

Visualization of conformational variability in the domains of long single-stranded RNA molecules

Jamie L. Gilmore^{1,*}, Aiko Yoshida¹, James A. Hejna² and Kunio Takeyasu^{1,3}

¹Laboratory of Plasma Membrane and Nuclear Signaling, Kyoto University Graduate School of Biostudies, Yoshida-konoe, Sakyo-ku, Kyoto 606-8501, Japan, ²Laboratory of Science Communication, Kyoto University Graduate School of Biostudies, Yoshida-konoe, Sakyo-ku, Kyoto 606-8501, Japan and ³Institute of Biotechnology, National Taiwan University, 4F, No. 81, Chang-Xing St, Taipei 106, Taiwan

Received September 08, 2016; Revised May 24, 2017; Editorial Decision May 25, 2017; Accepted June 04, 2017

ABSTRACT

We demonstrate an application of atomic force microscopy (AFM) for the structural analysis of long single-stranded RNA (>1 kb), focusing on 28S ribosomal RNA (rRNA). Generally, optimization of the conditions required to obtain three-dimensional (3D) structures of long RNA molecules is a challenging or nearly impossible process. In this study, we overcome these limitations by developing a method using AFM imaging combined with automated, MATLAB-based image analysis algorithms for extracting information about the domain organization of single RNA molecules. We examined the 5 kb human 28S rRNA since it is the largest RNA molecule for which a 3D structure is available. As a proof of concept, we determined a domain structure that is in accordance with previously described secondary structural models. Importantly, we identified four additional small (200–300 nt), previously unreported domains present in these molecules. Moreover, the single-molecule nature of our method enabled us to report on the relative conformational variability of each domain structure identified, and inter-domain associations within subsets of molecules leading to molecular compaction, which may shed light on the process of how these molecules fold into the final tertiary structure.

INTRODUCTION

The acquisition of high resolution three-dimensional (3D) structures of the ribosome in 2000 was considered an astounding breakthrough for science (1–4), and was recognized with the award of the 2009 Nobel Prize in Chemistry (5). To date, ribosomal RNA (rRNA) for the large subunit and small subunit of the ribosome in complex with their constituent ribosomal proteins are the longest RNA

molecules for which 3D structures have been obtained. Generally, the development of an effective experimental strategy that can yield 3D structures of a given RNA molecule is a time-consuming, painstaking process. Often, additional stabilizing factors, such as proteins or stabilizing mutations are necessary. To overcome these issues, we have developed a high-throughput technique for investigation of the global structure of long single-stranded RNA molecules (>1 kb), which can be easily applied to a range of RNA molecules. We have previously demonstrated the feasibility of this technology for imaging the domain architecture of partially folded RNA molecules by imaging in low salt conditions (6). With the development of automated MATLAB-based algorithms to estimate the range of nucleotides contained in each domain, and the use of deletion mutants to confirm the presence of structures in particular regions of the molecule, we have been able to identify conserved structural domains known to exist in the RNA genome of the Hepatitis C Virus (HCV) (7). In this study, we expanded the use of these methods to image and analyze the domain architecture of *Homo sapiens* 28S rRNA isolated from HeLa cells. Given that this molecule is very well characterized, it is ideal to test the validity of our method by comparing our analysis to existing structural models.

Our automated atomic force microscopy (AFM) analysis algorithm was able to identify a domain structure in 28S rRNA that correlates well with previously reported secondary structural models. Specifically, we identified five large structural domains (300–1000 nt) which overlap with regions of domain I, II, IV, V and VI of the six well-established canonical domains (8–12) (Supplementary Figure S1). In our assay, we also observed some variations from structural models. This includes the identification of four previously unreported smaller domains (200–300 nt). Two of these small domains are located in the 5' region of canonical domain II, and the other two comprise most of canonical domain III; this contrasts to most other models in which domain III is represented as a single domain structure. Also, the single-molecule nature of this method allowed us to characterize the conformational variability of each identi-

*To whom correspondence should be addressed. Tel: +81 75 753 7905; Fax: +81 75 753 7906; Email: gilmore.jamie.37m@st.kyoto-u.ac.jp

fied domain, as well as interdomain associations that occurred in some subsets of the molecules. Particularly, the Y-shaped expansion segments corresponding to the 3' region of canonical domain I and most of domain IV have a very reproducible morphology throughout all of the molecules, in contrast to regions corresponding to canonical domain 0 and the peptidyl transferase center (PTC), which were conformationally variable. Also, the structure corresponding to the 3' region of canonical domain II was observed to associate with the four flanking smaller domains in a subset of molecules to form a higher order domain structure, causing the molecules to have a higher degree of molecular compaction in comparison to the rest of the molecules. Possible reasons for these structural variations are discussed. The findings in this study demonstrate the potential of this technique for high-throughput identification and morphological analysis of the structural details in a variety of long RNA molecules such as viral RNA genomes or mRNA molecules.

MATERIALS AND METHODS

Isolation of RNA from HeLa cells

HeLa cells were grown in two 100 mm culture dishes in Dulbecco's modified Eagle's medium and maintained in a 37°C incubator with 5% CO₂. Once the cells reached confluency, the RNA was extracted according to the directions of the RNeasy mini Kit (Qiagen) and stored at -80°C. This sample preparation procedure entailed lysis of the cells with the chaotropic salt guanidine thiocyanate, followed by purification of the RNA by means of a silica membrane. The use of guanidine thiocyanate has been demonstrated to lyse cells while inactivating nucleases (13), and has been demonstrated to produce nucleic acid which is suitable for subsequent enzymatic treatments. Chaotropic salts have been shown to effectively dissociate proteins from ribosomal RNA (14). The ability of these salts to disrupt these complexes has been attributed to their ability to disrupt electrostatic interactions in addition to disrupting hydrophobic interactions through their lyotropic effects (changes to the H-bonded structure of water) which stabilize the interactions between ribosomal subunits. Furthermore, the sample was then treated with DNase, and then filtered through a silica membrane. With this method, the negatively charged RNA interacts with the hydroxyl groups of the silica membrane through salt bridges formed by the cations in solution; the membrane was then washed with ethanol, and RNA was subsequently eluted with low-salt nuclease-free water. The quality of the RNA was assessed by agarose gel electrophoresis to confirm the presence of 28S and 18S rRNA bands.

AFM sample preparation and imaging

Samples were prepared and imaged as described previously (6). The RNA was diluted about 700-fold to a concentration of 0.1 ng/μl with a final volume of 10 μl in 10 mM Tris-Cl, pH 8.5 containing no Mg²⁺ ions with a low ionic strength of around 2 mM. The RNA was then briefly heated to 65°C which should partially denature the less stable regions of

the molecule. The sample was then kept at room temperature for ~15 min while the mica surface was being modified with 10 μl 10 mM spermidine for 3.5 min and rinsed with ~2–3 ml water. The sample was then immediately deposited on the spermidine-treated mica for 3.5 min, rinsed with ~2–3 ml water and dried with a stream of N₂ gas. The principle of RNA interaction with the silica membrane is similar to the interaction of the RNA with the mica substrate, which is a kind of sheet silicate which cleaves into atomically flat layers with hydroxyl groups on the surface. In this case, the two-step modification procedure using the trivalent polyamine spermidine followed by deposition of the RNA sample allows the molecules to interact with the surface through electrostatic interactions. Since RNA folding is generally regarded as hierarchical (15), disruption of the Mg²⁺-dependent tertiary structure should not lead to dramatic changes in the secondary structure. This claim is supported by our observation that many of the structural elements appeared to correlate with those of the cryo-EM structure (Figure 3). After drying, the sample could then be immediately imaged in Tapping Mode™ with the Multimode AFM/Nanoscope system (Digital Instruments, Inc, USA) and rectangular silicon cantilevers with sharpened tetrahedral tips (OMCL AC160 TS, Olympus Corp, Japan).

AFM analysis

Images of each molecule were flattened and zoomed to an area of 300 × 300 nm² in the Nanoscope software version 5.31 (Veeco Instruments, Inc.). Then, each zoomed image was further processed in the Gwyddion software (16) using a Pygwy batch processing script which runs the Gwyddion Remove Scars, Plane Leveling, and Median Line Correction operations with features of the image with a mask applied to exclude features of the image with height values exceeding 0.05 nm from the data correction operations, and then saves each image as an ASCII data matrix (.txt) file. The text files were then opened in MATLAB as 512 × 512 matrices of height values using the `dlmread` command.

The images were then run through a series of automated algorithms using the MATLAB image processing toolbox R2014b (17) to analyze the structure of the RNA (Supplementary Figures S2–5). First, a height threshold of 0.05 nm was applied to select the region enclosing the molecule (Supplementary Figure S2 B and C). Then, the noisy edges of the molecule were smoothed by the application of 20 iterations of an edge-based active contours algorithm followed by application of the MATLAB `bwmorph` 'thin' operation (Supplementary Figure S2D and Figure 1B, middle) to generate skeletons of the molecule. Then, the longest end-to-end chain of the skeleton was identified by performing a quasi-euclidean geodesic distance transform from each endpoint in the molecule skeleton and deleting all branches not corresponding to the two identified endpoints corresponding to the longest chain, in addition to deleting any loops or branches that had the comparatively lower minimum height value (Supplementary Figure S2E and Figure 1B, right). This automated procedure worked well for many molecules, but sometimes the main chain trajectory was adjusted manually to match the domain pattern identified in Figure 2 (Supplementary Figure S3). To get the length of the main

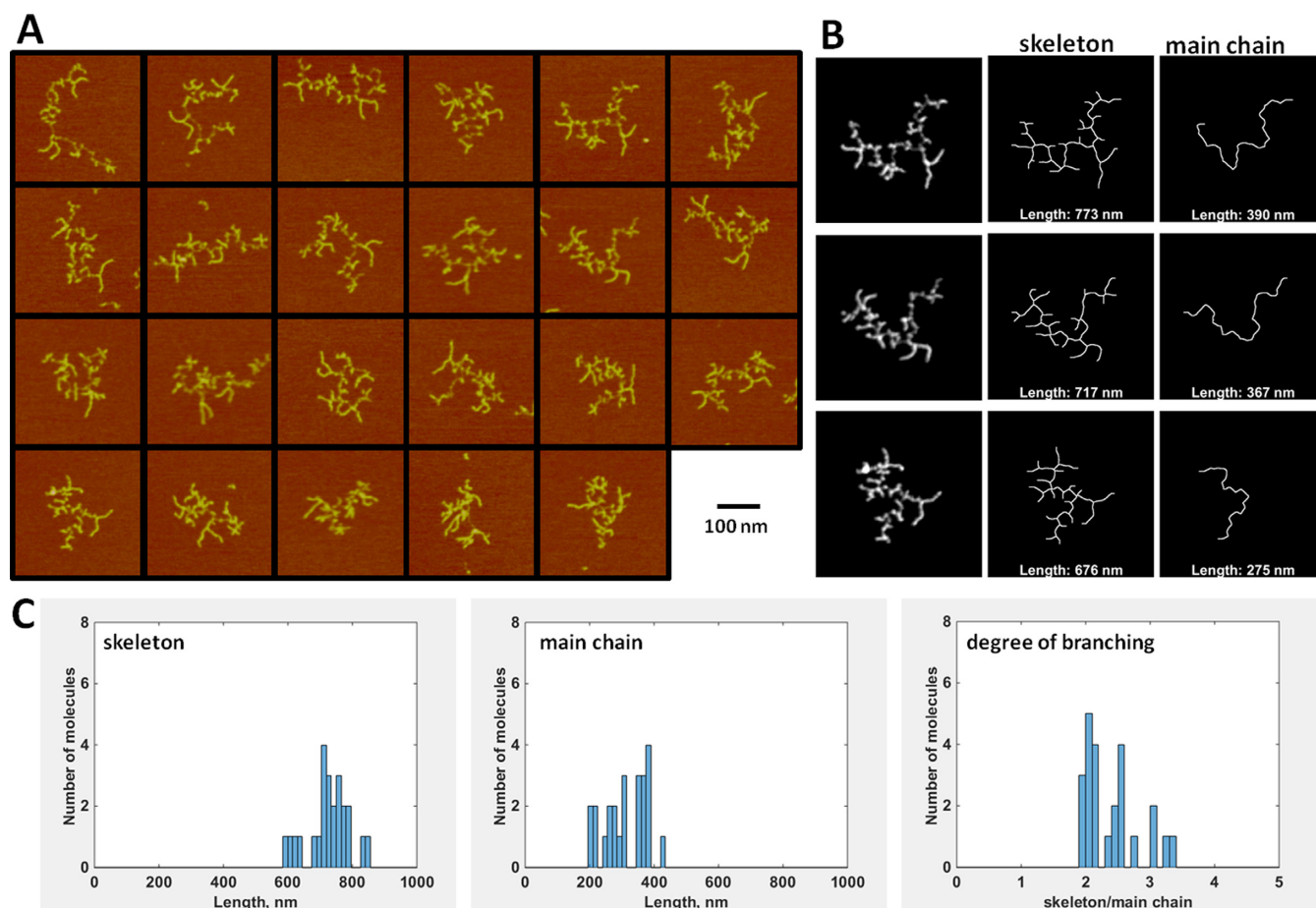


Figure 1. AFM imaging of 28S rRNA. **(A)** A gallery of $300 \times 300 \text{ nm}^2$ images of individual 28S rRNA molecules. **(B)** Selected images demonstrating the length of skeleton representations of each molecule in comparison to the longest end-to-end main chain. Lines are dilated to widths of three pixels for ease of visualization. **(C)** Histograms demonstrating the distribution of the lengths of the skeleton representations (left) and the main chains (middle) for $N = 23$ molecules. The degree of branching (right) was measured by taking the skeleton length divided by the main chain length.

chain, a quasi-euclidean geodesic distance transform was performed from one of the main chain endpoints and then converted to nm using a px/nm conversion factor, which was then used for the x-axis values in Figure 2A, middle. In order to generate the height profiles, the height value of the pixels directly underlying the main chain were used. To generate the volume profiles, it was necessary to account for the pixels lying adjacent to the main chain. To accomplish this, each pixel in the main chain was assigned a unique integer value by applying a chessboard geodesic distance transform (Supplementary Figure S2F), and then iteratively assigning four-connected pixels adjacent to the main chain the corresponding integer values until every pixel in the region of the molecule identified in Supplementary Figure S2B was assigned a value (Supplementary Figure S2G). The volume of each region corresponding to an integer value was calculated as the sum of all height values of the pixels in each integer-defined region, and the noisiness of the graphs was reduced by applying a 5 nm moving average to obtain the volume profile in Figure 2A, middle.

In order to assign the domains, a peak detection algorithm was applied to the height and volume profiles (Supplementary Figures S2H and 4) (18) using a slopethresh-

old of 0.005, a smoothwidth of 5, peakgroup of 1 and smoothtype of 1, and an ampthreshold of 0.3 for height profiles and of 2 for the volume profiles. Generally, height profiles are better for detecting small domains located along the trajectory of the main chain, and volume profiles are better at detecting domains branching from the main chain; peaks from both profiles were considered in order to avoid missing important structural features (Supplementary Figure S4A). To do this, all peaks from the height profiles were considered first, then taking any volume peak detected which was located more than 5 nm away from any height peak. Once the position of each peak along the molecular trajectory was recorded (Supplementary Figure S4B), the main chain pixel corresponding to each peak was labeled with the peak number to identify structures corresponding to each peak (Supplementary Figure S4C).

The nucleotide numbers were predicted according to the cumulative volume along the profiles by using a conversion factor generated by setting the total volume of each molecule to a value of 5070 nt, the known length of human 28S rRNA (Table 1 and Supplementary Figure S5). Domain boundaries were defined as the main chain pixel with the lowest height between two adjacent peaks or as re-

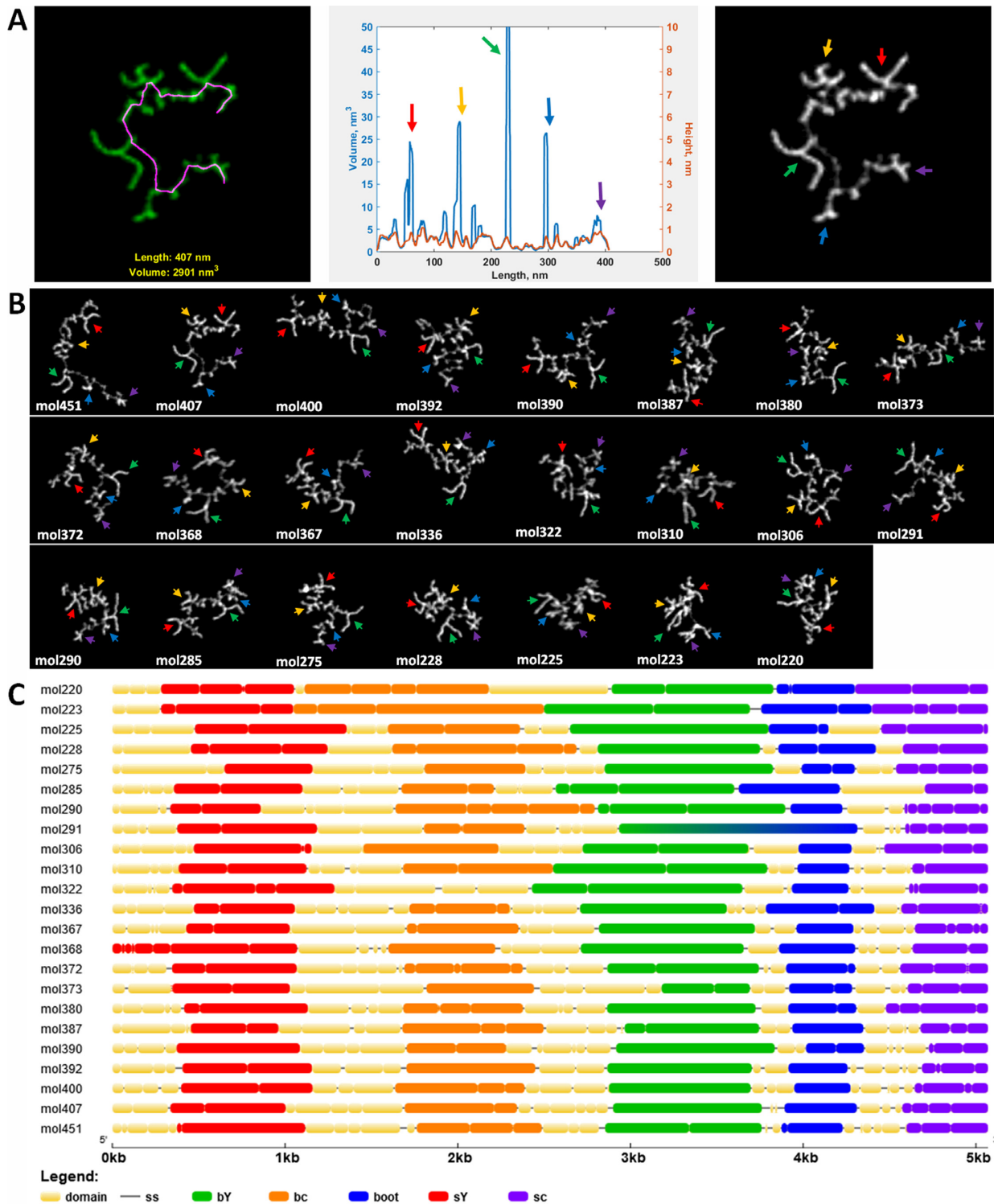


Figure 2. Domain identification in 28S rRNA. (A) A method of generating profiles along the main chain of the molecule (left) can be used to quantify the height of pixels immediately underlying the main chain or can be used to quantify the local volume along the main chain by summing pixels in the region of the molecule which are in the vicinity of each main chain pixel (middle). The correspondence of peaks in the volume profile to structures along the molecule are denoted by colored arrows (middle and right). (B) Analogous structures in images of 23 molecules are denoted by colored arrowheads as in (A). Here each molecule is named using a descriptor, 'mol', followed by the length of the main chain in nanometers. (C) The cumulative volume in the profiles was used to estimate the number of nucleotides in each domain by using a conversion factor assuming that the total volume of each molecule corresponds to the total number of nucleotides (5070 nt) in the molecule. The prominent domains identified in (A) and (B) were assigned names according to their general morphology: including sY (small Y)-red, bc (big claw)-orange, bY (big Y)-green, boot-blue and sc (small claw)-purple. Regions corresponding to other domain structures are shown in yellow in addition to detected single stranded regions represented as a line. The molecules are listed from top to bottom from the shortest (most compact) main chain length to the longest (most extended) main chain length, respectively.

Table 1. The predicted start nt, end nt, nt range and the number of molecules (*n*) used to obtain the values for each domain (\pm sd)

Domain	Start nt	End nt	nt range	<i>n</i>
sY	398 \pm 74	1107 \pm 99	722 \pm 129	22
bc	1687 \pm 84	2391 \pm 139	703 \pm 145	19
bY	2806 \pm 164	3734 \pm 81	932 \pm 169	21
boot	3900 \pm 86	4290 \pm 54	390 \pm 81	20
sc	4598 \pm 67	5070 \pm 0	471 \pm 66	21
sY \rightarrow bc #1	1084 \pm 83	1357 \pm 97	273 \pm 80	17
sY \rightarrow bc #2	1422 \pm 85	1687 \pm 62	264 \pm 63	16
bc \rightarrow bY #3	2386 \pm 100	2596 \pm 97	209 \pm 54	14
bc \rightarrow bY #4	2644 \pm 89	2844 \pm 101	199 \pm 73	15

The standard deviation for the end nt of the sc domain is 0 because the total cumulative volume of each molecule is assumed to correspond to the total number of nucleotides in the molecule (5070 nt).

gions where the main chain pixel fell below a height value of 0.3 nm, which were defined as single-stranded regions (Supplementary Figure S2I). Since single-stranded regions were predicted by height, yet averaged volume profiles were used to predict the nucleotides, the estimated number of nucleotides was unreasonably high when predicted using cumulative volume, especially when the single-stranded region was adjacent to a large structural domain. Instead, the length of the single-stranded region along the main chain was used to predict the nucleotide ranges of the single-stranded regions (Supplementary Figure S5). Based on a previous study of the HCV genome with a clearly discernible single-stranded polyU/UC region of 103 nt in the 3'UTR, it was found that the measured length was about half the number of nucleotides (6), so a rough estimate of 2 nt/nm was used. Further studies with other RNA molecules may help to refine this value in the future. The excess nucleotides from volume predictions in the single-stranded regions were assigned to the adjacent left and right domains proportional to the percentage volume for the left and right halves of the single-stranded region along the main chain, respectively. Diagrams of the predicted nucleotide ranges (Figure 2C) were generated using the Gene Structure Display Server (GSDS) (19).

Length measurements of the skeletons (Figure 1B, middle and C, left) were made by removing line pixels in a 9×9 px² area around each branchpoint of the skeleton, detecting each branch with connected component labeling, performing a quasi-euclidean geodesic distance transform from an endpoint of each branch followed by conversion of the pixel length to nm and then taking the sum of the length values for all branches to get the total length (Supplementary Figure S2J). To get the length value for individual domain structures (Table 2), the branches of these skeletons, along with the branches of skeletons generated from images with a 0.3 nm threshold applied to remove single-stranded regions (Supplementary Figure S2K and L) were used. To get the length value, the line from the images of the two skeletons which appeared to best approximate the overall shape of the domain structure of interest was used. Histograms of the length values of skeletons and main chains (Figure 1C) were generated in MATLAB.

RESULTS

AFM imaging of 28S ribosomal RNA

Images of ribosomal RNA were obtained by depositing total RNA extracted from HeLa cells on spermidine-treated mica. In order to observe the secondary structure of the molecules, they were prepared in low salt conditions. In contrast to proteins, where the secondary structure undergoes dramatic rearrangements upon formation of the tertiary structure, RNA folding is hierarchical, with the secondary structures forming stable autonomous structures independent of the addition of cations, which then coalesce into the tertiary structure upon the addition of Mg²⁺ ions (15). This means that we should be able to observe and characterize individual domains present in RNA molecules by omitting ions from the sample preparation buffer. Individual zoomed images of the 28S rRNA molecules were generated by selecting the molecules by height and area thresholding (Figure 1A). The volume values for the 28S rRNA ranged from 2760 to 3254 nm³ with an average volume of 3012 ± 145 nm³. These values correspond well with previously reported values (6), and are in accordance with the known length of 5070 nt for 28S rRNA.

The 28S rRNA molecules have a noticeable branched morphology in the images. To quantitate this, we compared the length of skeleton representations of the molecules generated by morphological thinning to the longest end-to-end main chains from those skeleton representations (Figure 1B and C; Supplementary Figures S2D,E and 3). The degree of branching was measured by dividing the length of the skeleton representations by the length of the main chains. We found that the skeleton representations were, on average, 2.3-fold longer than the main chains, suggesting that the main chain of the molecules comprised about 43% of the length of the entire molecule (Figure 1B and C). It is important to note that this procedure does not differentiate between single-stranded regions and more structured regions along the molecule, but is simply a measure of the overall branching pattern of the molecule as a whole.

Histograms of the skeleton and main chain lengths are dispersed over a wide range of values. The 28S skeletons range in length from 595–846 nm (average: 727 ± 64 nm) and the main chains range from 220–451 nm (average: 333 ± 67 nm) (Figure 1C). This is demonstrated in Figure 1A, in which the images are arranged with the most extended conformations in the top rows and the more compact conformations in the bottom rows.

Table 2. Length of various elongated structures on the 28S rRNA (\pm sd) and the number of molecules (n) used to obtain the values for each structure

Structure	Length (nm)	n
sY		
long branch	36 \pm 4	23
short branch	33 \pm 4	23
stalk	11 \pm 4	13
bY		
long branch	52 \pm 4	23
short branch	36 \pm 9	23
stalk	20 \pm 7	22
bc		
branch1 (5')	27 \pm 4	15
branch2	24 \pm 4	15
branch3 (3')	22 \pm 4	15
boot		
toe	16 \pm 4	17
heel	7 \pm 2	14
stalk	15 \pm 7	15
5'Y\rightarrowbc #1	12 \pm 3	10
5'Y\rightarrowbc #2	32 \pm 6	13
bc\rightarrowbY #3	27 \pm 6	6
bc\rightarrowbY #4	22 \pm 4	8

Measurements are made only for structures that could be clearly detected in the skeleton representations of the images, so more compact conformations of the structural features may not be represented. Skeleton representations were generated for both the entire molecule (selected using a height threshold of 0.05 nm) and of the individual domains (selected using a height threshold of 0.3 to filter out single stranded regions of the molecule) (see Supplementary Figure S2K and L). The length measurement was selected based on the branch shape that appeared to best follow the domain structure of interest, although there may be some anomalies, especially when part of the structure appears to fall in line with the main chain of the molecule.

mations on the bottom. This range of values suggests that some molecules are in more extended conformations than others. Thus, these images can likely shed light on the conformational variations of the molecules in addition to some of the initial steps of folding into the more compact 3D structures, provided that a method to recognize the arrangement of 'domains' along the trajectory of the molecules can be developed. We describe this method below.

Identification of domains in 28S rRNA molecules

To further disambiguate our AFM images, we developed systematic procedures for the identification of the reproducible domain architecture of the molecules. To do this, the main chains (Figures 1B, right and 2A, left) were used to generate profiles of the height and volume along the trajectory of the molecule (Figure 2A, middle), with the height profiles representing the height values of the pixels underlying the main chain and the volume profiles summing the height of pixels adjacent to the main chain to obtain the local volume (Supplementary Figure S2F and G). Compared to the height profiles, the volume profiles are able to account for the branches extended away from the main chain of the molecule. In this way, large structures in the molecule can be identified as prominent peaks in the volume profiles. The correspondence of volume profile peaks to structures in the RNA molecule is shown in Figure 2A. The structures were named the 'small Y' (sY), the 'big claw' (bc), the 'big Y' (bY), the 'boot' and the 'small claw' (sc). All molecules showed structures of similar morphology located along their trajectories (Figure 2B).

To confirm the nature of these structures, we used the cumulative volume from the volume profiles combined with methods of detecting profile peaks (Supplementary Figure S2H and 4) and single stranded regions (Supplementary

Figure S2I) in order to provide an estimate of the range and number of nucleotides comprising each domain (Figure 2C, Supplementary Figure S5 and Table 1). All molecules in Figure 2B and C are named according to the contour length of the main chain, which ranges from 220 nm (mol220) up to 451 nm (mol451) and are arranged from the most extended to their most compact conformations. It is evident that the domains in the extended molecules are well separated from one another, and the domains in the shortened molecules begin to coalesce with one another and become more variable in their predicted ranges. Generally, the molecules were highly structured with only 134 ± 60 nt (2.6%) of the inter-domain residues in each molecule predicted to have a single stranded morphology detectable in the images. This value ranged from 47–273 nt, with more compact molecules having a decreased number of single-stranded nucleotides (Supplementary Figure S6). It should be noted that the number of nucleotides is a rough estimate, assuming 2 nt/nm along the length of the main chain based on previous studies of a poly(U) region in HCV RNA (6) and that further studies may help to refine this conversion factor.

The 5' and 3' ends were identified by comparison to existing secondary structural models determined from cryo-EM structures of human ribosomal RNA (Supplementary Figure S1) (10,11). Labeling the secondary structural model with the predicted nucleotide range (Table 1) for each of the identified domains in Figure 2 shows good correlation between the overall domain shape and organization of helices in our AFM images (Figure 3A and B). Particularly, the two large Y-shaped structures (bY and sY) in our images correspond well to the expected range for the eukaryotic-specific expansion segments (ES), with the small Y structure corresponding to helix25/ES7 and the big Y structure corresponding to helix 63/ES27. This striking correlation with the known structure of human ribosomal RNA suggests

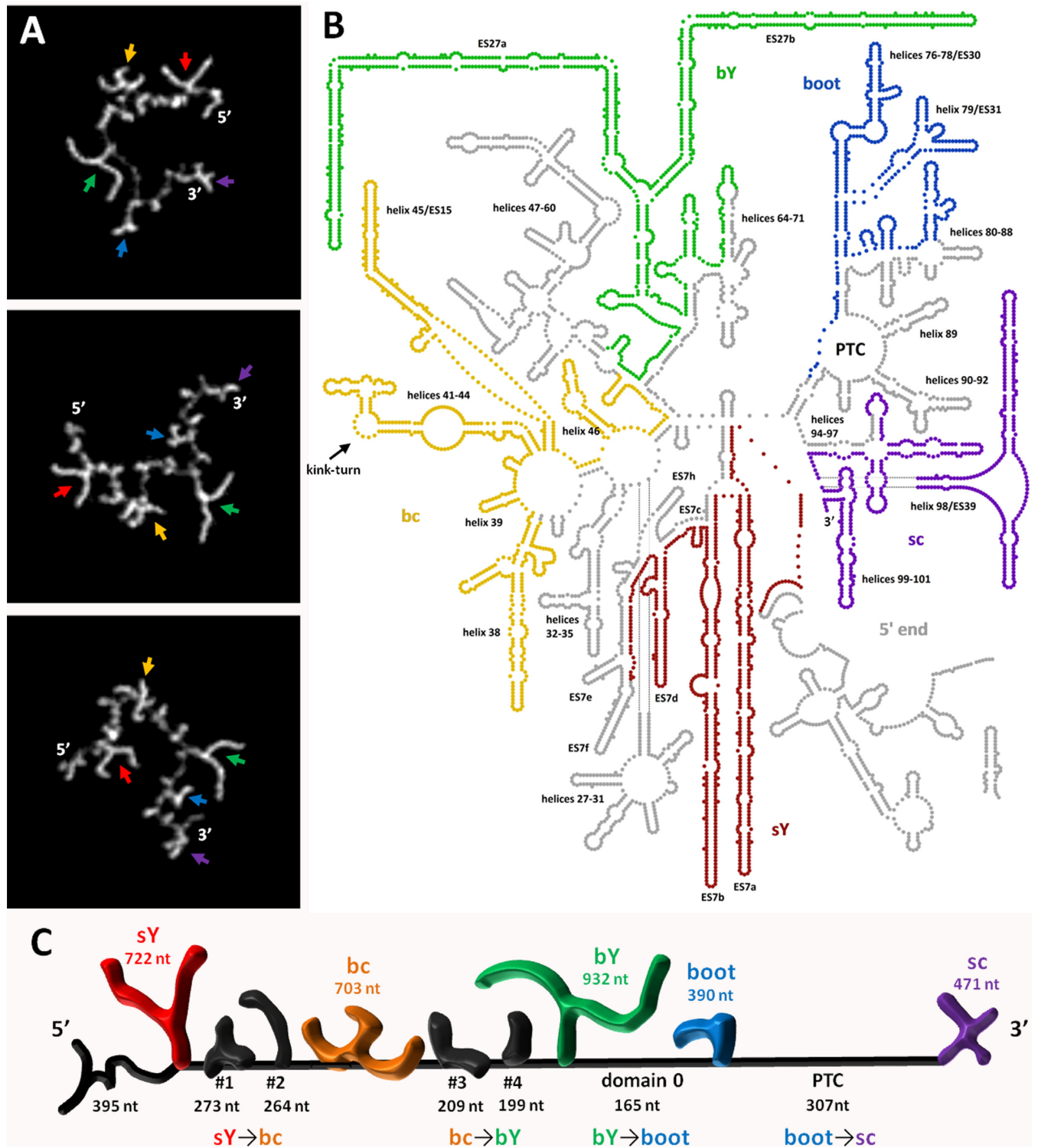


Figure 3. Comparison of AFM-based 28S rRNA structures to previously reported 3D-based structures. (A) Representative AFM images of the 28S rRNA with domains identified in Figure 2 (arrows). (B) Image of the 3D-based secondary structure of *Homo sapiens* 28S rRNA (11,12) generated with ribovision (10) showing how the predicted nucleotide ranges from our analysis (Table 1) coincide with predicted secondary structures of the 28S rRNA molecules from our AFM images shown on the right. Domains are color-coded as the arrows in (A). Helices and eukaryotic-specific expansion segments (ES) of interest are labeled, in addition to a kink-turn motif reported to exist within helix 42 (22,23) and the location of the peptidyl transferase center (PTC). For a more detailed description of the helices and nucleotide numbering, please refer to the aforementioned references (11,12). (C) Diagram of the conserved domains identified in AFM images and the predicted number of nucleotides comprising each domain. The five prominent domains identified in (A) are color-coded as in (A), in addition to 4 smaller domains shown in black.

that AFM is a viable method for obtaining structural information about the domain architecture of single-stranded RNA molecules.

To further investigate the conformations of each domain, each imaged molecule was divided into ten regions, comprising the five prominent domains (sY, bc, bY, boot and sc) (Figure 4B, D, F, H and J) and five interdomain regions (5' end, sY→bc, bc→bY, bY→boot, boot→sc) (Figure 4A, C, E, G and I). Within these regions, a total of nine reproducibly observed domain structures ranging in size from 199 to 932 nt (Figure 3C and Table 1) were identified. Three of them had calculated lengths in excess of 700 nt, including the two Y-shaped domains termed the big Y (bY) and the small Y (sY) (Figure 4B and F), in addition to the big claw (bc) domain (Figure 4D). The other two domains identified from the profiles are the boot (Figure 4H) and small claw (sc) (Figure 4J) with estimated sizes of 390 ± 81 and 471 ± 66 nt, respectively. Of the five interdomain regions, three of these appeared to be structured (Figure 4A, C and E). Notably, the region between the sY and bc structures and the region between the bc and bY domain each contained two smaller domains ranging in estimated size from ~200–300 nt (Figure 4C and E, arrows). The 5' end (Figure 4A) also appeared to form structures of varying morphology. The remaining two interdomain regions (Figure 4G and I) also had varying morphology ranging from relatively unstructured to being interspersed with some small structure(s). Further characteristics of each of these regions and their varying conformations are described below.

Conformational differences in large Y-shaped expansion segments

Of the domains identified in these molecules, two of the most readily identifiable are the highly conserved large Y-shaped domains, (Figure 4B and F). Both of these domains appeared to have two conspicuous branches which were attached to the main chain by a short stalk. The sY and bY have average sizes of 722 ± 129 and 932 ± 169 nt, respectively, with the sY located ~400 nt from the 5' end and the bY located ~1300 nt from the 3' end (Table 1). The predicted nucleotide range for both of these structures corresponds to ES, with the sY corresponding to helix 25/ES7 located in canonical domain I and the bY corresponding to helix 63/ES27 located in canonical domain IV (Figure 3 and Supplementary Figure S1). As an evolutionary point of interest, helix 25/ES7 has expanded from a 22 nucleotide helix in *Escherichia coli* to a 876 nt structure in *H. sapiens*, and helix 63/ES27 has expanded from a 45 nt helix in *E. coli* to a 718 nt structure in *H. sapiens* (12). Interestingly, helices c-h (nt 955–1285) of ES7 from the previously reported 3D-based structure are not fully encompassed within our predicted nucleotide range for the sY (Figure 3). The sY and bY domains also have the highest GC content (80.7 and 78.6%, respectively) of any of the 10 regions (Table 3). This most likely reflects the identity of these regions as expansion segments, as it has been reported that vertebrate expansion segments of rRNA tend to be GC-rich, in contrast to those of *Drosophila* which are generally AT-rich (20).

The conformations and identities of these structures were further characterized by measuring the length of each

branch. Since it is possible that the branches of the Y-shaped structures may lie in different conformations in the images, length measurements for these two structures were sorted based on the long and short branches (Table 2). For the sY structure, the measured branch lengths were 36 ± 4 and 33 ± 4 nm, which would correspond to branch lengths of 129 and 119 bp, respectively, assuming that these branches represent double-stranded A-form helices with a helical rise of 0.28 nm/bp. These values are in close agreement with the expected branch lengths based on the previously reported 3D-based secondary structures of ES7a (nt 455–701, 123 bp helix) and ES7b (nt 713–955, 121 bp helix) (Figure 3). For the bY structure, the measured branch lengths were 52 ± 4 and 36 ± 9 nm, corresponding to branch lengths of 186 and 129 bp, respectively. The value for the long branch is reasonably close to the expected length for branch ES7a (nt 2914–3279, 183 bp helix); however, the measured value for the shorter branch falls below the expected length for branch ES27b (nt 3285–3582, 149 bp) (Figure 3) from previously reported 3D-based secondary structures. The reason for this becomes clear upon examination of individual images of the bY structure (Figure 4F), in which some of these structures exhibit both branches with an elongated morphology (Figure 4Fi–iv) and others have one of the branches exhibiting a noticeably shortened morphology (Figure 4Fv–viii). Interestingly, ES27 has been reported to be particularly dynamic in 3D structures of the yeast ribosomal complex as well (21).

Both of the Y-shaped structures were connected to the main chain by a short stalk, with the sY having an average stalk length of 11 ± 4 nm (observed in $n = 13$ molecules, 57%) and the bY having a longer stalk length of 20 ± 7 nm (observed in $n = 22$ molecules, 96%) (Table 2). Also, the bY structures were generally observed to have both arms falling on the same side of the main chain (in $n = 22$ molecules, 96%) (Figure 4Fi–iv, vi–viii), and only one molecule observed to have the branches extending from opposite sides of the main chain (Figure 4Fv). In contrast, the sY was observed to have both branches extending from the same side of the main chain in slightly fewer molecules ($n = 19$ molecules, 83%) (Figure 4B, i–iv), with the branches falling on opposite sides in the remaining four molecules (17%) (Figure 4Bv and vi). These morphological differences are likely a result of how the molecules were oriented on the sample surface, and the shorter stalk length likely allows for the branches to more readily fall on opposite sides. Beyond these slight conformational differences, just two of the sY structures exhibited association with other domain structures, with one appearing to have the 5' end situated at the base of the sY structure (Figures 2B and C, mol368 and 4Bvii) and another appearing to associate with structures just 3' of the two arms (Figure 4Bviii).

A three-branch morphology in the big claw (bc) domain

The other large domain observed in these molecules is what we refer to as the big claw (703 ± 145 nt), which exhibits a morphology of three branches extending from a central region. This morphology was apparent in $n = 15$ molecules (65%) (Figure 4Di–vi), with $n = 3$ (13%) molecules having a dissociated morphology (Figure 4Dvii and viii), and the remaining molecules appearing to form a higher order do-

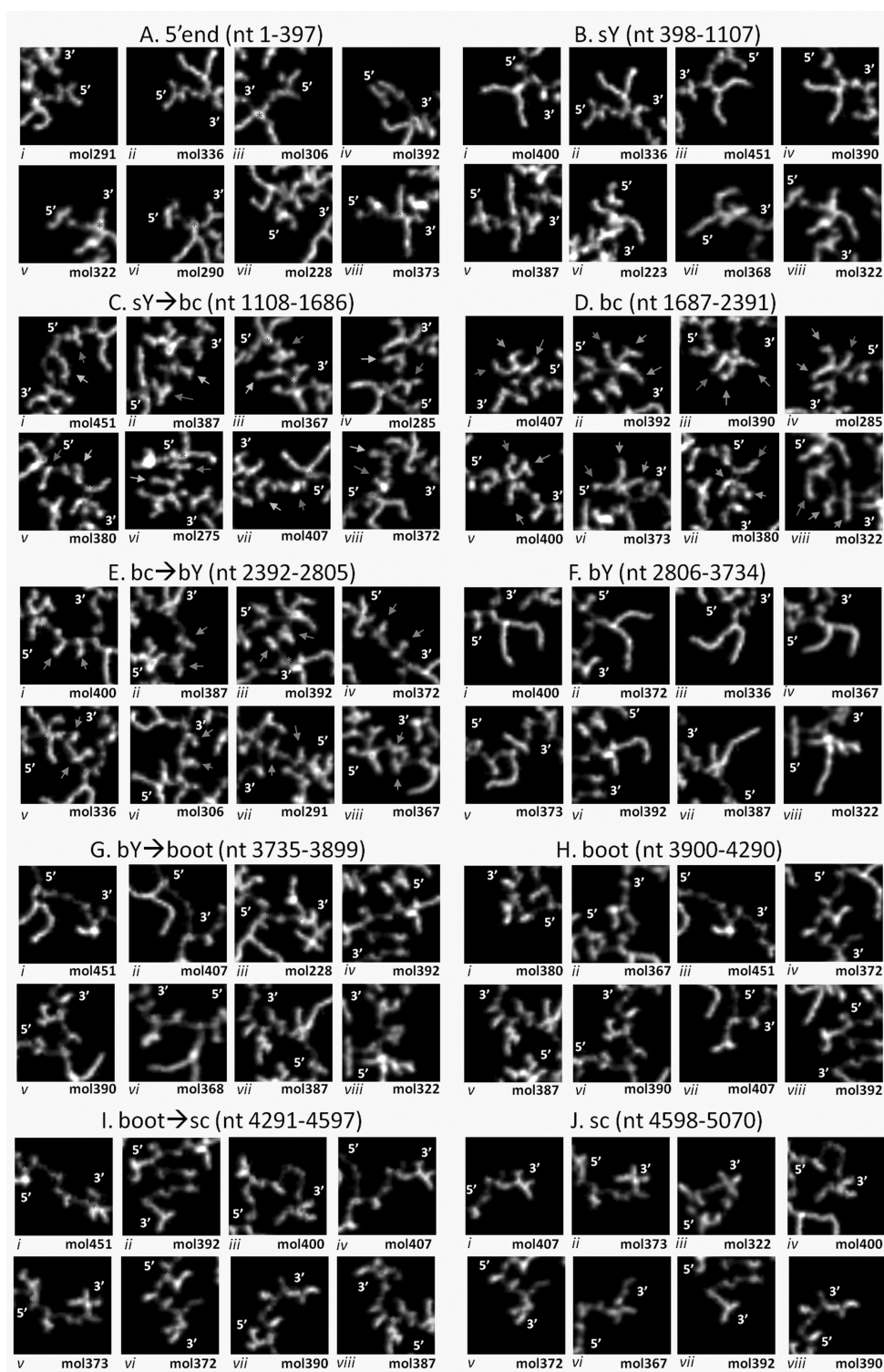


Figure 4. Representative zoomed images ($100 \times 100 \text{ nm}^2$) of each prominent domain identified in Figure 2 in addition to each interdomain region. The regions are as follows: (A) 5' end, (B) sY, (C) sY \rightarrow bc, (D) bc, (E) bc \rightarrow bY, (F) bY, (G) bY \rightarrow boot, (H) boot, (I) boot \rightarrow sc and (J) sc. The three branches of the bc (D) are labeled with colored arrowheads from the one which appears most 5' to the one which appears most 3'. In each set of interdomain images (A,C,E,G,I), the prominent domains are labeled with asterisks in order to delineate the interdomain boundaries. (A) Representative structures of the 5' end with the sY structure labeled with a red asterisk. (C) Representative structures of the sY \rightarrow bc region with the sY and bc structures labeled with red and yellow asterisks, respectively. The two main structures observed in this region are labeled with colored arrows. (E) Representative images of the bc \rightarrow bY region with the bc and bY structures labeled with yellow and green asterisks, respectively. The two main structures observed in this region are labeled with colored arrows. (G) Representative images of the bY \rightarrow boot region with the bY and boot structures labeled with green and blue asterisks, respectively. (I) Representative images of the boot \rightarrow sc region with the boot and sc structures labeled with blue and purple asterisks, respectively. The designation of each molecule, as originally described in Figure 2, is listed below each image.

Table 3. GC content for each of the regions displayed in Figure 4

Region	nt range	GC content
5'end	1–397	62.2%
sY	398–1107	80.7%
sY→bc	1108–1686	71.5%
bc	1687–2391	65.7%
bc→bY	2392–2805	65.7%
bY	2806–3734	78.6%
bY→boot	3735–3899	53.3%
boot	3900–4290	67.3%
boot→sc	4291–4597	47.9%
sc	4598–5070	67.9%

The nucleotide range for each region is based on predicted values in Table 1.

main structure in this region (discussed in more detail below) (Figure 5A). The nucleotide range calculated for these structures was nt 1687–2391, which corresponds to helices 38–46 of domain 2 (Figure 3 and Supplementary Figure S1). In images of the molecules with the three branch morphology, these branches can be observed to situate on the same side of the main chain in $n = 9$ (60%) (Figure 4Di–iv and vi), but also can be observed with branches oriented on different sides (Figure 4Dv). Usually, the three branches appear to originate from the main chain itself (Figure 4Di–v), but sometimes the structure can be observed to have a short stalk connecting the structure to the main chain (Figure 4Dvi). In two images, the branches of the structure appeared to be unassociated. In one molecule, the 5' branch appeared to be separated from the other two branches (Figure 4Dvii), and in another molecule, all three branches appeared to be dissociated from one another (Figure 4Dviii), which is the reason for this structure being unlabeled in molecule 322 of Figure 2C. Interestingly, in many of the molecules, the middle branch of the bc structure had a noticeable kink at the tip (Figure 4D, blue arrows). This kink likely reflects a well-known kink-turn motif which has been reported to exist within helix 42 (Figure 3) (22,23), and which has been reported to provide flexibility to the L7/L12 stalk domain in order to facilitate translocation of the tRNA and mRNA through the central cavity of the ribosome upon binding of EF-G in bacteria.

To further examine the identity and nature of the bc structure, length measurements of the branches were performed from the branch which appeared most 5' in the image to the one which appeared to be the most 3' in the image. The resulting measurements from the 5' arm to the 3' arm were 27 ± 4 , 24 ± 4 and 22 ± 4 nm (Table 2). If these structures form primarily A-form helices, each branch could be predicted to comprise about 96, 86 and 79 bp, respectively. It can be assumed that the three branches likely correspond to helix 38 (nt 1684–1852, 84 bp), helices 41–44 (nt 1899–2076, 89 bp) and helix 45/ES15 (nt 2078–2279, 100 bp) (Figure 3) of previously reported 3D-based secondary structures. With length measurement errors of 4 nm (14 bp) for each branch, the measured values for branches 1 and 2 are within error of the expected values, and the third branch has an observed value slightly shorter than the expected value. However, these measurements may vary due to the branches not being entirely A-form helices, by the orientation of the branches possibly criss-crossing in the images rather than

following the 5'→3' trajectory, or by the degree to which the base of the helix may extend from the central region of the structure. This structure has a moderately enriched GC-content of 65.7%, likely reflecting that the most 3' branch of this structure is an expansion segment (helix45/ES15) (Table 3).

A well-conserved boot-shaped structure corresponds to ES30 and ES31

The boot is another reproducibly observed domain which is predicted to contain 390 ± 81 nt and was identifiable as an individual domain structure in $n = 19$ molecules (83%). The calculated nucleotide range for this structure was nt 3900–4290, corresponding to the 5' region of canonical domain V (Figure 3 and Supplementary Figure S1). The dimensions of this domain include a 'toe' region with a length of 16 ± 4 nm, a 'heel' region with a length of 7 ± 2 nm and a 'stalk' of 15 ± 7 nm (Table 2). The stalk length is highly variable, with the structure often appearing very close to the main chain (Figure 4Hi–iv), and in various more extended conformations (Figure 4Hv–viii). The structures in this region likely include helices 75–79 (Figure 3). This region contains expansion segments 30 and 31 (ES30 and ES31), and overall, this region has undergone an increase from 167 nt in *E. coli* to 265 nt in *H. sapiens* (12). This relatively modest increase is consistent with the moderate GC content of 67.3%.

Conformational variability in the structures on the 5' and 3' ends

Although the 5' end was not identified as one of the prominent structures in the profiles, this region also displays a highly structured morphology in the images (Figure 4A) and can be said to encompass the nucleotides upstream of the sY structure (nt 1–397). The sc structure on the 3' end, however, was identified in the profiles and is predicted to contain 471 ± 66 nt ranging from nucleotides 4598 to 5070 (Table 1), suggesting that it comprises most of the region of canonical domain VI (Supplementary Figure S1). Both of these structures showed a number of different morphologies in the images. For example, the 3' sc structure was sometimes observed with three branches extending from the 3' end (Figure 4Ji–iii), sometimes observed with one of the three branches with an extended unfolded morphology (Figure 4Jiv–vi) and sometimes observed with one branch which appeared in line with the main chain and two

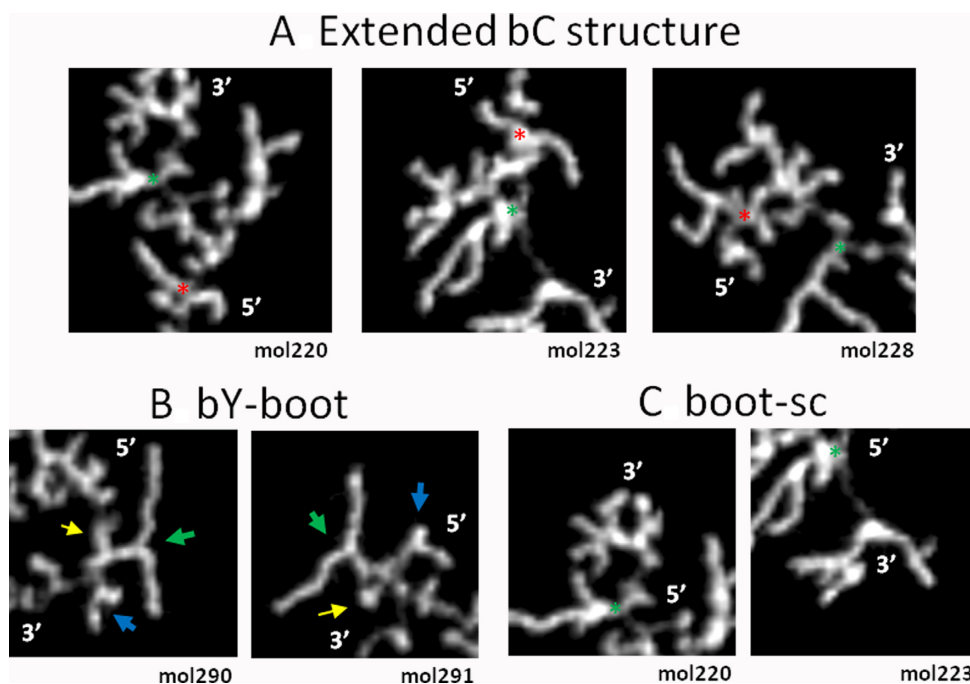


Figure 5. Representative zoomed images of the interdomain associations observed in some subsets of the molecules. (A) $170 \times 170 \text{ nm}^2$ images demonstrating a complex structure in the region where the bc structure is generally observed (see Figure 2B). The sY and bY structures bordering this structure are labeled with red and green asterisks, respectively. (B) $146 \times 146 \text{ nm}^2$ images demonstrating an association between the boot (blue arrow) and the bY (green arrow) structure (in contrast to the bY \rightarrow boot region observed in Figure 4G). An additional small structure extending opposite from the boot is also labeled with a yellow arrow. (C) $146 \times 146 \text{ nm}^2$ images demonstrating an association between the boot and the sc structure (in contrast to the boot \rightarrow sc region observed in Figure 4I). The upstream bY structure is labeled with a green asterisk.

branches extending from the 3' end (Figure 4Jvii and viii). This structure also contains a relatively large expansion segment (helix 98/ES39) (Figure 3), which underwent an evolutionary increase from a 15 nt helix in *E. coli* to a 232 nt structure in *H. sapiens* (12). The presence of this expansion segment within one of the branches of the sc region is likely the reason for the moderate GC content of 67.9%. The variable structure on the 5' end appeared to range between three different forms: a highly kinked structure with two major vertices (Figure 4Ai–iv); a ‘mushroom top’ shape at the 5' terminus with another small structure just 3' of it (Figure 4Av and vi); and a highly compact globular structure with an arm extending in the 5' direction (Figure 4Avii and viii). It is well established that the 5' end of the 28S rRNA forms a complex with 5.8S rRNA in eukaryotes. In our experiments, however, the RNA was diluted and heated prior to deposition on the mica for AFM imaging, so it was assumed that the 5.8S is not associated with the molecules in our images. Notably, the 5' end of one molecule (mol275) in Figure 2B and C had a measured volume that was much greater than that of the other molecules, causing the predicted start nucleotide for the subsequent sY domain to be pushed back to nt 649 (compared to the average start nt of 398 nt, Table 1), which may reflect the association of the 5.8S. However, it is an open question whether the 5' structures we observed have any functional relevancy.

Identification of novel small domains (200–300 nt)

Upon examination of the interdomain regions, two of them were observed to contain smaller structures ranging in size from 200–300 nt (Table 1, sY \rightarrow bc and bc \rightarrow bY domains). Two of the domains were located in the sY \rightarrow bc region (nt 1108–1686) (Figure 4C, arrows) and another two were located in the bc \rightarrow bY region (nt 2392–2805) (Figure 4E, arrows).

The two structures located between the sY and bc domains were observed in $n = 16$ (70%) molecules (Figure 4C, arrows). The structure located just 3' of the sY boundary (Figure 4C, red arrows) had a calculated nucleotide range of 1084–1357 (Table 1, sY \rightarrow bc #1). The bulk of this range corresponds to the third major branch of helix25/ES7 (ES7c–h). In the 28S secondary structure, this region is generally depicted as branching from the ES7b branch (Figure 3), so it is unexpected that this structure appears somewhat spatially separated from the sY structure. The morphology of this structure is variable, ranging from a 3-bladed pinwheel morphology (Figure 4Ci–iv), to an elongated structure (Figure 4Cv and vi), to a more compact globular structure (Figure 4Cvii and viii). The second major structure had a calculated nucleotide range of 1422–1687 (Table 1, sY \rightarrow bc #2), placing it within the 5' region of domain II (helices 27–35) (Figure 3). In contrast to the varying morphologies of domain #1, this domain had a consistent elongated morphology (Figure 4C, yellow arrows). In many of the molecules, an additional small region of structure could be observed between the two domains (Figure 4Cv–viii). Although these two structures were usually observed to coex-

ist together, one molecule (Figure 2B, mol368) appeared to have only a single structure corresponding to the nucleotide range for structure #1 and an unfolded region where structure #2 would be expected.

The two structures located between the bc and bY domains were observed in $n = 14$ (61%) molecules (Figure 4E, arrows). In some images, they appeared well separated from each other (Figure 4Ei–iv), whereas in others, they appeared connected by a structured region in line with the main chain (Figure 4Ev and vi), and occasionally the two domains formed a more compact morphology (Figure 4Eviii). The structure located 3' of the bc domain (Figure 4E, orange arrows) had a predicted nucleotide range of 2386–2596 (Table 1, bc→bY #3), corresponding to the 5' region of domain III, whereas the structure located 5' of the bY domain (Figure 4E, green arrows) had a predicted nucleotide range of 2644–2844 (Table 1, bc→bY #4), corresponding to the 3' region of domain III. Of all of the regions in the molecule, this one appears to deviate from the previously reported secondary structures more than any other; in the canonical model, this region should contain a single domain structure corresponding to domain III (Figure 3 and Supplementary Figure S1). Domain III of bacterial rRNA has even been reported to fold autonomously into its native conformation (24). It may be possible that the formation of this domain may be highly dependent on the presence of Mg^{2+} ions. Although these two independent structures usually coexisted in the observed molecules, in one molecule (Figure 2B, mol228) only structure #4 could be discerned, while structure #3 appeared to associate with the adjacent bc structure. In addition, in another molecule (Figures 2B, mol 322 and Figure 4Fviii), both domains appeared to associate with the adjacent bY structure.

Regions corresponding to domain 0 and the PTC are often unstructured

Further examination of the interdomain regions also revealed that two of them appear quite variable in the region between the bY and boot (nt 3735–3899) (Figure 4G), and the region between the boot and sc domains (Figure 4I). In both of these regions, the morphology ranged from relatively unstructured (Figure 4Gi–ii and Ii–ii), to having a single small structure (Figure 4Giii–vi and 4Iiii–v), or to having a single larger structure in the case of the bY→boot region (Figure 4Gvii–viii), or multiple small structures in the case of the boot→sc regions (Figure 4Ivi–viii). Interestingly, both of these regions also have the lowest GC content of any of the other regions (53.3% for bY→boot and 47.9% for boot→sc) (Table 3) and are widely considered to have evolved very early in the course of ribosomal evolution (12).

According to secondary structural models, the bulk of the nucleotides in the bY→boot region (nt 3735–3899) should contribute to the formation of domain 0 (Supplementary Figure S1 and Figure 3), so the observation that this region lacks well-defined, conserved structures is perhaps unsurprising since this region should later associate to form long range contacts with regions of the molecule located between the sY and bc structures to form the final tertiary structure. Predicted nucleotides for the boot→sc region (nt 4291–4597) should form the bulk of the PTC (Figure 3). It is

interesting that regions containing the expansion segments have very well-defined conserved structures in our images, yet some of the most functionally relevant structures of the ribosome do not readily form in these conditions.

Molecular compaction is caused by association of specific domains

As mentioned earlier, the distribution of length values in Figure 1C suggests that the molecules in these images have varying degrees of molecular compaction. Following the identification and morphological characterization of the individual domain structures, the conformational changes leading to this compaction could be further investigated. The main reason for the shortened morphology of some of the molecules appeared to be a higher order structure extending away from the main chain formed by an association of the bc structure with the four flanking smaller domains from the sY→bc and bc→bY regions (Figure 5A). This structure can be observed between the sY and bY structures, which are brought into noticeably closer proximity by the formation of this higher order structure. Molecules with this higher-order structure in Figure 5A had main chain lengths of 220, 223 and 228 nm, which are notably some of the shortest molecules in the dataset (Figure 1C, main chain). It is interesting to speculate that these structures could represent the partial formation of domain 0 (Supplementary Figure S1). All three molecules in Figure 5A have a noticeable loop structure, which may demonstrate the formation of long-range contacts between regions of the RNA molecule. Additional interdomain interactions were also observed between the bY-boot (Figure 5B) and the boot-sc (Figure 5C) structures, in contrast to the observation of the regions of variable structure which usually separate these domains (Figure 4G and I). Interestingly, the two molecules that exhibited the boot-sc structure also displayed the higher order bc structure (Figure 5A). The structure formed by association between the bY-boot structures had a morphology where the base of each domain appeared associated so as to orient the two domains perpendicularly to one another, with another small structure extending opposite from the boot (Figure 5B, yellow arrows). The two molecules that exhibited the bY-boot morphology had main chain lengths of 290 and 291 nm, which are moderately shortened compared to the other molecules in the data set (Figure 1C, main chain).

DISCUSSION

In a previous study, we reported on sample preparation procedures for observing secondary and tertiary structures of a variety of long single-stranded RNA molecules (>1 kb) by varying the Mg^{2+} concentration of the buffer (6,7). In this study, we demonstrated a high-throughput analytical technique to extract information about the secondary structural domains of the partially unfolded molecules in low salt buffer. This method enables the identification of domains located along the trajectory of the molecule and the estimation of nucleotides contained in each domain based on their relative volume. Furthermore, as proof-of-concept, we applied this technique to the 5 kb human 28S rRNA, the

largest ssRNA molecule with well-documented 3D structures (10–12). This analysis showed that many of the secondary structural domains that have previously been reported for this molecule remain intact in these molecules despite the removal of proteins and Mg^{2+} ions. Thus, since it is often difficult to obtain 3D structures for single-stranded RNA molecules of this size, our methods provide a viable alternative for high-throughput screening of structural domains present in a variety of biologically important RNA molecules.

Correlations with existing structural models

By mapping the nucleotide ranges predicted by our algorithms onto previously reported secondary structural models of 28S rRNA, we saw a striking correlation between the general shape and length of the branches in our images to the branches modeled in previous studies, even to the extent of being able to identify a sharp bend corresponding to a kink-turn motif which has been previously reported in 23S rRNA (22,23). We observed individual domain structures corresponding to five of the canonical domains: the 3' region of canonical domain I, the 3' region of canonical domain II, most of canonical domain IV, the 5' region of canonical domain V, and most of canonical domain VI. Overall, the structural detail we have been able to extract using this method seems to exceed that of previous electron microscopy studies (25,26). We were also able to observe flexible regions of the molecule which are difficult to observe in the 3D structures. For example, X-ray and cryo-EM structures have been unable to resolve ES27 due to its flexibility (27). However, with our single-molecule approach, ES27 can be clearly observed in all molecules as the big Y structure.

Variations from structural models

Although there is a high degree of correlation between the domain architecture observed in our images and structural models of ribosomal RNA from X-ray and Cryo-EM studies, some subtle deviations were observed. The main variations include the identification of four previously uncharacterized smaller domain regions that each contain ~200 nt flanking either side of the big claw structure (Tables 1 and 2, structures #1–4), and the lack of formation of the PTC and domain 0. Many of these deviations can likely be ascribed to the absence of Mg^{2+} ions from the sample preparation buffer. In the crystal structures of *Halobacterium marismortui* and *Thermus thermophilus* rRNA, there are four Mg^{2+} microcluster motifs, referred to as D1–D4 (28). Three of these surround the PTC, and the other is located within canonical domain III near the end of the polypeptide exit tunnel. In our analysis, the one canonical domain that did not have a single domain structure in our images was domain III, which was instead observed as two individual domains (Figure 4E; Tables 1 and 2, structures #3–4) each comprised of about 200 nt. Since Mg^{2+} microcluster D3 is reported to occur within this domain (28), this could indicate that the stability of this domain is more dependent on the presence of Mg^{2+} ions than the other domains. In contrast, in our images, the nucleotide range which makes up

the PTC (Figure 4I) appears quite variable in the amount of structure observed, suggesting that the other three microclusters (D1, D2 and D4) are critical for the formation of stable structure in this region. In addition, microcluster D2 bridges nucleotides of the PTC with nucleotides in canonical domain II, which also corresponds to one of the smaller domains (Tables 1 and 2, structure #2) recognized in our images. These results point to the differential dependence of the formation of some RNA domain structures on the presence of Mg^{2+} ions.

It is interesting that one Mg^{2+} -dependent region (Tables 1 and 2, structures #3–4) still contained domain structures which were reproducible in most of the molecules, while another one (PTC, Figure 4I) did not show the same reproducibility of structure. Although it is possible that these structures could represent a misfolded conformation, it is tempting to speculate that the new structures we observed could have a biological role. Numerous RNA molecules have been reported to have metastable structures which have a functional role, such as influencing viral infectivity, controlling the translation levels of proteins, or RNA editing (29–31). Given the complex multicompartiment processing and assembly process (32), the highly coordinated activities during the various steps of translation (33,34), and the various regulatory factors that might alter ribosome activity or localization (35), it is tempting to speculate that there could be some occasion for these structures to exist *in vivo*.

Conformational variations

One of the strengths of our single-molecule AFM approach is the ability to observe the molecules in a variety of conformations which may coexist within a population of molecules. Some of the interesting conformational variations observed in this study include an alternative conformation of the bY structure with one of the branches shortened (Figure 4F), a varying stalk length of the boot structure (Figure 4H) and multiple structural variations of domains on the 5' and 3' ends (Figure 4A and J). Interestingly, both the bY and the boot structures have been reported to be particularly dynamic. The bY, corresponding to ES27 has been reported to have two main conformations in which branch ES27a can be oriented either toward the L1 stalk or toward the tunnel exit (21), although only the structure oriented toward the L1 stalk has been observed so far in human ribosomes (27). Additionally, the boot structure corresponds to the flexible three way junction comprising the L1 stalk of the ribosome. The movements of the L1 stalk coordinate the release of deacylated tRNA from the ribosome (36,37). It is interesting to speculate that the conformational differences observed in our images may be related to the conformational dynamics reported from these other studies, suggesting that the propensity for these dynamic shifts may be present in the RNA molecule itself.

In addition to conformational differences within domains, associations between adjacent domains were also observed in the molecules, generally leading to molecular compaction in these subsets of molecules (Figure 5). Specifically, two of the conformations involve an association of the boot structure, corresponding to canonical domain V with either the bY, corresponding to canonical domain IV (Fig-

ure 5B) or with the sc, corresponding to canonical domain VI (Figure 5C). Generally, the PTC should be located between the boot and sc structures, which suggests that these molecules may represent formation of the PTC. It is also interesting that the molecules containing this structure also usually have an the extended bc structure (Figure 5A) which is made up of regions spanning domains II and III of the 28S structure. However, the alternative association between domains IV and V seems somewhat unexpected, as the region between the bY and boot structures is generally interceded by domain 0 (Supplementary Figure S1).

The use of AFM for RNA structural studies

The correlation between our structures and existing models demonstrates that this method is capable of extracting biologically relevant information about the secondary structural arrangement of ssRNA molecules. This suggests that this method can be used as a viable method for identification of structural domains within large RNA molecules such as mRNA molecules or viral RNA genomes, augmenting existing methods such as computational predictions (38) or chemical/enzymatic mapping (39). Also, since the domains identified by this method are likely to represent stable elements of the RNA, the ability to approximate the nucleotides comprising each domain can provide researchers with regions which can be further targeted for structural determination by techniques which are not amenable to working with conformationally diverse molecules. Additionally, the ability to visualize RNA in a variety of coexisting conformations makes it an excellent technique for identification of metastable intermediates which may not constitute the predominant conformation of the molecule. Additionally, automated data analysis algorithms allow for this technique to be developed into a high-throughput approach. Continued development could result in a molecular pattern recognition algorithm capable of sorting the molecules based on their varying conformations. Currently, our method is capable of performing RNA imaging experiments over the course of 1–2 days once a relatively low amount of RNA has been purified (~1–5 ng per sample). Then, current data analysis procedures generally require ~1–2 min/molecule. However, in the future, we hope to further automate later steps of the data analysis procedure and make improvements to existing algorithms in order to increase the speed of data analysis, increase the accuracy of nucleotide predictions and domains recognition and increase the structural information which can be extracted from these molecules.

Although much of the RNA structure correlates well with that of predicted models, the differences we observe suggest that additional factors, such as the addition of ions or proteins, can affect the structure of the molecule. Previously, we have demonstrated the ability to image RNA molecules in a variety of Mg^{2+} concentrations (6). However, the structures imaged in the absence of Mg^{2+} have extended conformations with individual domains readily visible along the trajectory of the molecules. For that reason, we have used these conditions to analyze the domain structure of these molecules as structural information can be more readily extracted. In the future, development of more advanced 3D-based algorithms could enable analysis of more

compact structures which can allow for piecing together the folding/assembly pathways of molecules when increasing concentrations of Mg^{2+} are used. In addition, the ability to image molecules with a high degree of conformational flexibility could allow us to investigate the effects of adding only one or a few proteins to the RNA at a time. In this way, ribosomal properties which are intrinsic to a particular domain of the RNA itself, as well as the effects of individual factors such as ions or proteins can be investigated in a stepwise fashion, so as to piece together the functional role that each component plays in the ribosome. These studies could further be expanded to study the dynamics of these interactions, or details of ribosomal folding/assembly pathways with the use of high speed Atomic Force Microscopy, in which molecules can be visualized in buffer solution at rates of ~1–2 frames/s (40,41).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Thanks to Prof. Osamu Chisaka for helpful discussion.

FUNDING

Grant-in-Aid for Scientific Research on Innovative Areas ‘Molecular basis of host cell competency in virus infection’ from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) Japan [#24115003 to K.T.]; Grant-in-aid for Japan Society for the Promotion of Science (JSPS) Fellowship [#P16390 to J.G.]. Funding for open access charge: Japan Society for the Promotion of Science (JSPS).

Conflict of interest statement. None declared.

REFERENCES

- Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
- Nissen,P., Hansen,J., Ban,N., Moore,P.B. and Steitz,T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289**, 920–930.
- Wimberly,B.T., Brodersen,D.E., Clemons,W.M. Jr, Morgan-Warren,R.J., Carter,A.P., Vornrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
- Schluzenzen,F., Tocilj,A., Zarivach,R., Harms,J., Gluehmann,M., Janell,D., Bashan,A., Bartels,H., Agmon,I., Franceschi,F. *et al.* (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, **102**, 615–623.
- Ehrenberg,M. (2009) *The Nobel Prize in Chemistry 2009-Advanced Information: Structure and Function of the Ribosome*. The Royal Swedish Academy of Sciences, Stockholm.
- Gilmore,J.L., Aizaki,H., Yoshida,A., Deguchi,K., Kumeta,M., Junghof,J., Wakita,T. and Takeyasu,K. (2014) Nanoimaging of ssRNA: genome architecture of the Hepatitis C Virus revealed by atomic force microscopy. *J. Nanomed. Nanotechnol.*, **S5**, 1–7.
- Gilmore,J.L. and Takeyasu,K. (2016) Commentary: can AFM reveal global viral RNA structure? *J. Pharm. Nanotechnol.*, **4**, 11–13.
- Noller,H.F., Kop,J., Wheaton,V., Brosius,J., Gutell,R.R., Kopylov,A.M., Dohme,F., Herr,W., Stahl,D.A., Gupta,R. *et al.* (1981) Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res.*, **9**, 6167–6189.

9. Petrov,A.S., Bernier,C.R., Hershkovits,E., Xue,Y., Waterbury,C.C., Hsiao,C., Stepanov,V.G., Gaucher,E.A., Grover,M.A., Harvey,S.C. *et al.* (2013) Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Res.*, **41**, 7522–7535.
10. Bernier,C.R., Petrov,A.S., Waterbury,C.C., Jett,J., Li,F., Freil,L.E., Xiong,X., Wang,L., Migliozi,B.L., Hershkovits,E. *et al.* (2014) RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss*, **169**, 195–207.
11. Petrov,A.S., Bernier,C.R., Gulen,B., Waterbury,C.C., Hershkovits,E., Hsiao,C., Harvey,S.C., Hud,N. V., Fox,G.E., Wartell,R.M. *et al.* (2014) Secondary structures of rRNAs from all three domains of life. *PLoS One*, **9**, 1–6.
12. Petrov,A.S., Bernier,C.R., Hsiao,C., Norris,A.M., Kovacs,N.A., Waterbury,C.C., Stepanov,V.G., Harvey,S.C., Fox,G.E., Wartell,R.M. *et al.* (2014) Evolution of the ribosome at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 10251–10256.
13. Boom,R., Sol,C.J., Salimans,M.M., Jansen,C.L. and Van Der Noordaa,J. (1990) Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.*, **28**, 495–503.
14. Damodaran,S. and Kinsella,J.E. (1984) Dissociation of yeast 80S ribosomes by chaotropic salts. *Biochim. Biophys. Acta*, **801**, 416–423.
15. Brion,P. and Westhof,E. (1997) Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137.
16. Nečas,D. and Klapetek,P. (2012) Gwyddion: an open-source software for SPM data analysis. *Cent. Eur. J. Phys.*, **10**, 181–188.
17. MathWorks (2014) *MATLAB and image processing toolbox release R2014b*. The Mathworks, Inc., Natick.
18. O'Haver,T. (2016) *Software for peak finding and measurement: findpeaks and iPeak*. In: A Pragmatic Introduction to Signal Processing. Createspace Independent Publishing Platform. pp. 74–84.
19. Hu,B., Jin,J., Guo,A.Y., Zhang,H., Luo,J. and Gao,G. (2015) GSDDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, **31**, 1296–1297.
20. Hancock,J.M. and Dover,G.A. (1988) Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. *Mol. Biol. Evol.*, **5**, 377–391.
21. Armache,J.-P., Jarasch,A., Anger,A.M., Villa,E., Becker,T., Bhushan,S., Jossinet,F., Habeck,M., Dindar,G., Franckenberg,S. *et al.* (2010) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 19748–19753.
22. Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
23. Rázga,F., Koča,J., Mokdad,A. and Šponer,J. (2007) Elastic properties of ribosomal RNA building blocks: molecular dynamics of the GTPase-associated center rRNA. *Nucleic Acids Res.*, **35**, 4007–4017.
24. Athavale,S.S., Gossett,J.J., Hsiao,C., Bowman,J.C., O'Neill,E., Hershkovitz,E., Preeprem,T., Hud,N. V., Wartell,R.M., Harvey,S.C. *et al.* (2012) Domain III of the T. thermophilus 23S rRNA folds independently to a near-native state. *RNA*, **18**, 752–758.
25. Wellauer,P.K. and Dawid,I.B. (1973) Secondary structure maps of RNA: processing of HeLa ribosomal RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 2827–2831.
26. Oostergetel,G.T., Wall,J.S., Hainfeld,J.F. and Boublik,M. (1985) Quantitative structural analysis of eukaryotic ribosomal RNA by scanning transmission electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 5598–5602.
27. Anger,A.M., Armache,J.P., Berninghausen,O., Habeck,M., Subklewe,M., Wilson,D.N. and Beckmann,R. (2013) Structures of the human and Drosophila 80S ribosome. *Nature*, **497**, 80–85.
28. Hsiao,C. and Williams,L.D. (2009) A recurrent magnesium-binding motif provides a framework for the ribosomal peptidyl transferase center. *Nucleic Acids Res.*, **37**, 3134–3142.
29. Loss,P., Schmitz,M., Steger,G. and Riesner,D. (1991) Formation of a thermodynamically metastable structure containing hairpin II is critical for infectivity of potato spindle tuber viroid RNA. *EMBO J.*, **10**, 719–727.
30. Møller-Jensen,J., Franch,T. and Gerdes,K. (2001) Temporal Translational Control by a Metastable RNA Structure. *J. Biol. Chem.*, **276**, 35707–35713.
31. Linnstaedt,S.D., Kasprzak,W.K., Shapiro,B.A. and Casey,J.L. (2009) The fraction of RNA that folds into the correct branched secondary structure determines hepatitis delta virus type 3 RNA editing levels. *RNA*, **15**, 1177–1187.
32. Henras,A.K., Plisson-Chastang,C., O'Donohue,M.F., Chakraborty,A. and Gleizes,P.E. (2015) An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA*, **6**, 225–242.
33. Lodmell,J.S. and Hennesly,S.P. (2003) Conformational dynamics within the ribosome. In: Lapointe,J and Brakier-Gingras,L (eds). *Translation Mechanisms*. Kluwer Academic/Plenum Publishers, Georgetown. pp. 264–279.
34. Fraser,C.S. and Doudna,J.A. (2007) Quantitative studies of ribosome conformational dynamics. *Q. Rev. Biophys.*, **40**, 163–189.
35. Xue,S. and Barna,M. (2012) Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.*, **13**, 355–369.
36. Cornish,P.V., Ermolenko,D.N., Staple,D.W., Hoang,L., Hickerson,R.P., Noller,H.F. and Ha.T. (2009) Following movement of the L1 stalk between three functional states in single ribosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 2571–2576.
37. Réblová,K., Šponer,J. and Lankaš,F. (2012) Structure and mechanical properties of the ribosomal L1 stalk three-way junction. *Nucleic Acids Res.*, **40**, 6290–6303.
38. Moss,W.N., Priore,S.F. and Turner,D.H. (2011) Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA*, **17**, 991–1011.
39. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W., Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
40. Gilmore,J.L., Suzuki,Y., Tamulaitis,G., Siksnyš,V., Takeyasu,K. and Lyubchenko,Y.L. (2009) Single-molecule dynamics of the DNA-EcoRII protein complexes revealed with high-speed Atomic Force Microscopy. *Biochemistry*, **48**, 10492–10498.
41. Suzuki,Y., Gilmore,J.L., Yoshimura,S.H., Henderson,R.M., Lyubchenko,Y.L. and Takeyasu,K. (2011) Visual analysis of concerted cleavage by type IIF restriction enzyme SfiI in subsecond time region. *Biophys. J.*, **101**, 2992–2998.