Visualizing the functional 3D shape and topography of long noncoding RNAs by single-particle atomic force microscopy and in-solution hydrodynamic techniques

Tina Uroda1,4, Isabel Chillón1, Paolo Annibale2, Jean-Marie Teulon3, Ombeline Pessey1, Manikandan Karuppasamy1, Jean-Luc Pellequer3 and Marco Marcia1✉

Long noncoding RNAs (lncRNAs) are recently discovered transcripts that regulate vital cellular processes, such as cellular differentiation and DNA replication, and are crucially connected to diseases. Although the 3D structures of lncRNAs are key determinants of their function, the unprecedented molecular complexity of lncRNAs has so far precluded their 3D structural characterization at high resolution. It is thus paramount to develop novel approaches for biochemical and biophysical characterization of these challenging targets. Here, we present a protocol that integrates non-denaturing lncRNA purification with in-solution hydrodynamic analysis and single-particle atomic force microscopy (AFM) imaging to produce highly homogeneous lncRNA preparations and visualize their 3D topology at ~15-Å resolution. Our protocol is suitable for imaging lncRNAs in biologically active conformations and for measuring structural defects of functionally inactive mutants that have been identified by cell-based functional assays. Once optimized for the specific target lncRNA of choice, our protocol leads from cloning to AFM imaging within 3–4 weeks and can be implemented using state-of-the-art biochemical and biophysical instrumentation by trained researchers familiar with RNA handling and supported by AFM and small-angle X-ray scattering (SAXS) experts.

Introduction

Long noncoding RNAs (lncRNAs) are non-protein-coding transcripts that regulate fundamental cellular processes, such as cell differentiation and replication1, and are directly connected to severe diseases, such as cancer and neurological or cardiovascular defects2. LncRNAs are thus of fundamental interest in a wide range of research areas, including epigenetics, neurobiology, oncology, plant biology and infection biology. However, the molecular mechanisms of lncRNAs are poorly characterized to date. For a quantitative and comprehensive mechanistic understanding of lncRNAs, characterizing their three-dimensional (3D) structures is now of utmost importance, because 3D structure has been shown to play a crucial role in lncRNA biological functions3,4. Specifically, many lncRNAs scaffold nuclear proteins, such as chromatin remodeling enzymes and transcription factors, or shape subnuclear bodies, such as speckles or paraspeckles5–7. Thus, it can be expected that lncRNA tertiary structures may guide specific and selective protein binding or ensure correct chromatin targeting and, consequently, efficient gene expression regulation1. Indeed, long-range tertiary interactions have already been identified in the lncRNAs RepA5, XIST6, and MEG310. However, high-resolution structural studies on lncRNAs are limited to the characterization of extremely small domains, that is, a 14-nt-long stem loop of the ~17,000-nt-long XIST11 and a 76-nt-long triple helix motif of the ~8,400-nt-long MALAT112. Instead, 3D structures have never been determined for any full-length lncRNA, because the size and complexity of these transcripts—which generally span 1,000–10,000 nucleotides (nt)—present unique and unprecedented challenges for biochemical and biophysical characterization4.
Development of the protocol

In our lab, we recently developed an approach that enabled us to visualize the structural organization of full-length lncRNAs. For our studies, we have integrated small-angle X-ray scattering (SAXS) and single-particle AFM imaging with a detailed functional characterization of the chemically probed lncRNA secondary structures. Optimization of our non-denaturing lncRNA purification pipeline with an enlarged set of transcription buffers and quality control assays (Bioanalyzer, electrophoresis, and static and dynamic light scattering (DLS)) was essential to obtain highly homogeneous conformations of our target. The identification of functional tertiary interactions via robust phenotypic assays provided further useful guidance for biophysical studies and, crucially, a rationale for designing mutants with perturbed structural architectures that served as invaluable test samples with which to benchmark our imaging method. Finally, specific screening of AFM surfaces, sample deposition procedures, and data processing methods—as we describe in detail in our protocol—enabled us to adapt AFM imaging to lncRNAs, which, to our knowledge, have never before been studied with such a method. Considering the AFM conditions that we have used, we can achieve a resolution of ~1.5 nm, which enables us to visualize the global topography of our targets, including their homogeneity, size, and compactness. We can clearly distinguish between different folding states of a wild-type lncRNA (unfolded, partially folded, and folded) and we can capture topological differences (i.e., differences in 3D shape) between lncRNA mutants that possess wild type versus disrupted long-range tertiary contacts.

In our work, we have focused on the human lncRNA called ‘maternally expressed gene 3’ (MEG3), which is an lncRNA that promotes neuronal differentiation and stimulates p53, preventing neurodevelopmental syndromes and intracranial tumors. Using our approach, we could specifically prove that evolutionarily conserved intramolecular long-range tertiary structure interactions called pseudoknots or ‘kissing loops’, which are required for MEG3-dependent p53 stimulation, are also strictly required for MEG3 folding. Our work established that physiologically relevant long-range RNA tertiary structure interactions guide the biological function of lncRNAs. In this protocol, we provide a detailed workflow that can be broadly applied to characterize the 3D topology of any lncRNA of interest and will thus enable the characterization of important molecular properties of many of these medically relevant targets.

Overview of the procedure

In this protocol, we describe how to study lncRNAs with a spectrum of complementary biophysical methods in solution, how to image lncRNAs with AFM, and how to correlate their structural features to functionally relevant conformations.

A conceptual overview of the procedure is shown in Fig. 1. First, the lncRNA of interest is produced by in vitro transcription (Steps 1–19). The folding conditions of the lncRNA are then determined by analytical ultracentrifugation (AUC) (Steps 20–38), and its homogeneity is characterized by gel electrophoresis (Steps 39A(i–viii)) and by dynamic and static light scattering (DLS and multi-angle laser light scattering (MALLS), Step 39B(i–xiii) and 39C(i–xy), respectively). Finally, the lncRNA is imaged in solution by SAXS (Steps 40–64) and on mica by AFM in dry mode (Steps 65–100). Importantly, we explain how to visualize the SAXS/AFM structure of the target in different folding conditions, including in compact, folded states at near-physiological concentrations of mono and divalent cations (~1–10 mM magnesium chloride, ~100–150 mM potassium chloride), which closely mimic the intracellular environment. We specifically describe how to accurately control folding by titrating different ionic conditions, particularly different concentrations of magnesium ions, which are essential lncRNA structural scaffolds.

The strengths of our protocol are that (i) we integrate hydrodynamic and AFM analyses, ensuring that the molecular shape and size adopted by the target on the AFM support is compatible with its dimensions in solution; (ii) we use known structured and unstructured RNAs as controls for benchmarking the protocol; and (iii) we can visualize and compare functionally active and inactive constructs of the same lncRNA target to correlate structural changes with specific functional states. The protocol is compatible with and complementary to chemical or enzymatic probing approaches, for example, selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE), as previously reported, which is a robust technique for mapping the secondary structural architecture of the target RNA. We strongly recommend that any data obtained using the approach described in this protocol be backed up and validated by functional experiments in cells. The relevant functional assays are, however, target specific and will therefore not be described in this protocol.
Sample production

Steps 1 and 2 & 6–12: DNA prep
- Cloning
- Linearization
- Purification
Pause: store linearized DNA at −20 °C

Steps 3–5 & 13–15: transcription
- Buffer screening
  → Checkpoint, Steps 10–12:
    Choose transcription buffer
  Pause: scale up later
- Scale up transcription

Steps 16–19: purification & verification
- DNA/protein digestion
- Rebuffering
- SEC
  Setup: pack column (\(t = 12 \text{ h}\))
  3 h

Steps 20–30: AUC (data acquisition)
- SAXS
  Steps 40–47: data acquisition
  Setup, Step 38C(i,ii)
  pack column (\(t = 12 \text{ h}\))
  Data acquisition
  Pause: analyze data later
- AFM imaging
  Steps 69–86: AFM (data acquisition)
  Setup, Steps 69–73:
  Prepare mica, mount cantilever (\(t = 0.5 \text{ h}\))
  Caution:
  Isolate from vibrations
  Calibrate piezo scanner
  Pause: process images later
- Data processing

Steps 31–38: AUC (data analysis)
- Checking for monodispersity
  Step 39A: electrophoresis
  Step 39B: DLS
  Step 39C: SEC-MALLS
  Setup, 39C(i,ii)
  pack column (\(t = 12 \text{ h}\))
  Data processing
  Pause: analyze data later

Steps 48–64: data processing
-检查折叠
  Checkpoint, Step 100: Specific folding?
- Suitable for ab initio shape determination?

Steps 50–56: AUC (data acquisition)
- Guinier plot
- PSD plot
- Checking for monodispersity
  Checkpoint, Step 36: choose Mg concentration

Steps 57–63: AUC (data analysis)
- Steps 65–68: AFM (sample preparation)
  Setup, Steps 69–73:
  Prepare mica, mount cantilever (\(t = 0.5 \text{ h}\))
  Caution:
  Isolate from vibrations
  Calibrate piezo scanner
  Pause: process images later
- Steps 87–94: AFM (image processing)
  Pause: analyze images later
- Steps 95–100: AFM (image analysis)
  Checkpoint, Step 19: bioanalyzer
  → Checkpoint, Steps 10–12: DNA prep
  \(t = 3 \text{ d}\)

Steps 101–120: purification & verification
- DNA/protein digestion
- Rebuffering
- SEC
  Setup: pack column (\(t = 12 \text{ h}\))
  3 h

Steps 121–127: AUC (data acquisition)
- SAXS
  Steps 40–47: data acquisition
  Setup, Step 38C(i,ii)
  pack column (\(t = 12 \text{ h}\))
  Data acquisition
  Pause: analyze data later
- AFM imaging
  Steps 69–86: AFM (data acquisition)
  Setup, Steps 69–73:
  Prepare mica, mount cantilever (\(t = 0.5 \text{ h}\))
  Caution:
  Isolate from vibrations
  Calibrate piezo scanner
  Pause: process images later
- Data processing

Steps 128–134: AUC (data analysis)
- Checking for monodispersity
  Step 39A: electrophoresis
  Step 39B: DLS
  Step 39C: SEC-MALLS
  Setup, 39C(i,ii)
  pack column (\(t = 12 \text{ h}\))
  Data processing
  Pause: analyze data later

Steps 135–141: AUC (data analysis)
- Steps 65–68: AFM (sample preparation)
  Setup, Steps 69–73:
  Prepare mica, mount cantilever (\(t = 0.5 \text{ h}\))
  Caution:
  Isolate from vibrations
  Calibrate piezo scanner
  Pause: process images later
- Steps 87–94: AFM (image processing)
  Pause: analyze images later
- Steps 95–100: AFM (image analysis)
  Checkpoint, Step 19: bioanalyzer
  → Checkpoint, Steps 10–12: DNA prep
  \(t = 3 \text{ d}\)

Steps 101–120: purification & verification
- DNA/protein digestion
- Rebuffering
- SEC
  Setup: pack column (\(t = 12 \text{ h}\))
  3 h

Steps 121–127: AUC (data acquisition)
- SAXS
  Steps 40–47: data acquisition
  Setup, Step 38C(i,ii)
  pack column (\(t = 12 \text{ h}\))
  Data acquisition
  Pause: analyze data later
- AFM imaging
  Steps 69–86: AFM (data acquisition)
  Setup, Steps 69–73:
  Prepare mica, mount cantilever (\(t = 0.5 \text{ h}\))
  Caution:
  Isolate from vibrations
  Calibrate piezo scanner
  Pause: process images later
- Data processing

Steps 128–134: AUC (data analysis)
- Checking for monodispersity
  Step 39A: electrophoresis
  Step 39B: DLS
  Step 39C: SEC-MALLS
  Setup, 39C(i,ii)
  pack column (\(t = 12 \text{ h}\))
  Data processing
  Pause: analyze data later

Steps 135–141: AUC (data analysis)
- Steps 65–68: AFM (sample preparation)
  Setup, Steps 69–73:
  Prepare mica, mount cantilever (\(t = 0.5 \text{ h}\))
  Caution:
  Isolate from vibrations
  Calibrate piezo scanner
  Pause: process images later
- Steps 87–94: AFM (image processing)
  Pause: analyze images later
- Steps 95–100: AFM (image analysis)
  Checkpoint, Step 19: bioanalyzer
  → Checkpoint, Steps 10–12: DNA prep
  \(t = 3 \text{ d}\)
Applications of the method

Our protocol is applicable to a vast number of newly discovered biological targets because it has recently been discovered that lncRNAs are numerous, especially in mammals. For instance, it is estimated that humans may express as many as 30,000 lncRNAs versus ~20,000 protein-coding genes. Moreover, recent studies on the secondary structure architecture of functional lncRNAs by chemical and enzymatic probing revealed that these transcripts are as complex as rRNAs or ribozymes and that their structures are evolutionarily conserved.

Furthermore, our protocol enables the visualization of functional and non-functional lncRNA conformations and folding states, which is of immediate interest to a broad spectrum of researchers, including epigeneticists, developmental biologists, plant biologists, and oncologists. Indeed, structured lncRNAs participate in diverse and fundamental biological processes. For instance, HOTAIR scaffolds Polycomb repressive complexes in trans for epigenetic differentiation of skin tissues; MEG3 participates in cell cycle regulation and tumor suppression; COOLAIR controls flowering and vernalization in plants; NEAT1 and lincRNA-p21 shape the expression landscape in response to environmental conditions modulating the stress response in paraspeckles; Braveheart regulates cardiomyocyte differentiation; and RepA, roX, and XIST ensure gene-dosage compensation during sex determination.

The profound medical implications of lncRNAs raise the prospects of gaining valuable insights from the applications of our method to the field of medicinal chemistry. Large investments are currently being made by private and academic labs to identify compounds designed to modulate RNA function for potential clinical use. In this context, our protocol could be used to screen the effects of currently being made by private and academic labs to identify compounds designed to modulate RNA handling—including scientists at early stages of their careers, that is PhD students, if appropriately supervised by experts when operating costly equipment—such as AUCs, SAXS beamlines, and AFM microscopes; and (iii) being implementable with affordable investments in house or via user-oriented facilities, such as synchrotron beamlines. The protocol is also compatible with parallel characterization of multiple lncRNA targets.

Comparison with other methods

Our protocol has the advantages of (i) integrating widely accessible biochemical and biophysical techniques; (ii) making use of instrumentation that can be operated by any researcher familiar with RNA handling—including scientists at early stages of their careers, that is PhD students, if appropriately supervised by experts when operating costly equipment—such as AUCs, SAXS beamlines, and AFM microscopes; and (iii) being implementable with affordable investments in house or via user-oriented facilities, such as synchrotron beamlines. The protocol is also compatible with parallel characterization of multiple lncRNA targets.

Moreover, our protocol is unique in the detailed description of the biochemical and biophysical pipeline for lncRNA production and for the experimental assessment of their purity and homogeneity. Complementarily to an approach recently used to image the lncRNA Braveheart by SAXS, our protocol reveals how integrated structural biology can powerfully enable the characterization of novel classes of large and difficult-to-handle biological macromolecules such as lncRNAs. In this respect, one of our protocol’s distinctive feature is the implementation of single-particle imaging by AFM, in addition to in-solution hydrodynamic analysis by SAXS, to visualize lncRNA functional conformations.
Single-particle AFM imaging has previously been used to visualize short nucleic acids and their ribonucleoprotein complexes, revealing their mechanistic details, interaction interfaces and conformational dynamics. For instance, AFM imaging has enabled morphological analysis of short structured motifs of genomic RNA of the turnip yellow mosaic virus (TYMV), internal ribosome entry site (IRES) elements of genomic RNA of hepatitis C virus (HCV), and de novo–designed small RNA nanostructures with potential biomedical applications. However, the length and complexity of lncRNAs present unique challenges with respect to such short, structured RNA motifs. The specific molecular properties of IncRNAs require completely different production, stabilization, and handling of the targets and have specific implications for data processing, as we describe in detail in our protocol.

Fig. 2 | Processing pipeline for AFM images. a–e. AFM scans are processed in Gwyddion with the following pipeline: raw images (a, Steps 85–86) are leveled by plane fitting (b, first fitting order, Step 88), followed by removal of polynomial background (c, second fitting order, Step 89), row alignment and horizontal scar correction (d, third fitting order, Step 90), and adjustment of the color scale (e, Step 92). e,f. Images should then be reported as either 2D (e) or 3D (f) representations (Step 94). xy scale bars in the top right corner of each panel = 400 nm; the z color scale bar is shown on the right of each panel.
**Fig. 3 | Power spectral density analysis (Steps 95–99).**

a. Graphical depiction of how characteristic length scales relate to the spatial frequency of specific features of the PSD. **a.** Schematic of a particle, characterized by an overall molecular size $X$ (corresponding to spatial frequency $\xi$) and by an intramolecular feature $Z$ (corresponding to spatial frequency $\psi$). **b.** The corresponding PSD of such a particle. **c, d.** PSD plots from simulated 60-nm (c) and 10-nm (d) particles. Convolution with a tip with a 10-nm curvature radius is simulated. Each plot is the average of 10 particles; error bars = SEM. Representative simulated particles are depicted in the bottom left inset in each plot. **e.** Selection of an ROI (yellow box) around an IncRNA particle in a prototypical AFM acquisition (Step 95). **f.** Detail of the IncRNA molecule within the ROI. xy scale bar, 100 nm. The z color scale bar applies to both e and f. **g.** PSD of the particle displayed in f, taken along the fast-scanning axis (x) (Step 96). **h.** Average PSD from 100 such ROIs, displayed with superimposed fits to the low-frequency plateau and the two regions of the $f^{-\alpha}$ decay (Steps 97–100). Intercepts at 25 nm and 80 nm relate to intermolecular domain size and overall molecule diameter (2D projected), respectively. Error bars = SEM.
Electron microscopy (EM) is an alternative approach to AFM for single-particle imaging of biological macromolecules—even at high resolution (i.e., 2–5 Å). This technique has recently gained momentum, especially since the advent of single-electron detectors. However, for still largely unclear reasons, EM has so far not been used successfully in the characterization of pure RNA molecules; that is, EM structures exist only for RNAs in complex with protein partners. We have attempted visualization of the lncRNA MEG3 by both negative-staining EM and cryo-EM, but we encountered technical challenges that we have not yet overcome. For instance, we have imaged MEG3 at increasing concentrations of magnesium in the same range as for AFM imaging (10–25 mM Mg\(^{2+}\)). By negative-staining EM, we could visualize individual particles on the grids, but the predominant aggregates prevented us from obtaining well-defined reference-free 2D class averages (Extended Data Fig. 1a,b). Furthermore, by cryo-EM on holey grids with and without continuous carbon support, we could observe only MEG3 aggregates (Extended Data Fig. 2). Different glow-discharging conditions (negative and positive polarity with varying current and duration) did not improve our EM imaging results. It thus seems apparent that specific EM optimization is required to image lncRNAs in the future. Certainly, it will probably be necessary to optimize grid preparation, that is, use of different support substrates such as graphene or graphene oxide, and vitrification, that is, blot-free and sample-spraying-based vitrification methods.
Besides grid preparation, a strong focus should, however, be applied to the optimization of the target. Production of minimal functional cores of lncRNAs, encompassing only the most highly structured regions will probably be necessary to obtain samples that ensure the highest image contrast, that are resistant to grid preparation and staining/freezing, and that yield particles that can be classified, averaged, and reconstructed in 3D. In this context, our current protocol will serve as an invaluable reference for optimizing targets and experimental conditions, which will crucially enable high-resolution lncRNA imaging by cryoEM in the coming years.

**Limitations of the approach**

Limitations of our protocol are the throughput of our analysis, the imaging resolution, and the fact that we currently cannot yet provide an atomic description of the 3D structures of our targets.

First, regarding throughput, the following considerations should be taken into account. With our approach, we have visualized ~100 target particles per condition. Although providing sufficient statistical sampling to visualize differences between lncRNA folding states\(^{10}\), this parameter could be beneficially improved, for example, by using high-speed microscopes\(^{57}\), to capture an even more complete spectrum of lncRNA conformations. Importantly, the sample preparation and biochemical characterization presented in our protocol (Steps 1–68) are fully compatible with higher-speed AFM imaging, but sample adsorption to the mica may require optimization, that is, of sample concentration and adsorption time (Steps 69–74). In addition, the operation of high-speed microscopes is different from that of the Multimode microscopes presented here (Steps 75–86) and will thus need to be carried out according to the manufacturer’s specifications and under the supervision of facility managers and AFM experts.

Second, regarding AFM image resolution, the following aspects should be considered. In AFM, ‘resolution’ can be defined in two ways. First, ‘sampling resolution’ determines the spacing between recorded data points. In our case, most images were acquired at a 1-µm scan size with 1,024 pixels (px) per line. Thus, our sampling resolution was 0.98 nm/px. Consequently, no molecular features smaller than 0.98 nm could be determined. Second, and analogously to the optical Rayleigh criterion, ‘image resolution’ can also be defined as half of the distance that can distinguish two different AFM topographic features. In this respect, at our sampling resolution (0.98 nm/px), we could easily determine distances between structural features of ~3 nm, and we thus estimate our image resolution to the value of ~1.5 nm. On the basis of these considerations, we can state that—although AFM does not offer atomic resolution for lncRNAs—it can unambiguously characterize their topography and their degree of structural compaction at nanometer resolution.

Last but not least, regarding the molecular description of our targets, it has to be considered that our lncRNA single-particle images are topographic AFM images that rely on contact-based AFM. On the one hand, contact-based AFM images are potentially affected by the convolution of sample topography and AFM tip geometry. Because the AFM tip geometry cannot be experimentally determined with sufficient precision for samples that are a few nanometers in height, true deconvolution methods cannot be applied to our image processing pipeline\(^{48}\). To reduce the tip-convolution effect, dedicated image processing techniques that are in development could be used instead\(^{49,50}\). However, deconvolution is not necessary for our protocol, whose main application is to image conformational differences of lncRNAs in their functional and non-functional states and in folded and unfolded forms. In this context, considering the height of lncRNA molecules on the mica (a few nanometers\(^{10}\)), the maximum convolution effect that our images suffer from is close to the nominal size of the AFM tip apex (2 nm, in our case). This small effect will not perturb the length measurement of lncRNA molecules, which are 30–85 nm in size, according to our SAXS, AUC, and power spectral density (PSD) analyses\(^{10}\). On the other hand, topographic AFM images are orientation dependent and cannot be used for reconstructing the atomic 3D coordinates of the target, especially not at our imaging resolution (~1.5 nm, see above). Nonetheless, lncRNA molecules will typically adsorb to the mica in all possible 3D orientations, so our AFM images capture different views of the lncRNA particles, and these views are accounted for in our PSD analysis. Moreover, our sample deposition and imaging procedures produce particles that have dimensions similar to those in solution, as determined by AUC and SAXS\(^{10}\), and it is known that viral particles or other nucleic acids are also adsorbed to the mica in conformations that closely mimic their conformations in solution (~10–15%)\(^{51}\). In the future, AFM topographic single-particle images could be used as experimental constrains to guide reconstitution of lncRNA 3D volumes by integrative structural biology methods, as can already be done for proteins\(^{52–54}\).
Experimental design

LncRNA selection

Our protocol is applicable to LncRNAs possessing a broad spectrum of lengths and sequence compositions. Indeed, our non-denaturing purification method is compatible with LncRNAs ranging from several hundred to a thousand nucleotides long\textsuperscript{10,25,26,33}, provided that the appropriate size-exclusion chromatography resin is used, as described in the 'Equipment setup' section.

For any new LncRNA of interest, we recommend performing secondary structure probing before the SAXS/AFM structural analysis\textsuperscript{13}. Chemical or enzymatic secondary structure probing methods have the potential to distinguish highly structured from loosely structured LncRNAs/LncRNA motifs\textsuperscript{17}. Loosely structured LncRNAs—of which the RepE motif of XIST could be one\textsuperscript{28}—are unlikely to be suitable candidates for topographic or shape analysis. However, because LncRNAs are typically modular, one could determine the boundaries between LncRNA domains by secondary structure probing and then focus on the analysis of only the most structured domains.

Finally, for any new target under investigation, we also advise developing functional assays as early as possible, this is, phenotypic assays in cells\textsuperscript{10} or in model organisms\textsuperscript{24,33}, or in vitro assays such as protein binding experiments\textsuperscript{26}. These assays will promote the design of biologically relevant mutants, ensuring that the structural analysis focuses on functionally meaningful features of the target and provides useful mechanistic insights.

LncRNA preparation and folding: important physical–chemical parameters and the key role of magnesium ions

Key parameters that affect LncRNA stability and folding are the temperature, the folding strategy, and the ionic conditions.

First, we strongly recommend carrying out LncRNA purification at room temperature (~20–25 °C). Cooling or freezing the LncRNA typically ‘traps’ the target in heterogeneous conformations, which compromise imaging.

Second, we strongly recommend purifying LncRNAs under non-denaturing conditions. Alternative protocols, which involve denaturation and refolding by heating and annealing and which are widely used for short RNAs, are typically more time consuming and produce lower yields. Most importantly, such protocols introduce heterogeneity into the sample, compromising imaging (Extended Data Fig. 3).

Third, it is essential to optimize ionic conditions experimentally for each new LncRNA target. It is emerging that LncRNAs can be studied at near-physiological concentrations of magnesium (1–10 mM)\textsuperscript{10,26,33}, but it is paramount to characterize the specific behavior of each target within this range of concentrations, because magnesium ions crucially determine the functional and structural properties of LncRNAs.

We recommend performing folding studies by titrating magnesium concentrations using sedimentation-velocity analytical ultracentrifugation (SV-AUC, Steps 20–38; ref.\textsuperscript{13}), which measures the velocity at which an RNA molecule sediments to the bottom of a closed compartment under high angular velocities\textsuperscript{46}. For our study, we used a Beckman Coulter Analytical XL-A/XL-I instrument equipped with an An-50 Ti analytical rotor, and Nanolytics Instruments cells and counterbalance (see also ref.\textsuperscript{13}). SV-AUC determines four important properties of the target LncRNA. First, SV-AUC determines the homogeneity of the sample at each magnesium concentration (Step 35). Coexisting species could be oligomers of the RNA molecule of interest, or unspecific aggregation products, and generally appear at high magnesium concentrations (Extended Data Fig. 4). Should these aggregated species dominate the particle distribution of the sample, further investigation of the folding conditions is necessary, because LncRNA aggregation will otherwise preclude biophysical characterization, particularly ab initio 3D shape determination by SAXS. Second, SV-AUC measures the sedimentation coefficient ($s$) and sedimentation coefficient distribution [$c(s)$] of the LncRNA of interest (Step 35). By appropriate conversion of the experimental sedimentation coefficient ($s$) into a theoretical sedimentation coefficient at standard conditions (i.e., in water at 20 °C, $s(20,w)$), it is possible to compare the hydrodynamic properties of different samples (Step 35). Third, AUC determines the LncRNA frictional ratio ($f/f_0$), which is indicative of the shape of the target LncRNA (axial ratios for oblate and prolate ellipsoid models, Step 35; ref.\textsuperscript{17}). The $f/f_0$ value can be calculated in SEDFIT, assuming that the partial specific volume and the hydration of the LncRNA are 0.53 mL/g and 0.59 g/g, respectively\textsuperscript{26}. Fourth, AUC determines the LncRNA hydrodynamic or Stokes radius ($R_H$), which is defined as the radius of an equivalent hard sphere diffusing at the same rate as the molecule under observation (Step 36). Compact folded molecules have smaller $R_H$ values, higher sedimentation.
coefficients, and a frictional ratio closer to 1 than unfolded ones. By specifically plotting $R_{H}$ values at increasing magnesium concentrations and fitting the graph to a Hill equation (Step 37), one can derive the Hill coefficient, which represents an estimate of the magnesium cooperativity for folding, and the $k_{Mg^{1/2}}$ value, which is the magnesium concentration at which the target molecule reaches 50% of its maximal compaction.

**LncRNA quality control**

Homogeneity and folding can additionally be tested by native gel electrophoresis, DLS, and MALLS (Step 39A/B/C).

Native gel electrophoresis (Step 39A(i–viii)) offers the advantage of a direct visualization of the migration pattern of the target and its potential to fold into alternative conformations. Our procedures describe how to analyze the folding state of an lncRNA such as MEG3 variant 1 (MEG3 v1), which possesses a $k_{Mg^{1/2}}$ value of 6.9 mM as experimentally determined by AUC. Conditions at lower magnesium concentrations will result in partially folded species with lower electrophoretic mobility and conditions at higher magnesium concentrations will result in folded species with higher electrophoretic mobility. The magnesium concentration values we have used for MEG3 v1 (2, 5 and 10 mM MgCl$_2$) can be used as initial references for magnesium concentrations but should be adjusted around the $k_{Mg^{1/2}}$ value determined by AUC for each lncRNA of interest.

DLS (Step 39B(i–xiii)) is performed to assess polydispersity of the target lncRNA preparation. DLS procedures that we describe here for lncRNA studies derive from well-established DLS protocols optimized for other biological macromolecules.

Finally, MALLS (Step 39C(iii–xv)) is based on the principle of proportionality between the intensity of the scattered light from biological macromolecules and their molar mass. For lncRNAs, we recommend coupling MALLS with size-exclusion chromatography (SEC-MALLS) to maximize sample homogeneity. Our procedure specifically describes how to perform SEC-MALLS using self-packed high-performance columns and an apparatus comprising a UV detector, a Wyatt Dawn HELEOS II light-scattering detector and a Wyatt Optilab T-rEX refractometer. The SEC-MALLS results (Step 39C(iii–xi)) inform about the experimental molar mass and polydispersity (Mw/Mn) of the target lncRNA. The experimental molar mass of the target should be similar ($\pm$5%) to the theoretical molar mass calculated from the target lncRNA sequence or a multiple n (integer number) thereof. If $n = 1$ the sample is monomeric, if $n > 1$, the sample is an oligomer. If the experimental molar mass is smaller than the calculated mass, this probably indicates sample degradation or incomplete transcription (see Troubleshooting section). If the experimental molar mass is larger than the calculated mass, but $n$ is not an integer, the sample is probably inhomogeneous and in equilibrium between multiple oligomeric states. Although this property may have biological relevance, it has to be carefully considered in the interpretation of AUC, SEC-SAXS, DLS, and AFM results. Moreover, a sample with monodisperse behavior will present a polydispersity value of 1. Any deviation will indicate polydispersity of the sample and should be optimized (see ‘Troubleshooting’ section). Monodispersity can also be visually detected by the overlap of the UV intensity, refraction index, and Rayleigh signals (or UV intensity and SAXS scattering intensity) peaks over the elution volume. A perfect overlap of the UV and the Rayleigh ratio curves indicates monodisperse behavior (Step 39C(xv)). In addition, the molar mass distribution of a monodisperse sample will be constant over the elution volume (Extended Data Figs. 3a–c and 4d).

**LncRNA 3D shape determination by SEC-SAXS**

Small-angle X-ray scattering (SAXS, Steps 40–64) is used to derive low-resolution information about the 3D shape of lncRNAs in solution. SAXS, like MALLS, is also based on the proportionality between the intensity of scattered light and the mass of the molecular targets. If the angular dependence of the scattered light is measured in the horizontal plane, it is possible to determine the size of the molecule. This size measurement is known as the radius of gyration ($R_g$) and is a measure of the size of the molecule weighted by the mass distribution around its center of mass (Steps 62 and 63). SAXS additionally informs about the maximal pairwise interatomic distance in the target lncRNA ($D_{max}$), and its overall volume (Steps 62 and 63). If the $R_g/D_{max}$ ratio is between 0.8 and 1.1, the lncRNA is globular, whereas if the $R_g/D_{max}$ ratio is >1.1, the sample is elongated.

As for MALLS, we specifically recommend coupling SAXS experiments with SEC, to maximize sample homogeneity. Homogeneity is particularly critical in SAXS, where small percentages of aggregates can affect the calculation of $R_g$ and prevent accurate ab initio shape determination. We performed SEC-SAXS experiments at the BioSAXS BM29 beamline at ESRF.
When ab initio shape determination is possible, the derived 3D volumes can be compared to atomistic models of the target lncRNA predicted in silico from experimentally determined secondary structure maps\(^\text{33}\). It has to be considered, however, that current software is not yet powerful enough to accurately predict high-resolution 3D structures and generally can process only lncRNA sequences up to a few hundred nucleotides long. Such predictions and the associated fit to SAXS-derived volumes can thus yield only qualitative information, not high-resolution 3D experimental models.

**LncRNA topographic imaging by AFM**

AFM (Steps 65–94) is used to visualize individual particles of the target lncRNA with the objective of capturing conformational differences in their functional and non-functional states or in different folding states. First, a large-view-field AFM image (5 × 5 μm, Step 83) gives an immediate estimation of sample homogeneity on the mica, on the density of the particles, and on the presence of any undesired features along with the target of interest (i.e., salt crystals, see ‘Troubleshooting’ section). If the large-view-field image is of good quality, one can proceed with higher-resolution scans (1 or 2 μm at 1,024 px per line, Step 85). From these scans, an immediate type of analysis consists of correlating particle size with the dimensions measured in solution by hydrodynamic techniques (i.e., \(D_{\text{max}}\) at 1,024 px per line, Step 85). From these scans, an immediate type of analysis consists of correlating particle size with the dimensions measured in solution by hydrodynamic techniques (i.e., \(D_{\text{max}}\) measured by SAXS, Steps 62 and 63). A more systematic acquisition of high-resolution scans should aim at imaging a large number of particles (i.e., at least ~100 particles per sample, per condition, as we did for MEG3 v1; ref. 10). On these particles, PSD analysis can then be carried out (see below).

The data acquisition procedures described below (Steps 75–86) are appropriate for a Multimode 8 Nanoscope V equipped with NanoScope 9.2 software but need to be customized for microscopes from different vendors. Readers should refer to the operational instructions for their specific instrument and crucially seek advice from the microscope operations manager or from an expert AFM microscopist before use. Importantly, on open-loop AFM systems such as Multimodes, the piezo scanner must be calibrated. Calibration should be done at the appropriate scanning size (≤5 μm) for imaging single molecules such as lncRNA. By default, AFM scanners are calibrated by manufacturers on the full piezo scanner length, which is ~100 μm in the case of Multimode J scanners. Refer to the manufacturer’s instruction manual for details on how to perform the dedicated calibration at lower voltages.

Finally, it must be considered that during adsorption to the mica, cations act as bridges between the lncRNA and the mica surface. Thus, the magnesium concentration used to image folded conformations of lncRNAs by AFM, should be higher than the \(k_{1/2}\text{Mg}\) value of the lncRNA of interest is too high (i.e., >25 mM), thick particulate may deposit on the mica, compromising the image quality (see ‘Troubleshooting’ section and Extended Data Fig. 5).

**LncRNA image processing using PSD analysis**

PSD analysis is performed on square AFM topographic images to extract characteristic length scales present and recurring in the image. Typically, these characteristic length scales associate to interparticle distances, particle size and any other topographic feature that recurs in the particles, such as protrusion, branches, or globular features. This ability of the PSD analysis derives from it being a form of Fourier analysis of the signal, leading to an emphasis on characteristic frequencies reappearing in the signal. The PSD is built by taking the squared modulus of the Fourier transform of the signal:

\[
\text{PSD}(f) = \frac{\Delta x}{L} (|\text{FTT}(x)|^2)
\]

Where \(\Delta x\) is the pixel size, \(L\) the size of the image, and FFT denotes the fast Fourier transform of the image, the result of an algorithm to compute a discrete, pixel-based version of the Fourier transform, which is otherwise a continuous function\(^{64}\). Notably, PSD units depend on the software used to calculate the PSD and on the dimensionality of the source data. For the calculation of characteristic distances, this scaling effect is completely immaterial. In our case, the software Gwyddion outputs PSD amplitudes as volumes; i.e., their units are \(\text{m}^3\).

In our specific case, for each line of the topographic image, an FFT is calculated, the corresponding PSD computed, and then the PSDs of all lines are averaged together to yield a global PSD of the image. The PSDs are collected along the fast scanning axis of the microscope to avoid artifacts due to line-to-line offset. Selecting small regions of interest (ROIs) surrounding individual lncRNA particles enables focusing on particle features rather than interparticle distances in the PSDs. In the case of a
periodic signal, the PSD displays a characteristic peak at the corresponding frequency, and the interpretation is straightforward. In the case of lncRNA particles, how can we relate PSD to particle size? We shall note here that qualitatively the PSD of a typical lncRNA particle displays a low-frequency plateau ($f$) followed by a power-law decay toward high frequency, of the form $f^{-\alpha}$. Intuitively, this means that there is a flat region outside the particle (low-frequency plateau) and then a knee where the sequence of higher frequencies of decremental density (following a power law), arising from the length scales associated with the actual particle, gives rise to the decay. Quantitatively, each change of slope of the power-law decay relates to a characteristic length scale associated with the particle (Fig. 3a). The intersection between the plateau and the power law decay defines the average maximal particle size (as highlighted by simulated data, Fig. 3b), whereas other changes of slope relate to intra-particle features.

The characteristic frequencies can be obtained by fitting the power-law-decay regime(s) of the PSD and looking at the intersections of the distinct power law regimes (including the low-frequency plateau).

Operationally, the PSD for all the particles imaged under the same experimental conditions is computed, and the average PSD for the specific experimental conditions is plotted against the spatial frequencies $\omega$, where $\omega = 2\pi f$. The corresponding spatial distances ($x$) can be easily calculated, taking into account that $\omega = \frac{2\pi}{x}$.

Although, in principle, it is always possible to measure the dimension of each particle individually, the rich variety of 2D projections of complex structural conformations, together with the bias induced by manual selection of profiles or contours, makes PSD analysis a powerful and rapid tool for assessing these structural parameters.

The PSD analysis can provide information at two levels. First, the presence of multiple slopes in the $f^{-\alpha}$ region provides immediate qualitative indication of a structured morphology of the lncRNA. Second, knowing the positions of the intercepts of the distinct power law decays enables estimation of the physical size of the structural domains.

Materials

**Biological materials**

- Expression plasmid coding for the target lncRNA downstream of a T7 promoter sequence and immediately upstream of a single restriction enzyme cutting site. The plasmid carries antibiotic resistance to ampicillin. In our example shown in this protocol we use plasmid pTU1, which can be obtained from the corresponding author upon request.

**Reagents**

- Acetic acid (VWR, cat. no. 20104.298) ! CAUTION This reagent is flammable and causes severe burns. Handle the reagent away from flames.
- Agar (Euromedex France, cat. no. 1329-D)
- Agarose (Euromedex, cat. no. D5-D)
- Ampicillin (Euromedex France, cat. no. EU0400-D) ! CAUTION This reagent irritates eyes, respiratory system, and skin and may cause sensitization upon inhalation and skin contact. Avoid all contact with eyes and skin, and handle the reagent under a laminar flow hood to avoid inhalation.
- ATP (GE Healthcare, cat. no. 27-1006-01)
- Boric acid (Euromedex France, cat. no. 5935) ! CAUTION This reagent is toxic upon inhalation, is toxic to the reproductive system, induces germ cell mutagenesis and is carcinogetic. Avoid all contact with eyes, skin and mouth and handle the reagent under a laminar flow hood to avoid ingestion or inhalation.
- BSA (molecular biology grade; New England BioLabs, cat. no. B9000S)
- Calcium chloride dihydrate (CaCl$_2$·2H$_2$O; Millipore, cat. no. 208291-250g) ! CAUTION This reagent irritates the eyes. Avoid all contact with eyes.
- CTP (GE Healthcare, cat. no. 27-1200-04) **CAUTION** This reagent irritates eyes, respiratory system, and skin. Avoid all contact with eyes and skin and handle the reagent under a laminar flow hood to avoid inhalation.
- Diethyl pyrocarbonate (DEPC; Sigma Aldrich, cat. no. 40718-25ML) **CAUTION** This reagent irritates eyes, respiratory system, and skin and is harmful if swallowed. Avoid all contact with eyes, skin, and mouth and handle the reagent under a laminar flow hood to avoid inhalation.
- Dithiothreitol (DTT; MP Biomedicals, cat. no. 04856126) **CAUTION** This reagent irritates eyes, respiratory system, and skin and is harmful if swallowed. Avoid all contact with eyes, skin, and mouth and handle the reagent under a laminar flow hood to avoid inhalation.
- EDTA (Carl Roth, cat. no. 80431) **CAUTION** This reagent irritates the eyes. Avoid all contact with eyes.
- Ethanol (VWR, cat. no. 20816.367) **CAUTION** This reagent is highly flammable. Handle the reagent away from flames.
- Formamide, deionized (Sigma-Aldrich, cat. no. F9037) **CAUTION** This reagent is toxic to the reproductive system, may induce fertility problems, damage the fetus, damage the cardiovascular system upon repeated or prolonged exposure and is susceptible to provoke cancer. Avoid all contact with eyes, skin and mouth and handle the reagent under a laminar flow hood to avoid inhalation.
- Glycerol (Euromedex France, cat. no. EU3550)
- GTP (GE Healthcare, cat. no. 27-2000-04)
- HEPES (Euromedex France, cat. no. 10-110-C)
- Hydrochloric acid (HCl, 37% (vol/vol); VWR, cat. no. 20252.290) **CAUTION** This reagent irritates the respiratory system, and causes burns. Avoid all contact with the skin and handle the reagent under a laminar flow hood to avoid inhalation.
- Luria Bertani (LB) medium (Dutscher, cat. no. 777495)
- Magnesium chloride hexahydrate (MgCl2·6H2O; Sigma-Aldrich, cat. no. 13152)
- Milli-Q water (ultrapure water, 18.2 MΩ·cm at 25 °C)
- MOPS (Sigma Aldrich, cat. no. M1254-100G) **CAUTION** This reagent irritates eyes, respiratory system, and skin. Avoid all contact with eyes and skin and handle the reagent under a laminar flow hood to avoid inhalation.
- One Shot Mach1 T1 phage-resistant chemically competent *Escherichia coli* cells (Thermo Fisher Scientific, cat. no. C862003) **CRITICAL** Store the cells at −80 °C for up to 1 year.
- Orange G dye (Sigma-Aldrich, cat. no. O3756-25G)
- Phenol/chloroform/isoamyl alcohol mixture (Sigma-Aldrich, cat. no. 77617-100ML) **CAUTION** This reagent is toxic upon inhalation, upon contact with skin, and if swallowed; it causes burns and shows limited evidence of a carcinogenic effect. It can cause serious damage to health upon prolonged exposure through inhalation, contact with skin, or swallowing, and the effects may be irreversible. Avoid all contact with eyes, skin, and mouth and handle the reagent under a laminar flow hood to avoid inhalation.
- Polyadenylic acid (polyA RNA; GE Healthcare, cat. no. 27-4110-01)
- Potassium hydroxide (KOH; VWR, cat. no. 26669.290) **CAUTION** This reagent is toxic upon inhalation, upon contact with skin, and if swallowed; it causes burns and is corrosive of metals. Avoid all contact with eyes, skin and mouth and handle the reagent under a laminar flow hood to avoid inhalation and ingestion.
- Proteinase K (Thermo Fisher Scientific, cat. no. 17916) **CAUTION** This reagent irritates eyes, respiratory system, and skin and may cause sensitization upon inhalation. Avoid all contact with eyes and skin and handle the reagent under a laminar flow hood to avoid inhalation.
- Quick-Load purple 2-log DNA ladder (New England BioLabs, cat. no. N0550S)
- RNase OUT (Life Technologies, cat. no. 10777019)
- RNase Remover (Bioassay; US Biological, cat. no. R2080)
- RNaseZap (Sigma-Aldrich, cat. no. R2020-250mL)
- Sephacryl S-500 HR resin (GE Healthcare, cat. no. 17-0613-01)
- Single-stranded RNA ladder (NEB, cat. no. N0362S)
- Sodium chloride (NaCl; Euromedex France, cat. no. 1112-A)
- Sodium hydroxide (NaOH; VWR, cat. no. 28244.295) **CAUTION** This reagent causes burns. Avoid all contact with skin.
- Spermidine (Sigma-Aldrich, cat. no. S0266-1G) **CAUTION** This reagent causes severe skin burns and eye damage. Avoid all contact with eyes and skin.
- SYBR Safe (Life Technologies, cat. no. S33102)
- T7 RNA polymerase (in-house produced; it can also be obtained commercially from various suppliers, e.g., New England Biolabs, cat. no. M0251S)

**CAUTION** This reagent causes burns. Avoid all contact with skin.

**CAUTION** This reagent causes severe skin burns and eye damage. Avoid all contact with eyes and skin.

**CAUTION** This reagent is highly flammable. Handle the reagent away from flames.

**CAUTION** This reagent is toxic to the reproductive system, may induce fertility problems, damage the fetus, damage the cardiovascular system upon repeated or prolonged exposure and is susceptible to provoke cancer. Avoid all contact with eyes, skin and mouth and handle the reagent under a laminar flow hood to avoid inhalation.
Tris base (Tris; Euromedex France, cat. no. 200923-A) **CAUTION** This reagent irritates eyes, respiratory system, and skin. Avoid all contact with eyes and skin and handle the reagent under a laminar flow hood to avoid inhalation.

Triton X-100 (Sigma-Aldrich, cat. no. T8787-100mL) **CAUTION** This reagent is harmful if swallowed, can cause serious damage to the eyes, is toxic to aquatic organisms, and may cause long-term adverse effects in the aquatic environment. Avoid all contact with eyes and mouth. Dispose of waste containing this reagent in dedicated bins to protect the environment.

TURBO DNase (Thermo Fisher Scientific, cat. no. AM2238)

Urea (Euromedex France, cat. no. EU0014-B)

UTP (Sigma-Aldrich, cat. no. U6750-1G)

XbaI restriction enzyme (New England BioLabs, cat. no. R0145S)

Equipment

- Agilent RNA 6000 Nano Kit (Agilent, cat. no. 5067-1511)
- Amicon Ultra-0.5 centrifugal filter unit (100 kDa; EMD Millipore, cat. no. UFC510024)
- Amicon Ultra-4 centrifugal filter unit (100 kDa; EMD Millipore, cat. no. UFC810024)
- An-50 Ti analytical rotor (Beckman Coulter, cat no. 363782)
- Analytical scale (Sartorius Entris 64I-1S; Dutscher, cat. no. 150632)
- Centrifugation cells with optical path length of 12 mm and titanium double-sector centerpieces (Nanolytics Instruments)
- Analytical ultracentrifuge (Optima AUC (A/I); Beckman Coulter, model no. B86437) **CRITICAL** We used the Beckman Coulter Analytical XL-A/XL-I models, which are no longer available. The product listed above is an appropriate substitute.
- Atomic force microscope equipped with a >100-μm piezoelectric scanner (Bruker, model no. Multimode 8)
- Centrifuge (Eppendorf, model no. 5424R; rotor no. FA-45-24-11)
- Centrifuge (Eppendorf, model no. 5804R; rotor no. A4-44)
- Cuvette washer (Sigma–Aldrich, cat. no. C1295)
- Double-sided adhesive pads for mounting samples on steel mounting disks (Bruker AFM Probes, cat. no. STKYDOT)
- Electrophoresis instrument (2100 Bioanalyzer; Agilent, model no. G2939B)
- Eppendorf tubes (Dutscher, cat. no. 033290)
- Falcon tubes (15 and 50 mL; Dutscher, cat. nos. 352096 and 352098)
- Filtration support for membranes (Nalgene, 500 mL, 45 mm; Dutscher, cat. no. 029312)
- Filters (0.22 μm; Ultrafree GV Durapore; Merck Millipore, cat. no. UFC30GV0S)
- FPLC (fast-protein liquid chromatography) instrument consisting of multiple pumps, sampler, UV detector at 260 nm, and fraction collector (AKTA FPLC purification system; GE Healthcare, model no. 28406268)
- Gel-casting apparatus
- GelDoc XR+ system with Image Lab Software (BioRad, cat. no. 1708195) with XcitaBlue conversion screen (BioRad, cat. no. 1708182)
- Hamilton syringes (50-, 100-, 250-, and 500-μL volumes; Dutscher, cat. nos. 074495, 074496, 074497, and 074492)
- Liquid chromatography empty columns (Tricorn column 10/300; GE Healthcare, model no. 28-4064-18)
- Low-volume quartz cuvettes (type 105.251, ultra-micro cells; Hellma, cat. no. 105-251-85-40)
- Membranes (0.22-μm filters, 45 mm; Merck Millipore; Dutscher, cat. no. 044121)
- Multi-angle static light-scattering (MALS) detection module (DAWN HELEOS II (Wyatt, model no. WH2-04)
- Oxide-sharpened Si3N4 microcantilevers of 115 μm length, pyramid tip, nominal tip radius of 2 nm, frequency of 70 kHz, nominal spring constant of k = 0.04 N/m (Bruker AFM Probes, cat. no. SCANASYST-AIR)
- Petri dishes (Carl Roth, cat. no. TA19.1)
- pH strips (4.5–10; Carl Roth, cat. no. C731.2)
- Straight pipette tips, Gel Load Greiner Bio-One (Dutscher, cat. no. 770291)
- Power supply for native agarose gels (BioRad, cat. no. 164-5070)
• Pump (Hitachi, model no. LaChrom Elite L-2130 HTA)
• Refractometer (Optitlab T-rEX; Wyatt, model no. WTREX-02)
• Resin (Superdex 400, 500, or 1000; GE Healthcare; cat no.17-0609-01, 17-0613-01, or 17-0476.01)
• Scotch Invisible Magic Tape (19 mm × 33 m; 3M, cat. no. 70005241826)
• Spectrophotometer (NanoDrop 2000; Thermo Fisher, model no. ND-2000)
• Stainless and magnetic steel disks (12-mm diameter; Ted Pella, cat. no. 16208)
• Stereomicroscope with C-W10×A/22 eyepiece (Nikon, cat. no. SMZ800)
• Tweezers
• UV detector (Hitachi, model no. LaChrom Elite L-2400)
• Vortex mixer (Carl Roth, model no. P505.1)

Software
• ASTRA v. 6.1 (https://www.wyatt.com/products/software/astra.html)
• ATSAS v.2.7.2-5 (https://www.embl-hamburg.de/biosaxs/download.html)65
• BsxCuBE v.6.03.11 (https://github.com/maaeli/BsxCuBE)
• DeStripe (http://biodev.cea.fr/destripe)66
• Gwyddion v.2.51 (http://gwyddion.net/download.php)67
• IgorPro v.8.03 (https://www.wavemeters.com/downloads)
• ISPyB (https://www.esrf.eu/ispyb and https://ispyb.github.io/ISPyB, version 5.4.5)68
• Microsoft Office v2013 or v2016 (https://www.office.com)
• PRIMUS (https://www.embl-hamburg.de/biosaxs/primus.html)
• Prism v.6.05 (https://www.graphpad.com/scientific-software/prism)
• SCATTER v.3.0g (https://www.esrf.eu/UsersAndScience/Experiments/CRG/BM26/SaxsWaxs/DataAnalysis/Scatter)
• Sedfit v.16.1c (http://www.analyticalultracentrifugation.com/download.htm)69
• Unicorn v.5.20 (https://www.gelifesciences.com/en/us/shop/chromatography/software/unicorn-7-p-05649)
• Zetasizer Nano S v.7.11 (https://www.malvernpanalytical.com/en/support/product-support/zetasizer-range/zetasizer-nano-range/zetasizer-nano-s)

Reagent setup

▲ CRITICAL All reagents must be prepared according to good RNA-handling practices. In particular, clean gloves should be worn at all times, and glassware should be baked in the oven at 180 °C for 4 h before use. Only DEPC-treated water should be used to prepare solutions and buffers. Immediately after preparation, all solutions are filtered through 0.22-μm membrane filters or 0.22-μm Ultrafree GV Durapore filters.

DEPC-treated water
This solution is 0.1% (vol/vol) DEPC in Milli-Q water. Measure 0.5 mL DEPC and adjust to a total volume of 500 mL with Milli-Q water. Incubate at 37 °C for 2 h and autoclave twice. This solution can be prepared in advance and stored at room temperature for several months.

Ribonucleotide stock solutions
These are 100 mM solutions of each ribonucleotide in DEPC-treated water. Weigh ~150 mg of each ribonucleotide powder and dissolve it in ~0.3 mL of DEPC-treated water. Adjust the pH by successively adding NaOH and measuring the pH with pH strips until reaching pH ~7.0. Measure the concentration in a NanoDrop spectrophotometer at 260 nm and adjust the concentration to 100 mM with DEPC-treated water, using the corresponding extinction coefficient values provided in Table 1. These solutions can be prepared in advance and stored at −20 °C for several months.

DTT, 1 M solution
Weigh 1.54 g DTT and dissolve in a total volume of 10 mL DEPC-treated water. This solution can be prepared in advance and stored in 100-μL aliquots at −20 °C for years. Avoid multiple freeze–thaw cycles.

MgCl₂, 1 M solution
Weigh 2.03 g MgCl₂·6H₂O and dissolve in a total volume of 10 mL DEPC-treated water. Correct for the decrease in the apparent density of the solution (due to the hygroscopicity of the salt) by weighing...
1 mL of solution on a precision balance and adding salt until reaching the weight for the corresponding concentration (1.07 g/mL at 20 °C). This solution can be prepared in advance and stored at 4 °C for several months.

**Tris-HCl, pH 7.5 and 8.0, 1 M solutions**
Weigh 60.55 g Tris base and dissolve in 450 mL deionized water and adjust the pH to 7.5 or 8.0 by adding HCl. Adjust the volume to 500 mL with deionized water. These solutions can be prepared in advance and stored at room temperature for several months.

**Spermidine, 2 M solution**
Weigh 2.9 g spermidine and dissolve in a total volume of 10 mL DEPC-treated water. This solution can be prepared in advance and stored as small aliquots at −20 °C for several months.

**NaCl, 5 M solution**
Weigh 146.1 g NaCl and dissolve in a total volume of 500 mL DEPC-treated water. This solution can be prepared in advance and stored at room temperature for months.

**KCl, 2 M solution**
Weigh 74.55 g KCl and dissolve in a total volume of 500 mL DEPC-treated water. This solution can be prepared in advance and stored at room temperature for months.

**EDTA-Na, pH 8.0, 0.5 M solution**
Dissolve 18.61 g EDTA in 80 mL DEPC-treated water and adjust the pH to 8.0 by adding 5 M NaOH. Adjust the volume to 100 mL with DEPC-treated water. This solution can be prepared in advance and stored at 4 °C for several months. EDTA disodium salt will not fully dissolve until the pH of the solution is adjusted to ~8.0 by the addition of NaOH.

**MOPS-K buffer, pH 6.5, 1 M solution**
Dissolve 20.93 g MOPS in 75 mL DEPC-treated water and adjust the pH to 6.5 by adding 4 M KOH. Adjust the volume to 100 mL with DEPC-treated water. This solution can be prepared in advance and stored at room temperature for several months. When diluted to 8 mM, the pH of this solution will decrease to ~6.0.

**Filtration buffer, 1×**
This solution is composed of 8 mM MOPS buffer, pH 6.5, 100 mM KCl, and 0.1 mM EDTA-Na, pH 8.0, in DEPC-treated water. Mix 4 mL 1 M MOPS buffer, pH 6.5, 25 mL 2 M KCl, and 0.1 mL 0.5 M EDTA-Na, pH 8.0, in a total volume of 500 mL DEPC-treated water. Filter the filtration buffer through a 0.22-µm filter using a vacuum pump before use in an FPLC system. This solution can be prepared in advance and stored at room temperature for several months.

**TE, pH 8, solution**
Tris-EDTA (TE) solution is 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA-Na, pH 8.0, in DEPC-treated water. Mix 1 mL 1 M Tris-HCl, pH 8.0, and 0.2 mL 0.5 M EDTA-Na, pH 8.0, in a total volume of 100 mL DEPC-treated water. This solution can be prepared in advance and stored at room temperature for several months.

### Table 1 | Molecular weights and extinction coefficients for ribonucleotides

| Ribonucleotide | MW (Da) | Extinction coefficient (ε, M⁻¹·cm⁻¹) |
|---------------|---------|-----------------------------------|
| ATP           | 507.18  | 15,346.6                          |
| CTP           | 483.15  | 7,623.2                           |
| GTP           | 523.18  | 9,929.7                           |
| UTP           | 484.14  | 11,751.8                          |
Transcription buffers (10×)
The solutions used for the various transcription buffers used in this protocol are listed in Table 2. All solutions for transcription buffers 1–4 can be prepared in advance and stored as small aliquots (i.e., 100 µL) at −20 °C for several months. BSA must be added fresh to the transcription reaction of transcription buffer 4.

| Component                  | Buffer 1 | Buffer 2 | Buffer 3 | Buffer 4 |
|----------------------------|----------|----------|----------|----------|
| MgCl₂                      | 220 mM   | 120 mM   | 100 mM   | 80 mM    |
| Tris-HCl, pH 8.0           | 400 mM   | 400 mM   | 400 mM   | 400 mM   |
| Spermidine                 | 20 mM    | 20 mM    | 20 mM    | —        |
| NaCl                       | —        | 100 mM   | —        | —        |
| KCl                        | —        | —        | —        | 1 M      |
| BSA, molecular biology grade | —      | —        | —        | 0.5 mg/mL |
| DTT                        | 100 mM   | —        | 100 mM   | 50 mM    |
| Triton X-100               | 0.1% (vol/vol) | 0.1% (vol/vol) | 0.1% (vol/vol) | —        |
| DEPC-treated water         | to 10 mL | 10 mL    | 10 mL    | 10 mL    |

*BSA should be added directly to the transcription reaction, as indicated in Table 3, not to the stock solution of 10× T7 transcription buffer 4.

Proteinase K storage buffer (1×)
This solution is 10 mM Tris-HCl, pH 7.5, 1 mM CaCl₂, and 40% (vol/vol) glycerol. Mix 0.1 mL 1 M Tris-HCl, pH 7.5, 10 µL 1 M CaCl₂, and 4 mL glycerol in a total volume of 10 mL DEPC-treated water. This solution can be prepared in advance and stored at 4 °C for several months.

CaCl₂, 1 M solution
Weigh 1.11 g CaCl₂ and dissolve in a total volume of 10 mL DEPC-treated water. This solution can be prepared in advance and stored at 4 °C for several months.

Proteinase K, 30 mg/mL suspension
This solution is a 30 mg/mL suspension of lyophilized proteinase K powder in proteinase K storage buffer. Weigh 30 mg proteinase K using an analytical scale and dissolve it in a total volume of 1 mL proteinase K storage buffer. This solution can be prepared in advance and stored at −20 °C for several months.

TAE buffer (10×)
Tris-acetate-EDTA (TAE) buffer is 400 mM Tris, 200 mM acetic acid, and 10 mM EDTA in ddH₂O. Weigh 48.5 g Tris base and dissolve it in 800 mL ddH₂O. Add 11.4 mL of glacial acetic acid and 20 mL of 0.5 M EDTA, pH 8.0. Fill to a total volume of 1 L with ddH₂O. This solution can be prepared in advance and stored at room temperature for several months.

Agarose, 1% (wt/vol) solution in 1× TAE buffer
Weigh 2 g agarose and dissolve it in a total volume of 200 mL of TAE buffer. Heat the solution in a microwave until the agarose has melted, swirling the beaker several times to dissolve any settled powder and gel pieces and to mix the agarose solution homogeneously! CAUTION The agarose solution may become overheated and foam when agitated.

Orange G solution, 2% (wt/vol)
Weigh 1 g orange G and dissolve it in 50 mL of DEPC-treated water. This solution can be prepared in advance and stored at room temperature for several months.

TB buffer (10×)
Tris-borate (TB) electrophoresis buffer (10×) solution is 890 mM Tris base and 890 mM boric acid. Weigh 26.95 g Tris base and 13.76 g boric acid and dissolve in a total volume of 250 mL.
DEPC-treated water. This solution can be prepared in advance and stored at room temperature for several months.

**RNA native loading dye (6×)**

This solution is 0.5% (wt/vol) orange G, 0.5× TB buffer, and 40% (wt/vol) sucrose in DEPC-treated water. Weigh 20 g sucrose and dissolve it in 30 mL of DEPC-treated water. Add 12.5 mL 2% (wt/vol) orange G, and 2.5 mL of 10× TB buffer. Fill to a total volume of 50 mL with DEPC-treated water. This solution can be prepared in advance and stored at 4 °C for several months.

**DNA native loading dye (6×)**

This solution is 0.5% (wt/vol) orange G, 10 mM EDTA and 50% (vol/vol) glycerol in DEPC-treated water. Mix 12.5 mL of 2% (wt/vol) orange G, 1 mL of 0.5 M EDTA-Na, pH 8.0, and 25 mL glycerol. Fill to a total volume of 50 mL with DEPC-treated water. This solution can be prepared in advance and stored at 4 °C for several months.

**AUC buffer (2×)**

Analytical ultracentrifugation (AUC) buffer is 100 mM HEPES, pH 7.4, 400 mM KCl, and 0.2 mM EDTA-Na, pH 8.0, in DEPC-treated water. Mix 15 mL 1 M HEPES, pH 7.4, 30 mL 2 M KCl, and 0.06 mL 0.5 M EDTA-Na, pH 8.0. Fill to a total volume of 150 mL with DEPC-treated water. This solution can be prepared in advance and stored at 4 °C for several months.

**HEPES buffer, pH 7.4, 1 M solution**

Weigh 59.58 g HEPES, dissolve in 200 mL DEPC-treated water and adjust the pH to 7.4 by adding KOH. Adjust the volume to 200 mL with DEPC-treated water. This solution can be prepared in advance and stored at 4 °C for several months.

**MgCl₂ solutions (5×)**

These solutions are 0.25, 5, 10, 62.5, 87.5, and 250 mM MgCl₂ in DEPC-treated water. Measure 2.5 μL, 50 μL, 100 μL, 625 μL, 875 μL, and 2.5 mL of 1 M MgCl₂ and mix each into a total volume of 10 mL DEPC-treated water. These solutions can be prepared in advance and stored at room temperature for several months.

**MgCl₂ solutions (10×)**

These solutions are 20, 50, and 100 mM MgCl₂ in DEPC-treated water. Measure 2 mL, 5 mL, and 10 mL 1 M MgCl₂ and mix each into a total volume of 100 mL DEPC-treated water. These solutions can be prepared in advance and stored at room temperature for several months.

**SYBR Safe staining solution (1×)**

This solution is 1:10,000 SYBR Safe in 1× TB buffer. Measure 5 μL of SYBR Safe dye into a total volume of 50 mL 1× TB buffer. This solution should be prepared fresh and can be reused up to three times.

**Equipment setup**

**FPLC purification system with a self-packed high-performance column**

Any FPLC system equipped with wavelength detection at 260 nm can be used. We use a Tricorn 10/300 empty high-performance column self-packed by gravity with Sephacryl S-500 HR resin for RNAs ranging from 1,000 to 2,000 nt long, although any new RNA being purified should ideally be tested on different size-exclusion resins to determine the best separation range. For additional references, we advise consulting the following website: gelifesciences.com/en/us/shop/chromatography/resins/size-exclusion.

Before the first use with RNA, the FPLC system must be thoroughly cleaned by passing successively 0.5 L filtered DEPC-treated water supplemented with RNaseZap and 0.5 L of pure DEPC-treated water. Before and after each use, the column should be equilibrated with 3 column volumes (CV) of filtration buffer and 3 CV of DEPC-treated water, respectively. The first time the self-packed column is connected to the system, the resin will compact slightly because of the increase in pressure. The adaptor of the column must then be screwed in until the resin is in contact with the top coarse filter before the column can be used.
**Procedure**

**Cloning, in vitro transcription, and lncRNA purification** ● **Timing** 3 d (cloning); 3 h (in vitro transcription); 3 h (puriﬁcation)

1. Clone the RNA of interest in a high-copy-number vector immediately downstream of a T7 promoter sequence. We have used the MEG3 v1 sequence as deposited in NCBI (NR_002766.2) and cloned it into a modiﬁed pBluescript vector by sequence and ligation-independent cloning (SLIC). We named this vector pTU110. For cloning, use LB Agar Petri dishes with the appropriate antibiotic (ampicillin, in the case of pTU1).

2. Linearize 100 µg vector with the appropriate restriction enzyme overnight. **PAUSE POINT** Linearized DNA can be stored at −20 °C for months.

3. Set up an initial screening of in vitro transcription reactions by mixing the linearized vector from Step 2 or 12 with T7 polymerase and various buffers (transcription buffers 1–4, see Table 2) in a total reaction volume of 25 µL per transcription condition. This initial screening will determine which transcription condition provides the highest yield and homogeneity for a given lncRNA (Table 3). At this stage, estimate the yield qualitatively, as described in Steps 4 and 5. **CRITICAL STEP** Transcription buffers 1–4 contain different magnesium and salt concentrations, and the choice of one over the others is empirical and depends on the particular lncRNA of interest. For example, for MEG3 v1 the highest transcription yield was obtained with transcription buffers 3 and 4 (Extended Data Fig. 6a).

4. After transcription, spin down the reaction at 20 °C and 21,130 g (15,000 r.p.m.) for 5 min to pellet the precipitated pyrophosphate.

5. Mix 10 µL of the transcription reaction with 6× RNA native loading dye and load it onto a 1% (wt/vol) native agarose gel containing 1:10,000 (vol/vol) SYBR Safe solution (3 µL for a 30-mL agarose gel), together with 5 µL of Quick-Load purple 2-log DNA ladder. Run the gel at 100 V for 30–45 min, depending of the size of your RNA. When the run is complete, visualize the gel under UV light, using a Gel Doc or Chemidoc device. Although the DNA marker will not provide accurate size estimations, it will serve as an indication of the level of compaction of the RNA transcribed under the different transcription buffers, the relative yield, and the presence of any degradation subproducts. Transcription yield can be considered satisfactory, that is, the yield is sufﬁcient for the applications described downstream (Steps 20–100), when the lncRNA band on the gel is more intense than each individual band of the DNA ladder (provided that the exact volumes of sample and ladder as described above have been loaded onto the gel). We routinely obtain sample bands that are 10–100 times more intense than those of the ladder. Examples of agarose gels for MEG3 transcription screenings are shown in Extended Data Fig. 6a. **CRITICAL STEP** When the transcription yield is sufﬁcient (see above), continue directly to Step 14. If the yield is low (i.e., sample bands on agarose gels no more intense than the DNA

### Table 3 | Screening of transcription buffers for in vitro transcription

| Component               | Buffer 1, 2 or 3 | Buffer 4 |
|-------------------------|------------------|---------|
|                         | C<sub>F</sub> | V   | C<sub>F</sub> | V   |
| Linearized plasmid     | —               | 10 µg| —              | 10 µg|
| 10× T7 transcription buffer | 1×      | 100 µL| 1×              | 100 µL|
| DTT, 1 M               | 10 mM           | 10 µL | 10 mM          | 10 µL|
| ATP, 100 mM            | 4 mM            | 40 µL | 4 mM           | 40 µL|
| CTP, 100 mM            | 4 mM            | 40 µL | 4 mM           | 40 µL|
| GTP, 100 mM            | 4 mM            | 40 µL | 4 mM           | 40 µL|
| UTP, 100 mM            | 4 mM            | 40 µL | 4 mM           | 40 µL|
| BSA, 2 mg/mL           | —               | —    | 50 µg/mL       | 25 µL|
| T7 RNA polymerase      | —               | 60 µL| —              | 60 µL|
| RNase OUT              | —               | 2 µL  | —              | 2 µL|
| DEPC-treated water     | —               | 1 mL  | —              | 1 mL|

C<sub>F</sub>, final concentration; V, volume.
ladder), we recommend purifying the linearized DNA as described in Steps 6–12 and repeating the initial screening of transcription conditions as described in Step 13. Purification of the linearized vector causes partial loss of template DNA but may increase the yield of RNA transcription.

**TROUBLESHOOTING**

6 *(Optional) Purification of the linearized DNA (Steps 6–13).* Extract the restriction reaction from Step 2 with one volume of phenol/chloroform/isoamyl alcohol (25:24:1 vol/vol/vol) followed by a second extraction with one volume of chloroform/isoamyl alcohol (24:1 vol/vol). For each extraction, mix the sample with an equal volume of solvent, vortex, and let decant on the bench until the water-soluble and solvent phases separate. Carefully aspirate the upper water-soluble phase containing the linearized DNA and transfer it to a clean tube.

▲ **CRITICAL STEP** Be conservative in the recovery of the water-soluble phase, especially in the second extraction. It is preferable to lose some DNA than to carry over chloroform/isoamyl alcohol contamination, which will inhibit the RNA polymerase and prevent transcription.

7 Add 3 volumes of ethanol to precipitate the DNA.

8 Spin down at 4 °C and 21,130g (15,000 r.p.m.) for 30 min.

9 Carefully decant and discard the supernatant.

10 Wash the pellet with 70% (vol/vol) ethanol and spin again in the same conditions for 5 min.

11 Carefully decant and discard the supernatant.

12 Dissolve the DNA in 50 µL TE solution. Measure absorbance at 260 nm using a NanoDrop spectrophotometer and derive the DNA concentration.

13 Repeat the in vitro transcription as described in Step 3.

■ **PAUSE POINT** Once optimal transcription conditions are identified, upscaling of transcription can be performed at a later date.

14 Repeat the transcription in the selected condition as described in Step 2, but in a larger volume, for example, 100 µL–1 mL.

15 After pyrophosphate precipitation (see Step 4), collect the supernatant in a clean Eppendorf tube.

16 Purify the resulting RNA following a non-denaturing protocol, as previously described13. Briefly, add 50 µL TURBO DNase to 1 mL transcription reaction and incubate for 30 min at 37 °C. After this step, add 50 µL protease K, 30 mg/mL suspension directly to the reaction tube and incubate for 45 min at 37 °C. Finally, add 26.4 µL of EDTA-Na, pH 8.0, to eliminate free magnesium, which will otherwise produce unspecific aggregation of the RNA in the next purification step.

17 Purify the transcribed lncRNA of unreacted nucleotides, digested protein peptides, protease K, and all other transcription components by filtering with Amicon Ultra-0.5 centrifugal filter units with a 100-kDa molecular weight cut-off and filtration buffer to concentrate and rebuffer the sample successively. We typically split the 1-mL transcription reaction among several filter units and concentrate for at least three successive rounds of centrifugation for 5 min at 2,348g (5,000 r.p.m.) and 20 °C.

18 Isolate a homogeneous form of the RNA of interest from other conformations, potential aggregates, and prematurely terminated transcripts by SEC using an FPLC purification system with a self-packed high-performance column equilibrated in filtration buffer. Use a constant flow rate of 0.5 mL/min to avoid changes in the compaction of the resin, which will decrease the reproducibility of the separation, and obtain fractions of 0.5 mL each. Keep the fraction corresponding to the highest peak of the chromatogram for further applications. Measure its concentration on a spectrophotometer using the Beer–Lambert law and the extinction coefficient appropriate to the lncRNA of interest, as calculated from its sequence and the extinction coefficients of individual nucleotides reported in Table 1.

■ **PAUSE POINT** Do not cool or freeze the RNA. Store at room temperature, if necessary. Follow-up experiments can be performed on different days but—although the RNA is stable at room temperature for several days—we advise using newly transcribed, fresh RNA each time.

**? TROUBLESHOOTING**

19 Check the integrity of the RNA with a 2100 Bioanalyzer using the RNA 6000 Nano Kit. To do so, prepare an aliquot in the range of 25–500 ng/µL and run it against the provided Agilent RNA 6000 ladder to confirm that a single peak, corresponding to the expected size of the target lncRNA, is visualized. An example of a Bioanalyzer trace of purified MEG3 is shown in Extended Data Fig. 6b.

▲ **CRITICAL STEP** Do not denature the RNA, for example, by gel electrophoresis or ethanol precipitation. Refolding of lncRNAs results in highly heterogeneous preparations (Extended Data Fig. 3).

■ **PAUSE POINT** Do not cool or freeze the RNA. Store at room temperature, if necessary. Follow-up experiments can be performed on different days but—although the RNA is stable at room temperature for several days—we advise using newly transcribed, fresh RNA each time.
Determination of folding conditions by AUC  

**Timing** 16 h (data acquisition); 6 h (data analysis)

20 **Data acquisition (Steps 20–30).** Set up a 1-mL transcription reaction and purify the lncRNA using the optimal transcription conditions determined in Steps 1–19. Select the fraction of the SEC chromatogram corresponding to the highest concentrations.

21 Dilute the lncRNA sample down to a concentration ranging from 0.2 to 1.2 absorbance units as measured at 260 nm (1-cm light path) to ensure a linear response of the system.

22 Set up a folding reaction by adding to the RNA 0.25 mL of 2× AUC buffer and 0.1 mL of 5× MgCl₂ solution in a final volume of 0.5 mL, so that the final magnesium concentration ranges from 0.01 to 50 mM (0.05, 1, 2, 7.5, 12.5, 17.5 and 50 mM for MEG3 folding assay). Incubate the RNA at 37 °C for 30 min.

23 For each sample, prepare a blank control containing a volume of filtration buffer instead of RNA and the same magnesium concentration as that of the corresponding experimental sample.

24 Clean each component of the analytical ultracentrifuge cells using RNaseZap and rinse first with ddH₂O and then with ethanol. Air-dry each component. Dry particularly well the centerpieces, which will be in direct contact with the sample. Place the cells and the counterbalance into the rotor, following the manufacturer’s instructions. Place 400 µL of each sample in the sample sector and 400 µL of corresponding buffer solution in the buffer sector of each analytical ultracentrifuge cell.

25 Place the rotor and the monochromator into the ultracentrifuge.

26 Perform radial calibration (one scan) at 20 °C and 66g (3,000 r.p.m.) by selecting the option *Radial calibration before first scan* in the XL-A/XL-I user interface software. At certain facilities, the calibration may be carried out regularly by the facility manager, in which case it is unnecessary to perform it at the beginning of each experiment.

27 Perform radial and wavelength scans (one scan each) at 20 °C and 66g (3,000 r.p.m.), recording the absorbance spectrum of each cell from 220 to 400 nm. These tests will show if the samples are correctly loaded; that is, the meniscus of all cells should be localized at ~6 cm in the An-50 Ti rotor using Nanalytics Instruments centrifuge cells with double-sector centerpieces (12-mm optical pathlength).

**Troubleshooting**

28 After the radial and wavelength scans, set the centrifuge speed to 0 r.p.m. and incubate the sample inside the ultracentrifuge for 1 h to allow for temperature equilibration.

29 Perform 100 scans at 20 °C and 4,550g (25,000 r.p.m.) indicating Velocity as Type of Scan, Absorbance as Optical System, and Continuous as Mode in the XL-A/XL-I user interface software. An example of an AUC scan for MEG3 v1 is provided in Extended Data Fig. 4a.

30 At the end of the scans, remove the monochromator and the rotor from the centrifuge, disassemble the cells, and wash each component first with an appropriate detergent, then with water, and finally with ethanol. If setting up a new run immediately, wash the cells and lenses first with RNaseZap, then with water, and finally with ethanol.

**Pause Point** Data analysis can be performed at a later date.

31 **Data analysis (Steps 31–38).** A detailed description of the data analysis procedure has been provided elsewhere. Briefly, when the run is finished, proceed to data analysis, which we routinely perform in SEDFIT. Alternatives to SEDFIT are DCDT+ (http://www.jphilo.mailway.com/dcdt+.html), SedAnal (http://www.sedanal.org/), Heteroanalysis (https://core.uconn.edu/files/auf/ha-help/HA-Help.htm), Sedberg (http://www.jphilo.mailway.com/sedberg.htm), and UltraScan (http://ultrascan.aucsolutions.com). The performance of some of these software alternatives for analyzing structured nucleic acids has been compared.

32 Load the files acquired by the XL-A/XL-I user interface software into the SEDFIT software. Set the meniscus and cell bottom limits to indicate the limits of the analysis.

33 Set up a Continuous c(s) distribution analysis.

34 Introduce the following parameters for the fitting: the molar mass of the target lncRNA; the range of s values to display, the resolution (100 is default); the partial specific volume, which is 0.53 mL/g for RNA molecules; the buffer density and viscosity as calculated by SEDNTERP; and the frictional coefficient initial value (we recommend starting with a value of 2 for lncRNA analysis).

35 Click on the Run and then on the Fit menu functions to start optimizing the parameters. Repeat the fit with optimized parameters, as described previously. Derive the sedimentation coefficient (s), the equivalent sedimentation coefficient at standard conditions (water, 20 °C) [s(20,w)], and the frictional coefficient (f/fo).
36 Calculate the Stokes radius ($R_H$) of the lncRNA at each magnesium concentration, using the calculator option in SEDFIT.

37 Represent the Stokes radii as a function of the magnesium concentration and fit the curve to a Hill equation, using Prism 6:

$$R_H = R_{H,0} + (R_{H,f} - R_{H,0}) \times \left( \frac{[\text{Mg}]^n}{K_{Mg} + [\text{Mg}]^n} \right)$$

where $R_H$ is the Stokes radius (in ångströms), $R_{H,0}$ and $R_{H,f}$ are the Stokes radii for unfolded and folded RNA, which correspond to the $R_H$ values determined at ~0 mM magnesium and at a concentration of magnesium higher than physiological (i.e., ≥50 mM), respectively. The $K_{Mg}$ value is the concentration of magnesium at which 50% of the RNA is folded, and $n$ is the Hill coefficient, which indicates the cooperativity of the folding transition. An example of a magnesium titration by AUC for MEG3 v1 is provided in Extended Data Fig. 4c.

38 To ensure folding in follow-up experiments, use a magnesium concentration higher than the $K_{Mg}$ calculated above.

▲ CRITICAL STEP  Accurate experimental determination of the magnesium concentrations required for folding the target lncRNA, as obtained via Steps 20–38, is extremely important because cations—and magnesium in particular—crucially affect folding of single-stranded structured RNAs.

Characterization of lncRNA monodispersity in solution

39 Characterize the purified RNA from Step 18 for monodispersity at the appropriate concentrations of magnesium (as determined in Step 38). For agarose gel electrophoresis, follow option A; for DLS, follow option B; for SEC-MALLS, follow option C.

(A) Characterization by agarose gel electrophoresis  ● Timing 2 h

(i) Set up a series of folding reactions by adding 20 µL of a 10× MgCl$_2$ solution in a final volume of 0.1 mL (0, 2, 5 and 10 mM MgCl$_2$ final concentrations for the MEG3 folding assay) to the RNA eluted from SEC (Step 18). Incubate the RNA at 37 °C for 30 min.

(ii) Prepare 1% (wt/vol) agarose gels in 1× TB buffer and 1× MgCl$_2$ solution (when necessary) to obtain the appropriate concentrations of magnesium ions (0, 2, 5 and 10 mM MgCl$_2$).

(iii) Mix 10 µL of the transcription reaction from Step 18 with RNA native loading dye (6×) supplied with the appropriate concentrations of magnesium.

(iv) Place a gel in the electrophoresis apparatus.

(v) Load samples into the wells. A Quick-Load purple 2-log DNA ladder can be used as loading control and to check for degradation subproducts if there is no need for a size reference marker. Alternatively, an ssRNA ladder can be used as a size standard on native agarose gels by using the provided 2× loading buffer, which does not denature RNA molecules.

(vi) Run the gel. Time and voltage should be adjusted to magnesium concentration. Indicatively, gels in TB with no Mg$^{2+}$ and with 2 mM Mg$^{2+}$ can be run for 45 min at 110 V, and gels in TB with 5 and 10 mM Mg$^{2+}$ can be run for 120 min at 80 V.

? TROUBLESHOOTING

(vii) Stain with 1× SYBR Safe staining solution for 1 h at room temperature before exposure.

(viii) Image without need of destaining. An example of agarose gels for MEG3 at different magnesium concentrations is shown in Extended Data Fig. 4e.

? TROUBLESHOOTING

(B) Characterization by DLS  ● Timing 1 h

(i) Wash a low-volume quartz cuvette extensively with 5 mL DEPC-treated water, followed by 5 mL 1% (vol/vol) RNaseZap in DEPC-treated water and 5 mL DEPC-treated water.

(ii) Dry the cuvette with ~5 mL of 70% (vol/vol) ethanol

▲ CRITICAL STEP For efficient cleaning, we recommend using a dedicated cuvette washer.

(iii) Terminate drying by air desiccation. Protect from dust.

(iv) Perform first measurement with filtered sample buffer. An error due to insufficient counts indicates cleanliness of the cuvette, which is thus ready to use.

(v) Filter the sample eluted from SEC (Step 18) using centrifugal filter units with 0.22-µm pore size.

(vi) Determine the precise concentration, using UV-Vis absorption on a spectrophotometer.
Pipette 45 μL of sample directly into the cuvette, avoiding formation of air bubbles.

**CRITICAL STEP** One can use straight pipette tips to reach the bottom of the cuvette more easily. Cuvettes of different volumes can be used, after ensuring compatibility with the operating instrument.

Analyze over a concentration range, for example, from 0.5 μM to 5.0 μM, starting from the highest concentration and diluting the sample at each measurement by adding filtered buffer to the cuvette, followed by gentle mixing. Perform all measurements at ambient temperature.

Open the Zetasizer Nano S software.

Configuring the new measurement under File → New → Measurement file.

Define buffer composition and experimental temperature for the sample.

Acquire the scan through Measure → Start SOP.

Export data through File → Export → Parameters → Export template → Frequency distribution. An example of a DLS trace of purified MEG3 is shown in Extended Data Fig. 3d.

**Characterization by SEC-MALLS**

- **Timing**
  - Step 39C(i,ii), 12 h; Step 39C(iii–xi), 3 h; Step 39C(xii–xv), 0.5 h
  - (i) **Data acquisition** (Step 39C(i–xi)). To avoid interference of any impurities of the buffers with the assay, filter 1 L of DEPC-treated water and all the required buffers through a 0.1-μm filter using a vacuum pump.
  - (ii) To remove impurities present in the resin, which could interfere with the scattered light, equilibrate the self-packed column on the same FPLC device used for RNA purification (see Step 18) for 12 h with filtration buffer at a 0.1-mL/min rate.
  - (iii) Before hanging the column on the MALLS system, wash the system (including the injection loop) at 5 mL/min with 20 mL DEPC-treated water, followed by 20 mL 1% (vol/vol) RNaseZap in DEPC-treated water, 20 mL DEPC-treated water, and 20 mL filtration buffer.
  - (iv) Connect the column and wash for additional 4 h in filtration buffer at a 0.4-mL/min rate. At this stage, the light scattering and refraction detectors should display a flat baseline.
  - (v) Prepare a sample of RNA eluted by SEC (Step 18) with a concentration ranging between 5 and 0.32 μM. This is an indicative concentration, which should be optimized for each target lncRNA. The sample should be filtered with a 0.22-μm Ultrafree GV Durapore filter to eliminate any aggregates. The goal is to obtain UV, scattering intensity, and index of refraction signals displaying sufficient signal-to-noise ratio without saturating the detectors.
  - (vi) After the equilibration of the column, connect the UV lamp, the refractometer, and the MALLS detector.
  - (vii) Purge the refractometer for 30 min to ensure stability of the signal. When the purge ends, set the refractometer to autozero.
  - (viii) Open the Astra software controller and create a new method with a 0.4-mL/min flow rate and the following run information:
    - Injection delay of 5 min, with no acquisition at the beginning of the run
    - Detector-washing step (COMET ultrasonic flow cell cleaner should be activated for 3 min) at the end of the run
    - The duration of the run must be calculated depending on the column used and the flow rate indicated (0.4-mL/min in this case). For example, for our Tricorn 10/30 column, the duration of passing 1 column volume (24.8 mL) at 0.4 mL/min is 62 min. When 8 min (injection delay plus duration of COMET cleaning step) is added, the total run time is 70 min.
  - (ix) Set the valve of the chromatography device in the LOAD position and click on RUN in the Astra software.
  - (x) Inject the lncRNA sample using a 100-μL Hamilton syringe
  - (xi) Move the valve to the INJECT position. The run will start and end automatically.

**PAUSE POINT** After the end of the run, data analysis can be performed at a later date.

- (xii) **Data processing** (Step 39C(xii–xv)). When the run is finished, in the Astra software set the baseline for the light scattering and refraction index signals.
  - (xiii) Find the scattering peaks either using the Autofind Peaks option or manually by eliminating unwanted peaks and dragging peak bars around the peaks of interest.
  - (xiv) Obtain the molar mass and radius of gyration from the MALLS detector, considering that the refractive index increment (dn/dc) for RNA is 0.17.
(xv) Export and plot the distributions of UV intensity, refraction index, Rayleigh ratio, and molar mass versus the elution volume. Examples of SEC-MALLS chromatograms for MEG3 are shown in Extended Data Figs. 3a–c and 4d.

? TROUBLESHOOTING

SAXS analysis ● Timing 3–4 h (data acquisition); 3–4 h (data analysis)

▲ CRITICAL To remove impurities in the resin, which could interfere with the scattered light, equilibrate the self-packed column on the same FPLC device used for RNA purification (see Step 39C(i,ii)) for 12 h with filtration buffer at a 0.1 mL/min rate. Generally, a column with a column volume of 25 mL should be suitable for loading the amount of RNA required for the SEC-SAXS experiments (0.5–5 μM, see below). The same self-packed columns as previously described for MALLS can be used.

▲ CRITICAL Alternative procedures for SAXS analysis of lncRNAs also exist and have recently been reported elsewhere, during peer review of this manuscript33.

40 Data acquisition (Steps 40–47). At the SAXS beamline, install onto the chromatography system a 500-μL loop and the shortest possible connectors between the end of the column, the UV cell, and the SAXS capillary.

41 Equilibrate the temperature of the SAXS capillary to 20 °C.

42 Before hanging the column on the SEC-SAXS system, wash the system (including the injection loop) at 5 mL/min with 20 mL DEPC-treated water, followed by 20 mL 1% (vol/vol) RNaseZap in DEPC-treated water, 20 mL DEPC-treated water, and 20 mL filtration buffer.

43 Connect the column and wash with an additional 1 column volume of filtration buffer at a 0.4-mL/min rate. At this stage, the small angle scattering detectors should display a flat baseline.

44 Filter the sample eluted from SEC (from Step 18) using centrifugal filter units with a 0.22-μm pore size.

45 Prepare different aliquots of pure lncRNA at concentrations of 0.5–5 μM.

46 Inject the sample.

47 Record scattering intensity and UV absorption at 260 nm. We generally measure 1 frame per s at 20 °C at ESRF BioSAXS beamline BM29 with a 700 × 700 μm2 nominal beam size at a flux of ~1.4 × 1012 photons/s at 12.5 keV (200 mA current).

PAUSE POINT Data analysis can be performed at a later date.

48 Data analysis (Steps 48–64). In ATSAS79, localize the raw scattering data files (.dat); these are text files that can be opened in a text editor, for example, Microsoft Excel, if needed. They contain three columns reporting scattering vector, experimental intensity, and experimental errors.

Among the raw data files, identify files to be used for background subtraction. These files should correspond to raw scattering data files with no signal, as judged from scattering intensity plots that show the scattering intensity of each frame versus the corresponding frame number.

50 Load background files in PRIMUS and click on Average.

▲ CRITICAL STEP Average and Merge perform similar operations with the following differences: Average requires the same number of data points in all frames and it is best used to combine frames that are all in the same scale; Merge has no limitations in terms of number of data points per frame, it scales the data and minimizes noise, and it is thus best used for frames having different intensities, that is, the frames collected during elution from a SEC-SAXS run.

51 Click on ‘Save’ and save as ‘averaged background file’.

52 Load all frames displaying scattering intensity.

▲ CRITICAL STEP Use only frames that correspond to the scattering of particles with constant Rg value.

53 Subtract ‘averaged background file’ from each scattering file.

! CAUTION Do not subtract ‘averaged background file’ from averaged scattering files

▲ CRITICAL STEP This step is better done with the following command in the ATSAS module:

```
DATOP: $ datop SUB sample.dat background.dat -o subtracted.dat
```

54 Average all subtracted files (see Step 50; use the same procedure as for the background files).

55 Click on ‘Save’ and save as ‘averaged data file’.

56 Open ‘averaged data file’ in PRIMUS.

57 Inspect the resulting processing curves. To compare different datasets, use the SCATTER program.

58 Define an appropriate Guinier region.

59 Restrict the data range from the beginning of the Guinier region until the highest resolution range at which the data are not yet too noisy (typically until ~3 nm−1 for MEG3).

60 Save the ‘cropped averaged data file’.
Open ‘cropped averaged data file’ in PRIMUS.

Determine $R_g$ and $D_{max}$ by optimizing smoothening factor alpha and $D_{max}$ to obtain the most accurate fit with the scattering curve.

Save the corresponding output file (.out) in which to plot relevant curves for reporting, for example, Log(I) versus s, and Guinier, Kratky, and pair distance distribution function ($P(r)$) plots. An example of SAXS data for purified MEG3 is shown in Extended Data Fig. 7a–c.

Ab initio shape determination (Step 64). Perform ab initio shape determination in the ATSAS module DAMMIF. Important output files produced by DAMMIF are:

- damfilt.pdb, which contains coordinate points present in all models.
- damaver.pdb, which contains coordinate points present in at least one model. damaver.pdb is always bigger than damfilt.pdb.
- damsel.log, which includes a cross-correlation analysis of all models. The model with the lowest normalized spatial discrepancy (NSD) value should be taken as the reference model, and it is the most accurate model to use for comparing with other techniques (e.g., AFM).
- damstart.pdb, which describes the occupancy of the coordinate points across generated models.

High-occupancy coordinates constitute the core of a flexible molecule; low-occupancy coordinates constitute the flexible regions.

An example of the shape of MEG3 determined using DAMMIF is shown in Extended Data Fig. 7d.

**CRITICAL STEP** DAMMIF should be used on samples displaying high homogeneity and unambiguous calculations of $R_g$ and $D_{max}$. Multiple runs with the same dataset testing different $D_{max}$ values from PRIMUS or multiple datasets of the same sample should result in similar structural models.

Atomic force microscopy sample preparation • **Timing 1 h**

Sample preparation (Steps 65–68). Prepare the target lncRNA in partially folded forms and fully folded forms as described above. After SEC (Step 18), incubate with appropriate magnesium concentration as for AUC (Step 38). We have used a concentration of 10 mM MgCl$_2$ to ensure folding of MEG3 v1 in our experiments$^{10}$.

**CRITICAL STEP** The concentration of magnesium is target dependent and needs to be determined experimentally, that is, by AUC, as described in Steps 20–38. Too-low magnesium concentrations (i.e., below the $K_{Mg}$ value determined in Step 38 by AUC) will not ensure folding. Too-high magnesium concentrations may induce sample aggregation (visible also in the AUC plots, Step 38) and formation of salt crystals on the mica, which would compromise data acquisition (see also Extended Data Fig. 5).

To obtain denatured samples, precipitate the target lncRNAs with isopropanol overnight at $-20 ^\circ$C, then resuspend it in deionized formamide, and finally dilute it with ethanol to reach the same final concentration as that of the samples diluted in buffer.

Use poly(A) RNA, dissolved in the same buffers as the target lncRNA, as negative control. A concentration of 0.3 µg/mL poly(A) RNA is generally appropriate for imaging.

Use a highly structured RNA as a positive control. We used the *Oceanobacillus iheyensis* group II intron ribozyme, which has been crystallized$^{71,76–78,80}$. Prepare the intron in the same way as the target lncRNA.

AFM data acquisition • **Timing 0.5 h (mica preparation and cantilever mounting); 0.5 h (sample adsorption to the mica); -2 d per condition (data acquisition)**

Mica preparation and cantilever mounting (Steps 69–73). Glue a mica disk to the steel disk, using a double-sided adhesive tape.

**CRITICAL STEP** We recommend the use of freshly cleaved non-derivatized mica, which ensured the best-quality image acquisition, in our case using the lncRNA MEG3. Mica derivatization with divalent metal ions (nickel) or silane chemistry (3-aminopropyltrihydroxy silane (APTES)) is also possible and would allow for stronger RNA adsorption, but in our case it yielded lower-quality images (see Extended Data Fig. 5d).

Position the AFM probe holder on a flat surface, exposing the tip fastening mechanism toward yourself.

Gently squeeze an AFM probe from the middle with the tweezers of choice, making sure to notice where the cantilevers you need are, because sometimes both sides of the probe contain cantilevers. We recommend using straight-edged blunt-ended tweezers.

Push down the spring-releasing mechanism from the probe holder and insert the probe beneath the fastening mechanism.
73 Make sure the probe is straight in the lodging space of the probe holder. A stereomicroscope is of particular help at this stage. Be careful not to touch the end where the cantilevers are located when moving the probe.

74 **Sample adsorption to the mica (Step 74)**. Adsorb the RNA sample to the mica. For partially and fully folded RNA samples, follow option A; for denatured RNA samples, follow option B.

(A) **Adsorption of partially and fully folded RNA samples to the mica**

(i) Dilute the target IncRNA in its partially or fully folded form (as determined by AUC in Step 38 and prepared experimentally in Step 65) to 1–10 nM.

(ii) Cleave the mica with Scotch tape.

(iii) Pipette 1–5 µL of RNA onto the freshly cleaved mica.

(iv) Incubate for 3 min.

(v) Wash with 2 mL of water (10 × 200-µL drop) to remove salt crystals.

(vi) Dry with nitrogen gas.

▲**CRITICAL STEP** Incubation and drying times are critical to ensuring optimal image quality (Extended Data Fig. 5).

▲**CRITICAL STEP** Proceed immediately (within minutes) to data acquisition. Longer storage of the mica may lead to formation of salt crystals or deposition of dust and impurities on the mica, which will compromise the image quality.

? **TROUBLESHOOTING**

(B) **Adsorption of denatured RNA samples to the mica**

(i) Dilute the denatured RNA as obtained in Step 64, to 0.1–1 nM.

(ii) Cleave the mica with Scotch tape.

(iii) Pipette 1–5 µL of RNA onto the freshly cleaved mica.

(iv) Incubate for 3 min.

(v) Dry with nitrogen gas.

▲**CRITICAL STEP** Proceed immediately (within minutes) to data acquisition. Longer storage of the mica may lead to formation of salt crystals or deposition of dust and impurities on the mica, which will compromise the image quality.

? **TROUBLESHOOTING**

75 **Data acquisition (Steps 75–86)**. On the microscope (e.g., a Multimode 8, Nanoscope V equipped with NanoScope v.9.2 software) insert the disk sample into the AFM piezo system. Make sure that the sample is well centered.

! **CAUTION** Make sure there is enough room between the sample and the probe holder so that the cantilever will not crash onto the sample surface; if needed, use the step motor to lower your sample position (or move your cantilever higher).

76 Install the AFM probe holder on top the sample and perform the laser alignment, following the manufacturer’s recommendations. Sometimes it is easier to align the laser without the sample, so this alignment step could be performed earlier in the protocol.

77 Focus your top-view camera to identify the surface of your sample and the position of your cantilever. If you are far away from the surface, use the step-down motor to lower the cantilever over the sample (it will avoid an unnecessarily long wait for engagement).

! **CAUTION** Keep a safe distance from the sample surface because if the cantilever crashes onto the sample (even without any visual defect of the cantilever), it is necessary to replace the cantilever.

78 Set up your AFM software with the appropriate imaging mode. For MEG3 analysis, we use peak-force tapping (PFT, ScanAsyst mode) in air. Imaging can also be performed by conventional tapping in air or by using the repulsive force regime and a correctly configured tapping mode in liquid, as described elsewhere.

79 Set your scan size and any offset values to 0.

80 Set the ScanAsyst mode to fully automatic.

81 Engage your cantilever.

82 When the cantilever touches the surface, slowly increase the scanning size to 200 nm.

83 Wait for a few seconds for the system to adjust the scanning parameters and then gradually increase the scanning size to 500 nm, 2 µm, and 5 µm.

▲**CRITICAL STEP** It is important to maintain a reasonable scanning resolution (number of nanometers per pixel). On a 5-µm scan size at 512 px per line, the scanning resolution is 9.8 nm/px, which is acceptable to identify appropriate locations of IncRNA molecules on the sample substrate. However, a scanning resolution of ~1 nm/px is required for topographical imaging.
84 If necessary, turn off some automated ScanAsyst options, such as setpoint or gain. In fact, it is usually necessary to have some control over the imaging parameters.

85 Image in PFT mode at an ~1-Hz rate (at piezo oscillations of 2 kHz), with 512- or 1,024-px sampling (depending on the scanning size). With a small scan size (≤2 µm), use a set-point of ~200 pN.

**CRITICAL STEP** When scanning at minimal forces, any vibration of the AFM can cause the AFM tip to lose contact with the sample surface. Use an efficient vibration-isolated AFM setup; for example, antivibration tables or heavy platforms supported by bungees. Acoustic isolation of the AFM can be efficiently achieved using an acoustic foam-coated casing for the microscope.

86 At high magnification, and because of the size of the samples (a few nanometers tall), the vertical range of the z piezoelectric scanner can be reduced (e.g., from 6 µm to 1 µm) to enhance the vertical resolution by the digitalization of the signal.

**PAUSE POINT** Image processing and data analysis can be performed at a later date.

**TROUBLESHOOTING**

**AFM image processing** ● **Timing** 6 h

87 Open raw images in Gwyddion. A detailed metrological analysis of biases in nano-object characterization with AFM and related techniques can be found elsewhere.

88 In Data process, level raw images by mean plane subtraction.

89 Remove polynomial background.

90 Perform row alignment and horizontal scar correction.

91 (Optional) If needed, remove stripe noise using DeStripe. Operationally, DeStripe can be used through its web-based graphical user interface after preparing a suitable input file. Instructions to prepare the input file and the web-based interface are located at http://biodev.cea.fr/destripe/.

**CRITICAL STEP** Stripe noise removal should be performed if visual inspection of the images suggests the presence of stripes.

92 Adjust the color scale.

93 (Optional) If desired to improve the graphical quality of the images, additional filtering options can be applied in Gwyddion, such as the non-linear smoothing noise reduction Kuwahara filter.

**CRITICAL STEP** Image filtering is particularly useful for smoothing the background (keeping the single molecules excluded from the filter).

94 Render the images either as 2D or 3D representations (Figs. 2e, f and 4).

**AFM image analysis** ● **Timing** 5–6 h

95 Open the AFM images using Gwyddion. Correct them according to the protocol described in Steps 87–94. Draw a square ROI of 256 px (corresponding to ~256 nm in the reported imaging conditions) around each particle (Fig. 3e), so that only one particle is present in the ROI. Save the ROI as a new image (Fig. 3f).

96 Calculate the PSD within the ROI of the image. If using Gwyddion, choose 1D Statistical Functions – PSD and choose the fast scan axis (x in our case) for the analysis (Fig. 3g).

97 Export the PSD function for the chosen ROI as a text file in ASCII format. The text file contains two columns, one for the spatial frequencies and the other for the PSD amplitudes.

98 Average the selected ROIs in the software of your choice. We typically select 100 ROIs and average them using Excel in Microsoft Office. For averaging, place the columns corresponding to the PSD amplitudes into a new Excel sheet (one column per ROI) and then average the values row by row.

99 Plot PSD versus spatial frequencies of individual particles or after averaging in the software of choice. We typically use IgorPro. In IgorPro, change both axes to log-scale. Use the cursors function in the plot to select what now appear as linear decays in the f⁻¹ region of the PSD (see equations in the ‘Experimental design’ section above). Use Analysis-Curve Fitting-power as the fitting function, selecting the option fit between cursors to fit into the desired range. Hold the offset to 0, effectively fitting Ax – b to the data, with A and b as free parameters. Repeat the procedure for the other linear ranges in the PSD, including the low-frequency plateau.

100 The intersection points between the linear fits yield the characteristic frequencies of the particles, namely average diameter and average size of any intra-particle feature, where present (Fig. 3h).

**TROUBLESHOOTING**
Troubleshooting

Table 4 | Troubleshooting table

| Step | Problem | Possible reason | Solution |
|------|---------|----------------|----------|
| 5    | Low yield (i.e., sample bands on agarose gels no more intense than the DNA ladder; see text for details; see also Extended Data Fig. 6a) | Vector sequence or transcription buffers are not optimized | Ensure that the transcription vector possesses at least two Gs immediately downstream of the T7 promoter, before the start of the target RNA sequence. Further modify the proposed transcription buffers by screening a wider range of magnesium concentrations. |
|      | RNA degradation can be observed on the agarose gel | RNase contamination in any of the reagents | Discard all reagents and prepare fresh working solutions using newly made DEPC water. |
|      | Target IncRNA migrates with unexpected electrophoretic mobility | Template DNA not digested, or cleaved at the wrong site by restriction endonuclease | Resequence the plasmid and ensure that the restriction enzyme chosen for 3′-site cleavage is unique. |
| 18   | Final SEC profile not homogeneous | LncRNA has aggregated during Amicon centrifugation | Use larger Amicon devices and concentrate less. Dialyze the sample against filtration buffer instead of rebuffering using Amicon devices. |
|      | Final SEC profile is not homogeneous | Unstructured LncRNA | Identify unstructured regions by secondary structure probing and remove them by cloning (ensuring functionality). |
| 27   | Meniscus displaced toward the cell bottom in acquired data | Cell leakage has occurred during the run | Repeat Steps 21–25, ensuring sealing in the filling holes, in the centerpiece face to the outside of the cell, and from sector to sector. Lubricate the screw-ring threads, increase the torque to 140 inch-pounds or, if necessary, replace damaged components. |
|      | Low absorbance | Absorbance data at 260 nm are filtered out | The monochromator 400-nm high-pass filter, which blocks all wavelengths <400 nm, is placed in the light path. To move the filter out of place, set the filter lever on the side of the monochromator parallel to the monochromator stem. |
| 35   | Strong diagonal line observed in the residual bitmap after fitting the data | There is a mismatch between the data and the calculated curve | The fitting needs to be refined by changing some of the parameters indicated in the model. |
| 39A(vi) | The gel heats up and partially melts; the run results in altered migration of bands and degradation | The agarose gel contains high magnesium concentrations (>10 mM) | Run the folded RNA in an agarose gel with low Mg concentration, and keep it constant for all folded RNAs under different Mg concentrations. Lower the voltage |
| 39A(viii) | Smear bands can be seen on the gel | Unstructured LncRNA | Identify unstructured regions by secondary structure probing and remove them by cloning (ensuring functionality). |
| 39C(xv) | UV/visible light scattering (LS)/refractive index (RI) signals do not overlap (see also Extended Data Figs. 3a–c and 4d) | LncRNA aggregates | If the SEC profile is homogenous, quantify amount of aggregates by AUC. If low, proceed with AFM imaging. Optimize ionic conditions or complement filtration buffer with additives (i.e., polyamines, chaotropic agents) to reduce polydispersity. |
|      | Molar mass of target lncRNA does not match the expected molar mass calculated from its sequence | Template DNA not digested, or cleaved at the wrong site by restriction endonuclease | Resequence the plasmid and ensure that the restriction enzyme chosen for 3′-site cleavage is unique. |
| 74A(vi), 74B(v) | Formation of large thick particulate on the mica (see also Extended Data Fig. 5) | Salt crystals | Increase the wash time and volume after deposition of the sample. Decrease the drying time. |
| 86   | Sample drift | Double-sided adhesive tape causes the drift | Adhere the mica to the steel disc using a two-part epoxy resin. Ensure sample stability and homogeneity by gel electrophoresis. Use clean AFM pipettes and solutions. Use fresh mica. |
|      | Particles inhomogeneous in size | LncRNA degradation | |

Table continued
Timing

The timing notes listed in the Procedures refer to the timing typically required by a trained researcher to implement the protocol after optimization of the experimental conditions required by the target lncRNA of interest. The timing is also critically dependent on the exact instruments used. We indicate the timing suitable to perform the experiments using the same instruments that we have used, as listed in the Materials section, under Equipment. AFM acquisition speed can be markedly improved using higher-speed AFMs. For AFM image acquisition, at a 1-Hz scanning rate, the acquisition of a single 512 px/line image takes 8.5 min, whereas for a single 1,024 px/line image, it takes 17 min. Supervision by AFM/SAXS/AUC experts is required when operating an AFM microscope, a SAXS synchrotron beamline, or an AUC.

Steps 1 & 2 and 6–12, cloning and linearization: 3 d
Steps 3–5 and 13–19, lncRNA non-denaturing purification from linearized DNA: 6 h
Steps 20–38, AUC: 16 h for data acquisition (per replica; each replica can analyze up to 7 magnesium concentrations), and 6 h for data analysis
Step 39A(i–viii), agarose gel electrophoresis: 2 h
Step 39B(i–xiii), DLS: 1 h
Step 39C(i,ii), column equilibration and system washing: 12 h and 4 h, respectively
Step 39C(iii–xv), SEC-MALLS: 3 h for data acquisition and 0.5 h for data analysis
Steps 40–64, SEC-SAXS: 3–4 h for data acquisition; 3–4 h for data analysis
Steps 65–86, AFM image acquisition: ~2 d per condition
Steps 87–100, AFM image processing and analysis: 6 h for processing and 6 h for PSD analysis (per sample and experimental condition)

Anticipated results

This protocol typically produces milligram amounts of lncRNA (i.e., 2–3 mg pure MEG3 v1 per 1-mL transcription reaction) via a non-denaturing purification method (Steps 1–19). Appropriate lncRNA folding conditions are determined by SV-AUC (Steps 20–38).

A successful SV-AUC run can be recognized by looking at the raw data distribution and residuals in the SEDFIT software. The heat map of the scans should display a gradient from blue to red colors, without a predominance of dark-blue scans (sample not completely sedimented) or red scans (in the latter case, fewer scans are enough to sediment the sample). If there has not been leakage, the meniscus should be located at about the 6-cm position, whereas the bottom of the cell should be at ~7.2 cm, when using a Beckman Coulter Analytical XL-A/XL-I instrument equipped with an An-50 Ti analytical rotor, and Nanolights Instruments cells and counterbalance. An example of SV-AUC data obtained for MEG3 v1 is shown in Extended Data Figs. 3a–c and 4d.

Folded lncRNAs should run as a single homogeneous sharp band on non-denaturing agarose gels. Additional sharp bands may also form, particularly at low magnesium concentrations; these probably indicate alternative discrete conformations of the sample (see ‘Troubleshooting’ section and Extended Data Fig. 4e). These bands should disappear, or at least substantially decrease in intensity, at higher
magnesium concentrations. If the purified lncRNA shows smearing on the gel, this is generally indicative of unstructured lncRNAs or lncRNAs prone to degradation. Unstructured or degraded samples are not suitable for further structural investigation. Sample homogeneity, purity, and integrity can also be assessed by DLS (Extended Data Fig. 3d) and SEC-MALLS (Extended Data Figs. 3a–c and 4d).

Homogeneous pure lncRNAs are then structurally analyzed in solution by SEC-SAXS, which enables determination of the $R_g$ and $D_{max}$ values, as well as globularity of the targets (Step 63, Extended Data Fig. 7a–c) and of their 3D shape (Step 64, Extended Data Fig. 7d).

Finally, the 3D topography of lncRNAs is visualized by AFM in different folding states (Steps 65–100; Fig. 4). Crucially, AFM analysis is guided by the use of specific controls. Polyadenylic acid serves as a control for unfolded RNAs, because although it forms agglomerates, polyA does not fold into a compact structure even in the presence of magnesium ions (Fig. 4). Instead, the group II intron ribozyme, which has been previously crystallized\textsuperscript{71,76,77}, can serve as a control for folded RNAs (Fig. 4). Ribosomal RNAs could be used as an alternative to the group II intron.

Specific structural features of the AFM topographic images can be analyzed by PSD analysis. The corner frequencies of the PSD (intersection between low-frequency plateau and $f^{-a}$ decay and intersections between different regions of the $f^{-a}$ decay) provide a measure of the diameter and overall size of lncRNA molecules and of their structural subdomains (Fig. 3h). To exemplify what should be anticipated when using PSD analysis, the following considerations should be made. Upon acquiring a set of high-resolution AFM pictures of lncRNA particles on mica substrates, one will probably have a collection of single-particle pictures such as the one depicted in Fig. 3d, which clearly displays a complex morphology. Qualitatively, the particle displays a central core, about 10 nm in diameter and higher than the rest, and three features branching out radially, at about 120° from each other, each spanning again about 10 nm outward. The entire particle can be circumscribed by a circle of radius $\sim$50 nm in this case. The PSD analysis provides a quantitative estimate of this morphological complexity (Fig. 3f). Here, the pooled data from hundreds of particles yield an average curve that exhibits a set of characteristic frequencies, at $\sim$80 nm and $\sim$25 nm, which respectively reflect the average lateral size of the entire lncRNA particle on the mica and the presence of smaller features, on the order of a few tens of nanometers, which give rise to the aforementioned internal morphology. Although more complex choices for characterizing the morphology of the particles are certainly possible, PSD analysis is thus a robust and straightforward way to appraise and quantitate the basic morphological features shared among a large dataset.

**Reporting Summary**

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

All data generated or analyzed during this study are included in the paper and its Supplementary Information and are available from the corresponding author on request. We have deposited in figshare an extensive set of AFM images, which we have used for calculation of the PSDs:

MEG3 v1 compact form: https://figshare.com/s/e73aff439aeaad75e456

MEG3 v1 intermediate form: https://figshare.com/s/a81600774ab9baed8932

MEG3 v1 denatured form: https://figshare.com/s/026d8eb31a8a4ab6910e9

MEG3 H11LpA mutant, compact form: https://figshare.com/s/666c9998f22d661d72f3

**References**

1. Mattick, J. S. & Rinn, J. L. Discovery and annotation of long noncoding RNAs. Nat. Struct. Mol. Biol. 22, 5–7 (2015).

2. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. Cell 136, 629–641 (2009).

3. Novikova, I. V., Hennelly, S. P. & Sanbonmatsu, K. Y. Sizing up long non-coding RNAs: do lncRNAs have secondary and tertiary structure? Bioarchitecture 2, 189–199 (2012).

4. Pyle, A. M. Looking at LncRNAs with the ribozyme toolkit. Mol. Cell 56, 13–17 (2014).

5. Kaneko, S., Son, J., Bonasio, R., Shen, S. S. & Reinberg, D. Nascent RNA interaction keeps PRC2 activity poised and in check. Genes Dev. 28, 1983–1988 (2014).

6. Sasaki, Y. T., Ideue, T., Sano, M., Mituyama, T. & Hirose, T. MENpsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc. Natl Acad. Sci. USA 106, 2525–2530 (2009).
7. Schmitt, A. M. et al. An inducible long noncoding RNA amplifies DNA damage signaling. *Nat. Genet.* 48, 1379–1376 (2016).
8. Liu, F., Somarowthu, S. & Pyle, A. M. Visualizing the secondary and tertiary architectural domains of lncRNA Rep3. *Nat. Chem. Biol.* 13, 282–289 (2017).
9. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165, 1267–1279 (2016).
10. Uroda, T. et al. Conserved pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Mol. Cell* 75, 982–995.e9 (2019).
11. Duszczyk, M. M., Wutz, A., Rybin, V. & Sattler, M. The Xist RNA A-repeat comprises a novel AUUG tetraloop fold and a platform for multimerization. *RNA* 17, 1973–1982 (2011).
12. Brown, J. A. et al. Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.* 21, 633–640 (2014).
13. Chillon, I. et al. Native purification and analysis of long RNAs. *Methods Enzymol.* 558, 3–37 (2015).
14. Kaneko, S. et al. Interactions between JARD2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol. Cell* 53, 290–300 (2014).
15. Zhou, Y. et al. Activation of p53 by MEG3 non-coding RNA. *J. Biol. Chem.* 282, 24731–24742 (2007).
16. Feig, A. L. & Uhlenbeck, O. C. in The RNA world (eds R. F. Gesteland, T. R. Cech, & J. F. Atkins) Ch. 12, 287–329 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1999).
17. Novikova, I. V., Dharap, A., Hennelly, S. P. & Sanbonmatsu, K. Y. 3S: shotgun secondary structure determination of long non-coding RNAs. *Methods* 63, 170–177 (2013).
18. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* 11, 959–965 (2014).
19. Volders, P. J. et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* 41, D246–D251 (2013).
20. Pennisi, E. Genomics. ENCODE project writes eulogy for junk DNA. *Science* 337, 1159–1161 (2012).
21. Novikova, I. V., Hennelly, S. P. & Sanbonmatsu, K. Y. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* 40, 5034–5051 (2012).
22. Chillon, I. & Pyle, A. M. Inverted repeat Alu elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function. *Nucleic Acids Res.* 44, 9462–9471 (2016).
23. Somarowthu, S. et al. HOTAIR forms an intricate and modular secondary structure. *Mol. Cell* 58, 353–361 (2015).
24. Zhou, Y., Zhang, X. & Klibanski, A. MEG3 noncoding RNA: a tumor suppressor. *J. Mol. Endocrinol.* 48, R45–R53 (2012).
25. Schmitt, A. M. et al. An inducible long noncoding RNA amplifies DNA damage signaling. *Nat. Genet.* 48, 1379–1376 (2016).
26. Liu, F., Somarowthu, S. & Pyle, A. M. Visualizing the secondary and tertiary architectural domains of lncRNA Rep3. *Nat. Chem. Biol.* 13, 282–289 (2017).
27. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165, 1267–1279 (2016).
28. Uroda, T. et al. Conserved pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Mol. Cell* 75, 982–995.e9 (2019).
29. Duszczyk, M. M., Wutz, A., Rybin, V. & Sattler, M. The Xist RNA A-repeat comprises a novel AUUG tetraloop fold and a platform for multimerization. *RNA* 17, 1973–1982 (2011).
30. Brown, J. A. et al. Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.* 21, 633–640 (2014).
31. Chillon, I. et al. Native purification and analysis of long RNAs. *Methods Enzymol.* 558, 3–37 (2015).
32. Kaneko, S. et al. Interactions between JARD2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol. Cell* 53, 290–300 (2014).
33. Zhou, Y. et al. Activation of p53 by MEG3 non-coding RNA. *J. Biol. Chem.* 282, 24731–24742 (2007).
34. Feig, A. L. & Uhlenbeck, O. C. in The RNA world (eds R. F. Gesteland, T. R. Cech, & J. F. Atkins) Ch. 12, 287–329 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1999).
35. Novikova, I. V., Dharap, A., Hennelly, S. P. & Sanbonmatsu, K. Y. 3S: shotgun secondary structure determination of long non-coding RNAs. *Methods* 63, 170–177 (2013).
36. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* 11, 959–965 (2014).
37. Volders, P. J. et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* 41, D246–D251 (2013).
38. Pennisi, E. Genomics. ENCODE project writes eulogy for junk DNA. *Science* 337, 1159–1161 (2012).
39. Novikova, I. V., Hennelly, S. P. & Sanbonmatsu, K. Y. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* 40, 5034–5051 (2012).
40. Chillon, I. & Pyle, A. M. Inverted repeat Alu elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function. *Nucleic Acids Res.* 44, 9462–9471 (2016).
41. Somarowthu, S. et al. HOTAIR forms an intricate and modular secondary structure. *Mol. Cell* 58, 353–361 (2015).
42. Zhou, Y., Zhang, X. & Klibanski, A. MEG3 noncoding RNA: a tumor suppressor. *J. Mol. Endocrinol.* 48, R45–R53 (2012).
43. Schmitt, A. M. et al. An inducible long noncoding RNA amplifies DNA damage signaling. *Nat. Genet.* 48, 1379–1376 (2016).
44. Liu, F., Somarowthu, S. & Pyle, A. M. Visualizing the secondary and tertiary architectural domains of lncRNA Rep3. *Nat. Chem. Biol.* 13, 282–289 (2017).
45. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165, 1267–1279 (2016).
46. Uroda, T. et al. Conserved pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Mol. Cell* 75, 982–995.e9 (2019).
Acknowledgements

We thank A. McCarthy and the beamline scientists at the BM29-BioSAXS beamline for their support during SAXS data collection; A. Leroy and C. Ebel (IBS Grenoble) for support with AUC; S. Acajjaoui and M. Soler Lopez (ESRF Grenoble) for support with DLS; and C. Mas (IBSG Grenoble) and M. Jamin (UGA Grenoble) for support with SEC-MALLS. We also thank all members of the Marcia lab for helpful discussions. Work in the Marcia lab was partly funded by the Agence Nationale de la Recherche (ANR-15-CE11-0003-01), the Agence Nationale de Recherche sur le Sida et les hépatites virales (ANRS, ECTZ18552), ITMO Cancer (18CN047-00), and the Fondation ARC pour la recherche sur le cancer (PA-20191209284). This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-05-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA). This work acknowledges the AFM platform at the IBS.

Author contributions

M.M. conceived the project and obtained funding; M.M. and I.C. supervised the research and wrote the initial draft of the manuscript; T.U., M.M., J.-M.T., and J.-L.P. developed the procedures for AFM image acquisition; T.U., I.C., and M.M. developed the procedures for lncRNA purification, AUC, and MALLS; T.U. and M.M. developed the procedures for DLS and SAXS experiments; M.K. carried out EM experiments; O.P. provided technical assistance and contributed to sample preparation; P.A., J.-L.P., and M.M. developed the procedures for AFM image analysis. All authors reviewed and edited the manuscript and approved the final draft.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.M.

Peer review information Nature Protocols thanks Yuri L. Lyubchenko, Neil H. Thomson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 9 December 2019; Accepted: 24 March 2020; Published online: 25 May 2020

Related links

Key reference using this protocol

Uroda, T. et al. Mol. Cell 75, 982-995 (2019): https://doi.org/10.1016/j.molcel.2019.07.025
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

☐ Confirmed
☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐ The statistical test(s) used AND whether they are one- or two-sided
☐ Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☐ A description of all covariates tested
☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
☐ Give P values as exact values whenever suitable.
☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection:
- chromatography: Unicorn 5.20 (GE Healthcare); MALLS: ASTRA 6.1 (Wyatt); DLS: Zetasizer Software 7.11 (Malvern Instruments);
- synchrotron data collection: iSpyB 5.4.5 (Delageniere 2011); AFM: NanoScope 9.2 (Bruker); SAXS: BsxCube (https://github.com/maaeli/BsxCube) + spec 6.03.11 (https://certif.com/spec.html);
- RNA extraction and quality assessment: QuBIT 3.0 Fluorometer APP 1.02 + MCU 0.21 (Life Technologies) and BioAnalyzer 2100 Expert B.02.08 SI648 (SR3) (Agilent Technologies);
- VIS-UV: Nanodrop 2000/2000C 1.6.198 (Thermo Fisher Scientific)

Data analysis:
- AUC: ProteomeLab XL-A/XL-I User Interface Software (Beckman Instruments), Sedfit 14.6e (Schuck 2000), GUSSI 1.08 (The University of Texas Southwestern Medical Center), Prism 6.05 for Windows (GraphPad);
- SANS: ATSAS 2.7.2.5 (Petoukhov 2012), BsxCube (https://github.com/maaeli/BsxCube, version spec 6.03.11); AFM: Gwyddion 2.45 and 2.51 (Necas 2011) and DeStripe (Chen 2011) and IgorPro (WaveMetrics);
- Cloning and primer design: CloneManager Professional Suite 6.00 (Sci Ed Central) and Primer Design 4.20 (Sci Ed Central)
- OligoAnalyzer 3.1 (Integrated DNA Technologies), figures and manuscript preparation: Microsoft Office 2016 (Microsoft)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated or analyzed during this study are included in the paper and its Supplementary Information and are available from the corresponding author on request.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

- Sample size
  - Sample size determination is not applicable to our study.
- Data exclusions
  - No data were excluded from the analysis.
- Replication
  - All biochemical and biophysical analysis including the AFM imaging could be reproduced successfully by independent researchers at distant time points (over the course of 2-3 years of research).
- Randomization
  - Not applicable to our study.
- Blinding
  - Not applicable to our study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a                             | n/a     |
| - Antibodies                    | - ChIP-seq |
| - Eukaryotic cell lines         | - Flow cytometry |
| - Palaeontology                 | - MRI-based neuroimaging |
| - Animals and other organisms   |         |
| - Human research participants   |         |
| - Clinical data                 |         |