# Single-chain models illustrate the 3D RNA folding shape during translation

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

[1]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California; [2]Department of Theoretical Physics, Jozef Stefan Institute, Ljubljana, Slovenia; and [3]Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Shenzhen, Guangdong, China

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. Here, we introduce the Three-dimensional RNA Illustration Program (TRIP) to model the three-dimensional RNA folding shape based on single-chain models and angle restriction of each bead component from previously reported single-molecule fluorescence in situ hybridization (smFISH) experimental data. This simulation method was able to recapitulate the mRNA conformation change of the translation activity and three-dimensional positional interaction between an organelle and its localized mRNAs as end-to-end distances. 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 top-down method will be used to interpret the aggregation mechanism of mRNA under different cellular conditions such as nucleolus and phase-separated granules.

SIGNIFICANCE  We presented that mRNA three-dimensional conformation can be approached as a simple single-chain bead model. Our model is able to characterize the generalized behavior and categorize the type of regime the single chain falls into, which in turn allows an improved understanding of mRNA's three-dimensional organization in the cell. We tested the translational state by modeling the mRNA conformation to be unfolded when the ribosome is associated. Furthermore, we applied our model to interpret mRNA conformation on the mitochondrial surface, which is experimentally difficult to address. Our top-down model introduces the first approximation on how polymer-like models can rapidly estimate end-to-end distances of mRNAs in the cellular condition and is useful for interpreting mRNA experiments devising end-to-end distances.

## 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 translational regulation and microRNA (miRNA)-mediated regulation ([1–3]). However, despite the importance of mRNA organization, there is little known about how ribonucleoprotein complexes (mRNPs) are organized as three-dimensional (3D) 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 eIF4F complex and the poly(A) binding protein PABPC1 ([2–5]). 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 (6).

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 (7). Each element of the RNA strand can be categorized by the different links of a polymer-like arrangement of beads, and it simplifies the analysis of the whole function of the RNA. In addition, it draws a relationship between the effect of confinement on the end-to-end distances of single-chain polymers and the stiffness of each fragment (8). Even though this relationship is based on arbitrary numbers, it is able to characterize the generalized behavior and categorize the type of regime the single chain falls into, which in turn allows an improved understanding of RNA conformation for the tackled cases and similar ones.

In this work, we applied the single-chain models for mRNA folding and developed the Three-Dimensional RNA Illustration Program (TRIP), a simplified model designed to rapidly interpret the end-to-end distances during translation. The 5′ to 3′ distance of mRNA when treated with puromycin, a known antibiotic that prematurely ends translation and disassembles polysomes, follows a Gaussian distribution with a skewness to the right (6). Furthermore, ribosome occupancy determines the compaction of mRNAs, and the number of ribosomes positively correlated with the length of the mRNA (6,9). When comparing our initial calculations with this single-molecule fluorescent in situ hybridization (smFISH) study, our findings gave an in silico demonstration of this conclusion by simulating a chain of mRNA with and without ribosomes attached. The 3D conformation of the mRNA is assembled by using a bent angle range of $\pm 140°$. Our single-chain models also recapitulated the distance between cytoplasmic mRNA and the mitochondria, which were previously analyzed in fixed cells by smFISH (10), and in live cells through an MS2-MCP method (11). We also found that integration of the information of the local secondary structure, as is the case of the tackled MS2-sequence, does not have much influence in our single-chain model. While there is little known about mRNA's 3D organization in the cell, our simulation introduces a first approximation on how polymer-like models can rapidly estimate end-to-end distances of mRNAs during translation in the cellular condition. In a nutshell, our very coarse model is useful for interpreting mRNA experiments devising end-to-end distances of the fragment.

## METHODS

### Modeling 3D RNA conformation

In this model, RNA molecules are considered single chains 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 nm (6). 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 3D Cartesian coordinates, and shifting them to connect the beads with each other. The confinement of the angles is applied to the polar angle $\theta$, which projection referenced to the y-z plane; and azimuthal angle $\varphi$, which projection referenced to the x-y plane. The set of $(\theta, \varphi)$ is randomly chosen from a defined bent angle range. The initial vector was randomly chosen as a given length r = 0.59 with a $(\theta, \varphi)$ set. The next angle was chosen based on both the previous $(\theta, \varphi)$ set and bent angle range.

$$\text{New } \theta \text{ range } = 90° - \text{previous } \theta \pm \text{angle range}$$

$$\text{New } \varphi \text{ range } = \text{previous } \varphi \pm \text{angle range}$$

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

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

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

$$z = r\cos(\theta)$$

After repeating the addition of all nucleotides, the final output gave the vector (x1, y1, z1), which was converted to 3D 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 2D 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 $\theta$; 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 $\varphi$. 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 2D 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-nt-long unfolded area. Based on the method used in simulating mRNA without ribosomes attached, we included an "add-ribosomes" function after each strand of mRNA is formed. Ribosomes were assumed in random positions with the RNA simulated between them using the simple flexible chain model. The resulting mRNA was a profile of coordinates that each nucleotide was in. We simulated the addition of ribosomes to let 30 continuous vectors in

the same direction 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

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

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, 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 3D distance using $d_{3-D} = \sqrt{(x1^2 + y1^2 + z1^2)}$ or 2D distance as the same process as the mRNA simulation without ribosomes.

## mRNA on mitochondria surface

For simulating an 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 3D 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 were generated in MATLAB from Mathworks.

## RESULTS

### Modeling 3D RNA folding by TRIP

mRNAs are biopolymers consisting of thousands of nucleotides connected to each other from the 5′ to 3′ ends. The RNA backbone is rotameric (12). For each residue along the RNA backbone, there are six angles ($\alpha$, $\beta$, $\gamma$, $\delta$, $\varepsilon$, $\zeta$) 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 3D RNA folding using TRIP, 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 (Fig. 1 A). The monomer length was set to be 0.59 nm based on an average distance between two nucleotides (13). The confinement of the angles is applied to the polar angle $\theta$, the angle projection referenced to the y-z plane, and azimuthal angle $\varphi$, the angle projection referenced to the x-y plane. The set of ($\theta$, $\varphi$) is randomly chosen from a defined angle range. We placed the first bead, the head of a single chain, at the origin of a spherical coordinate and let it be the 5′ end, and the next bead was randomly picked from the outer surface of the sphere with a radius of 0.59 so that the first monomer, as a vector in the coordinate system, is generated with a length and direction. Based on the first vector's direction, the next vector was created by setting the first bead as the origin and the second bead's direction depending on the first vector's direction within a confinement degree range (Fig. 1 B). This procedure was reiterated to create a strand of RNA.

While the monomer length was set, the angle range significantly affected the end-to-end distances. We used TRIP to visualize samples that were 100-nt long and applied different angle ranges. It was observed, especially at low angle ranges, such as 10°, that the RNA strand will extend far; when angle ranges increase, the RNA strand tends to compact (Fig. 1 C; Video S1). To identify the best angle range for TRIP, we compared the modified two-dimensional (2D) TRIP results (STAR Methods) with reported 2D 5′ to 3′ length for three different mRNAs: *MDN1*, *POLA1*, and *PRPF8* in non-translating conditions. Previous smFISH experiments tested the 5′ to 3′ distances in 3D and reported their result in 2D using image-flattening techniques. The reported data were 35.95 nm(*MDN1*), 33.15 nm (*POLA1*), and 34.19 nm (*PRPF8*) respectively (6). Because the measurement of length in smFISH is conducted by the distance between the centers of fluorescent signals from multiple probes, the nucleotides on the edge of the fluorescent signals may not be taken into account. To accurately input the experimental length for the simulation, we defined the net nucleotide number as the length between the centers of the FISH probes on the 5′ and 3′ ends. Different angle ranges were applied to the simulation for three mRNA references, with net nucleotide numbers 16,350 (*MDN1*), 4060 (*POLA1*), and 4969 (*PRPF8*). As shown in the graph, there is a rapid decrease in the TRIP length first seen around $\pm 10°$ to $\pm 30°$, then, at around $\pm 100°$, the
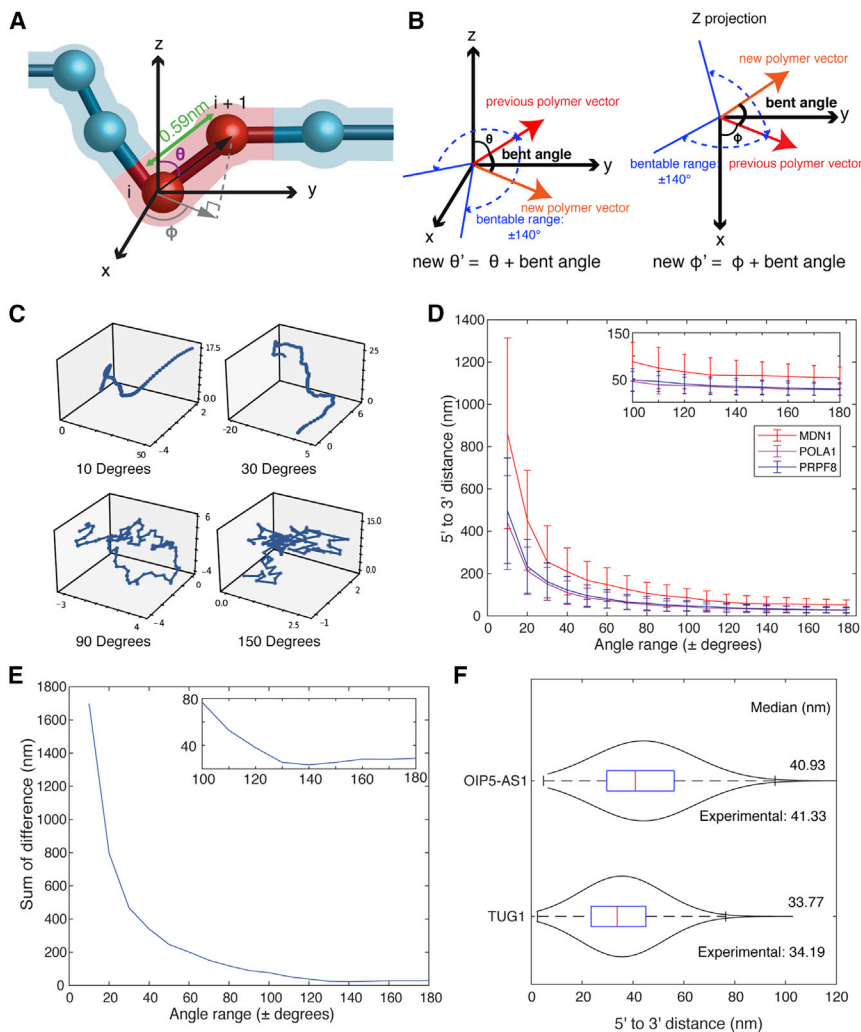
FIGURE 1 Simulating RNA 3D structure by RNA single-chain models. (*A*) Diagram of TRIP, where the bead-chain in red is the area of focus. φ is the azimuthal angle, which is the horizontal angle between the single chain and origin and measures the projection of the vector (indicated by black arrow) with reference to the +x axis. θ is the polar angle between the +z axis and the vector. The length between beads is 0.59 nm (6). (*B*) The TRIP simulation is shown in two different projections, with 3D Cartesian coordinates on the left, and Z projection on the right. The original vector is projected θ from the z axis and φ from the x axis, and the angle of the new single-chain vector is restricted by the 140° bendable range in both axes. (*C*) Visualization sample of 100-nt RNA in the absence of ribosomes. The blue line indicates the length in nanometers the mRNA moves in each direction. The angle ranges are plus and minus 10°, 30°, 90°, and 150° indicated respectively. (*D*) The relationship of 5′ to 3′ end-to-end distance simulated in 2D with respect to the chosen angle range for *MDN1* (red), *POLA1* (magenta), and *PRPF8* (blue) mRNAs. Each dot indicates the median 5′ to 3′ distance on the given angle range, with error bars indicating one standard deviation. The frame on the upper right shows a zoom-in figure at angle range ±100 to ±180. (*E*) Plot of the sum of error, which is calculated as the sum of the distances of the TRIP outputs of three mRNA in (*C*) at a certain angle range to the previously reported smFISH data (6). The frame on the upper right corner shows a zoom-in figure at angle range ±100 to ±180. (*F*) 5′ to 3′ end distance distribution of the non-translating RNAs *TUG1* and *OIP5-AS1* in 2D. 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 (6) are referenced below each plot.
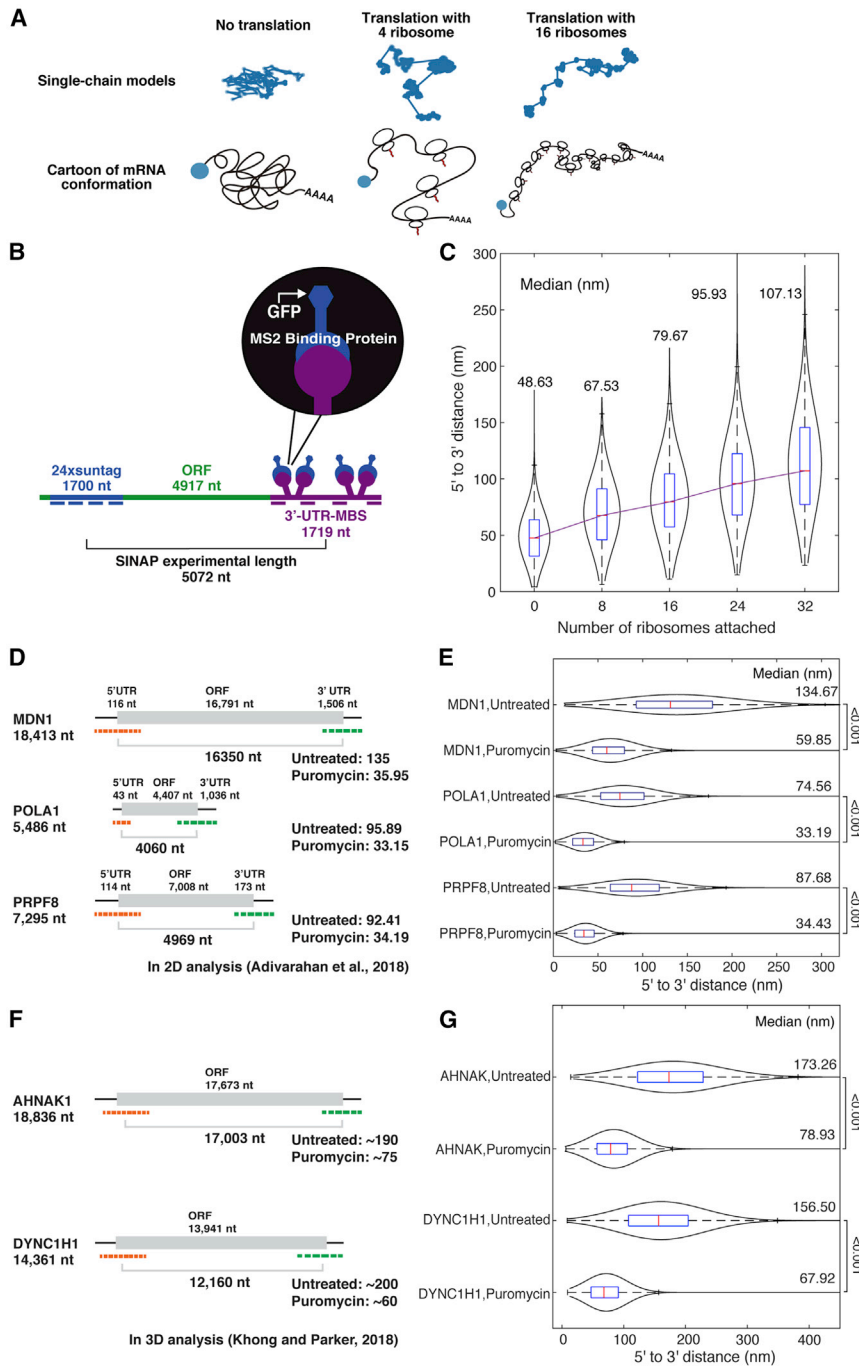
distances start to level off (Fig. 1 *D*). The results refer back to the observation that an increase in angle range leads to compaction. At around ±140°, the TRIP results showed high correspondence with the referenced values. As the angle further increases, the distances continue to level off but in a smaller magnitude (less fluctuation), indicating that the molecule becomes stable. To identify the best angle range, we analyzed the sum of the difference between the TRIP results and the previously reported smFISH data (6) in each angle range (Fig. 1 *E*). By magnifying the larger angle range region, we confirmed that, at ±140° angle ranges, the sum of the difference is the least among all the angle ranges we tested (Fig. 1 *E*). We further confirmed whether the ±140° angle range could explain long non-coding RNAs, which do not associate with ribosomes. Previous smFISH study showed that the long non-coding RNAs' end-to-end proximity was 41.33 nm

for *OIP5-AS1* and 34.49 nm for *TUG1* (6). The TRIP successfully recaptured these smFISH data as 40.93 nm for O*IP5-AS1* and 33.77 nm for *TUG1* respectively (Fig. 1 *F*).

## mRNA end-to-end distance positively correlates with numbers of ribosome attached

Ribosomes cover approximately 30 nt of mRNA and unfolded and flattened mRNA during translation (14). We tested whether translation influences the mRNA conformation using TRIP (Fig. 2 *A*). To analyze the relationship of translation activity and mRNA conformation, we used the 6636-nt-long SINAP mRNA, in which end-to-end proximity distance has been reported with a different number of ribosomal activities by smFISH (Fig. 2 *B*) (6). The 5072 net nucleotide numbers were analyzed using fluorescent FISH probes at its 5′ and 3′

FIGURE 2   For a Figure360 author presentation of this figure, see https://doi.org/10.1016/j.bpr.2022.100065.

The ribosome occupation explains the separation of 5′ end and 3′ end of mRNA in different mRNAs. (*A*) TRIP models the 3D RNA folding shape during no translation, mild translation, and intense translation (four ribosomes and 16 ribosomes in the case of cartoons). (*B*) Flat model of the SINAPs mRNA structure and length. The blue dashed line represents 24xsuntag FISH probes. The purple dashed line represents 3′-UTR-MBS FISH probes. The experimental length is the absolute mRNA length between the centers of the above two probes (6). (*C*) mRNA end-to-end distance positively correlates with numbers of ribosomes attached. Violin plots showing 5′ to 3′ distance distribution of the SINAP by TRIP simulation in the order of increasing translation activity in 3D. 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 top. The purple line connects the data by their medians linearly. (*D*) Diagram of the mRNAs used in the previously reported experiments (6). The bracket indicates the experimental length. The numbers indicated on the right exhibit the length in nanometers when cells are treated and untreated with puromycin in 2D. (*E*) Violin plots showing 5′ to 3′ distance distribution of *MDN1*, *POLA1*, and *PRPF8* mRNAs from the TRIP simulation for the cells treated and untreated with puromycin. 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 5′ to 3′ distance is simulated in 2D. The number indicated by brackets outside the graph is the KS test p value between the two sets of data. (*F*) Diagram of the mRNAs used in the previously reported experiments (9). The bracket indicates the experimental length. The numbers indicated on the right exhibit the length in nanometers when cells are treated and untreated with puromycin in 3D. (*G*) Violin plots showing 5′ to 3′ distance distribution of *AHNAK* and *DYNC1H1* mRNAs from the TRIP simulation for the cells treated and untreated with puromycin. 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 5′ to 3′ distance is simulated in 3D. The number indicated by brackets outside the graph is the KS test p value between the two sets of data.

ends for this mRNA (Fig. 2 *B*). The open reading frame length of this mRNA was 4917 nt long (Fig. 2 *B*). It has previously been shown that the maximum ribosome-bound number was 20 for 3129-nt-long open reading frame in SINAP mRNA (15). We estimated the maximum number of ribosomes for 4917-nt open reading frame as 32 ribosomes by calculating the ratio. We conducted the in silico experiment for five groups, each with 0, 8, 16, 24, and 32 ribosomes attached, similar to what was shown in the FISH experiment. We observed a stepwise increase in length with evenly distributed intervals according to increasing numbers

of ribosomes (Fig. 2 *C*). We also observed that the mRNA length distribution was a Gaussian distribution with a skewness toward the longer distance (Fig. 2 *C*). The relationship between distance and ribosome number suggests that more intense translation activity can increase the end-to-end distance, thus the molecule would be less compact since unfolded mRNA areas are greater.

## The ribosome occupation explains the separation of 5′ end and 3′ end of mRNA in different mRNAs

We further tested whether TRIP could recapitulate the impact of translation to end-to-end proximity distance using the three different mRNAs *MDN1*, *POLA1*, and *PRPF8* referring to the previously reported smFISH data (6). A significant difference of 5′ to 3′ proximity distance has been observed between the conditions of puromycin-treated and -untreated HEK293 cells (Fig. 2 *D*). These differences indicate that the translation contributes to the unfolding of mRNA conformations. We estimated the ribosome numbers as the average translational activity, which is the length of the open reading frame divided by 200 nt, where 200 is the average inter-ribosome distance on mRNA reported in human HEK293 and U2OS cells (16). Using this average number of ribosomes in all three mRNAs, we simulated the end-to-end proximity distance to compare with the previously reported smFISH results (Fig. 2 *D*) (6). In all three cases, the TRIP predicted a significant difference between non-translating and translating conditions (Fig. 2 *E*). This result matched what was expected from in vivo environments, in which the translation of these mRNAs resulted in the separation of the 5′ and 3′ ends (6). The distribution of TRIP results follows a Gaussian distribution with a skewness to the larger distance, meaning a larger density was found at lower values. The compaction effect is found often in non-translating conditions, where the 5′ to 3′ end-to-end distance is significantly reduced from translating conditions. Interestingly, the end-to-end proximity for *POLA1* mRNA at translation condition (95.89 nm) was reported longer than *PRPF8* (92.41 nm) in the referenced data, despite *PRPF8* having a longer open reading frame (6). However, the TRIP output predicted that *POLA1* untreated is lower than *PRPF8* untreated. This discrepancy might be due to *POLA1* possibly having higher translatability. We estimated that 40 ribosomes were required to reach the *POLA1* end-to-end proximity distance to the reported distance 95.13 nm by smFISH, which is 1.8 times greater translatability. In addition to comparing results shown in 2D space, we investigated the ribosomal occupation on 3D datasets. Similar experiments using smFISH probes to detect end-to-end distance were performed to demon-strate the end-to-end distances of two sample mRNAs with long open reading frames under translating conditions and puromycin conditions: AHNAK (17,673-nt open reading frame) and DYNC1N1 (13,941-nt open reading frame) (9). The median is shown from their cumulative distribution plots as AHNAK, translating conditions ~190 nm, and puromycin conditions ~75 nm; DYNC1H1, translating conditions ~200 nm, and puromycin conditions ~60 nm. AHNAK and DYNC1H1 were simulated with net nucleotides numbers 17,003 nt and 12,160 nt according to the position of smFISH probes used in previous research (9). The TRIP output showed a similar trend when it was in 2D; it predicted a significant difference between non-translating and translating conditions (Fig. 2 *F* and *G*). AHNAK1 in translating conditions results in a median that is similar (173.26 nm) to the experimental data. Nevertheless, for DYNC1H1 in translating conditions, the TRIP output shows a slight underestimation; the median (156.50 nm) is lower than the experimental data. This might be caused by the lower translatability of DYNC1H1. For both mRNAs in non-translating conditions, the TRIP outputs recapitulate the in vitro data: for AHNAK, the median gave 78.93 nm; for DYNC1H1, the median gave 67.92 nm (Fig. 2 *G*). In summary, the in silico TRIP experiment was able to recapitulate that mRNA in a non-translation state follows a geometrically compact shape, whereas translation significantly separates the 5′ and 3′ ends.

We compared the TRIP with scaling models in polymer theory; i.e., freely jointed chain (FJC) and excluded volume (EV) models in Fig. S1 (8). This shows that the TRIP simulations scale with the same trend of the theoretical models, as a function of the mRNA length, in particular for the mRNAs without ribosomes. However, for mRNAs with ribosomes (periodicity 200 nt), we observe that the slopes slightly differ (Fig. S1). These suggest that our results with ribosomes do not meet the minimal requirement necessary to apply the scaling laws of the FJC and EV models. In other words, for the six different mRNAs we tackle, every polymer blob will have roughly 200 nt, which results in small values of blobs (between 20 and 90). Hence, for such a small total number of blobs, the TRIP has a high capacity in interpreting the experimental dataset. We further tested how the different numbers of ribosomes can contribute to the end-to-end distance of different mRNAs (Fig. S2). This showed that, in the case of short mRNAs (*MS2*, *POLA1*, *PRPF8*), there are several scaling laws with slopes greater than 1 between the number of ribosomes and the end-to-end distance, unlike in longer mRNAs (*DYNC1H1*, *MDN1*, *AHNAK1*), which have slopes between 0.92 and 0.71. The latter shows a tendency toward the EV model. Even though we still need to consider specific cases, such as the formation of a

secondary structure, TRIP allows us to estimate the ribosome numbers quantitatively.

## Local secondary structure has minor effects in the studied cases

We tested how local secondary structures affect long mRNAs based on the lack of formation of circular structured mRNA. MS2 was used as a secondary structure in an effort to observe the way the mRNA folds. MS2 is a bacteriophage that consists of 3569 nt of single-stranded RNA. It is encoded with four proteins: the maturation protein, the lysis protein, the replicase protein, and the coat protein (17). The coat protein (MCP) is useful for the detection of RNA within living cells; MS2 binding sites (MBS) are inserted in the 3′ UTR of an mRNA of interest, and the co-expression of MCP fused with fluorescent proteins renders single mRNAs visible using fluorescence microscopy (18). Between the mRNA and MCP, there should be a secondary structure created by MS2. We predicted that, by incorporating the information of the secondary structure, we would be able to refine the predicted RNA length. Each MS2 stem-loop structure was composed of 17 nt. This implies that the mRNA had an absolute experimental length of 1787 nt and the structured mRNA length, which removed the 12 stem-loop sequences, was 1583 nt (Fig. 3 A). It has been observed that there was a mean observed distance of 48 nm between the MCP-open reading frame in 2D (19). While the addition of this secondary structure could provide a refined predicted mRNA length, our analysis reveals that base-pairing interactions only have minor effects (2.3 nm for 204 nt; i.e., 12 × 17 nt) on the end-to-end distance (Fig. 3 B), meaning that the effects of the MS2 secondary structure are less. This suggests that local secondary structure has minor effects on the tackled mRNA conformation.
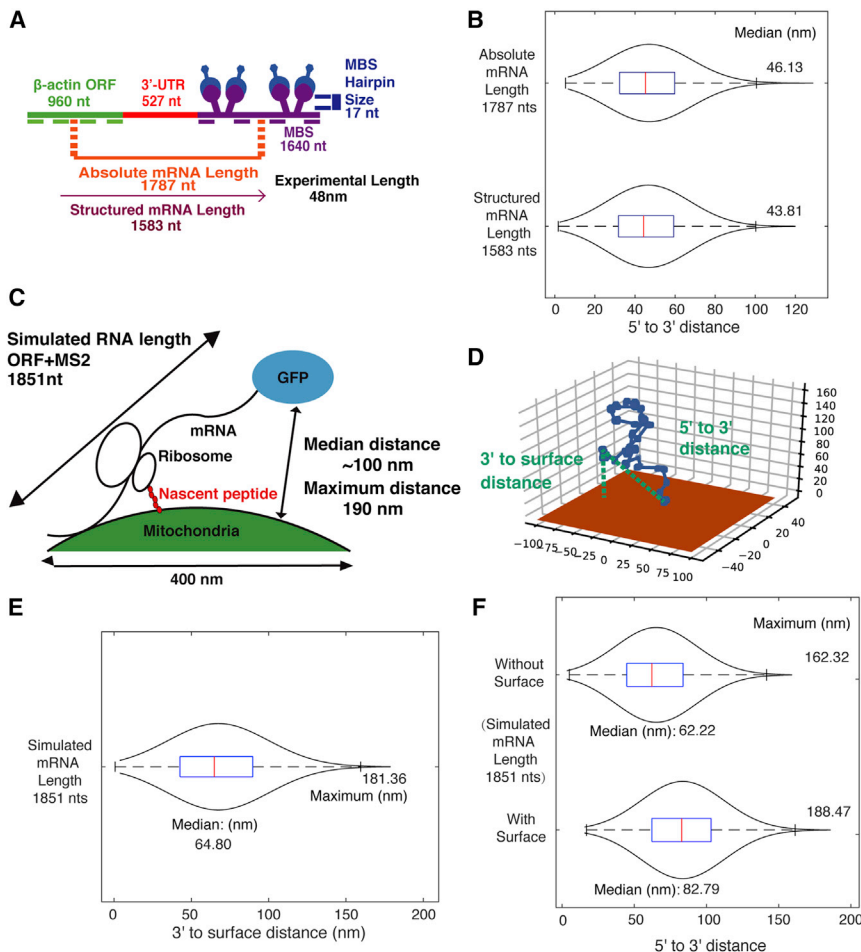
## Determination of the shape and length of a mitochondria-localized mRNA

Analyzing the proximity of nuclear-encoded mRNA to the mitochondrial outer membrane is challenged with long-term cycloheximide treatment by electron cryotomography (20). The distance between cytoplasmic mRNA and the mitochondria was also analyzed by visualizing single-molecule mRNA in fixed cells through smFISH (10), and in live cells using the MS2-MCP method (11). We tested whether the TRIP, which is a non-closed-loop structure model, could recapture and explain the previous observations. To model such mRNA conformation, we defined the beads (i.e., nucleotides) are not formed under the mitochondrial surface plane. To conduct the simulation, the x-y plane was set up as the surface, so that the result's z coordinate will never be less than

zero. This represents a strand of mRNA growing on the x-y surface. A previous study with *TIM50* mRNA, a mitochondrially localized mRNA (Fig. 3 C), observed that the distance to the mitochondria follows a Gaussian distribution skewed to the larger distance, and the median distance is ∼100 nm and maximum distance is defined at 190 nm (11). We compared these numbers with the in-silico modeling by TRIP. We found that 154 nt was the average distance between ribosomes in budding yeast, *Saccharomyces cerevisiae* (21,22), and that *TIM50* mRNA's open reading frame length is 1851 nt. With this information, we estimated that 12 (i.e., 1851/154) ribosomes exist per mRNA. The visualization results of TRIP showed that a strand of mRNA tail attached to the surface in orange color in the cartoon (z = 0) (Fig. 3 D). The TRIP results indicated the proximity distance between the mRNA 3′ and mitochondrial surface as a Gaussian distribution, with a skewness to the larger distance (Fig. 3 E). The median was at 64.80 nm and the maximum distance was 181.36 nm. The median value was far from the experimental data (11). It is possibly due to the inclusion of the non-mitochondria-associated mRNAs, which have the distances of more than 190 nm, in the experimental data. The maximum distance was close to the previously reported experimental threshold value of 190 nm. This was also consistent with the previous FISH experiments showing the distance between mRNA localized to mitochondria and ribosomal RNA of mitochondrial matrix measured as a range of 150–200 nm (10). These suggest that the mitochondrial-localized mRNA is also not forming a closed-loop structure. We further tested whether the addition of the surface by itself can directly contribute to the end-to-end distance of mRNA in TRIP, and found that this addition increases it (Fig. 3 F). This implies that the TRIP can be used to provide a primary assessment of the end-to-end distance of mRNAs in a variety of cellular constraints, which we cannot experimentally test.

## DISCUSSION

In this work, we presented that mRNA 3D conformation during translation can be approached as a simple single-chain bead model. We developed TRIP and identified that 140° is the restricted angle of each component of the single chain to reproduce the previously reported experimental dataset (Fig. 1). We further tested the translational state by modeling the mRNA conformation to be unfolded and flat when the ribosome is associated (Fig. 2). By analyzing these sets of MS2 structure, we showed that the secondary structure, especially in the case of local and minor percentage of base pairing in the mRNA fragments, would have a minimal effect on the mRNA end-to-end distance (Fig. 3 A

**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 open reading frame 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 2D and obtained from (19). (*B*) Violin plots showing 5′ to 3′ end distance distribution of the mRNA displayed in figure (*A*) with or without MS2 sequence in 2D. 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 (11). 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. Illustration of 3′ to surface and the 5′ to 3′ distance is indicated by green dotted line. (*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. (*F*) Violin plot showing 5′ to 3′ distance distribution of the experimental mRNA displayed in (*C*) with and without localization onto mitochondria surface. 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.

and *B*). Last, we applied TRIP to predict mRNA conformation on the mitochondrial surface, where it is experimentally difficult to address (Fig. 3 *C*–*F*). This methodology will be further applied to predict the aggregation mechanism for mRNA and long non-coding RNA in specific cell organizations, such as the nucleolus and phase-separated granules.

The six rotatable torsion angles of the RNA backbone were analyzed through vector quantization (23), multiresolution approach (24), and quality-filtering techniques (12) and have met in consensus. The analysis of the six torsion angles together provides quantifications of the nucleic acids' helical shape on an atomic level, and it also suggests that RNA can form various numbers of structures depending on the angles. However, at a coarser level, this has to provide the torsion angle between the whole nucleotides. In our model we assumed each nucleotide position as

beads, and, instead of accounting for microscopic torsion angles between each atomic bond, we defined our set of bent angles ($\theta$, $\varphi$) in a macroscopic view in which we assume and model each nucleotide broadly as observable clusters. The set of bent angles ($\theta$, $\varphi$) explained that the relative position of each neighboring nucleotide is bent within 140° positive or negative referring to the defined plane. We showed that an angle range of 140° still gives a variety of RNA lengths based on our model. Compared with double-helix structures like DNA, where each nucleotide is paired and forms a stable structure, RNA is relatively unstable and tends to compact or form secondary structures; thus, it may have smaller angles between each nucleotide. DNA, on average, has 10.5 bp/turn (25); geometrical analysis suggests that each angle between nucleotides on average is approximately 145.71° (i.e., $\frac{180° \times (10.5 - 2)}{10.5}$). Our bent angle analysis, which was converted to an

angle between nucleotides, ranges from 40° to 180°. Since our distribution was uniform in angle choosing, the statistical average angle between RNA was calculated to be 110°, which is less than that of double-helix structures.

Vanzi et al. (26) developed a method that used optical microscopy and infrared laser tweezers in order to measure the strength of the ribosome-polyuridylic acid (poly(U)) complex as well as the elastic properties of the poly(U). They found that, when modeled as a worm-like chain, the persistence length of poly(U) RNA when devoid of any secondary structures was $0.79 \pm 0.05$ nm. However, they also concluded from their results that poly(U) molecules can be bound at internal locations where natural mRNA cannot, so, while their results are valid for shorter genetic fragments, our work addresses the longer-chain behavior of mRNA. Another study used single-molecule stretching experiments in order to demonstrate that worm-like chain models better describe polymer elasticity compared with an FJC model for single-stranded RNAs that lack base-pairing and stacking interactions and conform to a random-coil structure (27). These results, however, only applied to single-stranded RNA, which has different properties and characteristics than mRNA. On average, the results collected from these two studies are consistent with those provided by TRIP. However, results produced by TRIP are less constrained due to the uniform distribution of allowed angles in successive nucleotides.

Intramolecular RNA-RNA base-pairing interactions are important for small 5′ to 3′ distances. Previous research argued that the distance between the two ends of single-stranded RNA molecules is small, under the assumption that there are approximately equal proportions of A, C, G, and U (28). A large number of circle diagrams associated with the secondary structures of many RNA molecules was also examined for different lengths and sequences (29). It was revealed that the first and last monomers were always within a few monomers of each other, suggesting that it is essentially impossible to not have a minimum of one set of base pairs that will bring the ends of an RNA together, Furthermore, it was observed that certain RNA molecules, such as viral genomes (30,31) and certain messenger RNAs, have been under selective pressure to maintain a small distance between the 5′ and 3′ ends of the molecule (28). However, this does not prove that biologically functional RNA molecules, such as viral genomes and certain messenger RNAs, have small 5′ to 3′ distance independent of sequence length (32). These findings were then experimentally confirmed by means of Förster resonance energy transfer (FRET) (33). Theoretical analyses of a number of randomized and natural RNA sequences suggest that the 5′ and 3′ ends of long RNAs (1000–10,000 nt) are always brought in the proximity of a few nanometers of each other regardless of RNA length and sequence because of the intrinsic propensity of RNA to form widespread intramolecular base-pairing interactions (33).

The evidence argues that ribosomes decompact mRNAs. An analysis of mammalian mRNP compaction by smFISH revealed that translating mRNPs are more extended compared with non-translating mRNAs (6,9). The fact that mRNAs under stressed conditions show similar compaction compared with non-translating conditions may symbolize the removal of ribosomes (9). Translating mRNPs are compacted relative to their contour length (6,9). This compaction is modeled by the formation of secondary structures of mRNA sequences between elongating ribosomes, which is supported by the spacing of ribosomes on mRNAs (34). Since the compaction of translating mRNPs is not affected by the inhibition of translation elongation, it is possible that stalled ribosomes on the open reading frame are sufficient to decompact mRNA (9,34). Due to spontaneous intramolecular RNA folding taking place rapidly, it can be assumed that mRNA sequences will collapse into RNA secondary structures, which are then unwound by each elongating ribosome (34). mRNA folding is also influenced by the binding of RNP-BP, including the ATP-dependent binding of DEAD/Delhi box proteins (34). These box proteins, specifically DHH1, catalyze the ATP-dependent folding and remodeling of RNA duplexes. Immunoprecipitation assays reveal that DHH1 interacts with a variety of proteins, such as heat shock proteins, mRNA binding proteins, initiations and elongation translation factors, ribosomal proteins, and metabolic proteins (35).

RNA's secondary structure is described by the configuration of the base pairings that are formed by the biopolymer. These secondary structures can be divided into stems, which, in naturally occurring RNA molecules, are made up of five consecutive base pairs, and loops, which connect or terminate the stems (36). While they locally form the same double-helical structure as DNA molecules, RNA molecules are single stranded, and therefore must fold back onto themselves to gain base pairings (36). The complexity of these secondary structures increases with length (37). We used the secondary structure formed by the MS2 bacteriophage, which is one of the simplest examples of a secondary structure in mRNA. Our analysis showed that, when the RNA secondary structure is formed in part of mRNA, which was MS2 structure in our experiment, it only has minor effects on the end-to-end distance of the whole mRNA (Fig. 3 B). While secondary structures are thought to have a huge impact on mRNA 3D conformation, this depends on the base-pairing percentage of

the total mRNA fragment. In particular for the tackled biopolymers, our analysis would point to base-pairing interactions only having minor effects on mRNA end-to-end distance.

## DATA AND MATERIALS AVAILABILITY

The code used for simulating the distance between 5′ and 3′ of mRNA and generating single-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).

## SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpr.2022.100065.

## AUTHOR CONTRIBUTIONS

## ACKNOWLEDGMENTS

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

1. Fabian, M. R., and N. Sonenberg. 2012. The mechanics of miRNA-mediated gene silencing: A look under the hood of miRISC. *Nat. Struct. Mol. Biol.* 19:586–593. https://doi.org/10.1038/nsmb.2296.

2. Imataka, H., A. Gradi, and N. Sonenberg. 1998. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J.* 17:7480–7489. https://doi.org/10.1093/emboj/17.24.7480.

3. Tarun, S. Z., and A. B. Sachs. 1996. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J.* 15:7168–7177. https://doi.org/10.1002/j.1460-2075.1996.tb01108.x.

4. Christensen, A. K., L. E. Kahn, and C. M. Bourne. 1987. Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. *Am. J. Anat.* 178:1–10. https://doi.org/10.1002/aja.1001780102.

5. Wells, S. E., P. E. Hillner, …, A. B. Sachs. 1998. Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell.* 2:135–140. https://doi.org/10.1016/s1097-2765(00)80122-7.

6. Adivarahan, S., N. Livingston, …, D. Zenklusen. 2018. Spatial organization of single mRNPs at different stages of the gene expression pathway. *Mol. Cell.* 72:727–738.e5. https://doi.org/10.1016/j.molcel.2018.10.010.

7. Wen, Q., and P. A. Janmey. 2011. Polymer physics of the cytoskeleton. *Curr. Opin. Solid State Mater. Sci.* 15:177–182. https://doi.org/10.1016/j.cossms.2011.05.002.

8. Doi, M. 1996. Introduction to Polymer Physics. Oxford: Clarendon Press.. translated by H. See.

9. Khong, A., and R. Parker. 2018. MRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. *J. Cell Biol.* 217:4124–4140. https://doi.org/10.1083/jcb.201806183.

10. Jourdren, L., T. Delaveau, …, M. Garcia. 2010. CORSEN, a new software dedicated to microscope-based 3D distance measurements: mRNA-mitochondria distance, from single-cell to population analyses. *RNA.* 16:1301–1307. https://doi.org/10.1261/rna.1996810.

11. Tsuboi, T., M. P. Viana, …, B. M. Zid. 2020. Mitochondrial volume fraction and translation duration impact mitochondrial mRNA localization and protein synthesis. *Elife.* https://doi.org/10.7554/eLife.57814.

12. Murray, L. J. W., W. B. Arendall, …, J. S. Richardson. 2003. RNA backbone is rotameric. *Proc. Natl. Acad. Sci. USA.* 100:13904–13909. https://doi.org/10.1073/pnas.1835769100.

13. Liphardt, J., B. Onoa, …, C. Bustamante. 2001. Reversible unfolding of single RNA molecules by mechanical force. *Science.* 292:733–737. https://doi.org/10.1126/science.1058498.

14. Lareau, L. F., D. H. Hite, …, P. O. Brown. 2014. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife.* 3. https://doi.org/10.7554/eLife.01257.

15. Wu, B., C. Eliscovich, …, R. H. Singer. 2016. Translation dynamics of single mRNAs in live cells and neurons. *Science.* 352:1430–1435. https://doi.org/10.1126/science.aaf1084.

16. Yan, X., T. A. Hoek, …, M. E. Tanenbaum. 2016. Dynamics of translation of single mRNA molecules in vivo. *Cell.* 165:976–989. https://doi.org/10.1016/j.cell.2016.04.034.

17. Fiers, W., R. Contreras, …, M. Ysebaert. 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: Primary and secondary structure of the replicase gene. *Nature.* 260:500–507. https://doi.org/10.1038/260500a0.

18. Tutucci, E., M. Vera, …, R. H. Singer. 2018. An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat. Methods.* 15:81–89. https://doi.org/10.1038/nmeth.4502.

19. Eliscovich, C., S. M. Shenoy, and R. H. Singer. 2017. Imaging mRNA and protein interactions within neurons. *Proc. Natl. Acad. Sci. USA.* 114:E1875–E1884. https://doi.org/10.1073/pnas.1621440114.

20. Gold, V. A., P. Chroscicki, …, A. Chacinska. 2017. Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO Rep.* 18:1786–1800. https://doi.org/10.15252/embr.201744261.

21. Hurowitz, E. H., and P. O. Brown. 2003. Genome-wide analysis of mRNA lengths in Saccharomyces cerevisiae. *Genome Biol.* 5:R2. https://doi.org/10.1186/gb-2003-5-1-r2.

22. Zenklusen, D., D. R. Larson, and R. H. Singer. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15:1263–1271. https://doi.org/10.1038/nsmb.1514.

23. Hershkovitz, E., G. Sapiro, …, L. D. Williams. 2006. Statistical analysis of RNA backbone. *IEEE ACM Trans. Comput. Biol. Bioinf.* 3:33–46. https://doi.org/10.1109/tcbb.2006.13.

24. Hsiao, C., S. Mohan, …, L. D. Williams. 2006. Single nucleotide RNA choreography. *Nucleic Acids Res.* 34:1481–1491. https://doi.org/10.1093/nar/gkj500.

25. Levitt, M. 1978. How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proc. Natl. Acad. Sci. USA.* 75:640–644. https://doi.org/10.1073/pnas.75.2.640.

26. Vanzi, F., Y. Takagi, …, Y. E. Goldman. 2005. Mechanical studies of single ribosome/mRNA complexes. *Biophys. J.* 89:1909–1919. https://doi.org/10.1529/biophysj.104.056283.

27. Seol, Y., G. M. Skinner, and K. Visscher. 2004. Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys. Rev. Lett.* 93:118102. https://doi.org/10.1103/physrevlett.93.118102.

28. Yoffe, A. M., P. Prinsen, …, A. Ben-Shaul. 2011. The ends of a large RNA molecule are necessarily close. *Nucleic Acids Res.* 39:292–299. https://doi.org/10.1093/nar/gkq642.

29. Fang, L. T. 2011. The end-to-end distance of RNA as a randomly self-paired polymer. *J. Theor. Biol.* 280:101–107. https://doi.org/10.1016/j.jtbi.2011.04.010.

30. Poblete, S., and H. V. Guzman. 2021. Structural 3d domain reconstruction of the rna genome from viruses with secondary structure models. *Viruses.* 13:1555. https://doi.org/10.3390/v13081555.

31. Poblete, S., A. Božič, …, H. V. Guzman. 2021. RNA Secondary Structures Regulate Adsorption of Fragments onto Flat Substrates. *ACS Omega.* 6:32823–32831. https://doi.org/10.1021/acsomega.1c04774.

32. Clote, P., Y. Ponty, and J.-M. Steyaert. 2012. Expected distance between terminal nucleotides of RNA secondary structures. *J. Math. Biol.* 65:581–599. https://doi.org/10.1007/s00285-011-0467-8.

33. Lai, W. J. C., M. Kayedkhordeh, …, D. N. Ermolenko. 2018. mRNAs and lncRNAs intrinsically form secondary structures with short end-to-end distances. *Nat. Commun.* 9:4328. https://doi.org/10.1038/s41467-018-06792-z.

34. Khong, A., and R. Parker. 2020. The landscape of eukaryotic mRNPs. *RNA.* https://doi.org/10.1261/rna.073601.119.

35. Marchat, L. A., S. I. Arzola-Rodríguez, …, C. Lopez-Camarillo. 2015. DEAD/DExH-Box RNA helicases in selected human parasites. *Kor. J. Parasitol.* 53:583–595. https://doi.org/10.3347/kjp.2015.53.5.583.

36. Bundschuh, R., and T. Hwa. 2002. Statistical mechanics of secondary structures formed by random RNA sequences. *Phys. Rev. E.* 65:031903. https://doi.org/10.1103/physreve.65.031903.

37. Yoffe, A. M., P. Prinsen, …, A. Ben-Shaul. 2008. Predicting the sizes of large RNA molecules. *Proc. Natl. Acad. Sci. USA.* 105:16153–16158. https://doi.org/10.1073/pnas.0808089105.