HHS Public Access

Author manuscript

Cell Rep Phys Sci. Author manuscript; available in PMC 2024 December 26.

Published in final edited form as:

Cell Rep Phys Sci. 2024 November 20; 5(11): . doi:10.1016/j.xcrp.2024.102264.

Sequence-dependent conformational preferences of disordered single-stranded RNA

Tong Wang[1,4], **Weiwei He**[2,3,4], **Suzette A. Pabit**[1], **Lois Pollack**[1,*], **Serdal Kirmizialtin**[2,3,5,*]

[1]School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA

[2]Chemistry Program, Science Division, New York University Abu Dhabi, Abu Dhabi 129188, UAE

[3]Department of Chemistry, New York University, New York, NY 10003, USA

[4]These authors contributed equally

[5]Lead contact

SUMMARY

Disordered single-stranded RNA (ssRNA) molecules, like their well-folded counterparts, have crucial functions that depend on their structures. However, since native ssRNAs constitute a highly heterogeneous conformer population, their structural characterization poses challenges. One important question regards the role of sequence in influencing ssRNA structure. Here, we adopt an integrated approach that combines solution-based measurements, including small-angle X-ray scattering (SAXS) and Förster resonance energy transfer (FRET), with experimentally guided all-atom molecular dynamics (MD) simulations, to construct structural ensembles of a 30-nucleotide RNA homopolymer (rU30) and a 30-nucleotide RNA heteropolymer with an A-/C-rich sequence. We compare the size, shape, and flexibility of the two different ssRNAs. While the average properties align with polymer-physics descriptions of flexible polymers, we discern distinct, sequence-dependent conformations at the molecular level that demand a more detailed representation than provided by polymer models. These findings emphasize the role of sequence in shaping the overall properties of ssRNA.

Graphical Abstract

*Correspondence: lp26@cornell.edu (L.P.), serdal@nyu.edu (S.K.).

AUTHOR CONTRIBUTIONS

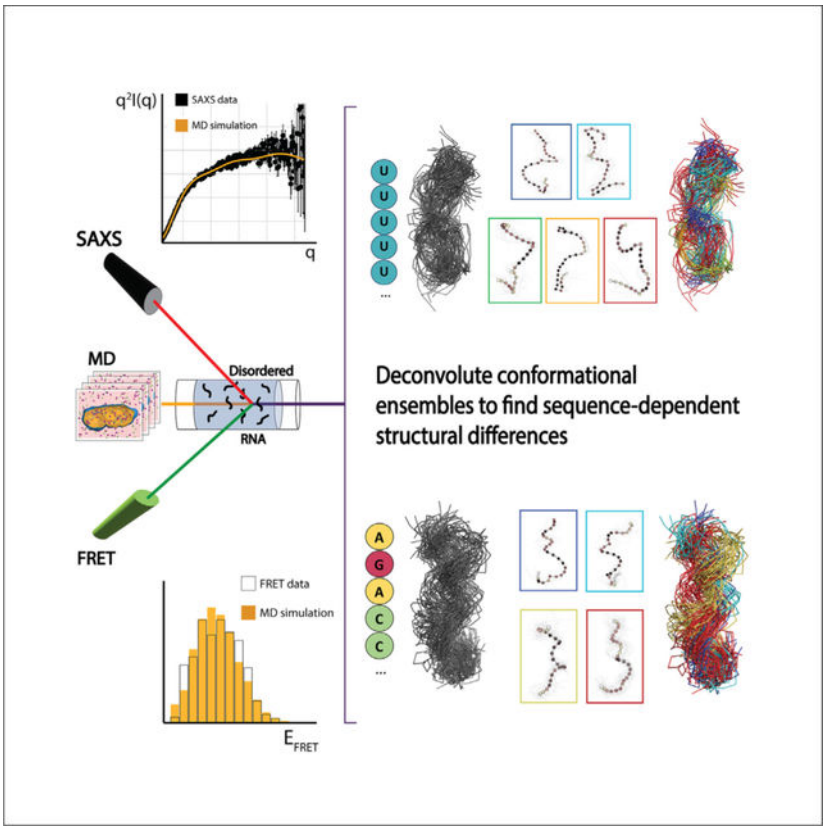
T.W. and S.A.P. carried out the experiments and analyzed the data. W.H. performed and analyzed the simulations. T.W. performed graph theoretic analysis. S.K. and L.P. conceived and designed the study. All authors contributed to integrating experimental and computational data, and to writing the manuscript. All authors edited and proofread the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrp.2024.102264.



Single-stranded RNAs (ssRNAs) are highly flexible molecules that play crucial roles in various biological processes, with their structure often influencing their function. Wang et al. compare a 30-nucleotide homopolymer and an A-/C-rich heteropolymer, finding that they have distinct conformational propensities that traditional polymer models may not capture.

INTRODUCTION

RNA macromolecules can adopt a wide variety of structures that facilitate their many functions. Some RNAs possess well-defined secondary and tertiary (folded) structures enabled by evolutionarily deliberate base-pairing schemes, such as the canonical examples of tRNA and ribozymes.[1–3] Other biologically important RNA regions lack base pairs, instead sampling a heterogeneous conformational ensemble in cells. Importantly, this disorder appears to be essential for function. Single-stranded RNA (ssRNA) regions are found in viral, long non-coding, and messenger RNA.[4–7] Disorder also enables variations in shape and surface electrostatics that create binding sites for proteins, small molecules, and other RNAs.[6,8–14] Finally, ssRNAs can phase separate to form both functional and pathological biomolecular condensates, where their ability to sample many configurations and interact at multiple sites allows for transient, multivalent contacts.[15–20] These roles are often specific to particular RNA sequences.

Despite the functional importance of disordered ssRNAs and the recognized connection between structure and function for most biomolecules, their sequence-dependent structures

remain elusive. Structural studies of disordered ssRNA are reminiscent of work on intrinsically disordered proteins (IDPs).[21,22] However, the more limited building blocks in the former, as well as the uniform negative charge of RNA chains, hint that the methods and properties used to study IDPs may not be directly applicable toward the study of ssRNA. Past efforts have revealed some biophysical properties of single-stranded nucleic acids.[23–25] Additional computational efforts are beginning to explore the conformational dynamics of ssRNA in greater detail, but they suffer from a lack of experimental validation.[26–30] Significant efforts have focused mainly on characterizing small (around 5-nucleotide) ssRNA motifs.[31–33] Longer (~30 nucleotides) homopolymeric chains have been explored in experiments and appear more complicated than Gaussian coils and worm-like chains; in particular, sequence-dependent base-stacking interactions in ssRNA present a challenge when applying polymer physics models.[34–36] Thus, while informative, single-value descriptors such as the Flory scaling parameter only offer a limited representation of flexible biopolymers.[37,38] Long heteropolymeric ssRNAs have rarely been considered structurally. As a first step, there is a need for approaches that characterize the conformations of dynamic ssRNA molecules.

A comprehensive characterization of ssRNA warrants combining information from multiple experimental and computational techniques to sample different aspects of the conformers within the structural ensemble. In this study, we employ small-angle X-ray scattering (SAXS) and single molecule Förster resonance energy transfer (FRET) to investigate the conformational properties of ssRNA. We determine the conformations of an adenine- (A) and cytosine (C)-rich 30-ribonucleotide heteropolymer and compare them with those of a homopolymer containing only uracil (U) bases. Although the two sequences show similar end-to-end distances via FRET measurements, SAXS suggests a more compact state for the mixed sequence. We use molecular dynamics (MD) simulations to generate conformational pools that satisfy the constraints of both SAXS and FRET and elucidate atomically detailed differences between the sequences. Validation as well as additional insight into helical content are provided by fluorescence correlation spectroscopy (FCS) and circular dichroism (CD). We utilize polymer-derived structural parameters to uncover the average structure; both shared features and variations among the sequences are investigated. Given that most of the elements within the conformational ensemble of highly flexible ssRNAs are concealed by random structures, we 3D classify the structural ensembles using network graphs that capture pairwise similarities between each conformation. This spectral clustering routine reveals sub-ensembles of well-defined conformations from the seemingly random ssRNA ensemble. These conformations are influenced by transient hydrogen bonding and base-stacking interactions, which play a crucial role in shaping the overall structure of the ssRNA molecules. This work represents a significant step toward developing a generalizable approach for generating and analyzing highly flexible sequences of ssRNAs. The outlined approach, systematic integration of multiple experimental and computational tools within a single framework, will assist in addressing how sequence influences the structure and, consequently, the function of highly flexible biomolecules.

RESULTS

rU30 and mixed-sequence ssRNA have distinct compaction states

We measured the sequence-dependent conformational preferences of our two ssRNA molecules using solution measurements. Interestingly, the two sequences display distinct structural properties when measured under identical solution (salt) conditions (Figure 1; Table 1). SAXS measurements report that the radius of gyration (R_g) of the mixed-sequence RNA was 16.4% lower than that of rU30, indicating greater compaction in the former. Further supporting this trend, the hydrodynamic radius (R_h) measured by FCS was found to be 13.5% lower in the mixed RNA compared to rU30 (Figure S1; Table S1). The R_g, describing the distribution of a macromolecule's mass, and the R_h, describing the aura of space that a macromolecule samples through diffusion, both provide complimentary information about the average size and compaction state of the structural ensemble. To describe the RNA-solvent interactions and the overall size of molecules, we employed the swollen Gaussian coil form factor to fit the Kratky plots. This model is commonly used to characterize disordered polymers.[39] Based on the Gaussian model, we determined a scaling factor of ($\nu \approx 0.55$) for rU30, suggesting a swollen random walk conformation. In contrast, the mixed sequence displayed $\nu \approx 0.49$, resembling a polymer in a theta solvent (Figure 1).

The experimental data (Table 1) suggest that the mixed-sequence RNA ensemble is on average more compact than the homopolymer rU30. For a deeper understanding of the differences between the two ensembles, we performed complementary single-molecule FRET measurements. The Förster radius was observed to be 2.45 Å greater for mixed ssRNA than rU30 (Figure S2; Tables S2–S4). This difference may result from the interaction of donor fluorophores with different bases. Indeed, despite a significantly greater apparent FRET efficiency measured for the mixed ssRNA, the increased Förster radius results in approximately equal end-to-end distances for both molecules (Table 1). The similarity of end-to-end distances, given differences in the R_g invites deeper investigation of the conformational ensembles, which we have accomplished by applying MD simulations.

Maximum entropy approach generates experimentally consistent ssRNA structural ensembles

The solution-based measurements provide quantitative evidence for sequence-dependent conformations of ssRNA. However, the information obtained from the polymer physics interpretation is insufficient to fully characterize the conformational differences and to understand how sequence dictates conformational propensities. To address these important questions, we leveraged a computational approach that utilizes atomistic details as well as explicit solvent RNA interactions. To ensure experimental consistency we sampled conformations committed to satisfy both SAXS and FRET constraints. The details of our workflow can be found in (Figure 2) and in the experimental procedures section.

MD simulations were conducted to generate conformational pools for each molecule using two different methods. The first involved running unbiased simulations, where conformational sampling was based on the empirical potential alone. In the second, we add the SAXS data to guide the conformational sampling. A third pool was also generated from

the latter by re-weighting the SAXS-guided conformational pools to ensure consistency with the FRET measurements (see experimental procedures and Figure 2 for detailed information). Full atomic coordinates of the conformational ensembles with their associated weights and a representative structure from each cluster are provided in the supplemental information for further analysis (Data S1 and Data S3).

The unbiased simulations yielded high errors, which can be attributed to insufficient sampling of flexible ssRNA conformations and imperfections in the force fields (Figures 3A and 3B). The application of SAXS-driven MD resulted in a notable reduction in the χ^2 value compared to brute force MD. Specifically, for rU30, the χ^2 value decreased from 15.46 to 0.96, and for the mixed sequence, it decreased from 8.14 to 2.42. (Figure 3B, blue vs. red). In our context, the χ^2 reports on the agreement between the experimental data and theoretical data computed from MD-derived conformers, where values closer to 1.0 indicate better agreement. Its full calculation is described in the supplemental experimental procedures section “computing SAXS from explicit-solvent MD simulations.” Despite the good agreement with the SAXS data, the SAXS-driven MD ensembles still exhibited significant deviation from the FRET experiments (Figure S3), indicating that SAXS and FRET observables provide mostly orthogonal constraints and unique information about the conformational states. Re-weighting the SAXS-driven MD pool using the FRET data exhibited good agreement with SAXS and FRET simultaneously (orange color) (Figures 3A–3D).

To further evaluate the accuracy of the pools, we computed each ssRNA’s ensemble-level R_g and R_h and compared them with independent measurements for cross validation. The agreement between these computed and measured parameters demonstrates the success of our methodology (Figures 3E and 3F).

rU30 and mixed-sequence RNA have different sizes and shapes

Having generated ssRNA conformational pools consistent simultaneously with experimental SAXS profiles and E_{FRET} histograms (Figure 3), we aimed to gain a more comprehensive understanding of the sequence-dependent global size and shape of the ssRNAs and their interactions with the solvent. For that purpose, we revisited the Scaling Law analysis introduced in Figure 1. Consistent with the previous analysis, we observed that the computational ensemble generated for rU30 shows a higher Flory exponent (ν) than for the mixed sequence, suggesting more swollen chain conformations for the homopolymer. While the analysis yielded the same trend, ν values obtained directly from simulations and those derived by fitting the SAXS curve to the Gaussian chain model are not the same (Figure 1 vs. Figure 4A). Direct computation of ν from simulations via scaling law relationships, exhibited $\nu \approx 0.76$, while the fitting procedure of the SAXS curve resulted in $\nu \approx 0.55$ for rU30. Similarly, for the mixed sequence, values obtained by the two approaches deviated significantly. As the measured SAXS curves are in excellent agreement with those computed from simulation, these deviations likely arise from the fact that the first method assumes a Gaussian coil homopolymer chain, whereas the second method does not.

To systematically characterize the conformations sampled by each sequence, we projected the RNA conformations onto tensors of the R_g and persistence and contour lengths (Figure 4). The projected parameters are represented as heat maps, illustrating their distribution within the conformational pool (Figure 4). We compared the shape and size of rU30 with the mixed sequence.

Based on this analysis, we conclude that the two ssRNA sequences explore different regions of conformational space, and rU30 exhibited higher contour and persistence lengths. We further found that rU30 and Mixed RNA differ in their distributions of Δ and S parameters (which are defined in the supplemental experimental procedures section “additional analyses of ssRNA structural ensembles” and describe the deviation of the ensemble from being spherical), further elucidating their different structural behaviors (Figures 4D and 4E vs. Figures 4H and 4I).

A-/C-rich mixed RNA is more helical than rU30

After determining the ensemble behaviors of both ssRNA constructs, we delved deeper into their unique properties and the local geometries that govern their overall ensembles. Firstly, we aimed to understand whether any local structures are present within the mixed-sequence ensemble, which contains many A and C bases. Consistent with previous studies, we observe that rU30 exhibits Gaussian coil behavior with limited secondary structure. Meanwhile, this study newly reveals that the C-rich sequence shows a strong propensity for helical structures. To our knowledge, there are no comparable studies of rC30. We further conducted circular dichroism (CD) experiments on the mixed ssRNA, using rU30 and rA30 as limiting cases representing mostly unstructured and structured helical strands (Figure 5A). Our results indicate that the degree of helicity in the mixed sequence is intermediate between these two extremes (Figure 5A). To further quantify this observation, we computed the orientational correlation function (OCF) between the chain phosphates in the conformational pools.[40] As depicted in Figure 5B, the OCF profiles computed for the mixed sequence and rU30 exhibit oscillatory behavior reflecting orientational order. The shorter periodicity observed in the mixed sequence reflects stronger correlations between local bases compared to rU30.

To provide additional context for the higher correlations observed in mixed sequence, we computed the tortuosity (degree of twist) and the amount of base-stacking events (Figures 5C and 5D). Our analysis reveals that the mixed ssRNA exhibits a larger tortuosity and a higher number of base-stacking events compared to rU30 (Figure 5D). This increase in stacking events is likely attributed to the greater prevalence of purine (A) bases in the mixed sequence. However, we did not observe significant differences in the base-base stacking preferences between the A-rich and C-rich regions of the mixed sequence (Figure S5), suggesting C bases also have a propensity to stack, supporting previous studies on short ssRNA.[41] The C-rich region of the mixed sequence also displayed a higher level of order compared to the homopolymeric rU30 in our study, as evidenced by its higher oscillatory periodicity (Figure S5A and Figure 5B).

We also utilized a contact map analysis to visualize the local structures of the two conformational ensembles (see supplemental experimental procedures section “contact map

analysis"). As shown in Figures S6A and S6B, A wider band along the diagonal direction indicates a higher number of contacts between adjacent/neighboring residues as observed in the mixed RNA compared to rU30. This observation is consistent with the greater helicity exhibited by the mixed RNA in comparison to rU30. Conversely, rU30 demonstrates frequent hairpin loop formation events at both ends of the chain (Figure S6).

Overall, these findings indicate a higher degree of helicity in the A-/C-rich mixed-sequence ssRNA, which is likely facilitated by an increased number of intra-strand interactions among the more diverse set of bases. Among these interactions, base stacking plays a prominent role and appears to be more prevalent in the mixed sequence. The helical behavior of the mixed ssRNA falls between the extremes represented by rU30 and rA30, suggesting that the inclusion of A and C bases enhances the occurrence of base stacking interactions compared to rU30, although not to the same extent as a sequence composed entirely of A bases.

Specific RNA conformations are sequence dependent

Ensemble averaging, which provides an average picture, has proven useful for benchmarking our findings against bulk measurements. However, it does not provide insights into how individual molecules within each sequence compare with one another. To explore in greater detail how the differences at the sequence level manifest, and to gain a deeper understanding of the structural underpinnings of our ensemble-level polymer metrics, we conducted 3D classification and averaging to determine sub-families of structures within the RNA ensembles. This allowed us to decompose each ensemble into a handful of characteristic conformations for tangible visualization in real space. To do so, we constructed network graphs based on spatial proximity followed by spectral clustering to identify distinct groups or subfamilies of similar structures within each ensemble (Figure 6); see experimental procedures for details. As a metric of spatial variability within each sub-family, we used the spatial variance at each phosphorus position, $\delta_{Ri} = \sqrt{\delta_{Xi}^2 + \delta_{Yi}^2 + \delta_{Zi}^2}$ where each component of the variance is defined as $\delta_\gamma = E\left[(\gamma - \mu_\gamma)^2\right]$ where $E[\,.\,]$ is the expectation value and μ_γ is the mean value of component γ. The spatial variance is found to be less than 50 $Å^2$ across the majority of the RNA backbones and the individual conformers are visually similar to the average conformers (Figure 6), indicating that the individual structures align well with each other and share similarities within their respective subfamilies. Higher spatial variability is observed in certain regions, particularly at the ends of the RNA molecules where the backbone can exhibit diverse orientations.

By examining the unique structural classes identified using our approach, we observe the impact of enhanced helicity of the A-/C-rich mixed-sequence ssRNA compared to rU30. The greater orientational variability along the backbone of the mixed RNA contributes to its lower persistence length in comparison to rU30. Additionally, the more pronounced helicity in the mixed RNA leads to its conformations having less radial deviation from its long axis, in contrast to rU30, which exhibits wider bends along the backbone due to longer featureless stretches. Furthermore, we observe an increased prevalence of strand fold-overs at the ends of the ssRNA chain in rU30, as reflected in the contact maps (Figure S6). The visualization also provides a real-space explanation for why the rU30 and mixed RNA conformations display comparable end-to-end distance distributions despite the lower R_g observed in mixed

RNA compared to rU30. Specifically, the ensemble of rU30 structures may exist within a larger sphere of mass as the molecules fluctuate and tumble in solution, leading to greater experimental values of R_g and R_h obtained through time-averaged measurements.

Compared to rU30, the structural ensemble of the mixed-sequence RNA exhibits a higher level of connectivity, as evidenced by the approximately 2-fold increase in closeness centrality and significantly greater degree centrality (Figure S7). In graph space, this indicates that the nodes in the mixed RNA ensemble are closer together, and there are more edges connecting the nodes. Specifically, the graph of mixed RNA contains 810 edges, compared to 506 edges in the rU30 graph, despite having the same number of nodes. This suggests that the structures within the mixed RNA ensemble are more similar to each other, while the rU30 homopolymer exhibits greater structural variability. This observation is further supported by the presence of outlying nodes in the rU30 graph, which primarily represent highly extended structures (Figure S8). In contrast, there are no outliers in the mixed RNA graph, indicating the absence of highly extended structures beyond the identified classes.

While graph theory enabled us to examine the connectivity of conformations according to their structural similarity, we also employed MD trajectories to characterize the network of kinetically accessible conformational transitions and time evolution of the structural metrics (Figures S9–S13 and S14–S18; supplemental experimental procedures section “additional analyses of ssRNA structural ensembles” for the details of connectivity network analysis). In addition, we provide a movie showcase the dynamics of each sequence in the supplemental information (Video S1). Trajectory analysis of rU30 reveals a dynamic exchange of conformations centered around four major structures, unveiling a highly interconnected network of transitions accessible within the simulation timescale. All conformations exhibit similar tendencies in base stacking, while measures of tortuosity and chain size (quantified by radius of gyration and end-to-end distance) display notable variations. Predominant clusters encompass both helical and random structures, showcasing collapsed as well as highly extended polymer conformations. Comparatively, mixed-sequence structures demonstrate heightened base stacking and helical content in contrast to rU30. Notably, in the mixed sequence, we observe a prominent conformational switch between bent and straight helical geometries. The conformational heterogeneity of mixed sequence is facilitated by excursions to high-energy states through clusters 3 and 4 (see Figure S14).

DISCUSSION

Structurally characterizing a highly disordered macromolecule, particularly ssRNA, requires experimental measurements that capture the full range of conformational variability of the molecule. Such measurements often come at the cost of resolution due to spatial averaging. In this study, we aim to overcome this limitation and venture beyond coarse-grained polymer models such as the Gaussian and worm-like chains which are commonly used to interpret experimental data. Instead, we examine the specific conformers of ssRNA in detail. To do this, we employ an integrated approach, generating all-atom ssRNA conformers by MD in a thermodynamically accurate maximum entropy framework and employing a combination of SAXS, FRET, FCS, and CD. These methods, when synthesized, provide a definitive

representation of the ssRNA conformers likely present in the solution at near-physiological ionic strength. Overall, the use of unique experimental metrics compounds the amount of useful information and results in a consistent description of the disordered system: MD provides physically plausible all-atom conformations, SAXS provides the distribution of paired distances and dictates the overall compaction state of the actual structural ensemble through the R_g, FCS validates this latter measurement by providing hydrodynamic radii, FRET offers end-to-end dimensions, and CD evaluates helical content. Recent characterizations of both disordered nucleic acids and IDPs also combine computational and experimental approaches to determine accurate structural ensembles, demonstrating the utility of an integrated approach toward characterizing disordered macromolecules.[21,22,42]

We apply our integrated approach to compare the solution structures of rU30 and a 30mer A-/C-rich mixed sequence (Figure 1A). As most biological RNA sequences contain a combination of purines and pyrimidines, studying mixed sequences is desirable. Tracts of repeated bases are often observed in physiologically relevant ssRNAs, such as those that are U-rich,[4,5,43,44] A-rich,[8,10] and C-rich sequences.[6,9] From our results, we find unique structural behaviors of A, C, and U tracts, developing the notion that base identity influences the average shape and size of the RNA molecules. Furthermore, it governs unique structural tendencies among sequences, which are likely to impact their functions.[45] This study extends previous descriptions of ssDNA and ssRNA, which ascribe the behavior of pyrimidine homopolymers like rU30 to resemble an unstructured Gaussian coil.[33,34]

We find that the structures of ssRNA are by no means random. We first notice the discrepancy in ν values of the ssRNA derived from fitting a swollen Gaussian coil form factor to the SAXS data (Figure 1B) vs. computing the internal scaling plots ($|i - j|$ vs. R) of the conformers in the pool directly (Figure 5A), which result in ν values that have the same trend but are 30–40% higher. This suggests that the Gaussian coil does not accurately model ssRNAs, supporting the effort to determine all-atom pools. Interestingly, the mixed sequence appears to have two distinct structural behaviors corresponding to its two distinct purine- (A) and pyrimidine (C)-rich sequence fragments (Figure S19). The C-rich portion has a more pronounced twist, while the A-rich segment appears less tortuous—these differences may reflect different types of helicity. These structural preferences are especially perceptible in cluster 0, and cluster 3 for the mixed sequence in Figure 6. Within the mixed sequence, the backbone conformation of the C-rich tract is different from the A-rich tract, and the presence of both tracts offers a glimpse as to why mixed ssRNA differs from both rU30 and rA30 in its secondary structure (Figure 5A). We note also the presence of G bases in the mixed ssRNA sequence, whose structural contribution is difficult to deduce as they are sparsely present and not in tracts. G bases are known to have a strong $\pi - \pi$ interaction potential, so it is conceivable that they would contribute to the elevated base stacking behavior in the mixed sequence.[46] The distinct behavior of A-/C-rich mixed ssRNA sequence within the same strand acknowledges the pivotal effect of sequence on ssRNA structure and the potential contribution to ssRNA design that our research offers.

Finally, this study combines data from SAXS and FRET to refine the ssRNA structural ensembles. Past work has questioned their compatibility, due to discrepancies between

Flory ν parameters obtained from both techniques and the observation that FRET dyes may artificially collapse some proteins.[47–49] In the current study, we observe that E_{FRET} reweighting of the ssRNA structural ensembles did not result in appreciable modifications in the R_g of the structural ensemble outside error margins (27.2 ± 3.4 Å to 25.8 Å for rU30 and 24.7 ± 2.2 Å to 23.3 Å for mixed). We also measured a significant difference in the R_g between rU30 and mixed ssRNA without an appreciable change in the R_{EE} (Figures 1C and 1E), demonstrating that the two metrics from SAXS and FRET are not necessarily coupled. Furthermore, we readily obtained ssRNA structural ensembles that agree simultaneously with SAXS and FRET. The results of this study demonstrate the possibility of integrating SAXS and FRET in a congruous manner, and the benefit of FRET providing an additional readout of the R_{EE} distribution. Similar conclusions were also arrived at in Gomes et al.[21] and Fuertes et al.[50]

Given that our method provides atomically detailed representations of these flexible molecules, we would be remiss to not address possible structural changes resulting from the presence of the dyes. To this end, we compare the structures determined by SAXS-driven MD with the FRET-data refined ensembles (Figure S20), along with a description of their similarities and differences. The salient features that characterize the ensemble, specifically the OCF and base stacking (Figure S21) deviate only slightly between the two differently derived ensembles; the distinct features and conformations that define ensembles for each construct are consistently reported by both methods. Differences between the ensembles mainly reflect the loss of the most extended structures from the SAXS pools, consistent with the absence of zero FRET (highly extended) states from our FRET measurements (Figure S3), and structures reveal no obvious dye-dye or dye-backbone interactions (Figure S20). This is, of course, subject to any limitations imposed by the sampling and the force fields used to create the pools. We also compared structural parameters obtained from unbiased and biased simulations. The average structural properties across three ensembles (unbiased, SAXS-driven, SAXS-FRET refined) were similar for both sequences, particularly in rU30. Differences between simulations were primarily attributed to variations in local base stacking, with brute force MD simulations showing an over-stabilization of base stacking, especially noticeable in mixed sequences. Biased simulations induced modifications to the conformational ensemble compared to unbiased simulations, particularly affecting global chain properties. FRET reweighting had minimal impact on the global structural parameters of rU30 but led to higher deviations in mixed sequences, with parameters S and Δ showing the largest differences between ensembles. Cross-validation of the ensemble's hydrodynamic radius revealed similar performance between the pools, suggesting that the conformations contributing to the overall behavior of this global property remain similar across the methodologies (Figure S22).

While we have modeled the conformational ensembles of ssRNA in detail, we must be careful not to overinterpret the results. We confidently provide the backbone conformations but further structural detail, for example orientation of the nucleotide bases, requires additional experimental validation. To obtain this level of detail, techniques such as NMR on short stretches of ssRNA molecules could be integrated into a future workflow. To better evaluate the influence of sequence on the structural dynamics of ssRNAs, sequences

with varying nucleobase compositions are desirable. However, such high-throughput studies are quite demanding with our current methodology, limiting its use to more targeted sequences. Moreover, ongoing improvements to MD force fields, especially those directed toward modeling electro-statically dense nucleic acids, will allow more realistic ssRNA conformations to be sampled and may also allow for the explicit account of phenomena such as base stacking.[51] Surpassing this hurdle would allow for the additional detailed modeling of constructs such as rA30 that are known to undergo extensive π–π stacking.[34]

In summary, by integrating multiple experimental techniques with all-atom MD simulations, we have constructed plausible structural ensembles of a ssRNA homopolymer (rU30) and heteropolymer (A-/C-rich mixed sequence) that agree simultaneously with SAXS, FRET, and FCS measurements. The two ssRNAs are distinct in their conformational landscapes, with the increased compaction of the mixed sequence stemming from enhanced base stacking leading to greater helicity. Through graph theory, we developed a method of 3D class averaging of the disordered structures and identified 4–5 main backbone conformations that describe the structural variability in each of the ensembles. The greater helicity in the mixed sequence relative to rU30 can be seen in real space through these backbone conformations, and distinct behavior arises in both the A- and C-rich halves of the molecule. Overall, we demonstrated that not all disordered RNA molecules are alike—disparate conformational landscapes and structural features are present between two distinct sequences and a generalized random walk model can fail to capture these subtle differences. The framework developed in this study is readily applicable to the detailed study of more ssRNA sequences, an important advance given the physiological relevance of these molecules.

EXPERIMENTAL PROCEDURES

ssRNA sample preparation

We identified a 30-nucleotide heteropolymer sequence with minimal secondary structure and intermolecular hybridization. This construct has the sequence 5′-AA GAAUAAAAGAG AAGCCACCCCACCCAGA-3′ and is referred to as “mixed” ssRNA. We also studied a 30-nucleotide poly-uridine RNA, which is amenable to solution experiments and is expected to have minimal structure.[33,34] While designing the mixed sequence, we limited our use of U bases to better contrast with rU30. To avoid any secondary and tertiary interactions, we also limited the use of repeated G bases to prevent quadruplex formation. We focused on A and C tracts due to their physiological relevance and unique interactions (see discussion). Unlabelled RNA samples for SAXS and RNA doubly labeled with 5′ cyanine-3 (cy3) dye and 3′ cyanine-5 (cy5) were purchased from Integrated DNA Technologies (Coralville, IA, USA). For SAXS experiments, unlabelled RNA was desalted and buffer exchanged into 100 mM NaCl, 10 mM MOPS pH 7.0, 20 μM EDTA using 3k MWCO 0.5mL centrifugal filters. For FRET measurements, doubly labeled RNA was diluted to 300p.m. concentration (in the single molecule regime) in the same 100 mM NaCl buffer used for SAXS. FRET samples were prepared in borosilicate chambered coverglass (Nu-Tek, Aberdeen MD, USA) that was passivated overnight at room temperature in 600 μL of 1 % casein to prevent non-specific

surface association. Before measurement, all RNA samples were annealed at 100 mM NaCl by heating the sample to 90°C for 3 min and snap cooling at 4°C for 20 min.

Solution X-ray scattering

SAXS measurements on the heteropolymeric RNA were performed on a BioXolver laboratory X-ray source (Xenocs, Grenoble France), averaging 10 frames, each with 120 s exposure time for each sample. The images were azimuthally integrated such that the scattering intensity I(q) is plotted as a function of the scattering vector q: $q = \frac{4\pi\ sin(\theta)}{\lambda}$, where λ is the wavelength of incident X-ray radiation and θ is one-half the scattering angle. Data points at q < 0.05 Å $^{-1}$ were acquired for multiple RNA concentrations. These curves were linearly extrapolated to zero concentration to eliminate inter-particle effects. SAXS data on homopolymeric poly-uridine RNA were acquired as previously described.[34] The radius of gyration (R_g) was determined by employing the Guinier approximation of Gaussian scattering at $qR_g < 1.3$ (Figure S23).

$$I(q) = I(0)e^{-q^2R_g^2/3}. \quad \text{(Equation 1)}$$

We also performed a dimensionless Kratky transformation of the axes to visualize geometric differences between the two disordered RNA constructs

$$(q, I(q)) \rightarrow \left(qR_g, I(q)\frac{q^2R_g^2}{I(0)}\right). \quad \text{(Equation 2)}$$

This approach normalizes out size information and amplifies differences at q values above those in the Guinier regime ($qR_g > 1.3$), allowing for a comparison of the overall shapes of the two constructs. Data reduction and processing were performed in BioXTAS RAW.[52] In addition, the SAXS data were fit to the semi-analytic form factor of a swollen Gaussian chain to obtain experimental estimates for the Flory scaling parameter ν, described previously.[39,53]

Förster resonance energy transfer

Dual color FRET measurements were performed using alternating laser excitation.[54] Photon traces were recorded in 30-x 30-s segments. Burst analysis was performed using in-house software written in MATLAB (Mathworks, Natick, MA, USA). Photon count channels were partitioned into 1 ms bins, intentionally longer than the dwell time of the molecules in the confocal volume (measured by FCS to be about 200 μs, see supplemental experimental procedures). Fluorescence intensity (I) was determined as the number of photons per time interval and was background corrected by subtracting the average I. The apparent FRET efficiency E_{app} was computed as

$$E_{app} = \frac{I_{A,532} - \alpha I_{D,532} - \delta I_{A,640}}{\gamma I_{D,532} + I_{A,532} - \alpha I_{D,532} - \delta I_{A,640}}.$$

(Equation 3)

Here, $I_{x,y}$ corresponds to the fluorescence intensity of x=cy3 donor (D) or cy5 acceptor (A) under y = 532nm (excites donor) or 640nm (excites acceptor) laser excitation. Along the second axis, the apparent stoichiometry ratio (S_{app}) of fluorophores was computed as

$$S_{app} = \frac{I_{A,532} - \alpha I_{D,532} - \delta I_{A,640}}{\frac{1}{\beta} I_{A,640} + \gamma I_{D,532} + I_{A,532} - \alpha I_{D,532} - \delta I_{A,640}}.$$

(Equation 4)

where α, δ, β, and γ are correction coefficients accounting for leakage of donor fluorescence, direct acceptor excitation, and differences in excitation and detection efficiencies of the optical system.

A dual channel burst search algorithm was employed to remove artifacts from fluorophore bleaching and blinking.[55] This left a 2D $E_{app} - S_{app}$ histogram with a single population at $S \in [0.3, 0.7]$ (the regime for doubly labeled molecules), which was fit to a single Gaussian function (Figure S2). Data from six independent measurements were pooled, and the mean E_{app} value was determined from the measurement series, with errors estimated by the standard error in the mean (Table S2).

To convert the measured apparent FRET efficiency into the absolute end-to-end ssRNA distance (R), we used the relation

$$< E_{app} > = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}.$$

(Equation 5)

Here, $< E_{app} >$ is given in Equation 3 and R_0 is the experimentally determined Förster radius. A more detailed account of the fluorescence instrumentation and experimental controls, as well as the experimental setups of FCS and CD, are provided in the supplemental experimental procedures and Figure S24.

Computational modeling of RNA structural ensembles based on SAXS and FRET data

We conducted all-atom brute force MD simulations, followed by multi-replica SAXS-driven MD simulations and ensemble refinement to derive pools consistent with experimental data. Each of the steps in the aforementioned approach is elaborated upon in the supplemental experimental procedures. Below, we provide a concise summary of our experimentally guided computational methodology for studying RNA structural ensembles.

First, we placed the RNA molecules in a simulation box ($12 \times 12 \times 16\ nm^3$) with explicit water and ions mimicking experimental conditions (Table S5). Molecular dynamics simulations were conducted using GROMACS 2018.5[56] with the Amber force field, incorporating χOL3 correction[57] representing the RNA. A 200 ns-long brute force MD simulation was conducted at 300K and sampled conformations were clustered using RMSD to create four replicas representing different regions of the conformational states.

Subsequently, we performed multi-replica SAXS-driven MD simulations (see Figures S25 and S26), which employed a hybrid energy function $E_{\text{hybrid}} = E_{\text{FF}} + E_{\text{SAXS}}$, where the energy from the MD force field (E_{FF}) was combined with a time-dependent penalty term for each replica system (E_{SAXS}), as defined in[58,59]:

$$E_{SAXS}(\mathbf{R}_1, \ldots, \mathbf{R}_N, I_{exp}, t) = \alpha(t) k_c N^{\Omega} \frac{k_B T}{n_q} \sum_{i=1}^{n_q} \frac{\left\{\left[\bar{I}_{comp}(q_i, \mathbf{R}_1, \ldots, \mathbf{R}_N, t)\right] - \left[I_{exp}(q_i)\right]\right\}^2}{\sigma^2(q_i)} \quad \text{(Equation 6)}$$

Here, $\mathbf{R}_1, \ldots \mathbf{R}_N$ represents the coordinates of the replicas, $\alpha(t)$ is a time-dependent function that gradually couples SAXS data to the Hamiltonian, and n_q is the number of scattering vectors of q. Here, the factor $k_c N^{\Omega}$ adjusts the weight of the SAXS potential E_{SAXS} relative to the force field potential E_{FF} while the summation measures the difference between the replica-averaged scattering amplitude computed from simulations, $\bar{I}_{comp}(q_i, \mathbf{R}_1, \ldots, \mathbf{R}_N, t)$, and the experimental scattering amplitude $I_{exp}(q_i)$. The difference is scaled by possible uncertainties. These uncertainties were modeled as an independent sum of experimental, statistical, and systematic errors, $\sigma^2(q_i) = \sigma^2_{exp}(q_i) + \sigma^2_{comp}(q_i) + \sigma^2_{buffer}(q_i)$. All snapshots with $\chi^2 < 2.5$ selected from the 400-ns-long simulation were then used to generate a conformational pool for each sequence. Details on the settings of the SAXS-driven MD simulation can be found in the supplemental experimental procedures.

Multi-replica SAXS-driven MD simulations yield SAXS-consistent conformations that follow a Boltzmann distribution within the modified potential energy function, with the probability distribution defined as $P_0(\mathbf{X}) \propto e^{-\beta V(\mathrm{X})}$, where X represents the sampled states and $\beta = 1/k_B T$. To minimally bias the probability distribution and generate an ensemble consistent with FRET experiments, we employed the Maximum Entropy reweighting method (see Figures S27 and S28).[60–63] In this approach, the ensemble is reweighted by maximizing Shannon's entropy between the prior and posterior distributions, given by $S[P][P_0] = -\int d\mathbf{X} P(\mathbf{X}) \ln \frac{P(\mathbf{X})}{P_0(\mathbf{X})}$, while satisfying the following conditions:

$$\begin{cases} P_{ME}(\mathbf{X}) = arg\ max\ S[P][P_0] \\ \int d\mathbf{X}\mathbf{f}(\mathbf{X})P(\mathbf{X}) = \mathbf{f}^{exp} \\ \int d\mathbf{X}P(\mathbf{X}) = 1 \end{cases}$$

(Equation 7)

where $\mathbf{f}(\mathbf{X})$ and $\mathbf{f}^{exp}$ are the FRET energy distributions computed from simulations and measured in experiments. To achieve this, we initially computed the FRET efficiencies for each conformation generated by SAXS-driven MD using the HandyFRET program,[64] and its particular implementation to RNA is described in the supplemental experimental procedures ("computing FRET efficiency from MD simulations").[51] Subsequently, we fit the data into a Gaussian function for both the simulated and experimental FRET data. To solve Equation 7, we used the code provided in ref. [29] after modifications.

Data analysis and visualization

We used the ensemble-weighted conformational pool to characterize both the global and local structural properties of the sequences. We computed various structural parameters, including Flory scaling law, global shape parameters (R_g,Δ,S), end-to-end, persistence and contour lengths (R_{EE}, L_P, L_C), the orientational correlation function of base pairs (OCF), and the base stacking propensity for each sequence. Furthermore, spectral clustering was employed to classify and visualize the distinct structural ensembles.

Namely, we calculated the scaling properties, represented as $R_{|i-j|} \propto |i-j|^{\upsilon}$, where υ denotes the scaling exponent, and $|i-j|$ represents the distance between phosphate groups of residue pairs.[65] To assess the global shape of the RNAs, we utilized the eigenvalues of the tensor of inertia, from which we compute R_g, Δ, S (see supplemental experimental procedures for details). Here, R_g quantifies the overall size of the chain, the parameter Δ quantifies the degree of deviation from spherical symmetry, and the parameter S characterizes whether the chain's shape is prolate ($S > 0$) or oblate ($S < 0$) ellipsoid. To determine the persistence length, we fit R_g to the worm-like chain model. Base stacking is monitored using Barnaba package[66] with default settings. We also utilized contact map analysis to elucidate local structural features. Finally, to evaluate structural correlations, we computed the orientation correlation function between base pairs (OCF)[34,40] A detailed description of each observable is described in the supplemental experimental procedures.

To visualize the disordered ssRNA conformational ensembles in real space, we employed a class averaging method based on spectral clustering. A visual overview of the method is shown in Figure S29. First, we aligned the select conformations using PyMOL v2.4.0 (Schrodinger, LLC) and computed the root-mean-square distance (RMSD) between each pair of structures after alignment.

These values were then input into a $N_{structures} \times N_{structures}$ symmetric matrix of pairwise RMSD distances. A binary, symmetrical adjacency matrix showing which structures are connected (sufficiently similar in space) was computed by setting an RMSD threshold of 7 Å and declaring that a pair of structures j is connected if $RMSD_j < RMSD_{thresh}$. The structures were then mapped to a graph space by using the adjacency matrix, containing information on which pairs of structures (nodes) are connected, as an input into the spectral clustering algorithm.[67] Briefly, we computed the Laplacian

$$L = D^{-1/2}AD^{-1/2}$$

(Equation 8)

where A is the adjacency matrix and D is a diagonal matrix determined from A, containing the number of connections at each node. The first k eigenvectors were then sampled from L to represent the number of subgroups present in the data. These k subgroups were identified through K-means clustering using MATLAB. The closeness and degree centrality metrics were utilized to quantify the degree of connectivity within the structural ensemble networks. Upon classifying the RNA ensembles into sub-families, the structures were coarse grained by sampling the positions of the phosphorus atoms along the backbone, further aligned by RMSD minimization, and averaged to determine the characteristic 3D chain conformation of each sub-family.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Serdal Kirmizialtin (serdal@nyu.edu).

Materials availability

This study did not generate new unique materials or reagents.

Data and code availability

All data related to the reported analyses are included in the paper and supplemental information. Atomic coordinates of the structural ensembles are provided in the supplemental information, and SAXS data are available on the Small Angle Scattering Biological DataBank (SASBDB) through the identifiers SASDSN5 and SASDFK9. Code for class averaging via spectral analysis is available at https://github.com/TongGeorgeWang/CASA-ToDiMo.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Dr. Dave Mathews for providing the sequence for the mixed ssRNA, Drs. Richard Gillilan and Qingqiu Huang for support with X-ray scattering experiments, and Drs. Warren Zipfel and Alex Plumridge for helpful discussions regarding FCS and FRET measurements. Computational research was carried out on the High Performance Computing resources at New York University Abu Dhabi. This work is based on research conducted at the Center for High-Energy X-ray Sciences (CHEXS), which is supported by the National Science Foundation (BIO, ENG and MPS Directorates) under award DMR-1829070, and the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by award 1-P30-GM124166-01A1 from the National Institute of General Medical Sciences, National Institutes of Health (NIH), and by New York State's Empire State Development Corporation (NYSTAR). This work benefited from measurements on a BioXolver X-ray Source from Xenocs, supported by NIH grant S10 OD028617. S.K. and W.H. are supported by AD181 faculty research grant and REF Grant 371. Work in the Pollack lab is supported by NIH grant R35-GM122514. T.W. acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC). Cette recherche a été financée en partie par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG).

REFERENCES

1. Cech TR, and Steitz JA (2014). The noncoding rna revolution—trashing old rules to forge new ones. Cell 157, 77–94. [PubMed: 24679528]
2. Nguyen K, and Whitford PC (2016). Capturing transition states for trna hybrid-state formation in the ribosome. J. Phys. Chem. B 120, 8768–8775. [PubMed: 27479146]
3. Levi M, Walak K, Wang A, Mohanty U, and Whitford PC (2020). A steric gate controls p/e hybrid-state formation of trna on the ribosome. Nat. Commun 11, 5706. [PubMed: 33177497]
4. Hackbart M, Deng X, and Baker SC (2020). Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. Proc. Natl. Acad. Sci. USA 117, 8094–8103. [PubMed: 32198201]
5. Schnell G, Loo Y-M, Marcotrigiano J, and Gale M Jr. (2012). Uridine composition of the poly-u/uc tract of hcv rna defines non-self recognition by rig-i. PLoS Pathog 8, e1002839. [PubMed: 22912574]
6. Makeyev AV, and Liebhaber SA (2002). The poly(c)-binding proteins: a multiplicity of functions and a search for mechanisms. RNA 8, 265–278. [PubMed: 12003487]
7. Teijaro JR, and Farber DL (2021). Covid-19 vaccines: modes of immune activation and future challenges. Nat. Rev. Immunol 21, 195–197. [PubMed: 33674759]
8. Tom JKA, Onuchic PL, and Deniz AA (2022). Short polya rna homopolymers undergo mg^{2+}-mediated kinetically arrested condensation. J. Phys. Chem. B 126, 9715–9725. [PubMed: 36378781]
9. Nakamoto MY, Lammer NC, Batey RT, and Wuttke DS (2020). hnrnpk recognition of the b motif of xist and other biological rnas. Nucleic Acids Res 48, 9320–9335. [PubMed: 32813011]
10. Torabi S-F, Chen Y-L, Zhang K, Wang J, DeGregorio SJ, Vaidya AT, Su Z, Pabit SA, Chiu W, Pollack L, and Steitz JA (2021). Structural analyses of an rna stability element interacting with poly(a). Proc. Natl. Acad. Sci. USA 118, e2026656118. [PubMed: 33785601]
11. Auweter SD, Oberstrass FC, and Allain FH-T (2006). Sequence-specific binding of single-stranded rna: is there a code for recognition? Nucleic Acids Res 34, 4943–4959. [PubMed: 16982642]
12. Shen C, Zhang D, Guan Z, Liu Y, Yang Z, Yang Y, Wang X, Wang Q, Zhang Q, Fan S, et al. (2016). Structural basis for specific single-stranded rna recognition by designer pentatricopeptide repeat proteins. Nat Commun 7, 11285. [PubMed: 27088764]
13. Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang HW, et al. (2013). Structural basis for the modular recognition of single-stranded rna by ppr proteins. Nature 504, 168–171. [PubMed: 24162847]
14. Messias AC, and Sattler M (2004). Structural basis of single-stranded rna recognition. Acc. Chem. Res 37, 279–287. [PubMed: 15147168]
15. Rhine K, Vidaurre V, and Myong S (2020). Rna droplets. Annu Rev Biophys 49, 247–265. [PubMed: 32040349]

16. Riback JA, Zhu L, Ferrolino MC, Tolbert M, Mitrea DM, Sanders DW, Wei M-T, Kriwacki RW, and Brangwynne CP (2020). Composition-dependent thermodynamics of intracellular phase separation. Nature 581, 209–214. [PubMed: 32405004]
17. Boeynaems S, Holehouse AS, Weinhardt V, Kovacs D, Van Lindt J, Larabell C, Van Den Bosch L, Das R, Tompa PS, Pappu RV, and Gitler AD (2019). Spontaneous driving forces give rise to protein-rna condensates with coexisting phases and complex material properties. Proc. Natl. Acad. Sci. USA 116, 7889–7898. [PubMed: 30926670]
18. Guo Q, Shi X, and Wang X (2021). Rna and liquid-liquid phase separation. Noncoding. RNA Res 6, 92–99. [PubMed: 33997539]
19. Chen H, Cui Y, Han X, Hu W, Sun M, Zhang Y, Wang P-H, Song G, Chen W, and Lou J (2020). Liquid–liquid phase separation by sars-cov-2 nucleocapsid protein and rna. Cell Res 30, 1143–1145. [PubMed: 32901111]
20. Wang B, Zhang L, Dai T, Qin Z, Lu H, Zhang L, and Zhou F (2021). Liquid–liquid phase separation in human health and diseases. Signal Transduct. Targeted Ther 6, 290.
21. Gomes G-NW, Krzeminski M, Namini A, Martin EW, Mittag T, Head-Gordon T, Forman-Kay JD, and Gradinaru CC (2020). Conformational ensembles of an intrinsically disordered protein consistent with nmr, saxs, and single-molecule fret. J. Am. Chem. Soc 142, 15697–15710. [PubMed: 32840111]
22. Goretzki B, Wiedemann C, McCray BA, Schäfer SL, Jansen J, Tebbe F, Mitrovic SA, Nöth J, Cabezudo AC, Donohue JK, et al. (2023). Crosstalk between regulatory elements in disordered trpv4 n-terminus modulates lipid-dependent channel activity. Nat Commun 14, 4165. [PubMed: 37443299]
23. Murphy MC, Rasnik I, Cheng W, Lohman TM, and Ha T (2004). Probing single-stranded dna conformational flexibility using fluorescence spectroscopy. Biophys. J 86, 2530–2537. [PubMed: 15041689]
24. McIntosh DB, Duggan G, Gouil Q, and Saleh OA (2014). Sequence-dependent elasticity and electrostatics of single-stranded dna: Signatures of base-stacking. Biophys. J 106, 659–666. [PubMed: 24507606]
25. Jacobson DR, McIntosh DB, Stevens MJ, Rubinstein M, and Saleh OA (2017). Single-stranded nucleic acid elasticity arises from internal electrostatic tension. Proc. Natl. Acad. Sci. USA 114, 5095–5100. [PubMed: 28461493]
26. Bergonzo C, Grishaev A, and Bottaro S (2022). Conformational heterogeneity of ucaauc rna oligonucleotide from molecular dynamics simulations, saxs, and nmr experiments. RNA 28, 937–946. [PubMed: 35483823]
27. Nguyen HT, Hori N, and Thirumalai D (2022). Condensates in rna repeat sequences are heterogeneously organized and exhibit reptation dynamics. Nat. Chem 14, 775–785. [PubMed: 35501484]
28. Bottaro S, Bussi G, and Lindorff-Larsen K (2021). Conformational ensembles of noncoding elements in the sars-cov-2 genome from molecular dynamics simulations. J. Am. Chem. Soc. 143, 8333–8343. [PubMed: 34039006]
29. Bernetti M, Hall KB, and Bussi G (2021). Reweighting of molecular simulations with explicit-solvent saxs restraints elucidates ion-dependent rna ensembles. Nucleic Acids Res 49, e84. [PubMed: 34107023]
30. Bernetti M, and Bussi G (2023). Integrating experimental data with molecular simulations to investigate RNA structural dynamics. Curr Op Struct Biol 78, 102503.
31. Zhao J, Kennedy SD, Berger KD, and Turner DH (2020). Nuclear magnetic resonance of single-stranded rnas and dnas of caau and ucaauc as benchmarks for molecular dynamics simulations. J. Chem. Theor. Comput 16, 1968–1984.
32. Zhao J, Kennedy SD, and Turner DH (2022). Nuclear magnetic resonance spectra and amber ol3 and roc-rna simulations of ucucgu reveal force field strengths and weaknesses for single-stranded rna. J. Chem. Theor. Comput 18, 1241–1254.
33. Chen H, Meisburger SP, Pabit SA, Sutton JL, Webb WW, and Pollack L (2012). Ionic strength-dependent persistence lengths of single-stranded rna and dna. Proc. Natl. Acad. Sci. USA 109, 799–804. [PubMed: 22203973]

34. Plumridge A, Andresen K, and Pollack L (2020). Visualizing disordered single-stranded rna: connecting sequence, structure, and electrostatics. J. Am. Chem. Soc 142, 109–119. [PubMed: 31804813]
35. Plumridge A, Meisburger SP, and Pollack L (2017). Visualizing single-stranded nucleic acids in solution. Nucleic Acids Res 45, e66. [PubMed: 28034955]
36. Meisburger SP, Sutton JL, Chen H, Pabit SA, Kirmizialtin S, Elber R, and Pollack L (2013). Polyelectrolyte properties of single stranded dna measured using saxs and single-molecule fret: Beyond the wormlike chain model. Biopolymers 99, 1032–1045. [PubMed: 23606337]
37. Song J, Gomes G-N, Gradinaru CC, and Chan HS (2015). An adequate account of excluded volume is necessary to infer compactness and asphericity of disordered proteins by förster resonance energy transfer. J. Phys. Chem. B 119, 15191–15202. [PubMed: 26566073]
38. O'Brien EP, Morrison G, Brooks BR, and Thirumalai D (2009). How accurate are polymer models in the analysis of förster resonance energy transfer experiments on proteins? J. Chem. Phys 130, 124903. [PubMed: 19334885]
39. Hammouda B (1993). Sans from homogeneous polymer mixtures: A unified overview. Polymer Characteristics, 87–133.
40. Ullner M, and Woodward CE (2002). Orientational correlation function and persistence lengths of flexible polyelectrolytes. Macromolecules 35, 1437–1445.
41. Tubbs JD, Condon DE, Kennedy SD, Hauser M, Bevilacqua PC, and Turner DH (2013). The nuclear magnetic resonance of cccc rna reveals a right-handed helix, and revised parameters for amber force field torsions improve structural predictions from molecular dynamics. Biochemistry 52, 996–1010. https://api.semanticscholar.org/CorpusID:26583474. [PubMed: 23286901]
42. Grotz KK, Nueesch MF, Holmstrom ED, Heinz M, Stelzl LS, Schuler B, and Hummer G (2018). Dispersion correction alleviates dye stacking of single-stranded dna and rna in simulations of single-molecule fluorescence experiments. J. Phys. Chem. B 122, 11626–11639. [PubMed: 30285443]
43. Tejedor AR, Garaizar A, Ramírez J, and Espinosa JR (2021). Rna modulation of transport properties and stability in phase-separated condensates. Biophys. J 120, 5169–5186. [PubMed: 34762868]
44. Banerjee PR, Milin AN, Moosa MM, Onuchic PL, and Deniz AA (2017). Reentrant phase transition drives dynamic substructure formation in ribonucleoprotein droplets. Angew. Chem 129, 11512–11517.
45. Zarudnaya MI, Kolomiets IM, Potyahaylo AL, and Hovorun DM (2019). Structural transitions in poly(a), poly(c), poly(u), and poly(g) and their possible biological roles. J. Biomol. Struct. Dyn 37, 2837–2866. [PubMed: 30052138]
46. Lech CJ, Heddi B, and Phan AT (2013). Guanine base stacking in g-quadruplex nucleic acids. Nucl Acids Res 41, 2034–2046. [PubMed: 23268444]
47. Yoo TY, Meisburger SP, Hinshaw J, Pollack L, Haran G, Sosnick TR, and Plaxco K (2012). Small-angle x-ray scattering and single-molecule fret spectroscopy produce highly divergent views of the low-denaturant unfolded state. J. Mol. Biol 418, 226–236. [PubMed: 22306460]
48. Watkins HM, Simon AJ, Sosnick TR, Lipman EA, Hjelm RP, and Plaxco KW (2015). Random coil negative control reproduces the discrepancy between scattering and fret measurements of denatured protein dimensions. Proc. Natl. Acad. Sci. USA 112, 6631–6636. [PubMed: 25964362]
49. Riback JA, Bowman MA, Zmyslowski AM, Plaxco KW, Clark PL, and Sosnick TR (2019). Commonly used fret fluorophores promote collapse of an otherwise disordered protein. Proc. Natl. Acad. Sci. USA 116, 8889–8894. [PubMed: 30992378]
50. Fuertes G, Banterle N, Ruff KM, Chowdhury A, Mercadante D, Koehler C, Kachala M, Estrada Girona G, Milles S, Mishra A, et al. (2017). Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in saxs vs. fret measurements. Proc. Natl. Acad. Sci. USA 114, 6342–6351.
51. He W, Naleem N, Kleiman D, and Kirmizialtin S (2022a). Refining the rna force field with small-angle x-ray scattering of helix–junction–helix rna. J. Phys. Chem. Lett 13, 3400–3408. [PubMed: 35404614]

52. Hopkins JB, Gillilan RE, and Skou S (2017). *Bioxtas raw*: improvements to a free open-source program for small-angle x-ray scattering data reduction and analysis. J. Appl. Crystallogr 50, 1545–1553. [PubMed: 29021737]
53. Martin EW, Hopkins JB, and Mittag T (2021). Small-angle x-ray scattering experiments of monodisperse intrinsically disordered protein samples close to the solubility limit. Methods Enzymol 646, 185–222. [PubMed: 33453925]
54. Kapanidis AN, Laurence TA, Lee NK, Margeat E, Kong X, and Weiss S (2005). Alternating-laser excitation of single molecules. Acc. Chem. Res 38, 523–533. [PubMed: 16028886]
55. N, K.A., Majumdar D, Heilemann M, Nir E, and Weiss S (2015). Alternating laser excitation for solution-based single-molecule fret. Cold Spring Harb Protoc 2015, 979–87. [PubMed: 26527772]
56. Hess B, Kutzner C, Van Der Spoel D, and Lindahl E (2008). Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theor. Comput 4, 435–447.
57. Zgarbová M, Otyepka M, Šponer J, Mládek A, Banáš P, Cheatham TE, and Jurečka P (2011). Refinement of the cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. J. Chem. Theor. Comput 7, 2886–2902.
58. Chen P. c., and Hub JS (2014). Validating solution ensembles from molecular dynamics simulation by wide-angle x-ray scattering data. Biophys. J 107, 435–447. [PubMed: 25028885]
59. He W, Henning-Knechtel A, and Kirmizialtin S (2022b). Visualizing rna structures by saxs-driven md simulations. Front. Bioinform 2, 781949. [PubMed: 36304317]
60. Pitera JW, and Chodera JD (2012). On the use of experimental observations to bias simulated ensembles. J. Chem. Theor. Comput 8, 3445–3451.
61. Cesari A, Reißer S, and Bussi G (2018). Using the maximum entropy principle to combine simulations and solution experiments. Computation 6, 15.
62. Bonomi M, Heller GT, Camilloni C, and Vendruscolo M (2017). Principles of protein structural ensemble determination. Curr. Opin. Struct. Biol 42, 106–116. [PubMed: 28063280]
63. Rangan R, Bonomi M, Heller GT, Cesari A, Bussi G, and Vendruscolo M (2018). Determination of structural ensembles of proteins: restraining vs reweighting. J. Chem. Theor. Comput 14, 6632–6641.
64. Walczewska-Szewc K, and Corry B (2014). Accounting for dye diffusion and orientation when relating fret measurements to distances: three simple computational methods. Phys. Chem. Chem. Phys 16, 12317–12326. [PubMed: 24824374]
65. Riback JA, Bowman MA, Zmyslowski AM, Knoverek CR, Jumper JM, Hinshaw JR, Kaye EB, Freed KF, Clark PL, and Sosnick TR (2017). Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. Science 358, 238–241. [PubMed: 29026044]
66. Bottaro S, Bussi G, Pinamonti G, Reißer S, Boomsma W, and Lindorff-Larsen K (2019). Barnaba: software for analysis of nucleic acid structures and trajectories. Rna 25, 219–231. [PubMed: 30420522]
67. Ng A, Jordan M, and Weiss Y (2001). On spectral clustering: Analysis and an algorithm. Adv Neural Information Processing Systems 14.

Highlights

- Coupling simulations with FRET and SAXS resolves ssRNA structures
- Graph theory-based class averaging identifies sequence-dependent conformational states
- SsRNAs of different sequence have similar lengths but different conformation states
- The A-/C-rich and rU30 ssRNAs have distinct conformations not explained by polymer models

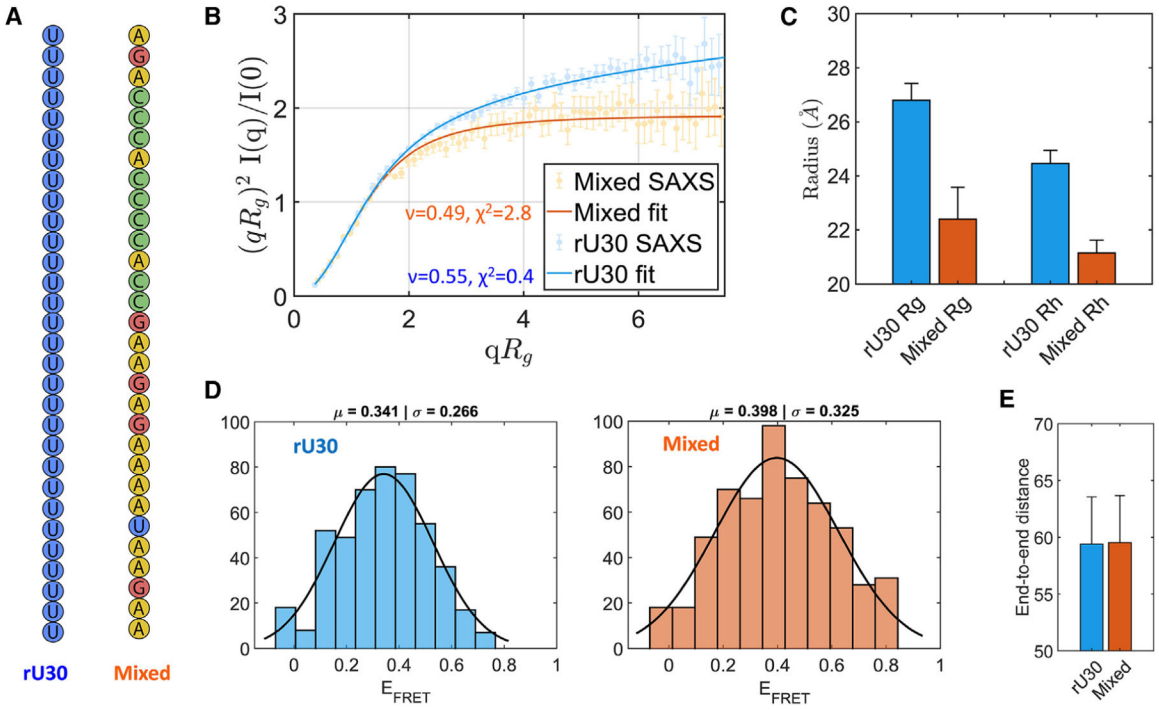


Figure 1. Experimental observables of rU30 and mixed-sequence ssRNA

(A and B) (A) rU30 homopolymer and mixed-sequence heteropolymer RNA sequences employed in this study, colored by nucleotide identity A, adenine; U, uracil; C, cytosine; G, guanine. For both RNA constructs, we show (B) SAXS measurements, plotted along dimensionless Kratky axes to emphasize differences in the overall shape of the structural ensembles. Errorbars denote experimental error in the SAXS measurement. Data are fit to the swollen Gaussian chain model.

(C) Experimentally determined radii of gyration and hydrodynamic radii. Errorbars represent errors in Guinier fits to determine R_g (see "solution X-ray scattering") and propagated error in R_h calculation (see supplemental experimental procedures section "determination of hydrodynamic radius").

(D) Histogram of FRET efficiency (E_{FRET}) values. Data from 3 out of 6 total experiments are pooled. Data are fited to a Gaussian function and the mean (μ) and standard deviation (σ) are shown at the top of the plots.

(E) FRET-derived end-to-end distances ± propagated error (see supplemental experimental procedures section "determination of Förster radii").

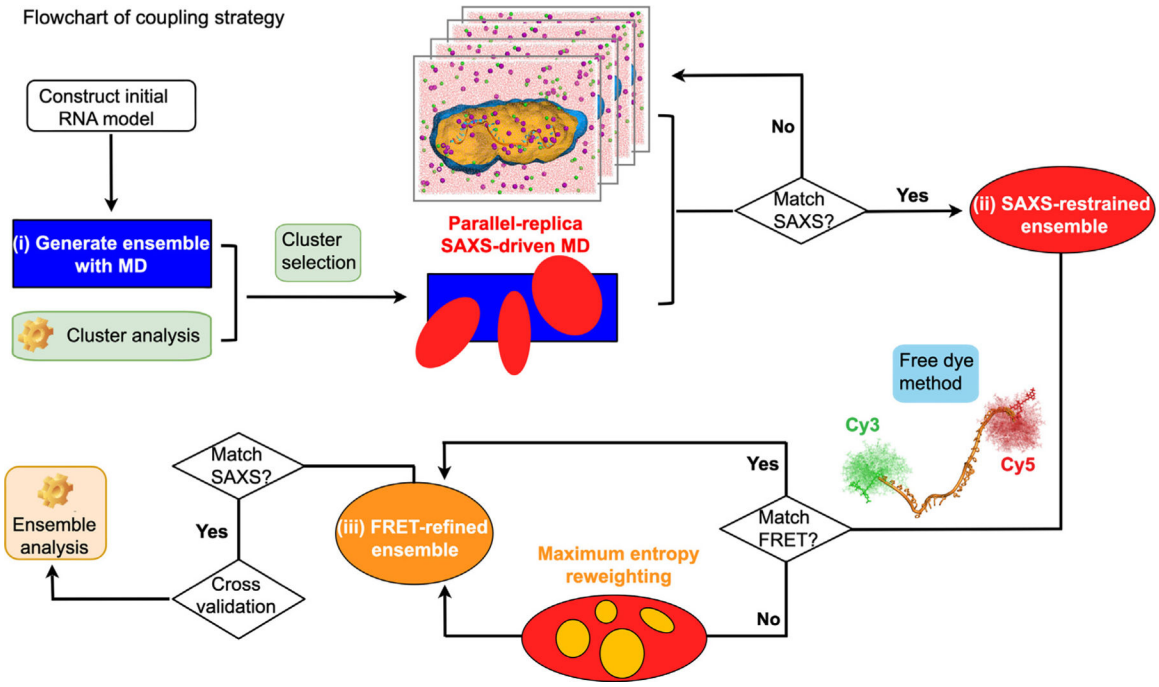


Figure 2. Computational approach for determining RNA conformational ensembles guided by SAXS and FRET data

The integrated approach begins with an RNA model. We then conduct explicit-solvent MD simulations to generate a pool of possible ssRNA conformations (blue block). Subsequently, a clustering analysis identifies accessible states, which are used to initiate parallel replica SAXS-driven MD simulations. During the SAXS-driven MD simulations, we explore conformations guided by SAXS and committed to the maximum entropy principle. This procedure produces a pool of conformations that satisfy the SAXS measurements, referred to as the SAXS-restrained ensemble (red ellipse). Next, we calculate the FRET profiles of the conformations in the SAXS-restrained ensemble (See supplemental experimental procedures section “computing FRET efficiency from MD simulations”). The conformations are re-weighted by the maximum entropy principle against the FRET data (experimental procedures). The re-weighted pool is named the FRET-refined ensemble (orange ellipse), which is later cross-validated against an independent experiment. The validated pool of RNA conformers was used for structural characterizations.

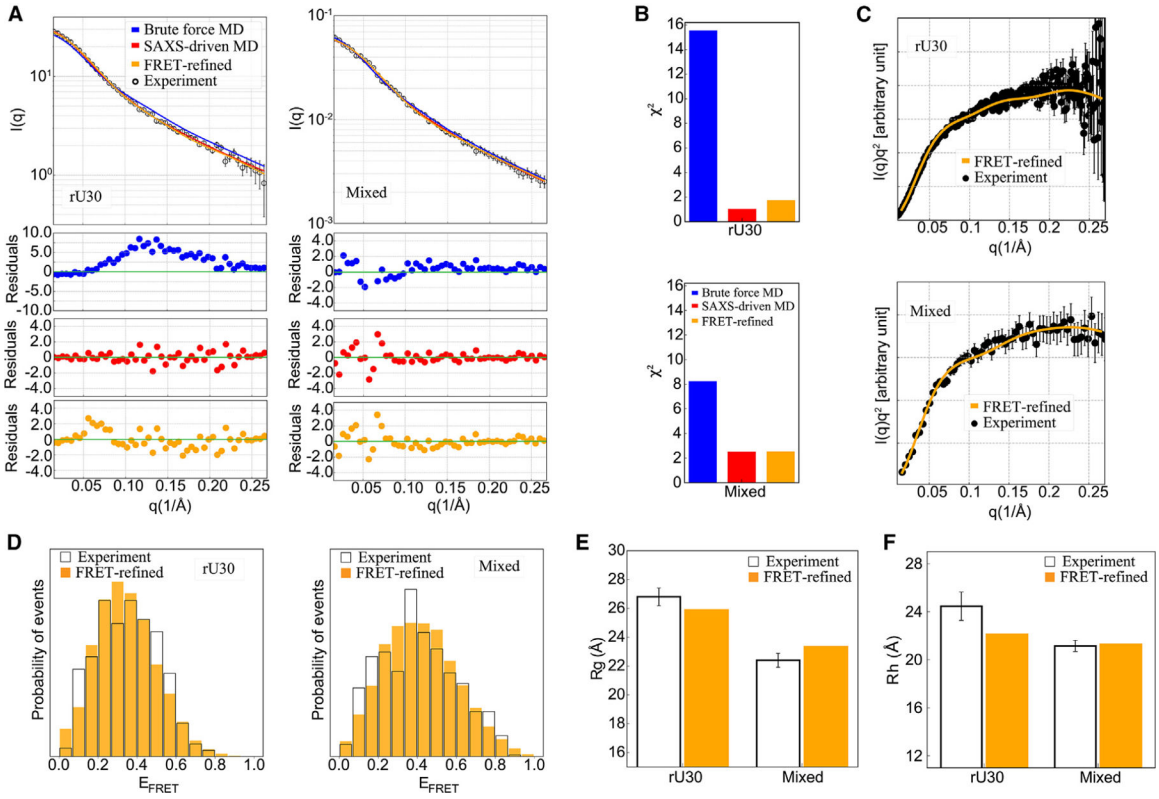


Figure 3. Agreement between the final RNA conformational ensemble with experimental measurements

(A) Agreement with measured SAXS profiles for rU30 and Mixed RNA (blue, brute force MD; red, SAXS-driven MD; orange, FRET-refined ensemble), and residuals between experimental and theoretical curves are shown below the plots.

(B) The agreement between the computed profiles and the SAXS data was quantified using the χ^2 metric.

(C) The SAXS profiles of the FRET-refined ensembles are shown on the Kratky axis to emphasize the mid-q regime.

(D) Agreement of FRET efficiency distributions computed directly from the final ensembles of structures with experimental measurements.

(E and F) Comparison between theoretical and experimentally determined Rg and Rh, and experimental errorbars are delineated in Figure 1C.

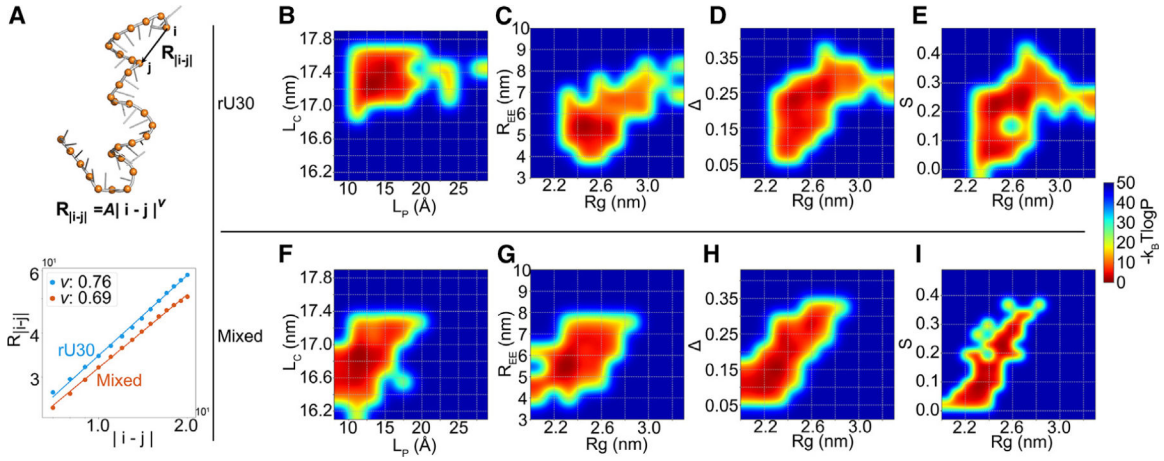


Figure 4. Shape features of single-stranded rU30 and Mixed RNA structural ensemble

(A) Flory's scaling law analysis for rU30 and Mixed, respectively.

(B–E) 2D parameterizations of the Persistence length (L_P), Contour length (L_C), Radius of gyration (Rg), End-to-end distance (R_{EE}), Spherical deviation of the ensemble (Δ), and Ellipsoidal nature of the ensemble (S). Pairs of these metrics are plotted for rU30. The color scale corresponds to the logarithm of the weighted probability.

(F–I) presents the same 2D parameterized L_P-L_C plot, Rg–R_{EE} plot, Rg-Δ plot, and Rg-S plots of Mixed RNA. A comparative analysis of brute force and SAXS-driven MD ensembles is provided in Figure S4. See supplemental experimental procedures section "additional analyses of ssRNA structural ensembles" for the precise definitions of these metrics.

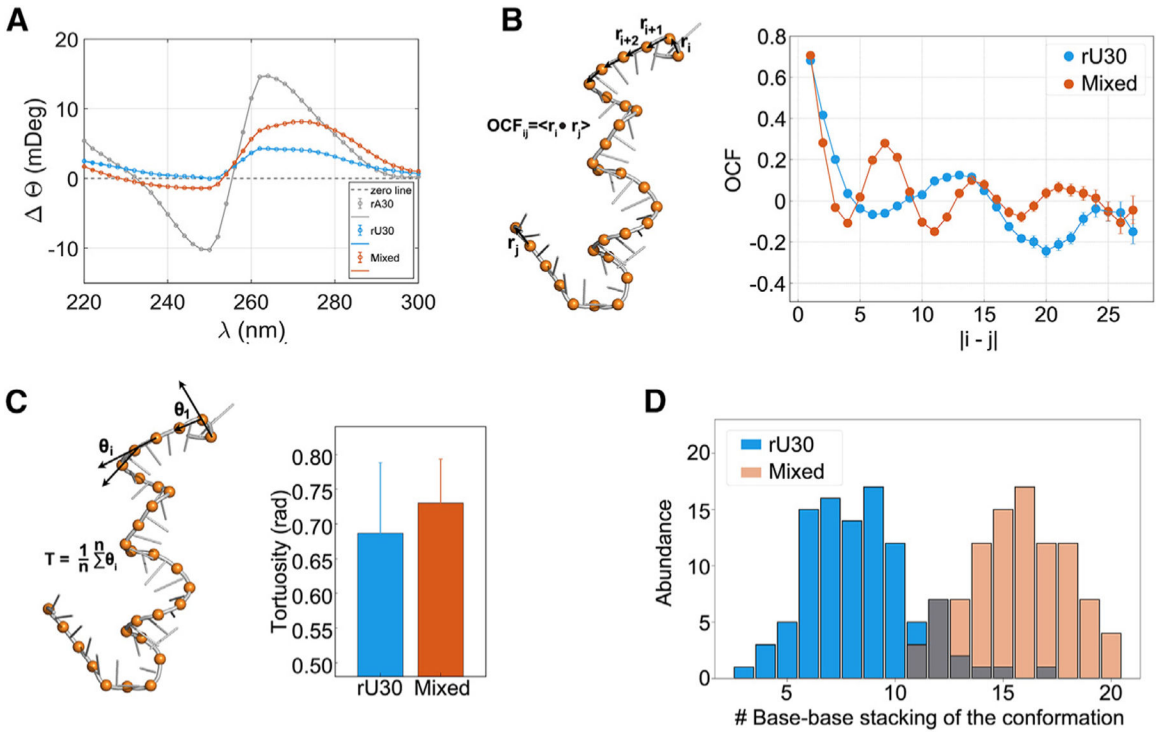


Figure 5. Order parameters describing the intra-chain structural properties of single-stranded rU30 and Mixed RNA

(A) Comparison of experimental circular dichroism spectra of rA30, rU30, and Mixed RNA, showing generally how helical each ssRNA species is.

(B) A comparison between the Orientational correlation function (OCF) curves of rU30 and Mixed RNA.

(C) Comparison of the tortuosity of rU30 structures with that of Mixed RNA.

(D) Histogram analysis of base-base stacking for the conformation pool of rU30 and Mixed. Metrics shown in B-D) are computed from the refined RNA structural ensembles and errorbars represent the variance within the datasets, while A) is derived from an experimental measurement of ssRNA samples.

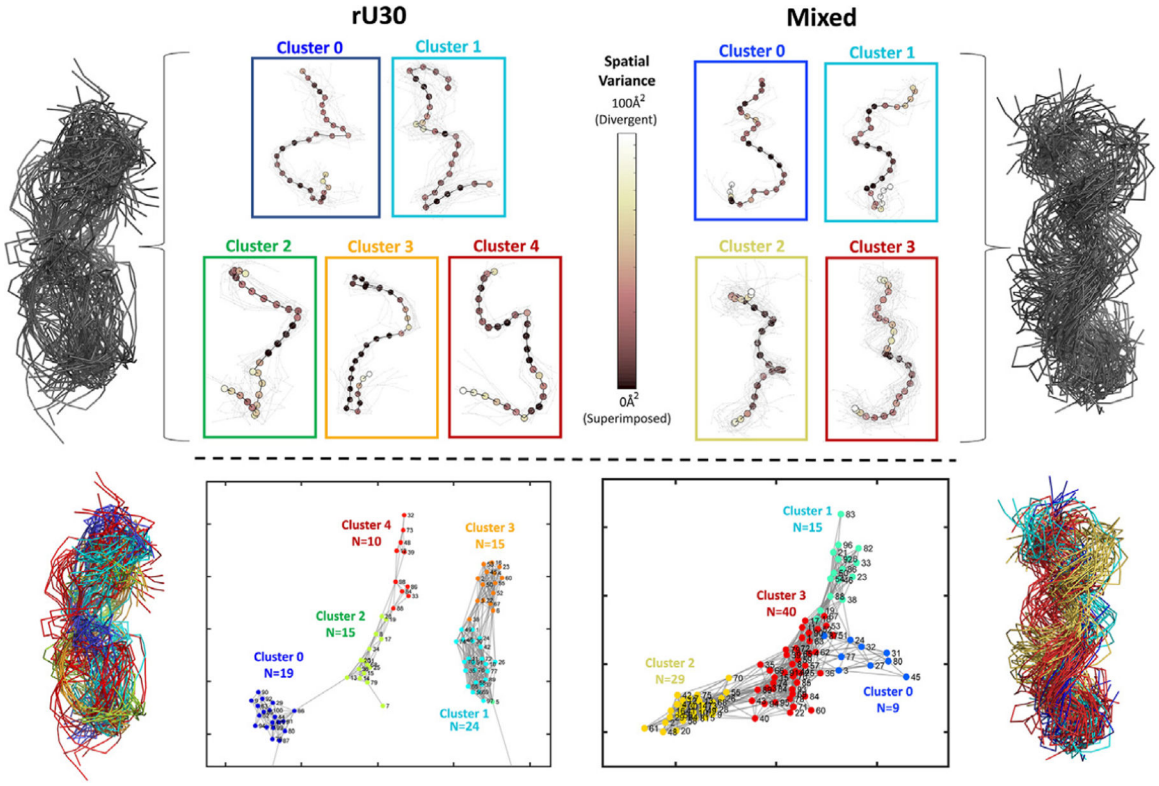


Figure 6. 3D classification and averaging identify 4–5 characteristic structural sub-families in rU30 and mixed-sequence ssRNA

K-means clustered network graphs of RNA conformers are shown after spectral clustering based on pairwise RMSD values. Above that, the averaged backbone conformation of each cluster is shown, colored by the spatial variance of the cluster average. Each dot in the backbone represents the mean position of a phosphorus atom. Every structure in each cluster is plotted along with the averaged structure in pale gray.

Table 1.

Experimental observables of homopolymeric and mixed-sequence ssRNA

Parameter	Technique	rU30	Mixed sequence
Radius of gyration	SAXS	26.80 ± 0.62 Å	22.40 ± 0.48 Å
Flory parameter	SAXS	0.551 ± 0.003	0.495 ± 0010
Diffusion coefficient	FCS	$89.24 \pm 4.31\ \mu m^2/s$	$103.20 \pm 2.27\ \mu m^2/s$
Hydrodynamic radius	FCS	24.46 ± 1.18 Å	21.15 ± 0.47 Å
Mean FRET efficiency	FRET	0.328 ± 0.007	0.390 ± 0.007
StD FRET efficiency	FRET	0.1810	0.2242
Förster radius	FRET	53.30 ± 3.70 Å	55.75 ± 3.87 Å
End to end distance	FRET	59.40 ± 4.16 Å	59.52 ± 4.16 Å