Title: Single chain models illustrate the 3D RNA folding shape during translation

Authors: Tianze Guo[1,†], Olivia L. Modi[1,†], Jillian Hirano[1,†], Horacio V. Guzman[2], Tatsuhisa Tsuboi[1,3,*]

Affiliations:

[1]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92023-0358, USA.

[2]Department of Theoretical Physics, Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

[3]Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Shenzhen 518055, China

† These authors contributed equally to this publication

* To whom correspondence should be addressed:

Tatsuhisa Tsuboi

Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School,

University Town of Shenzhen, Nanshan, Shenzhen 518055, China

Tel: +86-152-1679-5216

E-mail address: ttsuboi@sz.tsinghua.edu.cn

Abstract:

The three-dimensional conformation of RNA is important in the function and fate of the molecule. The common conformation of mRNA is formed based on the closed-loop structure and internal base pairings with the activity of the ribosome movements. However, recent reports suggest that the closed-loop structure might not be formed in many mRNAs. This implies that mRNA can be considered as a single polymer in the cell. We developed TRIP; *Three-dimensional RNA Illustration Program*, to model the three-dimensional RNA folding shape based on single-chain models. We identified the angle restriction of each bead component from previously reported single-molecule FISH experimental data. This simulation method was able to recapitulate the mRNA conformation change of the translation activity and three-dimensional positional interaction between organelle and its localized mRNAs. Within the analyzed cases base-pairing interactions only have minor effects on the three-dimensional mRNA conformation, and instead single-chain polymer characteristics have a more significant impact on the conformation. This method will be used to predict the aggregation mechanism of mRNA and long noncoding RNA in specific cellular conditions such as nucleolus and phase-separated granules.

Keywords

RNA folding, RNA secondary structure, translation, modelling, bead-chain models

Running Title

TRIP illustrates the 3D RNA folding shape

Introduction

RNAs are essential nucleic acids that convey genetic information, especially messenger RNAs (mRNA), which can be translated by ribosomes to produce proteins. RNA consists of single-stranded nucleic acid polymers that are connected to each other by forming phosphodiester bonds in the 5' to 3' direction. mRNA organization is important for many aspects of mRNA metabolism, particularly steps where different regions with (pre-) mRNAs are thought to communicate, such as splicing, translation regulation, or miRNA-mediated regulation (Fabian and Sonenberg, 2012; Imataka et al., 1998; Tarun and Sachs, 1996). However, despite the importance of mRNA organization, there is little known about how ribonucleoprotein complexes (mRNPs) are organized as three-dimensional assemblies. It has been observed that mRNAs exist in different levels of compaction depending on their translating state. These translating mRNAs are believed to exist in a closed-loop conformation where the 5' and 3' ends are brought together through the cap-binding elF4F complex and the poly(A) binding protein PABPC1 (Christensen et al., 1987; Imataka et al., 1998; Tarun and Sachs, 1996; Wells et al., 1998). Furthermore, since ribosomes will attach to mRNA during translation, this will add an unfolded and flat area to the strand, and the translation of mRNAs results in the separation of the 5' and 3' ends. This separation contentiously suggests that these RNAs are not translated in a stable closed-loop (Adivarahan et al., 2018).

In order to simulate the general pattern of RNA folding conformation, we adopted single-chain models and tested different angle restrictions. Such models were applied to break down the complicated structures of cytoskeletons and better understand aspects of the whole structure by considering smaller fragments (Wen and Janmey, 2011). Each element of the RNA strand can be

67 categorized by the different links of a polymer-like arrangement of beads, and it simplifies the
68 analysis of the whole function of the RNA. In addition, it draws a relationship between the effect
69 of confinement on the end-to-end distances of single-chain polymers and the stiffness of each
70 fragment (Doi, 1996). Even though this relationship is based on arbitrary numbers, it is able to
71 characterize the generalized behavior and categorize the type of regime the single-chain falls
72 into, which in turn allows an improved understanding of RNA conformation for the tackled cases
73 and similar.
74
75 In this work, we applied the single-chain models for mRNA folding and developed the *T*hree-
76 dimensional *R*NA *I*llustration *P*rogram (TRIP), a simplified model designed to rapidly interpret
77 the end-to-end distances. The 5' to 3' distance of mRNA when treated with Puromycin, a known
78 antibiotic that prematurely ends translation and disassembles polysomes, follows a Gaussian
79 distribution with a skewness to the right (Adivarahan et al., 2018). Furthermore, ribosome
80 occupancy determines the compaction of mRNAs; and the number of ribosomes positively
81 correlated with the length of the mRNA (Adivarahan et al., 2018). When comparing our initial
82 calculations with this single molecule Fluorescent in situ hybridization (smFISH) study, our
83 findings gave an in-silico demonstration of this conclusion by simulating a chain of mRNA with
84 and without ribosomes attached. Three-dimensional conformation of the mRNA is assembled by
85 using a bent angle range of ±140°. Our single-chain models also recapitulated the distance
86 between cytoplasmic mRNA and the mitochondria, which were previously analyzed in fixed
87 cells by smFISH (Jourdren et al., 2010), and in live cells through an MS2-MCP method (Tsuboi
88 et al., 2020). We also found that integration of the information of the local secondary structure,
89 as it is the case of the tackled MS2-sequence, does not have much influence in the single chain

model. While there is little known about mRNA's three-dimensional organization in the cell, our simulation introduces a first approximation on how polymer-like models can rapidly estimate end-to-end distances of mRNAs in the cellular condition. In a nutshell, our very coarse model shows being useful for interpreting mRNA experiments devising end-to-end distances of the fragment.

Results

Modeling 3D RNA folding by TRIP; *T*hree-dimensional *R*NA *I*llustration *P*rogram

mRNAs are biopolymers consisting of thousands of nucleotides connected to each other from the 5' to 3' ends. The RNA backbone is rotameric (Murray et al., 2003). For each residue along the RNA backbone, there are six angles (α, β, γ, δ, ε, ζ) that display the rotation of the six bonds in each nucleotide. This suggests that RNA molecules can adopt complicated shapes because of the different combinations of angles that may occur. The single-chain bead model can be utilized to simplify the biopolymer by disregarding the atomistic details and considering each monomer unit as a structureless segment with a given length defined by monomer units and angles between monomers. By modeling three-dimensional RNA folding using TRIP; *T*hree-dimensional *R*NA *I*llustration *P*rogram, we analyzed the end-to-end distances by taking into consideration two biopolymer's states one during non-translating condition and the other during translational elongation. These two states are similar to the conditions in which cells are treated with and without puromycin, translational inhibitor, prior to the formaldehyde fixation for the *in vivo* experimental measurements of end-to-end distance by the smFISH. In TRIP, we assumed single nucleotides as discrete monomers connected to each other (Figure 1A). The monomer length was set to be 0.59 nanometers based on an average distance between two nucleotides (Liphardt et al.,

113 2001). The confinement of the angles is applied to the polar angle θ, the angle projection
114 referenced to the y-z plane, and azimuthal angle φ, the angle projection referenced to the x-y
115 plane. The set of (θ, φ) is randomly chosen from a defined angle range. We placed the first bead,
116 the head of a single-chain, at the origin of a spherical coordinate and let it be the 5' end, and the
117 next bead was randomly picked from the outer surface of the sphere with a radius of 0.59 so that
118 the first monomer, as a vector in the coordinate system, is generated with a length and direction.
119 Based on the first vector's direction, the next vector was created by setting the first bead as the
120 origin and the second bead's direction depending on the first vector's direction within a
121 confinement degree range (Figure 1B). This procedure was reiterated to create a strand of RNA.
122
123 While the monomer length was set, the angle range significantly affected the end-to-end
124 distances. We used TRIP to visualize samples that were 100-nucleotides long and applied
125 different angle ranges. It was observed especially at low angle ranges, such as 10 degrees, that
126 the RNA strand will extend far; when angle ranges increase, the RNA strand tends to compact
127 (Figure 1C, Movie 1). To identify the best angle range for TRIP, we compared the modified 2D
128 TRIP results (Methods) to reported 2D 5' to 3' length for three different mRNAs: *MDN1*,
129 *POLA1*, and *PRPF8* in non-translating conditions. Previous smFISH experiments tested the 5' to
130 3' distances in 3D and reported their result in 2D using image flattening techniques. The reported
131 data were 35.95 nanometers (*MDN1*), 33.15 nanometers (*POLA1*), and 34.19 nanometers
132 (*PRPF8*) respectively (Adivarahan et al., 2018). Because the measurement of length in smFISH
133 is conducted by the distance between the centers of fluorescent signals from multiple probes, the
134 nucleotides on the edge of the fluorescent signals may not be taken into account. To accurately
135 input the experimental length for the simulation, we defined the net nucleotide number as the

136 length between the centers of the FISH probes on the 5' and 3' ends. Different angle ranges were
137 applied to the simulation for three mRNA references, with net nucleotides numbers 16350
138 (*MDN1*), 4060 (*POLA1*), and 4969 (*PRPF8*). As shown in the graph, there is a rapid decrease in
139 the TRIP length first seen around ±10 to ±30 degrees, then at around ±100 degrees, the distances
140 start to level off (Figure 1D). The results refer back to the observation that an increase in angle
141 range leads to compaction. At around ±140 degrees the TRIP results showed high
142 correspondence with the referenced values. To identify the best angle range, we analyzed the
143 sum of the difference between the TRIP results and the previously reported smFISH data
144 (Adivarahan et al., 2018) in each angle range (Figure 1E). By magnifying the larger angle range
145 region, we confirmed that at ±140 degrees angle ranges, the sum of the difference is the least
146 among all the angle ranges we tested (Figure 1E). We further confirmed if the ±140 degrees
147 angle range could explain non-coding RNAs, which do not associate with ribosomes. Previous
148 smFISH study showed that the non-coding RNAs' end-to-end proximity was 41.33 nm for *OIP5-*
149 *AS1* and 34.49 nm for *TUG1* (Adivarahan et al., 2018). The TRIP successfully recaptured these
150 smFISH data as 40.93 nm for O*IP5-AS1* and 33.77 nm for *TUG1* respectively (Figure 1F).
151
152 **mRNA end-to-end distance positively correlates with numbers of ribosome attached**
153 Ribosomes cover approximately 30 nucleotides of mRNA and unfolded and flattened mRNA
154 during translation (Lareau et al., 2014). We tested if translation influences the mRNA
155 conformation using TRIP. To analyze the relationship of translation activity and mRNA
156 conformation, we used the 6636-nucleotide long SINAP mRNA, in which end-to-end proximity
157 distance has been reported with a different number of ribosomal activities by smFISH (Figure
158 2B) (Adivarahan et al., 2018). The 5072 net nucleotides numbers were analyzed using

159 fluorescent FISH probes at its 5' and 3' ends for this mRNA (Figure 2B). The open reading
160 frame length of this mRNA was 4917 nucleotides long (Figure 2B). It has been shown that the
161 maximum ribosome-bound number was 20 for 3129 nucleotides long open reading frame in
162 SINAP mRNA previously (Wu et al., 2016). We estimated the maximum number of ribosomes
163 for 4917 nucleotides open reading frame as 32 ribosomes by calculating the ratio. We conducted
164 the *in silico* experiment for five groups, each with 0, 8, 16, 24, and 32 ribosomes attached,
165 similar to what was shown in the FISH experiment. We observed a stepwise increase in length
166 with evenly distributed intervals according to increasing numbers of ribosomes (Figure 2C). We
167 also observed that the mRNA length distribution was a Gaussian distribution with a skewness
168 toward the longer distance (Figure 2C). The relationship between distance and ribosome number
169 suggests that more intense translation activity can increase the end-to-end distance, thus the
170 molecule would be less compact since unfolded mRNA areas are greater.
171
172 **The ribosome occupation explains the separation of 5' end and 3' end of mRNA in different**
173 **mRNAs**
174 We further tested if TRIP could recapitulate the impact of translation to end-to-end proximity
175 distance using the three different mRNAs: *MDN1*, *POLA1*, and *PRPF8* mRNA referring to the
176 previously reported smFISH data (Adivarahan et al., 2018). A significant difference of 5' to 3'
177 proximity distance has been observed between the conditions of puromycin treated and untreated
178 HEK293 cells (Figure 2D). These differences indicate that the translation contributes to the
179 unfolding of mRNA conformations. We estimated the ribosome numbers as the average
180 translational activity which is the length of the open reading frame divided by 200 nucleotides,
181 where 200 is the average inter-ribosome distance on mRNA reported in human HEK293 and
182 U2OS cells (Yan et al., 2016). Using this average number of ribosomes in all three mRNAs, we

183 simulated the end-to-end proximity distance to compare with the previously reported smFISH
184 results (Figure 2D) (Adivarahan et al., 2018). In all three cases, the TRIP predicted a significant
185 difference between non-translating and translating conditions (Figure 2E). This result matched
186 what was expected from in-vivo environments, in which the translation of these mRNAs resulted
187 in the separation of the 5' and 3' ends (Adivarahan et al., 2018). The distribution of TRIP results
188 follows a Gaussian distribution with a skewness to the larger distance, meaning a larger density
189 was found at lower values. The compaction effect is found often in non-translating conditions,
190 where the 5' to 3' end-to-end distance is significantly reduced from translating conditions.
191 Interestingly, the end-to-end proximity for *POLA1* mRNA at translation condition (95.89 nm)
192 was reported longer than *PRPF8* (92.41 nm) in the referenced data, despite *PRPF8* having a
193 longer open reading frame (Adivarahan et al., 2018). However, the TRIP output predicted that
194 *POLA1*-untreated is lower than *PRPF8*-untreated. This discrepancy might be due to *POLA1*
195 possibly having higher translatability. We estimated that 40 ribosomes were required to reach the
196 *POLA1* end-to-end proximity distance to the reported distance 95.13 nanometers by smFISH,
197 which is 1.8 times greater translatability. In addition to comparing results shown in 2D space, we
198 further investigated the ribosomal occupation on 3D data sets. Similar experiments using
199 smFISH probes to detect end-to-end distance were performed to demonstrate the end to end
200 distances of two sample mRNAs with long open reading frames under translating conditions and
201 puromycin conditions: AHNAK (17673 nt ORF) and DYNC1N1 (13941 nt ORF) (Khong and
202 Parker, 2018). The median is shown from their cumulative distribution plots as: AHNAK,
203 translating conditions ~190nm, and puromycin conditions ~75nm; DYNC1H1, translating
204 conditions ~200nm, and puromycin conditions median ~60nm. AHNAK and DYNC1H1 were
205 simulated with net nucleotides numbers 17003 nts and 12160 nts according to the position of

206 smFISH probes used in previous research (Khong and Parker, 2018). The TRIP output showed a
207 similar trend when it was in 2D; it predicted a significant difference between non-translating and
208 translating conditions (Figure 2F, G). Moreover, for AHNAK in translating conditions, the
209 median is similar (173.26 nm) to the experimental data. For DYNC1H1 in translating conditions,
210 the TRIP output shows a slight underestimation; the median (156.50 nm) is lower than the
211 experimental data. This might be caused by the lower translatability of DYNC1H1. For both
212 mRNAs in non-translating conditions, the TRIP outputs recapitulate the in vitro data: for
213 AHNAK, the median gave 78.93nm; for DYNC1H1, the median gave 67.92nm (Figure 2G). In
214 summary, the *in-silico* TRIP experiment was able to recapitulate that mRNA in a non-translation
215 state follows a geometrically compact shape, whereas translation significantly separates the 5'
216 and 3' ends.
217
218 **Local Secondary structure has minor effects in the studied cases**
219 We tested how local secondary structures affect long mRNAs based on the lack of formation of
220 circular structured mRNA. MS2 was used as a secondary structure in an effort to observe the
221 way the mRNA folds. MS2 is a bacteriophage that consists of 3569-nucleotides of single-
222 stranded RNA. It is encoded with four proteins: the maturation protein, the lysis protein, the
223 replicase protein, and the coat protein (Fiers et al., 1976). The coat protein (MCP) is useful for
224 the detection of RNA within living cells. For mRNA detection, MS2 binding sites (MBS) are
225 inserted in the 3'UTR of an mRNA of interest, and the co-expression of MCP fused with
226 fluorescent proteins renders single mRNAs visible using wide-field epifluorescence microscopy
227 (Tutucci et al., 2018). Between the mRNA and MCP, there should be a secondary structure
228 created by MS2. We predicted that by incorporating the information of the secondary structure,

229 we would be able to refine the predicted RNA length. Each MS2 stem-loop structure was
230 comprised of 17-nucleotides. This insinuates that the mRNA had an absolute experimental length
231 of 1787-nucleotides and the structured mRNA length, which removed the 12 stem-loop
232 sequences, was 1583-nucleotides (Figure 3A). It has been observed that there was a mean
233 observed distance of 48 nanometers between the MCP-ORF in two-dimensional (Eliscovich et
234 al., 2017). While the addition of this secondary structure could provide a refined predicted
235 mRNA length, our analysis reveals that base-pairing interactions only have minor effects (2.3 nm
236 for 204 nt) on the shape (Figure 3B), meaning that the effects of the MS2 secondary structure are
237 less. This suggests that local secondary structure has minor effects on the tackled mRNA
238 conformation.
239
240 **Determination of the shape and length of a mitochondria localized mRNA**
241 Analyzing the proximity of nuclear-encoded mRNA to the mitochondrial outer membrane is
242 challenged with long-term cycloheximide treatment by electron cryotomography (Gold et al.,
243 2017). The distance between cytoplasmic mRNA and the mitochondria was also analyzed by
244 visualizing single-molecule mRNA in fixed cells through smFISH (Jourdren et al., 2010), and in
245 live cells using the MS2-MCP method (Tsuboi et al., 2020). We tested if the TRIP, which is a
246 non-closed loop structure model, could recapture and explain the previous observations. To
247 model such mRNA conformation, we defined the beads (i.e., nucleotides are not formed under
248 the mitochondrial surface plane). To conduct the simulation, the x-y plane was set up as the
249 "surface", so that the result's z-coordinate will never be less than zero. This represents a strand
250 of mRNA growing on the x-y surface. A previous study with *TIM50* mRNA, a mitochondrially-
251 localized mRNA (Figure 3C), has observed that the distance to the mitochondria follows a

252 Gaussian distribution skewed to the larger distance, and the maximum distance is defined at 190
253 nanometers (Tsuboi et al., 2020). Using this information, we conducted the in-silico modeling by
254 TRIP. 154 nucleotides were found as the average distance between ribosomes in budding yeast,
255 *Saccharomyces cerevisiae* (Hurowitz and Brown, 2003; Zenklusen et al., 2008) and that *TIM50*
256 mRNA's open reading frame length is 1851 nucleotides. With this information, we estimated
257 that 12 (=1851/154) ribosomes exist per mRNA. The visualization results of TRIP showed that a
258 strand of mRNA tail-attached to the surface in orange color ($z = 0$) (Figure 3D). The TRIP
259 results indicated the proximity distance between the mRNA 3' and mitochondrial surface as a
260 Gaussian distribution, with a skewness to the larger distance (Figure 3E). The median was at
261 63.27 nanometers and the maximum distance was 196.46 nanometers. The median value was far
262 off from the experimental data (Tsuboi et al., 2020). It is possibly due to the inclusion of the non-
263 associated mRNAs, which shows the distance of more than 190 nanometers, in the experimental
264 data. The maximum distance was close to the previously reported experimental threshold value
265 of 190 nanometers. This was also consistent with the previous FISH experiment showing the
266 distance between mRNA localized to mitochondria and ribosomal RNA of mitochondrial matrix
267 measured as 150nm to 200nm range (Jourdren et al., 2010). This suggests that the mitochondrial
268 localized mRNA is also not forming a closed loop structure.
269

270 **Discussion**

271 In this work, we presented that mRNA three-dimensional conformation can be approached as a
272 simple single-chain bead model. We developed TRIP; Three-dimensional RNA Illustration
273 Program and identified that 140 degrees is the restricted angle of each component of the single-
274 chain to reproduce the previously reported experimental data set (Figure 1). We further tested the

275 translational state by modeling the mRNA conformation to be unfolded and flat when the
276 ribosome is associated (Figure 2). By analysing these sets of MS2 structure, we showed that the
277 secondary structure, especially in the case of local and minor percentage of base paring in the
278 mRNA fragments, would have a minimum effect on the mRNA end-to-end distance (Figure 3A,
279 B). Lastly, we applied TRIP to predict mRNA conformation on the mitochondrial surface where
280 it is experimentally difficult to address (Figure 3C-E). This methodology will be further applied
281 to predict the aggregation mechanism for mRNA and noncoding RNA in specific cell
282 organizations, such as the nucleolus and phase-separated granules.
283
284 The six rotatable torsion angles of the RNA backbone were analyzed through vector quantization
285 (Hershkovitz et al., 2006), multiresolution approach (Hsiao et al., 2006), and quality-filtering
286 techniques (Murray et al., 2003) and have met in consensus. The analysis of the six torsion
287 angles together provides quantifications of the nucleic acids' helical shape on an atomic level,
288 and it also suggests that RNA can form various numbers of structures depending on the angles.
289 However, at a coarser level this has to provide the torsion angle between the whole nucleotides.
290 In our model we assumed each nucleotide position as beads, and instead of accounting for
291 microscopic torsion angles between each atomic bond, we defined our set of bent angles (θ, φ) in
292 a macroscopic view in which we assume and model each nucleotide broadly as observable
293 clusters. The set of bent angles (θ, φ) explained that the relative position of each neighboring
294 nucleotide is bent within 140 degrees positive or negative referring to the defined plane. We
295 showed that an angle range of 140 degrees still gives a variety of RNA lengths based on our
296 model. Compared to double helix structures like DNA, where each nucleotide is paired and
297 forms a stable structure, RNA is relatively unstable and tends to compact or form secondary

298 structures, thus, it may have smaller angles between each nucleotide. DNA on average has 10.5
299 bp/turn (Levitt, 1978), from geometrical analysis it suggests that each angle between nucleotides
300 on average is approximately 145.71° (= $\frac{180° \times (10.5-2)}{10.5}$). Our bent angle analysis, which was
301 converted to an angle between nucleotides, ranges from 40° to 180°. Since our distribution was
302 uniform in angle choosing, the statistical average angle between RNA was calculated to be 110°,
303 which is less than that of double helix structures.
304
305 The evidence argues that ribosomes decompact mRNAs. An analysis of mammalian mRNP
306 compaction by smFISH revealed that translating mRNPs are more extended compared to non-
307 translating mRNAs (Adivarahan et al., 2018; Khong and Parker, 2018). The fact that mRNAs
308 under stressed conditions show similar compaction compared to non-translating conditions may
309 symbolize the removal of ribosomes (Khong and Parker, 2018). Translating mRNPs are
310 compacted relative to its contour length (Adivarahan et al., 2018; Khong and Parker, 2018). This
311 compaction is modelled by the formation of secondary structures of mRNA sequences between
312 elongating ribosomes, which is supported by the spacing of ribosomes on mRNAs (Khong and
313 Parker, 2020). Since the compaction of translating mRNPs is not affected by the inhibition of
314 translation elongation, it is possible that stalled ribosomes on the open reading frame are
315 sufficient to decompact mRNA (Khong and Parker, 2018, 2020). Due to spontaneous
316 intramolecular RNA folding taking place rapidly, it can be assumed that mRNA sequences will
317 collapse into RNA secondary structures, which are then unwound by each elongating ribosome
318 (Khong and Parker, 2020). mRNA folding is also influenced by the binding of RNP-BP,
319 including the ATP dependent binding of DEAD/Delhi box proteins (Khong and Parker, 2020).
320 These box proteins, specifically DHH1, catalyse the ATP-dependent folding and remodelling of

321 RNA duplexes. Immunoprecipitation assays reveal that DHH1 interacts with a variety of
322 proteins, such as heat shock proteins, mRNA binding proteins, initiations and elongation
323 translation factors, ribosomal proteins, and metabolic proteins (Marchat et al., 2015).
324
325 RNA's secondary structure is described by the configuration of the base pairings that are formed
326 by the biopolymer. These secondary structures can be divided into stems, which in naturally
327 occurring RNA molecules are made up of five consecutive base pairs, and loops which connect
328 or terminate the stems (Bundschuh and Hwa, 2002). While they locally form the same double
329 helical structure as DNA molecules, RNA molecules are single-stranded, and therefore must fold
330 back onto themselves to gain base pairings (Bundschuh and Hwa, 2002). The complexity of
331 these secondary structures increase with length (Yoffe et al., 2008). We used the secondary
332 structure formed by the MS2 bacteriophage, which is one of the simplest examples of a
333 secondary structure. Our analysis showed MS2-structure only has minor effects on the shape
334 (Figure 3B).
335
336 Intramolecular RNA-RNA base pairing interactions are important for small 5'-3' distances.
337 Previous research argued that the distance between the two ends of single-stranded RNA
338 molecules is small, under the assumption that there are approximately equal proportions of A, C,
339 G, and U (Yoffe et al., 2011). A large number of circle diagrams associated with the secondary
340 structures of many RNA molecules was also examined for different lengths and sequences (Fang,
341 2011). It was revealed that the first and last monomers were always within a few monomers of
342 each other, suggesting that it is essentially impossible to not have a minimum of one set of base
343 pairs that will bring the ends of an RNA together, Furthermore, it was observed that certain RNA

molecules, such as viral genomes (Poblete and Guzman, 2021; Poblete et al., 2021) and certain messenger RNAs, have been under selective pressure to maintain a small distance between the 5' and 3' ends of the molecule (Yoffe et al., 2011). However, this does not prove that biologically functional RNA molecules, such as viral genomes and certain messenger RNAs, have small 5'-3' distance independent of sequence length (Clote et al., 2012). These findings were then experimentally confirmed by means of Förster resonance energy transfer (FRET). Theoretical analyses of a number of randomized and natural RNA sequences suggest that the 5' and 3' ends of long RNAs (1,000-10,000 nts) are always brought in the proximity of a few nanometres of each other regardless of RNA length and sequence because of the intrinsic propensity of RNA to form widespread intramolecular base pairing interactions (Lai et al., 2018). While secondary structures are thought to have a huge impact on mRNA three-dimensional conformation, this depends on the base-pairing percentage of the total mRNA fragment. In particular for the tackled biopolymers, our analysis would point towards base-pairing interactions only having minor effects on mRNA end-to-end distance.

Figure 1

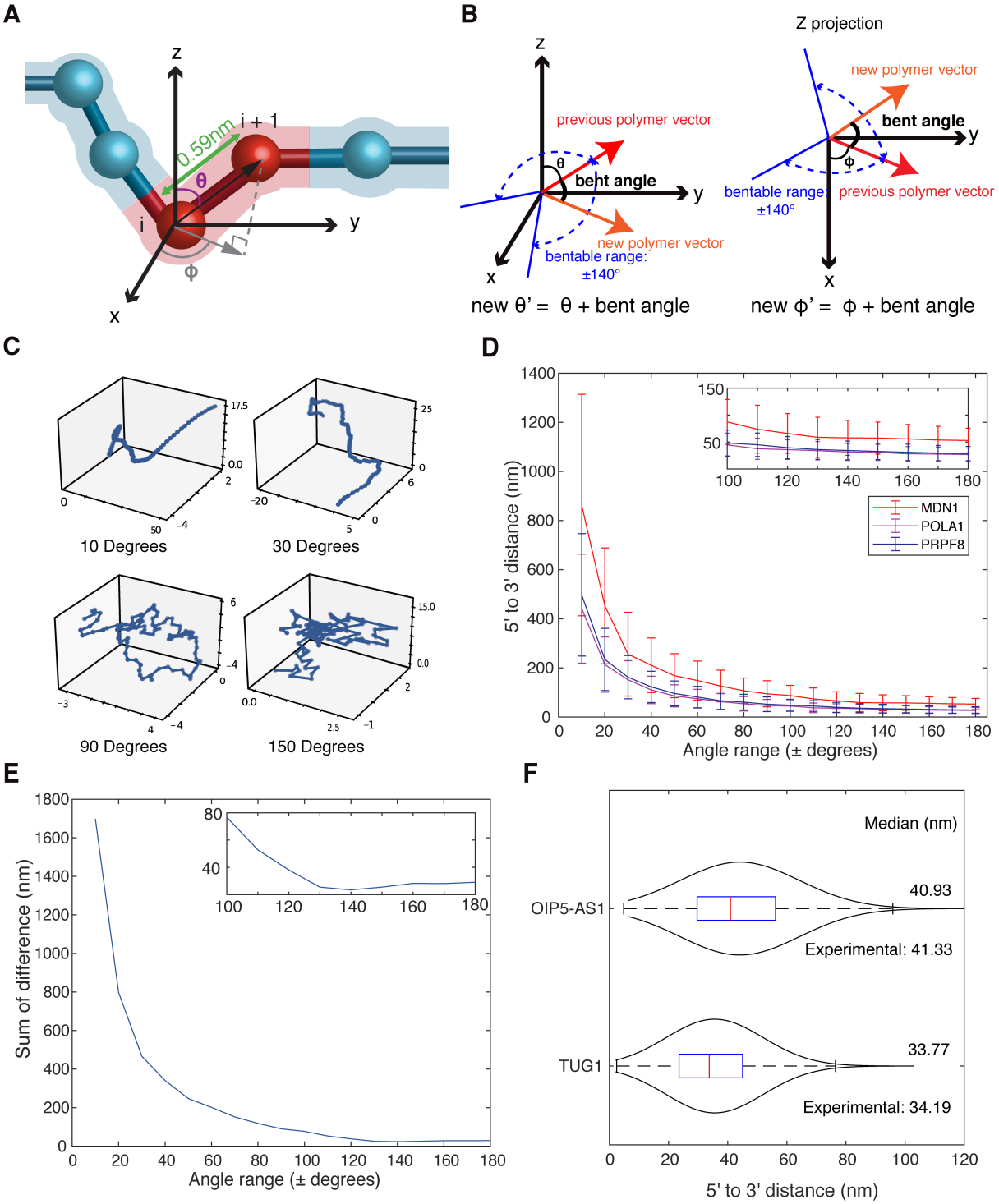
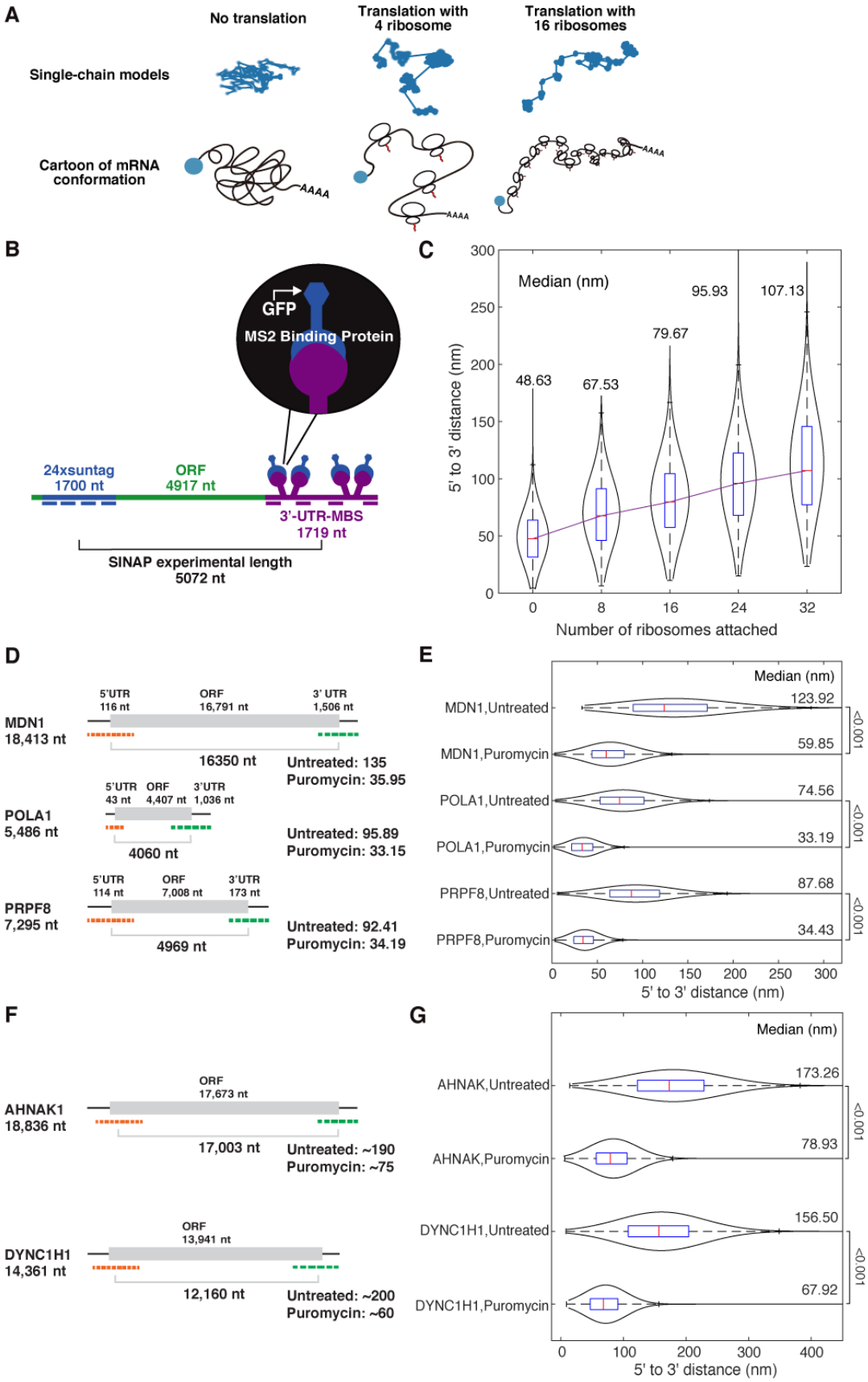


Figure 1 | Simulating RNA 3D structure by RNA single chain models

362 (A) Diagram of TRIP; *T*hree-dimensional *R*NA *Illustration P*rogram, where the bead-chain in
363 red is the area of focus. φ is the azimuthal angle, which is the horizontal angle between the
364 single-chain and origin and measures the projection of the vector (indicated in black arrow) with
365 reference to the +x axis. θ is the polar angle between the +z axis and the vector. The length
366 between beads is 0.59 nm (Adivarahan et al., 2018).
367 (B) The TRIP model is shown in two different projections, with three-dimensional Cartesian
368 coordinates on the left, and Z projection on the right. The original vector is projected θ from the
369 z axis and φ from the x axis, and the angle of the new single-chain vector is restricted by the 140
370 degree bendable range in both axes.
371 (C) Visualization sample of 100 nucleotides RNA in the absence of ribosomes. The blue line
372 indicates the length in nanometers the mRNA moves in each direction. The angle ranges are plus
373 and minus 10, 30, 90, 150 degrees indicated respectively.
374 (D) The relationship of 5' to 3' end-to-end distance simulated in two-dimensional with respect to
375 the chosen angle range for *MDN1* (red), *POLA1* (magenta), and *PRPF8* (blue) mRNAs. Each dot
376 indicates the median 5' to 3' distance on the given angle range, with error bars indicating one
377 standard deviation. The frame on the upper right shows a zoom-in figure at angle range ±100-
378 ±180.
379 (E) Plot of the sum of error, which is calculated as the sum of the distances of the TRIP outputs
380 of three mRNA in (C) at a certain angle range to the previously reported smFISH data
381 (Adivarahan et al., 2018). The frame on the upper right corner shows a zoom-in figure at angle
382 range ±100-±180.
383 (F) 5' to 3' end distance distribution of the non-translating RNAs *TUG1* and *OIP5-AS1* in two-
384 dimensional. The blue box inside the violin plot shows the first quartile, median (red), and third

quartile. The median from the simulation output is displayed on the right. The smFISH data (Adivarahan et al., 2018) is referenced below each plot.


A
No translation
Translation with 4 ribosome
Translation with 16 ribosomes
Single-chain models
Cartoon of mRNA conformation
AAAA
B
GFP
MS2 Binding Protein
24xsuntag 1700 nt
ORF 4917 nt
3'-UTR-MBS 1719 nt
SINAP experimental length 5072 nt
C
Median (nm)
48.63
67.53
79.67
95.93
107.13
5' to 3' distance (nm)
Number of ribosomes attached
D
MDN1 18,413 nt
5'UTR 116 nt
ORF 16,791 nt
3' UTR 1,506 nt
16350 nt
Untreated: 135
Puromycin: 35.95
POLA1 5,486 nt
5'UTR 43 nt
ORF 4,407 nt
3'UTR 1,036 nt
4060 nt
Untreated: 95.89
Puromycin: 33.15
PRPF8 7,295 nt
5'UTR 114 nt
ORF 7,008 nt
3'UTR 173 nt
4969 nt
Untreated: 92.41
Puromycin: 34.19
E
Median (nm)
MDN1,Untreated
123.92
MDN1,Puromycin
59.85
POLA1,Untreated
74.56
POLA1,Puromycin
33.19
PRPF8,Untreated
87.68
PRPF8,Puromycin
34.43
<0.001
5' to 3' distance (nm)
F
AHNAK1 18,836 nt
ORF 17,673 nt
17,003 nt
Untreated: ~190
Puromycin: ~75
DYNC1H1 14,361 nt
ORF 13,941 nt
12,160 nt
Untreated: ~200
Puromycin: ~60
G
Median (nm)
AHNAK,Untreated
173.26
AHNAK,Puromycin
78.93
DYNC1H1,Untreated
156.50
DYNC1H1,Puromycin
67.92
<0.001
5' to 3' distance (nm)

389 **Figure 2 | The ribosome occupation explains the separation of 5' end and 3' end of mRNA**
390 **in different mRNAs**
391 (A) Three-dimensional RNA Illustration Program models the three-dimensional RNA folding
392 shape during no translation, mild translation, and intense translation (4 ribosome and 16
393 ribosome in the case of cartoons).
394 (B) Flat model of the SINAPs mRNA structure and length. The blue dashed line represents
395 24xsuntag FISH probes. The purple dashed line represents 3'-UTR-MBS FISH probes. The
396 experimental length is the absolute mRNA length between the centers of the above two probes
397 (Adivarahan et al., 2018).
398 (C) mRNA end-to-end distance positively correlates with numbers of ribosomes attached. Violin
399 plots showing 5' to 3' distance distribution of the SINAP by TRIP simulation in the order of
400 increasing translation activity in three-dimensional. The blue box inside the violin plot shows the
401 first quartile, median (red), and third quartile. The median from the simulation output is
402 displayed on the top. The purple line connects the data by their medians linearly.
403 (D) Diagram of the mRNAs used in the previously reported experiments (Adivarahan et al.,
404 2018). The bracket indicates the experimental length. The numbers indicated on the right exhibit
405 the length in nanometers when cells are treated and untreated with puromycin in two-
406 dimensional.
407 (E) Violin plots showing 5' to 3' distance distribution of *MDN1*, *POLA1*, *PRPF8* mRNAs from
408 the TRIP simulation for the cells treated and untreated with puromycin. The blue box inside the
409 violin plot shows the first quartile, median (red), and third quartile. The median from the
410 simulation output is displayed on the right. The 5' to 3' distance is simulated in two-dimensional.

411 The number indicated by brackets outside the graph is the KS Test p-value between the two sets
412 of data.
413 (F) Diagram of the mRNAs used in the previously reported experiments (Khong and Parker,
414 2018). The bracket indicates the experimental length. The numbers indicated on the right exhibit
415 the length in nanometers when cells are treated and untreated with puromycin in three-
416 dimensional.
417 (G) Violin plots showing 5' to 3' distance distribution of *AHNAK* and *DYNC1H1* mRNAs from
418 the TRIP simulation for the cells treated and untreated with puromycin. The blue box inside the
419 violin plot shows the first quartile, median (red), and third quartile. The median from the
420 simulation output is displayed on the right. The 5' to 3' distance is simulated in three-
421 dimensional. The number indicated by brackets outside the graph is the KS Test p-value between
422 the two sets of data.
423
424

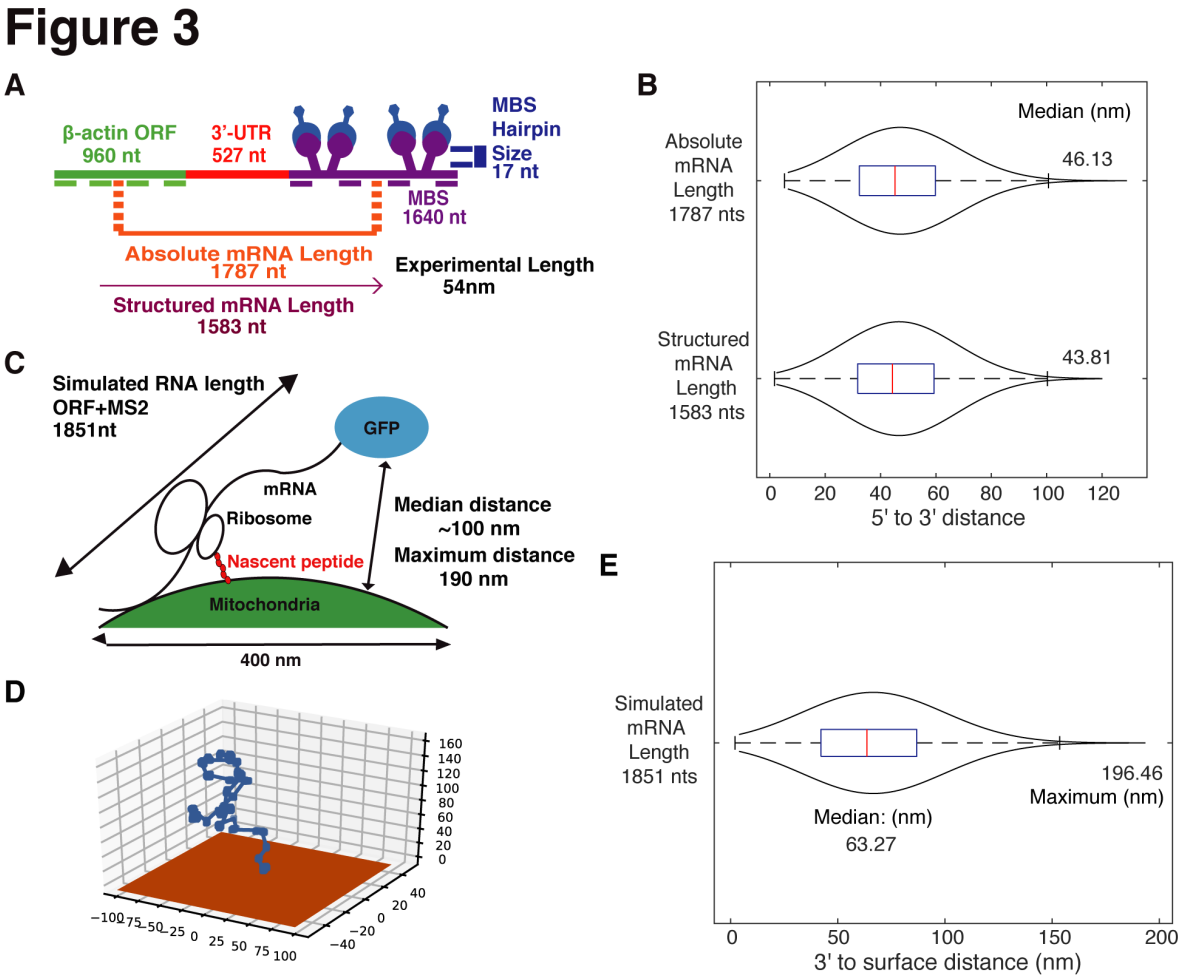


Figure 3 | Determination of the conformation and length of a mitochondria localized mRNA

(A) Flat model of the experimental mRNA structure and length. The green dashed line represents ORF probes. The purple dashed line represents RNA FISH MBS probes. The structured mRNA length is the absolute mRNA length without the MS2 stem-loop sequence. Measurements for mRNA were taken in two-dimensional and obtained from Eliscovich et. al., 2017.

(B) Violin plots showing 5' to 3' end distance distribution of the mRNA displayed in figure (A) with or without MS2 sequence in two-dimensional. The blue box inside the violin plot shows the

first quartile, median(red), and third quartile. The median from the simulation output is displayed on the right.

(C) Illustration of the distance between the mRNA and mitochondria in the previously reported experiments (Tsuboi et al., 2020). The indicated experimental distance was obtained from the association threshold between the mRNA and mitochondria.

(D) Visualization sample of an mRNA located onto mitochondrial surface (orange surface), the beads indicate nucleotides and long edges stand for a ribosome has attached.

(E) Violin plot showing 3' end to the mitochondria surface distance distribution of the experimental mRNA displayed in (C). The blue box inside the violin plot shows the first quartile, median (red), and third quartile. The median and maximum distances from the simulation output are displayed on the graph.

446 **Movie 1 | Simulating RNA 3D structure by RNA single chain models**

447 Visualization sample of 1-100 nucleotides RNA in the absence of ribosomes. The blue line
448 indicates the length in nanometers the mRNA moves in each direction. The angle ranges are plus
449 and minus 10, 30, 90, 150 degrees indicated respectively.

450

451 **Acknowledgments**

452 We thank Dr. T. Wiryaman and members of the Zid laboratory for helpful discussions and
453 feedback on the paper. We thank Dr. Y. Yang and Dr. B. Zid for critically reading the
454 manuscript. This work was supported by startup funds from Tsinghua-SIGS (to T.T.). T.T.
455 acknowledges support from a Japan Society for the Promotion of Science (JSPS) for a research
456 abroad fellowship and postdoctoral fellowship (18J00995), and Uehara Memorial Foundation for
457 research abroad fellowship. H.V.G. acknowledges financial support from the Slovenian Research
458 Agency ARRS (Funding No. P1-0055).

459

460 **Author contributions**

461 T.T. designed the study. T.G, O.M., and J.H. performed experiments T.G, O.M., and J.H.
462 performed analysis. H.V.G. advised on the research. T.G, O.M., J.H., H.V.G., and T.T. wrote the
463 manuscript. All authors discussed the results and commented on the manuscript.

464

465 **Data and materials availability**

466 The code used for simulating the distance between 5' and 3' of mRNA and generating single-
467 chain trajectory is available from https://github.com/paultianzeguo. Further information and

requests for resources, scripts, and reagents should be directed to and will be fulfilled by the lead contact, T.T. (ttsuboi@sz.tsinghua.edu.cn).

Methods

Modeling 3D RNA conformation

In this model, RNA molecules are considered single-chain consisting of multiple nucleotides and the atomistic details of each nucleotide were disregarded. Each nucleotide was considered as a structureless monomer of a given length r = 0.59 nanometers (Adivarahan et al., 2018). This model can be used to study a single-chain of beads with some prescribed restrictions between the monomers such as confinement of angles. We modeled the biopolymer by generating vectors that start from the origin, in three-dimensional Cartesian coordinates and shifting them to connect to beads with each other. The confinement of the angles is applied to the polar angle θ, the angle projection referenced to the y-z plane, and azimuthal angle φ, the angle projection referenced to the x-y plane. The set of (θ, φ) is randomly chosen from a defined bent angle range. The initial vector was randomly chosen as a given length r = 0.59 with a (θ, φ) set. The next angle was chosen based on both the previous (θ, φ) set and bent angle range.

new θ range = 90°-previous $\theta \pm$ angle range,

new φ range = previous $\varphi \pm$ angle range,

We repeated the process until the last nucleotide. In the process (r, θ, φ) was set and converted to (x, y, z), and added up to the previous vector in x, y, z coordinates. The conversion equation is:

$x = r \cos(\varphi)\sin(\theta)$

$y = r \sin(\varphi)\sin(\theta)$

$z = r \cos(\theta)$

After repeating the addition of all nucleotides, the final output gave the vector (x1, y1, z1), which was converted to three-dimensional distance using $d_{3-D} = \sqrt{(x1^2 + y1^2 + z1^2)}$. The experiment was repeated until it ran all the samples from a predefined sample number and graphed on a distribution plot. The sample number was chosen as 1000 for most of the analyses; for analyses on longer strands such as *MDN1*, 200 was used instead. For two-dimensional analyses, simple projection onto planes would have been biased; the spherical coordinates do not create a uniform length among the three possible projections onto the three planes. In the equation that converts the spherical coordinates to Cartesian coordinates, z was defined solely on the length r and the angle θ, however both x and y coordinates were dependent on the projection of r onto the z=0 plane, which is z= $r \sin \theta$, and the azimuth angle φ. Mathematically x and y components need to be multiplied by another trigonometry constant, which is less than one; in most of the cases x and y will be lower than its z counterpart. So, when we obtained the two-dimensional distance, we applied the approximation equation $d_{2-D} = \sqrt{\frac{2}{3}} \times d_{3-D}$, in which we approximate using an ideal condition when all the projections are the same, i.e., x1=y1=z1.

Simulation of mRNA with ribosomes attached

In simulating mRNA with ribosomes attached, we assumed ribosomes had a 30-nucleotides long unfolded area. Based on the method used in simulating mRNA without ribosomes attached, we included a “add ribosomes” function after each strand of mRNA is formed. The resulting “mRNA” was a profile of coordinates that each nucleotide was in. We simulate the addition of ribosomes to let 30 continuous vectors in the same direction to be connected once, which can be simplified as elongating one vector to be 30 times its original length, and the rest adjust to this change in position. Before the simulation starts, the number of nucleotides should be adjusted to

$$\text{number of experimental nucelotide} = \text{number of nucleotides} - 30 * \text{ribosome number}$$

because the simulation of adding ribosomes counts as 30 individual nucleotides. After the resulting mRNA with the desired number of nucleotides is generated, the "add ribosome" function will run the number of times equal to the ribosome number set, in order to insert ribosomes. The function will randomly pick one point in the coordinate array and that point would be multiplied 30 times and the resulting value will be added to the rest of the nucleotides' positions to adjust the following nucleotides' positions. In order to avoid picking the same vector and multiplying it by 30 again, the function would include a checking process that determines if the place already had a ribosome attached to it. After all the ribosomes have been added, the function will read the last point's coordinate (x1, y1, z1), and calculate its three-dimensional distance using $d_{3-D} = \sqrt{(x1^2 + y1^2 + z1^2)}$

or two-dimensional distance as the same process as the mRNA simulation without ribosomes.

mRNA on mitochondria surface

For simulating a mRNA strand on a surface, the x-y plane was chosen as the plane of reference. The simulation would start from the origin. We prevented the growing strand from going through the surface, i.e., $z < 0$. In order to accomplish this, we included an inspection process that checks if their z value is greater than zero after each new set of coordinates is formed.

Visualization of the 3D mRNA model

All three-dimensional mRNA model visualization graphs were generated by the python package mplot3d from matplotlib. In this simulation, we collected the array of x, y, and z values as the

loop function runs. It stores it in an array and uses it as input to the function mplot3d. All the other data-representing graphs and figures are generated in MATLAB from Mathworks.

538 **References**

539 Adivarahan, S., Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O.S., and
540 Zenklusen, D. (2018). Spatial Organization of Single mRNPs at Different Stages of the Gene
541 Expression Pathway. Mol. Cell *72*, 727-738.e5.

542 Bundschuh, R., and Hwa, T. (2002). Statistical mechanics of secondary structures formed by
543 random RNA sequences. Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.
544 *65*.

545 Christensen, A.K., Kahn, L.E., and Bourne, C.M. (1987). Circular polysomes predominate on the
546 rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary.
547 Am. J. Anat.

548 Clote, P., Ponty, Y., and Steyaert, J.-M. (2012). Expected distance between terminal nucleotides
549 of RNA secondary structures. J. Math. Biol. *65*, 581–599.

550 Doi, M. (Masao) (1996). Introduction to polymer physics / M. Doi ; translated by H. See.
551 (Oxford: Clarendon Press).

552 Eliscovich, C., Shenoy, S.M., and Singer, R.H. (2017). Imaging mRNA and protein interactions
553 within neurons. Proc. Natl. Acad. Sci. U. S. A.

554 Fabian, M.R., and Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: A
555 look under the hood of miRISC. Nat. Struct. Mol. Biol.

556 Fang, L.T. (2011). The end-to-end distance of RNA as a randomly self-paired polymer. J. Theor.
557 Biol. *280*, 101–107.

558 Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou,
559 W., Molemans, F., Raeymaekers, A., Van den Berghe, A., et al. (1976). Complete nucleotide
560 sequence of bacteriophage MS2 RNA: Primary and secondary structure of the replicase gene.
561 Nature.

562 Hershkovitz, E., Sapiro, G., Tannenbaum, A., and Williams, L.D. (2006). Statistical analysis of
563 RNA backbone. IEEE/ACM Trans. Comput. Biol. Bioinforma. *3*, 33–46.

564 Hsiao, C., Mohan, S., Hershkovits, E., Tannenbaum, A., and Williams, L.D. (2006). Single
565 nucleotide RNA choreography. Nucleic Acids Res. *34*, 1481–1491.

566 Hurowitz, E.H., and Brown, P.O. (2003). Genome-wide analysis of mRNA lengths in
567 Saccharomyces cerevisiae. Genome Biol. *5*.

568 Imataka, H., Gradi, A., and Sonenberg, N. (1998). A newly identified N-terminal amino acid
569 sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent
570 translation. EMBO J.

571 Jourdren, L., Delaveau, T., Marquenet, E., Jacq, C., and Garcia, M. (2010). CORSEN, a new
572 software dedicated to microscope-based 3D distance measurements: mRNA-mitochondria
573 distance, from single-cell to population analyses. RNA *16*, 1301–1307.

574 Khong, A., and Parker, R. (2018). MRNP architecture in translating and stress conditions reveals

575 an ordered pathway of mRNP compaction. J. Cell Biol.

576 Khong, A., and Parker, R. (2020). The landscape of eukaryotic mRNPs. RNA.

577 Lai, W.J.C., Kayedkhordeh, M., Cornell, E. V., Farah, E., Bellaousov, S., Rietmeijer, R., Salsi,
578 E., Mathews, D.H., and Ermolenko, D.N. (2018). mRNAs and lncRNAs intrinsically form
579 secondary structures with short end-to-end distances. Nat. Commun. *9*.

580 Lareau, L.F., Hite, D.H., Hogan, G.J., and Brown, P.O. (2014). Distinct stages of the translation
581 elongation cycle revealed by sequencing ribosome-protected mRNA fragments. Elife.

582 Levitt, M. (1978). How many base-pairs per turn does DNA have in solution and in chromatin?
583 Some theoretical calculations. Proc. Natl. Acad. Sci. U. S. A. *75*, 640–644.

584 Liphardt, J., Onoa, B., Smith, S.B., Tinoco, I.J., and Bustamante, C. (2001). Reversible unfolding
585 of single RNA molecules by mechanical force. Science *292*, 733–737.

586 Marchat, L.A., Arzola-Rodríguez, S.I., Hernandez-de la Cruz, O., Lopez-Rosas, I., and Lopez-
587 Camarillo, C. (2015). DEAD/DExH-Box RNA helicases in selected human parasites. Korean J.
588 Parasitol.

589 Murray, L.J.W., Arendall, W.B., Richardson, D.C., and Richardson, J.S. (2003). RNA backbone
590 is rotameric. Proc. Natl. Acad. Sci. *100*, 13904–13909.

591 Poblete, S., and Guzman, H. V. (2021). Structural 3d domain reconstruction of the rna genome
592 from viruses with secondary structure models. Viruses *13*.

593 Poblete, S., Božič, A., Kanduč, M., Podgornik, R., and Guzman, H. V. (2021). RNA Secondary
594 Structures Regulate Adsorption of Fragments onto Flat Substrates. ACS Omega *6*.

595 Tarun, S.Z., and Sachs, A.B. (1996). Association of the yeast poly(A) tail binding protein with
596 translation initiation factor eIF-4G. EMBO J.

597 Tsuboi, T., Viana, M.P., Xu, F., Yu, J., Chanchani, R., Arceo, X.G., Tutucci, E., Choi, J., Chen,
598 Y.S., Singer, R.H., et al. (2020). Mitochondrial volume fraction and translation duration impact
599 mitochondrial mRNA localization and protein synthesis. Elife.

600 Tutucci, E., Vera, M., Biswas, J., Garcia, J., Parker, R., and Singer, R.H. (2018). An improved
601 MS2 system for accurate reporting of the mRNA life cycle. Nat. Methods.

602 Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by
603 eukaryotic translation initiation factors. Mol. Cell.

604 Wen, Q., and Janmey, P.A. (2011). Polymer physics of the cytoskeleton. Curr. Opin. Solid State
605 Mater. Sci. *15*, 177–182.

606 Wu, B., Eliscovich, C., Yoon, Y.J., and Singer, R.H. (2016). Translation dynamics of single
607 mRNAs in live cells and neurons. Science (80-.). *352*, 1430 LP – 1435.

608 Yan, X., Hoek, T.A., Vale, R.D., and Tanenbaum, M.E. (2016). Dynamics of Translation of
609 Single mRNA Molecules In Vivo. Cell *165*, 976–989.

610 Yoffe, A.M., Prinsen, P., Gopal, A., Knobler, C.M., Gelbart, W.M., and Ben-Shaul, A. (2008).
611 Predicting the sizes of large RNA molecules. Proc. Natl. Acad. Sci. U. S. A. *105*, 16153–16158.

612 Yoffe, A.M., Prinsen, P., Gelbart, W.M., and Ben-Shaul, A. (2011). The ends of a large RNA
613 molecule are necessarily close. Nucleic Acids Res. *39*, 292–299.

614 Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative
615 modes of gene expression in yeast. Nat. Struct. Mol. Biol. *15*, 1263–1271.

616

617