

Title

RNA structures and dynamics with Å resolution revealed by x-ray free electron lasers

Authors

Kara A. Zielinski¹, Shuo Sui^{1†}, Suzette A. Pabit^{1‡}, Daniel A. Rivera², Tong Wang¹, Qingyue Hu¹, Maithri M. Kashipathy³, Stella Lisova³, Chris B. Schaffer², Valerio Mariani³, Mark S. Hunter³, Christopher Kupitz³, Frank R. Moss III^{3§}, Frédéric P. Poitevin³, Thomas D. Grant⁴, Lois Pollack^{1*}

Affiliations

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

² Meinig School of Biomedical Engineering, Cornell University; Ithaca NY 14853 USA

³ Linac Coherent Light Source, SLAC National Accelerator Laboratory; Menlo Park, CA 94025 USA

⁴ Department of Structural Biology, Jacobs School of Medicine and Biological Sciences; University at Buffalo, Buffalo, NY 14203 USA

†Present address: Pfizer Inc. Pearl River, NY

‡present address: Analytical Research and Development, Merck & Co., Inc., West Point, PA

§ present address: Altos Labs, Redwood City, CA.

*Corresponding author. Email: lp26@cornell.edu

Abstract

RNA macromolecules, like proteins, fold to assume shapes that are intimately connected to their broadly recognized biological functions; however, because of their high charge and dynamic nature, RNA structures are far more challenging to determine. We introduce an approach that exploits the high brilliance of x-ray free electron laser sources to reveal the formation and ready identification of Å scale features in structured and unstructured RNAs. Previously unrecognized structural signatures of RNA secondary and tertiary structures are identified through wide angle solution scattering experiments. With millisecond time resolution, we observe an RNA fold from a dynamically varying single strand through a base paired intermediate to assume a triple helix conformation. While the backbone orchestrates the folding, the final structure is locked in by base stacking. This method may help to rapidly characterize and identify structural elements in nucleic acids in both equilibrium and time-resolved experiments.

Teaser

Dynamically changing RNA structures are visualized in time-resolved, high-resolution solution scattering experiments.

Introduction

The majority of genomic DNA is transcribed into RNA, yet only a small fraction is translated into proteins (1). Although the biological roles of much of this untranslated RNA have yet to be elucidated, non-coding RNA is increasingly linked to vital cellular functions (2) and has great potential as a therapeutic (3). Advances in assigning biological roles to these transcripts, or in improving the targeted design of drugs, may be accelerated if their structures can be deduced. Macromolecular structures can be directly measured, simulated, or derived from sequence by solving the folding problem (4), using computational tools (5), or applying artificial intelligence (6). Because of the highly charged and dynamic nature of its backbone, as well as the similarity of its nucleotide building blocks, RNA structures are far more challenging to solve than protein structures, by either experiment (7-9) or all atom simulation (10, 11). The paucity of measured structures limits our ability to advance the field (12), especially when compared to the state-of-the-art in protein structure prediction (6).

X-ray scattering from biomolecules in solution has great potential to advance RNA structure characterization. Small angle X-ray scattering (SAXS) has been frequently applied to measure RNA structure (13-15) and informs about global, nanometer level molecular structures (16). Scattering to large angles (wide angle X-ray scattering, WAXS), enhances measurement resolution to the single Å (17, 18). Recent WAXS studies (19-23) have begun to connect distinctive peaks in scattering profiles with real space features of DNA and RNA, including backbone geometries, structural building blocks (duplexes), and base stacking.

Here, we demonstrate a new approach to characterizing RNA structures. Time-resolved WAXS studies, with millisecond time resolution and single Å spatial resolution, provide information about the transient structures that populate an RNA folding pathway. WAXS profiles, acquired milliseconds after initiation of folding, are interpreted by comparison with profiles of known motifs acquired in static measurements. These latter data were acquired at higher scattering angles (e.g. resolution) than reported in previous studies (19, 24) and display distinctive experimental signatures.

WAXS signals from biomolecules are 100-1000 times smaller than SAXS signals. Seconds-long exposures of typical millimeter or sub-millimeter sized solution samples, with volumes of 10's of microliters, are required for measurement at high flux synchrotron sources (19). Rapid fluidic mixers facilitate the millisecond scale time-resolved measurements relevant for detection of RNA folding reactions (25). X-ray illuminated sample volumes within these flow cells can be 100 to 1000 times smaller than used for equilibrium measurement (26), depending on the desired time resolution. Thus, these time-resolved WAXS measurements are beyond the current-state-of-the-art at synchrotrons.

X-ray free electron laser (XFEL) sources present a unique opportunity to illuminate micron scale, 10's of femtoliter sample volumes with high flux beams. At the Linac Coherent Light Source (LCLS, SLAC National Accelerator Lab), beam intensities at the sample are ~1000 times higher than at state-of-the-art synchrotron beamlines, enabling measurement of previously impossible-to-detect signals using time-resolved WAXS (27-30). We dramatically expand this field by introducing mixing injectors (31, 32) to perform the first, to our knowledge, chemically triggered time-resolved solution scattering experiments on biomolecules at an XFEL. This experiment follows the dynamic acquisition of secondary and tertiary structure of an unstructured, single strand of RNA, as it folds to a triple helical final state. Synchrotron acquired, equilibrium WAXS profiles of related RNAs, in single strand, duplex, and triplex molecular forms, allow us to interpret the distinctive experimental signatures revealed in the time-resolved experiments. These findings advance our understanding of how RNA molecules fold.

Results

High resolution features in solution scattering profiles provide incisive information about RNA's molecular state

The wide angle scattering regime, $q > 0.5 \text{ \AA}^{-1}$ ($q = 4\pi\sin\theta/\lambda$, where θ is half the scattering angle and λ is the x-ray wavelength) is information-rich for nucleic acid samples. Figure 1, panels A-C shows measured scattering profiles of three different RNA motifs, along with a cartoon representation of the structures they reflect. These curves were acquired at synchrotrons. Panel A shows WAXS profiles of RNA single strands, both unstructured chains of 30 uracil nucleotides, rU30, and minimally structured chains of 30 adenosine nucleotides, rA30. Structures derived from SAXS studies (measured for $q < 0.25 \text{ \AA}^{-1}$) of rU30 and rA30 (24) are shown below the WAXS profiles. Panel B shows the scattering profile of a designed duplex that terminates in a loop. Panel C shows scattering profiles of RNA triplexes with different lengths, constructed by adding a third, triplex forming strand to a hairpin duplex. Lower q portions of these data (for $q < 1.0 \text{ \AA}^{-1}$) were previously published (19) and best-fit structures were determined by comparison with all atom simulations. Measurements acquired at higher q , reported here, shows the distinct features of each profile that can be associated with structural patterns that characterize single, double, or triple-stranded RNA. These connections, highlighted by colored, dashed bars in panels B, C, E, and F, are explained in more detail below and in Figure S1.

Single strands. Figure 1 panel A shows the measured scattering profile of two single stranded constructs rU30 and rA30 in buffered solutions containing 100 mM NaCl. Single stranded RNA molecules are dynamic and highly flexible, thus their molecular structures are best recapitulated by an ensemble whose properties (and summed scattering profiles) are consistent with experimental measurements from a variety of probes (24). Typical structures from a best-fit ensemble are shown below the scattering profiles of panel A (24). The intensity of scattering from these disordered molecules decreases monotonically from low q to about $q=0.8 \text{ \AA}^{-1}$, where a broad peak is visible for both rU30 and rA30. This peak occupies a range of q values corresponding to length scales between about 3 and 8 $\text{\AA}$. As noted in prior work on DNA (23), well defined peaks in this range are derived from the phosphate backbone. In contrast, the broad peak seen here reflects the lack of structured elements on these length scales, hence the structural heterogeneity of the ensemble. Within it, signs of structure emerge for rA30, in the form of two loosely separated, broad peaks near $q= 1.1$ and $1.5 \text{ \AA}^{-1}$, but not for rU30, consistent with base stacking induced ordering present in the former and absent in the latter (24). Both curves decay at larger q .

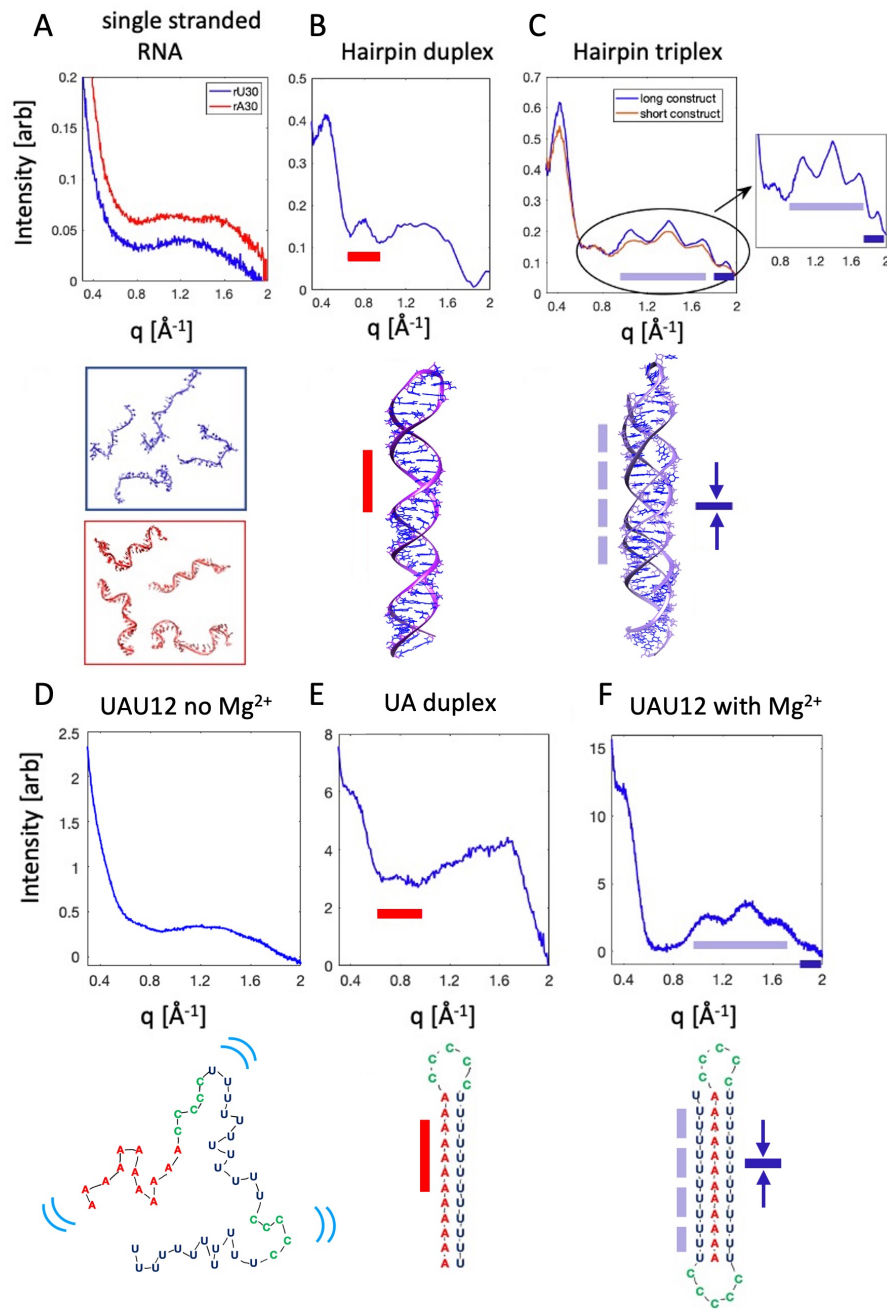


Figure 1. **WAXS features associated with structural motifs of RNA.** A) scattering profiles of single stranded constructs rU30 and rA30 shown as intensity vs. q . Representative structures in the experimentally determined ensemble of rU30 (top) and rA30 (bottom) are shown below. B) scattering profile of designed RNA hairpin duplex from Ref. (19), with a model of one duplex conformation shown below. C) scattering profile of RNA hairpin triplexes from Ref. (19) with a model of triplex from Ref. (19) shown below. The inset emphasizes the rich information content of the WAXS region. The colored lines in panels B and C link the features in the scattering profile with their corresponding real space structures. D) scattering profile of the $t=0$, Mg^{2+} free starting state of the time resolved experiment, acquired at LCLS, with a cartoon representation of unfolded UAU12. Blue lines indicate dynamic motion. E) scattering profile of designed hairpin duplex with UA base pairs, with a cartoon duplex shown below. F) scattering profile of folded UAU12 triplex, 1 second after the addition of Mg^{2+} to trigger folding, acquired at LCLS, with a cartoon model depicting the folded triplex shown below.

Duplexes. Figure 1 panel B shows the scattering profile of a 62 nucleotide RNA hairpin, consisting of a 29 base-paired duplex, terminated by a four nucleotide loop from Ref. (19). Prior studies examined the scattering out to $q = 1.25 \text{ \AA}^{-1}$ and used molecular dynamics simulations (19) in conjunction with machine learning (21) to confirm that the small, but notable peak at $q = 0.8 \text{ \AA}^{-1}$ is connected with the helical major groove, a repeated spatial dimension in the duplex structure. A red bar in the figure connects these reciprocal and real space features. At higher angle, $1 < q < 2 \text{ \AA}^{-1}$, a broad peak appears. Although better articulated than in panel A, the sharp peaks characteristic of DNA duplexes (23) are absent, suggesting that this duplex has some conformational variation. This observation is consistent with other studies of RNA duplexes (20, 22). A single molecular conformation from the simulations of Ref. (19) is pictured just below the profile.

Triplexes. Figure 1 panel C illustrates the unique and never-before-reported triplet of peaks associated with triple helical backbones, indicated by a purple bar in the figure. Measured scattering profiles from two triplexes are shown. Both were constructed by adding a triplex forming single strand to a hairpin duplex. The shorter/longer construct consists of loop terminated 17/29 base paired duplexes, with an added 12/24 nucleotide single strand that binds to form base triples over much of the length of the molecule. A model of this latter structure is shown just below the plot, taken from Ref. (19). Its increased length reinforces the features that uniquely identify the triplex structure by WAXS: three sharp peaks at $q = 1.0, 1.4$ and $1.7 \text{ \AA}^{-1}$. By analogy with the DNA fingerprinting studies of Ref. (23), and structural modeling (SI: *Modeling connects features in the scattering profile with molecular structures* and Figure S1), these peaks reflect the regular spacing of atoms along the backbone in a helical conformation. This cluster of three peaks is flanked by additional peaks on either side. At lower q , a weak reflection of the major groove peak can be seen. Figure S2 shows that this peak is substantially reduced and shifted to lower q between the duplex and triplex states. As previous x-ray fiber diffraction studies of a pure RNA triplex suggest that it is roughly cylindrical, lacking grooves (33), this reduced peak may reflect a small amount of duplex that is present in the construct shown below panel C. At higher q ($1.93 \text{ \AA}^{-1}$), the small but distinct peak (dark blue bar, Figure 1C) reflects the $3.26 \text{ \AA}$ distance between stacked base triples (Figure S1 and Ref. (19)). The sharpness of the peaks indicates that triplex molecular dimensions are well defined, with less conformational variation than either the duplex or single strands.

These curves reveal five distinct features of WAXS ($q > 0.5 \text{ \AA}^{-1}$) scattering profiles of RNA structures, only one of which has been previously identified (19). The peak at $q = 0.8 \text{ \AA}^{-1}$ reflects the existence of a major groove. The three peaks at $q = 1.0, 1.4$ and $1.7 \text{ \AA}^{-1}$ represent regularly positioned atoms in the three RNA backbones of the triplex and the tiny peak at $q = 1.9 \text{ \AA}^{-1}$ reflects stacking of base triples (Figure S1 and length scales extrapolated from structures of Ref. (19)). Additional information can be gleaned from lower q , SAXS features (SI: *Low q changes* and Figure S2). In particular, when unstructured single strands combine to form a duplex, a peak appears near $q = 0.4 \text{ \AA}^{-1}$. According to machine learning models (21), the position of this peak reflects the radius of the helical structure. It shifts slightly to lower q for the triplex relative to the duplex, reflecting the larger helical radius of the former. Although this work highlights higher angle features, the lower q peak also contributes to our understanding of RNA structure and serves as an important milestone in folding studies.

Starting and ending states of the mixing experiment. For folding studies, we used a 46 nucleotide RNA with sequence 12U-5C-12U-5C-12A. This construct is a long single strand of RNA, in contrast to the triplexes shown in panel C, which were each created from two separate RNA molecules. This design is based on a well-characterized triplex forming sequence (34, 35). We refer to it as UAU12, to highlight the capture of the A12 strand between the two U12 strands. Measured scattering profiles of the starting and ending states of the folding experiment are shown

in Figure 1, panels D and F. These data were acquired at the XFEL. In the initial, unfolded state the RNA is in a low ionic strength buffer. Folding is initiated by the addition of MgCl_2 .

Structural details about the RNA during the time resolved experiment can be gleaned by comparison with the profiles of the reference systems in panels A-C. For example, the similarity of the scattering profiles of panels D and A suggests that the initial state of the time resolved experiment is unstructured. This assessment is based on the absence of distinct features in its scattering profile. A cartoon that suggests this conformation is shown below the profile of panel D; here, the blue lines indicate dynamics.

Figure 1 panel E shows the measured scattering profile of a related hairpin duplex (12U-5C-12A, the first 29 nucleotides of the full UAU12 construct), folded by adding Mg^{2+} to the solution of single strands of this construct. This profile was acquired to identify a potential folding intermediate, a 12 base pair A-U duplex. Because RNA duplexes are length (20), sequence (36) and salt (11) dependent, it was critical to measure the structure of a molecule where an rA12 strand binds to an rU12 strand to form a 12 base pair duplex. The duplex of panel B is much longer, and of mixed sequence, hence will have better defined structures. Collections of rA and rU strands tend to form triplexes when combined, in addition to duplexes (37), but the mixed sequence 29 nucleotide strands formed duplexes: the scattering profile of panel E displays comparable features to the duplex profile of panel B. A peak detected near $q=0.8 \text{ \AA}^{-1}$, indicated by the red bar, reflects the formation of a duplex major groove (21, 22).

Finally, the ending state of the experiment (UAU12 with added Mg^{2+}), shown in panel F, displays the characteristic three-peak structure of a triplex (purple bar), as well as the higher q peak of panel C (blue bar). The major groove peak (at $q=0.8^{-1}$) is absent from this measured profile, consistent with third strand filling the major groove. The cartoon schematically indicates the conformation of the folded triplex.

Mixing injectors target time scales from the single millisecond to the single second

To capture WAXS profiles with sensitivity to millisecond scale conformational changes, time-resolved data were collected from micron-sized liquid jets at the CXI beamline at LCLS. Mixing injectors (schematically shown in Figure 2) were used to initiate the Mg^{2+} mediated folding of UAU12. A full description of the fabrication and operation of the injectors, including design parameters, can be found in Ref. (31). Three different injector geometries were used to access the broad range of time points of interest to this folding experiment: $t=0, 6, 10, 60, 100, 500$ and 1000 milliseconds. By varying the flow rates of the sample and buffer, each injector can be used to acquire data at multiple, closely spaced time points. On average, about 16,000 good quality single shot profiles are used to generate the averaged profile required to visualize the WAXS features from RNA flowing in a sheathed, micron-sized jet. With the current XFEL repetition rate of 120 Hz, it takes just over 2 minutes of beam exposure to acquire such a profile. Given that a data point requires both a sample-present (RNA in buffer) and a sample absent (buffer alone) exposure, a total of 4 minutes of data acquisition yields one time point. Importantly, each profile is scaled to account for the varying intensity of the XFEL pulse that created it. Details about profile acquisition, computation, normalization, selection, and background subtraction are provided in Methods and Figures S4-S7.

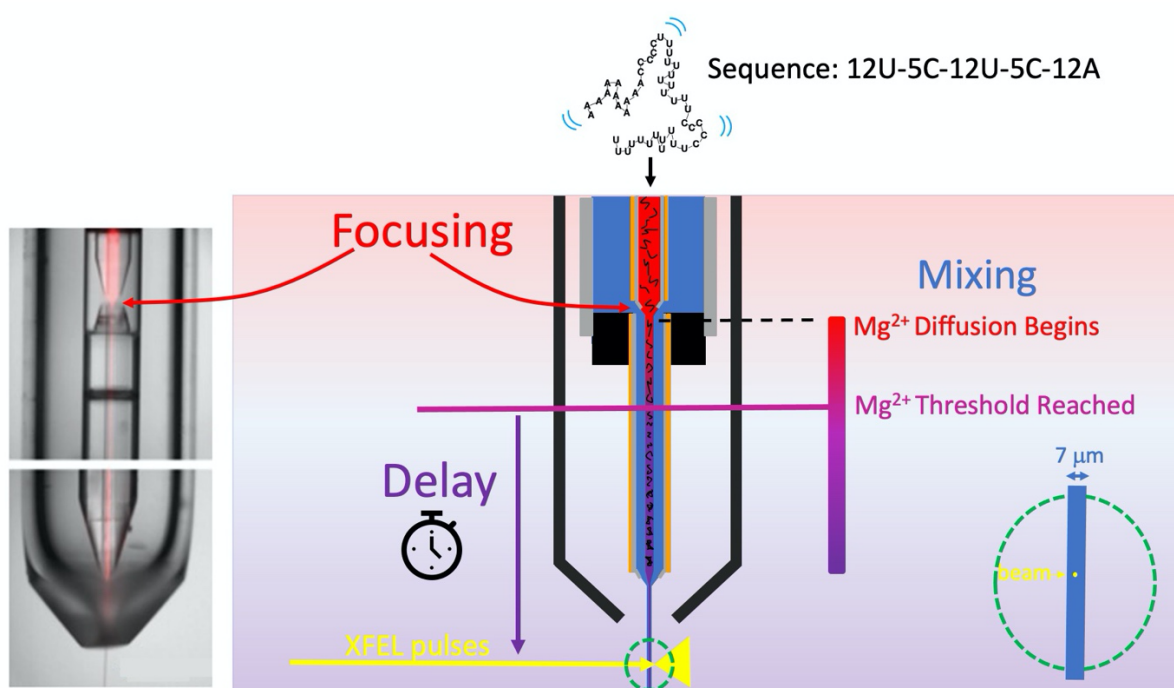


Figure 2: **Mixing injectors for time resolved studies at the XFEL.** A solution containing RNA in low ionic strength buffer (red liquid) is hydrodynamically focused into a thin cylindrical stream, which is surrounded/sheathed by a buffer containing additional Mg^{2+} ions (blue liquid). These streams flow coaxially through a constricted region, where mixing occurs by the rapid diffusion of Mg^{2+} ions into the thin, central, RNA-containing stream. The reaction is initiated once the $[\text{Mg}]$ reaches a carefully selected threshold (Figure S3). The reported time point for each measurement corresponds to the time it takes for the molecule to flow to reach the beam. This delay time is carefully determined for each experiment. The dashed green circle illustrates the intersection point of the beam with the jet, and is expanded at the bottom right of the figure to indicate the relative size of the jet (7 μm average over all time points probed) and the beam ($\sim 1 \mu\text{m}$ yellow dot). Detailed parameters for injectors, including dimensions, jet widths and flow conditions, can be found in Methods and Tables S1-2.

Time resolved studies reveal folding intermediates; a transient duplex forms before the triplex

Time-resolved experiments were performed at RNA concentrations of 1 mM and 0.5 mM. Different amounts of Mg^{2+} were required to quickly reach the reaction initiation threshold for each $[\text{RNA}]$ (SI: *Determining the proper $[\text{Mg}^{2+}]$ level to initiate the reaction* and Figure S3). Figure 3 shows the time series for the higher concentration experiment. These curves show the time progression of scattering profiles from the form shown in panel 1D to that of panel 1F. For this series, we acquired good quality data at $t = 6 \text{ ms}$, 10 ms, 60 ms and 1000 ms. The curves are shown along with the $t = 0$ state (acquired at $[\text{RNA}] = 0.5 \text{ mM}$). Four features are visible at the earliest time point acquired, 6 ms after the initiation of folding (orange). A shoulder appears near $q = 0.4 \text{ \AA}^{-1}$ and the major-groove associated peak appears near $q = 0.8 \text{ \AA}^{-1}$. Two higher q peaks also appear, though they are not well resolved. On this short time scale, the lower q data indicate that some fraction of the sample is in the duplex state. The appearance of triplex-associated peaks also suggests partial, but not yet full structuring of the triple helical backbone. It is unclear whether all molecules fold through a mandatory duplex intermediate, or whether some fold directly to triplex. This distinction will require measurements at shorter time points. Nonetheless, it is clear that the RNA shows some duplex and some weak triplex features within 6 ms of folding initiation.

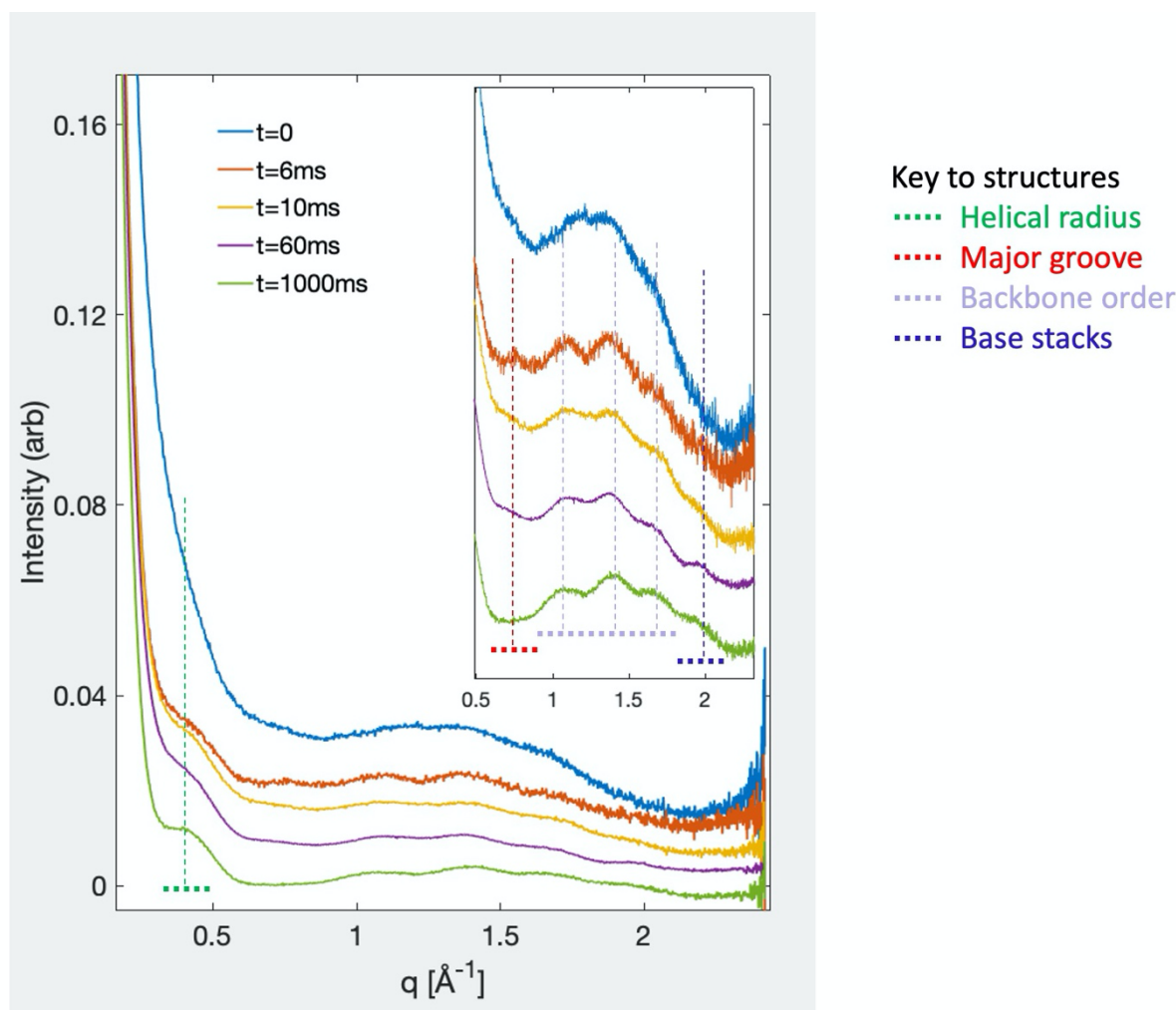


Figure 3 Time resolved WAXS data following the formation of an RNA triplex from a minimally structured single strand. The measured time series following folding of the UAU12 construct from single strand to triplex is shown for the initial ($t=0$) and final folded ($t=1000$ ms) states, along with three intermediate time points. Data were acquired 6, 10 and 60 ms after folding initiation. Distinct features in the scattering profiles, identified in conjunction with equilibrium measurements, modeling and machine learning studies shown in Figure 1 and indicated here with dashed lines and Figure 1's color code, aid in the interpretation of the dynamically evolving structures that populate the folding landscape. Features of disordered, partially ordered and fully ordered RNA are observed as the reaction proceeds. Together, they reveal the folding strategy of this molecule.

In the 10 ms profile (yellow), diminished duplex features are observed: the major groove peak is reduced. The triplex peaks are better articulated, and a small, but not quite resolvable, perturbation in intensity appears near the highest q , base stacking peak. Sixty ms after the initiation of folding (purple curve), the triplex peaks, as well as the base stacking peak are firmly established. These trends are further underscored at 1000 ms after folding (green). Here, the lower q shoulder near $q=0.4 \text{ \AA}^{-1}$ has settled into a form that resembles that of the static triplex (Figure S2). At this time point, the scattering profile displays all of the SAXS/WAXS peaks seen in an equilibrium, folded curve, acquired on our lab source (Figure S8). Deviations in the baseline levels can be explained by temperature variations (SI: *The effect of temperature on scattering profiles of nucleic acids* and Figures S9-S10). The signal-to-noise ratio of each curve depends on the flow rate of the sample, which determines the total amount of RNA in the beam-illuminated sample volume, and more critically on the beam power. In some cases, ice crystals formed in the jet (or at the nozzle) creating

some very intense diffraction spots on the detector. We attenuated the beam to protect against damage. The fraction of beam power used to collect data for each time point is provided in Tables S3-4.

Discussion

Overall, the RNA's folding strategy can be established directly from the curves shown in Figure 3. The triplex folds from a mostly unstructured single strand, through a duplex intermediate, to a state where the three backbone strands form the outline of the triplex. Finally, base stacking appears at longer times, locking the molecule into its final structure. This strategy is suggested in Figure 4.

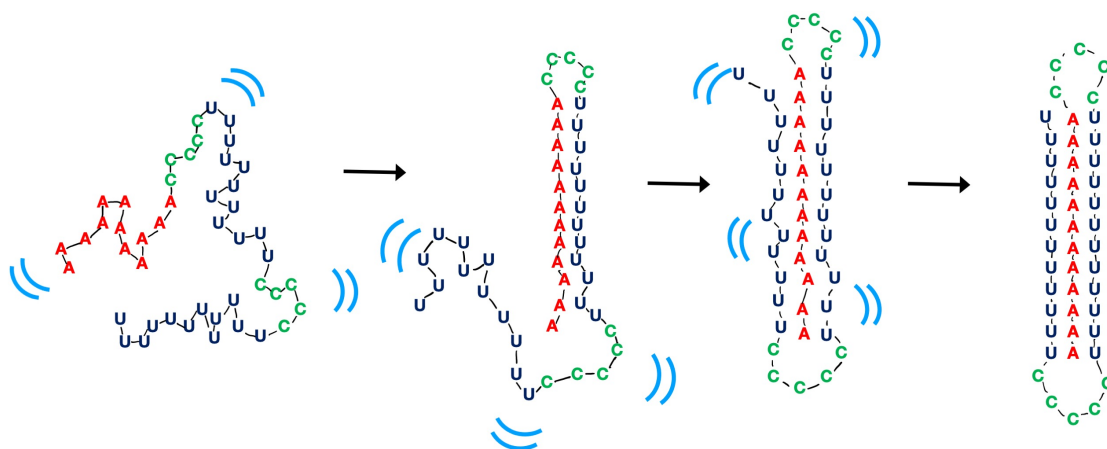


Figure 4. Schematic cartoon depicting folding dynamics from single strand to triplex as read directly off the scattering profiles. The left-most cartoon depicts the dynamic variation of the low ionic strength starting state: a mostly unstructured and dynamic (indicated by blue marks) single strand of 46 nucleotides. The next cartoon is a representation of a transiently detected state in which one 12 base paired A-U duplex forms. The duplex is identified by the ‘major groove feature’ that appears in the $t=6$ ms scattering profile. The third panel represents a state in which the backbones have formed a loose triplex structure, defined by the appearance of a triplet of peaks at higher q . The final structure, at right, appears to have a less dynamic structure, and shows strong evidence of stacking of base triples that lock the molecule into a more rigid structure. The sharp peaks of the WAXS profile suggest that the molecule is less dynamic in this state; the triplex structure is well-defined with much less variation than RNA duplex structure.

The order of appearance of all of the peaks is recapitulated by the measured time series at $[RNA]=0.5$ mM, where data points were acquired at 0, 6, 100 and 500 ms (SI: *RNA folding at reduced concentration: 0.5 mM* and Figure S11). Most significantly, the transient duplex intermediate again appears at 6 ms, and vanishes at all later times. Slight deviations were seen in the amount of triplex present at 6 ms, likely a result of mixing conditions.

This distinctive view of dynamic structural changes from single strand to triplex underscores the highly dynamic nature of the RNA backbone, even in the duplex state, until it captures a third strand in the proper triplex geometry. Once the proper backbone arrangement is assumed, the structure is locked into place by base stacking. This experiment provides direct evidence that the RNA backbone drives folding, a conclusion that was previously hypothesized by comparison with MD simulations (25), but not directly observed until now.

In summary, WAXS reveals numerous features of secondary and tertiary RNA structures, which are instantly identifiable as the data are acquired. Our studies reveal the dynamic appearance and disappearance of helical grooves, backbone ordering and base stacking. This structurally detailed information is trivially extracted by inspection of profiles of equilibrium (static) states, as well as from transient states during a folding reaction. When coupled with lower angle data that reveal the relative spatial arrangement of these structural elements, this work highlights the potential of solution scattering to uniquely and directly characterize both static and dynamic RNA structures.

Presently, less than 2000 RNA structures are deposited in the protein data bank. If a substantial database of structures can be built, either directly or by solving the folding problem, new deep learning techniques (12) can potentially provide a structural revolution in the RNA world comparable the one recently provided for proteins by Alpha Fold (6). High angle solution x-ray scattering reports distinct features in scattering profiles which, in conjunction with models (e.g. (19)), can be linked with real space RNA structures. When coupled with lower angle data that articulate the relative arrangement(s) of these motifs, this technique can help address this unmet need to characterize RNA structures.

In addition to enhancing structure solving methods, this work highlights the unique ability of highly brilliant, XFEL x-rays to provide the exceptional sensitivity needed to observe subtle structural changes in macromolecular systems. Although there is already much to learn from this now demonstrated technique, XFEL technology is rapidly advancing. The increasing pulse repetition rates at XFELs worldwide (from hundreds to millions of pulses per second (38)), coupled with the development of fast framing detectors to handle these data rates, will enable the visualization of even higher q (sharper spatial resolution) features in solution scattering. Once these upgrades are realized, atomic resolution in solution structures, at room temperature, and on biologically relevant time scales is imminent.

Materials and Methods

Nucleic acid samples

All RNA and DNA samples were purchased from Integrated DNA Technologies (Coralville, IA) as single strands. Molecular reconstitution, including buffers and annealing protocols are distinct for each sample, and are described in detail in SI: *Materials and Methods*.

Mixing Injectors

Mixing injectors (as described in (31, 39)) were designed and fabricated for the time-resolved experiments. These devices utilize a flow-focused diffusive mixer coupled to a Gas Dynamic Virtual Nozzle (GDVN) to initiate reactions just prior to producing a freely flowing liquid jet for data collection. Design parameters, fabrication details, and specifications for each mixer used are described in SI: *Materials and Methods*.

Data collection and analysis at XFEL

Time resolved RNA folding data, as well as data on the AT duplex were collected at the Coherent X-ray Imaging (CXI) instrument at the LCLS (40) (SLAC National Accelerator Lab, Menlo Park, CA), using the 1 micron sample environment. The mixing nozzles were loaded into the vacuum chamber on a nozzle rod (Standard Configuration 1). X-rays were delivered at 120 Hz frequency with a pulse energy of 2 mJ and beam size of $\sim 1 \mu\text{m}$ (FWHM). Scattering was collected on the

Jungfrau-4M detector (41). The X-ray energy was 6 keV and the detector was positioned 106 mm from the sample resulting in a q range of $0.12 - 2.4 \text{ \AA}^{-1}$. Calibration of the detector distance and geometry was performed using silver behenate. Newly developed protocols for acquiring fully background subtracted solution scattering data at XFELs are provided in SI: *Materials and Methods*.

Data collection and analysis at NSLS II

Solution X-ray scattering measurements on single stranded rU30 and rA30 RNA were performed at the 16-ID Life Science X-ray Scattering beamline at the National Synchrotron Light Source II (NSLS-II) of Brookhaven National Laboratory (42). Scattering from the small-angle ($q=0.01-0.3 \text{ \AA}^{-1}$) and wide-angle ($q=0.3-3.2 \text{ \AA}^{-1}$) regimes were read simultaneously using two Pilatus 1M detectors (Dectris, Switzerland, EU) arranged in series. The transmitted X-ray beam intensity was also recorded during each measurement. Centering and calibration of the beam on both detectors was performed using a silver behenate standard in BioXTAS RAW, as well as masking, radial averaging and buffer subtraction (43). Previously published data (19), reproduced here in Figure 1 panels B and C, were also acquired at NSLS II on the LiX beamline.

Data collection and analysis on lab source

Measurements on the rUA duplex hairpin, as well as on the temperature dependence of the DNA AT duplex were performed using a BioXolver with Genics source (Xenocs, Holyoke, MA) using the setting WAXS_STD. Data collected from the BioXolver are analyzed using BioXTAS RAW, as described above. The quoted sample temperatures, when noted, were achieved by temperature controlling the sample capillary.

Statistical Analysis

Time-resolved data were acquired from ~ 16000 frames for each sample present and sample absent (buffer) profile. The number of frames for each data point, for all data shown in Figures 3 and S1 **1** are provided in Tables S3 and S4. At a frame rate of 120 per second, this corresponds to just over two minutes of data acquisition for each condition.

References

1. I. Dunham *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
2. L. Statello, C.-J. Guo, L.-L. Chen, M. Huarte, Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Bio* **22**, 96-118 (2021).
3. M. Winkle, S. M. El-Daly, M. Fabbri, G. A. Calin, Noncoding RNA therapeutics — challenges and potential solutions. *Nat Rev Drug Discov* **20**, 629-651 (2021).
4. D. Herschlag, S. Bonilla, N. Bisaria, The Story of RNA Folding, as Told in Epochs. *CSH Perspect Biol* **10**, (2018).
5. B. Kuhlman, P. Bradley, Advances in protein structure prediction and design. *Nat Rev Mol Cell Bio* **20**, 681-697 (2019).
6. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
7. O. B. Becette *et al.*, Solution NMR readily reveals distinct structural folds and interactions in doubly ¹³C- and ¹⁹F-labeled RNAs. *Sci Adv* **6**, eabc6572 (2020).
8. S. L. Bonilla, J. S. Kieft, The promise of cryo-EM to explore RNA structural dynamics. *J Mol Biol* **434**, (2022).
9. A. Gomez, N. Toor, Selecting New RNA Crystal Contacts. *Structure* **26**, 1166-1167 (2018).
10. J. Sponer *et al.*, RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem Rev* **118**, 4177-4338 (2018).
11. D. Tan, S. Piana, R. M. Dirks, D. E. Shaw, RNA force field with accuracy comparable to state-of-the-art protein force fields. *P Natl Acad Sci USA* **115**, E1346-E1355 (2018).
12. R. J. L. Townshend *et al.*, Geometric deep learning of RNA structure. *Science* **373**, 1047-+ (2021).
13. W. A. Cantara, E. D. Olson, K. Musier-Forsyth, Analysis of RNA structure using small-angle X-ray scattering. *Methods* **113**, 46-55 (2017).
14. Y. Chen, L. Pollack, SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdiscip. Rev. RNA*, (2016).
15. X. Fang, J. R. Stagno, Y. R. Bhandari, X. Zuo, Y.-X. Wang, Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures. *Curr Opin Struc Biol* **30**, 147-160 (2015).
16. D. I. Svergun, M. H. J. Koch, Small-angle scattering studies of biological macromolecules in solution. *Rep Prog Phys* **66**, 1735-1782 (2003).
17. H. S. Cho, F. Schotte, V. Stadnytskyi, P. Anfinrud, Time-resolved X-ray scattering studies of proteins. *Curr Opin Struc Biol* **70**, 99-107 (2021).
18. L. Makowski, D. J. Rodi, S. Mandava, S. Devarapalli, R. F. Fischetti, Characterization of Protein Fold by Wide-Angle X-ray Solution Scattering. *J Mol Biol* **383**, 731-744 (2008).
19. Y.-L. Chen, W. He, S. Kirmizialtin, L. Pollack, Insights into the structural stability of major groove RNA triplexes by WAXS-guided MD simulations. *Cell Rep Physical Science* **3**, (2022).
20. Y. L. Chen, L. Pollack, Salt Dependence of A-Form RNA Duplexes: Structures and Implications. *J Phys Chem B* **123**, 9773-9785 (2019).
21. Y. L. Chen, L. Pollack, Machine learning deciphers structural features of RNA duplexes measured with solution X-ray scattering. *IUCrJ* **7**, 870-880 (2020).
22. W. He, Y. L. Chen, L. Pollack, S. Kirmizialtin, The structural plasticity of nucleic acid duplexes. *Sci Adv* **7**, (2021).
23. X. Zuo *et al.*, X-ray diffraction “fingerprinting” of DNA structure in solution for quantitative evaluation of molecular dynamics simulation. *P Natl Acad Sci USA* **103**, 3534-3539 (2006).

- 445 24. A. Plumridge, K. Andresen, L. Pollack, Visualizing Disordered Single-Stranded RNA:
446 Connecting Sequence, Structure, and Electrostatics. *J Am Chem Soc* **142**, 109-119 (2020).
- 447 25. A. Plumridge *et al.*, Revealing the distinct folding phases of an RNA three-helix junction.
448 *Nucleic Acids Res* **46**, 7354-7365 (2018).
- 449 26. A. Ghazal *et al.*, Recent advances in X-ray compatible microfluidics for applications in soft
450 materials and life sciences. *Lab Chip* **16**, 4263-4295 (2016).
- 451 27. D. Arnlund *et al.*, Visualizing a protein quake with time-resolved X-ray scattering at a free-
452 electron laser. *Nat Methods* **11**, 923-926 (2014).
- 453 28. E. H. Choi, Y. Lee, J. Heo, H. Ihee, Reaction dynamics studied via femtosecond X-ray
454 liquidography at X-ray free-electron lasers. *Chem Sci* **13**, 8457-8490 (2022).
- 455 29. M. Levantino *et al.*, Ultrafast myoglobin structural dynamics observed with an X-ray free-
456 electron laser. *Nat Commun* **6**, (2015).
- 457 30. R. Neutze, K. Moffat, Time-resolved structural studies at synchrotrons and X-ray free electron
458 lasers: opportunities and challenges. *Curr Opin Struc Biol* **22**, 651-659 (2012).
- 459 31. G. D. Calvey, A. M. Katz, L. Pollack, Microfluidic Mixing Injector Holder Enables Routine
460 Structural Enzymology Measurements with Mix-and-Inject Serial Crystallography Using X-ray
461 Free Electron Lasers. *Anal Chem* **91**, 7139-7144 (2019).
- 462 32. J. L. Olmos *et al.*, Enzyme intermediates captured "on the fly" by mix-and-inject serial
463 crystallography. *BMC Biology* **16**, (2018).
- 464 33. R. Chandrasekaran, A. Giacometti, S. Arnott, Structure of Poly (U)•Poly (A)•Poly (U). *J Biomol*
465 *Struc Dyn* **17**, 1023-1034 (2000).
- 466 34. C. E. Carr, I. Khutsishvili, L. A. Marky, Energetics, Ion, and Water Binding of the Unfolding of
467 AA/UU Base Pair Stacks and UAU/UAU Base Triplet Stacks in RNA. *J Phys Chem B* **122**,
468 7057-7065 (2018).
- 469 35. A. M. Soto, J. Loo, L. A. Marky, Energetic Contributions for the Formation of TAT/TAT,
470 TAT/CGC+, and CGC+/CGC+ Base Triplet Stacks. *J Am Chem Soc* **124**, 14355-14363 (2002).
- 471 36. J. D. Yesselman *et al.*, Sequence-dependent RNA helix conformational preferences predictably
472 impact tertiary structure formation. *P Natl Acad Sci USA* **116**, 16847-16855 (2019).
- 473 37. Y. Blagoi *et al.*, Phase equilibrium in poly(rA) · poly(rU) complexes with Cd²⁺ and Mg²⁺ ions,
474 studied by ultraviolet, infrared, and vibrational circular dichroism spectroscopy. *Biopolymers* **78**,
475 275-286 (2005).
- 476 38. W. Decking *et al.*, A MHz-repetition-rate hard X-ray free-electron laser driven by a
477 superconducting linear accelerator. *Nature Photonics* **14**, 391-397 (2020).
- 478 39. G. D. Calvey, A. M. Katz, C. B. Schaffer, L. Pollack, Mixing injector enables time-resolved
479 crystallography with high hit rate at X-ray free electron lasers. *Struct Dyn* **3**, (2016).
- 480 40. M. N. Liang *et al.*, The Coherent X-ray Imaging instrument at the Linac Coherent Light Source.
481 *J Synchrotron Radiat* **22**, 514-519 (2015).
- 482 41. F. Leonarski *et al.*, JUNGFR AU detector for brighter x-ray sources: Solutions for IT and data
483 science challenges in macromolecular crystallography. *Struct Dyn* **7**, (2020).
- 484 42. L. Yang *et al.*, Solution scattering at the Life Science X-ray Scattering (LiX) beamline. *J*
485 *Synchrotron Radiat* **27**, 804-812 (2020).
- 486 43. J. B. Hopkins, R. E. Gillilan, S. Skou, BioXTAS RAW: improvements to a free open-source
487 program for small-angle X-ray scattering data reduction and analysis. *J Appl Crystallogr* **50**,
488 1545-1553 (2017).
- 489 44. V. Mariani *et al.*, OnDA: online data analysis and feedback for serial X-ray imaging. *J Appl*
490 *Crystallogr* **49**, 1073-1080 (2016).
- 491 45. S. Skou, R. E. Gillilan, N. Ando, Synchrotron-based small-angle X-ray scattering of proteins in
492 solution. *Nature Protocols* **9**, 1727-1739 (2014).

46. Y. Feng *et al.*, *A single-shot intensity-position monitor for hard x-ray FEL sources* (SPIE Optical Engineering + Applications, SPIE, 2011), vol. 8140.
47. J. A. Sellberg *et al.*, Ultrafast X-ray probing of water structure below the homogeneous ice nucleation temperature. *Nature* **510**, 381-384 (2014).
48. N. Esmaeildoost *et al.*, Anomalous temperature dependence of the experimental x-ray structure factor of supercooled water. *The Journal of Chemical Physics* **155**, 214501 (2021).
49. H. Laksmono *et al.*, Anomalous Behavior of the Homogeneous Ice Nucleation Rate in “No-Man’s Land”. *The Journal of Physical Chemistry Letters* **6**, 2826-2832 (2015).
50. K. H. Kim *et al.*, Maxima in the thermodynamic response and correlation functions of deeply supercooled water. *Science* **358**, 1589-1593 (2017).
51. P.-C. Chen, Jochen S. Hub, Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys J* **107**, 435-447 (2014).
52. C. J. Knight, J. S. Hub, WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic Acids Res* **43**, W225-W230 (2015).
53. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
54. K. D. Collins, G. W. Neilson, J. E. Enderby, Ions in water: Characterizing the forces that control chemical processes and biological structure. *Biophysical Chemistry* **128**, 95-104 (2007).
55. M. L. McDermott, H. Vanselow, S. A. Corcelli, P. B. Petersen, DNA’s Chiral Spine of Hydration. *ACS Central Science* **3**, 708-714 (2017).
56. Y. Bai *et al.*, Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J Am Chem Soc* **129**, 14981-14988 (2007).

Acknowledgments

We acknowledge Weiwei He, Yen-Lin Chen and Serdal Kirmizialtin for providing pdb coordinates for previously published hairpin duplex and triplex structures shown in Figure 1. We also acknowledge useful discussions with Rick Kirian about solution scattering at XFELs. We thank Shirish Chodankar for assistance with data acquisition at NSLS II.

Funding

National Science Foundation through the BioXFEL STC award number 1231306 (LP)

National Science Foundation grant DBI-1930046 (LP)

National Institutes of Health grant R35-GM122514 (LP)

National Institutes of Health grant R01-GM133998 (TDG)

Natural Sciences and Engineering Research Council of Canada (NSERC) (TW)

U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515 (use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory)

National Institutes of Health grant P41-GM139687 (LCLS/MSH)

National Institutes of Health grant P30 GM133893, S10 OD012331, and BER- BO 070 (for work performed at the CBMS beam line LIX (16ID) at NSLS-II)

U.S. Department of Energy grant BES-FWP-PS001 (NSLS II)
National Institutes of Health grant S10 OD028617 (LP BioXolver X-ray Source)
National Institutes of Health grant S10 OD025079 (LCLS Jungfrau Detector)

Author Contributions

K.A.Z. designed, fabricated and operated the mixing injectors and assisted in data acquisition, sample preparation and writing the manuscript. S.S. characterized the UAU12 construct and the reaction conditions, and assisted in LCLS experiments measuring the dA-dT duplex. S.A.P. acquired data on the UAU12 constructs and the dA-dT duplexes on the BioXolver source and prepared sample for LCLS experiments. D.A.R. and C.B.S. fabricated critical components of the mixing injectors. T.W. acquired WAXS data on the single strand constructs using NSLS II. Q.H. acquired SWAXS data on the hairpin duplex using the BioXolver source. M.M.K. assisted with sample preparation during the beamtimes. S.L. assisted with injector operation during beamtimes. V.M. helped specialize data analysis software for this experiment. C.K. assisted with sample preparation and injector operation during beamtimes. M.S.H., F.M. and F.P.P. operated the CXI station during LCLS beamtimes. T.D.G. assisted in real-time data visualization and creating and modifying data analysis software during and after the beamtime. L.P. assisted in data acquisition, analyzed the data, wrote the manuscript, supervised and coordinated all aspects of the experiment.

Competing interests

Authors declare that they have no competing interests.

Data and materials availability

Raw data for all scattering profiles of RNA and DNA are provided in the supplementary materials or in the SASBDB as indicated. Atomic coordinates for model structures shown in Figure 1, also used for Figure S1, are provided.

Supplementary Materials

Supplementary text, figures, tables are available on-line.

Supplementary Materials for RNA structures and dynamics with Å resolution revealed by x-ray free electron lasers

Kara A. Zielinski *et al.*

*Corresponding author Lois Pollack. Email: lp26@cornell.edu

This PDF file includes:

- Supporting text
- Figures S1 to S11
- Tables S1 to S4
- Legends for Datasets S1 – S23

Other supporting materials for this manuscript include the following:

- Datasets S1 - S23
- Simulation generated pdb files for the two structures shown in Figure 1

Supporting Information Text

Sample Preparation

Single stranded RNA sample preparation

RNA samples were prepared as described in (24). Briefly, 30 nucleotide poly-uridine (rU30) and poly-adenine (rA30) RNA were purchased from Integrated DNA Technologies (Coralville, IA, USA), reconstituted in a storage buffer of 20 mM NaCl, 1 mM MOPS pH 7.0, 10 μ M EDTA, and subsequently desalted and buffer exchanged into 100 mM NaCl, 1 mM MOPS pH 7.0, 10 μ M EDTA for data collection. For samples containing magnesium, a small volume of 1M MgCl₂ was spiked into the sample shortly before the measurement. RNA concentrations of 600, 200, 50 μ M were prepared, with the highest concentration aimed toward obtaining sufficient wide-angle scattering signal. All RNA samples were annealed by heating to 90°C for 3 minutes and snap cooling at 4°C for 20 minutes before measurements. Molecular structures were reproduced from SASBDB files: SASDFB9 for rA30 and SASDFK9 for rU30.

Hairpin duplex and hairpin triplex preparation

These scattering profiles were reproduced from Reference (19) and the SASBDB entries SASDKX5 for the hairpin duplex in 5 mM Mg²⁺, and SASDK26 for the hairpin triplex in 5 mM Mg²⁺. Molecules were prepared as described in that publication; no new profiles were measured for this work.

A-U duplex preparation

The 29 nucleotide construct U12-C5-A12 was purchased from Integrated DNA Technologies (IDT) and reconstituted in low salt buffer (1mM NaMOPS, 50uM EDTA and 2mM NaCl). It was buffer exchanged into a Mg²⁺ containing buffer (1mM NaMOPS, 50uM EDTA, 2mM NaCl and 2.5 mM MgCl₂) three times using 3k MWCO spin column at 14k xg for 10min at 4°C. The sample was heated to 85°C for 3 mins and cooled on ice. Data shown in Figure 1 were acquired on a sample with [RNA]=1mM, at T=10°C; lower concentration samples acquired at room temperature show identical features in their scattering profiles.

Triplex construct preparation

The RNA construct, UAU12, with sequence 12U-5C-12U-5C-12A was purchased as a lyophilized powder from Integrated DNA Technologies, Inc. (Coralville, IA). The powder was reconstituted in a low salt buffer (1 mM NaMOPS, 2 mM NaCl, 50 μ M EDTA at pH 7) for 15 minutes to maximize the yield. The sample was then buffer exchanged six times using a 10 kDa cutoff filter by spinning at 3600 G for 15 minutes at 4°C to remove any impurities. A thermal heat block was used to anneal the sample by heating it to 92-94°C for 5 minutes. The sample was removed from the heat and cooled to room temperature in about 15 minutes. Finally, the sample concentration was measured by UV-Vis absorbance and adjusted to either 1 mM or 0.5 mM for data collection. During all sample handling, gloves were worn, and other precautions were taken to prevent RNase contamination.

A-T duplex preparation

AT duplexes were comprised by base pairing dT25 with dA25. Both strands were purchased from Integrated DNA Technologies, Inc. (Coralville, IA). They were annealed at 95°C for 5 minutes and slowly cooled to 22 °C in an hour. This duplex was buffer exchanged to a solution containing 150mM NaCl, 10mM NaMOPS and 50uM EDTA at pH7. The [DNA] for all experiments was 1 mM.

Mixer Design and Fabrication

The mixer portion of these devices is made of concentric capillaries; a central, sample-carrying fused silica capillary supply line (Polymicro Technologies, Phoenix, AZ) is held inside a larger glass tube (320 µm inner diameter, Polymicro Technologies, Phoenix, AZ) with custom Kapton centering spacers. The supply line is polished and beveled to a tip and an additional capillary, which acts as a mixing constriction and delay line, is held just downstream with a 50-100 µm gap. In this gap, the sample is flow-focused to a narrow stream and becomes fully sheathed by a Mg²⁺-containing buffer, which diffuses into the sample stream to initiate the reaction. With a single mixer, flowrates are varied to reach multiple timepoints, typically with a factor of two total range. The length and inner diameter of the mixing constriction is also changed to reach different timepoint ranges. A total of three different injectors were designed to reach timepoints from 6 ms to 1000 ms (Table S1).

The entire mixer is encapsulated in a larger glass shroud (750 µm inner diameter, Sutter Instrument), which is flame polished to create the nozzle opening. The end of the mixing constriction is held with another set of Kapton centering spacers, and as the liquid exits the capillary, it is exposed to a helium gas sheath (30 mg/min flowrate) which further thins and accelerates the stream to create a free-standing jet.

To achieve good signal-to-noise data for S/WAXS, it is crucial to have a pathlength that is as large as possible to maximize the scattering signal. Typical GDVN jets are only a few microns in diameter. To create thicker jets, closer to a target of a 10 µm diameter, the glass shrouds were flame polished to create wider openings of 110-150 µm, instead of the more typical 80-90 µm. Jet widths were measured 150 µm downstream of the nozzle opening at each flowrate condition with a liquid-jet imaging microscope equipped with backlight illumination and a Zyla CMOS camera (ANDOR, Oxford Instruments, Abingdon, UK; (39)). The overall nozzle geometry (tip of mixing constriction, shape of nozzle opening, distance between the constriction and the nozzle aperture) influence the size and stability of the jet, but generally wider apertures result in wider jets. Within a single nozzle, higher flowrates result in thicker jets. Additionally, since the sample is fully sheathed, only the center of the jet contains RNA. The following formula approximates the sample stream width:

$$d_{sam} = \frac{d_{jet} \sqrt{\dot{V}_{sam}}}{\sqrt{\dot{V}_{tot}}}$$

where d_{sam} is the sample stream diameter, d_{jet} is the total jet width for that flow condition, $\dot{V}_{sam}$ is the volumetric flowrate of the sample, and $\dot{V}_{tot}$ is the total volumetric flowrate (sample and buffer sheath combined). When designing the mixers, the sample flowrate was maximized to give as wide of a sample stream as possible, while balancing timing dispersion and sample consumption concerns. As the sample flowrate increases, it creates a wider sample stream, which increases the uncertainty in the measurement as the Mg^{2+} ions need to diffuse further for complete mixing. These ions, however, diffuse extremely quickly, so timing dispersion concerns were mostly limited to the 6 ms and 10 ms timepoints. Table S2 details the jet width for each measurement.

Data Collection and Real-Time Analysis at CXI

Acquiring the time series at CXI

The initial state (0 ms, before mixing) of the RNA was taken with a Mg-free buffer sheath of 1 mM NaMOPS, 2 mM NaCl, 50 μ M EDTA at pH 7. Similarly, the scattering profile of an AT DNA duplex was acquired with a Mg-free buffer sheath of 150mM NaCl, 10mM NaMOPS and 50uM EDTA at pH 7. Time-resolved data were acquired from RNA at either 1 mM or 0.5 mM in 1 mM NaMOPS, 2 mM NaCl, 50 μ M EDTA at pH 7. For mixing experiments, the sheath contained 1 mM NaMOPS, 2 mM NaCl, 50 μ M EDTA, 25 mM $MgCl_2$ at pH 7 for [RNA]= 1 mM experiments, or 1 mM NaMOPS, 2 mM NaCl, 50 μ M EDTA, 15 mM $MgCl_2$ at pH 7 for [RNA]=0.5 mM experiments. Background profiles for subtraction were acquired by flowing Mg-free buffer in the sample stream and the appropriate Mg^{2+} containing buffer in the sheath. High pressure reservoirs (1.2 mL for sample, Neptune Fluid Flow Systems, LLC., Knoxville, TN), 10-40 mL for buffer (KNAUER Variloop, KNAUER Wissenschaftliche Geräte GmbH, Berlin, Germany)) and home built switch boxes (using VICI valves, Valco Instruments Co. Inc., Houston, TX) facilitated an easy transition between sample and buffer.

Sample present and sample absent profiles were collected in succession. The RNA absent sample (buffer background) was acquired either immediately before or after the RNA present sample. In the best case, no adjustments were made at any time during these two experiments. Measurements at different [RNA] were interleaved to reduce nozzle changes. For the case of buffer background before sample data collection, a typical experiment proceeds as follows. The sample line was set to the Mg-free buffer, and the appropriate Mg-containing buffer was loaded into the sheath line. Flow rates were set for the desired time point. Once the jet stabilized, as assessed optically and with real time X-ray data analysis feedback with OM (see below), 5 minutes of data were collected. The sample stream was then switched to flow RNA, and flowed for about 5 minutes to purge the ~2 meter long supply line (necessitated by the CXI vacuum chamber design). The presence of RNA was assessed by the appearance of a visible scattering signal on the detector. Five minutes of data were collected from this point. After acquisition of the sample-present point, the flowrates were adjusted for the next timepoint, and another 5 minute acquisition began for the sample measurement of the next dataset. Following acquisition, the sample stream was switched to the Mg-free buffer and another 5 minute transition was allowed to clean the supply line. Subsequently, the buffer match was collected for 5 minutes. This buffer, sample, sample, buffer pattern allowed for maximum efficiency by minimizing buffer/sample transitions, while ensuring accurate buffer background matches. As discussed below, data quality depends on acquiring the buffer/sample sequence while flow conditions and beam position are as identical as possible.

Key to ensuring smooth data collection was the use of OnDA (online data analysis) Monitor (OM) (44). After subtraction of the detector dark image, each image was analyzed by an azimuthal integration around the beam center. Masks were created to remove spurious scattering flares from the jet as well as inactive pixels. Different masks were created for each run, as the jet widths were not constant, as discussed in Table S2.

This software provided real time scattering profiles, which helped judge jet stability and when supply lines were purged for buffer to sample transitions. Jet stability using the OM live-feedback software was assessed by visualizing radial stacks of WAXS profiles within seconds of X-ray illumination, monitoring continuity of the diffuse water peak at high-q values as well as stability of the overall scattering intensity.

Background subtraction and acquiring solution scattering data at an XFEL

Accurate solution scattering experiments require subtraction of a background (sample absent) scattering profile from a sample present scattering profile. Under normal conditions, where a fixed path sample cell is used (as is the case for all of the synchrotron and lab source data reported here), both profiles are acquired from samples that are sequentially loaded into the same sample holder. Corrections for any variations in beam intensity are made following standard practices, which usually involves scaling by the transmitted beam intensity (45). However, at the XFEL, the solutions are flowing in liquid jets whose size depends on the details of the injector design as well as the flow rates. In the absence of rarely occurring nozzle freezing events, the jet size remained stable for each sample condition, assessed via optical monitoring and using OM to view WAXS profiles and total scattering intensity. We developed the following protocol to perform accurate background subtraction from profiles acquired in flowing jets at the CXI beamline.

As a first step, it was important to assess the contribution to the scattering profile from the beamline alone (the beamline background). Because of the vacuum environment at CXI, the scattering from the chamber, and from the nozzle inside the chamber with its helium sheath, are more than a thousand times smaller than the scattering from the buffer jet (Figure S4), so this contribution is safely neglected.

To perform accurate buffer background requires careful normalization of each measured profile. Each profile must be individually scaled to account for all sources of variation. Three critical parameters were required to perform accurate background subtraction. We had to account for: the changing x-ray pulse intensity, variations in the sample thickness, to which the WAXS signal is directly proportional, as well as the sample temperature at the intersection point. We consider each of these factors in turn.

First, because of the stochastic way in which XFEL pulses are generated, pulse to pulse variations in energy are expected; most pulses fall within a factor of 4 of the mean intensity. Because the strength of the scattering is proportional to the pulse intensity that generates it, each profile must first be normalized by the intensity. For most solution scattering experiments, the profile is normalized to the transmitted beam intensity; however, because the flowing jet is so thin, transmission is nearly 100%. Thus, we recorded the intensity of each pulse using a Wave8 monitor which consists of a thin Si₃N₄ target and 8 photodiodes that collect backscattered X-rays. The X-ray pulse intensity is linearly proportional to the sum of the diode intensities. The beam position

is calculated by the relative intensities of the 8 diodes (46). This value was recorded for each trace, and each profile was normalized to account for variations.

Second, a fixed path length is ideally maintained in solution scattering experiments. Typically, a fixed thickness sample holder contains the sample, however our liquid jets are freely flowing. Thus, we must account for the jet's thickness at the point where it intersects the beam, especially because the jet naturally thins out as it approaches its droplet breakup region and because there can be small motions of the beam, resulting in illumination of a different spot on the jet (either at a different height or a different width, which can result in being off center of the jet). After each profile is scaled to account for variations in XFEL pulse intensity, it was subsequently scaled to account for sample thickness. We chose to scale each profile to the integrated intensity of the first water peak. With the water peak intensity maximum located near $q = 1.9 \text{ \AA}^{-1}$, we integrated the profile over a q range from 1.6 to $2.25 \text{ \AA}^{-1}$.

To validate this method for normalizing with sample thickness, we tested it at NSLS II because it assumes that the nucleic acid scattering is insignificant when integrated over this region. We acquired WAXS data on the single strands (rU30 and rA30) in a cell with a fixed sample thickness. Data were normalized either using the traditional approach (transmitted beam monitor) or by scaling to the integrated intensity of the water peak. Figure S5 shows that both normalization methods yielded results that are the same within noise fluctuation, suggesting that scattering derived from the RNA macromolecules is negligible in the $q \sim 2.0 \text{ \AA}^{-1}$ regime. Finally, proper background subtraction requires that the sample-present and sample-absent profiles are acquired at the same temperature. Because the liquid jet is flowing into the CXI vacuum chamber it is rapidly supercooled (47, 48) at a rate that depends both on jet speed (in this case ~ 10 – 50 m/s), thickness, and distance from nozzle. Evaporative cooling of the water occurs at rates of up to 10^7 K/s under these conditions (47, 49). Temperature has a dramatic effect on water scattering, it alters the water structure factor throughout the curve. The water temperature can be readily ascertained by measuring the position of the first water peak near $q = 2 \text{ \AA}^{-1}$, which reports on the structural ordering of water (50). Measurements of $I(q)$ at room temperature, acquired on our lab source, locate this peak at $q = 2.05 \text{ \AA}^{-1}$ at 25°C , and $q = 2.03 \text{ \AA}^{-1}$ at 9°C . At LCLS, the peak is found near $q = 1.9 \text{ \AA}^{-1}$. Comparing this to values from (47), which report the peak of the structure factor, we estimate that our sample temperatures are just below 260 K , in the supercooled regime.

Temperature also affects the water scattering profile at lower q . Here, the detailed shape of the scattering profile reflects changing thermodynamic parameters of the water including isothermal compressibility and correlation length (50). **Changes at lower q are particularly significant for our experiment, as they may be comparable in size to the time-resolved signals we seek to measure.**

As part of our data analysis algorithm, we first selected profiles whose water peaks fell within a given q range for each run. We imposed an additional requirement to limit variations in temperature through the run, perhaps resulting from relative motions of the beam and the jet. We integrated the intensity in the low q portion of the curve, (from 0.3 to $0.9 \text{ \AA}^{-1}$, a region where we detected variations in our control samples), and divided this sum by the integral over the water peak, discussed above. This ratio was computed for each beam and water peak height normalized profile, and over the course of a run, traced out a nearly gaussian curve. We selected curves for further analysis that fell within one sigma of the mean. Once this 'temperature' selection was made,

the previously scaled azimuthal traces are summed and an average is computed to represent the either ‘sample present’ or ‘sample absent—buffer background’ trace for each condition. Note that if the nozzle was moved or replaced in between the buffer and sample collection, the data could not be properly scaled due to variation in temperature, and the point was discarded.

Once both the buffer and sample curves are thickness normalized and are acquired at a single temperature, the data point for each condition is derived by direct subtraction of the buffer from that of the sample under identical flow and nozzle conditions. Typical profiles are shown in Figure S6. Once the subtracted profile was obtained, other standard corrections were applied to account for solid angle in the WAXS regime, as well as beam polarization (47). These are the curves shown in Figure 3 of the paper. Offsets were applied for display purposes.

For one of the data points reported in the paper ($t=60$ ms), we acquired 36920/37503 pulses for the RNA present/RNA absent sample, 36192/36734 had acceptable pulse energies, 34372/35717 intercepted the jet, and 22511/27539 were similar in temperature. Overall, 71% of the data acquired for this point met our criteria, and this corresponded to between 3 and 4 minutes of data acquisition for this condition. Figure S7 illustrates this process as a flow chart.

Modeling connects features in the scattering profiles with molecular structures.

To aid in the interpretation of the peaks that appear at high q features, we turned to models. Beginning with a .pdb file that results from our WAXS-driven MD results of (19) for the long hairpin triplex in Mg^{2+} we used WAXSIS (51, 52) to compute the scattering profiles out to $q = 2 \text{ \AA}^{-1}$. Results from these computations are shown in Figure S1. Panel a shows the profile computed from the full molecule, which is a 29 base pair hairpin interfaced with a 24 base triplex forming single strand, visualized using Chimera (53). The molecule is mostly in the triplex form, as shown to the right of panel a. The scattering profile was computed from a single frame of the simulation. Panel b shows the profile computed from the triplex piece of the full pdb file, extracted using Chimera. Note the appearance of four high q peaks near $q=1.05, 1.4, 1.7$ and $1.9 \text{ \AA}^{-1}$. Following the protocol introduced in Ref. (23), the profiles of just the backbone molecules (phosphates plus sugar) were computed and are shown in panel c. Note that the first three of the above referenced peaks persist, while the fourth, highest q peak is absent. As noted in Ref. (23), peaks in this region reflect the arrangements of the phosphorous and oxygen atoms along the backbone. Panel d shows the scattering profiles of the bases alone, and the highest of the four peaks of panel b is identified with base stacking, as concluded in Ref. (23) for a DNA duplex. Finally, panel e shows the experimental profile of the hairpin triplex, from Ref. (19) deposited as SASDK26 in the SASBDB.

Low q changes.

Although this paper focuses on the high q correlations that report RNA structural motifs, data at lower q are also informative. They reflect the appearance of features with larger spatial dimensions in the sample. A comparison of the data from the three model structures of Figure 1 (the single strand, duplex, and triplex), shown here as Figure S2, highlights significant changes in the scattering profiles in the form of a broad shoulder that appears just above $q=0.4 \text{ \AA}^{-1}$. Past machine learning studies of duplexes (21) associate this peak with the molecular radius, and it shifts between the duplex and triplex conformations. It is absent from the profiles of single stranded data,

which do not have a radius defined by the interaction of multiple strands. Importantly this shoulder shifts to lower q as the molecule transitions from mostly duplex to mostly triplex, reflecting the increasing radius of the triplex. The prominence of the peak depends on the exact molecular conformations which are dynamic and depend on length salt and sequence. Changes in this q range are seen in the XFEL data and correlate with the appearance of higher q features.

Determining the proper $[\text{Mg}^{2+}]$ level to initiate the reaction.

To find the proper amount of MgCl_2 required to initiate folding, we performed static titrations on our lab source, beginning with the low salt starting buffer and spiking in concentrated solutions of Mg in buffer. The goal was to determine the total Mg concentration required to fold the RNA at its different concentrations. At 0.5 mM RNA, the molecule appeared folded (assessed by examining the high q data, using a Kratky plot of Iq^2 vs. q), when the total Mg concentration equaled 5 mM. More Mg was required to fold the 1 mM RNA sample. In this case, nearly 10 mM of Mg (total) was introduced.

For mixing, we used a reaction threshold of 30%, meaning the sample was considered fully mixed when the concentration of Mg^{2+} at the center sample stream reached 30% of the level in the outer, sheathing solution. Thus, for the lower $[\text{RNA}]$, we considered the reaction initiated when the $[\text{Mg}]$ reached 4.5 mM (30% of the 15 mM MgCl_2 sheath), and for the higher $[\text{RNA}]$, we considered it initiated when $[\text{Mg}] = 7.5$ mM (30% of the 25 mM MgCl_2 sheath). It is plausible that the reaction was initiated at a slightly earlier time in the lower concentration sample, based on the increased ordering of the backbone visible at $t=6$ ms. Data are shown in Figure S3.

The effect of temperature on scattering profiles of nucleic acids.

The high q scattering of nucleic acids can be strongly impacted by the nature of the solvent around the biomolecule. As discussed above, the scattering of bulk water is temperature dependent and we must consider the temperature dependence of the hydration water around the biomolecule, whose interactions with the nucleic acid, or the ions around it (54), may be altered as a result of changes in the water structure. For example, it has been established that DNA has a chiral hydration spine (55), and the changing properties of liquid water, as described above, may alter the nature of this ‘bound water’, resulting in a temperature dependent change in the DNA scattering profile.

As a control, we measured the scattering profile of a model nucleic acid system, a DNA duplex consisting of a 25 nucleotide strand of A bases, that has been annealed with a 25 base strand composed of T bases. This A-T duplex displays numerous peaks in the high q region (23). To explore whether temperature affects the wide angle scattering profile of this A-T duplex, we performed experiments on the BioXolver lab source at fixed temperatures of 25°C and 9°C. Temperature dependent changes are detected both in the buffer scattering and in the DNA (as assessed after buffer subtraction). Figure S9 shows scattering of buffer acquired at 25°C and at 9.4°C. These profiles display a q dependent offset in the range $0.4 < q < 2.0 \text{ \AA}^{-1}$. Consistent with the above discussed reports of the temperature dependence of water scattering, made at both SSRL and LCLS, we measure changes in both the location of the water peak and in the scattering at lower values of q . Background subtracted profiles from the AT duplex, measured at both 9°C and 25°C are shown in Figure S10. Changes are detected above $q \sim 0.3 \text{ \AA}^{-1}$. A WAXS profile

acquired at LCLS, where the solvent is supercooled (water peak) is shown as the third line on the figure. A significant decrease is seen in the same q region, signaling a potential temperature dependent effect. This latter curve was acquired in a liquid jet from a mixing injector at CXI, though no mixing was triggered (there was no difference in the buffer around the DNA and the sheathing buffer).

Interestingly, the folded RNA trace (Figure S8) shows deviations when acquired at room temperature and at LCLS. In both cases, the profiles ‘match’ at low q and at the highest q , but deviate in a region that tracks with the length scales of duplex and triplex features.

RNA folding at reduced concentration: 0.5 mM.

Figure S1 1 shows the time series acquired with lower RNA concentration. Following the guidance of static Mg^{2+} titration experiments (Figure S3), and with the mixing threshold set at 30% (Methods), the $[\text{MgCl}_2]$ in the sheathing buffer is 15 mM when $[\text{RNA}]=0.5$ mM and 25 mM when $[\text{RNA}]=1$ mM. Because the exact composition of the ion atmosphere around the RNA is unknown, e.g. the number of bound vs. free Mg ions cannot be determined from this experiment (56), it is plausible that the free $[\text{Mg}]$ is higher for this time series, resulting in slight variations in folding kinetics between the two conditions. Nevertheless, the order of assembly is the same. Duplex features appear by the first measured time point. At later time points, the major groove peak disappears, and the triplex backbone structure strengthens. At longer times, the base stacking peak appears, locking in the structure. These trends are exactly recapitulated in the lower q peak ($0.4 \text{ \AA}^{-1}$): the loose shoulder becomes better defined on time scales longer than 100 ms.

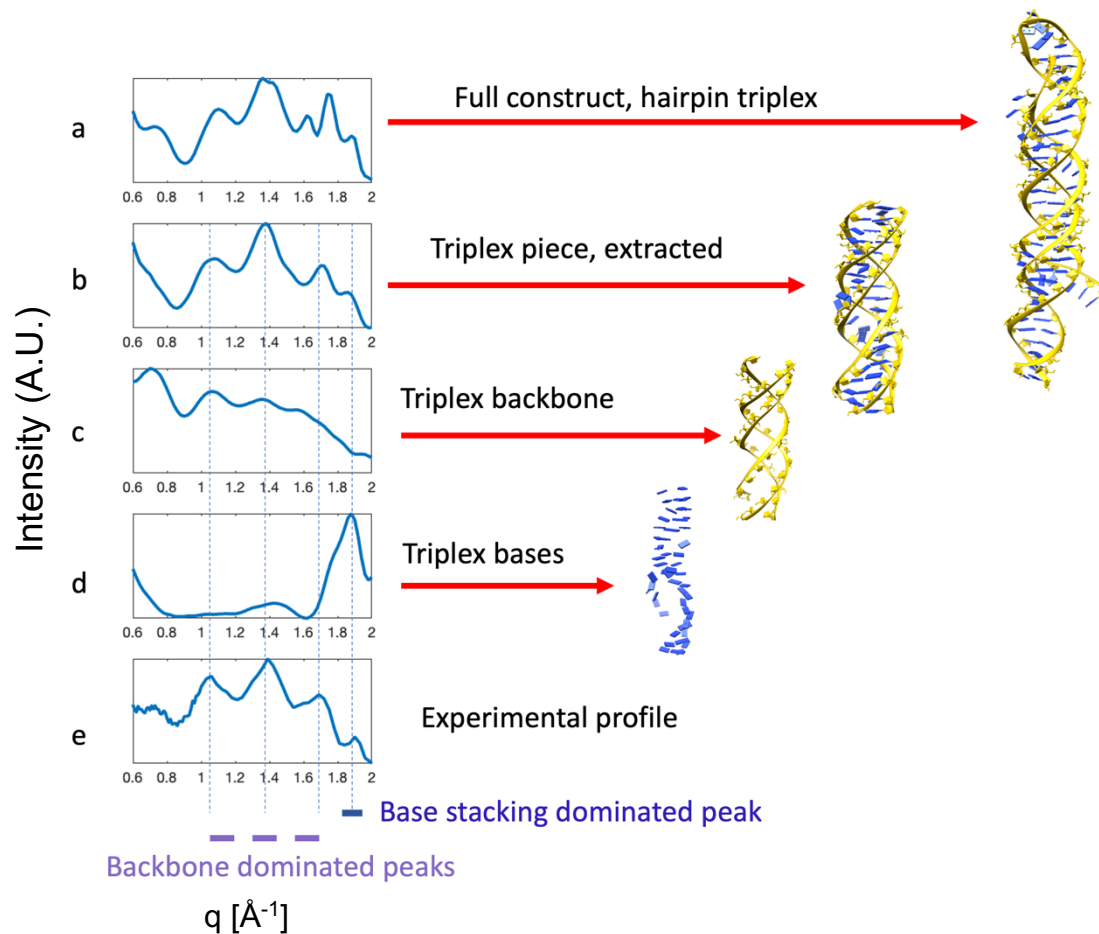


Figure S1. **Features of triplex scattering profiles can be connected with real space molecular structures.** As described and referenced in the text, profiles are computed via WAXSIS. The model structure shown here is the one pictured in Figure 1 panel C of the main text.

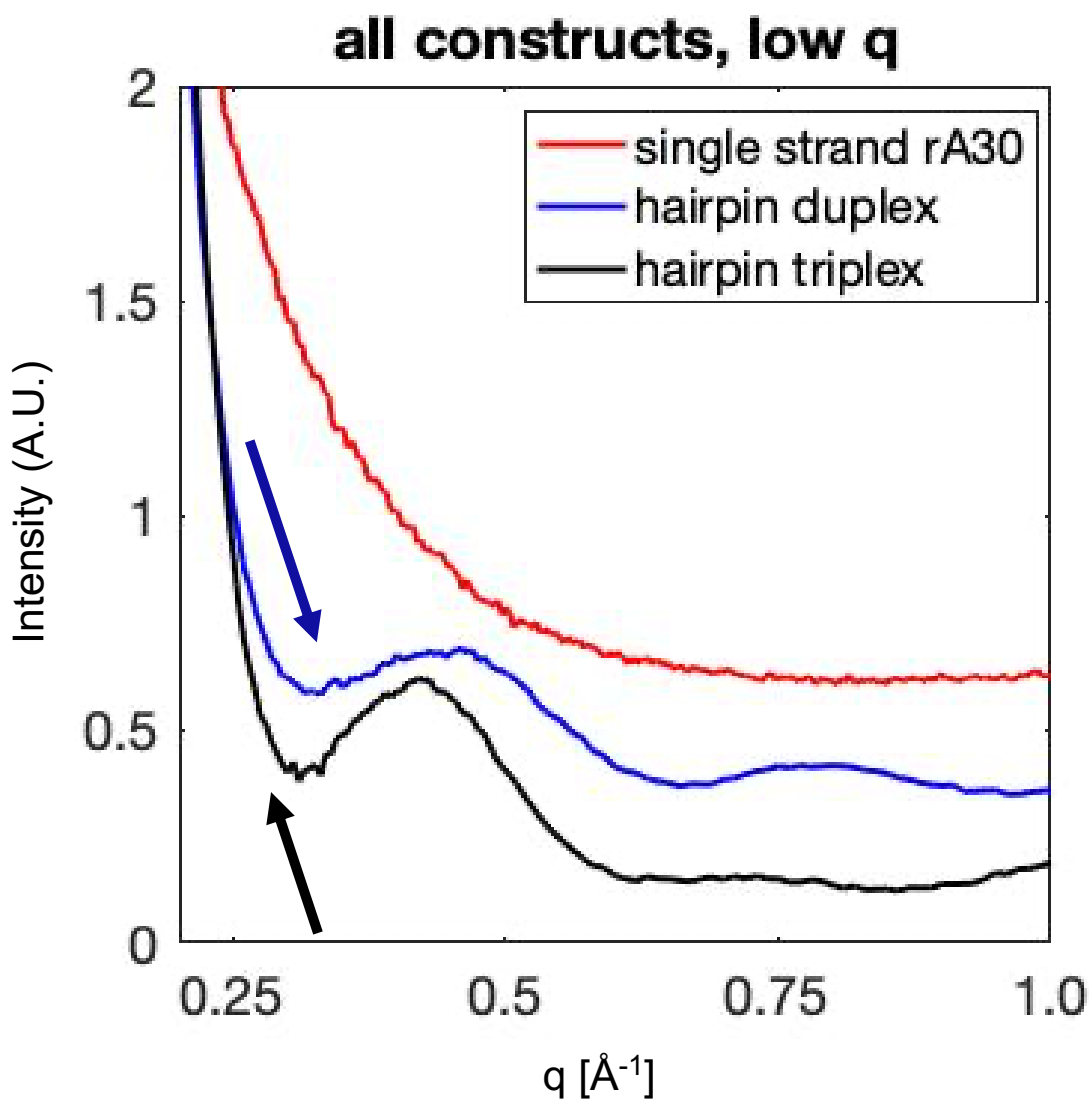


Figure S2. **An important low q feature reflects the molecular radius.** Arrows indicate regions which change with molecular conformation. Scattering from an unstructured single strand is shown in red, from a helical duplex in blue and a triplex in black. These profiles were shown in Figure 1 of the paper, we simply emphasize a different q region here.

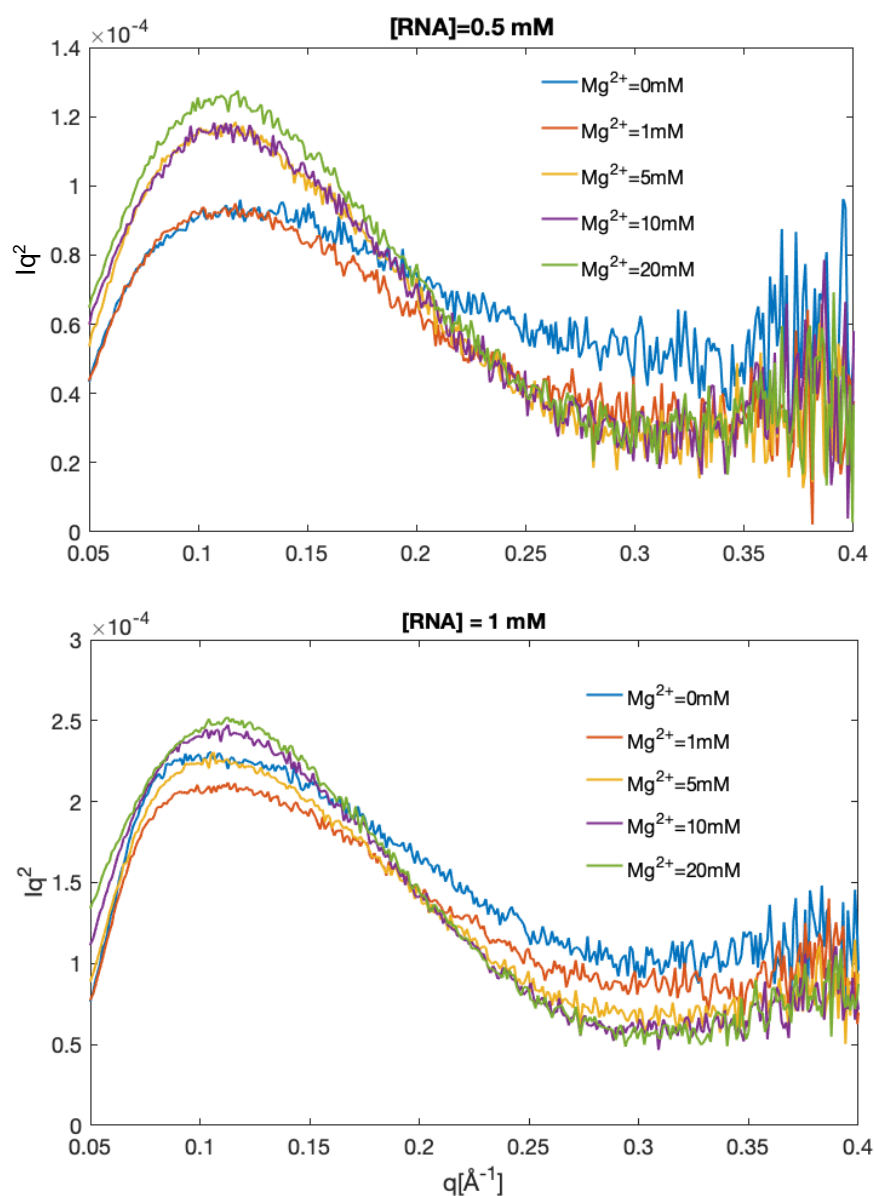


Figure S3. **The UAU12 construct has Mg^{2+} dependent folding.** Static data acquired with increasing $[\text{Mg}]$ acquired on the BioXolver source are shown. These profiles are displayed as Kratky plots of Iq^2 vs. q , which emphasize the changes that indicate folding. More folded molecules have better defined peaks near $q=0.12 \text{ \AA}^{-1}$ as well as lower values at large q . From these curves, we selected the concentrations of 4.5 mM and 7.5 mM as reaction initiation thresholds for $[\text{RNA}]$ at 0.5 mM and 1 mM.

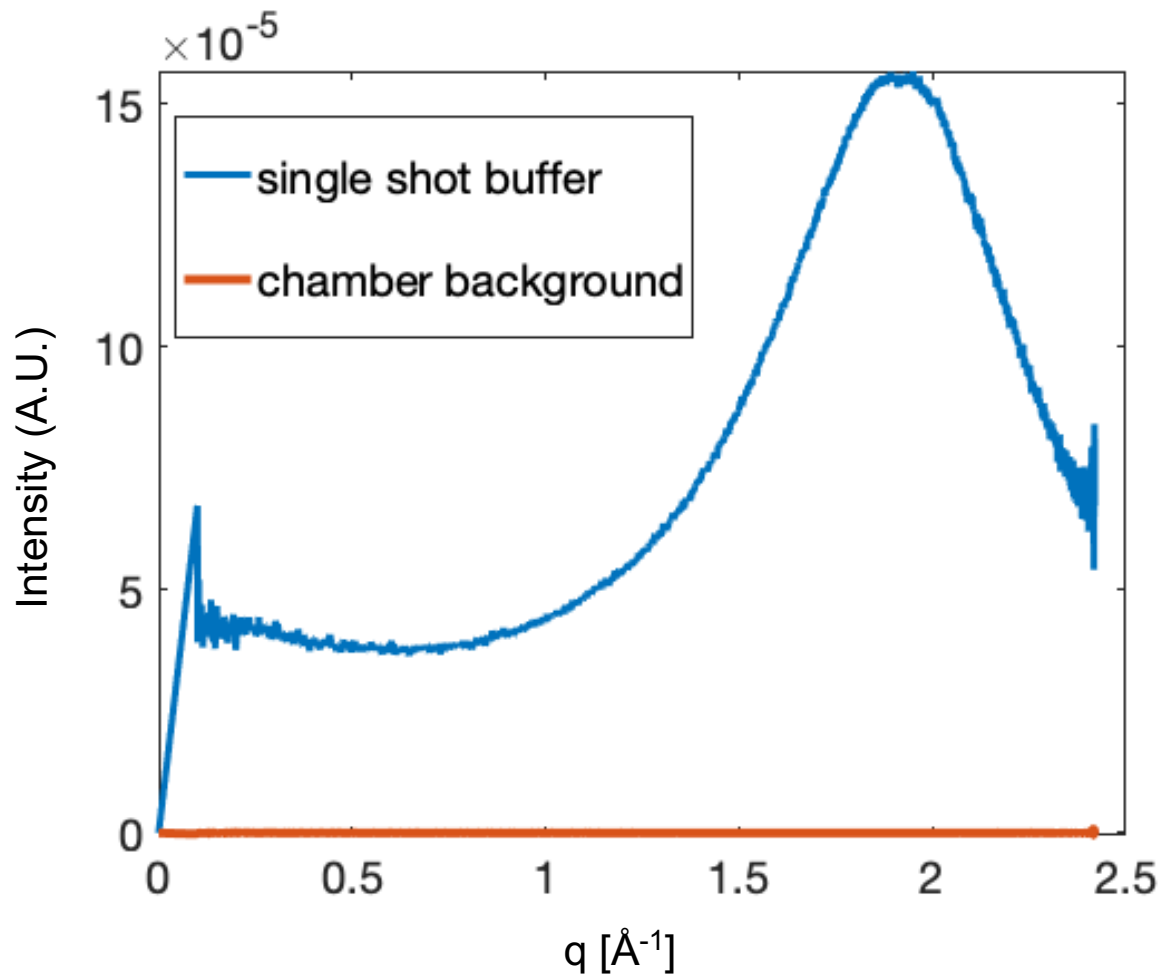


Figure S4. Scattering from the chamber background (brown), beam normalized, is compared to a single shot of buffer scattering (blue) also beam normalized. The small contribution from the background justifies its omission from our analysis.

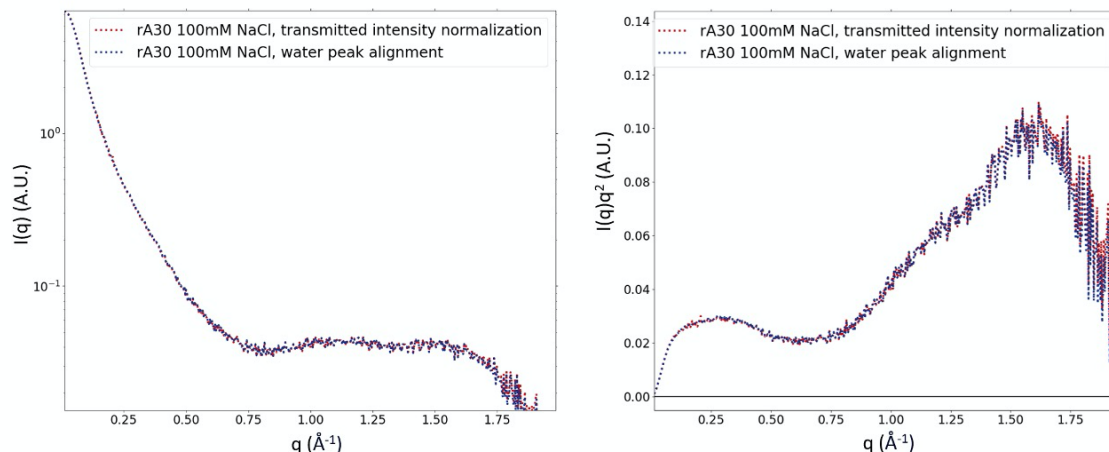


Figure S5. Water peak normalization is as effective as transmitted beam normalization. These plots show a comparison of scattering profiles of rA30 in 100mM NaCl, where alignment of the sample and buffer scattering has been performed using alignment of the average water peak scattering intensity versus normalization by the transmitted X-ray beam intensity. Scattering intensities are plotted in arbitrary units. Data are plotted on semilog (left) and Kratky-transformed (right) axes to visualize agreement of the two methods in both low and high q regimes. All other single stranded RNA data showed similar levels of agreement.

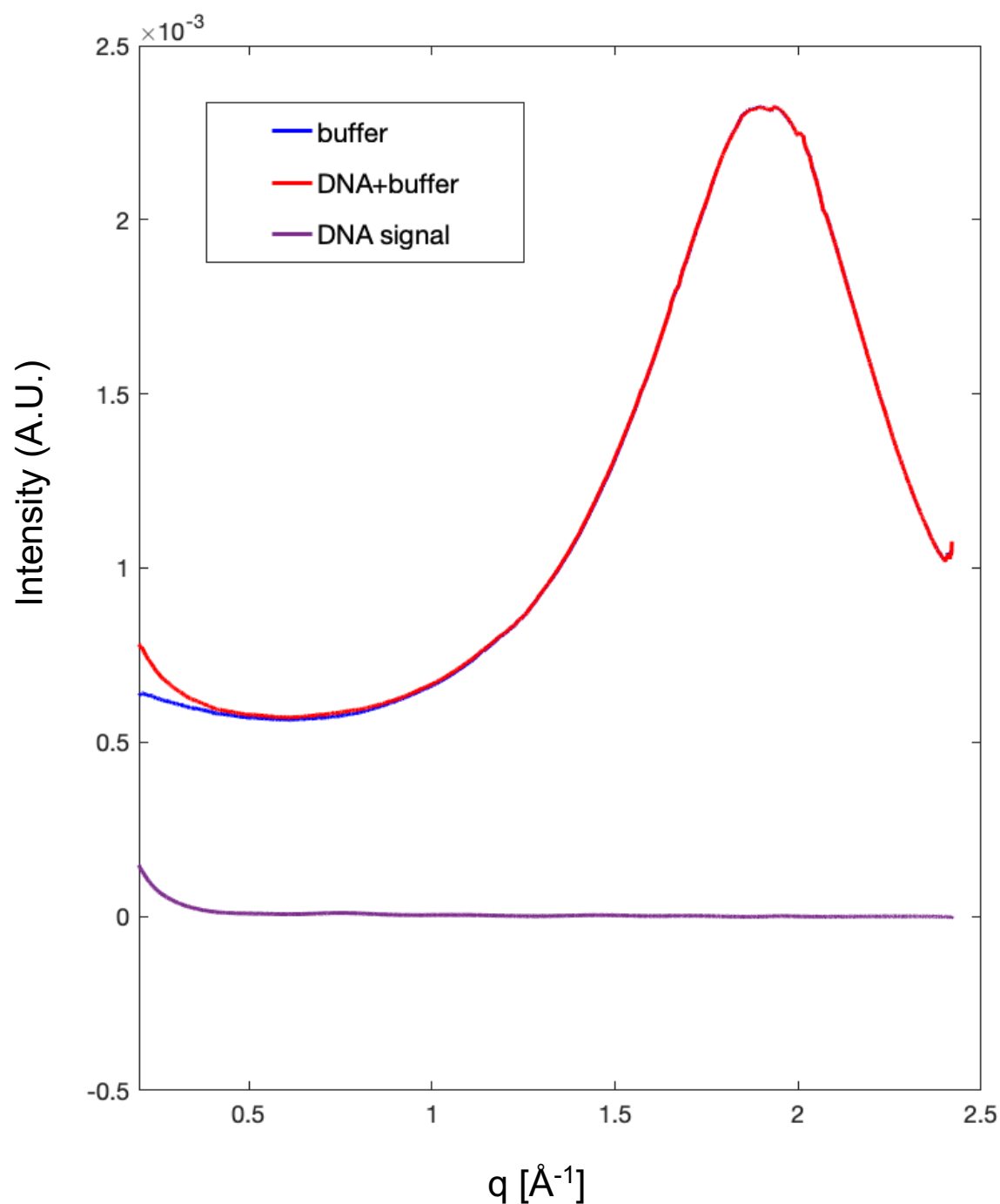


Figure S6. Extraction of the DNA signal occurs by buffer subtraction. This plot shows the beam and jet normalized buffer (blue), sample (red) and the difference (purple) from our control sample: the 25 base pair AT duplex. The subtracted curve is also shown in Figure S10.

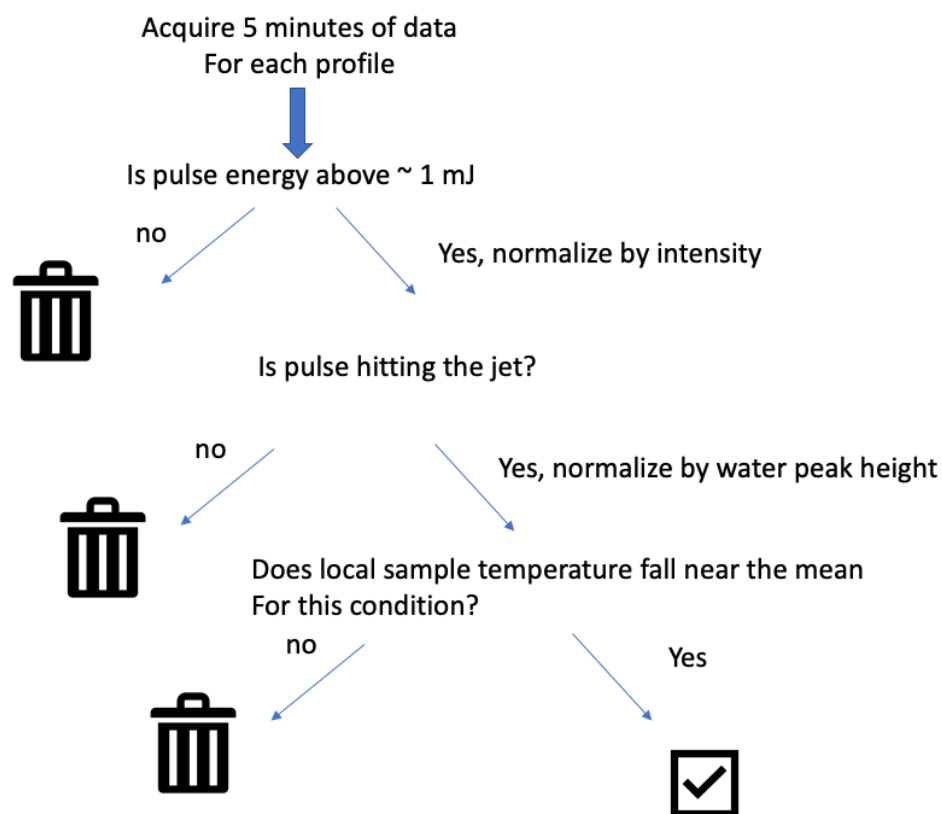


Figure S7. **Pipeline for data analysis.** This flowchart illustrates the pipeline for data analysis described in the text.

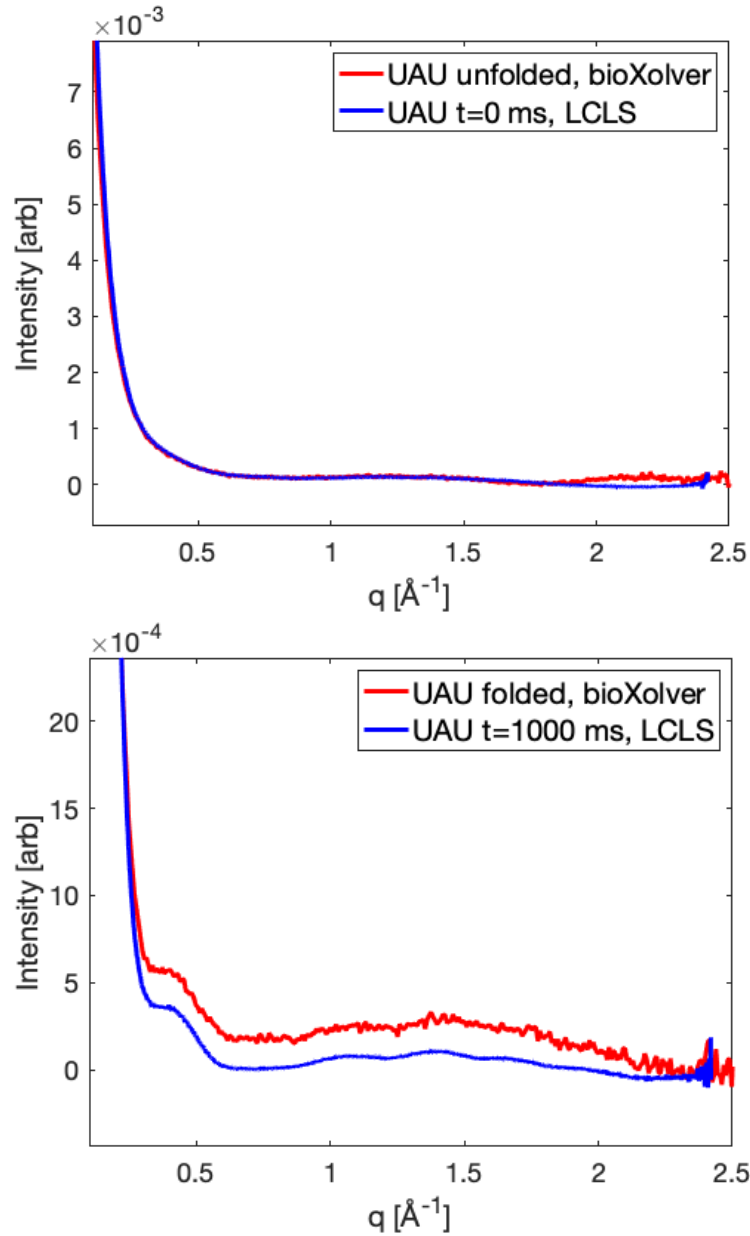


Figure S8. Comparison of scattering profiles of unfolded and folded states of the UAU12 triplex. These plots show scattering profiles of both single stranded UAU12 and triplex UAU12, acquired at room temperature on our lab source using a thick sample (red, exposure time 120 seconds), compared to LCLS (blue).

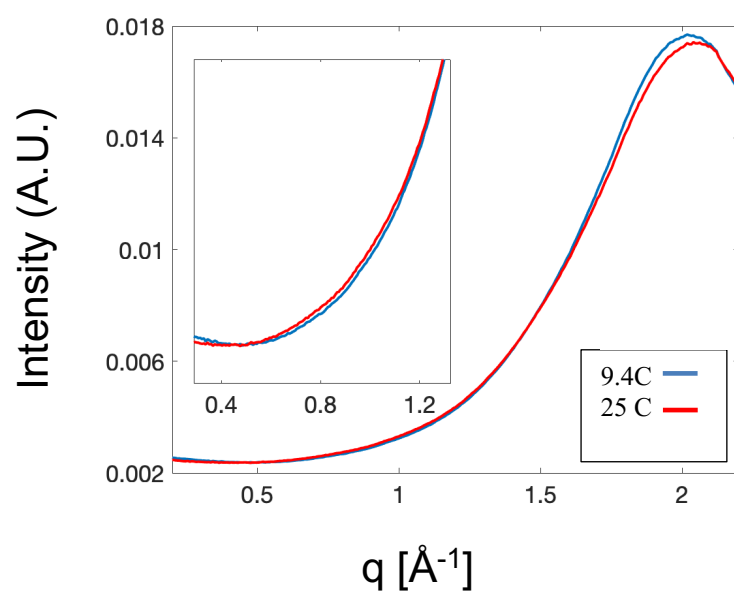


Figure S9. Temperature dependent buffer scattering. Two scattering profiles are shown that were acquired on the BioXolver lab source, at two temperatures: 9 and 25 C. These trends, especially the q regions where there are changes, exactly echo the conclusions of Ref. (47).

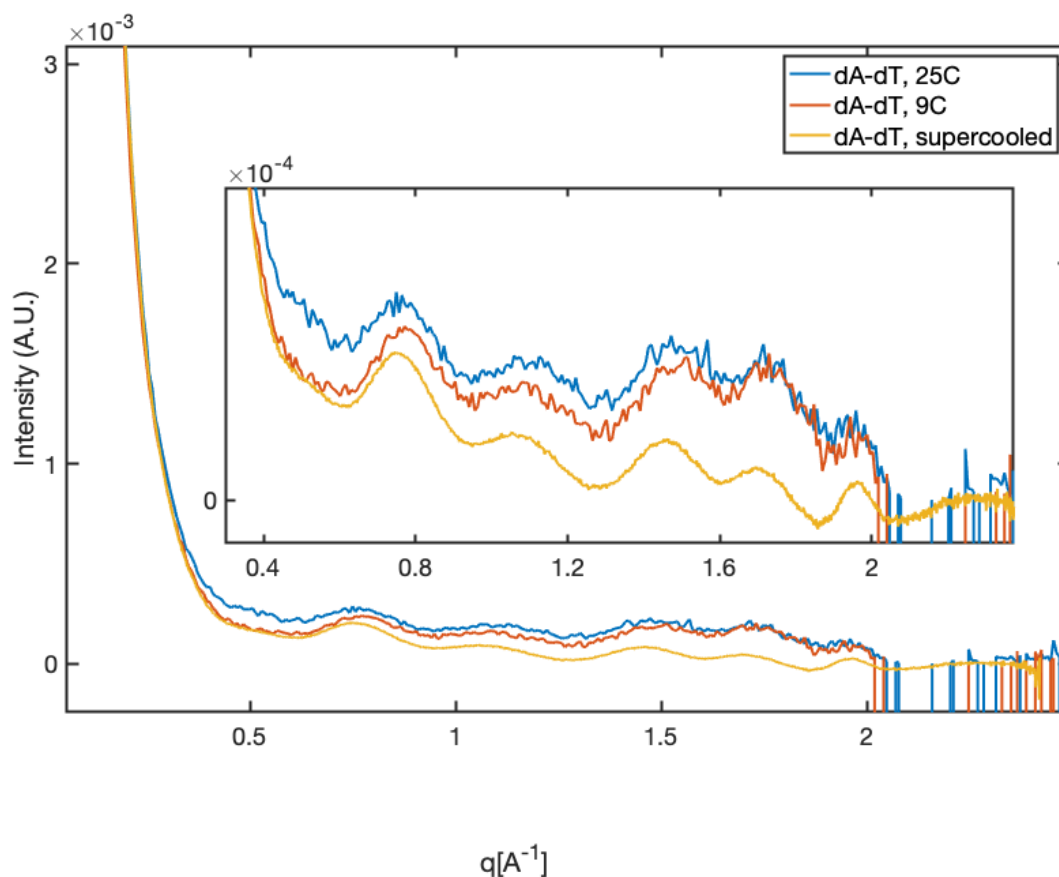


Figure S10. Temperature dependent scattering of a DNA duplex. The temperature dependent scattering of our control molecule, a 25 base pair DNA duplex made from joining dT25 with dA25 (two 25 nucleotide strands), is shown. Note the deviations from room temperature to 9 C, and again, to the LCLS measured state at a lower temperature. **Because they were acquired at different sources using different detectors, these data were scaled to match at the lowest and highest q values.**

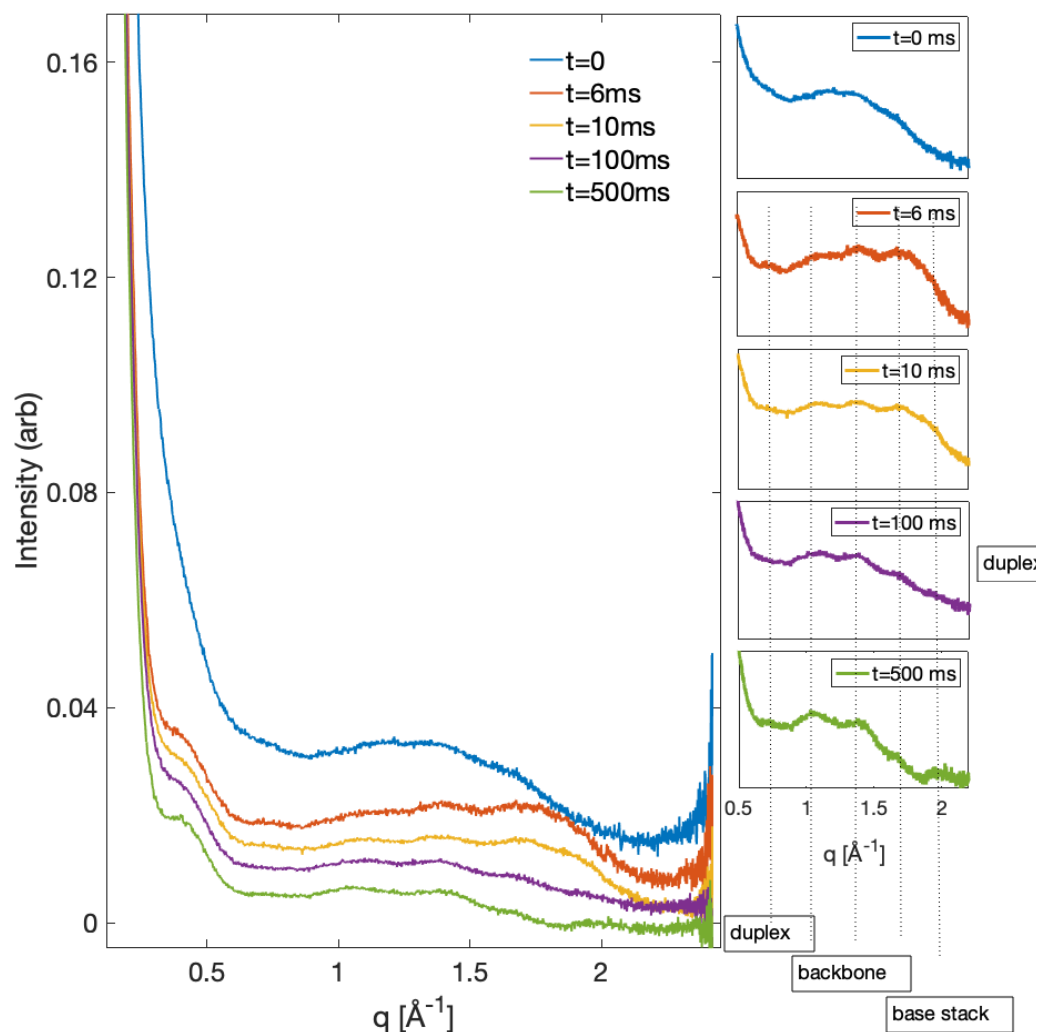


Figure S11. The time course of RNA folding for [RNA]= 0.5 mM. For these experiments, the sheath contains 15 mM Mg, and folding is considered initiated when the [Mg] reaches 4.5 mM. The order of appearance of the peaks recapitulates the data shown in the paper for [RNA] = 1 mM. A small, major groove associated peak appears at the fastest measurement time, $t = 6$ ms. Other peaks, representing the triplex backbone also appear and stabilize at longer reaction times. A high q peak, reflecting base stacking becomes prominent by 500 ms, though it is visible at earlier times. Note also that the low q behavior reproduces what is measured at the higher [RNA]. The high q background varies between this series and the previous. This may be a result either of differences in [Mg], variations in temperature due to changing salt conditions, or smaller signal size, due to the reduction in concentration. and are the result of either temperature variation or the more challenged signal strength for the lower [RNA] data set.

Supplementary Tables:

Timepoint ($\pm$ uncertainty; ms)	Constriction Length (mm)	Constriction Inner Diameter (μ m)	Sample Flowrate (μ L/min)	Sheath Flowrate (μ L/min)
6 $\pm$ 1.5	17.8	50	9.6	123.3
10 $\pm$ 2.9	17.8	50	10.3	67.3
60 $\pm$ 19.1	32	100	19.1	72.3
100 $\pm$ 28.3	32	100	16.5	41.7
500 $\pm$ 34.3	25**	100	20.6	80.4
1000 $\pm$ 96.6	25**	100	18.8	34.2

** For timepoints 500 ms and longer, an additional delay stage is added after the constriction

Table S1: Injector design and flow conditions for each timepoint

Timepoint ($\pm$ uncertainty; ms)	Nozzle Opening (μ m)	Total Jet Width (μ m)	Sample Width (μ m)	Total Flowrate (μ L/min)
0	127	10.2	4.6	101
6 $\pm$ 1.54	90	6.4	1.7	132.9
10 $\pm$ 2.90	90	3.8	1.4	77.6
60 $\pm$ 19.12	90	7.1	3.2	91.4
100 $\pm$ 28.27	90	5.1	2.7	58.2
500 $\pm$ 34.3	127	10.2	4.6	101
1000 $\pm$ 96.6	127	6.4	3.8	53

Table S2: Nozzle openings and jet widths for each timepoint

Timepoint	# frames sample	# frames buffer	Beam power
0	13166	10660	100%
6 $\pm$ 1.54	11448	16685	50%
10 $\pm$ 2.90	15316	15733	50%
60 $\pm$ 19.12	22511	27539	100%
1000 $\pm$ 96.6	12076	12600	100%

Table S3: Frame numbers and beam power for data of Figure 3.

Timepoint	# frames sample	# frames buffer	Beam power
0	13166	10660	100%
6 $\pm$ 1.54	13307	23102	90%
10 $\pm$ 2.90	27354	24655	90%
100 $\pm$ 28.27	4839	10135	100%
500 $\pm$ 34.3	15005	9981	90%

Table S4: Frame numbers and beam power for data of Figure S11.

Dataset S1-S23 (separate files).

All of the scattering profiles for RNA and DNA are uploaded as text files in the Data section (or their accession codes are given for those that are uploaded to the SASBDB).

Guide to raw data: all SAXS profiles related to structure of RNA or DNA

RNA or DNA molecule	File name or identifier
rA30	rA30
rU30	rU30
Mixed sequence duplex hairpin	SASBDB entry SASDKX5
Mixed sequence triplex hairpin short	SASBDB entry SADDKU5
Mixed sequence triplex hairpin long	SASBDB entry SASDK26
rUA hairpin duplex	rUA
Time resolved data in main text:	
	Time_resolved_data_t0
	Time_resolved_data_t6ms
	Time_resolved_data_t10ms
	Time_resolved_data_t60ms
	Time_resolved_data_t1000ms
SI:	
Mg titration, [RNA]=0.5 mM	
	Data_FigS3_top_0Mg
	Data_FigS3_top_1Mg
	Data_FigS3_top_5Mg
	Data_FigS3_top_10Mg
	Data_FigS3_top_20Mg
Mg titration, [RNA]=1 mM	
	Data_FigS3_bottom_0Mg
	Data_FigS3_bottom_1Mg
	Data_FigS3_bottom_5Mg
	Data_FigS3_bottom_10Mg
	Data_FigS3_bottom_20Mg
Lab source triplex	
Unfolded	SI_data_unfolded_Bioxolver
Folded	SI_data_folded_Bioxolver
Control duplex, dA-dT	
25°C	dAdT_25
9 °C	dAdT_9
Supercooled	dAdTsupercooled

Coordinate files (.pdb format) to support Figures 1 and S1 are uploaded for one conformation of the triplex and duplex constructs from simulations of Ref. 19.