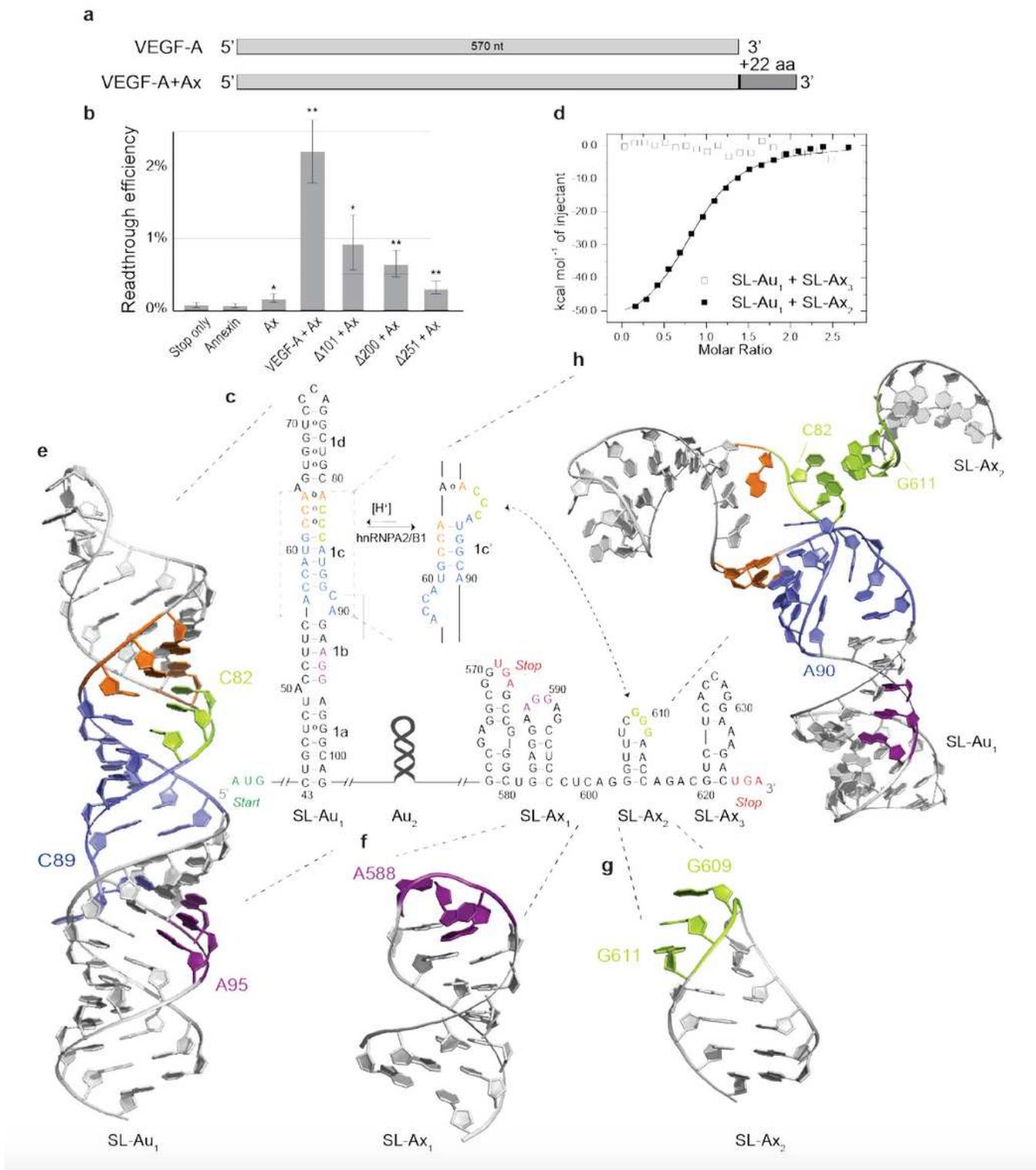
631 The authors declare no competing financial interests.

- 632 1. Schueren, F. & Thoms, S. Functional Translational Readthrough: A Systems Biology  
633 Perspective. *PLoS Genet* **12**, e1006196 (2016).
- 634 2. Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R. & Weissman, J.S. Ribosome profiling  
635 reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*.  
636 *Elife* **2**, e01179 (2013).
- 637 3. Schueren, F. et al. Peroxisomal lactate dehydrogenase is generated by translational  
638 readthrough in mammals. *Elife* **3**, e03640 (2014).
- 639 4. Stiebler, A.C. et al. Ribosomal readthrough at a short UGA stop codon context triggers  
640 dual localization of metabolic enzymes in Fungi and animals. *PLoS Genet* **10**, e1004685  
641 (2014).
- 642 5. Loughran, G. et al. Stop codon readthrough generates a C-terminally extended variant of  
643 the human vitamin D receptor with reduced calcitriol response. *J Biol Chem* **293**, 4434-  
644 4444 (2018).
- 645 6. Brierley, I. et al. An efficient ribosomal frame-shifting signal in the polymerase-encoding  
646 region of the coronavirus IBV. *EMBO J* **6**, 3779-85 (1987).
- 647 7. Yoshinaka, Y., Katoh, I., Copeland, T.D. & Oroszlan, S. Murine leukemia virus protease is  
648 encoded by the gag-pol gene and is synthesized through suppression of an amber  
649 termination codon. *Proc Natl Acad Sci U S A* **82**, 1618-22 (1985).
- 650 8. Steneberg, P. & Samakovlis, C. A novel stop codon readthrough mechanism produces  
651 functional Headcase protein in *Drosophila trachea*. *EMBO Rep* **2**, 593-7 (2001).
- 652 9. Curran, J.F. & Yarus, M. Use of tRNA suppressors to probe regulation of *Escherichia coli*  
653 release factor 2. *J Mol Biol* **203**, 75-83 (1988).
- 654 10. Jungreis, I. et al. Evolutionary Dynamics of Abundant Stop Codon Readthrough. *Mol Biol*  
655 *Evol* **33**, 3108-3132 (2016).
- 656 11. Eswarappa, S.M. et al. Programmed translational readthrough generates antiangiogenic  
657 VEGF-Ax. *Cell* **157**, 1605-18 (2014).
- 658 12. Xin, H., Zhong, C., Nudleman, E. & Ferrara, N. Evidence for Pro-angiogenic Functions of  
659 VEGF-Ax. *Cell* **167**, 275-284 e6 (2016).
- 660 13. Dever, T.E., Dinman, J.D. & Green, R. Translation Elongation and Recoding in Eukaryotes.  
661 *Cold Spring Harb Perspect Biol* **10**(2018).
- 662 14. Gesteland, R.F. & Atkins, J.F. Recoding: dynamic reprogramming of translation. *Annu Rev*  
663 *Biochem* **65**, 741-68 (1996).
- 664 15. Bidou, L., Rousset, J.P. & Namy, O. Translational errors: from yeast to new therapeutic  
665 targets. *FEMS Yeast Res* **10**, 1070-82 (2010).
- 666 16. Firth, A.E. & Brierley, I. Non-canonical translation in RNA viruses. *J Gen Virol* **93**, 1385-  
667 409 (2012).
- 668 17. Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J. & Brierley, I. A mechanical explanation of  
669 RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244-7  
670 (2006).
- 671 18. Barry, J.K. & Miller, W.A. A -1 ribosomal frameshift element that requires base pairing  
672 across four kilobases suggests a mechanism of regulating ribosome and replicase traffic  
673 on a viral RNA. *Proc Natl Acad Sci U S A* **99**, 11133-8 (2002).

- 674 19. Cimino, P.A., Nicholson, B.L., Wu, B., Xu, W. & White, K.A. Multifaceted regulation of  
675 translational readthrough by RNA replication elements in a tombusvirus. *PLoS Pathog* **7**,  
676 e1002423 (2011).
- 677 20. Xu, Y. et al. A Stem-Loop Structure in Potato Leafroll Virus Open Reading Frame 5 (ORF5)  
678 Is Essential for Readthrough Translation of the Coat Protein ORF Stop Codon 700 Bases  
679 Upstream. *J Virol* **92**(2018).
- 680 21. Naphine, S. et al. Protein-directed ribosomal frameshifting temporally regulates gene  
681 expression. *Nat Commun* **8**, 15582 (2017).
- 682 22. Naphine, S. et al. A novel role for poly(C) binding proteins in programmed ribosomal  
683 frameshifting. *Nucleic Acids Res* **44**, 5491-503 (2016).
- 684 23. Leung, D.W., Cachianes, G., Kuang, W.J., Goeddel, D.V. & Ferrara, N. Vascular  
685 endothelial growth factor is a secreted angiogenic mitogen. *Science* **246**, 1306-9 (1989).
- 686 24. Melincovici, C.S. et al. Vascular endothelial growth factor (VEGF) - key factor in normal  
687 and pathological angiogenesis. *Rom J Morphol Embryol* **59**, 455-467 (2018).
- 688 25. Loughran, G., Howard, M.T., Firth, A.E. & Atkins, J.F. Avoidance of reporter assay  
689 distortions from fused dual reporters. *RNA* **23**, 1285-1289 (2017).
- 690 26. Grentzmann, G., Ingram, J.A., Kelly, P.J., Gesteland, R.F. & Atkins, J.F. A dual-luciferase  
691 reporter system for studying recoding signals. *RNA* **4**, 479-86 (1998).
- 692 27. Cridge, A.G., Crowe-McAuliffe, C., Mathew, S.F. & Tate, W.P. Eukaryotic translational  
693 termination efficiency is influenced by the 3' nucleotides within the ribosomal mRNA  
694 channel. *Nucleic Acids Res* **46**, 1927-1944 (2018).
- 695 28. Zubradt, M. et al. DMS-MaPseq for genome-wide or targeted RNA structure probing in  
696 vivo. *Nat Methods* **14**, 75-82 (2017).
- 697 29. Vander Meulen, K.A., Davis, J.H., Foster, T.R., Record, M.T., Jr. & Butcher, S.E.  
698 Thermodynamics and folding pathway of tetraloop receptor-mediated RNA helical  
699 packing. *J Mol Biol* **384**, 702-17 (2008).
- 700 30. Hansen, J.L., Schmeing, T.M., Moore, P.B. & Steitz, T.A. Structural insights into peptide  
701 bond formation. *Proc Natl Acad Sci U S A* **99**, 11670-5 (2002).
- 702 31. Brown, J.A. Unraveling the structure and biological functions of RNA triple helices. *Wiley*  
703 *Interdiscip Rev RNA* **11**, e1598 (2020).
- 704 32. Jain, N., Morgan, C.E., Rife, B.D., Salemi, M. & Tolbert, B.S. Solution Structure of the HIV-  
705 1 Intron Splicing Silencer and Its Interactions with the UP1 Domain of Heterogeneous  
706 Nuclear Ribonucleoprotein (hnRNP) A1. *J Biol Chem* **291**, 2331-44 (2016).
- 707 33. Kooshapur, H. et al. Structural basis for terminal loop recognition and stimulation of pri-  
708 miRNA-18a processing by hnRNP A1. *Nat Commun* **9**, 2479 (2018).
- 709 34. Morgan, C.E. et al. The First Crystal Structure of the UP1 Domain of hnRNP A1 Bound to  
710 RNA Reveals a New Look for an Old RNA Binding Protein. *J Mol Biol* **427**, 3241-3257  
711 (2015).
- 712 35. Tolbert, M. et al. HnRNP A1 Alters the Structure of a Conserved Enterovirus IRES Domain  
713 to Stimulate Viral Translation. *J Mol Biol* **429**, 2841-2858 (2017).
- 714 36. Wu, B. et al. Molecular basis for the specific and multivariant recognitions of RNA  
715 substrates by human hnRNP A2/B1. *Nat Commun* **9**, 420 (2018).
- 716 37. Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. & Burd, C.G. hnRNP proteins and the  
717 biogenesis of mRNA. *Annu Rev Biochem* **62**, 289-321 (1993).

- 718 38. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in  
719 health and disease. *Hum Genet* **135**, 851-67 (2016).
- 720 39. Han, S.P., Tang, Y.H. & Smith, R. Functional diversity of the hnRNPs: past, present and  
721 perspectives. *Biochem J* **430**, 379-92 (2010).
- 722 40. Kwak, H., Park, M.W. & Jeong, S. Annexin A2 binds RNA and reduces the frameshifting  
723 efficiency of infectious bronchitis virus. *PLoS One* **6**, e24067 (2011).
- 724 41. Gao, F. & Simon, A.E. Multiple Cis-acting elements modulate programmed -1 ribosomal  
725 frameshifting in Pea enation mosaic virus. *Nucleic Acids Res* **44**, 878-95 (2016).
- 726 42. Wu, M.T. & D'Souza, V. Alternate RNA Structures. *Cold Spring Harb Perspect Biol*  
727 **12**(2020).
- 728 43. Vicens, Q., Kieft, J.S. & Rissland, O.S. Revisiting the Closed-Loop Model and the Nature of  
729 mRNA 5'-3' Communication. *Mol Cell* **72**, 805-812 (2018).
- 730 44. Yoffe, A.M., Prinsen, P., Gelbart, W.M. & Ben-Shaul, A. The ends of a large RNA molecule  
731 are necessarily close. *Nucleic Acids Res* **39**, 292-9 (2011).
- 732 45. Lai, W.C. et al. mRNAs and lncRNAs intrinsically form secondary structures with short  
733 end-to-end distances. *Nat Commun* **9**, 4328 (2018).
- 734 46. Adivarahan, S. et al. Spatial Organization of Single mRNPs at Different Stages of the  
735 Gene Expression Pathway. *Mol Cell* **72**, 727-738 e5 (2018).
- 736 47. Khong, A. & Parker, R. mRNP architecture in translating and stress conditions reveals an  
737 ordered pathway of mRNP compaction. *J Cell Biol* **217**, 4124-4140 (2018).
- 738 48. Harper, S. & Speicher, D.W. Purification of proteins fused to glutathione S-transferase.  
739 *Methods Mol Biol* **681**, 259-80 (2011).
- 740 49. D'Souza, V., Dey, A., Habib, D. & Summers, M.F. NMR structure of the 101-nucleotide  
741 core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol* **337**, 427-42  
742 (2004).
- 743 50. Norris, M., Fetler, B., Marchant, J. & Johnson, B.A. NMRFX Processor: a cross-platform  
744 NMR data processing program. *J Biomol NMR* **65**, 205-216 (2016).
- 745 51. Guntert, P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**,  
746 353-78 (2004).
- 747 52. Schwieters, C.D., Kuszewski, J.J., Tjandra, N. & Clore, G.M. The Xplor-NIH NMR molecular  
748 structure determination package. *J Magn Reson* **160**, 65-73 (2003).
- 749 53. Schwieters, C.D. & Clore, G.M. Using small angle solution scattering data in Xplor-NIH  
750 structure calculations. *Prog Nucl Magn Reson Spectrosc* **80**, 1-11 (2014).
- 751 54. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J.S. Genome-wide probing  
752 of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701-5  
753 (2014).
- 754 55. Dyer, K.N. et al. High-throughput SAXS for the characterization of biomolecules in  
755 solution: a practical approach. *Methods Mol Biol* **1091**, 245-58 (2014).
- 756 56. Forster, S., Apostol, L. & Bras, W. Scatter: software for the analysis of nano- and  
757 mesoscale small-angle scattering. *Journal of Applied Crystallography* **43**, 639-646 (2010).
- 758

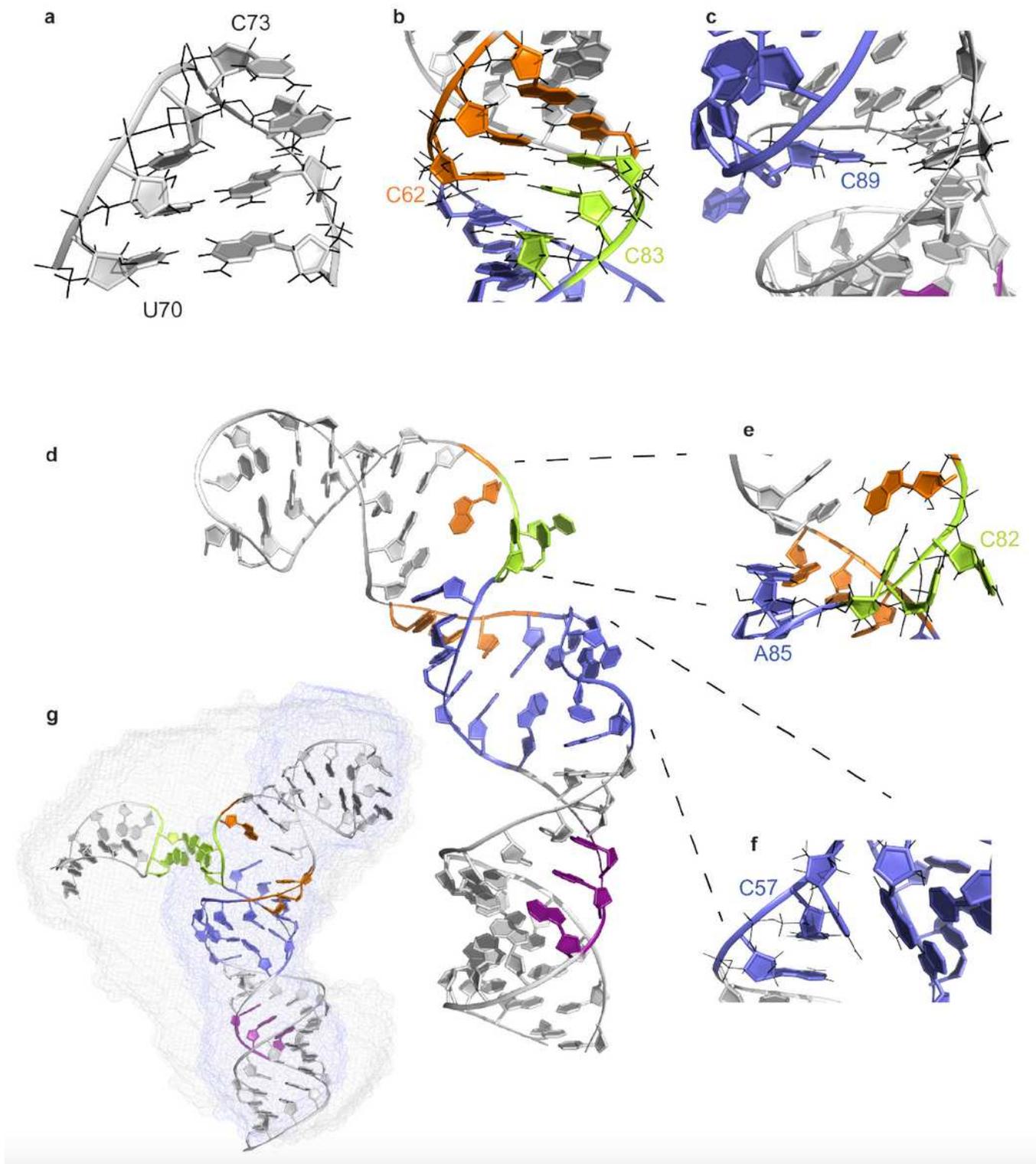
# Figures



**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

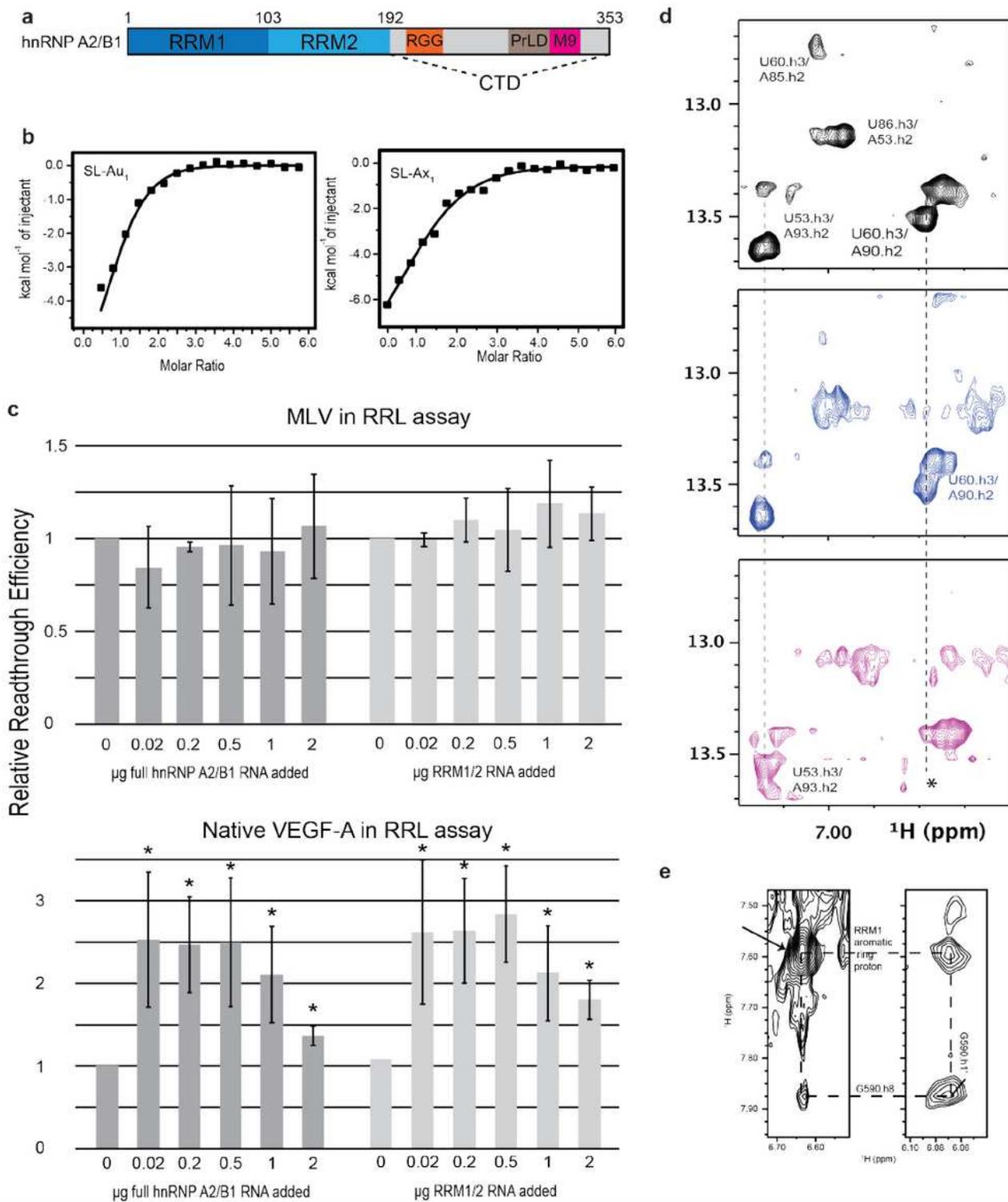
bars indicate standard error (n=3) and statistical tests were performed using a 2-tailed t-test (\* p < 0.05, \*\* p < 0.005). (c) The secondary structure reveals an extended stem-loop for SLAu 1, including an alternate conformation (bases shown in orange and blue for simplicity), as well as shorter stem-loops SL-Ax1-3 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-Au1 is able to specifically interact with SL-Ax2 (black squares), but not with SL-Ax3 (white squares). Tertiary folds are shown for SL-Au1 (PDB 7KUB) (e), SL-Ax1 (PDB 7KUC) (f), and SLAx 2 (PDB 7KUD) (g) with the same base coloring as in (c). (h) A model was created showing the docking interaction between SL-Au1 and SL-Ax2.



**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 A81CCA 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

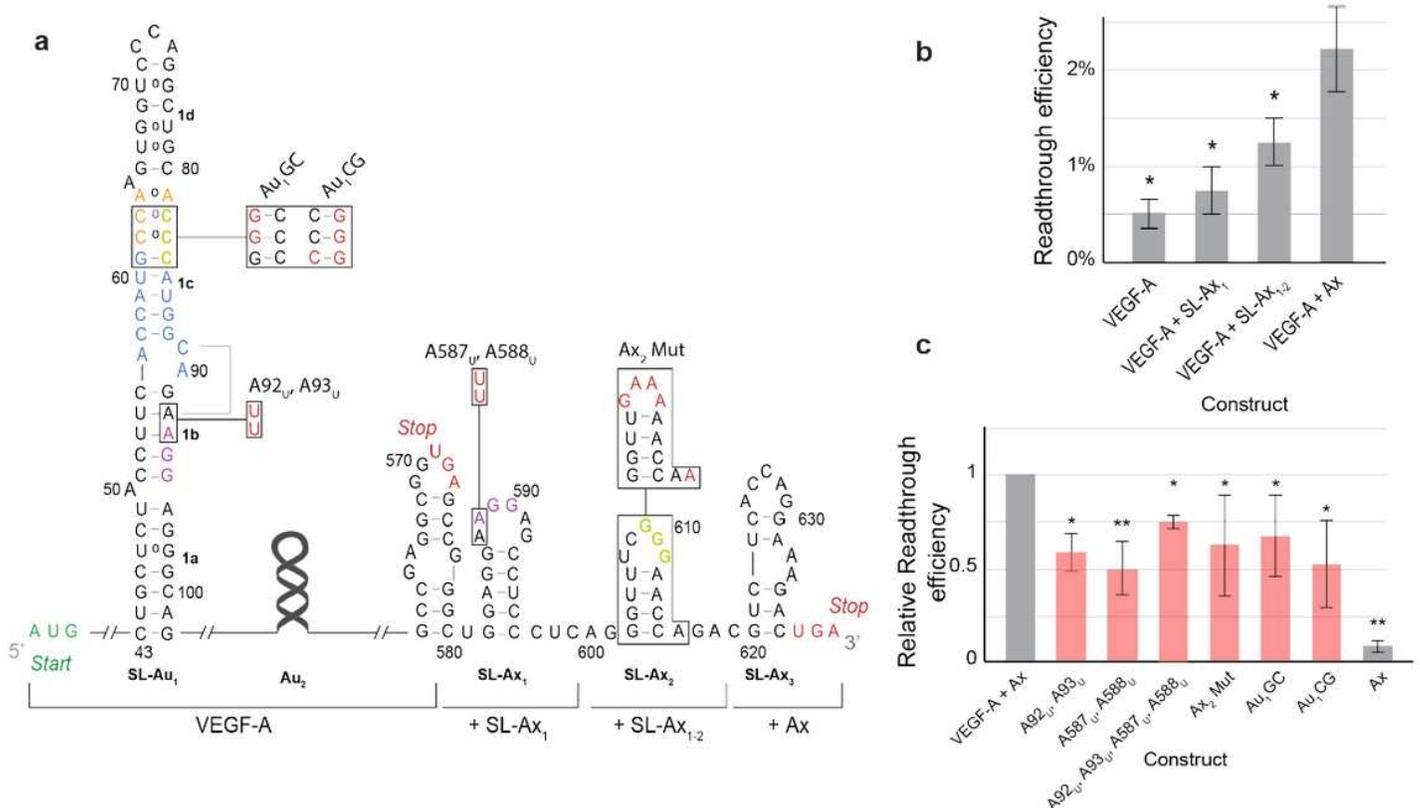
the bend described above. It also overlays with a model envelope (grey) of the SL-Au1:SL-Ax2 long-range interaction complex. Base coloring is as in Figure 1.



**Figure 3**

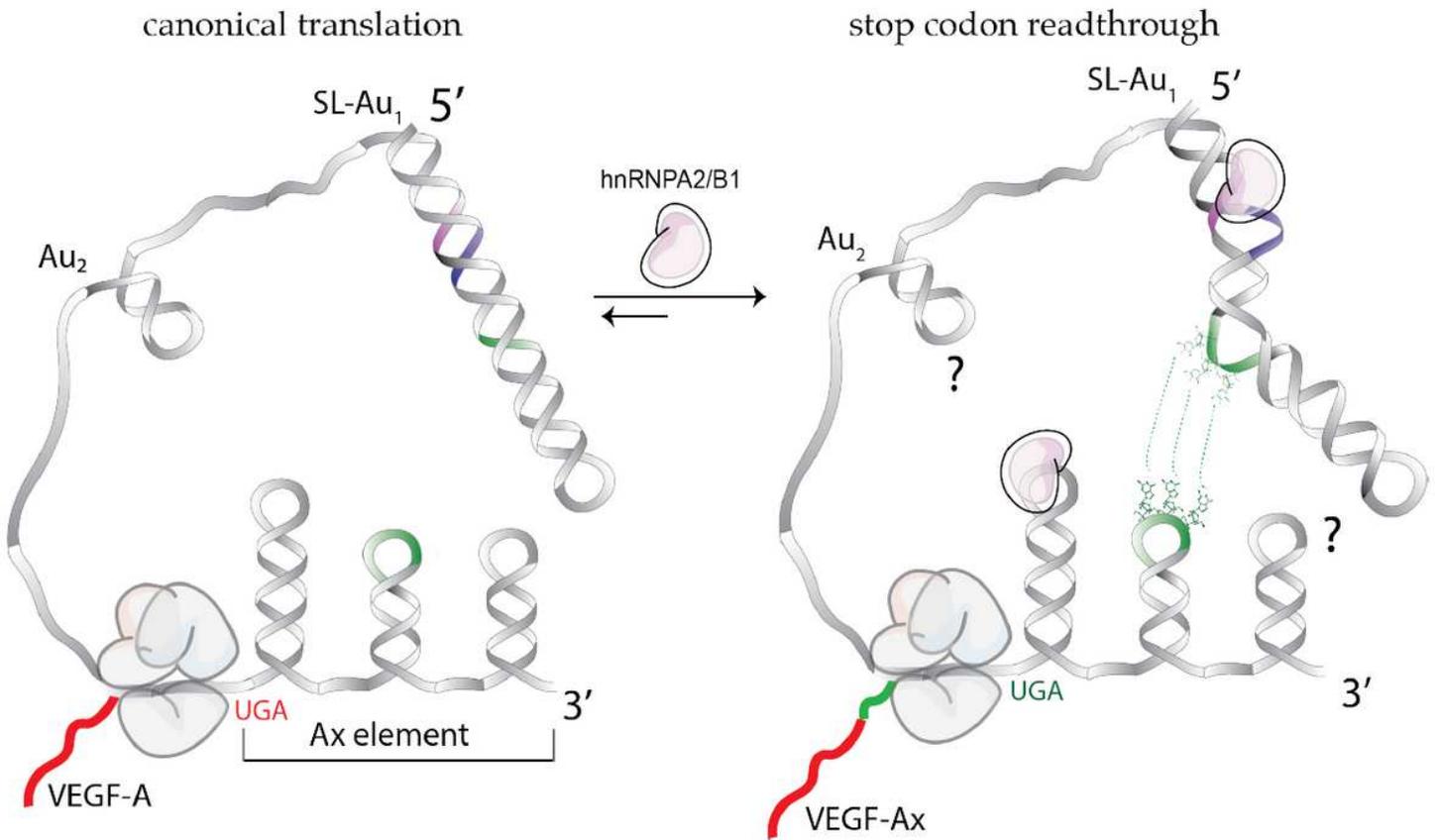
hnRNP A2/B1 binds to VEGF-A mRNA to help mediate readthrough. (a) Schematic of the hnRNP A2/B1 domain architecture showing its RNA-recognition motifs (RRM1 and RRM2) and C terminal domain (CTD), consisting of an Arg-Gly-Gly box (RGG), prion-like domain (PrLD), and M9 transport sequence. (b)

ITC binding studies of RRM1/2 to SL-Au1 (left) or SL-Ax1 (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-Au1 at 280 K shows the shifting of several 1H-1H NMR peaks of base pairs disrupted by protein binding. (e) Titration of RRM1 into SL-Ax1 at 311K results in 1H-1H NOE cross-peaks between G590 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.